Activity Report Paris 2017

Edition: 2018-02-19

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
B9.1.1. - E-learning, MOOC
B9.4.5. - Data science
B9.5.6. - Archeology, History
B9.5.8. - Linguistics
B9.5.10. - Digital humanities
B9.6. - Reproducibility
B9.7. - Knowledge dissemination
B9.7.1. - Open access
B9.7.2. - Open data

1. Personnel

Research Scientists

Benoît Sagot [Team leader, Inria, Researcher]
Pierre Boullier [Inria, Emeritus]
Laurent Romary [Inria, Senior Researcher, HDR]
Daniel Stökl Ben Ezra [EPHE, Senior Researcher]
Éric Villemonte de La Clergerie [Inria, Researcher]

Faculty Members

Marc Bui [Univ Vincennes-Saint Denis & EPHE, Professor]
Djamé Seddah [Univ Paris-Sorbonne, Associate Professor]

Technical Staff

Wigdan Abbas Mekki Medeni [Inria, from Apr 2017]
Achraf Azhar [Inria, from Nov 2017]
Elias Benaissa [Inria, from Apr 2017]

Luca Foppiano [Inria]
Tanti Kristanti [Inria, from Nov 2017]
Alba Marina Malaga Sabogal [Inria, from Sep 2017]
Héctor Martínez Alonso [Inria, until Nov 2017]
Stefan Pernes [Inria, from Mar 2017 until Aug 2017]
Marie Puren [Inria]
Charles Riondet [Inria]
Dorian Seillier [Inria]
Lionel Tadonfouet [Inria, from May 2017]
Émilia Verzeni [Inria, from Apr 2017]

PhD Students

Jack Bowers [Vienna Academy of Sciences]
Loïc Grobol [Ministère de l'Education Nationale, from Oct 2017]
Axel Herold [Berlin-Brandenburg Academy of Sciences]
Mohamed Khemakhem [Inria]
Mathilde Regnault [Ecole Normale Supérieure Paris, from Oct 2017]

Visiting Scientists

Basant Agarwal [ERCIM, from Aug 2017 until Sep 2017]
Daniel Dakota [Indiana University, until Jan 2017]

Administrative Assistant

Christelle Guiziou [Inria]

2. Overall Objectives

2.1. Overall Objectives

ALMAnaCH is a follow-up to the ALPAGE project-team, which came to an end at the end of 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 Paris-Diderot computational linguists 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 studying human sciences.

ALMAnaCH's scientific positioning therefore extends ALPAGE's. We remain committed to developing state-of-the-art natural language processing 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 now 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 with a new partner, namely the École Pratique des Hautes Études (EPHE). The EPHE is a leading institution in France in human sciences in general and in digital and computational humanities in particular. Two EPHE research directors, who have already been working together for some time in computational humanities, will be permanent members of the project-team: a philologist and a computer scientist, both specialists of computational approaches to philology and ancient language studies, in line with the above-mentioned scientific positioning.

3. Research Program

3.1. Overview and research strands

One of the main challenges in computational linguistics is **modelling and coping with language variation**. Language varies with respect to domain and genre (news wires, scientific literature, poetry, oral transcripts...), sociolinguistic factors (age, background, education; variation attested for instance on social media) and other dimensions (disabilities, for instance). But language is also in constant evolution at all time scales. 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 **ancient texts**.

ALMAnaCH tackles the challenge of language variation in two complementary directions.

3.1.1. Research strand 1

We focus on linguistic representations that are less affected by language variation. This first requires improving the **production of semantic representations (semantic parsing)**. This also involves investigating 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 the field of natural language processing (hereafter NLP). We have to identify, model and take advantage of each type of contextual information available. Addressing these issues enables the development of new lines of research related to conversational content. Applications thereof include chatbot-based systems and improved information and knowledge extraction algorithms. We especially focus our work on challenging datasets such as domain-specific texts and 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 forms. 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 on Old to all forms of Modern French, as well as Indo-European and Semitic languages in general). In addition, the noise introduced by OCR and HTR systems, especially in the context of historical documents, bears similarities with those 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**. Dealing with diachronic and language-encoding variation, as well as their interaction, is the main motivations behind the creation of a joint project-team between Inria and EPHE.

3.1.3. Research strand 3

These two first research strands rely on the availability of **language resources** (corpora, lexicons). The development of **raw corpora from original sources** is a domain of expertise of ALMAnaCH's EPHE members. The (manual, semi-automatic and automatic) development of **lexical resources** and **annotated corpora** is a domain of expertise of ALMAnaCH's Inria and Paris 4 members. This complementary expertise in language resource development (research strand 3) benefits to the whole team and beyond, and both feeds and benefits from the work of the other research strands.

3.2. Automatic Context-augmented Linguistic Analysis

This first research strand is centered 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: hybridation not only couples symbolic and

statistical approaches, but neural approaches as well. AI applications are twofold, notwithstanding the impact of language variation (for which 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 8.1). 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 organized 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 [55], [79] and one of the best for robust multilingual parsing as shown by our results in various shared tasks over the years [77], [28], [29]. 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 [28].

In particular, we continue to explore the hybridization 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 tool less sensitive to variation, more easily configurable, and self-adapting. Our short-terms goals is to explore techniques such multi-task learning (cite refs in soogard 2016-2017) to propose a joint model of tokenization, normalization, morphological analysis and syntactic analysis. We also explore adversarial learning, considering drastic variation we face in user generated content parsing and historical text processing, 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 challenge in NLP now: handling the context where a speech act takes place.

Indeed, there is a strong tendency in NLP to assume that each sentence is independent from both other sentences and its context of enunciation, in order 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 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 dialogue sequences, the addition of context-based features (namely the speakers information and the dialogue moves) was beneficial [59]. In the case of UGC, working accross sentence boundaries was explored for instance, with limited success, by [54] for document-wise parsing and by [68] for POS tagging.

[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

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 point carried out 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 datasets, 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 metadata 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 the resources used), 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 joint work with Patrice Lopez on named entity detection and linking in scientific texts. Through the co-supervision of a PhD funded by the LabEx EFL (on which 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 expectations, with also the emergence of conversational agents and digital assistants (such as Siri). The current approaches mostly rely on the design of scenari associated with very partial analysis of the requests to fill expected slots and to generate canned answers.

The next generations should rely on programs having a deeper understanding of the requests, being able to adapt to the specificities of the requesters, and providing less formatted answers. We believe that chabots 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 (to the various requester 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 decrease, sometimes dramatically, when faced with non-contemporary, non-edited or specialised texts. Despite the existence of domain-adaptation techniques and robust tools, for instance for trying to process social media texts, 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, and also regarding intertextuality, on which see below). 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 accesibility 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 sequencial nature of the problem at hand. We 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-resources languages (e.g. Old French texts), and OCR/HTR system outputs.

Nevertheless, the different types of language variation 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

We plan to explore computational models to deal with language variation. But it is important to start by getting more insights about language in general and about the way humans apprehend it. We 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. In this regard, ALMAnaCH memebrs 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 (for which 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 is information about processing evolutions (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 on our daily life through social media and such 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 linguistics behaviours. In particular, social medias 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 standardized 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 unstandardized historical sources. To define appropriate tools, descriptions of these varieties are needed. Yet such descriptions need tools to be validated. We address this circularity interdisciplinarily, by working both on linguistics descriptions and on NLP tool development. Recently, sociodemographic variables have been shown to bear a strong impact on NLP processing tools. This is why, in a first step, jointly with researchers involved in the ANR project SoSweet (ENS Lyon and Inria's Dante), we 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 instance of 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 [62] uses such models to automatically generate forms attested in (classical) Indo-European languages but is based on a idiosyncrasic and inacceptable 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 [53] and [58]. Such models rely on a good understanding of the phonetic evolution of the languages at hand.

In ALMAnaCH, we focus 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 are in direct interaction with sub-strand 3b (development of lexical resources). We 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, 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 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 are to be paired with semantic resources (wordnets, word embeddings, and other corpus-based statistical information). This makes it possible 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 the other (e.g. for studying the non-inherited layers in the Ancient Greek lexicon). In the longer term, we intend to 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 rely on etymological datasets and standards for representing etymological information, for which see Section 3.4.2.

3.3.4. Accessibility-related variation

Language variation does not always constitute an additional complexity in 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 [78]). 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 [63]. 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 [57], [73], our goal is 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 therefore investigate hybrid techniques involving rule-based, statistical and neural approaches based on parsing results (for an example of previous parsing-based work, see [51]). Lexical simplification, another aspect of text simplification [60], [64], is also to be investigted. [0]

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-extraxction suite 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.

[0] http://www.unapei.org/IMG/pdf/GuidePathways.pdf

[0] We have started a collaboration with Facebook's Parisian FAIR laboratory, 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.

3.3.5. Intertextual variation

Language variation is not restricted to language-internal dimensions such as the effects of sociolinguistic and diachronic factors. It also involves variation in the way a same content can be expressed. Detecting, analysing and qualifying this type of variation is a challenge that can be applied in different settings, such as the automatic study of intertextuality in ancient documents (different versions of a same myth, for instance), automatic comparison of documents dealing with the same facts and citations (e.g. journalistic articles and news wires), assessment of textual entailment, and automatic detection of plagiarism. In ALMAnaCH, we put an emphasis on the first two of these examples.

Intertextual comparison of close witnesses of the same text produces valuable data on orthographic, morphological or semantic equivalences and variance (textual criticism). Automatic parallel detection not only informs about the positive intertextuality between two sources (e.g. the use of Biblical quotations among Church Fathers or Rabbinic authors) but also reveal the differences in their use and transformation of the same textual material, and therefore the authorial strategies and politics.

In automatic language processing, it is customary to focus on similarities when dealing with distinct documents. Instead, we can focus on modelling what is idiosyncratic to a certain text, given a reference. This can allow, for instance, to identify whether an elided passage is relevant or not. Identifying such relevant omissions was one of the goals of the VerDi Project (on which see below).

3.4. Modelling and Development of Language Resources

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 is involved). It is therefore necessary to design a workflow 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. This project-team will therefore work on improving print OCR for some of these languages for this very aim (e.g. Syriac, Ge'ez, Armenian). When an ancient source is still unpublished (book, manuscript, stele, tablet...), and therefore available in raw (image) form, we intend to develop OCR / HTR techniques, at least for certain scripts (Hebrew, Coptic and Greek Uncials, Ge'ez), and construct a pipeline for historical manuscripts. Initial success for Hebrew and Latin manuscripts has been very comforting (ca. 3% CER). On the one hand, access to existing electronic corpora, especially epigraphic corpora (e.g. Aramaic, North and South Arabic), still have to be negotiated. On the other hand, data that has been produced directly in electronic form (e.g. on social media) is readily usable, but far from normalised. Of course, contemporary texts can be often gathered in very large volumes, as we already do within the ANR project SoSweet, but this results in specific issues.

An inventory of already available resources developed or used by ALMAnaCH members has been developed.

The team 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 want to ensure maximum compatibilities with available international standards for representing textual sources and their annotations. More precisely we consider TEI guidelines as well the standards produced by ISO committee TC 37/SC 4 as essential points of reference.

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

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 workflows for exploiting corpora, starting from a 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 process may be seen as progressivily enriching the documents with new layers of annotations produced by various NLP modules and possibility validated by users, preferably in a collaborative way. It relies on the use of clearly identified representation formats for the annotations, as advocated by ISO TC 37/SC 4 and TEI, but also on the existence of well-designed collaborative interfaces for browsing and validation. ALMAnaCH has been or is working on several of the NLP bricks needed for setting such a workflow, and has a solid expertise in the issues related to standardisation (of documents and annotations). However, putting all these elements in a unified workflow that is simple to deploy and configure remains to be done.

It should be noted that such workflows have also a large potential besides DH, for instance for valorising internal documentation (for a company) or exploring existing relationships between entities. [0]

3.4.2. Development of Lexical Resources

ALPAGE, the Inria predecessor 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 [4], [1]). Such resources play a crucial role in all NLP tools, as has been proven among other tasks for POS tagging [71], [25], [29] and parsing, and some of the lexical resource development are targeted towards the improvement of NLP tools. They 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 are also one of the primary sources of linguistic information for augmenting language models used in OCR systems for ancient scripts, and allow us to develop automatic annotation tools (e.g. POS taggers) for low-resourced languages (see already [30]), especially ancient languages. Finally, semantic lexicons such as wordnets play a crucial role in assessing lexical similarity and automating etymological research.

Therefore, an important effort towards the development of new morphological lexicons is intended, with a focus on ancient languages of interest. Following previous work by ALMAnaCH members, we try and leverage all existing resources whenever possible such as electronic dictionaries, OCRised dictionaries, both modern and ancient [70], [19], [26], [27], while using and developing (semi)automatic lexical information extraction techniques based on existing corpora [69], [74]. A new line of research consists in the integration of the diachronic axis by linking lexicons that are in diachronic relation with the one another thanks to phonetic and morphological change laws (e.g. XIIth century French with XVth century French and contemporary French). Another novelty is 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 [19], [26], [27].

An underlying effort for this research is 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 dicitonaries such as the Dictionnaire Universel de Basnage (preparation of an ANR project with the University of Grenoble-Alpes and Paris Sorbonne Nouvelle).

In the same way as we signalled the importance of standards for the representation of interoperable corpora and their annotations, we intend to keep making the best 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"

[0] In this regard, we have started preliminary discussions with Fujitsu Lab and with the International Consortium of Investigative Journalists.

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. [0] We consider that contributing to standards allows us to stabilize 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 are either be only morphosyntactic or 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 [76], [65], [75], [61]) and now participate to the creation of valuable resources originating from the historical domain genre.

4. Application Domains

4.1. Application domains of NLP and Computational Humanities

ALMAnaCH's research areas cover Natural Language Processing (nowadays recognised 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 include:

- Information extraction, information retreival, 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. Highlights of the Year

5.1. Highlights of the Year

- ALMAnaCH's submission to the 2017 CoNLL multilingual parsing shared task was ranked 3rd (out of 33) in part-of-speech tagging, and 6th (out of 33) in dependency parsing.
- Joint submissions of ALMAnaCH and Stanford University to the Extrinsic Parsing Evaluation campaign ranked 1st and 3rd.

6. New Software and Platforms

6.1. Enqi

- Author: Benoît Sagot
- Contact: Benoît Sagot

6.2. SYNTAX

KEYWORD: Parsing

[0] Defined in ISO 24613 part 1 (core model), 2 (Machine Readable Dictionaries) and 3 (Etymology).

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/

6.3. 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/

6.4. 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/

6.5. 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

6.6. 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

6.7. 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/

6.8. 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/

6.9. 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/

6.10. 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

6.11. 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

6.12. 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/

6.13. 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

6.14. 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

7. New Results

7.1. Standardisation of Natural Language data

Participants: Loïc Grobol, Laurent Romary, Stefan Pernes, Jack Bowers, Charles Riondet, Mohamed Khemakhem.

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 devlopping NLP components. Furthermore, our specific leading roles in the domain of standardisation in both the Parthenos [41] and EHRI [40] projects as well as in related initiatives (TEI consortium, ISO committee TC 37, COST action ENeL (European Network in e-Lexicography), DARIAH lexical working group) has allowed to make progress along the following lines:

- Contribution to the improvement of the TEI guidelines [15], [20] and in particular to the definition of an extension for stand-off annotation in the continuity of [52] [0]
- Editing an ISO standard on the annotation of reference phenomena in discourse [0] that intends to be 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 from the point of view of its implementation [18]
- Editing the draft for the future project ISO 24613-4, which, on the basis of the proposals made in [67], intends to provide a reference TEI based serialisation for the LMF model (comprising core model (ISO 24613-1), machine readable dictionary (ISO 24613-2) and etymology (ISO 24613-3, cf. below) modules). This work is also the basis for the output format of Grobid-dictionary [19]
- Editing the draft for the future project ISO 24613-4, which will provide the model for representing etymological information in dictionaries and lexical resources, on the basis of [11]. Preliminary experiments have been carried out in [26], [27] (see also section 7.10)
- Proposal of a modular specification of the TBX standard (ISO 30642) by means of a TEI ODD specification [24]
- Participation to a call for contribution to the future evolution of the archival standard EAC-CPF (Encoded Archival Context for Corporate Bodies, Persons, and Families), proposing to use the TEI ODD specification language [47]

[0] https://github.com/laurentromary/stdfSpec
[0] https://www.iso.org/standard/69658.html

7.2. Digital Humanities and Cultural Heritage

Participants: Stefan Pernes, Marie Puren, Charles Riondet, Laurent Romary, Dorian Seillier, Lionel Tadonfouet.

The very broad scope of Digital Humanities and Cultural Heritage is well represented in the latest works of the ALMAnaCH team, undertaken in various contexts (European and national research infrastructures and bilateral partnerships). However, the issues tackled always deal with interoperability, reusability and standardization:

- The "Data Reuse Charter"[33] project is carried by a large consortium of European infrastructures and institutions
- The “Standardization Survival Kit” (or SSK) [66] developed within the PARTHENOS project intends to show that proper data modelling and corresponding standards make digital content more sustainable and reusable. Arts and Humanities would be well-suited to taking up the technological prerequisites of standardization [41], as most technological domains have already done.
- A concrete application of what offers the SSK has been developed within the EHRI project, where we built a methodology for the management of heterogeneous archival sources—expressed in the EAD Encoded archival description format—in one single environment, namely a federated portal [40], [48]. This method is based on a specification and customisation method inspired from the TEI, i.e. the definition of project-specific subsets of the standard and the maintenance of both technical and editorial specifications within a single framework.
- the Time-US project aims to reconstruct the remuneration and time budgets of women and men working in the textile trades in four French industrial regions (Lille, Paris, Lyon, Marseille) in a long-term perspective. During the launch phase, the team has been active in the following domains:
  - Collection of primary sources. The Time-Us team works on a heterogenous corpus of French handwritten and printed sources spanning from the seventeenth to the twentieth century; it includes court decisions, petitions, police reports and files, and sociological surveys on living conditions of the working class.
  - Evaluation of technical solutions for image visualization, transcription and collaboration, such as Transkribus (https://transkribus.eu/Transkribus/). The Transkribus interface enables Humanities scholars to transcribe handwritten and printed historical sources, and offers a very powerful Handwritten Text Recognition engine.
  - Creation of an annotation schema in XML/TEI. As the corpus gathers together diverse historical sources, the definition of a light and flexible annotation schema is a major step to create data to train parsing models. This data take the form of annotated texts encoded in TEI (Text Encoding Initiative). The annotation process starts as a collaborative effort, in order to get a first dataset that will later be used to train and configure NLP tools. The current step also helps designing a precise annotation guide between the NLP people and historians, in particular to clarify their expectations.
  - Installation of a customized MediaWiki. Several digital projects have already taken into account the specific needs of historians in terms of image visualization, transcription and collaboration. But they do not address all the requirements of Humanities scholars working on primary sources, and the need of comprehensive Digital Humanities-based publishing systems is emerging. We have chosen to setup a specific digital workflow enabling historians and NLP experts to work together, namely a wiki under Mediawiki (http://timeusage.paris.inria.fr/mediawiki/index.php/Accueil) with the Transcribe Bentham transcription desk, adapted to our needs, and a TEI toolbar, specifically customized for tagging named entities and measures.
- Archives nationales

- In a complex of projects (eRabbinica, LAKME, NEH/DFG Mishna-Tosefta Synopsis) with different partners dealing with classical rabbinic literature in Middle Hebrew we thrive to create a critical edition with translation, linguistic annotation and lexicon of the Mishna (200k tokens, the hypotext of the Talmud). Hebrew, a script written from right to left and a highly agglutinative language, poses great challenges to encoding standards and demands the development of new technical solutions. No open source corpora exist for linguistically annotated texts in rabbinic Hebrew.
  - Building on ocropus HTR capacities, we have added our own layout analysis algorithms for column and line segmentation [35] that have proven very succesful for literary manuscripts for the tasks of aligning existing transcriptions of manuscripts with the word and character ROIs and for new transcriptions reaching similar results to transkribus but with a much easier complete control of the layout analysis.
  - With our partners at the University of Maryland we have produced a preliminary TEI transcription of the most important manuscript Kaufman A50 (https://raw.githubusercontent.com/umd-mith/mishnah/master/data/tei/S07326.xml). Further improvements are currently undertaken. We have been able to use this transcription to realign it with the manuscript glyphs.
  - We have produced preliminary transcriptions of two further manuscripts (Cambridge 450.2 and Parma A) that are in the process of TEIzation. A fourth manuscript (Munich Cod. Ebr. 95) is currently in treatment.
  - Our partners at Dicta, have produced a preliminary automatical linguistic annotation of a vulgate text of the Mishna with HMMs with data for lemma, POS and morphological analysis. In the LAKME project, we have now manually corrected 25k tokens (ca. 12 percent of the whole text) that will be used to train RNN to improve the current transcription of the remaining text and enter a human-machine dialogue to fully annotate the whole Mishna. The annotation will not only be the first open source annotation. It will also be considerably more detailed than the excellent but closed annotation of the Israel Academy of the Hebrew Language (http://maagarim.hebrew-academy.org.il/). The resulting system will enable us to annotate other texts such as Tosefta and Halakhic Midrashim for the upcoming Sofer Mahir (tachygraph) project.

7.3. Computational Humanities and ancient texts

Participants: Daniel Stökl Ben Ezra, Marc Bui.

In collaboration with Jérémie Bosom and Dogu Kaan Eraslan (PhD students (co-)supervised by Marc Bui at EPHE).

Ancient languages of interest: ancient Egyptian (hieroglyphics, hieratic, demotic) , ancient Greek, Aramaic, Elamite, biblical Hebrew, classical Arabic, Hán Nôm (ancient vietnamese), old Persian

Computational approaches in humanities makes it possible to address the problems encountered by philologists such as reading, analyze and archiving old texts in a systematic way. We based our research on algorithms, their implementations, and human expertise on ancient languages to automate these difficult tasks.

The research scope of 2017 was the work around historical document or manuscripts available in images. Our work program (or work in progress) includes:

- Document layout analysis for ancient manuscripts using computer vision techniques and machine learning
- Script identification taking into account the environment where the trace is located: image, artefact, noise due to deterioration of the medium of writing. By stacking auto-encoding neural networks in order our approach provides an alternative representation of the input data received.
- Text recognition (handwritten text recognition) by enhancing it with LSTM

- Palaeographic classification of manuscripts and ancient inscriptions. Classification of historical document images can be addressed through script identification, in that case, our proposed method is based on the use of Convolutional Auto-Encoders (CAE) stacked in several layers in order to obtain fine-grained features and automatically learn representations of the line of writing or drawing of script
- Cross language Information Retrieval and Information Retrieval applied to ancient languages.

7.4. Information Extraction with GROBID

Participants: Luca Foppiano, Mohamed Khemakhem, Laurent Romary.

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 developped into a rich information extration environment, and deployed in many Inria projects, but also national and international services, among which we can quote HAL. It is a central piece for our information extraction activities and we have been particularly active in 2017 in the following domains:

- General contributions to GROBID (https://github.com/kermitt2/grobid):
  - 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)
- Contribution to entity-fishing (https://github.com/kermitt2/nerd):
  - integration into the main open-access platform: EKT/OMP, OAPEN, OpenEdition, Gottinghen University Library Press, Ubiquity press
  - deployment in the DARIAH infrastructure via Huma-NUM
  - adding supported languages for Italian and Spanish
  - various fixes and refactoring
  - Creation of a specific client for Historical documents, combined with a POS-tagger that connect the found entities between them and with their structural context[34]
- Contribution to GROBID-Dictionaries [0]: the lexical GROBID extension has been implemented and tested on modern and multilingual dictionaries [19]. The architecture has been further developed and an extension for etymology has been plugged-in on the top of the existing models. First experiments on etymological samples have been carried out and more work is required on the features selection. In parallel, the output of the system is actively synchronised with the Standardisation initiatives such as TEI Lex0 and ISO 24613 (LMF). Usability has been enhanced as well by lightening the annotation process and simplifying the setup process of the tool. Such measures are going to unlock the workforce potential of different interested research partners to generate more annotated data required for feature engineering. A first user experiment has been carried out during a dedicated workshop at the Lexical Masterclass, where the new features have been tested

7.5. Multilingual POS-tagging and Parsing

Participants: Éric Villemonte de La Clergerie, Djamé Seddah, Benoît Sagot, Héctor Martínez Alonso.

Our participation in 2017 to two international shared tasks (CONLL UD and EPE—the latter in collaboration with Stanford University) led us to develop a new generation of statistical multilingual NLP tools, in particular for POS-tagging and for Parsing [29]. In particular, the CoNLL shared task involved 80+ datasets covering 50+ languages (including low-resource and no-resource languages) and, for some languages, various genres.

[0] https://github.com/MedKhem/grobid-dictionaries

For POS tagging, we have developed a new feature-based POS tagger, following our previous work on MElt [56], [72]. This new tagger, named alVWTagger, uses the Vowpal Wabbit system for training linear POS models, resulting in an important drop in training times. This has allowed us to better explore the feature set space based on development data for each and numerous ways to encode the information provided by external morphological lexicons, resulting in better tagging results. We also developed a derivative of this tagger for performing tokenisation and sentence segmentation. Experiments on the development sets of the CoNLL shared task allowed us to chose the best setting for each corpus between several configurations, by using the UDPipe baseline (provided by the shared task organisers) or alVWtagger for each of the 3 subtasks (tokenisation, segmentation in sentences, UPOS tagging). As a result, we ranked 3rd (out of 33 participants) in the UPOS tagging ranking of the CoNLL shared task, and 5th for the tokenisation subtask and 6th for the sentence segmentation substask. Moreover, later improvements in the parsing models resulted in alVWtagger being more often used than for the official run, with improved results (unofficial post-campaign ranking on UPOS tagging: 2nd/33).

In parallel, we have developed a neural POS tagger based on Barbara Plank's LSTM tagger, by exploring the impact of integrating lexical information extracted from morphological lexicons within the neural architecture. We showed that such information improves POS tagging on average [25]. A careful comparison of this neural tagger, alNNtagger, w.r.t. alVWtagger is yet to be carried out, but preliminary experiments tend to show that both taggers perform similarly on average. This is likely because POS tagging is a relatively easy task for which the manual design of adequate features is relatively easy. As a result, using a neural architecture, which has the advantage of learning the optimal features rather than relying on manually crafted ones, does not result in massive improvements as observed in many other NLP tasks and beyond.

For Parsing, DyALog-SR, a feature-based parser on top of DyALog system, was extended (into DyALog-SRNN) to integrate predictions proposed by deep neuronal layers, based on a global char LSTM and a word bi-LSTM. Based on the results of the CONLL UD shared task, further extensions were added to DyALog-SRNN, namely an adaptation of Stanford's winner system (based on a bi-affine prediction of word governors) and a version of the Maximum-Spanning Tree (MST) algorithm, allowing us to move from the 6th place (for parsing) to an unofficial post-campaign 4th place.

The new version DyALog-SRNN has preserved the functionality of DyALog-SR to produce (deep) dependency graphs rather than standard shallow dependency trees. This functionality was used during the EPE (Extrinsic Parsing Evaluation) shared task to test several dependency tree and graph representations for several downstream application tasks [28].

The goal of that collaboration with the Stanford NLP team was to evaluate the usability of several representations derived from English Universal Dependencies (UD), as well as the Stanford Dependencies (SD), Predicate Argument Structure (PAS), and DM representations. We further compared two parsing strategies: Directly parsing to graph-based dependency representations and a two-stage process of first parsing to surface syntax trees and then applying rule-based augmentations to obtain the final graphs. Our systems used advanced deep learning techniques on top of state-of-the-art preprocessing and part-of-speech tagging. Overall, our systems performed very well and our results were ranked first and third on that shared task (over more than 20 submitted systems). The main advantage of that shared task was to provide an extrinsic evaluation scenario which consisted in extracting relevant information for information retrieval from speech and biomedical data, as well as opinion mining.This showed the relevance of our approach and the interest of producing graph-based representations to downstream applications that were developed for tree-based structures.

In particular, it showed the interest of deeper syntactic representation instead of shallow ones. In parallel with these efforts, work was also carried out on the issues related to polylexical units in parsing [17]. Moreover, the *International Journal of Lexicography* has accepted a paper written in collaboration with three other European research centres on the interactions between NLP and lexicography on polylexical units (to appear in 2018).

7.6. Tweet processing

7.6.1.

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

In the context of the SoSweet and Parsiti ANR actions, we run various experiments on large amounts of tweets.

In a first experiment, around 20 millions tweets were normalized, and then parsed with FRMG. A first observation was that the current level of pre-parsing normalization was not sufficient to ensure a good parsing coverage with FRMG (around 67%, to be compared with around 93% on FTB journalistic texts), also leading to high parsing times because of correction strategies. However, error mining was tried to identify a first set of easy errors and further developments are planned to track errors more related to segmentation and normalization. Clustering and word embedding were also tried for lemmas relying on the dependency parse trees, again leading to semi-successful results due to the poor quality of the pre-parsing phases.

In a second experiment, we adapted our two clustering (DepCluster) and word embeddings (DepGlove) algorithms to take into account non-linguistic relations, such as the author-word relation (between an author and the words of her tweets). The algorithms were applied on raw tweets with only a basic tokenisation, and results produced on a month basis over 18 months (2016/02 to 2017/08). Several tools, with a special focus on Cytoscape, were tried to visualize the results as networks, in order to identify and explain communities.

7.7. Syntax modelling and treebank development

Participants: Djamé Seddah, Héctor Martínez Alonso, Benoît Sagot, Elias Benaissa, Wigdan Abbas Mekki Medeni, Émilia Verzeni.

In 2017, ALMAnaCH members have contributed to the *Universal Dependency* initiative [44]:

- Héctor Martínez Alonso has resumed his contribution to the *Universal Dependencies* (UD) initiative, with annotations and data evaluations for Catalan, Danish and Spanish datasets.
- Several ALMAnaCH members have worked on converting the French TreeBank into the UD model and format (paper to be presented in 2018) and on the automatic identification of syntactic structures in UD.

As part of the ANR Parsiti project (2016-2020), whose goal is to build the next generation of context-enhanced NLP tools, we are currently developing a parallel data set of user-generated content language pairs, French-English and North-African dialect Arabic-French. Each of those pairs contains highly non-canonical text, heavily contextualized. We built the translation pairs and are currently carrying out annotations at the morpho-syntactic level. None of these data set already exist, they will be first used for the evaluation of our current processing chains and then to bootstrap state-of-the-art models as part of their training data. 3 annotators are involved over a year long period (18 man.month, end in June 2018).

7.8. Context-Enhanced NLP tools building

Participants: Djamé Seddah, Julie Tytgat, Florian Gouret, Yann-Alan Pilatte.

The ANR Parsiti project also aims 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. These information can be of spatial temporal nature for example, and have been shown to improve Entity Linking over social media streams [0]. In our case, we decided to focus on a closed world scenario in order to study context and speech acts interaction. We built 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 organized and enable the modelling of the extra-linguistics context with different level of granularity. Recorded over many games sessions, we 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.

[0] fang2014entity

7.9. Quantitative and computational morphology

Participant: Benoît Sagot.

In 2017 we have resumed our work on empirical and computational morphology, although at a slower pace than during the previous years. Apart from the preparation of an issue of the *Morphology* journal on computational morphology as a guest editor, together with Olivier Bonami (LLF) [10], our work in this regard was threefold:

- Contribution to the development of a morphological lexicon, a small-scale POS-annotated corpus and a POS tagger (based on MElt) for Romansh Tuatschin, a variety of the Sursilvan dialect of Romansh (a Romance language spoken in Switzerland); this work is a collaboration with Géraldine Walther and Claudia Cathomas (University of Zurich) [30];
- Formal and quantitative work on the verbal morphological system of Khaling, a Kiranti (Sino-Tibetan) language from Nepal, following earlier work of ours [80], [81]; this is a collaboration with Géraldine Walther (University of Zurich) and Guillaume Jacques (CRLAO, CNRS);
- Preliminary work on the diachronic modelling of lexical information at the morphological and phonetic levels.

7.10. Creation, Extraction and Standardisation of Etymological Information

Participants: Jack Bowers, Mohamed Khemakhem, Laurent Romary, Benoît Sagot.

A new, important line of research in 2017 was the work around etymological information and resources. This work can be divided into three main dimensions:

- Standards for the representation of etymological information.
- Extraction of etymological resources from existing datasets. Two main resource types were exploited:
  - Digitalised legacy etymological dictionaries, using GROBID-dictionaries, in collaboration with the Berlin-Brandenburg Academy of Sciences. The output of the process is a TEI-structured dictionary (see module 7.4 for more details).
  - The English Wiktionary, from which structured, formalised etymological information was extracted and published (open-source) in the form of a database of lexemes (i.e. language/lemma/meaning triples) and an associated database of etymological relations (input lexeme(s)/output lexeme/type of relation) [26], [27].
- Etymological research (i.e. producing novel etymological hypotheses), in collaboration with Romain Garnier (Université de Limoges & Institut Universitaire de France) and, although to a lesser extent, Laurent Sagart (CRLAO, CNRS) [12], [37]. Although limited (for now), the contribution of computational models in our research is real; it allowed us to check the validity of the diachronic phonetic evolution model we have postulated for a new, hypothetical Indo-European language we suggest could have served as a source of borrowings for the ancestors of both Greek and Italic languages [12].

7.11. Automatic Detection of Coreference

Participants: Éric Villemonte de La Clergerie, Loïc Grobol.

In 2017, ALMAnaCH members have investigated coreference detection for French using machine learning and existing linguistic knowledge. Our efforts consisted in using insight gathered from deep and shallow parsers and standard machine learning approaches to detect entity mentions [31], adapting knowledge-poor deep-learning techniques for end-to-end coreference resolution to the case of oral French and researching new ways of exploiting structured such as parse trees in deep neural models.

7.12. Detecting omissions in journalistic texts

Participants: Héctor Martínez Alonso, Benoît Sagot.

In the journalistic genre that is characteristic of online news, editors make frequent use of citations as prominent information; yet these citations are not always in full. The reasons for leaving information out are often motivated by the political leaning of the news platform.

Existing approaches to the detection of political bias rely on bag-of-words models that examine the words present in the writings. In the context of the VerDI project (see below), we have resumed our work aimed at going beyond such approaches, which focus on what is said, by instead focusing on what is *ommited*. Thus, this method requires a pair of statements; an original one, and a shortened version with some deleted words or spans. The task is then to determine whether the information left out in the second statement conveys *substantial* additional information. If so, we consider that a certain statement pair presents an omission. To tackle this question, we used a supervised classification framework, for which we require a dataset of sentence pairs, each pair manually annotated for omission.

We had developed last year a small reference corpus for evaluation purposes, using and comparing both crowd and expert annotation. This corpus has allowed us to examine which features help automaticallly identify cases of omission. In 2017, we have finalized the annotation tools for the VerDI project [23], and published them online as free software (see below).

8. Bilateral Contracts and Grants with Industry

8.1. Industrial Collaborations

- **Verbatim Analysis**: this Inria start-up was co-created in 2009 by BS. It uses some of ALPAGE/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. A new Inria startup, **opensquare**, was co-created in December 2016 by BS with 2 senior specialists of HR consulting. It is dedicated to designing, carrying out and analysing employee surveys as well as 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 starting with Facebook's Parisian FAIR laboratory. It should start with a co-supervised (CIFRE) PhD thesis in collaboration with UNAPEI, the largest French federation of associations defending and supporting people with special needs and their families (the CIFRE application has just been submitted). This collaboration is expected to be part of 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 initiates a collaboration in 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 an engineer and 12 months for a research ingineer).
- **Science Miner**: ALMAnaCH (following 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.
- **Konverso**: A collaboration with this start-up is starting, focused on chatbots and text generation. One of our objectives with this collaborations is to initiate a larger initiative involving ALMAnaCH and several small companies, whose goal will be the development of open-source, NLP-enhanced

chatbot modules. This is because such developments are complex and would benefit from such a mutualisation initiative. In turn, an open-source chatbot engine would allow startups and ALMAnaCH to more rapidly develop and deploy high-performance application-specific chatbots. The first concrete outcome of this collaboration is our joint submission to the call for projects published by the DILA (French government agency) for exploring the relevance of deploying a chatbot on the public information plateform service-public.fr.

- There exists at least one formal collaboration between a company and EPHE involving future ALMAnaCH members. It involves **Insight-Signals**, an EPHE start-up that "designs data analytics and decision support systems that integrate the complexity of humans' behaviour and their interactions".
- **Trooclick**: A direct and active collaboration with this company is now strengthened by the "RAPID" ANR project VerDI on the automatic detection of omissions in news reports and other types of texts. This project will come to an end in February 2018.
- ALMAnaCH members have recently initiated discussions with other companies (Fujitsu, HyperLex, Fortia Financial Solutions...), 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 (Google, Toyota, Samsung...).

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. 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

9.1.2. 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]).

- **PSL project LAKME** (2015-2017, PI Thierry Poibeau [LATTICE]). Topic: language resource development for morphologically rich languages, especially Rabbinic Hebrew (syntactic level), Medieval French (morphological level) and some Finno-Ugric languages (to a lesser extent).
- **PSL Iris project SCRIPTA** This project emanates from the history and philology department of the EPHE (DSBE). It is directed by Andreas Stauder (EPHE) with Philip Huyse (EPHE) and Charlotte Schmid (EFEO). It unites the forces of a great number of researchers in PSL (EPHE, ENS, EHESS, ENC, Collège de France and in addition the IRHT) working on written texts in all its forms, on all kinds of material, from all periods and regions and has important digital and computational ambitions especially with regard to epigraphy, palaeography, digital editions and NLP.

9.1.3. Other National Initiatives

- **TGIR Huma-Num** ALPAGE was a member of the CORLI consortium on "corpora, languages and interactions" (BS is a member of the consortium's board), and ALMAnaCH is in the process of joining this consortium. With a joint funding of Huma-Num and the H2020 project Parthenos (on which see below), ALMAnaCH members have also co-organised a workshop on 3D techniques for Humanities in Bordeaux (December 2016).
- **Institut de Linguistique Française (ILF)**: ALPAGE was a member of this CNRS "federation". ALMAnaCH is in the process of joining this federation if possible, especially as BS is the scientific head of the "Corpus de Référence du Français" initiative, an ILF project whose other head is Franck Neveu and whose goal is to develop a French National Corpus, a resource that has been awaited for a long time.
- **Notary registers 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 gigitized 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.

9.2. European Initiatives

9.2.1. 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.

9.2.2. 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 is president of the board of director) 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

9.2.3. Collaborations with Major European Organizations

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

- University of Ljubljana (Darja Fišer) [wordnet development]
- University of Zürich, Switzerland (Géraldine Walther) [computational morphology, lexicons]
- Academy of Sciences, Berlin, Germany (Karl-Heinz Moerth) [lexicology]
- University of Fribourg, Switzerland [historical document analysis]
- University of Valencia, Spain [historical document analysis]
- University of Groningen, Netherlands [historical document analysis]
- University of Innsbruck, Austria [historical document analysis]

9.3. International Initiatives

9.3.1. International Partners

- **ANR-NSF project MCM-NL** (2016-2020, PI John Hale [Cornell University, USA], resp. for Inria Paris / ALMAnaCH: EVdLC) Topic: exploring correlations between data from neuro-imagery (fMRI, EEG) and data from NLP tools (mostly parsers). The data will come from "Le Petit Prince" read in French and English, and parsed with different parsers. Other partners: Cornell Univ., Univ. Michigan, Paris Saclay/Neurospin, Univ. Paris 8. Informal collaborations:
- **The SPMRL initiative** (Statistical Parsing of Morphologically Rich Languages): a worldwide network of internationally renowned teams that was initiated during the IWPT'09 conference ALPAGE organised in Paris, DS playing a leading role since then. Other institutions involved include the University of Heidelberg (Germany), Bar Ilan University (Israel), Potsdam University (Germany) and Indiana University (USA). The outcomes of this initiative include the successful SPMRL Workshop and Shared Task series hosted successively by NAACL-HLT (2010), IWPT (2011), ACL (2012), EMNLP (2013), CoLing (2014) and IWPT (2015), in which DS as well as other ALPAGE/ALMAnaCH members played an active role. DS also served as a co-editor of a special issue of Computational Linguistics on this topic.
- **Sofer Mahir ("fast scribe") project**. Joint work on the computational processing of Rabbinic Hebrew manuscripts involving DSBE: Nachum Dershowitz (Tel Aviv University, Israel), Moshe Koppel (DICTA, Bar Ilan University, Israel), Meni Adler (DICTA, Ben Gurion University, Israel), Michael Elhadad (Ben Gurion University, Israel) on the NLP side and Hayim Lapin (University of Maryland, USA), Tal Ilan (FU Berlin, Germany) Shamma Friedmann (Bar Ilan University, Israel) on morphological analysis of Rabbinic Hebrew, alignment of manuscript witnesses (textual criticism), finding parallels, aligning related but different texts (like the Gospels). This work is also connected to the LAKME project mentioned above.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Daniel Dakota (Indiana University, 4 months, until Jan 2017)
- Theresa Lynn (Dublin City University, 10 days in January 2017)
- Amir More (Open University of Israel, 10 days in April 2017)

9.4.1.1. Internships

- Basant Agarwal (ERCIM, Aug-Sep 2017)

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- DSBE: Co-Chair and coorganizer of the summerschool manuSciences '17, Fréjus, France, 10-15 September 2017.
- LR, DSBE: Coorganizers of the France/Israel DH conference, Jerusalem, 27 February-1 March 2017.
- DSBE: Co-chair and coorganizer of the workshop Cambridge-PSL-Lausanne, Paris, France, 11-12 March 2017.
- DSBE: Co-chair and coorganizer of the workshop FSP-Patrima-EPHE on Scientific Approaches to Inscribed Objects, Ministry of Culture and Communication, Paris, France, 24 January, 2017.

10.1.1.2. Member of the Organizing Committees

- LR, BS: Members of the Organizing Committee of the Lexical Data Masterclass, Berlin, 4-8 December 2017
- CR: Member of the Organizing Committee of the conference "La Part de l'ombre, Action clandestine et imaginaire du complot, XXe-XXIe siècles", Paris, France, 18-19 May 2017
- MP: Member of the Organizing Committee of the conference Text as a resource. Text mining in Historical Science, Paris, France, 29-30 June 2017.
- MP: Member of the Organizing Committee of the masterclass Penser/ Utiliser les données de la recherche, Paris, France, 25-29 September 2017.
- DSBE: Member of Organizing Committee for the Poster Session on Digital Humanities at the World Congress of Jewish Studies, Jerusalem, 6-10 August 2017.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- MB, Program chair of the 8th ACM-SoICT, 2017, Nha Trang, Vietnam.

10.1.2.2. Member of the Conference Program/Scientific/Reviewing Committee

- LR: Member of the Reviewing Committee of the following conferences: DHd 2017, DATeCH2017, ACL 2017, TOTh 2017, TPDL 2017, MDQual 2017
- EVdLC: Member of the Reviewing Committee of ACL'17 (Tagging, Chunking, Syntax and Parsing area), DaTeCH'17 (co-reviewer), TMPA-2017 (co-reviewer)
- EVdLC: Member of the Scientific, Programm or Reviewing Committee of EMNLP'17, DepLing'17, EPIA'17, LATA'18, Games4NLP'17, ToTH'17, LREC'18.

- DS: Member of the Scientific, Programm or Reviewing Committee of ACL'17, EMNLP'17, CoNLL'17, TALN2017, LREC2018, TLT #15, LAW 2017, W-NUT 2017
- DSBE: Member of the Program Committeee for HIP'17, member of the Reviewing Committee of HIP 2017
- BS: Member of the Program, Scientific or Reviewing Committe of the following conferences and workshops: *SEM 2017, CoNLL 2017, DeriMo 2017, EACL 2017, LAW XI (2017), LREC 2018, WRDTM 2017
- CR: Member of the Reviewing Committee of DH Nord and DH 2018
- MP: Member of the Reviewing Committee of the following conferences: DH Nord, Digital Humanities in the Nordic Countries, DH 2018

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- LR: Member of the Editorial Board of the following journals: *JDMDH*, *ACM JOCCH*
- LR: Member of the Scientific/Advisory Committee of the following journals: *Revue Humanités Numériques*, Journal of the TEI, *Language Resources and Evaluation*
- DSBE: Member of the Advisory Committee of the following journals: *Jewish Studies Quarterly*, *Henoch*, *Biblical Annals*, *Digital Paleography and Book History*, *Cultural Heritage Digitization*, *Marginalia*
- DSBE: Member of the Advisory Committee of the book series "Humanités numériques et patrimoine" (U. Grenoble, E. Pierazzo)
- BS, together with Olivier Bonami (LLF), was Guest Editor for a thematic issue of the journal *Morphology* on computational methods for descriptive and theoretical morphology [10].

10.1.3.2. Reviewer - Reviewing Activities

- BS: Reviewer for the following journals: *Computational Linguistics*, *Journal of Language Modelling*, *Arabian Journal for Science and Engineering*, *wék^{w}os*
- BS: Reviewer for the MIT Press
- DS: Computational Linguistics Journal, Transactions on Asian and Low-Resource Language Information Processing Journal

10.1.4. Invited Talks

- Héctor Martínez Alonso gave a talk at the Inria Paris Junior Seminar, Paris, France
- CR, "Traces de l'héroïsme, les plaques commémoratives de la Résistance parisienne", Plaques commémoratives dans l'espace parisien, Paris, France
- LF, CR, "Grobid for Humanities, when engineering meets History", DHIHA, Paris, France
- CR, "La Libération du Sud-Est parisien, un mouvement populaire, CLIO94, Créteil, France
- CR, "Traces de l'héroïsme, les plaques commémoratives de la Résistance parisienne", Plaques commémoratives dans l'espace parisien, Paris, France
- MP, "Fcailiter l'accès des chercheurs aux données patrimoniales. La Charte de réutilisation des données", séminaires Nouveaux champs d'étude en droit du patrimpine culturel, Le Mans, France, 14 December 2017.
- MP, "La TEI et les historiens. Le projet Time-Us : Travail, rémunération, textile et foyer (XVIIe-XXe siècles)", Edition électronique & TEI : enjeux, pratiques & perpectives, 4ème journée dayClic TEI, Le Mans, France
- DSBE: 'Automatic layout analysis and transcription of medieval manuscripts", Mondes anciens nouveaux regards, AnHIMA, Paris, 8 June 2017.
- DSBE: "Digital Humanities", University of Haifa, 10 June 2017.

- DSBE: “Automatic lectionary analysis with the database Thesaurus Antiquorum Lectionarium Ecclesiae Synagogaeque”, Hagiographico-homiletic collections in Greek, Latin and Oriental Manuscripts –Histories of Books and Text Transmission in a comparative perspective, Hamburg, 27 June 2017.
- DSBE: “Teaching DH and Jewish Studies”, Hamburg, 5 September 2017.
- DSBE: “Automatic layout analysis and transcription of medieval Hebrew manuscripts”, Jerusalem, 22 June 2017.
- DS: "Robust Data Driven Parsing of Deep Syntax Graph: Syntax Is Not Dead Yet", Naverlabs, ex-Xerox, Grenoble, December 8th, 2017.

10.1.5. Leadership within the Scientific Community

- LR: President of the board of directors of DARIAH
- LR: Member of the board of directors or the TEI consortium
- LR: President of ISO committee TC 37 (Language and terminology)
- 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 (since Dec. 2017)
- DS: Board member of the French NLP society (Atala, 2017-2020), program chair of the "journée d’études".
- Member of the ACL’s BIG (Broad Interest Group) Diversity group.

10.1.6. Scientific Expertise

- EVdLC: Reviewer for an European ERC proposal and for an European COST proposal
- BS: Reviewer for the ANR (CE32 committee)
- EVdLC: Member of the selection committee for the call for projects “Langues et numérique 2017” organized by DGLFLF (Délégation générale à la langue française et aux langues de France)
- DSBE: Reviewer for a proposal at the SNF.

10.1.7. Research Administration

- XY: [description]
- DSBE: Director of the DH Programme of the EPHE.
- 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

10.1.8. Teaching

Licence: Djamé Seddah, “Certificat Informatique et Internet”, 30h, L1-L2-L3, Université Paris Sorbonne, France

Licence: Djamé Seddah, “Programmation et algorithmique en Java”, 50h, L2, Université Paris Sorbonne, France

Licence: Loïc Grobol, “Informatique et Industries de la Langue”, 22h, L2, Université Sorbonne Nouvelle, France

Licence: Loïc Grobol, “Introduction aux Humanités Numériques”, 15h, L2, Université Sorbonne Nouvelle, France

Master: Djamé Seddah, “Modèles pour la linguistique computationnelle”, 36h, M2, Université Paris Sorbonne, France

Master: Djamé Seddah, “Modèles pour la linguistique computationnelle”,36h, M1, Université Paris Sorbonne, France

Master: Djamé Seddah, “Programmation générique et C++”, 26h, M1, Université Paris Sorbonne, France

Master: Djamé Seddah, “Programmation réseau et Java”, 26h, M1, Université Paris Sorbonne, France

Master: Djamé Seddah, “Traduction automatique”, 30h, M2, Université Paris Sorbonne, France

Master: Laurent Romary, “Governance challenges in setting up and running an ERIC”, 1h30, Webinar RItrain – Executive Master in Management of Research Infrastructures at University of Milano-Bicocca, 24 November 2017

Master: Marie Puren, Charles Riondet, “Formation à la TEI pour des documents historiques”, 3 heures, M1-M2, Université Aix-Marseille, France, 19 October 2017

Master: Marie Puren, “Valorisation de la recherche - Humanités numériques”, 4 X 3 heures, M2 Histoire, Université Versailles-Saint-Quentin, Rennes, France, January-March 2017.

Master/PhD: Daniel Stökl Ben Ezra, “Approches numériques aux textes du judaïsme ancien”, 26 heures, M2/PhD, EPHE

Master/PhD: Marc Bui, “Introduction à la programmation Python pour les chercheurs en SHS” (2x24h) niveau M1/M2/PhD, EPHE (EPHE students and Master humanités numériques ENC-ENS-EPHE students)

Master class, Daniel Stökl Ben Ezra, “Simple and Advanced Image Treatment for Manuscript Analysis”, 2h, manuSciences summer school, 11 September 2017

Master class: Laurent Romary, “Overview of lexical models and introduction to the TEI dictionary chapter”, 1h30, Lexical Master Class, Berlin, 5 December 2017

Master class: Laurent Romary, “Querying and presenting TEI dictionary data with XSLT”, 1h30, Lexical Master Class, Berlin, 6 December 2017

Master class: Marie Puren, “Data management practices and recommendations. Managing, sharing and preserving linguistic data”, 1h30, Lexical Master Class, Berlin, 7 December 2017

Master class: Mohamed Khemakhem, “GROBID-Dictionaries”, 4 X 3 heures , Lexical Master Class, Berlin, 5-7 December 2017

EPHE, Marc Bui, “Introduction à la conception des bases de données avec SQL” (24h)

EPHE, Marc Bui, “La mise en page avec LateX pour les chercheurs en SHS” (12h)

IES Inria, Laurent Romary, “Formation à la TEI, XPath et XSLT pour HAL”, 3 journées Centre Inria Paris

IES Inria, Marie Puren, Charles Riondet, “Gestion des données de la recherche”, 7 heures, Centre Inria Paris, France, 31 March 2017

INIST, Marie Puren, “Créer un plan de gestion des données” et “Les métadonnées dans un DMP”, 2h30, INIST, Vandoeuvre-Lès-Nancy, France, 6 July 2017.

INIST: Laurent Romary, “Formation à la TEI pour les documents scientifiques”, 2 X 2 journées, INIST, Vandoeuvre Lès Nancy, France

URFIST, Marie Puren, “Données de recherche : le Plan de Gestion des Données”, 1 journée, URFIST, Rennes, France, 1st June 2017

10.1.9. Supervision

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 Isabelle Tellier (main superviser), Éric de La Clergerie and Marco Dinarelli

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: 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”, October 2016, EPHE, 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: Mohamed Khemakhem, “Structuration automatique de dictionnaires à partir de modèles lexicaux standardisés”, September 2016, Paris Diderot, Laurent Romary

PhD in progress: Antony Perrot, “Qumran Opisthographs”, started in Oct 2015, EPHE, PSL, supervised by Daniel Stökl Ben Ezra

PhD in progress: Jérémie Bosom , “Big data, internet des objets, fouille de données: élaboration de services intelligents pour le pilotage industriel”, started in Oct 2015, EPHE, PSL, supervised by Marc Bui

PhD in progress: Dogu Kaan Eraslan, “Les relations entre Milet et l'Égypte à la Basse Époque (664-332 av. J.-C.)” (with a strong emphasis on computationa humanity approaches for both encoding information and extracting information, e.g. with neural image processing techniques on ancient documents), started in Oct 2015, EPHE, PSL, co-supervised by Michel Chauveau and Marc Bui

10.1.10. Juries

- EVdLC: member of the PhD committee for Jakub Waszczuk at University of Blois on June 26th (Title: “Leveraging MWEs in practical TAG parsing: towards the best of the two worlds”; Supervisors: Agata Savary and Yannick Parmentier)
- MB: member of the PhD committee for Karim Sayadi at Université Paris 6 en partenariat avec EPHE on March 28th (Title: “Classification du texte numérique et numérisé. Approche fondée sur les algorithmes d'apprentissage automatique”; Supervisor: Marc Bui)
- EVdLC: examiner and member of the Master 2 jury for Jean Argouarc'h (Master de sciences cognitives, Paris Diderot; title: “Semantic models for analysis of brain activation during naturalistic text listening”; Supervisor: C. Pallier)
- BS: member of the recruiting committee for a Maître de Conférences position (NLP) at Université Paris Sorbonne

10.2. Standardization activities

10.2.1. ISO TC 37/ SC4

- Mohamed Khemakhem and Laurent Romary: Project leaders of the ISO 24613-4 LMF “TEI Serialisation”
- Jack Bowers: Project leader of the ISO 24613-3 LMF “Etymology Extension”
- Éric de la Clergerie: Participation to AFNOR meetings, in relation with TC37/ SC4

10.2.2. COST ENEL & DARIAH

- Laurent Romary, Mohamed Khemakhem, Axel Herold and Jack Bowers: Experts of a joint lexical standardisation action “TEI Lex0: Towards Best TEI P5 Encoding Practices”

10.3. Popularization

- BS, jointly with Emmanuel Dupoux (EHESS & ENS), gave a talk on “NLP and AI” as part of the seminar on AI organised by the Institut de l’École Normale Supérieure (21 June 2017)
- EVdLC and BS presented ALMAnaCH’s research (as well as its spin-off opensquare) at the “Rencontres Inria-industries” (forum bringing together Inria and companies) (18 November 2017)
- DSBE and BS: participation to a meeting on Digital Humanities bringing together researchers from the PSL ComUE, the University of Cambridge (UK) and the EPFL (Switzerland) (11 May 2017)
- BS: with Jean Ponce, Isabelle Ryl and Hélène Robak, represented the Inria Paris research center at the forum organised for the 30th anniversary of the DRM (the French Military Intelligence Office) (23 March 2017)
- BS and DSBE: talks during the NLP edition of the “Paris Sciences et Data” conference series.
- EVdLC: talk at the opening of the Math Olympiades 2017 about “Une palette mathématique pour appréhender le langage” (Versailles, January 25 2017)
- EVdLC: talk at the GFII DIXIT Seminar on “IA et Traitement Automatique des Langues (TAL) : Quel panorama ?” (Paris, February 24th 2017); member of the organizing committee of the forum 2017 of the GFII and panelist of the session on“de l’IA washing à la réalité industrielle, quels sont les contours du renouveau actuel de l’IA ?” (December 5th, 2017)
- EVdLC: co-animator of a new GFII Working Group “Technologies de la Connaissance”, with a focus on AI
- Loïc Grobol: participation to the 18th meeting on “Culture & Jeux Mathématiques” organized by AMIES
- MP, LR : Webinar on “Humanities and Open Science: Workflows and tools for publishing, licensing, versioning, identifiers, archiving, software...” for the International Open Access Week, 26 October 2017.
- DS gave a talk on “From Noisy Questions to Minecraft Texts: Annotation Challenges in Extreme Syntax Scenarios” at the NLP Paris Meetup, Paris, November 22th, 2017.

11. 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, 35 [*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] 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.

[5] 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.

[6] 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.

[7] 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.

[8] R. TSARFATY, D. SEDDAH, S. KUEBLER, 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.

[9] 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

Articles in International Peer-Reviewed Journal

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

[11] J. BOWERS, L. ROMARY.*Deep encoding of etymological information in TEI*, in "Journal of the Text Encoding Initiative", August 2017, n^{o} 10, https://arxiv.org/abs/1611.10122 [*DOI :* 10.4000/JTEI.1643], https://hal.inria.fr/hal-01296498.

[12] R. GARNIER, B. SAGOT.*A shared substrate between Greek and Italic*, in "Indogermanische Forschungen", September 2017, vol. 122, n^{o} 1, p. 29-60 [*DOI :* 10.1515/IF-2017-0002], https://hal.inria.fr/hal-01621467.

[13] B. SAGOT.*Représentation de l'information sémantique lexicale : le modèle wordnet et son application au français*, in "Revue Française de Linguistique Appliquée", 2017, vol. XXII, https://hal.inria.fr/hal-01583995.

Invited Conferences

[14] A. BAILLOT.*Zahlenwahn oder Textliebe? Digitale Philologie als Disziplin und als Weltanschauung*, in "Machines / Maschinen Les 5O ans de l'AGES", Nantes, France, Association des Germanistes de l'Enseignement Supérieur, June 2017, https://halshs.archives-ouvertes.fr/halshs-01562486.

[15] L. ROMARY.*The Text Encoding Initiative as Infrastructure*, in "French-Israeli Symposium on Digital Humanities", Jerusalem, Israel, February 2017, https://hal.inria.fr/hal-01618017.

International Conferences with Proceedings

[16] M. CANDITO, B. GUILLAUME, G. PERRIER, D. SEDDAH.*Enhanced UD Dependencies with Neutralized Diathesis Alternation*, in "Depling 2017 - Fourth International Conference on Dependency Linguistics", Pisa, Italy, September 2017, https://hal.inria.fr/hal-01625466.

[17] M. CONSTANT, H. MARTINEZ ALONSO.*Benchmarking Joint Lexical and Syntactic Analysis on Multiword-Rich Data*, in "MWE 2017 - 13th Workshop on Multiword Expressions", Valencia, Spain, Association for Computational Linguistics, April 2017, p. 181 - 186, https://hal.inria.fr/hal-01677416.

[18] 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.

[19] 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.

[20] H. MARAOUI, K. HADDAR, L. ROMARY.*Encoding prototype of Al-Hadith Al-Shareef in TEI*, in "ICALP 2017 - The 6th International Conference on Arabic Language Processing", Fes, Morocco, October 2017, 14, https://hal.archives-ouvertes.fr/hal-01574543.

[21] H. MARTINEZ ALONSO, Ž. AGIĆ, B. PLANK, A. SØGAARD.*Parsing Universal Dependencies without training*, in "EACL 2017 - 15th Conference of the European Chapter of the Association for Computational Linguistics", Valencia, Spain, Association for Computational Linguistics, April 2017, vol. 1, p. 229 - 239, https://hal.inria.fr/hal-01677405.

[22] H. MARTINEZ ALONSO, B. PLANK.*When is multitask learning effective? Semantic sequence prediction under varying data conditions*, in "EACL 2017 - 15th Conference of the European Chapter of the Association for Computational Linguistics", Valencia, Spain, April 2017, p. 1-10, https://hal.inria.fr/hal-01677427.

[23] H. MARTÍNEZ ALONSO, A. DELAMAIRE, B. SAGOT.*Annotating omission in statement pairs*, in "11th Linguistic Annotation Workshop", Valencia, Spain, April 2017, p. 41-45, https://hal.inria.fr/hal-01584035.

[24] S. PERNES, L. ROMARY, K. WARBURTON.*TBX in ODD: Schema-agnostic specification and documentation for TermBase eXchange*, in "LOTKS 2017- Workshop on Language, Ontology, Terminology and Knowledge Structures", Montpellier, France, September 2017, https://hal.inria.fr/hal-01581440.

[25] B. SAGOT, H. MARTÍNEZ 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.

[26] B. SAGOT.*Construction automatique d'une base de données étymologiques à partir du wiktionary*, in "Traitement Automatique des Langues Naturelles 2017", Orléans, France, June 2017, https://hal.inria.fr/hal-01584013.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

National Conferences with Proceeding

[31] L. GROBOL, I. TELLIER, É. DE LA CLERGERIE, M. DINARELLI, F. LANDRAGIN. *Experiences in using deep and shallow parsing to detect entity mentions in oral French*, in "TALN 2017", Orléans, France, Actes de la 24e Conférence sur le Traitement Automatique des Langues Naturelles (TALN), Association pour le Traitement Automatique des Langues (ATALA), June 2017, https://hal.inria.fr/hal-01558711.

[32] D. SEDDAH, M. CANDITO.*Building a Question Treebank for French : The French QuestionBank*, in "ACor4French - Les corpus annotés du français", Orléans, France, Acts de l'Atelier ACor4French - Les corpus annotés du français, June 2017, https://hal.inria.fr/hal-01682869.

Conferences without Proceedings

[33] A. BAILLOT, M. PUREN, C. RIONDET, D. SEILLIER, L. ROMARY.*Access to cultural heritage data. A challenge for digital humanities*, in "Digital Humanities 2017", Montréal, Canada, August 2017, https://hal.archives-ouvertes.fr/hal-01582176.

[34] C. RIONDET, L. FOPPIANO.*GROBID for Humanities When engineering meets History*, in "Text as a Resource. Text Mining in Historical Science", Paris, France, Institut Historique Allemand, June 2017, https://hal.inria.fr/hal-01585693.

[35] M. SEURET, D. STÖKL BEN EZRA, M. LIWICKI.*Robust Heartbeat-based Line Segmentation Methods for Regular Texts and Paratextual Elements*, in "HIP 2017 - Proceedings of the 4th International Workshop on Historical Document Imaging and Processing", Kyoto, Japan, November 2017, https://hal.archives-ouvertes.fr/hal-01677054.

Scientific Books (or Scientific Book chapters)

[36] PARTHENOS (editor). *Digital 3D Objects in Art and Humanities: challenges of creation, interoperability and preservation. White paper: A result of the PARTHENOS Workshop held in Bordeaux at Maison des Sciences de l'Homme d'Aquitaine and at Archeovision Lab. (France), November 30th - December 2nd, 2016*, PARTHENOS, Bordeaux, France, May 2017, 71, https://hal.inria.fr/hal-01526713.

[37] R. GARNIER, L. SAGART, B. SAGOT.*Milk and the Indo-Europeans*, in "Language Dispersal Beyond Farming", M. ROBEETS, A. SAVALYEV (editors), John Benjamins Publishing Company, December 2017, p. 291–311 [*DOI :* 10.1075/Z.215.13GAR], https://hal.inria.fr/hal-01667476.

[38] D. STÖKL BEN EZRA.*The Mishnah into French: translation issues*, in "Studies in Mishnaic Hebrew and Related Dialects : Proceedings of the Yale Symposium, May 2014", E. B. ASHER, A. KOLLER (editors), Studies in Mishnaic Hebrew and Related Fields Proceedings of the Yale Symposium on Mishnaic Hebrew, May 2014, The Program in Judaic Studies, Yale University, December 2017, p. 349-367, https://hal.archives-ouvertes.fr/hal-01677074.

Books or Proceedings Editing

[39] A. DEGKWITZ, L. ROMARY (editors). *IFLA Satellite Meeting - Digital Humanities – Opportunities and Risks: Connecting Libraries and Research*, 2017, https://hal.inria.fr/hal-01643305.

Research Reports

[40] C. RIONDET, L. ROMARY, A. VAN NISPEN, K. J. RODRIGUEZ, M. BRYANT.*Report on Standards*, Inria Paris, March 2017, n^{o} D.11.4, https://hal.inria.fr/hal-01503235.

[41] L. ROMARY, P. BANSKI, J. BOWERS, E. DEGL'INNOCENTI, M. ĎURČO, R. GIACOMI, K. ILLMAYER, A. JOFFRES, F. KHAN, M. KHEMAKHEM, N. LARROUSSE, A. LITKE, M. MONACHINI, A. V. NISPEN, M. OGRODNICZUK, N. PAPADAKIS, G. PASTORE, S. PERNES, M. PUREN, C. RIONDET, M. SANZ, M. SANESI, P. SIOZOS, R. D. VALK.*Report on Standardization (draft)*, Inria, May 2017, n^{o} Deliverable 4.2, https://hal.inria.fr/hal-01560563.

[42] D. SEILLIER, A. BAILLOT, M. PUREN, C. RIONDET.*Survey on researchers requirements and practices towards Cultural Heritage institutions: Documentation and analysis*, Inria Paris, July 2017, https://hal.inria.fr/hal-01562860.

Scientific Popularization

[43] J. EDMOND, F. FISCHER, M. MERTENS, L. ROMARY.*The DARIAH ERIC: Redefining Research Infrastructure for the Arts and Humanities in the Digital Age*, in "ERCIM News", October 2017, n^{o} 111, https://hal.inria.fr/hal-01588665.

Other Publications

[44] J. NIVRE, Ž. AGIĆ, L. AHRENBERG, L. ANTONSEN, M. J. ARANZABE, M. ASAHARA, L. ATEYAH, M. ATTIA, A. ATUTXA, L. AUGUSTINUS, E. BADMAEVA, M. BALLESTEROS, E. BANERJEE, S. BANK, V. BARBU MITITELU, J. BAUER, K. BENGOETXEA, R. A. BHAT, E. BICK, V. BOBICEV, C. BÖRSTELL, C. BOSCO, G. BOUMA, S. BOWMAN, A. BURCHARDT, M. CANDITO, G. CARON, G. CEBIROĞLU ERYIĞIT, G. G. A. CELANO, S. CETIN, F. CHALUB, J. CHOI, S. CINKOVÁ, Ç. ÇÖLTEKIN, M. CONNOR, E. DAVIDSON, M. D. MARNEFFE, V. D. PAIVA, A. D. D. ILARRAZA, P. DIRIX, K. DOBROVOLJC, T. DOZAT, K. DROGANOVA, P. DWIVEDI, M. ELI, A. ELKAHKY, T. ERJAVEC, R. FARKAS, H. FERNANDEZ ALCALDE, J. FOSTER, C. FREITAS, K. GAJDOŠOVÁ, D. GALBRAITH, M. GARCIA, M. GÄRDENFORS, K. GERDES, F. GINTER, I. GOENAGA, K. GOJENOLA, M. GÖKIRMAK, Y. GOLDBERG, X. GÓMEZ GUINOVART, B. GONZÁLES SAAVEDRA, M. GRIONI, N. GRŪZĪTIS, B. GUILLAUME, N. HABASH, J. HAJIČ, J. HAJIČ JR., L. HÀ MỸ, K. HARRIS, D. HAUG, B. HLADKÁ, J. HLAVÁČOVÁ, F. HOCIUNG, P. HOHLE, R. ION, E. IRIMIA, T. JELÍNEK, A. JOHANNSEN, F. JØRGENSEN, H. KAŞIKARA, H. KANAYAMA, J. KANERVA, T. KAYADELEN, V. KETTNEROVÁ, J. KIRCHNER, N. KOTSYBA, S. KREK, V. LAIPPALA,

L. LAMBERTINO, T. LANDO, J. LEE, P. LÊ HỒNG, A. LENCI, S. LERTPRADIT, H. LEUNG, C. Y. LI, J. LI, K. LI, N. LJUBEŠIĆ, O. LOGINOVA, O. LYASHEVSKAYA, T. LYNN, V. MACKETANZ, A. MAKAZHANOV, M. MANDL, C. MANNING, C. MĂRĂNDUC, D. MAREČEK, K. MARHEINECKE, H. MARTÍNEZ ALONSO, A. MARTINS, J. MAŠEK, Y. MATSUMOTO, R. MCDONALD, G. MENDONÇA, N. MIEKKA, A. MISSILÄ, C. MITITELU, Y. MIYAO, S. MONTEMAGNI, A. MORE, L. MORENO ROMERO, S. MORI, B. MOSKALEVSKYI, K. MUISCHNEK, K. MÜÜRISEP, P. NAINWANI, A. NEDOLUZHKO, G. NEŠPORE-BĒRZKALNE, L. NGUYỄN THỊ, H. NGUYỄN THỊ MINH, V. NIKOLAEV, H. NURMI, S. OJALA, P. OSENOVA, R. ÖSTLING, L. ØVRELID, E. PASCUAL, M. PASSAROTTI, C. PEREZ, G. PERRIER, S. PETROV, J. PIITULAINEN, E. PITLER, B. PLANK, M. POPEL, L. PRETKALNIŅA, P. PROKOPIDIS, T. PUOLAKAINEN, S. PYYSALO, A. RADEMAKER, L. RAMASAMY, T. RAMA, V. RAVISHANKAR, L. REAL, S. REDDY, G. REHM, L. RINALDI, L. RITUMA, M. ROMANENKO, R. ROSA, D. ROVATI, B. SAGOT, S. SALEH, T. SAMARDŽIĆ, M. SANGUINETTI, B. SAULĪTE, S. SCHUSTER, D. SEDDAH, W. SEEKER, M. SERAJI, M. SHEN, A. SHIMADA, D. SICHINAVA, N. SILVEIRA, M. SIMI, R. SIMIONESCU, K. SIMKÓ, M. ŠIMKOVÁ, K. SIMOV, A. SMITH, A. STELLA, M. STRAKA, J. STRNADOVÁ, A. SUHR, U. SULUBACAK, Z. SZÁNTÓ, D. TAJI, T. TANAKA, T. TROSTERUD, A. TRUKHINA, R. TSARFATY, F. TYERS, S. UEMATSU, Z. UREŠOVÁ, L. URIA, H. USZKOREIT, S. VAJJALA, D. V. NIEKERK, G. V. NOORD, V. VARGA, E. V. D. L. CLERGERIE, V. VINCZE, L. WALLIN, J. N. WASHINGTON, M. WIRÉN, T. WONG, Z. YU, Z. ŽABOKRTSKÝ, A. ZELDES, D. ZEMAN, H. ZHU.*Universal Dependencies 2.1*, November 2017, LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University - Corpus - Project code: 15-10472S; Project name: Morphologically and Syntactically Annotated Corpora of Many Languages, https://hal.inria.fr/hal-01682188.

[45] M. PUREN, C. RIONDET, D. SEILLIER, L. ROMARY.*The Standardization Survival Kit (SSK): For a wider use of standards within Arts and Humanities*, July 2017, Digital Humanities Benelux Conference 2017, Poster, https://hal.archives-ouvertes.fr/hal-01587687.

[46] L. ROMARY, J. EDMOND.*Sustainability in DARIAH*, April 2017, 10, Sustainability of Digital Research Infrastructures for the Arts and Humanities, https://hal.inria.fr/hal-01516487.

[47] L. ROMARY, C. RIONDET.*Ongoing maintenance and customization of archival standards using ODD (EAC-CPF revision proposal)*, December 2017, EAC-CPF revision proposal, https://hal.inria.fr/hal-01677185.

[48] L. ROMARY, C. RIONDET.*Towards multiscale archival digital data*, September 2017, working paper or preprint, https://hal.inria.fr/hal-01586389.

[49] L. ROMARY.*How to Open up? (Digital) Libraries at the Service of (Digital) Scholars*, April 2017, Fiesole Collection Development Retreat, https://hal.inria.fr/hal-01513674.

[50] V. VANDEN DAELEN, J. EDMOND, P. LINKS, M. PRIDDY, L. REIJNHOUDT, V. TOLLAR, A. VAN NISPEN, C. HAUWAERT, C. RIONDET.*La publication durable digitale des guides d'archives de l'histoire du 20ème siècle*, November 2017, working paper or preprint, https://hal.inria.fr/hal-01632366.

References in notes

[51] 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.

[52] 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.

[53] 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^{o} 110, p. 4224–4229.

[54] 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.

[55] 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.

[56] P. DENIS, B. SAGOT.*Coupling an annotated corpus and a lexicon for state-of-the-art POS tagging*, in "Language Resources and Evaluation", 2012, vol. 46, n^{o} 4, p. 721-736 [*DOI :* 10.1007/s10579-012-9193-0], https://hal.inria.fr/inria-00614819.

[57] 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.

[58] 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.

[59] 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.

[60] 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.

[61] H. A. MARTÍNEZ, 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.

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

[63] 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.

[64] 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.

[65] 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.

[66] L. ROMARY, E. DEGL'INNOCENTI, K. ILLMAYER, A. JOFFRES, E. KRAIKAMP, N. LARROUSSE, M. OGRODNICZUK, M. PUREN, C. RIONDET, D. SEILLIER.*Standardization survival kit (Draft)*, Inria, October 2016, n° Deliverable 4.1, https://hal.inria.fr/hal-01513531.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

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

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

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] 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.

[78] 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.

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

[80] G. WALTHER, G. JACQUES, B. SAGOT.*Uncovering the inner architecture of Khaling verbal morphology*, in "3rd Workshop on Sino-Tibetan Languages of Sichuan", Paris, France, September 2013, https://hal.inria.fr/hal-00927278.

[81] G. WALTHER, G. JACQUES, B. SAGOT.*The Opacity-Compactness Tradeoff: Morphomic Features for an Economical Account of Khaling Verbal Inflection*, in "16th International Morphology Meeting (IMM 16)", Budapest, Hungary, May 2014, https://hal.inria.fr/hal-01114854.

Project-Team ALPINES

Algorithms and parallel tools for integrated numerical simulations

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

IN PARTNERSHIP WITH:
CNRS

Université Pierre et Marie Curie (Paris 6)

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
B9.4.2. - Mathematics
B9.4.3. - Physics

1. Personnel

Research Scientists

Laura Grigori [Team leader, Inria, Senior Researcher, HDR]
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]
Franck Houssen [Inria, from Apr 2017]
Ange Toulougoussou [Inria, until Feb 2017]
Pierre-Henri Tournier [CNRS]

PhD Students

Hussam Al Daas [Inria]
Alan Ayala Obregon [Inria]
Sebastien Cayrols [Inria]
Igor Chollet [Univ Pierre et Marie Curie, from Oct 2017]
Zakariae Jorti [IFPEN]
Pierre Marchand [Inria]
Van Thanh Nguyen [Inria, from Nov 2017]
Olivier Tissot [Inria]

Post-Doctoral Fellows

Jan Papez [Inria, from Mar 2017]
Amin Rafiei [Ambassade de France, from Nov 2017]

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, which includes devising adapted domain decomposition approaches. The final objective is to bring the state of the art on boundary integral equations closer to contemporary industrial needs.

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

The research of F. Nataf on inverse problems is rather new since this activity was started from scratch in 2007. Since then, several papers were published in international journals and conference proceedings. All our numerical simulations were performed in FreeFem++.
We focus on methods related to time reversal techniques. 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 main idea is to take advantage of the reversibility of wave propagation phenomena such as it occurs in acoustics, elasticity or electromagnetism in a non-dissipative unknown medium to back-propagate signals to the sources that emitted them. Number of industrial applications have already been developped: touchscreen, medical imaging, non-destructive testing and underwater communications. 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. Since then, numerous applications of this physical principle have been designed, see [Fink, Renversement du temps, ondes et innovation. Ed. Fayard, 2009] or for numerical experiments [Larmat et al., Time-reversal imaging of seismic sources and application to the great sumatra earthquake. Geophys. Res. Lett., 33, 2006] and references therein.

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 subregion 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

5.1.1. Awards and Recognitions

5.1.1.1. Laura Grigori elected Member of the SIAM Council

January 2018 - December 2020.

6. New Software and Platforms

6.1. FreeFem++

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

7. New Results

7.1. Communication avoiding algorithms for preconditioned iterative methods

Our group continues to work on algorithms for dense and sparse linear algebra operations that minimize communication, introduced in [1], [4]. An overview of communication avoiding algorithms for dense linear algebra operations is presented in [18]. During this year we focused on communication avoiding iterative methods and designing algorithms for computing rank revealing and low rank approximations of dense and sparse matrices.

Iterative methods are widely used in industrial applications, and in the context of communication avoiding algorithms, our research is related to increasing the scalability of Krylov subspace iterative methods. Indeed the dot products related to the orthogonalization of the Krylov subspace and performed at each iteration of the Krylov method require collective communication among all processors. This collective communication does not scale to very large number of processors, and thus is a main bottleneck in the scalability of Krylov subspace methods. Our research focuses on enlarged Krylov subspace methods, a new approach that we have introduced in the recent years [5] that consists of enlarging the Krylov subspace by a maximum of t vectors per iteration, based on a domain decomposition of the graph of the input matrix. The solution of the linear system is searched in the enlarged subspace, which is a superset of the classic subspace. The enlarged Krylov projection subspace methods lead to faster convergence in terms of iterations and parallelizable algorithms with less communication, with respect to Krylov methods.

In [20] we propose an algebraic method in order to reduce dynamically the number of search directions during block Conjugate Gradient iterations. Indeed, by monitoring the rank of the optimal step α_k it is possible to detect inexact breakdowns and remove the corresponding search directions. We also propose an algebraic criterion that ensures in theory the equivalence between our method with dynamic reduction of the search directions and the classical block Conjugate Gradient. Numerical experiments show that the method is both stable, the number of iterations with or without reduction is of the same order, and effective, the search space is significantly reduced. We use this approach in the context of enlarged Krylov subspace methods which reduce communication when implemented on large scale machines. The reduction of the number of search directions further reduces the computation cost and the memory usage of those methods.

In [19] 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.

7.2. Communication avoiding algorithms for low rank matrix approximation

Our work focuses on computing the low rank approximation of a sparse or dense matrix, while also minimizing communication, [3].

In [21] we introduce an URV Factorization with Random Orthogonal System Mixing. The unpivoted and pivoted Householder QR factorizations are ubiquitous in numerical linear algebra. A difficulty with pivoted Householder QR is the communication bottleneck introduced by pivoting. In this paper we propose using random orthogonal systems to quickly mix together the columns of a matrix before computing an unpivoted QR factorization. This method computes a URV factorization which forgoes expensive pivoted QR steps in exchange for mixing in advance, followed by a cheaper, unpivoted QR factorization. The mixing step typically reduces the variability of the column norms, and in certain experiments allows us to compute an accurate factorization where a plain, unpivoted QR performs poorly. We experiment with linear least-squares, rank-revealing factorizations, and the QLP approximation, and conclude that our randomized URV factorization behaves comparably to a similar randomized rank-revealing URV factorization, but at a fraction of the computational cost. Our experiments provide evidence that our proposed factorization might be rank-revealing with high probability.

7.3. Domain decomposition preconditioning for high frequency wave propagation problems

This work studies preconditioning the Helmholtz and Maxwell equations, where the preconditioner is constructed using two-level overlapping Additive Schwarz Domain Decomposition. The coarse space is based on the discretisation of the PDE on a coarse mesh. The PDE is discretised using finite-element methods of fixed, arbitrary order. The theoretical part of this work is the Maxwell analogue of a previous work for Helmholtz equation, and shows that for Maxwell problems with absorption, if the absorption is large enough and if the subdomain and coarse mesh diameters are chosen appropriately, then classical two-level overlapping Additive Schwarz Domain Decomposition preconditioning performs optimally – in the sense that GMRES converges in a wavenumber-independent number of iterations. An important feature of the theory is that it allows the coarse space to be built from low-order elements even if the PDE is discretised using high-order elements. This theory is presented in [24] and is illustrated by numerical experiments, which also (i) explore replacing the PEC boundary conditions on the subdomains by impedance boundary conditions, and (ii) show that the preconditioner for the problem with absorption is also an effective preconditioner for the problem with no absorption. The numerical results include two substantial examples arising from applications; the first (a problem arising in medical imaging from the Medimax ANR project) shows the robustness of the preconditioner against heterogeneity, and the second (scattering by a COBRA cavity) shows good scalability of the preconditioner with up to 3000 processors. The parallel implementation was done using FreeFem++ and HPDDM. We performed additional numerical studies of this two-level Domain Decomposition preconditioner for the Maxwell equations in [23], and for the Helmholtz equation (in 2D and 3D) in [25], where we also compare it to another two-level Domain Decomposition preconditioner where the coarse space is built by solving local eigenproblems on the interface between subdomains involving the Dirichlet-to-Neumann (DtN) operator.

7.4. First kind boundary integral formulation for the Hodge-Helmholtz equation

We adapt the variational approach to the analysis of first-kind boundary integral equations associated with strongly elliptic partial differential operators from [M. COSTABEL, *Boundary integral operators on Lipschitz domains: Elementary results*, SIAM J. Math. Anal., 19 (1988), pp. 613–626.] to the (scaled) Hodge-Helmholtz equation $\mathbf{curl}\,\mathbf{curl}\,\mathbf{u} - \eta\nabla\mathrm{div}\,\mathbf{u} - \kappa^2\mathbf{u} = 0, \eta > 0, \mathrm{Im}\,\kappa^2 \geq 0$, on Lipschitz domains in 3D Euclidean space, supplemented with natural complementary boundary conditions, which, however, fail to bring about strong ellipticity.

Nevertheless, a boundary integral representation formula can be found, from which we can derive boundary integral operators. They induce bounded and coercive sesqui-linear forms in the natural energy trace spaces for the Hodge-Helmholtz equation. We can establish precise conditions on η, κ that guarantee unique solvability of the two first-kind boundary integral equations associated with the natural boundary value problems for the Hodge-Helmholtz equations. Particular attention needs to be given to the case $\kappa = 0$.

7.5. Integral equation based optimized Schwarz method for electromagnetics

The optimized Schwarz method (OSM) is recognised as one of the most efficient domain decomposition strategies without overlap for the solution to wave propagation problems in harmonic regime. For the Helmholtz equation, this approach originated from the seminal work of Després, and led to the development of an abundent literature offering more elaborated but more efficient transmission conditions. Most contributions focus on transmission conditions based on local operators.

In recent years, F. Collino, P. Joly and M. Lecouvez introduced non-local transmission conditions that can drastically improve the convergence rate of OSM. The performance of this strategy seems to remain robust at high frequency. Such an approach was proposed only for the Helmholtz equation, and has still not been adapted to electromagnetics.

In this work we investigated such an approach for Maxwell's equations in a simple spherical geometry that allows explicit calculus by means of separation of variables. The transmission condition that we propose involves a non-local operator that is a dissipative counterpart of the so-called Electric Field integral operator (EFIE) which is a classical object in electromagnetic potential theory. We show that the iterative solver associated to our strategy converges at an exponential rate.

7.6. Quasi-local Multi-Trace formulations for electromagnetics

Multi-trace formulations (MTF) are a general methodology to derive first kind boundary integral formulations for harmonic wave scattering problems posed in multi-domain geometrical confgurations. There exists both a local and a global variant of MTF that only differ through the way transmission conditions are imposed across interfaces. Global MTF is easier to analyse but, from a computational viewpoint, local MTF appears more appealing because it looks computationally cheaper.

As regards local MTF, a decent stability theory has been developped for acoustic scalar wave propagation, but no such result as Garding inequality or uniform discrete inf-sup condition has been established so far for local MTF in the case of electromagnetics. Wether or not local MTF is stable for electromagnetics is actually an open question presently.

In this work, we have adopted a slightly modified version of local MTF where tranmission conditions are imposed by means of an operator that is non-local, but with a kernel whose support can be as small as desired. This so-called quasi-local MTF approach has previsouly been developped for acoustics and we adapted it to the case of electromagnetics. We could in particular prove a Garding inequality for quasi-local MTF applied to electromagnetics, and thus obtain uniform discrete inf-sup condition.

7.7. Domain decomposition preconditioning with approximate coarse solve

Convergence of domain decomposition methods relies heavily on the efficiency of the coarse space used in the second level. The GenEO coarse space has been shown to lead to a fully robust two-level Schwarz preconditioner which scales well over multiple cores [9], [2] as has been proved rigorously in [9]. The robustness is due to its good approximation properties for problems with highly heterogeneous material parameters. It is available in the finite element packages FreeFem++ [7], Feel++ [31] and recently in Dune [30] and is implemented as a standalone library in HPDDM [8]. But the coarse component of the preconditioner can ultimately become a bottleneck if the number of subdomains is very large and exact solves are used. It is therefore interesting to consider the effect of approximate coarse solves. In [28], robustness of GenEO methods is analyzed with respect to approximate coarse solves. Interestingly, the GenEO-2 method introduced in [6] has to be modified in order to be able to prove its robustness in this context.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- Contract with Total, February 2015 - February 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 - February 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.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

9.1.1.1. 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.

9.1.1.2. Medimax

ANR-MN (Modèles Numériques) October 2013 - September 2017

The main goal is the methodological and numerical development of a new robust inversion tool, associated with the numerical solution of the electromagnetic forward problem, including the benchmarking of different other existing approaches (Time Reverse Absorbing Condition, Method of Small-Volume Expansions, Level Set Method). This project involves the development of a general parallel open source simulation code, based on the high-level integrated development environment of FreeFem++, for modeling an electromagnetic direct problem, the scattering of arbitrary electromagnetic waves in highly heterogeneous media, over a wide frequency range in the microwave domain. The first applications considered here will be medical applications: microwave tomographic images of brain stroke, brain injuries, from both synthetic and experimental data in collaboration with EMTensor GmbH, Vienna (Austria), an Electromagnetic Medical Imaging company.

9.1.1.3. 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.

9.1.1.4. 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.

9.1.1.5. Soilμ-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 (μm) 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.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. NLAFET

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

Programm: H2020

Duration: November 2015 - November 2018

Coordinator: UMEÅUniversitet

Partners:

Science and Technology Facilities Council (United Kingdom)

Computer Science Department, UmeåUniversitet (Sweden)

Mathematics Department, The University of Manchester (United Kingdom)

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.

9.2.1.2. EXA2CT

Title: EXascale Algorithms and Advanced Computational Techniques

Programm: FP7

Duration: September 2013 - August 2016

Coordinator: IMEC

Partners:

Fraunhofer-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V (Germany)

Interuniversitair Micro-Electronica Centrum Vzw (Belgium)

Intel Corporations (France)

Numerical Algorithms Group Ltd (United Kingdom)

T-Systems Solutions for Research (Germany)

Universiteit Antwerpen (Belgium)

Universita della Svizzera italiana (Switzerland)

Université de Versailles Saint-Quentin-En-Yvelines. (France)

Vysoka Skola Banska - Technicka Univerzita Ostrava (Czech Republic)

Inria contact: Luc Giraud

Numerical simulation is a crucial part of science and industry in Europe. The advancement of simulation as a discipline relies on increasingly computing intensive models that require more computational resources to run. This is the driver for the evolution to exascale. Due to limits in the increase in single processor performance, exascale machines will rely on massive parallelism on and off chip, with a complex hierarchy of resources. The large number of components and the machine complexity introduce severe problems for reliability and programmability. The former of these will require novel fault-aware algorithms and support software. In addition, the scale of the numerical models exacerbates the difficulties by making the use of more complex simulation algorithms necessary, for numerical stability reasons. A key example of this is increased reliance on solvers. Such solvers require global communication, which impacts scalability, and are often used with preconditioners, increasing complexity again. Unless there is a major rethink of the design of solver algorithms, their components and software structure, a large class of important numerical simulations will not scale beyond petascale. This in turn will hold back the development of European science and industry which will fail to reap the benefits from exascale. The EXA2CT project brings together experts at the cutting edge of the development of solvers, related algorithmic techniques, and HPC software architects for programming models and communication. It will take a revolutionary approach to exascale solvers and programming models, rather than the incremental approach of other projects. We will produce modular open source proto-applications that demonstrate the algorithms and programming techniques developed in the project, to help boot-strap the creation of genuine exascale codes.

9.3. International Initiatives

9.3.1. Inria International Partners

9.3.1.1. Informal International Partners

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

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Ralf Hiptmair (ETH Zürich) came to visit Xavier Claeys for a sabbatical semester, from January to June 2017.
- Mahadevan Ganesh (Colorado School of Mines) came to visit Xavier Claeys from the 4th of July 2017 to 18th of July 2017.
- Carlos Jerez-Hanckes (Pontificia Universidad Catholica, Santiago, Chile) came to visit Xavier Claeys from the 3rd of December to the 16th of December 2017.

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

- Laura Grigori has spent 3 weeks at UC Berkeley, from July 21, 2016 to August 13, 2016.
- Xavier Claeys visited Catalin Turc (New Jersey Institute of Technology) from the 5th of November to the 14th of November 2017.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Frederic Hecht: Organized the 9th FreeFem++ days (December 2017, Paris)
- Xavier Claeys was main organiser (together with S. Fliss, B. Delourme and J. Diaz) of the 3 days colloquium "Waves Diffracted by Patrick Joly", in honour of P. Joly's 60th birthday, that took place in Saclay and Gif-sur-Yvette.
- Xavier Claeys and Frédéric Nataf were the main organisers (together with V. Dolean) of a workshop on "Numerical methods for wave propagation and applications" held in Jussieu campus of UPMC on 31st of August and 1st of September.

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

10.1.2.1.1. 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/.
- June 2013 – current. Area editor for Parallel Computing Journal, Elsevier
- January 2016 – current. Associate Editor, SIAM Journal on Scientific Computing.
- January 2017 – current. Associate Editor, SIAM Journal on Matrix Analysis and Applications.
- January 2016 – current. Editorial board, Numerical linear algebra with applications Journal, Wiley.

10.1.2.1.2. Frédéric Nataf

- 2014 – current. Associate Editor, Journal of Numerical Mathematics, de Gruyter.

10.1.3. Invited Talks

- Laura Grigori: Keynote speaker, 46th International Conference on Parallel Processing (ICPP-2017), http://www.icpp-conf.org/2017/index.php, Bristol, UK, August 14-17, 2017.
- Laura Grigori: Keynote speaker, Power-Aware Computing Workshop - PACO2017, https://www2.mpi-magdeburg.mpg.de/mpcsc/events/paco2017/index.php, July 2017, Germany.
- Laura Grigori: Plenary talk, Householder Symposium XX on Numerical linear algebra, http://www.math.vt.edu/HHXX/, June 2017, Virginia Tech.

- Laura Grigori: Invited plenary speaker, International Conference on Domain Decomposition Methods DD XXIV, http://www.ddm.org/dd24/invited.html, February 2017, Longyearbyen, Norway.

10.1.4. Leadership within the Scientific Community

- Laura Grigori: Chair of the SIAM SIAG on Supercomputing (SIAM special interest group on supercomputing), January 2016 - December 2017. Nominated by a Committee and elected by the members of this SIAG.
- Laura Grigori: Member of the PRACE (Partnership for Advanced Computing in Europe, http://www.prace-ri.eu/) Scientific Steering Committee, September 2016 - current.
- Laura Grigori: Steering committee member, Challenge 7: Information and communication society, ANR (Comité de Pilotage , Défi 7), November 2016 - September 2017.

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.
- Xavier Claeys was member of a hiring committee for a position of maître de conférence in section CNU 60/61 for the Institut de Recherche de Coordination Acoustique et Musicale (IRCAM) and Université Pierre-et-Marie Curie Paris 6, in Spring 2017.

10.1.6. Research Administration

- Laura Grigori: Member of the Director Committee (Comité Directeur) of GIS Geosciences francilienes, since November 2015.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

10.2.1.1. Xavier Claeys

- Academic year 2017-2018, total number of course hours: 192 hrs
  - Master 1: supervisoon of a student project for a group of 4 students in the curriculum Polytech, 80hrs, UPMC.
  - Master 1: Initiation to C++, 36 hrs of programming tutorials in C++, UPMC.
  - Master 1: Computational Linear Algebra, 32 hrs of lectures, UPMC.
  - Master 1: Practical programming of the finite element method, 12 hrs of lectures, UPMC.
  - Master 1: Approximation of EDPs, 24 hrs of programming tutorials in Python, UPMC.

10.2.1.2. Laura Grigori

- Spring 2017, 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/Spr2017.html, Master 2nd year, Mathematics & Applications, University Pierre and Marie Curie, 20 hours per year.
- Winter 2017, Participation in the course on High Performance Computing given at University Pierre and Marie Curie, Master 2nd year, Computer Science, intervention for 8 hours per year.

10.2.1.3. Frédéric Hecht

- Academic year 2016-2017, total number of course hours: 192 hrs
  - Master 1: Initiation au C++, 24hrs, M1, Université Pierre-et-Marie Curie Paris 6, France
  - Master 2: Des EDP à leur résolution par la méthode des éléments finis (MEF), 24hrs, M2, Université Pierre-et-Marie Curie Paris 6, France

- Master 2: Numerical methods for fluid mechanics, 10hrs, M2, Université Pierre-et-Marie Curie Paris 6, France
- Master 2: Calcul scientifique 3 / projet industriel FreeFem++, 28hrs, M2, Université Pierre-et-Marie Curie Paris 6, France
- Master 2: Ingénierie 1 / Logiciel pour la simulation (FreeFem++), 21hrs, M2, Université Pierre-et-Marie Curie Paris 6, France
- Master 2: Ingénierie 2 / Projet collaboratif, 21hrs, M2, Université Pierre-et-Marie Curie Paris 6, France

10.2.1.4. Frédéric Nataf

- Spring 2017: Course on *Domain Decomposition Methods*, Master 2nd year Mathematics & Applications, University Pierre and Marie Curie
- Winter 2017: Course on *Domain Decomposition Methods*, Master 2nd year, Mathematics & Applications, ENSTA and UVSQ

10.2.2. Supervision

- PhD in progress: Alan Ayala, since October 2015 (funded by NLAFET H2020 project), co-advisors Xavier Claeys and Laura Grigori.
- PhD in progress : Sebastien Cayrols, since October 2013 (funded by Maison de la simulation), adivsor Laura Grigori.
- PhD in progress: Hussam Al Daas, since February 2015 (funded by contract with Total), advisor 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 et 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

- Frédéric Hecht, President of the Jury for the PhD thesis of Alexandre LIMAR, Applied Mathematics, UTT, October 2017.
- Laura Grigori, President of the Jury for the PhD thesis of Chaoyu Quan, Applied Mathematics, UPMC, November 2017.
- Frédéric Hecht, President of the Jury for the PhD thesis of Roberto Molina, Applied Mathematics, UPMC, December 2017.

11. Bibliography

Major publications by the team in recent years

[1] 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.

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

[3] 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", 2018, In press.

[4] 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.

[5] 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.

[6] 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.

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

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

[9] 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

Articles in International Peer-Reviewed Journal

[10] 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", 2017 [*DOI :* 10.1002/JNM.2229], https://hal.archives-ouvertes.fr/hal-01328197.

[11] L. A. CHESNEL, X. CLAEYS, S. A. NAZAROV.*Small obstacle asymptotics for a 2D semi-linear convex problem*, in "Applicable Analysis", February 2017, 20 [*DOI :* 10.1080/00036811.2017.1295449], https://hal.archives-ouvertes.fr/hal-01427617.

[12] X. CLAEYS, R. HIPTMAIR, E. SPINDLER.*Second-kind boundary integral equations for electromagnetic scattering at composite objects*, in "Computers & Mathematics with Applications", December 2017, vol. 74, n^{o} 11, p. 2650 - 2670 [*DOI :* 10.1016/J.CAMWA.2017.08.014], https://hal.inria.fr/hal-01671147.

[13] X. E. CLAEYS, R. E. HIPTMAIR, E. E. SPINDLER.*Second Kind Boundary Integral Equation for Multi-Subdomain Diffusion Problems*, in "Advances in Computational Mathematics", October 2017, vol. 43, n^{o} 5, 26, https://hal.archives-ouvertes.fr/hal-01427625.

[14] 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, https://hal.archives-ouvertes.fr/hal-01216117.

[15] 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 Journal on Scientific Computing", February 2017, vol. 39, n^{o} 4, p. A1345 - A1365 [*DOI :* 10.1137/16M1060066], https://hal.archives-ouvertes.fr/hal-01278347.

[16] P.-H. TOURNIER, M. BONAZZOLI, V. DOLEAN, F. RAPETTI, F. HECHT, F. NATAF, I. ALIFERIS, I. EL KANFOUD, C. MIGLIACCIO, M. DE BUHAN, M. DARBAS, S. SEMENOV, C. PICHOT.*Numerical Modeling and High-Speed Parallel Computing: New Perspectives on Tomographic Microwave Imaging for Brain Stroke Detection and Monitoring*, in "IEEE Antennas and Propagation Magazine", October 2017, vol. 59, n^{o} 5, p. 98 - 110 [*DOI :* 10.1109/MAP.2017.2731199], https://hal.archives-ouvertes.fr/hal-01623106.

[17] 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", November 2017 [*DOI :* 10.1007/s00211-017-0925-3], https://hal.inria.fr/hal-01097662.

Scientific Books (or Scientific Book chapters)

[18] L. GRIGORI.*Introduction to Communication Avoiding Algorithms for Direct Methods of Factorization in Linear Algebra*, in "SEMA SIMAI Springer Series", M. MATEOS, P. ALONSO (editors), Computational Mathematics, Numerical Analysis and Applications, Springer, 2017, vol. 13, p. 153-185 [*DOI :* 10.1007/978-3-319-49631-3_4], https://hal.inria.fr/hal-01669315.

Research Reports

[19] H. AL DAAS, L. GRIGORI, P. HÉNON, P. RICOUX.*Enlarged GMRES for reducing communication*, Inria Paris, March 2017, n^{o} RR-9049, https://hal.inria.fr/hal-01497943.

[20] L. GRIGORI, O. TISSOT.*Reducing the communication and computational costs of Enlarged Krylov subspaces Conjugate Gradient*, Inria Paris, February 2017, n^{o} RR-9023, https://hal.inria.fr/hal-01451199.

Other Publications

[21] S. BECKER, J. FOLBERTH, L. GRIGORI.*URV Factorization with Random Orthogonal System Mixing*, 2017, working paper or preprint, https://hal.inria.fr/hal-01669915.

[22] I. BEN GHARBIA, M. BONAZZOLI, X. CLAEYS, P. MARCHAND, P.-H. TOURNIER.*Fast solution of boundary integral equations for elasticity around a crack network: a comparative study*, November 2017, working paper or preprint, https://hal.inria.fr/hal-01644518.

[23] 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*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01525438.

[24] M. BONAZZOLI, V. DOLEAN, I. G. GRAHAM, E. A. SPENCE, P.-H. TOURNIER.*Domain decomposition preconditioning for the high-frequency time-harmonic Maxwell equations with absorption*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01644011.

[25] M. BONAZZOLI, V. DOLEAN, I. G. GRAHAM, E. A. SPENCE, P.-H. TOURNIER.*Two-level preconditioners for the Helmholtz equation*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01525424.

[26] 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*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01298938.

[27] M. †. HADDAD † ‡, T. SAYAH, P.-H. TOURNIER, F. HECHT.*Parallel Computing Investigations for the Projection Method Applied to the Interface Transport Scheme of a Two-phase Flow by the Method of Characteristics*, February 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01470551.

[28] F. NATAF.*Mathematical Analysis of Robustness of Two-Level Domain Decomposition Methods with respect to Approximate Coarse Solves*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01573197.

[29] J. PAPEŽ, U. RÜDE, M. VOHRALÍK, B. WOHLMUTH.*Sharp algebraic and total a posteriori error bounds for h and p finite elements via a multilevel approach*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01662944.

References in notes

[30] M. BLATT, A. BURCHARDT, A. DEDNER, C. ENGWER, J. FAHLKE, B. FLEMISCH, C. GERSBACHER, C. GRÄSER, F. GRUBER, C. GRÜNINGER.*The distributed and unified numerics environment, version 2.4*, in "Archive of Numerical Software", 2016, vol. 4, n^{o} 100, p. 13–29.

[31] C. PRUD'HOMME.*A Domain Specific Embedded Language in C++ for automatic differentiation, projection, integration and variational formulations*, in "Scientific Programming", 2006, vol. 14, n^{o} 2, p. 81–110.

Project-Team ANGE

Numerical Analysis, Geophysics and Ecology

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

Université Pierre et Marie Curie (Paris 6)

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. - 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. - Littoral
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. Personnel

Research Scientists

Jacques Sainte-Marie [Team leader, CEREMA, Senior Researcher, HDR]
Vivien Mallet [Inria, Researcher, from Mar 2017]
Martin Parisot [Inria, Researcher]
Yohan Penel [CEREMA, Researcher]
Julien Salomon [Inria, Senior Researcher, from Oct 2017, HDR]

Faculty Members

Edwige Godlewski [Univ Pierre et Marie Curie, Professor, HDR]
Cindy Guichard [Univ Pierre et Marie Curie, Associate Professor]
Boris Haspot [Univ Dauphine Paris, Associate Professor, HDR]
Anne Mangeney [IPGP, Professor, HDR]

Julien Salomon [Univ Dauphine Paris, Associate Professor, until Sep 2017, HDR]

External Collaborators

Nina Aguillon [UPMC, from Jan 2017]
Nora Aïssiouene [UPMC, from Feb 2017]
Emmanuel Audusse [Univ Paris-Nord]
Bernard Di Martino [Univ Corse, HDR]
Nicole Goutal [EDF, HDR]

Technical Staff

Marius Guerard [Inria, from Mar 2017 until Aug 2017]
Janelle Hammond [Inria, from Nov 2017]
Fabien Souillé [Inria]
Jérémy Ledoux [Univ. Paris Dauphine]

PhD Students

Frédéric Allaire [Inria, from Nov 2017]
Léa Boittin [Inria]
Do Minh Hieu [Univ. Paris 13]
Virgile Dubos [Univ Pierre et Marie Curie, from Oct 2017]
Ngoc Bao Tran Le [IRSN]
Antoine Lesieur [Inria, from Oct 2017]
Hugo Martin [IPGP]
Ethem Nayir [Univ Pierre et Marie Curie, until Sep 2017]
Sebastian Reyes-Riffo [Univ. Paris Dauphine]
Jean Thorey [EDF, from Mar 2017 until Apr 2017]
Raphaël Ventura [Inria, from Mar 2017]
Fabien Wahl [Univ Pierre et Marie Curie]

Post-Doctoral Fellows

Bilal Al Taki [Inria, from Sep 2017]
Pierre-Olivier Lamare [Inria]

Visiting Scientist

Marie-Odile Bristeau [Retired]

Administrative Assistant

Maryse Desnous [Inria]

Other

Jim Pioche [SciWorks Technologies, from Sep 2017]

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. Also, efficient numerical resolutions of reduced-size models requires 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 [34], [36], [37]. 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 [1], [2] 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

5.1.1. Human resources

A major event in the year was the merging with CLIME which induces the incorporation of several new researchers (1 Researcher, 1 engineer, 2 PhD). CLIME research is naturally complementary to ANGE works insofar as it provides high level tools to improve modelling and numerical results.

Another fact is J. Salomon's arrival as a Senior Researcher.

5.1.2. Scientific activities

There has been major achievements within the team in the framework of dispersive models. An increased research activity is carried out with spanish collaborators (Univ. Sevilla, Córdoba and Málaga) supported by several project call fundings. This lead to a main publication [30]. In the aftermath of N. Aïssiouene's PhD thesis, a new PhD has been hired to go further in the design of robust and efficient numerical algorithms.

As detailed in Section 10.1.1.1, members of the team were involved in the organisation of a substantial number of scientific events, either in the framework of national initiatives (mainly funded by CNRS) or due to the expertise in the field. Members are is particularly involved in the mathematical community.

5.1.3. Awards

L. Boittin and F. Wahl were granted a SIAM Student Travel Award to attend SIAM GS 2017. F. Wahl also received a Young Researcher Scholarship to attend the 2017 SMAI conference.

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

7. New Results

7.1. Modelling of complex flows

7.1.1. Modelling and simulation of sediment transport

Participants: Emmanuel Audusse, Léa Boittin, Martin Parisot, Jacques Sainte-Marie.

Following previous works, a numerical scheme for the sediment layer is proposed and assessed. 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.

7.1.2. Modelling of photosynthesis through microalgae cultivation

Participants: Marie-Odile Bristeau, Jacques Sainte-Marie.

In collaboration with O. Bernard.

In the present multidisciplinary downscaling study, we reconstruct single cell trajectories in an open raceway and experimentally reproduce the according high frequency light pattern to observe its effect on the growth of Dunaliella salina. We show that the frequency of such a realistic signal plays a decisive role on the dynamics of photosynthesis, which reveal an unexpected photosynthetic response compared to that recorded under the on/off signals usually used in the literature. This study highlights the need for experiments with more realistic light stimuli in order to better understand microalgal growth at high cell density. We also propose an experimental protocol with simple piecewise constant, yet more realistic, light fluctuations.

7.1.3. Buoyancy modelling

Participants: Edwige Godlewski, Martin Parisot, Jacques Sainte-Marie, Fabien Wahl.

Firstly, the work of the previous year was completed and lead to the submission of an article [38]. More precisely the fixed point algorithm is rewritten using a new unknown. This allows to increase the numerical robustness and accuracy of the scheme. The proposed resolution is assessed on several stationary and non-stationary test cases with analytical solutions.

In the continuity of this work, the modelling of fluid-structure interaction resolution is added in the previous work in order to simulate floating structures for marine energy devices. In a first step only the vertical movement is studied, with no major scientific lock. In a second time the horizontal movement of the structure is considered and required a deeper analysis to ensure the entropy-stability at the discrete level.

7.1.4. A Free Interface Model for Static/Flowing Dynamics in Thin-Layer Flows of Granular Materials with Yield: Simple Shear Simulations and Comparison with Experiments

Participant: Anne Mangeney.

In collaboration with C. Lusso, F. Bouchut, A. Ern.

Flows of dense granular materials comprise regions where the material is flowing, and regions where it is static. In [15], we introduce two numerical methods to deal with the particular formulation of this model with a free interface. They are used to evaluate the respective role of yield and viscosity for the case of a constant source term, which corresponds to simple shear viscoplastic flows. Both the analytical solution of the inviscid model and the numerical solution of the viscous model (with a constant viscosity or the variable viscosity of the $\mu(I)$-rheology) are compared with experimental data.

7.1.5. Metamodelling of a road traffic assignment model

Participant: Vivien Mallet.

In collaboration with R. Chen, V. Aguiléra, F. Cohn, D. Poulet, F. Brocheton.

We proposed a metamodelling approach to design a close approximation to the traffic model, but with a very low computational cost. It consists in a dimensionality reduction of the model outputs by principal component analysis and a statistical emulation relying on regression and interpolation between training samples. A case study was carried out for the agglomeration of Clermont-Ferrand (France). Compared with traffic flow measurements, the performance of the metamodel is similar to that of the complete model during a one-month period, but the computational time decreases from 2 days on 110 cores to less than 1 minute on one core.

7.2. Assessments of models by means of experimental data and assimilation

7.2.1. Evaluation and calibration of mobile phones for noise monitoring application

Participants: Vivien Mallet, Raphaël Ventura.

In collaboration with V. Issarny, P-G. Raverdy, F. Rebhi.

The Ambiciti application was developed so as to acquire a larger control over the acquisition process by mobile phone sensors. Pink and narrowband noises were used to evaluate the phones' accuracy at levels ranging from background noise to 90 dB(A) inside the lab. Conclusions of this evaluation lead to the proposition of a calibration strategy that has been embedded in Ambiciti and applied to more than 50 devices during public events. In the perspective of citizens-driven noise sensing, in situ experiments were carried out, while additional tests helped to produce recommendations regarding the sensing context (grip, orientation, moving speed, mitigation, frictions, wind).

7.2.2. Assimilation of noise pollution data

Participants: Vivien Mallet, Raphaël Ventura.

In collaboration with P. Aumond, A. Can, V. Issarny.

We studied the generation of hourly noise maps in urban area at street resolution, based on temporally averaged simulation maps and mobile phone audio recordings. A data assimilation method produces an analysis noise map which is the so-called best linear unbiased estimator. We illustrated the method with a neighborhood-wide experiment.

Another work, lead by IFSTTAR, was dedicated to the spatial interpolation of point measurements collected at high density in Paris with a sound level meter. Compelling results were obtained with universal Kriging and a linear trend based on the distance to certain types of roads.

7.2.3. Granular and particle-laden flows: from laboratory experiments to field observations

Participant: Anne Mangeney.

In collaboration with R. Delannay, A. Valance, O. Roche and P. Richard.

A review article was written to provide an overview of dry granular flows and particle fluid mixtures, including experimental and numerical modelling at the laboratory scale, large scale hydrodynamics approaches and field observations. We also emphasize that the up-scaling from laboratory experiments to large scale geophysical flows still raises some theoretical physical challenges.

7.2.4. Continuum viscoplastic simulation of a granular column collapse on large slopes: $\mu(I)$ rheology and lateral wall effects

Participant: Anne Mangeney.

In collaboration with N. Martin, I. Ionescu, F. Bouchut and M. Farin.

We simulate here dry granular flows resulting from the collapse of granular columns on an inclined channel and compare precisely the results with laboratory experiments. The 2-D model is based on the so-called $\mu(I)$ rheology that induces a Drucker-Prager yield stress and a variable viscosity. We show that the use of a variable or a constant viscosity does not change significantly the results provided that these viscosities are of the same order. Finally, we observed that small-scale instabilities develop when refining the mesh.

7.3. Analysis of models in Fluid Mechanics

7.3.1. Analysis of the Riemann problem for a shallow water model with two velocities

Participants: Emmanuel Audusse, Edwige Godlewski, Martin Parisot.

In collaboration with N. Aguillon.

The question addressed in [24] is the hyperbolicity of a shallow water model with two velocities. The model is written in a nonconservative form and the analysis of its eigenstructure shows the possibility that two eigenvalues coincide. A definition of the nonconservative product is given which enables us to analyse the resonance and coalescence of waves. Eventually, we prove the well-posedness of the two dimensional Riemann problem with initial condition constant by half-plane.

7.3.2. Different formulations of an elliptic problem issued from geophysics

Participants: Cindy Guichard, Ani Miraçi, Yohan Penel, Jacques Sainte-Marie.

A simplified problem coming from [33] involving pressure and velocity unknowns is studied. Some weak formulations (conform or mixed) are derived and their well-posedness is analysed. These weak formulations are then discretised in a finite element framework with suitable discrete spaces.

7.4. Numerical methods for fluid flows

7.4.1. Kinetic entropy for the layer-averaged hydrostatic Navier-Stokes equations

Participants: Emmanuel Audusse, Marie-Odile Bristeau, Jacques Sainte-Marie.

In [26], the authors are interested in the numerical approximation of the hydrostatic free surface incompressible Navier-Stokes equations. By using a layer-averaged version of the equations, previous results obtained for shallow water system are extended. A vertically implicit / horizontally explicit finite volume kinetic scheme is designed that ensures the positivity of the approximated water depth, the well-balancing and a fully discrete energy inequality.

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

Participants: Marie-Odile Bristeau, Anne Mangeney, Jacques Sainte-Marie, Fabien Souillé.

In collaboration with S. Allgeyer, M. Vallée, R. Hamouda, D. Froger.

A stable and robust strategy is proposed to approximate incompressible hydrostatic Euler and Navier-Stokes systems with free surface. 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. We show that the model admits a kinetic interpretation, and we use this result to formulate a robust finite volume scheme for its numerical approximation.

7.4.3. Well balanced schemes for rotation dominated flows

Participants: Emmanuel Audusse, Do Minh Hieu, Yohan Penel.

In collaboration with P. Omnes.

In [27], we study the property of colocated Godunov type finite volume schemes applied to the linear wave equation with Coriolis source term. The purpose is to explain the bad behaviour of the classical scheme and to modify it in order to avoid accuracy issues around the geostophic equilibrium. We use tools from two communities: well-balanced schemes for the shallow water equation with topography and asymptotic preserving schemes for the low Mach model. CFL conditions that ensure the stability of fully discrete schemes are established. The extension to the nonlinear case is under study.

7.4.4. A two-dimensional method for a dispersive shallow water model

Participants: Nora Aïssiouene, Marie-Odile Bristeau, Anne Mangeney, Jacques Sainte-Marie.

In collaboration with C. Pares.

In [29], [6], we propose a numerical method for a two-dimensional dispersive shallow water system with topography [3]. A first approach in one dimension, based on a prediction-correction method initially introduced by Chorin-Temam has been presented in [33]. The prediction part leads to solving a shallow water system for which we use finite volume methods while the correction part leads to solving a mixed problem in velocity and pressure. From the variational formulation of the mixed problem proposed in [35], the idea is to apply a finite element method with compatible spaces to the two-dimensional problem on unstructured grids.

7.4.5. *Entropy-satisfying scheme for a hierarchy of dispersive reduced models of free surface flow*

Participant: Martin Parisot.

Article [32] is devoted to the numerical resolution in multidimensional framework of a hierarchy of reduced models of the free surface Euler equations. An entropy-satisfying scheme is proposed for the monolayer dispersive models [40] and [3]. 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. A work in progress focuses on the adaptation of the entropy-satisfying scheme to the layerwise models proposed in [30].

7.4.6. *A lateral coupling between river channel and flood plain with implicit resolution of shallow water equations*

Participant: Martin Parisot.

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

Multi-dimensional coupling in river hydrodynamics offers a convenient solution to properly model complex flow while limiting the computational cost and taking the advantage of most pre-existing models. The project aims to adapt the lateral interface coupling proposed in [39] to the implicit version and assess it with real data from the Garonne River.

7.4.7. *The discontinuous Galerkin gradient discretisation*

Participant: Cindy Guichard.

In collaboration with R. Eymard.

The Symmetric Interior Penalty Galerkin (SIPG) method, based on Discontinuous Galerkin approximations, is shown to be included in the Gradient Discretisation Method (GDM) framework. Therefore, it can take benefit from the general properties of the GDM, since we prove that it meets the main mathematical gradient discretisation properties on any kind of polytopal mesh. We illustrate this inheritance property on the case of the p-Laplace problem [13].

7.4.8. *Gradient-based optimization of a rotating algal biofilm process*

Participants: Pierre-Olivier Lamare, Jacques Sainte-Marie.

In collaboration with N. Aguillon, O. Bernard.

Here we focus on the optimal control of an innovative process where the microalgae are fixed on a support. They are thus successively exposed to light and dark conditions. The resulting growth can be represented by a dynamical system describing the denaturation of key proteins due to an excess of light. A PDE model of the Rotating Algal Biofilm is then proposed, representing local microalgal growth submitted to the time varying light. An adjoint-based gradient method is proposed to identify the optimal (constant) process folding and the (time varying) velocity of the biofilm.

7.4.9. *Method of reflections*

Participant: Julien Salomon.

In collaboration with G. Legendre, P. Laurent, G. Ciaramella, M. Gander, L. Halpern.

In [17], the authors carefully trace the historical development of the methods of reflections, give several precise mathematical formulations and an equivalence result with the alternating Schwarz method for two particles.

In [31], a general abstract formulation is proposed in a given Hilbert setting and the procedure is interpreted in terms of subspace corrections. The unconditional convergence of the sequential form is proven and a modification of the parallel one is proposed to make it unconditionally converging.

7.5. Modelling of environmental impacts and natural hazards

7.5.1. Numerical simulation of the 30–45 ka debris avalanche flow of Montagne Pelée volcano, Martinique: from volcano flank collapse to submarine emplacement

Participant: Anne Mangeney.

In collaboration with M. Brunet, L. Moretti, A. Le Friant, E.D. Fernandez Nieto, F. Bouchut.

We simulate here the emplacement of the debris avalanche generated by the last flank collapse event of Montagne Pelée volcano (30–45 ka), Martinique, Lesser Antilles. Our objective is to assess the maximum distance (i.e., runout) that can be reached by this type of debris avalanche as a function of the volume involved. This result provides new constraints on the emplacement processes of debris avalanches associated with these collapses which can drastically change the related hazard assessment such as the generated tsunami, in a region known for its seismic and volcanic risks.

7.5.2. Global sensitivity analysis and uncertainty quantification of on-road traffic emissions

Participant: Vivien Mallet.

In collaboration with R. Chen, V. Aguiléra, F. Cohn, D. Poulet, F. Brocheton.

Road traffic emissions of air pollutants depend on both traffic flow and vehicle emission factors. Global sensitivity analyses, especially the computation of Sobol' indices, were 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. Furthermore, the uncertainties in traffic flow and pollutant emissions were quantified by propagating into the model the uncertainties in the input parameters.

7.5.3. Uncertainty quantification in atmospheric dispersion of radionuclides

Participants: Ngoc Bao Tran Le, Vivien Mallet.

In collaboration with I. Korsakissok, R. Périllat, A. Mathieu, D. Didier.

In collaboration with IRSN, we investigated the uncertainties of the atmospheric-dispersion forecasts that are used during an accidental release of radionuclides like the Fukushima disaster. In order to quantify the uncertainties, Monte Carlo simulations and calibrations were carried out and coupled with ensemble meteorological forecasts from the European Centre for Medium-Range Weather Forecasts.

7.5.4. Simulation of air and noise pollution at high resolution and large scale

Participant: Vivien Mallet.

In collaboration with C. Pesin, P. Béal.

We developed fast surrogates for urban pollution models that they can be applied at global scale while preserving the street resolution, the main physical constraints and the performance against observational data. The surrogate models are based on the original models, machine learning algorithms and observational data.

7.6. Software developments

7.6.1. Improvements in the FRESHKISS3D code

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

Several tasks have been achieved in the FRESHKISS3D software:

- Reworked unittests and basic continuous integration
- Optimized IO functions
- Added compatibility with new mesh format
- Added generic run script that only takes yaml data as input
- Added validation cases and new example scripts
- Added paraview integrated post-processing scripts
- Reworked API and online documentation with sphinx
- Simplified dependencies and upgraded python to 3.6
- Worked on new numerical schemes:
  - Added implicit scheme for vertical exchanges terms
  - Reworked vertical viscosity scheme
  - Added new fluxes computations
  - Fixed vrious bugs (second order, viscosity, water state law)
  - Added vertical settling scheme on tracer (suspension models)
- Added 3D interpolator
- Added lagrangian particle tracking with reflexions on boundaries
- C++ Non-hydrostatic code (Nora) converted in cython (80%)
- Developement of a « Vilaine » package designed for SAUR/IAV/ANGE project

7.6.2. Numerical simulation of Free Surface Navier Stokes equations with Telemac 3D

Participants: Emmanuel Audusse, Nicole Goutal.

In collaboration with P. Quemar, A. Decoene, O. Lafitte, A. Leroy, C.T. Phan.

This work takes place in a joint project with EDF-LNHE (Laboratoire national d'hydraulique et d'environnement). The aim of the project is to understand the limitation of the actual numerical solution of the free surface Navier Stokes equations with software TELEMAC 3D and to propose new ways to handle important points as the advective part, the divergence free constraint, the coupling between velocity and hydrostatic pressure or the boundary conditions. A study of the mild-slope equation is also performed in order to obtain comparison solutions.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

A contract has been made (120.000 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.

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).

The ANR project ESTIMAIR includes the SME NUMTECH for a commercial deployment of the project results.

The EIT Digital project Env&You involves the SME NUMTECH and the startup Ambiciti, whose products rely on the results of this European project.

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".

J. Thorey's PhD thesis was funded by EDF R&D (CIFRE). The title of PhD thesis was: "Ensemble forecasting using sequential aggregation for photovoltaic power applications".

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. 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).

9.1.2. ANR Hyflo-Eflu (2016-2020)

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.

9.1.3. ANR MIMOSA (2014–2017)

Participants: Marie-Odile Bristeau, Anne Mangeney, Bernard Di Martino, Jacques Sainte-Marie.

Program: ANR Défi 1 "Gestion sobre des ressources et adaptation au changement climatique"

Project acronym: MIMOSA

Project title: MIcroseism modelling and Seismic Applications

Coordinator: Eleonore Stutzmann (IPGP)

Seismic noise is recorded by broadband seismometers in the absence of earthquakes. It is generated by the atmosphere-ocean system with different mechanisms in the different frequency bands. Even though some mechanisms have been known for decades, an integrated understanding of the noise in the broadband period band 1-300sec is still missing. Using novel theoretical, numerical and signal processing methods, this project will provide a unified understanding of the noise sources and quantitative models for broadband noise. Conversely, we will be able to interpret seismic noise in terms of ocean wave properties. This first analysis step will lead to the identification and characterisation of source events, which we will use to improve noise tomography, and seismic monitoring.

9.1.4. 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.

9.1.5. CNRS Moset (2016-2017)

Participants: Emmanuel Audusse, Martin Parisot.

CNRS project call: INSU Tellus

Project acronym: Moset

Project title: Modélisation des suspensions concentrées naturelles

Coordinator: Emmanuel Audusse

In collaboration with G. Antoine (EDF), S. Boyaval (LHSV), C. Le Bouteiller (Irstea), M. Jodeau (EDF).

Gathering mathematicians (numerical analysis) and geophysicists, this project focuses on the quantitative prediction of solid transport. This issue raises several questions about rheology when the sediment concentration is high enough. It is crucial for modelling the dynamics of suspension. The collaboration aims at assessing models by means of experimental data and at providing preliminary numerical results to evaluate the order of magnitude of constraints.

9.1.6. CNRS Simulations of free-surface flows (2017)

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

CNRS project call: PEPS JC

Project title: modélisation avancée et simulation d'écoulements à surface libre

Coordinator: Yohan Penel

Funding: 2.5k euros

In collaboration with E. Fernaández-Nieto.

Free-surface flows are extensively studied in the literature by means of simplified models (like the Shallow Water equations) due to the theoretical and numerical issues related to the Euler system. Intermediate models have then been derived to improve the accuracy and the physical relevance (e.g. taking into account hydrodynamic pressure or multilayer approaches). 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, ...).

9.1.7. 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 [39] to the implicit version and test it on real data for the Garonne River.

9.1.8. 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".

9.1.9. 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.

9.1.10. GdR EGRIN (2013–2017)

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.

9.1.11. ANR ESTIMAIR (2013-2017)

Participant: Vivien Mallet.

ANR project call: Modèles numériques

Project acronym: ESTIMAIR

Project title: Estimation d'incertitudes en simulation de la qualité de l'air à l'échelle urbaine

Coordinator: Vivien Mallet

Funding: 415k euros

The project aims to propagate uncertainties in a complete air quality modelling chain at urban scale, from road traffic assignment to air pollutant dispersion.

9.1.12. 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).

9.1.13. 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.

9.1.14. 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

9.1.15. ANR CARIB (2014-2017)

Participant: Anne Mangeney.

ANR project call: Simi6

Project acronym: CARIB

Project title: Fréquence et processus de mise en place des avalanches de débris tsunamigènes de l'arc des Petites Antilles : apport des forages de l'Expédition IODP 340 et impact en termes de risque

Coordinator: Anne Le Friant (IPGP)

Funding: 274k euros

9.1.16. 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)

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)

Participants: Vivien Mallet

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.

9.2.1.3. Env&You (2017)

Title: Env&You

Program: EIT Digital

Duration: January 2016 - December 2016

Coordinator: Inria (MiMove)

Partners: NUMTECH, Ambiciti, ForumVirium, TheCivicEngine

Inria contact: Valérie Issarny (Mimove project-team)

Participants: Vivien Mallet, Raphaël Ventura

Env&You aims at delivering the whole picture of urban pollution, from the individual exposure to neighborhood-by-neighborhood and day-to-day variation, to citisens and governments, informing their decisions for healthy urban living.

9.2.2. Collaborations with Major European Organisations

9.2.2.1. CNRS PICS NHML (2017-2019)

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 (CEREMA)

Other partners: IMUS (Sevilla, Spain)

Participants: Martin Parisot (Inria), Jacques Sainte-Marie (CEREMA), 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. Inria International Partners

9.3.1.1. Informal International Partners

Two collaborations with foreign colleagues are to be mentioned:

- 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.
- 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.

9.3.2. Participation in Other International Programs

9.3.2.1. PROCORE Hong-Kong (2016-2017)

Program: Hubert Curien PROCORE

Project title: time-parallelisation methods for control

Duration: 01/16-12/17

Coordinator: Felix Kwok (Univ. Hong-Kong)

Other partners: HKBU (Hong-Kong)

Funding: 5k euros

9.4. International Research Visitors

9.4.1. Visits to International Teams

9.4.1.1. Research Stays Abroad

- Y. Penel spent one month and a half (Mar.-Apr.) at the university of Sevilla (Spain) to collaborate with E. Fernández-Nieto.
- M. Parisot spent a week to Sevilla in April.

We also mention that M. Parisot spent four separate weeks at the university of Toulouse (CERFACS).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organising Committees

Y. Penel and J. Sainte-Marie organised (with E. Fernández-Nieto) the workshop "An overview on free-surface flows" that took place at Inria on November 13th-14th and that gathered 35 researchers (from France, US, Italy, Spain).

B. Di Martino, J. Sainte-Marie and A. Mangeney organised the 5th EGRIN summer school that took place at Cargèse from May, 29th to June, 2nd and that gathered 40 researchers.

E. Audusse was a member of the organising committee of the 8th edition of the conference "Finite-Volume for Complex Applications" (June, 12th to 16th, Lille).

M. Parisot organises the monthly ANGE seminar. The program for 2017 comprises: H. Martin (Jan), J. Sainte-Marie (Feb), V. Desveaux (Mar), V. Mallet (Apr), B. Al Taki (Sep), C. Cancès (Oct), R. Ventura (Nov), V. Duchêne (Dec).

J. Salomon co-organises the LJLL-Inria meetings (twice a month). L. Boittin co-organises the Junior Seminar at Inria–Paris.

E. Audusse, C. Guichard and Y. Penel organised the welcome session for newly recruited researchers in mathematics on behalf of national research institutions (CNRS, Inria, SFdS, SMAI, SMF) on January, 23rd.

We finally mention that M. Parisot and J. Salomon will organise a workshop entitled "Scientific computing and optimisation processes for renewable energies" at Inria on January 2018.

10.1.2. Journal

10.1.2.1. Reviewer - Reviewing Activities

Member	Journal
E. Audusse	M2AN, Water
C. Guichard	Springer Proc. Math., CRAS Mathematiques, Journal of Scientific Computing, Computers and Mathematics with Applications, Journal of Computational Physics, Numerical Methods for Partial Differential Equations
V. Mallet	JAMES, Atmospheric Environment
A. Mangeney	JGR, GRL, GJI
M. Parisot	ESAIM:ProcS, Springer Proc. Math., M2AN
Y. Penel	Springer Proc. Math., Journal of Computational Physics
J. Sainte-Marie	M2AN, Applied Mathematical Modelling, IJNMF, ANR, Journal of Scientific Computing
J. Salomon	CRAS, SIAM SISC, JMPA

10.1.3. Invited Talks

Conference	Location	Month	Members involved
Earth Sciences	Durham (UK)	January	A. Mangeney
DD24	Svalbard (Norway)	February	J. Salomon
Workshop on reduced order methods for wind and marine current power	Nice	February	J. Salomon
Entreprises pour l'environnement	Paris	March	V. Mallet
Clifford lectures	New-Orleans (USA)	April	J. Sainte-Marie
EGRIN (5th summer school)	Cargèse	June	E. Audusse, L. Boittin, A. Mangeney, M. Parisot, Y. Penel
SMAI 2017	Ronce-les-Bains	June	F. Wahl
Forecasting and risk management for renewable energy	Paris	June	V. Mallet
EGU Meeting "Environmental Seismology"	Ohlstadt (Germany)	June	A. Mangeney
InterNoise	Hong-Kong	August	V. Mallet
SIAM GS 2017	Erlangen (Germany)	September	E. Audusse, L. Boittin, F. Wahl
ENUMATH	Voss (Norway)	September	J. Salomon
Euromech colloquium 588	Toulouse	October	L. Boittin
Workshop EDP-Normandie	Caen	October	F. Wahl
Workshop on numerical schemes for low Mach number flows	Toulouse	November	Y. Penel
UMEMA	Torino (Italy)	November	V. Mallet
Numerical Methods for Shallow Water Equations and Related Models	Shenzhen (China)	November	J. Sainte-Marie
NumWave	Montpellier	December	E. Audusse, M. Parisot

Seminars	Date	Member
Clermont-Ferrand (Blaise Pascal)	January	B. Di Martino
Bordeaux (MathOcean)	February	E. Audusse, M. Parisot
Marne-la-Vallée (ENPC)	February	V. Mallet
Orléans (BRGM)	April	A. Mangeney
Tokyo (Earthquake Research Institute)	October	A. Mangeney
Saclay (CEA-LSCE)	October	V. Mallet

10.1.4. Leadership within the Scientific Community

Organisation	People	Duty
AMIES	E. Godlewski	Member of board
ANR	V. Mallet	Expert
CEREMADE lab council	J. Salomon	Member
CFEM	E. Godlewski	Director
EGRIN	E. Audusse	Correspondent (Paris 13)
	B. Di Martino	Correspondent (Corse)
	N. Goutal	Correspondent (EDF)
	C. Guichard	Correspondent (UPMC)
	B. Haspot	Correspondent (CEREMADE)
	A. Mangeney	Member of board
	M. Parisot	Correspondent (ANGE)
	J. Sainte-Marie	Scientific head
HCERES	A. Mangeney	Expert
LJLL	E. Godlewski	Deputy director
SMAI	Y. Penel	Member of board

A. Mangeney was a member of the hiring panel for a full professor position at Univ. Grenoble Alpes.

We also mention that V. Mallet and M. Parisot are members of the Inria committee of doctoral monitoring and that J. Salomon is in the committee for the next location of the centre Inria Paris (rue Barrault).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

PhD degree - J. Salomon, Variational inequalities, 6 hours (lectures), Mines ParisTech

Master's degree (M2) - E. Godlewski and J. Sainte-Marie, Hyperbolic models for complex flows and energy applications, 25 hours (lectures), Univ. Pierre et Marie Curie Paris 6

Master's degree (M2) - C. Guichard, Numerical methods for nonstationary PDEs, 6 hours (programming classes), Univ. Pierre et Marie Curie Paris 6

Master's degree (M2) - A. Mangeney and J. Sainte-Marie, Numerical methods in geosciences, 60 hours (lectures and programming classes), Univ. Paris Diderot Paris 7, IPGP

Master's degree (M2) - B. Di Martino, Mathematical modelling, 21 hours (lectures, example and programming classes), Univ. Corse

Master's degree (M2) - J. Salomon, Numerical methods for PDEs, 45 hours (lectures), Univ. Paris Dauphine

Master's degree (M2)- Y. Penel, Deterministic models in life sciences, 30h (lectures, example and programming classes), Univ. Paris Descartes

Master's degree (M2) - V. Mallet, Modelling of air quality, 4.5h (lectures), ENPC

Master's degree (M2) - V. Mallet, Simulation of atmospheric dispersion, 3h (programming classes), ENPC

Master's degree (M2) - V. Mallet, Data assimilation in geophysics, 7h (programming classes), ENPC

Engineering school (2nd year) - E. Audusse, ODEs: analysis and numerical simulation, 30 hours (lectures and example classes), Univ. Paris 13

Engineering school (2nd year) - E. Audusse, Finite difference method for PDEs, 21 hours (lectures), Univ. Paris 13

Master's degree (M1) - C. Guichard, Basis of numerical methods, 63 hours (programming classes), Univ. Pierre et Marie Curie Paris 6

Master's degree (M1) - A. Mangeney and J. Sainte-Marie, Modelling of gravity flows, 80 hours (lectures), Univ. Paris Diderot Paris 7, IPGP

Master's degree (M1) - B. Di Martino, Finite element methods, 18 hours (lectures and example classes), Univ. Corse

Master's degree (M1) - B. Di Martino, Risk modelling, 33 hours (lectures and example classes), Univ. Corse

Master's degree (M1) - J. Salomon, Numerical methods for PDEs, 85 hours (lectures and example classes), Univ. Paris Dauphine

Engineering school (1st year) - E. Audusse, Introduction to scientifi computing, 30 hours (lectures and example classes), Univ. Paris 13

Bachelor's degree (L3) - L. Boittin and C. Guichard, Numerical methods for ODEs, 101.5 hours (example and programming classes), Univ. Pierre et Marie Curie Paris 6

Bachelor's degree (L3) - C. Guichard, Python, 21 hours (programming classes), Univ. Pierre et Marie Curie Paris 6

Bachelor's degree (L3) - B. Di Martino, Linear numerical analysis, 54 hours (lectures and example classes), Univ. Corse

Bachelor's degree (L2) - B. Di Martino, Analysis, 54 hours (lectures and example classes), Univ. Corse

Bachelor's degree (L2) - Y. Penel and F. Wahl, Integration in 2 and 3 dimensions, 31 hours (lectures and example classes), Univ. Pierre et Marie Curie Paris 6

Bachelor's degree (L2) - C. Guichard, Some numerical approximations, 17 hours (lectures and programming classes), Univ. Pierre et Marie Curie Paris 6

Bachelor's degree (L1-L2) - B. Al Taki, Analysis and algebra, 12 hours (example classes), Univ. Pierre et Marie Curie Paris 6

Bachelor's degree (L1-L2) - Y. Penel, Linear algebra, ODEs, elementary probabilities, 20 hours (example classes), Univ. Pierre et Marie Curie Paris 6 (Roscoff)

Bachelor's degree (L1) - B. Di Martino, Mathematics, 18 hours (lectures and example classes), Univ. Corse

Bachelor's degree (L1) - F. Wahl, Sequences and integrals, linear algrebra, 54 hours (example classes), Univ. Pierre et Marie Curie Paris 6

Bachelor's degree (L1) - L. Boittin, Calculus, 38 hours (example classes), Univ. Pierre et Marie Curie Paris 6

Bachelor's degree (L1) - L. Boittin, Linear algebra, 18 hours (example classes), Univ. Pierre et Marie Curie Paris 6

Some members are responsible of educational pathways:

E. Audusse is the deputy director of the "Applied Mathematics and Scientific Computing" program of the SupGalilee engineering school.

E. Godlewski is the head of the "Mathematics for Industry" M.Sc. program of Univ. Pierre et Marie Curie Paris 6.

C. Guichard is the associated head of the "Mathematics and Programming" B. program of Univ. Pierre et Marie Curie Paris 6.

10.2.2. Supervision

PostDoc in progress -Bilal Al Taki, *Well-posedness multilayers Saint-Venant equation*, Inria, supervised by J. Sainte-Marie, B. Haspot and B. Di Martino, from 2017

PostDoc in progress - Guillaume Chérel, *Data assimilation for urban pollution*, Inria, supervised by V. Mallet, from 2016

PostDoc in progress - Virginie Durand, *Analysis of rockfalls and generated seismic signals at La Réunion*, Institut de Physique du Globe (Univ. Paris 7), supervised by A. Mangeney, from 2014

PostDoc in progress - Jannelle Hamond, *Uncertainty quantification applied to air quality simulation at urban scale*, Inria, supervised by V. Mallet, from 2017

PostDoc in progress - El Hadji Kone, *Numerical modelling of two-phase flows*, Institut de Physique du Globe (Univ. Paris 7), supervised by A. Mangeney (in collaboration with G. Narbona-Reina and E. Fernández-Nieto), from 2014

PostDoc in progress - Pierre-Olivier Lamarre, *Optimisation of the hydrodynamic regime in a raceway and lagrangian trajectories of algae*, Inria, supervised by J. Sainte-Marie (in collaboration with O. Bernard, BIOCORE), from 2016

PostDoc in progress - Sylvain Viroulet, *Numerical modelling of granular flows and generated forces on the topography*, Institut de Physique du Globe (Univ. Paris 7), supervised by A. Mangeney, from 2016

PhD - Do Minh Hieu, *Analyse mathématique et schémas volumes finis pour la simulation des écoulements quasi-géostrophiques à bas nombre de Froude*, Univ. Paris 13, supervised by E. Audusse and Y. Penel (in collaboration with S. Dellacherie and P. Omnes), defended on Dec. 17

PhD - Jean Thorey, *Prévision d'ensemble par agrégation séquentielle appliquée à la prévision de production d'énergie photovoltaïque*, Inria grant, supervised by V. Mallet (in collaboration with I. Herlin), defended on Sept. 17

PhD in progress - Frédéric Allaire, *Quantification du risque incendie par méta-modélisation de la propagation de feux de forêt*, Inria grant, supervised by V. Mallet and J. Sainte-Marie, from 2017

PhD in progress - Vincent Bachelet, *Granular flows and generated acoustic waves: a laboratory investigation*, Institut de Physique du Globe (Univ. Paris 7), supervised by A. Mangeney (in collaboration with J. De Rosny and R. Toussaint), from 2015

PhD in progress - Léa Boittin, *Modelling, analysis and efficient numerical resolution for erosion processes*, Univ. Pierre et Marie Curie Paris 6 (Inria grant), supervised by E. Audusse, M. Parisot and J. Sainte-Marie, from Jan. 16

PhD in progress - Pauline Bonnet, *Vers un catalogue des vêlages d'icebergs en région polaire par une approche couplée sismologie et modélisation mécanique*, Institut de Physique du Globe (Univ. Paris 7) & ENSAM-ParisTech, supervised by A. Mangeney (in collaboration with O. Castelnau and V. Yastrebof), from 2017

PhD in progress - Ruiwei Chen, *Uncertainty quantification in the simulation of traffic emissions*, ENPC, supervised by V. Mallet (in collaboration with V. Aguiléra, K. Sartelet), from 2014

PhD in progress - Virgile Dubos, *Numerical methods for the elliptic/parabolic parts of non-hydrostatic fluid models*, Univ. Pierre et Marie Curie Paris 6, supervised by C. Guichard, Y. Penel and J. Sainte-Marie, from 2017

PhD in progress - Nadia Jbili, *Contrôle optimal pour la résonance magnétique nucléaire*, Univ. Paris Dauphine, supervised by J. Salomon, from 2017

PhD in progress - Julian Kühnert, *Simulation of high frequency seismic waves*, Institut de Physique du Globe (Univ. Paris 7), supervised by A. Mangeney, from 2016

PhD in progress - Antoine Lesieur, *Estimation d'état et modélisation inverse appliquées à la pollution sonore en milieu urbain*, Inria grant, supervised by V. Mallet and J. Sainte-Marie, from 2017

PhD in progress - Hugo Martin, *Simulation of the coupling between seismic waves and granular flows*, Institut de Physique du Globe (Univ. Paris 7), supervised by A. Mangeney (in collaboration with Y. Maday), from 2016

PhD in progress - Ethem Nayir, *Approximation multi-vitesse des équations de Navier-Stokes hydrostatiques: Analyse mathématique et simulations numériques*, Univ. Pierre et Marie Curie Paris 6, supervised by E. Audusse, Y. Penel and J. Sainte-Marie, from 2014

PhD in progress - Nourelhouda Omrane, *Mathematical analysis and control of free-surface flows in variable domains*, Univ. Corse, supervised by B. Di Martino, from 2016

PhD in progress - Marc Peruzzetto, *Hazard assessment related to gravitational flows: from modelling to risk analysis*, Institut de Physique du Globe (Univ. Paris 7) & BRGM, supervised by A. Mangeney (in collaboration with G. Grandjean), from 2017

PhD in progress - Pierrick Quémar, *3D numerical simulations of environmental hydrolics: application to Telemac*, Univ. Paris 13, supervised by E. Audusse and N. Goutal (in collaboration with A. Decoene, O. Lafitte, A. Leroy and C. Tuân Phan), from 2016

PhD in progress - Sebastian Reyes-Riffo, *Mathematical methods for recovering marine energies*, Univ. Paris Dauphine, supervised by J. Salomon, from 2016

PhD in progress - Raphaël Ventura, *Estimation de la pollution sonore en milieu urbain par assimilation d'observations mobiles*, Inria grant, supervised by V. Mallet (in collaboration with I. Herlin), from 2014

PhD in progress - Fabien Wahl, *Modelling and analysis of interactions between free surface flows and floating objects*, Univ. Pierre et Marie Curie Paris 6, supervised by C. Guichard, E. Godlewski, M. Parisot and J. Sainte-Marie, from 2015

M2 internship - M Hamed Bouchiba, *Obtention du modèle Saint-Venant et stabilité des solutions faibles d'un système Navier-Stokes compressible*, Univ. Pierre et Marie Curie Paris 6, supervised by B. Di Martino and B. Haspot, Summer 2017

M2 internship - Anthony Guimpier, *Analyse d'instabilités gravitaires sur Mars*, IPGP, supervised by A. Mangeney (in collaboration with S. Conway, N. Mangold), Summer 2017

M2 internship - Ani Miraçi, *Formulation et étude numérique de la partie elliptique d'un problème couplé issu de la géophysique*, Univ. Pierre et Marie Curie Paris 6, supervised by C. Guichard, Y. Penel and J. Sainte-Marie, Summer 2017

M2 internship - Pablo Poulain, *Granular flows simulation and application to volcanic flank collapse*, IPGP, supervised by A. Mangeney (in collaboration with A. Le Friand, G. Boudon), Summer 2017

M2 internship - Laurie Supperamaniyen, *Dynamique éruptive du Piton de La Fournaise à partir d'enregistrements sismiques continus sur quelques éruptions*, IPGP, supervised by A. Mangeney (in collaboration with S. Vergniolle), Summer 2017

L3 internship - Jeanne Trinquier, *Tsunamath : module pédagogique présentant les travaux de modélisation et de simulation des tsunamis*, supervised by E. Audusse and J. Sainte-Marie, Summer 2017

10.2.3. Juries

Jan., PhD - J. Sainte-Marie: Amélie Simon (E. Centrale Lyon, *Modélisation des phénomènes de films liquides dans les turbines à vapeur*)

Feb., PhD - Y. Penel: Arthur Talpaert (École Polytechnique, *Simulation numérique directe de bulles sur maillage adaptatif avec algorithmes distribués*)

Mar., PhD - J. Sainte-Marie (referee): Nabil El Mocayd (Cerfacs – Univ. Toulouse, *La décomposition en polynôme du chaos pour l'amélioration de l'assimilation de données ensembliste en hydraulique fluviale*)

Mar., PhD - A. Mangeney (referee): Perrine Freydier (Univ. Grenoble Alpes – IRSTEA, *Dynamique interne au front d'écoulements à surface libre. Application aux laves torrentielles*)

May, PhD - J. Salomon: Francesc Fàbregas Flavià (École Centrale Nantes, *A Numerical Tool for the Frequency Domain Simulation of Large Clusters of Wave Energy Converters*)

May, PhD - V. Mallet: Chi Vuong NGuyen (École CEntrale Lyon, *Assimilation de données et couplage d'échelles pour la simulation de la dispersion atmosphérique en milieu urbain*)

Sept., PhD - V. Mallet: Jean Thorey (Univ. Pierre et Marie Curie, *Prévision d'ensemble par agrégation séquentielle appliquée à la prévision de production d'énergie photovoltaïque*)

Nov., PhD - J. Sainte-Marie (president): Charles Demay (Univ. Grenoble Alpes, *Modélisation et simulation des écoulements transitoires diphasiques eau-air dans les circuits hydrauliques*)

Nov., PhD - A. Mangeney (president): Thao Trinh (Univ. Rennes 1, *Mécanisme d'érosion et de déposition de l'écoulement granulaire sur un fond meuble*)

Dec., PhD - A. Mangeney (president): Julien Brondex (Univ. Grenoble Alpes – IGA, *Influence de l'endommagement et du frottement basal sur la dynamique de la ligne d'échouage*)

Dec., PhD - A. Mangeney: Alexis Bougouin (Institut National Polytechnique de Toulouse, *Étude expérimentale de l'effondrement d'une colonne fluide-grains*)

Dec., PhD - C. Guichard: Riad Sanchez (IFPEN – Univ. Paris-Saclay, *Techniques de bases réduites pour les écoulements diphasiques en milieux poreux*)

Dec., PhD - J. Sainte-Marie (referee): Athmane Bakhta (Univ. Paris-Est, *Modèles mathématiques et simulation numérique de dispositifs photovoltaïques*)

Dec., PhD - B. Di Martino: Réjane Fieschi (Univ. Corse, *Résolution d'équations d'ondes en dimension quelconque*)

Dec., PhD - E. Audusse and Y. Penel: Do Minh Hieu (Univ. Paris 13, *Analyse mathématique de schémas volume finis pour la simulation des écoulements quasi-géostrophiques à bas nombre de Froude*)

10.3. Popularisation

April - V. Mallet participated to the workshop Smart Cities organised by *Le Monde* in Lyon

May - L. Boittin and R. Ventura ran a stand on the occasion of the "salon de la culture et des jeux mathématiques" on behalf of AMIES

June - V. Mallet and R. Ventura helped the organising committee of the "noise walk" in Helsinki (as well as the November edition)

June-July - J. Trinquier's internship was dedicated to the development of an educational tool based on Tsunamath which incorporates modelling and simulation (supervision by E. Audusse and J. Sainte-Marie)

October - E. Audusse got involved in a TPE for students in 1re S

November - R. Ventura provided an association in Paris with noise reports

November - Y. Penel ran a stand during the ONISEP exhibition

December - L. Boittin, Y. Penel and F. Wahl helped the organisation at the "Math. Employment" show

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^{o} 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^{o} 3, p. 459-490, http://dx.doi.org/10.1142/S0218202511005118.

Publications of the year

Articles in International Peer-Reviewed Journal

[5] N. AGUILLON, F. LAGOUTIÈRE, N. SEGUIN.*Convergence of finite volumes schemes for the coupling between the inviscid Burgers equation and a particle*, in "Mathematics of Computation", 2017, vol. 86, p. 157-196, https://arxiv.org/abs/1412.0376 , https://hal.inria.fr/hal-01077311.

[6] N. AISSIOUENE, M.-O. BRISTEAU, E. GODLEWSKI, A. MANGENEY, C. PARÉS, J. SAINTE-MARIE.*Application of a combined finite element -finite volume method to a 2D non-hydrostatic shallow water problem*, in "Springer Proceedings in Mathematics & Statistics", 2017, https://hal.archives-ouvertes.fr/hal-01664481.

[7] M.-O. BRISTEAU, C. GUICHARD, B. DI MARTINO, J. SAINTE-MARIE.*Layer-averaged Euler and Navier-Stokes equations*, in "Communications in Mathematical Sciences", 2017, https://arxiv.org/abs/1509.06218 [*DOI :* 10.4310/CMS.2017.v15.n5.a3], https://hal.inria.fr/hal-01202042.

[8] C. CANCÈS, C. GUICHARD.*Numerical analysis of a robust free energy diminishing Finite Volume scheme for parabolic equations with gradient structure*, in "Foundations of Computational Mathematics", 2017, vol. 17, n^{o} 6, p. 1525-1584, https://arxiv.org/abs/1503.05649 , https://hal.archives-ouvertes.fr/hal-01119735.

[9] R. CHEN, V. AGUILERA, V. MALLET, F. COHN, D. POULET, F. BROCHETON.*A sensitivity study of road transportation emissions at metropolitan scale*, in "Journal of Earth Sciences and Geotechnical Engineering", 2017, vol. 7, n^{o} 1, https://hal.inria.fr/hal-01676006.

[10] R. DELANNAY, A. VALANCE, A. MANGENEY, O. ROCHE, P. RICHARD.*Granular and particle-laden flows: from laboratory experiments to field observations*, in "Journal of Physics D: Applied Physics", 2017, vol. 50, n^{o} 5, 40 [*DOI :* 10.1088/1361-6463/50/5/053001], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01481019.

[11] B. DI MARTINO, B. HASPOT, Y. PENEL.*Global stability of weak solutions for a multilayer Saint-Venant model with interactions between the layers*, in "Nonlinear Analysis: Hybrid Systems", November 2017, vol. 163, p. 177 - 200 [*DOI :* 10.1016/j.na.2017.07.010], https://hal.archives-ouvertes.fr/hal-01407886.

[12] R. EYMARD, P. FERON, C. GUICHARD.*Family of convergent numerical schemes for the incompressible Navier-Stokes equations*, in "Mathematics and Computers in Simulation", August 2017 [*DOI :* 10.1016/j.matcom.2017.08.003], https://hal.archives-ouvertes.fr/hal-01382924.

[13] R. EYMARD, C. GUICHARD.*Discontinuous Galerkin gradient discretisations for the approximation of second-order differential operators in divergence form*, in "Computational and Applied Mathematics", December 2017 [*DOI :* 10.1007/S40314-017-0558-2], https://hal.archives-ouvertes.fr/hal-01535147.

[14] M. LACHOWICZ, H. LESZCZYŃSKI, M. PARISOT.*Blow-up and global existence for a kinetic equation of swarm formation*, in "Mathematical Models and Methods in Applied Sciences", June 2017, vol. 27, n^{o} 6, 22 [*DOI :* 10.1142/S0218202517400115], https://hal.inria.fr/hal-01370006.

[15] C. LUSSO, F. BOUCHUT, A. ERN, A. MANGENEY.*A free interface model for static/flowing dynamics in thin-layer flows of granular materials with yield: simple shear simulations and comparison with experiments*, in "Applied Sciences", April 2017, vol. 7, n^{o} 4, 386 [*DOI :* 10.3390/APP7040386], https://hal-upec-upem.archives-ouvertes.fr/hal-00992309.

[16] C. LUSSO, A. ERN, F. BOUCHUT, A. MANGENEY, M. FARIN, O. ROCHE.*Two-dimensional simulation by regularization of free surface viscoplastic flows with Drucker-Prager yield stress and application to granular collapse*, in "Journal of Computational Physics", March 2017, vol. 333, p. 387-408 [*DOI :* 10.1016/J.JCP.2016.12.036], https://hal-upec-upem.archives-ouvertes.fr/hal-01133786.

[17] J. SALOMON, M. J. GANDER, G. CIARAMELLA, L. HALPERN.*Review of the Methods of Reflections*, in "Oberwolfach Reports", October 2017, p. 1-21 [*DOI :* 10.14760/OWP-2017-27], https://hal.archives-ouvertes.fr/hal-01659764.

[18] J. THOREY, V. MALLET, P. BAUDIN.*Online learning with the Continuous Ranked Probability Score for ensemble forecasting*, in "Quarterly Journal of the Royal Meteorological Society", January 2017, vol. 143, n^{o} 702, p. 521 - 529 [*DOI :* 10.1002/QJ.2940], https://hal.inria.fr/hal-01676007.

[19] R. VENTURA, V. MALLET, V. ISSARNY, P.-G. RAVERDY, F. REBHI.*Evaluation and calibration of mobile phones for noise monitoring application*, in "Journal of the Acoustical Society of America", November 2017, vol. 142, n^{o} 5, p. 3084 - 3093 [*DOI :* 10.1121/1.5009448], https://hal.inria.fr/hal-01676004.

International Conferences with Proceedings

[20] J. DRONIOU, R. EYMARD, T. GALLOUËT, C. GUICHARD, R. HERBIN.*An error estimate for the approximation of linear parabolic equations by the Gradient Discretization Method*, in "FVCA 2017 - International Conference on Finite Volumes for Complex Applications VIII", Lille, France, Finite Volumes for Complex Applications VIII - Hyperbolic, Elliptic and Parabolic Problems, 2017, https://hal.archives-ouvertes.fr/hal-01442921.

[21] R. EYMARD, C. GUICHARD.*DGM, an item of GDM*, in "FVCA 2017 - International Conference on Finite Volumes for Complex Applications VIII", Lille, France, Finite Volumes for Complex Applications VIII - Hyperbolic, Elliptic and Parabolic Problems, 2017, https://hal.archives-ouvertes.fr/hal-01442922.

[22] V. RAPHAËL, V. MALLET, V. ISSARNY, P.-G. RAVERDY, F. REBHI.*Estimation of urban noise with the assimilation of observations crowdsensed by the mobile application Ambiciti*, in "INTER-NOISE 2017 - the 46th International Congress and Exposition on Noise Control Engineering Taming Noise and Moving Quiet", Hong Kong, China, August 2017, https://hal.inria.fr/hal-01676010.

Books or Proceedings Editing

[23] E. AUDUSSE, S. DELLACHERIE, D. M. HIEU, P. OMNES, Y. PENEL (editors). *Godunov type scheme for the linear wave equation with Coriolis source term*, LMLFN 2015 – Low Velocity Flows – Application to Low Mach and Low Froude regimes, EDP Sciences, November 2017, vol. 58 [*DOI :* 10.1051/PROC/201758001], https://hal.archives-ouvertes.fr/hal-01254888.

Other Publications

[24] N. AGUILLON, E. AUDUSSE, E. GODLEWSKI, M. PARISOT.*Analysis of the Riemann Problem for a shallow water model with two velocities*, October 2017, working paper or preprint, https://hal.inria.fr/hal-01618722.

[25] 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 2017, working paper or preprint, https://hal.inria.fr/hal-01393147.

[26] E. AUDUSSE, M.-O. BRISTEAU, J. SAINTE-MARIE.*Kinetic entropy for the layer-averaged hydrostatic Navier-Stokes equations*, September 2017, working paper or preprint, https://hal.inria.fr/hal-01583511.

[27] E. AUDUSSE, D. MINH HIEU, P. OMNES, Y. PENEL.*Analysis of modified Godunov type schemes for the two-dimensional linear wave equation with Coriolis source term on cartesian meshes*, October 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01618753.

[28] P. AUMOND, A. CAN, V. MALLET, B. DE COENSEL, C. RIBEIRO, D. BOTTELDOOREN, C. LAVANDIER.*Acoustic mapping based on measurements: space and time interpolation*, 2017, In: Proceedings of INTER-NOISE 2017, 46th International Congress and Exposition on Noise Control Engineering. 2017, pp. 5, 707–5, 718, https://hal.inria.fr/hal-01676009.

[29] N. AÏSSIOUENE, M.-O. BRISTEAU, E. GODLEWSKI, A. MANGENEY, C. PARÉS, J. SAINTE-MARIE.*A two-dimensional method for a dispersive shallow water model*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01632522.

[30] 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*, May 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01324012.

[31] P. LAURENT, G. LEGENDRE, J. SALOMON.*On the method of reflections*, February 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01439871.

[32] M. PARISOT.*Entropy-satisfying scheme for a hierarchy of dispersive reduced models of free surface flow, Part I*, September 2017, working paper or preprint, https://hal.inria.fr/hal-01242128.

References in notes

[33] N. AISSIOUENE.*Numerical analysis and discrete approximation of a dispersive shallow water model*, Univ. Pierre et Marie Curie, Paris 6, 2016.

[34] 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.

[35] N. AÏSSIOUENE, M.-O. BRISTEAU, E. GODLEWSKI, J. SAINTE-MARIE.*A combined finite volume-finite element scheme for a dispersive shallow water system*, in "Networks & Heterogeneous Media", 2016, vol. 11, n^{o} 1.

[36] 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.

[37] 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.

[38] E. GODLEWSKI, M. PARISOT, J. SAINTE-MARIE, F. WAHL.*Congested shallow water type model: roof modelling in free surface flow*, September 2016, working paper or preprint, https://hal.inria.fr/hal-01368075.

[39] 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.

[40] A. GREEN, P. NAGHDI.*A derivation of equations for wave propagation in water of variable depth*, in "J. Fluid Mech.", 1976, vol. 78, n^{o} 2, p. 237–246.

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
A2.1.9. - 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. - 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
B1.1.10. - Mathematical biology
B1.1.11. - Systems 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. Personnel

Research Scientists

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

Jerome Feret [Inria, Researcher]

Faculty Member

Patrick Cousot [New York University, Emeritus, until Nov 2017]

Technical Staff

Ferdinanda Camporesi [Inria]
Yves-Stan Le Cornec [Inria, from Sep 2017]
Jiangchao Liu [Inria, from Dec 2017]
Kim Quyen Ly [Inria]

PhD Students

Andreea Beica [Ecole Normale Supérieure Paris]
Marc Chevalier [Ecole Normale Supérieure Lyon, from Sep 2017]
Hugo Illous [Ecole Normale Supérieure Paris]
Huisong Li [Inria]
Thibault Suzanne [Ecole Normale Supérieure Paris]

Post-Doctoral Fellow

Changhee Park [Inria, until Oct 2017]

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 became 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), and medical systems (pacemakers, surgery and patient monitoring systems) rely on complex software, which should be verified.

Beyond the field of embedded systems, other pieces of software may cause very significant harm in 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 be 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 expressing 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* [22]. 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* [22], which over-approximates 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.

Yet, the same principles can also be applied successfully to other goals. In particular, the abstract interpretation framework should be viewed 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 [21], 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. The analysis of biological models

One of our application domains, the analysis of biological models, is not a classical target of static analysis because it aims at analyzing models instead of programs. Yet, the analysis of biological models is closely intertwined with the other application fields of our group. Firstly, abstract interpretation provides a formal understanding of the abstraction process which is inherent to the modeling process. Abstract interpretation is also used to better understand the systematic approaches which are used in the systems biology field to capture the properties of models, until getting formal, fully automatic, and scalable methods. Secondly, abstract interpretation is used to offer various semantics with different grains of abstraction, and, thus, new methods to apprehend the overall behavior of the models. Conversely, some of the methods and abstractions which are developed for biological models are inspired by the analysis of concurrent systems and by security analysis. Lastly, the analysis of biological models raises issues about differential systems, stochastic systems, and hybrid systems. Any breakthrough in these directions will likely be very important to address the important challenge of the certification of critical systems in interaction with their physical environment.

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 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, monitoring typically rely on a *parallel* structure, where several threads are executed in parallel, 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. Biological systems

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 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 abstract interpretation-based analyses, 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 so as to identify the key elements in the dynamics of these models. The results of these analyses are sent to another tool, which is used to automatically simplify the 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.

5. Highlights of the Year

5.1. Highlights of the Year

The team obtained several strong results published in excellent international conferences, with high theoretical and applied impact (see detailed results). Among the theoretical results we underline those presented in conferences like Principles of programming languages POPL 2017, with the proposal of a novel and groundbreaking way to improve the precision and scalability of analyses performed with disjunctive abstract domains, using silhouette abstraction.

5.1.1. Awards

Patrick Cousot received the IEEE John Von Neumann Medal.

6. New Software and Platforms

6.1. 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/

6.2. 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/

6.3. 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/

6.4. 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

6.5. 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

6.6. 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

6.7. 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

6.8. OPENKAPPA

La platte-forme de modélisation OpenKappa
KEYWORDS: Model reduction - Simulation - Static analysis - Modeling - Systems Biology
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.

- 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/

6.9. 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

6.10. 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

6.11. 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

7. New Results

7.1. Memory Abstraction

7.1.1. Abstraction of arrays based on non contiguous partitions

Participants: Jiangchao Liu, Xavier Rival [correspondant].

In [9], we studied array abstractions.

Array partitioning analyses split arrays into contiguous partitions to infer properties of cell sets. Such analyses cannot group together non contiguous cells, even when they have similar properties. We proposed an abstract domain which utilizes semantic properties to split array cells into groups. Cells with similar properties will be packed into groups and abstracted together. Additionally, groups are not necessarily contiguous. This abstract domain allows to infer complex array invariants in a fully automatic way. Experiments on examples from the Minix 1.1 memory management demonstrated its effectiveness.

7.1.2. Semantic-Directed Clumping of Disjunctive Abstract States

Participants: Huisong Li, Francois Berenger, Bor-Yuh Evan Chang, Xavier Rival [correspondant].

In [16], we studied the semantic directed clumping of disjunctive abstract states.

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.

7.1.3. Relational Inductive Shape Abstraction

Participants: Hugo Illous, Matthieu Lemerre, Xavier Rival [correspondant].

In [13], we studied a relational inductive shape abstract domain.

Static analyses aim at inferring semantic properties of programs. While many analyses compute an over-approximation of reachable states, some analyses compute a description of the input-output relations of programs. In the case of numeric programs, several analyses have been proposed that utilize relational numerical abstract domains to describe relations. On the other hand, designing abstractions for relations over memory states and taking shapes into account is challenging. In this paper, we propose a set of novel logical connectives to describe such relations, which are inspired by separation logic. This logic can express that certain memory areas are unchanged, freshly allocated, or freed, or that only part of the memory was modified. Using these connectives, we build an abstract domain and design a static analysis that over-approximates relations over memory states containing inductive structures. We implement this analysis and report on the analysis of a basic library of list manipulating functions.

7.2. Static Analysis of JavaScript Code

7.2.1. Weakly Sensitive Analysis for Unbounded Iteration over JavaScript Objects

Participants: Yoonseok Ko, Xavier Rival [correspondant], Sukyoung Ryu.

In [14], 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.

7.2.2. Revisiting recency abstraction for JavaScript: towards an intuitive, compositional, and efficient heap abstraction

Participants: Jihyeok Park, Xavier Rival [correspondant], Sukyoung Ryu.

In [18], we studied recency abstractions and their use for the analysis of JavaScript programs.

JavaScript is one of the most widely used programming languages. To understand the behaviors of JavaScript programs and to detect possible errors in them, researchers have developed several static analyzers based on the abstract interpretation framework. However, JavaScript provides various language features that are difficult to analyze statically and precisely such as dynamic addition and removal of object properties, first-class property names, and higher-order functions. To alleviate the problem, JavaScript static analyzers often use recency abstraction, which refines address abstraction by distinguishing recent objects from summaries of old objects. We observed that while recency abstraction enables more precise analysis results by allowing strong updates on recent objects, it is not monotone in the sense that it does not preserve the precision relationship between the underlying address abstraction techniques: for an address abstraction A and a more precise abstraction B, recency abstraction on B may not be more precise than recency abstraction on A. Such an unintuitive semantics of recency abstraction makes its composition with various analysis sensitivity techniques also unintuitive. In this paper, we propose a new singleton abstraction technique, which distinguishes singleton objects to allow strong updates on them without changing a given address abstraction. We formally define recency and singleton abstractions, and explain the unintuitive behaviors of recency abstraction. Our preliminary experiments show promising results for singleton abstraction.

7.3. Astrée and AstréeA

7.3.1. Finding All Potential Run-Time Errors and Data Races in Automotive Software

Participants: Antoine Miné, Laurent Mauborgne, Xavier Rival [correspondant], Jerome Feret, Patrick Cousot, Daniel Kästner, Stephan Wilhelm, Christian Ferdinand.

Safety-critical embedded software has to satisfy stringent quality requirements. All contemporary safety standards require evidence that no data races and no critical run-time errors occur, such as invalid pointer accesses, buffer overflows, or arithmetic overflows. Such errors can cause software crashes, invalidate separation mechanisms in mixed-criticality software, and are a frequent cause of errors in concurrent and multi-core applications. The static analyzer ASTRÉE has been extended to soundly and automatically analyze concurrent software. This novel extension employs a scalable abstraction which covers all possible thread interleavings, and reports all potential run-time errors, data races, deadlocks, and lock/unlock problems. When the analyzer does not report any alarm, the program is proven free from those classes of errors. Dedicated support for ARINC 653 and OSEK/AUTOSAR enables a fully automatic OS-aware analysis. In [15], we give an overview

of the key concepts of the concurrency analysis and report on experimental results obtained on concurrent automotive software. The experiments confirm that the novel analysis can be successfully applied to real automotive software projects.

7.4. Static analysis of signaling pathways

7.4.1. Formal and exact reduction for differential models of signaling pathways in rule-based languages

Participant: Ferdinanda Camporesi.

The behavior of a cell is driven by its capability to receive, propagate and communicate signals. Proteins can bind together on some binding sites. Post- translational modifications can reveal or hide some sites, so new interactions can be allowed or existing ones can be inhibited.

Due to the huge number of different bio-molecular complexes, we can no longer derive or integrate ODE models. A compact way to describe these systems is supplied by rule-based languages. However combinatorial complexity raises again when one attempt to describe formally the behavior of the models. This motivates the use of abstractions.

In this PhD thesis, we propose two methods to reduce the size of the models, that exploit respectively the presence of symmetries between sites and the lack of correlation between different parts of the system. The symmetries relates pairs of sites having the same capability of interactions. We show that this relation induces a bisimulation which can be used to reduce the size of the original model. The information flow analysis detects, for each site, which parts of the system influence its behavior. This allows us to cut the molecular species in smaller pieces and to write a new system. Moreover we show how this analysis can be tuned with respect to a context.

Both approaches can be combined. The analytical solution of the reduced model is the exact projection of the original one. The computation of the reduced model is performed at the level of rules, without the need of executing the original model.

7.4.2. Translating BNGL models into Kappa our experience

Participant: Kim Quyen Ly [correspondant].

So as to test the Kappa development tools on more examples, we translated the models provided with the BNGL distribution, into Kappa. In [20], we report about our experience. The translation was quite straightforward except for few interesting issues that we detail here. Firstly the use of static analysis has exposed some glitches in the modelling of some pathways in the models of the BNGL distribution. We explain how static analysis has helped us to detect, locate, and correct these flaws. Secondly, expanding BNGL rules using equivalent sites into rules with uniquely identified sites is not so easy when one wants to preserve faithfully the kinetics of interactions. We recall the semantics of BNGL for equivalent sites, and explain how to perform such translation.

7.4.3. Using alternated sums to express the occurrence number of extended patterns in site-graphs

Participants: Ferdinanda Camporesi, Jerome Feret [correspondant].

Site-graph rewriting languages as Kappa or BNGL supply a convenient way to describe models of signaling pathways. Unlike classical reaction networks, they emphasise on the biochemical structure of proteins. In [10], we use patterns to formalise properties about bio-molecular species. Intentionally, a pattern is a part of a species, but extensionally it denotes the multi-set of the species containing this pattern (with the multiplicity). Thus reasoning on patterns allows to handle symbolically arbitrarily big (if not infinite) multi-sets of species. This is a key point to design fast simulation algorithms or model reduction schemes. In this paper, we introduce the notion of extended patterns. Each extended pattern is made of a classical pattern and of a set of potential bonds between pairs of sites. Extended patterns have positive (when at least one of the potential bonds is

realised) and negative (when none is realised) instances. They are important to express the consumption and the production of patterns by the rules that may break cycles in bio-molecular species by side-effects. We show that the number of positive (resp. negative) instances of extended patterns may be expressed as alternated sums of the number of occurrences of classical patterns.

7.4.4. KaDE: a Tool to Compile Kappa Rules into (Reduced) ODE Models

Participants: Ferdinanda Camporesi, Jerome Feret [correspondant], Kim Quyen Ly.

In [11], we introduce the tool KaDe, that may be used to compile models written in Kappa in ODE. Kappa is a formal language that can be used to model systems of biochemical interactions among proteins. It offers several semantics to describe the behaviour of Kappa models at different levels of abstraction. Each Kappa model is a set of context-free rewrite rules. One way to understand the semantics of a Kappa model is to read its rules as an implicit description of a (potentially infinite) reaction network. KaDE is interpreting this definition to compile Kappa models into reaction networks (or equivalently into sets of ordinary differential equations). KaDE uses a static analysis that identifies pairs of sites that are indistinguishable from the rules point of view, to infer backward and forward bisimulations, hence reducing the size of the underlying reaction networks without having to generate them explicitly. In [11], we describe the main current functionalities of KaDE and we give some benchmarks on case studies. A complete tutorial and more complete benchmarks may be found at the following url: http://www.di.ens.fr/~feret/CMSB2017-tool-paper/.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

Xavier Rival received a Facebook Faculty Award (2017).

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. 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: Jerome 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.

9.1.2. 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.

9.1.3. VeriFault

This was a PEPS project for one year, coordinated by Cezara Drăgoi, on the topic of fault-tolerant distributed algorithms. These algorithms are notoriously difficult to implement correctly, due to asynchronous communication and the occurrence of faults, such as the network dropping messages or computers crashing. Although fault-tolerant algorithms are at the core of critical applications, there are no automated verification techniques that can deal with their complexity. Due to the complexity distributed systems have reached, we believe it is no longer realistic nor efficient to assume that high level specifications can be proved when development and verification are two disconnected steps in the software production process. Therefore we propose to introduce a domain specific language that has a high-level control structure which focuses on the algorithmic aspects rather than on low-level network and timer code, and makes programs amendable to automated verification.

9.1.4. TGFSYSBIO

Title: Microznvironment and cancer: regulation of TGF-β 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: Jerome Feret

Abstract: Most cases of hepatocellular carcinoma (HCC) develop in cirrhosis resulting from chronic liver diseases and the Transforming Growth Factor β (TGF-β) 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-β 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-β 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-β. 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-β-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-β. 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-β signal.

The expected results are the first model of extracellular and intracellular TGF-β system that might permit to analyze the behaviors of TGF-β activity during the course of liver tumor progression and to identify new biomarkers and potential therapeutic targets.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

ASSUME, ITEA 3 project (Affordable Safe & Secure Mobility Evolution). Affordable Safe & Secure Mobility Evolution

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. MemCad

Type: IDEAS

Defi: Design Composite Memory Abstract Domains

Instrument: ERC Starting Grant

Objectif: Design Composite Memory Abstract Domains

Duration: October 2011 - September 2016

Coordinator: Inria (France)

Partner: None

Inria contact: Xavier Rival

Abstract: The MemCAD project aims at setting up a library of abstract domains in order to express and infer complex memory properties. It is based on the abstract interpretation frameworks, which allows to combine simple abstract domains into complex, composite abstract domains and static analyzers. While other families of abstract domains (such as numeric abstract domains) can be easily combined (making the design of very powerful static analyses for numeric intensive applications possible), current tools for the analysis of programs manipulating complex abstract domains usually rely on a monolithic design, which makes their design harder, and limits their efficiency. The purpose of the MemCAD project is to overcome this limitation.

Our proposal is based on the observation that the complex memory properties that need to be reasoned about should be decomposed in combinations of simpler properties. Therefore, in static analysis, a complex memory abstract domain could be designed by combining many simpler domains, specific to common memory usage patterns. The benefit of this approach is twofold: first it would make it possible to simplify drastically the design of complex abstract domains required to reason about complex softwares, hereby allowing certification of complex memory intensive softwares by automatic static analysis; second, it would enable to split down and better control the cost of the analyses, thus significantly helping scalability. As part of this project, we propose to build a static analysis framework for reasoning about memory properties, and put it to work on important classes of applications, including large softwares.

9.3. International Initiatives

9.3.1. Participation in Other International Programs

9.3.1.1. EXEcutable Knowledge

Title: EXEcutable Knowledge

Type: DARPA

Instrument: DARPA Program

Program: Big Mechanism

Duration: July 2014 - December 2017

Coordinator: Harvard Medical School (Boston, USA)

Partner: Inria Paris-Rocquencourt, École normale supérieure de Lyon Université Paris-Diderot,

Inria contact: Jerome Feret

Abstract: Our overarching objective is Executable Knowledge: to make modeling and knowledge representation twin sides of biological reasoning. This requires the definition of a formal language with a clear operational semantics for representing proteins and their interaction capabilities in terms of agents and rules informed by, but not exposing, biochemical and biophysical detail. Yet, to achieve Executable Knowledge we need to go further:

- Bridge the gap between rich data and their formal representation as executable model elements. Specifically, we seek an intermediate, but already formal, knowledge representation (meta-language) to express granular data germane to interaction mechanisms; a protocol defining which and how data are to be expressed in that language; and a translation procedure from it into the executable format.
- Implement mathematically sound, fast, and scalable tools for analyzing and executing arbitrary collections of rules.
- Develop a theory of causality and attendant tools to extract and analyze the unfolding of causal lineages to observations in model simulations.

We drive these technical goals with the biological objective of assembling rule-based models germane to Wnt signaling in order to understand the role of combinatorial complexity in robustness and control.

9.3.1.2. Active Context

Title: Active Context

Type: DARPA

Instrument: DARPA Program

Program: Communicating with Computers

Duration: July 2015 - December 2018

Coordinator: Harvard Medical School (Boston, USA)

Partner: University of California, (San Diego, USA), Inria Paris-Rocquencourt, École normale supérieure de Lyon Université Paris-Diderot,

Inria contact: Jerome Feret

Abstract: The traditional approach to the curation of biological information follows a philatelic paradigm, in which epistemic units based on raw or processed data are sorted, compared and catalogued in a slow and all too often insufficiently coordinated process aimed at capturing the meaning of each specimen in isolation. The swelling bounty of data generated by a systematic approach to biology founded on high-throughput technologies appears to have only intensified a sense of disconnected facts, despite their rendering as networks. This is all the more frustrating as the tide of static data (sequences, structures) is giving way to a tide of dynamic data about (protein-protein) interaction that want to be interconnected and understood (think annotated) in terms of process, i.e. a systemic approach.

The barrier is the complexity of studying systems of numerous heterogeneously interacting components in a rapidly evolving field of science. The complexity comes from two kinds of dynamically changing context: the internal dynamics of a biological system, which provide the context for assessing the meaning of a protein-protein interaction datum, and the external dynamics of the very fact base used to define the system in the first place. We propose the integration of dynamic modeling into the practice of bioinformatics to address these two dynamics by coupling them. The external dynamics is at first handled by a novel kind of two-layered knowledge representation (KR). One layer contextualizes proteins and their interactions in a structure that incrementally constructs, in an open-ended dialogue with the user, its own semantics by piecing together fragments of knowledge from a variety of sources tapped by the Big Mechanism program. The other layer is a model representation (MR) that handles and prioritizes the many executable abstractions compatible with the KR. The internal dynamics is handled not only by execution but also by addressing the impedance mismatch between the unwieldy formal language(s) required for execution and the more heuristic, high-level concepts that structure the modeling discourse with which biologists reason about molecular signaling systems. To the extent that we are successful on both ends, users will be able to effectively deploy modeling for curating the very fact base it rests upon, hopefully achieving self-consistency.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

9.4.1.1. Internships

Xavier Rival supervised the internship of Guillaume Cluzel (L3, École Normale Supérieure de Lyon), on the implementation of array abstract domains.

Xavier Rival supervised the internship of Sixiao Zhu (M1, École Polytechnique), on the integration of a three valued abstraction in MemCAD.

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

Xavier Rival visited KAIST (Korean Advanced Institute for Science and Technology) as an Invited Professor in November/December 2017.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Selection

10.1.1.1. Chair of Conference Program Committees

- Jerome Feret served as co-Chair of CMSB 2017 (Computational Methods in Systems Biology).
- Xavier Rival is serving as Chair of the Artifact Evaluation Committee of SAS 2018 (Static Analysis Symposium).

10.1.1.2. Member of the Conference Program Committees

- Jerome Feret served as a Member of the Program Committee of TMPA 2017 (Tools and Methods of Program Analysis).
- Jerome Feret served as a Member of the Program Committee of JOBIM 2017 (Journées Ouvertes en Biologie, Informatique et Mathématiques).
- Jerome Feret served as a Member of the Program Committee of SASB 2017 (Static Analysis and System Biology).
- Jerome Feret is serving as a Member of the Program Committee of LICS 2018 (Logic in Computer Science).
- Jerome Feret is serving as a Member of the Program Committee of VEMDP 2018 (Verification of Engineered Molecular Devices and Programs).
- Jerome Feret is serving as a Member of the Program Committee of SAS 2018 (Static Analysis Symposium).
- Jerome Feret is serving as a Member of the Program Committee of CMSB 2018 (Computational Methods in Systems Biology).
- Xavier Rival was a Member of the Program Committee of SAS 2017 (Static Analysis Symposium).
- Xavier Rival is serving as a Member of the Program Committee of SAS 2018 (Static Analysis Symposium).
- Xavier Rival was a Member of the Program Committee of Web Design, Analysis, Programming and Implementation of the WWW'18 Conference.
- Xavier Rival was a Member of the Extended Review Committee of PLDI 2018 (Programming Languages Design and Implementation).
- Cezara Dragoi was a member of Programming languages design and implementation PLDI'17.
- Cezara Dragoi was a member of Computer Aided Verification CAV'18.

10.1.1.3. Reviewer

- Jerome Feret served as reviewer for CONCUR 2017 (Concurrency Theory).
- Jerome Feret served as reviewer for LICS 2017 (Logic in Computer Science).
- Jerome Feret served as reviewer for VMCAI 2017 (Verification, Model Checking, and Abstract Interpretation).

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- Jerome Feret is a member of the editorial board of the Frontiers in Genetics journal and the Open Journal of Modeling and Simulation.
- Jerome 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 2018.
- Jerome Feret serves as Editor of an Issue of the IEEE/ACM Transactions on Computational Biology and Bioinformatics, that is composed of papers from CMSB 2016, 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 is expected to appear in 2018.

10.1.2.2. Reviewer - Reviewing Activities

- Jerome Feret served as a Reviewer for FMSD (Formal Methods in System Design).
- Jerome Feret served as a Reviewer for TCS (Theoretical Computer Sciences).
- Jerome Feret served as a Reviewer for TCBB (IEEE/ACM Transactions on Computational Biology and Bioinformatics).
- Jerome Feret served as a Reviewer for TCL (Transactions on Computational Logic).
- 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).

10.1.3. Invited Talks

- Jerome Feret gave a talk on automatic reduction of models of intra-cellular signaling pathways at the Dagstuhl Seminar on Algorithmic Cheminformatics (5–10 November 2017).
- Cezara Dragoi was invited speaker at the 6th South of England Regional Programming Language Seminar University College London, Workshop on Software Correctness and Reliability at ETH Zurich, and Workshop on Formal Reasoning in Distributed Algorithms (FRIDA), Wien Austria.

10.1.4. Leadership within the Scientific Community

Xavier Rival is a member of the IFIP Working Group 2.4 on Software implementation technology.

10.1.5. Research Administration

Jerome Feret and Xavier Rival are members of the Laboratory Council of DIENS.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence:

- Jerome Feret, and Marc Chevalier, Mathematics, 40h, L1, FDV Bachelor program (Frontiers in Life Sciences (FdV)), Université Paris-Descartes, France.
- Jerome Feret and Xavier Rival, “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.
- Cezara Dragoi “Les principes des langages de programmation”, 18h, L1, at École Polytechnique, France.

Master:

- Xavier Rival, “Verification” Lab Course, 20h, M1, École Polytechnique, France.
- Xavier Rival, “Protocol Safety and Verification”, 12h, M2, Advanced Communication Networks Master, France.
- Xavier Rival, “Program Analysis”, 24h, M2, Korea Advanced Institute for Science and Technology (KAIST), South Korea.
- Cezara Drăgoi, Jerome 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 Jerome Feret (with Jean Krivine), Computational Biology, 28h, M1. Interdisciplinary Approaches to Life Science (AIV), Master Program, Université Paris-Descartes, France.

10.2.2. Supervision

- PhD defended: Ferdinanda Camporesi, Formal and exact reduction for differential models of signaling pathways in rule-based languages. Defended the 23th of January, 2017 and supervised by Jerome Feret.
- PhD in progress: Marc Chevalier, Static analysis of Security Properties in Critical Embedded Software. started in 2017 and supervised by Jerome 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: Huisong Li, Disjunctive Shape Abstraction for Shared Data-Structures, started in 2014 and supervised by Xavier Rival
- PhD in progress: Jiangchao Liu, Static Analysis for Numeric and Structural Properties of Array Contents, started in 2014 and supervised by Xavier Rival

10.2.3. Juries

- Jerome Feret served as a member of the Jury of the PhD of Jean Coquet (Defended the 20th of December, 2017).
- 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).

10.2.4. Responabilities

- Jerome Feret is a member of the Monotoring Committee for PhD Studies (CSD) of Inria Paris.
- Jerome Feret is deputy head of studies of the Computer Science department of École normale supérieure.

10.2.5. Selection committees

- Jerome Feret was a member of the recruitment committee for an assistant professor in Evry University 2017.
- Jerome Feret is a member of the recruitment committee for an assistant professor in Paris-Diderot University 2018.

11. 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. DRĂGOI, 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, http://doi.acm.org/10.1145/1993498.1993566.

[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

Articles in International Peer-Reviewed Journal

[9] J. LIU, X. RIVAL.*An array content static analysis based on non-contiguous partitions*, in "Computer Languages, Systems and Structures", 2017, vol. 47, n^{o} 1, p. 104–129 [*DOI :* 10.1016/J.CL.2016.01.005], https://hal.inria.fr/hal-01399837.

International Conferences with Proceedings

[10] F. CAMPORESI, J. FERET.*Using alternated sums to express the occurrence number of extended patterns in site-graphs*, in "SASB 2017 - The Eighth International Workshop on Static Analysis for Systems Biology", New York, United States, J. YANG, J. A. BACHMAN (editors), Static Analysis and Systems Biology, Elsevier, August 2017, 18, To appear, https://hal.inria.fr/hal-01613603.

[11] F. CAMPORESI, J. FERET, K. Q. LY.*KaDE: A Tool to Compile Kappa Rules into (Reduced) ODE Models*, in "CMSB 2017 - 15th Conference on Computational Methods in Systems Biology", Darmstadt, Germany, J. FERET, H. KOEPPL (editors), Computational Methods in Systems Biology, Springer, September 2017, vol. 10545, p. 291-299, Tools paper track [*DOI :* 10.1007/978-3-319-67471-1_18], https://hal.inria.fr/hal-01613600.

[12] C. DRĂGOI, T. HENZINGER, D. ZUFFEREY.*PSYNC: A Partially Synchronous Language for Fault-Tolerant Distributed Algorithms*, in "POPL", Saint Petersburg , United States, January 2017, https://hal.inria.fr/hal-01434325.

[13] H. ILLOUS, M. LEMERRE, X. RIVAL.*A Relational Shape Abstract Domain*, in "NFM 2017 - 9th NASA Formal Methods Symposium", Moffett Field, United States, LNCS, Springer, April 2017, vol. 10227, p. 212-229 [*DOI :* 10.1007/978-3-319-57288-8_15], https://hal.inria.fr/hal-01648681.

[14] 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.

[15] D. KÄSTNER, A. MINÉ, A. SCHMIDT, H. HILLE, L. MAUBORGNE, S. WILHELM, X. RIVAL, J. FERET, P. COUSOT, C. FERDINAND.*Finding All Potential Run-Time Errors and Data Races in Automotive Software*, in "SAE world Congress", Detroit , United States, SAE Technical Paper, SAE International, April 2017, 9, https://hal.inria.fr/hal-01674831.

[16] 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.

[17] R. MONAT, A. MINÉ.*Precise Thread-Modular Abstract Interpretation of Concurrent Programs Using Relational Interference Abstractions*, in "Verification, Model Checking, and Abstract Interpretation (VMCAI) 2017", Paris, France, A. BOUAJJANI, D. MONNIAUX (editors), Lecture Notes in Computer Science, Springer, January 2017, vol. 10145, p. 386-404 [*DOI :* 10.1007/978-3-319-52234-0_21], https://hal.inria.fr/hal-01490178.

[18] J. PARK, X. RIVAL, S. RYU.*Revisiting Recency Abstraction for JavaScript Towards an Intuitive, Compositional, and Efficient Heap Abstraction*, in "SOAP 2017 - International Workshop on the State Of the Art in Java Program Analysis", Barcelona, Spain, June 2017, p. 1-6 [*DOI :* 10.1145/3088515.3088516], https://hal.inria.fr/hal-01648682.

Books or Proceedings Editing

[19] J. FERET, H. KOEPPL (editors). *Computational Methods in Systems Biology : 15th International Conference, CMSB 2017, Darmstadt, Germany, September 27–29, 2017, Proceedings*, Lecture Notes in Bioinformatics, Springer, France, September 2017, vol. 10545, 332 [*DOI :* 10.1007/978-3-319-67471-1], https://hal.inria.fr/hal-01613596.

Other Publications

[20] K. Q. LY.*Translating BNGL models into Kappa our experience*, August 2017, 4, Extended abstract, https://hal.inria.fr/hal-01613604.

References in notes

[21] 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.

[22] 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.

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
A2.1.8. - Synchronous languages
A2.1.10. - Domain-specific languages
A2.2.3. - Run-time systems
A2.2.4. - Parallel architectures
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. Personnel

Research Scientists

Liliana Cucu [Inria, Researcher, HDR]
Robert Davis [University of York UK, Chair]
Dumitru Potop-Butucaru [Inria, Researcher, HDR]
Yves Sorel [Inria, Senior Researcher]

External Collaborator

Laurent George [Univ Paris-Val de Marne, HDR]

Technical Staff

Irina-Mariuca Asavoae [Inria, until Aug 2017]
Mihail Asavoae [Inria, until Apr 2017]
Antoine Bertout [Inria, until Jul 2017]
Adriana Gogonel [Inria]
Fatma Jebali [Inria]
Tomasz Kloda [Inria, until Jul 2017]
Mehdi Mezouak [Inria]

PhD Students
Slim Ben-Amor [Inria]
Keryan Didier [Inria]
Cristian Maxim [Airbus]
Salah-Eddine Saidi [IFPEN]
Evariste Ntaryamira [Inria, Embassy of France at Burundi]
Walid Talaboulma [Inria]

Visiting Scientist
George Lima [Inria, from May 2017 until Jun 2017]

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 design 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 the functions of the functional specification 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 for every function its worst case execution times (WCET) and for every dependence its worst case communication times (WCCT). Both worst case execution and communication times are dependent of the platform. Using these worst case times, schedulability analyses are able to compute worst case response times and end-to-end worst case execution times in order to verify if real-time constraints, e.g. deadline, imposed by automatic control engineers, are met. Note that, unfortunately, automatic control engineers define these constraints whereas 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, their means of communication, 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 railways and avionics applications, the estimation of worst case times is relatively easy. For this purpose we use static analyses or techniques based on measurements for the WCETs for example. 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 (tasks) 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 [16]. 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 necessary 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 [18]. 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 [10].
- *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 [4]. 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 [13], 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 [3] 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 CAPACITES and ASSUME projects 9.1.2.1, 9.2.1.1).

4.2. Many-Core Embedded Architectures

Participants: Liliana Cucu, Keryan Didier, Dumitru Potop-Butucaru, Anselme Revuz, 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 CAPACITES and ASSUME projects 9.1.2.1, 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

Our team has hosted for the first time in France the 38th Real-Time Systems Symposium (RTSS'17) which is the flag conference of our research domain. All the members of team jointly participated to the big effort of ensuring an excellent edition.

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: Slim Ben-Amor, Liliana Cucu, Robert Davis, Mehdi Mezouak, Yves Sorel.

In the framework of the FUI CEOS project 9.1.1.1 we mainly investigated the PX4 autopilot free software program that was chosen by the partners to be implemented on the Pixhawk electronic board. This board will be installed in the multirotor drone that the project is intended to built. The board is based on a microcontroller which contains an ARM Cortex M4 microprocessor, timers, several sensors, accelerometer, gyroscope, magnetometer, barometer, and actuators, mainly four to eight electric motors depending on the level of redundancy.

We studied the existing source code of PX4 which consists of two main layers: the flight stack, which is an estimation and flight control system, and the middleware, which is a general robotics layer providing internal/external communications and hardware integration. This study allowed us to understand the general architecture of PX4. The flight stack is split into a set of threads communicating asynchronously through a micro object request broker messaging. In the CEOS project our team is in charge to guarantee that the drone will satisfy multiple real-time criticality levels. In order to be able to perform a real-time schedulability analysis on the PX4 autopilot, first we transformed this set of communicating threads into a task dependency graph. Second, we sought the period of each task starting from input tasks which read from sensors, to output tasks which write into actuators. The partners of the project chose to run PX4 on the NuttX OS which is open source, light-weight, efficient and very stable. It provides POSIX API and some form of real-time scheduling. Thus, we had to deeply understand the scheduler and the management of interruptions and time of NuttX. We plan to modify NuttX in order to support mixed-criticality applications using to start, online real-time scheduling, and then offline real-time scheduling.

Finally, always to perform the real-time schedulability analysis of PX4, we must estimate the worst execution time (WCET) of each task. This problem is very 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 probabilistic reasoning in characterizing the timing behavior (WCET, schedulability analyses) of mixed-criticality real-time applications [4].

7.2. Multiprocessor Real-Time Scheduling

Participants: Salah-Eddine Saidi, Yves Sorel.

During the third year of the PhD thesis of Salah Eddine Saidi, we focused on two aspects. First, we finalized our work on the parallelization on multi-core processors of FMI-based co-simulation of numerical models in order to accelerate its execution. Our approach, based on the transformation of FMU graphs into operation graphs which reveal more parallelism, comprises the following two steps: first acyclic orientation necessary for avoiding that some operations of a same model are executed in parallel and second multi-core offline scheduling of operations [5]. We proposed exact algorithms based on ILP (Integer Linear Programming) and heuristics for performing the acyclic orientation and the multi-core scheduling. Also, we proposed a random generator of synthetic co-simulations. Using these generated co-simulations, we compared the performances of the heuristics and the ILP-based exact algorithm for both the acyclic orientation and the scheduling in terms of execution time and quality of the obtained solution. Tests have been carried out for different sizes of co-simulation and different numbers of cores. Moreover, we compared the performance of our offline approach with an online scheduling approach based on the Intel TBB runtime library. This comparison was acheived by applying both approaches on an industrial use case which consists in a co-simulation of a four cylinder spark ignition engine. The various tests that we performed showed the efficieny of our proposed heuristics. Second, we focused on the parallelization of FMI-based co-simulation under real-time constraints. In particular, we were interested in HiL (Hardware-in-the-Loop) co-simulation where a part of the co-simulation is replaced by its real counterpart that 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 and ouputs of the real part are sampled periodically, sending and receving data to and from the simulated part. This 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 operations of the co-simulation (simulated part). In our ongoing work, we are focusing on multi-core scheduling of FMI-based co-simulation under real-time constraints. More precisely, we are working on a heuristic and an ILP-based algorithm that will enable the execution of the co-simulation on a multi-core processor while ensuring the defined real-time constraints are respected.

7.3. Principles of Probabilistic Composition

Participants: Slim Ben-Amor, Liliana Cucu, Adriana Gogonel, Cristian Maxim.

The statistical estimation of time parameters for real-time systems is proposed at two levels:

1. at program level and in this case we are dealing with timing analysis of programs that requires later appropriate probabilistic composition principles like reproducibility and representativity [3], [1]. For instance we have underlined in [14] the difficulties to ensure such properties for many-cores architectures.

   While we are proposing static analyses using worst-case bounds on the execution at instruction level for specialized architectures [2], we are interested also in proposing composition principles allowing to combine the timing impact of execution time variation factors, identified as a key open problem in the context of the timing analysis of programs while using the Extreme Value Theory [1]. Our composition solution is based on a Bayesian modeling that considers iteratively the inclusion of new factors while a representative measurement protocole is built [13] with respect to the reproducible Extreme Value Theory-based estimator that we have proposed.

2. at system level and in this case we are dealing with schedulability analysis of set of programs, a.k..a tasks, that requires appropriate composition principles like probabilistic independence while the dependence between tasks is taken into account. After proposing a first solution to the schedulability analysis of real-time probabilistic tasks in presence of precedence constraints on uniprocessor system [6], we explore the state of art of real-time scheduling on multiprocessor system and probabilistic real-time existing analysis. Our choice goes to partitioned multiprocessor scheduling to ensure the applicability of our previous results in the case of one processor. We have proposed a first optimal partitioning strategy based individual task utilization and we compare different tasks combinations that fit on a single processor following an utilization task ratio principle as partitioning choice. When assessing our method, a counter example of a possible optimality has appeared. Moreover this method has not an important improvement compared to existing partitioning strategies like best fit. Therefore we prepare the application of an existing solution to the bin packing problem [17] proposed in mathematics domain to partition real-time tasks on multiprocessor system in order to propose an appropriate probabilistic analysis.

   The exact schedulability analyses are often competing with statistical estimation of response time based on simulation and we propose such result in [9]. Such results allow to advance on the understanding of the notion of representativeness in the context of our problem that becomes today central in our community. The explosion of probabilistic schedulability analyses published in the last years have convinced us to join the book proposal of a Handbook on Real-Time Computing in order to integrate a comprehensive description of these analyses [4].

7.4. pWCET Estimation: a System Concern

Participants: Irina-Mariuca Asavoae, Mihail Asavoae, Slim Ben-Amor, Antoine Bertout, Liliana Cucu, Adriana Gogonel, Tomasz Kloda, Cristian Maxim, Walid Talaboulma.

From modelling to time validation, the design of an embedded system may benefit from a better utilisation of probabilities while providing means to prove their results. The arrival of new complex processors has made the time analysis of the programs more difficult while there is a growing need to integrate uncertainties from all levels of the embedded systems design. Probabilistic and statistical approaches are one possible solution and they require appropriate proofs in order to be accepted by both scientific community and industry. Such proofs cannot be limited at processor or program level and we plead for a system approach in order to take into account the possible interactions between different design levels by using the probabilistic formulation as compositional principle.

Our first arguments are provided by a valid statistical estimation of bounds on the execution time of a program on a processor. More precisely, the probabilistic worst-case execution time (pWCET) $\mathcal{C}$ of a program is an upper bound on all possible probabilistic execution times $\mathcal{C}_i$ for all possible execution scenarios $S_i, \forall i \geq 1$. According to EVT if the maximum of execution times of a program converges, then this maximum of the execution times $\mathcal{C}_i, \forall i \geq 1$ converges to one of the three possible Generalized Extreme Value (GEV) laws: Fréchet, Weibull and Gumbel corresponding to a shape parameter $\xi > 0$, $\xi < 0$, and $\xi = 0$, respectively. EVT has two different formulations: Generalized Extreme Value (GEV) and Generalized Pareto Distribution (GPD) and the difference between them is the way the extreme values are selected. GEV is based on the block maxima reasoning, grouping execution times by chronological groups (called blocks) and only the largest value of each group is considered as an extreme value. GPD is a method based on the threshold approach that considers only the values larger than the chosen threshold as extreme values. The voting procedure is based on the utilization of the both formulations of the EVT.

- **Block size estimation :** The GEV models obtained for different block sizes (BS), BS from 10 to $\frac{n}{10}$ are compared, where n is the cardinal of the trace of execution times. We compare the models fitting the extreme values corresponding to each choice of BS and the evolution of the shape parameter function of BS. We keep the BS that assures the best compromise between fitting the data and having a shape parameter within a stability interval of a range of shape parameters estimates. The way GEV models fit the data is analyzed within the tool by using a graphical method including the qqplot and the return level plot. We keep the GEV model corresponding to the shape parameter as the result of the aforementioned compromise and we compute the pWCET as the $1 - CDF$ (inverse of the cumulative distribution function) of the GEV.
- **Threshold level estimation :** The procedure is similar to the GEV procedure. All GPD models obtained for different threshold levels from 80% to 99%, are compared. In the same way as for GEV, we compare the models fitting the extreme values corresponding to each threshold and the evolution of the shape parameter function of threshold. At the end we keep the threshold level assuring the best compromise between fitting the data (graphical method) and having the shape parameter within a stability interval of a range of shape parameters estimates. We also consider the mean residual life plot (mean of excess) that may be consulted in case of a doubt between two different thresholds, we will prefer the threshold level such that the curve of mean of excess experiences linearity. We keep the GPD model corresponding to the shape parameter resulting from the aforementioned compromise and we compute the pWCET as the $1 - CDF$ of the GPD.
- **Comparing GEV and GPD pWCET estimates :** The comparison of the pWCET obtained with both methods, GEV and GPD is done graphically. Superposing the two curves allows to compare the distance between the two distributions. If an important difference is noticed, other GEV/GPD models are tested. In such cases calculating the pWCET estimate as a combination of GEV and GPD results is also recommended. A joint pWCET estimate is obtained by choosing for each probability the largest value between GEV and GPD . The tool implementing this method is available on line at `inria-rscript.serveftp.com` (requires a secured connection to be provided under request) [8].
- **Conditions of use :** The application of EVT requires to verify that the analyzed data are identically distributed, i.e., the execution times are following the same (unknown) probability distribution. That condition is tested before the analysis is started, and data is treated according to the test results. Another EVT applicability condition is the independence of the data. That condition is not mandatory in the sense that non-independent data can be analyzed. The case of dependent data can be split in two sub cases. The first one is where there are dependencies within the data, still the picked extremes values are independent. In that case the analysis will be done in the same way as for the independent data. The second case is the one where there are dependencies also between the extreme values. In that case one more step is added in the procedure. This step is the de-clustering process before applying GPD and the use of the index while GEV is applied.

During the second year of PhD thesis of Talaboulma Walid, we continued exploring solutions to WCET (Worst Case Execution Time) estimation and Real Time Scheduling on multiprocessors. WCET analysis done on a monoprocessor system (in isolation) can no longer be trusted to be accurate when we run our tasks on a

multiprocessor (two processors), the problem of Co-runner interference arises and this is due to contention in shared hardware, two processors share the same memory and contention will occur when a simultaneous access is done, thus delaying one of the request, and this can counter-intuitively make programs run longer in a multiprocessor than what the analysis predicted on a monoprocessor, leading to deadline misses. In [20] authors evaluate explicit reservation of cache memory to reduce the cache-related preemption delay observed when tasks share a cache in a preemptive multitasking hard real-time system. Another solution is presented in [19] by management of tasks shared resources access using performance counter to stop tasks when they exceed their allocated budget (for instance cache misses) and thus providing guarantees on global memory bandwidths, moreover in [15] some offline analysis is done using heuristics to find optimal time triggered schedules for shared memory access.

We propose in our work to generate 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 approach doesn't necessitate modifications of software layer, or recompilation of task code First we focus on simple tasks with few branches and simple memory access patterns as a proof of concept, and we choose a COTS (component of the shelf) platform with two complex processor cores. We intend to loosen those constraints when our analysis is matured. We use those profiles to account for co runners interference and add it to WCET value obtained in isolation, and then update our schedule, we can also insert idle times at correct scheduling events to decrease this interference, and in the future use a modified memory management system to pre-load specific memory areas into the cache and thus slide those access back in time to eliminate simultaneous memory access and converge toward an isolation WCET value.

7.5. 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, and Kalray.

Concurrent programming is notoriously difficult, especially in constrained embedded contexts. Threads, in particular, are wildly nondeterministic as a model of computation, and difficult to analyze in the general case. Fortunately, it is often the case that multi-threaded, semaphore-synchronized embedded software implements high-level functional specifications written in a deterministic data-flow language such as Scade or (safe subsets of) Simulink.

In many cases, the multi-threaded implementation of such specifications preserves a fundamentally dataflow structure, with specific rules on the way platform resources (shared memory, semaphores) are used. When this happens, the implementation is best represented as a dataflow synchronous program whose elements are mapped on the platform resources. Ensuring the correctness of such an implementation consists in ensuring that:

1. The dataflow program (without the mapping) implements the semantics of the functional specification. This analysis can be performed inside the dataflow model.
2. Once the mapping of program elements onto the platform resources [0] is performed, the execution of the platform (under platform semantics) implements the behavior of the dataflow program.

Together, the dataflow program and the mapping information form an *implementation model*. This model is strictly richer than the multi-threaded C code, which can be obtained through a pretty-printing of model parts. Exposing the internal data-flow structure of the implementation facilitates defining and establishing correctness, *e.g.* the correctness of the synchronization or memory coherence protocols synthesized during the implementation process. All analyses can be realized using efficient tools specific to the synchronous model. Finally, if manual inspection of the C multi-threaded code is required, such a representation can be used to enforce strict code structuring rules which facilitate understanding.

[0] Sequencing of blocks into threads executed by processors; code, stack and data variables to memory locations; synchronizations to semaphores, *etc.*

We proposed a language for describing such implementation models that expose the data-flow behavior hiding under the form of a multi-threaded program. The language allows the representation of efficient implementations featuring pipelined scheduling and optimized memory allocation and synchronization [12].

We also proposed a design and tool flow taking as input industrial specifications based on Lustre/Scade and automatically producing fully mapped parallel implementation models and implementations with hard real-time guarantees. The front-end of the flow implements properties facilitating the mapping, e.g., exposing the state of all nodes to memory optimization. To strictly enforce realtime guarantees, the offline mapping algorithms of the back-end consider all sources of interference, including concurrent memory accesses, coherence protocols and event-driven synchronization. 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.

7.6. Real-time Platform Modeling

Participants: Fatma Jebali, Dumitru Potop-Butucaru.

One key difficulty in embedded systems design is related to the existence of multiple models of the same system, at different abstraction levels, and used in various phases of the design flow. Usual models include *cycle-accurate, bit-accurate (CABA)* system models used to perform exact simulation for precision tuning, microarchitectural models used during WCET (*Worst-Case Execution Time*) analysis of sequential tasks, and high-level models used during WCRT (*Worst-Case Response Time*) analysis of the whole system. In current practice, these models are developed separately, and it is difficult to ensure (by extensive simulation) that they are consistent.

We explore the possibility of obtaining both a CABA and a WCET microarchitectural simulator from a single source, along with a formal consistency guarantee. This year we considered the timing abstraction issue: Both CABA and WCET simulators use a cycle-based execution model, but the cycle corresponds in one case to hardware clock cycles, and in the other to PC (program counter) advancement. We showed that for architectures satisfying a scheduling-independence property (known as in-order architectures) it is possible to produce from a single source both types of simulations (clock-driven and PC-driven), with a formal correctness guarantee. Preliminary results have been presented at the Synchron'07 workshop.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

The Airbus CIFRE grant which started on March 2014, provides full support for the PhD thesis of Cristian Maxim. The thesis concerns the statistical timing analysis while different variability factors are taken into account. The proposed methods are built on top of existing statistical approaches while proving appropriate programs for training these methods and thus learning from the history of the execution.

8.2. Bilateral Grants with Industry

The IFPEN grant which started on December 2014, 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, 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: Antoine Bertout, Liliana Cucu, Adriana Gogonel, Tomasz Kloda, Yves Sorel, Walid Talaboulma.

This project was started on September 2015. It targets the creation of a framework allowing to connect different existing methods while enriching the description with Waruna results. This framework allows timing analyses for different application domains like avionics, railways, medical, aerospace, automotive, etc.

9.1.2. PIA

9.1.2.1. CAPACITES

Participants: Liliana Cucu, Cristian Maxim, Dumitru Potop-Butucaru, Yves Sorel, Walid Talaboulma.

This project is funded by the LEOC Call (Logiciel Embarqué et Objets Connectés) of the national support programme Investissements d'Avenir. It was started on November 1st, 2014 with the kick-off meeting held on November, 12th 2014. The project cordinator is Kalray, and the objective of the project is to study the relevance of Kalray-style MPPA processor array for real-time computation in the avionic domain (with partners such as Airbus for instance). The PhD of Walid Talaboulma is funded on this contract.

9.1.2.2. DEPARTS

Participants: Liliana Cucu, Adriana Gogonel, Walid Talaboulma.

This project is funded by the BGLE Call (Briques Logicielles pour le Logiciel Embarqué) of the national support programme Investissements d'Avenir. Formally started on October 1st, 2012 with the kick-off meeting held on April, 2013 for administrative reasons. Research will target solutions for probabilistic component-based models, and a Ph.D. thesis should start at latest on September 2015. The goal is to unify in a common framework probabilistic scheduling techniques with compositional assume/guarantee contracts that have different levels of criticality.

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 a joint publication.

9.3. International Research Visitors

9.3.1. Visits of International Scientists

Professor George Lima (University of Baha, Brazil) visited us between May and June. His stay was dedicated the study of the utilization of extreme value theory on the problem of probabilistic estimation of worst case execution time bounds for a program on a processor.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Steering Committees

- 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.

10.1.1.2. Member of the Organizing Committees

- Liliana Cucu is Local Arrangement Chair of the 38th IEEE Real-time Systems Symposium (RTSS'17).

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- Liliana Cucu: RTAS, RTNS, WFCS
- Robert Davis: RTSS, RTAS, RTNS
- Adriana Gogonel: ACM RACS, WMC
- Dumitru Potop-Butucaru: ACSD, EMSOFT
- Yves Sorel: DASIP

10.1.2.2. 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. Reviewer - Reviewing Activities

All members of the team are regularly serving as reviewers for the main journals of our domain: Journal of Real-Time Systems, Journal of Systems Architecture, Leibniz Transactions on Embedded Systems, IEEE Transactions on Industrial Informatics, etc.

10.1.4. Invited Talks

- Liliana Cucu is keynote speaker at the 11th edition of CRTS, invited speaker at MMR'17 and MEFOSYLOMA seminar.

10.1.5. Leadership within the Scientific Community

Liliana Cucu and Rob Davis are the scientific organizers of the 2nd Dagstuhl seminar on mixed-criticality systems.

10.1.6. 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).

10.1.7. Research Administration

- Liliana Cucu-Grosjean is member of Inria Evaluation Commission, co-chair of Inria Committes 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: Liliana Cucu, Distributed Databases and Statistics in Computer Science, 64h, U. Dunarea de Jos, Romania (Invited Professor).

Master: Dumitru Potop Butucaru, A synchronous approach to the design of embedded real-time systems, 30h, M1, EPITA Engineering School, Paris France.

Master: Yves Sorel, Optimization of distributed real-time embedded systems, 38H, M2, University of Paris Sud, France.

Master: Yves Sorel, Synchronous languages and real-time scheduling, 9H, M2, University of Paris-Est Créteil, France.

Master: Yves Sorel, Correct by construction design of reactive systems, 18H, M2, ESIEE Engineering School, Noisy-Le-Grand, France.

10.2.2. Supervision

PhD: Cristian Maxim, End to end constraints using probabilistic approaches, UPMC, defended December 2017, supervised by Liliana Cucu.

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, Université Pierre et Marie Curie/EDITE, started November 2015, supervised by Dumitru Potop Butucaru.

PhD in progress: Cristian Maxim, End to end constraints using probabilistic approaches, UPMC, started March 2014, supervised by Liliana Cucu.

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, University of UMPC-Paris-Sorbonne, started December 2014, co-supervised by Nicolas Pernet (IFPEN) and Yves Sorel.

10.2.3. Juries

- Liliana Cucu is Phd reviewer for the thesis of Fabrice Guet, ONERA and ISAE, defended December 2017.
- Liliana Cucu is Phd reviewer for the thesis of Bader Alahmad, University of British Columbia, defended December 2017.
- Liliana Cucu is Phd jury member for the thesis of Romain Gratia, Telecom Paritech, defended January 2017.

10.3. Popularization

Popularization video of the probabilistic notions for mixed-criticality systems https://www.youtube.com/watch?v=sSJT4eGhS_A

11. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] S. J. GIL, I. BATE, G. LIMA, L. SANTINELLI, A. GOGONEL, L. CUCU-GROSJEAN.*Open Challenges for Probabilistic Measurement-Based Worst-Case Execution Time*, in "IEEE Embedded Systems Letters", June 2017, vol. 9, n^{o} 3, p. 69 - 72 [*DOI :* 10.1109/LES.2017.2712858], https://hal.archives-ouvertes.fr/hal-01633802.

[2] 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 / Real Time Systems; The Journal of Real-Time Systems", 2017, p. 1-82 [*DOI :* 10.1007/s11241-017-9295-2], https://hal.archives-ouvertes.fr/hal-01666091.

[3] 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.

International Conferences with Proceedings

[4] D. I. MAXIM, R. DAVIS, L. I. 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.

[5] S. E. SAIDI, N. PERNET, Y. SOREL.*Automatic parallelization of multi-rate fmi-based co-simulation on multi-core*, in "TMS/DEVS 2017 - Symposium on Theory of Modeling and Simulation", Virginia Beach, United States, ACM, April 2017, Article No. 5, https://hal.inria.fr/hal-01610268.

Conferences without Proceedings

[6] S. BEN-AMOR, D. MAXIM, L. CUCU.*Schedulability analysis of dependent probabilistic real-time tasks* , in "MAPSP 2017 - 13th Workshop on Models and Algorithms for Planning and Scheduling Problems", Seeon-Seebruck, Germany, RTNS '16 Proceedings of the 24th International Conference on Real-Time Networks and Systems, ACM, June 2017, p. 99-107 [*DOI :* 10.1145/2997465.2997499], https://hal.archives-ouvertes.fr/hal-01666138.

[7] L. CUCU-GROSJEAN, A. GOGONEL.*Probabilistic foundations for the time predictions of cyber-physical systems*, in "MMR 2017 - 10th International Conference on Mathematical Methods in Reliability", Grenoble, France, July 2017, https://hal.archives-ouvertes.fr/hal-01666293.

[8] A. GOGONEL, C. MAXIM, L. CUCU-GROSJEAN.*pWCET estimator for real-time systems*, in "RTSS 2017 - IEEE Real-Time Systems Symposium", Paris, France, December 2017, https://hal.archives-ouvertes.fr/hal-01666342.

[9] D. MAXIM, A. BERTOUT.*Analysis and Simulation Tools for Probabilistic Real-Time Systems*, in "8th International Workshop on Analysis Tools and Methodologies for Embedded and Real-time Systems (WATERS)", Dubrovnik, Croatia, June 2017, https://hal.archives-ouvertes.fr/hal-01552798.

Scientific Books (or Scientific Book chapters)

[10] 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.

Books or Proceedings Editing

[11] L. CUCU-GROSJEAN, R. DAVIS, S. K. BARUAH, Z. STEPHENSON (editors). *Mixed Criticality on Multicore / Manycore Platforms (Dagstuhl Seminar 17131)*, Schloss Dagstuhl, 2017, p. 70-98 [*DOI :* 10.4230/DAGREP.7.3.70], https://hal.archives-ouvertes.fr/hal-01666118.

Research Reports

[12] K. DIDIER, A. COHEN, A. GAUFFRIAU, A. GRAILLAT, D. POTOP-BUTUCARU.*Sheep in wolf's clothing: Implementation models for data-flow multi-threaded software*, Inria Paris, April 2017, n^{O} RR-9057, 31, https://hal.inria.fr/hal-01509314.

Patents and standards

[13] 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.

Other Publications

[14] A. REVUZ, L. CUCU-GROSJEAN.*Towards statistical estimation of worst case inter-core communications*, October 2017, JRWRTC 2017 - 11th Junior Researcher Workshop on Real-Time Computing, Poster, https://hal.inria.fr/hal-01666243.

References in notes

[15] M. BECKER, D. DASARI, B. NIKOLIC, B. AKESSON, V. NÉLIS, T. NOLTE.*Contention-Free Execution of Automotive Applications on a Clustered Many-Core Platform*, in "28th Euromicro Conference on Real-Time Systems, ECRTS 2016, Toulouse, France, July 5-8, 2016", 2016, p. 14–24, https://doi.org/10.1109/ECRTS.2016.14.

[16] J. DENNIS.*First Version of a Dataflow Procedure Language*, in "Lecture Notes in Computer Sci.", Springer-Verlag, 1975, vol. 19, p. 362-376.

[17] R. E. KORF.*A New Algorithm for Optimal Bin Packing*, in "Eighteenth National Conference on Artificial Intelligence", 2002, p. 731–736.

[18] 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.

[19] R. PELLIZZONI, E. BETTI, S. BAK, G. YAO, J. CRISWELL, M. CACCAMO, R. KEGLEY.*A Predictable Execution Model for COTS-Based Embedded Systems*, in "17th IEEE Real-Time and Embedded Technology and Applications Symposium, RTAS 2011, Chicago, Illinois, USA, 11-14 April 2011", 2011, p. 269–279, https://doi.org/10.1109/RTAS.2011.33.

[20] J. WHITHAM, N. C. AUDSLEY, R. I. DAVIS.*Explicit Reservation of Cache Memory in a Predictable, Preemptive Multitasking Real-time System*, in "ACM Trans. Embed. Comput. Syst.", April 2014, vol. 13, n^{o} 4s, p. 120:1–120:25, http://doi.acm.org/10.1145/2523070.

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

Université Pierre et Marie Curie (Paris 6)

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. Personnel

Research Scientists

Olivier Colliot [Team leader, CNRS, Senior Researcher, HDR]
Mario Chavez [CNRS, Researcher, until Aug 2017]
Fabrizio de Vico Fallani [Inria, Researcher, HDR]
Stanley Durrleman [Inria, Researcher]
Vincent Henry [Inria, Advanced Research Position, from Dec 2017]

Faculty Members

Anne Bertrand [Sorbonne University, Associate Professor]
Didier Dormont [Sorbonne University, Professor, HDR]
Benjamin Charlier [Assistant Professor with secondment, Univ de Montpellier]
Stephane Epelbaum [Hospital neurologist, Assistance publique/Hôpitaux de Paris, from Sep 2017]

External Collaborators

Mario Chavez [CNRS, from Sep 2017]
Marie Chupin [CNRS, from Sep 2017]
Denis Schwartz [INSERM]

Technical Staff

Chabha Azouani [ICM, until Sep 2017]
Michael Bacci [Inria, until Oct 2017]
Simona Bottani [ICM, from Feb 2017]
Marie Chupin [CNRS, until Aug 2017]
Sabrina Fontanella [ICM, until Oct 2017]
Clementine Fourrier [ICM, from Mar 2017]

Arnaud Marcoux [Inria, from Feb 2017]

PhD Students

Manon Ansart [INSERM]
Giulia Bassignana [INSERM, from Nov 2017]
Alexandre Bone [Univ Pierre et Marie Curie]
Tiziana Cattai [Inria, from Nov 2017]
Raphael Couronne [Inria, from Jul 2017]
Fanny Grosselin [Univ Pierre et Marie Curie]
Jeremy Guillon [Univ Pierre et Marie Curie]
Igor Koval [INSERM]
Thomas Lartigue [Inria, from Oct 2017]
Maxime Louis [Univ Pierre et Marie Curie]
Pascal Lu [Univ Pierre et Marie Curie]
Catalina Obando Forero [Inria]
Alexandre Routier [Univ Pierre et Marie Curie]
Jorge Samper Gonzalez [Inria]
Junhao Wen [CSC]
Wen Wei [Inria]

Post-Doctoral Fellows

Ninon Burgos [Inria]
Alexis Guyot [ICM, from Jun 2017]
Takoua Kaaouana [ICM, until Nov 2017]
Federico Battiston [Inria/CNRS, until May 2017]
Marie-Constance Corsi [Inria]

Administrative Assistants

Helene Milome [Inria]
Emmanuelle Mauduit [ICM, from Sep 2017]

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 University Pierre and Marie Curie.

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 (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

- Anne Bertrand spent a year half-time within the ARAMIS team, thanks to an Inria-APHP interface contract (i.e., "poste d'accueil"), from november 2016 to november 2017. At the end of this contract, she was appointed as an Assistant Professor of Radiology at Sorbonne University, on september 2017, allowing her to continue working 40% of her time within the ARAMIS team.
- Fabrizio De Vico Fallani was named associate editor of the journal Brain Topography
- Stanley Durrleman was nominated coordinator of the ICM Center of Neuroinformatics, and scientific manager of the ICM iCONICS core-facility on bioinformatics.
- The team has been awarded the projects SEMAPHORE, ATTACK and PredictICD under the "Big Brain Theory" program (ICM)

5.1.1. Awards

- Jeremy Guillon was awarded the best lighting presentation at the international conference on complex networks

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: Anatomy - Mesh - Automatic Learning - C++ - 3D modeling - 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.

- Participants: Junhao Wen, Jorge Samper Gonzalez, Alexandre Routier, Tristan Moreau, Arnaud Marcoux, Pascal Lu, Thomas Jacquemont, Jeremy Guillon, Olivier Colliot, Stanley Durrleman, Michael Bacci, Simona Bottani, Ninon Burgos, Sabrina Fontanella and Pietro Gori
- 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
- URL: http://www.clinica.run

6.4. Platforms

6.4.1. Platform Brain-computer interface

Our team has coordinated the implementation 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 funded project, and demos are currently being run by different researchers of the Institute. Such technological advance contributed to the scientific visibility of Inria and ICM with two TV reports (M6 and France 5).

7. New Results

7.1. Fiberprint: A subject fingerprint based on sparse code pooling for white matter fiber analysis

Participants: Kuldeep Kumar [Correspondant], Christian Desrosiers, Kaleem Siddiqi, Olivier Colliot, Matthew Toews.

White matter characterization studies use the information provided by diffusion magnetic resonance imaging (dMRI) to draw cross-population inferences. However, the structure, function, and white matter geometry vary across individuals. Here, we propose a subject fingerprint, called Fiberprint, to quantify the individual uniqueness in white matter geometry using fiber trajectories. We learn a sparse coding representation for fiber trajectories by mapping them to a common space defined by a dictionary. A subject fingerprint is then generated by applying a pooling function for each bundle, thus providing a vector of bundle-wise features describing a particular subject's white matter geometry. These features encode unique properties of fiber trajectories, such as their density along prominent bundles. An analysis of data from 861 Human Connectome Project subjects reveals that a fingerprint based on approximately 3000 fiber trajectories can uniquely identify exemplars from the same individual. We also use fingerprints for twin/sibling identification, our observations consistent with the twin data studies of white matter integrity. Our results demonstrate that the proposed Fiberprint can effectively capture the variability in white matter fiber geometry across individuals, using a compact feature vector (dimension of 50), making this framework particularly attractive for handling large datasets.

More details in [21].

7.2. Individual analysis of molecular brain imaging data through automatic identification of abnormality patterns

Participants: Ninon Burgos [Correspondant], Jorge Samper-González, Anne Bertrand, Marie-Odile Habert, Sébastien Ourselin, Stanley Durrleman, M. Jorge Cardoso, Olivier Colliot.

We introduce a pipeline for the individual analysis of positron emission tomography (PET) data on large cohorts of patients. This pipeline consists for each individual of generating a subject-specific model of healthy PET appearance and comparing the individual's PET image to the model via a novel regularised Z-score. The resulting voxel-wise Z-score map can be interpreted as a subject-specific abnormality map that summarises the pathology's topographical distribution in the brain. We then propose a strategy to validate the abnormality maps on several PET tracers and automatically detect the underlying pathology by using the abnormality maps as features to feed a linear support vector machine (SVM)-based classifier. We applied the pipeline to a large dataset comprising 298 subjects selected from the ADNI2 database (103 cognitively normal, 105 late MCI and 90 Alzheimer's disease subjects). The high classification accuracy obtained when using the abnormality maps as features demonstrates that the proposed pipeline is able to extract for each individual the signal characteristic of dementia from both FDG and Florbetapir PET data.

More details in [27].

7.3. Multilevel Modeling with Structured Penalties for Classification from Imaging Genetics data

Participants: Pascal Lu [Correspondant], Olivier Colliot.

In this paper, we propose a framework for automatic classification of patients from multimodal genetic and brain imaging data by optimally combining them. Additive models with unadapted penalties (such as the classical group lasso penalty or L_1-multiple kernel learning) treat all modalities in the same manner and can result in undesirable elimination of specific modalities when their contributions are unbalanced. To overcome this limitation, we introduce a multilevel model that combines imaging and genetics and that considers joint effects between these two modalities for diagnosis prediction. Furthermore, we propose a framework allowing to combine several penalties taking into account the structure of the different types of data, such as a group lasso penalty over the genetic modality and a L_2-penalty on imaging modalities. Finally , we propose a fast optimization algorithm, based on a proximal gradient method. The model has been evaluated on genetic (single nucleotide polymorphisms-SNP) and imaging (anatomical MRI measures) data from the ADNI database, and compared to additive models. It exhibits good performances in AD diagnosis; and at the same time, reveals relationships between genes, brain regions and the disease status.

More details in [33].

7.4. Towards Fully-reproducible Research on Classification of Alzheimer's Disease

Participants: Jorge Samper-González [Correspondant], Ninon Burgos, Sabrina Fontanella, Hugo Bertin, Marie-Odile Habert, Stanley Durrleman, Theodoros Evgeniou, Olivier Colliot.

In recent years, the number of papers on Alzheimer's disease classification has increased dramatically, generating interesting methodological ideas on the use machine learning and feature extraction methods. However, practical impact is much more limited and, eventually, one could not tell which of these approaches are the most efficient. While over 90% of these works make use of ADNI an objective comparison between approaches is impossible due to variations in the subjects included, image pre-processing, performance metrics and cross-validation procedures. In this paper, we propose a framework for reproducible classification experiments using multimodal MRI and PET data from ADNI. The core components are: 1) code to

automatically convert the full ADNI database into BIDS format; 2) a modular architecture based on Nipype in order to easily plug-in different classification and feature extraction tools; 3) feature extraction pipelines for MRI and PET data; 4) baseline classification approaches for unimodal and multimodal features. This provides a flexible framework for benchmarking different feature extraction and classification tools in a reproducible manner. Data management tools for obtaining the lists of subjects in AD, MCI converter, MCI non-converters, CN classes are also provided. We demonstrate its use on all (1519) baseline T1 MR images and all (1102) baseline FDG PET images from ADNI 1, GO and 2 with SPM-based feature extraction pipelines and three different classification techniques (linear SVM, anatomically regularized SVM and multiple kernel learning SVM). The highest accuracies achieved were: 91% for AD vs CN, 83% for MCIc vs CN, 75% for MCIc vs MCInc, 94% for AD-ABeta+ vs CN-ABeta- and 72% for MCIc-ABeta+ vs MCInc-ABeta+. The code is publicly available at https://gitlab.icm-institute.org/aramislab/AD-ML.

More details in [34].

7.5. Early Cognitive, Structural, and Microstructural Changes in Presymptomatic C9orf72 Carriers Younger Than 40 Years

Participants: Anne Bertrand [Correspondant], Junhao Wen, Sabrina Fontanella, Alexandre Routier, Stanley Durrleman, Olivier Colliot.

Presymptomatic carriers of chromosome 9 open reading frame 72 (C9orf72) mutation, the most frequent genetic cause of frontotemporal lobar degeneration and amyotrophic lateral sclerosis, represent the optimal target population for the development of disease-modifying drugs. Preclinical biomarkers are needed to monitor the effect of therapeutic interventions in this population. The aim of our study was to assess the occurrence of cognitive, structural, and microstructural changes in presymptomatic C9orf72 carriers. The PREV-DEMALS study is a prospective, multicenter, observational study of first-degree relatives of individuals carrying the C9orf72 mutation. Eighty-four participants entered the study between October 2015 and April 2017; 80 (95%) were included in cross-sectional analyses of baseline data. All participants underwent neuropsychological testing and magnetic resonance imaging; 63 (79%) underwent diffusion tensor magnetic resonance imaging. Gray matter volumes and diffusion tensor imaging metrics were calculated within regions of interest. Anatomical and microstructural differences between individuals who carried the C9orf72 mutation (C9+) and those who did not carry the C9orf72 mutation (C9-) were assessed using linear mixed-effects models. Data were analyzed from October 2015 to April 2017.Of the 80 included participants, there were 41 C9+ individuals (24 [59%] female; mean [SD] age, 39.8 [11.1] years) and 39 C9- individuals (24 [62%] female; mean [SD] age,45.2 [13.9] years). Compared with C9- individuals, C9+ individuals had lower mean (SD)praxis scores (163.4 [6.1] vs 165.3 [5.9]; P = .01) and intransitive gesture scores (34.9 [1.6] vs 35.7 [1.5]; P = .004), atrophy in 8 cortical regions of interest and in the right thalamus, and white matter alterations in 8 tracts. When restricting the analyses to participants younger than 40 years, compared with C9- individuals, C9+ individuals had lower praxis scores and intransitive gesture scores, atrophy in 4 cortical regions of interest and in the right thalamus, and white matter alterations in 2 tracts.Our work demontrates that cognitive, structural and microstructural alterations are detectable in young C9+ individuals. Early and subtle praxis alterations, underpinned by focal atrophy of the left supramarginal gyrus, may represent an early and nonevolving phenotype related to neurodevelopmental effects of C9orf72 mutation. White matter alterations reflect the future phenotype of frontotemporal lobar degeneration/amyotrophic lateral sclerosis, while atrophy appears more diffuse. Our results contribute to a better understanding of the preclinical phase of C9orf72

More details in [5].

7.6. Loss of brain inter-frequency hubs in Alzheimer's disease

Participants: Jeremy Guillon, Yohan Attal, Olivier Colliot, Valentina La Corte, Bruno Dubois, Denis Schwartz, Mario Chavez, Fabrizio de Vico Fallani [Correspondant].

Alzheimer's disease (AD) causes alterations of brain network structure and function. The latter consists of connectivity changes between oscillatory processes at different frequency channels. We proposed a multilayer network approach to analyze multiple-frequency brain networks inferred from magnetoencephalographic recordings during resting-states in AD subjects and age-matched controls. Main results showed that brain networks tend to facilitate information propagation across different frequencies, as measured by the multi-participation coefficient (MPC). However, regional connectivity in AD subjects was abnormally distributed across frequency bands as compared to controls, causing significant decreases of MPC. This effect was mainly localized in association areas and in the cingulate cortex, which acted, in the healthy group, as a true inter-frequency hub. MPC values significantly correlated with memory impairment of AD subjects, as measured by the total recall score. Most predictive regions belonged to components of the default-mode network that are typically affected by atrophy, metabolism disruption and amyloid-β deposition. We evaluated the diagnostic power of the MPC and we showed that it led to increased classification accuracy (78.39%) and sensitivity (91.11%). These findings shed new light on the brain functional alterations underlying AD and provide analytical tools for identifying multi-frequency neural mechanisms of brain diseases.

More details in [17].

7.7. A statistical model for brain networks inferred from large-scale electrophysiological signals

Participants: Catalina Obando, Fabrizio de Vico Fallani [Correspondant].

Network science has been extensively developed to characterize the structural properties of complex systems, including brain networks inferred from neuroimaging data. As a result of the inference process, networks estimated from experimentally obtained biological data represent one instance of a larger number of realizations with similar intrinsic topology. A modelling approach is therefore needed to support statistical inference on the bottom-up local connectivity mechanisms influencing the formation of the estimated brain networks. Here, we adopted a statistical model based on exponential random graph models (ERGMs) to reproduce brain networks, or connectomes, estimated by spectral coherence between high-density electroencephalographic (EEG) signals. ERGMs are made up by different local graph metrics, whereas the parameters weight the respective contribution in explaining the observed network. We validated this approach in a dataset of N ¼ 108 healthy subjects during eyes-open (EO) and eyes closed (EC) resting-state conditions. Results showed that the tendency to form triangles and stars, reflecting clustering and node centrality, better explained the global properties of the EEG connectomes than other combinations of graph metrics. In particular, the synthetic networks generated by this model configuration replicated the characteristic differences found in real brain networks, with EO eliciting significantly higher segregation in the alpha frequency band (8–13 Hz) than EC. Furthermore, the fitted ERGM parameter values provided complementary information showing that clustering connections are significantly more represented from EC to EO in the alpha range, but also in the beta band (14–29 Hz), which is known to play a crucial role in cortical processing of visual input and externally oriented attention. Taken together, these findings support the current view of the functional segregation and integration of the brain in terms of modules and hubs, and provide a statistical approach to extract new information on the (re)organizational mechanisms in healthy and diseased brains. More details in [23].

7.8. A Topological Criterion for Filtering Information in Complex Brain Networks

Participants: Fabrizio de Vico Fallani [Correspondant], Vito Latora, Mario Chavez.

In many biological systems, the network of interactions between the elements can only be inferred from experimental measurements. In neuroscience, non-invasive imaging tools are extensively used to derive either structural or functional brain networks in-vivo. As a result of the inference process, we obtain a matrix of values corresponding to a fully connected and weighted network. To turn this into a useful sparse network, thresholding is typically adopted to cancel a percentage of the weakest connections. The structural properties of the resulting network depend on how much of the inferred connectivity is eventually retained. However, how

to objectively fix this threshold is still an open issue. We introduce a criterion, the efficiency cost optimization (ECO), to select a threshold based on the optimization of the trade-off between the efficiency of a network and its wiring cost. We prove analytically and we confirm through numerical simulations that the connection density maximizing this trade-off emphasizes the intrinsic properties of a given network, while preserving its sparsity. Moreover, this density threshold can be determined a-priori, since the number of connections to filter only depends on the network size according to a power-law. We validate this result on several brain networks, from micro- to macro-scales, obtained with different imaging modalities. Finally, we test the potential of ECO in discriminating brain states with respect to alternative filtering methods. ECO advances our ability to analyze and compare biological networks, inferred from experimental data, in a fast and principled way.

More details in [11].

7.9. Preclinical Alzheimer's disease: a systematic review of the cohorts underlying the concept

Participants: Stéphane Epelbaum [Correspondant], Remy Genthon, Enrica Cavedo, Marie Odile Habert, Foudil Lamari, Geoffroy Gagliardi, Simone Lista, Marc Teichmann, Hovagim Bakardjian, Harald Hampel, Bruno Dubois.

Preclinical Alzheimer's disease (AD) is a relatively recent concept describing an entity characterized by the presence of a pathophysiological biomarker signature characteristic for AD in the absence of specific clinical symptoms. There is rising interest in the scientific community to define such an early target population mainly due to failures of all recent clinical trials despite evidence of biological effects on brain amyloidosis for some compounds. A conceptual framework has recently been proposed for this preclinical phase of AD. However, few data exist on this silent stage of AD. We performed a systematic review in order to investigate how the concept is defined across studies. The review highlights the substantial heterogeneity concerning the three main determinants of preclinical AD: "normal cognition", "cognitive decline" and "AD pathophysiological signature". We emphasize the need for a harmonized nomenclature of the preclinical AD concept and standardized population-based and case-control studies using unified operationalized criteria.

More details in [12].

7.10. Free and Cued Selective Reminding Test - accuracy for the differential diagnosis of Alzheimer's and neurodegenerative diseases: A large-scale biomarker-characterized monocenter cohort study (ClinAD)

Participants: Marc Teichmann [Correspondant], Stéphane Epelbaum, Dalila Samri, Marcel Levy Nogueira, Agnes Michon, Harald Hampel, Foudil Lamari, Bruno Dubois.

The International Working Group recommended the Free and Cued Selective Reminding Test (FCSRT) as a sensitive detector of the amnesic syndrome of the hippocampal type in typical Alzheimer's disease (AD). But does it differentiate AD from other neurodegenerative diseases? We assessed the FCSRT and cerebrospinal fluid (CSF) AD biomarkers in 992 cases. Experts, blinded to biomarker data, attributed in 650 cases a diagnosis of typical AD, frontotemporal dementia, posterior cortical atrophy, Lewy body disease, progressive supranuclear palsy, corticobasal syndrome, primary progressive aphasias, "subjective cognitive decline," or depression. The FCSRT distinguished typical AD from all other conditions with a sensitivity of 100% and a specificity of 75%. Non-AD neurodegenerative diseases with positive AD CSF biomarkers ("atypical AD") did not have lower FCSRT scores than those with negative biomarkers. The FCSRT is a reliable tool for diagnosing typical AD among various neurodegenerative diseases. At an individual level, however, its specificity is not absolute. Our findings also widen the spectrum of atypical AD to multiple neurodegenerative conditions.

More details in [13].

7.11. Parallel transport in shape analysis : a scalable numerical scheme

Participants: Maxime Louis, Alexandre Bône, Benjamin Charlier, Stanley Durrleman.

The analysis of manifold-valued data requires efficient tools from Riemannian geometry to cope with the computational complexity at stake. This complexity arises from the always-increasing dimension of the data, and the absence of closed-form expressions to basic operations such as the Riemannian logarithm. In this work, we adapted a generic numerical scheme recently introduced for computing parallel transport along geodesics in a Riemannian manifold to finite-dimensional manifolds of diffeomorphisms. We provided a qualitative and quantitative analysis of its behavior on high-dimensional manifolds, and investigated an application with the prediction of brain structures progression.

More details in [32].

7.12. Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks

Participants: Igor Koval, Jean-Baptiste Schiratti, Alexandre Routier, Michael Bacci, Olivier Colliot, Stéphanie Allassonnière, Stanley Durrleman.

We introduced a mixed-effects model to learn spatiotemporal patterns on a network by considering longitudinal measures distributed on a fixed graph. The data come from repeated observations of subjects at different time points which take the form of measurement maps distributed on a graph such as an image or a mesh. The model learns a typical group-average trajectory characterizing the propagation of measurement changes across the graph nodes. The subject-specific trajectories are defined via spatial and temporal transformations of the group-average scenario, thus estimating the variability of spatiotemporal patterns within the group. To estimate population and individual model parameters, we adapted a stochastic version of the Expectation-Maximization algorithm, the MCMC-SAEM. The model was used to describe the propagation of cortical atrophy during the course of Alzheimer's Disease. Model parameters show the variability of this average pattern of atrophy in terms of trajectories across brain regions, age at disease onset and pace of propagation. We showed that the personalization of this model yields accurate prediction of maps of cortical thickness in patients.

More details in [29]

7.13. Prediction of the progression of subcortical brain structures in Alzheimer's disease from baseline

Participants: Alexandre Bône, Maxime Louis, Alexandre Routier, Jorge Samper, Michael Bacci, Benjamin Charlier, Olivier Colliot, Stanley Durrleman.

We proposed a method to predict the subject-specific longitudinal progression of brain structures extracted from baseline MRI, and evaluated its performance on Alzheimer's disease data. The disease progression is modeled as a trajectory on a group of diffeomorphisms in the context of large deformation diffeomorphic metric mapping (LDDMM). We first exhibited the limited predictive abilities of geodesic regression extrapolation on this group. Building on the recent concept of parallel curves in shape manifolds, we then introduced a second predictive protocol which personalizes previously learned trajectories to new subjects, and investigate the relative performances of two parallel shifting paradigms. This design only requires the baseline imaging data. Finally, coefficients encoding the disease dynamics are obtained from longitudinal cognitive measurements for each subject, and exploited to refine our methodology which was demonstrated to successfully predict the follow-up visits.

More details in [28]

7.14. Prediction of amyloidosis from neuropsychological and MRI data for cost effective inclusion of pre-symptomatic subjects in clinical trials

Participants: Manon Ansart, Stéphane Epelbaum, Geoffroy Gagliardi, Olivier Colliot, Didier Dormont, Bruno Dubois, Harald Hampel, Stanley Durrleman.

We proposed a method for selecting pre-symptomatic subjects likely to have amyloid plaques in the brain, based on the automatic analysis of neuropsychological and MRI data and using a cross-validated binary classifier. By avoiding systematic PET scan for selecting subjects, it reduces the cost of forming cohorts of subjects with amyloid plaques for clinical trials, by scanning fewer subjects but increasing the number of recruitments. We validated our method on three cohorts of subjects at different disease stages, and compared the performance of six classifiers, showing that the random forest yields good results more consistently, and that the method generalizes well when tested on an unseen data set.

More details in [25]

7.15. Geodesic shape regression with multiple geometries and sparse parameters

Participants: James Fishbaugh, Stanley Durrleman, Marcel Prastawa, Guido Gerig.

Many problems in medicine are inherently dynamic processes which include the aspect of change over time, such as childhood development, aging, and disease progression. From medical images, numerous geometric structures can be extracted with various representations, such as landmarks, point clouds, curves, and surfaces. Different sources of geometry may characterize different aspects of the anatomy, such as fiber tracts from DTI and subcortical shapes from structural MRI, and therefore require a modeling scheme which can include various shape representations in any combination. In this paper, we present a geodesic regression model in the large deformation (LDDMM) framework applicable to multi-object complexes in a variety of shape representations. Our model decouples the deformation parameters from the specific shape representations, allowing the complexity of the model to reflect the nature of the shape changes, rather than the sampling of the data. As a consequence, the sparse representation of diffeomorphic flow allows for the straightforward embedding of a variety of geometry in different combinations, which all contribute towards the estimation of a single deformation of the ambient space. Additionally, the sparse representation along with the geodesic constraint results in a compact statistical model of shape change by a small number of parameters defined by the user. Experimental validation on multi-object complexes demonstrate robust model estimation across a variety of parameter settings. We further demonstrate the utility of our method to support the analysis of derived shape features, such as volume, and explore shape model extrapolation. Our method is freely available in the software package deformetrica which can be downloaded at www.deformetrica.org.

More details in [14]

7.16. A sub-Riemannian modular framework for diffeomorphism based analysis of shape ensembles

Participants: Barara Gris, Stanley Durrleman, Alain Trouvé.

Deformations, and diffeormophisms in particular, have played a tremendous role in the field of statistical shape analysis, as a proxy to measure and interpret differences between similar objects but with different shapes. Diffeomorphisms usually result from the integration of a flow of regular velocity fields, whose parameters have not enabled so far a full control of the local behaviour of the deformation. In this work, we propose a new mathematical and computational framework, in which diffeomorphisms are built on the combination of local deformation modules with few degrees of freedom. Deformation modules contribute to a global velocity field, and interact with it during integration so that the local modules are transported by the global diffeomorphic deformation under construction. Such modular diffeomorphisms are used to deform shapes and to provide the shape space with a sub-Riemannian metric. We then derive a method to estimate a Fréchet mean from a

series of observations, and to decompose the variations in shape observed in the training samples into a set of elementary deformation modules encoding distinctive and interpretable aspects of the shape variability. We show how this approach brings new solutions to long lasting problems in the fields of computer vision and medical image analysis. For instance, the easy implementation of priors in the type of deformations offers a direct control to favor one solution over another in situations where multiple solutions may fit the observations equally well. It allows also the joint optimisation of a linear and a non-linear deformation between shapes, the linear transform simply being a particular type of modules. The proposed approach generalizes previous methods for constructing diffeomorphisms and opens up new perspectives in the field of statistical shape analysis.

More details in [16]

7.17. A Bayesian Framework for Joint Morphometry of Surface and Curve meshes in Multi-Object Complexes

Participants: Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Cyril Poupon, Andreas Hartmann, Nicholas Ayache, Stanley Durrleman.

We present a Bayesian framework for atlas construction of multi-object shape complexes comprised of both surface and curve meshes. It is general and can be applied to any parametric deformation framework and to all shape models with which it is possible to define probability density functions (PDF). Here , both curve and surface meshes are modelled as Gaussian random varifolds , using a finite-dimensional approximation space on which PDFs can be defined. Using this framework , we can automatically estimate the parameters balancing data-terms and deformation regularity , which previously required user tuning. Moreover , it is also possible to estimate a well-conditioned covariance matrix of the deformation parameters. We also extend the proposed framework to data-sets with multiple group labels. Groups share the same template and their deformation parameters are modelled with different distributions. We can statistically compare the groups ' distributions since they are defined on the same space. We test our algorithm on 20 Gilles de la Tourette patients and 20 control subjects , using three sub-cortical regions and their incident white matter fiber bundles. We compare their morphological characteristics and variations using a single diffeomorphism in the ambient space. The proposed method will be integrated with the Deformetrica software package, publicly available at www.deformetrica.org.

More details in [15]

7.18. A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations

Participants: Jean-Baptiste Schiratti, Stéphanie Allassonnière, Olivier Colliot, Stanley Durrleman.

We propose a generic Bayesian mixed-effects model to estimate the temporal progression of a biological phenomenon from observations obtained at multiple time points for a group of individuals. The progression is modeled by continuous trajectories in the space of measurements. Individual trajectories of progression result from spatiotemporal transformations of an average trajectory. These transformations allow to quantify the changes in direction and pace at which the trajectories are followed. The framework of Rieman-nian geometry allows the model to be used with any kind of measurements with smooth constraints. A stochastic version of the Expectation-Maximization algorithm is used to produce produce maximum a posteriori estimates of the parameters. We evaluate our method using series of neuropsychological test scores from patients with mild cognitive impairments later diagnosed with Alzheimer's disease, and simulated evolutions of symmetric positive definite matrices. The data-driven model of the impairment of cognitive functions shows the variability in the ordering and timing of the decline of these functions in the population. We show also that the estimated spatiotemporal transformations effectively put into correspondence significant events in the progression of individuals.

More details in [40]

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Air-Liquide Medical Systems

Participants: Mario Chavez [Correspondant], Xavier Navarro.

Project title: Real-time characterisation of respiratory states from EEG

Funded in 2014

Amount: 370 K€

Coordinator: Thomas Similowski

Other partners: UPMC, Inserm UMR 1158

Abstract: The project aims at developing a real-time brain computer interface (BCI) for the monitoring of respiratory states from scalp EEG data of healthy volunteers and patients, recorded at the laboratory, hospital ward, operating room or intensive care units.

8.2. Bilateral Grants with Industry

8.2.1. 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

Funded 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 (μ- 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 μ-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

Amount: 633k€

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 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, Mario Chavez, Stanley Durrleman, Marie Chupin, Didier Dormont, Dominique Hasboun, Damien Galanaud, 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.

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 [Correspondant], 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 Big Brain Theory Program

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

Founded in 2016-2017

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.

9.1.3.5. IFR49-Internal Research projects

Participants: Mario Chavez [Correspondant], Fabrizio de Vico Fallani [Correspondant].

Project title: Exploring the impact and time frequency signature of rhythmic patterns of Transcranial Magnetic Stimulation (TMS) on network activity by Magneto-Encephalography (MEG)

Founded in 2014

Coordinator: Antoni Valero Cabre (ICM-team “Dynamiques Cérébrales, Plasticité et Rééducation”)

Other partners: TMS, EEG and MEG technical platforms of the ICM at the Hopital Pitié-Salptrière; and Service des Urgences Cérébro-Vasculaires de l’Hôpital Pitié-Salpêtrière, Paris.

The long-term goal of this project is to better understand the ability of non invasive neurostimulation to induce lasting local and distributed reorganization effects in the human brain to better plan and document therapies for patients. The short-term goal of this application is to develop a new mapping procedure to be able to capture and characterize in terms of oscillatory activity the lasting impact of repetitive Transcranial Magnetic Stimulation (TMS) on specific brain regions and associated networks.

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/

9.1.5. Other National Programs

9.1.5.1. Programme Hospitalier de Recherche Clinique (PHRC)

Participants: Olivier Colliot, Marie Chupin, Stanley Durrleman, Didier Dormont, Damien Galanaud.

- 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 (for this phase): 2016-2018

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

Participants: Stanley Durrleman, Raphael Couronné.

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. Inria International Partners

9.3.1.1. Informal International Partners

F. De Vico Fallani has a collaboration with the University Penn, Philadelphia, US (Prof. Danielle Bassett).

M. Chavez has different collaborations with the Mathematics Departement of the Queen Mary University of London, UK (Prof. V. Latora); and the Physics Department of the Universitat de Barcelona, Spain (Prof. Albert Diaz-Guilera)

F. De Vico Fallani has an enduring collaboration with the University Sapienza, Rome, Italy (Profs. Fabio and Claudio Babiloni) and with the IRCCS Fondazione Santa Lucia, Rome, Italy (M. Molinari and D. Mattia).

S. Durrleman has an enduring collaboration with professor Guido Gerig, Tandon School of Engineering, NYU. He is consultant for NIH Grant "4D shape analysis for modeling spatiotemporal change trajectories in Huntington's Disease "predict-HD".

O. Colliot has an enduring collaboration with the Center for Magnetic Resonance Research, University of Minnesota, USA (P-F Van de Moortele, T. Henry, M. Marjanska, K. Ugurbil) a leading center in 7T MRI.

S. Durrleman and O. Colliot have a collaboration with the Center for Medical Image Computing (CMIC) at University College London (UCL), London, UK (S. Ourselin, D. Alexander, M. Modat).

S. Durrleman has a collaboration with the department of Computer Science at New York University (NYU) (G. Gerig and J. Fishbaugh)

A. Bertrand has an enduring collaboration with professor Youssef Z. Wadghiri, head of the Preclinical Imaging Core, Center for Biomedical Imaging, NYU School of Medicine, New York, NY, USA.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Professor Tom Fletcher from the University of Utah visited ARAMIS from January 23 to January 27.

9.4.1.1. Internships

Kuldeep Kumar (Ecole de Technologie Supérieure, Montréal, Canada) is visiting ARAMIS from October 2016 to March 2017 under the MITACS programme.

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

Junhao Wen, PhD candidate, did a 3-month internship in the team of Hui Zhang, UCL, to develop pipelines of analysis for advanced diffusion MRI acquisitions (Neurite Orientation Dispersion and Density Imaging). This internship was funded by the ICM Carnot Program.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- S. Durrleman was co-chair of the 6th international workshop on the Mathematical Foundations of Computational Anatomy (MFCA) in Quebec City on September 14.

10.1.1.2. Member of the Organizing Committees

Fabrizio De Vico Fallani is member of the scientific board for the ICM conferences

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

O. Colliot was a member of the program committee of the Workshop on Patch-based Techniques in Medical Imaging (Patch-MI) held in conjunction with the MICCAI conference.

O. Colliot was a member of the program committee of SPIE Medical Imaging conference.

F. De Vico Fallani was member of the program committee of the Satellite on Brain networks, International Conference on Network Science (NetSci), Indianapolis, US, 2017

F. De Vico Fallani was member of the program committee of 5th International Workshop on Complex Networks and their Applications, Lyon, France, 2017

S. Epelbaum was a member of the French medical board on the management of Early phase Alzheimer's Disease, Issy les moulineaux, France, 2017

10.1.2.2. Reviewer

O. Colliot acted as a reviewer for the annual meeting of the Organization for Human Brain Mapping (OHBM).

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

O. Colliot is a member of the Editorial Board of Medical Image Analysis (Elsevier).

S. Durrleman is associate editor of IEEE Transactions on Medical Imaging (TMI)

10.1.3.2. Reviewer - Reviewing Activities

O. Colliot acted as a reviewer for NeuroImage, NeuroImage: Clinical, IEEE Trans Medical Imaging, Medical Image Analysis and Neurobiology of Aging.

S. Epelbaum acted as a reviewer for Alzheimer's & Dementia and the Journal of Alzheimer's disease.

A. Bertrand acted as a reviewer for European Radiology, Journal of Magnetic Resonance Imaging, Journal of Neuroradiology.

F. De Vico Fallani acted as a reviewer for Cerebral Cortex, Brain Topography, IEEE TBME/TNRSE, Neuroimage, Sci Rep, Brain Connectivity, PLOS ONE

10.1.4. Invited Talks

A. Bertrand gave an invited lecture at the French-Quebecois Colloquium of the National Academy of Medicine "Towards news preventive and therapeutic strategies for neurodegenerative diseases: the role of imaging" in Québec City, Québec, september 2017.

F. De Vico Fallani gave an invited lecture at the international workshop on complex networks in Lipari, Italy, september 2017.

F. De Vico Fallani gave an invited lecture at the Neuroscience School of Advanced Studies in Siena, Italy, April 2017.

O. Colliot gave an invited lecture at the ICM/Supelec Workshop 2017.

S. Durrleman gave an invited lecture at the BIOVISION congress in Lyon.

S. Durrleman gave an invited lecture at the conference "Topological and Geometrical Science of Information" at Centre International de Rencontre Mathématiques, Luminy

S. Durrleman gave an invited lecture for the "diplôme inter-universitaire: diagnostic et prise en charge de la Maladie d'Alzheimer et apparentée", Lille

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master: Olivier Colliot coordinates the module "Méthodes d'imagerie médicale" of the Master 2 in Computer Science of Université Pierre et Marie Curie.

Master: Olivier Colliot, Master in Computer Science (module Reseaux complexes), 4.5 hours (eqTD), Université Pierre et Marie Curie

Master: Olivier Colliot, Master in Computer Science, 4.5 hours (eqTD), Université Pierre et Marie Curie

Master: Stéphane Epelbaum, Master in Neuroscience, 4 hours (eqTD), Université Pierre et Marie Curie

Master: S. Durrleman, Master 2 "Mathematics, Vision and Learning", 31,5 jours (eqTD), ENS Paris-Saclay

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

Medical school: Anne Bertrand gives lectures in Neuroimaging of degenerative diseases and normal aging for residents in Radiology and Neurology, for Radiology technicians, for License students in Orthophony, and in various "University Diploma" medical programs (Neurogeriatrics, Neuroradiology, Alzheimer's Disease and related disorders, Neurovascular Imaging, Emergency-Stroke, Neuroresuscitation), for a total of 50 hours a year.

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

10.2.3. Juries

Olivier Colliot participated, as referee, to the PhD committee of Elaheh Moradi (TU Finland) (supervisor: J. Tohka).

Olivier Colliot participated, as examiner, to the HdR committee of Roberto Toro (Institut Pasteur).

Anne Bertrand participated, as examiner, to the PhD committee of Clémence Dudeffant (Université Paris Sud), 2017 (supervisor: Marc Dhenain).

S. Durrleman participated, as examiner, to the PhD committees of Jean Dumoncel (Université Paul Sabatier, Toulouse) and Pierre Roussillon (Université Paris Descartes).

10.3. Popularization

With the precious help of the communication department of the ICM, ARAMIS prepared and presented games on brain data analysis at the "Fête de la Science" 2017.

Fabrizio De Vico Fallani was invited speaker at the Neuroplanete conference held in Nice 2017. In that occasion, a live BCI experiment has been presented to the public. This event has been reported by France 3.

Fabrizio De Vico Fallani was invited speaker at the Palais Decouverte in Paris for the theme "Vers le meilleur des mondes" 2017.

Olivier Colliot was invited speaker at the Palais Decouverte in Paris for the theme "Vers le meilleur des mondes" 2017.

Stanley Durrleman and Stéphane Epelbaum were invited speakers at the "Open Brain Bar: Artificial intelligence and Alzheimer’s disease" event orginized by the ICM in Paris 2017.

Olivier Colliot was invited speaker for the "Olympiades académiques de mathématiques", Paris, 2017.

S. Durrleman gave a presentation for the 50th anniversary of Inria featuring 50 ERC laureates

S. Durrleman gave two interviews for the magazine Sciences et Avenir (1 video interview, one article)

S. Durrleman and S. Epelbaum took part in the program “Priorité Santé” of the radio Radio France International

S. Durrleman gave a video interview for FrenchWeb.fr

S. Durrleman gave an talk at the event S3 Odéon (www.s3odeon.fr)

ARAMIS was featured in a video of the Agence France Presse (AFP)

11. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] 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", 2017 [*DOI :* 10.1111/JOA.12745], https://hal-lirmm.ccsd.cnrs.fr/lirmm-01636811.

[2] A. BERTRAND, S. BRUNEL, M.-O. HABERT, M. SORET, S. JAFFRE, N. CAPEAU, L. BOURSEUL, I. DUFOUR-CLAUDE, A. KAS, D. DORMONT.*A New Fire Hazard for MR Imaging Systems: Blankets—Case Report*, in "Radiology", August 2017, vol. 000 [*DOI :* 10.1148/RADIOL.2017162921], https://hal.inria.fr/hal-01580514.

[3] A. BERTRAND, D. LECLERCQ, L. MARTINEZ-ALMOYNA, N. GIRARD, J.-P. STAHL, T. DE-BROUCKER.*MR imaging of adult acute infectious encephalitis*, in "Médecine et Maladies Infectieuses", 2017 [*DOI :* 10.1016/J.MEDMAL.2017.01.002], https://hal.archives-ouvertes.fr/hal-01490868.

[4] A. BERTRAND, S. STROËR, I. LE BER, M. TEICHMANN, D. DORMONT.*Structural magnetic resonance imaging in frontotemporal lobar dementia*, in "Gériatrie et Psychologie Neuropsychiatrie du Vieillissement", September 2017, vol. 15, n^{o} 3 [*DOI :* 10.1684/PNV.2017.0686], https://hal.inria.fr/hal-01583008.

[5] 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", December 2017 [*DOI :* 10.1001/JAMANEUROL.2017.4266], https://hal.inria.fr/hal-01654000.

[6] M. BOUDOT DE LA MOTTE, C. LOUAPRE, A. BERTRAND, P. REACH, C. LUBETZKI, C. PAPEIX, E. MAILLART.*Extensive white matter lesions after 2 years of fingolimod: progressive multifocal leukoencephalopathy or MS relapse?*, in "Multiple Sclerosis", 2017, vol. 23, n^{o} 4, p. 614 - 616 [*DOI :* 10.1177/1352458516682858], https://hal.archives-ouvertes.fr/hal-01562641.

[7] F. CACCIAMANI, C. TANDETNIK, G. GAGLIARDI, H. BERTIN, M. HABERT, M. HAMPEL, L. BOUKADIDA, M. RÉVILLON, S. EPELBAUM, B. DUBOIS.*Low Cognitive Awareness, but Not Complaint, is a Good Marker of Preclinical Alzheimer's Disease*, in "Journal of Alzheimer's disease : JAD", June 2017, vol. 59, n^{o} 2, p. 753-762 [*DOI :* 10.3233/JAD-170399], https://hal.archives-ouvertes.fr/hal-01562622.

[8] E. CAVEDO, M. J. GROTHE, O. COLLIOT, S. LISTA, M. CHUPIN, D. DORMONT, M. HOUOT, S. LEHÉRICY, S. TEIPEL, B. DUBOIS, H. HAMPEL.*Reduced basal forebrain atrophy progression in a randomized Donepezil trial in prodromal Alzheimer's disease*, in "Scientific Reports", September 2017, vol. 7, n° 1, 11706, Hippocampus Study Group [*DOI :* 10.1038/s41598-017-09780-3], https://hal.inria.fr/hal-01589622.

[9] R. COLLE, T. SEGAWA, M. CHUPIN, M. N. T. K. TRAN DONG, P. HARDY, B. FALISSARD, O. COLLIOT, D. DUCREUX, E. CORRUBLE.*Early life adversity is associated with a smaller hippocampus in male but not female depressed in-patients: a case–control study*, in "BMC Psychiatry", 2017, vol. 17, n° 1, 71 [*DOI :* 10.1186/s12888-017-1233-2], http://hal.upmc.fr/hal-01474139.

[10] C. CURY, J. GLAUNÈS, M. CHUPIN, O. COLLIOT.*Analysis of anatomical variability using diffeomorphic iterative centroid in patients with Alzheimer's disease*, in "Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization", 2017, vol. 5, n° 5, p. 350 - 358 [*DOI :* 10.1080/21681163.2015.1035403], https://hal.inria.fr/hal-01135304.

[11] F. DE VICO FALLANI, V. LATORA, M. CHAVEZ.*A Topological Criterion for Filtering Information in Complex Brain Networks*, in "PLoS Computational Biology", January 2017, vol. 13, n° 1, p. 1-18 [*DOI :* 10.1371/JOURNAL.PCBI.1005305], https://hal.inria.fr/hal-01443254.

[12] S. EPELBAUM, R. GENTHON, E. CAVEDO, M. HABERT, F. LAMARI, G. GAGLIARDI, S. LISTA, M. TEICHMANN, H. BAKARDJIAN, H. HAMPEL, B. DUBOIS.*Preclinical Alzheimer's disease: a systematic review of the cohorts underlying the concept*, in "Alzheimer's and Dementia", 2017, vol. 13, n° 4, p. 454-467 [*DOI :* 10.1016/J.JALZ.2016.12.003], https://hal.inria.fr/hal-01672859.

[13] S. EPELBAUM, M. TEICHMANN, D. SAMRI, M. L. NOGUEIRA, H. HAMPEL, F. LAMARI, B. DUBOIS, A. MICHON.*Free and Cued Selective Reminding Test – accuracy for the differential diagnosis of Alzheimer's and neurodegenerative diseases: A large-scale biomarker-characterized monocenter cohort study (ClinAD)*, in "Alzheimer's & Dementia: The Journal of the Alzheimer's Association", August 2017, vol. 13, n° 8, p. 913 - 923 [*DOI :* 10.1016/J.JALZ.2016.12.014], https://hal.inria.fr/hal-01672862.

[14] J. FISHBAUGH, S. DURRLEMAN, M. PRASTAWA, G. GERIG.*Geodesic shape regression with multiple geometries and sparse parameters*, in "Medical Image Analysis", July 2017, vol. 39, p. 1-17, https://hal.inria.fr/hal-01566280.

[15] P. GORI, O. COLLIOT, L. MARRAKCHI-KACEM, Y. WORBE, C. POUPON, A. HARTMANN, N. AYACHE, S. DURRLEMAN.*A Bayesian Framework for Joint Morphometry of Surface and Curve meshes in Multi-Object Complexes*, in "Medical Image Analysis", January 2017, vol. 35, p. 458-474 [*DOI :* 10.1016/J.MEDIA.2016.08.011], https://hal.inria.fr/hal-01359423.

[16] B. GRIS, S. DURRLEMAN, A. TROUVÉ.*A sub-Riemannian modular framework for diffeomorphism based analysis of shape ensembles*, in "SIAM Journal of Imaging Sciences", 2017, https://hal.archives-ouvertes.fr/hal-01321142.

[17] J. GUILLON, Y. ATTAL, O. COLLIOT, V. LA CORTE, B. DUBOIS, D. SCHWARTZ, M. CHAVEZ, F. DE VICO FALLANI.*Loss of brain inter-frequency hubs in Alzheimer's disease*, in "Scientific Reports", September 2017, vol. 7, n° 1, 10879, https://arxiv.org/abs/1701.00096 , https://hal.archives-ouvertes.fr/hal-01578419.

[18] 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", 2017 [*DOI :* 10.1016/J.NEUROL.2017.06.005], https://hal.archives-ouvertes.fr/hal-01562621.

[19] H. HAMPEL, S. O'BRYANT, S. DURRLEMAN, E. YOUNESI, K. ROJKOVA, V. ESCOTT-PRICE, J.-C. CORVOL, K. BROICH, B. DUBOIS, S. LISTA.*A Precision Medicine Initiative for Alzheimer's disease: the road ahead to biomarker-guided integrative disease modeling*, in "Climacteric", April 2017, vol. 20, n° 2, p. 107-118, https://hal.inria.fr/hal-01566281.

[20] T. JACQUEMONT, F. DE VICO FALLANI, A. BERTRAND, S. EPELBAUM, A. ROUTIER, B. DUBOIS, H. HAMPEL, S. DURRLEMAN, O. COLLIOT.*Amyloidosis and neurodegeneration result in distinct structural connectivity patterns in mild cognitive impairment*, in "Neurobiology of Aging", April 2017 [*DOI :* 10.1016/J.NEUROBIOLAGING.2017.03.023], https://hal.inria.fr/hal-01518785.

[21] K. KUMAR, C. DESROSIERS, K. SIDDIQI, O. COLLIOT, M. TOEWS.*Fiberprint: A subject fingerprint based on sparse code pooling for white matter fiber analysis*, in "NeuroImage", July 2017, vol. 158, p. 242 - 259 [*DOI :* 10.1016/J.NEUROIMAGE.2017.06.083], https://hal.inria.fr/hal-01562449.

[22] B. MATHON, F. BIELLE, S. SAMSON, O. PLAISANT, S. DUPONT, A. BERTRAND, R. MILES, V.-H. NGUYEN-MICHEL, V. LAMBRECQ, A. L. CALDERON-GARCIDUEÑAS, C. DUYCKAERTS, A. CARPENTIER, M. BAULAC, P. CORNU, C. ADAM, S. CLEMENCEAU, V. NAVARRO.*Predictive factors of long-term outcomes of surgery for mesial temporal lobe epilepsy associated with hippocampal sclerosis*, in "Epilepsia", 2017, vol. 58, n° 8, p. 1473–1485 [*DOI :* 10.1111/EPI.13831], https://hal.inria.fr/hal-01557196.

[23] C. OBANDO, F. DE VICO FALLANI.*A statistical model for brain networks inferred from large-scale electrophysiological signals*, in "Journal of the Royal Society Interface", March 2017, vol. 14, n° 128 [*DOI :* 10.1098/RSIF.2016.0940], https://hal.inria.fr/hal-01564939.

[24] C. SANCHES, A. ROUTIER, O. COLLIOT, M. TEICHMANN.*The structure of the mental lexicon: what primary progressive aphasias reveal*, in "Neuropsychologia", December 2017, vol. 109, p. 107-115 [*DOI :* 10.1016/J.NEUROPSYCHOLOGIA.2017.12.018], https://hal.inria.fr/hal-01672932.

International Conferences with Proceedings

[25] M. ANSART, S. EPELBAUM, G. GAGLIARDI, O. COLLIOT, D. DORMONT, B. DUBOIS, H. HAMPEL, S. DURRLEMAN.*Prediction of amyloidosis from neuropsychological and MRI data for cost effective inclusion of pre-symptomatic subjects in clinical trials*, in "Multimodal Learning for Clinical Decision Support", Quebec City, Canada, September 2017, https://hal.inria.fr/hal-01578422.

[26] N. BURGOS, J. SAMPER-GONZÁLEZ, A. BERTRAND, M.-O. HABERT, S. OURSELIN, S. DURRLEMAN, M. J. CARDOSO, O. COLLIOT.*Diagnosis of Alzheimer's Disease Through Identification of Abnormality Patterns in FDG PET Data*, in "30th Annual Congress of the European Association of Nuclear Medicine (EANM)", Vienna, Austria, October 2017, vol. 44, n° S2, p. 253 - 254 [*DOI :* 10.1007/S00259-017-3822-1], https://hal.inria.fr/hal-01632509.

[27] N. BURGOS, J. SAMPER-GONZÁLEZ, A. BERTRAND, M.-O. HABERT, S. OURSELIN, S. DURRLEMAN, M. J. CARDOSO, O. COLLIOT.*Individual Analysis of Molecular Brain Imaging Data Through Automatic Identification of Abnormality Patterns*, in "Computational Methods for Molecular Imaging - [MICCAI 2017 Satellite Workshop]", Quebec City, Canada, September 2017, https://hal.inria.fr/hal-01567343.

[28] A. BÔNE, M. LOUIS, A. ROUTIER, J. SAMPER, M. BACCI, B. CHARLIER, O. COLLIOT, S. DURRLEMAN.*Prediction of the progression of subcortical brain structures in Alzheimer's disease from baseline*, in "6th MICCAI Workshop on Mathematical Foundations of Computational Anatomy", Quebec City, Canada, September 2017, https://hal.archives-ouvertes.fr/hal-01563587.

[29] I. KOVAL, J.-B. SCHIRATTI, A. ROUTIER, M. BACCI, O. COLLIOT, S. ALLASSONNIÈRE, S. DURRLEMAN.*Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks*, in "Medical Image Computing and Computer Assisted Intervention", Quebec City, Canada, Medical Image Computing and Computer Assisted Intervention, September 2017, https://hal.archives-ouvertes.fr/hal-01540828.

[30] K. KUMAR, L. CHAUVIN, M. TOEWS, O. COLLIOT, C. DESROSIERS.*Multi-modal analysis of genetically-related subjects using SIFT descriptors in brain MRI*, in "Workshop on Computational Diffusion MRI, CDMRI 2017, MICCAI Workshop", Quebec, Canada, September 2017, https://hal.inria.fr/hal-01589647.

[31] K. KUMAR, P. GORI, B. CHARLIER, S. DURRLEMAN, O. COLLIOT, C. DESROSIERS.*White Matter Fiber Segmentation Using Functional Varifolds*, in "Workshop on Mathematical Foundations of Computational Anatomy, MFCA 2017, MICCAI Workshop", Québec, Canada, Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics, Lecture Notes in Computer Science, Springer, September 2017, vol. 10551, p. 92 - 100, https://hal.inria.fr/hal-01589649.

[32] M. LOUIS, A. BÔNE, B. CHARLIER, S. DURRLEMAN.*Parallel transport in shape analysis : a scalable numerical scheme*, in "Geometric Science of Information", Paris, France, November 2017, https://hal.archives-ouvertes.fr/hal-01565478.

[33] P. LU, O. COLLIOT.*Multilevel Modeling with Structured Penalties for Classification from Imaging Genetics data*, in "3rd MICCAI Workshop on Imaging Genetics (MICGen 2017)", Québec City, Canada, Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics, Springer, September 2017, vol. 1, n^{o} 21, p. 230-240, https://hal.inria.fr/hal-01578441.

[34] J. SAMPER-GONZALEZ, N. BURGOS, S. FONTANELLA, H. BERTIN, M.-O. HABERT, S. DURRLEMAN, T. EVGENIOU, O. COLLIOT.*Yet Another ADNI Machine Learning Paper? Paving The Way Towards Fully-reproducible Research on Classification of Alzheimer's Disease*, in "Machine Learning in Medical Imaging 2017", Quebec City, Canada, MLMI 2017, LNCS 10541, September 2017, 8, https://hal.inria.fr/hal-01578479.

Conferences without Proceedings

[35] N. BURGOS, J. SAMPER-GONZÁLEZ, M. J. CARDOSO, S. DURRLEMAN, S. OURSELIN, O. COLLIOT, S. M. OURSELIN.*Early Diagnosis of Alzheimer's Disease Using Subject-Specific Models of FDG-PET Data*, in "Alzheimer's Association International Conference", London, United Kingdom, July 2017, vol. 13, n^{o} 7, p. 1-2 [*DOI :* 10.1016/J.JALZ.2017.06.1618], https://hal.inria.fr/hal-01621383.

[36] S. STRÖER, L. FILLON, S. EPELBAUM, M. TEICHMANN, A. MICHON, L. MIGGLIACCIO, A. MENDES, M. CHUPIN, O. COLLIOT, J. AHDIDAN, C. CHRISTOFFERSEN, C. PEDERSEN, D. DORMONT, A. BERTRAND.*Étude quantitative des anomalies de signal flair de la substance blanche dans les pathologies neurodégénératives*, in "SFNR 2017 - 44ème Congrès de la Société Française de Neuroradiologie", Paris, France, March 2017, vol. 44, n^{o} 2, 1 [*DOI :* 10.1016/J.NEURAD.2017.01.011], https://hal.archives-ouvertes.fr/hal-01562645.

Other Publications

[37] V. ANQUETIL, A. BERTRAND, A. CAMUZAT, P. D. L. GRANGE, C. FOURNIER, V. BUÉE-SCHERRER, V. DERAMECOURT, N. SERGEANT, M. BARBIER, L. BUÉE, A. BRICE, C. DUYCKAERTS, I. LE BER.*Frontotemporal lobar degenerations, RNAopathy leading to proteinopathies*, 2017, vol. 13, n^{o} 7, AAIC, Poster [*DOI :* 10.1016/J.JALZ.2017.06.1982], https://hal.inria.fr/hal-01672304.

[38] M. LOUIS, B. CHARLIER, P. JUSSELIN, S. PAL, S. DURRLEMAN.*A Fanning Scheme for the Parallel Transport Along Geodesics on Riemannian Manifolds*, July 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01560787.

[39] C. OBANDO FORERO, F. DE VICO FALLANI.*A statistical model for brain networks inferred from large-scale electrophysiological signals*, June 2017, NetSci, Poster, https://hal.inria.fr/hal-01564952.

[40] J.-B. SCHIRATTI, S. ALLASSONNIERE, O. COLLIOT, S. DURRLEMAN.*A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations*, 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01540367.

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
B9.4.1. - Computer science
B9.8. - Privacy

1. Personnel

Research Scientists

David Pointcheval [Team leader, CNRS, Senior Researcher, HDR]
Michel Ferreira Abdalla [CNRS, Senior Researcher, HDR]
Georg Fuchsbauer [Inria, Researcher]
Hoeteck Wee [CNRS, Senior Researcher, HDR]

Faculty Member

Damien Vergnaud [Ecole Normale Supérieure Paris, Associate Professor, until Aug 2017, HDR]

PhD Students

Balthazar Bauer [Inria, from Sep 2017]
Florian Bourse [CNRS]
Jeremy Chotard [CNRS]
Geoffroy Couteau [CNRS, until Sep 2017]
Pierre-Alain Dupont [Ecole Normale Supérieure Paris, until Aug 2017]
Romain Gay [Ecole Normale Supérieure Paris]
Dahmun Goudarzi [CryptoExperts]
Chloe Hebant [CNRS, from Oct 2017]
Louiza Khati [ANSSI]
Pierrick Meaux [Inria]
Thierry Mefenza Nountu [Ecole Normale Supérieure Paris]
Michele Minelli [Ecole Normale Supérieure Paris]
Anca Nitulescu [CNRS]
Michele Orru [CNRS]
Alain Passelegue [Inria, until Jan 2017]
Razvan Rosie [Ecole Normale Supérieure Paris]
Melissa Rossi [Thales, from May 2017]
Quentin Santos [Orange Labs]

Post-Doctoral Fellows

Pooya Farshim [Ecole Normale Supérieure Paris]
Julia Hesse [CNRS]

Administrative Assistant
Nathalie Gaudechoux [Inria]

2. Overall Objectives

2.1. Presentation

Cryptographic algorithms are the equivalent of locks, seals, security stamps and identification documents on the Internet. They are essential to protect our on-line bank transactions, credit cards, medical and personal information and to support e-commerce and e-government. They come in different flavors. Encryption algorithms are essential 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) replace hand-written signatures in electronic transactions. A similar role can be played by MAC algorithms. Identification protocols allow to securely verify the identity of the party at the other end of the line. Therefore, cryptology is a research area with a high strategic impact for industries, individuals, and for the society as a whole. The research activity of the project-team CASCADE addresses the following topics, which cover most of the domains that are currently active in the international cryptographic community, but mainly in the public-key area:

1. Implementation of cryptographic and applied cryptography
2. Design and provable security
3. Theoretical and concrete 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 into 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 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 to get provable security, without such ideal assumptions, under new and possibly stronger computational assumptions. As a consequence, a cryptographer has to deal with the four following important steps, which are **all** our main goals:

computational assumptions, which are the foundations 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, with additional features, etc.

security proof, which consists in exhibiting a reduction.

3. Research Program

3.1. Randomness in Cryptography

Randomness is a key ingredient for cryptography. Random bits are necessary not only for generating cryptographic keys, but are also often an important part of cryptographic algorithms. In some cases, probabilistic protocols make it possible to perform tasks that are impossible deterministically. In other cases, probabilistic algorithms are faster, more space efficient or simpler than known deterministic algorithms. Cryptographers usually assume that parties have access to perfect randomness but in practice this assumption is often violated and a large body of research is concerned with obtaining such a sequence of random or pseudorandom bits.

One of the project-team research goals is to get a better understanding of the interplay between randomness and cryptography and to study the security of various cryptographic protocols at different levels (information-theoretic and computational security, number-theoretic assumptions, design and provable security of new and existing constructions).

Cryptographic literature usually pays no attention to the fact that in practice randomness is quite difficult to generate and that it should be considered as a resource like space and time. Moreover since the perfect randomness abstraction is not physically realizable, it is interesting to determine whether imperfect randomness is “good enough” for certain cryptographic algorithms and to design algorithms that are robust with respect to deviations of the random sources from true randomness.

The power of randomness in computation is a central problem in complexity theory and in cryptography. Cryptographers should definitely take these considerations into account when proposing new cryptographic schemes: there exist computational tasks that we only know how to perform efficiently using randomness but conversely it is sometimes possible to remove randomness from probabilistic algorithms to obtain efficient deterministic counterparts. Since these constructions may hinder the security of cryptographic schemes, it is of high interest to study the efficiency/security tradeoff provided by randomness in cryptography.

Quite often in practice, the random bits in cryptographic protocols are generated by a pseudorandom number generation process. When this is done, the security of the scheme of course depends in a crucial way on the quality of the random bits produced by the generator. Despite the importance, many protocols used in practice often leave unspecified what pseudorandom number generation to use. It is well-known that pseudorandom generators exist if and only if one-way functions exist and there exist efficient constructions based on various number-theoretic assumptions. Unfortunately, these constructions are too inefficient and many protocols used in practice rely on "ad-hoc" constructions. It is therefore interesting to propose more efficient constructions, to analyze the security of existing ones and of specific cryptographic constructions that use weak pseudorandom number generators.

The project-team undertakes research in these three aspects. The approach adopted is both theoretical and practical, since we provide security results in a mathematical framework (information theoretic or computational) with the aim to design protocols among the most efficient known.

3.2. 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 discrete log. This is somewhat 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 —in particular since they also offer features that other alternative public-key cryptosystems (such as MQ-based, code-based or hash-based schemes) cannot provide.

3.3. Advanced Encryption

Fully Homomorphic Encryption (FHE) is a very active research area. Let us just give one example illustrating the usefulness of computing on encrypted data: Consider an on-line patent database on which firms perform complex novelty queries before filing patents. With current technologies, the database owner might analyze the queries, infer the invention and apply for a patent before the genuine inventor. While such frauds were not reported so far, similar incidents happen during domain name registration. Several websites propose "registration services" preceded by "availability searches". These queries trigger the automated registration of the searched domain names which are then proposed for sale. Algorithms allowing arbitrary computations without disclosing their inputs (and/or their results) are hence of immediate usefulness.

In 2009, IBM announced the discovery of a FHE scheme by Craig Gentry. The security of this algorithm relies on worst-case problems over ideal lattices and on the hardness of the sparse subset sum problem. Gentry's construction is an ingenious combination of two ideas: a somewhat homomorphic scheme (capable of supporting many "logical or" operations but very few "ands") and a procedure that refreshes the homomorphically processed ciphertexts. Gentry's main conceptual achievement is a "bootstrapping" process in which the somewhat homomorphic scheme evaluates its own decryption circuit (self-reference) to refresh (recrypt) ciphertexts.

Unfortunately, it is safe to surmise that if the state of affairs remains as it is in the present, then despite all the theoretical efforts that went into their constructions, these schemes will never be used in practical applications.

Our team is looking at the foundations of these primitives with the hope of achieving a breakthrough that will allow them to be practical in the near future.

But FHE builds new ciphertexts from ciphertexts, and the initial user only can decrypt the result. A more recent primitive has be defined, under the name of "Functional Encryption". It allows users to encrypt messages and an authority to distribute functional decryption keys. The latter only allow recipient of ciphertexts to obtain an evaluation of the plaintexts, according to the functions associated to the functional decryption keys: this for example allows some people to have access to various aggregations of data, in clear, from an encrypted database.

While this functionality has initially been defined in theory, our team is actively working on designing concrete instantiations for practical purpose.

3.4. Security amidst Concurrency on the Internet

Cryptographic protocols that are secure when executed in isolation, can be 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 as well as 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,
2. and 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.

3.5. Electronic Currencies and the Blockchain

Electronic cash (e-cash) was first proposed in the 1980s but despite extensive research it has never been deployed on a large scale. Other means of digital payments have instead largely replaced physical cash. Common to all digital payments offered by banks and other payment providers is that they do not respect the citizens' right to privacy, which for legitimate purchases and moderate sums also includes their right of anonymous payments.

The rise of so-called decentralized currencies, such as Bitcoin and the numerous "alt-coins" inspired by it, have established a third way of payments in addition to physical cash, which offers privacy, and card and other electronic payments, which are traceable by its providers. The continuous growth of popularity and usage of this new kind of currencies, also called "cryptocurrencies" as their security and stability crucially relies on the use of cryptography, 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. In contrast, cryptocurrencies use money distribution as an incentive for widespread 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). Currently all efficient e-cash schemes require coins to be deposited at the bank once received; they are thus not transferable. Our goal is to propose efficient transferable e-cash schemes.

Another direction concerns (decentralized) cryptocurrencies whose adoption is continuously growing so that now even central banks, like the Swedish *Riksbank*, are considering issuing their own currency as a cryptocurrency. While systems like Bitcoin are perceived as offering anonymous payments, a line of research has shown that this is not the case. One of the major research challenges in this area is to devise schemes with an anonymity level comparable to that of physical cash. The currently proposed schemes either lack formal security analyses or they are inefficient due to the heavy-duty cryptography used. Our group works towards practical cryptocurrencies with formally analyzed privacy guarantees.

Cryptocurrencies rely on a decentralized data structure called the "blockchain", which has 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.

4.2. Hardware Security

Cryptography is only one component of information security, but it is a crucial component. Without cryptography, it would be impossible to establish secure communications between users over insecure networks like the Internet. In particular, public-key cryptography (invented by Diffie and Hellman in 1976) enables to establish secure communications between users who have never met physically before. One can argue that companies like E-Bay or Amazon could not exist without public-key cryptography. Since 30 years the theory of cryptography has developed considerably. However cryptography is not only a theoretical science; namely at some point the cryptographic algorithms must be implemented on physical devices, such as PCs, smart cards or RFIDs. Then problems arise: in general smart cards and RFIDs have limited computing power and leak information through power consumption and electro-magnetic radiations. Similarly a PC can be exposed to various computer viruses which can leak private informations to a remote attacker. Such information leakage can be exploited by an attacker; this is called a **side-channel attack**. It is well known that a cryptographic algorithm which is perfectly secure in theory can be completely insecure in practice if improperly implemented.

In general, countermeasures against side-channel attacks are heuristic and can only make a particular implementation resist particular attacks. Instead of relying on ad-hoc security patches, a better approach consists in working in the framework of **provable security**. The goal is to prove that a cryptosystem does not only resist specific attacks but can resist any possible side-channel attack. As already demonstrated with cryptographic protocols, this approach has the potential to significantly increase the security level of cryptographic products. Recently the cryptography research community has developed new security models to take into account these practical implementation attacks; the most promising such model is called the **leakage-resilient model**.

Therefore, our goal is to define new security models that take into account any possible side-channel attack, and then to design new cryptographic schemes and countermeasures with a proven security guarantee against side-channel attacks.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Conferences

- We organised the Eurocrypt '17 annual conference in Paris, from April 30 to May 4.

5.1.2. Awards

- Damien Vergnaud was nominated for a 5 year appointment as Junior Member of the Institut Universitaire de France
- Romain Gay received a Google PhD Fellowship.

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):

- More efficient constructions with lattices
- New e-cash constructions
- Advanced primitives for the privacy in the cloud
- Efficient functional encryption
- Various predicate encryption schemes

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. 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 within Academics

7.2.1. EnBiD

Title: Encryption for Big Data

Program: ANR JCJC

Duration: October 2014 – September 2018

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 even 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 2019

PI: David Pointcheval

The goal of the CryptoCloud project is to develop new interactive tools to provide privacy to 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 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 amount of IOT devices.

7.4. International Initiatives with Industry

7.4.1. CryPrivBC

Title: Cryptography for Privacy on 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. 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

- Melissa Chase (MSR Redmond)
- Huijia Rachel Lin (UCSB)
- Yuval Ishai (Technion)

- Stefano Tessaro (UCSB)
- Vinod Vaikuntanathan (MIT)

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events Organisation

8.1.1.1. Events and Activities

- a regular seminar is organized: http://www.di.ens.fr/CryptoSeminaire.html
- quarterly Paris Crypto Days (https://pariscryptoday.github.io) supported by CryptoCloud and aSCEND
- working group on lattices (http://perso.ens-lyon.fr/damien.stehle/LATTICE_MEETINGS.html), joint with ENS Lyon
- BibTeX database of papers related to Cryptography, open and widely used by the community (https://cryptobib.di.ens.fr)

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. Other Steering Committees

- steering committee of the Coding and Cryptography working group (GT-C2 - https://crypto.di.ens.fr/c2:main) of the *Groupe de Recherche Informatique Mathématique* (GDR-IM): Damien Vergnaud is the Head of this steering committee

8.1.1.4. 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 '17 – 14-17 February (San Francisco, California, USA): David Pointcheval
- TCC '17 – 12-15 November (Baltimore, Maryland, USA): Hoeteck Wee
- ICALP '17 – 10-14 July (Warsaw, Poland): Hoeteck Wee
- Euro S&P '17 – 26-28 Apr (Paris, France) : Hoeteck Wee
- PKC '17 – 28-31 March (Amsterdam, Netherlands): Hoeteck Wee, Georg Fuchsbauer
- ASIACRYPT '17 – 3-7 December (Hong Kong): Georg Fuchsbauer
- ACNS '17 – 10-12 July (Kanazawa, Japan): Georg Fuchsbauer
- Indocrypt '17 – 10-13 Dec (Chennai, India): Georg Fuchsbauer
- Africacrypt '17 – 24-26 May (Dakar, Senegal): Georg Fuchsbauer

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
- of *Journal of Cryptographic Engineering*: Damien Vergnaud

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

- Master: David Pointcheval, Jacques Stern, Damien Vergnaud, Introduction to Cryptology, M1, ENS
- Master: Michel Abdalla, David Pointcheval, Cryptography, M2, MPRI
- Master: Damien Vergnaud, Advanced Algebra and Applications to Cryptography, Ecole Centrale Paris
- Master: David Pointcheval, Cryptography, M2, ESIEA
- Bachelor: Georg Fuchsbauer, Cryptologie, 3e année, ESGI

8.2.2. Defenses

- PhD: Thierry Mefenza Nountu, Pseudo-Random Generators and Pseudo-Random Functions : Cryptanalysis and Complexity Measures, ENS, November 28th, 2017 (Supervisor: Damien Vergnaud)
- PhD: Geoffroy Couteau, Zero-Knowledge Proofs for Secure Computation, ENS, November 30th, 2017 (Supervisor: David Pointcheval & Hoeteck Wee)
- PhD: Pierrick Méaux, Hybrid Fully Homomorphic Framework, ENS, December 8th, 2017 (Supervisor: Vadim Lyubashevsky & David Pointcheval)
- HdR: Céline Chevalier, UC-Secure Protocols using Smooth Projective Hash Functions, ENS, December 11th, 2017 (Supervisor: David Pointcheval)
- PhD: Florian Bourse, Functional Encryption for Inner-Product Evaluations, ENS, December 13th, 2017 (Supervisors: Michel Abdalla & David Pointcheval)

8.2.3. Supervision

- PhD in progress: Raphael Bost, Symmetric Searchable Encryption, from 2014, David Pointcheval (with Pierre-Alain Fouque, at Rennes)
- PhD in progress: Rafael Del Pino, Lattice-Based Cryptography – Complexity and Ideal-Lattices, from 2014, Vadim Lyubashevsky
- PhD in progress: Aurélien Dupin, Multi-Party Computations, from 2015, David Pointcheval (with Christophe Bidan, at Rennes)
- PhD in progress: Pierre-Alain Dupont, Secure Communications, from 2015, David Pointcheval
- PhD in progress: Romain Gay, Functional Encryption, from 2015, Michel Abdalla & Hoeteck Wee
- PhD in progress: Dahmun Goudarzi, Secure and Fast Cryptographic Implementation for Embedded Devices, from 2015, Damien Vergnaud
- PhD in progress: Louiza Khati, Disk Encryption Modes, from 2015, Damien Vergnaud
- PhD in progress: Michele Minelli, Increased efficiency and functionality through lattice-based cryptography, from 2015, Michel Abdalla & Hoeteck Wee
- PhD in progress: Anca Nitulescu, Verifiable Outsourced Computations, from 2015, David Pointcheval
- PhD in progress: Razvan Rosie, Practical Functional Encryption Schemes For the Cloud, from 2015, Michel Abdalla & Hoeteck Wee

- PhD in progress: Quentin Santos, Advanced Cryptography from a Blockchain, from 2015, David Pointcheval
- 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)

8.2.4. Committees

- PhD Afonso Delerue Arriaga. *Private Functional Encryption* – Luxembourg University - Luxembourg – January 17th, 2017: Michel Abdalla (Examiner)
- HdR Maria Naya-Plasencia. *Symmetric Cryptography for Long- Term Security* – Université Paris VI – France – May 5th, 2017: David Pointcheval (Reviewer)
- PhD Benjamin Richard. *Study of 3-Party Authentication and Key-Derivation Protocols* – Université Rennes 1 - France – August 30th, 2017: Michel Abdalla (Reviewer)
- PhD Thierry Mefenza Nountu. *Pseudo-Random Generators and Pseudo-Random Functions: Cryptanalysis and Complexity Measures* – ENS - France – November 28th, 2017: Damien Vergnaud (Supervisor)
- PhD Geoffroy Couteau. *Zero-Knowledge Proofs for Secure Computation* – ENS - France – November 30th, 2017: David Pointcheval & Hoeteck Wee (Supervisors)
- PhD Britta Hale. *Low-Latency Key Exchange and Secure Channels* – NTNU - Norway – December 5th, 2017: Michel Abdalla (Reviewer)
- PhD Pierrick Méaux. *Hybrid Fully Homomorphic Framework* – ENS - France – December 8th, 2017: David Pointcheval (Supervisor)
- HdR Céline Chevalier. *UC-Secure Protocols using Smooth Projective Hash Functions* – ENS - France – December 11th, 2017: David Pointcheval
- PhD Florian Bourse. *Functional Encryption for Inner-Product Evaluations* – ENS - France – December 13th, 2017: Michel Abdalla & David Pointcheval (Supervisors)

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^{o} 6, p. 43:1–43:35.

[10] V. LYUBASHEVSKY, T. PREST.*Quadratic Time, Linear Space Algorithms for Gram-Schmidt Orthogonalization and Gaussian Sampling in Structured Lattices*, in "Advances in Cryptology – Proceedings of Eurocrypt '15 (1)", E. OSWALD, M. FISCHLIN (editors), Lecture Notes in Computer Science, Springer, 2015, vol. 9056, p. 789-815.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] F. BOURSE.*Functional Encryption for Inner-Product Evaluations*, Université de recherche Paris Sciences et Lettres, December 2017, https://hal.archives-ouvertes.fr/tel-01665276.

[12] G. COUTEAU.*Zero-Knowledge Proofs for Secure Computation*, PSL research University, November 2017, https://hal.inria.fr/tel-01668125.

[13] T. MEFENZA NOUNTU.*Pseudo-Random Generators and Pseudo-Random Functions: Cryptanalysis and Complexity Measures*, Paris Sciences et Lettres, November 2017, https://hal.inria.fr/tel-01667124.

[14] P. MÉAUX.*Hybrid fully homomorphic framework*, Université de recherche Paris Sciences et Lettres, December 2017, https://hal.archives-ouvertes.fr/tel-01665358.

Articles in International Peer-Reviewed Journal

[15] M. ABDALLA, M. BELLARE, G. NEVEN.*Robust Encryption*, in "Journal of Cryptology", June 2017 [*DOI :* 10.1007/s00145-017-9258-8], https://hal.archives-ouvertes.fr/hal-01538113.

[16] C. CARLET, P. MÉAUX, Y. ROTELLA.*Boolean functions with restricted input and their robustness; application to the FLIP cipher*, in "IACR Transactions on Symmetric Cryptology", 2017, vol. 2017, n° 3, p. 192–227 [*DOI :* 10.13154/TOSC.V2017.I3.192-227], https://hal.inria.fr/hal-01633506.

[17] P. FARSHIM, L. KHATI, D. VERGNAUD.*Security of Even-Mansour Ciphers under Key-Dependent Messages*, in "IACR Transactions on Symmetric Cryptology", 2017, vol. 2017, n° 2, p. 84-104 [*DOI :* 10.13154/TOSC.V2017.I2.84-104], https://hal.inria.fr/hal-01613782.

[18] P. FARSHIM, C. ORLANDI, R. ROŞIE.*Security of Symmetric Primitives under Incorrect Usage of Keys*, in "IACR Transactions on Symmetric Cryptology", 2017 [*DOI :* 10.13154/TOSC.V2017.I1.449-473], https://hal-ens.archives-ouvertes.fr/hal-01470885.

[19] T. MEFENZA, D. VERGNAUD.*Polynomial Interpolation of the Naor-Reingold Pseudo-Random Function*, in "Applicable Algebra in Engineering, Communication and Computing", June 2017, vol. 28, n° 3, p. 237-255 [*DOI :* 10.1007/S00200-016-0309-4], https://hal.inria.fr/hal-01550044.

[20] D. VERGNAUD.*Comment on 'Attribute-Based Signatures for Supporting Anonymous Certification' by N. Kaaniche and M. Laurent (ESORICS 2016)* , in "The Computer Journal (Oxford)", 2017 [*DOI :* 10.1093/COMJNL/BXX058], https://hal.inria.fr/hal-01613813.

International Conferences with Proceedings

[21] M. ABDALLA, F. BENHAMOUDA, D. POINTCHEVAL.*Removing Erasures with Explainable Hash Proof Systems*, in "Public-Key Cryptography - PKC 2017 - 20th International Conference on Practice and Theory in Public-Key Cryptography", Amsterdam, Netherlands, S. FEHR (editor), Springer, March 2017, vol. 10174, n° Part I, p. 151-174 [*DOI :* 10.1007/978-3-662-54365-8_7], https://hal.inria.fr/hal-01471896.

[22] M. ABDALLA, R. GAY, M. RAYKOVA, H. WEE.*Multi-Input Inner-Product Functional Encryption from Pairings*, in "EUROCRYPT 2017 - Advances in Cryptology", Paris, France, Lecture Notes in Computer Science, April 2017, vol. 10210, p. 601-626 [*DOI :* 10.1007/978-3-319-56620-7_21], https://hal.archives-ouvertes.fr/hal-01524105.

[23] M. AMBRONA, G. BARTHE, R. GAY, H. WEE.*Attribute-Based Encryption in the Generic Group Model: Automated Proofs and New Constructions*, in "ACM Conference on Computer and Communications Security (CCS) 2017", Dallas, United States, October 2017, https://hal.inria.fr/hal-01599851.

[24] S. BELAID, F. BENHAMOUDA, A. PASSELÈGUE, E. PROUFF, A. THILLARD, D. VERGNAUD.*Private Multiplication over Finite Fields*, in "Advances in Cryptology - CRYPTO 2017", Santa Barbara, United States, J. KATZ, H. SHACHAM (editors), Lecture Notes in Computer Science, Springer, August 2017, vol. 10403, p. 397-426 [*DOI :* 10.1007/978-3-319-63697-9_14], https://hal.inria.fr/hal-01613773.

[25] F. BENHAMOUDA, F. BOURSE, H. LIPMAA.*CCA-Secure Inner-Product Functional Encryption from Projective Hash Functions*, in "PKC 2017 - IACR International Workshop on Public Key Cryptography", Amsterdam, Netherlands, Lecture Notes in Computer Science, Springer, March 2017, vol. 10175, p. 36-66 [*DOI :* 10.1007/978-3-662-54388-7_2], https://hal.archives-ouvertes.fr/hal-01613546.

[26] M. BEUNARDEAU, A. CONNOLLY, H. FERRADI, R. GÉRAUD, D. NACCACHE, D. VERGNAUD.*Reusing Nonces in Schnorr Signatures: (and Keeping It Secure...)*, in "Computer Security - ESORICS 2017", Oslo, Norway, S. N. FOLEY, D. GOLLMANN, E. SNEKKENES (editors), Lecture Notes in Computer Science,

Springer, September 2017, vol. 10492, p. 224-241 [*DOI :* 10.1007/978-3-319-66402-6_14], https://hal.inria.fr/hal-01613794.

[27] A. BOLDYREVA, S. CHEN, P.-A. DUPONT, D. POINTCHEVAL.*Human Computing for Handling Strong Corruptions in Authenticated Key Exchange*, in "CSF 2017 - 30th IEEE Computer Security Foundations Symposium", Santa Barbara, CA, United States, Proceedings of the 30th IEEE Computer Security Foundations Symposium (CSF '17), IEEE, August 2017, p. 159 - 175 [*DOI :* 10.1109/CSF.2017.31], https://hal.inria.fr/hal-01628797.

[28] E. BOYLE, G. COUTEAU, N. GILBOA, Y. ISHAI, M. ORRÙ.*Homomorphic Secret Sharing: Optimizations and Applications*, in "CCS '17 - ACM SIGSAC Conference on Computer and Communications Security", Dallas, United States, ACM, October 2017, https://hal.inria.fr/hal-01614451.

[29] J. CHOTARD, D. H. PHAN, D. POINTCHEVAL.*Homomorphic-Policy Attribute-Based Key Encapsulation Mechanisms*, in "20th International Conference on Information Security (ISC '17)", Ho Chi Minh, Vietnam, Proceedings of the 20th International Conference on Information Security (ISC '17), Springer Verlag, November 2017, vol. Lecture Notes in Computer Science, n^{o} 10599 [*DOI :* 10.1007/978-3-319-69659-1_9], https://hal.inria.fr/hal-01609278.

[30] G. COUTEAU, T. PETERS, D. POINTCHEVAL.*Removing the Strong RSA Assumption from Arguments over the Integers*, in "EUROCRYPT 2017 - 36th Annual International Conference on the Theory and Applications of Cryptographic Techniques", Paris, France, April 2017 [*DOI :* 10.1007/978-3-319-56614-6_11], https://hal.inria.fr/hal-01471901.

[31] P.-A. DUPONT, D. POINTCHEVAL.*Functional Encryption with Oblivious Helper*, in "ASIA CCS'17", Abu Dhabi, United Arab Emirates, April 2017 [*DOI :* 10.1145/3052973.3052996], https://hal.inria.fr/hal-01470375.

[32] C. ELISABETTA ZAIRA BALTICO, D. CATALANO, D. FIORE, R. GAY.*Practical Functional Encryption for Quadratic Functions withApplications to Predicate Encryption*, in "Advances in Cryptology – CRYPTO 2017", Santa Barbara, United States, Springer, August 2017, vol. LNCS, n^{o} 10401, p. 67-98, https://hal.inria.fr/hal-01599768.

[33] G. FUCHSBAUER, R. GAY, L. KOWALCZYK, C. ORLANDI.*Access Control Encryption for Equality, Comparison, and More*, in "Public-Key Cryptography - PKC 2017 - 20th International Conference on Practice and Theory in Public-Key Cryptography", Amsterdam, Netherlands, March 2017, p. 88-118 [*DOI :* 10.1007/978-3-662-54388-7_4], https://hal.inria.fr/hal-01470315.

[34] R. GAY, D. HOFHEINZ, L. KOHL.*Kurosawa-Desmedt Meets Tight Security*, in "CRYPTO 2017 - Advances in Cryptology", Santa Barbara, United States, LNCS, Springer, August 2017, vol. 10403, p. 133-160 [*DOI :* 10.1007], https://hal.inria.fr/hal-01599757.

[35] D. GOUDARZI, M. RIVAIN.*How Fast Can Higher-Order Masking Be in Software?* , in "Advances in Cryptology - EUROCRYPT 2017", Paris, France, Advances in Cryptology - EUROCRYPT 2017, April 2017, vol. Lecture Notes in Computer Science, n^{o} 10210, https://hal.archives-ouvertes.fr/hal-01494061.

[36] D. GOUDARZI, M. RIVAIN, D. VERGNAUD, S. VIVEK.*Generalized Polynomial Decomposition for S-boxes with Application to Side-Channel Countermeasures*, in "Cryptographic Hardware and Embedded Systems - CHES 2017", Taipei, Taiwan, W. FISCHER, N. HOMMA (editors), Lecture Notes in Computer Science,

Springer, September 2017, vol. 10529, p. 154-171 [*DOI :* 10.1007/978-3-319-66787-4_8], https://hal.inria.fr/hal-01613764.

[37] R. GÉRAUD, D. NACCACHE, R. ROŞIE.*Twisting Lattice and Graph Techniques to Compress Transactional Ledgers*, in "SecureComm 2017 - 13th EAI International Conference on Security and Privacy in Communication Networks", Niagara Falls, Canada, October 2017, p. 1-20, https://hal.inria.fr/hal-01668213.

[38] L. KHATI, N. MOUHA, D. VERGNAUD.*Full Disk Encryption: Bridging Theory and Practice*, in "CT-RSA 2017 - RSA Conference Cryptographers' Track", San Francisco, United States, H. HANDSCHUH (editor), Lecture Notes in Computer Science, Springer, February 2017, vol. 10159, p. 241–257 [*DOI :* 10.1007/978-3-319-52153-4_14], https://hal.inria.fr/hal-01403418.

[39] T. LIU, V. VAIKUNTANATHAN, H. WEE.*Conditional Disclosure of Secrets via Non-linear Reconstruction*, in "Advances in Cryptology - CRYPTO 2017", Santa Barbara, United States, August 2017 [*DOI :* 10.1007/978-3-319-63688-7_25], https://hal.archives-ouvertes.fr/hal-01619913.

[40] D. POINTCHEVAL, O. SANDERS, J. TRAORÉ.*Cut Down the Tree to Achieve Constant Complexity in Divisible E-Cash*, in "Public-Key Cryptography - PKC 2017 - 20th International Conference on Practice and Theory in Public-Key Cryptography", Amsterdam, Netherlands, S. FEHR (editor), Springer, March 2017, vol. 10174, n^{o} Part I, p. 61-90 [*DOI :* 10.1007/978-3-662-54365-8_4], https://hal.inria.fr/hal-01471876.

[41] D. POINTCHEVAL, G. WANG.*VTBPEKE: Verifier-based Two-Basis Password Exponential Key Exchange*, in "ASIA CCS'17", Abu Dhabi, United Arab Emirates, April 2017 [*DOI :* 10.1145/3052973.3053026], https://hal.inria.fr/hal-01471737.

[42] M. ROSSI, M. HAMBURG, M. HUTTER, M. E. MARSON.*A Side-Channel Assisted Cryptanalytic Attack Against QcBits*, in "CHES 2017 - Conference on Cryptographic Hardware and Embedded Systems", Taipei, Taiwan, Cryptographic Hardware and Embedded Systems - CHES 2017, September 2017, vol. Lecture Notes in Computer Science, n^{o} 10529, 22 [*DOI :* 10.1007/978-3-319-66787-4_1], https://hal.inria.fr/hal-01614569.

Books or Proceedings Editing

[43] P. FARSHIM, E. SIMION (editors). *Innovative Security Solutions for Information Technology and Communications*, Springer, Romania, August 2017, https://hal-ens.archives-ouvertes.fr/hal-01671056.

Research Reports

[44] M. ABDALLA, F. BENHAMOUDA, A. PASSELÈGUE.*Algebraic XOR-RKA-Secure Pseudorandom Functions from Post-Zeroizing Multilinear Maps*, IACR Cryptology ePrint Archive, June 2017, n^{o} Report 2017/500, https://hal.inria.fr/hal-01667132.

[45] M. ABDALLA, F. BENHAMOUDA, D. POINTCHEVAL.*On the Tightness of Forward-Secure Signature Reductions*, IACR Cryptology ePrint Archive, August 2017, n^{o} Report 2017/746, https://hal.inria.fr/hal-01667150.

[46] M. ABDALLA, D. CATALANO, D. FIORE, R. GAY, B. URSU.*Multi-Input Functional Encryption for Inner Products: Function-Hiding Realizations and Constructions without Pairings*, IACR Cryptology ePrint Archive, October 2017, n^{o} Report 2017/972, https://hal.inria.fr/hal-01667169.

[47] F. BOURSE, M. MINELLI, M. MINIHOLD, P. PAILLIER.*Fast Homomorphic Evaluation of Deep Discretized Neural Networks*, IACR Cryptology ePrint Archive, November 2017, n^{o} Report 2017/1114, https://hal.archives-ouvertes.fr/hal-01665330.

[48] J. CHOTARD, E. DUFOUR SANS, D. POINTCHEVAL, D. H. PHAN.*Decentralized Multi-Client Functional Encryption for Inner Product*, IACR Cryptology ePrint Archive, October 2017, n^{o} 2017/989, https://hal.archives-ouvertes.fr/hal-01668020.

[49] P.-A. DUPONT, J. HESSE, D. POINTCHEVAL, L. REYZIN, S. YAKOUBOV.*Fuzzy Authenticated Key Exchange*, IACR Cryptology ePrint Archive, November 2017, n^{o} 2017/1111, https://hal.archives-ouvertes.fr/hal-01668008.

Team COML

Cognitive Machine Learning

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
A5.9. - Signal processing
A5.9.1. - Sampling, acquisition
A5.9.2. - Estimation, modeling
A5.9.3. - Reconstruction, enhancement
A5.9.4. - Signal processing over graphs
A5.9.5. - Sparsity-aware processing
A5.9.6. - Optimization tools
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. Personnel

Faculty Member

Emmanuel Dupoux [Team leader, Ecole Normale Supérieure Paris, Professor, from May 2017, HDR]

Technical Staff

Mathieu Bernard [Inria, from Nov 2017]
Julien Karadayi [Ecole Normale Supérieure]
Xuan Nga Cao [Ecole Normale Supérieure]

PhD Students

Maria Julia Carbajal [Ecole Normale Supérieure Paris]
Adriana Carolina Guevara Rukoz [Ecole Normale Supérieure Paris]
Neil Zeghidour [CIFRE Facebook; Ecole Normale Supérieure Paris]

Rahma Chaabouni [CIFRE Facebook; Ecole Normale Supérieure Paris, from Sep 2017]
Ronan Riochet [Ecole Normale Supérieure Paris, from Sep 2017; part time with the WILLOW team]
Elin Larsen [Ecole Normale Supérieure Paris, from Oct 2017]

Post-Doctoral Fellow

Bogdan Ludusan [CNRS, until Oct 2017]

Administrative Assistants

Catherine Urban [Ecole Normale Supérieure]
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., [72], [75], [80], [71], [77]). 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)[15]; 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) [74].

The general aim of CoML, following the roadmap described in [15], 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 [21], [40], [54], 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 [78].

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 [79] 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* [73] , i.e., functional components within a more global cognitive architecture [76]. 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 [46], [28].

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. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

BEST PAPERS AWARDS :
[67] **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**. B. LUDUSAN, R. MAZUKA, M. BERNARD, A. CRISTIA, E. DUPOUX.

6. New Software and Platforms

6.1. 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

6.2. 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

6.3. 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

6.4. 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

7. New Results

7.1. 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:

- Recent works have explored deep architectures for learning multimodal speech representation (e.g. audio and images, articulation and audio) in a supervised way. In [63], we investigate the role of combining different speech modalities, i.e. audio and visual information representing the lips' movements, in a weakly-supervised way using Siamese networks and lexical same-different side information. In particular, we ask whether one modality can benefit from the other to provide a richer representation for phone recognition in a weakly supervised setting. We introduce mono-task and multi-task methods for merging speech and visual modalities for phone recognition. The mono-task learning consists in applying a Siamese network on the concatenation of the two modalities, while the multi-task learning receives several different combinations of modalities at train time. We show that multi-task learning enhances discriminability for visual and multimodal inputs while minimally impacting auditory inputs. Furthermore, we present a qualitative analysis of the obtained phone embeddings, and show that cross-modal visual input can improve the discriminability of phonetic features which are visually discernable (rounding, open/close, labial place of articulation), resulting in representations that are closer to abstract linguistic features than those based on audio only.
- In [67] , we explore the role of speech register and prosody for the task of word segmentation. Since these two factors are thought to play an important role in early language acquisition, we aim to quantify their contribution for this task. We study a Japanese corpus containing both infant- and adult-directed speech and we apply four different word segmentation models, with and without knowledge of prosodic boundaries. The results showed that the difference between registers is smaller than previously reported and that prosodic boundary information helps more adult- than infant-directed speech.
- Phonemic segmentation of speech is a critical step of speech recognition systems. In [68], we propose a novel unsupervised algorithm based on sequence prediction mod- els such as Markov chains and recurrent neural networks. Our approach consists in analyzing the error profile of a model trained to predict speech features frame- by-frame. Specifically, we try to learn the dynamics of speech in the MFCC space and hypothesize boundaries from lo- cal maxima in the prediction error. We evaluate our system on the TIMIT dataset, with improvements over similar methods.

- In [70], we describe a new challenge aimed at discovering subword and word units from raw speech. This challenge is the follow-up to the Zero Resource Speech Challenge 2015. It aims at constructing systems that generalize across languages and adapt to new speakers. The design features and evaluation metrics of the challenge are presented and the results of seventeen models are discussed.

7.2. Evaluation of AI algorithms

Machine learning algorithms are typically evaluated in terms of end-to-end tasks, but it is very often difficult to get a grasp of how they achieve these tasks, what could be their break point, and more generally, how they would compare to the algorithms used by humans to do the same tasks. This is especially true of Deep Learning systems which are particularly opaque. The team develops evaluation methods based on psycholinguistic/linguistic criteria, and deploy them for systematic comparison of systems.

- What is the information captured by neural network models of language? In [66], we address this question in the case of character-level recurrent neural language models. These models do not have explicit word representations; do they acquire implicit ones? We assess the lexical capacity of a network using the lexical decision task common in psycholinguistics: the system is required to decide whether or not a string of characters forms a word. We explore how accuracy on this task is affected by the architecture of the network, focusing on cell type (LSTM vs. SRN), depth and width. We also compare these architectural properties to a simple count of the parameters of the network. The overall number of parameters in the network turns out to be the most important predictor of accuracy; in particular, there is little evidence that deeper networks are beneficial for this task.

7.3. Learnability relevant descriptions of linguistic corpora

Evidently, infants are acquiring their language based on whatever linguistic input is available around them. The extent of variation that can be found across languages, cultures and socio-economic background provides strong constraints (lower bounds on data, higher bounds on noise, and variation and ambiguity) for language learning algorithms.

- In [60], we provide an estimation of how frequently, and from whom, children aged 0-11 years (Ns between 9 and 24) receive one-on-one verbal input among Tsimane forager-horticulturalists of lowland Bolivia. Analyses of systematic daytime behavioral observations reveal $<$ 1 min per daylight hour is spent talking to children younger than 4 years of age, which is 4 times less than estimates for others present at the same time and place. Adults provide a majority of the input at 0–3 years of age but not afterward. When integrated with previous work, these results reveal large cross-cultural variation in the linguistic experiences provided to young children. Consideration of more diverse human populations is necessary to build generalizable theories of language acquisition.
- In [69], we provide a new measure of how the acoustic realizations of a given phonetic segment are affected by coarticulation with the preceding and following phonetic context. While coarticulation has been extensively studied using descriptive phonetic measurements, little is known about the functional impact of coarticulation for speech processing, and in particular, learnability. Here, we use DTW-based similarity defined on raw acoustic features and ABX scores to derive a measure of the effect of coarticulation on phonetic discriminability. This measure does not rely on defining segment-specific phonetic cues (formants, duration, etc.) and can be applied systematically and automatically to any segment in large scale corpora. We illustrate our method using stimuli in English and Japanese. We replicate some well-known results, i.e., stronger anticipatory than perseveratory coarticulation and stronger coarticulation for lax/short vowels than for tense/long vowels. We then quantify for the first time the impact of coarticulation across different segment types (like vowels and consonants).

7.4. 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).

- In [61] we aim to quantify the relative contributions of phonetic categories and acoustic detail on phonotactically induced perceptual vowel epenthesis in Japanese listeners. A vowel identification task tested whether a vowel was perceived within illegal consonant clusters and, if so, which vowel was heard. Cross-spliced stimuli were used in which vowel coarticulation present in the cluster did not match the quality of the flanking vowel. Two clusters were used, /hp/ and /kp/, the former containing larger amounts of resonances of the preceding vowel. While both flanking vowel and coarticulation influenced vowel quality, the influence of coarticulation was larger, especially for /hp/.
- In [64], we explore the well documented example of vowel epenthesis, a phenomenon in which non-existent vowels are hallucinated by listeners, for stimuli containingr illegal consonantal sequences. As reported in previous work, this occurs in Japanese (JP) and Brazilian Portuguese (BP), languages for which the 'default' epenthetic vowels are /u/ and /i/, respectively. In a perceptual experiment, we corroborate the finding that the quality of this illusory vowel is language-dependent, but also that this default choice can be overridden by coarticulatory information present on the consonant cluster. In a second step, we analyse recordings of JP and BP speakers producing 'epenthesized' versions of stimuli from the perceptual task. Results reveal that the default vowel corresponds to the vowel with the most reduced acoustic characteristics, also the one for which formants are acoustically closest to formant transitions present in consonantal clusters. Lastly, we model behavioural responses from the perceptual experiment with an exemplar model using dynamic time warping (DTW)-based similarity measures on MFCCs.
- A range of computational approaches have been used to model the discovery of word forms from continuous speech by infants. Typically, these algorithms are evaluated with respect to the ideal 'gold standard' word segmentation and lexicon. These metrics assess how well an algorithm matches the adult state, but may not reflect the intermediate states of the child's lexical development. In [65], we set up a new evaluation method based on the correlation between word frequency counts derived from the application of an algorithm onto a corpus of child-directed speech, and the proportion of infants knowing the words according to parental reports. We evaluate a representative set of 4 algorithms, applied to transcriptions of the Brent corpus, which have been phonologized using either phonemes or syllables as basic units. Results show remarkable variation in the extent to which these 8 algorithm-unit combinations predicted infant vocabulary, with some of these predictions surpassing those derived from the adult gold standard segmentation. We argue that infant vocabulary prediction provides a useful complement to traditional evaluation; for example, the best predictor model was also one of the worst in terms of segmentation score, and there was no clear relationship between token or boundary F-score and vocabulary prediction.
- A central assumption of most computational models of language acquisition is the reliance on statistical processes. This would predict that the frequency of particular sounds or contrasts in a given language should have a massive effect on perception. Surprisingly , this has not up to now been put to empirical test. In [62], we elucidated indicators of frequency-dependent perceptual attunement in the brain of 5–8-month-old Dutch infants. We tested the' discrimination of tokens containing a highly frequent [haet-he:t] and a highly infrequent [hYt-hø:t] native vowel contrast as well as a non-native [htˆ-hæt] vowel contrast in a behavioral visual habituation paradigm (Experiment 1). Infants discriminated both native contrasts similarly well, but did not discriminate the non-native contrast. We sought further evidence for subtle differences in the processing of the two native contrasts using near-infrared spectroscopy and a within-participant design (Experiment 2). The neuroimaging data did not provide additional evidence that responses to native contrasts are modulated by frequency of exposure. These results suggest that even large differences in exposure to a native contrast may not directly translate to behavioral and neural indicators of perceptual attunement, raising the possibility that frequency of exposure does not influence improvements in discriminating native contrasts.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

- Grant from MSR (Zero Resources Challenge, 2017) - 5K€
- AWS Grant (Zero Resources Challenge, 2017) - 20K€

9. Partnerships and Cooperations

9.1. 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

9.2. National Initiatives

9.2.1. 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€)

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

ERC Advanced Grant (BOOTPHON, PI: E. Dupoux, Budget 2.4M€).

9.4. International Initiatives

9.4.1. Informal International Partners

- Johns Hopkins University, Baltimore, USA: S. Kudanpur, H. Hermanksy
- RIKEN Institute, Tokyo, Japan: R. Mazuka

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Valentina Gliozzi (Professor, Univ. di Torino, Visiting Professor Spring 2017)

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Zero Resource Challenge 2017, held as a special session of EEE ASRU 2017, Okinawa.

10.1.1.2. Member of the Organizing Committees

Executive committee of SIGMORPHON (Association for Computational Linguistics Special Interest Group, http://www.sigmorphon.org/).

10.1.2. Scientific Events Selection

10.1.2.1. Reviewer

Invited editor for international conferences: Interspeech, NIPS, ACL, etc. (around 5-10 papers per conferences, 2 conferences per year)

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Member of the editorial board of: Mathématiques et Sciences Humaines, L'Année Psychologique, Frontiers in Psychology.

10.1.3.2. 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)

10.1.4. Invited Talks

- Learning in Machines and Brains (CIFAR) invited talk, 2017, Paris.
- CBMM (Center for Brain Mind and Machine) Workshop on Speech representation, perception and recognition. Invited talk. Feb 02-03, 2017, MIT

10.1.5. Scientific Expertise

E. Dupoux is invited expert for ERC, ANR, and other granting agencies (around 2 per year).

10.1.6. Research Administration

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).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence : E. Dupoux, "Introduction to the cognitive science of language", 8h, L2, PSL, France

Master : E. Dupoux, "Theoretical Cognitive Science: Connections and symbols", 8h, M1/M2, PSL,Paris 5, Paris France

Master : E. Dupoux, "Cognitive Engineering", 80h, M2, ITI-PSL, Paris France

Doctorat : E. Dupoux, "Computational models of cognitice development", 32 h, Séminaire EHESS, Paris France

10.2.2. Supervision

Six PhD theses are currently conducted in the team. Two are programmed to be defended in september 2018.

10.2.3. Juries

E. Dupoux participated in the PhD jury of Martin Felipe Perez-Guevara on novembre 29th, 2017 at UPMC (supervisor: C. Pallier).

10.3. Popularization

E. Dupoux talked in two general public conferences on speech recognition, one organised par France is AI (oct 17, 2017) and the other by Paris Sciences& Data (dec 7, 2017). The public was composed of entrepreneurs in machine learning. He also gave a training course in speech and language technology organised by the Institut de l'Ecole Normale (june 21) aimed at information technology professionals. He gave three interview on the limits of deep learning in general public outlets (Le Monde, La Recherche [0], Usine Nouvelle).

N. Zeghidour did a high level presentation of AI to 30 students and 8 teachers from Sciences-Po and Ecole 42 during the Policy Innovation Lab [0] following the visit of Facebook's COO. He has been a Mentor (technical advisor) for 9 European start-ups during a pitching event for the IBM Watson AI Xprize [0]. He animated a booth on AI at the Platform Meetup Paris of Facebook. He gave an interview to Libération included in a 100 page special issue on AI [0]. He taught a class on Speech Recognition as part of the Facebook AI Masterclass, a class for developers located at Ecole 42 and broadcasted in developer circles over 25 cities of Europe, Middle-Est and Africa with 850 people attending, followed by a live Q&A session with several of these cities over video-conference.

11. Bibliography

Major publications by the team in recent years

[1] C. BERGMANN, A. CRISTIA, E. DUPOUX.*Discriminability of sound contrasts in the face of speaker variation quantified*, in "Proceedings of the 38th Annual Conference of the Cognitive Science Society", 2016, p. 1331-1336.

[2] M. BUON, E. DUPOUX, P. JACOB, P. CHASTE, M. LEBOYER, T. ZALLA.*The role of causal and intentional reasoning in moral judgment in individuals with High Functioning Autism*, in "Journal of Autism and Developmental Disorders", 2013, vol. 43, n° 2, p. 458-70 [*DOI :* 10.1016/J.COGNITION.2012.09.006].

[3] M. BUON, P. JACOB, E. LOISSEL, E. DUPOUX.*A non-mentalistic cause-based heuristic in human social evaluations*, in "Cognition", 2013, vol. 126, n° 2, p. 149-155.

[4] M. BUON, P. JACOB, S. MARGULES, I. BRUNET, M. DUTAT, D. CABROL, E. DUPOUX.*Friend or foe? Early social evaluation of human interactions*, in "PloS One", 2014, vol. 9, n° 2, e88612.

[5] J. CARBAJAL, A. DAWUD, R. THIOLLIÈRE, E. DUPOUX.*The 'Language Filter' Hypothesis: Modeling Language Separation in Infants using I-vectors*, in "EPIROB 2016", 2016, p. 195-201.

[6] J. CARBAJAL, R. FÉR, E. DUPOUX.*Modeling language discrimination in infants using i-vector representations*, in "Proceedings of the 38th Annual Conference of the Cognitive Science Society", 2016, p. 889-896.

[7] L. CLERET DE LANGAVANT, C. JACQUEMOT, A.-C. BACHOUD-LÉVI, E. DUPOUX.*The second person in 'I'-'you'-'it' triadic interactions*, in "Behavioral and Brain Sciences", 2013, vol. 36, n° 416-417.

[0] http://www.larecherche.fr/lintelligence-artificielle-a-lassaut-des-labos
[0] http://www.sciencespo.fr/public/en/policy-lab/public-policy-incubator
[0] https://ai.xprize.org/
[0] http://www.liberation.fr/voyage-au-coeur-de-lIA/2017/12/20/en-kiosque-notre-hors-serie-voyage-au-coeur-de-l-ia_1617896

[8] A. CRISTIA, E. DUPOUX, Y. HAKUNA, S. LLOYD-FOX, M. SCHUETZE, J. KIVITS, T. BERGVELT, M. VAN GELDER, L. FILIPPIN, S. CHARRON, Y. MINAGAWA-KAWAI.*An online database of infant functional Near InfraRed Spectroscopy studies: A community-augmented systematic review.*, in "PLoS One", 2013, vol. 8, n^{o} 3, e58906.

[9] A. CRISTIA, Y. MINAGAWA-KAWAI, N. EGOROVA, J. GERVAIN, L. FILIPPIN, D. CABROL, E. DUPOUX.*Neural correlates of infant dialect discrimination: A fNIRS study*, in "Developmental Science", 2014, vol. 17, n^{o} 4, p. 628-635.

[10] A. CRISTIA, Y. MINAGAWA-KAWAI, I. VENDELIN, D. CABROL, E. DUPOUX.*Responses to vocalizations and auditory controls in the human newborn brain*, in "Plos One", 2014, vol. 9, n^{o} 12, e115162.

[11] E. DUNBAR, E. DUPOUX.*Geometric constraints on human speech sound inventories*, in "Frontiers in Psychology", 2016, vol. 7, n^{o} 1061 [*DOI :* 10.3389/FPSYG.2016.01061].

[12] E. DUNBAR, G. SYNNAEVE, E. DUPOUX.*Quantitative methods for comparing featural representations*, in "ICPhS", 2015, paper number 1024.

[13] E. DUPOUX.*Towards Quantitative Studies of Early Cognitive Development*, in "Autonomous Mental Development Technical Committee Newsletter", 2014, vol. 11, n^{o} 1, p. 10-11.

[14] E. DUPOUX.*Category Learning in Songbirds: top-down effects are not unique to humans*, in "Current Biology", 2015, vol. 25, n^{o} 16, p. R718-R720.

[15] E. DUPOUX.*Cognitive Science in the era of Artificial Intelligence: A roadmap for reverse-engineering the infant language-learner*, in "Cognition", 2018.

[16] A. FOURTASSI, B. BOERSCHINGER, M. JOHNSON, E. DUPOUX.*WhyisEnglishsoeasytosegment*, in "Proceedings of the 4th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2013)", Sofia, Bulgaria, ACL, 2013, p. 1-10.

[17] A. FOURTASSI, E. DUNBAR, E. DUPOUX.*Self Consistency as an Inductive Bias in Early Language Acquisition*, in "Proceedings of the 36th Annual Meeting of the Cognitive Science Society", CogSci, 2014, p. 469-474.

[18] A. FOURTASSI, E. DUPOUX.*A corpus-based evaluation method for Distributional Semantic Models*, in "Proceedings of ACL-SRW 2013", Sofia, Bulgaria, ACL, 2013, p. 165-171.

[19] 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].

[20] A. FOURTASSI, E. DUPOUX.*The role of word-word co-occurrence in word learning*, in "Proceedings of the 38th Annual Conference of the Cognitive Science Society", 2016, p. 662-667.

[21] 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", Bal-

timore, Maryland, Association for Computational Linguistics, 2014, vol. 2, p. 1-6 [*DOI :* 10.3115/v1/P14-2001].

[22] K. GVOZDIC, S. MOUTIER, E. DUPOUX, M. BUON.*Priming Children's Use of Intentions in Moral Judgement with Metacognitive Training*, in "Frontiers in Language Sciences", 2016, vol. 7, n° 190 [*DOI :* 10.3389/FPSYG.2016.00190].

[23] H. HERMANSKY, L. BURGET, J. COHEN, E. DUPOUX, N. FELDMAN, J. GODFREY, S. KHUDANPUR, M. MACIEJEWSKI, S. H. MALLIDI, A. MENON, T. OGAWA, V. PEDDINTI, R. ROSE, R. STERN, M. WIESNER, K. VESELY.*Towards machines that know when they do not know: Summary of work done at 2014 Frederick Jelinek memorial workshop in Prague*, in "ICASSP-2015 (IEEE International Conference on Acoustics Speech and Signal Processing)", Brisbane, Australia, 19-24 April 2015, p. 5009-5013.

[24] A. JANSEN, E. DUPOUX, S. GOLDWATER, M. JOHNSON, S. KHUDANPUR, K. CHURCH, N. FELDMAN, H. HERMANSKY, F. METZE, R. ROSE, M. SELTZER, P. CLARK, I. MCGRAW, B. VARADARAJAN, E. BENNETT, B. BORSCHINGER, J. CHIU, E. DUNBAR, A. FOURTASSI, D. HARWATH, C.-Y. LEE, K. LEVIN, A. NOROUZIAN, V. PEDDINTI, R. RICHARDSON, T. SCHATZ, S. THOMAS.*A summary of the 2012 JH CLSP Workshop on zero resource speech technologies and models of early language acquisition*, in "ICASSP-2013 (IEEE International Conference on Acoustics Speech and Signal Processing)", Vancouver, BC, Canada, IEEE, May 2013, p. 8111-8115 [*DOI :* 10.1109/ICASSP.2013.6639245].

[25] M. JOHNSON, A. CHRISTOPHE, K. DEMUTH, E. DUPOUX.*Modelling function words improves unsupervised word segmentation*, in "Proceedings of the 52nd Annual meeting of the ACL", ACL, 2014, p. 282–292 [*DOI :* 10.3115/v1/P14-1027].

[26] M. JOHNSON, J. PATER, R. STAUB, E. DUPOUX.*Sign constraints on feature weights improve a joint model of word segmentation and phonology*, in "NAACL HLT 2015", ACL, 2015, p. 303-313 [*DOI :* 10.3115/v1/N15-1034].

[27] 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.

[28] 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. 1-11 [*DOI :* 10.18653/v1/S16-2001].

[29] B. LUDUSAN, A. CARANICA, H. CUCU, A. BUZO, C. BURILEANU, E. DUPOUX.*Exploring multi-language resources for unsupervised spoken term discovery*, in "Speech Technology and Human-Computer Dialogue (SpeD), 2015 International Conference on", 2015, p. 1-6.

[30] B. LUDUSAN, A. CRISTIA, A. MARTIN, R. MAZUKA, E. DUPOUX.*Learnability of prosodic boundaries: Is infant-directed speech easier?*, in "Journal of the Acoustical Society of America", 2016, vol. 140, n° 2, p. 1239-1250.

[31] B. LUDUSAN, E. DUPOUX.*Towards Low Resource Prosodic Boundary Detection*, in "Proceedings of International Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU'14)", St Petersburg, Russia, SLTU, May 14-16 2014, p. 231-237.

[32] B. LUDUSAN, E. DUPOUX.*A multilingual study on intensity as a cue for marking prosodic boundaries*, in "ICPhS", 2015, e982.

[33] B. LUDUSAN, E. DUPOUX.*Automatic syllable segmentation using broad phonetic class information*, in "SLTU-2016 Procedia Computer Science", 2016, vol. 81, p. 101-106.

[34] B. LUDUSAN, E. DUPOUX.*The role of prosodic boundaries in word discovery: Evidence from a computational model*, in "Journal of the Acoustical Society of America", 2016, vol. 140, n^{o} 1, EL1.

[35] B. LUDUSAN, G. GRAVIER, E. DUPOUX.*Incorporating Prosodic Boundaries in Unsupervised Term Discovery*, in "Proceedings of Speech Prosody", Dublin, Ireland, 2014, vol. 7, p. 939-943.

[36] B. LUDUSAN, A. ORIGLIA, E. DUPOUX.*Rhythm-Based Syllabic Stress Learning without Labelled Data*, in "Proceedings of Statistical Language and Speech Processing -SLSP 2015", 2015, p. 185-196.

[37] B. LUDUSAN, A. SEIDL, E. DUPOUX, A. CRISTIA.*Motif discovery in infant- and adult-directed speech*, in "Proceedings of CogACLL2015", ACL, 2015, p. 93-102 [*DOI :* 10.18653/v1/W15-2413].

[38] B. LUDUSAN, G. SYNNAEVE, E. DUPOUX.*Prosodic boundary information helps unsupervised word segmentation*, in "NAACL HLT 2015", ACL, 2015, p. 953-963 [*DOI :* 10.3115/v1/N15-1096].

[39] B. LUDUSAN, M. VERSTEEGH, A. JANSEN, G. GRAVIER, X.-N. CAO, M. JOHNSON, E. DUPOUX.*Bridging the gap between speech technology and natural language processing: an evaluation toolbox for term discovery systems*, in "Proceedings of LREC 2014", Some place, LREC, 2014, p. 560-567.

[40] 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].

[41] A. MARTIN, T. SCHATZ, M. VERSTEEGH, K. MIYAZAWA, R. MAZUKA, E. DUPOUX, A. CRISTIA.*Mothers speak less clearly to infants: A comprehensive test of the hyperarticulation hypothesis*, in "Psychological Science", 2015, vol. 26, n^{o} 3, p. 341-347 [*DOI :* 10.1177/0956797614562453].

[42] E. MICHON, E. DUPOUX, A. CRISTIA.*Salient dimensions in implicit phonotactic learning*, in "INTERSPEECH-2015", 2015, p. 2665-2669.

[43] Y. MINAGAWA-KAWAI, A. CRISTIA, B. LONG, I. VENDELIN, Y. HAKUNO, M. DUTAT, L. FILIPPIN, D. CABROL, E. DUPOUX.*Insights on NIRS sensitivity from a cross-linguistic study on the emergence of phonological grammar*, in "Frontiers in Language Sciences", April 16 2013, vol. 4, n^{o} 170 [*DOI :* 10.3389/fpsyg.2013.00170].

[44] C. NGON, A. MARTIN, E. DUPOUX, D. CABROL, S. PEPERKAMP.*Nonwords, nonwords, nonwords: Evidence for a proto-lexicon during the first year of life*, in "Developmental Science", 2013, vol. 16, n^{o} 1, p. 24-34 [*DOI :* 10.1111/j.1467-7687.2012.01189.x].

[45] T. OGAWA, S. H. MALLIDI, E. DUPOUX, J. COHEN, N. FELDMAN, H. HERMANSKY.*A new efficient measure for accuracy prediction and its application to multistream-based unsupervised adaptation*, in "ICPR", 2016.

[46] 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.

[47] T. SCHATZ, V. PEDDINTI, X.-N. CAO, F. BACH, H. HYNEK, E. DUPOUX.*Evaluating speech features with the Minimal-Pair ABX task (II): Resistance to noise*, in "INTERSPEECH-2014", International Speech Communication Association, 2014, p. 915-919.

[48] G. SYNNAEVE, I. DAUTRICHE, B. BOERSCHINGER, M. JOHNSON, E. DUPOUX.*Unsupervised word segmentation in context*, in "Proceedings of 25th International Conference on Computational Linguistics (CoLing)", CoLing, 2014, p. 2326-2334.

[49] G. SYNNAEVE, E. DUPOUX.*In Depth Deep Beliefs Networks for Phone Recognition*, in "Poster presented in NIPS-2013", NIPS, 2013.

[50] G. SYNNAEVE, E. DUPOUX.*Weakly Supervised Multi-Embeddings Learning of Acoustic Models*, in "ICLR Workshop", 2015, ArXiv 1412.6645 [cs.SD].

[51] G. SYNNAEVE, E. DUPOUX.*A temporal coherence loss function for learning unsupervised acoustic embeddings*, in "SLTU-2016 Procedia Computer Science", ISCA-ITRW, 2016, vol. 81, p. 95-100.

[52] G. SYNNAEVE, T. SCHATZ, E. DUPOUX.*Phonetics embedding learning with side information*, in "IEEE Spoken Language Technology Workshop", IEEE, 2014, p. 106 - 111 [*DOI :* 10.1109/SLT.2014.7078558].

[53] G. SYNNAEVE, M. VERSTEEGH, E. DUPOUX.*Learning words from images and speech*, in "NIPS Workshop on Learning Semantics", Montreal, Canada, 2014.

[54] 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.

[55] M. VERSTEEGH, X. ANGUERA, A. JANSEN, E. DUPOUX.*The Zero Resource Speech Challenge 2015: Proposed Approaches and Results*, in "SLTU-2016 Procedia Computer Science", ISCA-ITRW, 2016, vol. 81, p. 67-72.

[56] M. VERSTEEGH, R. THIOLLIÈRE, T. SCHATZ, X.-N. CAO, X. ANGUERA, A. JANSEN, E. DUPOUX.*The Zero Resource Speech Challenge 2015*, in "INTERSPEECH-2015", 2015, p. 3169-3173.

[57] N. ZEGHIDOUR, G. SYNNAEVE, N. USUNIER, E. DUPOUX.*Joint Learning of Speaker and Phonetic Similarities with Siamese Networks*, in "INTERSPEECH-2016", ISCA, 2016, p. 1295-1299.

[58] N. ZEGHIDOUR, G. SYNNAEVE, M. VERSTEEGH, E. DUPOUX.*A Deep Scattering Spectrum - Deep Siamese Network Pipeline For Unsupervised Acoustic Modeling*, in "ICASSP-2016", IEEE, 2016, p. 4965-4969.

[59] R. DE DIEGO-BALAGUER, C. SCHRAMM, I. REBEIX, E. DUPOUX, A. DURR, A. BRICE, P. CHARLES, L. CLERET DE LANGAVANT, K. YOUSSOV, C. VERNY, V. DAMOTTE, J.-P. AZULAY, C. GOIZET, C. SIMONIN, C. TRANCHANT, P. MAISON, A. RIALLAND, D. SCHMITZ, C. JACQUEMOT, B. FONTAINE,

A.-C. BACHOUD-LÉVI.*COMT Val158Met Polymorphism Modulates Huntington's Disease Progression*, in "Plos One", 2016, vol. 11, n^{o} 9, e0161106 [*DOI :* 10.1371/JOURNAL.PONE.0161106].

Publications of the year

Articles in International Peer-Reviewed Journal

[60] A. CRISTIA, E. DUPOUX, M. GURVEN, J. STIEGLITZ.*Child-Directed Speech Is Infrequent in a Forager-Farmer Population: A Time Allocation Study*, in "Child Development", 2017 [*DOI :* 10.1111/CDEV.12974], https://hal.inria.fr/hal-01687336.

[61] A. GUEVARA-RUKOZ, I. LIN, M. MORII, Y. MINAGAWA-KAWAI, E. DUPOUX, S. PEPERKAMP.*Which epenthetic vowel? Phonetic categories versus acoustic detail in perceptual vowel epenthesis*, in "Journal of the Acoustical Society of America", August 2017, vol. 142, n^{o} 2, p. EL211 - EL217 [*DOI :* 10.1121/1.4998138], https://hal.inria.fr/hal-01687489.

[62] S. TSUJI, P. FIKKERT, Y. MINAGAWA-KAWAI, E. DUPOUX, L. FILIPPIN, M. VERSTEEGH, P. HAGOORT, A. CRISTIA.*The more, the better? Behavioral and neural correlates of frequent and infrequent vowel exposure*, in "Developmental Psychobiology", July 2017, vol. 59, n^{o} 5, p. 603 - 612 [*DOI :* 10.1002/DEV.21534], https://hal.inria.fr/hal-01687403.

International Conferences with Proceedings

[63] R. CHAABOUNI, E. DUNBAR, N. ZEGHIDOUR, E. DUPOUX.*Learning Weakly Supervised Multimodal Phoneme Embeddings*, in "Interspeech 2017", Stockholm, Sweden, ISCA, 2017 [*DOI :* 10.21437/INTERSPEECH.2017-1689], https://hal.inria.fr/hal-01687415.

[64] A. GUEVARA-RUKOZ, E. PARLATO-OLIVEIRA, S. YU, Y. HIROSE, S. PEPERKAMP, E. DUPOUX.*Predicting Epenthetic Vowel Quality from Acoustics*, in "Interspeech 2017", Stockholm, Sweden, ISCA, 2017 [*DOI :* 10.21437/INTERSPEECH.2017-1735], https://hal.inria.fr/hal-01687378.

[65] E. LARSEN, E. DUPOUX, A. CRISTIA.*Relating Unsupervised Word Segmentation to Reported Vocabulary Acquisition*, in "Interspeech 2017", Stockholm, Sweden, ISCA, 2017 [*DOI :* 10.21437/INTERSPEECH.2017-937], https://hal.inria.fr/hal-01687534.

[66] G. LE GODAIS, T. LINZEN, E. DUPOUX.*Comparing Character-level Neural Language Models Using a Lexical Decision Task*, in "Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers", Valencia, Spain, Association for Computational Linguistics, April 2017 [*DOI :* 10.18653/v1/E17-2020], https://hal.inria.fr/hal-01687583.

[67] *Best Paper*
B. LUDUSAN, R. MAZUKA, M. BERNARD, A. CRISTIA, E. DUPOUX.*The Role of Prosody and Speech Register in Word Segmentation: A Computational Modelling Perspective*, in "Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)", Vancouver, Canada, Association for Computational Linguistics, July 2017 [*DOI :* 10.18653/v1/P17-2028], https://hal.inria.fr/hal-01687451.

[68] M. PAUL, O. RASANEN, R. THIOLLIÈRE, E. DUPOUX.*Blind Phoneme Segmentation With Temporal Prediction Errors*, in "Proceedings of ACL 2017, Student Research Workshop", Vancouver, Canada, Association

for Computational Linguistics, July 2017, https://arxiv.org/abs/1608.00508 [*DOI :* 10.18653/v1/P17-3011], https://hal.inria.fr/hal-01687524.

[69] T. SCHATZ, R. TURNBULL, F. BACH, E. DUPOUX.*A Quantitative Measure of the Impact of Coarticulation on Phone Discriminability*, in "Interspeech 2017", Stockholm, Sweden, ISCA, 2017 [*DOI :* 10.21437/INTERSPEECH.2017-1306], https://hal.inria.fr/hal-01687436.

Books or Proceedings Editing

[70] E. DUNBAR, X.-N. CAO, J. BENJUMEA, J. KARADAYI, M. BERNARD, L. BESACIER, X. ANGUERA, E. DUPOUX (editors). *The Zero Resource Speech Challenge 2017*, 2017, https://arxiv.org/abs/1712.04313 - IEEE ASRU (Automatic Speech Recognition and Understanding) 2017. Okinawa, Japan, https://hal.inria.fr/hal-01687504.

References in notes

[71] 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.

[72] 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.

[73] 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.

[74] 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.

[75] C. LU, X. TANG.*Surpassing human-level face verification performance on LFW with GaussianFace*, in "arXiv preprint arXiv:1404.3840", 2014.

[76] 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.

[77] 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.

[78] 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.

[79] A. M. TURING.*Computing machinery and intelligence*, in "Mind", 1950, vol. 59, n^{o} 236, p. 433–460.

[80] 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.

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. Personnel

Research Scientists

François Baccelli [Inria, Senior Researcher, part time, HDR]
Bartlomiej Blaszczyszyn [Inria, acting team leader in 2017, Senior Researcher, adjunct professor at DI ENS since September 2017, HDR]
Ana Busic [Inria, Researcher]
Francesco Caltagirone [Inria, Starting Research Position, until Jan 2017]
Marc Lelarge [on leave at Safran in 2017, Researcher, HDR]

Faculty Members

Anne Bouillard [Ecole Normale Supérieure Paris, Associate Professor, until Jan 2017, HDR]
Jocelyne Elias [Univ René Descartes, Associate Professor, by delegation, until Aug 2017]

External Collaborators

Anne Bouillard [Nokia, from Feb 2017, HDR]
Pierre Bremaud [Prof. Emeritus]
Marc Olivier Buob [Bell Labs (Alcatel)]
Josu Doncel [University of the Basque Country, Jul 2017]
Mir Omid Haji Mirsadeghi [Sharif University, Tehran]
Fabien Mathieu [Nokia, HDR]
Christelle Rovetta [Univ Pierre et Marie Curie, from Jul 2017 until Aug 2017]

Technical Staff

Antoine Brochard [Inria, from Apr 2017]

Holger Keeler [Inria, from Oct 2017, granted by HUAWEI TECHNOLOGIES FRANCE]
Eric Tramel [Inria, until Mar 2017]

PhD Students

Arnaud Cadas [PSL, from Oct 2017]
Lennart Gulikers [Inria, until Nov 2017]
Md Umar Hashmi [PSL]
Dalia-Georgiana Herculea [Bell Labs (Nokia)]
Alexandre Hollocou [Ministère de la Défense]
Quentin Le Gall [Orange, from Oct 2017]
Leo Miolane [Ecole polytechnique]
Christelle Rovetta [Univ Pierre et Marie Curie, until Jun 2017]
Sébastien Samain [Inria]
Rémi Varloot [Inria]

Post-Doctoral Fellow

Arpan Mukhopadhyay [Inria, until Feb 2017]

Visiting Scientists

Venkatachalam Anantharam [UC Berkeley, from Jun 2017 until Jul 2017]
Prabir Barooah [University of Florida, from May 2017 until Jun 2017]
Milan Bradonjic [Nokia, until Jan 2017]
Adithya Munegowda Devraj [University of Florida, from Aug 2017 until Sep 2017]
Christian Hirsch [LMU Munich, Sep 2017]
Yuting Ji [Stanford, Oct 2017]

Administrative Assistant

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 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 in collaboration with Ch. Bordenave [IMT Toulouse], L. Massoulié [MSR-Inria].

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

B. Blaszczyszyn has just been appointed ENS adjunct professor in September 2017.

6. New Software and Platforms

6.1. CloNES

CLOsed queueing Networks Exact Sampling
FUNCTIONAL DESCRIPTION: Clones is a Matlab toolbox for exact sampling of closed queueing networks.

- Participant: Christelle Rovetta
- Contact: Christelle Rovetta
- URL: http://www.di.ens.fr/~rovetta/Clones/index.html

6.2. Platforms

6.2.1. CapRadio

Cellular network dimensioning toolbox *CapRadio* is being developed by Orange in a long-term collaboration between TREC/DYOGENE represented by B. Blaszczyszyn, and Orange Labs, represented by M. K. Karray. This year it has been enriched by the results of the contract titled "Scheduling effect on the distribution of QoS over cells in 4G wireless cellular networks"; cf 8.1.1.

7. New Results

7.1. Reversibility and further properties of FCFS infinite bipartite matching

[3] The model of FCFS infinite bipartite matching was introduced in Caldentey, Kaplan, & Weiss Adv. Appl. Probab., 2009. In this model, there is a sequence of items that are chosen i.i.d. from a finite set C and an independent sequence of items that are chosen i.i.d. from a finite set S, and a bipartite compatibility graph G between C and 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 & Weiss, Operations Research, 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 the current paper, 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.

7.2. Point-map-probabilities of a point process and Mecke's invariant measure equation

[4] A compatible point-shift F maps, in a translation invariant way, each point of a stationary point process Φ to some point of Φ. It is fully determined by its associated point-map, f, which gives the image of the origin by F. It was proved by J. Mecke that if F is bijective, then the Palm probability of Φ is left invariant by the translation of $-f$. The initial question motivating this paper is the following generalization of this invariance result: in the nonbijective case, what probability measures on the set of counting measures are left invariant by the translation of $-f$? The point-map-probabilities of Φ are defined from the action of the semigroup of point-map translations on the space of Palm probabilities, and more precisely from the compactification of the orbits of this semigroup action. If the point-map-probability exists, is uniquely defined and if it satisfies certain continuity properties, it then provides a solution to this invariant measure problem. Point-map-probabilities are

objects of independent interest. They are shown to be a strict generalization of Palm probabilities: when F is bijective, the point-map-probability of Φ boils down to the Palm probability of Φ. When it is not bijective, there exist cases where the point-map-probability of Φ is singular with respect to its Palm probability. A tightness based criterion for the existence of the point-map-probabilities of a stationary point process is given. An interpretation of the point-map-probability as the conditional law of the point process given that the origin has F-pre-images of all orders is also provided. The results are illustrated by a few examples.

7.3. Gibbsian on-line distributed content caching strategy for cellular networks

[7] 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 imple 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.

7.4. State estimation for the individual and the population in mean field control with application to demand dispatch

[10] This paper concerns state estimation problems in a mean field control setting. In a finite population model, the goal is to estimate the joint distribution of the population state and the state of a typical individual. The observation equations are a noisy measurement of the population. The general results are applied to demand dispatch for regulation of the power grid, based on randomized local control algorithms. In prior work by the authors it is shown that local control can be designed so that the aggregate of loads behaves as a controllable resource, with accuracy matching or exceeding traditional sources of frequency regulation. The operational cost is nearly zero in many cases. The information exchange between grid and load is minimal, but it is assumed in the overall control architecture that the aggregate power consumption of loads is available to the grid operator. It is shown that the Kalman filter can be constructed to reduce these communication requirements, and to provide the grid operator with accurate estimates of the mean and variance of quality of service (QoS) for an individual load.

7.5. Distributed spectrum management in TV white space networks

[11] In this paper, we investigate the spectrum management problem in TV White Space (TVWS) Cognitive Radio Networks using a game theoretical approach, accounting for adjacent-channel interference. TV Bands Devices (TVBDs) compete to access available TV channels and choose idle blocks that optimize some objective function. Specifically, the goal of each TVBD is to minimize the price paid to the Database operator and a cost function that depends on the interference between unlicensed devices. We show that the proposed TVWS management game admits a potential function under general conditions. Accordingly, we use a Best Response algorithm to converge in few iterations to the Nash Equilibrium (NE) points. We evaluate the performance of the proposed game, considering both static and dynamic TVWS scenarios and taking into account users' mobility. Our results show that at the NE, the game provides an interesting tradeoff between efficient TV spectrum use and reduction of interference between TVBDs.

7.6. A spectral method for community detection in moderately sparse degree-corrected stochastic block models

[12] We consider community detection in degree-corrected stochastic block models. We propose a spectral clustering algorithm based on a suitably normalized adjacency matrix. We show that this algorithm consistently recovers the block membership of all but a vanishing fraction of nodes, in the regime where the lowest degree is of order log(n) or higher. Recovery succeeds even for very heterogeneous degree distributions. The algorithm does not rely on parameters as input. In particular, it does not need to know the number of communities.

7.7. Non-backtracking spectrum of degree-corrected stochastic block models

[25] Motivated by community detection, we characterise the spectrum of the non-backtracking matrix B in the Degree-Corrected Stochastic Block Model. Specifically, we consider a random graph on n vertices partitioned into two asymptotically equal-sized clusters. The vertices have i.i.d. weights $\{\phi_u\}_{u=1}^n$ with second moment $\Phi^{(2)}$. The intra-cluster connection probability for vertices u and v is $\frac{\phi_u\phi_v}{n}a$ and the inter-cluster connection probability is $\frac{\phi_u\phi_v}{n}b$. We show that with high probability, the following holds: The leading eigenvalue of the non-backtracking matrix B is asymptotic to $\rho = \frac{a+b}{2}\Phi^{(2)}$. The second eigenvalue is asymptotic to $\mu_2 = \frac{a-b}{2}\Phi^{(2)}$ when $\mu_2^2 > \rho$, but asymptotically bounded by $\sqrt{\rho}$ when $\mu_2^2 \leq \rho$. All the remaining eigenvalues are asymptotically bounded by $\sqrt{\rho}$. As a result, a clustering positively-correlated with the true communities can be obtained based on the second eigenvector of B in the regime where $\mu_2^2 > \rho$. In a previous work we obtained that detection is impossible when $\mu_2^2 < \rho$, meaning that there occurs a phase-transition in the sparse regime of the Degree-Corrected Stochastic Block Model. As a corollary, we obtain that Degree-Corrected Erdős-Rényi graphs asymptotically satisfy the graph Riemann hypothesis, a quasi-Ramanujan property. A by-product of our proof is a weak law of large numbers for local-functionals on Degree-Corrected Stochastic Block Models, which could be of independent interest.

7.8. A spectral algorithm with additive clustering for the recovery of overlapping communities in networks

[13] This paper presents a novel spectral algorithm with additive clustering designed to identify overlapping communities in networks. The algorithm is based on geometric properties of the spectrum of the expected adjacency matrix in a random graph model that we call stochastic blockmodel with overlap (SBMO). An adaptive version of the algorithm, that does not require the knowledge of the number of hidden communities, is proved to be consistent under the SBMO when the degrees in the graph are (slightly more than) logarithmic. The algorithm is shown to perform well on simulated data and on real-world graphs with known overlapping communities.

7.9. Optimal geographic caching in cellular networks with linear content coding

[14] We state and solve a problem of the optimal geographic caching of content in cellular networks, where linear combinations of contents are stored in the caches of base stations. We consider a general content popularity distribution and a general distribution of the number of stations covering the typical location in the network. We are looking for a policy of content caching maximizing the probability of serving the typical content request from the caches of covering stations. The problem has a special form monotone sub-modular set function maximization. Using dynamic programming, we find a deterministic policy solving the problem. We also consider two natural greedy caching policies. We evaluate our policies considering two popular stochastic geometric coverage models: the Boolean one and the Signal-to-Interference-and-Noise-Ratio one, assuming Zipf popularity distribution. Our numerical results show that the proposed deterministic policies are in general not worst than some randomized policy considered in the literature and can further improve the total hit probability in the moderately high coverage regime.

7.10. Online mobile user speed estimation: performance and tradeoff considerations

[15] This paper presents an online algorithm for mobile user speed estimation in 3GPP Long Term Evolution (LTE)/LTE-Advanced (LTE-A) networks. The proposed method leverages on uplink (UL) sounding reference signal (SRS) power measurements performed at the base station, also known as eNodeB (eNB), and remains effective even under large sampling period. Extensive performance evaluation of the proposed algorithm is carried out using field traces from realistic environment. The on-line solution is proven highly efficient in terms of computational requirement, estimation delay, and accuracy. In particular, we show that the proposed algorithm can allow for the first speed estimation to be obtained after 10 seconds and with an average speed underestimation error of 14 kmph. After the first speed acquisition, subsequent speed estimations can be obtained much faster (e.g., each second) with limited implementation cost and still provide high accuracy.

7.11. Self-similarity in urban wireless networks: Hyperfractals

[18] In this work we study a Poisson patterns of fixed and mobile nodes distributed on straight lines designed for 2D urban wireless networks. The particularity of the model is that, in addition to capturing the irregularity and variability of the network topology, it exploits self-similarity, a characteristic of urban wireless networks. The pattern obeys to " Hyperfractal " measures which show scaling properties corresponding to an apparent dimension larger than 2. The hyperfractal pattern is best suitable for capturing the traffic over the streets and highways in a city. The scaling effect depends on the hyperfractal dimensions. Assuming radio propagation limited to streets, we prove results on the scaling of routing metrics and connectivity graph.

7.12. Optimizing spatial throughput in device-to-device networks

[19] Results are presented for optimizing device-to-device communications in cellular networks, while maintaining spectral efficiency of the base-station-to-device downlink channel. We build upon established and tested stochastic geometry models of signal-to-interference ratio in wireless networks based on the Poisson point process, which incorporate random propagation effects such as fading and shadowing. A key result is a simple formula, allowing one to optimize the device-to-device spatial throughput by suitably adjusting the proportion of active devices. These results can lead to further investigation as they can be immediately applied to more sophisticated models such as studying multi-tier network models to address coverage in closed access networks.

7.13. Demand dispatch with heterogeneous intelligent loads

[20] A distributed control architecture is presented that is intended to make a collection of heterogeneous loads appear to the grid operator as a nearly perfect battery. Local control is based on randomized decision rules advocated in prior research, and extended in this paper to any load with a discrete number of power states. Additional linear filtering at the load ensures that the input-output dynamics of the aggregate has a nearly flat input-output response: the behavior of an ideal, multi-GW battery system.

7.14. Energy savings for virtual MISO in fractal sensor networks

[21] We design a model of wireless terminals, i.e. transmitters and receivers, obtained from a Poisson point process with support in an embedded fractal map. The terminals form a virtual MISO (Multiple Input Single Output) system with successful reception under SNR (signal-to-noise ratio) capture condition in a single hop transmission. We show that if we omit antennas cross sections, the energy needed to broadcast a packet of information tends to zero when the density of transmitters and receivers increases. This property is a direct consequence of the fact that the support map is fractal and would not hold if the terminal distribution were Poisson uniform, as confirmed by simulations. The result becomes invalid if the cross sections overlap or if we consider a masking effect due to antennas, which would imply an extremely large density of terminals. In the case where the cross sections of the transmitters have a non-zero value, the energy has a non-zero limit which decays to zero when the cross sections tend to zero.

7.15. Distributed control of a fleet of batteries

[22] Battery storage is increasingly important for grid-level services such as frequency regulation, load following, and peak-shaving. The management of a large number of batteries presents a control challenge: How can we solve the apparently combinatorial problem of coordinating a large number of batteries with discrete, and possibly slow rates of charge/discharge? The control solution must respect battery constraints, and ensure that the aggregate power output tracks the desired grid-level signal. A distributed stochastic control architecture is introduced as a potential solution. Extending prior research on distributed control of flexible loads, a randomized decision rule is defined for each battery of the same type. The power mode at each time-slot is a randomized function of the grid-signal and its internal state. The randomized decision rule is designed to maximize idle time of each battery, and keep the state-of-charge near its optimal level, while ensuring that the aggregate power output can be continuously controlled by a grid operator or aggregator. Numerical results show excellent tracking, and low stress to individual batteries.

7.16. Exact Computation and bounds for the coupling time in queueing systems

[23] This paper is a work in progress on the exact computation and bounds of the expected coupling time for finite-state Markov chains. We give an exact formula in terms of generating series. We show how this may help to bound the expected coupling time for queueing networks.

7.17. An online disaggregation algorithm and its application to demand control

[24] The increase of renewable energy has made the supply-demand balance of power more complex to handle. Previous approach designed randomized controllers to obtain ancillary services to the power grid by harnessing inherent flexibility in many loads. However these controllers suppose that we know the consumption of each device that we want to control. This introduce the cost and the social constraint of putting sensors on each device of each house. Therefore, our approach was to use Nonintrusive Appliance Load Monitoring (NALM) methods to solve a disaggregation problem. The latter comes down to estimating the power consumption of each device given the total power consumption of the whole house. We started by looking at the Factorial Hierarchical Dirichlet Process-Hidden Semi-Markov Model (Factorial HDP-HSMM). In our application, the total power consumption is considered as the observations of this state-space model and the consumption of each device as the state variables. Each of the latter is modelled by an HDP-HSMM which is an extension of a Hidden Markov Model. However, the inference method proposed previously is based on Gibbs sampling and has a complexity of $O(T2N + TN2)$ where T is the number of observations and N is the number of hidden states. As our goal is to use the randomized controllers with our estimations, we wanted a method that does not scale with T. Therefore, we developed an online algorithm based on particle filters. Because we worked in a Bayesian setting, we had to infer the parameters of our model. To do so, we used a method called Particle Learning. The idea is to include the parameters in the state space so that they are tied to the particles. Then, for each (re)sampling step, the parameters are sampled from their posterior distribution with the help of Bayesian sufficient statistics. We applied the method to data from Pecan Street. Using their Dataport, we have collected the power consumption of each device from about a hundred houses. We selected the few devices that consume the most and that are present in most houses. We separated the houses in a training set and a test set. For each device of each house from the training set, we estimated the operating modes with a HDP-HSMM and used these estimations to compute estimators of the priors hyperparameters. Finally we applied the particle filters method to the test houses using the computed priors. The algorithm performs well for the device with the highest power consumption, the air compressor in our case. We will discuss ongoing work where we apply the "Thermo-statically Controlled Loads" example using our estimations of this air compressor's operating modes.

7.18. Multiple local community detection

[26] Community detection is a classical problem in the field of graph mining. We are interested in local community detection where the objective is the recover the communities containing some given set of nodes, called the seed set. While existing approaches typically recover only one community around the seed set, most nodes belong to multiple communities in practice. In this paper, we introduce a new algorithm for detecting multiple local communities, possibly overlapping, by expanding the initial seed set. The new nodes are selected by some local clustering of the graph embedded in a vector space of low dimension. We validate our approach on real graphs, and show that it provides more information than existing algorithms to recover the complex graph structure that appears locally.

7.19. A Streaming Algorithm for Graph Clustering

[27] We introduce a novel algorithm to perform graph clustering in the edge streaming setting. In this model, the graph is presented as a sequence of edges that can be processed strictly once. Our streaming algorithm has an extremely low memory footprint as it stores only three integers per node and does not keep any edge in memory. We provide a theoretical justification of the design of the algorithm based on the modularity function, which is a usual metric to evaluate the quality of a graph partition. We perform experiments on massive real-life graphs ranging from one million to more than one billion edges and we show that this new algorithm runs more than ten times faster than existing algorithms and leads to similar or better detection scores on the largest graphs.

7.20. Discrete probability models and methods: probability on graphs and trees, markov chains and random fields, entropy and coding

[28] The emphasis in this book is placed on general models (Markov chains, random fields, random graphs), universal methods (the probabilistic method, the coupling method, the Stein-Chen method, martingale methods, the method of types) and versatile tools (Chernoff's bound, Hoeffding's inequality, Holley's inequality) whose domain of application extends far beyond the present text. Although the examples treated in the book relate to the possible applications, in the communication and computing sciences, in operations research and in physics, this book is in the first instance concerned with theory. The level of the book is that of a beginning graduate course. It is self-contained, the prerequisites consisting merely of basic calculus (series) and basic linear algebra (matrices). The reader is not assumed to be trained in probability since the first chapters give in considerable detail the background necessary to understand the rest of the book.

7.21. Distributed control design for balancing the grid using flexible loads

[29] nexpensive 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 this chapter 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.

7.22. Un classificateur non-supervisé utilisant les complexes simpliciaux avec une application à la stylométrie

[30] Un classificateur non-supervisé utilisant les complexes simpliciaux (avec une application à la stylométrie). Nous nous proposons au cours des quelques pages de ce rapport de présenter au lecteur ce que sont les complexes simpliciaux ainsi qu'une de leurs possibles (et nombreuses !) applications : en classification non-supervisée. Les complexes simpliciaux peuvent s'appréhender comme une généralisation des graphes ; un graphe étant la donnée d'un ensemble de sommets ainsi que d'une relation de voisinage entre des paires de ces sommets (deux points sont voisins si une arête les relie). Les complexes simpliciaux permettent de rendre compte de relations de voisinage plus élaboré (et faisant notamment intervenir un nombre arbitraire de points ; pas seulement deux). La classification non supervisée est une branche du vaste domaine de l'apprentissage automatique. Etant donné un échantillon de données (le plus souvent des points de l'espace euclidien R^d), elle consiste a regrouper ces données en différentes classes de sorte que les données d'une même classe présentent des similarités entre elles tandis que deux données appartenant à deux classes distinctes soient dissemblables. Le présent rapport s'articulera donc en deux parties : la première introduira au lecteur non forcément familier cette notion de complexe simplicial d'un point de vue théorique. On l'illustrera ensuite avec la présentation des complexes de Cech et certaines propriétés mathématiques qui en font un outil puissant et pratique (la théorie de Morse permet, par exemple, de manier ces complexes de différentes façons). On verra encore quelques résultats des complexes simpliciaux aléatoires (c'est-à-dire que les sommets sont des points générés aléatoirement) dans le cas des régimes dits surcritiques jus-tifiant certains algorithmes d'apprentissage de variétés (une des multiples applications promises des complexes simpliciaux). Enfin, nous présenterons très succinctement l'homologie persistante...

7.23. Phase transitions, optimal errors and optimality of message-passing in generalized linear models

[31] We consider generalized linear models where an unknown nn-dimensional signal vector is observed through the successive application of a random matrix and a non-linear (possibly probabilistic) componentwise function. We consider the models in the high-dimensional limit, where the observation consists of $m \times n$ points, and $m/n \to \alpha$ where α stays finite in the limit $m, n \to \infty$. This situation is ubiquitous in applications ranging from supervised machine learning to signal processing. A substantial amount of work suggests that both the inference and learning tasks in these problems have sharp intrinsic limitations when the available data become too scarce or too noisy. Here, we provide rigorous asymptotic predictions for these thresholds through the proof of a simple expression for the mutual information between the observations and the signal. Thanks to this expression we also obtain as a consequence the optimal value of the generalization error in many statistical learning models of interest, such as the teacher-student binary perceptron, and introduce several new models with remarquable properties. We compute these thresholds (or "phase transitions") using ideas from statistical physics that are turned into rigorous methods thanks to a new powerful smart-path interpolation technique called the stochastic interpolation method, which has recently been introduced by two of the authors. Moreover we show that a polynomial-time algorithm refered to as generalized approximate message-passing reaches the optimal generalization performance for a large set of parameters in these problems. Our results clarify the difficulties and challenges one has to face when solving complex high-dimensional statistical problems.

7.24. Lecture notes on random geometric models — random graphs, point processes and stochastic geometry

[32] The goal of this sequence of lessons is to provide quick access to some popular models of random geometric structures used in many applications: from communication networks, including social, transportation, wireless networks, to geology, material sciences and astronomy. The course is composed of the following 15 lessons: (1) Bond percolation on the square lattice, (2) Galton-Watson tree, (3) Erdős-Rényi graph — emergence of the giant component, (4) Graphs with a given node degree distribution, (5) Typical nodes and random unimodular graphs, (6) Erdős-Rényi graph — emergence of the full connectivity, (7) Poisson point process,

(8) Point conditioning and Palm theory for point processes, (9) Hard-core point processes, (10) Stationary point processes and mass transport principle, (11) Stationary Voronoi tessellation, (12) Ergodicity and point-shift invariance, (13) Random closed sets, (14) Boolean model and coverage processes, (15) Connectedness of random sets and continuum percolation. Usually, these subjects are presented in different monographs: random graphs (lessons 2–6), point processes (7-12), stochastic geometry (13-14), with percolation models presented in lesson 1 and 15 often addressed separately. Having them in one course gives us an opportunity to observe some similarities and even fundamental relations between different models. Examples of such connections are:

- Similar phase transitions regarding the emergence of big components observed in different discrete, lattice and continuous euclidean models (lessons 1–4, 15).
- Single isolated nodes being the last obstacle in the emergence of the full connectivity in some discrete and euclidean graphs exhibiting enough independence (lessons 6, 15).
- A mass transport principle as a fundamental property for unimodular random graphs and Palm theory for stationary point processes; with both theories seeking to define the typical node/point of a homogeneous structure (lessons 5, 10–12).
- Poisson-Galton-Watson tree and Poisson process playing a similar role in the theory of random graphs and point processes, respectively: for both models independence and Poisson distribution are the key assumptions, both appear as natural limits, and both rooted/conditioned to a typical node/point preserve the distribution of the remaining part of the structure (lessons 2,5, 7–8).
- Size biased sampling appearing in several, apparently different, conditioning scenarios, as unimodular trees (lesson 5), Palm distributions for point process (lesson 8), zero cell of the stationary tessellations (lessons 11).

The goal of this series of lectures is to present some spectrum of models and ideas. When doing this, we sometimes skip more technical proof details, sending the reader for them to more specialised monographs. Some theoretical and computer exercises are provided after each lesson to let the reader practice his/her skills. Regarding the prerequisites, the reader will benefit from having had some prior exposure to probability and measure theory, but this is not absolutely necessary.

The content of the course has been evolving while the author teaches it within the master programme *Probabilité et modelés aléatoires* at the University Pierre and Marie Curie in Paris. The present notes were thoroughly revised when the author was presenting them as a specially appointed professor at the School of Computing, Tokyo Institute of Technology, in the autumn term 2017.

7.25. Energy trade-offs for end-to-end communications in urban vehicular networks exploiting an hyperfractal model

[34] We present 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 $O(n1-\alpha/(dm-1))$ where $\alpha < 1$ and d $m > 2$ 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.

7.26. Fundamental limits of low-rank matrix estimation: the non-symmetric case

[36] We consider the high-dimensional inference problem where the signal is a low-rank symmetric matrix which is corrupted by an additive Gaussian noise. Given a probabilistic model for the low-rank matrix, we compute the limit in the large dimension setting for the mutual information between the signal and the observations, as well as the matrix minimum mean square error, while the rank of the signal remains constant. We also show that our model extends beyond the particular case of additive Gaussian noise and we prove an universality result connecting the community detection problem to our Gaussian framework. We unify and generalize a number of recent works on PCA, sparse PCA, submatrix localization or community detection by computing the information-theoretic limits for these problems in the high noise regime. In addition, we show that the posterior distribution of the signal given the observations is characterized by a parameter of the same dimension as the square of the rank of the signal (i.e. scalar in the case of rank one). Finally, we connect our work with the hard but detectable conjecture in statistical physics.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. CRE with Orange

One year contract titled "Scheduling effect on the distribution of QoS over cells in 4G wireless cellular networks" between Inria and Orange Labs have been realized in 2017. It is a part of the long-term collaboration between TREC/DYOGENE, represented by B. Blaszczyszyn and Orange Labs, represented by M. K. Karray, for the development of analytic tools for the QoS evaluation and dimensioning of operator cellular networks. The developed solutions are implemented in Orange dimensioning toolbox *CapRadio* 6.2.1. Antoine Brochard was hired by Inria as a research engineer thanks to this contract.

8.1.2. 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. It aims 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.

8.1.3. Contract with the Ministry of Defense

The contract supports a PhD student Alexandre Hollocou hired in 2015, co-advised by M. Lelarge.

8.1.4. 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.

8.1.5. CIFRE with Orange

Contract with Orange started in 2017 for the co-advising by B. Blaszczyszyn of a PhD student of Orange, Quentin Le Gall.

8.2. Bilateral Grants with Industry

8.2.1. Google Tides

Ana Busic and Sean Meyn received jointly in 2015 a Google Faculty Research Award for their research on Distributed Control for Renewable Integration in Smart Communities. The corresponding grant allowed us to cover some part of the scholarship of the PhD student **Sebastien Samain**. in 2017.

9. Partnerships and Cooperations

9.1. Regional Initiatives

DYOGENE is associated to the Laboratory of Information, Networking and Communication Sciences (LINCS) http://www.lincs.fr/ co-founded in 2010 by Inria, Institut Mines-Télécom and UPMC, with Bell Labs Nokia (formerly Alcatel-Lucent) and SystemX joining it as strategic partners in 2011 and 2014, respectively. The LINCS is dedicated to research and innovation in the domains of future information and communication networks, systems and services.

9.2. National Initiatives

9.2.1. 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 Pierre Calka [Université de Rouen], Viet Chi Tran [Université de Lille] 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.

9.2.2. GdR IM

Members of Dyogene participate in GdR-IM (Informatique-Mathématiques), https://www.gdr-im.fr/, working groups ALEA and SDA2 (Systèmes dynamiques, Automates et Algorithmique).

9.2.3. 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

9.2.4. PGMO

Gaspard Monge Program for Optimization and Operations Research project Decentralized control for renewable integration in smart-grids (2015-17). PI: A. Busic.

9.2.5. ANR MARMOTE

Markovian Modeling Tools and Environments - coordinator: Alain Jean-Marie (Inria Maestro); local coordinator (for partner Inria Paris-Rocquencourt): A. Bušić; Started: January 2013; Duration: 48 months; partners: Inria Paris-Rocquencourt (EPI DYOGENE), Inria Sophia Antipolis Méditerranée (EPI MAESTRO), Inria Grenoble Rhône-Alpes (EPI MESCAL), Université Versaillese-St Quentin, Telecom SudParis, Université Paris-Est Creteil, Université Pierre et Marie Curie.

The aim of the project was to realize a modeling environment dedicated to Markov models. One part developed the Perfect Simulation techniques, which allow one to sample from the stationary distribution of the process. A second one developed parallelization techniques for Monte Carlo simulation. A third one developed numerical computation techniques for a wide class of Markov models. All these developments were integrated into a programming environment allowing the specification of models and their solution strategy. Several applications have been studied in various scientific disciplines: physics, biology, economics, network engineering.

The project terminated in October 2017.

9.2.6. 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.

9.3. International Initiatives

9.3.1. PARIS

Title: Probabilistic Algorithms for Renewable Integration in Smart Grid

International Partner (Institution - Laboratory - Researcher):

University of Florida (United States) - Laboratory for Cognition & Control in Complex Systems - Sean Meyn.

Start year: 2015

See also: http://www.di.ens.fr/~busic/PARIS/

The importance of statistical modeling and probabilistic control techniques in the power systems area is now evident to practitioners in both the U.S. and Europe. Renewable generation has brought unforeseen volatility to the grid that require new techniques in distributed and probabilistic control. In a series of recent papers the two PIs have brought together their complementary skills in optimization, Markov modeling, simulation, and stochastic networks that may help to solve some pressing open problems in this area. This new research also opens many exciting new scientific questions.

9.3.2. Inria International Partners

9.3.2.1. Informal International Partners

- O. Mirsadeghi [Sharif University, Tehran],
- V. Anantharam [UC Berkeley],
- D. Yogeshwaran [Indian Statistical Institute].

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Venkat Anantharam [UC Berkeley, from Jun 2017 until Jul 2017]
- Prabir Barooah [University of Florida, from May 2017 until Jun 2017]
- Milan Bradonjic [Nokia, until Jan 2017]

- Adithya Munegowda Devraj [University of Florida, from Aug 2017 until Sep 2017]
- Christian Hirsch [LMU Munich, Sep 2017]
- Yuting Ji [Stanford, Oct 2017]
- Marc Olivier Buob [Bell Labs (Alcatel)]
- Josu Doncel [University of the Basque Country, Jul 2017]
- Mir Omid Haji Mirsadeghi [Sharif University, Tehran]

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

- B. Blaszczyszyn, October 1st – December 15th, Specially Appointed Professor at The School of Computing, Tokyo Institute of Technology.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Conference Program Committees

- Ana Busic: QEST 2017.

10.1.1.2. Reviewer

All members of the team act as reviewers for numerous conferences.

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

François Baccelli: Bernoulli, Queueing Systems.

10.1.2.2. Reviewer - Reviewing Activities

All members of the team act as reviewers for numerous scientific journals.

10.1.3. Invited Talks

- François Baccelli
  - WiOpt, Paris, May 2017, keynote speaker,
  - WITMSE, Paris, September 2017; plenary speaker.
  - 2nd Symposium on Spatial Networks, Oxford, September 2017; invited talk.
- Bartek Blaszczyszyn
  - LINCS Scientific Committee Workshop, Paris, June 2017; invited talk.
  - Tokyo Institute of Technology, November 2017; invited talk.
  - Kyushu Unuversuty, November 2017, invited talk.
- Ana Busic
  - Tutorial speaker at YEQT XI: “Winterschool on Energy Systems”, December, 2017; https://www.eurandom.tue.nl/event/yeqt-xi-winter-school-on-energy-systems/
  - Inviter speaker at Second Conference on the Mathematics of Energy Markets Wolfgang Pauli Institute, Vienna, July, 2017; https://www.mn.uio.no/math/english/research/groups/store/events/conferences/second-conference-on-the-mathematics-of-energy-mar.html
  - Inviter speaker at PDE and Probability Methods for Interactions, Sophia Antipolis (France), March 2017; https://project.inria.fr/pde2017/

10.1.4. Leadership within the Scientific Community

Since 2014: A. Busic is leading (with E. Hyon, Univ. Paris Ouest Nanterre et LIP6) the working group GdT COSMOS (Stochastic Control and Optimization, Modeling, and Simulation) of CNRS research network GdR RO; http://gdrro.lip6.fr/?q=node/78. This includes scientific organization of 2 one-day workshops per year (30-50 participants).

10.1.5. Research Administration

A. Busic : Member of the Conseil du DI ENS.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence: A. Busic (Cours) and Sébastien Samain (TD) **Structures et algorithmes aléatoires** 80heqTD, L3, ENS, France.

Licence: B. Blaszczyszyn (Cours) **Théorie de l'information et du codage** 24 heqTD, L3, ENS, France.

Master: A. Busic (Cours + TD) **Fondements de la modélisation des réseaux** 18 heqTD, L3, ENS, France.

Master: Bartlomiej Blaszczyszyn (Cours) **Processus ponctuels, graphes aléatoires et géeométrie stochastique** 39heqTD, M2 Probabilités et Modèles Aléatoires, UPMC, France

Master: Ana Busic and Marc Lelarge (Cours) et Rémi Varloot (TD) **Modéles et algorithmes de réseaux**, 50 heqTD, M1, ENS, Paris, France.

Master: B. Blaszczyszyn **Introduction to Spatial Stochastic Modeling: Random Graphs, Point Processes and Stochastic Geometry** (School of Computing, Tokyo Institute of Technology), 15 lessons 1h30.

10.2.2. Supervision

PhD: Lennart Gulikers "Trouver des algorithmes permettant de détecter des communautés dans des réseaux sociaux. Analyse de *Stochastic Block Model* et ses extensions" since 2015, defense 13 November 2017, going to Safran, supervised by Marc Lelarge with Laurent Massoulié.

PhD: Christelle Rovetta "Applications of perfect sampling to queuing networks and random generation of combinatorial objects", since December 2013, defense 20 June 2017, co-advised by Anne Bouillard and Ana Busic.

PhD in progress: Léeo Miolane, since 2016, supervised by Marc Lelarge

PhD in progress: Dalia-Georgiana Herculea, since October 2016 co-advised by B. Blaszczyszyn, E. Altman and Ph. Jacquet

PhD in progress: Md Umar Hashmi, Decentralized control for renewable integration in smartgrids, from December 2015, co-advised by A. Busic and M. Lelarge

PhD in progress: Alexandre Hollocou, since December 2015, supervised by Marc Lelarge with Thomas Bonald

PhD in progress: Sébastien Samain, Monte Carlo methods for performance evaluation and reinforcement learning, from November 2016, supervised by A. Busic

PhD in progress: Quentin Le Gall, since October 2017, co-supervised by B. Blaszczyszyn and E. Cali (Orange).

PhD in progress: Arnaud Cadas [PSL] "Dynamic matching models", since 2017, supervised by A. Busic.

10.2.3. Juries

- B. Blaszczyszyn, PhD defense of Frederik Mallmann-Trenn (ENS).
- A. Busic
  - Co-president of the Commission Emplois Scientifiques (CES) at Inria Paris (hiring committee for PhD, post-doc and visiting professor positions);

– Member of the hiring committee for Assistant Professor at Ecole Normale Supérieure and IUT Orsay.

11. Bibliography

Major publications by the team in recent years

[1] F. BACCELLI, B. BŁASZCZYSZYN.*Stochastic Geometry and Wireless Networks, Volume I — Theory*, Foundations and Trends in Networking, NoW Publishers, 2009, vol. 3, No 3–4, p. 249–449.

[2] F. BACCELLI, B. BŁASZCZYSZYN.*Stochastic Geometry and Wireless Networks, Volume II — Applications*, Foundations and Trends in Networking, Now Publishers, 2009, vol. 4, No 1–2, p. 1–312.

Publications of the year

Articles in International Peer-Reviewed Journal

[3] I. ADAN, A. BUSIC, J. MAIRESSE, G. WEISS.*Reversibility and further properties of FCFS infinite bipartite matching*, in "Mathematics of Operations Research", December 2017, https://arxiv.org/abs/1507.05939 - 27 pages, 12 figures [*DOI :* 10.1287/MOOR.2017.0874], https://hal.inria.fr/hal-01273897.

[4] F. BACCELLI, M.-O. HAJI-MIRSADEGHI.*Point-map-probabilities of a point process and Mecke's invariant measure equation*, in "Annals of Probability", May 2017, vol. 45, n^{o} 3, p. 1723 - 1751 [*DOI :* 10.1214/16-AOP1099], https://hal.inria.fr/hal-01673194.

[5] F. BACCELLI, F. MATHIEU, I. NORROS.*Mutual Service Processes in Euclidean Spaces: Existence and Ergodicity*, in "Queueing Systems", April 2017, vol. 86, n^{o} 1-2, p. 95 - 140 [*DOI :* 10.1007/s11134-017-9524-3], https://hal.inria.fr/hal-01535925.

[6] M. BROZOVIC, C. DANTEC, J. DARDAILLON, D. DAUGA, E. FAURE, M. GINESTE, A. LOUIS, M. NAVILLE, K. R. NITTA, J. PIETTE, W. REEVES, C. SCORNAVACCA, P. SIMION, R. VINCENTELLI, M. BELLEC, S. B. AICHA, M. FAGOTTO, M. GUEROULT-BELLONE, M. HAEUSSLER, E. JACOX, E. K. LOWE, M. MENDEZ, A. ROBERGE, A. STOLFI, R. YOKOMORI, C. T. BROWN, C. CAMBILLAU, L. CHRISTIAEN, F. DELSUC, E. J. DOUZERY, R. DUMOLLARD, T. KUSAKABE, K. NAKAI, H. NISHIDA, Y. SATOU, B. SWALLA, M. VEEMAN, J.-N. VOLFF, P. LEMAIRE.*ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets*, in "Nucleic Acids Research", November 2017 [*DOI :* 10.1093/NAR/GKX1108], https://hal.archives-ouvertes.fr/hal-01636650.

[7] A. CHATTOPADHYAY, B. BŁASZCZYSZYN, H. P. KEELER.*Gibbsian On-Line Distributed Content Caching Strategy for Cellular Networks*, in "IEEE Transactions on Wireless Communications", November 2017, https://arxiv.org/abs/1610.02318 [*DOI :* 10.1109/TWC.2017.2772911], https://hal.inria.fr/hal-01401010.

[8] Y. CHEN, A. BUSIC, 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", 2017, https://hal.archives-ouvertes.fr/hal-01672476.

[9] Y. CHEN, A. BUSIC, S. MEYN.*Estimation and Control of Quality of Service in Demand Dispatch*, in "IEEE Transactions on Smart Grid", 2017, https://hal.archives-ouvertes.fr/hal-01672458.

[10] Y. CHEN, A. BUSIC, S. P. MEYN.*State Estimation for the Individual and the Population in Mean Field Control With Application to Demand Dispatch*, in "IEEE Transactions on Automatic Control", March 2017, vol. 62, n° 3, p. 1138 - 1149 [*DOI :* 10.1109/TAC.2016.2572880], https://hal.archives-ouvertes.fr/hal-01508107.

[11] J. ELIAS, F. MARTIGNON, L. CHEN, M. KRUNZ.*Distributed Spectrum Management in TV White Space Networks*, in "IEEE Transactions on Vehicular Technology", May 2017, vol. 66, n° 5, p. 4161 - 4172 [*DOI :* 10.1109/TVT.2016.2597866], https://hal.inria.fr/hal-01350583.

[12] L. GULIKERS, M. LELARGE, L. MASSOULIÉ.*A spectral method for community detection in moderately sparse degree-corrected stochastic block models*, in "Advances in Applied Probability", June 2017, vol. 49, n° 03, p. 686 - 721 [*DOI :* 10.1017/APR.2017.18], https://hal.archives-ouvertes.fr/hal-01622731.

[13] E. KAUFMANN, T. BONALD, M. LELARGE.*A Spectral Algorithm with Additive Clustering for the Recovery of Overlapping Communities in Networks*, in "Journal of Theoretical Computer Science (TCS)", 2017, https://arxiv.org/abs/1506.04158 , https://hal.archives-ouvertes.fr/hal-01163147.

International Conferences with Proceedings

[14] J. ELIAS, B. BŁASZCZYSZYN.*Optimal Geographic Caching in Cellular Networks with Linear Content Coding*, in "WIOPT/CCDWN", Paris, France, Proc. of WIOPT/CCDWN, May 2017, https://hal.inria.fr/hal-01505881.

[15] M. HADDAD, D.-G. HERCULEA, C. S. CHEN, E. ALTMAN, V. CAPDEVIELLE.*Online Mobile User Speed Estimation: Performance and Tradeoff Considerations*, in "IEEE Consumer Communications and Networking Conference (CCNC)", Las Vegas, United States, January 2017, https://hal.inria.fr/hal-01380734.

[16] M. U. HASHMI, A. MUKHOPADHYAY, A. BUSIC, J. ELIAS.*Optimal Control of Storage under Time Varying Electricity Prices*, in "IEEE International Conference on Smart Grid Communications", Dresden, Germany, October 2017, https://hal.archives-ouvertes.fr/hal-01672483.

[17] P. JACQUET, D. POPESCU.*Self-similar Geometry for Ad-Hoc Wireless Networks: Hyperfractals*, in "3rd conference on Geometric Science of Information", Paris, France, Société Mathématique de France, November 2017, https://hal.inria.fr/hal-01561828.

[18] P. JACQUET, D. POPESCU.*Self-similarity in urban wireless networks: Hyperfractals*, in "Workshop on Spatial Stochastic Models
for Wireless Networks (SpaSWiN)", Paris, France, May 2017, https://hal.inria.fr/hal-01498987.

[19] P. KEELER, B. BŁASZCZYSZYN, P. MÜHLETHALER.*Optimizing spatial throughput in device-to-device networks*, in "WIOPT/SPASWIN 2017 - Workshop on Spatial Stochastic Models
for Wireless Networks", Paris, France, IEEE, May 2017, 5, https://arxiv.org/abs/1612.09198 - 6 pages, 4 figures. Submitted, https://hal.inria.fr/hal-01505044.

[20] J. MATHIAS, A. BUSIC, S. MEYN.*Demand Dispatch with Heterogeneous Intelligent Loads*, in "50th Annual Hawaii International Conference on System Sciences (HICSS)", Waikoloa, HI, United States, Proc. of 55th Hawaii International Conference on System Sciences (HICSS), January 2017, https://arxiv.org/abs/1610.00813 - Extended version of paper to appear in Proc. 50th Annual Hawaii International Conference on System Sciences (HICSS), 2017, https://hal.archives-ouvertes.fr/hal-01423485.

Conferences without Proceedings

[21] M. BRADONJIC, P. JACQUET, D. POPESCU.*Energy Savings for Virtual MISO in Fractal Sensor Networks*, in "55th Annual Allerton Conference on Communication, Control, and Computing", Urbana-Champaign, United States, Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, October 2017, https://hal.archives-ouvertes.fr/hal-01591112.

[22] A. BUSIC, M. U. HASHMI, S. MEYN.*Distributed control of a fleet of batteries*, in "ACC 2017 - American Control Conference", Seattle, United States, May 2017, p. 1-6, https://hal.archives-ouvertes.fr/hal-01656735.

[23] A. BUSIC, S. SAMAIN.*Exact Computation and Bounds for the Coupling Time in Queueing Systems*, in "VALUETOOLS 2017 - 11th EAI International Conference on Performance Evaluation Methodologies and Tools", Venice, Italy, December 2017, p. 1-2 [*DOI :* 10.1145/3150928.3150965], https://hal.archives-ouvertes.fr/hal-01672449.

[24] A. CADAS, A. BUSIC.*An online disaggregation algorithm and its application to demand control*, in "PGMO Days 2017", Saclay, France, November 2017, 1, https://hal.archives-ouvertes.fr/hal-01672479.

[25] L. GULIKERS, M. LELARGE, L. MASSOULIÉ.*Non-Backtracking Spectrum of Degree-Corrected Stochastic Block Models*, in "ITCS 2017 - 8th Innovations in Theoretical Computer Science", Berkeley, United States, January 2017, https://arxiv.org/abs/1609.02487 , https://hal.archives-ouvertes.fr/hal-01622719.

[26] A. HOLLOCOU, T. BONALD, M. LELARGE.*Multiple Local Community Detection*, in "IFIP WG 7.3 Performance 2017 conference - International Symposium on Computer Performance, Modeling, Measurements and Evaluation 2017", New York City, United States, 2017, https://hal.archives-ouvertes.fr/hal-01625444.

[27] A. HOLLOCOU, J. MAUDET, T. BONALD, M. LELARGE.*A Streaming Algorithm for Graph Clustering*, in "NIPS 2017 - Wokshop on Advances in Modeling and Learning Interactions from Complex Data", Long Beach, United States, December 2017, p. 1-12, https://hal.archives-ouvertes.fr/hal-01639506.

Scientific Books (or Scientific Book chapters)

[28] P. BREMAUD.*Discrete Probability Models and Methods: Probability on Graphs and Trees, Markov Chains and Random Fields, Entropy and Coding*, Springer Probability Theory and Stochastic Modelling, Springer, 2017, vol. 78, 559 [*DOI :* 10.1007/978-3-319-43476-6], https://hal.inria.fr/hal-01505040.

[29] Y. CHEN, M. U. HASHMI, J. MATHIAS, A. BUSIC, S. MEYN.*Distributed Control Design for Balancing the Grid Using Flexible Loads*, in "IMA Volume on the Control of Energy Markets and Grids", 2017, p. 1-26, https://hal.archives-ouvertes.fr/hal-01656726.

Research Reports

[30] L. HAUSEUX, B. BLASZCZYSZYN.*Un classificateur non-supervisé utilisant les complexes simpliciaux (avec une application à la stylométrie)*, Inria Paris, September 2017, https://hal.archives-ouvertes.fr/hal-01597846.

Other Publications

[31] J. BARBIER, F. KRZAKALA, N. MACRIS, L. MIOLANE, L. ZDEBOROVÁ.*Phase Transitions, Optimal Errors and Optimality of Message-Passing in Generalized Linear Models*, October 2017, https://arxiv.org/abs/1708.03395 - 35 pages, 3 figures, https://hal-cea.archives-ouvertes.fr/cea-01614258.

[32] B. BŁASZCZYSZYN.*Lecture Notes on Random Geometric Models — Random Graphs, Point Processes and Stochastic Geometry*, December 2017, p. 1-199, Lecture, https://hal.inria.fr/cel-01654766.

[33] L. HAUSEUX, F. LE ROUX.*Polynomial entropy of Brouwer homeomorphisms*, December 2017, https://arxiv.org/abs/1712.01502 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01653488.

[34] P. JACQUET, D. POPESCU, B. MANS.*Energy Trade-offs for end-to-end Communications in Urban Vehicular Networks exploiting an Hyperfractal Model*, January 2018, working paper or preprint, https://hal.inria.fr/hal-01674685.

[35] M. LELARGE, L. MIOLANE.*Fundamental limits of symmetric low-rank matrix estimation*, November 2017, https://arxiv.org/abs/1611.03888 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01648368.

[36] L. MIOLANE.*Fundamental limits of low-rank matrix estimation: the non-symmetric case*, November 2017, https://arxiv.org/abs/1702.00473 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01648369.

[37] P. MOYAL, A. BUSIC, J. MAIRESSE.*A product form and a sub-additive theorem for the general stochastic matching model*, 2017, https://arxiv.org/abs/1711.02620 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01672482.

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. - Architectures, systems and networks
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
A1.2.8. - Network security
A1.2.9. - Social Networks
A1.4. - Ubiquitous Systems
A1.6. - Green Computing
A2.3. - Embedded and cyber-physical systems
A3.4. - Machine learning and statistics
A3.5. - Social networks
A4.1. - Threat analysis
A4.4. - Security of equipment and software
A4.6. - Authentication
A4.7. - Access control
A6.1. - Mathematical Modeling
A6.1.2. - Stochastic Modeling (SPDE, SDE)
A8.2. - Optimization
A8.8. - Network science
A8.9. - Performance evaluation
A8.11. - Game Theory
A9.2. - Machine learning
A9.6. - Decision support

Other Research Topics and Application Domains:

B4.2. - Nuclear Energy Production
B4.3. - Renewable energy production
B5.1. - Factory of the future
B5.9. - Industrial maintenance
B6.3.2. - Network protocols
B6.3.3. - Network Management
B6.4. - Internet of things
B7.2. - Smart travel
B7.2.1. - Smart vehicles
B7.2.2. - Smart road
B8.1.2. - Sensor networks for smart buildings

1. Personnel

Research Scientists

Paul Muhlethaler [Team leader, Inria, Senior Researcher, HDR]
Pascale Minet [Inria, Researcher, HDR]
Thomas Watteyne [Inria, Starting Research Position]

External Collaborators

Nadjib Achir [Univ Paris-Nord, HDR]
Nadjib Ait Saadi [Univ Paris-Est Marne La Vallée]
Selma Boumerdassi [CNAM]
Samia Bouzefrane [CNAM]
Mohamed Elhadad [Univ René Descartes Paris, until Sep 2017]
Philippe Jacquet [Bell Labs (Nokia), HDR]
Anis Laouiti [France Telecom, HDR]
Dana Marinca [Univ de Versailles Saint-Quentin-en-Yvelines]
Malisa Vucinic [University of Montenegro, from Oct 2017]

Technical Staff

Ines Khoufi [Inria]

PhD Students

Keoma Brun-Laguna [Inria]
Jonathan Munoz [GridBee, Inria]
Nesrine Ben Hassine [Inria, until Sep 2017]
Younes Bouchaala [Vedecom, until Sep 2017]
Abdallah Sobehy [Télécom Sud-Paris, from Oct 2017]
Amar Abane [CNAM]

Post-Doctoral Fellows

Ziran Zhang [Inria]
Tengfei Chang [Inria]
Rémy Leone [Inria]
Malisa Vucinic [Inria, until Sep 2017]
Ehsan Ebrahimi Khaleghi [Inria, until Jun 2017]

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 focus 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. Generalities

EVA inherits its expertise in designing algorithms and protocols from HiPERCOM2 (e.g. OLSR). EVA also inherit know-how in modeling, simulation, experimentation and standardization. Through this know-how and experience, the results obtained are both far-reaching and useful.

3.2. Physical Layer

We plan to study how advanced physical layers can be used in low-power wireless networks. For instance, collaborative techniques such as multiple antennas (e.g. the Massive MIMO technology) can improve communication efficiency. The idea is to use a 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 will pursue 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 will focus particularly on scheduled medium access in the context of deterministic industrial networks; this will involve optimizing the joint time slot and channel assignment. Distributed approaches will be considered, and the EVA team will determine their limits in terms of reliability, latency and throughput. Furthermore, adaptivity to application or environment changes will be 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 IEEE802.15.4 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.

[0] Highway Addressable Remote Transducer

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.

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 (subcarriers) 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

1. **Pascale Minet**, **Paul Muhlethaler** and Ines Khoufi received the best paper award for their paper "Coded Slotted Avoidance in a Wireless Network: Models and Simulations" at PEMWN 2017.
2. SolSystem selected as one of the 10 testbeds at the IoT Solutions World Congress, Barcelona, Spain, 3-5 October 2017.
3. SmartMesh IP awarded "Internet of Things Product of the Year" at the Annual Creativity in Electronics (ACE) Awards, 6 December 2017. (Note: this is not a personal award)

5.1.2. 6TiSCH Standardization Virtually Completed

Time Synchronized Channel Hopping (TSCH) is a Medium-Access Control (MAC) technique in which nodes synchronize, and a schedule orchestrates all communication in the network. Inria-EVA created the IETF 6TiSCH Working Group in 2013. The goal of 6TiSCH is to get the best of both world by combining TSCH ("industrial" performance) and the ease of use of IPv6 through the IETF upper stack (6LoWPAN, RPL, CoAP). Since the creation of 6TiSCH in October 2013, **Thomas Watteyne** co-chairs the working group, helps drive its technical developments, and coaches authors and authors technical documents. 6TiSCH also encompasses and important security aspect, where we look how to enable nodes to join a network efficiently, which includes mutual authentication between node and network. The 6TiSCH security solution if based on PSK, and relies on AES-128 CCM*.

421 people now follow the 6TiSCH activities through its mailing-list, with a healthy mix of industrial and academic contributors. In 2017, 6TiSCH has produced 2 RFCs, 6 working group document in the process of being published, and various individual submissions. The working group has met 3 times in person during 2017, tens of times through Webex. Inria-EVA co-organized a 6TiSCH interop event (attended by 15 entities) in July 2017. 6TiSCH is now supported by all major open-source implementations (OpenWSN , Contiki , RIOT, TinyOS), and several companies are building commercial product lines with it. 6TiSCH has been playing a real role of catalyst for the academic low-power wireless community, which has now mostly moved towards TSCH/6TiSCH.

5.1.3. Over 1,000 Sensors Deployed on 3 Continents

Inria-EVA uses SmartMesh IP as a low-power wireless building block for building end-to-end solutions. Deploying real networks allows Inria-EVA to do system-level cross-disciplinary research. Inria-EVA oversees over 1,000 sensors deployed on 3 continents:

- http://snowhow.io/. Monitoring the snowmelt process in the California Sierra Nevada. 945 sensors deployed in 21 networks. Collaboration with UC Berkeley Prof. Steven Glaser.
- http://www.savethepeaches.com/. Predicting frost events in peach orchards. 120 sensors deployed in Mendoza, Argentina. Collaboration with local agronomy/networking teams
- http://smartmarina.org/. Monitoring the occupancy and per-boat water/electricity consumption of the 3rd largest marina in Europe (Cap d'Agde, 4300 boats). Inria-EVA is working on turning this activity into a startup company.

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

6.8. Platforms

6.8.1. SolSystem

In collaboration with University College London and IBM, we have designed a cloud-based low-power network management solution called SolSystem. It serves as a "control tower" for the networks we deploy, allowing us to manage both the network and data produced by those networks. It is architected following the micro-service principle, and we are in the process of switching all of our deployments to that interface. Fig. 1 gives an example of the visualization the SolSystem web interface gives us.

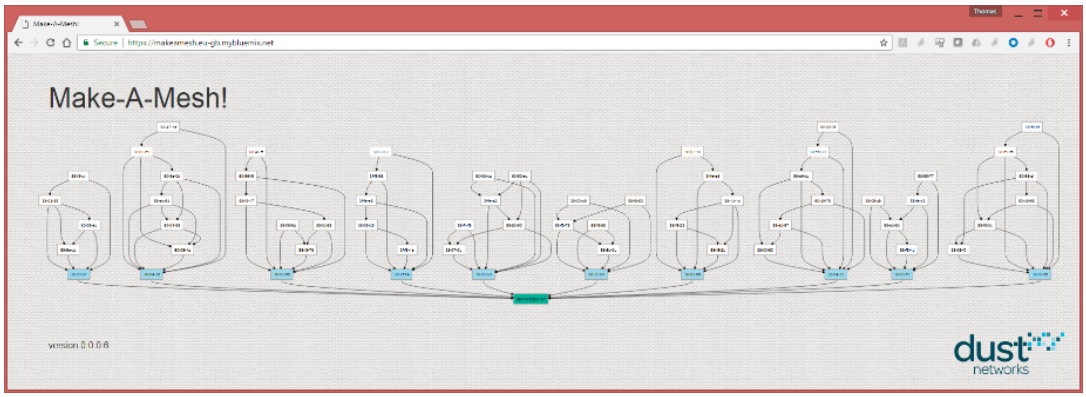


Figure 1. Topological view of the canopy network deployed across the Robert's building at University College London from February to April 2017, using SolSystem.

6.8.2. OpenMote B

In collaboration with OpenMote (http://www.openmote.com/), we have designed the OpenMote B platform. This board contains both a CC2538 IEEE802.15.4 radio, and an AT86RF215 IEEE802.15.4g radio, offering communication on both 2.4 GHz and sub-GHz frequency bands, 4 modulations schemes, and data rates from 50 kbps to 800 kbps. The first prototypes (shown in Fig. 2) started being tested in December 2017.

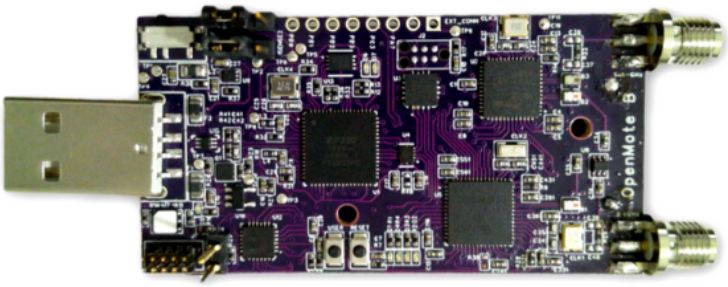


Figure 2. The OpenMote B. sub-GHz radio (and antenna connector) on top, 2.4 GHz radio on the bottom.

7. New Results

7.1. 6TiSCH Standardization and Benchmarking

7.1.1. Minimal Security Solution

Participants: Malisa Vucinic, Thomas Watteyne.

The 6TiSCH standardization effort had, until 2017, a big gap: security. Thanks to the work of Malisa Vucinic, this gap is now filled, with the publication of the Minimal Security solution (draft-ietf-6tisch-minimal-security). Here is a summary of what has been implemented and tested:

- Two implementations of the OSCORE protocol, formerly known as OSCOAP, specified in draft-ietf-core-object-security-03, in C and in Python, supporting both client and server roles, as part of the OpenWSN stack. Updated the test suite of the Python implementation with OSCOAP functional tests.
- Two implementations of Simple Join Protocol for 6TiSCH, specified in draft-ietf-6tisch-minimal-security-03, in C supporting the role of a pledge and in Python, supporting the role of JRC. Written unit tests for the implemented CBOR decoder in C.
- Simulation of the join process in 6TiSCH simulator. Extended the simulator to support shared cells, downwards RPL routing and join traffic. Tested the two implementations of Simple Join Protocol/OSCOAP using the F-Interop tools.

7.1.2. OpenWSN Fresh with full 6TiSCH Support

Participants: Tengfei Chang, Thomas Watteyne.

Thanks to the incredible work of Tengfei Chang, the OpenWSN project was refocused on being the lead reference 6TiSCH implementation. “OpenWSN Fresh” was a 2017 program to separate the protocol stack implementation from the rest of the OpenWSN code, and to have full standards-compliance.

7.1.3. First F-Interop 6TiSCH Interop Event

Participants: Remy Leone, Tengfei Chang, Malisa Vucinic, Thomas Watteyne.

The 6TiSCH WG organized an interoperability event co-located with the IETF meeting in Prague in July 2017. The interop tests focused on the minimal security framework and the 6top protocol. OpenWSN was used as the reference implementation, and F-Interop tools were demonstrated.

7.1.4. Agile Networking

Participants: Jonathan Munoz, Thomas Watteyne.

Today's low-power wireless devices typically consist of a micro-controller and a radio. The most commonly used radios are IEEE802.15.4 2.4GHz, IEEE802.15.4g sub-GHz and LoRA (SemTech) compliant. Radios offer a different trade-off between range and data-rate, given some energy budget. To make things more complex, standards such IEEE802.15.4g include different modulations schemes (2-FSK, 4-FSK, O-QPSK, OFDM), further expanding the number of options.

The main idea behind agile networking is to redefine a low-power wireless device as having multiple radios, which it can possibly use at the same time. That is, in a TSCH context, for each frame a node sends, it can change the radio it is using, and its setting. If the next hop is close, it sends the frame with a fast data rate thereby reducing the radio on-time and the energy consumption. If the next hop is far, it uses a slower data rate.

We recently design the OpenMote B within the OpenMote company. This board contains both a CC2538 IEEE802.15.4 radio, and an AT86RF215 IEEE802.15.4g radio, offering communication on both 2.4GHz and sub-GHz frequency bands, 4 modulations schemes, and data rates from 50 kbps to 800 kbps. The first prototypes started being tested in December 2017.

The second challenge is to redesign the protocol stack in a standards-compliant way. We are working with Jonathan Munoz on a 6TiSCH design in which neighbor discovery happens independently on each radio, and the same neighbor node can appear as many times in the neighbor table as it has radios. The goal is to standardize an "Agile 6TiSCH" profile, without having to touch the core specifications. This is been implemented in OpenWSN. The next step is to evaluate the performance of the solution on an 80-node OpenMote B testbed we are putting together. We hope to show that a single device running the same stack can satisfy both building-size and campus-size deployment, with the same industrial requirements.

7.2. SolSystem Deployments

SolSystem (http://solsystem.io/) is a complete sensor-to-cloud solution, which the Inria-EVA team uses to federate the different real-world deployments it is conducting.

7.2.1. SmartMarina

Participants: Ziran Zhang, Keoma Brun-Laguna, Thomas Watteyne.

Marinas are quickly evolving from sailing spots to floating neighborhoods. It is now common for people to live on their boat year-round, and for boats to be rented for just a week-end through online platforms. Today, living or staying on a boat is often cheaper that buying or renting an apartment. Similarly, in coastal areas, the marina is often the center of the city, so an ideal location for lodging. As a result, the trend is not going to end any time soon. Today's marinas are tomorrow's smart cities.

And as the marina is evolving, so are its needs.

- From a marina management point of view, automatic mooring management and electricity/water monitoring allows personnel to free up to welcome visitors and focus entirely on their well-being.
- Year-road boat owners and occasional marina visitors now can enjoy new services, from increased mooring availability to remote monitoring and alerts about the state of their boat.

The combination of embedded micro-controllers, low-power wireless communication and sensors/actuators offers tremendous opportunities for marinas. Off-the-shelf "Internet of Things" technology can now be used to detect the presence of boats in moorings, track usage of water and electricity on a per-boat basis, track a boat in real-time as it enters the marina, etc. Because no wires need to be installed – neither for power, nor communication – installation can be done in a matter of hours in a peal-and-stick fashion. Pontoons can be moved, rearranged or removed, without having to worry about the smart devices mounted on it.

The goal of the SmartMarina project (http://smartmarina.org/) is to build a system composed of sensors deployed all over the marina, and advanced software to monitor the occupation of moorings, and the electricity and water consumption on each spot. The result is a system that allows more efficient management and new services. The first sensor was installed in April 2017, and the Inria-EVA team is looking at turning this activity into a startup company.

7.2.2. SaveThePeaches

Participants: Keoma Brun-Laguna, Thomas Watteyne.

In 2013, 85% of the peach production in the Mendoza region (Argentina) was lost because of frost. Because less fruit was produced in the region, 600.000 less work days were needed to process the harvest between November 2013 and March 2014, a reduction in work force of 10.600 people. Across the Mendoza region, frost has caused a loss of revenue of 950 million Argentine pesos roughly 100 million USD in the peach business alone.

A frost event happens when the temperature is so low that the crops cannot recover their tissue or internal structure from the effects of water freezing inside or outside the plant. For the peach production, a critical period is when the trees are in bloom and fruit set (Aug./Sept. in Mendoza), during which the temperature needs to be kept above 3 C. Even a few hours below that temperature causes flowers to fall, preventing fruits to grow.

Because of the huge economic impact, countermeasures exist and are used extensively. Today, virtually all industrial peach orchards are equipped with a small number of meteorological stations which monitor temperature and humidity. If the temperature drops dangerously low, the most effective countermeasures is to install a number of furnaces in the orchard (typically coalfueled) and fly helicopters above the orchard to distribute the heat and avoid cold spots. This countermeasure is effective, but suffers from false positives (the helicopters are called in, but there is no frost event) and false negatives (the meteorological stations don't pick up a frost event happening in some part of the orchard).

What the SaveThePeaches project (http://www.savethepeaches.com/) has developed in 2016-2017 is a dense 120-sensor real-time monitoring solution deployed in the orchard, and feeding a frost prediction model. A node is the size of a deck of cards, is self-contained and battery-operated. When switched on, nodes form a multi-hop low-power wireless network, automatically. Rather than being installed at a fixed location, these nodes can be hung directly in the trees. A network is deployed in an orchard in a matter of hours, and if needed, sensing points can be moved to improve the accuracy of the prediction model in minutes. We use machine learning and pattern recognition to build an micro-climate predictive model by continuously analyzing the gathered sensor data in real time. This model generates early frost warnings. Ones demonstrated, the solution can be extended to other crops, and other regions.

7.2.3. SnowHow

Participants: Keoma Brun-Laguna, Thomas Watteyne.

Between 2012 and 2015, California suffered from the highest water drought since recordings started in this state. Up to 2/3 of its water resources are coming from the Sierra Nevada snowpack. Understanding the effect of the droughts on the mountain snowpack is crucial.

Historically, the study of mountain hydrology and the water cycle has been largely observational, with variables extrapolated from a few infrequent manual measurements. Low-power wireless mesh networking technology has evolved significantly over recent years. With this technology, a node is the size of a deck of cards, is self-contained and battery-operated. When switched on, nodes form a multi-hop low-power wireless network, automatically. Next-generation hydrologic science and monitoring requires real-time, spatially distributed measurements of key variables including: soil moisture, air/soil temperature, snow depth, and air relative humidity.

The SnowHow project (http://snowhow.io/) provides these measurements by deploying low-power mesh networks across the California Sierra Nevada. Off-the-shelf commercial solutions are available today which offer >99.999% end-to-end data reliability and a decade of battery lifetime. A new wireless network can be deployed in a couple of hours and report sensor data minutes after it was measured.

7.3. 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.

7.3.1. Deployment of autonomous and mobile wireless sensor nodes

Participants: Ines Khoufi, Pascale Minet.

This work was done in collaboration with Nadia Boufares (ENSI, University of Manouba, Tunisia) and Leila Saidane (ENSI, University of Manouba, Tunisia).

Wireless Sensor Networks (WSNs) are used in a wide range of applications due to their monitoring and tracking abilities. Depending on the applications goals, sensor nodes are deployed either in a two-dimensional (2D) area or in a three-dimensional (3D) area. In addition, WSN deployment can be either in a distributed or a centralized manner. In 2017, we were interested in a fully distributed deployment of WSN in several 3D-flat-surface configurations using autonomous and mobile nodes. Our goal was to ensure full 3D flat surfaces coverage and maintain network connectivity for these surfaces. To reach our goal we proposed 3D-DVFA-FSC, a distributed deployment algorithm based on virtual forces strategy to move sensor nodes over different 3D-flat-surface shapes. Initially, nodes were randomly deployed. Full coverage was reached in the given configurations and maintained up to the end of simulation. We also evaluated the total distance traveled by nodes. Simulation results show that sensor nodes still move even when full 3D-surface coverage is reached. This is due to the node oscillations problem. This problem will be tackled in our future work. We will also focus on how to stop nodes when full coverage is reached and consider 3D surface complex shapes where the challenges of coverage and connectivity are more complicated. This work was presented at the IWCMC 2017 conference, see [15].

7.3.2. Collision avoidance on shared slots in wireless slotted networks

Participants: Ines Khoufi, Pascale Minet, Paul Muhlethaler.

We propose an analysis of slotted based protocols designed for devices of the Internet of Thing (IoT). In contrast to other TDMA-based protocols this scheme uses a random technique to access shared slots which presents similarities with CSMA protocols. In practice the transmissions are scheduled in a given back-off window of slots whose duration allows the transmission of a packet and its acknowledgment. Therefore this protocol can be analyzed according to the methodology introduced by Bianchi for the IEEE 802.11 protocol even if the protocol studied differs in many aspects. The model we use is also particular because we succeed in obtaining a Markov model even if the scheme used to send a packet (in a node) may depend on the transmission of the previous packet(s). We distinguish two protocols. In the first one, at the initial stage or after a successful transmission, the packets are transmitted without any back-off, whereas in the second protocol each transmission is always preceded by the count down of a random back-off. Extensive simulations show a very good match between the model and the simulation results, see [22]. For moderate medium load, the protocol performing a backoff before each transmission outperforms the TSCH protocol, when the number of neighboring nodes is greater than or equal to 8. For a smaller number of neighboring nodes, the TSCH protocol provides a higher throughput. For high medium load, the TSCH protocol provides the highest normalized throughput at the cost of some unfairness in the transmission opportunities.

7.3.3. Security in the OCARI wireless sensor network

Participant: Pascale Minet.

Wireless Sensor Networks and Industrial Internet of Things use smart, autonomous and usually limited capacity devices in order to sense and monitor industrial environments. The devices in a wireless sensor network are managed by a controller, also called CPAN, which should authenticate them before they join the network. OCARI is a promising wireless sensor network technology providing optimized protocols in order to reduce the energy consumption and support pedestrian mobility. However, it needs to be secured against the different threats, especially those that concern confidentiality, data integrity, and entities authentication. This challenge was addressed in a joint work with Mohammed Tahar Hammi (Telecom ParisTech), Erwan Livolant (AFNet, Boost technologies), Patrick Bellot (Telecom ParisTech), Ahmed Serhouchni (Telecom ParisTech) and **Pascale Minet** (Inria). The main results have been published in two papers.
A robust mutual authentication is the challenge addressed in the paper presented at the ICMWT 2017 conference [28]. We proposed a lightweight, robust, and energy efficient WSN mutual authentication protocol. This protocol is especially designed to be implemented on devices with low storage and computing capacities. It has been implemented on OCARI. All nodes wanting to access the network should be authenticated at the MAC sub-layer of OCARI. This solution provides a protection against "replay attacks", because the exchanged OTPs are based on random numbers, therefore, they are valid only for one transaction. Using the blacklisting mechanism we can secure our systems against "some DoS" attacks. Finally it is flexible and does not decrease the scalability of the system, and can be deployed in different WSNs technologies, while keeping the same level of robustness. In our future work we aim to ensure the confidentiality of the transmitted messages exchanged after the MAC sub-layer association and authentication procedure. And thus we will have a secure system which ensures the "Confidentiality", "Integrity, and "Authentication" services.
In the paper presented at CSNet 2017 ([27] , we designed a security protocol that enables to secure most of the WSNs thanks to its lightness and energy efficiency. It ensures a mutual authentication of the communicating entities and a protection of both the integrity and the confidentiality of the exchanged data. The "personalization" mechanism solves the problem of the internal identity usurpation. The proposed key management allows a safe and secure keys exchange between the concerned entities. Furthermore, this protocol provides a very fast establishment of a secure channel based on a robust, fast, and lightweight symmetric encryption algorithm (AES GCM/CCM). Finally, this solution is resilient against the cryptanalysis and the replay attacks. In our future works, we aim to create a secure communicating system between different CPANs and to facilitate a secure migration of devices from a network managed by a CPAN to a network managed by another CPAN.

7.3.4. Security in Wireless Sensor Networks

Participants: Selma Boumerdassi, Paul Muhlethaler.

Sensor networks are often used to collect data from the environment where they are located. These data can then be transmitted regularly to a special node called a *sink*, which can be fixed or mobile. For critical data (like military or medical data), it is important that sinks and simple sensors can mutually authenticate so as to avoid data to be collected and/or accessed by fake nodes. For some applications, the collection frequency can be very high. As a result, the authentication mechanism used between a node and a sink must be fast and efficient both in terms of calculation time and energy consumption. This is especially important for nodes which computing capabilities and battery lifetime are very low. Moreover, an extra effort has been done to develop alternative solutions to secure, authenticate, and ensure the confidentiality of sensors, and the distribution of keys in the sensor network. Specific researches have also been conducted for large-scale sensors. At present, we work on an exchange protocol between sensors and sinks based on low-cost shifts and xor operations. After this publication, we have been working on the performance evaluation of the solution to determine the memory overhead together with both computing and communication latencies.

7.3.5. Massive MIMO Cooperative Communications for Wireless Sensor Networks

Participants: Nadjib Achir, Paul Muhlethaler.

This work is done in collaboration with Mérouane Debbah (Supelec, France).

The objective of this work is to propose a framework for massive MIMO cooperative communications for Wireless Sensor Networks. Our main objective is to analyze the performances of the deployment of a large number of sensors. This deployment should cope with a high demand for real time monitoring and should also take into account energy consumption. We have assumed a communication protocol with two phases: an initial training period followed by a second transmit period. The first period allows the sensors to estimate the channel state and the objective of the second period is to transmit the data sensed. We start analyzing the impact of the time devoted to each period. We study the throughput obtained with respect to the number of sensors when there is one sink. We also compute the optimal number of sinks with respect to the energy spent for different values of sensors. This work is a first step to establish a complete framework to study energy efficient Wireless Sensor Networks where the sensors collaborate to send information to a sink. Currently, we are exploring the multi-hop case.

7.4. Industry 4.0 and Wireless Sensor Networks

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.

Several standards have been designed for industrial wireless sensor (IoT) networks such as WirelessHart and ISA100. Both of them are based on the IEEE 802.15.4 standard for the lower layers. More recently, Time Slotted Channel Hopping (TSCH) which is specified in amendment e of the IEEE 802.15.4 standard, uses a time slotted medium access operating on several channels simultaneously. In addition, radio perturbations are mitigated by frequency hopping. TSCH supports star and mesh topologies, as well as multi-hop communication. It has been designed for process automation, process control, equipment monitoring and more generally the Internet of Things. It is a candidate technology for the Industry 4.0. In fact, Industry 4.0 will use more and more the on-demand manufacturing in a highly fexible and widespread environment. Different supply chains located in various regions need to coordinate their actions in a real-time basis with high fidelity. The IoT communicating in a wireless manner will play a major role to achieve this target. Time Slotted Channel Hopping (TSCH) networks are emerging as a promising technology for the Internet of Things and the Industry 4.0 where ease of deployment, reliability, short latency, flexibility and adaptivity are required. However, the strong requirements in terms of short latency and high reliability of such applications are obstacles to its penetration in the Industry 4.0. That is why in 2017 we made three contributions dealing with:

- how to quickly build a TSCH network;
- how to increase the reliability of end-to-end communications;
- how to efficiently schedule the transmissions made for data gathering.

7.4.1. Building an IEEE 802.15.4e TSCH network

Participants: Ines Khoufi, Pascale Minet.

The IEEE 802.15.4e amendment has been designed to meet the requirements of industrial applications with regard to the wireless sensor networks supporting them. Because of its scheduled medium access and multichannel transmissions, the TSCH mode has received much attention. 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 DBA, a Deterministic Beacon Advertisement algorithm that ensures a regular transmission of beacons without collisions. The goal

of DBA is to ensure that beacons are transmitted on all frequencies used by the TSCH network, regularly and without collision. With DBA, the exact value for the maximum time for a joining node to detect a beacon can be computed easily. We use the NS3 Simulator to evaluate this time as well as the the number of message losses, considering different network topologies (star or multihop). Simulation results show that DBA clearly outperforms existing solutions such as Random Vertical and Random Horizontal, two algorithms existing in the state of the art. In addition, DBA is able to provide the exact value of the maximum joining time.These results have been presented at the EUCASS 2017 conference, see [31].

7.4.2. Increasing the reliability of an IEEE 802.15.4e TSCH network

Participants: Ines Khoufi, Pascale Minet.

Our goal is to improve reliability of data gathering in such wireless sensor networks. We present three redundancy patterns to build a reliable path from a source to a destination. The first one is the well-known two node-Disjoint paths. The second one is based on a Triangular pattern, and the third one on a Braided pattern. The reliability provided by each pattern, the delivery time and the overhead in terms of the number of transmissions generated by each pattern as well as the amount of energy consumed by an end-to-end transmission allows us to conclude that the Braided pattern provides the highest reliability but with an overhead approximately twice the overhead of the Disjoint-path pattern and 4/3 the overhead of the Triangular pattern. These performance results are corroborated by simulations performed with NS3 for various configurations. This result has been presented at the NCA 2017 conference ([21]).

7.4.3. Scheduling transmissions in an IEEE 802.15.4e TSCH network

Participants: Ines Khoufi, Pascale Minet.

TSCH provides a multichannel slotted medium access ruled by a periodic schedule and supports multihop communications. This schedule is repeated every slotframe. A slotframe consists of a set of cells, each cell is identified by a (time slot offset, channel offset) pair. The size of a timeslot (e.g. 10 ms by default) allows the transmission of a point-to-point frame and its immediate acknowledgment. The schedule defines for each cell the nodes allowed to transmit and those that should receive. The channel offset is translated into a physical channel depending on the channel hopping sequence of the TSCH network. Channel hopping allows the TSCH to increase its robustness against external perturbations of the radio signal.

In the paper presented at VTC-Fall 2017 [20], we study how applications with data delivery constraints can be supported by a TSCH network. We first propose a framework based on a multislotframe that allows the coexistence of Data Slotframes and Control Slotframes. We then determine a lower bound on the minimum number of slots required to perform data gathering, taking into account the number of channels, the number of interfaces of the sink, the number of packets generated by each sensor node as well as the number of children of the sink. These feasibility conditions are established for two cases: with spatial reuse and without. We propose a debt-based scheduler that for simple topologies, provides a schedule minimizing the slotframe size. We determine the conditions for which an increase in the number of channels or sink's interfaces leads to a shorter data delivery delay. We compare the number of slots needed by data gathering with and without spatial reuse for small configurations. Finally, we consider a network configuration representative of an industrial application and evaluate the performance of the TSCH network in terms of data delivery delay and queue size for each sensor node, using the NS-3 simulator, where the multislotframe has been integrated. Simulation results Simulation results showed that the maximum theoretical delivery delay is never exceeded and the number of messages in the Transmit queue of each sensor node remains small. In addition, the debt-based scheduler builds a valid schedule with the minimum number of slots for the industrial application considered. we can conclude that TSCH with its time-slotted and multichannel medium access provides an efficient support for data gathering.

7.5. Machine Learning for an efficient and dynamic management of network resources and services

7.5.1. Machine Learning in Networks

Participants: Nesrine Ben Hassine, Dana Marinca, Pascale Minet.

This work was done in collaboration with Dominique Barth (UVSQ) .

Content Delivery Networks (CDNs) are faced with an increasing and time varying demand of video contents. Their ability to promptly react to this demand is a success factor. Caching helps, but the question is: which contents to cache? We need to know which resources are needed before they are requested. This anticipation is made possible by using prediction computed by learning techniques.

Machine learning techniques can be used to improve the quality of experience for the end users of Content Delivery Networks (CDNs). In a CDN, the most popular video contents are cached near the end-users in order to minimize the contents delivery latency. Classically, machine learning techniques are classified as supervised or unsupervised. In 2017, we addressed two challenges:

- as a supervised learning, the use of prediction techniques based on regression to evaluate the future popularity of video contents in order to decide which ones should be cached. The popularity of a video content is evaluated by the number of daily requests for this content.
- as an unsupervised learning, the use of clustering techniques to put together videos with similar features. This clustering will reduce the number of prediction methods, called experts, used to provide an accurate prediction.

7.5.2. Prediction of video content popularity

Participants: Nesrine Ben Hassine, Dana Marinca, Pascale Minet.

This work was done in collaboration with Dominique Barth (UVSQ).

We consider various experts, coming from different fields (e.g. statistics, control theory). To evaluate the accuracy of the experts' popularity predictions, we assess these experts according to three criteria: cumulated loss, maximum instantaneous loss and best ranking. The loss function expresses the discrepancy between the prediction value and the real number of requests. We use real traces extracted from YouTube to compare different prediction methods and determine the best tuning of their parameters. The goal is to find the best trade-off between complexity and accuracy of the prediction methods used.
We also show the importance of a decision maker, called forecaster, that predicts the popularity based on the predictions of a selection of several experts. The forecaster based on the best K experts outperforms in terms of cumulated loss the individual experts' predictions and those of the forecaster based on only one expert, even if this expert varies over time.
The paper presented at the Wireless days 2017 conference ([29] is the result of a joint work done in collaboration with Ruben Milocco (Universidad Nacional Comahue, Buenos Aires, Argentina) and Selma Boumerdassi (CNAM, Paris). We focused on predicting the popularity of video contents using Auto-Regressive Moving Average (ARMA) methods applied on a sliding window. These predictions are used to put the most popular video contents into caches. After having identified the parameters of ARMA experts, we compare them with an expert predicting the same number of requests as the previous day. Results show that ARMA experts improve the accuracy of the predictions. Nevertheless, there is no ARMA model that provides the best prediction for all the video contents over all their lifetime. We combine these statistical experts with a higher level of experts, called forecasters. By combining the experts prediction, some forecasters succeed in predicting more accurate values which helped to increase the hit ratio while keeping a correct update ratio. Hence, improving the accuracy of the predictions succeeds in improving the hit ratio. To summarize, we proposed an original solution combining the predictions of several ARMA models. This solution achieves a better Hit Ratio and a smaller Update Ratio than the classical Least Frequently Used (LFU) caching technique.

7.5.3. Clustering of video contents

Participants: Nesrine Ben Hassine, Pascale Minet.

With regard to video content clustering, we proposed an original solution based on game theory that was presented at the CCNC 2017 conference ([30]. This is a joint work with Mohammed-Amine Koulali (Mohammed I University Oujda, Morocco), Mohammed Erradi (Mohammed I University Rabat, Morocco), Dana Marinca (University of Versailles Saint-Quentin) and Dominique Barth (University of Versailles Saint-Quentin). Game theory is a powerful tool that has recently been used in networks to improve the end users' quality of experience (e.g. decreased response time, higher delivery rate). In this paper, the original idea

consists in using game theory in the context of Content Delivery Networks (CDNs) to organize video contents into clusters having similar request profiles. The popularity of each content in the cluster can be determined from the popularity of the representative of the cluster and used to store the most popular contents close to end users. A group of experts and a decision-maker predict the popularity of the representative of the cluster. This considerably reduces the number of experts used. More precisely, we model the clustering problem as a hedonic coalition formation game where the players are the video contents. We proved that this game always converges to a stable partition consisting of different clusters. We determined the best size of the observation window and showed that the play order minimizing the maximum distance to the representative of the cluster is the Rich-to-Poor order, whatever the number of video contents in the interval [20; 200]. The complexity of the coalition game remains very light. Convergence is obtained in a small number of rounds (i.e. less than 35 rounds for 200 video contents). We compare the results of this approach with the clustering obtained by the K-means algorithm, using real traces extracted from YouTube. We also evaluate the complexity of the proposed algorithm. The coalition game outperforms K-means in terms of the average and maximum distances to the representative of the cluster. The execution time is also in favor of the coalition game when the number of contents is higher than or equal to 50. Furthermore, the coalition game can be used to quickly determine the best value of K that is required as an input parameter of the K-means algorithm. Simulation results show that the coalition game provides very good performances.

7.6. Protocols and Models for Wireless Networks - Application to VANETs

7.6.1. Protocols for VANETs

7.6.1.1. TRPM: a TDMA-aware routing protocol for multi-hop communications in VANETs

Participants: Mohamed Elhadad Or Hadded, Paul Muhlethaler, Anis Laouiti.

The main idea of TRPM is to select the next hop using the vehicle position and the time slot information from the TDMA scheduling. Like the GPSR protocol, we assume that each transmitting vehicle knows the position of the packet's destination. In TRPM, the TDMA scheduling information and the position of a packet's destination are sufficient to make correct forwarding decisions at each transmitting vehicle. Specifically, if a source vehicle is moving in area x_i, the locally optimal choice of next hop is the neighbor geographically located in area x_{i+1} or x_{i-1} according to the position of the packet's destination. As a result, the TDMA slot scheduling obtained by DTMAC can be used to determine the set of next hops that are geographically closer to the destination. In fact, each vehicle that is moving in the area x_i can know the locally optimal set of next hops that are located in adjacent areas x_{i+1} or x_{i-1} by observing the set of time slots $S_{(i+3)\%3}$ or $S_{(i+1)\%3}$, respectively. We consider the same example presented above when vehicle G as the destination vehicle that will broadcast a message received from vehicle A. As shown in Figure 3, only two relay vehicles are needed to ensure a multi-hop path between vehicle A and G (one relay node in the area x_2 and another one in the area x_3).

In the following, the DTMAC protocol has been used by the vehicles to organize the channel access. The TDMA slot scheduling obtained by DTMAC is illustrated in Figure 3. Firstly, vehicle A forwards a packet to B, as vehicle A uses its frame information to choose a vehicle that is accessing the channel during the set S_1. Upon receiving the packet for forwarding, vehicle B will choose by using its frame information a vehicle that's accessing the channel during the set of time slots S_2 (say vehicle D). Then, vehicle D will forward the packet to G, as G is moving in area x_4 (accessing the channel during the set S_0) and it is the direct neighbor of vehicle D. By using DTMAC as the MAC layer, we can note that the path A-B-D-G is the shortest, in terms of the number of hops as well as the end-to-end delay which is equal to 6 time slots (2 time slots between t_0 and t_2 as t_2 is the transmission slot for vehicle B, then 2 time slots between t_2 and t_4 as t_4 is the transmission slot for vehicle D and finally 2 time slots between t_4 and t_0 as t_0 is the transmission slot in which vehicle G will broadcast the message received from vehicle A).

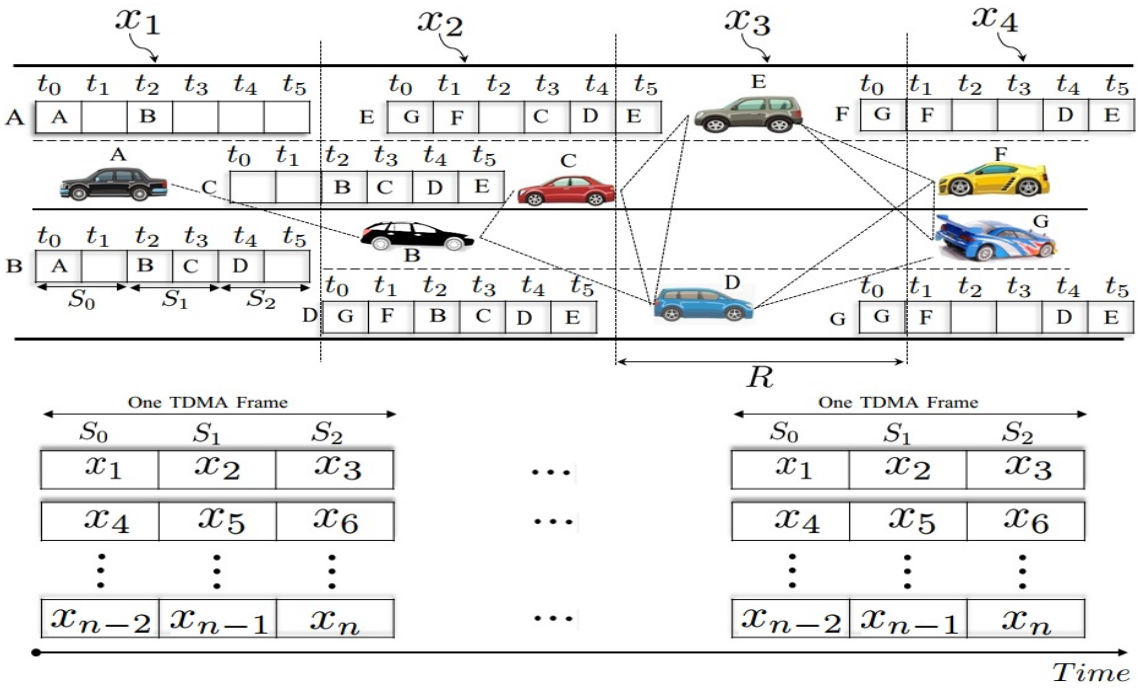


Figure 3. VANET network using DTMAC scheduling scheme.

The idea of TRPM [16] is the following . Whenever a vehicle i accessing the channel during the set S_k wants to send/forward an event-driven safety message, it constructs two sets of candidate forwarders based on its frame information FI as follows, where $TS(j)$ indicates the time slot reserved by vehicle j.

- $A_i = \{j \in N(i) \mid TS(j) \in S_{(k+1)\%3}\}$ *// The set of vehicles that are moving in the adjacent right-hand area.*
- $B_i = \{j \in N(i) \mid TS(j) \in S_{(k+2)\%3}\}$ *// The set of vehicles that are moving in the adjacent left-hand area.*

Each source vehicle uses the position of a packet's destination and the TDMA scheduling information to make packet forwarding decisions. In fact, when a source vehicle i is moving behind the destination vehicle, it will select a next hop relay that belongs to set B_i; when the transmitter is moving in front of the destination vehicle, it will select a forwarder vehicle from those in set A_i. For each vehicle i that will send or forward a message, we define the normalized weight function WHS (Weighted next-Hop Selection) which depends on the delay and the distance between each neighboring vehicle j. WHS is calculated as follows:

$$WHS_{i,j} = \alpha * \frac{\Delta t_{i,j}}{\tau} - (1-\alpha) * \frac{d_{i,j}}{R} \qquad (1)$$

Where:

- τ is the length of the TDMA frame (in number of time slots).
- j is one of the neighbors of vehicle i, which represents the potential next hop that will relay the message received from vehicle i.
- $\Delta t_{i,j}$ is the gap between the sending slot of vehicle i and the sending slot of vehicle j.
- $d_{i,j}$ is the distance between the two vehicles i and j, and R is the communication range.
- α is a weighted value in the interval $[0, 1]$ that gives more weight to either distance or delay. When α is high, more weight is given to the delay. Otherwise, when α is small, more weight is given to the distance.

We note that the two weight factors $\frac{\Delta t_{i,j}}{\tau}$ and $\frac{d_{i,j}}{R}$ are in conflict. For simplicity, we assume that all the factors should be minimized. In fact, the multiplication of the second weight factor by (-1) allows us to transform a maximization to a minimization. Therefore, the forwarding vehicle for i is the vehicle j that is moving in an adjacent area for which $WHS_{i,j}$ is the lowest value.

The simulation results reveal that our routing protocol significantly outperforms other protocols in terms of average end-to-end delay, average number of relay vehicles and the average delivery ratio.

We have developed an analytical model to evaluate the packet loss rate and the end-to-end delay for safety messages transmitted in vehicular networks over long distances when TRPM is used as a routing protocol, see refhadded:hal-01617924. Comparisons of realistic simulation results, carried out using ns-2 and MOVE/SUMO, and analytical results show that the analytical model proposed provides close approximations for the end-to-end delay and packet loss rate for the different scenarios considered.

7.6.1.2. Trust-CTMAC: A Trust Based Scheduling Algorithm

Participants: Mohamed Elhadad Or Hadded, Paul Muhlethaler, Anis Laouiti.

In Vehicular Ad hoc NETworks, communication is possible both between the vehicles themselves and between the vehicles and the infrastructure. These applications need a reliable and secure broadcast system that takes into consideration the security issues in VANETs, the high speed of nodes and the strict QoS requirements. For these reasons, we propose a trust-based and centralized TDMA-based MAC protocol. Our solution will permit Road Side Units (RSUs) to manage time slot assignment by avoiding malicious nodes and by minimizing message collision. The experiments carried out and the results obtained prove the effectiveness of our approach.

We present a trust based centralized TDMA scheduling mechanism which aims to isolate and prevent malicious vehicles from accessing the channel. This is accomplished by serving only the slot reservation requests of vehicles that have trust values greater than a trust threshold. In Trust-CTMAC, each RSU maintains additional data structure called Trust Counters Table (TCT) and Malicious Vehicles Table (MVT) for all vehicles within its communication range based on the list of properties shown in Table 1. The TCT and the FI information are periodically broadcasted by the RSU for ecah time interval of $100ms$. So each vehicle can identify and isolate malicious vehicles among all neighboring nodes based on the TCT information received from its RSU, which can protect the radio channel from any potential damage caused by the malicious vehicles. An RSU declares a vehicle as a malicious node if the corresponding trust value falls below a trust threshold.

Table 1. Threat lists that are checked in our trust platform

Threat Name	Description	Level
Message Saturation	A huge number of a vehicle packets do not include any form of identification information	3 (high)
False GNSS (Global Navigation Satellite System) Signals	A vehicle is sending messages with false geographic information	3 (high)
Slot reservation attack	A vehicle requests different slots during the same frame	3 (high)
Malicious MAC behavior	A vehicle is sending data in another slot different to its reserved one	4 (Critical)
Malicious isolation	Some vehicle functionalities are disabled (create, process, receive and send messages) caused by the installation of a malware	3 (high)
Denial of access to incoming messages	A vehicle may be unlinked if it receives a huge number of messages.	4 (Critical)
Frame information poisoning	The frame information is falsified by a vehicle	3 (high)
Identity spoofing	A vehicle is using a wrong node type in order to act as an RSU	3 (high)

7.6.1.3. A Flooding-Based Location Service in VANETs

Participants: Selma Boumerdassi, Paul Muhlethaler.

This work has been done in collaboration with Eric Renault, Telecom Sud Paris.

We have designed and analyzed a location service for VANETs; such a service can be used in Location-based routing protocols for VANETs. Our protocol is a proactive flooding-based location service that drastically reduces the number of update packets sent over the network as compared to traditional flooding-based location services. This goal is achieved by partially forwarding location information at each node. A mathematical model and some simulations are proposed to show the effectiveness of this solution. Cases for 1D, 2D and 3D spaces are studied for both deterministic and probabilistic forwarding decisions. We compare our protocol with the Multi-Point Relay (MPR) technique which is used in the OLSR protocol and determine the best technique according to the network conditions.

7.6.2. Models for Wireless Networks and VANETs

7.6.2.1. Performance analysis of IEEE 802.11 broadcast schemes with different inter-frame spacings

Participants: Younes Bouchaala, Paul Muhlethaler, Nadjib Achir.

This work has been done in collaboration with Oyunchimeg Shagdar (Vedecom).

We have started to build a model which analyzes the performance of IEEE 802.11p managing different classes of priorities. The differentiation of traffic streams is obtained with different inter-frame spacings: AIFSs (for Arbitration Inter Frame Spacings) and with different back-off windows: CWs (for Collision Windows). This model is based on a Markov model where the state is the remaining number of idle slots that a packet of a given class has to wait before transmission. However, in addition to this Markov model for which we compute a steady state we also consider the Markov chain which counts the number of idle slots after the smallest AIFS. As a matter of fact the probability these states are not evenly distributed since with different AIFSs the arrival rate is not constant when the number of idle slots experienced after the smallest AIFS varies. The resolution of the steady state of these two inter-mixed Markov chains lead to non linear and intertwined equations that can be easily solved with a software such as Maple. With the model we have obtained, we can compute the delivery rate of packets of different classes and show the influence of system parameters: AIFSs and CWs. The preliminary results show a a very strong influence of different AIFSs on the performance for each traffic streams, see [13].

7.6.2.2. Model and optimization of CSMA

Participants: Younes Bouchaala, Paul Muhlethaler, Nadjib Achir.

This work has been done in collaboration with Oyunchimeg Shagdar (Vedecom).

We have studied the maximum throughput of CSMA in scenarios with spatial reuse. The nodes of our network form a Poisson Point Process (PPP) of a one- or two-dimensional space. The one-dimensional PPP well represents VANETs. To model the effect of Carrier Sense Multiple Access (CSMA), we give random marks to our nodes and to elect transmitting nodes in the PPP we choose the nodes with the smallest marks in their neighborhood, this is the Matern hardcore selection process. To describe the signal propagation, we use a signal with power-law decay and we add a random Rayleigh fading. To decide whether or not a transmission is successful, we adopt the Signal-over-Interference Ratio (SIR) model in which a packet is correctly received if its transmission power divided by the interference power is above a capture threshold. We assume that each node in our PPP has a random receiver at a typical distance. We choose the average distance to its closest neighbor. We also assume that all the network nodes always have a pending packet. With these assumptions, we analytically study the density of throughput of successful transmissions and we show that it can be optimized with the carrier-sense threshold. The model makes it possible to analytically compute the performance of a CSMA system and gives interesting results on the network performance such as the capture probability when the throughput is optimized, and the effect on a non-optimization of the carrier sense threshold on the throughput. We can also study the influence of the parameters and see their effects on the overall performance. We observe a significant difference between 2D an 1D networks.

We have built two models to compare the spatial density of successful transmissions of CSMA and Aloha. To carry out a fair comparison, we optimize both schemes by adjusting their parameters. For spatial Aloha, we can adapt the transmission probability, whereas for spatial CSMA we have to find the suitable carrier sense threshold. The results obtained show that CSMA, when optimized, outperforms Aloha for nearly all the parameters of the network model values and we evaluate the gain of CSMA over Aloha. We also find interesting results concerning the effect of the model parameters on the performance of both Aloha and CSMA. The closed formulas we have obtained provide immediate evaluation of performance, whereas simulations may take minutes to give their results, see [14]. Even if Aloha and CSMA are not recent protocols, this comparison of spatial performance is new and provides interesting and useful results.

For Aloha networks, when we study transmissions over the average distance to the closest neighbor, the optimization does not depend on the density of nodes, which is a very interesting property. Thus in Aloha networks, the density of successful transmissions easily scales linearly in λ when we vary λ whereas in CSMA networks the protocol must be carefully tuned to obtain this scaling.

With CSMA, we have also shown that this density of throughput (when optimized) scales with the density of nodes if we study the throughput is measured between the nodes to their closest neighbors. We have mathematically justified this property.

7.6.2.3. Adaptive CSMA

Participants: Nadjib Achir, Younes Bouchaala, Paul Muhlethaler.

This work has been done in collaboration with Oyunchimeg Shagdar (Vedecom).

Using the model we have built for CSMA, we have shown that when optimized with the carrier sense detection threshold P_{cs}, the probability p^* of transmission for a node in the CSMA network does not depend on the density of nodes λ. In other words when the CSMA is optimized to obtain the largest density of successful transmissions (communication from nodes to their neighbors), p^* is constant. We have verified this statement on several examples and we think that a formal proof of this remark is possible using scaling arguments. The average access delay is a direct function of the probability of transmission p. Thus the average delay when the carrier sense detection threshold is optimized is a constant D_{target} which does not depend on λ. A stabilization algorithm which adapts P_{cs} to reach the D_{target} can thus be envisioned. Another stabilization algorithm adapts P_{cs} so that the mean number of neighbors of a node is N_{target} a given number of nodes which only depends on the network parameters and not on the network density. A third stabilization algorithm adapts P_{cs} so that the channel busy ratio (CBR) is near a given target.

We have justified theoretically all these algorithms and simulated their behavior. The simulations well justify the theoretical analysis.

7.6.2.4. Optimizing spatial throughput in device-to-device networks

Participants: Bartek Blaszczyszyn, Paul Keeler, Paul Muhlethaler.

Results are presented for optimizing device-to-device communications in cellular networks, while maintaining spectral efficiency of the base-station-to-device downlink channel. We build upon established and tested stochastic geometry models of signal-to-interference ratio in wireless networks based on the Poisson point process, which incorporate random propagation effects such as fading and shadowing. A key result is a simple formula, allowing one to optimize the device-to-device spatial throughput by suitably adjusting the proportion of active devices, see [19]. These results can lead to further investigation as they can be immediately applied to more sophisticated models such as studying multi-tier network models to address coverage in closed access networks.

7.6.2.5. Model and analysis of Coded Slotted Aloha (CSA) with capture

Participants: Ebrahimi Khaleghi, Cedric Adjih, Paul Muhlethaler.

This work has been done in collaboration with Amira Alloum, Nokia Bell Labs.

Motivated by scenario requirements for 5G cellular networks, we have studied one among the protocols candidate to the massive random access: the family of random access methods known as Coded Slotted ALOHA (CSA). Recent body of research has explored aspects of such methods in various contexts, but one aspect has not been fully taken into account: the impact of the path loss, which is a major design constraint in long-range wireless networks. We have explored the behavior of CSA, by focusing on the path loss component correlated to the distance to the base station. Path loss provides opportunities for capture, improving the performance of CSA. We have revised methods for estimating CSA behavior. We have provided bounds of performance and derived the achievable throughput. We have extensively explore the key parameters, and their associated gain (experimentally). Our results has shed light on the open question of the optimal distribution of repetitions in actual wireless networks.

7.6.2.6. Mobility Prediction in Vehicular Networks : An Approach through Hybrid Neural Networks under Uncertainty

Participants: Soumya Banerjee, Samia Bouzefrane, Paul Muhlethaler.

Conventionally, the exposure regarding knowledge of the inter vehicle link duration is a significant parameter in *Vehicular Networks* to estimate the delay during the failure of a specific link during the transmission. However, the mobility and dynamics of the nodes is considerably higher in a smart city than on highways and thus could emerge a complex random pattern for the investigation of the link duration, referring all sorts of uncertain conditions. There are existing link duration estimation models, which perform linear operations

under linear relationships without imprecise conditions. Anticipating, the requirement to tackle the uncertain conditions in *Vehicular Network*s, this paper presents a hybrid neural network-driven mobility prediction model. The proposed hybrid neural network comprises a *Fuzzy Constrained Boltzmann machine (FCBM*), which allows the random patterns of several vehicles in a single time stamp to be learned. The several dynamic parameters, which may make the contexts of *Vehicular Networks* uncertain, could be vehicle speed at the moment of prediction, the number of leading vehicles, the average speed of the leading vehicle, the distance to the subsequent intersection of traffic roadways and the number of lanes in a road segment. In this paper, a novel method of hybrid intelligence is initiated to tackle such uncertainty. Here, *the Fuzzy Constrained Boltzmann Machine (FCBM)* is a stochastic graph model that can learn joint probability distribution over its visible units (say n) and hidden feature units (say m). It is evident that there must be a prime driving parameter of the holistic network, which will monitor the interconnection of weights and biases of *the Vehicular Network* for all these features. The highlight of this paper is that the prime driving parameter to control the learning process should be a fuzzy number, as fuzzy logic is used to represent the vague and and uncertain parameters. Therefore, if uncertainty exists due to the random patterns caused by vehicle mobility, the proposed Fuzzy Constrained Boltzmann Machine could remove the noise from the data representation. Thus, the proposed model will be able to predict robustly the mobility in VANET, referring any instance of link failure under *Vehicular Network* paradigm.

7.6.3. Reliable routing architecture

Participants: Mohamed Hadded, Anis Laouiti, Paul Muhlethaler.

Flooding scheme represents one of the fundamental operation in wireless mesh networks. It plays an important role in the design of network and application protocols. Many existing flooding solutions have been studied to address the flooding issues in mesh networks. However, most of them are not able to operate efficiently where there are network equipment failures. In this work, we consider nodes failures and we build the flooding tree the maximum expectation of the throughput (taking into account the potential unavailability of certain nodes). After a formal stochastic definition of the problem, we show how to use a tabu search algorithm, to solve this optimization problem.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. CNES contract

Participants: Pascale Minet, Ines Khoufi.

Inria and CNES co-funded a study of one year in the framework of the CNES Launchers Research and Technology program. This study deals with the improvement and performance evaluation of a solution of wireless sensor networks in a spatial environment, based on the IEEE 802.15.4e standard of TSCH (Time Slotted Channel Hopping).
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:

- Avoiding collisions on shared slots: see the PEMWN 2017 conference.
- Building an IEEE 802.15.4e TSCH network: see the EUCASS 2017 publication.
- Increasing the reliability of an IEEE 802.15.4e TSCH network: see the NCA 2017 publication.
- Scheduling transmissions in an IEEE 802.15.4e TSCH network: see the VTC-Fall 2017 publication.

9. Partnerships and Cooperations

9.1. National Initiatives

- EVA has a collaboration with Vedecom. **Paul Muhlethaler** supervises Younes Bouchaala's PhD funded by Vedecom. This PhD aims at studying 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 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 in European Programs, Except FP7 & H2020

9.2.2.1. Collaborations with Major European Organizations

Inria-EVA has collaboration in 2017 with ETSI (the European Telecommunications Standards Institute) to organize the F-Interop 6TiSCH Interop Event in July 2017 in Prague.

9.3. International Initiatives

9.3.1. Inria International Labs

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

9.3.2.1. 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 IEEE802.15.4e (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.

9.3.2.2. 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.

9.3.3. Inria International Partners

9.3.3.1. 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.

9.3.3.2. 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 2017.

9.3.4. Participation in Other International Programs

9.3.4.1. International Initiatives

- **PEACH**
- Title: PrEcision Agriculture through Climate researcH
- International Partners (Institution - Laboratory - Researcher):
  - Universidad Diego Portales (Chile) - Diego Dujovne
  - Universidad Tecnológica de Mendoza (Argentina) - Gustavo Mercado
- Duration: 2016 - 2017

- In 2013, 85% of the peach production in the Mendoza region (Argentina) was lost because of frost. Because less fruit was produced in the region, 600.000 less work days were needed to process the harvest between November 2013 and March 2014, a reduction in work force of 10.600 people. Across the Mendoza region, frost has caused a loss of revenue of 950 million Argentine pesos - roughly 100 million USD - in the peach business alone. A frost event happens when the temperature is so low that the crops cannot recover their tissue or internal structure from the effects of water freezing inside or outside the plant. For the peach production, a critical period is when the trees are in bloom and fruit set (Aug./Sept. in Mendoza), during which the temperature needs to be kept above -3 C. Even a few hours below that temperature causes flowers to fall, preventing fruits to grow. Because of the huge economic impact, countermeasures exist and are used extensively. Today, virtually all industrial peach orchards are equipped with a small number of meteorological stations which monitor temperature and humidity. If the temperature drops dangerously low, the most effective countermeasures is to install a number of furnaces in the orchard (typically coal-fueled) and fly helicopters above the orchard to distribute the heat and avoid cold spots. This countermeasure is effective, but suffers from false negatives (the helicopters are called in, but there is no frost event) and false positives (the meteorological stations don't pick up a frost event happening in some part of the orchard). What is missing is a dense real-time monitoring solution deployed in the orchard, and feeding a frost prediction model. For this, having a couple of meteorological stations doesn't provide the measurement density needed. Frost events are micro-climatic: cold and hot air have a different density, wind blows irregularly between the trees, so different parts of an orchard are affected very differently by frost. What is needed are a large number of sensing points (humidity, temperature, wind speed), at different elevations, throughout the orchard. Low-power wireless mesh networking technology has evolved significantly over recent years. With this technology, a node is the size of a deck of cards, is self-contained and battery-operated. When switched on, nodes form a multi-hop low-power wireless network, automatically. Off-the-shelf commercial solutions are available today which offer >99.999% end-to-end data reliability and a decade of battery lifetime. Rather than being installed at a fixed location, these nodes can be hung directly in the trees. A network is deployed in an orchard in a matter of hours, and if needed, sensing points can be moved to improve the accuracy of the prediction model in minutes. And this solution is cheap, too: for the price one meteorological station, one can build 10 low-power wireless mesh sensing nodes. We use machine learning and pattern recognition to build an micro-climate predictive model by continuously analyzing the gathered sensor data in real time. This model generates early frost warnings. If successful, the solution can be extended to other crops, and other regions. The goal of this project is to dramatically increase the predictability of frost events in peach orchards by using dense monitoring using low-power wireless mesh networking technology. The project is designed to be completed in 24-month, and involves: (1) building a dense sensing solution based on off-the-shelf networking and sensing products, (2) developing accurate frost prediction models based on the sensing data gathered, (3) conducting real-world deployments on peach orchards in the Mendoza region. This project brings together world experts in agronomic and networking fields in a symbiotic manner. Perfectly in line with the philosophy of STIC-AmSud, the teams are already conducting cutting-edge research in their respective fields the funding we are applying for would enable the teams to collaborate together in a cross-disciplinary manner.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

1. *David Burnett* (*UC Berkeley*), Micro-Motes, collaboration with **Thomas Watteyne**, 30 November 2017.
2. *Prof. Xavi Vilajosana* (*UOC/OpenMote*), OpenMote B, 6TiSCH, collaboration with **Thomas Watteyne** and Tengfei Chang, 20 November – 19 December 2017.
3. *Pablo Modernell* (*UOC*), F-Interop, collaboration with Remy Leone and **Thomas Watteyne**, 20–27 November 2017.

4. *Malisa Vucinic* (*U Montenegro*), 6TiSCH Security, collaboration with **Thomas Watteyne**, 06-24 November 2017.
5. *Carlos Oroza* (*UC Berkeley*), Machine-Learning Based Placement Strategy, collaboration with **Thomas Watteyne**, 18 October – 06 November 2017.
6. *Prof. Xavi Vilajosana* (*UOC/OpenMote*), OpenMote B, the greatest thing since sliced bread, collaboration with **Thomas Watteyne** and Tengfei Chang, 19–20 September 2017.
7. *Felipe Lallane* (*Inria Chile*), Exploiring collaboration opportunities with Inria-Chile around IoT, collaboration with **Thomas Watteyne**, 19–20 June 2017.
8. *Cristina Cano* (*UOC, Barcelona*), Wireless Coexistence, collaboration with **Thomas Watteyne**, 16 May 2017.
9. *Ryan Grammenos* (*Univ. College London*), Machine Learning for 6TiSCH networks, collaboration with Keoma Brun-Laguna and **Thomas Watteyne**, 15–19 May 2017.
10. *Craig Schindler* (*UC Berkeley*), Industrial Process Control with 6TiSCH, collaboration with Tengfei Chang and **Thomas Watteyne**, 9–19 May 2017.
11. *Pedro Henrique Gomez* (*USC*), Exploiting Diversity in 6TiSCH Networks, collaboration with Tengfei Chang and **Thomas Watteyne**, 5 June – 9 July 2017.
12. *Prof. Diego Dujovne* (*UDP, Chile*), Advanced Scheduling in 6TiSCH networks, collaboration with **Thomas Watteyne**, 5–22 July 2017.
13. *Prof. Steven Glaser* (*UC Berkeley*), Real-time real-world remote sensing, collaboration with Ziran Zhang, Keoma Brun-Laguna, **Thomas Watteyne**, 27 May – 3 June 2017.
14. *Prof. Xavi Vilajosana* (*UOC/OpenMote*), OpenWSN core-team meet-up, collaboration with **Thomas Watteyne** and Tengfei Chang, 3–7 April 2017.

9.4.2. Internships

1. **Felipe Moran Correa Meyer**, sub-100 μs synchronization and sub-m RTLS with SmartMesh IP (ENSTA), September 2017 – August 2018.
2. **Fatima Adda**, simulation of active signaling in TDMA networks (Paris VI), March-August 2017.
3. **Nasr Khouaja Mohamed Hassine**, positioning with wireless networks (ENSTA), April-June 2017.

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.
- Keoma Brun-Laguna spent summer 2017 with the Dust Networks product team at Analog Devices in Silicon Valley as part of an internship.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organization

10.1.1.1. General Chair, Scientific Chair

- **Thomas Watteyne** co-chaired the IOT2SUSTAIN invited workshop (https://tinyurl.com/iot2sustain), which brought together 22 researcher (half UCL, half Inria) to brainstorm on the state of IoT research today, and identify research directions. This event was organized on the UCL campus in London, 6-7 July 2017.

- **Thomas Watteyne** co-chaired the IETF98 (Chicago, remotely), IETF99 (Prague) and IETF100 (Singapore) 6TiSCH Working group meetings.
- **Pascale Minet** was general co-chair with Leila Saidane from ENSI (Tunisia) of the PEMWN 2017 conference, the 6th IFIP 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/pemwn2017/). This conference was held at CNAM in Paris, the 28th, 29th and 30th of November 2017. It was sponsored by Inria, CNAM and ENSI. The organization co-chairs were Samia Bouzefrane and Selma Boumerdassi. Three tutorials were given:
  - *Dynamic Resource Allocation and Network Optimization in C-RAN* by Rami Langar, University of Paris-Est Marne-la vallée, (UPEM), France.
  - *Getting your Hands Dirty With the Industrial IoT!* by Thomas Watteyne, Inria, France.
  - *Machine learning: From supervised to unsupervised* by Francis Bach, Inria, France.

  Twenty papers have been selected by the technical program committee and presented during the three days of the PEMWN 2017 conference.
- **Samia Bouzefrane** was general co-chair with Mehammed Daoui (University of Tizi-Ouzou, Algeria) and Damien Sauveron (Univesity of Limoges, France) of the MSPN 2017 conference, the 3rd IFIP international conference on Mobile and Secure Programming Networks, technically co-sponsored by IFIP WG 11.2 (see http://cedric.cnam.fr/workshops/mspn2017/). This conference was held at CNAM in Paris, the 29th and 30th of June 2017. It was sponsored by CNAM, University of Limoges and SAFRAN. Two tutorials were given:
  - *Enabling emergent mobile systems in the IoT: Functional and QoS interoperability aspects at the middleware layer* by Nikolaos Georgantas, MIMOVE team (Inria, Paris), France.
  - *Identity Management for Internet connected objects* by Hanene Maupas, SAFRAN, France.

  Seventeen papers have been selected by the technical program committee and presented during the two days of the MSPN 2017 conference

10.1.1.2. Member of the Organizing Committees

- **Paul Muhlethaler** organized the DGA Inria workshop on Telecommunication and networking "Cloud, cloudlet & MANET" in May 2017. The EVA team presented its activities: Remy Leone the results of the project F-Interop, Malisa Vucinic the results of the project ARMOUR, **Pascale Minet** the model and simulation results on collision avoidance on shared slots in TSCH, **Thomas Watteyne** presented the standardization issues in 6TiSCH, **Paul Muhlethaler** presented results on adaptive CSMA techniques.
- **Kevin Bonny** was member of the organizing committee of the international conference PEMWN 2017.

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 2017.
- **Samia Bouzefrane** was track chair of ANT, The 8th International Conference on Ambient Systems, Networks and Technologies, May 2017.
- **Anis Louiti** was in Steering committee member of MobileHealth Workshop 2017.

10.1.2.2. Member of the Conference Program Committees

- **Paul Muhlethaler**:
  - ITST 2017,
  - Mownet 2017, International Conference on Selected Topics in Mobile & Wireless Networking, May 2017.

  - PEMWN 2017, 6th International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks, November 2017.
  - Wireless Days, IFIP/IEEE Wireless Days 2017, 29 - 31 March 2017 Porto, Portugal.
  - WiOpt 2017, Paris, France, 15th - 19th May, 2017.
- **Pascale Minet**:
  - CoRes 2017, May 2017.
  - ETFA 2017, 22th IEEE International Conference on Emerging Technologies & Factory Automation, September 2017.
  - IINTEC 2017, the IEEE International Conference on Internet of Things, Embedded Systems and Communications, October 2017.
  - IUCC 2017, the 16th IEEE International Conference on Ubiquitous Computing and Communications, December 2017.
  - MoWNet 2017, International Conference on Selected Topics in Mobile & Wireless Networking, May 2017.
  - PEMWN 2017, 6th International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks, November 2017.
  - PECCS 2017, 7th International conference on Pervasive and Embedded Computing and Communication Systems, July 2017.
  - WINCOM 2017, 3rd International Conference on Wireless Networks and Mobile Communications, October 2015.
  - Wireless Days, IFIP/IEEE Wireless Days 2017, 29 - 31 March 2017 Porto, Portugal.
  - WiSEE 2017, 5th IEEE International Conference on Wireless for Space and Extreme Environments, October 2017.
- **Thomas Watteyne**:
  - IFIP/IEEE International Symposium on Integrated Network Management, workshop on Future Networks for Secure Smart Cities (FNSSC), 2017.
  - IEEE International Conference on Communications (ICC), Selected Areas in Communications (SAC), 2015, 2016, 2017.
- **Samia Bouzefrane**:
  - PEMWN, 6th International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks, November 2017.
  - VECoS, 11th International Conference on Verification and Evaluation of Computer and Communication Systems, August 2017.
  - MobiSecServ, the third International Conference On Mobile And Secure Services, February 2017.
  - EUSPN, The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks, September 2017.
- **Anis Louiti**:
  - AIST 2017,
  - ITST 2017,
  - Mownet 2017,
  - PEMWN 2017,
  - SMARTCOMP 2017
  - Aintec 2017

10.1.2.3. Member of the Editorial Boards

Nadjib Achir was guest editor of the special issue “Planning and Deployment of Wireless Sensor Networks”, of the International Journal of Distributed Sensor Networks.

10.1.2.4. Reviewer (Journals)

- **Paul Muhlethaler**
  - Reviewer Annals of telecommunications
  - Reviewer IEEE Transactions on Wireless Communications
  - Reviewer IEEE Transactions on Vehicular Technology
  - Reviewer IEEE Transactions on Information Theory
  - Reviewer International Journal of Distributed Sensor Networks. Hindawi.
- **Pascale Minet**
  - Annals of Telecommunications,
  - Ad Hoc Networks,
  - Computer Communications,
  - Computer Networks,
- **Thomas Watteyne**
  - ACM Transactions on Sensor Networks, 2017.
- 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)

10.1.2.5. Reviewer (Book proposals)

Samia Bouzefrane was reviewer for the two book proposals in the CRC Press Taylor and Francis.

10.1.3. Invited Talks

- *(tutorial)* **Thomas Watteyne**. Getting your Hands Dirty With the Industrial IoT! International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN), Paris, France, 28 November 2017.
- *(talk)* **Thomas Watteyne**. Construire son Reseau IoT en 10min. Inria-Paris, 12 March 2018.
- *(talk)* **Thomas Watteyne**. Exciting Developments in Industrial IoT, and its Potential Applications to Smart Agriculture. INRA-Inria Workshop, Rennes, France, 19 December 2017
- *(talk)* **Thomas Watteyne**. Low Power Wireless Solutions for Industry 4.0: Products, Standardization, Research and Example Deployments. Industry 4.0 Predictive Analytics and Forecasting: Research and Applications. Siemens Corporate Technology, Munich, Germany, 14-15 September 2017.
- *(tutorial)* **Thomas Watteyne** and Sami Malek. Wireless Bootcamp with NoeMote and SolSystem. UC Berkeley, 22-23 August 2017.
- *(talk)* **Thomas Watteyne**. OpenWSN Fresh: the 6TiSCH Reference Protocol Stack. Journees Scientifiques Inria, Sofia Antipolis, 15 June 2017.
- *(panel)* **Thomas Watteyne**. "Enough with the visions, Industrial IoT is here Today!", on the panel "the path to successful IoT: from need of standards to security threads". IEEE ICC, Paris, France, 22 May 2017.
- *(talk)* **Thomas Watteyne**. Building the Internet of (Important) Things, Universidad Diego Portales, Santiago, Chile, 5 May 2017.
- *(talk)* **Thomas Watteyne**. Researching and Deploying 6TiSCH Networks, Inria-Chile, 5 May 2017.
- *(talk)* **Thomas Watteyne**. Deploying 6TiSCH Networks. RIOT seminar, 13 April 2017.

10.1.4. Leadership within the Scientific Community

Thomas Watteyne co-chairs the IETF 6TiSCH standardization group.

10.1.5. Scientific Expertise

Thomas Watteyne regularly consults with major player in the (Industrial) IoT space.

10.1.6. Research Administration

Thomas Watteyne is 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 : Nesrine Ben Hassine : Machine learning techniques for network resource allocation. University of Paris Saclay. **Pascale Minet** and Dana Marinca.
- PhD : Younes Bouchaala : Handling Safety Messages in Vehicular Ad-hoc NETworks. University of Paris Saclay. **Paul Muhlethaler**.

- PhD (in progress) : 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 évaluation de la dissémination des informations dans la 5G. Eric Renault and **Paul Muhlethaler**.

10.2.3. Juries

- HdR:
  - Riadh Dhaou, "IoT composé de réseaux terrestres et par satellite au service des Smart Cities : mobilité et hétérogénéité", Institut National Polytechnique de Toulouse, November 2017, **Pascale Minet** reviewer.
  - Hanen Hidoudi, "Contributions à l'amélioration des communications dans les réseaux sans fil multi-sauts", University of Toulouse, November 2017, **Pascale Minet** examiner.
- PhD:
  - Alexandre Laubé, "Agrégation de trafic pour réduire la consommation énergétique globale dans les réseaux sans fil multi-sauts", University of Paris Saclay, September 2017, Anis Laouiti examiner.
  - Bilal M. Maaz, "Allocation des ressources radio dans les réseaux sans-fil de la 5G", University of Paris-Saclay, prepared at the University of Versailles-Saint Quentin, March 2017, **Pascale Minet** examiner.
  - Laurent Reynaud, "Stratégie de mobilité optimisée s pour la tolérance aux perturbations dans les r'eseaux sans fil", University of Lyon, March 2017, **Pascale Minet** examiner.
  - Ahmed Amari, "Specification and analysis of AeroRing - a full duplex Ethernet ring network for new generation avionics systems", University of Toulouse - ISAE, September 2017, **Pascale Minet** reviewer.
  - Florian Grandhomme, "Etudes de protocoles de routage dynamique externe de type BGP dans un environnement réseaux tactiques ad hoc mobiles : faisabilité et performances", University of Rennes 1, November 2017, **Pascale Minet** reviewer.
  - 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, December 2017, **Pascale Minet** reviewer.
  - Aravinthan Gopalasingham, " SDN Based Service Oriented Control Approach For Future Radio Access Networks", Télécom Sud-Paris, June 2017, **Paul Muhlethaler** examiner,
  - Boutheina Dab, "Optimization of Routing and Wireless Resource Allocation in Hybrid Data Center Networks", Paris VI, July 2017, **Paul Muhlethaler** examiner.
  - David Alvarez Corrales, "Cooperative Communications in very large cellular Networks", Télécom ParisTech, November 2017, **Paul Muhlethaler** examiner.
  - Younes Bouchaala, "Handling Safety Messages in Vehicular Ad-hoc NETworks", UVSQ, December 2017, **Paul Muhlethaler** examiner.

10.3. Popularization

Outreach is very important for Inria-EVA, and it brings enormous visibility to the research done in the team. We use it to target a audience different that the one targetted through purely scientific publications.

10.3.1. Web presence

Within the Inria-EVA team, we are regular Twitter contributor through a dozen Twitter accounts, promoting the activities of Inria and Inria-Paris.

Inria-EVA also maintains the following Inria websites:

- https://team.inria.fr/eva/
- http://www.openwsn.org/
- http://smartmarina.org/
- http://www.savethepeaches.com/
- http://www.snowhow.io/
- http://www.solsystem.io/
- http://www.headsup.tech/

10.3.2. Tradeshows

- **Thomas Watteyne**, Ziran Zhang, Felipe Moran. IoT Solutions World Congress 2017, Barcelona, presenting SolSystem.
- **Thomas Watteyne**, Ziran Zhang. VIVATech 2017, Paris, Inria booth, presenting SolSystem.

10.3.3. In The News

- *(French)* SolSystem : une solution "sensor-to-cloud" clés en main, InriaInnovation, 10 November 2017 (**Thomas Watteyne**).
- Interview by L'Esprit Sorcier about SmartMarina, Fete de la Science, 8 October 2017 (**Thomas Watteyne**).
- Presenting the Smart Marina project at the VIVA Tech trade-show, Inria@Silicon Valley Newsletter, 19 September 2017 (**Thomas Watteyne**).
- *(French, TV)* Cap d'Agde : Une puce innovante pour gérer le port de plaisance, TV Sud, aired 20 June 2017 (**Thomas Watteyne**).
- *(French)* 1ère mondiale au Cap d'Agde, des capteurs détectent les entrées les sorties de bateaux, Atout Nautic, 20 June 2017 (**Thomas Watteyne**).
- *(French)* Le port du Cap d'Agde connecté, une première mondiale, France Bleu, 20 June 2017 (**Thomas Watteyne**).
- Interview with Thomas Watteyne "The new use of marinas requires rethinking their operation and imagining a new offer of services", Marine & Oceans, June 2017 (**Thomas Watteyne**).
- *(Spanish)* Experto en IoT de Inria visita Inria Chile, Inria Chile, 9 May 2017 (**Thomas Watteyne**).
- *(Spanish)* Proyecto SmartMarina reinventa los puertos utilizando IoT, Inria Chile, 5 May 2017 (**Thomas Watteyne**).
- The EVA team reinvents the Smart Marina, Inria.fr, 26 April 2017 (**Thomas Watteyne**).
- FBF supports SmartMarina project in Cap d'Agde, France. France Berkeley Fund new, 21 April 2017 (**Thomas Watteyne**).
- *(French)* Le Cap d'Agde – le Port en Passe d'Etre Connecte ! Herault Tribune, 19 April 2017 (**Thomas Watteyne**).
- *(French, video)* Objets connectés : des capteurs intelligents pour mesurer l'environnement, YouTube InriaChannel, March 2017 (**Thomas Watteyne**).
- *(French)* Le numérique peut-il sauver l'agriculture? Inriality, February 2017 (**Thomas Watteyne**).

10.3.4. Miscellaneous Activities

The Inria-EVA team is running a permanent demo in the Inria-Paris demo showroom.

We worked together with the A/V team on the following videos:

- SmartMarina [English] https://www.youtube.com/watch?v=LUcLE8D0RbM
- SmartMarina [French] https://www.youtube.com/watch?v=CwgyCmJvyuw
- interview [French] https://www.youtube.com/watch?v=zsbS310YVe0
- SolSystem [French] https://www.youtube.com/watch?v=juQGnGX5OGs
- SaveThePeaches [English] https://www.youtube.com/watch?v=_qGSH8l0Vkk
- SaveThePeaches [French] https://www.youtube.com/watch?v=cZvGw7DyIzI
- HeadsUp! [English] https://www.youtube.com/watch?v=51mIibi-tDs

The Inria-EVA team very regularly hosts visiting companies interested in using the technologu developed.

11. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] F. ADELANTADO, X. VILAJOSANA, P. TUSET, B. MARTINEZ, J. MELIA-SEGUI, T. WATTEYNE.*Understanding the Limits of LoRaWAN*, in "IEEE Communications Magazine", June 2017, https://hal.inria.fr/hal-01444572.

[2] T. CHANG, T. WATTEYNE, X. VILAJOSANA, Q. WANG.*CCR: Cost-Aware Cell Relocation in 6TiSCH Networks*, in "Transactions on Emerging Telecommunications Technologies", October 2017, https://hal.inria.fr/hal-01533174.

[3] P. H. GOMES, T. WATTEYNE, B. KRISHNAMACHARI.*MABO-TSCH: Multi-hop And Blacklist-based Optimized Time Synchronized Channel Hopping*, in "Transactions on emerging telecommunications technologies", October 2017, https://hal.inria.fr/hal-01555429.

[4] K. M. JOSEPH, T. WATTEYNE, B. KERKEZ.*Awa: Using Water Distribution Systems to Transmit Data*, in "Transactions on emerging telecommunications technologies", October 2017, https://hal.inria.fr/hal-01548381.

[5] S. MALEK, F. AVANZI, K. BRUN-LAGUNA, T. MAURER, C. OROZA, P. HARTSOUGH, T. WATTEYNE, S. D. GLASER.*Real-time Alpine Measurement System Using Wireless Sensor Networks*, in "Sensors", December 2017, vol. 17, n^{o} 11, p. 1-30 [*DOI :* 10.3390/s17112583], https://hal.inria.fr/hal-01630303.

[6] C. OROZA, Z. ZHANG, T. WATTEYNE, S. D. GLASER.*A Machine-Learning Based Connectivity Model for Complex Terrain Large-Scale Low-Power Wireless Deployments*, in "IEEE Transactions on Cognitive Communications and Networking", November 2017, https://hal.inria.fr/hal-01571215.

[7] X. VILAJOSANA, B. MARTINEZ, T. WATTEYNE, I. VILAJOSANA.*On the Suitability of 6TiSCH for Wireless Seismic Data Streaming*, in "Wiley Internet Technology Letters", December 2017, https://hal.inria.fr/hal-01651949.

[8] X. VILAJOSANA, K. PISTER, T. WATTEYNE.*Minimal IPv6 over the TSCH Mode of IEEE 802.15.4e (6TiSCH) Configuration*, in "Internet Engineering Task Force RFC series", May 2017, n^{o} RFC8180, https://hal.inria.fr/hal-01531205.

[9] M. VUCINIC, T. WATTEYNE, X. VILAJOSANA.*Broadcasting Strategies in 6TiSCH Networks*, in "Internet Technology Letters", December 2017, https://hal.inria.fr/hal-01630316.

[10] M. VUČINIĆ, M. KRÓL, †. BAPTISTE JONGLEZ, T. COLADON, B. TOURANCHEAU.*Trickle-D: High Fairness and Low Transmission Load with Dynamic Redundancy*, in "IEEE internet of things journal", October 2017, https://hal.archives-ouvertes.fr/hal-01653203.

[11] T. WATTEYNE, P. TUSET, X. VILAJOSANA, S. POLLIN, B. KRISHNAMACHARI.*Teaching Communication Technologies and Standards for the Industrial IoT? Use 6TiSCH!*, in "IEEE Communications Magazine", September 2017, vol. 55, n^{o} 5, p. 132-137 [*DOI :* 10.1109/MCOM.2017.1700013], https://hal.inria.fr/hal-01485425.

International Conferences with Proceedings

[12] K. AVRACHENKOV, P. JACQUET, J. K. SREEDHARAN.*Hamiltonian System Approach to Distributed Spectral Decomposition in Networks*, in "nDS 2017 - 10th International Workshop on Multidimensional (nD) Systems", Zielona Gora, Poland, September 2017, https://hal.inria.fr/hal-01646881.

[13] Y. BOUCHAALA, P. MUHLETHALER, N. ACHIR.*Analysis of the IEEE 802.11 EDCF scheme for broadcast traffic: Application for VANETs*, in "2017 Wireless Days", Porto, Portugal, IEEE, March 2017, p. 252 - 257 [*DOI :* 10.1109/WD.2017.7918156], https://hal.archives-ouvertes.fr/hal-01617895.

[14] Y. BOUCHAALA, P. MUHLETHALER, O. SHAGDAR, N. ACHIR.*Optimized Spatial CSMA for VANETs: A Comparative Study using a Simple Stochastic Model and Simulation Results*, in "CCNC 2017. 8-11 january 2017. Las Vegas", Las Vegas, United States, Proceedings of CCNC 2017, January 2017, https://hal.archives-ouvertes.fr/hal-01379978.

[15] N. BOUFARES, P. MINET, I. KHOUFI †, L. SAIDANE.*Covering a 3D flat surface with autonomous and mobile wireless sensor nodes*, in "IWCMC 2017 - the 13th International Wireless Communications and Mobile Computing Conference", Valencia, Spain, June 2017, https://hal.archives-ouvertes.fr/hal-01640508.

[16] M. HADDED, P. MUHLETHALER, A. LAOUITI, L. AZOUZ SAIDANE.*TDMA-aware Routing Protocol for Multi-hop Communications in Vehicular Ad Hoc Networks*, in "WCNC 2017 - IEEE Wireless Communications and Networking Conference", San Francisco , United States, March 2017, https://hal.archives-ouvertes.fr/hal-01441264.

[17] H. JIANG, Z. BRODARD, T. CHANG, A. BOUABDALLAH, N. MONTAVONT, G. TEXIER, P. THUBERT, T. WATTEYNE, G. PAPADOPOULOS.*Competition: Controlled Replication for Higher Reliability and Predictability in Industrial IoT Networks*, in "International Conference on Embedded Wireless Systems and Networks (EWSN)", Uppsala, Sweden, February 2017, https://hal.inria.fr/hal-01664764.

[18] H. JIANG, Z. BRODARD, T. CHANG, A. BOUABDALLAH, N. MONTAVONT, G. TEXIER, P. THUBERT, T. WATTEYNE, G. PAPADOPOULOS.*Dependability Competition: Controlled Replication for Higher Reliability and Predictability in Industrial IoT Networks*, in "EWSN 2017 : International Conference on Embedded Wireless Systems and Networks", Uppsala, Sweden, ACM, February 2017, p. 282 - 283, https://hal.archives-ouvertes.fr/hal-01638297.

[19] P. KEELER, B. BŁASZCZYSZYN, P. MÜHLETHALER.*Optimizing spatial throughput in device-to-device networks*, in "WIOPT/SPASWIN 2017 - Workshop on Spatial Stochastic Models

for Wireless Networks", Paris, France, IEEE, May 2017, 5, https://arxiv.org/abs/1612.09198 - 6 pages, 4 figures. Submitted, https://hal.inria.fr/hal-01505044.

[20] I. KHOUFI, P. MINET, B. RMILI.*Scheduling transmissions with latency constraints in an IEEE 802.15.4e TSCH network*, in "VTC 2017 - IEEE 86th Vehicular Technology Conference", Toronto, Canada, September 2017, https://hal.archives-ouvertes.fr/hal-01636656.

[21] P. MINET, I. KHOUFI, A. LAOUITI.*Increasing Reliability of a TSCH Network for the Industry 4.0*, in "16th IEEE International Symposium on Network Computing and Applications (NCA 2017)", Boston, United States, November 2017, https://hal.archives-ouvertes.fr/hal-01637085.

[22] P. MINET, P. MUHLETHALER, I. KHOUFI.*Collision Avoidance on Shared Slots in a Wireless Slotted Network: Models and Simulations*, in "PEMWN 2017 - 6th IFIP International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks", Paris, France, November 2017, https://hal.archives-ouvertes.fr/hal-01636646.

[23] P. MUHLETHALER, Y. BOUCHAALA, S. †. OYUNCHIMEG, N. ACHIR.*Evaluating the Gain of Directional Antennas in Linear VANETs using Stochastic Geometry*, in "PEMWN 2017 - 6th IFIP International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks", Paris , France, November 2017, p. 1-7, https://hal.archives-ouvertes.fr/hal-01617937.

[24] G. PAPADOPOULOS, T. MATSUI, P. THUBERT, G. TEXIER, T. WATTEYNE, N. MONTAVONT.*Leapfrog Collaboration: Toward Deterministic and Predictable in Industrial-IoT Applications*, in "IEEE International Conference on Communications (ICC)", Paris, France, May 2017, https://hal.inria.fr/hal-01451339.

[25] C. SCHINDLER, T. WATTEYNE, X. VILAJOSANA, K. PISTER.*Implementation and Characterization of a Multi-hop 6TiSCH Network for Experimental Feedback Control of an Inverted Pendulum*, in "WiOpt 2017 - 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks", Paris, France, IEEE (editor), IEEE, May 2017 [*DOI :* 10.23919/WIOPT.2017.7959925], https://hal.inria.fr/hal-01485523.

Conferences without Proceedings

[26] M. HADDED, P. MUHLETHALER, A. LAOUITI.*Performance evaluation of a TDMA-based multi-hop communication scheme for reliable delivery of warning messages in vehicular networks*, in "IWCMC 2017 - 13th International Wireless Communications and Mobile Computing Conference", Valencia , Spain, June 2017, p. 1029 - 1034 [*DOI :* 10.1109/IWCMC.2017.7986427], https://hal.archives-ouvertes.fr/hal-01617924.

[27] M. T. HAMMI, E. LIVOLANT, P. BELLOT, A. SERHROUCHNI, P. MINET.*A Lightweight IoT Security Protocol*, in "1st Cyber Security in Networking Conference (CSNet2017)", Rio de Janeiro, Brazil, October 2017, https://hal.archives-ouvertes.fr/hal-01640510.

[28] M. T. HAMMI, E. LIVOLANT, P. BELLOT, A. SERHROUCHNI, P. MINET.*A Lightweight Mutual Authentication Protocol for the IoT*, in "ICMWT 2017 - iCatse International Conference on Mobile and Wireless Technology", Kuala Lumpur, Thailand, June 2017 [*DOI :* 10.1007/978-981-10-5281-1], https://hal.archives-ouvertes.fr/hal-01640511.

[29] N. B. HASSINE, R. MILOCCO, P. MINET.*ARMA based Popularity Prediction for Caching in Content Delivery Networks*, in "IFIP Wireless Days 2017", Porto, Portugal, March 2017, https://hal.archives-ouvertes.fr/hal-01636975.

[30] N. B. HASSINE, P. MINET, M.-A. KOULALI, M. ERRADI, D. MARINCA, D. BARTH.*Coalition Game for Video Content Clustering in Content Delivery Networks*, in "the 14th Annual IEEE Consumer Communications and Networking Conference, CCNC 2017", Las Vegas, United States, January 2017, https://hal.archives-ouvertes.fr/hal-01636959.

[31] I. KHOUFI, P. MINET, B. RMILI.*Beacon Advertising in an IEEE 802.15.4e TSCH Network for Space Launch Vehicles*, in "EUCASS 2017 - 7th European Conference for Aeronautics and Aerospace Sciences", Milano, Italy, July 2017, https://hal.archives-ouvertes.fr/hal-01636659.

Books or Proceedings Editing

[32] N. MITTON, H. CHAOUCHI, T. NOEL, T. WATTEYNE, A. GABILLON, P. CAPOLSINI (editors). *Interoperability, safety and security in IoT : second international conference, InterIoT 2016 and third international conference, SaSeIoT 2016*, LNICST, Springer, 2017, vol. 190, 139, https://hal.archives-ouvertes.fr/hal-01647101.

Other Publications

[33] T. CHANG, T. WATTEYNE, Q. WANG, X. VILAJOSANA.*Demo: Scheduling Function Zero on a 6TiSCH Network*, February 2017, International Conference on Embedded Wireless Systems and Networks (EWSN), Poster, https://hal.inria.fr/hal-01419913.

[34] P. JACQUET, D. POPESCU, B. MANS.*Information Dissemination in Vehicular Networks in an Urban Hyperfractal Topology*, December 2017, https://arxiv.org/abs/1712.04054 - KEYWORDS: DTN; Wireless Networks; Broadcast; Fractal; Vehicular Networks; Urban networks, https://hal.inria.fr/hal-01662286.

[35] T. MATSUI, G. PAPADOPOULOS, P. THUBERT, T. WATTEYNE, N. MONTAVONT.*Poster: 4th Industrial Revolution: Toward Deterministic Wireless Industrial Networks*, February 2017, International Conference on Embedded Wireless Systems and Networks (EWSN), Poster, https://hal.inria.fr/hal-01419907.

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
A2.1.2. - Object-oriented programming
A2.1.3. - 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
A2.2.3. - Run-time systems
A2.2.4. - Parallel architectures
A2.4. - 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
B9.4.1. - Computer science

1. Personnel

Research Scientists

Xavier Leroy [Team leader, Inria, Senior Researcher]
Umut Acar [Carnegie Mellon University & Inria, Advanced Research Position]
Damien Doligez [Inria, Researcher]
Fabrice Le Fessant [Inria, Researcher, until Sep 2017]
Jean-Marie Madiot [Inria, Researcher]
Luc Maranget [Inria, Researcher]
Michel Mauny [Inria, Senior Researcher]
François Pottier [Inria, Senior Researcher, HDR]

Michael Rainey [Inria, Starting Research Position]
Didier Rémy [Inria, Senior Researcher, HDR]

Technical Staff

Sébastien Hinderer [Inria, Research Engineer, 80%]
Thomas Blanc [Inria, from Apr 2017 until Sep 2017]

PhD Students

Vitalii Aksenov [Inria]
Armaël Guéneau [Université Paris Diderot]
Naomi Testard [Inria]
Thomas Williams [ENS Paris]

Post-Doctoral Fellows

Gergö Barany [Inria, from Mar 2017]
Adrien Guatto [Inria, from Sep 2017]

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 [45]. 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 [47] 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 [43], [41], 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 [50], 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 [48]. 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 languages 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 [46] and enforcement of data confidentiality through type-based inference of information flow and noninterference properties [49].

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 Caml language family has an impact on the teaching of programming. Caml Light 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 2017, Jacques-Henri Jourdan received the "prix du GDR GPL" (http://gdr-gpl.cnrs.fr/node/284) for his dissertation, entitled "Verasco: a Formally Verified C Static Analyzer". Jacques-Henri was a Ph.D. student in the Gallium team, advised by Xavier Leroy.

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: In 2017, we have continued to work on a complete reimplementation of the proof manager. One objective is a cleaner interaction with the TLA$^+$ front-ends, in particular SANY, the standard parser and semantic analyzer. The reimplementation is also necessary for extending the scope of the fragment of TLA$^+$ that is handled by TLAPS, in particular full temporal logic and module instantiation.

- Participants: Damien Doligez, Stephan Merz and Martin Riener
- 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], Prashanth Mundkur [SRI International].

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:

- The support for 64-bit target processors that was initiated last year was improved and released as part of version 3.0 of CompCert. CompCert has been supporting 64-bit integer arithmetic since 2013. However, pointers and memory addresses were still assumed to be 32 bits wide. CompCert 3.0 lifts this restriction by parameterizing the compiler over the bit width of memory addresses. This required extensive changes throughout the back-end compiler passes and their correctness proofs.
- The x86 code generator, initially 32-bit only, was extended to handle 64-bit x86 as well. This is the first instantiation of the generic support for 64-bit target architectures mentioned above. This extension greatly improves the usability and performance of CompCert on servers and PCs, where x86 64-bit is the dominant architecture.
- Support for the RISC-V processor architecture was added to CompCert. Prashanth Mundkur contributed a prototype port targeting 32-bit RISC-V. Xavier Leroy extended this port to target 64-bit RISC-V as well and to integrate it in CompCert 3.1. While not commercially available yet, the RISC-V architecture is used in many academic verification projects.
- Several minor optimizations were added to address inefficiencies observed in AbsInt's customer code. The most notable one is the optimization of leaf functions to avoid return address reloads.
- Error and warning messages were improved and made more like those of GCC and Clang. Command-line flags were added to control which warning to emit and which warnings to treat as fatal errors.

We released version 3.0 of CompCert in February 2017 incorporating support for 64-bit architectures, and version 3.1 in August 2017 incorporating the other enhancements listed above.

Two papers describing industrial uses of CompCert for critical software were written, with Daniel Kästner from AbsInt as lead author. The first paper [24] was presented at the 2017 symposium of the British Safety-Critical Systems Club. The second paper [23] will be presented in January 2018 at the ERTS congress. 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.

7.1.2. A verified model of register aliasing in CompCert

Participants: Gergö Barany, Xavier Leroy.

In the setting of the ASSUME ITEA3 project, Gergö Barany and Xavier Leroy are working on implementing a CompCert back-end for the Kalray MPPA processor architecture. This architecture features pervasive register aliasing: each of its 64-bit registers can also be accessed as two separate 32-bit halves. The ARM architecture's floating-point register file is similarly aliased. Modifying a superregister invalidates the data stored in subregisters and vice versa; this behavior was not yet modeled in CompCert's semantics.

Integrating subregister aliasing in CompCert involved re-engineering much of its 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 is nearing completion.

7.1.3. Random program generation for compiler testing

Participant: Gergö Barany.

Randomized testing is a powerful tool for finding bugs in compilers. In a project aimed at finding missed compiler optimizations, Gergö Barany wanted to use such random testing techniques, but found that the standard random C program generator, Csmith, generates very large amounts of dead code. This is code whose results are never used and that can therefore be removed by the compiler.

The presence of large amounts of dead code prevents testing optimizations: almost all of the code is trivially removed by compilers' dead code elimination passes. Gergö resolved this problem by designing a new approach to random program generation. The new generator generates code backwards and performs a simultaneous liveness analysis of the program to rule out the generation of dead code. Its practical evaluation shows that it is much more efficient than Csmith at generating programs that compile to large amounts of machine code with a much more varied instruction mix than Csmith-generated code. In addition, the new generator is much faster than Csmith, because it is designed to work in a single, linear pass, without generating invalid states that cause backtracking. This work resulted in the development of the `ldrgen` tool, and was presented at LOPSTR 2017 [34].

7.1.4. 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 started work on adapting compiler correctness testing techniques for 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. The comparison is done by a custom binary analysis tool that Gergö developed for this purpose. This tool assigns scores to programs according to various 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. New criteria can be added using a simple plug-in system. If the binaries generated by different compilers are assigned different scores, the input program is considered interesting, and it is reduced to a minimal test case using an off-the-shelf program reducer (C-Reduce).

This automated process often results in small, simple examples of missed optimizations: optimizations that compilers should be able to perform, but that they failed to apply for various reasons. Gergö found previously unreported missing arithmetic optimizations, as well as individual cases of unnecessary register spilling, missed opportunities for register coalescing, dead stores, redundant computations, and missing instruction selection patterns. Several of these missed optimization issues were reported and fixed in the GCC, Clang, and CompCert compilers. An article describing this work is currently under review, and work is in progress on other binary analysis techniques that can find further missed optimizations.

7.1.5. Towards a verified compilation stack for concurrent programs

Participants: Jean-Marie Madiot, Andrew Appel [Princeton University].

The verified compiler CompCert compiles programs from C to assembly while preserving their semantics, thus allowing formal reasoning on source programs, which is much more tractable than reasoning on assembly code. It is however limited to sequential programs, running as one thread on one processor. Jean-Marie Madiot is working to extend CompCert to shared-memory concurrency *and* to provide users with techniques to reason and prove properties about concurrent programs.

Concurrent Separation Logic is used to reason about source programs and prove their correctness with respect to a "concurrent permission machine". The programs are compiled by a concurrency-aware version of CompCert. As of 2017, this has been done for the x86 architecture only.

This project is a continuation of a collaboration with Andrew Appel's team at Princeton University. Appel's team has been working for several years on the "Verified Software Toolchain" project, which provides users with tools to establish properties of sequential programs. Jean-Marie Madiot has been extending the program logic to shared-memory concurrency and developing a new proof of concurrent separation logic that is both formalised and usable in this setting. A paper has been submitted and rejected and is being improved.

Jean-Marie Madiot is now also working on a more general adaptation of CompCert to the reasoning principles of concurrency, and started a collaboration to adapt it to architectures other than x86 (see Section 7.3.4).

7.1.6. Verified compilation of Lustre

Participants: Xavier Leroy, Timothy Bourke [team Parkas], Lélio Brun [team Parkas], Pierre-Évariste Dagand [team Whisper], Marc Pouzet [team Parkas], Lionel Rieg [Yale University].

The Velus project of team Parkas develops a compiler for the Lustre reactive language that generates CompCert Clight intermediate code and is proved correct using the Coq proof assistant. A paper describing the Velus compiler and its verification was presented at the conference PLDI 2017 [20]. Xavier Leroy contributed to the verification of the final pass of Velus, the one that translates from the Obc object-oriented intermediate language of Velus to the Clight C-like, early intermediate language of CompCert. The correctness proof of this pass captures the shape of memory states during execution using formulas from separation logic. The separation logic assertions for CompCert memory states used in this proof come from a library that Xavier Leroy developed last year to help revise the proof of the "stacking" pass of CompCert, and that Timothy Bourke and Xavier Leroy later extended with a "magic wand" operator.

7.2. Language design and type systems

7.2.1. Refactoring with ornaments in ML

Participants: Thomas Williams, Didier Rémy.

Thomas Williams and Didier Rémy continued working on ornaments for program refactoring and program transformation in ML. Ornaments have been introduced as a way of describing changes in data type definitions that preserve the recursive structure but can reorganize, add, or drop 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.

This year, Williams and Rémy continued working on the description of the lifting algorithm: 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 into 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. A paper has been accepted for presentation at POPL 2018 [14]. A research report gives more technical details [30].

On the practical side, several families of case studies have been explored, including refactoring and code specialization, as so as to make certain existing invariants apparent, or so as to use more efficient data structures. We improved the user interface of the prototype implementation so as to make it easier to write useful examples. We are currently developing a new version of the prototype that will handle most of the OCaml language.

7.3. Shared-memory parallelism

7.3.1. The Linux Kernel Memory Model

Participants: Luc Maranget, Jade Alglave [University College London–Microsoft Research, UK], 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 optimizations 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 progress as regards languages. To wit, a two-year effort to define a weak memory model for the Linux Kernel has yielded an article in the *Linux Weekly News* online technical magazine [31], and a scholarly paper accepted for publication at the *Architectural Support for Programming Languages and Operating Systems* (ASPLOS) conference in March 2018. While targeting different audiences, both articles describe a formal model that defines how Linux programs are supposed to behave. The model is of course a CAT model, hence is understood by the **herd** simulator (Section 7.3.3) that allows programmers to experiment and develop an intuition. The model has been tested against hardware and refined in consultation with maintainers. Finally, the ASPLOS article formalizes the *fundamental law of the Read-Copy-Update synchronization mechanism*, and proves that one of its implementations satisfies this law.

For the record, Luc Maranget also co-authored a paper that has been presented at POPL 2017 [22]. This work, which we described last year, is joint work with many researchers, including S. Flur and other members of P. Sewell's team (University of Cambridge) as well as M. Batty (University of Kent). Moreover, Luc Maranget still interacts with the Cambridge team, mostly by providing tests and performing comparisons between his axiomatic models and the operational models developed by this team.

7.3.2. ARMv8 and RISC-V memory models

Participants: Will Deacon [ARM Ltd], Luc Maranget, Jade Alglave [University College London–Microsoft Research, UK].

Jade Alglave and Luc Maranget helped Will Deacon, an engineer at ARM Ltd., who developed a model for the ARMv8 64-bit processor. Will wrote a CAT model, which ARM uses internally as a specification. (CAT is the domain-specific language for describing memory models and is understood by the **herd** simulator; see Section 7.3.3.) ARM's official documentation presents a natural language transliteration of the CAT model.

Luc Maranget also joined the RISC-V consortium (https://riscv.org/) as an individual and as a member of the memory model group. He takes part in the development of the memory model of this open architecture, mostly by writing CAT models and reviewing tests that will be part of the documentation. A CAT model will be part of the next version (V2.3) of the User-Level ISA Specification.

7.3.3. Improvements to the diy tool suite

Participants: Luc Maranget [**contact**], Jade Alglave [University College London–Microsoft Research, UK].

The **diy** suite (for "Do It Yourself") 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, ARM, generic (LISA) assembler, or a subset of the C language that can thus be generated from concise specifications, run on hardware, or simulated on top of memory models. Test results can be handled and compared using additional tools.

This year's new features are a model for the Linux Kernel developed as a collaborative effort (see Section 7.3.1) and an ongoing RISC-V model transliterated by Luc Maranget from the model elaborated by the RISC-V committee which Luc Maranget joined this year (see Section 7.3.2). Those new models were made possible due to significant extensions of **diy**, such as a new tool chain for RISC-V and the extension of the macro system so as to handle most of the memory-model-related macros used by Linux kernel developers.

7.3.4. Towards formal software verification with respect to weak memory models

Participants: Jean-Marie Madiot, Jade Alglave [University College London & Microsoft Research Cambridge], Simon Castellan [Imperial College London].

Past research efforts on weak memory models have provided both academia and industry with very efficient tools to precisely describe memory models and to carefully test them on a wide variety of architectures. While these models give us a good understanding of complex *hardware* behaviors, exploiting them to formally guarantee the good behavior of *software* remains practically out of reach.

A difficulty is that weak memory models are described in terms of properties of graphs of execution candidates. Because graphs are far from the usual way of defining programming language semantics, because execution candidates are not defined formally, and because existing proofs of "data-race freedom" (DRF) theorems are hard to fathom and formally imprecise, there is a strong demand in the programming language community for a formal account of weak memory models.

In 2017, Jean-Marie Madiot started a collaboration with weak memory model expert Jade Alglave and concurrent game semantics researcher Simon Castellan to tackle these problems. The idea is to have a formal description, using partial-order techniques similar to the ones used in game semantics, of execution candidates. On the other side, a given model of shared memory is then described in terms of partial orders, and the composition of those partial orders provides the final possible executions of a given program in a given architecture. This should yield a formal semantics for programs in a weak memory setting, and should allow proving a DRF theorem so as to connect this semantics to more standard sequentially consistent semantics. A success in this direction would finally allow tractable verification of concurrent programs, particularly in combination with Madiot's ongoing work on a generalization to concurrency of the CompCert certified compiler (see Section 7.1.5).

7.3.5. Granularity control for parallel programs

Participants: Umut Acar, Vitaly Aksenov, Arthur Charguéraud, Adrien Guatto, Mike Rainey, Filip Sieczkowski.

The DeepSea team focused this year on the development of techniques for controlling granularity in parallel programs. Granularity control is an essential problem because creating too many tasks may induce overwhelming overheads, while creating too few tasks may harm the ability to process tasks in parallel. Granularity control turns out to be especially challenging for nested parallel programs, i.e., programs in which parallel constructs such as fork-join or parallel-loops can be arbitrarily nested. Two different approaches were investigated.

The first approach is based on the use of asymptotic complexity functions provided by the programmer, combined with runtime measurements to estimate the constant factors that apply. Combining these two sources of information allows to predict with reasonable accuracy the execution time of tasks. Such predictions may be used to guide the generation of tasks, by sequentializing computations of sufficiently-small size. An analysis is developed, establishing that task creation overheads are indeed bounded to a small fraction of the total runtime. These results builds upon prior work by the same authors [39], extending it with a carefully-designed algorithm for ensuring convergence of the estimation of the constant factors deduced from the measures, even in the face of noise and cache effects, which are taken into account in the analysis. The approach is demonstrated on a range of benchmarks taken from the state-of-the-art PBBS benchmark suite. A paper describing the results is under preparation.

The second approach is based on an instrumentation of the runtime system. The idea is to process parallel function calls just like normal function calls, by pushing a frame on the stack, and only subsequently promoting these frames as threads that might get scheduled on other cores. The promotion of frames takes place at regular time intervals, which is why we named this approach *heartbeat scheduling*. Unlike prior approaches such as *lazy scheduling*, in which promotion is guided by the workload of the system, heartbeat scheduling can be proved to induce only small scheduling overheads, and to not asymptotically reduce the amount of parallelism inherent in the program. The theory behind the approach is formalized in Coq. It is also implemented through instrumented C++ programs, and evaluated on PBBS benchmarks. A paper describing this approach was submitted to an international conference.

7.3.6. Non-zero indicators: a provably-efficient, concurrent data structure

Participants: Umut Acar, Mike Rainey.

This work, conducted in collaboration with Naama Ben David from Carnegie Mellon University, investigates the design and analysis of an implementation of a concurrent data structure called *non-zero indicator*. This data structure plays a crucial role in the scheduling of nested parallel programs: it is used to handle dependency resolution among parallel tasks. Concretely, a non-zero indicator is initialized with value 1, and it supports the following two concurrent operations, which may be invoked by threads that have knowledge that the counter is non-zero: (1) atomically increase the counter by one unit, and (2) atomically decrease the counter by one unit, and detect whether the counter reaches zero. While a trivial implementation can be set up using an atomic operation on a shared memory cell (e.g., fetch-and-add), the key challenge is to design a non-zero indicator that scales well to hundreds if not thousands of threads, without suffering from contention.

Prior work leverages dynamic tree data structures to tame contention [42]. Yet, such prior work, as well as most concurrent data structures in general, are analyzed empirically, omitting asymptotic bounds on their efficiency. In this work, we propose a new variant of a tree-based non-zero indicator implementation, for which we are able to present a formal analysis establishing bounds on the worst-case contention of concurrent updates. Our analysis is the first to achieve relevant bounds of this kind. Furthermore, we demonstrate in practice that our proposal improves scalability, compared with a naive fetch-and-add atomic counter, and also compared with the original tree-based data structure. Our work was presented at PPoPP [16].

7.3.7. Efficient sequence data structures for ML

Participants: Arthur Charguéraud, Mike Rainey.

The use of sequence containers, including stacks, queues, and double-ended queues, is ubiquitous in programming. When the maximal number of elements to be stored is not known in advance, containers need to grow dynamically. For this purpose, most ML programs rely on either lists or vectors. These data structures are inefficient, both in terms of time and space usage. In this work, we investigate the use of data structures based on *chunks*, adapting ideas from some of our prior work implemented in C++ [38]. Each chunk stores items in a fixed-capacity array. All chunks are linked together to represent the full sequence. These chunk-based structures save a lot of memory and generally deliver better performance than classic container data structures for long sequences. We measured a 2x speedup compared with vectors, and up to a 3x speedup compared with long lists. This work was presented at the ML Family Workshop [36]. Generalization of this work to double-ended sequences and to persistent sequences is under progress.

7.3.8. A parallel algorithm for the dynamic trees problem

Participants: Umut Acar, Vitaly Aksenov.

Dynamic algorithms are used to compute a property of some data while the data undergoes changes over time. Many dynamic algorithms have been proposed, but nearly all of them are sequential.

In collaboration with Sam Westrick (Carnegie Mellon University), Umut Acar and Vitaly Aksenov investigated the design of an efficient parallel dynamic tree data structure. This data structure supports four operations, namely insertion and deletion of vertices and edges; these operations can be executed in parallel. The proposed data structure is work-efficient and highly parallel. A preliminary version of this work was presented in a brief announcement at SPAA 2017 [15].

7.3.9. A concurrency-optimal binary search tree

Participant: Vitaly Aksenov.

In joint work with Vincent Gramoli (IT School of Information Technologies, Sydney), Petr Kuznetsov (Telecom ParisTech), Anna Malova (Washington University in St Louis), and Srivatsan Ravi (Purdue University), Vitaly Aksenov proposed a concurrency-optimal implementation of binary search trees. Concurrency-optimality means that the data structure allows all interleavings of the underlying sequential implementation, except those that would violate linearizability. Aksenov and co-authors show that none of the state-of-the-art concurrent binary search trees are concurrency-optimal, and they experimentally verify that the new concurrency-optimal binary search tree is competitive with known implementations. This work was presented at Euro-Par 2017 [17].

7.4. The OCaml language and system

7.4.1. The OCaml system

Participants: Damien Doligez, Xavier Leroy, Luc Maranget, David Allsop [Cambridge University], Florian Angeletti, Alain Frisch [Lexifi], Jacques Garrigue [University of Nagoya], Sébastien Hinderer [SED], Nicolás Ojeda Bär [Lexifi], Thomas Refis [Jane Street], Gabriel Scherer [team Parsifal], Mark Shinwell [Jane Street], Leo White [Jane Street], Jeremy Yallop [Cambridge University].

This year, we released four versions of the OCaml system: versions 4.04.1 and 4.04.2 are minor releases that fix about 16 issues; versions 4.05.0 and 4.06.0 are major releases that introduce some new features, many improvements in usability and performance, and fix about 100 issues. The most important new features are:

- Character strings are now immutable (read-only) by default. This completes the evolution of OCaml towards immutable strings that started in 2014 with the introduction of a compile-time option to separate text-like read-only strings from array-like read-write byte sequences. This option is now the default, making OCaml programs safer and clearer.
- Extensions of the "destructive substitution" operator over module signatures (*sig* `with type t :=` ...) to make it more general and more widely usable.
- Support for the UTF8 encoding of Unicode characters in strings was improved with the introduction of an escape `\u{XXXX}` in string literals, and more importantly with a complete overhaul of the OCaml interface for Windows system calls that make them compatible with UTF8-encoded Unicode.
- An alternate register allocator based on linear scan was added and can be selected to reduce compilation times.

On the organization side, we switched to a deadline-based release cycle whereby a major release occurs at a set date with the features that are ready by that date, instead of waiting for a set of new features to be ready. Releases 4.05.0 and 4.06.0 were produced in this manner at 6-months intervals. Damien Doligez and Gabriel Scherer served as release managers.

Sébastien Hinderer worked on integrating `ocamltest`, the compiler's test driver he developed last year, in the 4.06 release of OCaml. He migrated a large part of the test suite from the former Makefile-based infrastructure to `ocamltest`. He also started to rewrite OCaml's build system so that the compiler can be built in parallel as much as its dependencies allow.

We have improved our Continuous Integration infrastructure by taking advantage of Jenkins features such as configuration matrices, adding five new architectures (ARM-64, Fedora, FreeBSD, PPC64-LE, Ubuntu), and upgrading to the latest version of MacOS. Our testing is now done on all of the major architectures that are officially supported by OCaml.

7.4.2. Type-checking the OCaml intermediate languages

Participants: Pierrick Couderc [ENSTA-ParisTech & OCamlPro], Grégoire Henry [OCamlPro], Fabrice Le Fessant, Michel Mauny.

This work aims at designing and implementing a consistency checker for the type-annotated abstract syntax trees (TASTs) produced by the OCaml compiler. When presented as inference rules, the different cases of this TAST checker can be read as the rules of the OCaml type system. Proving the correctness of (part of) the checker would prove the soundness of the corresponding part of the OCaml type system. A preliminary report on this work has been presented at the 17th Symposium on Trends in Functional Programming (TFP 2016).

In 2017, Pierrick Couderc formalized the consistency checker, and wrote a Coq proof of its correctness. The dissertation is being written, and Pierrick's Ph.D. defense should take place at the beginning of 2018.

7.4.3. Optimizing OCaml for satisfiability problems

Participants: Sylvain Conchon [LRI, Univ. Paris Sud], Albin Coquereau [ENSTA-ParisTech], Mohamed Iguernelala [OCamlPro], Fabrice Le Fessant, Michel Mauny.

This work aims at improving the performance of the Alt-Ergo SMT solver, implemented in OCaml. For safety reasons and to ease reasoning about its algorithms, the implementation of Alt-Ergo uses as much as possible 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.

We therefore explored the replacement of Alt-Ergo's default, functional, SAT solver by an imperative CDCL solver. In a first step, we reimplemented a C++ version of miniSAT in OCaml. A comparison of their respective performance showed that the OCaml version is slower and has more cache misses.

In a second step, we studied the use of the imperative miniSAT-like SAT solver in Alt-Ergo. The integration is actually not immediate because of the interaction between this solver and both the theories and the quantifier instantiation engines of Alt-Ergo. In fact, although the default (functional) SAT solver of Alt-Ergo is not as effective as a CDCL solver for reasoning on pure Boolean problems, its smart interaction with theories and instantation engines makes it quite effective in the context of program verification.

7.4.4. Type compatibility checking for dynamically-loaded OCaml data

Participants: Florent Balestrieri [ENSTA-ParisTech], Michel Mauny.

The SecureOCaml project (FUI 18) aims at enhancing the OCaml language and environment in order to make it more suitable for building secure applications, following the recommendations published by the French ANSSI in 2013. Florent Balestrieri (ENSTA-ParisTech) represents ENSTA-Paristech in this project for 2016 and 2017.

The first year has been dedicated to designing and producing an effective OCaml implementation that checks whether a memory graph – typically the result obtained by unmarshalling some data – is compatible with a given OCaml type, following the algorithm designed by Henry *et al.* in 2012. Because the algorithm requires a runtime representation of OCaml types, Florent Balestrieri implemented a library for generic programming in OCaml. This library was presented at the OCaml Users and Developers Workshop in 2016 [40]; an extended version of this paper has been submitted [33]. He also implemented a type-checker which, when given a type and a memory graph, checks whether the former could be the type of the latter. In 2017, Florent Balestrieri implemented a prototype type-checker for OCaml bytecode.

7.4.5. Visitors

Participant: François Pottier.

Traversing and transforming abstract syntax trees that involve name binding is notoriously difficult to do in a correct, concise, modular, customizable manner. In 2017, François Pottier addressed this problem in the setting of OCaml by proposing visitor classes as partial, composable descriptions of the operations that one wishes to perform on abstract syntax trees. By combining auto-generated visitor classes (which have no knowledge of binding) and hand-written visitor classes (each of which knows about a specific binding construct, a specific representation of names, and/or a specific operation on abstract syntax trees), a wide range of operations can be defined. A syntax extension for OCaml has been released under the name `visitors` and this work has been presented at the conference ICFP 2017 [13].

7.4.6. Improvements in Menhir

Participant: François Pottier.

In 2017, François Pottier incorporated several improvements, proposed by Frédéric Bour, to the Menhir parser generator. Many functions were added to Menhir's incremental API, which (at runtime) allows inspecting and updating the parser's state from the outside. A new library, MenhirSdk, was introduced, which (at compile-time) allows inspecting the grammar and the automaton constructed by Menhir. Together, these improvements allow new features to be programmed outside of Menhir; the advanced error recovery mode implemented in the Merlin IDE is an example.

François Pottier also improved the termination test that takes place before parameterized symbols are expanded away. The new test, it is hoped, should reject the grammar if and only if expansion would not terminate. This improves the expressive power of the grammar description language.

7.5. Software specification and verification

7.5.1. Formal reasoning about asymptotic complexity

Participants: Armaël Guéneau, Arthur Charguéraud, François Pottier.

For several years, Arthur Charguéraud and 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. An extended version of their work on the UnionFind algorithm has appeared in the *Journal of Automated Reasoning* [11]. In this work, the complexity bounds that are established involve explicit constants: for instance, the complexity of *find* is $2\alpha(n) + 4$.

Armaël Guéneau, who is supervised by Arthur Charguéraud and François Pottier, is working on relaxing this approach so as to use asymptotic bounds: e.g., the advertised complexity of *find* should be $O(\alpha(n))$. The challenge is to give a formal account of the O notation and of its properties and to develop techniques that make asymptotic reasoning as convenient in Coq as it seemingly is on paper.

For that purpose, this year, Armaël Guéneau developed two Coq libraries. A first library gives a formal definition of the O notation, provides proofs for many commonly used lemmas, as well as a number of tactics that automate the application of these lemmas. A second library implements a simple yet very useful mechanism, allowing the user to delay and collect proof obligations in Coq scripts. Using these libraries, Armaël extended the CFML tool with support for making asymptotic time complexity claims as part of specifications. He developed tactics that perform (guided) inference and resolution of recursive equations for the cost of recursive programs.

Armaël evaluated this framework on several small-scale case studies, namely simple algorithms such as binary search, selection sort, and the Bellman-Ford algorithm. This work has been accepted for publication at the conference ESOP 2018.

7.5.2. Revisiting the CPS transformation and its implementation

Participant: François Pottier.

While preparing an MPRI lecture on the CPS transformation, François Pottier did a machine-checked proof of semantic correctness for Danvy and Filinski's properly tail-recursive, one-pass, call-by-value CPS transformation.

He proposed a new first-order, one-pass, compositional formulation of the transformation. He pointed out that Danvy and Filinski's simulation diagram does not hold in the presence of `let` and proved a slightly more complex diagram, which involves parallel reduction. He suggested representing variables as de Bruijn indices and showed that, thanks to state-of-the-art libraries such as Autosubst, this does not represent a significant impediment to formalization. Finally, he noted that, given this representation of terms, it is not obvious how to efficiently implement the transformation. To address this issue, he proposed a novel higher-order formulation of the CPS transformation, proved that it is correct, and informally argued that it runs in time $O(n \log n)$.

This work has been submitted for publication in a journal.

7.5.3. Zenon

Participant: Damien Doligez.

This year, Damien Doligez did maintenance work on Zenon: updating to the latest version of OCaml and fixing a few bugs. He also started work on adding a few minor features, such as inductive proofs for mutually inductive types.

7.5.4. TLA+

Participants: Damien Doligez, Leslie Lamport [Microsoft Research], 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 [44], and to build tools for writing TLA+ specifications and mechanically checking the proofs.

Damien is still working on a new version of TLAPS and has started writing a formal description of the semantics of TLA+.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. The Caml Consortium

Participants: Xavier Leroy [**contact**], Damien Doligez, 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 Caml. 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 16 member companies:

- Aesthetic Integration
- Ahrefs
- Be Sport
- Bloomberg
- CEA
- Citrix
- Dassault Aviation
- Docker
- Esterel Technologies
- Facebook
- Jane Street
- Kernelyze LLC
- LexiFi
- Microsoft
- OCamlPro
- SimCorp

For a complete description of this structure, refer to http://caml.inria.fr/consortium/. Xavier Leroy chairs the scientific committee of the Consortium.

8.1.2. The OCaml Foundation

Participant: Michel Mauny.

Throughout 2017, Michel Mauny has been preparing the project of an OCaml Foundation, which should support OCaml in a more efficient way than the existing Caml Consortium could do, thanks to the facilities and flexibility provided by the recently created Inria Foundation. The goal is to raise enough funds to effectively support the development and evolution of OCaml, and to animate and grow its user and teaching communities.

8.1.3. Scientific Advisory for OCamlPro

Participant: Fabrice Le Fessant.

OCamlPro is a startup company founded in 2011 by Fabrice Le Fessant to promote the use of OCaml in the industry, by providing support, services and tools for OCaml to software companies. OCamlPro performs a lot of research and development, in close partnership with academic institutions such as IRILL, Inria and Univ. Paris Sud, and is involved in several collaborative projects with Gallium, such as the Bware ANR, the Vocal ANR and the Secur-OCaml FUI.

Since 2011, Fabrice Le Fessant has been a scientific advisor at OCamlPro, as part of a collaboration contract for Inria, to transfer his knowledge on the internals of the OCaml runtime and the OCaml compilers. Fabrice has left Inria in October 2017 to join OCamlPro on a full-time position.

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, Naomi Testard.

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”.

The library will be made available to all OCaml programmers and will be of particular interest to implementors of safety-critical OCaml programs, such as Coq, Astrée, Frama-C, CompCert, Alt-Ergo, as well as new projects. By offering verified program components, our work will provide the essential building blocks that are needed to significantly decrease the cost of developing new formally verified programs.

9.1.2. FUI Projects

9.1.2.1. Secur-OCaml

Participants: Damien Doligez, Fabrice Le Fessant.

The "Secur-OCaml" project (2015–2018) is 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 is invoved in this project to integrate security features in the OCaml language, to build a new independant interpreter for the language, and to update the recommendations for developers issued by the former LaFoSec project of ANSSI.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. Deepsea

Participants: Umut Acar, Vitalii 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: 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.

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.
- Cambridge University and Microsoft Research Cambridge: 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 participated in the program committee of the ACM symposium on Principles of Programming Languages (POPL 2018), of the European Symposium on Programming (ESOP 2018), and of the second Principles of Secure Compilation workshop (PRISC 2018).

Jean-Marie Madiot was a member of the program committee of the Interaction and Concurrency Experience Workshop (ICE 2017).

Michel Mauny was a member of the program committee for Trends in Functional Programming in Education (TFPIE 2017).

François Pottier was program chair of the ACM SIGPLAN Workshop on Higher-Order Programming with Effects (HOPE 2017) and a member of the program committee of the Journées Françaises des Langages Applicatifs (JFLA 2018).

Mike Rainey was a member of the program committee for the IEEE International Parallel and Distributed Processing Symposium (IPDPS 2018).

Didier Rémy was a member of the program commitee of the International Symposium on Functional and Logic Programming (FLOPS 2018).

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 June 2017, he was on the editorial board for the Research Highlights column of Communications of the ACM.

Michel Mauny is 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

Until September 2017, Xavier Leroy was an appointed member of Inria's *Commission d'Évaluation*. He participated in the following Inria hiring committees: *jury d'admissibilité DR2* and *jury d'admissibilité CR1*.

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*.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Didier Rémy is Inria's delegate in the pedagogical team of the MPRI (*Master Parisien de Recherche en Informatique*).

Master: Luc Maranget, "Semantics, languages and algorithms for multi-core programming", 18 HETD, M2 (MPRI), Université Paris Diderot, France.

Master: Michel Mauny, "Principles of Programming Languages", 32 HETD, M1, ENSTA-ParisTech, France.

Master: François Pottier and Didier Rémy, "Functional programming and type systems", 18 + 18 HETD, M2 (MPRI), Université Paris Diderot, France.

Licence: Armaël Guéneau, "Initiation à la programmation" (TP), "Projet informatique" (TP), "Concepts informatiques" (TD), "Langages et automates" (TD), 64 HETD, L1 and L2, Université Paris Diderot, France.

Licence: Thomas Williams, "Projet informatique" (TD), "Programation orientée objet et interfaces graphiques" (TD/TP), 64 HETD, L2, Université Paris Diderot, France.

10.2.2. Supervision

M1: Danny Willems, Université de Mons, supervised by François Pottier.

PhD in progress: Vitalii Aksenov, "Parallel Dynamic Algorithms", Université Paris Diderot, since September 2015, supervised by Umut Acar (co-advised with Anatoly Shalyto, ITMO University of Saint Petersburg, Russia).

PhD in progress: Pierrick Couderc (ENSTA-ParisTech & OCamlPro), "Typage modulaire du langage intermédiaire du compilateur OCaml," Université Paris-Saclay, since December 2014, supervised by Michel Mauny, Grégoire Henry (OCamlPro) and Fabrice Le Fessant.

PhD in progress: Albin Coquereau (ENSTA-ParisTech), "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, supervised 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, supervised by Arthur Charguéraud and François Pottier.

PhD in progress: Naomi Testard, "Reasoning about Effect Handlers and Cooperative Concurrency", Université Paris Diderot, since January 2017, supervised by François Pottier.

PhD in progress: Thomas Williams, "Putting Ornaments into practice", Université Paris Diderot, since September 2014, supervised by Didier Rémy.

10.2.3. Juries

Xavier Leroy was on the Ph.D. committees of Quentin Carbonneaux (Yale University, August 2017) and of Gabriel Radanne (University Paris Diderot, November 2017).

François Pottier was a reviewer for the Ph.D. thesis of Sandro Stucki (École Polytechnique Fédérale de Lausanne, September 2017). He was a member of the jury for the GDR GPL dissertation award (*prix de thèse du GDR GPL*).

10.3. Popularization

Xavier Leroy wrote a popularization article describing the hunt for a hardware bug in Intel processors, which was published by the Web news site *The Next Web* [32].

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^{o} 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^{o} 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

Articles in International Peer-Reviewed Journal

[11] 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.

[12] J.-H. JOURDAN, F. POTTIER.*A Simple, Possibly Correct LR Parser for C11*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", September 2017, vol. 39, n° 4, p. 1 - 36 [*DOI :* 10.1145/3064848], https://hal.archives-ouvertes.fr/hal-01633123.

[13] F. POTTIER.*Visitors unchained*, in "Proceedings of the ACM on Programming Languages", August 2017, vol. 1, n° ICFP, p. 1 - 28 [*DOI :* 10.1145/3110272], https://hal.inria.fr/hal-01670735.

[14] 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

[15] U. A. ACAR, V. AKSENOV, S. WESTRICK.*Brief Announcement: Parallel Dynamic Tree Contraction via Self-Adjusting Computation*, in "The 29th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA '17)", Washington, United States, July 2017 [*DOI :* 10.1145/3087556.3087595], https://hal.inria.fr/hal-01664903.

[16] U. A. ACAR, N. BEN-DAVID, M. RAINEY.*Contention in Structured Concurrency: Provably Efficient Dynamic Non-Zero Indicators for Nested Parallelism*, in "22nd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming", Austin, United States, February 2017 [*DOI :* 10.1145/3018743.3018762], https://hal.inria.fr/hal-01416531.

[17] V. AKSENOV, V. GRAMOLI, P. KUZNETSOV, A. MALOVA, S. RAVI.*A Concurrency-Optimal Binary Search Tree*, in "23rd International European Conference on Parallel and Distributed Computing - Euro-Par 2017", Santiago de Compostella, Spain, August 2017, https://arxiv.org/abs/1702.04441 , https://hal.inria.fr/hal-01664898.

[18] T. BALABONSKI, P. COURTIEU, L. RIEG, S. TIXEUIL, X. URBAIN.*Certified Gathering of Oblivious Mobile Robots: survey of recent results and open problems*, in "Formal Methods for Industrial Critical Systems and Automated Verification of Critical Systems (FMICS/AVOCS)", Turin, Italy, Lecture Notes in Computer Science, Springer, September 2017, vol. 10471, p. 165-181 [*DOI :* 10.1007/978-3-319-67113-0_11], http://hal.upmc.fr/hal-01549942.

[19] G. BARANY, J. SIGNOLES.*Hybrid Information Flow Analysis for Real-World C Code*, in "TAP 2017 - 11th International Conference on Tests & Proofs", Marburg, Germany, Springer, July 2017, vol. 10375, p. 23-40 [*DOI :* 10.1007/978-3-319-61467-0_2], https://hal.inria.fr/hal-01658653.

[20] 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.

[21] A. CHARGUÉRAUD, F. POTTIER.*Temporary Read-Only Permissions for Separation Logic*, in "Proceedings of the 26th European Symposium on Programming (ESOP 2017)", Uppsala, Sweden, April 2017, https://hal.inria.fr/hal-01408657.

[22] S. FLUR, S. SARKAR, C. PULTE, K. NIENHUIS, L. MARANGET, K. E. GRAY, A. SEZGIN, M. BATTY, P. SEWELL.*Mixed-size Concurrency: ARM, POWER, C/C++11, and SC*, in "44th ACM SIGPLAN Symposium on Principles of Programming Languages (POPL 2017)", Paris, France, ACM, January 2017, https://hal.inria.fr/hal-01413221.

[23] 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 - Embedded Real Time Software and Systems", Toulouse, France, 3AF, SEE, SIE, January 2018, https://hal.inria.fr/hal-01643290.

[24] D. KÄSTNER, X. LEROY, S. BLAZY, B. SCHOMMER, M. SCHMIDT, C. FERDINAND.*Closing the Gap – The Formally Verified Optimizing Compiler CompCert*, in "SSS'17: Safety-critical Systems Symposium 2017", Bristol, United Kingdom, Developments in System Safety Engineering: Proceedings of the Twenty-fifth Safety-critical Systems Symposium, CreateSpace, February 2017, p. 163-180, https://hal.inria.fr/hal-01399482.

[25] F. POTTIER.*Verifying a Hash Table and Its Iterators in Higher-Order Separation Logic*, in "Certified Programs and Proofs", Paris, France, Proceedings of the 6th ACM SIGPLAN Conference on Certified Programs and Proofs (CPP 2017), January 2017, https://hal.inria.fr/hal-01417102.

[26] M. RAAB, G. BARANY.*Challenges in Validating FLOSS Conguration*, in "OSS 2017 - The 13th International Conference on Open Source Systems", Buenos Aires, Argentina, OSS 2017: Open Source Systems: Towards Robust Practices, Springer, May 2017, vol. 496, p. 101-114 [*DOI :* 10.1007/978-3-319-57735-7_11], https://hal.inria.fr/hal-01658595.

[27] M. RAAB, G. BARANY.*Introducing Context Awareness in Unmodified, Context-unaware Software*, in "ENASE 2017 - 12th International Conference on Evaluation of Novel Approaches to Software Engineering", Porto, Portugal, April 2017, p. 1-8, https://hal.inria.fr/hal-01658620.

Research Reports

[28] X. LEROY, D. DOLIGEZ, A. FRISCH, J. GARRIGUE, D. RÉMY, J. VOUILLON.*The OCaml system release 4.06: Documentation and user's manual*, Inria, November 2017, p. 1-726, https://hal.inria.fr/hal-00930213.

[29] X. LEROY.*The CompCert C verified compiler: Documentation and user's manual: Version 3.1*, Inria, August 2017, p. 1-68, https://hal.inria.fr/hal-01091802.

[30] T. WILLIAMS, D. RÉMY.*A Principled Approach to Ornamentation in ML*, Inria, November 2017, https://hal.inria.fr/hal-01628060.

Scientific Popularization

[31] J. ALGLAVE, L. MARANGET, P. MCKENNEY, A. STERN, A. PARRI.*A formal kernel memory-ordering model (Part 1 and 2)*, April 2017, Article published in the online magazine "Linux Weekly News" (LWN), available on the web at https://lwn.net/Articles/718628 and https://lwn.net/Articles/720550, https://hal.inria.fr/hal-01668178.

[32] X. LEROY.*How I found a crash bug with hyperthreading in Intel's Skylake processors*, July 2017, News article at The Next Web (https://tnw.to/2tJ08uM), https://hal.inria.fr/hal-01620870.

Other Publications

[33] F. BALESTRIERI, M. MAUNY.*Generic Programming in OCAML*, March 2017, working paper or preprint, https://hal.inria.fr/hal-01664286.

[34] G. BARANY.*Liveness-Driven Random Program Generation*, December 2017, https://arxiv.org/abs/1709.04421 - Pre-proceedings paper presented at the 27th International Symposium on Logic-Based Program Synthesis and Transformation (LOPSTR 2017), Namur, Belgium, 10-12 October 2017 (arXiv:1708.07854), https://hal.inria.fr/hal-01658563.

[35] A. CHARGUÉRAUD, J.-C. FILLIÂTRE, M. PEREIRA, F. POTTIER.*VOCAL – A Verified OCAml Library*, September 2017, ML Family Workshop 2017, https://hal.inria.fr/hal-01561094.

[36] A. CHARGUÉRAUD, M. RAINEY.*Efficient Representations for Large Dynamic Sequences in ML*, September 2017, ML Family Workshop, Poster, https://hal.inria.fr/hal-01669407.

[37] M. RAAB, G. BARANY.*Introducing Context Awareness in Unmodified, Context-unaware Software*, December 2017, https://arxiv.org/abs/1702.06806 - working paper or preprint, https://hal.inria.fr/hal-01658638.

References in notes

[38] U. A. ACAR, A. CHARGUÉRAUD, M. RAINEY.*Theory and Practice of Chunked Sequences*, A. S. SCHULZ, D. WAGNER (editors), Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, p. 25–36, https://doi.org/10.1007/978-3-662-44777-2_3.

[39] U. A. ACAR, A. CHARGUÉRAUD, M. RAINEY.*Oracle-Guided Scheduling for Controlling Granularity in Implicitly Parallel Languages*, in "Journal of Functional Programming", November 2016, vol. 26 [*DOI :* 10.1017/S0956796816000101], https://hal.inria.fr/hal-01409069.

[40] F. BALESTRIERI, M. MAUNY.*Generic Programming in OCaml*, in "OCaml 2016 - The OCaml Users and Developers Workshop", Nara, Japan, September 2016, https://hal.inria.fr/hal-01413061.

[41] 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.

[42] F. ELLEN, Y. LEV, V. LUCHANGCO, M. MOIR.*SNZI: Scalable NonZero Indicators*, in "Proceedings of the Twenty-sixth Annual ACM Symposium on Principles of Distributed Computing", PODC '07, 2007, p. 13–22.

[43] 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.

[44] 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.

[45] X. LEROY, D. DOLIGEZ, J. GARRIGUE, D. RÉMY, J. VOUILLON.*The Objective Caml system, documentation and user's manual – release 4.02*, Inria, August 2014, http://caml.inria.fr/pub/docs/manual-ocaml-4.02/.

[46] 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.

[47] B. C. PIERCE.*Types and Programming Languages*, MIT Press, 2002.

[48] 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.

[49] 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.

[50] 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.

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:

B1.1.8. - Evolutionnary biology
B1.1.11. - Systems biology
B6.3.2. - Network protocols
B6.3.4. - Social Networks
B7.2. - Smart travel

1. Personnel

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

Yacine Boufkhad [Univ Denis Diderot, Associate Professor]
Pierre Charbit [Univ Denis Diderot, Associate Professor]
Fabien de Montgolfier [Univ Denis Diderot, Associate Professor]
Hugues Fauconnier [Univ Denis Diderot, Associate Professor, HDR]
Carole Delporte-Gallet [Univ Denis Diderot, Professor, HDR]
Michel Habib [Univ Denis Diderot, Professor, HDR]

PhD Students

Simon Collet [ERC BDA]
Lucas Boczkowski [ERC BDA]
Laurent Feuilloley [ENS Cachan]
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. New Software and Platforms

5.1. big-graph-tools

FUNCTIONAL DESCRIPTION: Gang is developping a software for big graph manipulation. A preliminary library offering diameter and skeleton computation is available at https://who.rocq.inria.fr/Laurent.Viennot/dev/big-graph-tools/. 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/

5.2. 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/

6. New Results

6.1. Graph and Combinatorial Algorithms

6.1.1. Induced Matching algorithms

In [21] we study the maximum induced matching problem on a graph G. Induced matchings correspond to independent sets in $L^2(G)$, the square of the line graph of G. The problem is NP-complete on bipartite graphs. In this work, we show that for a number of graph families with forbidden vertex orderings, almost all forbidden patterns on three vertices are preserved when taking the square of the line graph. That is, given a graph class $\mathcal{G}$ characterized by a vertex ordering, and a graph $G = (V, E) \in \mathcal{G}$ with a corresponding vertex ordering σ of V, one can produce (in linear time in the size of G) an ordering on the vertices of $L^2(G)$, that shows that $L^2(G) \in \mathcal{G}$. This result gives alternate closure proofs for the $L^2(\bullet)$ closure operation. Furthermore, these orderings on $L^2(G)$ can be exploited algorithmically to compute a maximum induced matching for graphs belonging to $\mathcal{G}$ faster. We illustrate this latter fact in the second half of the paper where we focus on cocomparability graphs, a large graph class that includes interval, permutation, and trapezoid graphs, and we present the first $O(mn)$ time algorithm to compute a maximum weighted induced matching on G; an improvement from the best known $O(n^4)$ time algorithm for the unweighted case.

6.1.2. The LexBFS cycle on cocomparability graphs

Since its introduction to recognize chordal graphs by Rose, Tarjan, and Lueker, Lexicographic Breadth First Search (LexBFS) has been used to come up with simple, often linear time, algorithms on various classes of graphs. These algorithms, called multi-sweep algorithms, compute a number of LexBFS orderings $\sigma_1, ..., \sigma_k$, where σ_i is used to break ties for σ_{i+1}, we write $LexBFS^+(\sigma_i) = \sigma_{i+1}$. For instance, Corneil et al. gave a linear time multi-sweep algorithm to recognize interval graphs [SODA 1998], Kratsch et al. gave a certifying recognition algorithm for interval and permutation graphs [SODA 2003]. Since the number of LexBFS orderings for a graph is finite, after some fixed number of $^+$ sweeps, we will eventually loop in a sequence of $\sigma_1, ..., \sigma_k$ vertex orderings such that $\sigma_{i+1} = LexBFS^+(\sigma_i)$ modulo k.

In [13] we introduce and study this new graph invariant, *LexCycle*(G), defined as the maximum length of a cycle of vertex orderings obtained via a sequence of LexBFS$^+$. In this work, we focus on graph classes with small LexCycle. We give evidence that a small LexCycle often leads to linear structure that has been exploited algorithmically on a number of graph classes. In particular, we show that for proper interval, interval, co-bipartite, domino-free cocomparability graphs, as well as trees, there exists two orderings σ and τ such that $\sigma = LexBFS^+(\tau)$ and $\tau = LexBFS^+(\sigma)$. One of the consequences of these results is the simplest algorithm to compute a transitive orientation for these graph classes.

It was conjectured by Stacho [2015] that LexCycle is at most the asteroidal number of the graph class, we disprove this conjecture by giving a construction for which the LexCycle(G) grows polynomially in the asteroidal number of G.

6.1.3. Approximation Strategies for Generalized Binary Search in Weighted Trees

In [15], we have considered the following generalization of the binary search problem. A search strategy is required to locate an unknown target node t in a given tree T. Upon querying a node v of the tree, the strategy receives as a reply an indication of the connected component of $T \setminus \{v\}$ containing the target t. The cost of querying each node is given by a known non-negative weight function, and the considered objective is to minimize the total query cost for a worst-case choice of the target.

Designing an optimal strategy for a weighted tree search instance is known to be strongly NP-hard, in contrast to the unweighted variant of the problem which can be solved optimally in linear time. Here, we show that weighted tree search admits a quasi-polynomial time approximation scheme: for any $0 < \varepsilon < 1$, there exists a $(1+\varepsilon)$-approximation strategy with a computation time of $n^{O(\log n/\varepsilon^2)}$. Thus, the problem is not APX-hard, unless $NP \subseteq DTIME(n^{O(\log n)})$. By applying a generic reduction, we obtain as a corollary that the studied problem admits a polynomial-time $O(\sqrt{\log n})$-approximation. This improves previous $\widehat{O}(\log n)$-approximation approaches, where the $\widehat{O}$-notation disregards $O(\mathrm{poly}\log\log n)$-factors.

6.1.4. The Dependent Doors Problem: An Investigation into Sequential Decisions without Feedback

In [24] we introduce the *dependent doors problem* as an abstraction for situations in which one must perform a sequence of possibly 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.

6.2. Distributed Computing

6.2.1. Robust Detection in Leak-Prone Population Protocols

In [10], we aim to design population protocols for the problem of detecting a signal in the presence of faults, motivated by scenarios of chemical computation. In contrast to electronic computation, chemical computation is noisy and susceptible to a variety of sources of error, which has prevented the construction of robust complex systems. To be effective, chemical algorithms must be designed with an appropriate error model in mind. Here we consider the model of chemical reaction networks that preserve molecular count (population protocols), and ask whether computation can be made robust to a natural model of unintended "leak" reactions. Our definition of leak is motivated by both the particular spurious behavior seen when implementing chemical reaction networks with DNA strand displacement cascades, as well as the unavoidable side reactions in any implementation due to the basic laws of chemistry. We develop a new "Robust Detection" algorithm for the problem of fast (logarithmic time) single molecule detection, and prove that it is robust to this general model of leaks. Besides potential applications in single molecule detection, the error-correction ideas developed here might enable a new class of robust-by-design chemical algorithms. Our analysis is based on a non-standard hybrid argument, combining ideas from discrete analysis of population protocols with classic Markov chain techniques.

6.2.2. Minimizing Message Size in Stochastic Communication Patterns: Fast Self-Stabilizing Protocols with 3 bits

In [12] 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 that uses ℓ-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.

6.2.3. The ANTS Problem

In [6] we introduce the *Ants Nearby Treasure Search (ANTS)* problem, which models natural cooperative foraging behavior such as that performed by ants around their nest. In this problem, k probabilistic agents, initially placed at a central location, collectively search for a treasure on the two-dimensional grid. The treasure is placed at a target location by an adversary and the agents' goal is to find it as fast as possible as a function of both k and D, where D is the (unknown) distance between the central location and the target. We concentrate on the case in which agents cannot communicate while searching. It is straightforward to see that the time until at least one agent finds the target is at least $\Omega(D + D^2/k)$, even for very sophisticated agents, with unrestricted memory. Our algorithmic analysis aims at establishing connections between the time complexity and the initial knowledge held by agents (*e.g.*, regarding their total number k), as they commence the search. We provide a

range of both upper and lower bounds for the initial knowledge required for obtaining fast running time. For example, we prove that $\log\log k + \Theta(1)$ bits of initial information are both necessary and sufficient to obtain asymptotically optimal running time, *i.e.,*$O(D + D^2/k)$. We also we prove that for every $0 < \epsilon < 1$, running in time $O(\log^{1-\epsilon} k \cdot (D + D^2/k))$ requires that agents have the capacity for storing $\Omega(\log^{\epsilon} k)$ different states as they leave the nest to start the search. To the best of our knowledge, the lower bounds presented in this paper provide the first non-trivial lower bounds on the memory complexity of probabilistic agents in the context of search problems.

We view this paper as a "proof of concept" for a new type of interdisciplinary methodology. To fully demonstrate this methodology, the theoretical tradeoff presented here (or a similar one) should be combined with measurements of the time performance of searching ants.

6.2.4. Breathe before Speaking: Efficient Information Dissemination despite Noisy, Limited and Anonymous Communication

Distributed computing models typically assume reliable communication between processors. While such assumptions often hold for engineered networks, e.g., due to underlying error correction protocols, their relevance to biological systems, wherein messages are often distorted before reaching their destination, is quite limited. In this study we take a first step towards reducing this gap by rigorously analyzing a model of communication in large anonymous populations composed of simple agents which interact through short and highly unreliable messages.

In [9] we focus on the broadcast problem and the majority-consensus problem. Both are fundamental information dissemination problems in distributed computing, in which the goal of agents is to converge to some prescribed desired opinion. We initiate the study of these problems in the presence of communication noise. Our model for communication is extremely weak and follows the push gossip communication paradigm: In each round each agent that wishes to send information delivers a message to a random anonymous agent. This communication is further restricted to contain only one bit (essentially representing an opinion). Lastly, the system is assumed to be so noisy that the bit in each message sent is flipped independently with probability $1/2 - \epsilon$, for some small $\epsilon > 0$.

Even in this severely restricted, stochastic and noisy setting we give natural protocols that solve the noisy broadcast and the noisy majority-consensus problems efficiently. Our protocols run in $O(\log n/\epsilon^2)$ rounds and use $O(n \log n/\epsilon^2)$ messages/bits in total, where n is the number of agents. These bounds are asymptotically optimal and, in fact, are as fast and message efficient as if each agent would have been simultaneously informed directly by an agent that knows the prescribed desired opinion. Our efficient, robust, and simple algorithms suggest balancing between silence and transmission, synchronization, and majority-based decisions as important ingredients towards understanding collective communication schemes in anonymous and noisy populations.

6.2.5. Parallel Search with no Coordination

In [23] we consider a parallel version of a classical Bayesian search problem. k agents are looking for a treasure that is placed in one of finately many boxes according to a known distribution p. The aim is to minimize the expected time until the first agent finds it. Searchers run in parallel where at each time step each searcher can "peek" into a box. A basic family of algorithms which are inherently robust is *non-coordinating* algorithms. Such algorithms act independently at each searcher, differing only by their probabilistic choices. We are interested in the price incurred by employing such algorithms when compared with the case of full coordination.

We first show that there exists a non-coordination algorithm, that knowing only the relative likelihood of boxes according to p, has expected running time of at most $10 + 4(1 + \frac{1}{k})^2 T$, where T is the expected running time of the best fully coordinated algorithm. This result is obtained by applying a refined version of the main algorithm suggested by Fraigniaud, Korman and Rodeh in STOC'16, which was designed for the context of linear parallel search.

We then describe an optimal non-coordinating algorithm for the case where the distribution p is known. The running time of this algorithm is difficult to analyse in general, but we calculate it for several examples. In the case where p is uniform over a finite set of boxes, then the algorithm just checks boxes uniformly at random among all non-checked boxes and is essentially 2 times worse than the coordinating algorithm. We also show simple algorithms for Pareto distributions over M boxes. That is, in the case where $p(x) \sim 1/x^b$ for $0 < b < 1$, we suggest the following algorithm: at step t choose uniformly from the boxes unchecked in $\{1, ..., \min(M, \lfloor t/\sigma \rfloor)\}$, where $\sigma = b/(b + k - 1)$. It turns out this algorithm is asymptotically optimal, and runs about $2 + b$ times worse than the case of full coordination.

6.2.6. Wait-free local algorithms

When considering distributed computing, reliable message-passing synchronous systems on the one side, and asynchronous failure-prone shared-memory systems on the other side, remain two quite independently studied ends of the reliability/asynchrony spectrum. The concept of locality of a computation is central to the first one, while the concept of wait-freedom is central to the second one. In [2] we propose a new DECOUPLED model in an attempt to reconcile these two worlds. It consists of a synchronous and reliable communication graph of n nodes, and on top a set of asynchronous crash-prone processes, each attached to a communication node. To illustrate the DECOUPLED model, the paper presents an asynchronous 3-coloring algorithm for the processes of a ring. From the processes point of view, the algorithm is wait-free. From a locality point of view, each process uses information only from processes at distance $O(\log * n)$ from it. This local wait-free algorithm is based on an extension of the classical Cole and Vishkin's vertex coloring algorithm in which the processes are not required to start simultaneously.

6.2.7. Immediate t-resilient Snapshot

An immediate snapshot object is a high level communication object, built on top of a read/write distributed system in which all except one processes may crash. It allows each process to write a value and obtains a set of pairs (process id, value) such that, despite process crashes and asynchrony, the sets obtained by the processes satisfy noteworthy inclusion properties. Considering an n-process model in which up to t processes are allowed to crash, [14] is on the construction of t-resilient immediate snapshot objects.

6.2.8. Decidability classes for mobile agents computing

In [7], we establish a classification of decision problems that are to be solved by mobile agents operating in unlabeled graphs, using a deterministic protocol. The classification is with respect to the ability of a team of agents to solve decision problems, possibly with the aid of additional information. In particular, our focus is on studying differences between the decidability of a decision problem by agents and its verifiability when a certificate for a positive answer is provided to the agents (the latter is to the former what NP is to P in the framework of sequential computing). We show that the class MAV of mobile agents verifiable problems is much wider than the class MAD of mobile agents decidable problems. Our main result shows that there exist natural MAV-complete problems: the most difficult problems in this class, to which all problems in MAV are reducible via a natural mobile computing reduction. Beyond the class MAV we show that, for a single agent, three natural oracles yield a strictly increasing chain of relative decidability classes.

6.2.9. Distributed Detection of Cycles

Distributed property testing in networks has been introduced by Brakerski and Patt-Shamir (2011), with the objective of detecting the presence of large dense sub-networks in a distributed manner. Recently, Censor-Hillel et al. (2016) have shown how to detect 3-cycles in a constant number of rounds by a distributed algorithm. In a follow up work, Fraigniaud et al. (2016) have shown how to detect 4-cycles in a constant number of rounds as well. However, the techniques in these latter works were shown not to generalize to larger cycles C_k with $k \geq 5$. In [19], we completely settle the problem of cycle detection, by establishing the following result. For every $k \geq 3$, there exists a distributed property testing algorithm for C_k-freeness, performing in a constant number of rounds. All these results hold in the classical CONGEST model for distributed network computing. Our algorithm is 1-sided error. Its round-complexity is $O(1/\epsilon)$ where $\epsilon \in (0, 1)$ is the property testing parameter measuring the gap between legal and illegal instances.

6.2.10. What Can Be Verified Locally?

In [18], we are considering *distributed network computing*, in which computing entities are connected by a network modeled as a connected graph. These entities are located at the nodes of the graph, and they exchange information by message-passing along its edges. In this context, we are adopting the classical framework for *local distributed decision*, in which nodes must collectively decide whether their network configuration satisfies some given boolean predicate, by having each node interacting with the nodes in its vicinity only. A network configuration is accepted if and only if every node individually accepts. It is folklore that not every Turing-decidable network property (e.g., whether the network is planar) can be decided locally whenever the computing entities are Turing machines (TM). On the other hand, it is known that every Turing-decidable network property can be decided locally if nodes are running *non-deterministic* Turing machines (NTM). However, this holds only if the nodes have the ability to guess the identities of the nodes currently in the network. That is, for different sets of identities assigned to the nodes, the correct guesses of the nodes might be different. If one asks the nodes to use the same guess in the same network configuration even with different identity assignments, i.e., to perform *identity-oblivious* guesses, then it is known that not every Turing-decidable network property can be decided locally.

We show that every Turing-decidable network property can be decided locally if nodes are running *alternating* Turing machines (ATM), and this holds even if nodes are bounded to perform identity-oblivious guesses. More specifically, we show that, for every network property, there is a local algorithm for ATMs, with at most 2 alternations, that decides that property. 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 level k contains those tasks solvable with ATMs with k 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 properties 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*.

6.2.11. Certification of Compact Low-Stretch Routing Schemes

On the one hand, the correctness of routing protocols in networks is an issue of utmost importance for guaranteeing the delivery of messages from any source to any target. On the other hand, a large collection of *routing schemes* have been proposed during the last two decades, with the objective of transmitting messages along short routes, while keeping the routing tables small. Regrettably, all these schemes share the property that an adversary may modify the content of the routing tables with the objective of, e.g., blocking the delivery of messages between some pairs of nodes, without being detected by any node.

In [17], we present a simple *certification* mechanism which enables the nodes to locally detect any alteration of their routing tables. In particular, we show how to locally verify the stretch-3 routing scheme by Thorup and Zwick [SPAA 2001] by adding certificates of $\widetilde{O}(\sqrt{n})$ bits at each node in n-node networks, that is, by keeping the memory size of the same order of magnitude as the original routing tables. We also propose a new *name-independent* routing scheme using routing tables of size $\widetilde{O}(\sqrt{n})$ bits. This new routing scheme can be locally verified using certificates on $\widetilde{O}(\sqrt{n})$ bits. Its stretch is 3 if using handshaking, and 5 otherwise.

6.2.12. Error-Sensitive Proof-Labeling Schemes

Proof-labeling schemes are known mechanisms providing nodes of networks with *certificates* that can be *verified* locally by distributed algorithms. Given a boolean predicate on network states, such schemes enable to check whether the predicate is satisfied by the actual state of the network, by having nodes interacting with their neighbors only. Proof-labeling schemes are typically designed for enforcing fault-tolerance, by making sure that if the current state of the network is illegal with respect to some given predicate, then at least one node will detect it. Such a node can raise an alarm, or launch a recovery procedure enabling the system to return to a legal state. We introduce *error-sensitive* proof-labeling schemes. These are proof-labeling schemes which guarantee that the number of nodes detecting illegal states is linearly proportional to the edit-distance between the current state and the set of legal states. By using error-sensitive proof-labeling schemes, states which are far from satisfying the predicate will be detected by many nodes, enabling fast return to legality. In [20], we

provide a structural characterization of the set of boolean predicates on network states for which there exist error-sensitive proof-labeling schemes. This characterization allows us to show that classical predicates such as, e.g., acyclicity, and leader admit error-sensitive proof-labeling schemes, while others like regular subgraphs don't. We also focus on *compact* error-sensitive proof-labeling schemes. In particular, we show that the known proof-labeling schemes for spanning tree and MST, using certificates on $O(\log n)$ bits, and on $O(\log^2 n)$ bits, respectively, are error-sensitive, as long as the trees are locally represented by adjacency lists, and not by a pointer to the parent.

6.2.13. Distributed Property Testing

In [16], we designed distributed testing algorithms of graph properties in the CONGEST model [Censor-Hillel et al. 2016], especially for testing subgraph-freeness. Testing a given property means that we have to distinguish between graphs having the property, and graphs that are ϵ-far from having it, meaning that one must remove an ϵ-fraction of the edges to obtain it. We established a series of results, among which:

- Testing H-freeness in a constant number of rounds, for any graph H that can be transformed into a tree by removing a single edge. This includes, e.g., cycle-freeness for any constant cycle, and K_4-freeness. As a byproduct, we give a deterministic CONGEST protocol determining whether a graph contains a fixed tree as a subgraph.
- For cliques K_k with $k \geq 5$, we show that K_k-freeness can be tested in $O(\left(\frac{m}{\epsilon}\right)^{\frac{1}{2}+\frac{1}{k-2}})$ rounds, where m is the number of edges in the network graph.
- We describe a general procedure for converting ϵ-testers with $f(D)$ rounds, where D denotes the diameter of the graph, to work in $O((\log n)/\epsilon) + f((\log n)/\epsilon)$ rounds, where n is the number of processors of the network. We then apply this procedure to obtain an ϵ-tester for testing whether a graph is bipartite.

These protocols extend and improve previous results of [Censor-Hillel et al. 2016] and [Fraigniaud et al. 2016].

6.3. Models and Algorithms for Networks

6.3.1. Analysis of Multiple Random Walks on Paths and Grids

In [22], we derive several new results on multiple random walks on "low-dimensional" graphs. First, inspired by an example of a weighted random walk on a path of three vertices given by Efremenko and Reingold, we prove the following dichotomy: as the path length n tends to infinity, we have a super-linear speed-up w.r.t. the cover time if and only if the number of walks k is equal to 2. An important ingredient of our proofs is the use of a continuous-time analogue of multiple random walks, which might be of independent interest. Finally, we also present the first tight bounds on the speed-up of the cover time for any d-dimensional grid with $d \geq 2$ being an arbitrary constant, and reveal a sharp transition between linear and logarithmic speed-up.

6.3.2. Decomposing a Graph into Shortest Paths with Bounded Eccentricity

In [11], we introduce the problem of hub-laminar decomposition which generalizes that of computing a shortest path with minimum eccentricity (MESP). Intuitively, it consists in decomposing a graph into several paths that collectively have small eccentricity and meet only near their extremities. The problem is related to computing an isometric cycle with minimum eccentricity (MEIC). It is also linked to DNA reconstitution in the context of metagenomics in biology. We show that a graph having such a decomposition with long enough paths can be decomposed in polynomial time with approximated guaranties on the parameters of the decomposition. Moreover, such a decomposition with few paths allows to compute a compact representation of distances with additive distortion. We also show that having an isometric cycle with small eccentricity is related to the possibility of embedding the graph in a cycle with low distortion.

6.3.3. Individual versus collective cognition in social insects

The concerted responses of eusocial insects to environmental stimuli are often referred to as collective cognition at the level of the colony. To achieve collective cognition, a group can draw on two different sources: individual cognition and the connectivity between individuals. Computation in neural networks, for example, is attributed more to sophisticated communication schemes than to the complexity of individual neurons. The case of social insects, however, can be expected to differ. This is because individual insects are cognitively capable units that are often able to process information that is directly relevant at the level of the colony. Furthermore, involved communication patterns seem difficult to implement in a group of insects as they lack a clear network structure. In [5] we discusses links between the cognition of an individual insect and that of the colony. We provide examples for collective cognition whose sources span the full spectrum between amplification of individual insect cognition and emergent group-level processes.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. 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 Nidhi Hegde. An ADR (joint research action) is dedicated to content centric networks and forwarding information verification. The PhD thesis of Leonardo Linguaglossa was funded by this contract.

This collaboration is developed inside the Alcatel-Lucent and Inria joint research lab.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. 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.

8.2. National Initiatives

8.2.1. 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.

8.2.2. ANR MultiMod

Participants: Adrian Kosowski, Laurent Viennot.

David Coudert (Sophia Antipolis) leads this project. L. Viennot coordinates locally. The project begins 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.

8.2.3. 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.

8.2.4. ANR Distancia

Participants: Pierre Charbit, Michel Habib, Laurent Viennot.

Victor Chepoi (Univ. Marseille) leads this project. P. Charbit coordinates locally. The project begins 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.

8.2.5. 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.

8.3. European Initiatives

8.3.1. 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.

8.3.2. 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.

8.4. International Initiatives

8.4.1. 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/).

8.4.2. Inria International Partners

8.4.2.1. 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).

Pierluigi Crescenzi (University of Florence, Italy) is a frequent visitor to the team and maintains an active research collaboration with Gang team members (Pierre Fraigniaud).

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).

8.5. International Research Visitors

8.5.1. Visits of International Scientists

[chercheurs invités, profs invités (via université), Les internships sont à mettre dans la subsection suivante.]

Sergio Rajsbaum (UNAM-Mexico) was invited for two months (May-June).

Eli Gafni visited the team for one month (mid-June to mid-July).

Lalla Mouatadid visited the group for 2 weeks in 2017. She is finishing her PhD in computer. Science at University of Toronto, under the supervision of prof. Derek Corneil and Alan Borodin.

8.5.2. Visits to International Teams

Carole Delporte-Gallet and Hugues Fauconnier have visited 2x10 days Sergio Rajsbaum at UNAM (Mexico) in September and November 2017.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Selection

9.1.1.1. Member of Conference Program Committees

- Adrian Kosowski: ICALP 2017, MFCS 2018, SIROCCO 2018.
- Carole Delporte-Gallet: ICDCN 2017, PODC 2017, IPDPS 2017, OPODIS 2017
- Pierre Fraigniaud: WWW 2017, SPAA 2017, DISC 2017.
- Hugues Fauconnier: SSS 2017, Algotel 2017
- Michel Habib: WG 2018
- Amos Korman: ICALP 2017
- Laurent Viennot: FCT 2017

9.1.2. Journal

9.1.2.1. Member of 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).
- Pierre Fraigniaud is a member of the Editorial Board of Fundamenta Informaticae (FI).
- Adrian Kosowski is a member of the Editorial Board of Mathematical Foundations of Computing (AIMS MFOC)

9.1.3. Invited Talks

Carole Delporte, The Basics of Distributed Computing: consensus. 9th edition of the International Spring School on Distributed Systems(METIS 2017), May 2017.

Hugues Fauconnier, The Basics of Distributed Computing: Shared Memory. 9th edition of the International Spring School on Distributed Systems (METIS 2017), May 2017.

Carole Delporte: Workshop IWDCMR 2017 in honor of Michel Raynal, Détecteurs de défaillance, May 2017.

Michel Habib: IPM Combinatorics and Computing Conference 2017 (IPMCCC2017), Téhéran, May 2017

Michel Habib: Journée Charles Hermite Complexité : Théorie des graphes et théorie des nombres, Nancy, November 2017.

Adrian Kosowski: ADGA 2017 workshop (co-located with DISC 2017), Vienna, October 2017.

9.1.4. Scientific Expertise

Adrian Kosowski was an expert panel member for grant panel PE6 of the National Science Center, Poland (Spring 2017).

Carole Delporte-Gallet was in the recruting jury of "Informaticien" at Assemblee Nationale.

Michel Habib was member of the COS (Comité d'Orientation Stratégique) of the Labex Archimede Marseille. One meeting in April 2017.

9.1.5. Research Administration

Pierre Fraigniaud is director of the Institute de Recherche en Informatique Fondamentale (IRIF).

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.

Michel Habib is member of the Administration Council of University Paris Diderot.

9.2. Teaching - Supervision - Juries

9.2.1. 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, Sécurité informatique, 36h, L3, Univ. 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 avancée, 24h, Ecole Centrale Supélec Paris, M2

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 and Michel Habib, 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, 3h, M2 MPRI, Univ. Paris Diderot

Licence: Michel Habib, Algorithmique, 45h, L, ENS Cachan

Master: Michel Habib, Algorithmique avancée, 24h, M1, Univ. Paris Diderot

Master: Michel Habib, Mobilité, 33h, M2, Univ. Paris Diderot

Master: Michel Habib, Méthodes et algorithmes pour l'accès à l'information numérique, 16h, M2, Univ. Paris Diderot

Master: Michel Habib, Algorithmique de graphes, 12h, M2, 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

9.2.2. Supervision

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: Lucas Boczkowski (co-advised by Amos Korman and Iordanis Kerenidis). Title of thesis is: "Computing with Limited Resources in Uncertain Environments". 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: Laurent Feuilloley (advised by Pierre Fraigniaud). Title of thesis is: "Synchronous Distributed Computing". Started September 2015.

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: Léo Planche (co-advised by Étienne Birmelé and Fabien de Montgolfier). Title if thesis is : "Classification de collections de graphes". Started October 2015.

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): Emmanuel Arrighi (advised by Laurent Viennot). (March-August 2017) Title of report: "Schéma d'étiquetage de graphe pour le pré-calcul de plus courts chemins".

9.2.3. Juries

Laurent Viennot was on the jury committee of the HDR thesis of Mauro Sozio (Telecom Paristech) "Algorithms for Making Sense of Massive Dynamic Graphs" at Paris-Saclay University, July 2017.

Michel Habib was referee and on the jury committee of the HDR thesis of Mauro Sozio (Telecom Paristech) "Algorithms for Making Sense of Massive Dynamic Graphs" at Paris-Saclay University, July 2017.

Michel Habib was on the jury committee of the HDR thesis of Reza Naserasr "Projective Cubes, a coloring point of view" at Paris Diderot University, july 2017.

Michel Habib was referee and on the jury of the PhD thesis of Kaoutar Ghazi "Heurisitques et conjectures à propos de la 2-dimension des ordres", Clermont-Ferrand, october 2017.

Michel Habib was referee and on the jury of the HDR thesis of Phan Thi Ha Duong "Chip Firing Game and related models: algebraic structures and enumerative combinatorics" at Paris Diderot University, december 2017.

Carole Delporte-Gallet was on the jury committee of the thesis of Florent Chevrou " Formalisation of Asynchronous Interactions" Toulouse University, november 2017.

Hugues Fauconnier was referee and on the jury of the PhD thesis of Matoula Petrolia "Distribuer Shared memory in failure prone message passing systèmes", Nantes University, october 2017.

9.3. Popularization

- Amos Korman wrote the atricle “Conseils d’une fourmi: Ne me prenez pas trop au sérieux !” published in the French journal *interstices* in 10.7.2017.
- An article about a paper by Amos Korman, Adrian Kosowski and Lucas Boczkowski was published in *CNRS news* in 10.1.2017.
- An article called “Les fourmis, génies de l’orientation” about a paper by Amos Korman, Adrian Kosowski and Lucas Boczkowski was published in *Le Monde* in 30.1.2017.
- Laurent Viennot is “commissaire d’exposition” for the permanent exposition on “Informatique et sciences du numérique” at Palais de la déécouverte in Paris (opening in February 2018).

10. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] E. BAMPAS, L. GĄSIENIEC, N. HANUSSE, D. ILCINKAS, R. KLASING, A. KOSOWSKI, T. RADZIK.*Robustness of the Rotor–Router Mechanism*, in "Algorithmica", July 2017, vol. 78, n^{o} 3, p. 869-895 [*DOI :* 10.1007/s00453-016-0179-y], https://hal.inria.fr/hal-01416012.

[2] A. CASTAÑEDA, C. DELPORTE-GALLET, ·. HUGUES FAUCONNIER, S. RAJSBAUM, M. RAYNAL.*Making Local Algorithms Wait-Free: the Case of Ring Coloring*, in "Theory of Computing Systems", May 2017, p. 1-22 [*DOI :* 10.1007/s00224-017-9772-y], https://hal.archives-ouvertes.fr/hal-01672723.

[3] J. CZYZOWICZ, D. DERENIOWSKI, L. GĄSIENIEC, R. KLASING, A. KOSOWSKI, D. PAJĄK.*Collision-Free Network Exploration*, in "Journal of Computer and System Sciences", 2017, vol. 86, p. 70-81 [*DOI :* 10.1016/j.jcss.2016.11.008], https://hal.inria.fr/hal-01416026.

[4] J. CZYZOWICZ, L. GĄSIENIEC, A. KOSOWSKI, E. KRANAKIS, D. KRIZANC, N. TALEB.*When Patrolmen Become Corrupted: Monitoring a Graph Using Faulty Mobile Robots*, in "Algorithmica", 2017, vol. 79, n^{o} 3, p. 925-940 [*DOI :* 10.1007/s00453-016-0233-9], https://hal.inria.fr/hal-01416010.

[5] O. FEINERMAN, A. KORMAN.*Individual versus collective cognition in social insects*, in "The Journal of Experimental Biology", 2017, vol. 220, p. 73 - 82, https://arxiv.org/abs/1701.05080 [*DOI :* 10.1242/jeb.143891], https://hal.inria.fr/hal-01432718.

[6] O. FEINERMAN, A. KORMAN.*The ANTS problem*, in "Distributed Computing", June 2017, https://arxiv.org/abs/1701.02555 [*DOI :* 10.1007/s00446-016-0285-8], https://hal.inria.fr/hal-01430372.

[7] P. FRAIGNIAUD, A. PELC.*Decidability classes for mobile agents computing*, in "Journal of Parallel and Distributed Computing", November 2017, vol. 109, p. 117-128 [*DOI :* 10.1016/J.JPDC.2017.04.003], https://hal.inria.fr/hal-01674623.

[8] R. KLASING, A. KOSOWSKI, D. PAJĄK, T. SAUERWALD.*The multi-agent rotor-router on the ring: a deterministic alternative to parallel random walks*, in "Distributed Computing", 2017, vol. 30, n° 2, p. 127-148 [*DOI :* 10.1007/S00446-016-0282-Y], https://hal.inria.fr/hal-01416011.

[9] A. KORMAN, O. FEINERMAN, B. HAEUPLER.*Breathe before speaking: efficient information dissemination despite noisy, limited and anonymous communication*, in "Distributed Computing", October 2017, vol. 30, n° 5, p. 339 - 355 [*DOI :* 10.1007/S00446-015-0249-4], https://hal.inria.fr/hal-01672280.

International Conferences with Proceedings

[10] D. ALISTARH, B. DUDEK, A. KOSOWSKI, D. SOLOVEICHIK, P. UZNANSKI.*Robust Detection in Leak-Prone Population Protocols*, in "DNA 2017 - 23rd International Conference DNA Computing and Molecular Programming", Austin, TX, United States, DNA Computing and Molecular Programming, Springer, September 2017, vol. 10467, p. 155-171, https://arxiv.org/abs/1706.09937 [*DOI :* 10.1007/978-3-319-66799-7_11], https://hal.inria.fr/hal-01669203.

[11] E. E. BIRMELÉ, F. DE MONTGOLFIER, L. PLANCHE, L. VIENNOT.*Decomposing a Graph into Shortest Paths with Bounded Eccentricity*, in "28th International Symposium on Algorithms and Computation (ISAAC 2017)", Phuket, Thailand, December 2017 [*DOI :* 10.4230/LIPICS.ISAAC.2017.15], https://hal.inria.fr/hal-01671718.

[12] L. BOCZKOWSKI, A. KORMAN, E. NATALE.*Minimizing Message Size in Stochastic Communication Patterns: Fast Self-Stabilizing Protocols with 3 bits*, in "ACM-SIAM Symposium on Discrete Algorithms (SODA17)", Barcelona, Spain, January 2017, https://arxiv.org/abs/1602.04419 - 28 pages, 4 figures, https://hal.inria.fr/hal-01447435.

[13] P. CHARBIT, M. HABIB, L. MOUATADID, R. NASERASR.*A New Graph Parameter to Measure Linearity*, in "COCOA 2017 - 11th Annual International Conference on Combinatorial Optimization and Applications", Shanghai, China, International Conference on Combinatorial Optimization and Applications, Springer, December 2017, vol. 10628, p. 154-168 [*DOI :* 10.1007/978-3-319-71147-8_11], https://hal.inria.fr/hal-01672521.

[14] C. DELPORTE-GALLET, H. FAUCONNIER, S. RAJSBAUM, M. RAYNAL.*t-résilient snapshot immédiat* , in "ALGOTEL 2017 - 19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Quiberon, France, Aspects Algorithmiques des Télécommunications, May 2017, https://hal.archives-ouvertes.fr/hal-01517320.

[15] D. DERENIOWSKI, A. KOSOWSKI, P. UZNANSKI, M. ZOU.*Approximation Strategies for Generalized Binary Search in Weighted Trees*, in "ICALP 2017 - 44th International Colloquium on Automata, Languages, and Programming", Warsaw, Poland, LIPIcs, 2017, vol. 80, p. 84:1–84:14, https://arxiv.org/abs/1702.08207 [*DOI :* 10.4230/LIPICS.ICALP.2017.84], https://hal.inria.fr/hal-01470935.

[16] G. EVEN, O. FISCHER, P. FRAIGNIAUD, T. GONEN, R. LEVI, M. MEDINA, P. MONTEALEGRE, O. DENNIS, R. OSHMAN, I. RAPAPORT, I. TODINCA.*Three Notes on Distributed Property Testing*, in "DISC 2017 - 31st International Symposium on Distributed Computing", Vienna, France, October 2017, p. 1-30, https://hal.inria.fr/hal-01674664.

[17] P. FRAIGNIAUD, B. ALKIDA.*Certification of Compact Low-Stretch Routing Schemes*, in "DISC 2017 - 31st International Symposium on Distributed Computing", Vienna, Austria, October 2017, https://hal.inria.fr/hal-01674656.

[18] P. FRAIGNIAUD, O. DENNIS, B. ALKIDA, G. D'ANGELO.*What Can Be Verified Locally?*, in "STACS 2017 - 34th International Symposium on Theoretical Aspects of Computer Science", Hannover, Germany, March 2017, p. 1-13 [*DOI :* 10.4230/LIPICS.STACS.2017.8], https://hal.inria.fr/hal-01674650.

[19] P. FRAIGNIAUD, O. DENNIS.*Distributed Detection of Cycles*, in "SPAA 2017 - 29th ACM Symposium on Parallelism in Algorithms and Architectures", Washington, United States, July 2017, p. 153-162 [*DOI :* 10.1145/3087556.3087571], https://hal.inria.fr/hal-01674646.

[20] P. FRAIGNIAUD, L. FEUILLOLEY.*Error-Sensitive Proof-Labeling Schemes*, in "DISC 2017 - 31st International Symposium on Distributed Computing", Vienna, Austria, October 2017, p. 1-15 [*DOI :* 10.4230/LIPICS.DISC.2017.16], https://hal.inria.fr/hal-01674660.

[21] M. HABIB, L. MOUATADID.*Maximum Induced Matching Algorithms via Vertex Ordering Characterizations*, in "ISAAC 2017 - 28th International Symposium on Algorithms and Computation", Phuket, Thailand, December 2017, p. 1-12 [*DOI :* 10.4230/LIPICS.ISAAC.2017.43], https://hal.inria.fr/hal-01672520.

[22] A. IVASKOVIC, A. KOSOWSKI, D. PAJĄK, T. SAUERWALD.*Multiple Random Walks on Paths and Grids*, in "STACS 2017 - 34th Symposium on Theoretical Aspects of Computer Science", Hannover, Germany, March 2017, vol. 66, p. 44:1–44:14 [*DOI :* 10.4230/LIPICS.STACS.2017.44], https://hal.inria.fr/hal-01669223.

[23] A. KORMAN, Y. RODEH.*Parallel Search with no Coordination*, in "24th International Colloquium on Structural Information and Communication Complexity (SIROCCO)", Porquerolles, France, June 2017, https://arxiv.org/abs/1705.05704 , https://hal.inria.fr/hal-01523506.

[24] A. KORMAN, Y. RODEH.*The Dependent Doors Problem: An Investigation into Sequential Decisions without Feedback*, in "The 44th International Colloquium on Automata, Languages, and Programming (ICALP)", Warsaw, Poland, July 2017, https://arxiv.org/abs/1704.06096 , https://hal.inria.fr/hal-01511050.

[25] A. KOSOWSKI, L. VIENNOT.*Beyond Highway Dimension: Small Distance Labels Using Tree Skeletons*, in "SODA 2017 - 28th ACM-SIAM Symposium on Discrete Algorithms", Barcelona, Spain, January 2017, https://arxiv.org/abs/1609.00512 , https://hal.inria.fr/hal-01359084.

Scientific Popularization

[26] T. F. DRUMOND, L. VIENNOT, T. VIÉVILLE, V. FRANÇOIS.*Jouez avec les neurones de la machine*, Le Monde, October 2017, p. 1-3, Le but de ce blog est de parler d'informatique, de communiquer sur ce qu'est vraiment l'informatique en tant que science et technique, ses définitions, ses progrès, ses dangers, ses questionnements, ses succès et impacts, ses enjeux, ses métiers, son enseignement…, https://hal.inria.fr/hal-01620451.

[27] O. FEINERMAN, A. KORMAN.*Conseils d'une fourmi : Ne me prenez pas trop au sérieux !*, in "Interstices", July 2017, https://hal.inria.fr/hal-01616341.

Other Publications

[28] E. E. BIRMELÉ, F. DE MONTGOLFIER, L. PLANCHE, L. VIENNOT.*Decomposing a Graph into Shortest Paths with Bounded Eccentricity*, September 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01511357.

[29] B. DUDEK, A. KOSOWSKI.*Universal Protocols for Information Dissemination Using Emergent Signals*, November 2017, https://arxiv.org/abs/1705.09798 - working paper or preprint, https://hal.inria.fr/hal-01503359.

Project-Team MAMBA

Modelling and Analysis for Medical and Biological Applications

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

IN PARTNERSHIP WITH:
CNRS

Université Pierre et Marie Curie (Paris 6)

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. - Mathematical Modeling
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling (SPDE, SDE)
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

Other Research Topics and Application Domains:

B1. - Life sciences
B1.1. - Biology
B1.1.2. - Molecular biology
B1.1.3. - Cellular biology
B1.1.7. - Immunology
B1.1.8. - Evolutionnary biology
B1.1.9. - Bioinformatics
B1.1.10. - 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
B9.5.4. - Management science

1. Personnel

Research Scientists
Marie Jauffret [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, from Oct 2017]

Faculty Members
Stephane Mischler [Univ Paris Dauphine, Professor, in delegation at Inria from Sep 2017]
Benoît Perthame [Sorbonne Université, Professor]
Alexander Lorz [Sorbonne Université, Assistant Professor]

External Collaborators
Ismael González Valverde [University of Zaragoza, from Sep 2017 until Nov 2017]
Noémie Boissier [Leibniz Institute IfADo, from Jul 2017]
Géraldine Cellière [Lixoft, Zürich]
Ján Eliaš [Univ Paris-Sud]
Adeline Fermanian [Ecole Nationale des Ponts et Chaussées, from Aug 2017]
Lorena Romero Medrano [Leibniz Institute IfADo, Oct 2017]
Jieling Zhao [Leibniz Institute IfADo, from Sep 2017]

Technical Staff
Lorena Romero Medrano [Inria, from Nov 2017]
Paul Van Liedekerke [Inria]
Yi Yin [Inria]

PhD Students
Noémie Boissier [Inria, until Jun 2017]
Federica Bubba [Sorbonne Université, from Oct 2017]
Julia Delacour [Ecole Normale Supérieure Lyon, from Sep 2017]
Hugo Martin [Univ Pierre et Marie Curie]
Mathieu Mézache [Inria]
Camille Pouchol [Sorbonne Université]
Andrada Quillas Maran [Sorbonne Université, until May 2017]
Teresa Taing [Sorbonne Université, until Sep 2017]
Martin Strugarek [Ecole Nationale des Ponts et Chaussées]

Post-Doctoral Fellows
Cécile Carrère [Sorbonne Université, from Oct 2017]
Xinran Ruan [Inria, from Dec 2017]

Administrative Assistant

Kevin Bonny [Inria]

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 two kinds of 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. Modelling 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 broaden its subjects as its individuals develop their own research agendas. It aims at developing models, simulations and numerical algorithms to solve questions from life sciences involving dynamics of phenomena encountered in biological systems such as protein intra-cellular spatio-temporal dynamics, cell motion, early embryonic development, multicelluar growth, wound healing and liver regeneration, cancer evolution, healthy and tumour growth control by pharmaceuticals, protein polymerisation occurring in neurodegenerative disorders, 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 UPMC (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, UPMC, 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, are key points in our project.

Our main objective is the creation, investigation and transfer of new models, methods 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 re-organised the presentation of our research program in five main axes, three methodological, and two application-driven axes. In more details, these research axes are the following.

Axis 1 (methodological) is devoted to works 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 (methodological) is devoted to reaction and motion equations for living systems. It aims at describing biological phenomena such as tumour growth, chemotaxis and wound healing.

Axis 3 (methodological) tackles the question of model and parameter identification, combining stochastic and deterministic approaches and inverse problem methods in nonlocal and multi-scale models.

Axis 4 (applicative) 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.

Axis 5 (applicative) is devoted to growth, evolution and regeneration in populations and tissues. It involves protein aggregation and fragmentation models for neurodegenerative diseases (prion, Alzheimer), organ modelling, mainly of the liver, its damages induced by toxic molecules, and its regeneration after toxic insult. Newcomers in this applicative field are epidemiological modelling of propagation of insect vector-borne diseases by reaction-diffusion equations and of their optimal control, bacterial growth and wound healing.

3.2. Research axis 1: analysis and control for population dynamics

Personnel

Pierre-Alexandre Bliman, Jean Clairambault, Marie Doumic, Alexander Lorz, Benoît Perthame

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 behaviour of such structured population equations, see e.g. [113], [59], [79], [68]. 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 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 behaviour 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 [65], 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 [77], 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 [62].

We also enriched the models by taking into account a nucleation term, modeling the spontaneous formation of large polymers out of monomers [122]. 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 axis 4, 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 behaviour 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 [109], [106], [105], [107]. Space may be added as a complementary structure variable provided that something is known of the (Cartesian) geometry of the population [108], which is seldom the case.

Mathematical models of infectious diseases
These 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.

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 [116], followed by [112], [101], [117], [67]. This field also yields a beautiful illustration for the capacity of the team to combine and compare stochastic and PDE modelling (see axis 3), in [72].

Collaborations

- Nucleation, growth and fragmentation equations: **Juan Calvo**, university of Granada, came for two one-month visits, **Miguel Escobedo**, University of Bilbao (see also 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.
- 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 axis 4). **Emmanuel Trélat**, UPMC professor, member of LJLL and of the CAGE Inria team, is the closest Mamba collaborator for optimal control.
- Estimation and identification of epidemiological models: **Maria Soledad Aronna**, Fundação Getulio Vargas, Brazil; **Alain Rapaport**, INRA-Montpellier; **Abderrahmane Iggidr**, Inria Nancy-Grand Est
- Neural networks: **Delphine Salort,** Professor UPMC, Laboratory for computations and quantification in biology, and **Patricia Reynaud**, University of Nice, **Maria Cáceres**, university of Granada.

3.3. Research axis 2: reaction and motion equations for living systems

Personnel

Luis Almeida, Casimir Emako-Kazianou, Alexander Lorz, Benoît Perthame, Nicolas Vauchelet.

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 benefits from the ERC Advanced Grant 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.

Mathematical modelling for bacterial chemotaxis.
Chemotaxis is the phenomenon in which cells direct their motion in response to a chemical signal present in their environment. Our unique expertise is on mathematical aspects of the kinetic equations which describe the run and tumble motion of bacteria and their asymptotic analysis.

An interdisciplinary collaboration with biophysicists from Institut Curie has been successful on experimental observations concerning the interaction between two species of bacteria and emergence of travelling bands [86]. The mathematical models used in this work are derived in [51] thanks to a diffusive limit of a kinetic system with tumbling modulation along the path. A numerical investigation of this limit is provided in [87]. These works enter into the framework of the PhD of Casimir Emako-Kazianou [88]. Recently, we have been able to derive rigorously such kinetic models from a more sophisticated equation incorporating internal variable when cells adapt rapidly to changes in their environment [119].

Aggregation equation.
In the mathematical study of collective behaviour, 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 [100]. The extension to higher dimensions has been studied in [70]. 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 behaviour of such weak measure-valued solutions [93], [99].

Free boundary problems for tumour growth.
Fluid dynamic equations are now commonly used to describe tumour growth with two main classes of models: those which describe tumour growth through the dynamics of the density of tumoral cells subjected to a mechanical stress; those describing the tumour 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 [115] for a simple model. This result has motivated the use of another strategy based on viscosity solutions, leading to similar results, in [102].

Since more realistic systems are used in the analysis of medical images, we have extended these studies to include active motion of cells in [114], viscosity in [120] and proved regularity results in [110]. 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 [118], see also 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 [103].

Collaborations

- Institut Curie, joint work with Axel Buguin on bacterial models for chemotaxis.
- Shanghai Jiao Tong University, joint publications with Min Tang on bacterial models for chemotaxis and free boundary problems for tumour 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 tumour growth models.

3.4. Research axis 3: Model and parameter identification combining stochastic and deterministic approaches in nonlocal and multi-scale models

Personnel

Marie Doumic, Dirk Drasdo, Aurora Armiento, Thibault Bourgeron, Rebecca Chisholm, Tommaso Lorenzi

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 [82], TRAIL treatment of HeLa cells [61], growth of multicellular spheroids [98], blood detoxification after drug-induced liver damage [124], [92].

This naturally led to increasingly combine methods from various fields: image analysis, statistics, probability, numerical analysis, PDEs, ODEs, agent-based modelling 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 variablity 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 Philippe Robert (Inria RAP) in probability, 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 [121], [81], 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 [65]. 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) [35]. 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 [80], a work which inspired also very recently other groups in statistics and probability [62], [95] and was the basis for Adélaïde Olivier's Ph.D thesis [111], [96] and of some of her more recent works [21] (see also axis 5).

Model identification for growing multicellular spheroids

For multicellular spheroids growing under different conditions, first an agent-based model on an unstructured lattice for one condition has been developed and then stepwise extended for each additional condition which could not be captured by the present model state [98] [97] (axis 5). The multicellular dynamics has been mimicked by a master equation, intracellular processes by ODEs, and extracellular molecular transport processes by partial differential equations. The model development was based almost completely on bright field image sequences, whereby image segmentation parameter identification was performed by investigation of sensitivity and specificity of the segmentation with a biologist expert serving as gold standard. The Akaike and Bayesian Information Criteria were used to evaluate whether parameters introduced due to the model extension led to a significant increase of the information. It turned out that the final model could predict the outcome of growth conditions not considered in model development.
A similar stepwise strategy has been recently performed to identify the pressure and strain constraints of growing multicellular cell populations subject to mechanical stress. Here, a novel deformable cell model has been used that permits to display cell shape changes explicitly [127].

Data assimilation and stochastic modelling 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 [52], 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 [89], was devoted to the stochastic modelling and analysis of protein aggregation, compared both with the deterministic approach traditionnally developed in Mamba [122] and with experiments.

Model identification in liver regeneration

Based on successful model predictions for models addressing different aspects of liver regeneration, we extracted a general workflow on how modelling can inform liver disease pathogenesis [82]. Liver has a complex micro-architecture ensuring its function (axis 5). Hence many clinical questions require quantitative characterisation of micro-architecture in normal liver and during degeneration or regeneration processes which has been performed using the software TiQuant ([91] see software) or other tools we generated. Disease specific and personal information can be used to build a list of hypotheses on the question of interest, which then can be systematically implemented in mathematical models and in simulation runs tested against the data. As confocal micrographs only display part of lobules, statistically representative lobules were constructed to permit definition of boundary conditions for flow and transport.

In order to compare data and model results quantitatively, quantitative measures characterising the processes under study have to be defined, and measured in both experiment and model. Models in a context where micro-architecture is important were based on agent-based models representing each hepatocyte as well as blood vessels explicitly. They were parameterised by measurable parameters as for those physiologically relevant ranges can be identified, and systematic simulated parameter sensitivity analyses can be performed. Movement of each cell was then mimicked by an equation of motion, describing the change of position as a function of all forces on that cell including active migration. If the best model disagreed with the data, the underlying hypotheses were considered incomplete or wrong and were modified or complemented. Models quantitatively reproducing data were either used to predict so far unknown situations, or the key mechanisms were directly challenged by our experimental partners. Along this line, two unrecognised mechanisms could be identified (axis 5).

Statistical methods decide on subsequently validated mechanism of ammonia detoxification

To identify the mechanisms involved in ammonia detoxification [92], 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.

Collaborations

- **Philippe Robert**, Inria Rap, for the stochastic process modelling [90]
- **Marc Hoffmann,** Université Paris-Dauphine, for the statistical approach to growth and division processes [80], **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 [57],[1]
- **Jan G. Henstler group**, IfADo, Dortmund (Germany), **Irene Vignon-Clementel** (Inria, REO), others for Liver regeneration, ammonia detoxification.
- **Kai Breuhahn group**, DKFZ Heidelberg (Germany), **Pierre Nassoy**, Univ. Bordeaux, for multicellular tumor growth

3.5. Research axis 4: Focus on cancer

Personnel

Luis Almeida, Thibault Bourgeron, Cécile Carrère, Rebecca Chisholm, Jean Clairambault, Marie Doumic, Dirk Drasdo, Sarah Eugène, Paul Van Liedekerke, Tommaso Lorenzi, Alexander Lorz, Benoît Perthame, Yi Yin

Project-team positioning

The MAMBA team designs and analyses mathematical models of tumour 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, modelling 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 antagonising 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 analysing models of cell populations structured in continuous phenotypes, relevant for the description of the behaviour 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 modelling 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 tumour 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

Molecular modelling towards theoretical optimisation of anticancer drug delivery
The protein p53, guardian of the genome and tumour suppressor, has been the object of Ján Eliaš's PhD thesis [85], defended in September 2015, and of articles in 2014 and 2017 [83], [84], [11]. Based on an original intracellular spatial PDE model of the protein dynamics, it allows for the prediction of biologically observed oscillations of p53 nuclear concentrations in case of (e.g. radiotherapy- or anticancer drug-induced) damage to the DNA. In parallel, in [75], that for us concluded works initiated by a fruitful collaboration with Francis Lévi (retired from CNRS 2014), we associate pharmacokinetics-pharmacodynamics of anticancer drugs, their action on the cell cycle at the cell population level, and optimisation algorithms to maximise their combined action under the constraint of preserving healthy tissue integrity.

Modelling Acute Myeloid Leukaemia (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 [54], [55], [56], [64], [76], [53]. These works study the stability of the haematopoietic system and its possible restabilisation 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 modelling with a evolutionary perspective on tumour heterogeneity, is documented in a series of articles [73], [74], [105], [106], [108]. 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)

Senescence modelling 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 [66].

Biomechanically mediated growth control of cancer cells

Mechanical feedback has been identified as a key regulator of tissue growth, by which external signals are transduced into a complex intracellular molecular machinery. Using multi-scale computational modelling of multicellular growth in two largely different experimental settings with the same tumour cell line we were able to show that the cellular growth response on external mechanical stress is surprisingly quantitatively predictable. For this purpose, the mechanical parameters of a center-based agent-based model were calibrated with a deformable agent-based cell model, which displays cell shape and hence can deal with high cell compressions. The cell cycle progression function was calibrated with findings of population growth in an elastic capsule. The emerging model was able to correctly predict the growth response both for modified stresses in a capsule as well as the growth response in a different experimental setting [128], [127].

Model identification for TRAIL treatment

Repetitive administration of TRAIL (TNF-Related Apoptosis Induced-Ligand) on HeLa cells produces characteristic resistance pattern in time that can be explained by cell-to-cell variability in the protein composition. The TRAIL signal transduction pathway is one of the best-studied apoptosis pathways and hence permits detailed comparisons with data. Within a stochastic model of gene expression coupled to transcription and translation to the pathway members, we were able to quantitatively explain the resistance pattern. An important challenge was in parameter identification at each of the level for numerous proteins, whereby the most sensitive parameter was to correctly capture short-lived proteins in the TRAIL toxicity pathway as those mainly determine the regeneration of protein distribution in the cell population and thereby may generate strong stochastic fluctuations [61], [60].

Radiotherapy

In close cooperation with M. Herrero (U. Complutense, Madrid) we have explored by extensive computer simulations using an agent-based model the consequences of spatially inhomogeneous x-ray irradiation in cancer treatment. The model predicted that in the case of different competing sub-populations, namely cancer stem cells with unlimited division capacity, and cancer cells with limited division capacity, inhomogeneous radiation focusing higher doses at the tumour center and lower doses at the tumour periphery should outperform homogeneous irradiation [104]. Cancer stem cells are believed to have a longer cell cycle duration than cancer cells, and are less radiosensitive than cancer cells, which is why they often survive radiation and lead to tumour relapse.

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 (2 articles in common [74], [108]), **Michèle Sabbah** (2 PhD theses in common) at Annette Larsen's lab, 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: one funded Inria PRE project in common.
- Telomere shortening: **Teresa Teixeira** and **Zhou Xu** (IBCP, Paris), **Philippe Robert** (Inria RAP).
- TRAIL treatment: **Gregory Batt,** Inria Saclay and Inst. Pasteur (France)

3.6. Research axis 5: Growth, evolution and regeneration in populations and tissues

Personnel

Luis Almeida, Pierre-Alexandre Bliman, Marie Doumic, Dirk Drasdo, Benoît Perthame, Nicolas Vauchelet

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. 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 modelling 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 (TU Dortmund), and P. Segers group (Leuven). Different from these, Mamba has a strong focus on spatial-temporal modelling on the histological scale, integration of molecular processes in each individual cell, and single-cell (agent) based models. Works by Schliess [124], [92] have been highlighted in editorials.

Mathematical modelling 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 (UPMC) and Monique Chyba (Hawaii, USA) in control, and Suzanne Sindi (USA) for the modelling of the yeast prion. We have interactions with all these groups and organised 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 [69], 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 [78], [90], Sarah Eugène's Ph.D subject (co-supervised by Philippe Robert) [89].

In collaboration with Tom Banks first [58], [57] and then Philippe Moireau, we developed quantitative comparisons between model and data. Through data assimilation and statistical methods [52], we proposed new models and mechanisms.

Dengue fever
The spread of certain strains of the intracellular parasitic bacterium Wolbachia in populations of mosquitoes Aedes aegypti drastically reduces their competence as vector of dengue and other severe mosquito-borne viral diseases. In the absence of a vaccine, or of any preventive or curative treatment, the release of mosquitoes deliberately infected in laboratory by this bacterium has been recently considered a promising tool to control these diseases. Technically the situation can be described by a bistable model, and the issue consists in moving from a Wolbachia-free equilibrium to a fully contaminated equilibrium.

When implementing such a method, an important issue concerns the spatial propagation of the mosquitoes: on releasing infected mosquitoes in a given domain (which can be part of a city), the hope is to invade the whole area. The study of this propagation phenomena falls into the study of existence of travelling waves.

Wound healing
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.

Liver regeneration
An integrated model, coupling a spatial-temporal model of liver regeneration after drug-induced damage to a compartment model of detoxification blood from ammonia, identified the lack of an ammonia detoxifying reaction in the biochemical consensus scheme [124]. Hyperammonia is the most frequent reason for death due to acute liver failure in UK and USA. The spatial model represents liver micro-architecture in a group of liver lobules, the repetitive anatomical and functional units of liver, mimicking each hepatocyte as single agent and blood vessels as a network of chains of spherical objects. This model had previously predicted the subsequently validated orientation of dividing hepatocytes along the liver capillaries as order mechanism. It was here coupled to ODEs for metabolites participating in the zonated ammonia metabolism by calculating the volume of each liver lobule zone with time during regeneration after drug induced damage, which is an input parameter for the detoxification compartment model. Experiments triggered by the model predictions could identify later a candidate ammonia sink mechanism which in a follow-up work [92] could be shown to be the most likely mechanism compared with alternative explanations (see axis 3). This mechanism could be validated, and led to a possible therapy option in treatment of hyperammonemia.
The models have been further expanded towards true multilevel-multiscale models that include molecular HGF control of cell cycle progression (unpublished) and ammonia detoxification (Géraldine Cellière's PhD thesis, 2016; Noémie Boissier's PhD thesis, 2018). In these models, the intracellular models were executed in each individual hepatocyte, and transport of molecules with blood were simulated. Blood flow was modelled by Poiseuille law in the entire capillary network. Further conditions could be identified, under which standard pharmacokinetics-pharmacodynamics (PKPD) models fail to predict the correct dynamics and need to be replaced by spatial temporal models representing organ microarchitecture. The model has further been extended towards bile flow.

[0] ravasio:hal-01245750, vedula:hal-01298859

Toxicity extrapolation from in vitro to in vivo
In vivo toxicity prediction from in vitro data is a major objective in toxicology as it permits bypassing animal experiments. The multilevel-multi scale approach outlined above has been used to explore a strategy to predict the in vivo damage of paracetamol (acetaminophen) from in vitro experimental data. Model simulations and data obtained so far strongly suggest that the prediction is quantitative, if the time development of the toxicity in vitro is displayed (this is so far not common), differences in the concentration kinetics of drug metabolising enzymes in vitro are measured, and micro-architecture is determined (Géraldine Cellière's PhD thesis [71]). Common strategies in toxicology based on relating the maximum drug concentration or area under the drug concentration - time curve between in vitro and in vivo damage could be shown to fail.

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 [123], 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.

Collaborations

- Dengue control by releasing Wolbachia infected mosquitoes **Maria Soleda Aronna, F.C. Coelho** (Fundação Getulio Vargas, Brazil); **D. Villela, C. Struchiner** (Fiocruz, Brazil); **Jorge Zubelli** (IMPA, Brazil); **Alain Rapaport** (INRA-Montpellier), **Y. Dumont** (CIRAD-Montpellier); **Ch. Schaerer** (UNA, Paraguay).
- 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), **Philippe Moireau** (M3DISIM) and **Philippe Robert** (Rap) in Inria
- bacterial growth and division: **Lydia Robert**, UPMC (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).

4. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Awards

Benoît Perthame has been elected member of the Académie des Sciences, in the section "Physique, mécanique, informatique".

4.1.2. Personnel

Marie Doumic has prolonged for one more year her sabbatical at WPI (Vienna, Austria, 2016-2018).

Diane Peurichard has been hired as Chargée de Recherche classe normale in Mamba, beginning in October 2017.

5. New Software and Platforms

5.1. 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

5.2. 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

5.3. Platforms

TiQuant and TiSim The software for tissue image analysis (**Ti**ssue **Quant**ifier) and simulation (**Ti**ssue **Sim**ulator) has been enriched. In more details,

5.3.1. TiQuant

TiQuant [94], [91] is implemented in portable object-oriented JSO C++. The GUI is based on QT and supports real-time visualisation using OpenGL. TiQuant is embedded in the tissue modelling framework CellSys and thus is tightly linked with TiSim, a versatile and efficient simulation environment for tissue models. TiQuant provides an interface to VolView and further complements its functionality by linking to the open-source libraries ITK and VTK (itk/vtk.org). The image/volume processing chains currently implemented in TiQuant for example include techniques to segment conduit and cell segmentation from 3D confocal micrographs of liver tissue based on the Adaptive Otsu Thresholding method and a number of morphological operators. TiQuant was currently extended by a machine-learning component, largely replacing the manual image-processing pipeline.

5.3.2. TiSim

TiSim permits agent-based simulations of multicellular systems. It is modular, in object-oriented ISO C++, the GUI based on Qt and OpenGL, while also allowing for batch mode runs. The software permits multi-scale simulations by integration of molecular pathways (for signalling, metabolisms, drug) into each individual cell. Applications so far are monolayer growth, multicellular spheroids, liver regeneration, TRAIL-treatment simulations. It has an SBML interface. In a largely finished follow-up version it will integrate a deformable cell model by triangulation of cell surface, deformable rod models, extracellular matrix and vascular flow and transport. TiSim can be directly fed by structures synthesised from processed image data from TiQuant.

Impact: The tool is used by our collaborators in liver biology, medicine and toxicology. We recently trained a PhD student from P. Segers (Ghent Univ.) on TiQuant and from T. Hillen (Univ. Alberta, Ca) on TiSim and organised a workshop on benchmarking and comparing agent-based models and tools (workshop Leipzig, volet 5).

6. New Results

6.1. Analysis and control for population dynamics

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 [33].

Cell population dynamics and its control

The question of optimal control of the population dynamics, that naturally arises when dealing with anticancer drug delivery optimisation, has been specifically the object of [24], work led in common with E. Trélat (LJLL and Inria team CAGE) and published in the *J. Maths. Pures Appl*.

The asymptotic behaviour of interacting populations in a nonlocal Lotka-Volterra way is also, independently of any control, studied for two populations in this article, and for many in [49].

Mathematical models of infectious diseases

First results in this subject (which is new for the team) have been obtained for elementary models including a model of vector-borne disease [31], [29].

6.2. Reaction and motion equations for living systems

Mathematical modelling for chemotaxis

A new kinetic model of chemotaxis for angiogenesis has been developed [22].

Aggregation equation.

Based on the approach relying on weak measure-valued solutions [100], an extension to a model for two species in interaction has been proposed in [12].

Free boundary problems for tumour growth.

Motivated by numerical observations from D. Drasdo using agent-based modelling, the article [17] studies the interfaces between two cell populations described by continuous models with different motilities and recovers interface instabilities.

6.3. Model and parameter identification combining stochastic and deterministic approaches in nonlocal and multi-scale models

Data assimilation and stochastic modelling for protein aggregation

Following Carola Kruse's post-doc [57], in collaboration with Tom Banks, Aurora Armiento's Ph.D [1], co-supervised with Philippe Moireau, was devoted to the question of adapting data assimilation strategies to the specific context and difficulties of protein aggregation.

In parallel with the statistical approach to growth and division processes, the deterministic approach has been continued in collaboration with Magali Tournus [35].

Estimating cellularity and tumour heterogeneity from Diffusion-Weighted MRI based on histological data

In [25] we developed, in close collaboration with the University of Heidelberg and DKFZ, together with I. Vignon-Clementel (Inria team REO), a procedure to estimate tumour heterogeneity and cellularity from Diffusion-Weighted Imaging (DWI) with calibration using histological data. The estimate is based on the intravoxel incoherent motion (IVIM) model that relates the DWI signal to water diffusion within each image voxel, as well as on an image processing and analysis procedure we developed for automated cell counting in large histological samples after tumour removal. We recently showed that biopsies routinely taken are likely to be sufficient to construct a calibration curve to relate DWI diffusion coefficient to cell density, and thus to infer the whole tumour heterogeneity. The biopsies have to be taken in regions of largely different diffusion values.

6.4. Focus on cancer

Modelling Acute Myeloid Leukaemia (AML) and its control by anticancer drugs by PDEs and Delay Differential equations

The collaboration with the DISCO team at Inria-Saclay has been continued in conference papers [26], [27]. In one of these papers, the concept of *dormancy* in cancer as a state of coexistence between tumour and healthy stem cell populations is studied using a new model.

Adaptive dynamics setting to model and circumvent evolution towards drug resistance in cancer by optimal control

This topic, main subject in Camille Pouchol's ongoing PhD thesis, has already been mentioned about Axis 1. It has led to the publication [24].

The general question of drug resistance in cancer, from biological observations to mathematical modelling and optimal control, has been reviewed in [14], [15] and presented in various international conferences and workshops.

Senescence modelling by telomere shortening

This work, following Sarah Eugène's PhD thesis, has been continued in collaboration with Zhou Xu at IBPC [13].

6.5. Growth, evolution and regeneration in populations and tissues

Amyloid disease

With Wei-Feng Xue in Canterbury, we continued to investigate the intrinsic variability among identical experiments of nucleation [78], [90], with recent results in [13].

Making use of data assimilation and statistical methods [52], we proposed new models and mechanisms and most recently we predicted the existence of several coexisting species of protein fibrils [2].

Dengue fever
The release of Wolbachia-infected mosquitoes in Dengue infested zones and the study of their propagation may be represented by spatial reaction-diffusion models. When implementing such a method, an important issue concerns the spatial propagation of the mosquitoes: on releasing infected mosquitoes in a given domain (which can be part of a city), the hope is to invade the whole area. The study of this propagation phenomena falls into the study of existence of travelling waves. We proposed in [125] a mathematical model to study such phenomena and have simplified it to recover a well-known simple bistable system for which existence of traveling wave is known. The study of the probability of success of spatial invasiveness has been performed in [126], and [41] is devoted to the blocking of the propagation in heterogeneous environment presenting strong enough population gradient. In the previous works, the invasion is installed by large enough impulsive deliveries. Another approach, consisting in igniting the propagation by feedback control, has been studied in [63], [6].

Toxicity extrapolation from in vitro to in vivo
The investigation of this field has been continued by Géraldine Cellière, leading to her PhD defense in June 2017 [71].

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

Industrial contract with SANOFI on the modelling of employees population dynamics and turnover.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

8.1.1.1. 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.

8.1.1.2. ANR 2014-2017 IFLOW

Eric Vibert, Hopital Paul Brousse (coordinator). Partners: Inria REO, Hopital Toulouse, Dirk Drasdo. Objectives are simulation of liver perfusion after partial hepatectomy with and without therapeutic manipulations to improve patients survival after PHx.

8.1.1.3. 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.

8.1.1.4. ANR InTelo 2017-2020

Telomere dynamics, headed by Teresa Teixeira (IBPC, Paris).

8.1.2. ITMO Cancer 2016 - 2020, HTE call (heterogeneity of tumours in their ecosystems)

8.1.2.1. ITMO Cancer EcoAML

Early leukaemogenesis in Acute Myelogenous Leukaemia (AML), 8 teams headed by François Delhommeau (CDR St Antoine, Paris).

8.1.2.2. ITMO Cancer MoGlImaging

Treatment-induced treatment resistance and heterogeneity in glioblastoma, 8 teams headed by Elizabeth Moyal (INSERM, Toulouse).

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

Research axis 1 (population dynamics): The ERC Starting Grant SKIPPER^{AD} (Marie Doumic, 2014-2018) supported and was the guideline for the study of nucleation, growth and fragmentation equations.

Benoît Perthame has obtained in April 2017 the ERC Advanced Grant ADORA (Asymptotic approach to spatial and dynamical organisations)

8.2.2. Collaborations with Major European Organisdations

German BMBF: LiverSimulator (Dirk Drasdo, 2014 - 2017)

8.3. International Initiatives

8.3.1. Participation in International Programs

CAPES/COFECUB project “Modelling innovative control methods for dengue fever" (Bliman)
STIC AmSud project “MOSTICAW- MOdelling the Spread and (opTImal) Control of Arboviroses by Wolbachia" (2016-2017) (Bliman)..

ECOS-Nord project “New methods for controlling epidemics of dengue fever and arboviroses" (2017-2019) (Bliman)

(See below)

8.3.1.1. International Initiatives

MOSTICAW

Title: MOdelling the Spread and (opTImal) Control of Arboviroses by Wolbachia

International Partners (Institution - Laboratory - Researcher):

Universidad de Buenos Aires (Argentina) - Hernán G. Solari

Universidad de Chile (Chile) - Carlos Conca

Universidade Federal Fluminense (Brazil) - Max Souza

Duration: 2016 - 2017

Start year: 2016

The spread of certain strains of the intracellular parasitic bacterium Wolbachia in populations of mosquitoes Aedes aegypti drastically reduces their competence as vector of dengue and other severe mosquito-borne viral diseases known as arboviral infections. In absence of vaccine, or of preventive or curative treatment, the release of mosquitoes infected by the bacterium has been recently considered a promising tool to control these diseases, and experimental introductions in wild populations are currently under way in Brazil and Colombia. A key question about this method concerns the effective strategies of release of the infected mosquitoes in the field that can be applied with limited cost to reach the desired state of complete exclusion of Wolbachia-free mosquitoes. The mathematical study of central topics is the core of this project. The scientific questions to be

addressed during this project are related to the study of the dynamic and control of the key invasion mechanism on finite-dimensional compartmental models; and to specific focus on the spatial aspects, achieved through more elaborate models (PDE, models on interaction graphs, stochastic models). We further propose to elaborate on the risks involved in the spreading of Wolbachia, implementing in mathematical models critical analysis, complex systems (R. García) and a complexity aware epistemology (E. Morin) in contrast with the instrumental reason (Horkheimer).

8.3.1.2. International Initiatives

C17M01

Title: New methods for the control of epidemics of dengue and arboviroses

International Partner (Institution - Laboratory - Researcher):

Universidad del Valle (Colombia) - Olga Vasilieva

Duration: 2017 - 2019

Start year: 2017

8.4. International Research Visitors

8.4.1. Internships

September 2016-January 2017: Julie Favre, M1 student at EPFL (Zürich), research internship report [39]

8.4.2. Visits to International Teams

8.4.2.1. Sabbatical programme

Doumic Marie

Date: Sep 2016 - Jul 2018

Institution: Wolfgang Pauli Institute, Vienna (Austria)

8.4.2.2. Research Stays Abroad

P.-A. Bliman is still a professor at Funadação Getulio Vargas, Rio de Janeiro, Brazil, and makes frequent stays there.

9. Dissemination

9.1. Promoting Scientific Activities

Editorial activities, scientific boards, research administration

P.-A. Bliman is member of the Scientific committee of the ANR program “Environnement, pathogènes et maladies émergentes ou ré-émergentes - One health".

J. Clairambault is member of the expert group of ITMO Cancer, representative of Inria (since 2008) and member of the bureau of the interdisciplinary doctoral programme “Interfaces pour le Vivant” (IPV) at Sorbonne Université.

M. Doumic is member of the expert group of ITMO BMSV, representative of Inria (since 2014).

D. Drasdo is head of a research team, until June 2017 co-localised at Interdisciplinary Center for Bioinformatics, Univ. Leipzig, and since July 2017 co-localised at Leibniz Institute for Work-environment IfADo, Dortmund. He is member of the boards of TheScientificWorldJOURNAL and Royal Society open science (UK), J. Theor. Biol. and member of CaSyM expert committee for EU Horizon 2020.

B. Perthame is Chief Editor of Acta ApplicandæMathematicæ(Springer-Nature) (since october 2017), Editor of De Gruyter Series in Mathematics and Life Sciences and of Frontiers in Mathematical Sciences (Birkhaüser), and member of the scientific board for the ECMTB Conference 2018.

9.2. Teaching - Supervision - PhD and HDR defence committees

9.2.1. Teaching

We indicate here only the courses given by scientific staff members who do not have a teaching position. Benoît Perthame gives courses at UPMC.

Luis Ameida is in charge of the Major MathBio of the speciality "Mathematics of modelling", M2 level, UPMC.

- *Luis Almeida, 2017*

  Tissue growth (with Delphine Salort, IBPS). UPMC M2 course, Paris **20 h**

- *Pierre-Alexandre Bliman, 2017*

  Analysis, Graduate cycle, School of Applied Mathematics, Fundação Getulio Vargas, Rio de Janeiro, Brazil **60 h**

  and Calculus III, Graduate cycle, School of Applied Mathematics, Fundação Getulio Vargas, Rio de Janeiro, Brazil **30 h**

  and Control theory, Graduate cycle, School of Industrial Management, Université Mohamed 6 Polytechnique, Ben Guerir, Morocco **6 h**

- *Jean Clairambault, 2017*

  International course on stem cells, UPMC, September 2017 **2 h**

  and Spring school on systems biology, Instituto Gulbenkian de Ciência, Lisbon, May 2017 **6 h**

- *Marie Doumic, 2017*

  Direct and inverse problems in population dynamics (with P. Moireau, Inria M3DISIM). UPMC M2 course, Paris **24 h / yr**

- *Dirk Drasdo, 2017*

  Agent-based models of tissue organisation, UPMC M2 course, Paris **24 h / yr**

  University of Rome: Tutorial, 2h: TiSim: A modelling tool for multicellular simulations

9.2.2. Supervision

9.2.2.1. PhD defences in 2017

- Aurora Armiento, "Inverse problems and data assimilation methods applied on protein polymerisation", UPMC, begun September 2013, supervision by M. Doumic and Ph. Moireau (Inria Saclay, M3DISIM team), PhD defence January 2017, UPMC
- Giulia Fabrini "Numerical methods for optimal control problems with biological applications", supervision by L. Almeida and P. Bagnerini (University of Genova, Italy), PhD defence April 2017, Univ. Genova
- Walid Djema, "Understanding Cell Dynamics in Cancer from Control and Mathematical Biology Standpoints: Particular Insights into the Modelling and Analysis Aspects in Hematopoietic Systems and Leukemia", supervision by C. Bonnet (DISCO, Saclay), J. Clairambault, and F. Mazenc (DISCO, Saclay), PhD defence November 2017, L2S, Gif/Yvette

9.2.2.2. Ongoing PhD theses

- PhD in progress: Noémie Boissier (since November 2013, PhD defence in February 2018), supervision by D. Drasdo and I. Vignon-Clementel
- PhD in progress: Julia Delacour (since September 2017), supervision by Marie Doumic and Christian Schmeiser (WPI, Vienna)
- PhD in progress: Adrian Friebel, "Software of image processing and analysis of liver tissue at histological scales", supervision by D. Drasdo and S. Hoehme

- PhD in progress: Ghassen Haddad, “Optimisation of cancer treatments”, UPMC in co-tutela with ENIT Tunis, begun October 2015, supervision by J. Clairambault and S. Ben Miled (Tunis)
- PhD in progress: Hugo Martin, “New structured population models for bacterial growth", begun October 2016, supervision by M. Doumic in co-tutela with Pierre Gabriel (Versailles)
- PhD in progress: Mathieu Mézache, begun October 2016, “Oscillatory dynamics in protein aggregation", supervision by M. Doumic in co-tutela with Human Rezaei (INRA)
- PhD in progress: Johannes Neitsch, “Growth and regeneration modelling based on an agent-based model with deformable cells”, (since June 2011), supervision by D. Drasdo and P. Van Liedekerke
- PhD in progress: Pastor Pérez-Estigarribia, Universidad Nacional de Asunción, Paraguay, supervision by C. Schaerer and P.-A. Bliman
- PhD in progress: Camille Pouchol, “Modelling interactions between tumour cells and adipocytes in breast cancer”, UPMC, begun September 2015, supervision by J. Clairambault, M. Sabbah, and E. Trélat
- PhD in progress: Antonin Prunet, UPMC, begun October 2014, supervision by L. Almeida and M. Sabbah
- PhD in progress: Andrada Quillas Maran, “Modelling early leukaemogenesis”, UPMC, begun March 2014, supervision by J. Clairambault, F. Delhommeau and B. Perthame
- PhD in progress: Martin Strugarek, “Structured population dynamics for transmissible diseases”, UPMC, begun October 2015, supervision by N. Vauchelet and B. Perthame
- PhD in progress: Cécile Taing, UPMC, begun October 2014, supervision by A. Lorz and B. Perthame

9.2.3. PhD and HDR defence committees

- Luis Almeida: Giulia Fabrini, “Numerical methods for optimal control problems with biological applications”, PhD defence April 26, 2017, Genova (Italy)
- Jean Clairambault: Cécile Carrère, “Prise en compte de l’hétérogénéité tumorale dans l’optimisation d’une chimiothérapie : contrôle optimal, analyse théorique et numérique”, PhD defence October 6, 2017, Marseille
- Jean Clairambault: Sébastien Benzekry (HDR), “Contributions in Mathematical Oncology: When Theory Meets Reality”, HDR defence November 13, 2017, Bordeaux
- Jean Clairambault: Walid Djema, “Understanding Cell Dynamics in Cancer from Control and Mathematical Biology Standpoints: Particular Insights into the Modelling and Analysis Aspects in Hematopoietic Systems and Leukemia”, PhD defence November 21, 2017, Gif/Yvette
- Marie Doumic: Aurora Armiento, “Inverse problems and data assimilation methods applied on protein polymerisation”, UPMC, PhD defence January 13, 2017, UPMC
- Marie Doumic: Apollos Besse, “Modélisation mathématique de la leucémie myéloïde chronique", PhD defence July 6, University Lyon 1
- Marie Doumic: Aline Marguet, “Processus de branchement pour des populations structurées et estimateurs pour la division cellulaire”, PhD defence November 27, 2017, Ecole Polytechnique
- Benoît Perthame: Giulia Fabrini, “Numerical methods for optimal control problems with biological applications”, PhD defence April 26, 2017, Genova (Italy)

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] A. ARMIENTO.*Inverse problems and data assimilation methods applied to protein polymerisation*, Université Paris 7 - Diderot, January 2017, https://hal.inria.fr/tel-01447286.

Articles in International Peer-Reviewed Journal

[2] 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° 7 [*DOI :* 10.1371/JOURNAL.PONE.0180538], https://hal.archives-ouvertes.fr/hal-01574346.

[3] H. T. BANKS, M. DOUMIC-JAUFFRET, C. KRUSE.*A numerical scheme for the early steps of nucleation-aggregation Models*, in "Journal of Mathematical Biology", January 2017, vol. 74, n° 1-2, p. 259-287 [*DOI :* 10.1007/S00285-016-1026-0], https://hal.inria.fr/hal-00954437.

[4] A. A. BHAYA, P.-A. A. BLIMAN, G. NIEDU, F. A. PAZOS.*A cooperative conjugate gradient method for linear systems permitting efficient multi-thread implementation*, in "Computational and Applied Mathematics", 2017, p. 1–28 [*DOI :* 10.1007/S40314-016-0416-7], https://hal.inria.fr/hal-01558765.

[5] 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.

[6] P.-A. BLIMAN, N. VAUCHELET.*Establishing Traveling Wave in Bistable Reaction-Diffusion System by Feedback*, in "IEEE Control Systems Letters", 2017, vol. 1, n° 1, p. 62 - 67 [*DOI :* 10.1109/LCSYS.2017.2703303], https://hal.inria.fr/hal-01558631.

[7] 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.

[8] M. BURGER, A. LORZ, M.-T. WOLFRAM.*Balanced Growth Path Solutions of a Boltzmann Mean Field Game Model for Knowledge Growth*, in "Kinetic and Related Models ", March 2017, https://arxiv.org/abs/1602.01423 [*DOI :* 10.3934/KRM.2017005], https://hal.archives-ouvertes.fr/hal-01267078.

[9] 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, http://hal.upmc.fr/hal-01674142.

[10] M. DOUMIC, B. PERTHAME, E. RIBES, D. SALORT, N. TOUBIANA.*Toward an integrated workforce planning framework using structured equations*, in "European Journal of Operational Research", April 2017, vol. 262, https://arxiv.org/abs/1607.02349 [*DOI :* 10.1016/J.EJOR.2017.03.076], https://hal.inria.fr/hal-01343368.

[11] J. ELIAS.*Positive effect of Mdm2 on p53 expression explains excitability of p53 in response to DNA damage*, in "Journal of Theoretical Biology", April 2017, vol. 418, p. 94-104, 1 year long embargo for free article distribution [*DOI :* 10.1016/J.JTBI.2017.01.038], https://hal.inria.fr/hal-01443268.

[12] C. EMAKO, J. LIAO, N. VAUCHELET.*Synchronising and non-synchronising dynamics for a two-species aggregation model*, in "Discrete and Continuous Dynamical Systems - Series B (DCDS-B)", August 2017, vol. 22, n° 6, p. 2121 - 2146, https://arxiv.org/abs/1505.07659 [*DOI :* 10.3934/DCDSB.2017088], https://hal.archives-ouvertes.fr/hal-01157578.

[13] S. EUGENE, T. BOURGERON, Z. XU.*Effects of initial telomere length distribution on senescence onset and heterogeneity*, in "Journal of Theoretical Biology", January 2017, vol. 413, 8, https://arxiv.org/abs/1606.06842 , https://hal.inria.fr/hal-01378596.

[14] A. GOLDMAN, M. KOHANDEL, J. CLAIRAMBAULT.*Integrating Biological and Mathematical Models to Explain and Overcome Drug Resistance in Cancer, Part 1: Biological Facts and Studies in Drug Resistance*, in "Current Stem Cell Reports", August 2017 [*DOI :* 10.1007/s40778-017-0097-1], https://hal.inria.fr/hal-01558477.

[15] A. GOLDMAN, M. KOHANDEL, J. CLAIRAMBAULT.*Integrating Biological and Mathematical Models to Explain and Overcome Drug Resistance in Cancer, Part 2: From Theoretical Biology to Mathematical Models*, in "Current Stem Cell Reports", August 2017 [*DOI :* 10.1007/s40778-017-0098-0], https://hal.inria.fr/hal-01558479.

[16] C. JOURDANA, P. PIETRA, N. VAUCHELET.*Hybrid coupling of a one-dimensional Energy-Transport Schrödinger system*, in "Monatshefte für Mathematik", December 2017, vol. 184, n^{o} 4, p. 563–596 [*DOI :* 10.1007/s00605-016-1008-8], https://hal.archives-ouvertes.fr/hal-01052415.

[17] T. LORENZI, A. LORZ, B. PERTHAME.*On interfaces between cell populations with different mobilities*, in "Kinetic and Related Models ", March 2017, vol. 10, n^{o} 1, p. 299-311, https://hal.inria.fr/hal-01257180.

[18] A. LORZ, B. PERTHAME, C. TAING.*Dirac concentrations in a chemostat model of adaptive evolution*, in "Chinese Annals of Mathematics - Series B", March 2017, http://hal.upmc.fr/hal-01255449.

[19] A. MELLET, B. PERTHAME, F. QUIROS.*A Hele-Shaw Problem for Tumor Growth*, in "Journal of Functional Analysis", 2017, vol. 273, p. 3061-3093, https://arxiv.org/abs/1512.06995 , http://hal.upmc.fr/hal-01241309.

[20] H. MOUNDOYI, A. MOUSSA, B. PERTHAME, B. SARELS.*Analytical examples of diffusive waves generated by a traveling wave*, in "Applicable Analysis", April 2017 [*DOI :* 10.1080/00036811.2017.1314463], http://hal.upmc.fr/hal-01404972.

[21] 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, https://arxiv.org/abs/1602.06970 [*DOI :* 10.3934/KRM.2017019], https://hal.archives-ouvertes.fr/hal-01274529.

[22] N. OUTADA, N. VAUCHELET, T. AKRID, M. KHALADI.*From Kinetic Theory of Multicellular Systems to Hyperbolic Tissue Equations: Asymptotic Limits and Computing*, in "Mathematical Models and Methods in Applied Sciences", 2017, https://arxiv.org/abs/1610.03290 , https://hal.archives-ouvertes.fr/hal-01378301.

[23] 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, https://arxiv.org/abs/1706.05796 [*DOI :* 10.1016/J.PHYSD.2017.05.005], http://hal.upmc.fr/hal-01541093.

[24] 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", October 2017, https://arxiv.org/abs/1612.04698 [*DOI :* 10.1016/J.MATPUR.2017.10.007], https://hal.archives-ouvertes.fr/hal-01673589.

[25] 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.

International Conferences with Proceedings

[26] W. DJEMA, C. BONNET, J. CLAIRAMBAULT, F. MAZENC, P. HIRSCH, F. DELHOMMEAU.*Analysis of a Model of Dormancy in Cancer as a State of Coexistence Between Tumor and Healthy Stem Cells*, in "ACC 2017 - American Control Conference", Seattle, United States, IEEE, May 2017, p. 5135-5140 [*DOI :* 10.23919/ACC.2017.7963751], https://hal.inria.fr/hal-01677927.

[27] W. DJEMA, H. ÖZBAY, C. BONNET, E. FRIDMAN, F. MAZENC, J. CLAIRAMBAULT.*Analysis of Blood Cell Production under Growth Factors Switching*, in "IFAC 2017 - 20th World Congress of the International Federation of Automatic Control", Toulouse, France, Elsevier, July 2017, vol. 50, n^{o} 1, p. 13312-13317 [*DOI :* 10.1016/J.IFACOL.2017.08.1331], https://hal.inria.fr/hal-01677914.

Scientific Books (or Scientific Book chapters)

[28] F. BERTAUX, D. DRASDO, G. BATT.*System modeling of receptor-induced apoptosis*, in "TRAIL, Fas Ligand, TNF and TLR3 in Cancer", O. MICHEAU (editor), Resistance to Targeted Anti-Cancer Therapeutics, 2017, n^{o} 12, https://arxiv.org/abs/1712.06822 , https://hal.inria.fr/hal-01667015.

[29] P.-A. BLIMAN, B. D 'AVILA BARROS.*Interval Observers for SIR Epidemic Models Subject to Uncertain Seasonality*, in "Lecture Notes in Control and Information Sciences", 2017, vol. 471, 9 [*DOI :* 10.1007/978-3-319-54211-9_3], https://hal.inria.fr/hal-01567474.

Other Publications

[30] L. 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*, October 2017, Conference Biomat 2017 - 17th International Symposium on Mathematical and Computational Biology, https://hal.inria.fr/hal-01618357.

[31] M. S. ARONNA, P.-A. BLIMAN.*Interval observer for uncertain time-varying SIR-SI epidemiological model of vector-borne disease*, March 2017, https://arxiv.org/abs/1703.07083 - working paper or preprint, https://hal.inria.fr/hal-01493078.

[32] P.-A. BLIMAN, N. VAUCHELET.*Establishing Traveling Wave in Bistable Reaction-Diffusion System by Feedback*, May 2017, https://arxiv.org/abs/1703.00672 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01480833.

[33] J. CALVO, M. DOUMIC, B. PERTHAME.*Long-time asymptotics for polymerization models*, July 2017, https://arxiv.org/abs/1707.09777 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01570292.

[34] M. DOUMIC.*Simulation of the critical fragmentation equation with binary fission*, April 2017, Linked to the preprint hal-01510960, https://hal.archives-ouvertes.fr/medihal-01510970.

[35] M. DOUMIC, M. ESCOBEDO, M. TOURNUS.*Estimating the division rate and kernel in the fragmentation equation*, April 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01501811.

[36] M. DOUMIC, M. MEZACHE, B. PERTHAME, E. RIBES, D. SALORT.*Strategic Workforce Planning and sales force : a demographic approach to productivity*, February 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01449812.

[37] M. DOUMIC, B. VAN BRUNT.*Explicit Solution and Fine Asymptotics for a Critical Growth-Fragmentation Equation*, April 2017, https://arxiv.org/abs/1704.06087 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01510960.

[38] J. ELIAS.*Trend to equilibrium for a reaction-diffusion system modelling reversible enzyme reaction*, January 2017, https://arxiv.org/abs/1610.07172 - 20 pages, https://hal.inria.fr/hal-01443266.

[39] J. FAVRE.*Design and identification on biological data of a dedicated model of the interactions between haematopoietic stem/progenitor cells and supporting stroma*, Ecole Polytechnique Fédérale de Lausanne (EPFL) ; Inria Paris Research Centre, MAMBA Team, F-75012, Paris, France ; Sorbonne Universités, UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France ; CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France, March 2017, https://hal.inria.fr/hal-01500920.

[40] C. HENDERSON, B. PERTHAME, P. E. SOUGANIDIS.*Super-linear propagation for a general, local cane toads model*, May 2017, https://arxiv.org/abs/1705.04029 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01520145.

[41] G. NADIN, M. STRUGAREK, N. VAUCHELET.*Hindrances to bistable front propagation: application to Wolbachia invasion*, January 2017, https://arxiv.org/abs/1701.05381 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01442291.

[42] S. NORDMANN, B. PERTHAME, C. TAING.*Dynamics of concentration in a population model structured by age and a phenotypical trait*, March 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01493068.

[43] 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.

[44] B. PERTHAME, E. RIBES, D. SALORT.*Career plans and wage structures: a mean field game approach*, January 2018, working paper or preprint, http://hal.upmc.fr/hal-01674630.

[45] B. PERTHAME, W. SUN, M. TANG.*The fractional diffusion limit of a kinetic model with biochemical pathway*, September 2017, https://arxiv.org/abs/1709.03308 - working paper or preprint, http://hal.upmc.fr/hal-01584754.

[46] 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.

[47] B. PERTHAME, S. YASUDA.*Stiff-response-induced instability for chemotactic bacteria and flux-limited Keller-Segel equation*, January 2018, https://arxiv.org/abs/1703.08386 - working paper or preprint, http://hal.upmc.fr/hal-01494963.

[48] 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, http://hal.upmc.fr/hal-01686461.

[49] C. POUCHOL, E. TRÉLAT.*Global stability with selection in integro-differential Lotka-Volterra systems modelling trait-structured populations*, April 2017, https://arxiv.org/abs/1702.06187 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01470722.

[50] M. STRUGAREK, L. DUFOUR, N. VAUCHELET, L. 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.

References in notes

[51] L. ALMEIDA, C. EMAKO, N. VAUCHELET.*Existence and diffusive limit of a two-species kinetic model of chemotaxis*, in "Kinetic and Related Models ", June 2015 [*DOI :* 10.3934/KRM.2015.8.359], https://hal.archives-ouvertes.fr/hal-00980594.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.

[57] 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.

[58] H. T. BANKS, M. DOUMIC-JAUFFRET, 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.

[59] 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.

[60] F. BERTAUX, S. HOEHME, W. WEENS, B. GRASL-KRAUPP, J. G. HENGSTLER, D. DRASDO.*Model prediction and validation of an order mechanism controlling the spatio-temporal phenotype of early hepatocellular carcinoma*, October 2016, working paper or preprint, https://hal.inria.fr/hal-01426629.

[61] 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.

[62] J. BERTOIN, A. R. WATSON.*Probabilistic aspects of critical growth-fragmentation equations*, in "Advances in Applied Probability", 9 2015.

[63] P.-A. BLIMAN, M. S. ARONNA, F. C. COELHO, M. DA SILVA.*Global stabilizing feedback law for a problem of biological control of mosquito-borne diseases*, in "54th IEEE Conference on Decision and Control", Osaka, Japan, Proc. of the 54th IEEE Conference on Decision and Control, December 2015, https://hal.inria.fr/hal-01261162.

[64] 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.

[65] 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.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] G. CELLIÈRE.*Multi-scale modeling of hepatic drug toxicity and its consequences on ammonia detoxification*, Université Paris 6 - Pierre et Marie Curie, July 2017.

[72] 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.

[73] 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.

[74] R. H. CHISHOLM, T. LORENZI, A. LORZ, A. K. LARSEN, L. N. 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.

[75] J. CLAIRAMBAULT, O. FERCOQ.*Physiologically structured cell population dynamic models with applications to combined drug delivery optimisation in oncology*, in "Mathematical Modelling of Natural Phenomena", 2016, 22, V2 d'un dépôt précédemment effectué sous la référence clairambault:hal-01321536v1 [*DOI :* 10.1051/MMNP/201611604], https://hal.inria.fr/hal-01413791.

[76] 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.

[77] 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.

[78] 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.

[79] 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.

[80] 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.

[81] 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.

[82] 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.

[83] J. ELIAS, J. CLAIRAMBAULT.*Reaction-diffusion systems for spatio-temporal intracellular protein networks: a beginner's guide with two examples*, in "Computational and structural biotechnology journal", June 2014, 11 [*DOI :* 10.1016/J.CSBJ.2014.05.007], https://hal.inria.fr/hal-00957344.

[84] J. ELIAS, L. DIMITRIO, J. CLAIRAMBAULT, R. NATALINI.*Dynamics of p53 in single cells: physiologically based ODE and reaction-diffusion PDE models*, in "Physical Biology", July 2014, 22, Phys. Biol. 11 (2014) 045001 [*DOI :* 10.1088/1478-3975/11/4/045001], https://hal.inria.fr/hal-00859412.

[85] J. ELIAS.*Mathematical model of the role and temporal dynamics of protein p53 after drug-induced DNA damage*, Université Pierre et Marie Curie - Paris VI, September 2015, https://tel.archives-ouvertes.fr/tel-01237604.

[86] C. EMAKO, C. GAYRARD, A. BUGUIN, L. NEVES DE ALMEIDA, N. VAUCHELET.*Traveling Pulses for a Two-Species Chemotaxis Model*, in "PLoS Computational Biology", April 2016, vol. 12, n^{o} 4, e1004843 [*DOI :* 10.1371/JOURNAL.PCBI.1004843], https://hal.archives-ouvertes.fr/hal-01302632.

[87] C. EMAKO, M. TANG.*Well-balanced and asymptotic preserving schemes for kinetic models* , March 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01265029.

[88] C. EMAKO KAZIANOU.*Study of two-species chemotaxis models*, Université Pierre et Marie Curie - Paris VI, March 2016, https://tel.archives-ouvertes.fr/tel-01365414.

[89] 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.

[90] S. EUGENE, W.-F. XUE, P. ROBERT, M. DOUMIC-JAUFFRET.*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.

[91] A. FRIEBEL, J. NEITSCH, T. JOHANN, S. HAMMAD, D. DRASDO, S. HOEHME.*TiQuant: software for tissue analysis, quantification and surface reconstruction*, in "Bioinformatics", June 2015, vol. 31, n^{o} 19, p. 3234-3236 [*DOI :* 10.1093/BIOINFORMATICS/BTV346], https://hal.inria.fr/hal-01257137.

[92] 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.

[93] 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.

[94] S. HAMMAD, S. HOEHME, A. FRIEBEL, I. VON RECKLINGHAUSEN, A. OTHMAN, B. BEGHER-TIBBE, R. REIF, P. GODOY, T. JOHANN, A. VARTAK, K. GOLKA, P. O. BUCUR, E. VIBERT, R. MARCHAN, B.

CHRIST, S. DOOLEY, C. MEYER, I. ILKAVETS, U. DAHMEN, O. DIRSCH, J. BÖTTGER, R. GEBHARDT, D. DRASDO, J. G. HENGSTLER.*Protocols for staining of bile canalicular and sinusoidal networks of human, mouse and pig livers, three-dimensional reconstruction and quantification of tissue microarchitecture by image processing and analysis.*, in "Archives of Toxicology", May 2014, vol. 88, n° 5, p. 1161-1183 [*DOI :* 10.1007/S00204-014-1243-5], https://hal.inria.fr/hal-01110657.

[95] 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.

[96] 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.

[97] N. JAGIELLA.*Parameterization of Lattice-Based Tumor Models from Data.*, Université Pierre et Marie Curie - Paris VI, September 2012, https://tel.archives-ouvertes.fr/tel-00779981.

[98] 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° 2, e1004412 [*DOI :* 10.1371/JOURNAL.PCBI.1004412], http://hal.upmc.fr/hal-01244593.

[99] F. JAMES, N. VAUCHELET.*Numerical methods for one-dimensional aggregation equations*, in "SIAM Journal on Numerical Analysis", 2015, vol. 53, n° 2, p. 895-916 [*DOI :* 10.1137/140959997], https://hal.archives-ouvertes.fr/hal-00955971.

[100] 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° 3, p. 1355-1382, https://hal.archives-ouvertes.fr/hal-00803709.

[101] M.-J. KANG, B. PERTHAME, D. SALORT.*Dynamics of time elapsed inhomogeneous neuron network model*, in "Comptes Rendus Mathématique", September 2015, n° 353, p. 1111-1115 [*DOI :* 10.1016/J.CRMA.2015.09.029], http://hal.upmc.fr/hal-01241300.

[102] 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.

[103] 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° 5, p. 871-882, 13 pages, https://hal.archives-ouvertes.fr/hal-00939013.

[104] J. C. LOPEZ ALFONSO, N. JAGIELLA, L. NÚÑEZ, M. HERRERO, D. DRASDO.*Estimating Dose Painting Effects in Radiotherapy: AMathematical Model*, in "PLoS ONE", February 2014, vol. 9, n° 2, 22 [*DOI :* 10.1371/JOURNAL.PONE.0089380], https://hal.inria.fr/hal-01109675.

[105] 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° 1, 43 [*DOI :* 10.1186/S13062-016-0143-4], https://hal.inria.fr/hal-01355314.

[106] 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.

[107] 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.

[108] 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.

[109] 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.

[110] A. MELLET, B. PERTHAME, F. QUIROS.*A Hele-Shaw Problem for Tumor Growth*, December 2015, working paper or preprint, http://hal.upmc.fr/hal-01241309.

[111] A. OLIVIER.*Statistical analysis of growth-fragmentation models*, Université Paris Dauphine - Paris IX, November 2015, https://hal.archives-ouvertes.fr/tel-01235239.

[112] 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.

[113] B. PERTHAME.*Transport equations in biology*, Frontiers in Mathematics, Birkhäuser Verlag, Basel, 2007, x+198.

[114] 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.

[115] 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.

[116] 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.

[117] 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.

[118] 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.

[119] B. PERTHAME, M. TANG, N. VAUCHELET.*Derivation of the bacterial run-and-tumble kinetic equation from a model with biochemical pathway*, in "Journal of Mathematical Biology", 2016, http://hal.upmc.fr/hal-01131101.

[120] 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.

[121] 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.

[122] 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.

[123] 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.

[124] 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.

[125] M. STRUGAREK, N. VAUCHELET.*Reduction to a single closed equation for 2 by 2 reaction-diffusion systems of Lotka-Volterra type*, in "SIAM Journal on Applied Mathematics", 2016, vol. 76, n^{o} 5, p. 2060-2080, https://hal.archives-ouvertes.fr/hal-01264980.

[126] M. STRUGAREK, N. VAUCHELET, J. ZUBELLI.*Quantifying the Survival Uncertainty of Wolbachia-infected Mosquitoes in a Spatial Model* *, August 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01355118.

[127] 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.

[128] 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.

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 (SPDE, SDE)
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 biology
B4.3.4. - Solar Energy
B5.3. - Nanotechnology
B5.5. - Materials
B9.4.2. - Mathematics
B9.4.3. - Physics
B9.4.4. - Chemistry

1. Personnel

Research Scientists

Claude Le Bris [Team leader, École Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Sébastien Boyaval [École Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Éric Cancès [École Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Virginie Ehrlacher [École Nationale des Ponts et Chaussées, Researcher]
Frédéric Legoll [École Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Tony Lelièvre [École Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Antoine Levitt [Inria, Researcher]
Gabriel Stoltz [École Nationale des Ponts et Chaussées, Senior Researcher, HDR]

PhD Students

Grégoire Ferré [École Nationale des Ponts et Chaussées]
Marc Josien [Ministère de la Transition écologique et solidaire]
Adrien Lesage [École Nationale des Ponts et Chaussées, from Oct 2017]
Boris Nectoux [École Nationale des Ponts et Chaussées and Inria, until Nov 2017]
Mouad Ramil [École Nationale des Ponts et Chaussées, from Oct 2017]
Pierre-Loïk Rothé [École Nationale des Ponts et Chaussées]

Julien Roussel [École Nationale des Ponts et Chaussées]
Laura Silva Lopes [École Nationale des Ponts et Chaussées]
Sami Siraj-Dine [Université Paris-Est, from Oct 2017]
Pierre Terrier [Ministère de la Transition écologique et solidaire]

Post-Doctoral Fellows

Dena Kazerani [Inria, from Oct 2017]
Pierre Monmarché [Inria, until Aug 2017]

Administrative Assistants

Kevin Bonny [Inria, from Jun 2017 until Sep 2017]
Sabrine Boumizy [Inria, from Oct 2017]
Sarah Le [Inria, from Jan 2017 until May 2017]

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

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. Such problems are described in the following sections.

4. Application Domains

4.1. Electronic structure of large systems

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.

4.2. 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.*7×10^{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?"

4.3. 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.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

E. Cancès was awarded the 2017 Dargelos Prize from the Alumni of Ecole Polytechnique.

6. New Software and Platforms

6.1. simol

KEYWORDS: Molecular simulation - Quantum chemistry - Statistical physics - C++ - OpenMP
FUNCTIONAL DESCRIPTION: Molecular simulation software written in C++

- Contact: Gabriel Stoltz

7. New Results

7.1. Electronic structure calculations

Participants: 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. We also insist on rigorously studying current materials of technological interest.

7.1.1. Mathematical analysis

In [42], E. Cancès and N. Mourad performed a detailed study of the extended Kohn-Sham model for atoms subjected to cylindrically-symmetric external potentials. In particular, they computed the occupied and unoccupied energy levels of all the atoms of the first four rows of the periodic table for the reduced Hartree-Fock (rHF) and the extended Kohn-Sham Xα models. These results allowed them to test numerically the assumptions on the negative spectra of atomic rHF and Kohn-Sham Hamiltonians used in their previous theoretical works on density functional perturbation theory and pseudopotentials. Interestingly, they observed accidental degeneracies between s and d shells or between p and d shells at the Fermi level of some atoms.

7.1.2. Numerical analysis

E. Cancès has pursued his long-term collaboration with Y. Maday (UPMC) on the numerical analysis of linear and nonlinear eigenvalue problems. Together with G. Dusson (UMPC), B. Stamm (UMPC), and M. Vohralík (Inria SERENA), they have designed a posteriori error estimates for conforming numerical approximations of the Laplace eigenvalue problem with a homogeneous Dirichlet boundary condition [15]. In particular, upper and lower bounds for any simple eigenvalue are given. These bounds are guaranteed, fully computable, and converge with the optimal speed to the exact eigenvalue. In [41], this analysis is extended to all standard numerical methods, including nonconforming discontinuous Galerkin, and mixed finite element approximations or arbitrary polynomial degree.

It is often claimed that error cancellation plays an essential role in quantum chemistry and first-principle simulation for condensed matter physics and materials science. Indeed, while the energy of a large, or even medium-size, molecular system cannot be estimated numerically within chemical accuracy (typically 1 kcal/mol or 1 mHa), it is considered that the energy difference between two configurations of the same system can be computed in practice within the desired accuracy. In [14], E. Cancès and G. Dusson initiated the quantitative study of discretization error cancellation. Discretization error is the error component due to the fact that the model used in the calculation (e.g. Kohn-Sham LDA) must be discretized in a finite basis set to be solved by a computer. They first reported comprehensive numerical simulations showing that errors on energy differences are indeed significantly smaller than errors on energies, but that these two quantities asymptotically converge at the same rate when the energy cut-off goes to infinity. They then analyzed a simple one-dimensional periodic Schrödinger equation with Dirac potentials, for which analytic solutions are available. This allowed them to explain the discretization error cancellation phenomenon on this test case with quantitative mathematical arguments.

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 the speed of convergence for a number of schemes in the metallic case, and 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, after the work [16] with E. Cancès, G. Panati (Rome) and G. Stoltz was published. With H. Cornean (Aalborg), D. Gontier (Dauphine) and D. Monaco (Rome), they introduced a mathematical definition of Wannier functions for metals, used routinely in materials science but not studied theoretically until now. They proved that, under generic assumptions, there exists a set of localized Wannier functions that span a given set of bands, even if this set is not isolated from the others [50]. With A. Damle (Cornell) and L. Lin (Berkeley), they proposed an efficient numerical method for the computation of maximally-localized Wannier functions in metals, and showed on the example of the free electron gas that they are not in general exponentially localized. 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.

7.1.3. New materials

As an external collaborator of the MURI project on 2D materials (PI: M. Luskin), E. Cancès has collaborated with P. Cazeaux (Kansas) and M. Luskin (University of Minnesota) on the computation of the electronic and optical properties of multilayer 2D materials. In particular, they have adapted the C^*-algebra framework for aperiodic solids introduced by J. Bellissard and collaborators, to the case of tight-binding models of incommensurate (and possibly disordered) multilayer systems [13].

The optimal design of new crystalline materials to achieve targeted electronic properties is a very important issue, in particular for photovoltaic applications. In the context of a collaboration with IRDEP, A. Bakhta (CERMICS), V. Ehrlacher and D. Gontier (Dauphine) studied the following inverse problem in [37]: given desired functions defined over the Brillouin zone of a crystalline structure, is it possible to compute a periodic potential so that the first bands of the associated periodic Schrödinger operator are as close as possible to these functions? Theoretical results were obtained for the corresponding variational problem in one dimension for the first band, and it appears from the mathematical analysis that the potential has to belong to a Borel measure space. In addition, a numerical method has been developped to solve the resulting optimization problem where the different discretization parameters are adjusted throughout the calculation, which leads to significant computational gains.

7.2. Computational Statistical Physics

Participants: Grégoire Ferré, Frédéric Legoll, Tony Lelièvre, Pierre Monmarché, Boris Nectoux, 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 of materials, using concepts of statistical physics (thermodynamic ensembles and molecular dynamics). The contributions of the team can be divided into four main topics: (i) the computation of thermodynamic quantities by sampling the canonical measure; (ii) the sampling of the stationary measure of non-equilibrium systems (namely non-reversible dynamics); (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.

7.2.1. Sampling of the canonical measure, free energy calculations and adaptive biasing techniques

The work by T. Lelièvre and G. Stoltz, together with G. Fort (Toulouse) and B. Jourdain (CERMICS), on the study of a dynamics similar to the well-tempered metadynamics has been published [19]. This dynamics can be seen as an extension of the so-called self-healing umbrella sampling method, with a partial biasing of the dynamics only. In particular, the authors proposed a version which leads to much shorter exit times from metastable states (accelerated well-tempered metadynamics).

In [29], T. Lelièvre, in collaboration with C. Chipot (Nancy), T. Zhao, H. Fu, X. Shao, and W. Cai (Nankai University) proposed a new version of the adaptive biasing force (ABF) technique, which is well suited for the computation of free energy landscapes in high dimensions. In addition, V. Ehrlacher, T. Lelièvre and P. Monmarché are currently developping a tensorized version of the ABF algorithms. As in the usual ABF algorithm, the objective is still to compute in an adaptive way (through MCMC computations) the free energy A of a molecular system, which is a function of given reaction coordinates. To keep in memory an approximation of A requires a numerical grid of size m^d where d is the number of reaction coordinates and m is the number of points in a 1-d grid. This prevents d to be larger than 4. To allow for larger number of reaction coordinates, A is approximated as a sum of tensor products of functions of only one variable which only requires a memory of size Nmd, where N is the number of tensor products used in the approximation.

In [53], G. Stoltz and E. Vanden-Eijnden (Courant Institute) have studied the properties of the temperature accelerated molecular dynamics method. This dynamics provides a way to compute the free energy. It consists in introducing an extended variable into the system, coupled to the chosen reaction coordinate, and evolving at a higher temperature in order to alleviate metastable behavior, while the dynamics of the system at lower temperature is accelerated. G. Stoltz and E. Vanden-Eijnden proved in particular that the law of the dynamics converges exponentially fast to the steady-state, with a rate which is dictated by the Poincaré inequality of the effective dynamics on the free energy surface at higher temperature. This work was performed while E. Vanden-Eijnden was spending two months as an Inria invited professor in the project-team.

7.2.2. Sampling of out-of-equilibrium dynamics

Together with A. Iacobucci and S. Olla (Univ. Dauphine), G. Stoltz studied in [20] the convergence to the steady-state of nonequilibrium Langevin dynamics, by a perturbative approach based on hypocoercive techniques developed for equilibrium Langevin dynamics. The Hamiltonian and overdamped limits (corresponding respectively to frictions going to zero or infinity) were carefully investigated. In particular, the maximal magnitude of admissible perturbations are quantified as a function of the friction. Numerical results based on a Galerkin discretization of the generator of the dynamics confirmed the relevance of the theoretical lower bounds on the spectral gap.

J. Roussel and G. Stoltz have proven the consistency of the Galerkin method for hypocoercive operators in [52]. This method allows to solve Poisson problems related to the Fokker-Planck equation very efficiently for small-dimensional systems, even if the dynamics is hypocoercive, as is the case for the Langevin dynamics for example. J. Roussel and G. Stoltz showed in particular the exponential convergence of the semigroup associated with the projected generator and provide error estimates for the solution of the numerical method, under assumptions that are proven to hold for a toy model. The authors illustrated these results by numerical experiments. In addition, an ongoing work 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 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.

G. Ferré is currently working on sampling problems and rare event estimates, in particular with nonequilibrium methods. During this year, he focused on a range of methods related to the estimation of rare event probabilities, mostly based on Feynman-Kac semigroups. These processes correspond to stochastic differential equations whose trajectories are weighted, which is a form of importance sampling. This project resulted in a work on the discretization of such processes (error estimates on ergodic properties, with G. Stoltz), and led to the study of adaptive techniques, with H. Touchette (Stellenbosch). These two works will lead to publications in a close future. This research also raises questions on the long-time stability of Feynman-Kac semigroups, an issue partially covered by the litterature. G. Ferré is currently addressing this subject with G. Stoltz and M. Rousset (Inria Rennes). Other long-term projects are ongoing: one on exclusion processes with M. Simon (Inria Lille), and one on random matrices and Coulomb Gases with D. Chafai (Dauphine).

7.2.3. 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. We 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.

To analyze accelerated dynamics algorithms (and in particular the Temperature Accelerated Dynamics algorithm), one needs to show that the exit event from a metastable state for the Langevin or overdamped Langevin dynamics can be approximated by a kMC model parameterized by the Eyring-Kramers laws. In [45], G. Di Gesu, T. Lelièvre and B. Nectoux, together with D. Le Peutrec (Université de Paris Saclay), used the quasi-stationary distribution approach in order to justify the use of kinetic Monte Carlo models parameterized by the Eyring-Kramers formulas to describe exit events from metastable states. The proof is based on tools from semi-classical analysis.

Concerning the AMS technique, two recent contributions showed the interest of this approach in different applicative fields. In [51], L. Silva Lopes and T. Lelièvre analyzed the performance of the AMS method for biological systems on a simple test case: the alanine dipeptide. The interest of the method was demonstrated on this simple example: it enables to compute transition rates, to sample transition paths, and to compute reactive fluxes between two metastable states. In [26], T. Lelièvre in collaboration with H. Louvin (CEA), E. Dumonteil (IRSN), M. Rousset (Inria Rennes) and C.M. Diop (CEA) implemented the AMS method in the framework of nuclear safety. The idea was to use the AMS method to compute neutron fluxes in strongly absorbing media, for shielding applications. The method has been implemented in Tripoli 4, and gives very interesting results compared to the classical exponential biasing approach, in particular for neutron branching processes.

7.2.4. Coarse-graining

In [25], F. Legoll and T. Lelièvre, in collaboration with S. Olla (Dauphine), analyzed the error introduced when deriving an effective dynamics for a stochastic process in large dimension on a few degrees of freedom using a projection approach *à la Zwanzig*. More precisely, a pathwise error estimate was obtained, which is an improvement compared to a previous result by F. Legoll and T. Lelièvre where only the marginal in times were considered. This analysis is also useful to obtain quantitative estimate for some averaging procedure on two-scale dynamics.

G. Stoltz developed new numerical methods to stabilize the time discretization of generalizations of Langevin dynamics, more precisely dissipative particle dynamics with energy conservation (DPDE) and smoothed dissipative particle dynamics (SDPD). The latter case was studied with a PhD student, Gérôme Faure (CEA/DAM and CERMICS). These two models describe mesoscopic systems of particles with two global invariants: energy and momentum. The numerical schemes are obtained as the composition of a Verlet integration of the deterministic part of the dynamics, and successive integration of the pairwise fluctuation-dissipation dynamics. These elementary dynamics are the one which need to be stabilized because too large timesteps can lead to negative internal energies of the particles. The idea of the methods is to rewrite the elementary 8-dimensional fluctuation-dissipation dynamics as effective reversible one-dimensional dynamics on the relative velocities, which can then be Metropolized; see [27] for DPDE and [18] for SDPD.

In [28], a joint work with Manuel Athènes, Thomas Jourdan (CEA/Saclay SRMP) and Gilles Adjanor (EDF R&D, MMC), G. Stoltz and P. Terrier presented a coupling algorithm for cluster dynamics. Rate equation cluster dynamics (RECD) is a mean field technique where only defect concentrations are considered. It consists in solving a large set of ODEs (one equation per cluster type) governing the evolution of the concentrations. Since clusters might contain up to million of atoms or defects, the number of equations becomes very large. Therefore solving such a system of ODEs becomes computationally prohibitive as the cluster sizes increase. Efficient deterministic simulations propose an approximation of the equations for large clusters by a single Fokker-Planck equations. Nevertheless this approach is still limited by the number of equations to solve in the case of complex materials. Fully stochastic simulations see the RECD as a master equation, hence reducing

the number of equations to solve to the number of stochastic particles, but are limited by the high frequency of certain events. The proposed algorithm is based on a splitting of the dynamics and combines deterministic and stochastic approaches. It is generic (allowing different stochastic approaches such as a jump process or a Langevin dynamics based on the Fokker-Planck approximation) and is highly parallelizable. The accuracy of this new algorithm is illustrated in a case of vacancy clustering of materials under thermal ageing. Numerical analysis of the algorithm shows that the errors due to the splitting (a standard Lie-Trotter splitting) and due to the stochastic approaches decrease according to the theory, *i.e.* respectively linearly with the time step and as $N^{-1/2}$, N being the number of stochastic particles. The error due to the Fokker-Planck approximation is currently under study.

7.3. Homogenization

Participants: Virginie Ehrlacher, Marc Josien, Claude Le Bris, Frédéric Legoll, Adrien Lesage, Pierre-Loïk Rothé.

7.3.1. Deterministic non-periodic systems

In homogenization theory, members of the project-team have pursued their systematic study of perturbations of periodic problems (by local and nonlocal defects). This has been done in two different directions. For linear elliptic equations, they have first, in collaboration with X. Blanc (Paris Diderot) and P-L. Lions (Collège de France), provided a more versatile proof on local defects, and also extended their analysis to advection-diffusion equations. Second, they have also provided more details on the quality of approximation achieved by their theory. These are works in preparation with X. Blanc and M. Josien (Matherials). On the other hand, they have approached the same perturbation problem but for nonlinear equations. The specific case considered is that of viscosity solutions of Hamilton-Jacobi equations, and the work has been completed in collaboration with Pierre Cardaliaguet (Paris Dauphine) and Panagiotis Souganidis (University of Chicago), see [43]. To the best knowledge of the authors, this is the first time such a perturbation has been studied for this type of nonlinear equations.

7.3.2. 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.

In addition, a question of interest is to describe how the oscillatory solution u_ϵ fluctuates around its effective behavior (which is given by the homogenized limit u^*). This question is investigated in the PhD thesis of P.-L. Rothé. Results have been obtained for a weakly stochastic framework (with a periodic coefficient and a small random perturbation). It has been shown that, at the first order, the fluctuations are at the scale $\epsilon^{-\frac{d}{2}}$. Furthermore when ϵ is small, the localized fluctuations (characterized by a test function g) of u_ϵ are Gaussian. The corresponding variance depends on the localization function g and on a fourth order tensor Q. A numerical approach has been designed to approximate Q and its convergence has been proven. Numerical experiments in more general settings (full stochastic case) following the same approach have been performed. The results are promising.

7.3.3. 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).

The MsFEM has been introduced almost 20 years ago. However, even in simple deterministic cases, there are still some open questions, for instance concerning multiscale advection-diffusion equations. Such problems are possibly advection dominated and a stabilization procedure is therefore required. How stabilization interplays with the multiscale character of the equation is an unsolved mathematical question worth considering for numerical purposes.

During the year, the final writing of the various works performed in the context of the PhD thesis of F. Madiot has been completed. The comparison of the various MsFEM approaches has been documented in [24]. The case of an advection-diffusion equation with a dominating convection in a perforated domain is completely studied in [47]. For the latter equation, the approach based on the introduction of the invariant measure has been described, tested and studied in [48].

One of the perspectives of the team, through the thesis of A. Lesage, is the development of a multiscale finite element method for thin heterogeneous plates. The fact that one of the dimension of the domain of interest scales typically like the typical size of the heterogeneities within the material induces theoretical and practical difficulties that have to be carefully taken into account.

7.3.4. Dislocations

In the context of the PhD thesis of M. Josien, some results have been obtained regarding the modeling and numerical simulation of dislocations. Plastic properties of crystals are due to dislocations, which are thus objects of paramount importance in materials science. The geometrical shape of dislocations may be described by (possibly time-dependent) nonlinear integro-differential equations (e.g. the Weertman equation and the dynamical Peierls-Nabarro equation), involving non-local operators.

In collaboration with C. Le Bris, F. Legoll and Y.-P. Pellegrini (CEA-DAM), M. Josien has first focused on the steady state regime (the Weertman equation), and has designed a numerical method for approximating its solution. This relies on a preconditioned scheme based on a dynamical system that integrates differently the linear nonlocal terms (by means of the Fourier transform) and the nonlinear local terms. The numerical scheme is described in [21]. M. Josien has mathematically studied the Weertman equation. In particular, under physically relevant hypotheses, it has been shown in [46] that the equation is the long-term limit of a dynamical system, namely exactly that which has been used for the numerical approximation. The time-dependent regime of a dislocation involves an integrodifferential equation with memory kernel (the so-called Dynamic Peierls-Nabarro equation). M. Josien is currently working on possible numerical approaches to solve it, and is writing a code that is intended to be used in some simple physical test cases. A special effort is devoted to the memory aspect of this equation, using techniques designed by Ch. Lubich and collaborators.

7.4. 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 multi-phase flowing materials, like liquid suspensions of particles or stratified air-water flows, and
- to propose efficient algorithms for the computation of flow solutions, mainly for the many applications in the hydraulic engineering context.

The analysis of heterogeneous flow models for the paradigmatic complex fluids of Maxwell type has been pursued [38], [34], in particular for gravity flows with a free surface (natural in the hydraulic engineering context). It is planned to pursue the analysis with other fluids, and obtain thereby mathematically-sound models for the erosion of sediment. Dena Kazerani has recently started working on that goal, in the context of the ongoing ANR JCJC project SEDIFLO of S. Boyaval with E. Audusse (Paris 13), A. Caboussat (Genève), A. Lemaitre (ENPC) and M. Parisot (Inria ANGE).

Even for Newtonian fluids like water, the simpler models that are currently used do not always produce satisfactory numerical results in the hydraulic engineering context, especially because the data that is used to perform numerical predictions is uncertain. Considering that some model uncertainties induce (stochastic) parametric variations like material heterogeneities, S. Boyaval pursued his analysis of new fast algorithms to compute many PDE solutions for many parameter values in the (uncertain) hydraulic engineering context [30], [54].

7.5. Various topics

Participant: Virginie Ehrlacher.

In the context of a collaboration with EDF, V. Ehrlacher, together with A. Benaceur, A. Ern (CERMICS) and S. Meunier (EDF) has developed in [35] a new reduced basis methodology for parabolic nonlinear systems of equations which enables to significantly reduce the computational time of the offline phase of the method.

V. Ehrlacher, with T. Boiveau, A. Ern (CERMICS) and A. Nouy (Centrale Nantes), has developed a new global space-time unconditionally stable approximation scheme for linear parabolic equations, which relies on the Lions-Magenes formulation of such partial differential equations, in [39]. Such a formulation is perfectly adapted for the use of tensor methods to approximate the solution of these equations at a significantly lower computational cost, based on the separation of space and time variables. Different greedy algorithms to compute this tensor approximation of the solution are compared on numerical testcases using several formulations including the new proposed one. The new approach enables to define a provably convergent algorithm with better approximation properties than the other methods.

8. Bilateral Contracts and Grants with Industry

8.1. Contracts and grants with Industry

Many research activities of the project-team are conducted in close collaboration with private or public companies: CEA, SANOFI, IRDEP, EDF, IFPEN. 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.

9. Partnerships and Cooperations

9.1. 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.
- G. Stoltz is the PI of the ANR project COSMOS (2014-2018) 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.
- E. Cancès is a member of the ANR project BECASIM (2013-2017), PI: I. Danaila (Université de Rouen). This project is concerned with the numerical simulation of Bose-Einstein condensates.
- F. Legoll is a member of the ANR project CINE-PARA (2015-2019), PI: Y. Maday, UPMC. This project is concerned with parallel-in-time algorithms.

Members of the project-team are participating in the following GdR:

- CORREL (correlated methods in electronic structure computations),
- EGRIN (gravity flows),
- MANU (MAthematics for NUclear applications)
- MASCOT-NUM (stochastic methods for the analysis of numerical codes),
- MEPHY (multiphase flows)
- DYNQUA (time evolution of quantum systems, with applications to transport problems, nonequilibrium systems, etc.),
- REST (theoretical spectroscopy),
- CHOCOLAS (experimental and numerical study of shock waves).

The project-team is involved in two Labex, namely the Labex Bezout (started in 2011) and the Labex MMCD (started in 2012).

9.2. European Initiatives

The ERC consolidator Grant MSMATH (ERC Grant Agreement number 614492, PI T. Lelièvre) is running (it started in June 2014).

9.3. International Initiatives

The *Germaine de Staël* grant awarded to S. Boyaval (from CampusFrance Hubert-Curien program) has been used in 2017 to pursue the collaboration with A. Caboussat (Lausanne) about 3D numerical simulations of free-surface flows.

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, Université de Lyon and Université Aix-Marseille. The LIA is renewed for 4 years, starting January 1st, 2018.

10. Dissemination

10.1. 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-), Communications in Mathematical Sciences (2011-), SIAM Multiscale Modeling and Simulation (2012-), and the Journal of Computational Mathematics (2017-),
- was a member of the executive committee of the CEA-EDF-Inria schools in applied mathematics and computer science (2010-July 2017),
- is co-organizing the IMA Long Program on Multiscale Mathematics and Computing in Science and Engineering, 2017-2018.

V. Ehrlacher

- is a member of the "Conseil d'Enseignement et de Recherche" of Ecole des Ponts,
- has co-organized the Oberwolfach workshop on "Applications of Optimal Transportation in the Natural Sciences", January 2017 (with J.-D. Benamou and D. Matthes),
- has co-organized a minisymposium on "Numerical methods for electronic structure calculations" at the SIAM CSE conference, February 2017 (with B. Stamm, L. Lin and C. Yang),
- has co-organized the IPAM workshop on "Uncertainty Quantification for Stochastic Systems and Applications", November 2017 (with M. Katsoulakis, T. Lelièvre, P. Plechac, A. Stuart and D. Trinkle).

G. Ferré and J. Roussel co-organize the working group J-PSI (Jeunes chercheurs en physique statistique et interactions) at IHP, which aims at stimulating interactions between PhD students and post-docs coming from different institutions in Paris and working on the analysis of models in statistical physics.

C. Le Bris is editor-in-chief of Applied Mathematics Research Express (2003-2017). He 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-) and Journal de Mathématiques Pures et Appliquées (2009-).

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,
- the "Comité d'experts" for the Fondation de Recherche pour l'Aéronautique et l'Espace,
- the "International Scientific Advisory Committee" of the Centre de Recherche Mathématique, Université de Montréal,
- 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.

He is the president of the strategic committee of the Institut des Sciences du calcul et des données, Sorbonne Universités.

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-).

T. Lelièvre

- is editor-in-chief of ESAIM: Proceedings and Surveys (with D. Chafai, C. Imbert and P. Lafitte),
- is a member of the "Conseil d'Administration" of SMAI and École des Ponts,
- has co-organized the Journées EDP-Probas at Institut Henri Poincaré (with F. Malrieu),
- has co-organized the IPAM Long Program on "Complex High-Dimensional Energy Landscapes", September 11th - December 15th 2017 (with C. Clementi, G. Henkelman, R. Hennig, M. Luskin, N. Marom, P. Plechac and C. Schuette),
- has co-organized the ICTS program on "Large deviation theory in statistical physics: Recent advances and future challenges", August 14th - October 13th 2017 (with A. Ayyer, F. den Hollander, A. Dhar, J.P. Garrahan, C. Jarzynski, M. Krishnapur, S. Sabhapandit and H. Touchette),
- has co-organized with Florent Malrieu and Pierre-André Zitt the workshop "Piecewise Deterministic Markov Processes and sampling", January 25-27th, 2017,
- has co-organized with C. Chipot and G. Stoltz the "Rencontre Math-Industrie simulation moléculaire dans l'industrie pharmaceutique", at IHP on 28th April 2017,
- has co-organized with A. Jentzen the Stochastic Computation Workshop at FoCM 2017, Barcelona, July 10th-12th, 2017.

G. Stoltz

- is a member of the scientific council of UNIT (Université Numérique Ingénierie et Technologie),
- has co-organized the IHP trimester "Stochastic Dynamics Out of Equilibrium", Spring 2017 (with G. Giacomin, S. Olla, E. Saada and H. Spohn).

10.2. Teaching - Supervision - Juries

The members of the project-team have taught the following courses:

- Licence: Outils mathématiques pour l'ingénieur, 15h, L3, École des Ponts (E. Cancès, V. Ehrlacher, M. Josien, F. Legoll, T. Lelièvre),
- Licence: Analyse et calcul scientifique, 30h, L3, École des Ponts (G. Ferré, A. Levitt, M. Josien, G. Stoltz),
- Licence: Projet de 1ère Année, 6h, L3, École des Ponts (J. Roussel, P. Terrier),
- Licence: Mathématiques, 36h, L1, Université Paris Dauphine (G. Ferré, J. Roussel),
- Licence: Optimisation, 15h, L3, École des Ponts (A. Levitt),
- Licence: Équations aux dérivées partielles et éléments finis, 15h, L3, École des Ponts (F. Legoll, A. Levitt),
- Licence: Méthodes numériques pour les problèmes en grande dimension, 17h30, L3, École des Ponts (V. Ehrlacher, S. Boyaval),
- Licence: Maths 1 et 2, 9h, L3, École des Mines (G. Stoltz),
- Licence: méthodes pour la grande dimension, L3, École des Ponts (V. Ehrlacher 10h, S. Boyaval 5h),
- Licence: hydraulique numérique, 15h, L3, École des Ponts (S. Boyaval),
- Master: Modéliser Programmer Simuler, 28 h, M1, École des Ponts (T. Lelièvre),
- Master: Analyse variationnelle des équations aux dérivées partielles, 32h, École Polytechnique (T. Lelièvre),
- Master: Aléatoire, 32h, École Polytechnique (T. Lelièvre),
- Master: Simulation moléculaire, 6h, UVSQ (T. Lelièvre, G. Stoltz),
- Master: Analyse de Fourier, 15h, M1, École des Ponts (V. Ehrlacher, A. Levitt, G. Stoltz),
- Master: Partial differential equations, 21h, M1, École des Ponts (E. Cancès),
- Master: Control of dynamical systems, 16h, M1, École Polytechnique (E. Cancès),
- Master: Projet du département IMI, 12h, M1, École des Ponts (J. Roussel, G. Ferré),
- Master: Analyse spectrale et application aux Équations aux dérivées partielles, 36h, M1, École des Ponts (F. Legoll, V. Ehrlacher),
- Master: Spectral theory and variational methods, 10h, M2, UPMC (E. Cancès),
- Master: Méthodes de quantification des incertitudes en ingénierie, 18h, M2, École des Ponts (V. Ehrlacher),
- Master: Simulation moléculaire en sciences des matériaux, 6h, M1, École des Ponts (V. Ehrlacher),
- Master: Introduction to computational statistical physics, 20h, M2, UPMC (G. Stoltz),
- Master: Méthodes numériques probabilistes, 36 h, M2, UPMC (T. Lelièvre),
- Master: Problèmes multiéchelles, aspects théoriques et numériques , 20h, M2, UPMC (F. Legoll).

The following Habilitation thesis has been defended in the group at École des Ponts:

- Sébastien Boyaval, Topics in the numerical modelling of flows, Université Paris-Est, defended on December 21 2017.

The following PhD theses have been defended in the group at École des Ponts:

- Athmane Bakhta, Modélisation and simulation for photovoltaic applications, Université Paris-Est, École des Ponts, defended on December 19th, 2017, supervised by E. Cancès and T. Lelièvre, co-supervised by V. Ehrlacher,
- Gérôme Faure, Multiscale methods for the simulation of shock and detonation waves, Université Paris-Est, École des Ponts and CEA/DAM, defended on November 29th 2017, supervised by G. Stoltz and J.-B. Maillet (CEA/DAM),
- Alessandra Iacobucci, Nonequilibrium steady-states of rotor and oscillator chains, defended on October 20th 2017, University Paris Dauphine, supervised by S. Olla (Dauphine) and G. Stoltz,
- Henri Louvin, Development of adaptive variance reduction methods for Monte Carlo particle transport, Ecole Doctorale PHENIICS, defended on October 12th, supervised by Check Diop (CEA) and T. Lelièvre,
- Boris Nectoux, Spectral analysis and semi-classical analysis for metastability in molecular dynamics, Université Paris-Est, École des Ponts, defended on November 20th, supervised by T. Lelièvre and E. Cancès.

The following PhD theses are ongoing in the group at École des Ponts:

- Amina Benaceur, Thèse CIFRE EDF, started January 1st, 2016, supervised by A. Ern, co-supervised by V. Ehrlacher, in collaboration with G. Blatman (EDF) and S. Meunier (EDF),
- 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 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 (UPMC) 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,
- Marc Josien, Multiscale approaches for materials science, started September 1st, 2015, supervised by C. Le Bris,
- 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),
- Julien Roussel, Variance reduction techniques for nonequilibrium systems, Université Paris-Est, École des Ponts, started September 1st, 2015, supervised by G. Stoltz,
- 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, Rare event simulation and applications to biological systems, 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 2017, supervised by E. Cancès, C. Fermanian and co-supervised by A. Levitt,
- Pierre Terrier, Reduced models for defect migration in metals, Université Paris-Est, École des Ponts and CEA Saclay, started September 1st, 2015, supervised by G. Stoltz and M. Athènes (CEA).

Project-team members have participated in the following PhD juries:

- S. Boyaval, PhD of Riad Sanchez (“Application des techniques de bases réduites à la simulation des écoulements poreux”), defended at IFPEN in December 2017,
- E. Cancès, PhD of Geneviève Dusson, defended at UPMC in October 2017,
- E. Cancès, PhD of Chaoyu Quan, defended at UPMC in November 2017,
- V. Ehrlacher, PhD of Xianglong Duang (“Transport optimal et diffusion de courants”), defended at Université Paris-Saclay in September 2017,
- V. Ehrlacher, PhD of Eleonora Musharbash (“Dynamical Low Rank approximation for PDEs with random parameters”), defended at EPFL in May 2017,
- V. Ehrlacher, PhD of Julien Ricaud (“Symétrie et brisure de symétrie pour certains problèmes non linéaires”), defended at Université de Cergy-Pontoise in June 2017,
- V. Ehrlacher, PhD of Pierre-Éric Allier (“Contrôle d’erreur pour et par les modèles réduits PGD”), defended at ENS Paris-Saclay in November 2017,
- V. Ehrlacher, PhD of Quentin Ayoul-Guilmard (“Méthodes numériques pour la prise en compte de défauts aléatoires en mise en forme de composites quasi-périodiques”), defended at Ecole Centrale Nantes in December 2017,
- T. Lelièvre, referee for the PhD of Arthur Talpaert (“Simulation numérique directe de bulles sur maillage adaptatif avec algorithmes distribuées”, defended at Ecole Polytechnique in February 2017,
- T. Lelièvre, PhD of Romain Poncet (“Méthodes numériques pour la simulation d’équations aux dérivées partielles stochastiques non-linéaires en condensation de Bose-Einstein”, defended at Ecole Polytechnique in October 2017,
- T. Lelièvre, president of the jury for the PhD of Manon Baudel (“Théorie spectrale pour des applications de Poincaré aléatoires”), defended at Université d’Orléans in December 2017,
- T. Lelièvre, referee for the PhD of Riad Sanchez (“Application des techniques de bases réduites à la simulation des écoulements en milieux poreux”), defended at Université Paris Saclay in December 2017,
- G. Stoltz, referee for the PhD of Romain Poncet (“Méthodes numériques pour la simulation d’éequations aux dérivées partielles stochastiques non-linéaires en condensation de Bose-Einstein”), defended at École Polytechnique in October 2017,
- G. Stoltz, referee for the PhD of Viviana Letizia (“Modèles microscopiques pour la loi de Fourier”), defended at Université Paris Dauphine in December 2017.

Project-team members have participated in the following habilitation jury:

- T. Lelièvre, HDR of Fabio Pietrucci (“Inventing general simulation methods to study the transformations of matter”), defended at UPMC on December 1st 2017.

10.3. Conference participation

Members of the project-team have delivered lectures in the following seminars, workshops and international conferences:

- S. Boyaval, Finite-Element for Flows, Roma, April 2017,
- S. Boyaval, RWTH AICS colloquim, Aachen, November 2017,
- S. Boyaval, weekly seminar of Collège de France, Paris, December 2017,
- E. Cancès, workshop on New trends in Mathematical Physics at the interface of Analysis and Probability, University College London, England, February 2017,
- E. Cancès, SIAM conference on Scientific Computing, Atlanta, Georgia, February 2017,
- E. Cancès, Mathematical Physics seminar, Université Paris Dauphine, March 2017,

- E. Cancès, weekly seminar of the Mathematics department, University of Metz, March 2017,
- E. Cancès, workshop on Wavelet and Tensor Methods for Partial Differential Equations, Berlin, May 2017,
- E. Cancès, IMA workshop on Mathematical Modeling of 2D Materials, Minneapolis, Minnesota, May 2017,
- E. Cancès, weekly seminar of the Mathematics department, Peking University, Beijing, China, June 2017,
- E. Cancès, workshop on Focus activity on quantum and kinetic problems, Beijing, China, June 2017,
- E. Cancès, BIRS workshop on Mathematical and Numerical Methods for Time-Dependent Quantum Mechanics from Dynamics to Quantum Information, Oaxaca, Mexico, August 2017,
- E. Cancès, workshop on Advances in mathematical modelling and numerical simulation of superfluids, University of Rouen, August 2017,
- E. Cancès, Colloquium lecture, University of Kansas, Lawrence, Kansas, September 2017,
- E. Cancès, workshop on Multiscale Theory and Computation, Minneapolis, Minnesota, September 2017,
- E. Cancès, MOANSI workshop, Aachen, Germany, October 2017,
- V. Ehrlacher, Demi-journée d'échange Labex Bézout/EADS, Marne-la-Vallée, October 2017,
- V. Ehrlacher, Seminar Institut für Numerische Simulation, Bonn, Allemagne, November 2017,
- V. Ehrlacher, IPAM workshop on "Uncertainty Quantification for Stochastic Systems and Applications", Los Angeles, California, November 2017,
- V. Ehrlacher, MORTECH 2017 (keynote lecture), Sevilla, Spain, November 2017,
- V. Ehrlacher, Oberwolfach workshop on "Multiscale and High-Dimensional Problems", Oberwolfach, Germany, April 2017,
- V. Ehrlacher, Conference in honor of Y. Maday's 60th birthday, Roscoff, May 2017,
- V. Ehrlacher, SIAM CSE conference, Atlanta, Georgia, February 2017,
- G. Ferré, Young researchers' seminar, IHP semester, "Stochastic dynamics out of equilibrium", Paris, June 2017,
- G. Ferré, Large deviation theory in statistical physics, ICTS, Bengalore, September 2017,
- M. Josien, CAMP Seminar, University of Chicago, April 2017,
- M. Josien, SciCADE Conference, Bath, September 2017,
- M. Josien, Séminaire de Physique Mathématique-EDP, Institut de Mathématiques de Bordeaux, December 2017,
- D. Kazerani, weekly seminar, Orléans, October 2017,
- D. Kazerani, Post-doc days of IHES, Orsay, October 2017,
- D. Kazerani, weekly seminar IRD, Paris, October 2017,
- C. Le Bris, Conference in honor of Yvon Maday's 60th birthday, May 2017,
- C. Le Bris, Conference in honor of Patrick Joly's 60th birthday, August 2017,
- C. Le Bris, Workshop HPC, Institut d'Etudes Scientifiques de Cargèse, September 2017,
- C. Le Bris, Multiscale Modeling, Theory, and Computation, Conference in honor of Mitchell Luskin's 65th birthday, Minneapolis, September 2017,
- C. Le Bris, Homogenization Theory and Applications, Weierstrass Institute Berlin, October 2017,
- C. Le Bris, BIRS Workshop on "Computational Uncertainty Quantification", Banff International Research Station (BIRS), Canada, October 2017,

- C. Le Bris, Séminaire d'Automatique du plateau de Saclay, June 2017,
- C. Le Bris, Forum Teratec, July 2017,
- F. Legoll, Workshop stochastic homogenization, Bonn, February 2017,
- F. Legoll, UNECECOMP Conference, Rhodes, June 2017,
- F. Legoll, ADMOS Conference, Verbania, June 2017,
- F. Legoll, CIMPA Summer school on multiscale methods, Lucknow, India, July 2017,
- F. Legoll, USNCCM Conference, Montreal, July 2017,
- F. Legoll, COMPLAS 2017 Conference, Barcelona, September 2017,
- F. Legoll, Scicade conference, Bath, September 2017,
- F. Legoll, IMA program on multiscale mathematics, Minneapolis, September 2017,
- F. Legoll, séminaire Université de Genève, October 2017,
- F. Legoll, MORTECH 2017 conference, Sevilla, November 2017,
- F. Legoll, IPAM Program, Los Angeles, November, 2017,
- T. Lelièvre, workshop on Multiscale methods for stochastic dynamics, Geneva, February 2017,
- T. Lelièvre, Séminaire du Laboratoire de Chimie Physique, Université Paris-Sud, March 2017,
- T. Lelièvre, CECAM workshop "Exploiting finite-size effects in simulations", UPMC, April 2017,
- T. Lelièvre, CIRM workshop "interactions EDP/probabilités - équations cinétiques, temps long et propagation du chaos", Marseille, April 2017,
- T. Lelièvre, Colloquium Lorrain de Mathématiques, Université de Nancy, April 2017,
- T. Lelièvre, IHP trimester on Stochastic Dynamics Out of Equilibrium, Paris, April 2017,
- T. Lelièvre, CECAM workshop "Beyond Kd's: New computational methods to address challenges in drug discovery", EPFL, Lausanne June 2017,
- T. Lelièvre, Séminaire de probabilités, ENS Lyon, June 2017,
- T. Lelièvre, "Multiscale Theory and Computation Conference", University of Minneapolis, September 2017,
- T. Lelièvre, "Quasistationary Distributions: Analysis and Simulation", University of Paderborn, September 2017,
- T. Lelièvre, Colloquium du laboratoire Dieudonné, Université Nice Sophia Antipolis, October 2017,
- T. Lelièvre, "Workshop Stochastic Sampling and Accelerated Time Dynamics on Multidimensional Surfaces", IPAM, Los Angeles, October 2017,
- T. Lelièvre, Workshop "Bridging Scales in Molecular Biology", Mathematics & Physical Sciences conference of the Simons Foundation, New York, November 2017,
- T. Lelièvre, workshop "Mathématiques pour la neutronique", GDR MANU, Paris, November 2017,
- T. Lelièvre, Mathematisches Kolloquium RWTH Aachen University, Aachen, December 2017,
- A. Levitt, Young researchers working group, UPMC, January 2017,
- A. Levitt, Chemistry colloquium, Cornell, New York, February 2017,
- A. Levitt, THEOS seminar, Cornell, New York, February 2017,
- A. Levitt, SIAM CSE conference, Atlanta, Georgia, February 2017,
- A. Levitt, Conference in honor of Y. Maday's 60th birthday, Roscoff, May 2017,
- A. Levitt, Scalable solvers group seminar, Lawrence Berkeley National Lab, California, June 2017,
- A. Levitt, CCP17, Paris, July 2017,
- A. Levitt, Density Functional Theory and Beyond, Warwick, July 2017,
- A. Levitt, Mathematical physics summer school, Zurich, July 2017,

- A. Levitt, ICJ seminar, Lyon, November 2017,
- P. Monmarché, Groupe de travail Prob., Théo. Erg. et Systèmes Dynamiques, LMRS, Rouen, January 2017,
- P. Monmarché, Workshop PDMP et sampling, ENPC, Marne-la-Vallée, January 2017,
- P. Monmarché, Conférence PDE/Probability Interactions: Kinetic Equations, CIRM, Marseille, April 2017,
- P. Monmarché, Seminar of the Department of Statistics, University of Oxford, May 2017,
- P. Monmarché, Groupe de travail de probabilités, Université Paris 5, May 2017,
- P. Monmarché, Trimestre IHP dynamiques hors équilibre, Institut Henri Poincaré, Paris, June 2017,
- B. Nectoux, Worskhop "Interactions EDP/probabilités : équations cinétiques, temps long et propagation du chaos", CIRM, April 2017,
- B. Nectoux, SciCADE, university of Bath, UK, Septembre 11-15, 2017,
- B. Nectoux, Workshop "Quasi-stationary distribution: analysis and simulation", Paderborn, September 2017,
- P.-L. Rothé, SciCADE 2017 Conference, Bath, UK, September 2017,
- P.-L. Rothé, USNCCM14, 14th U.S. National Congress on Computational Mechanics, Montreal, Canada, July 2017,
- P.-L. Rothé, Congrès SMAI 2017, La Tremblade, June 2017,
- J. Roussel, Young researchers' seminar, IHP semester "Stochastic dynamics out of equilibrium", Paris, June 2017,
- J. Roussel, ICL Seminar, London, November 2017,
- L. Silva Lopes, "Hands-on" Workshop on Enhanced Sampling and Free-Energy Calculation, Urbana-Champaign, Illinois, September 2017,
- G. Stoltz, seminar at Army Research Laboratory, Aberdeen Proving Grounds, February 2017,
- G. Stoltz, seminar at University of Massachussetts, February 2017,
- G. Stoltz, seminar at University of Geneva, March 2017,
- S. Siraj-Dine, Density Functional Theory and Beyond, Warwick, July 2017,
- P. Terrier, The MRS Spring Meeting & Exhibit, Phoenix, April 2017,
- P. Terrier, Séminaire des doctorants du LAMFA, Amiens, December 2017.

Members of the project-team have delivered the following series of lectures:

- E. Cancès, The mathematics of quantum chemistry, 9h, GDR CORREL winter school, Paris, January 2017,
- E. Cancès, Density Functional Theory: Models and numerical methods, 4h, Beijing, China, June 2017,
- E. Cancès, Mathematical aspects of electronic structure theory, 3h, Aussois, France, June 2017,
- E. Cancès, Mathematical structure of quantum mechanics, 3h, Heidelberg, Germany, October 2017,
- T. Lelièvre, Lectures on "Algorithms for computational statistical physics", 3h, ICTS, Bangalore, August 2017,
- T. Lelièvre, Tutorial on "Sampling efficiently metastable dynamics: algorithms and mathematical analysis", 2h, IPAM, Los Angeles, September 2017,
- A. Levitt, Numerical analysis of periodic quantum systems, 2h, Aalborg, Denmark, June 2017,
- G. Stoltz, From a microscopic description of matter to a macroscopic one on a computer: computational statistical physics, 6h, CIMPA Summer School on Multiscale Computational Methods and Error Control, IIT Kanpur, India, July 2017,
- Random homogenization, theoretical and numerical aspects, 6h, CIMPA Summer School on Multiscale Computational Methods and Error Control, IIT Kanpur, India, July 2017.

Members of the project-team have presented posters in the following seminars, workshops and international conferences:

- G. Ferré, Complex high-dimensional energy landscapes, UCLA, Los Angeles, October 2017,
- G. Ferré, Numerical Aspects of Nonequilibrium dynamics, IHP semester "Stochastic dynamics out of equilibrium", Paris, April 2017,
- G. Ferré, Trends and Advances in Monte Carlo Sampling Algorithms, Duke University, Durham (North Carolina), December 2017,
- D. Kazerani, colloque EDP Normandie, Caen, October 2017,
- B. Nectoux, Workshop "Dynamiques stochastiques hors d'équilibre", CIRM, April 2017,
- P.-L. Rothé, colloque EDP Normandie, Caen, October 2017,
- J. Roussel, workshop "Trends and Advances in Monte Carlo Sampling Algorithms", SAMSI (Duke University), December 2017,
- J. Roussel, Numerical Aspects of Nonequilibrium dynamics, IHP semester "Stochastic dynamics out of equilibrium", Paris, April 2017,
- L. Silva Lopes, Beyond Kd's: New computational methods to address challenges in drug discovery, Lausanne, Switzerland, June 2017,
- L. Silva Lopes, CEMRACS 2017: Numerical methods for stochastic models: control, uncertainty quantification, mean-field, Marseille, July, 2017,
- L. Silva Lopes, 11th Triennial Congress of the World Association of Theoretical and Computational Chemistry, Munich, Germany, August 2017,
- L. Silva Lopes, Stochastic Sampling and Accelerated Time Dynamics on Multidimensional Surfaces, Los Angeles, California, October 2017,
- P. Terrier, SMAI 2017, La Tremblade, June 2017.

Pierre Terrier has won the best poster award at SMAI 2017.

Members of the project-team have participated (without giving talks nor presenting posters) in the following seminars, workshops and international conferences:

- M. Josien, colloque EDP Normandie, Caen, October 2017,
- Mouad Ramil, PDE/Probability Interactions: Kinetic Equations, Long time and Propagation of Chaos at CIRM, Marseille, April 2017
- Mouad Ramil, Workshop on Quasi-Stationary distributions, Paderborn, September 2017
- P.-L. Rothé, Winter School on Numerical Analysis of Multiscale Problems, Hausdorff Research Institute for Mathematics, Bonn, Germany, January 2017,
- J. Roussel, CEMRACS, CIRM, July 2017,
- L. Silva Lopes, IPAM Long Program on "Complex High-Dimensional Energy Landscapes", Los Angeles, California, September-November 2017.

10.4. Software development and contributions

- A. Levitt has added methods for optimization on Riemannian manifolds to the Optim.jl optimization library, see https://github.com/JuliaNLSolvers/Optim.jl.
- A. Levitt has published an implementation of the method developed in [16] to construct Wannier functions, see https://github.com/antoine-levitt/wannier.
- In the framework of the PhD of Laura Silva Lopes, L. Silva Lopes and T. Lelièvre have implemented a new tutorial on the NAMD code in order to popularize the Adaptive Multilevel Splitting method among the practitioners, see http://www.ks.uiuc.edu/Training/Tutorials/namd/ams-tutorial/tutorial-AMS.pdf.

- J. Roussel and G. Stoltz have added new features to the Simol code, in particular concerning the use of control variates.

10.5. Popularization

- G. Ferré gave a talk about statistical physics and its applications to undergraduate students at Lycée Pierre Corneille, Rouen, in November 2017.
- A. Levitt is a member of the editorial board of Interstices, Inria's popularization website.
- P. Monmarché gave a talk about mathematics and music to high school students at lycée Pablo Picasso, Avion, in May 2017.
- P. Monmarché participated to Les Matinales de la Recherche de l'ENPC and presented a poster about his work to the students of ENPC in April 2017.
- G. Stoltz participated to Les Matinales de la Recherche de l'ENPC and gave a talk about his work to the staff of ENPC in April 2017.
- G. Stoltz, together with Gilles Buisson (Ecole des Ponts), published a contribution to the proceedings of QPES 2017 (Questions de Pédagogie dans l'Enseignement Supérieur), on his teaching experience involving flipped classrooms organized at the level of a complete class of first year students at Ecole des Ponts. See G. Buisson and G. Stoltz, La classe inversée à grande échelle en école d'ingénieur, Actes du colloque QPES 2017, 633-640.

11. 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

Articles in International Peer-Reviewed Journal

[7] X. ANTOINE, A. LEVITT, Q. TANG.*Efficient spectral computation of the stationary states of rotating Bose-Einstein condensates by the preconditioned nonlinear conjugate gradient method*, in "Journal of Computational Physics", August 2017, vol. 343, p. 92-109, https://arxiv.org/abs/1611.02045 [*DOI :* 10.1016/J.JCP.2017.04.040], https://hal.archives-ouvertes.fr/hal-01393094.

[8] 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, https://arxiv.org/abs/1509.01348 [*DOI :* 10.1007/S40072-017-0105-6], https://hal.archives-ouvertes.fr/hal-01192862.

[9] F. AVIAT, A. LEVITT, B. STAMM, Y. MADAY, P. REN, J. W. PONDER, L. LAGARDERE, J.-P. PIQUEMAL.*Truncated Conjugate Gradient (TCG): an optimal strategy for the analytical evaluation of the many-body polarization energy and forces in molecular simulations*, in "Journal of Chemical Theory and Computation", January 2017, vol. 13, n^{o} 1, p. 180–190 [*DOI :* 10.1021/ACS.JCTC.6B00981], https://hal.archives-ouvertes.fr/hal-01395833.

[10] A. BAKHTA, V. EHRLACHER.*Global existence of bounded weak solutions to degenerate cross-diffusion equations in moving domain*, in "ESAIM: Mathematical Modelling and Numerical Analysis", 2017, https://hal.archives-ouvertes.fr/hal-01238636.

[11] N. BERGLUND, G. DI GESÙ, H. WEBER.*An Eyring–Kramers law for the stochastic Allen–Cahn equation in dimension two*, in "Electronic Journal of Probability", April 2017, vol. 22, n^{o} 41, p. 1-27 [*DOI :* 10.1214/17-EJP60], https://hal.archives-ouvertes.fr/hal-01304559.

[12] X. BLANC, M. JOSIEN.*From the Newton equation to the wave equation : the case of shock waves*, in "AMRX", February 2017, https://arxiv.org/abs/1605.00974v1 [*DOI :* 10.1093/AMRX/ABX001], https://hal.archives-ouvertes.fr/hal-01314690.

[13] E. CANCÈS, P. CAZEAUX, M. LUSKIN.*Generalized Kubo Formulas for the Transport Properties of Incommensurate 2D Atomic Heterostructures*, in "Journal of Mathematical Physics", May 2017, vol. 58, n^{o} 6, p. 1-29 [*DOI :* 10.1063/1.4984041], https://hal.inria.fr/hal-01403588.

[14] E. CANCÈS, G. DUSSON.*Discretization error cancellation in electronic structure calculation: toward a quantitative study*, in "ESAIM: Mathematical Modelling and Numerical Analysis", September 2017, vol. 51, n^{o} 5, p. 1617 - 1636 [*DOI :* 10.1051/M2AN/2017035], http://hal.upmc.fr/hal-01435054.

[15] 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 Journal on Numerical Analysis", September 2017, vol. 55, n^{o} 5, p. 2228-2254 [*DOI :* 10.1137/15M1038633], https://hal.inria.fr/hal-01194364.

[16] É. CANCÈS, A. LEVITT, G. PANATI, G. STOLTZ.*Robust determination of maximally-localized Wannier functions*, in "Physical Review B : Condensed matter and materials physics", February 2017, vol. 95, https://arxiv.org/abs/1605.07201 [*DOI :* 10.1103/PHYSREVB.95.075114], https://hal.archives-ouvertes.fr/hal-01323700.

[17] V. EHRLACHER, D. LOMBARDI.*A dynamical adaptive tensor method for the Vlasov-Poisson system*, in "Journal of Computational Physics", 2017, vol. 319, p. 285-306, https://hal.archives-ouvertes.fr/hal-01335507.

[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. FORT, B. JOURDAIN, T. LELIÈVRE, G. STOLTZ.*Self-Healing Umbrella Sampling: Convergence and efficiency*, in "Statistics and Computing", January 2017, vol. 27, n^{o} 1, p. 147–168, https://arxiv.org/abs/1410.2109 [*DOI :* 10.1007/S11222-015-9613-2], https://hal.archives-ouvertes.fr/hal-01073201.

[20] A. IACOBUCCI, S. OLLA, G. STOLTZ.*Convergence rates for nonequilibrium Langevin dynamics*, in "Annales mathématiques du Quebec", October 2017, https://arxiv.org/abs/1702.03685 [*DOI :* 10.1007/S40316-017-0091-0], https://hal.archives-ouvertes.fr/hal-01467131.

[21] 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", 2017, https://arxiv.org/abs/1704.04489 [*DOI :* 10.1002/NME.5723], https://hal.archives-ouvertes.fr/hal-01510158.

[22] C. LE BRIS, F. LEGOLL.*Examples of computational approaches for elliptic, possibly multiscale PDEs with random inputs*, in "Journal of Computational Physics", 2017, vol. 328, p. 455-473, https://arxiv.org/abs/1604.05061 , https://hal.archives-ouvertes.fr/hal-01304040.

[23] 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 , https://hal.archives-ouvertes.fr/hal-01420187.

[24] C. LE BRIS, F. LEGOLL, F. MADIOT.*A numerical comparison of some Multiscale Finite Element approaches for advection-dominated problems in heterogeneous media*, in "ESAIM: Mathematical Modelling and Numerical Analysis", 2017, vol. 51, n^{o} 3, p. 851-888, https://arxiv.org/abs/1511.08453 - DOI: http://dx.doi.org/10.1051/m2an/2016057, https://hal.archives-ouvertes.fr/hal-01235642.

[25] F. LEGOLL, T. LELIÈVRE, S. OLLA.*Pathwise estimates for an effective dynamics*, in "Stochastic Processes and their Applications", 2017, vol. 127, n^{o} 9, p. 2841-2863, https://arxiv.org/abs/1605.02644 , https://hal.archives-ouvertes.fr/hal-01314221.

[26] H. LOUVIN, E. DUMONTEIL, T. LELIEVRE, M. ROUSSET, C. M. DIOP.*Adaptive Multilevel Splitting for Monte Carlo particle transport*, in "EPJ Web of Conferences", 2017, vol. 153, p. 1-9 [*DOI :* 10.1051/EPJCONF/201715306006], https://hal.archives-ouvertes.fr/hal-01661012.

[27] G. STOLTZ.*Stable schemes for dissipative particle dynamics with conserved energy*, in "Journal of Computational Physics", March 2017, vol. 340, p. 451-469, https://arxiv.org/abs/1612.04154 [*DOI :* 10.1016/J.JCP.2017.03.059], https://hal.archives-ouvertes.fr/hal-01416108.

[28] P. TERRIER, M. ATHÈNES, T. JOURDAN, G. ADJANOR, G. STOLTZ.*Cluster dynamics modelling of materials: a new hybrid deterministic/stochastic coupling approach*, in "Journal of Computational Physics", December 2017, vol. 350, p. 280-295, https://arxiv.org/abs/1610.02949 [*DOI :* 10.1016/J.JCP.2017.08.015], https://hal.archives-ouvertes.fr/hal-01378916.

[29] T. ZHAO, H. FU, T. LELIÈVRE, X. SHAO, C. CHIPOT, W. CAI.*The Extended Generalized Adaptive Biasing Force Algorithm for Multidimensional Free-Energy Calculations*, in "Journal of Chemical Theory and Computation", 2017, vol. 13, n° 4, p. 1566 - 1576 [*DOI :* 10.1021/ACS.JCTC.7B00032], https://hal.archives-ouvertes.fr/hal-01587011.

International Conferences with Proceedings

[30] Q. H. TRAN, G. ENCHÉRY, R. SANCHEZ, S. BOYAVAL.*A Reduced-Basis Approach to Two-Phase Flow in Porous Media*, in "FVCA 2017", Lille, France, Finite Volumes for Complex Applications VIII - Hyperbolic, Elliptic and Parabolic Problems, June 2017 [*DOI :* 10.1007/978-3-319-57394-6_50], https://hal-enpc.archives-ouvertes.fr/hal-01587116.

Conferences without Proceedings

[31] L. CHAMOIN, F. LEGOLL.*Certified computations with PGD model reduction in the MsFEM framework*, in "14th National Congress on Computational Mechanics - USNCCM14", Montréal, Canada, July 2017, https://hal.archives-ouvertes.fr/hal-01581791.

[32] F. LEGOLL, L. CHAMOIN.*A posteriori error estimation for multiscale computations based on MsFEM*, in "14th National Congress on Computational Mechanics - USNCCM14", Montréal, Canada, July 2017, https://hal.archives-ouvertes.fr/hal-01581797.

[33] F. LEGOLL, L. CHAMOIN.*Multiscale computations based on MsFEM: model reduction and goal-oriented a posteriori error estimation*, in "8th International Conference on Adaptive Modeling and Simulation - ADMOS 2017", Verbania, Italy, June 2017, https://hal.archives-ouvertes.fr/hal-01583144.

Scientific Books (or Scientific Book chapters)

[34] S. BOYAVAL.*A Finite-Volume discretization of viscoelastic Saint-Venant equations for FENE-P fluids*, in "Finite Volumes for Complex Applications VIII - Hyperbolic, Elliptic and Parabolic Problems. FVCA 2017. Springer Proceedings in Mathematics & Statistics", C. CANCÈS, P. OMNES (editors), Springer, 2017, vol. 200, p. 163-170, https://arxiv.org/abs/1701.04639 [*DOI :* 10.1007/978-3-319-57394-6_18], https://hal-enpc.archives-ouvertes.fr/hal-01433712.

Other Publications

[35] A. BENACEUR, V. EHRLACHER, A. ERN, S. MEUNIER.*A progressive reduced basis/empirical interpolation method for nonlinear parabolic problems* , November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01599304.

[36] A. BAKHTA, V. EHRLACHER.*Cross-diffusion systems with non-zero flux and moving boundary conditions*, September 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01397682.

[37] A. BAKHTA, V. EHRLACHER, D. GONTIER.*Numerical reconstruction of the first band(s) in an inverse Hill's problem*, September 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01591133.

[38] J. W. BARRETT, S. BOYAVAL.*Finite Element Approximation of the FENE-P Model*, April 2017, working paper or preprint, https://hal-enpc.archives-ouvertes.fr/hal-01501197.

[39] T. BOIVEAU, V. EHRLACHER, A. ERN, A. NOUY.*Low-rank approximation of linear parabolic equations by space-time tensor Galerkin methods*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01668316.

[40] S. BOYAVAL.*Derivation and numerical approximation of hyperbolic viscoelastic flow systems: Saint-Venant 2D equations for Maxwell fluids*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01661269.

[41] E. CANCÈS, G. DUSSON, Y. MADAY, B. STAMM, M. VOHRALÍK.*Guaranteed and robust a posteriori bounds for Laplace eigenvalues and eigenvectors: a unified framework*, March 2017, working paper or preprint, https://hal.inria.fr/hal-01483461.

[42] E. CANCÈS, N. MOURAD.*A numerical study of the extended Kohn-Sham ground states of atoms*, February 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01454677.

[43] P. CARDALIAGUET, C. LE BRIS, P. E. SOUGANIDIS.*Perturbation problems in homogenization of hamilton-jacobi equations*, January 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01435744.

[44] L. CHAMOIN, F. LEGOLL.*A posteriori error estimation and adaptive strategy for the control of MsFEM computations*, September 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01586825.

[45] G. DI GESÙ, T. LELIÈVRE, D. LE PEUTREC, B. NECTOUX.*Sharp asymptotics of the first exit point density*, June 2017, https://arxiv.org/abs/1706.08728 - 146 pages, https://hal.archives-ouvertes.fr/hal-01548737.

[46] M. JOSIEN.*Mathematical properties of the Weertman equation*, September 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01589690.

[47] C. LE BRIS, F. LEGOLL, F. MADIOT.*Multiscale Finite Element methods for advection-dominated problems in perforated domains*, October 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01624049.

[48] C. LE BRIS, F. LEGOLL, F. MADIOT.*Stable approximation of the advection-diffusion equation using the invariant measure*, 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01367417.

[49] D. LE PEUTREC.*On Witten Laplacians and Brascamp-Lieb's inequality on manifolds with boundary*, February 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01349786.

[50] A. LEVITT, H. D. CORNEAN, D. GONTIER, D. MONACO.*Localised Wannier functions in metallic systems*, December 2017, https://arxiv.org/abs/1712.07954 - 18 pages, 4 figures, https://hal.archives-ouvertes.fr/hal-01671848.

[51] L. J. S. LOPES, T. LELIÈVRE.*Analysis of the Adaptive Multilevel Splitting method with the alanine di-peptide's isomerization*, August 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01574847.

[52] J. ROUSSEL, G. STOLTZ.*Spectral methods for Langevin dynamics and associated error estimates*, February 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01470251.

[53] G. STOLTZ, E. VANDEN-EIJNDEN.*Longtime convergence of the Temperature-Accelerated Molecular Dynamics Method*, August 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01578911.

[54] K. VEROY, M. GREPL, M. KAERCHER, S. BOYAVAL.*Reduced basis approximation and a posteriori error bounds for 4D-Var data assimilation*, July 2017, working paper or preprint, https://hal.inria.fr/hal-01556304.

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. - Mathematical Modeling
A6.1.2. - Stochastic Modeling (SPDE, SDE)
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
B9.5.3. - Economy, Finance
B9.9. - Risk management

1. Personnel

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]

Faculty Members

Vlad Bally [Univ Paris-Est Marne La Vallée, Professor, HDR]
Benjamin Jourdain [Ecole Nationale des Ponts et Chaussées, Professor, HDR]
Damien Lamberton [Univ Paris-Est Marne La Vallée, Professor, HDR]

External Collaborators

Jean-Philippe Chancelier [Ecole Nationale des Ponts et Chaussées]
Oleg Kudryavtsev [Rostov University (Russia), from Jul 2017 until Aug 2017]
Céline Labart [Université de Savoie, from Mar 2017 until Aug 2017]
Jérôme Lelong [ENSIMAG, HDR]
Antonino Zanette [University of Udine,Italy]

PhD Students

Rui Chen [Autre entreprise publique]
Rafael Coyaud [Ecole Polytechnique, from Oct 2017]

Administrative Assistant

Martine Verneuille [Inria]

2. Overall Objectives

2.1. Overall Objectives

MathRisk is a joint Inria project-team with ENPC (CERMICS Laboratory) and the University Paris Est Marne-la-Vallée (UPEMLV, LAMA Laboratory), located in Paris and Marne-la-Vallée.

http://www.inria.fr/en/teams/mathrisk. Mathrisk is based on the former Mathfi project team. Mathfi was founded in 2000, and was devoted to financial mathematics. The project was focused on advanced stochastic analysis and numerical techniques motivated by the development of increasingly complex financial products. Main applications concerned evaluation and hedging of derivative products, dynamic portfolio optimization in incomplete markets, and calibration of financial models.

2.1.1. Crisis, deregulation, and impact on the research in finance

The starting point of the development of modern finance theory is traditionally associated to the publication of the famous paper of Black and Scholes in 1973 [54]. Since then, in spite of sporadic crises, generally well overcome, financial markets have grown in a exponential manner. More and more complex exotic derivative products have appeared, on equities first, then on interest rates, and more recently on credit markets. The period between the end of the eighties and the crisis of 2008 can be qualified as the "golden age of financial mathematics": finance became a quantitative industry, and financial mathematics programs flourished in top universities, involving seminal interplays between the worlds of finance and applied mathematics. During its 12 years existence, the Mathfi project team has extensively contributed to the development of modeling and computational methods for the pricing and hedging of increasingly complex financial products.

Since the crisis of 2008, there has been a critical reorientation of research priorities in quantitative finance with emphasis on risk. In 2008, the "subprime" crisis has questioned the very existence of some derivative products such as CDS (credit default swaps) or CDOs (collateralized debt obligations), which were accused to be responsible for the crisis. The nature of this crisis is profoundly different from the previous ones. It has negatively impacted the activity on the exotic products in general, - even on equity derivative markets-, and the interest in the modeling issues for these products. The perfect replication paradigm, at the origin of the success of the Black and Scholes model became unsound, in particular through the effects of the lack of liquidity. The interest of quantitative finance analysts and mathematicians shifted then to more realistic models taking into account the multidimensional feature and the incompleteness of the markets, but as such getting away from the "lost paradi(gm)" of perfect replication. These models are much more demanding numerically, and require the development of hedging risk measures, and decision procedures taking into account the illiquidity and various defaults.

Moreover, this crisis, and in particular the Lehman Brothers bankruptcy and its consequences, has underlined a systemic risk due to the strong interdependencies of financial institutions. The failure of one of them can cause a cascade of failures, thus affecting the global stability of the system. Better understanding of these interlinkage phenomena becomes crucial.

At the same time, independently from the subprime crisis, another phenomenon has appeared: deregulation in the organization of stock markets themselves. This has been encouraged by the Markets in Financial Instruments Directive (MIFID) which is effective since November, 1st 2007. This, together with the progress of the networks, and the fact that all the computers have now a high computation power, have induced arbitrage opportunities on the markets, by very short term trading, often performed by automatic trading. Using these high frequency trading possibilities, some speculating operators benefit from the large volatility of the markets. For example, the flash crash of May, 6 2010 has exhibited some perverse effects of these automatic speculating trading strategies. These phenomena are not well understood and the theme of high frequency trading needs to be explored.

To summarize, financial mathematics is facing the following new evolutions:

- the complete market modeling has become unsatisfactory to provide a realistic picture of the market and is replaced by incomplete and multidimensional models which lead to new modeling and numerical challenges.
- quantitative measures of risk coming from the markets, the hedging procedures, and the lack of liquidity are crucial for banks,
- uncontrolled systemic risks may cause planetary economic disasters, and require better understanding,
- deregulation of stock markets and its consequences lead to study high frequency trading.

The project team MathRisk is designed to address these new issues, in particular dependence modeling, systemic risk, market microstructure modeling and risk measures. The research in modeling and numerical analysis remain active in this new context, motivated by new issues.

The MathRisk project team develops the software Premia dedicated to pricing and hedging options and calibration of financial models, in collaboration with a consortium of financial institutions. https://www.rocq.inria.fr/mathfi/Premia/index.html.

The MathRisk project is part of the Université Paris-Est **"Labex" BÉZOUT**.

3. Research Program

3.1. Dependence modeling

Participants: Aurélien Alfonsi, Benjamin Jourdain, Damien Lamberton, Bernard Lapeyre.

The volatility is a key concept in modern mathematical finance, and an indicator of the market stability. Risk management and associated instruments depend strongly on the volatility, and volatility modeling has thus become a crucial issue in the finance industry. Of particular importance is the assets *dependence* modeling. The calibration of models for a single asset can now be well managed by banks but modeling of dependence is the bottleneck to efficiently aggregate such models. A typical issue is how to go from the individual evolution of each stock belonging to an index to the joint modeling of these stocks. In this perspective, we want to model stochastic volatility in a *multidimensional* framework. To handle these questions mathematically, we have to deal with stochastic differential equations that are defined on matrices in order to model either the instantaneous covariance or the instantaneous correlation between the assets. From a numerical point of view, such models are very demanding since the main indexes include generally more than thirty assets. It is therefore necessary to develop efficient numerical methods for pricing options and calibrating such models to market data. As a first application, modeling the dependence between assets allows us to better handle derivatives products on a basket. It would give also a way to price and hedge consistensly single-asset and basket products. Besides, it can be a way to capture how the market estimates the dependence between assets. This could give some insights on how the market anticipates the systemic risk.

3.2. Liquidity risk

Participants: Aurélien Alfonsi, Agnès Sulem, Antonino Zanette.

The financial crisis has caused an increased interest in mathematical finance studies which take into account the market incompleteness issue and the liquidity risk. Loosely speaking, liquidity risk is the risk that comes from the difficulty of selling (or buying) an asset. At the extreme, this may be the impossibility to sell an asset, which occurred for "junk assets" during the subprime crisis. Hopefully, it is in general possible to sell assets, but this may have some cost. Let us be more precise. 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. Instead, 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), this 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.

3.2.1. Long term liquidity risk.

On a long-term perspective, illiquidity can be approached via various ways: transactions costs [46], [47], [53], [58], [61], [73], [70], delay in the execution of the trading orders [74], [72], [55], trading constraints or restriction on the observation times (see e.g. [60] and references herein). As far as derivative products are concerned, one has to understand how delta-hedging strategies have to be modified. This has been considered for example by Cetin, Jarrow and Protter [71]. We plan to contribute on these various aspects of liquidity risk modeling and associated stochastic optimization problems. Let us mention here that the price impact generated by the trades of the investor is often neglected with a long-term perspective. This seems acceptable since the investor has time enough to trade slowly in order to eliminate its market impact. Instead, when the investor wants to make significant trades on a very short time horizon, it is crucial to take into account and to model how prices are modified by these trades. This question is addressed in the next paragraph on market microstructure.

3.2.2. Market microstructure.

The European directive MIFID has increased the competition between markets (NYSE-Euronext, Nasdaq, LSE and new competitors). As a consequence, the cost of posting buy or sell orders on markets has decreased, which has stimulated the growth of market makers. Market makers are posting simultaneously bid and ask orders on a same stock, and their profit comes from the bid-ask spread. Basically, their strategy is a "round-trip" (i.e. their position is unchanged between the beginning and the end of the day) that has generated a positive cash flow.

These new rules have also greatly stimulated research on market microstructure modeling. From a practitioner point of view, the main issue is to solve the so-called "optimal execution problem": given a deadline T, what is the optimal strategy to buy (or sell) a given amount of shares that achieves the minimal expected cost? For large amounts, it may be optimal to split the order into smaller ones. This is of course a crucial issue for brokers, but also market makers that are looking for the optimal round-trip.

Solving the optimal execution problem is not only an interesting mathematical challenge. It is also a mean to better understand market viability, high frequency arbitrage strategies and consequences of the competition between markets. For example when modeling the market microstructure, one would like to find conditions that allow or exclude round trips. Beyond this, even if round trips are excluded, it can happen that an optimal selling strategy is made with large intermediate buy trades, which is unlikely and may lead to market instability.

We are interested in finding synthetic market models in which we can describe and solve the optimal execution problem. A. Alfonsi and A. Schied (Mannheim University) [48] have already proposed a simple Limit Order Book model (LOB) in which an explicit solution can be found for the optimal execution problem. We are now interested in considering more sophisticated models that take into account realistic features of the market such as short memory or stochastic LOB. This is mid term objective. At a long term perspective one would like to bridge these models to the different agent behaviors, in order to understand the effect of the different quotation mechanisms (transaction costs for limit orders, tick size, etc.) on the market stability.

3.3. Contagion modeling and systemic risk

Participants: Benjamin Jourdain, Agnès Sulem.

After the recent financial crisis, systemic risk has emerged as one of the major research topics in mathematical finance. 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. Ideally, one would like to be able to find capital requirements (such as the one proposed by the Basel committee) that ensure that the probability of multiple defaults is below some level.

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. [49] and A. Minca [65]) or mean field interaction models (Garnier-Papanicolaou-Yang [59]). The recent approach in [49] seems very promising. It describes the financial network approach as a weighted directed graph, in which nodes represent financial institutions and edges the exposures between them. Distress propagation in a financial system may be modeled as an epidemics on this graph. In the case of incomplete information on the structure of the interbank network, cascade dynamics may be reduced to the evolution of a multi-dimensional Markov chain that corresponds to a sequential discovery of exposures and determines at any time the size of contagion. Little has been done so far on the *control* of such systems in order to reduce the systemic risk and we aim to contribute to this domain.

3.4. Stochastic analysis and numerical probability

3.4.1. Stochastic control

Participants: Vlad Bally, Jean-Philippe Chancelier, Marie-Claire Quenez, Agnès Sulem.

The financial crisis has caused an increased interest in mathematical finance studies which take into account the market incompleteness issue and the default risk modeling, the interplay between information and performance, the model uncertainty and the associated robustness questions, and various nonlinearities. We address these questions by further developing the theory of stochastic control in a broad sense, including stochastic optimization, nonlinear expectations, Malliavin calculus, stochastic differential games and various aspects of optimal stopping.

3.4.2. Optimal stopping

Participants: Aurélien Alfonsi, Benjamin Jourdain, Damien Lamberton, Agnès Sulem, Marie-Claire Quenez.

The theory of American option pricing has been an incite for a number of research articles about optimal stopping. Our recent contributions in this field concern optimal stopping in models with jumps, irregular obstacles, free boundary analysis, reflected BSDEs.

3.4.3. Simulation of stochastic differential equations

Participants: Benjamin Jourdain, Aurélien Alfonsi, Vlad Bally, Damien Lamberton, Bernard Lapeyre, Jérôme Lelong, Céline Labart.

Effective numerical methods are crucial in the pricing and hedging of derivative securities. The need for more complex models leads to stochastic differential equations which cannot be solved explicitly, and the development of discretization techniques is essential in the treatment of these models. The project MathRisk addresses fundamental mathematical questions as well as numerical issues in the following (non exhaustive) list of topics: Multidimensional stochastic differential equations, High order discretization schemes, Singular stochastic differential equations, Backward stochastic differential equations.

3.4.4. Monte-Carlo simulations

Participants: Benjamin Jourdain, Aurélien Alfonsi, Damien Lamberton, Vlad Bally, Bernard Lapeyre, Ahmed Kebaier, Céline Labart, Jérôme Lelong, Antonino Zanette.

Monte-Carlo methods is a very useful tool to evaluate prices especially for complex models or options. We carry on research on *adaptive variance reduction methods* and to use *Monte-Carlo methods for calibration* of advanced models.

This activity in the MathRisk team is strongly related to the development of the Premia software.

3.4.5. Malliavin calculus and applications in finance

Participants: Vlad Bally, Arturo Kohatsu-Higa, Agnès Sulem, Antonino Zanette.

The original Stochastic Calculus of Variations, now called the Malliavin calculus, was developed by Paul Malliavin in 1976 [63]. It was originally designed to study the smoothness of the densities of solutions of stochastic differential equations. One of its striking features is that it provides a probabilistic proof of the celebrated Hörmander theorem, which gives a condition for a partial differential operator to be hypoelliptic. This illustrates the power of this calculus. In the following years a lot of probabilists worked on this topic and the theory was developed further either as analysis on the Wiener space or in a white noise setting. Many applications in the field of stochastic calculus followed. Several monographs and lecture notes (for example D. Nualart [66], D. Bell [52] D. Ocone [68], B. Øksendal [75]) give expositions of the subject. See also V. Bally [50] for an introduction to Malliavin calculus.

From the beginning of the nineties, applications of the Malliavin calculus in finance have appeared : In 1991 Karatzas and Ocone showed how the Malliavin calculus, as further developed by Ocone and others, could be used in the computation of hedging portfolios in complete markets [67].

Since then, the Malliavin calculus has raised increasing interest and subsequently many other applications to finance have been found [64], such as minimal variance hedging and Monte Carlo methods for option pricing. More recently, the Malliavin calculus has also become a useful tool for studying insider trading models and some extended market models driven by Lévy processes or fractional Brownian motion.

We give below an idea why Malliavin calculus may be a useful instrument for probabilistic numerical methods.

We recall that the theory is based on an integration by parts formula of the form $E(f'(X)) = E(f(X)Q)$. Here X is a random variable which is supposed to be "smooth" in a certain sense and non-degenerated. A basic example is to take $X = \sigma\Delta$ where Δ is a standard normally distributed random variable and σ is a strictly positive number. Note that an integration by parts formula may be obtained just by using the usual integration by parts in the presence of the Gaussian density. But we may go further and take X to be an aggregate of Gaussian random variables (think for example of the Euler scheme for a diffusion process) or the limit of such simple functionals.

An important feature is that one has a relatively explicit expression for the weight Q which appears in the integration by parts formula, and this expression is given in terms of some Malliavin-derivative operators.

Let us now look at one of the main consequences of the integration by parts formula. If one considers the *Dirac* function $\delta_x(y)$, then $\delta_x(y) = H'(y-x)$ where H is the *Heaviside* function and the above integration by parts formula reads $E(\delta_x(X)) = E(H(X-x)Q)$, where $E(\delta_x(X))$ can be interpreted as the density of the random variable X. We thus obtain an integral representation of the density of the law of X. This is the starting point of the approach to the density of the law of a diffusion process: the above integral representation allows us to prove that under appropriate hypothesis the density of X is smooth and also to derive upper and lower bounds for it. Concerning simulation by Monte Carlo methods, suppose that you want to compute $E(\delta_x(y)) \sim \frac{1}{M}\sum_{i=1}^{M} \delta_x(X^i)$ where $X^1, ..., X^M$ is a sample of X. As X has a law which is absolutely continuous with respect to the Lebesgue measure, this will fail because no X^i hits exactly x. But if you are able to simulate the weight Q as well (and this is the case in many applications because of the explicit form mentioned above) then you may try to compute $E(\delta_x(X)) = E(H(X-x)Q) \sim \frac{1}{M}\sum_{i=1}^{M} E(H(X^i-x)Q^i)$. This basic remark formula leads to efficient methods to compute by a Monte Carlo method some irregular quantities as derivatives of option prices with respect to some parameters (the *Greeks*) or conditional expectations, which appear in the pricing of American options by the dynamic programming). See the papers by Fournié et al [57] and [56] and the papers by Bally et al., Benhamou, Bermin et al., Bernis et al., Cvitanic et al., Talay and Zheng and Temam in [62].

L. Caramellino, A. Zanette and V. Bally have been concerned with the computation of conditional expectations using Integration by Parts formulas and applications to the numerical computation of the price and the Greeks (sensitivities) of American or Bermudean options. The aim of this research was to extend a paper of Reigner and Lions who treated the problem in dimension one to higher dimension - which represent the real challenge

in this field. Significant results have been obtained up to dimension 5 [51] and the corresponding algorithms have been implemented in the Premia software.

Moreover, there is an increasing interest in considering jump components in the financial models, especially motivated by calibration reasons. Algorithms based on the integration by parts formulas have been developed in order to compute Greeks for options with discontinuous payoff (e.g. digital options). Several papers and two theses (M. Messaoud and M. Bavouzet defended in 2006) have been published on this topic and the corresponding algorithms have been implemented in Premia. Malliavin Calculus for jump type diffusions - and more general for random variables with locally smooth law - represents a large field of research, also for applications to credit risk problems.

The Malliavin calculus is also used in models of insider trading. The "enlargement of filtration" technique plays an important role in the modeling of such problems and the Malliavin calculus can be used to obtain general results about when and how such filtration enlargement is possible. See the paper by P. Imkeller in [62]). Moreover, in the case when the additional information of the insider is generated by adding the information about the value of one extra random variable, the Malliavin calculus can be used to find explicitly the optimal portfolio of an insider for a utility optimization problem with logarithmic utility. See the paper by J.A. León, R. Navarro and D. Nualart in [62]).

A. Kohatsu Higa and A. Sulem have studied a controlled stochastic system whose state is described by a stochastic differential equation with anticipating coefficients. These SDEs can be interpreted in the sense of *forward integrals*, which are the natural generalization of the semi-martingale integrals, as introduced by Russo and Valois [69]. This methodology has been applied for utility maximization with insiders.

4. Application Domains

4.1. Financial Mathematics, Insurance

The applications domains 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. New Software and Platforms

5.1. 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, Jacques Printems and Jérôme Lelong
- Partners: Inria - Ecole des Ponts ParisTech - Université Paris-Est
- Contact: Agnes Sulem
- URL: http://www.premia.fr

5.2. Platforms

5.2.1. Development of the quantitative platform Premia in 2017

- Premia 18 has been registered at the Agence pour la Protection des Programmes APP (IDDN.FR.001.190010.014.S.C.2001.000.31000)

- Premia 19 has been delivered to the Consortium Premia on March 16th. It contains the following new algorithms : Risk Management, Model Risk, Insurance

- XVA simulation on GPUs using Nested Monte Carlo. L. Abbas Turki
- Model-independent bounds for option prices a mass transport approach. M. Beiglböck, P. H. Labordère, F. Penkner
  Finance Stochastics Volume 17, 2013.
- Model-Independent Pricing of Asian Options via Optimal Martingale. F Stebegg
- Pricing and Hedging GMWB in the Heston and in the Black-Scholes with Stochastic Interest Rate Models. L. Goudenege, A. Molent, A. Zanette

Equity Derivatives

- Quantization meets Fourier: a new technology for pricing options. G. Callegaro, L. Fiorin, M. Grasselli.
- Efficient unbiased simulation scheme for the SABR stochastic volatility model. B. Chen, C.W. Oosterlee J.A.M van der Weide.
  International Journal of Theoretical and Applied Finance. Vol. 15, No. 2 (2012)
- An efficient Monte Carlo method for discrete variance contracts. N.Merener L.Vicchi
  The Journal of Computational Finance. Vol. 16, Issue 4, 2013.
- The evaluation of barrier option prices under stochastic volatility. C.Chiarella, B.Kang, G.H.Meyer
  Computers and Mathematics with Applications 64, 2012.
- Path-Dependent Volatility. J.Guyon
  Risk Magazine, October 2014.
- Cross-Dependent Volatility. J.Guyon
  Risk Magazine, March 2016
- Option pricing and implied volatilities in a 2-hypergeometric stochastic volatility model. N. Privault Q. She
  Applied Mathematics Letters, 53, 2016
- Option price with stochastic volatility for both fast and slow mean-reverting regimes. Q. Zhang, J. Han, and M. Gao.
  C. R. Math. Acad. Sci. Paris, 351(9-10), 2013.
- Application of the improved fast Gauss transform to option pricing under jump-diffusion processes. T. Sakuma and Y. Yamada

The Journal of Computational Finance, Volume 18, Issue 2, 2014.

- Features of the Russian derivatives market volatility index development taking into account possible price jumps. A. Grechko, O. Kudryavtsev
  Theory of Probability and Its Applications-SIAM, 2017 61:3 (2017), to appear
- On the application of spectral filters in a Fourier option pricing technique. C.W. Oosterlee M.J. Ruijter, M. Versteegh
  The Journal of Computational Finance, Volume 19, Issue 1, 2015.
- Unbiased simulation of stochastic differential equations. P.H. Labordere X. Tan, N. Touzi

We have benefited from the help of the engineer Cedric Doucet, supervised by Jérôme Lelong, for designing non regression tests for Premia.

6. New Results

6.1. Systemic risk

Participants: Agnès Sulem, Andreea Minca [Cornell University], Rui Chen.

We have studied optimal connectivity of a large financial network in presence of growth and contagion [27]. We obtained asymptotic results for the magnitude of default contagion in a large financial system with intrinsic recovery features in the framework of a random network. We have moreover added a game component to the model, allowing institutions to choose their optimal linkages in order to maximize their final profits, given their initial states and estimated survival probabilities.

6.2. Optimal stopping for Backward stochastic (partial) differential equations with jumps

Agnès Sulem, Rui Chen and R. Dumitrescu have addressed the problem of optimal stopping for general mean-field backward stochastic differential equations driven by a Brownian motion and an independent Poisson random measure. Existence, uniqueness, comparison and dual representation results have been obtained. Links with reflected mean-field BSDEs have been established and application to global dynamic risk measure theory has been investigated.

American options in markets with imperfections and default have been studied by Agnès Sulem, M.C. Quenez and R. Dumitrescu [28].

6.3. Approximation of Martingale Optimal Transport problems

With J. Corbetta, A. Alfonsi and B. Jourdain study sampling methods preserving the convex order for two probability measures μ and ν on $\mathbf{R}^d$, with ν dominating μ. When $(X_i)_{1\leq i\leq I}$ (resp. $(Y_j)_{1\leq j\leq J}$) are independent and identically distributed according μ (resp. ν), in general $\mu_I = \frac{1}{I}\sum_{i=1}^{I}\delta_{X_i}$ and $\nu_J = \frac{1}{J}\sum_{j=1}^{J}\delta_{Y_j}$ are not rankable for the convex order. They investigate modifications of μ_I (resp. ν_J) smaller than ν_J (resp. greater than μ_I) in the convex order and weakly converging to μ (resp. ν) as $I, J \to \infty$. They first consider the one dimensional case $d = 1$, where, according to Kertz and Rösler, the set of probability measures with a finite first order moment is a lattice for the increasing and the decreasing convex orders. Given μ and ν in this set, they define $\mu \vee \nu$ (resp. $\mu \wedge \nu$) as the supremum (resp. infimum) of μ and ν for the decreasing convex order when $\int_{\mathbf{R}} x\mu(dx) \leq \int_{\mathbf{R}} x\nu(dx)$ and for the increasing convex order otherwise. This way, $\mu \vee \nu$ (resp. $\mu \wedge \nu$) is greater than μ (resp. smaller than ν) in the convex order. They give efficient algorithms permitting to compute $\mu \vee \nu$ and $\mu \wedge \nu$ (and therefore $\mu_I \vee \nu_J$ and $\mu_I \wedge \nu_J$) when μ and ν are convex combinations of Dirac masses. In general dimension, when μ and ν have finite moments of order $\rho \geq 1$, they define the projection $\mu \curlywedge_\rho \nu$ (resp. $\mu \curlyvee_\rho \nu$) of μ (resp. ν) on the set of probability measures dominated by ν (resp. larger than μ) in the convex order for the Wasserstein distance with index ρ. When $\rho = 2$, $\mu_I \curlywedge_2 \nu_J$ can be computed efficiently

by solving a quadratic optimization problem with linear constraints. It turns out that, in dimension $d = 1$, the projections do not depend on ρ and their quantile functions are explicit in terms of those of μ and ν, which leads to efficient algorithms for convex combinations of Dirac masses. Last, they illustrate by numerical experiments the resulting sampling methods that preserve the convex order and their application to approximate Martingale Optimal Transport problems.

With V. Ehrlacher, D. Lombardi and R. Coyaud, A. Alfonsi has started to develop and analyze numerical methods to approximate the optimal transport between two probability measures.

6.4. Numerical methods for Asset-Liability Management

With A. Cherchali, A. Alfonsi is working on obtaining a model for the Asset-Liability Management (ALM) of insurance companies. The purpose is to use this model to develop Monte-Carlo methods to approximate the SCR (Solvency Capital Requirement).

6.5. American options

With Giulia Terenzi, D. Lamberton has been working on American options in Heston's model. 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) have been obtained, as well as some results on monotonicity and regularity properties of the price function. A paper on this topic has just been submitted.

6.6. Stochastic Analysis and Malliavin calculus

- Invariance principles for stochastic polynomials [40].

  With L. Caramellino (Roma), V. Bally has studied invariance principles for stochastic polynomials. This is a generalization of the classical invariance principle from the Central Limit Theorem, of interest in U-statistics. The main contribution concerns convergence in total variation distance, using an abstract variant of Malliavin calculus for general random variables which verify a Doeblin type condition.

- Convergence in distribution norms in the Central Limit Theorem and Edgworth expansions [39] (V. Bally, L. Caramellino and G. Poly).

  The convergence in "distribution norms" represents an extension of the convergence in total variation distance which permits to take into account some singular phenomenons. The main tool is the abstract Malliavin calculus mentioned above. Several examples are given in the paper and an outstanding application concerns the estimates of the number of roots of trigonometric polynomials. considered in a second paper; see [40].

- Bolzmann equation and Piecewise Deterministic Markov Processes. (see [41], [37]). In collaboration with D. Goreac and V. Rabiet, V. Bally has studied the regularity of the semigroup of $PDMP's$ and, as an application estimates of the distance between two such semigroups. An interesting example is given by the two dimensional homogeneous Bolzmann equation. Furthermore, V. Bally obtained some exponential estimates for the function solution of this equation.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Grants with Industry

- Consortium PREMIA, Natixis - Inria
- Consortium PREMIA, Crédit Agricole CIB - Inria

7.2. Bilateral Contracts with Industry

- Chair Ecole Polytechnique-ENPC-UPMC-Société Générale "Financial Risks" of the Risk fondation . Participants: A. Alfonsi, B. Jourdain, B. Lapeyre
- AXA Joint Research Initiative on Numerical methods for the ALM, from September 2017 to August 2020, Participant: A. Alfonsi.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

- ANR Cosmos 2015-2018, Participant: B. Jourdain ; Partners : Ecole des Ponts, Telecom, Inria Rennes and IBPC

8.1.2. Competitivity Clusters

Pôle Finance Innovation.

8.2. International Initiatives

8.2.1. 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).

8.3. International Research Visitors

- Oleg Kudryavtsev, Rostov University (Russia)
- Martino Grasselli, Padova University,

8.3.1. Visits of International Scientists

8.3.1.1. Internships

- Adel Cherchali (June to August 2017): Multilevel Monte-Carlo methods for nested expectations. Supervisor: A. Alfonsi.

- Zeqi Chen (ENSTA), May -July , Supervisor: A. Zanette

- Mohamed Homed, April-September, Supervisor: A. Zanette

- Xinglong Tian (ENSTA), May-July 2017, Supervisor: A. Zanette

- Sebastien Villette, April-October, Supervisor: A. Zanette

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Organizing Committees

- A. Alfonsi:

  - Co-organizer of the conference "Advances in Financial Mathematics", 10-13 January 2017, https://fin-risks2017.sciencesconf.org/.

  - Co-organizer of the working group seminar of MathRisk "Méthodes stochastiques et finance".

- J. Lelong :

  - Member of the organizing committee of Les journées de Probabilités, 2017, France, Aussois.

  - Member of the organizing committee of CEMRACS, 2017, Marseille.

  - Member of the organizing committee of Les journées SMAI MODE 2018, Grenoble.

- A. Sulem

  Co-organizer of the seminar Inria-MathRisk /Université Paris 7 LPMA "Numerical probability and mathematical finance". https://www.lpma-paris.fr/mathfipronum/gt

9.1.2. Scientific Events Selection

9.1.2.1. Member of the Conference Program Committees

B. Jourdain: Member of the scientific committee of the congrès SMAI 2017

9.1.2.2. Reviewer

A. Sulem: Reviewer for *Mathematical Reviews*

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

- D. Lamberton

  Associate editor of

  - Mathematical Finance,
  - Associate editor of ESAIM Probability & Statistics

- A. Sulem

  Associate editor of

  - 2011- Present: *Journal of Mathematical Analysis and Applications (JMAA)*
  - 2009- Present: *International Journal of Stochastic Analysis (IJSA)*
  - 2008- Present: *SIAM Journal on Financial Mathematics (SIFIN)*

9.1.3.2. Reviewer - Reviewing Activities

The members of the team reviewed numerous papers for many journals in probability, finance, stochastic control, applied mathematics, ...

9.1.4. Invited Talks

- A. Alfonsi

  - "Maximum Likelihood Estimation for Wishart processes", conference on "Mathematics of Quantitative Finance", Oberwolfach, March 1.

  - "Optimal Execution in a Hawkes Price Model and Calibration", Market Microstructure and High Frequency Data June 1-3, The University of Chicago, June 2.

  - "Maximum Likelihood Estimation for Wishart processes", Recent Developments in Numerical Methods with Applications in Statistics and Finance Mannheim, Germany, June 9.

  - "Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems." New York, conference in honour of Jim Gatheral, NYU, October 15.

  - "Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems." GT CMAP-ENSTA-ENSAE, November 27.

- V. Bally

  - LMS-EPSRC Durham Symposium 10-20 July. Regularity for the solution of jump equations using an interpolation method.

  - Conference of Stochastic Processes and their Applications (SPA2017) 24-28 July. Regularity for the solution of jump equations using an interpolation method.

  - Workshop on Piecewise Deterministic Markov Processes, 29.05-2.06 Gaussian noise versus Poisson Point Measures in PDMP's.

- B. Jourdain:

  - Workshop Stochastic Sampling and Accelerated Time Dynamics on Multidimensional Surfaces, IPAM, Los Angeles, 16-20 October : Convergence and efficiency of adaptive importance sampling techniques with partial biasing

  - Workshop Singular McKean-Vlasov dynamics, Sophia-Antipolis, 14-15 September : Existence to calibrated regime-switching local volatility model

  - Summer school CEMRACS, Marseille, 17-21 July : The Metropolis-Hastings algorithm, introduction and optimal scaling of the transient phase

  - Workshop Stochastic Computation, FOCM 2017, Barcelona, 10-12 July : Strong convergence properties of the Ninomiya-Victoir scheme and applications to multilevel Monte Carlo methods

  - Workshop BSDEs SPDEs, Edinburgh, 3-7 July : Existence to calibrated regime-switching local volatility model

  - Conference PDE/Probability Interactions : Kinetic Equations, Large Time and Propagation of Chaos, Marseille, 18-22 April : On a stochastic particle approximation of the Keller-Segel equation

  - Conference PDE and probability methods for interaction, Sophia-Antipolis, 30-31 March : Evolution of the Wasserstein distance between the marginals of two Markov processes

  - Applied Mathematics seminar of the Collège de France, February 24 : Multitype sticky particles and diagonal hyperbolic systems

  - Conference Advances in Financial Mathematics, Paris, 10-13 January : Existence to a calibrated regime-switching local volatility model

- J. Lelong

  High Performance Computing session during CEMRACS: 4 hours of lectures and 6 hours of hands–on sessions, Marseille.

- A. Sulem

  - Plenary talk at the Congrès SMAI 2017, 8ème biennale française des Mathématiques Appliquées et Industrielles, Ronce-les-bains, June 20017, http://smai.emath.fr/smai2017/index.php

- "Recent advances in financial mathematics", conference organised by "Financial Risks Chair", Paris, Janvier 2017. https://fin-risks2017.sciencesconf.org/program

- Worshop on *Optimal Stopping in Complex environments*, Bielefeld University, December 18-20 2017,
https://sites.google.com/view/imwworkshop17/

- Workshop on "Asymptotics of Stochastic Dynamics", University of Swansea, August, 29-21, 2017

- Simulation of Stochastic graphs and applications symposium, International Conference on Monte Carlo techniques, Paris, July 5-8, 2017 https://montecarlo16.sciencesconf.org

9.1.5. Scientific Expertise

B. Jourdain: Member of the Scientific Advisory Board of the Center for interdisciplinary Research in Biology, Collège de France : March 1st and 2nd 2017

9.1.6. Research Administration

- A. Alfonsi

  - Deputy director of the CERMICS since April 2017.

  - 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 technology development, Inria Paris

  - Corresponding member of the comité opérationel d’évaluation des risques légaux et éthniques (COERLE) at Inria Paris research center

  - Member of the Committee for Inria international Chairs

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence :

- A. Alfonsi

  ‘Probabilités”, first year course at the Ecole des Ponts.

- B. Jourdain

  - "Introduction to probability theory", 1st year, Ecole Polytechnique

  - "Mathematical finance", 2nd year ENPC

- V. Bally

  Hilbertien Analysis L3 (36h)

Master

- A. Alfonsi:

  - “Traitement des données de marché : aspects statistiques et calibration”, lecture for the Master at UPEMLV.

  - “Mesures de risque”, Master course of UPEMLV and Paris VI.

  - Professeur chargé de cours at Ecole Polytechnique.

- V. Bally

  - Interest rates (20h) M2 filière finance

  - Malliavin calculus and applications in finance (30h) M2 filière finance

  - "Risk analysis " M2 filière actuariat (45h)

- 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".

Doctorat :

A. Sulem : "Stochastic Control with Applications to Mathematical Finance", International summer school in "Financial Mathematics and Actuarial Science", **Doctoral lectures**, (30 heures), Atlantic Association for Research in the Mathematical Sciences, University of Prince Edward Island (UPEI), Canada, July https://aarms.math.ca/summer-school

9.2.2. Supervision

HdR : J. Lelong, Quelques contributions aux méthodes numériques probabilistes et à la modélisation stochastique, Université Grenoble-Alpes, September 2017.

- PhD in progress :

  - Adel Cherchali, “Numerical methods for the ALM”, funded by Fondation AXA, starting from September 2017, Supervisor: A. Alfonsi

  - Rafaël Coyaud, “Deterministic and 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.

  - Marouen Iben Taarit , “ On CVA and XVA computations ”, CIFRE Natixis/ENPC, Supervisor: Bernard Lapeyre

  - Giulia Terenzi , "American options in complex financial models", Université Paris-Est Marne-la-Vallée, Supervisors: Damien Lamberton and Lucia Caramellino, from University Tor Vergata, Rome

  - Alexandre Zhou (started November 2015) "Analysis of stochastic particle methods applied to finance", Supervisor: B.Jourdain

  - Oumaima Bencheikh (started November 2017) "Acceleration of probabilistic particle methods", Supervisor: B. Jourdain

9.2.3. Juries

- Damien Lamberton

  "Opponent" for the PhD thesis defense of Hannah Dyrssen (student of Erik Ekstrom) at Uppsala University (Sweden), May 2017.

- Benjamin Jourdain

  Referee for the PhD thesis and participation to the jury for the defense of the PhD thesis of

  - Victor Reutenauer, defended on March 22, University Côte d'Azur

  - Daphné Giorgy, defended on June 2, University Pierre and Marie Curie

  - Radu Maftei, defended on December 14, University Côte d'Azur

- Agnès Sulem

  - Participation to the committee for the recrutment of a Professeur in "applied mathematics, finance and numerical probability", Laboratoire de probabilités (LPMA), Université Paris VI, 2017
  - Participation to the committee for the recrutment of a Assistant professor in "économy, finance et game theory", Université Paris-Dauphine, 2017.

  - Participation to the jury (Chair) for the defense of the PhD thesis of Amine Ismail, *Robust modeling of volatility and application to derivatives pricing and portfolio optimization*, December 15 2017, Université Paris-Diderot - Paris 7.

  - Participation to the jury (Chair) for the defense of the PhD thesis of Jiang Pu, *Contrôle optimal et applications en finance: exécution optimale, couverture d'options et choix de portefeuille*, September 25 2017, Université Paris-Diderot - Paris 7.

10. Bibliography

Major publications by the team in recent years

[1] L. ABBAS-TURKI, B. LAPEYRE.*American options by Malliavin calculus and nonparametric variance and bias reduction methods*, in "SIAM J. Financ. Math.", 2012, vol. 3, n^{o} 1, p. 479-510.

[2] A. AHDIDA, A. ALFONSI.*Exact and high order discretization schemes for Wishart processes and their affine extensions*, in "Annals of Applied Probability", 2013, vol. 23, n^{o} 3, p. 1025-1073 [*DOI :* 10.1214/12-AAP863], http://hal.inria.fr/hal-00491371.

[3] A. ALFONSI.*High order discretization schemes for the CIR process: Application to affine term structure and Heston models*, in "Stochastic Processes and their Applications", 2010, vol. 79, p. 209-237, http://www.ams.org/journals/mcom/2010-79-269/S0025-5718-09-02252-2/home.html.

[4] A. ALFONSI.*Affine diffusions and related processes: simulation, theory and applications*, Bocconi and Springer Series, Mathematics statistics, finance and economics, Springer, 2015.

[5] A. ALFONSI, B. JOURDAIN, A. KOHATSU-HIGA.*Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme*, in "Annals of Applied Probability", 2014, https://hal-enpc.archives-ouvertes.fr/hal-00727430.

[6] A. ALFONSI, A. SCHIED.*Optimal Trade Execution and Absence of Price Manipulations in Limit Order Book Models*, in "SIAM J. Finan. Math.", 2010, vol. 1, n^{o} 1, p. 490-522, http://epubs.siam.org/doi/abs/10.1137/090762786.

[7] H. AMINI, A. MINCA, A. SULEM.*Control of interbank contagion under partial information*, in "SIAM Journal on Financial Mathematics", December 2015, vol. 6, n^{o} 1, 24, https://hal.inria.fr/hal-01027540.

[8] V. BALLY, N. FOURNIER.*Regularization properties od the 2D homogenuos Bolzmann equation without cutoff*, in "PTRF", 2011, n^{o} 151, p. 659-670.

[9] M. JEUNESSE, B. JOURDAIN.*Regularity of the American put option in the Black-Scholes model with general discrete dividends*, in "Stochastic Processes and their Applications", 2012, vol. 112, p. 3101-3125, http://hal.archives-ouvertes.fr/hal-00633199.

[10] B. JOURDAIN.*Probabilités et statistique*, Ellipses, 2009.

[11] D. LAMBERTON, M. MIKOU.*Exercise boundary of the American put near maturity in an exponential Lévy model*, in "Finance and Stochastics", 2013, vol. 17, n^{o} 2, p. 355-394.

[12] D. LAMBERTON, M. ZERVOS.*On the optimal stopping of a one-dimensional diffusion*, in "Electronic Journal of Probability", 2013, vol. 18, n^{o} 34, p. 1-49.

[13] M.-C. QUENEZ, A. SULEM.*BSDEs with jumps, optimization and applications to dynamic risk measures*, in "Stochastic Processes and their Applications", March 2013, vol. 123, n^{o} 8, p. 3328-3357, https://hal.inria.fr/hal-00709632.

[14] 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://hal.inria.fr/hal-00773708.

[15] A. SULEM.*Numerical Methods implemented in the Premia Software*, March-April 2009, vol. 99, Special issue of the Journal "Bankers, Markets, Investors", Introduction by Agnès Sulem (Ed) and A. Zanette.

[16] B. ØKSENDAL, A. SULEM.*Applied Stochastic Control of Jump Diffusions*, Universitext, Second Edition, Springer, Berlin, Heidelberg, New York, 257 pages 2007.

[17] B. ØKSENDAL, A. SULEM.*Singular stochastic Control and Optimal stopping with partial information of Itô-Lévy processes*, in "SIAM J. Control & Optim.", 2012, vol. 50, n^{o} 4, p. 2254–2287, http://epubs.siam.org/doi/abs/10.1137/100793931.

[18] B. ØKSENDAL, A. SULEM, T. ZHANG.*Singular Control and Optimal Stopping of SPDEs, and Backward SPDEs with Reflection*, in "Mathematics of Operations Research", June 2013, https://hal.inria.fr/hal-00919136.

Publications of the year

Articles in International Peer-Reviewed Journal

[19] 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.

[20] A. AL GERBI, B. JOURDAIN, E. CLÉMENT.*Asymptotics for the normalized error of the Ninomiya-Victoir scheme*, in "Stochastic Processes and their Applications", September 2017, https://arxiv.org/abs/1601.05268 , https://hal-enpc.archives-ouvertes.fr/hal-01259915.

[21] 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.

[22] 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.

[23] 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, https://arxiv.org/abs/1509.01348 [*DOI :* 10.1007/S40072-017-0105-6], https://hal.archives-ouvertes.fr/hal-01192862.

[24] V. BALLY, L. CARAMELLINO.*Convergence and regularity of probability laws by using an interpolation method*, in "Annals of Probability", 2017, vol. 45, n^{o} 2, https://hal-upec-upem.archives-ouvertes.fr/hal-01109276.

[25] 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.

[26] M. BRIANI, L. CARAMELLINO, A. ZANETTE.*A hybrid approach for the implementation of the Heston model.*, in "IMA Journal of Management Mathematics", October 2017, vol. 28, n^{o} 4, p. 467-500, https://arxiv.org/abs/1307.7178 [*DOI :* 10.1093/IMAMAN/DPV032], https://hal.archives-ouvertes.fr/hal-00916440.

[27] 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.

[28] R. DUMITRESCU, M.-C. QUENEZ, A. SULEM.*American Options in an Imperfect Complete Market with Default*, in "ESAIM: Proceedings and Surveys", 2017, p. 1 - 10, https://hal.inria.fr/hal-01614741.

[29] 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^{o} 1, p. 532 - 559 [*DOI :* 10.1137/16M1109102], https://hal.inria.fr/hal-01614758.

[30] G. FORT, B. JOURDAIN, T. LELIÈVRE, G. STOLTZ.*Self-Healing Umbrella Sampling: Convergence and efficiency*, in "Statistics and Computing", January 2017, vol. 27, n^{o} 1, p. 147–168, https://arxiv.org/abs/1410.2109 [*DOI :* 10.1007/S11222-015-9613-2], https://hal.archives-ouvertes.fr/hal-01073201.

[31] 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^{o} 5, p. 2807-2861, https://arxiv.org/abs/1507.01087 , https://hal-enpc.archives-ouvertes.fr/hal-01171481.

[32] M. GAUDENZI, A. ZANETTE.*Fast binomial procedures for pricing Parisian/ParAsian options*, in "Computational Management Science", 2017, vol. 14, n^{o} 3, p. 313-331 [*DOI :* 10.1007/s10287-017-0278-5], https://hal.archives-ouvertes.fr/hal-01632859.

[33] Y. HU, B. ØKSENDAL, A. SULEM.*Singular mean-field control games*, in "Stochastic Analysis and Applications", June 2017, vol. 35, n^{o} 5, p. 823 - 851 [*DOI :* 10.1080/07362994.2017.1325745], https://hal.inria.fr/hal-01614747.

[34] B. LAPEYRE, E. QUINET.*A Simple GDP-based Model for Public Investments at Risk*, in "Journal of Benefit-Cost Analysis", 2017, vol. 8, n^{o} 01, p. 91 - 114 [*DOI :* 10.1017/BCA.2017.5], https://hal.archives-ouvertes.fr/hal-01666574.

Scientific Books (or Scientific Book chapters)

[35] 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 [*DOI :* 10.1007/978-3-319-65313-6_3], https://hal-enpc.archives-ouvertes.fr/hal-01670011.

Other Publications

[36] A. ALFONSI, J. CORBETTA, B. JOURDAIN.*Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems*, September 2017, https://arxiv.org/abs/1709.05287 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01589581.

[37] V. BALLY.*Upper bounds for the function solution of the homogenuous 2D Bolzmann equation with hard potential*, September 2017, https://arxiv.org/abs/1710.00695 - working paper or preprint, https://hal-upec-upem.archives-ouvertes.fr/hal-01593131.

[38] 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.

[39] V. BALLY, L. CARAMELLINO, G. POLY.*Convergence in distribution norms in the CLT for non identical distributed random variables*, January 2017, https://arxiv.org/abs/1606.01629 - working paper or preprint, https://hal-upec-upem.archives-ouvertes.fr/hal-01413548.

[40] V. BALLY, L. CARAMELLINO, G. POLY.*Non universality for the variance of the number of real roots of random trigonometric polynomials*, 2017, https://arxiv.org/abs/1711.03316 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01634848.

[41] V. BALLY, D. GOREAC, V. RABIET.*Regularity and Stability for the Semigroup of Jump Diffusions with State-Dependent Intensity*, July 2017, https://arxiv.org/abs/1707.02713 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01558741.

[42] R. DUMITRESCU, B. ØKSENDAL, A. SULEM.*Stochastic control of mean-field SPDEs with jumps*, May 2017, working paper or preprint, https://hal.inria.fr/hal-01527225.

[43] M. IBEN TAARIT, B. LAPEYRE.*A Forward Solution for Computing Risk-Neutral Derivatives Exposure*, December 2017, working paper or preprint [*DOI :* 10.2139/SSRN.2353308], https://hal.archives-ouvertes.fr/hal-01667100.

[44] B. JOURDAIN, A. KEBAIER.*Non-asymptotic error bounds for The Multilevel Monte Carlo Euler method applied to SDEs with constant diffusion coefficient*, August 2017, https://arxiv.org/abs/1708.07064 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01577874.

[45] A. KEBAIER, J. LELONG.*Coupling Importance Sampling and Multilevel Monte Carlo using Sample Average Approximation*, July 2017, https://arxiv.org/abs/1510.03590 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01214840.

References in notes

[46] M. AKIAN, J. MENALDI, A. SULEM.*On an Investment-Consumption model with transaction costs*, in "SIAM J. Control and Optim.", 1996, vol. 34, p. 329-364.

[47] M. AKIAN, A. SULEM, M. TAKSAR.*Dynamic optimisation of long term growth rate for a portfolio with transaction costs - The logarithmic utility case*, in "Mathematical Finance", 2001, vol. 11, p. 153-188.

[48] A. ALFONSI, A. SCHIED.*Optimal Trade Execution and Absence of Price Manipulations in Limit Order Book Models*, in "SIAM J. Finan. Math.", 2010, vol. 1, p. 490-522.

[49] H. AMINI, R. CONT, A. MINCA.*Resilience to Contagion in Financial Networks*, in "Mathematical Finance", 2013.

[50] V. BALLY.*An elementary introduction to Malliavin calculus*, Inria, Rocquencourt, February 2003, n^{o} 4718, http://hal.inria.fr/inria-00071868.

[51] V. BALLY, L. CARAMELLINO, A. ZANETTE.*Pricing American options by a Monte Carlo method using a Malliavin calculus approach*, in "Monte Carlo methods and applications", 2005, vol. 11, n^{o} 2, p. 97–133.

[52] D. BELL.*The Malliavin Calculus*, Pitman Monographs and Surveys in Pure and Applied Math., Longman and Wiley, 1987, n^{o} 34.

[53] T. BIELECKI, J.-P. CHANCELIER, S. PLISKA, A. SULEM.*Risk sensitive portfolio optimization with transaction costs*, in "Journal of Computational Finance", 2004, vol. 8, p. 39-63.

[54] F. BLACK, M. SCHOLES.*The pricing of Options and Corporate Liabibilites*, in "Journal of Political Economy", 1973, vol. 81, p. 637-654.

[55] I. ELSANOSI, B. ØKSENDAL, A. SULEM.*Some Solvable Stochastic control Problems with Delay*, in "Stochastics and Stochastics Reports", 2000.

[56] E. FOURNIÉ, J.-M. LASRY, J. LEBUCHOUX, P.-L. LIONS.*Applications of Malliavin calculus to Monte Carlo methods in Finance, II*, in "Finance & Stochastics", 2001, vol. 2, n^{o} 5, p. 201-236.

[57] E. FOURNIÉ, J.-M. LASRY, J. LEBUCHOUX, P.-L. LIONS, N. TOUZI.*An application of Malliavin calculus to Monte Carlo methods in Finance*, in "Finance & Stochastics", 1999, vol. 4, n^{o} 3, p. 391-412.

[58] N. C. FRAMSTAD, B. ØKSENDAL, A. SULEM.*Optimal Consumption and Portfolio in a Jump Diffusion Market with Proportional Transaction Costs*, in "Journal of Mathematical Economics", 2001, vol. 35, p. 233-257.

[59] J. GARNIER, G. PANANICOLAOU, T.-W. YANG.*Large deviations for a mean field model of systemic risk*, 2012, Manuscript, arXiv:1204.3536.

[60] P. GASSIAT, H. PHAM, M. SIRBU.*Optimal investment on finite horizon with random discrete order flow in illiquid markets*, in "International Journal of Theoretical and Applied Finance", 2010, vol. 14, p. 17-40.

[61] Y. KABANOV, M. SAFARIAN.*Markets with Transaction Costs: Mathematical Theory*, Springer Verlag, 2009.

[62] D. LAMBERTON, B. LAPEYRE, A. SULEM.*Application of Malliavin Calculus to Finance*, in "special issue of the journal Mathematical Finance", January 2003.

[63] P. MALLIAVIN.*Stochastic calculus of variations and hypoelliptic operators*, in "Proc. Inter. Symp. on Stoch. Diff. Equations", Kyoto, Wiley 1978, 1976, p. 195-263.

[64] P. MALLIAVIN, A. THALMAIER.*Stochastic Calculus of variations in Mathematical Finance*, Springer Finance, Springer Verlag, 2006.

[65] A. MINCA.*Modélisation mathématique de la contagion de défaut; Mathematical modeling of financial contagion*, Université Pierre et Marie Curie, Paris 6, September 5 2011.

[66] D. NUALART.*The Malliavin Calculus and Related Topics*, Springer–Verlag, 1995.

[67] D. OCONE, I. KARATZAS.*A generalized representation formula with application to optimal portfolios*, in "Stochastics and Stochastic Reports", 1991, vol. 34, p. 187-220.

[68] D. OCONE.*A guide to the stochastic calculus of variations*, in "Stochastic Analysis and Related Topics", H. KOERZLIOGLU, S. ÜSTÜNEL (editors), Lecture Notes in Math.1316, 1987, p. 1-79.

[69] F. RUSSO, P. VALLOIS.*Stochastic calculus with respect to continuous finite quadratic variation processes*, in "Stochastics and Stochastics Reports", 2000, vol. 70, p. 1–40.

[70] A. SULEM.*Dynamic Optimisation for a mixed Portfolio with transaction costs*, in "Numerical methods in Finance", 1997, p. 165-180, edited by L.C.G. Rogers and D.Talay, Cambridge University Press, Publications of the Newton Institute.

[71] U. ÇETIN, R. JARROW, P. PROTTER.*Liquidity risk and arbitrage pricing theory*, in "Finance and Stochastics", 2004, vol. 8.

[72] B. ØKSENDAL, A. SULEM, T. ZHANG.*Optimal control of stochastic delay equations and time-advanced backward stochastic differential equations*, in "Advances in Applied Probability", 2011, vol. 43, p. 572-596.

[73] B. ØKSENDAL, A. SULEM.*Optimal Consumption and Portfolio with both fixed and proportional transaction costs: A Combined Stochastic Control and Impulse Control Model*, in "SIAM J. Control and Optim.", 2002, vol. 40, p. 1765-1790.

[74] B. ØKSENDAL, A. SULEM.*Optimal stochastic impulse control with delayed reaction*, in "Applied Mathematics and Optimization", 2008, vol. 58, p. 243-255.

[75] B. ØKSENDAL.*An Introduction to Malliavin Calculus with Applications to Economics*, in "Lecture Notes from a course given 1996 at the Norwegian School of Economics and Business Administration (NHH)", September 1996, NHH Preprint Series.

Team MIMOVE

Middleware on the Move

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 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. Personnel

Research Scientists

Nikolaos Georgantas [Team leader, Inria, Researcher]
Vassilis Christophides [Inria, Advanced Research Position, from Apr 2017]
Renata Cruz Teixeira [Inria, Senior Researcher, from Apr 2017, HDR]
Valérie Issarny [Inria, Senior Researcher, HDR]

Faculty Member

Françoise Sailhan [CNAM, Associate Professor, CRCT, until Aug 2017]

External Collaborators

Rafael Angarita Arocha [Autre entreprise privée, from Sep 2017]
Zied Ben Houidi [Bell Labs (Alcatel), from Apr 2017]
Sara El Aouad [Autre entreprise privée, from May 2017]
Diego Neves Da Hora [Autre entreprise privée, from May 2017]

Françoise Sailhan [CNAM, from Sep 2017]

Technical Staff

Rachit Agarwal [Inria]
Rafael Angarita Arocha [Inria, until Sep 2017]
Georgios Bouloukakis [Inria, from Apr 2017]
Cong Kinh Nguyen [Autre entreprise privée]
Patient Ntumba [Inria]
Fadwa Rebhi [Inria, until Jul 2017]
Otto Tavares Nascimento [Inria, from Mar 2017]

PhD Students

Georgios Bouloukakis [Inria, until Mar 2017]
Fethi Dilmi [Inria, from Oct 2017]
Yifan Du [Inria, from Oct 2017]
Sara El Aouad [Autre entreprise privée, Apr 2017]
Diego Neves Da Hora [Inria, Apr 2017]
Radha Pallavali [Inria]
Sarah Wassermann [Inria, from Oct 2017]

Post-Doctoral Fellows

Eya Ben Ahmed [Inria, until Feb 2017]
Francesco Bronzino [Inria, from Apr 2017]
Bruno Lefevre [Inria, from Apr 2017]

Administrative Assistants

Maryse Desnous [Inria]
Nathalie Gaudechoux [Inria]

2. Overall Objectives

2.1. Overall Objectives

This year, Inria Muse team joined MiMove. We provide in Sections 2, 3 and 4 a first overview of our common research vision in the new MiMove team.

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 [2], [4].

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 [8], 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 outside. Our goal is to develop adaptative 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. Big data stream mining and processing

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. Advanced analysis of big data streams from sensors embedded in the environment and wearable or mobile user devices is bound to become a key area of data mining research as the number of applications requiring such processing increases. However, the *high data speed (velocity)* in conjunction with the *low data quality (veracity)* of IoT data streams challenges traditional Machine-Learning (ML) approaches assuming that a good quality training set is available a priori to learn models that may be effectively applied to new data collected under very similar conditions. As previous work has observed, data quality issues are detrimental to data analysis [0]. Good quality training data are typically the result of a thorough *data pre-processing* comprising data aggregation/integration, data cleaning/normalization, data dimensionality reduction, etc. The offline nature of these data engineering tasks represent nowadays one of the biggest technical barriers for supporting a high-value data analytics in real-time for various IoT settings (e.g., residential, industrial, urban, etc.). Furthermore, existing techniques for data quality management are usually agnostic of the analytical process that is to be applied on the data. For this reason, analysts either

[0] US National Research Council. 2013. Frontiers in Massive Data Analysis. The National Academies Press. http://www.nap.edu/openbook.php?record id=18374

clean everything, which is impossible for Big Data, or clean random subsets and hope for the best. We are interested in studying the following research questions: (a) what specific characteristics of the data quality (e.g., incomplete, extreme or erroneous values) led to the improvement, or lack thereof, in the performance of a ML algorithm (e.g., regression, classification)? (b) how we can identify influential data that are both unusual in the predictor variables and do not follow the general trend of the data relative to a prediction? (c) to what extend we can automate data pre-processing tasks (in particular cleaning) for specific streaming data analytics scenarios?

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 intend to focus on, further leveraging our research on emergent mobile distributed systems, large-scale mobile sensing & actuation, and mobile social crowd-sensing.

In a first step, we concentrate on the following specialized applications, which we investigate in close collaboration with other researchers, in particular as part of the dedicated Inria Project Lab *CityLab@Inria*:

- **Democratization of urban data for healthy cities.** The objective here is to integrate the various urban data sources, especially by way of crowd-Xing, to better understand city nuisances 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).
- **Socially-aware urban mobility.** Mobility within mega-cities is known as one of the major challenges to face urgently due to the fact that today's mobility patterns do not scale and to the negative effect on the environment and health. It is our belief that mobile social and physical sensing may significantly help in promoting the use of public transport, which we have started to investigate through empirical study based on the development and release of dedicated apps.
- **Social applications.** Mobile applications are being considered by sociologists as a major vehicle to actively involve citizens and thereby prompt them to become activists. This is especially studied with the Social Apps Lab at UC Berkeley. Our objective is to study such a vehicle from the ICT perspective and in particular elicit relevant middleware solutions to ease the development and development of such "*civic apps*".

Acknowledging the need for collaborative research in the application domain of smart cities, MiMove is heavily involved and actually leading CityLab@Inria [0]. The Inria Project Lab CityLab is focused on the study of ICT solutions promoting social sustainability in smart cities, and involves the following Inria project-teams in addition to MiMove: CLIME, DICE, FUN, MYRIADS, SMIS, URBANET and WILLOW. CityLab further involves 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. We note that Valérie Issarny acts as scientific manager of Inria@SiliconValley and is currently visiting scholar at CITRIS at UC Berkeley. In this context, MiMove researchers are working closely with colleagues of UC Berkeley, including researchers from various disciplines interested in smart cities (most notably sociologists).

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 "Interne 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 causes of potential problems and—ideally—ways to fix them. We intend to develop a set of easy-to-use home network diagnosis tools that can reliably identify performance and functionality shortcomings rooted in the home. The development of home network 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, 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. Third, 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 conceiving methods for two application scenarios: (i) when the end user in the home deploys our diagnostic tools either on the home gateway (the gateway often combines a DSL/cable modem and an access point; it connects the home network to the ISP) or on devices connected to the home network and (ii) when ISPs collect measurements from homes of subscribers and then correlate these measurements to help identify problems.

Assisting end users. We are developing algorithms to determine whether network performance problems lie inside or outside the home network. Given that the home gateway connects the home with the rest of the Internet, we are designing an algorithm (called *HoA*) that analyzes traffic that traverses the gateway to distinguish access link and home network bottlenecks. A measurement vantage point on the gateway is key for determining if the performance bottleneck lies within the home network or the access ISP, but we also need to deploy diagnosis tools in end-devices. First, some users may not want (or not know how) to deploy a new home gateway in their homes. Second, some problems will be hard to diagnose with only the vantage point of the gateway (for example, when a device cannot send traffic or when the wireless is poor in certain locations of a home). We can obtain more complete visibility by leveraging *multiple* measurement nodes around the home, potentially including the home gateway, all participating jointly in the measurement task. We have an ongoing project to realize a home network analyzer as a web-based measurement application built on top of our team's recently developed browser-based measurement platform, *Fathom*. To integrate the home gateway in the analyzer, we plan to engage the BISmark Project. BISmark already provides a web server as well as

[0] http://citylab.inria.fr

extensive configurability, allowing us to experiment freely with both passive as well as active measurements. We must develop a home network analyzer that can first discover the set of devices connected to the home network that can collaborate on the diagnosis task. We will then develop tomography algorithms to infer where performance problems lie given measurements taken from the set of available vantage points.

Assisting Internet Service Providers (ISPs). Our discussions with several large access ISPs reveal that service calls are costly, ranging from $9–25 per call, and as many as 75% of service calls from customers are usually caused by problems that have nothing to do with the ISP. Therefore, ISPs are eager to deploy techniques to assist in home network diagnosis. In many countries ISPs control the home gateway and set-top-boxes in the home. We plan to develop more efficient mechanisms for home users to report trouble to their home ISP and consequently reduce the cost of service calls. This project is in collaboration with Technicolor and Portugal Telecom. Technicolor is a large manufacturer of home gateways and set-top-boxes. Portugal Telecom is the largest broadband access provider in Portugal. Technicolor already collects data from 200 homes in Portugal. We are working with the data collected in this deployment together with controlled experiments to develop methods to diagnose problems in the home wireless.

4.3. Quality of Experience

An increasing number of residential users consume online services (e.g., VoD, Web browsing, or Skype) in their everyday activities (e.g., for education or entertainment purposes), using a variety of devices (e.g., tablets, smartphones, laptops). A high Quality of Service (QoS) is essential for sustaining the revenue of service providers, carriers, and device manufactures. Yet, the perceived Quality of Experience (QoE) of users is far from perfect e.g., videos that get stalled or that take a long time to load. Dissatisfied users may change Internet Service Providers (ISPs) or the online services. Hence, the incentives for measuring and improving QoE in home networks are high while mapping network and application QoS to QoE is a challenging problem. In this work we have focused in measuring several network Quality-of-Service (QoS) metrics, such as latency and bandwidth, both in residential Wi-Fi as well as broadband networks, homes are using for connecting to the Internet.

The WiFi Context. Residential Wi-Fi performance, however, is highly variable. Competing Wi-Fi networks can cause contention and interference while poor channel conditions between the station and the access point (AP) can cause frame losses and low bandwidth. In some cases, the home Wi-Fi network can bottleneck Internet access. While problems in the Wi-Fi network may affect several network QoS metrics, users will typically only notice a problem when poor Wi-Fi affects the QoE of Internet applications. For example, a Wi-Fi network with low bandwidth may go unnoticed unless the time to load Web pages increases significantly. A user observing degraded QoE due to Wi-Fi problems may mistakenly assume there is a problem with the Internet Service Provider (ISP) network. Our discussions with residential ISPs confirm that often customers call to complain about problems in the home Wi-Fi and not the ISP network.

Prior work has focused on QoS metrics for some applications (e.g., on-line video, Web browsing, or Skype) with no attempt to identify when Wi-Fi quality affects QoE. We are particularly interested in assisting ISPs to predict when home Wi-Fi quality degrades QoE. ISPs can use this system to detect customers experiencing poor QoE to proactively trigger Wi-Fi troubleshooting. ISPs often control the home AP, so we leverage Wi-Fi metrics that are available on commercial APs. Detecting when Wi-Fi quality degrades QoE using these metrics is challenging. First, we have no information about the applications customers are running at any given time. ISPs avoid capturing per-packet traffic traces from customers, because of privacy considerations and the overload of per-packet capture. Thus, we must estimate the effect of Wi-Fi quality on QoE of popular applications, which most customers are likely to run. In this context, we study Web as a proof of concept, as a large fraction of home traffic corresponds to Web. Second, application QoE may be degraded by factors other than the Wi-Fi quality (e.g., poor Internet performance or an overloaded server). Although a general system to explain any QoE degradation would be extremely helpful, our monitoring at the AP prevents us from having the end-to-end view necessary for such general task. Instead, we focus on identifying when Wi-Fi quality degrades QoE. Finally, Wi-Fi metrics available in APs are coarse aggregates such as the average PHY rate or the fraction of busy times. It is open how to effectively map these coarse metrics into QoE.

Predicting QoE. Clearly, different actors in the online service chain (e.g., video streaming services, ISPs) have different incentives and means to measure and affect the user QoE. Uncovering statistically equivalent subsets of QoS metrics across and within levels provides actionable knowledge for building QoE predictors. To achieve this goal, we leverage recent advances on feature selection algorithms to exploit available experimental evidence of the joint probability distributions of QoE/QoS metrics. This type of statistical reasoning will enable us to determine local causal relationships between a target QoE variable, seen as effect, and multiple QoS metrics across or within levels, seen as causes. Such data-driven analysis is justified by the multiplicity of dependencies that exist between network or application QoS metrics as different adaptation mechanisms (e.g., TCP congestion avoidance, HTTP bitrate adaptation) are activated at each level in real life. Building optimal predictors based on (eventually several) probabilistically minimal subsets of features opens the way for a principled comparison of the predictors.

4.4. Crowd-sourced Information Filtering and Summarization

With the explosion of the People-centric Web, there is a proliferation of crowd-sourced content either under the form of qualitative reviews (mainly textual) and quantitative ratings (as 5 star ratings) regarding diverse products or services or under the form of various "real-time" feedback events (e.g., re-tweets, replies, likes, clicks, etc.) on published web content (ranging from traditional news, TV series, and movies to specialized blogs and posts shared over social networks). Such content captures the wisdom of the crowd and is valuable information source for building collaborative filtering systems and text summarization tools coping with information overload. For example, they can assist users to pick the most interesting web pages (e.g. Delicious) or to choose which movie to watch next (e.g. Netflix).

Implicit Feedback in Communities of a Place. We are initially interested in addressing one of the main limitation of collaborative filtering systems namely, the strong user engagement required to provide the necessary input (e.g., regarding their friends, tags or sites of preference) which is usual platform specific (i.e., for a particular social network, tagging, or bookmark system). The lack of user engagement translates into cold start and data sparsity. To cope with this limitation, we are developing a system called WeBrowse that passively observes network traffic to extract user clicks (i.e., the URLs users visit) for group of people who live, study, or work in the same place. Examples of such communities of a place are: (i) the students of a campus, (ii) the people living in a neighbourhood or (iii) researchers working in the same site. WeBrowse then promotes the hottest and most popular content to the community members sharing common interests.

Personalized Review Summarization. Finally, we are interested in helping people to take informed decisions regarding their shopping or entertainment activities. The automated summarization of a review corpus (for example, movie reviews from Rotten Tomatoes or IMDB; or restaurant reviews from Yelp) aims to assist people to form an opinion regarding a product/service of interest, by producing a coherent summary that is helpful and can be easily assimilated by humans. We are working on review summarisation methods that combine both objective (i.e., related to the review corpus) and subjective (i.e., related to the end-user interests) interestingness criteria of the produced reviews. In this respect we are exploiting domain models (e.g., Oscar's merit categories for movies) to elicit user preferences and mine the aspects of products/services actually commented in the textual sentences of reviews. For example, different summaries should be produced when a user is more interested in the actors' performance rather than the movie story. We are particularly interested in extracting automatically the signatures of aspects (based on a set of seed terms) and rank review sentences on their importance and relevance w.r.t. the aspects they comment. Last but not least we are optimizing the automatically constructed summary w.r.t. to a number of criteria such as the number of the length of included sentences from the original reviews, the polarity of sentiments in the described aspects, etc.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

- R. Teixeira was selected to appear in the 2017 list of "N2Women: Stars in Computer Networking and Communications".
- The AppCivist project, which is a joint initiative between the Social Apps Lab at UC Berkeley and the MiMove team at Inria, won the 2016-17 Chancellor's Award for Public Service in the category of Campus-Community Partnership in collaboration with the City of Vallejo [20].

BEST PAPERS AWARDS :
[19] **IEEE International Conference on Collaboration and Internet Computing**. R. ANGARITA, N. GEORGANTAS, V. ISSARNY.

6. New Software and Platforms

6.1. SoundCity - Ambiciti

KEYWORDS: Crowd-sensing - Mobile application
FUNCTIONAL DESCRIPTION: Is your exposure to noise too high on certain days? How is air pollution in your street? Will air quality improve in the next hours? Do you want to measure the noise pollution on the way between your home and your office? What pollution levels are considered harmful for your health? Ambiciti (previously SoundCity) provides answers to these questions and many others.

Noise pollution. Ambiciti (previously called SoundCity) measures the actual noise levels to which you are exposed. Ambiciti can monitor noise levels throughout the day and inform you about your instantaneous, hourly and daily exposures. If you want to contribute to the improvement of the noise map in your city, you can anonymously send your measurements.

Air pollution. Ambiciti computes the air quality index in your region or at the exact location where you stand. You can also access to forecasts and find information about the main pollutants. Depending on your location, you may have access to hourly air quality maps, at street resolution, in real time and for the next two days. Currently, only Paris (France) enjoys such fine maps, but other cities are on the way to be included.

Since 2017, the software is exclusively licensed to the Ambiciti start-up company.

- Authors: Fadwa Rebhi, Pierre-Guillaume Raverdy, Cong Kinh Nguyen, Rajiv Bhatia, Valérie Issarny and Vivien Mallet
- Partners: Ambientic - The Civic Engine
- Contact: Valérie Issarny

6.2. 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.

Social Entities. Social entities can be humans or systems. They can create a profile in the USNB and link it with their OSNSs identities. Social entities can also choose the OSNS identity they want to use when contacted through the USNB or specify one or more OSNS identities for message reception concerning specific events or senders.

Personae. Personae are USNB entities interacting with users within concrete OSNSs or systems, achieving interoperability between heterogeneous OSNSs. New personae can be developed, registered in the USNB, discovered and used to include additional OSNSs.

Privacy & Security. The USNB is designed to be as less intrusive as possible. It does not ask users their credentials nor any kind of authorization concerning their OSNS accounts.

- Authors: Rafael Angarita Arocha, Nikolaos Georgantas and Valérie Issarny
- Contact: Valérie Issarny
- URL: https://gitlab.inria.fr/usnb/universal-social-network-bus

6.3. 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/

6.4. 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

6.5. HostView

KEYWORDS: Quality of Experience - Network monitoring
FUNCTIONAL DESCRIPTION: End-host performance monitoring and user feedback reporting.

- Participants: Anna-Kaisa Pietilainen, Francesco Bronzino, George Rosca and Renata Cruz Teixeira
- Contact: Renata Cruz Teixeira
- URL: https://github.com/inria-muse/hostview-win

6.6. 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://github.com/sidutta/vsb

7. New Results

7.1. Living with Interpersonal Data: Observability and Accountability in the Age of Pervasive ICT

Participants: Murray Goulden (University of Nottingham), Peter Tolmie (University of Nottingham), Richard Mortier (University of Cambridge), Tom Lodge (University of Nottingham), Anna-Kaisa Pietilainen (Google), Renata Teixeira

The Internet of Things, alongside existing mobile digital technologies, herald a world in which pervasive sensing constantly captures data about us. Simultaneous with this technology programme are moves by policymakers to shore up the digital economy, through the legislating of new models of data management. These moves seek to give individuals control and oversight of their personal data. Within shared settings the consequences of these changes are the large-scale generation of interpersonal data, generated by and acting on the group rather than individual. We consider how such systems create new forms of observability and hence accountability amongst members of the home, and draw on the work of Simmel (1906) and Goffman (1971) to explore how these demands are managed. Such management mitigates the more extreme possibilities for domestic monitoring posited by these systems, yet without careful design there remains a considerable danger of unanticipated negative consequences.

7.2. Predicting the effect of home Wi-Fi quality on QoE

Participants: Diego da Hora (Telecom Paris Tech), Karel van Doorselaer (Technicolor), Koen van Oost (Technicolor), Renata Teixeira

Poor Wi-Fi quality can disrupt home users' internet experience, or the Quality of Experience (QoE). Detecting when Wi-Fi degrades QoE is extremely valuable for residential Internet Service Providers (ISPs) as home users often hold the ISP responsible whenever QoE degrades. Yet, ISPs have little visibility within the home to assist users. Our goal is to develop a system that runs on commodity access points (APs) to assist ISPs in detecting when Wi-Fi degrades QoE. Our first contribution is to develop a method to detect instances of poor QoE based on the passive observation of Wi-Fi quality metrics available in commodity APs (e.g., PHY rate). We use support vector regression to build predictors of QoE given Wi-Fi quality for popular internet applications. We then use K-means clustering to combine per-application predictors to identify regions of Wi-Fi quality where QoE is poor across applications. We call samples in these regions as poor QoE samples. Our second contribution is to apply our predictors to Wi-Fi metrics collected over one month from 3,479 APs of customers of a large residential ISP. Our results show that QoE is good on the vast majority of samples of the deployment, still we find 11.6% of poor QoE samples. Worse, approximately 21% of stations have more than 25% poor QoE samples. In some cases, we estimate that Wi-Fi quality causes poor QoE for many hours, though in most cases poor QoE events are short.

7.3. Narrowing the gap between QoS metrics and Web QoE using Above-the-fold metrics

Participants: Diego da Hora (Telecom Paris Tech), Alemnew Sheferaw Asrese (Aalto University), Vassilis Christophides, Renata Teixeira, Dario Rossi (Telecom Paris Tech)

Page load time (PLT) is still the most common application Quality of Service (QoS) metric to estimate the Quality of Experience (QoE) of Web users. Yet, recent literature abounds with proposals for alternative metrics (e.g., Above The Fold, SpeedIndex and variants) that aim at better estimating user QoE. The main purpose of this work is thus to thoroughly investigate a mapping between established and recently proposed objective metrics and user QoE. We obtain ground truth QoE via user experiments where we collect QoS metrics over 3,000 Web accesses annotated with explicit user ratings in a scale of 1 to 5, which we make available to the community. In particular, we contrast domain expert models (such as ITU-T and IQX) fed with a single QoS metric, to models trained using our ground-truth dataset over multiple QoS metrics as features. Results of our experiments show that, albeit very simple, ex- pert models have a comparable accuracy to machine learning approaches. Furthermore, the model accuracy improves considerably when building per-page QoE models, which may raise scalability concerns as we discuss.

7.4. Performance Modeling of the Middleware Overlay Infrastructure of Mobile Things

Participants: Georgios Bouloukakis, Nikolaos Georgantas, Valérie Issarny.

Internet of Things (IoT) applications consist of diverse Things (sensors and devices) in terms of hardware resources. Furthermore, such applications are characterized by the Things' mobility and multiple interaction types, such as synchronous, asynchronous, and streaming. Middleware IoT protocols consider the above limitations and support the development of effective applications by providing several Quality of Service features. These features aim to enable application developers to tune an application by switching different levels of response times and delivery success rates. However, the profusion of the developed IoT protocols and the intermittent connectivity of mobile Things, result to a non-trivial application tuning. In this work, we model the performance of the middleware overlay infrastructure using Queueing Network Models. To represent the mobile Thing's connections/disconnections, we model and solve analytically an ON/OFF queueing center. We apply our approach to streaming interactions with mobile peers. Finally, we validate our model using simulations. The deviations between the performance results foreseen by the analytical model and the ones provided by the simulator are shown to be less than 5%.

7.5. USNB: Enabling Universal Online Social Interactions

Participants: Rafael Angarita, Nikolaos Georgantas, Valérie Issarny.

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. This work is concerned with relieving users from such a burden by enabling universal online social interactions. The contributions of this work span: (1) a model of the universal social network bus (USNB) for OSNS interoperability; (2) a prototype for universal online social interactions that builds upon the proposed model; and (3) a preliminary experimental evaluation involving 50 participants. Results show that people are positive about the solution as they are able to reach out a larger community of users independently of the OSNSs they use.

7.6. Opportunistic Multiparty Calibration for Robust Participatory Sensing

Participants: Françoise Sailhan, Valérie Issarny, Otto Tavares Nascimento.

While bringing massive-scale sensing at low cost, mobile participatory sensing is challenged by the low accuracy of the sensors embedded in and/or connected to the smartphones. The mobile measurements that are collected need to be corrected so as to accurately match the phenomena being observed. This paper addresses this challenge by introducing a multi-hop, multiparty calibration method that operates in the background in an automated way. Using our method, sensors that are within a relevant sensing (and communication) range coordinate so that the observations of the participating (previously) calibrated sensors serve calibrating the other participants. As a result, our method is particularly well suited for participatory sensing within crowd meetings, as as for instance within public spaces. Our solution leverages multivariate linear regression, together with robust regression so as to discard the measurements that are of too low quality for being meaningful. To the best of our knowledge, we are the first to introduce a multiparty calibration algorithm, while previous work in the area focused on pairwise calibration. This work further introduces a supporting prototype implemented over Android, and related experiment in the context of noise sensing. We show that the proposed multiparty calibration system enhances the accuracy of the mobile noise sensing application.

7.7. Extracting usage patterns of home IoT devices

Participants: Vassilis Christophides, Gevorg Poghosyan (Insight Centre for Data Analytics), Ioannis Pefkianakis (Hewlett Packard Labs), Pascal Le Guyadec (Technicolor)

We have initially investigated how data analytics for Machine-to-Machine (M2M) data (connectivity, performance, usage) produced by connected devices in residential Intranet of Things, could support novel *home automation services* that enrich the living experience in smart homes. We have investigated new data mining techniques that go beyond binary association rule mining for traditional market basket analysis, considered by previous works. We design a multidimensional pattern mining framework, which collects raw data from operational home gateways, it discretizes and annotates the raw data, it produces traffic usage logs which are fed in a multidimensional association rule miner, and finally it extracts home residents' habits. Using our analysis engine, we extract complex device co-usage patterns of 201 residential broadband users of an ISP, subscribed to a n-play service. Such fine-grained device usage patterns provide valuable insights for emerging use cases, such as adaptive usage of home devices (aka horizontal integration of things). Such use cases fall within the wider area of human-cognizant Machine-to-Machine communication aiming to predict user needs and complete tasks without users initiating the action or interfering with the service. While this is not a new concept, according to Gartner cognizant computing is a natural evolution of a world driven not by devices but collections of applications and services that span across multiple devices, in which human intervention becomes as little as possible, by analyzing past human habits. To realize this vision, we are interested in co-usage patterns featuring spatio-temporal information regarding the context under which devices have been actually used in homes. For example, a network extender which is currently turned off, could be turned on at a certain day period (e.g., evening) when it has been observed to be highly used along with other devices (e.g., a laptop or a

tablet). Alternatively, the identification of frequent co-usage of particular devices at a home (say iPhone with media player), could be used by a things' recommender to advertise the same set of devices at another home (say another iPhone user could be interested in a media player).

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

"Monitoring and diagnosis of Internet QoE", Google Faculty Award to R. Teixeira (Inria) and D. Choffnes (Northeastern University), 2017.

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 [correspondent], Fadwa Rebhi.

- **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, URBANET, WILLOW
- **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, Romain Rouvoy.

- **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 CHOReVOLUTION

Participants: Nikolaos Georgantas [correspondent], Valérie Issarny [correspondent].

Name: CHOReVOLUTION – *Automated Synthesis of Dynamic and Secured Choreographies for the Future Internet*

URL: http://www.chorevolution.eu

Type: Research & Innovation Action (ICT)

Topic: Tools and Methods for Software Development

Period: [January 2015 - December 2017]

Partners: CEFRIEL (Italy), Inria MiMove, OW2 Consortium (France), Thales Communications S.A. (France) [**coordinator**], Università degli Studi dell'Aquila (Italy) [**scientific leader**], Softeco Sismat SrL (Italy), Tirasa (Italy), Viktoria Swedish ICT (Sweden).

The Future Internet (FI) represents an age of unprecedented opportunities for social, economic, and business growth thanks to the global scale connectivity of the virtual as well as of the physical world. This indeed opens up a wealth of innovative and revolutionary real-life scenarios, as for instance illustrated by the smarter cities perspectives where envisioned scenarios significantly ease daily human activities and give support for the growth of new markets and employment opportunities. However, leveraging the FI for the development of innovative software applications remain a challenging task even though major enablers are readily available by ways of service-oriented and cloud computing. It is in particular our vision that enabling the choreography of FI services shall play a significant role in the provisioning of innovative applications. However, existing choreography-based service composition approaches are rather static and are poorly suited to the need of the FI that is a highly dynamic networking environment, further bringing together highly heterogeneous services ranging from Thing- to Business-based services that span different security domains. As a result, the technology is not mature enough for market take-up. CHOReVOLUTION elevates the Readiness Level of existing choreography technologies in order to drop the dynamism and cross-organization security barriers via the automated synthesis of dynamic and secured choreographies in the FI. To meet its objectives, CHOReVOLUTION undertakes both research and innovation tasks. The former concentrates on choreography modelling, synthesis, adaptation, service bus, security, and cloud; the latter focus on industrial validation, development support and integration platform, and the establishment of a CHOReVOLUTION community and market take- up. Last but not least CHOReVOLUTION outcomes are assessed by experimenting with new applications in the field of Intelligent Transportation Systems.

9.2.1.2. H2020 ICT FIESTA-IoT

Participants: Valérie Issarny [correspondent], Nikolaos Georgantas [correspondent].

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.

9.2.2. Collaborations in European Programs, Except FP7 & H2020

9.2.2.1. EIT Digital Env&You

Participant: Valérie Issarny [correspondent].

Name: Env&You – *Personalizing environmental science for your home, your neighborhood and your life*

URL: http://ambiciti.io

Period: [January 2017 - December 2017]

Partners: Ambiciti (F), Forum Virium Helsinki (FI), Inria CLIME/ANGE, Inria MIMOVE [**coordinator**], NumTech (F), TheCivicEngine (USA).

There is a clear, and probably increasing, desire from the citizens to better know their individual exposure to pollution. Partial solutions exist to the exposure data problem but each focuses on one or another domain of information – crowdsourcing exposure, translating government open data to usable consumer information, harnessing social media information, harnessing biometrics – what is unique about Env&You is that we assimilate a multi-dimensional picture of exposure and provide the integrated information to citizen, government, and business use (spanning: B2G, B2B and B2C business cases).

9.2.2.2. EIT Digital CivicBudget

Participants: Valérie Issarny [correspondent], Nikolaos Georgantas [correspondent].

Name: CivicBudget – *Software platform supporting Internet-based participatory budgeting campaigns*

Period: [January 2017 - December 2017]

Partners: CITRIS@UCB (USA), Inria MIMOVE, MissionsPubliques (F) [**coordinator**], Nexus (DE), and TU Berlin (DE).

Many cities in Europe and the U.S.A, and around the world, commit a percentage of their annual budget (often 5%) to implement citizen-proposed projects through a process called Participatory Budgeting (PB). However, supporting urban-scale participatory budgeting campaigns is greatly challenged as it still principally relies on physical meetings. CivicBudget addresses this challenge by leveraging latest ICT so as to promote urban-scale inclusion. CivicBudget fosters a new and inclusive urban public sphere of citizenship. It is especially designed for community groups and activists who want to participate in the PB process. City governments will also be able to promote its use. CivicBudget will facilitate the mobilization of residents both to promote their proposals and to monitor their progress through the PB process to implementation.

9.3. International Initiatives

9.3.1. Inria International Labs

Valérie Issarny acts as scientific manager of the Inria@Silicon Valley program (https://project.inria.fr/inria-siliconvalley/) since summer 2013; she is visiting scholar at the EECS Department of University of California, Berkeley, and hosted by CITRIS.

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 (United States) - Nick Feamster

Start year: 2017

Website: 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

Participant: Nikolaos Georgantas [correspondent].

Title: *Adaptive enactment of service choreographies*

International Partner: Universidade Federal de Goiás (UFG), Brazil - Fabio Costa

Start year: 2016

Website: 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 and his student Arash Molavi): we are working on monitoring and diagnosing Internet QoE.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

9.4.1.1. Internships

Kushagra Singh (from Jun 2017 until Jul 2017)

Internship funded by H2020 CHOReVOLUTION project.

Subject: *Towards correction of outliers in spatial dataset*

Institution: Indraprastha Institute of Information Technology (IIIT) Delhi (India)

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

- Valérie Issarny is visiting scholar at the EECS Department at UC Berkeley; she is hosted by CITRIS in the context of which she carries out collaborative research in the area of smart cities and acts as scientific coordinator of the Inria@SiliconValley program.
- Rachit Agarwal was visiting professor at Fundacao Getulio Vargas (FGV), Rio De Janerio, Brazil (from Jun 2017 until Aug 2017). He was hosted at EMAp (Escola de Matematica Aplicada) department within FGV. He taught a Network Science course to Master's students.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Valérie Issarny is co-chair of the BIS'17 workshop, which is the yearly workshop organized by Inria@SiliconValley to present the state of scientific collaborations and to work on proposals for future ambitious joint projects.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Valérie Issarny is PC co-chair of ICSE-SEIS'2018 - The Software Engineering in Society Track of the ICSE'18 conference.
- R. Teixeira, Co-chair of the program committee of ACM IMC 2017.
- R. Teixeira, Co-chair of the program committee of ACM/ISOC ANRW 2017.

10.1.2.2. Member of the Conference Program Committees

- R. Teixeira, PC member of 2017 ACM SIGCOMM conference.
- Vassilis Christophides, PC member of 34th IEEE International Conference on Data Engineering.
- Vassilis Christophides, PC member of 2nd IFIP/IEEE International Workshop on Analytics for Network and Service Management (AnNet 2017)
- Nikolaos Georgantas is PC member of the following international conferences: FiCloud'17, ICSE'17 Poster Track, SAC'17 &'18, WETICE'17, AmI'17, SOSE'17&'18
- Nikolaos Georgantas is PC member of the following international workshops: CASA'17, MRT'17, SERENE'17, IoT-ASAP'17.
- Valérie Issarny is PC member of the following International Conferences: AIMS'17 & 18, EUROSYS'17, FASE'17 & 18, ICDCS'18, ICSE-SEIS'17, ICSE'18, ISEC'17, SEAMS'17.
- Valérie Issarny is PC member of the following international workshops: AMS'18, InterIoT'17.
- R. Agarwal, PC member of 10th International Conference of Contemporary Computing.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Nikolaos Georgantas is associate editor of the International Journal of Ambient Computing and Intelligence (IJACI);
- Valérie Issarny is associate editor of the IEEE transactions on Software Engineering.
- Valérie Issarny is associate editor of the IEEE Transactions on Services Computing.
- Valérie Issarny is associate editor of the Springer JISA Journal of Internet Services and Applications.

10.1.4. Invited Talks

- R. Teixeira, invited talk at Politecnico di Torino, "Home network QoE diagnosis", October 2017.
- R. Teixeira, jounées scientifiques Inria, "Home network QoE diagnosis", June 2017.
- R. Teixeira, LINCS workshop, "WeBrowse: a passive crowdsourced content discovery platform for communities of a place", June 2017.
- R. Teixeira, keynote speaker at the Simpósio Brasileiro de Redes de Computadores (Brazilian networking conference), "Home network QoE diagnosis", May 2017.

- R. Teixeira, tutorial at the Simpósio Brasileiro de Redes de Computadores (Brazilian networking conference), “Internet Measurements: Bandwidth”, May 2017.
- R. Teixeira, tutorial at the Simpósio Brasileiro de Redes de Computadores (Brazilian networking conference), “Internet Traffic Measurements”, May 2017.
- R. Agarwal: “Extracting mobility information from CDRs and using it towards achieving enhanced dissemination in wireless networks”, EMAp FGV seminar, 3 August 2017, Rio De Janerio, Brazil.
- Vassilis Christophides, “Web-scale Blocking, Iterative and Progressive Entity Resolution”, with K. Stefanidis, V. Efthymiou ICDE 2017.
- Vassilis Christophides, “Web-scale Entity Resolution”, Invited lecture at the 3rd EGC Winter School (Ecole EGC), 23 & 24 January 2017, Grenoble, France.
- Nikolaos Georgantas, “Enabling emergent mobile systems in the IoT: Functional and QoS interoperability aspects at the middleware layer”, Invited keynote talk at MSPN’17 (International Conference on Mobile, Secure and Programmable Networking), June 2017.

10.1.5. Leadership within the Scientific Community

- V. Issarny, Council member, ACM Europe (since July 2017).
- R. Teixeira, Vice-chair of ACM SIGCOMM (July 2013 – June 2017).
- R. Teixeira, Chair of the ACM SIGCOMM Industrial Liaison Board (October 2013 – June 2017).
- V. Christophides, Member of the EDBT Association (since 2014).

10.1.6. Scientific Expertise

- R. Teixeira, Reviewer of H2020 projects: CogNet, SUPERFLUIDITY, MONROE, MAMI.
- R. Teixeira, Technical advisory board of the project "Mapping of Broadband Services in Europe".
- Valérie Issarny is elected member of the *Commission d’Evaluation Inria*.
- Valérie Issarny is member of the Inria DR2 (Research Director position) selection committee.

10.1.7. Research Administration

- Valérie Issarny is scientific coordinator of Inria@Silicon Valley and CityLab@Inria;
- Nikolaos Georgantas is member of the PhD monitoring committee at Inria Paris.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master : R. Teixeira, “Methodology for research in networking”, 10h eqTD, M2, UPMC, France

Master : R. Teixeira, “Network Metrology”, 22h CM, M2, UPMC, France

Master: Rachit Agarwal, “Network Science”, semester course at Fundacao Getulio Vargas (FGV), Rio De Janerio, Brazil

Master: Rachit Agarwal, “Urban data democratization and its application to access urban concepts” as part of “Gestion de données ambiantes et internet des objets”, 9 hours eqTD, Niveau M2, UVSQ, France

E-learning

Valérie Issarny, MOOC *Villes intelligentes : défis technologiques et sociétaux* , 5 weeks, 4,085 registered students in the session of March 2017.

E-learning

Mooc : R. Teixeira, “Internet Measurements: a Hands-on Introduction”, 5 weeks in FUN platform, 1,204 registered students.

10.2.2. Supervision

PhD : Diego da Hora, “Predicting Home Wi-Fi QoE from Passive Measurements on Commodity Access Points”, UPMC, 27/04/2017, R. Teixeira and K. van Doorselaer.

PhD: Georgios Bouloukakis, “Enabling Emergent Mobile Systems in the IoT: from Middleware-layer Communication Interoperability to Associated QoS Analysis”, UPMC, 01/08/2017, Nikolaos Georgantas and Valérie Issarny.

PhD: Amel Belaggoun, “Adaptabilité et reconfiguration des systèmes temps-réel embarquées”, UPMC, 10/10/2017, Valérie Issarny and Ansgar Radermacher (CEA-LISE).

PhD: Vassilis Efthimiou, “Entity resolution in the Web of Data”, University of Crete, 27/09/2017, V. Christophides.

PhD in progress: F. Dilmi, “End-to-end monitoring and diagnosis of video Quality of Experience”, October 2017, R. Teixeira.

PhD in progress: S. Wassermann, “Passive analysis and opitimization of Internet Quality of Experience”, October 2017, R. Teixeira.

PhD in progress: S. El Aouad, “Building a personalized summary from movie reviews”, May 2013, V. Christophides, R. Teixeira, P. Perez.

PhD in progress : Radha Pallavali, “Dynamic adaptation of middleware-layer protocols for emergent mobile systems”, UPMC, November 2016, Nikolaos Georgantas and Valérie Issarny.

PhD in progress: Yifan Du, “In-network collaborative crowd-Xing”, UPMC, October 2017, Valérie Issarny and Françoise Sailhan.

10.2.3. Juries

Nikolaos Georgantas was examiner for the PhD thesis of Adrien Carteron, *Une approche événementielle pour le développement de services multi-métiers dédiés à l’assistance domiciliaire*, defended in December 2017, Université de Bordeaux, France, advised by Charles Consel and Nic Volanschi.

Valerie Issarny was examiner for the HDR of Sonia Ben Mokhtar, *Contributions To Building Reliable Distributed Systems*, defended in December 2017, INSA Lyon, France.

11. Bibliography

Major publications by the team in recent years

[1] S. BEN MOKHTAR, D. PREUVENEERS, N. GEORGANTAS, V. ISSARNY, Y. BERBERS.*EASY: Efficient SemAntic Service DiscoverY in Pervasive Computing Environments with QoS and Context Support*, in "Journal of Systems and Software, Special Issue on Web Services Modelling and Testing", 2008, vol. 81, n^{o} 5, p. 785-808.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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° 4, p. 1025–1038, https://hal.inria.fr/hal-01097439.

[7] 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.

[8] 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.

[9] 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° 1, p. 23-45 [*DOI :* 10.1007/s13174-011-0021-3], http://hal.inria.fr/inria-00588753/en.

[10] 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

[11] G. BOULOUKAKIS.*Enabling Emergent Mobile Systems in the IoT: from Middleware-layer Communication Interoperability to Associated QoS Analysis*, Inria Paris, August 2017, https://hal.inria.fr/tel-01592623.

[12] D. DA HORA.*Predicting Home Wi-Fi QoE from Passive Measurements on Commodity Access Points*, Université Paris 6 (UPMC), April 2017, https://hal.inria.fr/tel-01670997.

Articles in International Peer-Reviewed Journal

[13] 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° 2, p. 1 - 21 [*DOI :* 10.1145/3041025], https://hal.inria.fr/hal-01505879.

[14] M. GOULDEN, P. TOLMIE, R. MORTIER, T. LODGE, A.-K. PIETILAINEN, R. TEIXEIRA.*Living with interpersonal data: Observability and accountability in the age of pervasive ICT*, in "New Media and Society", April 2017 [*DOI :* 10.1177/1461444817700154], https://hal.inria.fr/hal-01516136.

[15] G. SCAVO, Z. BEN HOUIDI, S. TRAVERSO, R. TEIXEIRA, M. MELLIA.*WeBrowse: Leveraging User Clicks for Content Discovery in Communities of a Place*, in "Proceedings of the ACM on Human-Computer Interaction", November 2017, vol. 1, n° CSCW, p. 93:1-93:24 [*DOI :* 10.1145/3134728], https://hal.inria.fr/hal-01663712.

[16] R. TRONCY, G. RIZZO, A. JAMESON, O. CORCHO, J. PLU, E. PALUMBO, J. C. BALLESTEROS HERMIDA, A. SPIRESCU, K.-D. KUHN, C. BARBU, M. ROSSI, I. CELINO, R. AGARWAL, C. SCANU, M. VALLA, T. HAAKER.*3cixty: Building comprehensive knowledge bases for city exploration*, in "Journal of Web Semantics", 2017 [*DOI :* 10.1016/J.WEBSEM.2017.07.002], https://hal.inria.fr/hal-01556271.

[17] R. VENTURA, V. MALLET, V. ISSARNY, P.-G. RAVERDY, F. REBHI.*Evaluation and calibration of mobile phones for noise monitoring application*, in "Journal of the Acoustical Society of America", November 2017, vol. 142, n° 5, p. 3084 - 3093 [*DOI :* 10.1121/1.5009448], https://hal.inria.fr/hal-01676004.

Invited Conferences

[18] B. BILLET, V. ISSARNY, G. TEXIER.*Composing Continuous Services in a CoAP-based IoT*, in "AIMS - 6th IEEE International Conference on AI & Mobile Services", Honolulu, United States, June 2017, https://hal.inria.fr/hal-01519132.

International Conferences with Proceedings

[19] *Best Paper*
R. ANGARITA, N. GEORGANTAS, V. ISSARNY.*USNB: Enabling Universal Online Social Interactions*, in "IEEE International Conference on Collaboration and Internet Computing", San Jose, United States, October 2017, https://hal.inria.fr/hal-01591757.

[20] R. ANGARITA, N. GEORGANTAS, C. PARRA, J. HOLSTON, V. ISSARNY.*Leveraging the Service Bus Paradigm for Computer-mediated Social Communication Interoperability*, in "International Conference on Software Engineering (ICSE), Software Engineering in Society (SEIS) Track", Buenos Aires, Argentina, May 2017, https://hal.inria.fr/hal-01485213.

[21] G. BOULOUKAKIS, N. GEORGANTAS, A. KATTEPUR, V. ISSARNY.*Timeliness Evaluation of Intermittent Mobile Connectivity over Pub/Sub Systems*, in "ICPE 2017 - 8th ACM/SPEC International Conference on Performance Engineering", L'Aquila, Italy, ICPE 2017, April 2017, https://hal.inria.fr/hal-01415893.

[22] G. BOULOUKAKIS, I. MOSCHOLIOS, N. GEORGANTAS, V. ISSARNY.*Performance Modeling of the Middleware Overlay Infrastructure of Mobile Things*, in "IEEE International Conference on Communications", Paris, France, May 2017, https://hal.inria.fr/hal-01470328.

[23] V. CHRISTOPHIDES, V. EFTHYMIOU, O. HASSANZADEH, M. RODRIGUEZ-MURO.*Matching Web Tables with Knowledge Base Entities: From Entity Lookups to Entity Embeddings*, in "ISWC 2017 - 16th International Semantic Web Conference", Vienna, Austria, The Semantic Web – ISWC 2017, October 2017, vol. 10587, p. 260-277 [*DOI :* 10.1007/978-3-319-68288-4_16], https://hal.inria.fr/hal-01664023.

[24] V. CHRISTOPHIDES, G. POGHOSYAN, I. PEFKIANAKIS, P. LE GUYADEC.*Extracting usage patterns of home IoT devices*, in "ISCC 2017 - 22nd IEEE Symposium on Computers and Communications", Heraklion, Crete, Greece, IEEE, July 2017, p. 1-7 [*DOI :* 10.1109/ISCC.2017.8024707], https://hal.inria.fr/hal-01664015.

[25] V. CHRISTOPHIDES, K. STEFANIDIS.*Web-Scale Blocking, Iterative and Progressive Entity Resolution*, in "ICDE 2017 - 33rd IEEE International Conference on Data Engineering", San Diego, CA, United States, IEEE, April 2017, p. 1-4 [*DOI :* 10.1109/ICDE.2017.214], https://hal.inria.fr/hal-01664035.

[26] 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.

[27] 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.

[28] C. PARRA, C. ROHAUT, M. MAECKELBERGH, V. ISSARNY, J. HOLSTON.*Expanding the Design Space of ICT for Participatory Budgeting* , in "Communities and Technologies 2017", Troyes, France, ACM, June 2017, https://hal.inria.fr/hal-01519127.

[29] F. SAILHAN, V. ISSARNY, O. TAVARES NASCIMENTO.*Opportunistic Multiparty Calibration for Robust Participatory Sensing*, in "MASS 2017 - IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems", Orlando, United States, October 2017, https://hal.inria.fr/hal-01599377.

Scientific Books (or Scientific Book chapters)

[30] M. SERRANO, A. GYRARD, M. BONIFACE, P. GRACE, N. GEORGANTAS, R. AGARWAL, P. BARNAGHI, F. CARREZ, B. ALMEIDA, T. TEIXEIRA, P. COUSIN, F. LE GALL, M. BAUER, E. KOVACS, L. MUNOZ, L. SANCHEZ, J. SOLDATOS, N. KEFALAKIS, I. ABAITUA FERNÁNDEZ-ESCÁRZAGA, J. E. ECHEVARRIA CUENCA, R. STEINKE, M. HAUSWIRTH, J. KIM, J. YUN.*Cross-Domain Interoperability Using Federated Interoperable Semantic IoT/Cloud Testbeds and Applications: The FIESTA-IoT Approach*, in "Building the Future Internet through FIRE 2016 FIRE Book: A Research and Experiment based Approach", River Publishers, 2017, https://hal.inria.fr/hal-01556289.

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] V. RAPHAËL, V. MALLET, V. ISSARNY, P.-G. RAVERDY, F. REBHI.*Estimation of urban noise with the assimilation of observations crowdsensed by the mobile application Ambiciti*, 2017, In: Proceedings of INTER-NOISE 2017, 46th International Congress and Exposition on Noise Control Engineering. 2017, pp. 5, 741–5, 748, https://hal.inria.fr/hal-01676010.

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
B9.4.2. - Mathematics
B9.4.3. - Physics
B9.4.4. - Chemistry
B9.5.3. - Economy, Finance

1. Personnel

Research Scientists

Jean-David Benamou [Team leader, Inria, Senior Researcher, HDR]
Vincent Duval [Inria, Senior Researcher]
Thomas Gallouët [Inria, Researcher, from Sep 2017]
Irene Waldspurger [CNRS, Researcher]

Faculty Members

Guillaume Carlier [Univ Dauphine Paris, Professor, HDR]
François-Xavier Vialard [Délégation Univ Dauphine Paris, Associate Professor]
Yohann de Castro [Délégation Université Paris-Saclay Associate Professor]

External Collaborators

Yann Brenier [CNRS]
Jean Louet [Univ Paris-Sud, until Aug 2017]
Quentin Merigot [CNRS]
Luca Nenna [Univ Paris-Sud, until Sep 2017]
Gabriel Peyre [CNRS, HDR]

PhD Students

Paul Catala [Ecole Normale Supérieure Paris, from Oct 2017]
Lenaic Chizat [Univ Dauphine Paris, until Nov 2017]
Quentin Denoyelle [Univ Dauphine Paris]
Aude Genevay [Univ Dauphine Paris]
Lucas Martinet [INSA Rouen, from Nov 2017]
Marco Masoero [Univ Dauphine Paris]
Miao Yu [Univ Denis Diderot Paris]

Post-Doctoral Fellows

Roman Andreev [Univ Pierre et Marie Curie, until Jun 2017]
Jean-Baptiste Courbot [Université Paris Sciences et Lettres, from Nov 2017]

Guillaume Garrigos [CNRS, from Oct 2017]
Andrea Natale [Inria, from Nov 2017]

Visiting Scientist
Gwendoline de Bie [Ecole normale supérieure, from Jun 2017]

Administrative Assistant
Martine Verneuille [Inria]

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 [123]). This engineering problem consists in minimizing the transport cost between two given mass densities. In the 40's, Kantorovich [131] 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 [176], [175], 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 [141] 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 [82] 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 [111] [139] [57] [58] and the references therein. Geometric approaches based on Laguerre diagrams and discrete data [148] 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 cost has been established by McCann and Gangbo [122]. 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 [152] 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 [41]. 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 [151].

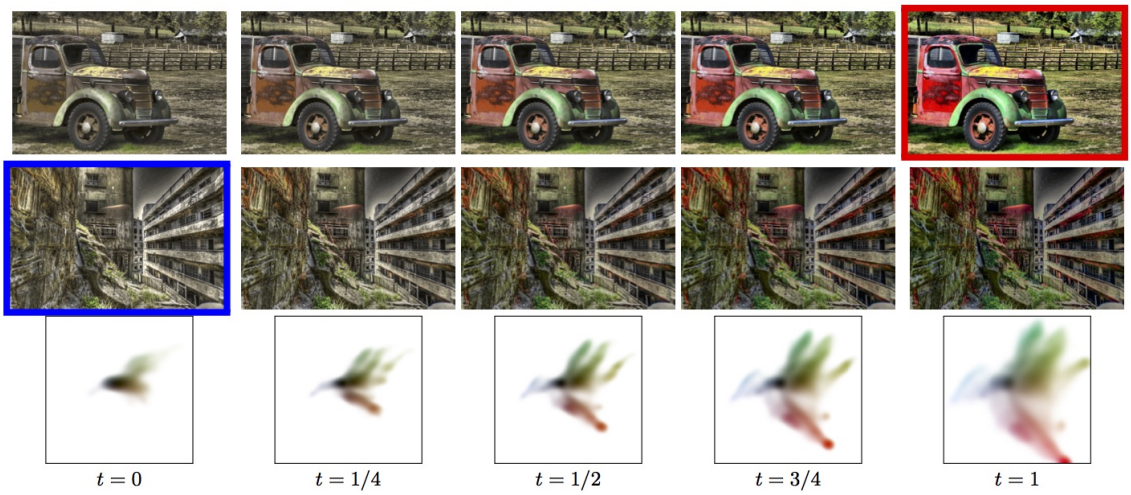


Figure 1. Example of color transfer between two images, computed using the method developed in [54], see also [169]. The image framed in red and blue are the input images. Top and middle row: adjusted image where the color of the transported histogram has been imposed. Bottom row: geodesic (displacement) 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 [128], [107], [106], [51], [138], mesh adaptation [137], the reconstruction of the early mass distribution of the Universe [119], [75] in Astrophysics, and the numerical

optimisation of reflectors following the Optimal Transportation interpretation of Oliker [83] and Wang [177]. 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], [121] ...Recently, there has been a spread of interest in applications of OT methods in imaging sciences [67], statistics [63] and machine learning [109]. 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 [133] 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 [69].

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 [129]. 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 [130], [100] and a model of congested crowd motion proposed by Maury, Santambrogio and Roudneff-Chupin [144]. 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 in involves taking the derivative in the Wasserstein sense of the relevant energy, which in turns 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 to define 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 [178]. 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], [127]. 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 problem through geodesic

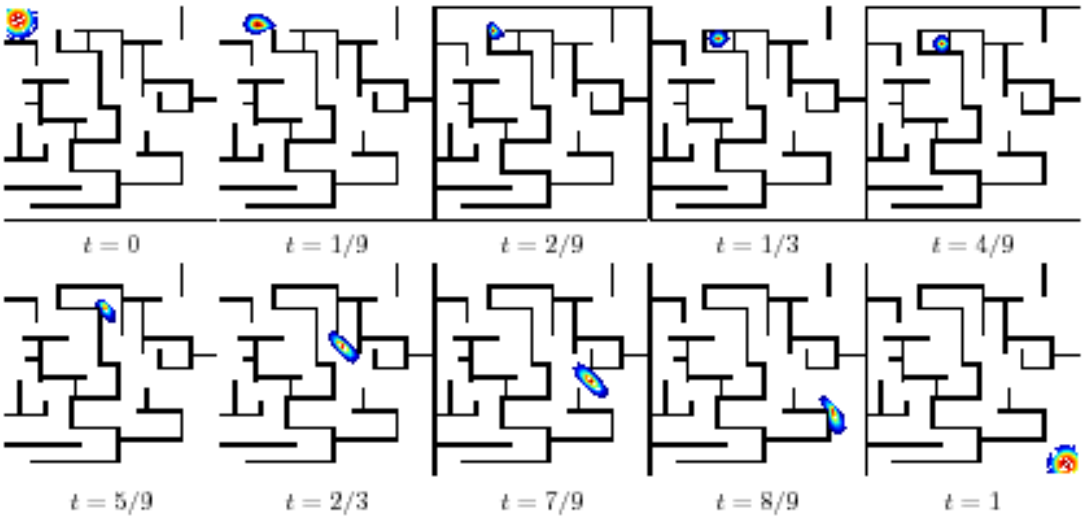


Figure 2. Examples of displacement interpolation (geodesic for optimal transport) according to a non-Euclidean Riemannian metric (the mass is constrained to move inside a maze) between to input Gaussian distributions. Note that the maze is dynamic: its topology change over time, the mass being "trapped" at time $t = 1/3$.

shooting method [145], 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 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) [84].

2.4. Sparsity in Imaging

2.4.1. Sparse ℓ^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 [143] and learned representation [179]).

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 ℓ^1 norm introduced independently in imaging by Donoho and co-workers under the name "Basis Pursuit" [103] and in statistics by Tibshirani [170] 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 [85], which make use of randomized forward operators and ℓ^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], [115], [112] 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 [163], [79]. 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 [170] and machine learning [43]. 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) [44]. 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 [136], [172] 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's 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], [160]). 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 [109]). 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 [161] reformulated the problem as a variational problem subject to a convexity constraint. For more general models, and using ideas from Optimal Transportation, Carlier [89] considered the more general c-convexity constraint and proved a general existence result. Using the formulation of [89] McCann, Figalli and Kim [116] gave conditions under which the principal agent problem can be written as an infinite dimensional convex variational problem. The important results of [116] 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], [149], [146].
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 [92] 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], [121] 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], [59], [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 [90] 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 usual 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) [129] 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 [129]), the porous medium equations, the granular media equation, just to give a few examples. It also finds application in image processing [78]. 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 [98])
- 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 [65]
- 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.

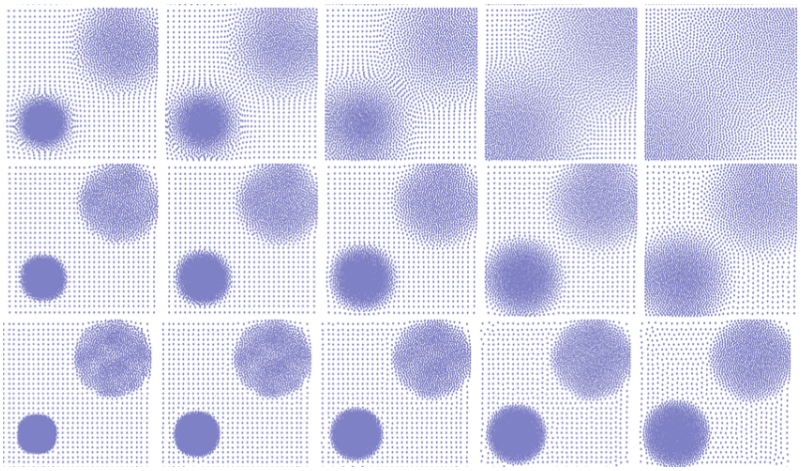


Figure 4. Example of non-linear diffusion equations solved with a JKO flow [55]. The horizontal axis shows the time evolution minimizing the functional $\int \frac{\rho^\alpha}{\alpha-1}$ on the density ρ (discretized here using point clouds, i.e. sum of Diracs' with equal mass). Each row shows a different 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 [59].
Goals: The case of general networks/anisotropies is still not well understood, general Γ-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 in Section 4.2.

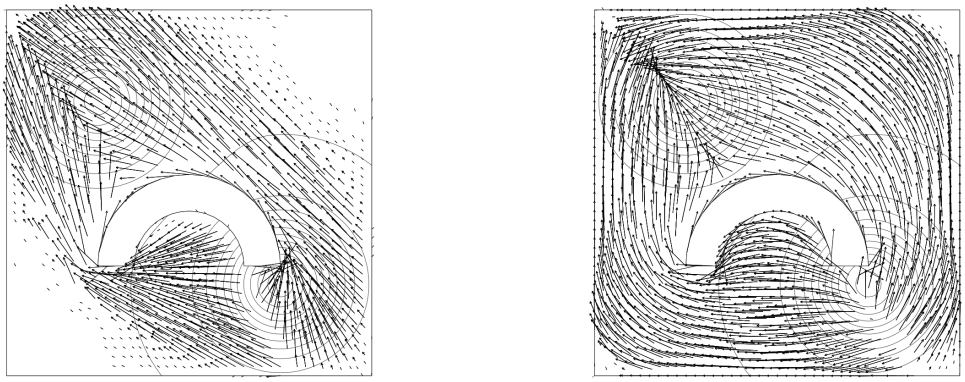


Figure 5. Monge and Wardrop flows of mass around an obstacle [52]. the source/target mass is represented by the level curves. Left : no congestion, Right : congestion.

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 [81] or tracking elastic organ movements [167]. Previous attempts [140], [157] 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 [133] with an extremely wide range of potential applications [124]. 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 [88] 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 [40].

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 [144] 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 [40], [77].
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) [173], [76], [171]. Our team has already studied flows and geodesics over non-Riemannian shape spaces, which allows for piecewise smooth deformations [102].

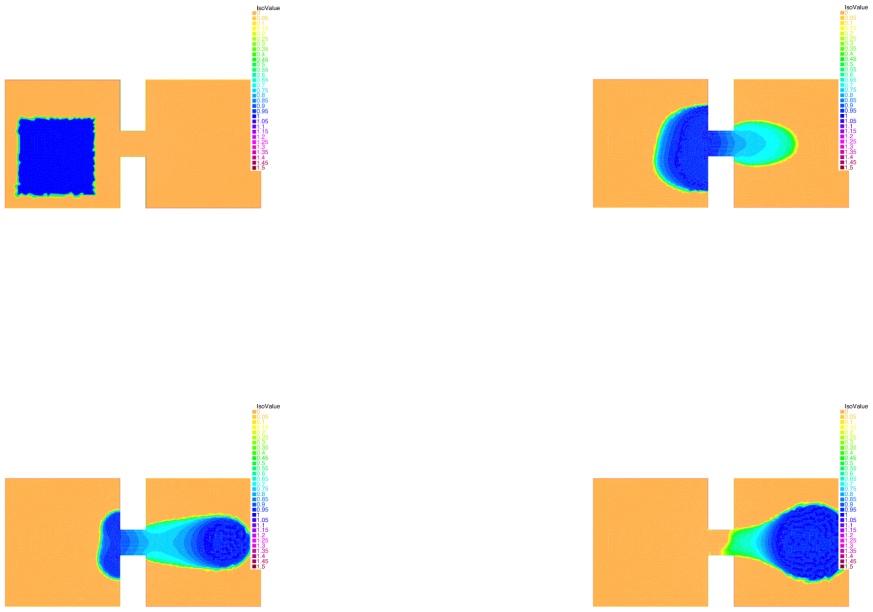


Figure 6. Example of crowd congestion with density dependent speed. The macroscopic density, at 4 different times, of people forced to exit from one room towards a meeting point in a second room.

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 [166], [167] 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 [135], 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 [174] 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 [117]. 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 [117], [118].
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 [172] 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 [66]) for discrete (finite dimensional) problems. We extended this result in a simple infinite dimensional setting, namely sparse regularization of Radon measures for deconvolution [112]. A deep understanding of those continuous inverse problems is crucial to analyze the behavior of their discrete counterparts, and in [113] 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

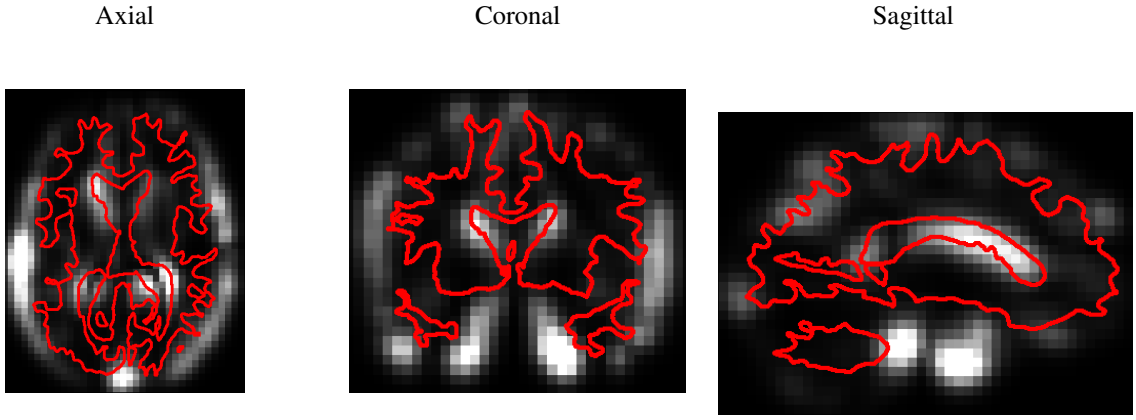


Figure 7. Learning Riemannian metrics in diffeomorphic image matching to capture the brain variability: a diagonal operator that encodes the Riemannian metric is learnt on a template brain out of a collection of brain images. The values of the diagonal operator are shown in greyscale. The red curves represent the boundary between white and grey matter. For more details, we refer the reader to [174], which was a first step towards designing effective and robust metric learning algorithms.

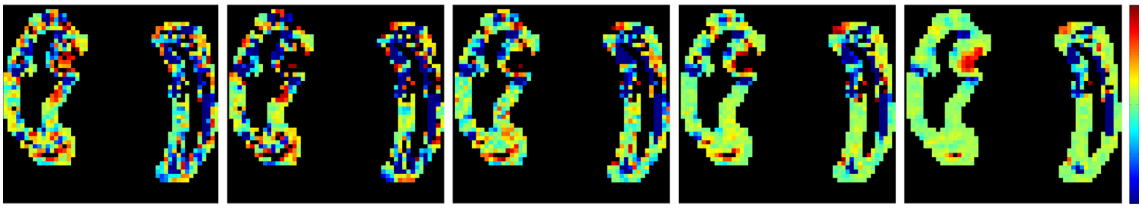


Figure 8. Statistics on initial momenta: In [117], we compared several intersubject transport methodologies to perform statistics on longitudinal evolutions. These longitudinal evolutions are represented by an initial velocity field on the shapes boundaries and these velocity fields are then compared 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 [91], [66] 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], [42], [99]. 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 [112] 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 [156], [114]. This requirement is also at the heart of recently proposed models of cortical processing [155]. 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 [147], [114]. G. Peyré has done several contributions on the definition of geometric wavelets transform and directional texture models, see for instance [156]. Dario Prandi has recently applied methods from sub-Riemannian geometry to image restoration [68].
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 [60]) 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 [126]. 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 [165]. Typical approaches [132] try to recovered 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 [112].
Goals: We aim at extending the theoretical performance analysis obtained for sparse measures [112] to the set of piecewise regular 2-D and 3-D functions. Some interesting previous work of C. Poon et al [158] (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 [39].

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 [81], [154] 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 [142], which are connected to so-called metamorphoses models in the LDDMM framework [61].
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 [81], [154] 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 [58] for the Monge-Ampère equation and [146] for general non-linear variational problems. Both offer convergence guarantees and are amenable to fast numerical resolution techniques such as Newton solvers. Since [58] 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 [120], a new different and more accurate approach has been proposed by Mirebeau, Benamou and Collino [56]. 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 [149] and J-M. Mirebeau [146]. 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 [129]) 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 [55] 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 Γ-convergence from discrete to continuous is surprisingly over-looked in imaging sciences. To the best of our knowledge, our recent work [112], [113] 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 Γ-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 [112], [113] 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 [112] to generic partly-smooth convex regularizers routinely used in imaging science and machine learning, a prototypal example being the nuclear norm (see [172] 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], [112]. 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 [112] 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 [134]. 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 [163]), 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], [150], congested optimal transport [80] and to sparse regularizations (see for instance [160] 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 [150]. We have also recently developed new proximal schemes that can cope with non-smooth composite objectives functions [160].
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.

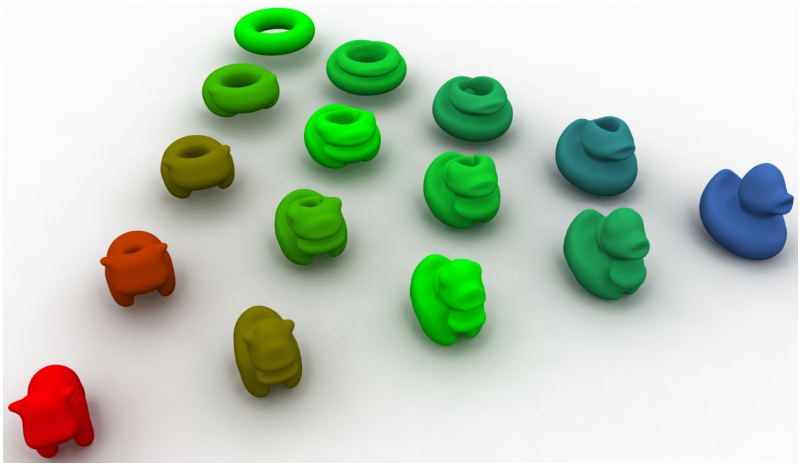


Figure 9. Example of barycenter between shapes computed using optimal transport barycenters of the uniform densities inside the 3 extremal shapes, computed as detailed in [169]. Note that the barycenters are not in general uniform distributions, and we display them as the surface defined by a suitable 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 [109]. 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 [92] 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 [169]. 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. Application Domains

4.1. Freeform Optics

Following the pioneering work of Caffarelli and Oliker [83], Wang [177] has shown that the inverse problem of freeforming a *convex* reflector which sends a prescribed source to a target intensity is a particular instance of Optimal Transportation. This is a promising approach to automatize the industrial design of optimised energy efficient reflectors (car/public lights for instance). We show in figure 10 the experiment setting and one of the first numerical simulations produced by the ADT Mokabajour.

A third specific topic is the use of optimal transport models in *non-imaging optics*. Light intensity here plays the role of the source/target prescribed mass and the transport map defines the physical shape of specular reflector or refracting lense achieving such a transformation. This models have been around since the works of Oliker and Wang in the 90's. Recent numerical progresses indicate that OT may have an important industrial impact in the design of optical elements and calls for further modelisation and analysis.

The method developed in [58] has been used by researchers of TU Eindhoven in collaboration with Philips Lightning Labs to compute reflectors [159] in a simplified setting (directional light source). Another approach, based on a geometric discretization of Optimal Transportation has been developed in [8], and is able to handle more realistic conditions (punctual light source).

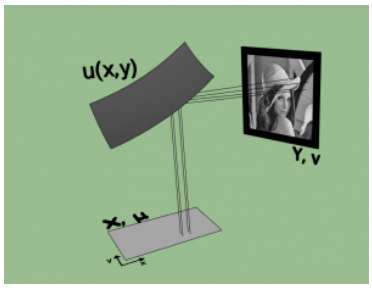
Solving the exact Optimal Transportation model for the Reflector inverse problem involves a generalized Monge-Ampère problem and is linked to the open problem of c-convexity compatible discretization we plan to work on. The corresponding software development is the topic of the ADT Mokabajour.

4.1.1. Software and industrial output.

See section 4.3 below for softwares. These methods will clearly become mainstream in reflector design but also in lense design [162]. The industrial problems are mainly on efficiency (light pollution) and security (car head lights) based on free tailoring of the illumination. The figure below is an extreme test case where we exactly reproduce an image. They may represent one of the first incursion on PDE discretisation based methods into the field of non-imaging optics.

4.2. Metric learning for natural language processing

The analysis of large scale datasets to perform un-supervised (clustering) and supervised (classification, regression) learning requires the design of advanced models to capture the geometry of the input data. We believe that optimal transport is a key tool to address this problem because (i) many of these datasets are composed of histograms (social network activity, image signatures, etc.) (ii) optimal transport makes use of a ground metric that enhances the performances of classical learning algorithms, as illustrated for instance in [109].



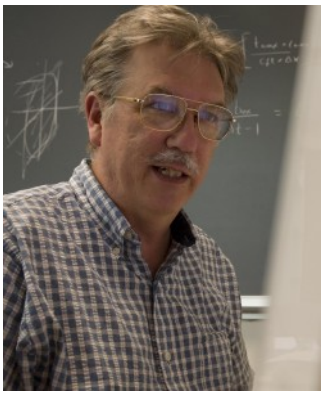



Figure 10. A constant source to a prescribed image (center). The reflector is computed (but not shown) and a resimulation using ray tracing shows the image reflected by the computed reflector.

Some of the theoretical and numerical tools developed by our team, most notably Wasserstein barycenters [41], [67], are now becoming mainstream in machine learning [63], [109]. In its simplest (convex) form where one seeks to only maximize pairwise wasserstein distances, metric learning corresponds to the congestion problem studied by G. Carlier and collaborators [97], [70], and we will elaborate on this connection to perform both theoretical analysis and develop numerical schemes (see for instance our previous work [59]).

We aim at developing novel variational estimators extending classification regression energies (SVM, logistic regression [125]) and kernel methods (see [168]). One of the key bottleneck is to design numerical schemes to learn an optimal metric for these purpose, extending the method of Marco Cuturi [108] to large scale and more general estimators. Our main targeted applications is natural language processing. The analysis and processing of large corpus of texts is becoming a key problems at the interface between linguistic and machine learning [45]. Extending classical machine learning methods to this field requires to design suitable metrics over both words and bag-of-words (i.e. histograms). Optimal transport is thus a natural candidate to bring innovative solutions to these problems. In a collaboration with Marco Cuturi (Kyoto University), we aim at unleashing the power of transportation distances by performing ground distance learning on large database of text. This requires to lift previous works on distance on words (see in particular [153]) to distances on bags-of-words using transport and metric learning.

4.3. Physics and Astrophysic

The Brenier interpretation of the generalized solutions of Euler equations in the sense of Arnold is an instance of multi-marginal optimal transportation, a recent and expanding research field which also appears in DFT (see chemistry below). Recent numerical developments in OT provide new means of exploring these class of solutions.

In the years 2000 and after the pioneering works of Otto, the theory of *many-particle systems* has become "geometrized" thanks to the observed intimate relation between the geometric theory of geodesic convexity in the Wasserstein distance and the proof of entropy dissipation inequalities that determine the trend to equilibrium. The OT approach to the study of equilibration is still an extremely active field, in particular the various recently established connections to sharp functional inequalities and isoperimetric problems.

4.4. Chemistry

Figure 11. Examples of two histogram (bag-of-words) extracted from the congress speech of US president. In this application, the goal is to infer a meaningful metric on the words of the english language and lift this metric to histogram using OT technics.

The treatment of *chemical reactions* in the framework of OT is a rather recent development. The classical theory must be extended to deal with the transfer of mass between different particle species by means of chemical reactions.

A promising and significant recent advance is the introduction and analysis of a novel metric that combines the pure transport elements of the Wasserstein distance with the annihilation and creation of mass, which is a first approximation of chemical reactions. The logical next challenge is the extension of OT concepts to vectorial quantities, which allows to rewrite cross-diffusion systems for the concentration of several chemical species as gradient flows in the associated metric. An example of application is the modeling of a *chemical vapor deposition process*, used for the manufacturing of thin-film solar cells for instance. This leads to a degenerate cross-diffusion equations, whose analysis — without the use of OT theory — is delicate. Finding an appropriate OT framework to give the formal gradient flow structure a rigorous meaning would be a significant advance for the applicability of the theory, also in other contexts, like for biological multi-species diffusion.

A very different application of OT in chemistry is a novel approach to the understanding of *density functional theory* (DFT) by using optimal transport with "Coulomb costs", which is highly non convex and singular. Albeit this theory shares some properties with the usual optimal transportation problems, it does not induce a metric between probability measures. It also uses the multi-marginal extension of OT, which is an active field on its own right.

4.5. Biology

OT methods have been introduced in biology via gradient flows in the Wasserstein metric. Writing certain *chemotaxis* systems in variational form allowed to prove sharp estimates on the long time asymptotics of the bacterial aggregation. This application had a surprising payback on the theory: it lead to a better understanding and novel proofs of important functional inequalities, like the logarithmic Hardy-Littlewood-Sobolev inequality. Further applications followed, like transport models for species that avoid over-crowding, or cross-diffusion equations for the description of *biologic segregation*. The inclusion of dissipative cross-diffusion systems into the framework of gradient flows in OT-like metrics appears to be one of the main challenges for the future development of the theory. This extension is not only relevant for biological applications, but is clearly of interest to participants with primary interest in physics or chemistry as well.

Further applications include the connection of OT with game theory, following the idea that many selection processes are based on competition. The ansatz is quite universal and has been used in other areas of the *life sciences* as well, like for the modeling of personal income in economics.

Another application of our methods is the use of inverse problems in measure spaces for microscopy imaging. The *Single Molecule Microscopy Imaging* techniques such as PALM [62] or STORM [164] have yielded a breakthrough in fluorescence microscopy, improving the typical resolution of conventional microscopes (250 nm) by an order of magnitude (20 nm). These techniques convert the problems of full image reconstruction into a family of sparse spike reconstructions. Our variational methods, which take advantage of the sparsity of the signals to recover, are much more powerful than the usual methods used by biologists for sparse recovery. They promise to release the full potential of PALM and STORM in terms of resolution and speed of acquisition.

4.6. Medical Imaging

Applications of variational methods are widespread in medical imaging and especially for diffeomorphic image matching. The formulation of large deformation by diffeomorphisms consists in finding geodesics on a group of diffeomorphisms. This can be seen as a non-convex and smoothed version of optimal transport where a correspondence is sought between objects that can be more general than densities. Whereas the diffeomorphic approach is well established, similarity measures between objects of interest are needed in order to drive the optimization. While being crucial for the final registration results, these similarity measures are often non geometric due to a need of fast computability and gradient computation. However, our team pioneered the use of entropic smoothing for optimal transport which gives fast and differentiable similarity measures that take into account the geometry. Therefore, we expect an important impact on this topic, work still in progress. This example of application belongs to the larger class of inverse problems where a geometric similarity measure such as optimal transport might enhance notably the results. Concerning this particular application, potential interactions with the Inria team ARAMIS and also the team ASCLEPIOS can leverage new proposed similarity measure towards a more applicative impact.

4.7. Economics

Recent years have seen intense cross-fertilization between OT and various problems arising in economics. The principal-agent problem with adverse selection is particularly important in modern microeconomics, mathematically it consists in minimizing a certain integral cost functional among the set of c-concave functions, this problem is convex under some conditions related to the MTW regularity theory for OT as shown in the important paper [116]. Other examples of fruitful interactions between mathematical economics concern multi-marginal OT and multi-populations matching [93], or games with a continuum of agents and Cournot-Nash equilibria [64]. The team has as strong expertise, both numerical and theoretical in the field of variational problems subject to a convexity constraint and their applications to the principal-agent problem. Our expertise in numerical OT and entropic regularization will also enable us to develop efficient solvers for realistic matching and hedonic pricing models.

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/

5.3. Platforms

5.3.1. MABV2

A 2D Julia implementation of the algorithm described in [25]. https://gforge.inria.fr/scm/browser.php?group_id=9995

6. New Results

6.1. Optimal transport for diffeomorphic registration

J. Feydy and B. Charlier and G. Peyré and F-X. Vialard

[18]

This paper introduces the use of unbalanced optimal transport methods as a similarity measure for diffeomorphic matching of imaging data. The similarity measure is a key object in diffeomorphic registration methods that, together with the regularization on the deformation, defines the optimal deformation. Most often, these similarity measures are local or non local but simple enough to be computationally fast. We build on recent theoretical and numerical advances in optimal transport to propose fast and global similarity measures that can be used on surfaces or volumetric imaging data. This new similarity measure is computed using a fast generalized Sinkhorn algorithm. We apply this new metric in the LDDMM framework on synthetic and real data, fibres bundles and surfaces and show that better matching results are obtained.

6.2. Quantum Optimal Transport for Tensor Field Processing

G. Peyré and L. Chizat and F-X. Vialard and J. Solomon

[18]

This article introduces a new notion of optimal transport (OT) between tensor fields, which are measures whose values are positive semidefinite (PSD) matrices. This "quantum" formulation of OT (Q-OT) corresponds to a relaxed version of the classical Kantorovich transport problem, where the fidelity between the input PSD-valued measures is captured using the geometry of the Von-Neumann quantum entropy. We propose a quantum-entropic regularization of the resulting convex optimization problem, which can be solved efficiently using an iterative scaling algorithm. This method is a generalization of the celebrated Sinkhorn algorithm to the quantum setting of PSD matrices. We extend this formulation and the quantum Sinkhorn algorithm to compute barycenters within a collection of input tensor fields. We illustrate the usefulness of the proposed approach on applications to procedural noise generation, anisotropic meshing, diffusion tensor imaging and spectral texture synthesis.

6.3. The Camassa-Holm equation as an incompressible Euler equation: a geometric point of view

T. Gallouët and F-X. Vialard

[35]

The group of diffeomorphisms of a compact manifold endowed with the L2 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 L2-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 L2 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 S1, solutions to the standard CH thus give particular solutions of the incompressible Euler equation on a group of homeomorphisms of R2 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.

6.4. Minimal convex extensions and finite difference discretization of the quadratic Monge-Kantorovich problem

J-D. Benamou and V. Duval

[25]

We designed an adaptation of the MA-LBR scheme [4] to the Monge-Ampère 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.

6.5. Phase retrieval for wavelet transforms

I. Waldspurger

[15]

This article describes an algorithm for solving a particular phase retrieval problem, with important applications in audio processing: the reconstruction of a function from the modulus of its wavelet transform. Previous algorithms for this problem were either unreliable in certain regimes, or too slow to be applied to large-dimensional audio signals. Ours relies on a new reformulation of the phase retrieval problem, that involves the holomorphic extension of the wavelet transform. Numerical results, on audio and non-audio signals, show it allows precise reconstruction, and is stable to noise. Its complexity is linear in the size of the unknown signal, up to logarithmic factors. It can thus be applied to large signals.

6.6. Phase retrieval with random Gaussian sensing vectors by alternating projections

I. Waldspurger

[16]

We consider the phase retrieval problem that consists in reconstructing a vector from its phaseless scalar products with sensing vectors independently sampled from complex normal distributions. In the previous two years, several new non-convex algorithms have been introduced to solve it, and have been proven to succeed with high probability. In this work, we show that the same success guarantees hold true for the oldest and most well-known phase retrieval algorithm, namely alternating projections (Gerchberg-Saxton), provided that it is carefully initialized. We conjecture that this result is still true when no special initialization procedure is used, and present numerical experiments that support this conjecture.

6.7. Exponential decay of scattering coefficients

I. Waldspurger

[19]

The scattering transform is a deep representation, defined as a cascade of wavelet transforms followed by the application of a complex modulus. In her PhD, the author showed that, under some conditions on the wavelets, the norm of the scattering coefficients at a given layer only depends on the values of the signal outside a frequency band whose size is exponential in the depth of the layer. This article succintly describes this result, and generalizes it by removing one of the assumptions on the wavelets (namely the weak analyticity condition).

6.8. Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm

J-D. Benamou and G. Carlier and L. Nenna

[24]

Starting from Brenier's relaxed formulation of the incompressible Euler equation in terms of geodesics in the group of measurepreserving diffeomorphisms, we propose a numerical method based on Sinkhorn?s algorithm for the entropic regularization of optimal transport. We also make a detailed comparison of this entropic regularization with the so-called Bredinger entropic interpolation problem. Numerical results in dimension one and two illustrate the feasibility of the method.

6.9. A Characterization of the Non-Degenerate Source Condition in Super-Resolution

V. Duval

[34]

This article deals with the Basis Pursuit (or LASSO) for measures for for the super-resolution problem, *i.e.* retrieving the fine details of a signal or an image. If the signal is made of M non-negative Dirac masses, under some assumptions on the measurement process, it is possible to exactly recover the signal from $2M$ observations, regardless of the minimum distance between the spikes. We study the stability to noise of such a reconstruction, and we propose a characterization of the *Non-Degenerate Source Condition* which is an almost necessary and sufficient for the stability of the support (the number and locations of the reconstructed spikes). The case of Laplace and Gaussian measurements are studied in detail.

6.10. A Low-Rank Approach to Off-The-Grid Sparse Deconvolution

P. Catala, V. Duval and G. Peyré

[28].

We propose a new solver for the sparse spikes deconvolution problem over the space of Radon measures. A common approach to off-the-grid deconvolution considers semidefinite (SDP) relaxations of the total variation (the total mass of the absolute value of the measure) minimization problem. The direct resolution of this SDP is however intractable for large scale settings, since the problem size grows as f_c^{2d} where f_c is the cutoff frequency of the filter and d the ambient dimension. Our first contribution introduces a penalized formulation of this semidefinite lifting, which has low-rank solutions. Our second contribution is a conditional gradient optimization scheme with non-convex updates. This algorithm leverages both the low-rank and the convolutive structure of the problem, resulting in an $O(f_c^d \log(f_c))$ complexity per iteration. Numerical simulations are promising and show that the algorithm converges in exactly r steps, r being the number of Diracs composing the solution.

6.11. Approximate Optimal Designs for Multivariate Polynomial Regression

Y. De Castro

[110].

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.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. Optimal Transport applied to altimetric data dynamic interpolation

(S. Legrand V. Duval L. Chizat J-D. Benamou).

This collaboration between CLS and and funded by CNES intends to test on Column of Tropospheric Humidity data Optimal transportation interpolation techniques for balanced and unbalanced data.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

J-D. Benamou is the coordinator of the ANR ISOTACE (Interacting Systems and Optimal Transportation, Applications to Computational Economics) ANR-12-MONU-0013 (2012-2016). The consortium explores new numerical methods in Optimal Transportation AND Mean Field Game theory with applications in Economics and congested crowd motion. Check https://project.inria.fr/isotace/.
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. 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. Gallouët is member of the ANR GEOPOR Scientific topic: geometrical approach, based on Wasserstein gradient flow, for multiphase flows in porous media. Theory and Numerics.

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

- J-D. Benamou is a member of the ITN ROMSOC (Nov. 2017-Nov.2021).
- Andrea Natale has a PRESTIGE Post-Doc Fellowship.

8.3. International Research Visitors

8.3.1. Visits of International Scientists

The following people visited MOKAPLAN during 2016.

- Alfred Galichon (Courant), Teresa Radice (Naples), Gaoyue Guo (Oxford) visited G. Carlier at inria in 2017
- Simone di Marino (Pisa)

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Organizing Committees

G. Carlier has corganized CMO-BIRS 17w5093.

J-D. Benamou has co-organized Brenier 60.
J-D. Benamou has co-organized MFO workshop 1705 (February).

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

Guillaume Carlier is in the board of Journal de l'école Polytechnique, Applied Mathematics and optimization (since 2016) and Mathematics and financial economics, with Filippo Santambrogio and Thierry Champion he co-edited a special issues of RICAM Series devoted to optimal transport. G. Peyré is editor for SIAM Journal of Imaging Sciences and Springer Journal of Mathematical Imaging and Vision. He co-edited a special issues of RICAM Series devoted to inverse problems.

9.1.2.2. Reviewer - Reviewing Activities

The members of the team are frequently reviewing papers in SIIMS (SIAM Journal on Imaging Sciences), JMAA (Journal of Mathematical Analysis and Applications), IPol (Image Processing Online), JVCI (Journal of Visual Communication and Image Representation), COCV, M2AN ... Discrete and computational geometry, Journal of the London Math Society, JOTA, JCP, "Information and Inference: A Journal of the IMA", JMIV, Optimization Letters, PAMI, SIAM optimization and control, IPMI, MICCAI (leading conferences in medical imaging), IEEE Information Theory, ICLR, NIPS, ICML (important machine learning conferences).

9.1.3. Invited Talks

Y. De Castrogave talks at Séminaire d'Informatique de l'Ecole Normale Supérieure, Lyon, and Séminaire de Probabilités de l'Ecole Normale Supérieure, Lyon, Groupe de Travail "Gaussian Process" Université Jean Monnet, St-Etienne, Séminaire de Probabilités de Lille, Séminaire de Probabilités et Statistique de Liège, Séminaire de Probabilités et Statistique de Versailles, LMV, Séminaire de Statistique de Toulouse, IMT, Groupe de Travail "Sequential Structured Statistical Learning", IHES, Cambridge Statistics Seminar, Cambridge, UK.

G. Carlier gave talks in Banff, Victoria, Naples, Le Teich, Toulouse (conference in honor of P. Cattiaux and C. Léonard), Paris (conference in honor of Y. Brenier, functional analysis seminar at IMJ and Game theory seminar at IHP), PGMO Days Paris Saclay.

J-D. Benamou was invited speaker at FOCM (Barcelona, July), CEMRACS (CIRM, July), Conf. in Honor of P. Joly (Gif, September) SPO (IHP, October) .

I. Waldspurger gave talks at Journées EDP (Roscoff, June), and at workshops on phase retrieval (Minneapolis, August) and on generative models, parameter learning and sparsity (Cambridge, October). She also gave a mini-course at Journées de géométrie algorithmique (Aussois, December).

V. Duval has given talks at the SPOC seminar (Université de Dijon, January) and Statistics seminar (Télécom ParisTech, September).

F-X. Vialard gave talks at MIT, csail, in the medical imaging group, in Chapell Hill University (April), workshop on applied geometric mechanics (Darryl Holm's anniversary) in Madrid (July), Classic and Stochastic Approaches to Mathematical Fluid Dynamics at Imperial College (September), workshop in Cambridge about growth and form (November), and about mathematics for imaging (December), Geometric Functional Data Analysis Workshop in Tallahassee (September), GMO (Paris-Saclay).

T. Gallouët gave a talk at the ANEDP seminar of Paris Sud University (December 2017).

9.1.4. Research Administration

J-D. Benamou is an elected member of the "Conseil Académique" of the PSL COMUE.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master : V. Duval, Student project supervision, October 2017 to March 2018, M2, INSA Rouen-Normandie

Licence : I. Waldspurger, Analyse 1, Université Paris-Dauphine, $72h$.

Master : Y. De Castro, Master 1 course on Statistics at Orsay.

9.2.2. Supervision

Internship :

PhD in progress : Miao Yu, *Optimal Transport distances and Geophyscial imaging* J-D. Benamou (co-direction J.-P. Vilotte, IPGP).

PhD in progress : Paul Catala, *Low-rank approaches for off-the-grid superresolution*, October 2016, G. Peyré and V. Duval.

PhD in progress : Lucas Martinet *Multi-Marginal OT* Oct. 2017 , J-D. Benamou.

PhD in progress : Aude Genevay, *Optimal Transport for Machine Learning* , october 2015, G. Peyré.

PhD in progress : Quentin Denoyelle, *Off-the-grid super-resolution: theory, algorithms and applications in fluorescence imaging*, October 2014, G. Peyré and V. Duval.

Postdoc completed:

PhD completed: Lenaic Chizat

Postdoc in progress : A. Natale (Inria/Prestige)

Postdoc in progress : J.B. Courbot (PSL IRIS, in collaboration with LMD, ENS).

PhD in progress : Ernesto Araya, *Measures on graphs* , Y. De Castro

9.2.3. Juries

J-D. Benamou F-X. Vialard were in the PhD comittee of Lenaic Chizat (Paris-Dauphine, November).
J-D. Benamou was in the HDR comittee of F-X. Vialard (Paris-Dauphine, December).
G. Carlier and F-X. Vialard were in the PhD comittee of Xianglong Duan (École Polytechnique, September).
G. Carlier was in the HDR commitee of Francisco Silva (Limoges) and Daniela Tonon (Dauphine), in the PhD comittee of Fatima Al Reda (Orsay), Xianglong Duan (Poolytechnique), Van-Thanh Nguyen (Limoges) and Luigia Ripani (Lyon). G. Carlier was president of the commitee for the PGMO PhD. award.

10. 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° 5, p. 1583–1592.

Publications of the year

Articles in International Peer-Reviewed Journal

[10] G. CARLIER, V. DUVAL, G. PEYRÉ, B. SCHMITZER.*Convergence of Entropic Schemes for Optimal Transport and Gradient Flows*, in "SIAM Journal on Mathematical Analysis", April 2017, vol. 49, n° 2, https://arxiv.org/abs/1512.02783 [*DOI :* 10.1137/15M1050264], https://hal.archives-ouvertes.fr/hal-01246086.

[11] C. DOSSAL, V. DUVAL, C. POON.*Sampling the Fourier transform along radial lines*, in "SIAM Journal on Numerical Analysis", November 2017, vol. 55, n° 6, https://arxiv.org/abs/1612.06752 [*DOI :* 10.1137/16M1108807], https://hal.inria.fr/hal-01421265.

[12] V. DUVAL, G. PEYRÉ.*Sparse Regularization on Thin Grids I: the LASSO*, in "Inverse Problems", March 2017, vol. 33, n° 5 [*DOI :* 10.1088/1361-6420/aa5e12], https://hal.archives-ouvertes.fr/hal-01135200.

[13] V. DUVAL, G. PEYRÉ.*Sparse Spikes Super-resolution on Thin Grids II: the Continuous Basis Pursuit*, in "Inverse Problems", August 2017, vol. 33, n° 9 [*DOI :* 10.1088/1361-6420/aa7fce], https://hal.archives-ouvertes.fr/hal-01389956.

[14] J. LIANG, J. M. FADILI, G. PEYRÉ.*Local Convergence Properties of Douglas–Rachford and Alternating Direction Method of Multipliers*, in "Journal of Optimization Theory and Applications", March 2017, vol. 172, n° 3, p. 874-913 [*DOI :* 10.1007/s10957-017-1061-z], https://hal.archives-ouvertes.fr/hal-01658848.

[15] I. WALDSPURGER.*Phase retrieval for wavelet transforms*, in "IEEE Transactions on Information Theory", 2017, https://hal.archives-ouvertes.fr/hal-01645088.

[16] I. WALDSPURGER.*Phase retrieval with random Gaussian sensing vectors by alternating projections*, in "IEEE Transactions on Information Theory", 2017, https://hal.archives-ouvertes.fr/hal-01645081.

Invited Conferences

[17] Y. DE CASTRO, Y. GOUDE, G. HÉBRAIL, J. MEI.*Recovering Multiple Nonnegative Time Series From a Few Temporal Aggregates*, in "ICML 2017 - 34th International Conference on Machine Learning", Sydney, Australia, August 2017, p. 1-9, https://arxiv.org/abs/1610.01492 , https://hal.inria.fr/hal-01686437.

International Conferences with Proceedings

[18] J. FEYDY, B. CHARLIER, F.-X. VIALARD, G. PEYRÉ.*Optimal Transport for Diffeomorphic Registration*, in "MICCAI 2017", Quebec, Canada, Proc. MICCAI 2017, September 2017, https://arxiv.org/abs/1706.05218 , https://hal.archives-ouvertes.fr/hal-01540455.

Conferences without Proceedings

[19] I. WALDSPURGER.*Exponential decay of scattering coefficients*, in "SampTA 2017 - Sampling Theory and Applications", Tallinn, Estonia, July 2017, https://hal.archives-ouvertes.fr/hal-01645078.

Scientific Books (or Scientific Book chapters)

[20] M. BERGOUNIOUX, J.-B. CAILLAU, T. HABERKORN, G. PEYRÉ, C. SCHNÖRR (editors). *Variational methods in imaging and geometric control*, Radon Series on Comput. and Applied Math., de Gruyter, January 2017, n^{o} 18, https://hal.archives-ouvertes.fr/hal-01315508.

Other Publications

[21] M. AGUEH, G. CARLIER, N. IGBIDA.*On the minimizing movement with the 1-Wasserstein distance*, February 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01467979.

[22] R. ANDREEV.*Preconditioning the augmented Lagrangian method for instationary mean field games with diffusion*, May 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01301282.

[23] J.-M. AZAÏS, Y. DE CASTRO, Y. GOUDE, G. HÉBRAIL, J. MEI.*Nonnegative matrix factorization with side information for time series recovery and prediction*, January 2018, https://arxiv.org/abs/1709.06320 - working paper or preprint, https://hal.inria.fr/hal-01686429.

[24] J.-D. BENAMOU, G. CARLIER, L. NENNA.*Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm*, October 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01621311.

[25] J.-D. BENAMOU, V. DUVAL.*Minimal convex extensions and finite difference discretization of the quadratic Monge-Kantorovich problem*, October 2017, https://arxiv.org/abs/1710.05594 - working paper or preprint, https://hal.inria.fr/hal-01616842.

[26] 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.

[27] G. CARLIER, C. POON.*On the total variation Wasserstein gradient flow and the TV-JKO scheme*, March 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01492343.

[28] P. CATALA, V. DUVAL, G. PEYRÉ.*A Low-Rank Approach to Off-The-Grid Sparse Deconvolution*, December 2017, https://arxiv.org/abs/1712.08800 - working paper or preprint, https://hal.inria.fr/hal-01672896.

[29] L. CHIZAT, G. PEYRÉ, B. SCHMITZER, F.-X. VIALARD.*Scaling Algorithms for Unbalanced Transport Problems*, January 2017, https://arxiv.org/abs/1607.05816 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01434914.

[30] Y. DE CASTRO, J.-M. AZAÏS, S. MOURAREAU.*Testing Gaussian Process with Applications to Super-Resolution*, January 2018, https://arxiv.org/abs/1706.00679 - 34 pages, 5 figures; this new version essentially put more emphasis on the comparison between grid methods and "grid-less" methods, https://hal.inria.fr/hal-01686434.

[31] Y. DE CASTRO, F. GAMBOA, D. HENRION, R. HESS, J.-B. LASSERRE.*Approximate Optimal Designs for Multivariate Polynomial Regression*, October 2017, To appear at Annals of Satistics, https://hal.laas.fr/hal-01483490.

[32] L. DE PASCALE, J. LOUET.*A study of the dual problem of the one-dimensional L-infinity optimal transport problem with applications*, August 2017, https://arxiv.org/abs/1704.02730 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01504249.

[33] S. DI MARINO, J. LOUET.*The entropic regularization of the Monge problem on the real line*, March 2017, https://arxiv.org/abs/1703.10457 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01498732.

[34] V. DUVAL.*A characterization of the Non-Degenerate Source Condition in Super-Resolution*, December 2017, https://arxiv.org/abs/1712.06373 - working paper or preprint, https://hal.inria.fr/hal-01665805.

[35] T. GALLOUËT, F.-X. VIALARD.*The Camassa-Holm equation as an incompressible Euler equation: a geometric point of view*, December 2017, https://arxiv.org/abs/1609.04006 - To appear in Journal of Differential Equations, 26 pages, https://hal.archives-ouvertes.fr/hal-01363647.

[36] K. LOUNICI, K. MEZIANI, G. PEYRÉ.*Adaptive sup-norm estimation of the Wigner function in noisy quantum homodyne tomography*, March 2017, https://arxiv.org/abs/1506.06941 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01491197.

[37] M. SEIDL, S. D. MARINO, A. GEROLIN, L. NENNA, K. J. H. GIESBERTZ, P. GORI-GIORGI.*The strictly-correlated electron functional for spherically symmetric systems revisited*, February 2017, working paper or preprint, https://hal.inria.fr/hal-01469822.

[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.

References in notes

[39] 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.

[40] 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.

[41] M. AGUEH, G. CARLIER.*Barycenters in the Wasserstein space*, in "SIAM J. Math. Anal.", 2011, vol. 43, n^{o} 2, p. 904–924.

[42] 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.

[43] 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.

[44] 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.

[45] M. BATES.*Models of natural language understanding*, in "Proceedings of the National Academy of Sciences", 1995, vol. 92, n^{o} 22, p. 9977-9982.

[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, Q. MÉRIGOT, É. OUDET.*Discretization of functionals involving the Monge-Ampère operator*, HAL, July 2014, https://hal.archives-ouvertes.fr/hal-01056452.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] M. BENNING, M. BURGER.*Ground states and singular vectors of convex variational regularization methods*, in "Meth. Appl. Analysis", 2013, vol. 20, p. 295–334.

[61] B. BERKELS, A. EFFLAND, M. RUMPF.*Time discrete geodesic paths in the space of images*, in "Arxiv preprint", 2014.

[62] E. BETZIG, G. H. PATTERSON, R. SOUGRAT, O. W. LINDWASSER, S. OLENYCH, J. S. BONIFACINO, M. W. DAVIDSON, J. LIPPINCOTT-SCHWARTZ, H. F. HESS.*Imaging Intracellular Fluorescent Proteins at Nanometer Resolution*, in "Science", 2006, vol. 313, n^{o} 5793, p. 1642–1645 [*DOI :* 10.1126/SCIENCE.1127344], http://science.sciencemag.org/content/313/5793/1642.

[63] J. BIGOT, T. KLEIN.*Consistent estimation of a population barycenter in the Wasserstein space*, in "Preprint arXiv:1212.2562", 2012.

[64] A. BLANCHET, G. CARLIER.*Optimal Transport and Cournot-Nash Equilibria*, in "Mathematics of Operations Resarch", 2015, to appear.

[65] 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.

[66] J. BLEYER, G. CARLIER, V. DUVAL, J.-M. MIREBEAU, G. PEYRÉ.*A Γ-Convergence Result for the Upper Bound Limit Analysis of Plates*, in "arXiv preprint arXiv:1410.0326", 2014.

[67] 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/.

[68] 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/.

[69] 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.

[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] M. BURGER, M. DIFRANCESCO, P. MARKOWICH, M. T. WOLFRAM.*Mean field games with nonlinear mobilities in pedestrian dynamics*, in "DCDS B", 2014, vol. 19.

[78] M. BURGER, M. FRANEK, C. SCHONLIEB.*Regularized regression and density estimation based on optimal transport*, in "Appl. Math. Res. Expr.", 2012, vol. 2, p. 209–253.

[79] M. BURGER, S. OSHER.*A guide to the TV zoo*, in "Level-Set and PDE-based Reconstruction Methods, Springer", 2013.

[80] 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.

[81] 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.

[82] 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.

[83] 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.

[84] C. CANCERITOGLU.*Computational Analysis of LDDMM for Brain Mapping*, in "Frontiers in Neuroscience", 2013, vol. 7.

[85] E. CANDES, M. WAKIN.*An Introduction to Compressive Sensing*, in "IEEE Signal Processing Magazine", 2008, vol. 25, n^{o} 2, p. 21–30.

[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] 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.

[89] 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.

[90] G. CARLIER, V. CHERNOZHUKOV, A. GALICHON.*Vector Quantile Regression*, Arxiv 1406.4643, 2014.

[91] 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/.

[92] G. CARLIER, X. DUPUIS.*An iterated projection approach to variational problems under generalized convexity constraints and applications*, In preparation, 2015.

[93] G. CARLIER, I. EKELAND.*Matching for teams*, in "Econom. Theory", 2010, vol. 42, n^{o} 2, p. 397–418.

[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] G. CARLIER, F. SANTAMBROGIO.*A continuous theory of traffic congestion and Wardrop equilibria*, in "Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)", 2011, vol. 390, n^{o} Teoriya Predstavlenii, Dinamicheskie Sistemy, Kombinatornye Metody. XX, p. 69–91, 307–308.

[98] 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.

[99] 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.

[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° 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° 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° 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° 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° 1, p. 20–31.

[108] M. CUTURI, D. AVIS.*Ground Metric Learning*, in "J. Mach. Learn. Res.", January 2014, vol. 15, n° 1, p. 533–564, http://dl.acm.org/citation.cfm?id=2627435.2627452.

[109] 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.

[110] Y. DE CASTRO, F. GAMBOA, D. HENRION, R. HESS, J. LASSERRE.*Approximate Optimal Designs for Multivariate Polynomial Regression*, October 2017, Accepted to Annals of Statistics, https://hal.laas.fr/hal-01483490.

[111] 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° 13-16, p. 1344–1386.

[112] 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.

[113] V. DUVAL, G. PEYRÉ.*Sparse Spikes Deconvolution on Thin Grids*, HAL, 2015, n° 01135200, http://hal.archives-ouvertes.fr/hal-01135200.

[114] J. FEHRENBACH, J.-M. MIREBEAU.*Sparse Non-negative Stencils for Anisotropic Diffusion*, in "Journal of Mathematical Imaging and Vision", 2014, vol. 49, n° 1, p. 123-147, http://dx.doi.org/10.1007/s10851-013-0446-3.

[115] 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.

[116] A. FIGALLI, R. MCCANN, Y. KIM.*When is multi-dimensional screening a convex program?*, in "Journal of Economic Theory", 2011.

[117] 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.

[118] 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.

[119] 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.

[120] 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.

[121] 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.

[122] 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.

[123] 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.

[124] 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.

[125] T. HASTIE, R. TIBSHIRANI, J. FRIEDMAN.*The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York Inc., New York, NY, USA, 2001.

[126] G. HERMAN.*Image reconstruction from projections: the fundamentals of computerized tomography*, Academic Press, 1980.

[127] D. D. HOLM, J. T. RATNANATHER, A. TROUVÉ, L. YOUNES.*Soliton dynamics in computational anatomy*, in "NeuroImage", 2004, vol. 23, p. S170–S178.

[128] 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.

[129] 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.

[130] 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.

[131] L. KANTOROVITCH.*On the translocation of masses*, in "C. R. (Doklady) Acad. Sci. URSS (N.S.)", 1942, vol. 37, p. 199–201.

[132] 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.

[133] 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.

[134] 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.

[135] 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.

[136] A. S. LEWIS.*Active sets, nonsmoothness, and sensitivity*, in "SIAM Journal on Optimization", 2003, vol. 13, n^{o} 3, p. 702–725.

[137] 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.

[138] 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).

[139] 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.

[140] D. LOMBARDI, E. MAITRE.*Eulerian models and algorithms for unbalanced optimal transport*, in "Preprint hal-00976501", 2013.

[141] 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.

[142] J. MAAS, M. RUMPF, C. SCHONLIEB, S. SIMON.*A generalized model for optimal transport of images including dissipation and density modulation*, in "Arxiv preprint", 2014.

[143] S. G. MALLAT.*A wavelet tour of signal processing*, Third, Elsevier/Academic Press, Amsterdam, 2009.

[144] 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.

[145] 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.

[146] J.-M. MIREBEAU.*Adaptive, Anisotropic and Hierarchical cones of Discrete Convex functions*, in "Preprint", 2014.

[147] 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.

[148] Q. MÉRIGOT.*A multiscale approach to optimal transport*, in "Computer Graphics Forum", 2011, vol. 30, n^{o} 5, p. 1583–1592.

[149] 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.

[150] 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/.

[151] B. PASS, N. GHOUSSOUB.*Optimal transport: From moving soil to same-sex marriage*, in "CMS Notes", 2013, vol. 45, p. 14–15.

[152] B. PASS.*Uniqueness and Monge Solutions in the Multimarginal Optimal Transportation Problem*, in "SIAM Journal on Mathematical Analysis", 2011, vol. 43, n^{o} 6, p. 2758-2775.

[153] J. PENNINGTON, R. SOCHER, C. MANNING.*Glove: Global Vectors for Word Representation*, in "Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)", Association for Computational Linguistics, 2014, p. 1532–1543.

[154] 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^{o} 1, p. 93-127, http://dx.doi.org/10.1007/s00205-013-0704-y.

[155] J. PETITOT.*The neurogeometry of pinwheels as a sub-riemannian contact structure*, in "Journal of Physiology-Paris", 2003, vol. 97, n^{o} 23, p. 265–309.

[156] G. PEYRÉ.*Texture Synthesis with Grouplets*, in "Pattern Analysis and Machine Intelligence, IEEE Transactions on", April 2010, vol. 32, n^{o} 4, p. 733–746.

[157] 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^{o} 1, p. 335–358.

[158] C. POON.*Structure dependent sampling in compressed sensing: theoretical guarantees for tight frames*, in "Applied and Computational Harmonic Analysis", 2015.

[159] C. PRINS, J.H.M. TEN. THIJE BOONKKAMP, J. VAN . ROOSMALEN, W.L. IJZERMAN, T.W. TUKKER.*A numerical method for the design of free-form reflectors for lighting applications*, in "External Report, CASA Report, No. 13-22", 2013, http://www.win.tue.nl/analysis/reports/rana13-22.pdf.

[160] H. RAGUET, J. FADILI, G. PEYRÉ.*A Generalized Forward-Backward Splitting*, in "SIAM Journal on Imaging Sciences", 2013, vol. 6, n^{o} 3, p. 1199–1226 [*DOI :* 10.1137/120872802], http://hal.archives-ouvertes.fr/hal-00613637/.

[161] J.-C. ROCHET, P. CHONÉ.*Ironing, Sweeping and multi-dimensional screening*, in "Econometrica", 1998.

[162] J. RUBINSTEIN, G. WOLANSKY.*Intensity control with a free-form lens*, in "J Opt Soc Am A Opt Image Sci Vis.", 2007, vol. 24.

[163] L. RUDIN, S. OSHER, E. FATEMI.*Nonlinear total variation based noise removal algorithms*, in "Physica D: Nonlinear Phenomena", 1992, vol. 60, n^{o} 1, p. 259–268, http://dx.doi.org/10.1016/0167-2789(92)90242-F.

[164] M. J. RUST, M. BATES, X. ZHUANG.*Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)*, in "Nature methods", 2006, vol. 3, n^{o} 10, p. 793–796.

[165] O. SCHERZER, M. GRASMAIR, H. GROSSAUER, M. HALTMEIER, F. LENZEN.*Variational Methods in Imaging*, Springer, 2008.

[166] 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.

[167] 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.

[168] B. SCHÖLKOPF, A. J. SMOLA.*Learning with kernels : support vector machines, regularization, optimization, and beyond*, Adaptive computation and machine learning, MIT Press, 2002, http://www.worldcat.org/oclc/48970254.

[169] 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.

[170] R. TIBSHIRANI.*Regression shrinkage and selection via the Lasso*, in "Journal of the Royal Statistical Society. Series B. Methodological", 1996, vol. 58, n^{o} 1, p. 267–288.

[171] A. TROUVÉ, F.-X. VIALARD.*Shape splines and stochastic shape evolutions: A second order point of view*, in "Quarterly of Applied Mathematics", 2012.

[172] 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/.

[173] 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.

[174] 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.

[175] C. VILLANI.*Topics in optimal transportation*, Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003, vol. 58, xvi+370.

[176] 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.

[177] 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.

[178] 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.

[179] 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.

Project-Team MYCENAE

Multiscale dYnamiCs in neuroENdocrine AxEs

RESEARCH CENTER
Paris

THEME
Modeling and Control for Life Sciences

Table of contents

Project-Team MYCENAE

Creation of the Project-Team: 2014 January 01, end of the Project-Team: 2017 December 31

Keywords:

Computer Science and Digital Science:

A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling (SPDE, SDE)
A6.1.3. - Discrete Modeling (multi-agent, people centered)
A6.1.4. - Multiscale modeling
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.3. - Probabilistic methods
A6.3.1. - Inverse problems
A6.3.4. - Model reduction

Other Research Topics and Application Domains:

B1.1.3. - Cellular biology
B1.1.4. - Developmental biology
B1.1.10. - Mathematical biology
B1.2.1. - Understanding and simulation of the brain and the nervous system
B2.2.2. - Nervous system and endocrinology

1. Personnel

Research Scientists

Frédérique Clément [Team leader, Inria, Senior Researcher, HDR]
Jonathan Touboul [Inria, Researcher, detached from Corps des Mines, HDR, until Apr 2017]

External Collaborators

Jean-Pierre Françoise [Univ Pierre et Marie Curie]
Marie Postel [Univ Pierre et Marie Curie]
Alexandre Vidal [Univ d'Evry Val d'Essonne]

PhD Students

Richard Bailleul [CIRB,until Apr 2017]
Yi Cui [UPMC,until Apr 2017]
Frédérique Robin [Inria]

Administrative Assistant

Martine Verneuille [Inria]

2. Overall Objectives

2.1. Overall Objectives

MYCENAE (Multiscale dYnamiCs in neuroENdocrine AxEs) is a project-team dedicated to mathematical neuroendocrinology and mathematical neuroscience. We are interested in the modeling, analysis and simulation of multiscale in time and/or space dynamics in the fields of neuroscience, endocrinology and physiology. Our main research topics are the followings:

- Numerical and theoretical studies of slow-fast systems with complex oscillations
- Non conservative transport equations for cell population dynamics
- Macroscopic limits of stochastic neural networks and neural fields

3. Research Program

3.1. Project team positioning

The main goal of MYCENAE is to address crucial questions arising from both Neuroendocrinology and Neuroscience from a mathematical perspective. The choice and subsequent study of appropriate mathematical formalisms to investigate these dynamics is at the core of MYCENAE's scientific foundations: slow-fast dynamical systems with multiple time scales, mean-field approaches subject to limit-size and stochastic effects, transport-like partial differential equations (PDE) and stochastic individual based models (SIBM).

The scientific positioning of MYCENAE is on the way between Mathematical Biology and Mathematics: we are involved both in the modeling of physiological processes and in the deep mathematical analysis of models, whether they be (i) models developed (or under development) within the team (ii) models developed by collaborating teams or (iii) benchmark models from the literature.

Our research program is grounded on previous results obtained in the framework of the REGATE (REgulation of the GonAdoTropE axis) Large Scale Initiative Action and the SISYPHE project team on the one hand, and the Mathematical Neuroscience Team in the Center for Interdisciplinary Research in Biology (Collège de France), on the other hand. Several of our research topics are related to the study and generalization of 2 master models: a 4D, multiscale in time, nonlinear model based on coupled FitzHugh-Nagumo dynamics that has proved to be a fruitful basis for the study of the complex oscillations in hypothalamic GnRH dynamics [27], [26], and a nD, multiscale in space, system of weakly-coupled non conservative transport equations that underlies our approach of gonadal cell dynamics [28],[8]. Most our topics in mathematical neuroscience deal with the study of complex oscillatory behaviors exhibited either by single neurons or as emergent macroscopic properties of neural networks, from both a deterministic and stochastic viewpoint.

3.2. Numerical and theoretical studies of slow-fast systems with complex oscillations

In dynamical systems with at least three state variables, the presence of different time scales favors the appearance of complex oscillatory solutions. In this context, with (at least) two slow variables MixedMode Oscillations (MMO) dynamics can arise. MMOs are small and large amplitude oscillations combined in a single time series. The last decade has witnessed a significant amount of research on this topic, including studies of folded singularities, construction of MMOs using folded singularities in combination with global dynamics, effects of additional time scales, onset of MMOs via singular Hopf bifurcations, as well as generalization to higher dimensions. In the same period, many applications to neuroscience emerged [9]. On the other hand, bursting oscillations, another prototype of complex oscillations can occur in systems with (at least) two fast variables. Bursting has been observed in many biological contexts, in particular in the dynamics of pancreatic cells, neurons, and other excitable cells. In neuronal dynamics a burst corresponds to a series of spikes, interspersed with periods of quiescent behavior, called inter-burst intervals. We are interested in systems combining bursting, MMOs and canards. One of the interesting directions is torus canards, which are canard-like structures occurring in systems combining canard explosion with fast rotation [5]. Torus canards help understand transitions from spiking or MMO dynamics to bursting. Another study on the boundary of bursting and MMOs is the work of [36] on the so-called plateau bursting. A major challenge in this direction is to gain a complete understanding of the transition from "3 time scales" to "2 fast/ 1 slow" (bursting) and then to "1 fast/ 2 slow (MMOs)". Also, a key challenge that we intend to tackle in the next few years is that of large dynamical systems with many fast and many slow variables, which additionally are changing in time and/or in phase space. We aim to pursue this research direction both at theoretical and computational level, using numerical continuation approaches based on the location of unstable trajectories by using fixed point methods, rather than simulation, to locate trajectories.

3.3. Non conservative transport equations for cell population dynamics

Models for physiologically-structured populations can be considered to derive from the so-called McKendrick-Von Foerster equation or renewal equation that has been applied and generalized in different applications of population dynamics, including ecology, epidemiology and cell biology. Renewal equations are PDE transport equations that are written so as to combine conservation laws (e.g. on the total number of individuals) with additional terms related to death or maturation, that blur the underlying overall balance law [33]. Renewal equations can be deployed only in contexts where a deterministic and continuous formalization is suitable to describe the populations. In the case of low or very low number of individuals, the probabilistic nature of the events driving the population dynamics has to be accounted for. In that context, the formalism initially developed in the framework of ecological modeling (see e.g. [30]) is particularly interesting since it can bridge the gap between branching processes and renewal equations for structured populations.

The development of ovarian follicles is a tightly-controlled physiological and morphogenetic process, that can be investigated from a middle-out approach starting at the cell level.

To describe the terminal stages of follicular development on a cell kinetics basis and account for the selection process operated amongst follicles, we have developed a multiscale model describing the cell density in each follicle, that can be roughly considered as a system of weakly-coupled, non conservative transport equations with controlled velocities and source term. Even if, in some sense, this model belongs to the class of renewal equations for structured populations, it owns a number of specificities that render its theoretical and numerical analysis particularly challenging: 2 structuring variables (per follicle, leading as a whole to $2n$D system), control terms operating on the velocities and source term, and formulated from moments of the unknowns, discontinuities both in the velocities and density on internal boundaries of the domain representing the passage from one cell phase to another. On the theoretical ground, the well-posedness (existence and uniqueness of weak solutions with bounded initial data) has been established in [12], while associated control problems have been studied in the framework of hybrid optimal control [6]. On the numerical ground, the formalism dedicated to the simulation of these hyperbolic-like PDEs is that of finite volume method. Part of the numerical strategy consists in combining in the most efficient way low resolution numerical schemes (such as the first-order Godunov scheme), that tend to be diffusive, with high resolution schemes (such as the Lax Wendroff second-order scheme), that may engender oscillations in the vicinity of discontinuities [2], with a critical choice of the limiter functions. The 2D finite volume schemes are combined with adaptive mesh refinement through a multi-resolution method [4] and implemented in a problem-specific way on parallel architecture [1].

To describe the first stages of follicular development, which only involve a few cells, we call to a stochastic and discrete formalism. We have designed an individual-based, stochastic model [7] embedding specific laws of morphodynamics, which leads to a multiscale model where individual cells are endowed with a non-zero size and occupy space partitions of predefined sizes, which can bear a limit rate of overcrowding.

3.4. Macroscopic limits of stochastic neural networks and neural fields

The coordinated activity of the cortex is the result of the interactions between a very large number of cells. Each cell is well described by a dynamical system, that receives non constant input which is the superposition of an external stimulus, noise and interactions with other cells. Most models describing the emergent behavior arising from the interaction of neurons in large-scale networks have relied on continuum limits ever since the seminal work of Wilson and Cowan and Amari [37], [25]. Such models tend to represent the activity of the network through a macroscopic variable, the population-averaged firing rate.

In order to rationally describe neural fields and more generally large cortical assemblies, one should yet base their approach on what is known of the microscopic neuronal dynamics. At this scale, the equation of the activity is a set of stochastic differential equations in interaction. Obtaining the equations of evolution of the effective mean-field from microscopic dynamics is a very complex problem which belongs to statistical physics. As in the case of the kinetic theory of gases, macroscopic states are defined by the limit of certain quantities as the network size tends to infinity. When such a limit theorem is proved, one can be ensured that large networks are well approximated by the obtained macroscopic system. Qualitative distinctions between

the macroscopic limit and finite-sized networks (finite-size effects), occurs in such systems. We have been interested in the relevant mathematical approaches dealing with macroscopic limits of stochastic neuronal networks, that are expressed in the form of a complex integro-differential stochastic implicit equations of McKean-Vlasov type including a new mathematical object, the spatially chaotic Brownian motion [15].

The major question consists in establishing the fundamental laws of the collective behaviors cortical assemblies in a number of contexts motivated by neuroscience, such as communication delays between cells [14], [13] or spatially extended areas, which is the main topic of our current research. In that case additional difficulties arise, since the connection between different neurons, as well as delays in communications, depend on space in a correlated way, leading to the singular dependence of the solutions in space, which is not measurable.

4. Application Domains

4.1. Introduction

MYCENAE addresses rather "upstream" questions in neuroendocrinology and neuroscience. Nevertheless, MYCENAE's expected results can contribute to more applied issues in these fields, mainly by helping understand the mechanisms underlying physiological and pathological processes and also by designing new concepts for biomedical data analysis. MYCENAE thematics are related to societal issues concerning endocrine disruptors, reproductive biotechnologies, and neurological diseases, especially in case of pathological synchronizations encountered in epilepsy and Parkinson's disease.

4.2. Neuroendocrinology and Neuroscience

We are interested in the complex dynamical processes arising within neuroendocrine axes, with a special focus on the reproductive (hypothalamo-pituitary-gonadal) axis. This axis can be considered as the paragon of neuroendocrine axes, since it both concentrates all remarkable dynamics that can be exhibited by these axes and owns its unique specificities, as gonads are the only organs that host germ cells. Since, in neuroendocrine axes, neural systems are embedded within endocrine feedback loops and interact with peripheral organs, one also needs to get interested in the peripheral dynamics to be able to "close the loop" and account for the effect of peripheral inputs on neural dynamics. In the case of the HPG axis, these dynamics are especially complex, because they involve developmental processes that occur even in adult organisms and combine the glandular function of the gonads with their gametogenic function.

Neuroendocrinology is thus a scientific field at the interface between Neuroscience, Endocrinology and Physiology (and even of Developmental Biology in the case of the HPG axis). On a neuroscience ground, mathematical neuroendocrinology is specifically interested in endocrine neurons, which have the uncommon ability of secreting neurohormones into the blood stream. Neuroendocrine networks are characterized by the emergence of very slow rhythms (on the order of an hour), finite size effects due to their relative small number of neurons (on the order of a few thousands for the Gonadotropin-Releasing-Hormone network) and neuroanatomical particularities, that impact the way they can synchronize and desynchronize. On a physiological ground, gonadal cell biology raises specific cell biology issues on more than one account. First, the gonads are the only organs sheltering the germ cell lines (corresponding to oogenesis in ovaries and spermatogenesis in testes). Hence, the two modes of cell division, mitosis and meiosis are encountered in these tissues. Second, there are intricate interactions between the gonadal somatic cells (granulosa cells in the ovaries, sertoli cells in the testes) and the germ cells. Third, the control of gonadal cell populations is exerted within endocrine feedback loops involving both the hypothalamus and pituitary, which results naturally in multiscale population dynamics coupled with hormonally-controlled cell kinetics.

MYCENAE's research topics in mathematical neuroscience deal with complex oscillations, synchronization and plasticity.

We study (i) the emergence of network-level behaviors from individual dynamics of excitable cells (mainly neurons, but not exclusively, as the pituitary cells belong to the family of excitable cells): complete synchronization or synchronization of specific events, effect of the recruitment rate in the synchronization process, dependence on the neuro-anatomical and functional coupling properties; (ii) the control of the different possible configurations of the network depending on external (e.g. daylength) and/or internal inputs (e.g. metabolic status), at the source of plasticity processes in cognitive (vision learning) or neuroendocrine systems (differential sensitivity to gonadal steroids and peptides across the different steps of the reproductive life); (iii) the encoding of neuro-hormonal signals as complex oscillations, on the electrical, ionic (calcium dynamics) and secretory levels; and (iv) the decoding of these signals by their target neuronal or non-neuronal cells.

More recently, we have been interested into developmental biology issues in neurosciences: neurogenesis and brain development. The anatomical and functional organization of the nervous system, and especially the brain, is highly structured and tightly regulated. The surface of the cortex, its thickness, but also the size and shape of the brain areas associated to the different sensory or motor areas are very reliable quantities across different individuals. In collaboration with different teams of biologists, we develop and investigate models of the development of the brain, at different time and spatial scale.

The biological relevance of our modeling and model-based signal analysis approaches is grounded on our network of collaborations with teams of experimentalist biologists. In particular, we have long standing collaborations with the UMR 6175 (INRA-CNRS-Université François Rabelais-Haras Nationaux) "Physiologie de la Reproduction et des Comportements" that covers most our research topics in reproductive neuroendocrinology. We have especially close links with the Bingo (Integrative Biology of the ovary) and Bios (Biology and Bioinformatics of Signaling Systems) teams, which were partners of the REGATE LSIA. We have been jointly investigating issues relative to terminal or basal follicular development [7], [8], analysis of neurosecretory patterns [16] and modeling of GPCR (G-Protein Coupled Receptors) signaling networks [10]. We also have special links with the Center for Interdisciplinary Research in Biology (CIRB, Collège de France), headed by Alain Prochiantz, that help us get a better understanding of how the brain connectivity develops and how it is functionally organized. An instance of a recent collaborative work is the study of the organization of spatial frequencies in the primary visual cortex [34].

5. Highlights of the Year

5.1. Highlights of the Year

- We have completed in [17] our series of studies [8], [12], [6], [2], [4], [3] on the mathematical and numerical analysis of our multiscale model of structured cell populations in terminally developing ovarian follicles.
- We have completed in [19] our series of studies [27], [26], [35], [29], [32] on the mathematical and numerical analysis of our model of GnRH pulse and surge generator.

6. New Software and Platforms

6.1. DynPeak

KEYWORDS: Biology - Health - Physiology
SCIENTIFIC DESCRIPTION: DynPeak is an algorithm for pulse detection and frequency analysis in hormonal time series.

- Participants: Alexandre Vidal, Claire Médigue, Frédérique Clément, George Rosca, Qinghua Zhang and Serge Steer
- Partner: INRA
- Contact: Frédérique Clément
- URL: https://team.inria.fr/mycenae/en/software/

7. New Results

7.1. Numerical and theoretical studies of slow-fast systems with complex oscillations

7.1.1. Coupled multiple timescale dynamics in populations of endocrine neurons: Pulsatile and surge patterns of GnRH secretion

Participants: Elif Köksal Ersöz, Alexandre Vidal, Frédérique Clément.

We have finalized the study of a 6D extension of our model of GnRH pulse and surge generator, which has now been published [19]. The gonadotropin releasing hormone (GnRH) is secreted by hypothalamic neurons into the pituitary portal blood in a pulsatile manner. The alternation between a frequency-modulated pulsatile regime and the ovulatory surge is the hallmark of the GnRH secretion pattern in ovarian cycles of female mammals. In this work, we aimed at modeling additional features of the GnRH secretion pattern: the possible occurrence of a two-bump surge ("camel surge") and an episode of partial desynchronization before the surge. We have proposed a six-dimensional extension of a former four-dimensional model with three timescale and introduced two mutually-coupled, slightly heterogenous GnRH subpopulations (secretors) regulated by the same slow oscillator (regulator). We have considered two types of coupling functions between the secretors, including dynamic state-dependent coupling, and we have used numerical and analytic tools to characterize the coupling parameter values leading to the generation of a two-bump surge in both coupling cases. We have revealed the impact of the slowly varying control exerted by the regulator onto the pulsatile dynamics of the secretors, which leads to dynamic bifurcations and gives rise to desynchronization. To assess the occurrence time of desynchronization during the pulsatile phase, we have introduced asymptotic tools based on quasi-static and geometric approaches, as well as analytic tools based on the H-function derived from phase equation and numerical tracking of period-doubling bifurcations. We discuss the role of coupling parameters in the two-bump surge generation and the speed of desynchronization.

7.1.2. Wild oscillations in a nonlinear neuron model with resets

Participants: Jonathan Rubin [University of Pittsburgh], Justyna Signerska-Rynkowska, Jonathan Touboul, Alexandre Vidal.

We have finalized the work undergone in a series of two studies, where we have investigated the mechanisms by which complex oscillations are generated in a class of nonlinear dynamical systems with resets modeling the voltage and adaptation of neurons. These studies have been published as a two-part article [21], [22].

The first study [21] presents a mathematical analysis showing that the system can support bursts of any period as a function of model parameters, and that are organized in a period-incrementing structure. In continuous dynamical systems with resets, such period-incrementing structures are complex to analyze. In the present context, we have used the fact that bursting patterns correspond to periodic orbits of the adaptation map that governs the sequence of values of the adaptation variable at the resets. Using a slow-fast approach, we have shown that this map converges towards a piecewise linear discontinuous map whose orbits are exactly characterized. That map shows a period-incrementing structure with instantaneous transitions. We have further shown that the period-incrementing structure persists for the full system with non-constant adaptation, yet the transitions are more complex. We have also established the presence of chaos at the transitions.

The second study [22] shows that these neuron models can generically display a form of mixed-mode oscillations (MMOs), which are trajectories featuring an alternation of small oscillations with spikes or bursts (multiple consecutive spikes). The mechanism by which these are generated relies fundamentally on the hybrid structure of the flow: invariant manifolds of the continuous dynamics govern small oscillations, while discrete resets govern the emission of spikes or bursts, contrasting with classical MMO mechanisms in ordinary differential equations involving more than three dimensions and generally relying on a timescale separation. The decomposition of mechanisms reveals the geometrical origin of MMOs, allowing a relatively simple classification of points on the reset manifold associated to specific numbers of small oscillations. We

have shown that the MMO pattern can be described through the study of orbits of a discrete adaptation map, which is singular as it features discrete discontinuities with unbounded left- and right-derivatives. We have studied the orbits of the map via rotation theory for circle maps and elucidated in detail complex behaviors arising in the case where MMOs display a single small oscillation per cycle.

7.1.3. Studies of the Petrov module for a family of generalized Liénard integrable systems

Participants: Lucile Megret [UPMC], Jean-Pierre Françoise [UPMC].

In [20], we have used the Lambert function in order to study a family of integrable generalized Liénard equations X_f which display a center. We have first proven a conjugation lemma inside a continuum of nested periodic orbits. Then we have deduced an explicit operator of Gelfand-Leray associated with the Hamiltonian of equation X_f. Afterwards, we have provided a generating family for the associated Petrov module. Finally, by using the Lambert function, we have studied the monotonicity of the Abelian integral of this generating family's elements.

7.2. Non conservative transport equations for cell population dynamics

7.2.1. Dimensional reduction of a multiscale model based on long time asymptotics

Participants: Frédérique Clément, Frédéric Coquel [CMAP], Marie Postel, Kim Long Tran.

We have finalized the study on the dimensional reduction of our multiscale model of terminal follicle development, which has now been published [17]. We have considered a class of kinetic models for which a moment equation has a natural interpretation. We have shown that, depending on their velocity field, some models lead to moment equations that enable one to compute monokinetic solutions economically. We have detailed the example of a multiscale structured cell population model, consisting of a system of 2D transport equations. The reduced model, a system of 1D transport equations, is obtained from computing the moments of the 2D model with respect to one variable. The 1D solution is defined from the solution of the 2D model starting from an initial condition that is a Dirac mass in the direction removed by reduction. For arbitrary initial conditions, we have compared 1D and 2D model solutions in asymptotically large time. Finite volume numerical approximations of the 1D reduced model can be used to compute the moments of the 2D solution with proper accuracy, both in the conservative and non conservative framework. The numerical robustness is studied in the scalar case, and a full scale vector case is presented.

7.2.2. Analysis and calibration of a linear model for structured cell populations with unidirectional motion : application to the morphogenesis of ovarian follicles

Participants: Frédérique Clément, Frédérique Robin, Romain Yvinec [INRA].

We have analyzed a multi-type age dependent model for cell populations subject to unidirectional motion, in both a stochastic and deterministic framework [23]. Cells are distributed into successive layers; they may divide and move irreversibly from one layer to the next. We have adapted results on the large-time convergence of PDE systems and branching processes to our context, where the Perron-Frobenius or Krein-Rutman theorem can not be applied. We have derived explicit analytical formulas for the asymptotic cell number moments, and the stable age distribution. We have illustrated these results numerically and we have applied them to the study of the morphodynamics of ovarian follicles. We have proven the structural parameter identifiability of our model in the case of age independent division rates. Using a set of experimental biological data, we have estimated the model parameters to fit the changes in the cell numbers in each layer during the early stages of follicle development.

This work has been undergone in the framework of the PhD of Frédérique Robin. It has been the matter of a poster at ReprosSciences2017 [24] (April 10-12) and of an oral presentation (Dynamiques de populations cellulaires structurées) at the annual meeting (September 27-29) of GDR MaMovi (Mathématiques Appliquées à la MOdélisation du VIvant).

7.2.3. Mathematical modeling of progenitor cell populations in the mouse cerebral cortex

Participants: Frédérique Clément, Alice Karam [IBPS], Matthieu Perez, Marie Postel, Sylvie Schneider-Maunoury [IBPS].

We have finalized the study of our PDE-based model of structured cell populations during the development of cerebral cortex. The model accounts for three main cell types: apical progenitors (APs), intermediate progenitors (IPs), and neurons. Each cell population is structured according to the cell age distribution. Since the model describes the different phases of the cell division cycle, we could derive the numeric equivalents of many of the experimental indexes measured in experimental setups, including classical mitotic or labeling indexes targeting the cells in phase S or mitosis, and more elaborated protocols based on double labeling with fluorescent dyes. We have formulated a multi-criterion objective function which enables us to combine experimental observations of different nature and to fit the data acquired in the framework of the NeuroMathMod project (Sorbonne-Universités Émergence call with IBPS, Institut de Biologie Paris Seine). Great efforts have been put on the experimental side to provide the model with the quantitative values of cell numbers for both progenitors and neurons. With the retrieved parameters, the model can provide useful information not supplied by the data, such as the cell origin of neurons (direct neurogenesis from AP or IPgenic neurogenesis) and the proportion of IPs cells undergoing several rounds of cell cycles. In addition, we have compared the cell dynamics patterns observed in wild-type mice with respect to mutant mice used as an animal model of human ciliopathies.

In the framework of the internship of Matthieu Perez (INSA Rouen, co-supervised by Frédérique Clément and Marie Postel), we have investigated numerically the link between our deterministic, PDE-based model of progenitor and neuron cell dynamics, and possible stochastic counterparts inspired from previous work in the team [31]. The deterministic approach is averaged with respect to the deterministic one, since it does not account for the trajectories of individual cells, yet it describes in more details the progression of cells within the cell cycle since it explicitly embeds the structuring of the cell cycle into different phases. The work has consisted in comparing the main model outputs (numbers of progenitors and neurons as a function of time) obtained by numerical simulations based on characteristics, on the deterministic side, or Gillespie algorithms, on the stochastic side. A proper strategy had to be settled to deal with the main difficulties raised by this comparison, namely the time-varying rates involved in the stochastic transition rates from one cell type to another, and the matching between the average stochastic rates and the deterministic rates ruling cell kinetics, especially the cell cycle duration.

8. Partnerships and Cooperations

8.1. European Initiatives

Together with our BIOS INRA partner, we have participated in a synergistic way in the proposal EVE (*In-Silico Safety and Efficacy Assessment of Reproductive Endocrinology Treatments*) submitted to the H2020-SC1-2016-2017 call (Personalised Medicine), whose PI was Enrico Tronci (Sapienza, Roma).

8.2. National Initiatives

8.2.1. ANR

Jonathan Touboul is member of the Kibord (KInetic models in Biology Or Related Domains) project obtained in 2014.

He is also PI of the projects “Mathematical modeling of synaptic plasticity” (with Laurent Venance, CIRB) funded as an interdisciplinary structuring project of INSB (Institut des Sciences Biologiques in CNRS) and “Altering Fear Memory” (with Sidney Wiener, CIRB and Karim Benchenane, ESPCI) funded by the PSL Labex MemoLife.

8.2.2. National Networks

- GdR REPRO (F. Clément is member of the direction board)
- MIA REM network: Réduction de modèles (PI Béatrice Laroche, INRA Jouy)

8.2.3. National Collaborations

- UMR Physiologie de la Reproduction et des Comportements, INRA Centre- Val de Loire (Bios and Bingo teams)
- Université Pierre & Marie Curie (UPMC)
  - Jacques-Louis Lions Laboratory, Pierre & Marie Curie University (Jean-Pierre Françoise, Marie Postel)
  - Developmental Biology Laboratory, Institut de Biologie Paris Seine (IBPS), Pierre & Marie Curie University (Alice Karam, Sylvie Schneider Maunoury), in the framework of the NeuroMathMod, Sorbonne-Universités Émergence call
- Center for Interdisciplinary Research in Biology (CIRB), Collège de France (Alain Prochiantz, Marie Manceau, Laurent Venance)

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Organizing Committees

- Reprosciences 2017, April 10-12, Tours, co-organized by Frédérique Clément, Yves Combarnous, Florian Guillou, Joëlle Cohen-Tannoudji and François Vialard.

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

Jonathan Touboul participates in the editorial boards of *Plos One* and *Frontiers in neuronal circuits*

9.1.2.2. Reviewer - Reviewing Activities

Journal of Ovarian research, Journal of Mathematical Biology, SIAM Journal on Applied Dynamical Systems

9.1.3. Invited Talks

Frédérique Clément gave a talk dedicated to "'Multiscale modeling in reproductive and developmental biology : A middle-out, cell dynamics-based approach" during the Journées Scientifiques Inria, June, 14-16, Sophia-Antipolis.

Our study on the analysis and calibration of a linear model for structured cell populations with unidirectional motion was the matter of an invited presentation, given by Romain Yvinec, at the annual meeting of GDR MaMovi , September 27-29, Villeurbanne (https://gdr-mamovi-2017.sciencesconf.org).

9.1.4. Scientific Expertise

Frédérique Clément belongs to the expert board of the BCDE (Cell Biology, Development and Evolution) ITMO (Multi OrganizationThematic Institute) of the French National Alliance for Life and Health Sciences Aviesan.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Frédérique Robin gave lectures in the framework of the UFR de Mathématiques de Jussieu (L2 and L3 level)

UE P2.2M003 Vector analysis (L3) 85h

UE E.2M261 Integer and integral series with parameters (L2), 20h

UE 3M100 Initiation to Python language (L3), 15h

9.2.2. Supervision

PhD in progress : Richard Bailleul. Modeling of the developmental mechanisms underlying the formation of color and appendage patterns in birds, since September 2015. Université Pierre & Marie Curie (ED515), supervisors: Benoît Perthame, Marie Manceau and Jonathan Touboul (funded by the ERC starting grant of Marie Manceau)

PhD in progress: Yi Cui. Role of Pax6 in neurodevelopment: experiments and models, since September 2014, Université Pierre & Marie Curie (ED158), supervisors: Jonathan Touboul, Alain Prochiantz and Alessandra Pierani

PhD in progress: Frédérique Robin. Multiscale modeling of the morphodynamics in ovarian follicles, since October 2016, Université Pierre & Marie Curie (ED386), supervisors: Frédérique Clément and Romain Yvinec (INRA)

Master degree: Matthieu Perez. Simulations de modèles déterministes et stochastiques de la neurogenèse corticale. INSA Rouen, supervisors: Frédérique Clément and Marie Postel

9.3. Popularization

Frédérique Robin participated in a regional committee in the framework of the “Tournoi français des jeunes mathématiciennes et mathématiciens” (TFJM).

Book chapter in the new edition of the Encyclopedia of Endocrine Diseases
D. Monniaux, V. Cadoret, F. Clément, R. Dalbies-Tran, S. Elis, S. Fabre, V. Maillard, P. Monget, and S. Uzbekova
Editors Ilpo Huhtaniemi and Luciano Martini, Section Editor Sophie Christin-Maitre
Folliculogenesis (2017) p1-26.

10. Bibliography

Major publications by the team in recent years

[1] B. AYMARD, F. CLÉMENT, F. COQUEL, M. POSTEL.*Numerical simulation of the selection process of the ovarian follicles*, in "ESAIM Proc.", 2012, vol. 28, p. 99-117.

[2] B. AYMARD, F. CLÉMENT, F. COQUEL, M. POSTEL.*A numerical method for cell dynamics; kinetic equations with discontinuous coefficients*, in "SIAM J. Sci. Comput.", 2013, vol. 35, p. A2442-A2468.

[3] B. AYMARD, F. CLÉMENT, D. MONNIAUX, M. POSTEL.*Cell-kinetics based calibration of a multiscale model of structured cell populations in ovarian follicles*, in "SIAM J. Appl. Math.", 2016, vol. 76, n^{o} 4, p. 1471–1491.

[4] B. AYMARD, F. CLÉMENT, M. POSTEL.*Adaptive mesh refinement strategy for a non conservative transport problem*, in "ESAIM Math. Model. Numer. Anal.", 2014, vol. 48, n^{o} 5, p. 1381-1412.

[5] J. BURKE, M. DESROCHES, A. BARRY, T. KAPER, M. KRAMER.*A showcase of torus canards in neuronal bursters*, in "J. Math. Neurosci.", 2012, vol. 2.

[6] F. CLÉMENT, J.-M. CORON, P. SHANG.*Optimal control of cell mass and maturity in a model of follicular ovulation*, in "SIAM J. Control Optim.", 2013, vol. 51, n^{o} 2, p. 824-847.

[7] F. CLÉMENT, P. MICHEL, D. MONNIAUX, T. STIEHL.*Coupled somatic cell kinetics and germ cell growth: multiscale model-based insight on ovarian follicular development*, in "Multiscale Model. Simul.", 2013, vol. 11, n^{o} 3, p. 719-746.

[8] F. CLÉMENT, D. MONNIAUX.*Multiscale modelling of follicular selection*, in "Prog. Biophys. Mol. Biol.", 2013, vol. 113, p. 398-408.

[9] M. DESROCHES, J. GUCKENHEIMER, B. KRAUSKOPF, C. KUEHN, H. OSINGA, M. WECHSELBERGER.*Mixed-mode oscillations with multiple time scales*, in "SIAM Rev.", 2012, vol. 54, p. 211–288.

[10] D. HEITZLER, G. DURAND, A. RIZK, S. AHN, J. KIM, J. VIOLIN, L. DUPUY, C. GAUTHIER, V. PIKETTY, P. CRÉPIEUX, A. POUPON, F. CLÉMENT, F. FAGES, R. LEFKOWITZ, E. REITER.*Competing G protein-coupled receptor kinases balance G protein and β-arrestin signaling*, in "Mol. Syst. Biol.", 2012, vol. 8, n^{o} 590.

[11] M. KRUPA, A. VIDAL, F. CLÉMENT.*A network model of the periodic synchronization process in the dynamics of calcium concentration in GnRH neurons*, in "J. Math. Neurosci.", 2013, vol. 3, 4.

[12] P. SHANG.*Cauchy problem for multiscale conservation laws: Application to structured cell populations*, in "J. Math. Anal. Appl.", 2013, vol. 401, n^{o} 2, p. 896-920.

[13] J. TOUBOUL.*Limits and dynamics of stochastic neuronal networks with random delays*, in "J. Stat. Phys.", 2012, n^{o} 149, p. 569–597.

[14] J. TOUBOUL.*Mean-Field equations for stochastic firing-rate neural fields with delays: derivation and noise-induced transitions*, in "Phys. D", 2012, vol. 241, p. 1223–1244.

[15] J. TOUBOUL.*Propagation Of Chaos In Neural Fields*, in "Ann. Appl. Probab.", 2014, vol. 24, n^{o} 3, p. 1298–1327.

[16] A. VIDAL, Q. ZHANG, C. MÉDIGUE, S. FABRE, F. CLÉMENT.*DynPeak: An algorithm for pulse detection and frequency analysis in hormonal time series*, in "PloS One", 2012, vol. 7, e39001.

Publications of the year

Articles in International Peer-Reviewed Journal

[17] F. CLÉMENT, F. COQUEL, M. POSTEL, K. L. TRAN.*Dimensional Reduction of a Multiscale Model Based on Long Time Asymptotics*, in "Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal", 2017, vol. 15, n^{o} 3, p. 1198 - 1241 [*DOI :* 10.1137/16M1062545], http://hal.upmc.fr/hal-01325275.

[18] E. KÖKSAL ERSÖZ, M. DESROCHES, M. KRUPA.*Synchronization of weakly coupled canard oscillators*, in "Physica D: Nonlinear Phenomena", June 2017, vol. 349, p. 46-61 [*DOI :* 10.1016/J.PHYSD.2017.02.016], https://hal.inria.fr/hal-01558897.

[19] E. KÖKSAL ERSÖZ, A. VIDAL, F. CLÉMENT.*Coupled multiple timescale dynamics in populations of endocrine neurons: Pulsatile and surge patterns of GnRH secretion*, in "SIAM Journal on Applied Dynamical Systems", 2017, p. 1-29, https://hal.inria.fr/hal-01666251.

[20] L. MÉGRET.*Studies of the Petrov Module for a Family of Generalized Liénard Integrable Systems*, in "Qualitative Theory of Dynamical Systems", 2017, vol. 20, n° 2, p. 1-21 [*DOI :* 10.1007/S12346-017-0250-3], https://hal.inria.fr/hal-01571808.

[21] J. RUBIN, J. SIGNERSKA-RYNKOWSKA, J. TOUBOUL, A. VIDAL.*Wild oscillations in a nonlinear neuron model with resets: (I) Bursting, spike adding and chaos*, in "Discrete and Continuous Dynamical Systems - Series B (DCDS-B)", September 2017, vol. 22, p. 3967-4002, https://arxiv.org/abs/1611.02740 , https://hal.inria.fr/hal-01416002.

[22] J. SIGNERSKA-RYNKOWSKA, J. TOUBOUL, A. VIDAL, J. RUBIN.*Wild oscillations in a nonlinear neuron model with resets: (II) Mixed-mode oscillations*, in "Discrete and Continuous Dynamical Systems - Series B", 2017, vol. 22, p. 4003-4039, https://arxiv.org/abs/1509.08282 , https://hal.inria.fr/hal-01256368.

Other Publications

[23] F. CLÉMENT, F. ROBIN, R. YVINEC.*Analysis and calibration of a linear model for structured cell populations with unidirectional motion : Application to the morphogenesis of ovarian follicles*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01666373.

[24] F. CLÉMENT, F. ROBIN, R. YVINEC.*Multiscale modeling of the morphogenesis of ovarian follicles: a cell-dynamics based approach*, April 2017, 75, 2. Journées du GdR 3606 Repro , Poster, https://hal.archives-ouvertes.fr/hal-01606594.

References in notes

[25] P. BRESSLOFF.*Spatiotemporal dynamics of continuum neural fields*, in "J. Phys. A: Math. Theor.", 2012, vol. 45, p. 033001–.

[26] F. CLÉMENT, A. VIDAL.*Foliation-based parameter tuning in a model of the GnRH pulse and surge generator*, in "SIAM J. Appl. Dyn. Syst.", 2009, vol. 8, n° 4, p. 1591–1631.

[27] F. CLÉMENT, J.-P. FRANÇOISE.*Mathematical modeling of the GnRH-pulse and surge generator*, in "SIAM J. Appl. Dyn. Syst.", 2007, vol. 6, p. 441-456.

[28] N. ECHENIM, D. MONNIAUX, M. SORINE, F. CLÉMENT.*Multi-scale modeling of the follicle selection process in the ovary*, in "Math. Biosci.", 2005, vol. 198, p. 57-79.

[29] S. FERNÁNDEZ-GARCÍA, M. DESROCHES, M. KRUPA, F. CLÉMENT.*A Multiple Time Scale Coupling Of Piecewise Linear Oscillators. Application To A Neuroendocrine System*, in "Siam Journal On Applied Dynamical Systems", 2015, vol. 14, n° 2, 31.

[30] N. FOURNIER, S. MÉLÉARD.*A microscopic probabilistic description of a locally regulated population and macroscopic approximations*, in "Ann. Appl. Probab", 2004, vol. 14, p. 119-144.

[31] B. FRERET-HODARA, Y. CUI, A. GRIVEAU, L. VIGIER, Y. ARAI, J. TOUBOUL, A. PIERANI.*Enhanced Abventricular Proliferation Compensates Cell Death in the Embryonic Cerebral Cortex*, in "Cereb. Cortex", 2016, p. 1-18.

[32] E. KÖKSAL-ERSÖZ, M. DESROCHES, M. KRUPA, F. CLÉMENT.*Canard-Mediated (De)Synchronization in Coupled Phantom Bursters*, in "SIAM J. Appl. Dyn. Syst.", 2016, vol. 15, n^{o} 1, p. 580-608.

[33] B. PERTHAME.*Transport equations in biology*, Frontiers in Mathematics, Birkhauüser, Basel, 2007.

[34] J. RIBOT, A. ROMAGNONI, C. MILLERET, D. BENNEQUIN, J. TOUBOUL.*Pinwheel-dipole configuration in cat early visual cortex*, 2014, p. 63–73.

[35] A. VIDAL, F. CLÉMENT.*A dynamical model for the control of the GnRH neurosecretory system*, in "J. Neuroendocrinol.", 2010, vol. 22, p. 1251–1266.

[36] T. VO, R. BERTRAM, J. TABAK, M. WECHSELBERGER.*Mixed mode oscillations as a mechanism for pseudo-plateau bursting*, in "J. Comput. Neurosci.", 2010, vol. 28, n^{o} 3, p. 443–458.

[37] H. WILSON, J. COWAN.*Excitatory and inhibitory interactions in localized populations of model neurons*, in "Biophys. J.", 1972, vol. 12, p. 1–24.

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

Keywords:

Computer Science and Digital Science:

A1.1.1. - Multicore, Manycore
A1.1.3. - Memory models
A2.1.1. - Semantics of programming languages
A2.1.3. - Functional programming
A2.1.6. - Concurrent programming
A2.1.8. - Synchronous languages
A2.2.2. - Memory models
A2.2.3. - Run-time systems
A2.2.4. - Parallel architectures
A2.2.5. - GPGPU, FPGA, etc.
A2.2.6. - 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
B9.4.1. - Computer science
B9.4.2. - Mathematics

1. Personnel

Research Scientists

Timothy Bourke [Inria, Starting Research Position]
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

Michael Kruse [Inria]
Oleksandr Zinenko [Inria]

PhD Students

Guillaume Baudart [Ecole Normale Supérieure Paris, until Mar 2017]
Ulysse Beaugnon [Ecole Normale Supérieure Paris]
Lelio Brun [Ecole Normale Supérieure Paris]
Robin Morisset [Inria, until Feb 2017]
Chandan Reddy Gopal [Inria]
Jie Zhao [National Digital Switching System Engineering and Technological Research Center, Zhengzhou, China]

Post-Doctoral Fellow

Guillaume Iooss [Ecole Normale Supérieure Paris]

Administrative Assistant

Christine Anocq [Inria]

2. Overall Objectives

2.1. Overall Objectives

The 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 (VÃ©lus, *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 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.

[0] http://www.esterel-technologies.com/products/scade-suite
[0] http://www.mathworks.com/products/simulink
[0] https://www.modelica.org

- 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

4.1.1. Awards

- Francesco Zappa Nardelli received the *Most Influential ICFP Paper Award* for 2007 paper “Ott: Effective Tool Support for the Working Semanticist” (http://www.sigplan.org/Awards/ICFP/).

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).

- 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. Compiler Optimisations for Multicore Architectures

Participants: Robin Morisset, Francesco Zappa Nardelli.

Robin has completed his research work on sound optimisations for modern multicore architectures. This covered optimisations that can be expressed inside the semantics of the C11/C++11 programming language, as well as optimisations that can be expressed only at the harware level. In particular we have shown how partial redundancy elimination (PRE) can be instantiated to perform *provably correct* fence elimination for multi-threaded programs running on top of the x86, ARM and IBM Power relaxed memory models. We have implemented our algorithm in the x86, ARM and Power backends of the `LLVM` compiler infrastructure. The optimisation does not induce an observable overhead at compile-time and can result in up-to 10% speedup on some benchmarks.

This work has been published in CC 2017 [10]. The implementation of the optimisations will be submitted for inclusion in the LLVM compiler suite.

Robin Morisset completed this line of research and defended his PhD Thesis in April 2017.

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. The work on subtyping is under submission to an international conference. This line of research will be pursued in the next year, studying method dispatch and type inference.

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. We have submitted this work to an international conference.

6.4. Symbolic Simulation for a timed-automaton subset of Zélus

Participants: Guillaume Baudart, Timothy Bourke, Marc Pouzet.

Synchronous languages like Lustre are ideal for programming an important class of embedded controllers. Their discrete model of time and deterministic semantics facilitate the precise expression of reactive behaviors. That said, many systems are naturally modeled using physical timing constraints that almost inevitably involve some 'timing nondeterminism' due to tolerances in requirements or uncertainties in implementations. Conversely, such constraints are readily modeled using Timed Automata, and simulated symbolically in Uppaal, but large-scale discrete-time behaviors are more cumbersome to express in such tools.

In this work, we combined existing techniques and data structures for Timed Safety Automata with typing and compilation techniques for synchronous languages to develop a novel programming language where discrete reactive logic can be mixed with nondeterministic continuous-time features. In particular, we developed an extension of Lustre and a specialization of Zélus for modeling real-time reactive systems, proposed a symbolic simulation scheme based on 'sweeping', and showed how to implement it via source-to-source compilation. A type system, based on that of Zélus, ensures the correct composition of discrete-time and continuous-time elements.

Our proposal has been implemented using the Zélus compiler and a small library of operations on Difference-Bound Matrices (DBMs). Unlike the work around Uppaal, we do not address verification or treat industrial case studies. A future direction could be to verify programs in our 'extended version of Lustre' by either generating C code and using the highly-tuned Uppaal DBM library, or combining symbolic techniques for Lustre programs with those for Timed Automata.

This work was presented at FDL 2017 [5]. A prototype implementation is available online.

This work is also described with extended examples in Baudart's PhD thesis [1] which was defended in March of 2017.

6.5. 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 modelling 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 set of benchmarks and evaluated the Worst Case Execution time of code generated by our compiler with that of code generated by the academic Heptagon and Lustre v6 compilers. This work also required numerous improvements to the parser and elaborator. We also tested the compiler on an industrial example in the context of the ASSUME project. We completed the end-to-end theorem showing that the dataflow semantics of input programs is preserved by the assembly language semantics generated by our compiler combined with the CompCert compiler. This work was presented in June at PLDI [8].

In the latter half to the year we worked on extending the compiler to accept nodes with clocked arguments, treating non-normalized Lustre, and adding a modular reset to the language.

To accept clocked arguments, we extended the semantic model, developed a richer encoding of the clock system, added a new invariant to forbid non-trivial sub-clocked expressions, and adapted the correctness proof. An unexpected complication was the need to pass undefined variables in function call arguments: this required changes to our intermediate Obc language and introduces minor technical difficulties in the translation to Clight which requires that variables be defined. This work is now almost complete.

To treat non-normalized Lustre, we introduced new syntactic and semantic definitions, updated the parser, and complete reworked the elaboration and type-checking passes. We developed many small Lustre programs to confirm our understanding of the language and test the updated front-end; this also revealed several bugs in other academic Lustre compilers. This work is now complete. The next step is to implement the normalization pass to connect the new front-end to the existing compilation passes.

The work on modular resets continues as part of L. Brun's PhD thesis. This year we developed a novel semantic model for modular resets and started considering how to generate provably correct code.

In collaboration with Pierre-Évariste Dagand (CNRS), Lionel Reig (Collège de France), and Xavier Leroy (Inria, GALLIUM team).

6.6. Zélus: Synchronous Languages + Ordinary Differential Equations

Participants: Timothy Bourke, Marc Pouzet.

Zélus is a synchronous language extended with Ordinary Differential Equations (ODEs) to model systems with complex interactions between discrete-time and continuous-time dynamics. It shares the basic principles of Lustre with features from Lucid Synchrone (type inference, hierarchical automata, and signals). The compiler is written in OCaml and is structured as a series of source-to-source and traceable transformations that ultimately yield statically scheduled sequential code. Continuous components are simulated using off-the-shelf numerical solvers (here Sundials CVODE) and, for the moment, two built-in solvers (ode23 and ode45).

Zélus is used to experiment with new techniques for building hybrid modelers like Simulink/Stateflow and Modelica on top of a synchronous language. The language exploits novel techniques for defining the semantics of hybrid modelers, it provides dedicated type systems to ensure the absence of discontinuities during integration and the generation of sequential code. In particular, all discrete computations must be aligned to zero-crossing events; programs with causality loops and uninitialized values are statically rejected.

This year we added arrays with iterators and statically expanded higher-order functions to the language. Both extensions required adapting the existing type and causality systems, and extending the compilation algorithms. These extensions allowed us to show that a fairly large set of blocks from the Simulink standard library can be programmed in a precise, purely functional language using stream equations, hierarchical automata, Ordinary Differential Equations (ODEs), and deterministic synchronous parallel composition. Although some blocks cannot be expressed as they mix discrete-time and continuous-time signals in unprincipled ways; they are statically rejected by the type checker. This work was presented at EMSOFT in October [9]

Our work on analyzing causality loops in hybrid systems modelers was published in the NAHS journal [2].

In collaboration with B. Caillaud and A. Benveniste (Inria Rennes); and F. Carcenac, B. Pagano, and C. Pasteur (ANSYS/Esterel Technologies).

6.7. Compiling synchronous languages for multi-processor implementations

Participants: Timothy Bourke, Albert Cohen, Guillaume Iooss, Marc Pouzet.

Working together with industrial partners in the context of the ASSUME project.

We spent a week in Toulouse working at Airbus on their use case and our front-end tools. We can now treat the case and generate code for Lopht (AOSTE team), which, in turn, generates executable code for the Kalray MPPA. We have also advanced significantly on two use cases provided by Safran. The first one is similar to the Airbus use case. The second one is more preliminary, it revealed the need for more general iterators to better express FFT algorithms.

We have made solid progress on a language extension for expressing and manipulating harmonic clocks. In particular, we derive a scheduling problem from the clock constraints in a program and we are working on automatically calculating their initial phases.

We have written an import tool that transforms graphs of dependencies between several Lustre components scheduled with different harmonic periods into a monolithic Lustre program. We are working on a hyper-scheduling transformation that generates a single step function running at the slowest period and that contains multiple instances of the faster tasks with annotations to ensure they execute at the correct time.

In collaboration (this year) with Dumitru Potop-Butucaru and Keryan Didier (Inria, AOSTE team); Jean Souyris and Adrien Gauffriau (Airbus); Philippe Baufreton et Jean-Marie Courtelle (Safran).

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).

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

ANR/CHIST-ERA DIVIDEND project, 2013-2018.

8.1.2. Investissements d'avenir

Sys2Soft contract (Briques Génériques du Logiciel Embarqué). Partenaire principal: Dassault-Systèmes, etc. 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. FP7 & H2020 Projects

8.2.1.1. Eurolab-4-HPC

Title: EuroLab-4-HPC: Foundations of a European Research Center of Excellence in High Performance Computing Systems

Programm: H2020

Duration: September 2015 - September 2017

Coordinator: CHALMERS TEKNISKA HOEGSKOLA AB

Inria contact: Albert Cohen

Europe has built momentum in becoming a leader in large parts of the HPC ecosystem. It has brought together technical and business stakeholders from application developers via system software to exascale systems. Despite such gains, excellence in high performance computing systems is often fragmented and opportunities for synergy missed. To compete internationally, Europe must bring together the best research groups to tackle the longterm challenges for HPC. These typically cut across layers, e.g., performance, energy efficiency and dependability, so excellence in research must target all the layers in the system stack. The EuroLab-4-HPC project's bold overall goal is to build connected and sustainable leadership in high-performance computing systems by bringing together the different and leading performance orientated communities in Europe, working across all layers of the system stack and, at the same time, fuelling new industries in HPC.

8.2.1.2. TETRACOM

Title: Technology Transfer in Computing Systems

Programm: FP7

Duration: September 2013 - August 2016

Coordinator: RHEINISCH-WESTFAELISCHE TECHNISCHE HOCHSCHULE AACHEN

Inria contact: Albert Cohen

The mission of the TETRACOM Coordination Action is to boost European academia-to-industry technology transfer (TT) in all domains of Computing Systems. While many other European and national initiatives focus on training of entrepreneurs and support for start-up companies, the key differentiator of TETRACOM is a novel instrument called Technology Transfer Project (TTP). TTPs help to lower the barrier for researchers to make the first steps towards commercialisation of their research results. TTPs are designed to provide incentives for TT at small to medium scale via partial funding of dedicated, well-defined, and short term academia-industry collaborations that bring concrete R&D results into industrial use. This will be implemented via competitive Expressions-of-Interest (EoI) calls for TTPs, whose coordination, prioritization, evaluation, and management are the major actions of TETRACOM. It is expected to fund up to 50 TTPs. The TTP activities will be complemented by Technology Transfer Infrastructures (TTIs) that provide training, service, and dissemination actions. These are designed to encourage a larger fraction of the R&D community to engage in TTPs, possibly even for the first time. Altogether, TETRACOM is conceived as the major pilot project of its kind in the area of Computing Systems, acting as a TT catalyst for the mutual benefit of academia and industry. The projects primary success metrics are the number and value of coordinated TTPs as well as the amount of newly introduced European TT actors. It is expected to acquire around more than 20 new contractors over the project duration. TETRACOM complements and actually precedes the use of existing financial instruments such as venture capital or business angels based funding.

8.2.1.3. EMC2

Title: Embedded Multi-Core Systems for Mixed Criticality Applications in Dynamic and Changeable Real-Time Environments

Programm: FP7

Duration: April 2014 - March 2017

Coordinator: Infineon Technologies

Inria contact: Albert Cohen

'Embedded systems are the key innovation driver to improve almost all mechatronic products with cheaper and even new functionalities. Furthermore, they strongly support today's information society as inter-system communication enabler. Consequently boundaries of application domains are alleviated and ad-hoc connections and interoperability play an increasing role. At the same time, multi-core and many-core computing platforms are becoming available on the market and provide a breakthrough for system (and application) integration. A major industrial challenge arises facing (cost) efficient integration of different applications with different levels of safety and security on a single computing platform in an open context. The objective of the EMC^2 project (Embedded multi-core systems for mixed criticality applications in dynamic and changeable real-time environments) is to foster these changes through an innovative and sustainable service-oriented architecture approach for mixed criticality applications in dynamic and changeable real-time environments. The EMC2 project focuses on the industrialization of European research outcomes and builds on the results of previous ARTEMIS, European and National projects. It provides the paradigm shift to a new and sustainable system architecture which is suitable to handle open dynamic systems. EMC^2 is part of the European Embedded Systems industry strategy to maintain its leading edge position by providing solutions for: . Dynamic Adaptability in Open Systems . Utilization of expensive system features only as Service-on-Demand in order to reduce the overall system cost. . Handling of mixed criticality applications under real-time conditions . Scalability and utmost flexibility . Full scale deployment and management of integrated tool chains, through the entire lifecycle Approved by ARTEMIS-JU on 12/12/2013 for EoN. Minor mistakes and typos corrected by the Coordinator, finally approved by ARTEMIS-JU on 24/01/2014. Amendment 1 changes approved by ECSEL-JU on 31/03/2015.'

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$).

8.4. International Research Visitors

8.4.1. Visits of International Scientists

8.4.1.1. Internships

Alex Susu from Polytechnica di Bucarest spent a 3 months internship in the Fall.

8.4.2. Visits to International Teams

8.4.2.1. Sabbatical programme

Francesco Zappa Nardelli, from Feb. 1st, 2017 to July. 29th, 2017 has been on sabbatical leave at Northeastern University, Boston, USA, invited by Prof. Jan Vitek.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. General Chair, Scientific Chair

- Albert Cohen was the General Chair of PLDI 2017.

9.1.2. Scientific Events Selection

9.1.2.1. Member of the Conference Program Committees

- Timothy Bourke was a member of the PC of EMSOFT 2017.
- Timothy Bourke was a member of the PC of the Modelica Conference 2017.
- Timothy Bourke was a member of the PC of SCOPES 2017.
- Timothy Bourke was a member of the Student Research Competition panel of PLDI 2017.
- Francesco Zappa Nardelli was a member of the PC of POPL 2017.
- Francesco Zappa Nardelli was a member of the PC of ECOOP 2017.
- Marc Pouzet was a member of the PC of SCOPES 2017, EOOLT 2017, FADL 2017.
- Albert Cohen was a PC member of CGO 2018, Supercomputing 2017, PACT 2017.
- Albert Cohen was the area co-chair for programming models at IPDPS 2018.

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

- Albert Cohen is associate editor of the ACM Transactions on Architecture and Code Optimization

9.1.3.2. Reviewer - Reviewing Activities

- Timothy Bourke was a reviewer for the Springer Real-Time Systems Journal.

- Timothy Bourke was a reviewer for IEEE Transactions on Software Engineering.

9.1.4. Invited Talks

- Timothy Bourke was invited to talk about the seL4 project at the Forum Méthodes Formelles: “Méthodes formelles et cyber-sécurité” in Toulouse in January 2017.
- Timothy Bourke was invited to speak on the “Verified Compilation of Lustre” at the University of Birmingham in March 2017.
- Francesco Zappa Nardelli is an invited speaker at the Entropy Workshop, January 2018.
- Francesco Zappa Nardelli was an invited speaker at Dagsthul Seminar 17502 “Testing and Verification of Compilers”, December 2017.
- Francesco Zappa Nardelli was an invited speaker at the ETH Workshop on Software Correctness and Reliability, October 2017.
- Marc Pouzet was an invited speaker of the GT OVSTR Digicosme (CEA - Telecom Paris), in April 2017; the GDR Glace (part of GPL), in June 2017.

9.1.5. Leadership within the Scientific Community

Albert Cohen is a steering committee member of the PLDI, PPoPP and Compiler Construction conferences.

9.1.6. Scientific Expertise

Albert Cohen has been a visiting scientist at Facebook Artificial Intelligence Research.

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

Licence: F. Zappa Nardelli: “Concurrent Programming" (L3), PCs, 32h, E´cole Polytechnique, France

Master : M. Pouzet & T. Bourke: “Synchronous Systems” (M2), Lectures and TDs, MPRI, France

Master: M. Pouzet : “Synchronous reactive Languages” (M2), Lectures. Master Comasic (Polytechnique), France.

Master: T. Bourke participated in reviewing the 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 : Guillaume Baudart, 3rd year, supervised by T. Bourke and M. Pouzet. This thesis was defended in March 2017.
- PhD in progress : Ulyssse Beaugnon, 3rd year, supervised by A. Cohen and M. Pouzet.
- PhD in progress : Lélio Brun, 2nd year, supervised by T. Bourke and M. Pouzet.
- PhD : Robin Morisset, 3rd year, supervised by F. Zappa Nardelli. This thesis was defended in April 2017.
- PhD in progress : Chandan Reddy, 3rd year, supervised by A. Cohen.
- PhD in progress : Jie Zhao, 3rd year, supervised by A. Cohen.

9.2.3. Juries

Francesco Zappa Nardelli was jury member of the PhD thesis of Yannick Zakowski, ENS Rennes, Dec 2017.

Francesco Zappa Nardelli will be jury member of the PhD thesis of Francois Ginraud, Grenoble, Jan 2018.

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] G. BAUDART.*A synchronous approach to quasi-periodic systems*, PSL Research University, March 2017, https://tel.archives-ouvertes.fr/tel-01507595.

Articles in International Peer-Reviewed Journal

[2] A. BENVENISTE, T. BOURKE, B. CAILLAUD, B. PAGANO, M. POUZET.*A Type-based Analysis of Causality Loops in Hybrid Systems Modelers*, in "Nonlinear Analysis: Hybrid Systems", November 2017, vol. 26, p. 168–189 [*DOI :* 10.1016/J.NAHS.2017.04.004], https://hal.inria.fr/hal-01549183.

[3] I. LLOPARD, C. FABRE, A. COHEN.*A From a Formalized Parallel Action Language to its Efficient Code Generation*, in "ACM Transactions on Embedded Computing Systems (TECS)", January 2017 [*DOI :* 10.1145/0000000.0000000], https://hal.inria.fr/hal-01425140.

Invited Conferences

[4] J.-L. COLAÇO, B. PAGANO, M. POUZET.*Scade 6: A Formal Language for Embedded Critical Software Development*, in "TASE 2017 - 11th International Symposium on Theoretical Aspects of Software Engineering", Nice, France, September 2017, p. 1-10, https://hal.inria.fr/hal-01666470.

International Conferences with Proceedings

[5] G. BAUDART, T. BOURKE, M. POUZET.*Symbolic Simulation of Dataflow Synchronous Programs with Timers*, in "12th Forum on Specification and Design Languages (FDL 2017)", Vérone, Italy, Electronic Chips & System Design Initiative (ECSI), September 2017, https://hal.inria.fr/hal-01575621.

[6] U. BEAUGNON, A. POUILLE, M. POUZET, J. PIENAAR, A. COHEN.*Optimization Space Pruning without Regrets*, in "CC 2017 - 26th International Conference on Compiler Construction", Austin, TX, United States, Proceedings of the International Conference on Compiler Construction, ACM Press, February 2017, p. 34-44 [*DOI :* 10.1145/3033019.3033023], https://hal.inria.fr/hal-01655602.

[7] A. BENVENISTE, B. CAILLAUD, H. ELMQVIST, K. GHORBAL, M. OTTER, M. POUZET.*Structural Analysis of Multi-Mode DAE Systems*, in "Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control, HSCC 2017", Pittsburgh, PA, United States, April 2017 [*DOI :* 10.1145/3049797.3049806], https://hal.inria.fr/hal-01521918.

[8] 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.

[9] 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.

[10] R. MORISSET, F. ZAPPA NARDELLI.*Partially Redundant Fence Elimination for x86, ARM and Power processors*, in "International Conference on Compiler Construction (CC)", Austin, United States, February 2017, https://hal.inria.fr/hal-01423612.

[11] J. ZHAO, A. COHEN.*A general compilation algorithm to parallelize and optimize counted loops with dynamic data-dependent bounds*, in "IMPACT 2017 - 7th International Workshop on Polyhedral Compilation Techniques", Stockholm, Sweden, January 2017, p. 1-10, https://hal.inria.fr/hal-01657608.

[12] R. VON HANXLEDEN, T. BOURKE, A. GIRAULT.*Real-Time Ticks for Synchronous Programming*, in "FDL 2017 - 12th Forum on Specification and Design Languages", Vérone, Italy, Electronic Chips & System Design Initiative (ECSI), September 2017, https://hal.inria.fr/hal-01575629.

National Conferences with Proceeding

[13] T. BOURKE, P.-E. DAGAND, M. POUZET, L. RIEG.*Vérification de la génération modulaire du code impératif pour Lustre*, in "JFLA 2017 - Vingt-huitième Journées Francophones des Langages Applicatifs", Gourette, France, January 2017, https://hal.inria.fr/hal-01403830.

Conferences without Proceedings

[14] A. SUSUNGI, A. COHEN, C. TADONKI.*More Data Locality for Static Control Programs on NUMA Architectures*, in "IMPACT 2017 - 7th International Workshop on Polyhedral Compilation Techniques IMPACT 2017", Stockholm, Sweden, January 2017, 11, https://hal-mines-paristech.archives-ouvertes.fr/hal-01529354.

Research Reports

[15] K. DIDIER, A. COHEN, A. GAUFFRIAU, A. GRAILLAT, D. POTOP-BUTUCARU.*Sheep in wolf's clothing: Implementation models for data-flow multi-threaded software*, Inria Paris, April 2017, n° RR-9057, 31, https://hal.inria.fr/hal-01509314.

[16] O. ZINENKO, S. VERDOOLAEGE, C. REDDY, J. SHIRAKO, T. GROSSER, V. SARKAR, A. COHEN.*Unified Polyhedral Modeling of Temporal and Spatial Locality*, Inria Paris, November 2017, n° RR-9110, 41, https://hal.inria.fr/hal-01628798.

Other Publications

[17] A. SUSUNGI, N. A. RINK, J. CASTRILLÓN, I. HUISMANN, A. COHEN, C. TADONKI, J. STILLER, J. FRÖHLICH.*Towards Compositional and Generative Tensor Optimizations*, October 2017, ACM SIGPLAN conference on Systems, Programming, Languages and Applications: Software for Humanity (SPLASH), Poster, https://hal-mines-paristech.archives-ouvertes.fr/hal-01666818.

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
A2.1.3. - 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. Personnel

Research Scientists
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
Thierry Coquand [University of Gothenburg, Professor, Inria International Chair]
Pierre Letouzey [Univ Paris Diderot (Paris 7), Associate Professor]
Yann Régis-Gianas [Univ Paris Diderot, Associate Professor]

External Collaborators
Arnaud Spiwack [Tweag I/0, engineer]
Samuel Mimram [Ecole Polytechnique, associate professor]
Claudia Faggian [IRIF, CNRS, Researcher]
Philippe Malbos [Université Lyon I, associate professor]
Nicolas Tabareau [Inria Nantes, Researcher]

Technical Staff
Daniel de Rauglaudre [Inria]
Thierry Martinez [Inria]

PhD Students
Amina Doumane [Univ Paris Diderot, until June 2017]
Étienne Miquey [Univ Paris Diderot, until Nov 2017]
Jovana Obradović [Univ Paris Diderot, until Aug 2017]
Maxime Lucas [Univ Paris Diderot, until Dec 2017]
Thibaut Girka [Mitsubishi]
Guillaume Claret [Univ Paris Diderot]

Cyprien Mangin [Univ Paris Diderot]
Théo Zimmermann [Univ Paris Diderot]
Cédric Ho Thanh [Univ Paris Diderot, since Sep 2017]

Post-Doctoral Fellows

Cyrille Chenavier [ATER Univ Paris Diderot, until Aug 2017]
Eric Finster [ERC CoqHoTT, from Jan 2017]
Kailiang Ji [Inria, from Dec 2017]
Exequiel Rivas Gadda [Inria, from Dec 2017]

Administrative Assistants

Lindsay Polienor [Inria]
Sandrine Vergès [Inria]

2. Overall Objectives

2.1. Overall Objectives

The research conducted in πr^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 [82] 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 [76], then by Howard and de Bruijn at the end of the 60's [94], [113], 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 λ-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 [73], [74] stemmed out – a formalism that is both a logic and a programming language and that is at the source of the Coq system [112].

3.1.2. Towards the calculus of constructions

The λ-calculus, defined by Church [71], 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 λ-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 λ-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 [61].

To explain the Curry-Howard correspondence, it is important to distinguish between intuitionistic and classical logic: following Brouwer at the beginning of the 20^{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) λ-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 [103].

In 1985, Coquand and Huet [73], [74] in the Formel team of Inria-Rocquencourt explored an alternative approach based on Girard-Reynolds' system F [83], [107]. 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 [75] 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 five 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 by employees of Inria, the CNAM and Paris 7.

In the last five 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 and wiki), made its development process public, starting to use public pull-requests to track the work of developers, organizing yearly hackatons/coding-sprints for the dissemination of expertise and developers & users meetings like the Coq Workshop and CoqPL, and, perhaps more annectodicaly, retransmitting Coq working groups on a public youtube channel.

This move was made possible thanks to the hiring of Maxime Dénès in 2016 as an Inria research engineer (in Sophia-Antipolis), and the work of Matej Košík (1-year research engineer) whose work involved making the development process more predictable, streamlined and to provide a higher level of quality to the whole system, while relieving researchers from some time-consuming software development tasks. Maxime Dénès is also working in collaboration with Yves Bertot to develop the Coq consortium, which aims to become the incarnation of the global Coq community and 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), Deducteam (Saclay) 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 hosts 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 version Coq 8.7.1 had 46 contributors (counted from the start of 8.7 development), while 8.6 had 38 contributors.

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 [63], the more advanced Certified Programming with Dependent Types book by Chlipala [70] 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 now 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 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 [84] 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 [100] and Reynolds [106] and started to be studied in an abstract way in the 80's by Felleisen *et al* [80], leading to Parigot's $\lambda\mu$-calculus [104], 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 λ-calculus is Landin's SECD machine [99] and Krivine's abstract machine for call-by-name evaluation [96], [95]. 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 [81]: any side-effect expressible in monadic style (and this covers references, exceptions, states, dynamic bindings, ...) can be simulated in λ-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, ∞-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* ∞*-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 [60]) 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 [78]).

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 [93], it has been observed recently that the so-called "identity types" are naturally equiped with a structure of ∞-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 ressemblance with the fundamental ∞-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 [111], and independently by Burroni [69], 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 [97], [102], [85], [86], and beyond: Petri nets [88] and formal proofs of classical and linear logic have been expressed in this framework [87]. 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 [64].

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 [58] and Squier [109], [108] 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 [12], [13], [90], [91], [92].

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 [12] and algebras [47]. 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 [98], [59]). 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 [89][47].

4. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Awards

Amina Doumane was awarded the The Kleene Award for Best Student Paper at the LICS 2017 conference, for her work on "Constructive Completeness for the Linear-Time mu-Calculus". She also received in January 2018 the prize of the Journal La Recherche for the same paper.

Amina Doumane was awarded the Gilles Kahn 2017 prize for her PhD thesis entitled "On the infinitary proof theory of logics with fixed points" supervised by Alexis Saurin, David Baelde and Pierre-Louis Curien.

Ludovic Patey was awarded the Prix Thiessé de Rosemont / Demassieux 2017 for his PhD thesis "Les mathématiques à rebours de théorèmes de type Ramsey", supervised by Laurent Bienvenu and Hugo Herbelin.
BEST PAPERS AWARDS :
[37] **Conference on Logic in Computer Science 2017**. A. DOUMANE.
[] **On the infinitary proof theory of logics with fixed points**.

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: Version 8.7 features a large amount of work on cleaning and speeding up the code base, notably the work of Pierre-Marie Pédrot on making the tactic-level system insensitive to existential variable expansion, providing a safer API to plugin writers and making the code more robust.

New tactics: Variants of tactics supporting existential variables "eassert", "eenough", etc. by Hugo Herbelin. Tactics "extensionality in H" and "inversion_sigma" by Jason Gross, "specialize with" accepting partial bindings by Pierre Courtieu.

Cumulative Polymorphic Inductive Types, allowing cumulativity of universes to go through applied inductive types, by Amin Timany and Matthieu Sozeau.

The SSReflect plugin by Georges Gonthier, Assia Mahboubi and Enrico Tassi was integrated (with its documentation in the reference manual) by Maxime Dénès, Assia Mahboubi and Enrico Tassi.

The "coq_makefile" tool was completely redesigned to improve its maintainability and the extensibility of generated Makefiles, and to make "_CoqProject" files more palatable to IDEs by Enrico Tassi.

A lot of other changes are described in the CHANGES file.
NEWS OF THE YEAR: Version 8.7 was released in October 2017 and version 8.7.1 in December 2017, development started in January 2017. This is the second release of Coq developed on a time-based development cycle. Its development spanned 9 months from the release of Coq 8.6 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 scientific advance in this version is the integration of cumulative inductive types in the system. More practical advances in stability, performance, usability and expressivity of tactics were also implemented, resulting in a mostly backwards-compatible but appreciably faster and more robust release. Much work on plugin extensions to Coq by the same development team has also been going on in parallel, including work on JSCoq by Emilio JG Arias, Ltac 2 by P.M-Pédrot, which required synchronised changes of the main codebase. In 2017, the construction of the Coq Consortium by Yves Bertot and Maxime Dénès has greatly advanced and is now nearing its completion.

- 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 8.7.1
- 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/

6. New Results

6.1. Effects in proof theory and programming

Participants: Hugo Herbelin, Étienne Miquey, Yann Régis-Gianas, Alexis Saurin.

6.1.1. A classical sequent calculus with dependent types

Dependent types are a key feature of type systems, typically used in the context of both richly-typed programming languages and proof assistants. Control operators, which are connected with classical logic along the proof-as-program correspondence, are known to misbehave in the presence of dependent types [14], unless dependencies are restricted to values. As a step in his work to develop a sequent-calculus version of Hugo Herbelin's dPA_ω system [16], Étienne Miquey proposed a sequent calculus with classical logic and dependent types. His calculus—named dL—is an extension of the $\mu\tilde{\mu}$-calculus with a syntactical restriction of dependent types to the fragment of *negative-elimination free* proofs. The corresponding type system includes a list of explicit dependencies, which maintains type safety. He showed that a continuation-passing style translation can be derived by adding delimited continuations, and how a chain of dependencies can be related to a manipulation of the return type of these continuations. This work has been presented at ESOP 2017 [39].

6.1.2. Normalisation and realisability interpretation of call-by-need with control

The call-by-need evaluation strategy is an evaluation strategy of the λ-calculus which evaluates arguments of functions only when needed, and, when needed, shares their evaluations across all places where the argument is needed. The call- by-need evaluation is for instance at the heart of a functional programming language such as Haskell. A continuation-passing-style semantics for call-by-need, de facto giving a semantics to control operators, was proposed in the 90s by Okasaki, Lee and Tarditi. However, this semantics does not ensure normalisation of simply-typed call-by-need evaluation, thus failing to ensure a property which holds in the simply-typed call-by-name and call-by-value cases. Étienne Miquey and Hugo Herbelin have been considering a call-by-need λ-calculus due to Ariola et al. for which they proved the normalisation by means of a realisability interpretation. Incidentally, the variant of realisability they proposed allows to define realisers as pairs of a term and a substitution. This paves the way to give interpretation of calculus with global and mutable memory. This work has been accepted for publication at the FOSSACS 2018 conference.

6.1.3. A sequent calculus with dependent types for classical arithmetic

In 2012, Hugo Herbelin showed that classical arithmetic in finite types extended with strong elimination of existential quantification proves the axiom of dependent choice. Getting classical logic and choice together without being inconsistent is made possible by: (1) constraining strong elimination of existential quantification to proofs that are essentially intuitionistic; (2) turning countable universal quantification into an infinite conjunction of classical proofs, which are evaluated along a call-by-need evaluation strategy, so as to extract from them intuitionistic contents that complies to the intuitionistic constraint put on strong elimination of existential quantification.

Relying on its sequent calculus with dependent types and its realisability interpretation for call-by-need with control, Étienne Miquey proposed in his thesis a sequent calculus with the same computational features [24]. His calculus therefore also allows for the direct definition of proof terms for the axioms of countable and dependent choices. The proofs of normalisation and soundness are made through a realisability interpretation of the calculus, which is obtained by using Danvy's methodology of semantics artifacts.

6.1.4. Reverse mathematics of Gödel's completeness theorem

Charlotte Barot, under the supervision of Hugo Herbelin, studied the relative intuitionistic strength of Gödel's completeness theorem, the ultrafilter lemma, and different forms of the Fan Theorem, as a way to transfer computational contents of proofs from one to the other theorems.

6.1.5. A theory of effects and resources

Arnaud Spiwack, in collaboration with Jean-Philippe Bernardy, Mathieu Boespflug, Ryan R. Newton and Simon Peyton-Jones, developed an extension of the type system of Haskell with linear types The work is to be presented at POPL'18.

In collaboration with Thomas Letan (Agence Nationale pour la Sécurité des Systèmes Informatiques), Yann Régis-Gianas studied how free monads can be used to develop modular implementations and proofs of effectful systems. This proof technique is applied to the formal study of architectural attacks on IBM PC like architectures.

6.1.6. Classical realisability and implicative algebras

Étienne Miquey has been working with Alexandre Miquel in Montevideo on the topic of implicative algebras. Implicative algebras are an algebraisation of the structure needed to develop a realisability model. In particular, they give rise to the usual ordered combinatory algebras and thus to the triposes used to model classical realisability. An implicative algebra is given by an implicative structure (which consists of a complete semi-lattice with a binary operation $\rightarrow$) together with a separator containing the element interpreted as true in the structure. Following the work of Guillaume Munch-Maccagnoni on focalisation and classical realisability, Étienne Miquey gave alternative presentations within structures based on other connectives rather than $\rightarrow$, namely disjunctive algebras (based on negation, "par") and conjunctive algebras (negation, tensor). Such connectives correspond to the decomposition of the arrow according to the strategy of evaluation (call-by-name/call-by-value). In particular, he showed that disjunctive algebras were particular cases of implicative algebras; and that conjunctive algebras can be obtained by duality from disjunctive algebras. Besides, Étienne Miquey has formalised the theory of implicative algebras (resp. disjunctive, conjunctive) in Coq.

6.2. Reasoning and programming with infinite data

Participants: Amina Doumane, Yann Régis-Gianas, Alexis Saurin.

This theme is part of the ANR project Rapido (see the National Initiatives section).

6.2.1. Proof theory of infinitary and circular proofs

In collaboration with David Baelde and Guilhem Jaber, Amina Doumane and Alexis Saurin extended the proof theory of infinite proofs for fixpoint logics by relaxing the validity condition necessary to distinguish sound proofs from invalid ones. In CSL 2016, Baelde, Doumane and Saurin proved cut-elimination and focalisation for infinite proofs for $\mu MALL$ with a validity condition inspired from the acceptance condition of parity automata (or the winning condition of parity games). However, this validity condition rules out lots of proofs which are computational sound and does not account for the cut-axiom interaction in sequent proofs.

With Jaber, they relaxed the 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, approximating productivity. If the decidability of the validity condition in the most general case is still open, it allows for decidable restrictions which are still useful in the sense they allow for a much more flexible writing of circular proofs (or, through the proofs-as-programs bridge, circular progams). Cut-elimination is obtained in two steps, combining CSL 2016 result with a technique for "straightening" bouncing threads, that is perfoming just the necessary amount of cut-elimination to recover straight threads, the two results are combined thanks to a compression lemma, a standard result from infinitary rewriting ensuring that a transfinite strongly converging sequence can be turned into an ω-indexed strongly converging sequence. Preliminary results were presented at the Types 2017 conference.

6.2.2. Automata theory meets proof theory: completeness of the linear time mu-calculus.

Amina Doumane extended her previous results with David Baelde, Lucca Hirschi and Alexis Saurin proving a constructive completeness theorem for the full linear-time μ-calculus, while the previous results only captured a fragment of the linear-time mu-calculus expressing all inclusions of Büchi automata suitably encoded as formulas.

In order to achieve this tour de force (for which her publication at LICS 2017 received the Kleene award of the best student paper [37], see Highlights of the year), she identified several fragments of the linear-time mu-calculus corresponding to various classes of ω-automata and proved completeness of those classes by using circular proof systems and finitisation of the infinite proofs in the Kozen's usual axiomatisation (see paragraph on finitising circular proofs for more details).

6.2.3. 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 μ and ν statements.

In the setting of fixpoint logics with circular proofs, several investigations were carried on in the team:

- firstly, in the setting of the usual validity condition for circular proofs of $\mu MALL$, Doumane extended in her PhD thesis a translatibility criterion for finitising circular proofs which was first used in joint work with Baelde, Saurin and Hirschi and later applied to the full linear-time mu-calculus in her LICS 2017 paper. Her translatibility criterion abstracts the proof scheme for finitising circular proofs and is not formulated with respect to a specific fragment of the logic, but with respect to conditions allowing finitisation of the cycles.
- Secondly, Nollet, working with Saurin and Tasson, recently proposed 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. 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.

6.2.4. Co-patterns

In collaboration with Paul Laforgue (Master 2, University Paris 7), Yann Régis-Gianas developed an extension of OCaml with copatterns. Copatterns generalize standard ML patterns for algebraic datatypes: While a pattern-matching destructs a finite value defined using a constructor, a copattern-matching creates an infinite computation defined in terms of its answers to observations performed by the evaluation context. They exploits the duality between functions defined by pattern matching and functions that define codata by copattern-matching, going from the second to the first by introducing a well-typed inversion of control which is a purely local syntactic transformation. This result shows that copattern-matching can be added with no effort to any programming language equipped with second-order polymorphism and generalized algebraic datatypes. This work has been published in the proceeding of PPDP'17. A short paper has also been accepted at JFLA'18.

6.2.5. Streams, classical logic and the ordinal λ-calculus

Polonsky and Saurin defined an extension of infinitary λ-calculi allowing transfinite iteration of abstraction and ordinal sequences of applications, Λ^o, and established a standardisation theorem for this calculus. The $\Lambda\mu$-calculus can be embedded in this calculus, as well as Saurin's full Stream hierarchy: as a consequence, they obtain a uniform framework to investigate this family of calculi and provide uniform proofs of important results such a standardisation.

6.2.6. Theory of fixpoints in the lambda-calculus

In collaboration with Manzonetto, Polonsky and Simonsen, Saurin studied two long-standing conjectures on fixpoints in the λ-calculus: the "fixpoint property" and the "double-fixpoint conjecture". The former asserts that every λ-term admits either a unique or an infinite number of β-distinct fixpoints while the second,

formulated by Statman, says that there is no fixpoint satisfying $Y\delta = Y$ for $\delta = \lambda y, x.x(yx)$. They proved the first conjecture in the case of open terms and refute it in the case of sensible theories (instead of β). Moreover, they provide sufficient conditions for both conjectures in the general case. Concerning the double-fixpoint conjecture, they propose a proof technique identifying two key properties from which the results would follow, while they leave as conjecture to prove that those actually hold. Those results are currently submitted to a journal [53].

6.3. Effective higher-dimensional algebra

Participants: Cyrille Chenavier, Pierre-Louis Curien, Yves Guiraud, Cédric Ho Thanh, Maxime Lucas, Philippe Malbos, Samuel Mimram, Jovana Obradović, Matthieu Sozeau.

6.3.1. Higher linear rewriting

Yves Guiraud and Philippe Malbos have completed a four-year long collaboration with Eric Hoffbeck (LAGA, Univ. Paris 13), 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 [69] 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 [68] and Poincaré-Birkhoff-Witth bases [105]. Then, they transposed the construction of [12], based on an extension of Squier's theorem [108] 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 has just been submitted for publication [47].

Cyrille Chenavier has continued his work on reduction operators, a functional point of view on rewriting in associative algebras initiated by Berger [62], on which his PhD thesis was focused [4]. First, using the lattice structure of the reduction operators, he gave a new algebraic characterisation of confluence, and developed a new algorithm for completion, based on an iterated use of the meet-operation of the lattice [28]. Then he related this completion procedure to Faugère's F4 completion procedure for noncommutative Gröbner bases [79]. Finally, he gave a construction of a linear basis of the space of syzygies of a set of reduction operations, and used this work to optimise his completion procedure [45].

6.3.2. Cubical higher algebra

Maxime Lucas, supervised by Yves Guiraud and Pierre-Louis Curien, has applied the rewriting techniques of Guiraud and Malbos [92] to prove coherence theorems for bicategories and pseudofunctors. He obtained a coherence theorem for pseudonatural transformations thanks to a new theoretical result, improving on the former techniques, that relates the properties of rewriting in 1- and 2-categories [31]. Then he has transposed to a cubical setting, and improved, the results of [12]. This first involved a deep foundational work on the connections between globular and cubical higher categories [51], generalising several already known links in a unique theoretical setting [66], [67], [57], [110]. Then, he could prove Squier's theorem, giving a construction of a polygraphic resolution of monoids in the category of cubical Gray monoids [50]. All these results are contained in his PhD thesis, that was successfully defended in December 2017 [23].

6.3.3. Coherent Presentations of Monoidal categories

Presentations of categories are a well-known algebraic tool to provide descriptions of categories by means of generators, for objects and morphisms, and relations on morphisms. Pierre-Louis Curien and Samuel Mimram have generalised this notion, in order to consider situations where the objects are considered modulo an equivalence relation, which is described by equational generators. When those form a convergent (abstract) rewriting system on objects, there are three very natural constructions that can be used to define the category which is described by the presentation: one consists in turning equational generators into identities (i.e. considering a quotient category), one consists in formally adding inverses to equational generators (i.e. localising the category), and one consists in restricting to objects which are normal forms. Under suitable coherence conditions on the presentation, the three constructions coincide, thus generalising celebrated results on presentations of groups. Those conditions are then extended to presentations of monoidal categories [29].

6.3.4. Categorified cyclic operads

The work of Pierre-Louis Curien and Jovana Obradović on categorified cyclic operads has been condtionally accepted in the Journal Applied Categorical Structures [46]. The revision will include a careful treatment of weakened identity laws, as well of weakened equivariance laws. It will also include the details of an example and an illustration of the work. The example involves a generalisation of profunctors, and the application is to the notion of anti-cyclic operad, which they prove to be "sign-coherent".

6.3.5. Syntactic aspects of hypergraph polytopes

In collaboration with Jelena Ivanović, Pierre-Louis Curien and Jovana Obradović have introduced an inductively defined tree notation for all the faces of polytopes arising from a simplex by truncations, that allows them to view inclusion of faces as the process of contracting tree edges. This notation instantiates to the well-known notations for the faces of associahedra and permutohedra. Various authors have independently introduced combinatorial tools for describing such polytopes. In this work, the authors build on the particular approach developed by Došen and Petrić, who used the formalism of hypergraphs to describe the interval of polytopes from the simplex to the permutohedron. This interval was further stretched by Petrić to allow truncations of faces that are themselves obtained by truncations, and iteratively so. The notation applies to all these polytopes, and this fact is illustrated by showing that it instantiates to a notation for the faces of the permutohedron-based associahedra, that consists of parenthesised words with holes. In their work, Pierre-Louis Curien, Jovana Obradović and Jelena Ivanović also explore links between polytopes and categorified operads, as a follow-up of another work of Došen and Petrić, who had exhibited some families of hypergraph polytopes (associahedra, permutohedra, and hemiassociahedra) describing the coherences, and the coherences between coherences etc., arising by weakening sequential and parallel associativity of operadic composition. Their work is complemented with a criterion allowing to recover the information whether edges of these "operadic polytopes" come from sequential, or from parallel associativity. Alternative proofs for some of the original results of Došen and Petrić are also given. A paper containing this material has been accepted in the Journal Homotopy and Related Structure [32].

6.3.6. Opetopes

Opetopes are a formalisation of higher many-to-one operations leading to one of the approaches for defining weak ω-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 has developped a corresponding syntax, which he presented at the workshop "Categories for homotopy and rewriting" (CIRM, September 2017).

Cédric Ho Thanh started his PhD work around opetopes in September 2017. His first contributions include a careful embedding of opetopic sets into polygraphs, and a (finite) critical pair lemma for opetopic sets. Indeed, opetopic sets seem to delimit a subset of polygraphs in which the basics of rewriting theory can be developped, without the anomalies already observed by Lafont and others happening, like the existence of a possibly infinite set of critical pairs in a rewriting system specified by finitely many rules. Opetopes are tree-like and hence first-order-term-like and that is the intuitive reason why these anomalies are avoided.

6.3.7. Higher Garside theory

Building on [9], Yves Guiraud is currently finishing with Matthieu Picantin (IRIF, Univ. Paris 7) a work that generalises already known constructions such as the bar resolution, several resolutions defined by Dehornoy and Lafont [77], 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.

6.3.8. Foundations and formalisation of higher algebra

Yves Guiraud has started a collaboration with Marcelo Fiore (Univ. Cambridge) on the foundations of higher-dimensional categories, with the aim to define a general notion of polygraphs for various notions of algebraic structures. This is based on seeing higher categories as n-oids in a specific n-oidal category (a category with n monoidal structures with exchange morphisms between them). With that point of view, a good notion of polygraph can be iteratively defined for monoids in any monoidal category with pullbacks, which is a sufficiently general setting for most purposes.

Eric Finster, Yves Guiraud and Matthieu Sozeau have started to explore the links between combinatorial higher algebra and homotopy type theory, two domains that describe computations with a homotopical point of view. Their first goal is to formalise the rewriting methods of [12] and [10] in homotopy type theory, establishing a first deep connection between the two fields. This direction will be explored further by Antoine Allioux, a PhD student co-directed by Guiraud and Sozeau, starting in February 2018.

6.4. Incrementality

Participants: Thibaut Girka, Yann Régis-Gianas, Kostia Chardonnet.

In collaboration with Colin Gonzalez, Yann Régis-Gianas developed BLACS, a programming framework that applies differential functional programming techniques to the implementation of asynchronous spreadsheets for big data.

In collaboration with Lelio Brun (ENS), Yann Régis-Gianas developed DeltaCoq, a library for certified incremental functional programming. A paper is in preparation.

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 will be submitted to ICFP 2018.

6.4.2. Difference languages

In collaboration with David Mentré (Mitsubishi), Thibaut Girka and Yann Régis-Gianas developed a metatheoretical framework to develop verifiable difference languages in Coq. Such formal differences capture semantic differences between close programs. This work appeared in PPDP'17 [38].

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, Cyprien Mangin, Théo Zimmermann.

6.5.1. Homotopy type theory

Hugo Herbelin worked on the computational contents of extensional equality in type theory. Exploiting the idea introduced in Cohen, Coquand, Huber and Mörtberg's Cubical Type Theory of equality as abstraction over a geometrical dimension, he developed a direct-style system of notations for a scoped iterated parametricity semantics. The resulting logic respects equivalence of types by construction, thus providing a simple computational content to the key axiom of Homotopy Type Theory, namely the axiom of univalence.

6.5.2. Proof irrelevance and Homotopy Type Theory

Gaëtan Gilbert (PhD student of N. Tabareau, Gallinette and M. Sozeau, started in 2016) is studying the integration of a new notion of propositions, called *strict* propositions, in the calculus of inductive constructions. This new sort dubbed sProp supports definitional proof-irrelevance (two proofs of a strict proposition are always convertible), while maintaining compatibility with Univalence or Uniqueness of Identity Proofs. The goal of this work is to provide a more comfortable programming experience in the system by allowing more proofs to be identified "for free" during conversion. This should have an impact both on programming with dependent types (avoiding issues with coercions during typechecking) and for the development of homotopy type theory (avoiding "trivial" transports of equality proofs on natural numbers for example). Gaëtan Gilbert has developed a prototype version integrating this extension in Coq.

6.5.3. Extensionality and Intensionality in Type Theory

Théo Winterhalter (internship co-advised by Matthieu Sozeau and Andrej Bauer in 2017, now PhD student at Inria Nantes, co-advised by Nicolas Tabareau and Matthieu Sozeau) studied a translation from extensional to intensional type theory during his internship with Matthieu Sozeau and a general framework for formalising variants of type theory previously with Andrej Bauer at the University of Ljubljana in Slovenia. They developed a revised version of the translation by Nicolas Oury which doesn't require the use of John Major equality nor suspicious axioms associated to it. It results in a mixed translation that can transport derivations of extensional type theory into intensional 2-level type theory (with an original, syntactic presentation of the latter). This allows in principle to use the convenience of the reflection rule of equality in proofs while being able to derive decorated terms checkable by the kernel of a 2-level variant of Coq: one where both a univalent equality and a strict equality with uniqueness of identity proofs can cohabit. They are working on a Coq formalisation of this result using the Template-Coq framework, which will be extracted to a translation plugin to provide this facility in Coq itself.

6.5.4. Dependent pattern-matching

Cyprien Mangin developed a new simplification engine on top of the Equations plugin. This simplification engine is similar to the one of Cockx [72], allowing an interpretation of dependent pattern-matching that is independent of axioms like UIP or Univalence. While refining the implementation, he also designed a few optimisations allowing for a smarter compilation scheme, in terms of the required properties of the objects and the size of the generated proofs. Matthieu Sozeau concentrated on making the treatment of recursive functions more robust and complete, leading to the first tool of this kind for Coq that can handle both mutual and nested structurally recursive functions along with nested well-founded definitions. The elimination principle generation part of the system was adapted accordingly, putting the tool in good position to replace the previous Function tool of Coq that supports neither dependent pattern-matching nor nested fixpoints. Matthieu Sozeau developed a number of examples showcasing the tool, the largest one having actually been first developed by a student of the MPRI 2.7.2 course. An article presenting this tool and the smart case analysis method is in revision [52]. Version 1.0 of the system was released in December 2017. Cyprien Mangin gave a demo / presentation of the tool at the EUTypes Type Theory Tools workshop in January 2017 and will present a poster and demonstration of the new version at PEPM 2018 in Los Angeles.

Thierry Martinez started 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. Transferring theorems along isomorphisms

Following his work on theorem transfer along (iso)morphisms, Théo Zimmermann has started to explore more fundamental aspects that are connected to it: the concept of logical relation, which was originally invented to prove behavioral equivalence of programs and served to formalise parametricity, seems, following Hermida, Reddy and Robinson, to correspond to a very generic relational notion of morphism that was precisely the one needed for transfer lemmas.

6.5.6. Unification

Matthieu Sozeau has developped a complete reimplementation of the basic tactics of Coq in terms of the type-inference unification algorithm of Coq. This work is scheduled to be integrated in part in the 8.8 version of Coq due next year. It should provide a clean slate for development of the 9 series of Coq relying solely on an algorithm close to the one studied with Beta Ziliani in [22].

6.5.7. Cumulativity for Inductive Types

Together with Amin Timany (PhD student of Bart Jacobs at KU Leuven), Matthieu Sozeau developed an extension of the Calculus of Inductive Constructions featuring cumulativity for inductive types [43]. This extension is useful for developments using universe polymorphism like Category Theory formalisations and the HoTT library [35] but also crucial to develop syntactic program translations that add structures to types, as advocated by Boulier et al [65], requiring to validate the cumulativity rule on sigma types. They showed the relative consistency of this extension of the calculus using a set-theoretic model, inspired by the one of Lee and Werner [101] for proof-irrelevance. This extension is integrated in the 8.7 release of Coq and involved a large amount of design and implementation work in particular in relation with the unification strategy used in presence of subtyping and delta reduction, extending the framework studied in [33]. An article describing this work is in revision.

6.6. Formalisation work

Participants: Jean-Jacques Lévy, Daniel de Rauglaudre.

6.6.1. Proofs of algorithms on graphs

Jean-Jacques Lévy and Chen Ran (a PhD student of the Institute of Software, Beijing, visiting the Toccata team 9 months until April 2017) pursue their work about formal proofs of algorithms. Their goal is to provide proofs of algorithms which ought to be both checked by computer and easily 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. There are mainly two algorithms: one by Kosaraju (1978) working in two phases (some formal proofs of it have already been achieved by Pottier with Coq and by Théry and Gonthier with Coq-SSReflect), one by Tarjan (1972) working in a single pass.

Their proofs use a first-order logic with definitions of inductive predicates. This logic is the one defined in the Why3 system (research-team Toccata, Saclay). They widely use automatic provers interfaced with Why3. A minor part of these proofs is also achieved in Coq. The difficulty of this approach is to combine automatic provers and intuitive design.

In 2017, the same proofs were fully completed in Coq-ssreflect by Cohen and Théry, and in Isabelle-HOL by Merz, both proofs with the assistance of J.-J. Lévy. A Fstar proof is also under development. These proofs are between a factor 4 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.

Part of this work (Tarjan 1972) was presented at JFLA 2017 in Gourette [40]. A more comprehensive version was presented at the VSTTE 2017 conference in Heidelberg [36]. Scripts of proofs can be found at http://jeanjacqueslevy.net/why3.

6.6.2. Banach-Tarski Paradox

Banach-Tarski Paradox states that, if we admit the axiom of choice, a sphere is equidecomposable into two spheres identical to the initial one. The equidecomposability is a property of geometric objects: two objects (sets) are equidecomposable if they can be partitioned into a same finite number of sets, and each set of the first object is mapped to a set of the second object by only rotations and translations. In other words, one breaks the first object into a finite number of pieces, and with them, one reconstructs the second object. Its pen and paper proof was done in 1924 by Banach and Tarski.

The formal proof was completed this year by Daniel de Rauglaudre, after 9 months, with a result of about 10000 lines of Coq. A paper about it was published in JFR (Journal of Formalized Reasoning) [34].

6.6.3. Univalence for Free

Together with E. Tanter at Inria Chile and N. Tabareau at Inria Nantes, Matthieu Sozeau developed the theory and implementation of an ad-hoc version of univalence. This axiom at the basis of Homotopy Type Theory morally says that all constructions of type theory are invariant under equivalence, which for programming purposes means invariance by isomorphism. Using a carefuly designed variant of the parametriciy translation for type theory, they can show that indeed all type constructors of type theory, except indexed inductive types with non-hset indices respect univalence. In practice, this leads to a type-class based framework for constructing the proofs that values of a given type do indeed transport equivalences/isomorphisms correctly, relying on univalence itself only for universes and in well-delimited places. An article about this work is in revision [56].

6.6.4. 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 for the Coq programming language down to CompCert C-light which provides itself a certified compilation path to assembly language. The compiler can already be run and most phases are proven correct. As part of this work, Matthieu Sozeau took the lead of the Template-Coq library development originally developed by Gregory Malecha and extended it. Template-Coq provides quoting and unquoting facilities for Coq's kernel syntax and environment to Coq, allowing to reason on the actual definitions checked by the Coq system in Coq itself. For CertiCoq, the quoted type of Coq terms corresponds to its frontend language. The plugin can however be used in many other ways, notably to implement certified syntactic translations from Coq (or extended theories) to Coq, and to develop plugins to the Coq system in Coq itself. Together with Nicolas Tabareau and Simon Boulier in Nantes and Abhishek Anand at Cornell University, they are developing a general plugin for certified meta-programming in the system. It will be presented at CoqPL'18 [41]. Matthieu Sozeau worked in particular on reimplementing the basic typing and conversion algorithms of Coq inside Coq itself, providing a mechanised specification of the implementation of the system that can be used to verify arbitrarily large parts of it. The type inference algorithm developed there is also useful to help writting program translations on the "forgetful" kernel syntax.

7. Partnerships and Cooperations

7.1. National Initiatives

Alexis Saurin (coordinator) and Yann Régis-Gianas 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.

Pierre-Louis Curien (coordinator), Yves Guiraud (local coordinator), Philippe Malbos and Samuel Mimram are members of the four-year Cathre ANR project (January 2014 to December 2017). This project investigates the general theory of higher-dimensional rewriting, the development of a general-purpose library for higher-dimensional rewriting, and applications in the fields of combinatorial linear algebra, combinatorial group theory and theoretical computer science. This project is joint with mathematicians and computer scientists from LAGA (Univ. Paris 13), LIX (École Polytechnique), ICJ (Univ. Lyon 1 and Univ. Saint-Étienne), I2M (Univ. Aix-Marseille) and IMT (Univ. Toulouse 3). The project Cathre provided the funding for the PhD of Maxime Lucas.

Pierre-Louis Curien, Yves Guiraud, Hugo Herbelin, Philippe Malbos, Samuel Mimram and Alexis Saurin are members of the GDR Informatique Mathématique, in the Géocal (Geometry of computation) and LAC (Logic, algebra and computation) working groups.

Pierre-Louis Curien, Yves Guiraud (local coordinator), Philippe Malbos, Samuel Mimram 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.

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 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).

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. Inria International Labs

7.3.1.1. Other 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 will officially start 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 Hasuoi (NII, Tokyo). The full name of the project is Concurrent, Resourceful and Effectful 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 a 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 Inria-CAS project (a new programme of Inria with the 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, starting in December 2017.

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, and Universidad Nacional de Córdoba.

7.3.4. Participation in Other International Programs

Pierre-Louis Curien participates to the ANR International French-Chinese project LOCALI (Logical Approach to Novel Computational Paradigms), coordinated by Gilles Dowek (Deducteam). This project ended in July 2017.

7.4. International Research Visitors

7.4.1. Visits of International Scientists

John Baez (University of California River Side) visited the team for a week in November 2017.

Marcelo Fiore (University of Cambridge) visited the team for two weeks in February 2017.

Jovana Obradović (now a postdoc at Charles University, Prague) visited the team from December 1 to December 10 2017.

Amin Timany (KU Leuven, Belgium) visited the team for two months in March-April 2017 and collaborated with Matthieu Sozeau on the design and implementation of cumulative inductive types in Coq.

7.4.2. Visits to International Teams

7.4.2.1. Research Stays Abroad

Pierre-Louis Curien visited East China Normal University for a month in June 2017 (collaborations with Yuxin Deng and Min Zhang). Pierre-Louis Curien and Jovana Obradović visited the Institute of Mathematics of the Serbian Academy of Sciences in Belgrade in July 2017 (collaboration with Zoran Petrić).

Jean-Jacques Lévy visited the Institute of Software of Chinese Academy of Sciences (ISCAS) in December 2017 (project VIP and on-going work with Ran Chen) during 2 weeks. He gave talks at ISCAS hosted by Ying Jiang, and during a third week at ECNU Shanghai hosted by Min Zhang, USTC Suzhou (University of Science and Technology of China) hosted by Xinyu Feng, Nankai University in Tianjin hosted by Chunfu Jia.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events Organisation

8.1.1.1. General Chair, Scientific Chair

Pierre-Louis Curien is organising a Day of Hommage to the memory of Maurice Nivat on Fevruary 6, 2018, at University Paris 7.

8.1.1.2. Member of the Organising Committees

Yann Régis-Gianas was multimedia chair of the organising committee of POPL 2017 that took place in Paris in January 2017.

Yves Guiraud, Philippe Malbos and Samuel Mimram have organised the third edition of the Higher-Dimensional Rewriting and Algebra (HDRA) workshop of the Formal Structures for Computation and Deduction conference (FSCD), held in Oxford in September 2017.

Yves Guiraud and Samuel Mimram, with Dimitri Ara (Univ. Aix-Marseille) have organised the "Categories in Homotopy and Rewriting" one-week international conference, at the CIRM, in Marseille, in September 2017. As the closing conference of the Cathre ANR project, focused on higher-dimensional algebra, this conference attracted 80 participants, working in category theory, algebraic topology, logic and theoretical computer science.

Matthieu Sozeau co-organised with Nicolas Tabareau the 3rd Coq Implementors Workshop in Le Croisic, France, June 12-16 2017. It included presentations from developers, both from France and abroad and a large amount of hacking.

8.1.2. Scientific Events Selection

8.1.2.1. Member of the Conference Program Committees

Hugo Herbelin was a member of the program committees of the conference FSCD'17, of the TYPES'17 venue, as well as of the PxTP'17 and CoqPL'18 workshops.

Matthieu Sozeau was member of the program committee of CoqPL'17.

Yann Régis-Gianas was member of the program committee of JFLA'18.

Alexis Saurin was a member of the program committee of the workshop Coinduction in Type Theory which took place in Chambéry from the 3rd to the 6th of July, 2017

Alexis Saurin was a member of the program committee of the workshop on Trends in Linear Logic and Applications which took place in Oxford on the 3rd of September 2017 as a satellite event of FSCD conference.

8.1.2.2. Member of the Conference Steering Committees

Hugo Herbelin was a member of the steering committee of the conference *Formal Structures for Computation and Deduction* (FSCD) until September 2017.

Hugo Herbelin is a member of the steering committee of the conference *TYPES*.

Pierre-Louis Curien is member of the steering committee of the international workshop Games for Logic and Programming Languages (GaLop).

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 an invited talk at the Conference Categories for Homotopy Theory and Rewriting on "A syntactic approach to polynomial functors, polynomial monads and opetopes" (September 2017).

Yann Régis-Gianas gave an invited talk at the Conference for Trends in Functional Programming In Education on the OCaml MOOC (April 2017).

8.1.5. Scientific Expertise

Pierre-Louis Curien and Yves Guiraud have been members of the "Comité de sélection" for an assistant professor position in mathematical foundations of computer science at the University Paris 7 (spring 2017).

8.1.6. Research Administration

Pierre-Louis Curien, Hugo Herbelin and Yves Guiraud are members of the scientific council of the Computer Science deparment of University Paris 7.

Yves Guiraud is the head of the "Preuves, Programmes and Systèmes (PPS)" pole of the IRIF laboratory (since April 2016), a member of the IRIF council (January 2016 - December 2017), and of the IRIF direction council (since September 2017).

Pierre-Louis Curien is a member of the Scientific Council of the CIRM (Centre International de Rencontres Mathématiques.

8.1.7. Presentation of papers

Hugo Herbelin gave a talk at the GT Geocal-Lac on a proof of Gödel's completeness theorem using side-effects.

Hugo Herbelin gave a talk at the PPS pole days on the intuitionistic reverse mathematics properties of Gödel's completeness theorem.

Étienne Miquey gave a talk at the ESOP conference in Uppsala (Sweden) on "A classical sequent calculus with dependent types" (April 2017).

Jovana Obradović gave a talk at the Conference Topology in Ecuador (Galapagos Islands) on "Categorified cyclic operads" (August 2017).

Théo Zimmermann gave a talk on Coq's Prolog and application to defining semi-automatic tactics at the TTT'17 workshop (Type-Theory Based Tools).

Cyprien Mangin gave a talk on Equations at the TTT'17 workshop.

Matthieu Sozeau gave a talk on Coq 8.6 at the CoqPL'17 workshop in January 2017.

8.1.8. Talks in seminars

Pierre-Louis Curien gave talks at NII (Hasuo's Lab, Tokyo) and at Shanghai Jiaotong University on "Trialgebraic structures on faces of some families of polytopes" (October-November 2017).

Théo Zimmermann gave a talk on transfer of isomorphisms at the Deducteam seminar (April 2017).

8.1.9. Attendance to conferences, workshops, schools,...

Hugo Herbelin attended TTT'17, TYPES'17 and FSCD'17.

Cyrille Chenavier, Maxime Lucas, Philippe Malbos and Samuel Mimram attended the HDRA workshop in Oxford (September 2017).

Cyrille Chenavier, Pierre-Louis Curien, Yves Guiraud, Maxime Lucas, Philippe Malbos and Samuel Mimram attended the "Categories in Homotopy and Rewriting" conference in Marseille (September 2017).

Hugo Herbelin attended the conferences Types 2017 in Budapest (Hungaria, May), the Coq implementors workshop in Le Croisic (June), the FSCD conference in Oxford (UK, September). He attended various workshops of the POPL'17 and FSCD'17 conference, including TTT'17 and CoqPL'17. He participated to the Big Proofs seminar in Cambridge (July).

Jean-Jacques Lévy participated to CAV and VSTTE 2017 conferences, Heidelberg, Germany, July 21-28 where his co-author Ran Chen presented the article [36].

Matthieu Sozeau attended POPL'17, CPP'17 and CoqPL'17 in Paris (January), TYPES'17 in Budapest (Hungaria, May) and the Coq implementors workshop in Le Croisic (June).

Étienne Miquey attended the conference ESOP in Uppsala (Sweden) in April 2017.

Pierre-Louis Curien and Jovana Obradović attended the conference Topology in Ecuador (Galapagos Islands) in August 2017, and the conference Geometry and combinatorics of associativity in Dublin in October 2017.

Pierre-Louis Curien and Claudia Faggian attended the CRECOGI meeting in Ito (Japan), and Claudia Faggian gave a talk on "Proof techniques in Probabilistic Reduction Systems" (November 2017).

Théo Zimmermann attended the TTT'17 workshop (January 2017) and the Coq implementors workshop in Le Croisic (June 2017).

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 2017 were Pierre-Marie Pédrot, Guilhem Jaber, Francesco A. Genco, Ludovic Patey.

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 (IRIF) 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).

Master: Hugo Herbelin teaches the course on the proof-as-program correspondence for classical logic and beyond at the LMFI.

Master: Pierre Letouzey teaches two short courses to the LMFI Master 2 students : "Models of programming" 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).

Master: Yves Guiraud gave a course on the applications of rewriting methods in algebra in the M2 Mathématiques Fondamentales of Lyon (Univ. Lyon 1 and ENS Lyon).

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 taught a course at the EJCP'17 summer school in Toulous in June 17, on an introduction to interactive theorem proving.

Master: Alexis aurin 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.

8.2.2. Supervision

Internship: Hugo Herbelin has supervised the M2 internship of Charlotte Barot.

Internship: Yann Régis-Gianas has supervised the L3 internship of Kostia Chardonnet.

Internship: Yann Régis-Gianas has supervised the M2 internship of Colin Gonzalez.

PhD (completed) Amina Doumane, supervised by Alexis Saurin, David Baelde and Pierre-Louis Curien, successfully defended in June 2017 (title: On the infinitary proof theory of logics with fixed points).

PhD (completed): Étienne Miquey, co-supervised by Hugo Herbelin and Alexandre Miquel, Réalisabilité classique et effets de bords, September 2014, successfully defended in November 2017.

PhD (completed): Jovana Obradović, supervised by Pierre-Louis Curien, successfully defended in Septemver 2017 (title: Cyclic operads: syntactic, algebraic and categorified aspects).

PhD (completed): Maxime Lucas, supervised by Yves Guiraud and Pierre-Louis Curien, successfully defended in December 2017 (title: Cubical categories for homotopy and rewriting).

PhD in progress: Guillaume Claret, Programmation avec effets en Coq, (started in September 2012), supervised by Hugo Herbelin and Yann Régis-Gianas, the dissertation was completed in 2015 but Guillaume Claret moved in the meantime to a private company and the defense has been delayed to 2018.

PhD in progress: Thibaut Girka, Differential semantics (started in January 2014), supervised by Roberto Di Cosmo and Yann Régis-Gianas.

PhD in progress: Cyprien Mangin, Dependent Pattern-Matching, induction-induction and higher inductive types (started in September 2015), supervised by Matthieu Sozeau and Bruno Barras

PhD in progress: Théo Zimmermann (started in September 2016), supervised by Hugo Herbelin.

PhD starting: Cédric Ho Thanh, on Opetopes for higher-dimensional rewriting and koszulity, supervised by Pierre-Louis Curien and Samuel Mimram.

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 Samuel Mimram and Yves Guiraud.

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 referee for the Habilitation of Russ Harmer (ENS Lyon, May 2017).

Pierre-Louis Curien and Alexis Saurin were members of the jury of the thesis of Amina Doumane (Paris 7, June 2017).

Pierre-Louis Curien was president of the jury of the Habilitation of Matthieu Picantin (Paris 7, July 2017).

Pierre-Louis Curien was president of the jury of the thesis of Simon Castellan (ENS Lyon, July 2017).

Pierre-Louis Curien was member of the jury of the thesis of Jovana Obradović (Paris 7, September 2017).

Pierre-Louis Curien was referee for the Habilitation of Paul-André Melliès (Paris 7, November 2017).

Pierre-Louis Curien was president of the jury of Habilitation of Damiano Mazza (Paris 13, December 2017).

Pierre-Louis Curien was referee for the Habilitation of Samuele Giraudo (Marne la Vallée, December 2017).

Pierre-Louis Curien was member of the jury of the thesis of Nicolas Ninin (Paris Saclay, December 2017).

Pierre-Louis Curien was president of the jury of the thesis of Luc Pellissier (Paris 13, December 2017).

Pierre-Louis Curien was referee for the thesis of Christopher Nguyen (Macquarie University, Sydney, December 2017).

Pierre-Louis Curien and Yves Guiraud were members of the jury of the thesis of Maxime Lucas (Paris 7, December 2017).

8.3. Popularization

Hugo Herbelin wrote with Sandrine Blazy and Pierre Castéran an introduction to Coq for engineers edited by Techniques de l'Ingénieur.

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] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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/.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] M. SOZEAU, N. OURY.*First-Class Type Classes*, in "Theorem Proving in Higher Order Logics, 21st International Conference, TPHOLs 2008, Montreal, Canada, August 18-21, 2008. Proceedings", O. A. MOHAMED, C. MUÑOZ, S. TAHAR (editors), Lecture Notes in Computer Science, Springer, 2008, vol. 5170, p. 278-293.

[22] 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

[23] M. LUCAS.*Cubical categories for homotopy and rewriting*, Université Paris 7, Sorbonne Paris Cité, December 2017, https://hal.archives-ouvertes.fr/tel-01668359.

[24] É. MIQUEY.*Classical realizability and side-effects*, Université Sorbonne Paris Cité - Université Paris Diderot (Paris 7) ; Universidad de la República - Montevideo, Uruguay, November 2017, https://hal.inria.fr/tel-01653733.

[25] J. OBRADOVIC.*Cyclic operads: syntactic, algebraic and categorified aspects*, Université Paris Diderot - Paris 7 - Sorbonne Paris Cité, September 2017, https://hal.archives-ouvertes.fr/tel-01676983.

Articles in International Peer-Reviewed Journal

[26] J.-P. BERNARDY, M. BOESPFLUG, R. R. NEWTON, S. PEYTON JONES, A. SPIWACK.*Linear Haskell: practical linearity in a higher-order polymorphic language*, in "Proceedings of the ACM on Programming Languages", December 2017, vol. 2, n^{o} POPL, p. 1-29, https://arxiv.org/abs/1710.09756 [*DOI :* 10.1145/3158093], https://hal.archives-ouvertes.fr/hal-01673536.

[27] S. BLAZY, P. CASTÉRAN, H. HERBELIN.*L'Assistant de Preuve Coq Table des matières*, in "Techniques de l'Ingenieur", August 2017, https://hal.inria.fr/hal-01645486.

[28] C. CHENAVIER.*Reduction Operators and Completion of Rewriting Systems*, in "Journal of Symbolic Computation", 2017, https://arxiv.org/abs/1605.00174 , https://hal.archives-ouvertes.fr/hal-01325907.

[29] P.-L. CURIEN, S. MIMRAM.*Coherent Presentations of Monoidal Categories*, in "Logical Methods in Computer Science", September 2017, vol. 13, n^{o} 3, p. 1-38, https://arxiv.org/abs/1705.03553 [*DOI :* 10.23638/LMCS-13(3:31)2017], https://hal.inria.fr/hal-01662524.

[30] F. LOULERGUE, W. BOUSDIRA, J. TESSON.*Calculating Parallel Programs in Coq using List Homomorphisms*, in "International Journal of Parallel Programming", 2017, vol. 45, n^{o} 2, p. 300-319 [*DOI :* 10.1007/s10766-016-0415-8], https://hal.inria.fr/hal-01159182.

[31] M. LUCAS.*A coherence theorem for pseudonatural transformations*, in "Journal of Pure and Applied Algebra", 2017, vol. 221, n^{o} 5, p. 1146-1217, https://arxiv.org/abs/1508.07807 [*DOI :* 10.1016/J.JPAA.2016.09.005], https://hal.archives-ouvertes.fr/hal-01191867.

[32] J. OBRADOVIC, P.-L. CURIEN, J. IVANOVIC.*Syntactic aspects of hypergraph polytopes*, in "Journal of Homotopy and Related Structures", 2017, https://hal.archives-ouvertes.fr/hal-01669490.

[33] 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.

[34] D. DE RAUGLAUDRE.*Formal Proof of Banach-Tarski Paradox*, in "Journal of Formalized Reasoning", October 2017, vol. 10, n^{o} 1, p. 37-49 [*DOI :* 10.6092/ISSN.1972-5787/6927], https://hal.archives-ouvertes.fr/hal-01673378.

International Conferences with Proceedings

[35] A. BAUER, G. JASON, P. LUMSDAINE, M. SHULMAN, M. SOZEAU, B. SPITTERS.*The HoTT Library: A Formalization of Homotopy Type Theory in Coq*, in "CPP'17", Paris, France, CPP'17, ACM, January 2017, 9 [*DOI :* 10.1145/3018610.3018615], https://hal.inria.fr/hal-01421212.

[36] R. CHEN, J.-J. LÉVY.*A Semi-automatic Proof of Strong connectivity*, in "9th Working Conference on Verified Software: Theories, Tools and Experiments (VSTTE)", Heidelberg, Germany, July 2017, https://hal.inria.fr/hal-01632947.

[37] *Best Paper*
A. DOUMANE.*Constructive completeness for the linear-time μ-calculus*, in "Conference on Logic in Computer Science 2017", Reykjavik, Iceland, June 2017, https://hal.archives-ouvertes.fr/hal-01430737.

[38] T. GIRKA, D. MENTRÉ, Y. RÉGIS-GIANAS.*Verifiable Semantic Difference Languages*, in "International Symposium on Principles and Practice of Declarative Programming", Namur, Belgium, October 2017 [*DOI :* 10.1145/3131851.3131870], https://hal.inria.fr/hal-01653283.

[39] É. MIQUEY.*A Classical Sequent Calculus with Dependent Types* , in "26th European Symposium on Programming", Uppsala, Sweden, April 2017, https://hal.inria.fr/hal-01375977.

National Conferences with Proceeding

[40] R. CHEN, J.-J. LÉVY.*Une preuve formelle de l'algorithme de Tarjan-1972 pour trouver les composantes fortement connexes dans un graphe*, in "JFLA 2017 - Vingt-huitièmes Journées Francophones des Langages Applicatifs", Gourette, France, Vingt-huitièmes Journées Francophones des Langages Applicatifs, January 2017, https://hal.inria.fr/hal-01422215.

Conferences without Proceedings

[41] A. ANAND, S. BOULIER, N. TABAREAU, M. SOZEAU.*Typed Template Coq – Certified Meta-Programming in Coq*, in "The Fourth International Workshop on Coq for Programming Languages", Los Angeles, CA, United States, January 2018, https://hal.inria.fr/hal-01671948.

[42] T. ZIMMERMANN, H. HERBELIN.*Coq's Prolog and application to defining semi-automatic tactics*, in "Type Theory Based Tools", Paris, France, January 2017, https://hal.archives-ouvertes.fr/hal-01671994.

Research Reports

[43] A. TIMANY, M. SOZEAU.*Consistency of the Predicative Calculus of Cumulative Inductive Constructions (pCuIC)*, KU Leuven, Belgium ; Inria Paris, October 2017, n^{o} RR-9105, 30, Version 2 fixes some typos from version 1, https://hal.inria.fr/hal-01615123.

Other Publications

[44] C. CHENAVIER.*A Lattice Formulation of the F 4 Completion Procedure*, March 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01489200.

[45] C. CHENAVIER.*Syzygies among reduction operators*, August 2017, https://arxiv.org/abs/1708.08709 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01578555.

[46] 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.

[47] Y. GUIRAUD, E. HOFFBECK, P. MALBOS.*Convergent presentations and polygraphic resolutions of associative algebras*, December 2017, 65 pages, https://hal.archives-ouvertes.fr/hal-01006220.

[48] H. HERBELIN, É. MIQUEY.*Normalization and continuation-passing-style interpretation of simply-typed call-by-need λ-calculus with control*, July 2017, working paper or preprint, https://hal.inria.fr/hal-01570987.

[49] N. JEANNEROD, Y. RÉGIS-GIANAS, R. TREINEN.*Having Fun With 31.521 Shell Scripts*, April 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01513750.

[50] M. LUCAS.*A cubical Squier's theorem*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01662132.

[51] M. LUCAS.*Cubical (ω,p)-categories*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01662127.

[52] C. MANGIN, M. SOZEAU.*Equations reloaded*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01671777.

[53] G. MANZONETTO, A. POLONSKY, A. SAURIN, J. G. SIMONSEN.*The Fixed Point Property and a Technique to Harness Double Fixed Point Combinators* , December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01672846.

[54] É. MIQUEY, H. HERBELIN.*Realizability interpretation and normalization of typed call-by-need λ-calculus with control*, October 2017, working paper or preprint, https://hal.inria.fr/hal-01624839.

[55] É. MIQUEY.*A Classical Sequent Calculus with Dependent Types (Extended Version)*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01519929.

[56] N. TABAREAU, É. TANTER, M. SOZEAU.*Equivalences for Free!: Univalent Parametricity for Effective Transport*, July 2017, working paper or preprint, https://hal.inria.fr/hal-01559073.

References in notes

[57] F. A. AL-AGL, R. BROWN, R. STEINER.*Multiple categories: the equivalence of a globular and a cubical approach*, in "Adv. Math.", 2002, vol. 170, n^{o} 1, p. 71–118, https://doi.org/10.1006/aima.2001.2069.

[58] D. J. ANICK.*On the Homology of Associative Algebras*, in "Trans. Amer. Math. Soc.", 1986, vol. 296, n^{o} 2, p. 641–659.

[59] D. ARA, F. MÉTAYER.*The Brown-Golasiński Model Structure on strict ∞-groupoids revisited*, in "Homology, Homotopy and Applications", 2011, vol. 13, n^{o} 1, p. 121–142.

[60] J. BAEZ, A. CRANS.*Higher-dimensional algebra. VI. Lie 2-algebras*, in "Theory Appl. Categ.", 2004, vol. 12, p. 492–538.

[61] H. P. BARENDREGT.*The Lambda Calculus: Its Syntax and Semantics*, North Holland, Amsterdam, 1984.

[62] R. BERGER.*Confluence and Koszulity*, in "J. Algebra", 1998, vol. 201, n^{o} 1, p. 243–283.

[63] Y. BERTOT, P. CASTÉRAN.*Interactive Theorem Proving and Program Development Coq'Art: The Calculus of Inductive Constructions*, Springer, 2004.

[64] 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.

[65] 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.

[66] R. BROWN, P. J. HIGGINS.*The equivalence of ∞-groupoids and crossed complexes*, in "Cahiers Topologie Géom. Différentielle", 1981, vol. 22, n^{o} 4, p. 371–386.

[67] R. BROWN, P. J. HIGGINS.*The equivalence of ω-groupoids and cubical T-complexes*, in "Cahiers Topologie Géom. Différentielle", 1981, vol. 22, n^{o} 4, p. 349–370.

[68] 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.

[69] 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.

[70] 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.

[71] A. CHURCH.*A set of Postulates for the foundation of Logic*, in "Annals of Mathematics", 1932, vol. 2, p. 33, 346-366.

[72] J. COCKX, D. DEVRIESE, F. PIESSENS.*Pattern matching without K*, in "Proceedings of the 19th ACM SIGPLAN international conference on Functional programming, Gothenburg, Sweden, September 1-3, 2014", 2014, p. 257–268, http://doi.acm.org/10.1145/2628136.2628139.

[73] T. COQUAND.*Une théorie des Constructions*, University Paris 7, January 1985.

[74] 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.

[75] 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.

[76] H. B. CURRY, R. FEYS, W. CRAIG.*Combinatory Logic*, North-Holland, 1958, vol. 1, §9E.

[77] 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.

[78] P. DELIGNE.*Action du groupe des tresses sur une catégorie*, in "Invent. Math.", 1997, vol. 128, n^{o} 1, p. 159–175.

[79] J.-C. FAUGÉRE.*A new efficient algorithm for computing Gröbner bases (F_4)*, in "J. Pure Appl. Algebra", 1999, vol. 139, n^{o} 1-3, p. 61–88, Effective methods in algebraic geometry (Saint-Malo, 1998), https://doi.org/10.1016/S0022-4049(99)00005-5.

[80] 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.

[81] 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.

[82] G. GENTZEN.*Untersuchungen über das logische Schließen*, in "Mathematische Zeitschrift", 1935, vol. 39, p. 176–210,405–431.

[83] 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.

[84] 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.

[85] Y. GUIRAUD.*Présentations d'opérades et systèmes de réécriture*, Univ. Montpellier 2, 2004.

[86] Y. GUIRAUD.*Termination Orders for 3-Dimensional Rewriting*, in "Journal of Pure and Applied Algebra", 2006, vol. 207, n^{o} 2, p. 341–371.

[87] Y. GUIRAUD.*The Three Dimensions of Proofs*, in "Annals of Pure and Applied Logic", 2006, vol. 141, n^{o} 1–2, p. 266–295.

[88] Y. GUIRAUD.*Two Polygraphic Presentations of Petri Nets*, in "Theoretical Computer Science", 2006, vol. 360, n^{o} 1–3, p. 124–146.

[89] 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.

[90] Y. GUIRAUD, P. MALBOS.*Higher-dimensional categories with finite derivation type*, in "Theory Appl. Categ.", 2009, vol. 22, n^{o} 18, p. 420-478.

[91] 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.

[92] Y. GUIRAUD, P. MALBOS.*Coherence in monoidal track categories*, in "Math. Structures Comput. Sci.", 2012, vol. 22, n^{o} 6, p. 931–969.

[93] 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.

[94] 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.

[95] J.-L. KRIVINE.*A call-by-name lambda-calculus machine*, in "Higher Order and Symbolic Computation", 2005.

[96] J.-L. KRIVINE.*Un interpréteur du lambda-calcul*, 1986, Unpublished.

[97] Y. LAFONT.*Towards an Algebraic Theory of Boolean Circuits*, in "Journal of Pure and Applied Algebra", 2003, vol. 184, p. 257-310.

[98] Y. LAFONT, F. MÉTAYER, K. WORYTKIEWICZ.*A Folk Model Structure on Omega-Cat*, in "Advances in Mathematics", 2010, vol. 224, n^{o} 3, p. 1183–1231.

[99] P. LANDIN.*The mechanical evaluation of expressions*, in "The Computer Journal", January 1964, vol. 6, n^{o} 4, p. 308–320.

[100] P. LANDIN.*A generalisation of jumps and labels*, UNIVAC Systems Programming Research, August 1965, n^{o} ECS-LFCS-88-66, Reprinted in Higher Order and Symbolic Computation, 11(2), 1998.

[101] G. LEE, B. WERNER.*Proof-irrelevant model of CC with predicative induction and judgmental equality*, in "Logical Methods in Computer Science", 2011, vol. 7, n^{o} 4.

[102] P. MALBOS.*Critères de finitude homologique pour la non convergence des systèmes de réécriture de termes*, Univ. Montpellier 2, 2004.

[103] P. MARTIN-LÖF.*A theory of types*, University of Stockholm, 1971, n^{o} 71-3.

[104] 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.

[105] S. B. PRIDDY.*Koszul resolutions*, in "Trans. Amer. Math. Soc.", 1970, vol. 152, p. 39–60.

[106] 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.

[107] 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.

[108] 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.

[109] 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.

[110] R. STEINER.*Omega-categories and chain complexes*, in "Homology Homotopy Appl.", 2004, vol. 6, n^{o} 1, p. 175–200, http://projecteuclid.org/euclid.hha/1139839551.

[111] R. STREET.*Limits Indexed by Category-Valued 2-Functors*, in "Journal of Pure and Applied Algebra", 1976, vol. 8, p. 149–181.

[112] T. C. D. TEAM.*The Coq Proof Assistant, version 8.7.1*, December 2017, https://doi.org/10.5281/zenodo.1133970.

[113] 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

Université Pierre et Marie Curie (Paris 6)

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. - Verification, reliability, certification
A4.3. - Cryptography
A4.3.1. - Public key cryptography
A4.3.4. - Quantum Cryptography
A5.10.1. - Design
A6.1. - 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
B9.4.1. - Computer science
B9.4.2. - Mathematics
B9.8. - Privacy

1. Personnel

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 [UPMC, Associate Professor]
Daniel Lazard [UPMC, Emeritus Professor, HDR]
Ludovic Perret [UPMC, Associate Professor, HDR]
Guénaël Renault [UPMC, Associate Professor, on leave at ANSSI, HDR]
Mohab Safey El Din [UPMC, Professor, HDR]

External Collaborators
Emmanuel Prouff [ANSSI, Associate Member, HDR]
Victor Magron [CNRS, Researcher, from Oct. 2017]

Technical Staff
Jocelyn Ryckeghem [UPMC, from Apr. 2017 until Dec. 2017]

PhD Students
Ivan Bannwarth [UPMC, until Aug. 2017]
Matías Bender [Inria]
Olive Chakraborty [UPMC, from May 2017]
Nagardjun Chinthamani Dwarakanath [UPMC, from Dec. 2017]
Solane El Hirch [UPMC, from June 2017]
Thi Xuan Vu [UPMC, from Oct. 2017]

Post-Doctoral Fellows
Amine Mrabet [UPMC, ATER, from Sept. 2017]
Kaie Kubjas [UPMC, Post-Doctoral fellow, on leave at MIT, from Sept. 2017]

Administrative Assistants
Kevin Bonny [Inria]
Georgette Bonpapa [UPMC, Assistant, until July 2017]
Virginie Collette [Inria]
Irphane Khan [UPMC, Assistant]
Azzeddine Saidani [Inria]

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 who 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 FGBlibrary 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 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.

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

Dongming Wang has been elected as a Member of the Academia Europaea.

Elias Tsigaridas was awarded an ANR "Jeune Chercheur Grant". The title of the project is GALOP (Games through the lens of ALgebra and OPptimization)

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

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. 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

5.8. 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. Linear Algebra for Computing Gröbner Bases of Linear Recursive Multidimensional Sequences

The so-called Berlekamp – Massey – Sakata algorithm computes a Gröbner basis of a 0-dimensional ideal of relations satisfied by an input table. It extends the Berlekamp – Massey algorithm to n-dimensional tables, for $n > 1$.

In [1], we investigate this problem and design several algorithms for computing such a Gröbner basis of an ideal of relations using linear algebra techniques. The first one performs a lot of table queries and is analogous to a change of variables on the ideal of relations.

As each query to the table can be expensive, we design a second algorithm requiring fewer queries, in general. This FGLM-like algorithm allows us to compute the relations of the table by extracting a full rank submatrix of a *multi-Hankel* matrix (a multivariate generalization of Hankel matrices).

Under some additional assumptions, we make a third, adaptive, algorithm and reduce further the number of table queries. Then, we relate the number of queries of this third algorithm to the *geometry* of the final staircase and we show that it is essentially linear in the size of the output when the staircase is convex. As a direct application to this, we decode n-cyclic codes, a generalization in dimension n of Reed Solomon codes.

We show that the multi-Hankel matrices are heavily structured when using the LEX ordering and that we can speed up the computations using fast algorithms for quasi-Hankel matrices. Finally, we design algorithms for computing the generating series of a linear recursive table.

6.1.2. In-depth comparison of the Berlekamp – Massey – Sakata and the Scalar-FGLM algorithms: the non adaptive variants

In [22], we compare thoroughly the BERLEKAMP – MASSEY – SAKATA algorithm and the SCALAR-FGLM algorithm, which compute both the ideal of relations of a multidimensional linear recurrent sequence.

Suprisingly, their behaviors differ. We detail in which way they do and prove that it is not possible to tweak one of the algorithms in order to mimic exactly the behavior of the other.

6.1.3. Resultants and Discriminants for Bivariate Tensor-product Polynomials

Optimal resultant formulas have been systematically constructed mostly for *unmixed polynomial systems*, that is, systems of polynomials which all have the same support. However, such a condition is restrictive, since *mixed systems* of equations arise frequently in practical problems. In [16] we present a square, *Koszul-type* matrix expressing the resultant of arbitrary (mixed) bivariate *tensor-product systems*. The formula generalizes the classical Sylvester matrix of two univariate polynomials, since it expresses a map of *degree one*, that is, the entries of the matrix are simply coefficients of the input polynomials. Interestingly, the matrix expresses a primal-dual multiplication map, that is, the tensor product of a univariate multiplication map with a map expressing derivation in a dual space. Moreover, for tensor-product systems with more than two (affine) variables, we prove an impossibility result: no universal degree-one formulas are possible, unless the system is unmixed. We also present applications of the new construction in the computation of discriminants and mixed discriminants as well as in solving systems of bivariate polynomials with tensor-product structure.

6.1.4. Sparse Rational Univariate Representation

In [15] we present explicit worst case degree and height bounds for the rational univariate representation of the isolated roots of polynomial systems based on mixed volume. We base our estimations on height bounds of resultants and we consider the case of 0-dimensional, positive dimensional, and parametric polynomial systems.

Multi-homogeneous polynomial systems arise in many applications. In [11], we provide bit complexity estimates for representing the solutions of these systems. These are the best currently known bounds. 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 is finite.

We do not only obtain bounds but an algorithm is also given for solving such systems. We give bit complexity estimates 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 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.

6.1.5. Improving Root Separation Bounds

Let f be a polynomial (or polynomial system) with all simple roots. The root separation of f is the minimum of the pair-wise distances between the complex roots. A root separation bound is a lower bound on the root separation. Finding a root separation bound is a fundamental problem, arising in numerous disciplines. In [7] we present two new root separation bounds: one univariate bound, and one multivariate bound. The new bounds improve on the old bounds in two ways: (1) The new bounds are usually significantly bigger (hence better) than the previous bounds. (2) The new bounds scale correctly, unlike the previous bounds. Crucially, the new bounds are not harder to compute than the previous bounds.

6.1.6. Accelerated Approximation of the Complex Roots and Factors of a Univariate Polynomial

The known algorithms approximate the roots of a complex univariate polynomial in nearly optimal arithmetic and Boolean time. They are, however, quite involved and require a high precision of computing when the degree of the input polynomial is large, which causes numerical stability problems. In [8] we observe that these difficulties do not appear at the initial stages of the algorithms, and in our present paper we extend one of these stages, analyze it, and avoid the cited problems, still achieving the solution within a nearly optimal complexity estimates, provided that some mild initial isolation of the roots of the input polynomial has been ensured. The resulting algorithms promise to be of some practical value for root-finding and can be extended to the problem of polynomial factorization, which is of interest on its own right. We conclude with outlining such an extension, which enables us to cover the cases of isolated multiple roots and root clusters.

6.1.7. Nearly optimal computations with structured matrices

In [9] we estimate the Boolean complexity of multiplication of structured matrices by a vector and the solution of nonsingular linear systems of equations with these matrices. We study four basic and most popular classes, that is, Toeplitz, Hankel, Cauchy and Vandermonde matrices, for which the cited computational problems are equivalent to the task of polynomial multiplication and division and polynomial and rational multipoint evaluation and interpolation. The Boolean cost estimates for the latter problems have been obtained by Kirrinnis, except for rational interpolation. We supply them now as well as the Boolean complexity estimates for the important problems of multiplication of transposed Vandermonde matrix and its inverse by a vector. All known Boolean cost estimates for such problems rely on using Kronecker product. This implies the d-fold precision increase for the d-th degree output, but we avoid such an increase by relying on distinct techniques based on employing FFT. Furthermore we simplify the analysis and make it more transparent by combining the representations of our tasks and algorithms both via structured matrices and via polynomials and rational functions. This also enables further extensions of our estimates to cover Trummer's important problem and computations with the popular classes of structured matrices that generalize the four cited basic matrix classes, as well as the transposed Vandermonde matrices. It is known that the solution of Toeplitz, Hankel, Cauchy, Vandermonde, and transposed Vandermonde linear systems of equations is generally prone to numerical stability problems, and numerical problems arise even for multiplication of Cauchy, Vandermonde, and transposed Vandermonde matrices by a vector. Thus our FFT-based results on the Boolean complexity of these important computations could be quite interesting because our estimates are reasonable even for more general classes of structured matrices, showing rather moderate growth of the complexity as the input size increases.

6.1.8. Sliding solutions of second-order differential equations with discontinuous right-hand side

In [2], we consider second-order ordinary differential equations with discontinuous right-hand side. We analyze the concept of solution of this kind of equations and determine analytical conditions that are satisfied by typical solutions. Moreover, the existence and uniqueness of solutions and sliding solutions are studied.

6.1.9. Sparse FGLM algorithms

Given a zero-dimensional ideal $I \subset \mathbb{K}[x_1, ..., x_n]$ of degree D, the transformation of the ordering of its Gröbner basis from DRL to LEX is a key step in polynomial system solving and turns out to be the bottleneck of the whole solving process. Thus it is of crucial importance to design efficient algorithms to perform the change of ordering. The main contributions of [3] are several efficient methods for the change of ordering which take advantage of the sparsity of multiplication matrices in the classical *FGLM* algorithm. Combining all these methods, we propose a deterministic top-level algorithm that automatically detects which method to use depending on the input. As a by-product, we have a fast implementation that is able to handle ideals of degree over $40,000$. Such an implementation outperforms the *Magma* and *Singular* ones, as shown by our experiments. First for the shape position case, two methods are designed based on the Wiedemann algorithm: the first is probabilistic and its complexity to complete the change of ordering is $O(D(N_1 + n \log D))$, where N_1 is the number of nonzero entries of a multiplication matrix; the other is deterministic and computes the LEX Gröbner basis of $\sqrt{I}$ via Chinese Remainder Theorem. Then for the general case, the designed method is characterized by the Berlekamp–Massey–Sakata algorithm from Coding Theory to handle the multi-dimensional linearly recurring relations. Complexity analyses of all proposed methods are also provided. Furthermore, for generic polynomial systems, we present an explicit formula for the estimation of the sparsity of one main multiplication matrix, and prove its construction is free. With the asymptotic analysis of such sparsity, we are able to show for generic systems the complexity above becomes $O(\sqrt{6/n\pi} D^{2+\frac{n-1}{n}})$.

6.2. Solving Systems over the Reals and Applications

6.2.1. Answering connectivity queries in real algebraic sets

A roadmap for a semi-algebraic set S is a curve which has a non-empty and connected intersection with all connected components of S. Hence, this kind of object, introduced by Canny, can be used to answer connectivity queries (with applications, for instance, to motion planning) but has also become of central importance in effective real algebraic geometry, since it is used in higher-level algorithms. In [10], we provide a probabilistic algorithm which computes roadmaps for smooth and bounded real algebraic sets. Its output size and running time are polynomial in $(nD)^{n \log d}$, where D is the maximum of the degrees of the input polynomials, d is the dimension of the set under consideration and n is the number of variables. More precisely, the running time of the algorithm is essentially subquadratic in the output size. Even under our assumptions, it is the first roadmap algorithm with output size and running time polynomial in $(nD)^{n \log d}$.

6.2.2. Polynomial optimization and semi-definite programming

In [6], we describe our freely distributed Maple library spectra, for Semidefinite Programming solved Exactly with Computational Tools of Real Algebra. It solves linear matrix inequalities, a fundamental object in effective real algebraic geometry and polynomial optimization, with symbolic computation in exact arithmetic and it is targeted to small-size, possibly degenerate problems for which symbolic infeasibility or feasibility certificates are required.

The positive semidefinite rank of a convex body C is the size of its smallest positive semi-definite formulation. In [5], 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.

6.2.3. The Complexity of an Adaptive Subdivision Method for Approximating Real Curves

In [14] we present the first complexity analysis of the algorithm by Plantinga and Vegter for approximating real implicit curves and surfaces. This approximation algorithm certifies the topological correctness of the output using both subdivision and interval arithmetic. In practice, it has been seen to be quite efficient; our goal is to quantify this efficiency. We focus on the subdivision step (and not the approximation step) of the Plantinga and Vegter algorithm. We begin by extending the subdivision step to arbitrary dimensions. We provide *a priori* worst-case bounds on the complexity of this algorithm both in terms of the number of subregions constructed and the bit complexity for the construction. Then, we use continuous amortization to derive adaptive bounds on the complexity of the subdivided region. We also provide examples showing our bounds are tight.

6.3. Solving Systems in Finite Fields, Applications in Cryptology and Algebraic Number Theory.

6.3.1. Private Multiplication over Finite Fields

The notion of privacy in the probing model, introduced by Ishai, Sahai, and Wagner in 2003, is nowadays frequently involved to assess the security of circuits manipulating sensitive information. However, provable security in this model still comes at the cost of a significant overhead both in terms of arithmetic complexity and randomness complexity. In [13], we deal with this issue for circuits processing multiplication over finite fields. Our contributions are manifold. Extending the work of Belaïd, Benhamouda, Passelègue, Prouff, Thillard, and Vergnaud at Eurocrypt 2016, we introduce an algebraic characterization of the privacy for multiplication in any finite field and we propose a novel algebraic characterization for non-interference (a stronger security notion in this setting). Then, we present two generic constructions of multiplication circuits in finite fields that achieve non-interference in the probing model. The second proposal achieves a linear complexity in terms of randomness consumption. This complexity is proved to be almost optimal. Eventually, we show that our constructions can always be instantiated in large enough finite fields.

6.3.2. Convolutional Neural Networks with Data Augmentation Against Jitter-Based Countermeasures - Profiling Attacks Without Pre-processing

In the context of the security evaluation of cryptographic implementations, profiling attacks (aka Template Attacks) play a fundamental role. Nowadays the most popular Template Attack strategy consists in approximating the information leakages by Gaussian distributions. Nevertheless this approach suffers from the difficulty to deal with both the traces misalignment and the high dimensionality of the data. This forces the attacker to perform critical preprocessing phases, such as the selection of the points of interest and the temporal realignment of measurements. Some software and hardware countermeasures have been conceived exactly to create such a misalignment. In [17], we propose an end-to-end profiling attack strategy based on Deep Learning algorithms combined with Data Augmentation strategies.

6.3.3. Submissions to the NIST Post-Quantum Standardization Process

We have submitted three cryptosystems to the current process leads by NIST for standardizing post-quantum public-key algorithms.

6.3.3.1. GeMSS

The acronym stands for a Great Multivariate Signature Scheme [18]. As suggested by its name, *GeMSS* is a multivariate-based signature scheme producing small signatures. It has a fast verification process, and a medium/large public-key. *GeMSS* is in direct lineage from QUARTZ and borrows some design rationale of the Gui multivariate signature scheme. The former schemes are built from the *Hidden Field Equations* crypotsystem (HFE) by using the so-called minus and vinegar modifiers. It is fair to say that HFE and its variants, are the most studied schemes in multivariate cryptography. QUARTZ produces signatures of 128 bits for a security level of 80 bits and was submitted to the *Nessie Ecrypt* competition for public-key signatures. In contrast to many multivariate schemes, no practical attack has been reported against QUARTZ. This is remarkable knowing the intense activity in the cryptanalysis of multivariate schemes.

GeMSS is a faster variant of QUARTZ that incorporates the latest results in multivariate cryptography to reach higher security levels than QUARTZ whilst improving efficiency.

6.3.3.2. DualModeMS

DualModeMS [20] is a multivariate-based signature scheme with a rather peculiar property. Its public-key is small whilst the signature is large. This is in sharp contrast with traditional multivariate signature schemes based on the so-called *Matsumoto and Imai* (MI) principle, such as QUARTZ or Gui, that produce short signatures but have larger public-keys.

DualModeMS is based on the method proposed by A. Szepieniec, W. Beullens, and B. Preneel at PQC'17 where they present a generic technique permitting to transform any (MI-based multivariate signature scheme into a new scheme with much shorter public-key but larger signatures. This technique can be viewed as a *mode of operations* that offers a new flexibility for MI-like signature schemes. Thus, we believe that *DualModeMS* could also be useful for others multivariate-based signature candidates proposed to NIST.

6.3.3.3. CPFKM

CPFKM [19] is a based on the problem of solving a system of noisy non-linear polynomials, also known as the PoSSo with Noise Problem. Our scheme largely borrows its design rationale from key encapsulation schemes based on the Learning With Errors (LWE) problem and its derivatives. The main motivation of building this scheme is to have a key exchange and encapsulation scheme based on the hardness of solving system of noisy polynomials.

6.3.4. The Point Decomposition Problem over Hyperelliptic Curves: toward efficient computations of Discrete Logarithms in even characteristic

Computing discrete logarithms is generically a difficult problem. For divisor class groups of curves defined over extension fields, a variant of the Index-Calculus called Decomposition attack is used, and it can be faster than generic approaches. In this situation, collecting the relations is done by solving multiple instances of

the Point m-Decomposition Problem (PDP_m). An instance of this problem can be modelled as a zero-dimensional polynomial system. Solving is done with Gröbner bases algorithms, where the number of solutions of the system is a good indicator for the time complexity of the solving process. For systems arising from a PDP_m context, this number grows exponentially fast with the extension degree. To achieve an efficient harvesting, this number must be reduced as much as possible. Extending the elliptic case, we introduce in [4] a notion of Summation Ideals to describe PDP_m instances over higher genus curves, and compare to Nagao's general approach to PDP_m. In even characteristic we obtain reductions of the number of solutions for both approaches, depending on the curve's equation. In the best cases, for a hyperelliptic curve of genus g, we can divide the number of solutions by $2^{(n-1)(g+1)}$. For instance, for a type II genus 2 curve defined over $\mathbb{F}_{293}$ whose divisor class group has cardinality a near-prime 184 bits integer, the number of solutions is reduced from 4096 to 64. This is enough to build the matrix of relations in around 7 days with 8000 cores using a dedicated implementation.

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.

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 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.

  `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 project is certainly the biggest industrial project ever organized in quantum-safe cryptography. RISQ is one of few projects accepted in the call Grands Défis du Numérique 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.

8.3. European Initiatives

8.3.1. FP7 & H2020 Projects

8.3.1.1. A3

Type: PEOPLE

Instrument: Career Integration Grant

Duration: May 2013 - Apr. 2017

Coordinator: Jean-Charles Faugère

Partner: Institut National de Recherche en Informatique et en Automatique (Inria), France

Inria contact: Elias Tsigaridas

Abstract: The project Algebraic Algorithms and Applications (A3) is an interdisciplinary and multidisciplinary project, with strong international synergy. It consists of four work packages The first (Algebraic Algorithms) focuses on fundamental problems of computational (real) algebraic geometry: effective zero bounds, that is estimations for the minimum distance of the roots of a polynomial system from zero, algorithms for solving polynomials and polynomial systems, derivation of non-asymptotic bounds for basic algorithms of real algebraic geometry and application of polynomial system solving techniques in optimization. We propose a novel approach that exploits structure and symmetry, combinatorial properties of high dimensional polytopes and tools from mathematical physics. Despite the great potential of the modern tools from algebraic algorithms, their use requires a combined effort to transfer this technology to specific problems. In the second package (Stochastic Games) we aim to derive optimal algorithms for computing the values of stochastic games, using techniques from real algebraic geometry, and to introduce a whole new arsenal of algebraic tools to computational game theory. The third work package (Non-linear Computational Geometry), we focus on exact computations with implicitly defined plane and space curves. These are challenging problems that commonly arise in geometric modeling and computer aided design, but they also have applications in polynomial optimization. The final work package (Efficient Implementations) describes our plans for complete, robust and efficient implementations of algebraic algorithms.

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, 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 Initiatives

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

8.4.1.1. GOAL

Title: Geometry and Optimization with ALgebraic methods.

International Partner (Institution - Laboratory - Researcher):

University of California Berkeley (United States) - Dept. of Mathematics - Bernd Sturmfels

Start year: 2015

See also: http://www-polsys.lip6.fr/GOAL/index.html

Polynomial optimization problems form a subclass of general global optimization problems, which have received a lot of attention from the research community recently; various solution techniques have been designed. One reason for the spectacular success of these methods is the potential impact in many fields: data mining, big data, energy savings, etc. More generally, many areas in mathematics, as well as applications in engineering, biology, statistics, robotics etc. require a deeper understanding of the algebraic structure of their underlying objects.

A new trend in the polynomial optimization community is the combination of algebraic and numerical methods. Understanding and characterizing the algebraic properties of the objects occurring in numerical algorithms can play an important role in improving the efficiency of exact methods. Moreover, this knowledge can be used to estimate the quality (for example the number of significant digits) of numerical algorithms. In many situations each coordinate of the optimum is an algebraic number. The degree of the minimal polynomials of these algebraic numbers is the Algebraic Degree of the problem. From a methodological point of view, this notion of Algebraic Degree emerges as an important complexity parameter for both numerical and the exact algorithms. However, algebraic systems occurring in applications often have special algebraic structures that deeply influence the geometry of the solution set. Therefore, the (true) algebraic degree could be much less than what is predicted by general worst case bounds (using Bézout bounds, mixed volume, etc.), and would be very worthwhile to understand it more precisely.

The goal of this proposal is to develop algorithms and mathematical tools to solve geometric and optimization problems through algebraic techniques. As a long-term goal, we plan to develop new software to solve these problems more efficiently. These objectives encompass the challenge of identifying instances of these problems that can be solved in polynomial time with respect to the number of solutions and modeling these problems with polynomial equations.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

- May – July 2017, Delaram Kahrobaei, Professor, CUNY, NYC, USA

8.5.1.1. Internships

- May – July 2017, Kelsey Horan, PhD student, CUNY, NYC, USA.
- Apr. – Nov. 2017, Eliane Koussa, Université de Versailles
- Apr. – Aug. 2017, Pascal Fong, Université de Versailles

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Organizing Committees

Emmanuel Prouff was a member of the organization committee of Eurocrypt 2017 (Paris, France, 2017, April 30 - May 4).

Jean-Charles Faugère and Ludovic Perret were members of the organization committee of the Quantum-Safe Cryptography for Industry (Paris, France, 2017, April 30).

9.1.2. Scientific Events Selection

9.1.2.1. Chair of Conference Program Committees

Mohab Safey El Din was PC Chair of the International Symposium on Symbolic and Algebraic Computation (ISSAC), Kaiserslautern, Germany, 2017.

Jean-Charles Faugère was PC co-Chair of the International workshop on Parallel Symbolic Computation (PASCO), Kaiserslautern, Germany, 2017.

9.1.2.2. Member of the Conference Program Committees

Ludovic Perret was a member of the program-committee of

- 20th IACR International Conference on Practice and Theory of Public-Key Cryptography (PKC'17), Amsterdam, 28-31 March 2017

Emmanuel Prouff was a member of the steering committees of the following conferences

- Conference on Cryptographic Hardware and Embedded Systems 2017 (CHES 2017) (Taipei, Taiwan, 2017, Sept. 25-28);
- Smart Card Research and Advanced Application Conference (CARDIS 2017) (Lugano, Switzerland, 2017, Nov. 13-17).

Guénaël Renault was a member of the program committee of

- 7th Int'l Conference on Mathematical Aspects of Computer and Information Sciences (MACIS) 2017;

Elias Tsigaridas was a member of the program committees of the following conferences

- 7th Int'l Conference on Mathematical Aspects of Computer and Information Sciences (MACIS) 2017;
- 19th International Workshop on Computer Algebra in Scientific Computing (CASC) 2017.

Dongming Wang was a member of the program committees of the following conferences

- 11th International Workshop on Automated Deduction in Geometry (ADG 2016) (Strasbourg, France, June 27-29, 2016);
- 8th International Symposium on Symbolic Computation in Software Science (SCSS 2017) (Gammarth, Tunisia, 2017, April 6-9).

Dongming Wang was a member of the steering committees of the following conferences

- International Conference on Mathematical Aspects of Computer and Information Sciences (MACIS),
- International Symposium on Symbolic Computation in Software Science (SCSS).

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

Ludovic Perret is an Associate Editor for:

- Designs, Codes and Cryptography (Springer, Berlin).
- The Computer Journal (Oxford University Press)
- Groups, Complexity, Cryptology (De Gruyter)

Emmanuel Prouff is an Associate Editor of the Journal of Cryptographic Engineering (Springer, Berlin).

Mohab Safey El Din is an Associate Editor of the Journal of Symbolic Computation.

Dongming Wang has the following editorial activities:

- Editor-in-Chief and Managing Editor for the journal
  Mathematics in Computer Science (published by Birkhäuser/Springer, Basel).
- Executive Associate Editor-in-Chief for the journal
  SCIENCE CHINA Information Sciences (published by Science China Press, Beijing and Springer, Berlin).
- Member of the Editorial Boards for the
  - 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),
- Member of the International Advisory Board for the Communications of JSSAC (Japan Society for Symbolic and Algebraic Computation) (published by JSSAC).

9.1.4. Invited Talks

Jean-Charles Faugère was a plenary invited speaker at the

- SIAM Conference on Applied Algebraic Geometry, Atlanta (August 2017).

Ludovic Perret was invited speaker at the

- HEXATRUST Summer school 2017 (Paris, September 2017)

Emmanuel Prouff was invited speaker at the

- Journées du GDR-IM (Montpellier, France, 2017, Mar. 14-16).
- Aix-Marseille Cyber Security Forum (AMUSEC) (Marseille, France, 2017, Oct. 12-13).

Guénaël Renault was invited speaker at the

- Third French-Japanese Meeting on Cybersecurity (Tokyo, Japan, April 2017)

Mohab Safey El Din was invited speaker at:

- The mini-symposium on Euclidean Distance Degree at the 2017 SIAM Conference on Applied Algebraic Geometry, Atlanta, USA 2017;
- The math. seminar of the University of Dortmund, Germany;
- The Berlin-Leipzig Seminar on Algebra, Geometry and Combinatorics, Germany;
- The mini-symposium on Numeric and Symbolic Convex Programming for Polynomial Optimization, at the PGMO days, Saclay, France.

Dongming Wang invited speaker at the

- 7th International Conference on Mathematical Aspects of Computer and Information Sciences (Vienna, Austria, 2017, Nov. 15-17).
- 19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (Timisoara, Romania, 2017, Sept. 21-24).
- 5th Summer School in Symbolic Computation (Nanning, China, 2017, July 16-22).

9.1.5. Scientific Expertise

Mohab Safey El Din was evaluator for the FWF International Program (Austrian funding agency).

Jean-Charles Faugère was the head of the hiring Committee for an associate professor in Grenoble.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Jérémy Berthomieu had the following teaching activities:

Master : Computation Modeling, 35 hours, M1, Université Pierre-et-Marie-Curie, France.

Master : In charge of Basics of Algebraic Algorithms, 73 hours, M1, Université Pierre-et-Marie-Curie & Polytech' UPMC, France.

Master : Introduction to Security, 20 hours, M1, Université Pierre-et-Marie-Curie, France.

Master : Projects supervision, 8 hours, M1, Université Pierre-et-Marie-Curie, France.

Licence : Introduction to Algorithmics, 40,5 hours, L2, Université Pierre-et-Marie-Curie, France.

Licence : Representations and Numerical Methods, 38,5 hours, L2, Université Pierre-et-Marie-Curie, France.

Licence : Projects supervision, 10 hours, L2, Université Pierre-et-Marie-Curie, France.

Jean-Charles Faugère had the following teaching activities:

Master: Fundamental Algorithms in Real Algebraic Geometry, 13,5 hours, M2, ENS de Lyon, France.

Master : Polynomial Systems solving, 12 hours, M2, MPRI, France.

Ludovic Perret is teaching a full service (192 hours), balanced between master and licence in cryptography, complexity and introduction to algorithms.

Mohab Safey El Din had the following teaching activities:

Master : In charge of Modeling and problems numerical and symbolic solving through MAPLE and MATLAB software, 36 hours, M1, Université Pierre-et-Marie-Curie, France

Master : In charge of Introduction to polynomial system solving, 48 hours, M2, Université Pierre-et-Marie-Curie, France

Master : In charge of the Security, Reliability and Numerical Efficiency Program in Master, 40 hours, M1 and M2, Université Pierre-et-Marie-Curie, France

Licence : Introduction to Cryptology, 20 hours, L3, Université Pierre-et-Marie-Curie, France

9.2.2. Supervision

PhD in progress : Ivan Bannwarth, Fast algorithms for studying real algebraic sets, started in Sept. 2014, Mohab Safey El Din.

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 : Eleonora Cagli, Analysis and interest points research in the attacks by observation context, Emmanuel Prouff.

PhD in progress : Loïc Masure, Recognition and Side Channel Analysis, Emmanuel Prouff.

PhD in progress, Olive Chakraborty, Design and Analysis of Post-Quantum Schemes, started in May 2017, Jean-Charles Faugère and Ludovic Perret.

PhD in progress, Nagardjun Chinthamani Dwarakanath, Design and Analysis of Fully Homomorphic Schemes, started in Dec. 2017, Jean-Charles Faugère and Ludovic Perret.

PhD in progress, Solane El Hirch Design and Analysis of Post-Quantum Schemes, started in June 2017, Jean-Charles Faugère and Ludovic Perret.

PhD in progress, Xuan Vu. Algorithms for solving structured semi-algebraic systems, started in October 2017, Jean-Charles Faugère and Mohab Safey El Din.

9.2.3. Juries

Emmanuel Prouff was examiner in the PhD committee of N. Bruneau and M. Dugardin and in the HDR committees of J.-M. Dutertre and N. El Mrabet.

Guénaël Renault was referee in the Phd committee of T. Mefenza.

9.3. Popularization

The activity of POLSYS in post-quantum cryptography has been covered in several large audience magazines:

- "Enfin! La révolution quantique", L'Usine Nouvelle, November 2017.
- "QUANTIQUE : THE NEXT BIG THING(K)", L'Informaticien, November 2017.
- "L'ORDINATEUR QUANTIQUE VA-T-IL METTRE À MAL LA CYBERSÉCURITÉ MONDIALE?", Bouygues Blog, October 2017.

Ludovic Perret is member of the Cloud Security Alliance (CSA) quantum-safe security working group. In particular, he contributed to the following documents:

- B. Huttner, J. Melia, G. Carter, L. Perret and L. Wilson. "Applied Quantum-Safe Security", Feb. 2017.
- B. Huttner, J. Melia, G. Carter, L. Perret and L. Wilson. "Quantum Safe Security Glossary", January 2017.

Ludovic Perret is also member of the quantum-safe cryptography specification group of the European Telecommunications Standards Institute (ETSI) where is the referee for a document on quantum-safe signatures.

Since May 2010, Daniel Lazard is engaged in a strong edition work on the English Wikipedia (more than 6 000 contributions, including vandalism revert and talk pages). Initially focused on the themes of POLSYS, these contributions were later enlarged to general algebra and algebraic geometry, because many elementary articles require to be expanded to be useful as a background for computer algebra. Examples of articles that have been subject of major editing: "System of polynomial equations" (created), "Computer algebra", "Algebra", "Algebraic geometry", "Polynomial greatest common divisor", "Polynomial factorization", "Finite field", "Hilbert series and Hilbert polynomial",...

For the year 2017, this contribution amounts to about $2,000$ edits on the English Wikipedia.

Mohab Safey El Din was invited by FMJH to present and popularize symbolic and algebraic computation to Master students in Mathematics following the curricula proposed by Univ. Paris-Saclay.

10. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] J. BERTHOMIEU, B. BOYER, J.-C. FAUGÈRE.*Linear Algebra for Computing Gröbner Bases of Linear Recursive Multidimensional Sequences*, in "Journal of Symbolic Computation", November 2017, vol. 83, n^{o} Supplement C, p. 36-67, Special issue on the conference ISSAC 2015: Symbolic computation and computer algebra [*DOI :* 10.1016/J.JSC.2016.11.005], https://hal.inria.fr/hal-01253934.

[2] C. E. L. DA SILVA, P. R. DA SILVA, A. JACQUEMARD.*Sliding solutions of second-order differential equations with discontinuous right-hand side*, in "Mathematical Methods in the Applied Sciences", September 2017, vol. 40, n^{o} 14, p. 5295 - 5306 [*DOI :* 10.1002/MMA.4387], https://hal-univ-bourgogne.archives-ouvertes.fr/hal-01609363.

[3] J.-C. FAUGÈRE, C. MOU.*Sparse FGLM algorithms*, in "Journal of Symbolic Computation", May 2017, vol. 80, n^{o} 3, p. 538 - 569 [*DOI :* 10.1016/J.JSC.2016.07.025], https://hal.inria.fr/hal-00807540.

[4] J.-C. FAUGÈRE, A. WALLET.*The Point Decomposition Problem over Hyperelliptic Curves: toward efficient computations of Discrete Logarithms in even characteristic*, in "Designs, Codes and Cryptography", 2017 [*DOI :* 10.1007/S10623-017-0449-Y], https://hal.inria.fr/hal-01658573.

[5] 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", 2017, p. 1-14, https://hal.inria.fr/hal-01657849.

[6] D. HENRION, S. NALDI, M. SAFEY EL DIN.*SPECTRA -a Maple library for solving linear matrix inequalities in exact arithmetic*, in "Optimization, Methods and Software", 2017, https://arxiv.org/abs/1611.01947 - Significantly extended version, https://hal.laas.fr/hal-01393022.

[7] A. HERMAN, H. HONG, E. TSIGARIDAS.*Improving Root Separation Bounds*, in "Journal of Symbolic Computation", 2017, https://hal.inria.fr/hal-01456686.

[8] V. Y. PAN, E. TSIGARIDAS.*Accelerated Approximation of the Complex Roots and Factors of a Univariate Polynomial*, in "Theoretical Computer Science", June 2017, To appear, https://hal.inria.fr/hal-01105267.

[9] V. Y. PAN, E. TSIGARIDAS.*Nearly optimal computations with structured matrices*, in "Theoretical Computer Science", June 2017, https://hal.inria.fr/hal-01105263.

[10] M. SAFEY EL DIN, É. SCHOST.*A nearly optimal algorithm for deciding connectivity queries in smooth and bounded real algebraic sets*, in "Journal of the ACM (JACM)", 2017, vol. 63, n^{o} 6, p. 48:1–48:37, https://arxiv.org/abs/1307.7836v2 - Major revision, accepted for publication to Journal of the ACM [*DOI :* 10.1145/2996450], https://hal.inria.fr/hal-00849057.

[11] M. SAFEY EL DIN, É. SCHOST.*Bit complexity for multi-homogeneous polynomial system solving Application to polynomial minimization*, in "Journal of Symbolic Computation", 2017, p. 1-32, https://arxiv.org/abs/1605.07433 [*DOI :* 10.1016/J.JSC.2017.08.001], https://hal.inria.fr/hal-01319729.

[12] 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

[13] S. BELAID, F. BENHAMOUDA, A. PASSELÈGUE, E. PROUFF, A. THILLARD, D. VERGNAUD.*Private Multiplication over Finite Fields*, in "Advances in Cryptology - CRYPTO 2017", Santa Barbara, United States, J. KATZ, H. SHACHAM (editors), Lecture Notes in Computer Science, Springer, August 2017, vol. 10403, p. 397-426 [*DOI :* 10.1007/978-3-319-63697-9_14], https://hal.inria.fr/hal-01613773.

[14] M. BURR, S. GAO, E. TSIGARIDAS.*The Complexity of an Adaptive Subdivision Method for Approximating Real Curves*, in "ISSAC 2017 - International Symposium on Symbolic and Algebraic Computation", Kaiserslautern, Germany, July 2017, 8 [*DOI :* 10.1145/3087604.3087654], https://hal.inria.fr/hal-01528392.

[15] A. MANTZAFLARIS, É. SCHOST, E. TSIGARIDAS.*Sparse Rational Univariate Representation*, in "ISSAC 2017 - International Symposium on Symbolic and Algebraic Computation", Kaiserslautern, Germany, July 2017, 8 [*DOI :* 10.1145/3087604.3087653], https://hal.inria.fr/hal-01528377.

[16] A. MANTZAFLARIS, E. TSIGARIDAS.*Resultants and Discriminants for Bivariate Tensor-product Polynomials*, in "ISSAC 2017 - International Symposium on Symbolic and Algebraic Computation", Kaiserslautern, Germany, Mohab Safey El Din, July 2017, 8 [*DOI :* 10.1145/3087604.3087646], https://hal.inria.fr/hal-01525560.

Conferences without Proceedings

[17] E. CAGLI, C. DUMAS, E. PROUFF.*Convolutional Neural Networks with Data Augmentation against Jitter-Based Countermeasures*, in "Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th International Conference", Taipei, Taiwan, September 2017, https://hal.archives-ouvertes.fr/hal-01661212.

Research Reports

[18] A. CASANOVA, J.-C. FAUGÈRE, G. MACARIO-RAT, J. PATARIN, L. PERRET, J. RYCKEGHEM.*GeMSS: A Great Multivariate Short Signature*, UPMC - Paris 6 Sorbonne Universités ; Inria Paris Research Centre, MAMBA Team, F-75012, Paris, France ; LIP6 - Laboratoire d'Informatique de Paris 6, December 2017, p. 1-4, https://hal.inria.fr/hal-01662158.

[19] O. CHAKRABORTY, J.-C. FAUGÈRE, L. PERRET.*CFPKM : A Key Encapsulation Mechanism based on Solving System of non-linear multivariate Polynomials 20171129*, UPMC - Paris 6 Sorbonne Universités ; Inria Paris ; CNRS, December 2017, https://hal.inria.fr/hal-01662175.

[20] J.-C. FAUGÈRE, L. PERRET, J. RYCKEGHEM.*DualModeMS: A Dual Mode for Multivariate-based Signature 20170918 draft*, UPMC - Paris 6 Sorbonne Universités ; Inria Paris ; CNRS, December 2017, https://hal.inria.fr/hal-01662165.

Patents and standards

[21] L. PERRET, J.-C. FAUGÈRE.*Mise en Oeuvre Optimisée du HFE*, January 2017, n^{o} WO 2017001809 A1, https://hal.inria.fr/hal-01668254.

Other Publications

[22] J. BERTHOMIEU, J.-C. FAUGÈRE.*In-depth comparison of the Berlekamp – Massey – Sakata and the Scalar-FGLM algorithms: the non adaptive variants*, May 2017, working paper or preprint, https://hal.inria.fr/hal-01516708.

[23] B. BONNARD, O. COTS, J.-C. FAUGÈRE, A. JACQUEMARD, J. ROUOT, M. SAFEY EL DIN, T. VERRON.*Algebraic-geometric techniques for the feedback classification and robustness of the optimal control of a pair of Bloch equations with application to Magnetic Resonance Imaging*, 2017, submitted, https://hal.inria.fr/hal-01556806.

[24] L. BUSÉ, A. MANTZAFLARIS, E. TSIGARIDAS.*Matrix formulae for Resultants and Discriminants of Bivariate Tensor-product Polynomials*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01654263.

[25] I. Z. EMIRIS, B. MOURRAIN, E. TSIGARIDAS.*Separation bounds for polynomial systems*, February 2017, working paper or preprint, https://hal.inria.fr/hal-01105276.

[26] D. HENRION, S. NALDI, M. SAFEY EL DIN.*Real root finding for low rank linear matrices*, October 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01159210.

[27] V. MAGRON, M. SAFEY EL DIN, M. SCHWEIGHOFER.*Algorithms for Weighted Sums of Squares Decomposition of Non-negative Univariate Polynomials*, June 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01538729.

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
A2.1.3. - Functional programming
A2.1.7. - Distributed programming
A2.1.11. - Proof languages
A2.2. - Compilation
A2.2.1. - Static analysis
A2.2.3. - Run-time systems
A2.4. - 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
B9.8. - Privacy

1. Personnel

Research Scientists

Karthikeyan Bhargavan [Team leader, Inria, Senior Researcher, HDR]
Amal Ahmed [Inria, Advanced Research Position, from Sep 2017]
David Baelde [Ecole Normale Supérieure Cachan, Researcher, until Aug 2017]
Bruno Blanchet [Inria, Senior Researcher, HDR]
Harry Halpin [Inria, Starting Research Position]
Catalin Hritcu [Inria, Researcher]

External Collaborators

David Baelde [Ecole Normale Supérieure Cachan, from Sep 2017]
Theo Laurent [Ecole Normale Supérieure Paris, from Aug 2017]
Jonathan Protzenko [Microsoft Research]

Technical Staff

Danel Ahman [Inria, from Apr 2017 until Sep 2017]
Gergely Bana [Inria, until Jan 2017, granted by FP7 ERC CIRCUS project]
Victor Dumitrescu [Inria]
Guglielmo Fachini [Inria]
Natalia Kulatova [Inria, until Oct 2017]
Tomer Libal [Inria, until Jul 2017]
Marc Sylvestre [Inria]

PhD Students

Benjamin Beurdouche [Inria]
Nadim Kobeissi [Inria]
Natalia Kulatova [Inria, from Nov 2017]
Kenji Maillard [Ecole Normale Supérieure Paris]
Marina Polubelova [Inria, from Sep 2017]
Jean-Karim Zinzindohoué [Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer]

Post-Doctoral Fellows

Danel Ahman [Inria, from Oct 2017]
Marco Stronati [Inria]

Visiting Scientists

Ana Evans [University of Virginia, from Apr 2017 until Aug 2017]
David Evans [University of Virginia, from Apr 2017 until Aug 2017]
Lucca Hirschi [Ministère de l'Enseignement Supérieur et de la Recherche, until Sep 2017]
Jake Silverman [Inria, from Jun 2017 until Aug 2017]
Aaron Weiss [Northeastern University, from Sep 2017]

Administrative Assistants

Anna Bednarik [Inria]
Helene Milome [Inria]
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 OCaml, F#, or C for execution.

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 security-oriented programming language and type system F*. 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 computationally sound type system F* for applications written in F#. 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 compilers for realistic programming languages, both low-level (the C language) and high-level (ML and F*, a dependently-typed variant). These compilers 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 compilers will 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.

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 2014) 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, the 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 dependent type system 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 [59]; it has motivated a serious research effort on the formal analysis of cryptographic protocols, starting with [57] 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* [54]. ProVerif also distinguishes itself from other tools by the variety of cryptographic primitives it can handle, defined by rewrite rules or by some equations, and the variety of security properties it can prove: secrecy [52], [43], correspondences (including authentication) [53], and observational equivalences [51]. 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 is the only tool that proves equivalences for an unbounded number of sessions.

Using ProVerif, it is now possible to verify large parts of industrial-strength protocols, such as TLS [48], JFK [44], and Web Services Security [50], 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 30 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 [55], results which were extended by formal analysis work using a Dolev-Yao style model of the standard [56]. 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# [48]. 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 such as typechecking. F7 [46] is a refinement typechecker for F#, developed jointly at Microsoft Research Cambridge and Inria. It implements a dependent type-system that allows us to specify security assumptions and goals as first-order logic annotations directly inside the program. It has been used for the modular verification of large web services security protocol implementations [49]. F* (see below) is an extension of F7 with higher-order kinds and a certifying typechecker. The cryptographic protocol implementations verified using F7 and F* already represent the largest verified cryptographic applications to our knowledge.

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 already designed the automatic tool CryptoVerif, which generates proofs by sequences of games. Much work is still needed in order to develop this approach, so that it is applicable to more protocols. We also plan to design and implement techniques for proving implementations of protocols secure in the computational model, by generating them from CryptoVerif specifications that have been proved secure, or by automatically extracting CryptoVerif models from implementations.

A different approach is to directly verify cryptographic applications in the computational model by typing. A recent work [58] 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 Designed for Program Verification

F* [60] is a verification system for ML programs developed collaboratively by Inria and Microsoft Research. ML types are extended with logical predicates that can conveniently express precise specifications for programs (pre- and post- conditions of functions as well as stateful invariants), including functional correctness and security properties. The F* typechecker implements a weakest-precondition calculus to produce first-order logic formulas that are automatically discharged using the Z3 SMT solver. The original F* implementation has been successfully used to verify nearly 50,000 lines of code, including cryptographic protocol implementations, web browser extensions, cloudhosted web applications, and key parts of the F* typechecker and compiler (itself written in F*). F* has also been used for formalizing the semantics of other languages, including JavaScript and a compiler from a subset of F* to JavaScript, and TS*, a secure subset of TypeScript. Programs verified with F* can be extracted to F#, OCaml, C, and JavaScript and then efficiently executed and integrated into larger code bases.

The latest version of F* is written entirely in F*, and bootstraps in OCaml and F#. It is open source and under active development on GitHub. A detailed description of this new F* version is available in a POPL 2016 paper [62] and a POPL 2017 one [22]. We continue to evolve and develop F* and we use it to develop large case studies of verified cryptographic applications, such as miTLS.

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 compilers for realistic programming languages, both low-level (the C language) and high-level (ML and F*, a dependently-typed variant). These compilers 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 compilers will 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 experimentally evaluate and carefully optimize the efficiency of our secure compilers on realistic workloads and standard benchmark suites. We will use property-based testing and formal verification to provide high confidence that our compilers are indeed secure. Formally, we will construct machine-checked proofs of full abstraction with respect to a secure high-level semantics. This strong property complements compiler correctness and ensures that no machine-code attacker can do more harm to securely compiled components than a component in the secure source language already could.

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 [47] [47] and TS* [61], 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 the 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.

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 such as POPL (2), IEEE S&P (2), ACM CCS (1), IEEE CSF (1), ICFP (1), PETS (1), and IEEE Euro S&P (2).
- Bruno Blanchet published a paper on the applied pi calculus in the prestigious Journal of the ACM.
- The HACL* verified cryptographic library developed in our group was integrated into Mozilla Firefox 57 and is being actively used by hundreds of millions of users around the world.
- We organized the second edition of the IEEE Euro S&P Conference in Paris, which was attended by over 200 security researchers from around the world.

5.1.1. Awards

- Karthikeyan Bhargavan, Bruno Blanchet, and Nadim Kobeissi won a Distinguished Paper award at IEEE S&P 2017 .
- Catalin Hritcu was awarded a new DARPA SSITH grant called HOPE with DRAPER Labs.
- Antoine Delignat-Lavaud received an “accessit” for the prix de thèse GDR GPL 2016.

BEST PAPERS AWARDS :
[24] **38th IEEE Symposium on Security and Privacy**. K. BHARGAVAN, B. BLANCHET, N. KOBEISSI.

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: We made several case studies using CryptoVerif (Signal, TLS 1.3 Draft 18, ARINC 823 avionic protocol) and have made a few technical improvements.

- Participants: Bruno Blanchet and David Cadé
- Contact: Bruno Blanchet
- Publications: 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 improved the display of attacks, in particular by showing the computations performed by the attacker to obtain the messages sent in the attack, and by explaining why the found trace breaks the considered security property. He also developed an interactive simulator that allows the user to run the protocol step by step. We also made several case studies using this tool (Signal, TLS 1.3 Draft 18, ARINC 823 avionic protocol).

- 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 - 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. Verification of Security Protocols in the Symbolic Model

Participants: Bruno Blanchet, Marc Sylvestre.

The applied pi calculus is a widely used language for modeling security protocols, including as a theoretical basis of PROVERIF. However, the seminal paper that describes this language [45] does not come with proofs, and detailed proofs for the results in this paper were never published. Martín Abadi, Bruno Blanchet, and Cédric Fournet wrote detailed proofs of all results of this paper. This work appears in the Journal of the ACM [12].

Marc Sylvestre improved the display of attacks in ProVerif, in particular by showing the computations performed by the attacker to obtain the messages sent in the attack, and by explaining why the found trace breaks the considered security property. He also developed an interactive simulator that allows the user to run the protocol step by step. The extended tool is available at http://proverif.inria.fr.

7.2. Symbolic and Computational Verification of Signal

Participants: Karthikeyan Bhargavan, Bruno Blanchet, Nadim Kobeissi.

We proposed a novel methodology that allows protocol designers, implementers, and security analysts to collaboratively verify a protocol using automated tools. The protocol is implemented in ProScript, a new domain-specific language that is designed for writing cryptographic protocol code that can both be executed within JavaScript programs and automatically translated to a readable model in the applied pi calculus. This model can then be analyzed symbolically using ProVerif to find attacks in a variety of threat models. The model can also be used as the basis of a computational proof using CryptoVerif, which reduces the security of the protocol to standard cryptographic assumptions. If ProVerif finds an attack, or if the CryptoVerif proof reveals a weakness, the protocol designer modifies the ProScript protocol code and regenerates the model to enable a new analysis. We demonstrated our methodology by implementing and analyzing two protocols: a variant of the popular Signal Protocol and TLS 1.3 Draft-18.

In our analysis of Signal, we used ProVerif and CryptoVerif to find new and previously-known weaknesses in the protocol and suggest practical countermeasures. Our ProScript protocol code is incorporated within the current release of Cryptocat, a desktop secure messenger application written in JavaScript. Our results indicate that, with disciplined programming and some verification expertise, the systematic analysis of complex cryptographic web applications is now becoming practical [33].

7.3. Symbolic and Computational Verification of TLS 1.3

Participants: Karthikeyan Bhargavan, Bruno Blanchet, Nadim Kobeissi.

We also applied our verification methodology to TLS 1.3, the next version of the Transport Layer Security (TLS) protocol. Its clean-slate design is a reaction both to the increasing demand for low-latency HTTPS connections and to a series of recent high-profile attacks on TLS. The hope is that a fresh protocol with modern cryptography will prevent legacy problems; the danger is that it will expose new kinds of attacks, or reintroduce old flaws that were fixed in previous versions of TLS. The protocol is nearing completion, and the working group has appealed to researchers to analyze the protocol before publication. We responded by presenting a comprehensive analysis of the TLS 1.3 Draft-18 protocol.

We seeked to answer three questions that had not been fully addressed in previous work on TLS 1.3: (1) Does TLS 1.3 prevent well-known attacks on TLS 1.2, such as Logjam or the Triple Handshake, even if it is run in parallel with TLS 1.2? (2) Can we mechanically verify the computational security of TLS 1.3 under standard (strong) assumptions on its cryptographic primitives? (3) How can we extend the guarantees of the TLS 1.3 protocol to the details of its implementations?

To answer these questions, we used our methodology for developing verified symbolic and computational models of TLS 1.3 hand-in-hand with a high-assurance reference implementation of the protocol. We presented symbolic ProVerif models for various intermediate versions of TLS 1.3 and evaluated them against a rich class of attacks to reconstruct both known and previously unpublished vulnerabilities that influenced the current design of the protocol. We presented a computational CryptoVerif model for TLS 1.3 Draft-18 and proved its security. We presented RefTLS, an interoperable implementation of TLS 1.0-1.3 in ProScript and automatically analyzed its protocol core by extracting a ProVerif model from its typed JavaScript code [24], [37]. This work was awarded the Distinguished Paper award at IEEE S&P 2017.

7.4. Verification of Avionic Security Protocols

Participant: Bruno Blanchet.

Within the ANR project AnaStaSec, we studied an air-ground avionic security protocol, the ARINC823 public key protocol [41]. We verified this protocol both in the symbolic model of cryptography, using ProVerif, and in the computational model, using CryptoVerif. While this study confirmed the main security properties of the protocol (entity and message authentication, secrecy), we found several weaknesses and imprecisions in the standard. We proposed fixes for these problems. This work appears in [27], [38].

We also verified the ATN Secure Dialogue protocol (ICAO 9880-IV [42]), which is currently under development. We verified it using ProVerif and CryptoVerif. While we confirmed the main security properties of the intended protocol, we found several incoherences, weaknesses, and imprecisions in the draft standard. We proposed fixes for these problems. We presented this work to the ICAO Secure Dialogue Subgroup (September 2017).

7.5. Design and Verification of next-generation protocols: identity, blockchains, and messaging

Participants: Harry Halpin, George Danezis [University College London], Carmela Troncoso [IMDEA].

We continued work on next-generation protocols via the NEXTLEAP project in 2017. The work started in 2016 to define the principles of design of decentralized protocols and a paper was published in the Privacy Enhancing Techologies Symposium as "Systematizing Decentralization and Privacy: Lessons from 15 years of research and deployments", which systematized over 180 papers from p2p to blockchains. We formally defined decentralization in terms of a distributed system operating in an adversarial environment, which we hope will be a foundational contribution to the field. NEXTLEAP also published a paper in ARES 2017 on how these principles can be applied to secure messaging systems, including the work of Prosecco on formalizing secure messaging as presented in EuroS&P 2017. NEXTLEAP had a successful launch event at Centre Pompidou, colocated with Eurocrypt, which was attended by a panel of prominent cryptographers (Phil Rogaway, Moti Yung, Tanja Lange, Daniel Bernstein) and members of the European Commission and European Parliament, attracting over 100 members of the general public to hear about Prosecco's research.

Building on the work on identity started in 2017, we finished the design of ClaimChain, the privacy-enhanced blockchain-based identity system, and work started on a F* implementation and scalability simulations. Unlike most blockchain systems that are public and are essentially replicated state machines, Claimchains use VRFs for privacy and do not require global consensus, instead allowing private linking between Claimchains and gossiping to maintain local consensus on secret material. We believe that this design may be the first workable approach to decentralizing PKI. Claimchains also use Merkle Trees for efficiency, and some of this library may end up as generally useful for F* programming after more development in 2018. Claimchain has yet to be published in an academic venue, but it has already attracted considerable interest and was presented in the popular CCC security conference in Leipzig Germany. We also continued to raise the bar on security and privacy, hosting the first ever workshop on "Security and Privacy on the Blockchain" at EuroS&P 2017, which was sponsored by Blockstream. We expect the first formally verified blockchain system based on this design to be finished in 2018.

Another aspect of building next-generation protocols is to evaluate their usability. Prior studies have shown that users typically do not understand encryption and are even hostile to open-source code. However, these studies are typically done with students drawn for a general population, and in response Prosecco, in co-operation with sociologists from CNRS/Sorbonne, have started the largest-ever study of high-risk users from countries as diverse as Ukraine, Russia, Egypt and Tunisia. Preliminary results were presented at the European Usable Security (EuroUSEC) workshop, and already have attracted considerable attention from developers of secure messaging applications such as Signal and Briar. We hope that our findings on how users actually do group messaging and key verification will lead to changes in the underlying protocols.

Lastly, we continue to work with standards bodies in order to do security and privacy analysis of new protocols. For example, we have started formalizing W3C Web Authentication and inspecting its privacy properties, and our work on the lack of security in Semantic Web standards led to "Semantic Insecurity: Security and the Semantic Web" at ISWC 2017. Work on the security and privacy properties of the W3C Encrypted Media Extension led to an invited keynote at SPACE 2017.

Next year, we will finalize ClaimChain and add on the mix-network we have been developing over the last year, leading to a metadata-resistant and decentralized secure messaging application. We will work on spreading awareness of the importance of formally verified open standards as being necessary for the future of security, rather than closed-source solutions that may have backdoors and dangerous bugs that could cause severe

economic damage if not fixed. To this end, we will work with ECRYPT CSA on the IACR Summer School of Societal and Business Impact of Cryptography, colocated with Real-World Crypto 2018, and co-organize an event at the European Commission and Parliament.

7.6. The F* programming language

Participants: Danel Ahman, Benjamin Beurdouche, Karthikeyan Bhargavan, Barry Bond [Microsoft Research], Tej Chajed [MIT], Antoine Delignat-Lavaud [Microsoft Research], Victor Dumitrescu, Cédric Fournet [Microsoft Research], Catalin Hritcu, Qunyan Mangus [Microsoft Research], Markulf Kohlweiss [Microsoft Research], Kenji Maillard, Asher Manning [McGill University], Guido Martínez [CIFASIS-CONICET Rosario], Zoe Paraskevopoulou [Princeton University], Clément Pit-Claudel [MIT], Jonathan Protzenko [Microsoft Research], Tahina Ramananandro [Microsoft Research], Aseem Rastogi [Microsoft Research], Jared Roesch [University of Washington], Nikhil Swamy [Microsoft Research], Christoph M. Wintersteiger [Microsoft Research], Santiago Zanella-Béguelin [Microsoft Research].

F* 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 OCaml, F#, or C for execution.

The latest version of F* is written entirely in F*, and bootstraps in OCaml and F#. It is open source and under active development on http://github.com/FStarLang/FStar. A detailed description of this new F* version is available in a series of POPL papers [62], [22], [14].

The main ongoing use case of F* is building a verified, drop-in replacement for the whole HTTPS stack in Project Everest [25]. This includes verified implementations of TLS 1.2 and 1.3 including the underlying cryptographic primitives. Moreover, while F* is extracted to OCaml by default, we have devised a subset of F* that can be compiled to C for efficiency [18].

We released two versions of the software this year.

7.7. Micro-Policies

Participants: Arthur Azevedo de Amorim [University of Pennsylvania], Chris Casinghino [Draper Labs], André Dehon [University of Pennsylvania], Catalin Hritcu, Théo Laurent [ENS Paris], Benjamin Pierce [University of Pennsylvania], Howard Shrobe [MIT], Greg Sullivan [Dover Microsystems], Andrew Tolmach [Portland State University].

This year we obtained a new DARPA grant called SSITH/HOPE on “Advanced New Hardware Optimized for Policy Enforcement, A New HOPE”. This grant is in the process of starting and our contribution will focus on devising a high-level micro-policy language and investigating micro-policies targetting today’s most severe security vulnerabilities.

7.8. HACL*: A Verified Modern Cryptographic Library

Participants: Jean Karim Zinzindohoue, Karthikeyan Bhargavan, Jonathan Protzenko [Microsoft Research], Benjamin Beurdouche.

HACL* is a verified portable C cryptographic library that implements modern cryptographic primitives such as the ChaCha20 and Salsa20 encryption algorithms, Poly1305 and HMAC message authentication, SHA-256 and SHA-512 hash functions, the Curve25519 elliptic curve, and Ed25519 signatures.

HACL* is written in the F* programming language and then compiled to readable C code using the KreMLin tool [18]. The F* source code for each cryptographic primitive is verified for memory safety, mitigations against timing side-channels, and functional correctness with respect to a succinct high-level specification of the primitive derived from its published standard. The translation from F* to C preserves these properties and the generated C code can itself be compiled via the CompCert verified C compiler or mainstream compilers like GCC or CLANG. When compiled with GCC on 64-bit platforms, our primitives are as fast as the fastest pure C implementations in OpenSSL and Libsodium, significantly faster than the reference C code in TweetNaCl, and between 1.1x-5.7x slower than the fastest hand-optimized vectorized assembly code in the SUPERCOP benchmark test-suite.

HACL* implements the NaCl cryptographic API and can be used as a drop-in replacement for NaCl libraries like Libsodium and TweetNaCl. HACL* provides the cryptographic components for a new mandatory ciphersuite in TLS 1.3 and is being developed as the main cryptographic provider for the miTLS verified implementation. Primitives from HACL* have now been integrated within Mozilla's NSS cryptographic library. Our results show that writing fast, verified, and usable C cryptographic libraries is now practical.

This work appeared at the ACM CCS conference [36] and all our software is publicly available and in active development on GitHub.

7.9. miTLS: A Verified TLS Implementation

Participants: Karthikeyan Bhargavan, Antoine Delignat-Lavaud [Microsoft Research], Cédric Fournet [Microsoft Research], Markulf Kohlweiss [Microsoft Research], Jianyang Pan, Jonathan Protzenko [Microsoft Research], Aseem Rastogi [Microsoft Research], Nikhil Swamy [Microsoft Research], Santiago Zanella-Béguelin [Microsoft Research], Jean Karim Zinzindohoue.

The record layer is the main bridge between TLS applications and internal sub-protocols. Its core functionality is an elaborate authenticated encryption: streams of messages for each sub-protocol (handshake, alert, and application data) are fragmented, multiplexed, and encrypted with optional padding to hide their lengths. Conversely, the sub-protocols may provide fresh keys or signal stream termination to the record layer.

Compared to prior versions, TLS 1.3 discards obsolete schemes in favor of a common construction for Authenticated Encryption with Associated Data (AEAD), instantiated with algorithms such as AES-GCM and ChaCha20-Poly1305. It differs from TLS 1.2 in its use of padding, associated data and nonces. It encrypts the content-type used to multiplex between sub-protocols. New protocol features such as early application data (0-RTT and 0.5-RTT) and late handshake messages require additional keys and a more general model of stateful encryption.

As part of the miTLS project, we built and verified a reference implementation of the TLS record layer and its cryptographic algorithms in F*. We reduced the high-level security of the record layer to cryptographic assumptions on its ciphers. Each step in the reduction is verified by typing an F* module; when the step incurs a security loss, this module precisely captures the corresponding game-based security assumption.

We computed concrete security bounds for the AES-GCM and ChaCha20-Poly1305 ciphersuites, and derived recommended limits on sent data before re-keying. Combining our functional correctness and security results, we obtained the first verified implementation of the main TLS 1.3 record ciphers. We plugged our implementation into an existing TLS library and confirmed that the combination interoperates with Chrome and Firefox, and thus that experimentally the new TLS record layer (as described in RFCs and cryptographic standards) is provably secure.

This work appeared at IEEE S&P 2017 [26] and our verified software is publicly available and actively developed on GitHub.

7.10. A Cryptographic Analysis of Content Delivery of TLS

Participants: Karthikeyan Bhargavan, Ioana Boureanu [University of Surrey], Pierre-Alain Fouque [University of Rennes 1/IRISA], Cristina Onete [University of Rennes 1/IRISA], Benjamin Richard [Orange Labs Chatillon].

The Transport Layer Security (TLS) protocol is designed to allow two parties, a client and a server, to communicate securely over an insecure network. However, when TLS connections are proxied through an intermediate middlebox, like a Content Delivery Network (CDN), the standard end-to-end security guarantees of the protocol no longer apply.

As part of the SafeTLS project, we investigated the security guarantees provided by Keyless SSL, a CDN architecture currently deployed by CloudFlare that composes two TLS 1.2 handshakes to obtain a proxied TLS connection. We demonstrated new attacks that show that Keyless SSL does not meet its intended security goals. We argued that proxied TLS handshakes require a new, stronger, 3-party security definition, and we presented one.

We modified Keyless SSL and proved that our modifications guarantee this notion of security. Notably, we showed that secure proxying in TLS 1.3 is computationally lighter and requires simpler assumptions on the certificate infrastructure than our proposed fix for Keyless SSL. Our results indicate that proxied TLS architectures, as currently used by a number of CDNs, may be vulnerable to subtle attacks and deserve close attention [39].

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/Antique, Inria/Celtique, Airbus Operations SAS, AMOSSYS, CEA-LIST, TrustInSoft

Duration: January 2015 - December 2018.

Coordinator: Jérôme Féret, Inria Antique (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, Univesité 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.

8.2. 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 Bhargavn, Inria

Abstract: 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: This new ERC-funded project called SECOMP1 is aimed at leveraging emerging hardware capabilities for fine-grained protection to build the first, efficient secure compilers for realistic programming languages, both low-level (the C language) and high-level (F*, a dependently-typed ML variant). These compilers 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 compilers will 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

Programme: 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 can 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 Labs

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:

- **F***: Microsoft Research (Cambdridge, Redmond), IMDEA (Madrid)
- **TLS analysis**: Microsoft Research (Cambridge), Mozilla, University of Rennes
- **Web Security**: Microsoft Research (Cambridge, Redmond), Imperial College (London), University of Stuttgart
- **Micro-Policies**: University of Pennsylvania, Portland State University

8.3.2. Participation in Other International Programs

8.3.2.1. International Initiatives

Title: Advanced New Hardware Optimized for Policy Enforcement, A New HOPE

Program: DARPA SSITH

Duration: January 2016 - December 2018

Coordinator: Charles Stark, Draper Laboratory

Participants: Catalin Hritcu

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.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

- Claudia Diaz from KUL visited the group from 1-2 March and gave a seminar "Designing Mix-nets"
- Peter Schwabe visited Inria Paris on 11 April; he gave a seminar: From NewHope to Kyber.
- Joseph Bonneau (Stanford University) visited Inria on 20 April 2017, he gave a seminar: Public randomness, blockchains and proofs-of-delay
- Stefan Ciobaca (Alexandru Ioan Cuza University of Iai, Romania) visited Inria Paris on 15 May 2017; he gave a seminar: The RMT Tool for Rewriting Modulo Theories.
- Ana Nora Evans (University of Virginia) joined Inria as a Visiting Scientist Apr–Aug 2017; she gave a seminar: Using Verified Software Fault Isolation for a Formally Secure Compiler.
- David Evans (University of Virginia) joined Inria as a Visiting Scientist Apr–Aug 2017; he gave a seminar: Can Machine Learning Work in the Presence of Adversaries?
- Jean Yang (CMU) visited Inria Paris on 6 June 2017; she gave a seminar: Policy-Agnostic Programming for Database-Backed Applications.
- Amal Ahmed (Northeastern University) joined Inria as a Visiting Professor from September 2017; she gave a seminar: Prosecco Seminars: Compositional Compiler Verification for a Multi-Language World.
- Aaron Weiss (Northeastern University) joined Inria as a Visiting Scientist from September 2017.
- Amin Timany (KU Leuven) visited Inria Paris 6-8 December 2017; he gave a seminar: A Logical Relation for Monadic Encapsulation of State: Proving contextual equivalences in the presence of runST.
- Eric Rescorla visited Prosecco to discuss the design of TLS 1.3.

8.4.1.1. Internships

- Benjamin Lipp: Dec 2017 until May 2018, supervised by B. Blanchet, K. Bhargavan, and H. Halpin
- Iness Ben Guirat: Masters student 2017, supervised by H. Halpin
- Carmine Abate (University of Trento): Dec 2017 until May 2018
- William Bowman (Northeastern University): Oct 2017 until Dec 2017
- Keith Cannon (American University Paris): Mar 2017 until Sep 2017
- Théo Laurent (ENS Paris): Mar 2017 until Aug 2017
- Clément Pit-Claudel (MIT): Jul 2017 until Oct 2017

8.4.2. Visits to International Teams

- Catalin Hritcu, October 8-13, 2017, Aarhus University, Denmark.
- Catalin Hritcu, October 16-17, 2017, MPI-SWS, Saarbrucken, Germany.
- Catalin Hritcu, December 18, 2017, University of Iasi, Romania.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. General Chair, Scientific Chair

- Prosecco organized the 2nd IEEE European Symposium on Security and Privacy in Paris, 26-28 April 2017. Catalin Hritcu was General Chair, Bruno Blanchet was Finance Chair, and Karthikeyan Bhargavan was Local arrangements Chair.
- Harry Halpin co-chaired the IEEE Security and Privacy on the Blockchain workshop, colocated with IEEE EuroS&P, on 29 April 2017.
- Catalin Hritcu is Artifact Evaluation Co-Chair of POPL 2018
- Catalin Hritcu created a New Workshop on Principles of Secure Compilation (PriSC) colocated with POPL 2017 and 2018. He is PC Chair for PriSC 2018.
- Prosecco organized a Project Everest Workshop at Inria Paris, 2 October 2017
- Prosecco organized an ESOP PC workshop at Inria Paris, 15 December 2017 Workshop at POPL: 13 January 2018, Los Angeles, USA

9.1.2. Scientific Events Selection

9.1.2.1. Member of the Conference Program Committees

- Bruno Blanchet was PC member at TAP 2017.
- Harry Halpin was a PC member for ISWC 2017 and WWW 2017.
- Catalin Hritcu was PC member at ESOP 2018 and EuroS&P 2018
- Karthikeyan Bhargavan was a PC member at ACM CCS 2017-18, IEEE S&P 2017-18, POST 2018.

9.1.2.2. Reviewer

- Harry Halpin served as a reviewer for LatinCrypt, AsiaCrypt, JAIST, TCS

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

- Bruno Blanchet gave an invited talk at the workshop on Models and Tools for Security Analysis and Proofs, 2017.
- Bruno Blanchet gave an invited talk at the workshop TLS:DIV (TLS 1.3: Design, Implementation & Verification), 2017.
- Bruno Blanchet gave an invited talk at the workshop TMSP (Trends in Mechanized Security Proofs), 2017.
- Bruno Blanchet gave an invited talk at the Summer Research Institute, EPFL, 2017.
- Harry Halpin gave an invited talk at SPACE 2017
- Harry Halpin gave an invited talk at Conference on Privacy and Data Protection, January 2017.
- Harry Halpin gave an invited talk at RightsCon, March 2017.
- Harry Halpin gave an invited talk at E-CRYPT Cryptosympsium, March 2017.
- Harry Halpin gave an invited talk at La Firma Digital, July 2017.
- Harry Halpin gave an invited talk at Google, October 2017.
- Harry Halpin gave an invited talk at IMMWorld, November 2017.
- Harry Halpin gave an invited talk at Boston University Law School, November 2017.
- Harry Halpin gave an invited talk at University of North Carolina-Chapel Hill, December 2017.
- Harry Halpin gave a keynote talk at Security, and Privacy, and Cryptographic Engineering, December 2017.

- Catalin Hritcu was an invited speaker at TFP 2017
- Catalin Hritcu gave talks at Infoiasi, ESOP PC Workshop, Everest Workshop, TFP (Keynote), FADEx 2017, EuroS&P 2017, Université Clermont Auvergne, University Paris-Sud.
- Karthikeyan Bhargavan gave a keynote at ACNS 2017, Kanazawa, Japan.
- Karthikeyan Bhargavan gave an invited talk at Apple, Cupertino, USA.

9.1.5. 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.
- Karthikeyan Bhargavan advises the TLS working group at the IETF and consults for Mozilla, Apple, and Microsoft Research.
- Catalin Hritcu consilts for Microsoft Research and the DARPA SSITH/HOPE grant.

9.1.6. Research Administration

- Bruno Blanchet is a member of the Inria hiring committee for PhD, post-docs, and *délégations* (*Commision des Emplois Scientifiques*, CES).

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

- Master: Catalin Hritcu, Cryptographic protocols: formal and computational proofs, 31.5h equivalent TD, master M2 MPRI, université Paris VII, France
- Doctorat: Catalin Hritcu: Verifying Cryptographic Implementations with F* at Computer-aided security proofs summer school. Aarhus, Denmark, October, 2017
- Doctorat: Catalin Hritcu: Verifying Cryptographic Implementations with F* course at Models and Tools for Cryptographic Proofs summer school, Nancy, France, July 2017
- Master: Karthikeyan Bhargavan, Cryptographic protocols: formal and computational proofs, 31.5h equivalent TD, master M2 MPRI, université Paris VII, France
- Master: Karthikeyan Bhargavan, Protocol Verification and Safety, 18h equivalent TD, master ACN, Ecole Polytechnique et Telecom ParisTech, France

9.2.2. Supervision

- PhD: Evmorfia-Iro Bartzia, *A formalization of elliptic curves for cryptography*, Université Paris-Saclay, February 2017. Co-supervised by Pierre-Yves Strub and Karthikeyan Bhargavan.
- PhD in progress: Kenji Maillard, *Semantic Foundations for F**, started January 2017, supervised by Catalin Hritcu and Karthikeyan Bhargavan
- PhD in progress: Jean Karim Zinzindohoue, *A Verified Cryptographic Libary*, supervised by Karthikeyan Bhargavan
- PhD in progress: Nadim Kobeissi, 2015-, *Verified Web Security Applicaitons*, supervised by Karthikeyan Bhargavan
- PhD in progress: Benjamin Beurdouche, 2016-, *Verified Cryptographic Protocols for the Internet of Things*, supervised by Karthikeyan Bhargavan
- PhD in progress: Natalia Kulatova, 2017-, *Verified Hardware Security Devices*, co-supervised by Karthikeyan Bhargavan and Graham Steel
- PhD in progress: Marina Polybelova, 2017-, *Verified Cryptographic Web Applications*, supervised by Karthikeyan Bhargavan

- PhD in progress: Yaëlle Vincont, 2017-, *Software Security: combining fuzzing and symbolic execution for vulnerability detection*, co-supervised by Karthikeyan Bhargavan and Sebastien Bardin

9.2.3. Juries

- Bruno Blanchet was reviewer of Lucca Hirschi's PhD thesis.
- Harry Halpin served on the PhD jury of Evo Busseniers (Vrije Universitat Bruxelles)

9.3. Popularization

- Karthikeyan Bhargavan, Benjamin Beurdouche, Jean Karim Zinzindohoue published a paper in the Communications of the ACM.

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] 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)", Paris, France, ACM, 2017, p. 515-529, https://arxiv.org/abs/1608.06499 [*DOI :* 10.1145/3009837.3009878], https://hal.archives-ouvertes.fr/hal-01424794.

[3] R. BARDOU, R. FOCARDI, Y. KAWAMOTO, L. SIMIONATO, G. STEEL, J.-K. TSAY.*Efficient Padding Oracle Attacks on Cryptographic Hardware*, in "CRYPTO", 2012, p. 608–625.

[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.

[6] B. BLANCHET.*A Computationally Sound Mechanized Prover for Security Protocols*, in "IEEE Transactions on Dependable and Secure Computing", 2008, vol. 5, n° 4, p. 193–207, Special issue IEEE Symposium on Security and Privacy 2006.

[7] 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.

[8] C. HRITCU, M. GREENBERG, B. KAREL, B. C. PIERCE, G. MORRISETT.*All Your IFCException Are Belong to Us*, in "IEEE Symposium on Security and Privacy (Oakland)", 2013, p. 3–17.

[9] 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, http://doi.acm.org/10.1145/2994620.2994637.

[10] Y. JUGLARET, C. HRITCU, A. AZEVEDO DE AMORIM, B. ENG, B. C. PIERCE.*Beyond Good and Evil: Formalizing the Security Guarantees of Compartmentalizing Compilation*, in "29th IEEE Symposium on Computer Security Foundations (CSF)", IEEE Computer Society Press, July 2016, p. 45–60 [*DOI :* 10.1109/CSF.2016.11], http://arxiv.org/abs/1602.04503.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] E.-I. BARTZIA.*A formalization of elliptic curves for cryptography*, Université Paris-Saclay, February 2017, https://pastel.archives-ouvertes.fr/tel-01563979.

Articles in International Peer-Reviewed Journal

[12] 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^{o} 1, p. 1 - 103 [*DOI :* 10.1145/3127586], https://hal.inria.fr/hal-01636616.

[13] 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.

[14] 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.

[15] K. BHARGAVAN, B. BEURDOUCHE, A. DELIGNAT-LAVAUD, C. FOURNET, M. KOHLWEISS, A. PIRONTI, P.-Y. STRUB, J. K. ZINZINDOHOUE.*A messy state of the union*, in "Communications of the ACM", January 2017, vol. 60, n^{o} 2, p. 99 - 107 [*DOI :* 10.1145/3023357], https://hal.inria.fr/hal-01673714.

[16] W. J. BOWMAN, Y. CONG, N. RIOUX, A. AHMED.*Type-Preserving CPS Translation of* Σ *and* Π *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.

[17] 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", 2017, https://arxiv.org/abs/1711.03050 [*DOI :* 10.1145/3158137], https://hal.inria.fr/hal-01646765.

[18] J. PROTZENKO, J. ZINZINDOHOUÉ, A. RASTOGI, T. RAMANANANDRO, P. WANG, S. ZANELLA-BÉGUELIN, A. DELIGNAT-LAVAUD, C. HRIŢCU, K. BHARGAVAN, C. FOURNET, N. SWAMY.*Verified Low-Level Programming Embedded in F**, in "Proceedings of the ACM on Programming Languages", September 2017, vol. 1, n^{o} ICFP, p. 17:1–17:29, https://arxiv.org/abs/1703.00053 [*DOI :* 10.1145/3110261], https://hal.archives-ouvertes.fr/hal-01672706.

[19] C. TRONCOSO, M. ISAAKIDIS, G. DANEZIS, H. HALPIN.*Systematizing Decentralization and Privacy: Lessons from 15 Years of Research and Deployments*, in "Proceedings on Privacy Enhancing Technologies", October 2017, vol. 2017, n^{o} 4, p. 307 - 329 [*DOI :* 10.1515/POPETS-2017-0056], https://hal.inria.fr/hal-01673295.

Invited Conferences

[20] H. HALPIN.*The Crisis of Standardizing DRM: The Case of W3C Encrypted Media Extensions*, in "SPACE 2017 - Seventh International Conference on Security, Privacy, and Applied Cryptography Engineering", Goa, India, Lecture Notes in Computer Science, Springer, December 2017, vol. 10662, p. 10-29 [*DOI :* 10.1007/978-3-319-71501-8_2], https://hal.inria.fr/hal-01673296.

[21] G. LEURENT, K. BHARGAVAN.*On the Practical (In-)Security of 64-bit Block Ciphers*, in "ESC 2017 - Early Symmetric Crypto", Canach, Luxembourg, January 2017, https://hal.inria.fr/hal-01105128.

International Conferences with Proceedings

[22] 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)", Paris, France, ACM, 2017, p. 515-529, https://arxiv.org/abs/1608.06499 [*DOI :* 10.1145/3009837.3009878], https://hal.archives-ouvertes.fr/hal-01424794.

[23] D. AHMAN, T. UUSTALU.*Taking Updates Seriously*, in "Proceedings of the 6th International Workshop on Bidirectional Transformations co-located with The European Joint Conferences on Theory and Practice of Software - ETAPS 2017", Uppsala, Sweden, April 2017, p. 59–73, https://hal.archives-ouvertes.fr/hal-01672736.

[24] *Best Paper*
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.

[25] K. BHARGAVAN, B. BOND, A. DELIGNAT-LAVAUD, C. FOURNET, C. HAWBLITZEL, C. HRITCU, S. ISHTIAQ, M. KOHLWEISS, R. LEINO, J. LORCH, K. MAILLARD, J. PAIN, 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)", Asilomar, CA, United States, May 2017 [*DOI :* 10.4230/LIPIcs.SNAPL.2017.1], https://hal.archives-ouvertes.fr/hal-01672707.

[26] 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 "SP 2017 - 38th IEEE Symposium on Security and Privacy", San Jose, United States, May 2017, p. 463-482 [*DOI :* 10.1109/SP.2017.58], https://hal.inria.fr/hal-01674096.

[27] B. BLANCHET.*Symbolic and Computational Mechanized Verification of the ARINC823 Avionic Protocols*, in "30th IEEE Computer Security Foundations Symposium", Santa Barbara, United States, August 2017, p. 68-82 [*DOI :* 10.1109/CSF.2017.7], https://hal.inria.fr/hal-01575861.

[28] K. CAIRNS, H. HALPIN, G. STEEL.*Security Analysis of the W3C Web Cryptography API*, in "Proceedings of Security Standardisation Research (SSR)", Gaithersberg, United States, Lecture Notes in Computer Science (LNCS), Springer, December 2017, vol. 10074, p. 112 - 140 [*DOI :* 10.1007/978-3-319-49100-4_5], https://hal.inria.fr/hal-01426852.

[29] 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.

[30] H. HALPIN.*A Roadmap for High Assurance Cryptography*, in "FPS 2017 - 10th International Symposium on Foundations & Practice of Security", Nancy, France, October 2017, p. 1-9, https://hal.inria.fr/hal-01673294.

[31] H. HALPIN.*NEXTLEAP: Decentralizing Identity with Privacy for Secure Messaging*, in "ARES 2017 - 12th International Conference on Availability, Reliability and Security", Reggio Calabria, Italy, ACM, August 2017, p. 1-10 [*DOI :* 10.1145/3098954.3104056], https://hal.inria.fr/hal-01673292.

[32] H. HALPIN.*Semantic Insecurity: Security and the Semantic Web*, in "Society, Privacy and the Semantic Web - Policy and Technology (PrivOn 2017)", Vienna, Austria, October 2017, https://hal.inria.fr/hal-01673291.

[33] 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.

[34] N. KOBEISSI, K. BHARGAVAN, B. BLANCHET.*Automated Verification for Secure Messaging Protocols and Their Implementations: A Symbolic and Computational Approach*, in "EuroS&P 2017 - 2nd IEEE European Symposium on Security and Privacy", Paris, France, A. SABELFELD, M. SMITH (editors), IEEE, April 2017, p. 435 - 450 [*DOI :* 10.1109/EUROSP.2017.38], https://hal.inria.fr/hal-01583009.

[35] L. LAMPROPOULOS, D. GALLOIS-WONG, C. HRIŢCU, J. HUGHES, B. C. PIERCE, L.-Y. XIA.*Beginner's Luck: A Language for Random Generators*, in "44th ACM SIGPLAN Symposium on Principles of Programming Languages (POPL)", Paris, France, ACM, 2017, p. 114-129, https://arxiv.org/abs/1607.05443 [*DOI :* 10.1145/3009837.3009868], https://hal.archives-ouvertes.fr/hal-01424793.

[36] J.-K. ZINZINDOHOUÉ, K. BHARGAVAN, J. PROTZENKO, B. BEURDOUCHE.*HACL * : A Verified Modern Cryptographic Library*, in "ACM Conference on Computer and Communications Security (CCS)", Dallas, United States, October 2017, https://hal.inria.fr/hal-01588421.

Research Reports

[37] K. BHARGAVAN, B. BLANCHET, N. KOBEISSI.*Verified Models and Reference Implementations for the TLS 1.3 Standard Candidate*, Inria Paris, May 2017, n° RR-9040, 51, https://hal.inria.fr/hal-01528752.

[38] B. BLANCHET.*Symbolic and Computational Mechanized Verification of the ARINC823 Avionic Protocols*, Inria Paris, May 2017, n° RR-9072, 40, https://hal.inria.fr/hal-01527671.

Other Publications

[39] K. BHARGAVAN, I. BOUREANU, C. ONETE, P.-A. FOUQUE, B. RICHARD.*Content delivery over TLS: a cryptographic analysis of keyless SSL*, IEEE, April 2017, p. 600-615, EuroS&P 2017 - 2nd IEEE European Symposium on Security and Privacy [*DOI :* 10.1109/EUROSP.2017.52], https://hal.inria.fr/hal-01673853.

[40] H. HALPIN, M. PIEKARSKA.*Introduction to Security and Privacy on the Blockchain*, April 2017, 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&P Workshops 2017), https://hal.inria.fr/hal-01673293.

References in notes

[41] *ARINC SPECIFICATION 823P1: DATALINK SECURITY, PART 1 â ACARS MESSAGE SECURITY*, December 2007.

[42] *ICAO Doc 9880: Manual on Detailed Technical Specifications for the Aeronautical Telecommunication Network (ATN) using ISO/OSI Standards and Protocols, Part IV B — Security Services, Third edition (Proposed Draft)*, May 2017.

[43] 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.

[44] 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.

[45] M. ABADI, C. FOURNET.*Mobile Values, New Names, and Secure Communication*, in "28th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL'01)", London, United Kingdom, ACM Press, January 2001, p. 104–115.

[46] J. BENGTSON, K. BHARGAVAN, C. FOURNET, A. D. GORDON, S. MAFFEIS.*Refinement types for secure implementations*, in "ACM Trans. Program. Lang. Syst.", 2011, vol. 33, n^{o} 2, 8.

[47] K. BHARGAVAN, A. DELIGNAT-LAVAUD, S. MAFFEIS.*Language-Based Defenses Against Untrusted Browser Origins*, in "Proceedings of the 22th USENIX Security Symposium", 2013.

[48] 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, 3:1.

[49] K. BHARGAVAN, C. FOURNET, A. D. GORDON.*Modular Verification of Security Protocol Code by Typing*, in "ACM Symposium on Principles of Programming Languages (POPL'10)", 2010, p. 445–456.

[50] K. BHARGAVAN, C. FOURNET, A. D. GORDON, N. SWAMY.*Verified Implementations of the Information Card Federated Identity-Management Protocol*, in "Proceedings of the ACM Symposium on Information, Computer and Communications Security (ASIACCS'08)", ACM Press, 2008, p. 123–135.

[51] 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.

[52] B. BLANCHET.*An Efficient Cryptographic Protocol Verifier Based on Prolog Rules*, in "14th IEEE Computer Security Foundations Workshop (CSFW'01)", 2001, p. 82–96.

[53] B. BLANCHET.*Automatic Verification of Correspondences for Security Protocols*, in "Journal of Computer Security", July 2009, vol. 17, n^{o} 4, p. 363–434.

[54] 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.

[55] J. CLULOW.*On the Security of PKCS#11*, in "CHES", 2003, p. 411-425.

[56] S. DELAUNE, S. KREMER, G. STEEL.*Formal Analysis of PKCS#11 and Proprietary Extensions*, in "Journal of Computer Security", November 2010, vol. 18, n^{o} 6, p. 1211-1245.

[57] D. DOLEV, A. YAO.*On the security of public key protocols*, in "IEEE Transactions on Information Theory", 1983, vol. IT–29, n^{o} 2, p. 198–208.

[58] C. FOURNET, M. KOHLWEISS, P.-Y. STRUB.*Modular Code-Based Cryptographic Verification*, in "ACM Conference on Computer and Communications Security", 2011.

[59] R. NEEDHAM, M. SCHROEDER.*Using encryption for authentication in large networks of computers*, in "Communications of the ACM", 1978, vol. 21, n^{o} 12, p. 993–999.

[60] N. SWAMY, J. CHEN, C. FOURNET, P.-Y. STRUB, K. BHARGAVAN, J. YANG.*Secure distributed programming with value-dependent types*, in "16th ACM SIGPLAN international conference on Functional Programming", 2011, p. 266-278.

[61] 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.

[62] N. SWAMY, C. HRIŢCU, C. KELLER, A. RASTOGI, A. DELIGNAT-LAVAUD, S. FOREST, K. BHARGAVAN, C. FOURNET, P.-Y. STRUB, M. KOHLWEISS, J.-K. ZINZINDOHOUE, S. ZANELLA-BÉGUELIN.*Dependent Types and Multi-Monadic Effects in F**, in "43rd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL)", Unknown, United States, ACM, 2016, p. 256-270, https://hal.archives-ouvertes.fr/hal-01265793.

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

Université Pierre et Marie Curie (Paris 6)

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. - Mathematical Modeling
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling (SPDE, SDE)
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
B9.8. - Privacy

1. Personnel

Research Scientists

Mazyar Mirrahimi [Team Leader, Inria, Senior Researcher, HDR]
Alain Sarlette [Inria, Researcher]

Faculty Members

Mallet François [Université Pierre et Marie Curie, Associate professor]
Zaki Leghtas [Ecole Nationale Supérieure des Mines de Paris, Associate professor]
Pierre Rouchon [Ecole Nationale Supérieure des Mines de Paris, Professor, HDR]

External Collaborator

Benjamin Huard [ENS Lyon, Professor, HDR]

PhD Students

Remi Azouit [Ecole Nationale Supérieure des Mines de Paris, PhD Student, until Oct 2017]
Gerardo Cardona Sanchez [Ecole Nationale Supérieure des Mines de Paris, PhD Student]
Lucas Verney [ENS Paris, PhD Student]
Lescanne Raphael [ENS Paris]
Jeremie Guillaud [Inria, from Sep 2017]

Post-Doctoral Fellows

Zibo Miao [Inria, Post-Doctoral Fellow]
Paolo Forni [Ecole Nationale Supérieure des Mines de Paris, from Oct 2017]

Administrative Assistant

Martine Verneuille [Inria, Assistant]

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 [89], [93]. 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 [66]. 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 [63], [57] 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 [65]. Through a recent experimental work [98], 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 [94] has recently led to a first experimental realization of a full quantum error correcting code improving the coherence time of quantum information [6]. 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 the 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 system's state. The concept of *quantum non-demolotion* (QND) measurement has played a crucial role in understanding and resolving this difficulty [40]. In the context of cavity quantum electro-dynamics (cavity QED) with Rydberg atoms [59], a first experiment on continuous QND measurements of the number of microwave photons was performed by the group at Laboratoire Kastler-Brossel (ENS) [58]. Later on, this

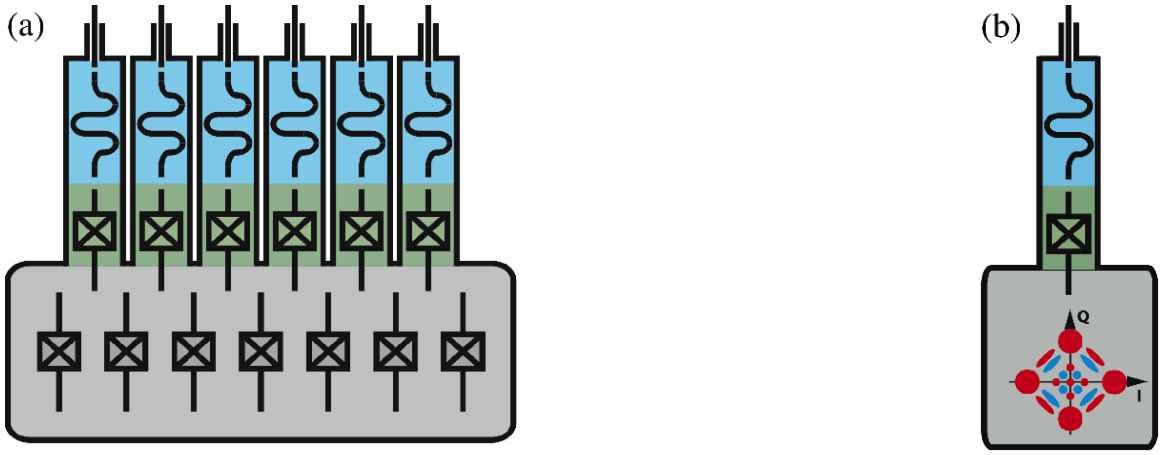


Figure 1. (a) A protected logical qubit consisting of a register of many qubits: here, we see a possible architecture for the Steane code [93] consisting of 7 qubits requiring the measurement of 6 error syndromes. In this sketch, 7 transmon qubits in a high-Q resonator and the measurement of the 6 error 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.

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 [49], [31], [92], [33]. 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) [48], recent advances in quantum-limited amplifiers [83], [96] have opened doors to high-fidelity non-demolition measurements and real-time feedback for superconducting qubits [60]. 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 [96], [82], [42]. 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 [73].

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 [80] and the closely related coherent feedback [71] 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 [56], single-qubit state stabilization [75], and the creation [35] and stabilization [64], [70][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 [56], [68], [46]. The experimental results based on these protocols have illustrated the efficiency of the approach [56][9]. Through these experiments, we exploit the strong dispersive interaction [87] between superconducting qubits and a single low-Q cavity mode playing the role of a dissipative reservoir. Applying some continuous-wave (cw) microwave drives with well-chosen fixed frequencies, amplitudes, and phases, we engineer an effective interaction Hamiltonian which evacuates entropy from the qubits when an eventual perturbation occurs: 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 [74], [85], [61][5].

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 [84], [31], [91], [86], [92], [33][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 [33]: 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 [99]; 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 [78]. 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 [60], [42].

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 [36] and is related to quantum trajectories [44], [47]. A modern and mathematical exposure of the diffusive models is given in [34]. In [100] 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 [84], [91]. This stability result is extended to a large class of continuous-time filters in [32]. Further efforts are required to characterize asymptotic and exponential stability. Estimations of convergence rates are available only for quantum non-demolition measurements [37]. Parameter estimations based on measurement data of quantum trajectories can be formulated within such quantum filtering framework [51], [76].

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 [41] that post-selection statistics and "past quantum" state analysis [52] 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 [80], [71]. 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] [56], 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 [79]). 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., [30][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] [56], [74] 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 [55], [54]. We will start with [88] and [81] where, based on a theorem due to Birkhoff [38], 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 [95] and by our recent generalized extension of the latter synchronizing interactions to quantum systems [72].

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 [77], 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 [97], [69] with experimental limitations. The characterization of Kraus maps to stabilize any types of states has also been established [39], 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 [59], [53]. 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 [67].

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 [43] for the adiabatic elimination of low-Q cavity). Contrarily 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 [50] and invariant manifold techniques [45] 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 [62] 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 [56][9] [68].

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

- Rémi Azouit (supervisor: Pierre Rouchon; co-supervisor: Alain Sarlette) has successfully defended his PhD thesis on October 27th and is now moving as a postdoc to Sherbrooke University. This thesis provides a systematic approach towards model reduction through adiabatic elimination for open quantum systems.
- Joachim Cohen (supervisor: Mazyar Mirrahimi) has successfully defended his PhD thesis on February 2nd. This thesis provides a roadmap for future experiments on autonomous hardware efficient quantum error correction with superconducting circuits.

[0] Here the partiality means that no single quantum measurement is capable of providing the complete information on the state of the system.

5.1.1. Awards

- Mazyar Mirrahimi has received the “Inria-Academie des Sciences young researcher award 2017”.
- Pierre Rouchon has received the “Grand Prix IMT-Academie ses Sciences 2017”.

6. New Results

6.1. Quantum Walks and accelerated mixing algorithms

Participants: A. Sarlette

This major line of work has been pursued together with S.Apers (UGent) and F.Ticozzi (U.Padova), in an attempt to distinguish what is "necessarily" quantum in such models, and what could be explained by memory effects which we could mimic with just classical dynamic controllers. We hence have a series of papers on both sides (quantum and non-quantum): the conference papers are published, the journal papers will be for 2018.

In [19], we investigate under which conditions a higher-order Markov chain, or more generally a Markov chain on an extended state space, can mix faster than a standard Markov chain on a graph of interest. We find that, depending on the constraints on the dynamics, two very different scenarios can emerge: under strict invariance of the target marginal and for general initialization of the lifted chain no speedup is possible; on the other hand, if these requirements are both relaxed, the lifted dynamics can achieve mixing in a time that corresponds to the diameter of the graph, which is optimal.

In [20], we establish a discrete-geometric bound on the convergence speed of mixing with *any* local stochastic process, under the key assumption that it leaves the target distribution invariant at each time. These processes include classical algorithms, any quantum algorithms, as well as possibly other strategies that obey the non-signalling criterion of probability transmission. We explicitly give the bound in terms of isoperimetric inequalities. We illustrate how this general result leads to new bounds on convergence times beyond the explicit Markovian setting. Mixing is essentially concerned with the discrete-time spreading of a distribution along the edges of a graph. In essence we establish that even by exploiting global information about the graph and allowing a very general use of this information, this spreading can still not be accelerated beyond the so-called *conductance bound*. An upcoming journal paper will discuss which assumption changes do lead to faster algorithms, and argue how relevant they are for practical applciations.

In [26], we give a preview on our specific results about Quantum walks. Quantum walks have been linked to acceleration in various information processing tasks, and proposed as a possible model for quantum-enhanced behavior in biological systems. These links and acceleration claims have been made with various levels of detail. Here we consider discrete-time quantum walks, and focus on the task of mixing, i.e., distributing the state over a graph. Previous papers have observed that the so-called coined quantum walks can accelerate mixing on certain graphs with respect to the optimal classical Markov chain. We here show that the same speedup can be attained with a classical process, if a similar classical coin is added. We establish a precise correspondence between the mixing performance of quantum walks and such " lifted walks " for all (finite) graphs, and thereby improve known bounds on quantum walk mixing time. We conclude that the advantage of quantum walks with respect to classical processes is not in the mixing speed of the optimal design. However, a notable quantum advantage might reside in the fact that the mixing speed obtained with suboptimal designs, due to for instance limited graph knowledge, appears to be generically faster. The journal version is being finalized and will be sumbitted before the end of 2017.

6.2. String Stability towards Leader thanks to Asymmetric Bidirectional Controller

Participants: A. Sarlette

This result published in [21] is the result of an investigation of classical (non-quantum) distributed and coupled systems and their fundamental limitations – a sequel of A.Sarlette's previous line of work. It deals with the problem of string stability of interconnected systems with double-integrator open loop dynamics (e.g. acceleration-controlled vehicles). We analyze an asymmetric bidirectional linear controller, where each vehicle is coupled solely to its immediate predecessor and to its immediate follower with different gains in these two directions. We show that in this setting, unlike with unidirectional or symmetric bidirectional controllers, string stability can be recovered when disturbances act only on a small (N-independent) set of leading vehicles. This improves existing results from the literature with this assumption. We also indicate that string stability with respect to arbitrarily distributed disturbances cannot be achieved with this controller.

A journal version is in preparation where we essentially close the subject, on a discrete-controller version:
- we will show that no local digital controller whatsoever (including nonlinearity, local communication,...) can achieve the academic property of string stability for infinite length chains and with bounded noise/disturbance on *each* member of the chain, and this implies serious consequences for practical behaviors of finite-length chains.
- conversely, we give the equivalent of the above result to show that if one is concerned mainly about the noise/disturbance acting on the leader (boundary condition of the chain), then indeed our above result achieves all existing variants of the string stability definitions.

6.3. Towards generic adiabatic elimination for bipartite open quantum systems

Participants: R. Azouit, A. Sarlette, P. Rouchon (and F. Chittaro, visitor in 2016)

The paper [12] is the main paper summarizing the results of the PhD thesis of R.Azouit. We give a theoretical method, with a directly applicable recipe for the physicists who would want to use it, and with examples worked out on applications that experimentalists (e.g. in the partner group at Yale U.) are actually considering nowadays.

We consider a composite open quantum system consisting of a fast subsystem coupled to a slow one. Using the timescale separation, we develop an adiabatic elimination technique to derive at any order the reduced model describing the slow subsystem. The method, based on an asymptotic expansion and geometric singular perturbation theory, ensures the physical interpretation of the reduced second-order model by giving the reduced dynamics in a Lindblad form and the state reduction in Kraus map form. We give explicit second-order formulas for Hamiltonian or cascade coupling between the two subsystems. These formulas can be used to engineer, via a careful choice of the fast subsystem, the Hamiltonian and Lindbald operators governing the dissipative dynamics of the slow subsystem.

6.4. Deterministic submanifolds and analytic solution of the quantum stochastic differential master equation describing a monitored qubit

Participants: A. Sarlette, P. Rouchon

In the paper [18], we study the stochastic differential equation (SDE) associated with a two-level quantum system (qubit) subject to Hamiltonian evolution as well as unmonitored and monitored decoherence channels. The latter imply a stochastic evolution of the quantum state (density operator), whose associated probability distribution we characterize. We first show that for two sets of typical experimental settings, corresponding either to weak quantum non demolition measurements or to weak fluorescence measurements, the three Bloch coordinates of the qubit remain confined to a deterministically evolving surface or curve inside the Bloch sphere. We explicitly solve the deterministic evolution, and we provide a closed-form expression for the probability distribution on this surface or curve. Then we relate the existence in general of such deterministically evolving submanifolds to an accessibility question of control theory, which can be answered with an explicit algebraic criterion on the SDE. This allows us to show that, for a qubit, the above two sets of weak measurements are essentially the only ones featuring deterministic surfaces or curves.

This paper was motivated by a striking experimental observation of Ph.Campagne-Ibarcq (group of Benjamin Huard - now at ENS Lyon and still collaborator). It appears to be actually quite general, and to generalize to higher-dimensional systems than the qubit. We are working on this extension, time permitting (as we have no student support currently), to publish a complete story about relevant experimental systems where the QSDE can be modeled in a very low-dimensional manifold.

6.5. Loss-tolerant parity measurement for distant quantum bits

Participants: A. Sarlette, M. Mirrahimi

This work, published in [17], [24], is part of the major line of work led by M.Mirrahimi about stabilizing distant entangled states. The latter are a major building block in quantum information technology, thanks to their ability to enable quantum teleportation. They are supposed to play a major 'quantum-bus-type' role in some of the most promising quantum computing architectures.

In this paper, we propose a scheme to measure the parity of two distant qubits, while ensuring that losses on the quantum channel between them does not destroy coherences within the parity subspaces. This capability enables deterministic preparation of highly entangled qubit states whose fidelity is not limited by the transmission loss. The key observation is that for a probe electromagnetic field in a particular quantum state, namely a superposition of two coherent states of opposite phases, the transmission loss stochastically applies a near-unitary back-action on the probe state. This leads to a parity measurement protocol where the main effect of the transmission losses is a decrease in the measurement strength. By repeating the non-destructive (weak) parity measurement, one achieves a high-fidelity entanglement in spite of a significant transmission loss.

6.6. Discrete-time reservoir engineering with entangled bath and stabilizing squeezed states

Participants: Z. Miao and A. Sarlette

The paper [15] is the first result of a line of work that we try to establish about the possible use of "time-structured reservoirs" towards stabilizing more complicated states of quantum systems. In particular, we here analyze a setting where reservoir items (qubits) are entangled over discrete time, and we show how it stabilizes squeezed states of a quantum harmonic oscillator. The parameters of the stabilized state can be tuned at will, in tradeoff with the convergence speed. The squeezing direction is determined by the phase of entanglement, thus allowing to distinguish genuine entanglement from mere classical correlations.

This work has allowed to identify the following lines for future research:
- first check time-varying, non-entangled reservoir inputs: from the same mathematical model, it appears that they can also stabilize squeezed states.
- provide a proof, on a non-trivial setting, of the specific benefit of entangled inputs: i.e. show how they achieve stabilization of some interesting states which are not accessible with any non-entangled inputs.
- laying the premises of possible approaches to studying continuous-time reservoir inputs which are entangled over time. This is currently an open question even from the modeling perspective.

6.7. Observing a quantum Maxwell demon at work

Participants: R. Azouit, B. Huard and P. Rouchon

The results of this section were published [14]

In apparent contradiction to the laws of thermodynamics, Maxwell's demon is able to cyclically extract work from a system in contact with a thermal bath exploiting the information about its microstate. The resolution of this paradox required the insight that an intimate relationship exists between information and thermodynamics. Here, this Maxwell demon experiment tracks the state of each constituent both in the classical and quantum regimes. The demon is a microwave cavity that encodes quantum information about a superconducting qubit and converts information into work by powering up a propagating microwave pulse by stimulated emission. Thanks to the high level of control of superconducting circuits, direct measurements (combined with maximum-likelihood estimation techniques inspired by [90]) give the extracted work and entropy remaining in the demon's memory. This experiment provides an enlightening illustration of the interplay of thermodynamics with quantum information.

6.8. Asymptotic expansions of Laplace integrals for quantum state tomography

Participant: P. Rouchon (with his former PhD student P. Six)

The results of this section were published in [25].

Bayesian estimation of a mixed quantum state can be approximated via maximum likelihood (MaxLike) estimation when the likelihood function is sharp around its maximum. Such approximations rely on asymptotic expansions of multi-dimensional Laplace integrals. When this maximum is on the boundary of the integration domain, as it is the case when the MaxLike quantum state is not full rank, such expansions are not standard. We provide here such expansions, even when this maximum does not belong to the smooth part of the boundary, as it is the case when the rank deficiency exceeds two. These expansions provide, aside the MaxLike estimate of the quantum state, confidence intervals for any observable. They confirm the formula proposed and used without precise mathematical justifications by the authors in an article published in Physical Review A in 2016 [90].

6.9. Generating higher order quantum dissipation from lower order parametric processes

Participant: M. Mirrahimi (and S. Mundhada, visitor from Yale in 2016)

The results of this section were published in [16].

Stabilization of quantum manifolds is at the heart of error-protected quantum information storage and manipulation. Nonlinear driven-dissipative processes achieve such stabilization in a hardware efficient manner. Josephson circuits with parametric pump drives implement these nonlinear interactions. In this work, we propose a scheme to engineer a four-photon drive and dissipation on a harmonic oscillator by cascading experimentally demonstrated two-photon processes. This would stabilize a four-dimensional degenerate manifold in a superconducting resonator. We analyze the performance of the scheme using numerical simulations of a realizable system with experimentally achievable parameters. This theoretical work, initiated by Shantanu Mundhada during his visit to Inria in 2016, is currently investigated experimentally at Yale.

6.10. Degeneracy-preserving quantum nondemolition measurement of parity-type observables for cat qubits

Participant: J. Cohen, M. Mirrahimi

The results of this section were published in [13] and correspond to an important chapter of J. Cohen's thesis [11].

A central requirement for any quantum error correction scheme is the ability to perform quantum nondemolition measurements of an error syndrome, corresponding to a special symmetry property of the encoding scheme. It is in particular important that such a measurement does not introduce extra error mechanisms, not included in the error model of the correction scheme. In this work, we ensure such a robustness by designing an interaction with a measurement device that preserves the degeneracy of the measured observable. More precisely, we propose a scheme to perform continuous and quantum nondemolition measurement of photon-number parity in a microwave cavity. This corresponds to the error syndrome in a class of error correcting codes called the cat codes, which have recently proven to be efficient and versatile for quantum information processing. In our design, we exploit the strongly nonlinear Hamiltonian of a high-impedance Josephson circuit, coupling a high-Q storage cavity mode to a low-Q readout one. By driving the readout resonator at its resonance, the phase of the reflected or transmitted signal carries directly exploitable information on parity-type observables for encoded cat qubits of the high-Q mode. This important result has defined a new line of experimental research persued by the experimentalists of the Quantic team and Yale university.

7. Partnerships and Cooperations

7.1. Regional Initiatives

7.1.1. Emergences-Ville de Paris program, ENDURANCE project

In the framework of the Ville de Paris program “EMERGENCES”, Zaki Leghtas has received a funding for his research program "Multi-photon processes in superconducting circuits for quantum error correction". This grant of 232k euros over 4 years will complement the ANR project of the same name obtained last year. Using this funding, we will purchase all the microwave and nano-fabrication equipment and consumables for the experiment based at ENS.

7.1.2. DIM SIRTEQ, PhD fellowship

In the framework of the project “DIM SIRTEQ Domaine d’intérêt Majeur: Science et Ingénierie Quantique” of Ile de France Region, we have received 18 months of PhD fellowship. This completes the funding from ANR GEARED of the PhD thesis of J. Guillaud, who has started his PhD under the supervision of M. Mirrahimi and P. Rouchon in September 2017.

7.1.3. Programme Math-PSL, Postdoctoral fellowship

In the framework of the programme Math-PSL of PSL Research University, we have resceived a 12 month postdoctoral fellowship. Paolo Forni has been hired as a postdoc on this funding.

7.2. National Initiatives

7.2.1. 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. The participants of the project are Mazyar Mirrahimi, François Mallet (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). The entangled states are difficult to generate and sustain as interaction with a noisy environment leads to rapid loss of their unique quantum properties. Through Geared we intend to investigate different complementary approaches to master the entanglement of microwave photons coupled to quantum superconducting circuits.

7.2.2. 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.

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. We are in particular working on the proper combination of two model reduction techniques in their experimental context: adiabatic elimination and Rotating-Wave Approximation. An ANR-JCJC project has been deposited by Alain Sarlette on this subject, with Benjamin Huard as external supporting collaborator.

Partner 2: University of Padova

Alain Sarlette has been pursuing a fruitful collaboration with the group of Francesco Ticozzi on dynamical systems aspects of quantum systems. Common work on the theory of quantum random walks is being finalized and we are working out a concrete plan about next possible steps.

Partner 3: Ghent University.

A. Sarlette is collaborating with applied mathematicians interested in quantum control at his former institution UGent (Dirk Aeyels, Lode Wylleman, Gert De Cooman) in the framework of thesis co-supervisions. Two students are in their last year PhD, in particular Simon Apers is finalizing a thesis centered around Quantum Walks, also in collaboration with Partner 2. A master student in applied physics has started an internship in 2017.

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, L. Verney, J. Guillaud and M. Mirrahimi visited Yale for respectively, 2, 3 and 4 months.

7.5. International Research Visitors

7.5.1. Visits of International Scientists

P. S. Pereira da Silva (Escola Politécnica, PTC, University of SaoPaulo, Brazil) made a 2-week visit (July 3 to July 14) 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 TAQUILLA associate team, Mazyar Mirrahimi spent four months in the Quantronics Laboratory of Michel H. Devoret and in the Rob Schoelkopf Lab at Yale University. Also, in this same framework Jérémie Guillaud and Lucas Verney spent respectively three months and two months in the same group.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Journal

8.1.1.1. Member of the Editorial Boards

Pierre Rouchon is member of the editorial board of Annual Reviews in Control (since 2016).

Mazyar Mirrahimi was a guest editor for the journal "Quantum Science and Technology" (Institute Of Physics, 2016-2017), Special issue on “Quantum coherent feedback and quantum reservoir engineering”.

8.1.1.2. Reviewer - Reviewing Activities

Zaki Leghtas served as a referee for Physical Review Journals.

Mazyar Mirrahimi served as a referee for Physical Review Journals.

Pierre Rouchon has been a reviewer for several automatic control and dynamical systems journals and conferences.

Alain Sarlette has been a reviewer for several automatic control and dynamical systems journals and conferences.

8.1.2. Invited Talks

Mazyar Mirrahimi, July 2017, ICTS (Workshop Open Quantum Systems), Bangalore, India.

Mazyar Mirrahimi, June 2017, 22ème conférence Claude Itzykson, CEA Saclay, France.

Mazyar Mirrahimi, June 2017, CIFAR Workshop on Quantum Cavities, Jouvence, Quebec, Canada.

Mazyar Mirrahimi, May 2017, L2S, Supelec, France.

Mazyar Mirrahimi, April 2017, Conference of Optical Society of America, Quantum Information and Measurement, Paris, France.

Mazyar Mirrahimi, March 2017, CMAP, Ecole Polytechnique, France.

Mazyar Mirrahimi, February 2017, UVSQ, France.

Pierre Rouchon, November 2017, Control and Optimization Conference on the occasion of Frédéric Bonnans 60th birthday, Palaiseau, France.

Pierre Rouchon, June 2017, 22ème conférence Claude Itzykson, CEA Saclay, France.

Pierre Rouchon, April 2017, workshop on Quantum Control Theory: Mathematical Aspects and Physical Applications, TUM-IAS, Garching, Germany.

Pierre Rouchon, April 2017, 4th Workshop on Quantum Non-Equilibrium Dynamics, University of Nottingham, UK.

Alain Sarlette, July 2017, Praqcsys: Principles and Applications of Control in Quantum Systems, Seattle, USA.

Alain Sarlette, June 2017, L2S, Supelec, France.

Rémi Azouit, February 2017, Sherebrooke University, Canada.

.

8.1.3. Scientific Expertise

Mazyar Mirrahimi is a member of the Technical Committee on "Distributed Parameter Systems" in IFAC (International Federation of Automatic Control).

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.

Pierre Rouchon was a member of the scientific committee of PRACQSYS 2017

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Zaki Leghtas taught a course on Quantum Mechanics at Paris Sciences et Lettres (40 hours).

Zaki Leghtas taught a course on Quantum Mechanics and Statistical Physics at Mines ParisTech (12 hours).

Zaki Leghtas taught a course on Complex Analysis at Mines ParisTech (10 hours).

Mazyar Mirrahimi and Pierre Rouchon have given a course (20 hours) entitled "UE : Analyse et contrôle de systèmes quantiques " in the "Master de sciences et technologies, mention mathématiques et applications, Université Pierre et Marie Curie".

Mazyar Mirrahimi is hired as a professeur chargé de cours à temps partiel of Applied Mathematics at Ecole Polytechnique. His teaching will start during winter 2018.

Mazyar Mirrahimi has given TDs of the courses on Probabilities and on Stochastic Processes at Ecole des Mines de Paris.

Pierre Rouchon gave a course on "Cryptographie, théorie des nombres et information quantique" at Mines ParisTech (24 hours).

Pierre Rouchon gave a course on "Modelling, simulation and feedback of open quantum systems" in the PSL-Master, PSL-IT, IQ Ingénierie Quantique (12 hours).

Alain Sarlette has given a master course on “Probabilistic robotics” at Ghent University (30 hours) and has given TDs of the courses on Probabilities and on Stochastic Processes at Ecole des Mines de Paris.

Alain Sarlette has given a quantum-related lecture in the course on Stochastic Processes (5 hours), Ecole des Mines de Paris.

8.2.2. Supervision

PhD in progress: Raphael Lescanne. ENS. “Engineering Multi-Photon Dissipation In Superconducting Circuits For Quantum Error Correction”. September 2016. (advisors: Zaki Leghtas and Benjamin Huard).

PhD: Rémi Azouit. Mines Paristech. “Adiabatic elimination for open quantum systems”. 2014-2017. (advisors: Pierre Rouchon and Alain Sarlette), Defended on Oct 2017.

PhD in progress: Gerardo Cardona. Mines ParisTech. "Beyond static gains in analog quantum feedback control". Nov 2016 (advisors: Pierre Rouchon and Alain Sarlette).

PhD in progress: Alain Sarlette is co-supervising 3 PhD students with his former institution UGent (Simon Apers, Zhifei Zhang, Arash Farnam). Simon Apers is working on (quantum) network algorithms accelerations and intends to address other quantum control questions.

PhD: Joachim Cohen. ENS. “Autonomous quantum error correction with superconducting circuits”. 2013-2017 (advisor: Mazyar Mirrahimi), Defended on Feb 2017.

PhD in progress: Lucas Verney. ENS. “Robust quantum information processing with superconducting circuits”. Sept 2016. (advisors: Zaki Leghtas and Mazyar Mirrahimi).

PhD in progress: Jérémie Guillaud. ENS. “Modular architecture for quantum information processing”. Sept 2017. (advisors: Mazyar Mirrahimi and Pierre Rouchon).

8.2.3. Juries

Mazyar Mirrahimi was a member the PhD defense committees of Serguei Fedortchenko (Jury president, University Paris Diderot).

Pierre Rouchon was a referee for the PhD thesis of Muhammad Emzi, Australian National University, and for the Habilitation thesis of Marco Caponigro, UMPC.

Alain Sarlette was a jury member for the PhD of Stavros Lopatatzidis (UGent, Belgium) and of Bram Vervisch (UGent, Belgium).

8.3. Popularization

Mazyar Mirrahimi has been interviewed by Le Monde for a dossier on quantum information.

Mazyar Mirrahimi gave an invited talk on “Quantum computing” at the CRiP’s ITES Innovation Summit at Deauville, France in Mars 2017.

Mazyar Mirrahimi gave an invited talk on “Quantum computing” at X-Creation (X-Drahi) in May 2017.

Pierre Rouchon was invited to give a talk "Contrôle des systèmes: du classique au quantique", Journée d’inauguration du programme PSL-maths, 19 October 2017 at ENS-Paris.

Alain Sarlette has been speaking at inria-organized dissemination events:
- 03/07 "fresh from the labs" talk about quantum technology hardware (at Boston Consulting Group, Paris – team gamma)
- 06-08/06 Keynote speech at Journées DGDT
- 10/07 presentation at inria-Paris-labs visit by high-level managers and stakeholders

Alain Sarlette is answering questions about quantum control and quantum computing on the website "ik-heb-een-vraag.be" where Flemish layman can ask questions to scientific experts.

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] N. COTTET, S. JEZOUIN, L. BRETHEAU, P. CAMPAGNE-IBARCQ, Q. FICHEUX, J. ANDERS, A. AUFFÈVES, R. AZOUIT, P. ROUCHON, B. HUARD.*Observing a quantum Maxwell demon at work*, in "Proceedings of the National Academy of Sciences of the United States of America ", July 2017, vol. 114, n° 29, p. 7561 - 7564 [*DOI :* 10.1073/PNAS.1704827114], https://hal.archives-ouvertes.fr/hal-01626961.

[5] 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° 6224, p. 853-857 [*DOI :* 10.1126/SCIENCE.AAA2085], https://hal.inria.fr/hal-01240210.

[6] 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.

[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

Doctoral Dissertations and Habilitation Theses

[11] J. COHEN.*Autonomous quantum error correction with superconducting qubits*, PSL Research University, February 2017, https://tel.archives-ouvertes.fr/tel-01545186.

Articles in International Peer-Reviewed Journal

[12] R. AZOUIT, F. C. CHITTARO, A. SARLETTE, P. ROUCHON.*Towards generic adiabatic elimination for bipartite open quantum systems*, in "Quantum Science and Technology", 2017, vol. 2, n° 4, p. 1-15, https://arxiv.org/abs/1704.00785 [*DOI :* 10.1088/2058-9565/AA7F3F], https://hal.inria.fr/hal-01634588.

[13] 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, https://arxiv.org/abs/1611.01219 - 25 pages, 7 figures [*DOI :* 10.1103/PHYSREVLETT.119.060503], https://hal.inria.fr/hal-01437156.

[14] N. COTTET, S. JEZOUIN, L. BRETHEAU, P. CAMPAGNE-IBARCQ, Q. FICHEUX, J. ANDERS, A. AUFFÈVES, R. AZOUIT, P. ROUCHON, B. HUARD.*Observing a quantum Maxwell demon at work*, in "Proceedings of the

National Academy of Sciences of the United States of America ", 2017, vol. 114, n^{o} 29, p. 7561-7564 [*DOI :* 10.1073/PNAS.1704827114], https://hal.archives-ouvertes.fr/hal-01626961.

[15] Z. MIAO, A. SARLETTE.*Discrete-time reservoir engineering with entangled bath and stabilising squeezed states*, in "Quantum Science and Technology", 2017, vol. 2, n^{o} 3, p. 1-20, https://arxiv.org/abs/1704.07881 [*DOI :* 10.1088/2058-9565/AA7CE8], https://hal.inria.fr/hal-01634586.

[16] S. O. MUNDHADA, A. GRIMM, S. TOUZARD, U. VOOL, S. SHANKAR, M. H. DEVORET, M. MIRRAHIMI.*Generating higher order quantum dissipation from lower order parametric processes*, in "Quantum Science and Technology", May 2017, https://arxiv.org/abs/1612.04341 - 9 pages, 5 figures [*DOI :* 10.1088/2058-9565/AA6E9D], https://hal.inria.fr/hal-01437303.

[17] A. SARLETTE, M. MIRRAHIMI.*Loss-tolerant parity measurement for distant quantum bits*, in "Physical Review A", 2017, vol. 95, n^{o} 3 [*DOI :* 10.1103/PHYSREVA.95.032329], https://hal.inria.fr/hal-01395590.

[18] A. SARLETTE, P. ROUCHON.*Deterministic submanifolds and analytic solution of the quantum stochastic differential master equation describing a monitored qubit*, in "Journal of Mathematical Physics", 2017, vol. 58, n^{o} 6, p. 1-28, https://arxiv.org/abs/1603.05402 [*DOI :* 10.1063/1.4984587], https://hal-mines-paristech.archives-ouvertes.fr/hal-01635290.

International Conferences with Proceedings

[19] S. APERS, A. SARLETTE, F. TICOZZI.*When Does Memory Speed-up Mixing?*, in "IEEE Conference on Decision and Control", Melbourne, Australia, December 2017, https://hal.inria.fr/hal-01634630.

[20] S. APERS, F. TICOZZI, A. SARLETTE.*Bounding the convergence time of local probabilistic evolution*, in "Geometric Science of Information, Second International Conference, GSI 2017", Paris, France, F. NIELSEN, F. BARBARESCO (editors), LNCS - Lecture Notes in Computer Science, Springer, November 2017, vol. 10589, https://hal.inria.fr/hal-01634611.

[21] A. FARNAM, A. SARLETTE.*String Stability towards Leader thanks to Asymmetric Bidirectional Controller*, in "IFAC World congres 2017", Toulouse, France, July 2017, https://arxiv.org/abs/1603.05498 - interactive presentation at 2017 IFAC World Congress, Toulouse [*DOI :* 10.1016/J.IFACOL.2017.08.1673], https://hal.inria.fr/hal-01634578.

Conferences without Proceedings

[22] R. AZOUIT, F. CHITTARO, A. SARLETTE, P. ROUCHON.*Structure-preserving adiabatic elimination for open bipartite quantum systems*, in "2017 IFAC World Congress", Toulouse, France, July 2017 [*DOI :* 10.1016/J.IFACOL.2017.08.2000], https://hal.archives-ouvertes.fr/hal-01394422.

[23] P. COMBES, F. MALRAIT, P. MARTIN, P. ROUCHON.*Modeling and identification of synchronous reluctance motors*, in "Electric Machines and Drives Conference (IEMDC), 2017 IEEE International", Miami, United States, May 2017, https://hal-mines-paristech.archives-ouvertes.fr/hal-01636712.

[24] A. SARLETTE, M. MIRRAHIMI.*Fault tolerant remote parity detection and entanglement stabilization*, in "OMA-Quantum Information and Measurement", Paris, France, April 2017, https://hal.inria.fr/hal-01634629.

Scientific Books (or Scientific Book chapters)

[25] P. SIX, P. ROUCHON.*Asymptotic Expansions of Laplace Integrals for Quantum State Tomography*, in "Feedback Stabilization of Controlled Dynamical Systems", Lecture Notes in Control and Information Sciences, Springer, March 2017, vol. 473 [*DOI :* 10.1007/978-3-319-51298-3_12], https://hal-mines-paristech.archives-ouvertes.fr/hal-01528082.

Other Publications

[26] S. APERS, A. SARLETTE, F. TICOZZI.*Fast Mixing with Quantum Walks vs. Classical Processes*, January 2017, Quantum Information Processing (QIP) 2017, Poster, https://hal.inria.fr/hal-01395592.

[27] N. DIDIER, J. GUILLAUD, S. SHANKAR, M. MIRRAHIMI.*Remote entanglement stabilization for modular quantum computing*, November 2017, https://arxiv.org/abs/1703.03379 - 5 pages, 4 figures, https://hal.inria.fr/hal-01652766.

[28] 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*, November 2017, https://arxiv.org/abs/1709.05425 - 10 pages, 11 figures (incl. Supplementary Information), https://hal.inria.fr/hal-01652773.

[29] 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*, November 2017, working paper or preprint, https://hal.inria.fr/hal-01652771.

References in notes

[30] S. ATTAL, A. JOYE, C.-A. PILLET (editors). *Open Quantum Systems III: Recent Developments*, Springer, Lecture notes in Mathematics 1880, 2006.

[31] 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.

[32] 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.

[33] 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.

[34] A. BARCHIELLI, M. GREGORATTI.*Quantum Trajectories and Measurements in Continuous Time: the Diffusive Case*, Springer Verlag, 2009.

[35] 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.

[36] V. BELAVKIN.*Quantum stochastic calculus and quantum nonlinear filtering*, in "Journal of Multivariate Analysis", 1992, vol. 42, n^{o} 2, p. 171–201.

[37] 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.

[38] G. BIRKHOFF.*Extensions of Jentzch's theorem*, in "Trans. Amer. Math. Soc.", 1957, vol. 85, p. 219–227.

[39] S. BOLOGNANI, F. TICOZZI.*Engineering stable discrete-time quantum dynamics via a canonical QR decomposition*, in "IEEE Trans. Autom. Control", 2010, vol. 55.

[40] V. BRAGINSKI, F. KHALILI.*Quantum Measurements*, Cambridge University Press, 1992.

[41] 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.

[42] 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.

[43] H. CARMICHAEL.*Statistical Methods in Quantum Optics 2: Non-Classical Fields*, Spinger, 2007.

[44] H. CARMICHAEL.*An Open Systems Approach to Quantum Optics*, Springer-Verlag, 1993.

[45] J. CARR.*Application of Center Manifold Theory*, Springer, 1981.

[46] J. COHEN, M. MIRRAHIMI.*Dissipation-induced continuous quantum error correction for superconducting circuits*, in "Phys. Rev. A", 2014, vol. 90, 062344.

[47] 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.

[48] M. H. DEVORET, A. WALLRAFF, J. MARTINIS.*Superconducting Qubits: A Short Review*, 2004, arXiv:cond-mat/0411174.

[49] 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.

[50] N. FENICHEL.*Geometric singular perturbation theory for ordinary differential equations*, in "J. Diff. Equations", 1979, vol. 31, p. 53–98.

[51] 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.

[52] 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.

[53] C. GARDINER, P. ZOLLER.*Quantum Noise*, third, Springer, 2010.

[54] S. GAUBERT, Z. QU.*Checking the strict positivity of Kraus maps is NP-hard*, in "arXiv:1402.1429", 2014.

[55] 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.

[56] 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.

[57] D. GOTTESMAN, A. KITAEV, J. PRESKILL.*Encoding a qubit in an oscillator*, in "Phys. Rev. A", 2001, vol. 64, 012310.

[58] 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.

[59] S. HAROCHE, J.-M. RAIMOND.*Exploring the Quantum: Atoms, Cavities and Photons*, Oxford University Press, 2006.

[60] 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.

[61] 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.

[62] T. KATO.*Perturbation Theory for Linear Operators*, Springer, 1966.

[63] E. KNILL, R. LAFLAMME, G. MILBURN.*A scheme for efficient quantum computation with linear optics*, in "Nature", 2001, vol. 409, 46.

[64] 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.

[65] 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.

[66] 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.

[67] Z. LEGHTAS, A. SARLETTE, P. ROUCHON.*Adiabatic passage and ensemble control of quantum systems*, in "J. Phys. B", 2011, vol. 44, 154017.

[68] 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.

[69] J.-S. LI, N. KHANEJA.*Ensemble control of Bloch equations*, in "IEEE Trans. Autom. Control", 2009, vol. 54, p. 528–536.

[70] 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.

[71] S. LLOYD.*Coherent quantum feedback*, in "Phys. Rev. A", 2000, vol. 62, 022108.

[72] L. MAZZARELLA, A. SARLETTE, F. TICOZZI.*Consensus for quantum networks: from symmetry to gossip iterations*, in "IEEE Trans. Automat. Control", 2014, in press.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] H. NURDIN, M. JAMES, I. PETERSEN.*Coherent quantum LQG control*, in "Automatica", 2009, vol. 45, p. 1837–1846.

[78] 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.

[79] D. PETZ.*Monotone Metrics on matrix spaces*, in "Linear Algebra and its Applications", 1996, vol. 244, p. 81–96.

[80] 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.

[81] 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.

[82] 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.

[83] 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.

[84] 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.

[85] 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.

[86] 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.

[87] 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.

[88] 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.

[89] P. SHOR.*Scheme for reducing decoherence in quantum memory*, in "Phys. Rev. A", 1995, vol. 52, p. 2493–2496.

[90] P. SIX, P. CAMPAGNE-IBARCQ, I. DOTSENKO, A. SARLETTE, B. HUARD, P. ROUCHON.*Quantum state tomography with noninstantaneous measurements, imperfections, and decoherence*, in "Phys. Rev. A", 2016, vol. 93, 012109.

[91] 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.

[92] A. SOMARAJU, M. MIRRAHIMI, P. ROUCHON.*Approximate stabilization of infinite dimensional quantum stochastic system*, in "Reviews in Mathematical Physics", 2013, vol. 25, 1350001.

[93] A. STEANE.*Error Correcting Codes in Quantum Theory*, in "Phys. Rev. Lett", 1996, vol. 77, n^{o} 5.

[94] 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.

[95] J. TSITSIKLIS.*Problems in decentralized decision making and computation*, in "PhD Thesis, MIT", 1984.

[96] 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.

[97] L. VIOLA, E. KNILL, S. LLOYD.*Dynamical decoupling of open quantum system*, in "Phys. Rev. Lett.", 1999, vol. 82, p. 2417-2421.

[98] 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.

[99] 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.

[100] R. VAN HANDEL.*The stability of quantum Markov filters*, in "Infin. Dimens. Anal. Quantum Probab. Relat. Top.", 2009, vol. 12, p. 153–172.

Team RAP2

Réseaux, algorithmes et probabilités

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
Networks and Telecommunications

Table of contents

Team RAP2

Creation of the Team: 2017 January 01, end of the Team: 2017 December 31

Keywords:

Computer Science and Digital Science:

A6.1.2. - Stochastic Modeling (SPDE, SDE)
A8.8. - Network science
A8.9. - Performance evaluation

Other Research Topics and Application Domains:

B6.3.2. - Network protocols

1. Personnel

Research Scientists

Nicolas Broutin [Inria, Researcher, until Aug 2017, HDR]
Christine Fricker [Inria, Researcher, HDR]
Philippe Robert [Team leader, Inria, Senior Researcher, HDR]

External Collaborators

Yousra Chabchoub [ISEP]
Pierre Nicodeme [Université Paris XIII]

PhD Students

Renaud Dessalles [INRA, until Feb 2017]
Veronica Quintuna Rodriguez [Telecom Bretagne]
Wen Sun [Inria]
Guilherme Thompson [Inria]

Post-Doctoral Fellow

Davit Martirosyan [Inria]

Visiting Scientist

Ravi Rasendra Mazumdar [University of Waterloo, until Feb 2017]

Administrative Assistant

Nelly Maloisel [Inria]

2. Overall Objectives

2.1. Overall Objectives

The research team RAP2 (Networks, Algorithms and Communication Networks) was created in 2004 on the basis of a long standing collaboration between engineers at Orange Labsin Lannion and researchers from Inria Paris. The initial objective was to formalize and expand this fruitful collaboration.

At Orange Labsin Lannion, the members of the team are experts in the analytical modeling of communication networks as well as on some of the operational aspects of network management concerning traffic measurements on ADSL networks, for example.

At Inria Paris, the members of RAP have a recognized expertise in modeling methodologies applied to stochastic models of communication networks.

RAP2 also has the objective of developing new fundamental tools to investigate *probabilistic* models of complex communication networks. We believe that mathematical models of complex communication networks require a deep understanding of general results on stochastic processes. The two fundamental domains targeted are:

1. Design and analysis of algorithms for communication networks.
2. Analysis of scaling methods for Markov processes: fluid limits and functional limit theorems.

From the very beginning, it has been decided that RAP would focus on a number of particular issues over a period of three or four years. The general goal of the collaboration with Orange Labs is to develop, analyze and optimize algorithms for communication networks. The design of algorithms to allocate resources in large distributed systems is currently investigated in the framework of this collaboration:

3. Research Program

3.1. Scaling of Markov Processes

The growing complexity of communication networks makes it more difficult to apply classical mathematical methods. For a one/two-dimensional Markov process describing the evolution of some network, it is sometimes possible to write down the equilibrium equations and to solve them. The key idea to overcome these difficulties is to consider the system in limit regimes. This list of possible renormalization procedures is, of course, not exhaustive. The advantages of these methods lie in their flexibility to various situations and to the interesting theoretical problems they raised.

A fluid limit scaling is a particularly important means to scale a Markov process. It is related to the first order behavior of the process and, roughly speaking, amounts to a functional law of large numbers for the system considered.

A fluid limit keeps the main characteristics of the initial stochastic process while some second order stochastic fluctuations disappear. In “good” cases, a fluid limit is a deterministic function, obtained as the solution of some ordinary differential equation. As can be expected, the general situation is somewhat more complicated. These ideas of rescaling stochastic processes have emerged recently in the analysis of stochastic networks, to study their ergodicity properties in particular.

3.2. Design and Analysis of Algorithms

Data Structures, Stochastic Algorithms

The general goal of the research in this domain is of designing algorithms to analyze and control the traffic of communication networks. The team is currently involved in the design of algorithms to allocate bandwidth in optical networks and also to allocate resources in large distributed networks. See the corresponding sections below.

The team also pursues analysis of algorithms and data structures in the spirit of the former Algorithms team. The team is especially interested in the ubiquitous divide-and-conquer paradigm and its applications to the design of search trees, and stable collision resolution protocols.

3.3. Structure of random networks

This line of research aims at understanding the global structure of stochastic networks (connectivity, magnitude of distances, etc) via models of random graphs. It consists of two complementary foundational and applied aspects of connectivity.

RANDOM GRAPHS, STATISTICAL PHYSICS AND COMBINATORIAL OPTIMIZATION. The connectivity of usual models for networks based on random graphs models (Erdős–Rényi and random geometric graphs) may be tuned by adjusting the average degree. There is a *phase transition* as the average degree approaches one, a *giant* connected component containing a positive proportion of the nodes suddenly appears. The phase of practical interest is the *supercritical* one, when there is at least a giant component, while the theoretical interest lies at the *critical phase*, the break-point just before it appears.

At the critical point there is not yet a macroscopic component and the network consists of a large number of connected component at the mesoscopic scale. From a theoretical point of view, this phase is most interesting since the structure of the clusters there is expected (heuristically) to be *universal*. Understanding this phase and its universality is a great challenge that would impact the knowledge of phase transitions in all high-dimensional models of *statistical physics* and *combinatorial optimization*.

RANDOM GEOMETRIC GRAPHS AND WIRELESS NETWORKS. The level of connection of the network is of course crucial, but the *scalability* imposes that the underlying graph also be *sparse*: trade offs must be made, which required a fine evaluation of the costs/benefits. Various direct and indirect measures of connectivity are crucial to these choices: What is the size of the overwhelming connected component? When does complete connectivity occur? What is the order of magnitude of distances? Are paths to a target easy to find using only local information? Are there simple broadcasting algorithms? Can one put an end to viral infections? How much time for a random crawler to see most of the network?

NAVIGATION AND POINT LOCATION IN RANDOM MESHES. Other applications which are less directly related to networks include the design of improved navigation or point location algorithms in geometric meshes such as the Delaunay triangulation build from random point sets. There the graph model is essentially fixed, but the constraints it imposes raise a number of challenging problems. The aim is to prove performance guarantees for these algorithms which are used in most manipulations of the meshes.

4. New Results

4.1. Resource Allocation in Large Data Centres

Participants: Christine Fricker, Philippe Robert, Guilherme Thompson, Veronica Quintuna Rodriguez.

With the emergence of new networking paradigms such as Cloud Computing and related technologies (Fog Computing, VNF, etc.) new challenges in understanding, modelling and improving systems relying on these technologies arise. Our research goal is to understand how the stochastic nature of the access to these systems affects their performance, and to design algorithms which can improve global performance using local information. This research is made in collaboration with Fabrice Guillemin, from Orange Labs.

Building up from the results previously obtained by this team, we have extend our research towards more complex systems, investigating the behaviour of multi-resource systems, which are globally stable but local congested, a problem that naturally arises from the decentralization of resources. We investigate a cooperation scheme between processing facilities, where congestion-maker clients, the one with the largest demand the locally congested resource are systematically forwarded to the another data centre when some threshold on the occupation level is reached. These thresholds are chosen to anticipate sufficiently in advance potential shortages of any resource in any data centre. After providing some convergence results, we are able to express the performance of the system in terms of the invariant distribution of an inhomogeneous random walk on the plane. We derive optimal threshold parameters, improving the performance of the distributed Cloud Computing system in such a way that it approaches the efficiency of a centralised system. Currently, a document is being prepared for publication, but the main results are presented in G. Thompson's PhD Document [2].

4.2. Ressource allocation in vehicle sharing systems

Participants: Christine Fricker, Yousra Chabchoub.

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 to return them. 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 problem is to understand the system behavior and how to manage these systems in order to improve the allocation of both resources to users. This work is in collaboration with El Sibai Rayane (ISEP), Plinio Santini Dester (École Polytechnique), Hanène Mohamed (Université Paris-Ouest), and Danielle Tibi (Université Paris Diderot).

4.2.1. Stochastic modelling of bike-sharing systems

The goal is to derive the stationary behavior of the state process in a quite general model: number of bikes in the stations and in routes between two stations. Our stochastic model is the first one taking into account the finite number of spots at the stations. The basic model for bike-sharing systems comes within the framework of closed networks with two types of nodes: single server/finite capacity nodes and infinite servers/infinite capacity nodes. The effect of local saturation is modeled by generalized blocking and rerouting procedures, under which, as a key argument, the stationary state is proved to have product-form. For a class of large closed Jackson networks submitted to capacity constraints, asymptotic independence of the nodes in normal traffic phase is proved at stationarity under mild assumptions, using a Local Limit Theorem. The limiting distributions of the queues are explicit. In the Statistical Mechanics terminology, the equivalence of ensembles - canonical and grand canonical - is proved for specific marginals. This widely extends the existing results on heterogeneous bike-sharing systems. The grand canonical approximation can then be used for adjusting the total number of bikes and the capacities of the stations to the expected demand. [12]

4.2.2. Local load balancing policies.

Recently we investigated some load balancing algorithms for stochastic networks to improve the bike sharing system behavior. We focus on the choice of the least loaded station among two to return the bike, the so called Power of choice. Nevertheless, in real systems, this choice is local. Thus the main challenge is to deal with the choice between two neighboring stations.

For that, a set of N queues, with a local choice policy, is studied. When a customer arrives at queue i, he joins the least loaded queue between queues i and $i+1$. When the load tends to zero, we obtain an asymptotic for the stationary distribution of the number of customers at a queue. The main result is that, in equilibrium, queue lengths decay geometrically when ρ tends to 0, N fixed. It allows to compare local choice, no choice and *Power of choice*. The local policy changes the exponential decay with respect to no choice but does not lead to an improvement (double exponential tail decay) comparable to the random choice model. [19].

For a bike-sharing homogeneous model, we study a deterministic cooperation between the stations, two by two. Analytic results are achieved in an homogeneous bike-sharing model. They concern the mean-field limit as the system is large, and its equilibrium point. Results on performance mainly involve an original closed form expression of the stationary blocking probability and new tight bounds for the mean of the total number of customers in the classical join-the-shortest-queue model. These results are compared by simulations with the policy where the users choose the least loaded between two neighboring stations. It turns out that, because of randomness, the choice between two neighbours gives better performance than grouping stations two by two.

It relies on new results for the classical system of two queues under the join-the-shortest-queue policy. We revisited the study of the stationary distribution. A simple analytical solution is proposed. Using standard generating function arguments, a simple expression of the blocking probability is derived, which as far as we know is original. Furthermore, from the balance equations, all stationary probabilities are obtained as explicit combinations of those of states $(0,k)$ for $0 \leq k \leq K$. The blocking probability is also obtained for a variant with two queues under JSQ, where the constraint is on the total capacity of the system.

This extends to the infinite capacity and asymmetric cases, i.e., when the queues have different service rates. For the initial symmetric finite capacity model, the stationary probabilities of states $(0, k)$ can be obtained recursively from the blocking probability. In the other cases, they are implicitly determined through some functional equation that characterizes their generating function. For the infinite capacity symmetric model, we provide an elementary proof of a result by Cohen which gives the solution of the functional equation in terms of an infinite product with explicit zeroes and poles. See [9].

We use data, trip data (trips collected in a month) obtained from JCDecaux and reports on station status collected as open data, to test local choice policy. Indeed we designed and tested a new method that globally improves the distribution of the resources (bikes and docks) among the stations. It relies on a local small change in user behaviors, by adapting their trips to resource availability around their departure and arrival stations. Results show that, even with a partial 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. The key of our study is to detect spatial outliers, objects having a behavior significantly different from their spatial neighbors, in a context where neighbors are heavily correlated. Moran scatterplot is a well-known method that exploits similarity between neighbors in order to detect spatial outliers. We proposed an improved version of Moran scatterplot, using a robust distance metric called Gower similarity. Using this new version of Moran scatterplot, we identified many spatial outliers stations (often with much more available bikes, or with much more empty docks during the day) in Velib. For the occupancy data set obtained by modifiying trips, the number of spatial outliers drastically decreases. See [18].

4.3. Scaling Methods

Participants: Davit Martirosyan, Philippe Robert, Wen Sun.

4.3.1. Large Unreliable Stochastic Networks

The reliability of a large distributed system is studied. The framework is a system where files are stored on servers. When one of these servers breaks down, all files on it are lost. We assume that these files could be retrieved immediately and re-allocated among other servers while the failed server restarts but empty. It is a reasonable assumption since the failure rate is quit small comparing to an effective recovery mechanism. It is also assumed that each server is connected with a subset of servers in the system. When it breaks down, files on it are re-allocated on the servers that in this subset, following a given policy. Our main interest is the influence on the loads due to two allocation algorithms: the " Random Choice" (RC) policy and the "Power of d Choices" (PoC) policy.

- (RC) Each copy join a server in the subset at random.
- (PoC) Each copy chooses d servers in the subset at random, and joins the least loaded one.

The asymptotic behaviors of these two policies are investigated through mean field models. We have show that when the number of servers getting large, the load of each server can be approached by a linear (resp. non-linear) Markov process for RC (resp. PoC) policy. The equilibrium distributions of these asymptotic processes are also given.

For the case $d = 2$ and all the servers are connected, see the paper [15]. This is a joint work with Inria/UPMC Team Regal. For a generalized case, there is a paper in preparation.

4.3.2. Bandwidth Allocation in Large Data Center

We are investigating a problem of efficient resource allocation in a large data center. In our model, the following is assumed. Each job that should be treated arrives to an M/M/C queue and is placed in it if the latter is not exhausted. Otherwise, it is sent to another queue for the possible implementation with the help of a certain canal, whose size is finite. A mean-field or the so called chaoticity result is established. Informally speaking, we show that the stochastic process that describes the evolution of our system converges to a non-random limit. We then study the stability properties of this limiting process and prove that it has a unique equilibrium that attracts exponentially all solutions that are issued from its small neighborhood. Moreover, we also show that if the size of the canal is infinite (i.e., the jobs go freely to another queue when not served), the

uniqueness for the fixed point problem is not guaranteed and, depending on some physical parameters, one can have no solution, a unique solution or two solutions. This phenomenon is quite surprising and it seems that is was not observed before. We also investigate the stability of equilibrium points. Some techniques used in our proofs come from theories developed in the context of PDEs.

4.4. Stochastic Models of Biological Networks

Participants: Renaud Dessalles, Philippe Robert, Wen Sun.

4.4.1. Stochastic Modelling of self-regulation in the protein production system of bacteria.

This is a collaboration with Vincent Fromion from INRA Jouy-en-Josas, which started in December 2013.

In prokaryotic cells (e.g. E. Coli. or B. Subtilis) the protein production system has to produce in a cell cycle (i.e. less than one hour) more than 10^6 molecules of more than 2500 kinds, each having different level of expression. The bacteria uses more than 67% of its resources to the protein production. Gene expression is a highly stochastic process: bacteria sharing the same genome, in a same environment will not produce exactly the same amount of a given protein. Some of this stochasticity can be due to the system of production itself: molecules, that take part in the production process, move freely into the cytoplasm and therefore reach any target in the cell after some random time; some of them are present in so much limited amount that none of them can be available for a certain time; the gene can be deactivated by repressors for a certain time, etc. We study the integration of several mechanisms of regulation and their performances in terms of variance and distribution. As all molecules tends to move freely into the cytoplasm, it is assumed that the encounter time between a given entity and its target is exponentially distributed.

4.4.1.1. Models with Cell Cycle

Usually, classical models of protein production do not explicitly represent several aspects of the cell cycle: the volume variations, the division and the gene replication. Yet these aspects have been proposed in literature to impact the protein production. We have therefore proposed a series of "gene-centered" models (that concentrates on the production of only one type of protein) that integrates successively all the aspects of the cell cycle. The goal is to obtain a realistic representation of the expression of one particular gene during the cell cycle. When it was possible, we analytically determined the mean and the variance of the protein concentration using Marked Poisson Point Process framework.

We based our analysis on a simple model where the volume changes across the cell cycle, and where only the mechanisms of protein production (transcription and translation) are represented. The variability predicted by this model is usually assimilated to the "intrinsic noise" (i.e. directly due to the protein production mechanism itself). We then add the random segregation of compounds at division to see its effect on protein variability: at division, every mRNA and every protein has an equal chance to go to either of the two daughter cells. It appears that this division sampling of compounds can add a significant variability to protein concentration. This effect directly depends on the relative variance (Fano factor) of the protein concentration: this effect is stronger as the relative variance is low. The dependence on the relative variance can be explained by considering a simplified model. With parameters deduced from real experimental measures, we estimate that the random segregation of compounds can double the variability of the genes with the lowest relative variance.

Finally, we integrate the gene replication to the model: at some point in the cell cycle, the gene is replicated, hence doubling the transcription rate. We are able to give analytical expressions for the mean and the variance of protein concentration at any moment of the cell cycle; it allows to directly compare the variance with the previous model with division. We show that gene replication has little impact on the protein variability: an environmental state decomposition shows that the part of the variance due to gene replication represents only at most 2% of the total variability predicted by the model.

Finally, we have investigated other possible sources of variability by presenting other simulations that integrate some specific aspects: variability in the production of RNA-polymerases and ribosomes, uncertainty in the division and DNA replication decisions, etc. None of the considered aspects seems to have a significant impact on the protein variability.

In the end, these results are compared to the real experimental measure of protein variability. It appears that the models with cell cycle presented above tend to underestimate the protein variability especially for highly expressed proteins. See Dessalles [1] and Dessalles et al. [17]

4.4.2. Stochastic Modelling of Protein Polymerization

This is a collaboration with Marie Doumic, Inria MAMBA team. The first part of our work focuses on the study of the polymerization of protein. This phenomenon is involved in many neurodegenerative diseases such as Alzheimer's and Prion diseases, e.g. mad cow. In this context, it consists in the abnormal aggregation of proteins. Curves obtained by measuring the quantity of polymers formed in in vitro experiments are sigmoids: a long lag phase with almost no polymers followed by a fast consumption of all monomers. Furthermore, repeating the experiment under the same initial conditions leads to somewhat identical curves up to translation. After having proposed a simple model to explain this fluctuations, we studied a more sophisticated model, closer to the reality. We added a conformation step: before being able to polymerize, proteins have to misfold. This step is very quick and remains at equilibrium during the whole process. Nevertheless, this equilibrium depends on the polymerization which is happening on a slower time scale. The analysis of these models involves stochastic averaging principles.

We have also investigated a more detailed model of polymerisation by considering the the evolution of the number of polymers with different sizes $(X_i(t))$ where $X_i(t)$ is the number of polymers of size i at time t. By assuming that the transitions rates are scaled by a large parameter N, it has been shown that, in the limit, the process $(X_i^N(t))$ is converging to the solution of Becker-Döring equations as N goes to infinity. For another model including nucleation, we have given an asymptotic description of the lag time at the first and second order. These results are obtained in particular by proving stochastic averaging theorems.

4.4.3. Central Limit Theorems

We have investigate the fluctuations of the stochastic Becker-Döring model of polymerization when the initial size of the system converges to infinity. A functional central limit problem is proved for the vector of the number of polymers of a given size. It is shown that the stochastic process associated to fluctuations is converging to the strong solution of an infinite dimensional stochastic differential equation (SDE) in a Hilbert space. We haveproved that, at equilibrium, the solution of this SDE is a Gaussian process. The proofs are based on a specific representation of the evolution equations, the introduction of a convenient Hilbert space and several technical estimates to control the fluctuations, especially of the first coordinate which interacts with all components of the infinite dimensional vector representing the state of the process. See Sun [21]

4.4.4. Study of the Nucleation Phenomenon

We have investigated a new stochastic model describing the time evolution of a polymerization process. The initial state of the system consists only of isolated monomers. We study the *lag time* of the polymerization process, that is, the first instant when a fraction of the initial monomers is polymerized, i.e. the fraction of monomers used in the polymers. The mathematical model includes *a nucleation property*: polymers with a size below some threshold n_c, the size of the nucleus, are quickly fragmented into smaller polymers. For a size greater than n_c, the fragmentation still occurs but at a smaller rate. A scaling approach is used, by taking the volume N of the system as a scaling parameter. If $n_c \geq 3$, under quite general assumptions on the way polymers are fragmented, we prove a limit theorem for the instant T^N of creation of the first "stable" polymer, i.e. a polymer of size n_c. It is proved that the distribution of T^N/N^{n_c-3} converges to an exponential distribution. We also show that, if $n_c \geq 4$, then the lag time has the same order of magnitude as T^N and, if $n_c = 3$, it is of the order of $\log N$. An original feature of our model is the significant variability (asymptotic exponential distribution) proved for the instants associated to polymerization. This is a well known phenomenon observed in the experiments in biology but it has not been really proved in appropriate mathematical models up to now. The results are proved via a series of (quite) delicate technical estimates for occupations measures on fast time scales associated to the first n_c coordinates of the corresponding Markov process. Extensive Stochastic calculus with Poisson processes, several coupling arguments and classical results from continuous branching processes theory are the main ingredients of the proofs.

5. Bilateral Contracts and Grants with Industry

5.1. Bilateral Contracts with Industry

- Contrat de recherche externalisé avec ORANGE SA "Scheduling Global OS". Duration three years 2014-2017.
- PhD grant CJS (Contrat Jeune Scientifique) Frontières du vivant of INRA for Renaud Dessalles.
- PhD grant from Fondation Sciences Mathématiques de Paris for Wen Sun.
- PhD grant from Brazilian Government for Guilherme Thompson.
- CELTIC+ Contract “SENDATE”.

6. Partnerships and Cooperations

6.1. International Research Visitors

6.1.1. Visits of International Scientists

- Louigi Addario-Berri (McGill)
- Gabor Lugosi (ICREA and Pompeu Fabra)
- Cyril Marzouk (Paris Sud)
- Minmin Wang (Buenos Aires)
- S. Foss (Heriot-Watt University, UK)
- V. Gupta (University of Chicago, USA)

6.1.2. Visits to International Teams

- *Nicolas Broutin* visited the computer science department of McGill University (Canada), the CRM in Montreal, the mathematics institute in Nice and the university Aix-Marseille.
- *Wen Sun* has visited the Division of Applied Mathematics in Brown University to work with Kavita Ramanan, 07-16 Nov. 2017.

7. Dissemination

7.1. Promoting Scientific Activities

7.1.1. Journal

7.1.1.1. Member of the Editorial Boards

- *Nicolas Broutin* is member of the steering committee of the international meeting on analysis of algorithms (AofA).
- *Philippe Robert* is Associate Editor of the Book Series “Mathématiques et Applications” edited by Springer Verlag and Associate Editor of the journal “Queueing Systems, Theory and Applications”.

7.1.1.2. Member of the Conference Program Committees

- *Philippe Robert* has been in the TPC of Performance’2017, Caching and Delivery in Wireless Networks Workshop (CCDWN)’2017 and Stochastic Networks’2018.

7.1.2. Conferences

- *Nicolas Broutin* has given lectures at the probability seminar in Marseille on Jan 9, the probability seminar in Nice on March 21, and the annual days of the ANR project GRAAL on May 10.

- *Christine Fricker* has given a talk "Analysis of large scale closed networks with reservation" at the 2017 INFORMS 19th Applied Probability Conference, Northwestern University, USA on 10-12 July, 2017. *Christine Fricker*gave a talk at the day "Filles et math: une équation lumineuse" for 100 high school girls organised by Animath and Femmes et maths (19/12/2017).
- *Wen Sun* has given a talk "A Large Scale Network with Moving Servers" on 5-9 June 2017 at the Sigmetrics workshop MAMA in Urbana-Champaign, Illinois, USA and a talk "An asymptotic study of allocation policies in a large balls and bins model" at the 2017 INFORMS 19th Applied Probability Conference, Northwestern University, USA on 10-12 July, 2017.
- *Guilherme Thompson* has given a talk "Studying an offloading policy for multi-resource Cloud services under Kelly's Regime" at the 2017 INFORMS 19th Applied Probability Conference, Northwestern University, USA on 10-12 July, 2017. *Philippe Robert* has given a talk "Asymptotics of Stochastic Protein Assembly Models" at the PDMP workshop in Seillac, France, June 2017. He has given a talk "Analysis of a Stochastic Model of Replication in Large Distributed Storage Systems: A Mean-Field Approach" at the ACM Sigmetrics conference, Urbana-Champaign, USA on 5-9 June 2017, and also a talk "A Large Scale Analysis of Unreliable Stochastic Networks" at the 2017 INFORMS 19th Applied Probability Conference, Northwestern University, USA on 10-12 July, 2017. He has given a talk "Stochastic Models of Gene Expression" at the SBF-Kolloquium, Institut für Mathematik, Universität Potsdam, October 2017. He gave a seminar "Asymptotic Studies of Large Distributed systems with Failures" at Inria Rhône-Alpes (Polaris team).

7.1.3. Scientific Expertise

- *Christine Fricker* is member of the jury of agrégation

7.2. Teaching - Supervision - Juries

7.2.1. Teaching

- Master : *Nicolas Broutin* Master Parisien de Recherche en Informatique (MPRI), in the course 2.15 on Analysis of Algorithms.
- *Philippe Robert* Master, Probabilités et Applications, UPMC

7.2.2. Supervision

- *Renaud Dessalles* [17] (January 2017) and *Guilherme Thompson* [2] (December 2017) have defended their PhD thesis.
- *Christine Fricker* got her habilitation to supervise research on 20/11/2017. The manuscript is entitled "Stochastic Networks". Reviewers: Alexandre Proutière (professor, KTH Stockohlm), Isi Mitrani (professor, Newcastle University), Laurent Massoulié (research position, Inria-Microsoft Research leader). Jury: Alexandre Proutière (professor, KTH Stockohlm), Laurent Massoulié (Inria-Microsoft Research leader), reviewers, Irina Kurkova (professor, UPMC), Laurent Decreusefond (professor, Telecom Paristech), Carl Graham (research position, CNRS, Ecole Polytechnique).

7.2.3. Juries

Philippe Robert has been a member of the juries of PhD defenses by R. Dessalles (Ecole Polytechnique, January 2017), Y. Petot (Nancy, October 2017), G. Thompson (UPMC< December 2017) and V. Bœuf (Saclay, December 2017).

8. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] R. DESSALLES.*Stochastic models for protein production: the impact of autoregulation, cell cycle and protein production interactions on gene expression*, École Polytechnique, January 2017, https://hal.inria.fr/tel-01482087.

[2] G. R. THOMPSON.*Stochastic Models for Resource Allocation in Large Distributed Systems*, Université Pierre et Marie Curie, December 2017, https://hal.inria.fr/tel-01661815.

Articles in International Peer-Reviewed Journal

[3] L. ADDARIO-BERRY, N. BROUTIN, C. GOLDSCHMIDT, G. MIERMONT.*The scaling limit of the minimum spanning tree of the complete graph*, in "Annals of Probability", 2017, vol. 45, p. 3075–3144, https://arxiv.org/abs/1301.1664 , https://hal.inria.fr/hal-00773360.

[4] 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", 2017, https://arxiv.org/abs/1608.08743 , https://hal.archives-ouvertes.fr/hal-01359208.

[5] N. BROUTIN, R. J. KANG.*Bounded monochromatic components for random graphs*, in "Journal of Combinatorics", 2017, https://arxiv.org/abs/1407.3555 - 20 pages, 1 figure, https://hal.archives-ouvertes.fr/hal-01056126.

[6] N. BROUTIN, C. MAILLER.*And/or trees: a local limit point of view*, in "Random Structures and Algorithms", 2017, https://arxiv.org/abs/1510.06691 , https://hal.inria.fr/hal-01220794.

[7] N. BROUTIN, M. WANG.*Reversing the cut tree of the Brownian continuum random tree*, in "Electronic Journal of Probability", August 2017, vol. 22, n^{o} 80, p. 1-23, https://arxiv.org/abs/1408.2924 , https://hal.archives-ouvertes.fr/hal-01056124.

[8] R. DESSALLES, V. FROMION, P. ROBERT.*A Stochastic Analysis of Autoregulation of Gene Expression*, in "Journal of Mathematical Biology", March 2017, https://arxiv.org/abs/1509.02045 , https://hal.inria.fr/hal-01203076.

[9] P. S. DESTER, C. FRICKER, D. TIBI.*Stationary analysis of the shortest queue problem*, in "Queueing Systems", 2017, vol. 87, n^{o} 3-4, p. 211-243, https://arxiv.org/abs/1704.06442 [*DOI :* 10.1007/s11134-017-9556-8], https://hal.inria.fr/hal-01666312.

[10] S. EUGENE, T. BOURGERON, Z. XU.*Effects of initial telomere length distribution on senescence onset and heterogeneity*, in "Journal of Theoretical Biology", January 2017, vol. 413, 8, https://arxiv.org/abs/1606.06842 , https://hal.inria.fr/hal-01378596.

[11] C. FRICKER, F. GUILLEMIN, P. ROBERT, G. R. THOMPSON.*Allocation Schemes of Ressources with Downgrading*, in "Advances in Applied Probability", April 2017, vol. 49, n^{o} 2, 22, https://arxiv.org/abs/1604.00894 [*DOI :* 10.1017/APR.2017.15], https://hal.inria.fr/hal-01301291.

[12] C. FRICKER, D. TIBI.*Equivalence of ensembles for large vehicle-sharing models*, in "The Annals of Applied Probability : an official journal of the institute of mathematical statistics", 2017, vol. 27, n^{o} 2, p. 883-916, https://arxiv.org/abs/1507.07792 , https://hal.inria.fr/hal-01203789.

[13] A. GANGULY, K. RAMANAN, P. ROBERT, W. SUN.*A Large-Scale Network with Moving Servers*, in "ACM SIGMETRICS Performance Evaluation Review", October 2017, vol. 45, n^{o} 2, p. 42 - 44 [*DOI :* 10.1145/3152042.3152057], https://hal.inria.fr/hal-01644146.

Articles in National Peer-Reviewed Journal

[14] P. ROBERT.*Jacques Neveu et les modèles probabilistes de réseaux*, in "Matapli", March 2017, vol. 112, 7, https://hal.inria.fr/hal-01382215.

International Conferences with Proceedings

[15] W. SUN, V. SIMON, S. MONNET, P. ROBERT, P. SENS.*Analysis of a Stochastic Model of Replication in Large Distributed Storage Systems: A Mean-Field Approach*, in "ACM Sigmetrics 2017- International Conference on Measurement and Modeling of Computer Systems", Urbana-Champaign, Illinois, United States, ACM, June 2017, p. 51–51, https://arxiv.org/abs/1701.00335 [*DOI :* 10.1145/3078505.3078531], https://hal.inria.fr/hal-01494235.

Other Publications

[16] V. BOEUF, P. ROBERT.*A Stochastic Analysis of a Network with Two Levels of Service*, August 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01583704.

[17] R. DESSALLES, V. FROMION, P. ROBERT.*Models of protein production with cell cycle*, November 2017, working paper or preprint, https://hal.inria.fr/hal-01643441.

[18] R. EL SIBAI, Y. CHABCHOUB, C. FRICKER.*Using spatial outliers detection to assess balancing mechanisms in bike sharing systems*, December 2017, Soumis à The 32-nd IEEE International Conference on Advanced Information Networking and Applications, https://hal.inria.fr/hal-01666345.

[19] C. FRICKER, P. S. DESTER, H. MOHAMED.*Balancing queues with a random local choice algorithm*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01666326.

[20] P. ROBERT, W. SUN.*On the Asymptotic Distribution of Nucleation Times of Polymerization Processes*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01672800.

[21] W. SUN.*A Functional Central Limit Theorem for the Becker-Döring model*, October 2017, https://arxiv.org/abs/1710.04059 - 18 pages [*DOI :* 10.04059], https://hal.inria.fr/hal-01616039.

Project-Team REGAL

Large-Scale Distributed Systems and Applications

IN COLLABORATION WITH: Laboratoire d'informatique de Paris 6 (LIP6)

IN PARTNERSHIP WITH:
CNRS

Université Pierre et Marie Curie (Paris 6)

RESEARCH CENTER
Paris

THEME
Distributed Systems and middleware

Table of contents

Project-Team REGAL

Creation of the Project-Team: 2005 July 01, end of the Project-Team: 2017 December 31

Keywords:

Computer Science and Digital Science:

A1.1.1. - Multicore, Manycore
A1.1.6. - Cloud
A1.1.7. - Peer to peer
A1.1.9. - Fault tolerant systems
A1.1.13. - Virtualization
A1.3. - Distributed Systems
A1.6. - Green Computing
A2.6. - Infrastructure software
A2.6.1. - Operating systems
A2.6.2. - Middleware
A2.6.3. - Virtual machines
A3.1.3. - Distributed data
A3.1.8. - Big data (production, storage, transfer)
A7.1. - Algorithms

Other Research Topics and Application Domains:

B4.5. - Energy consumption
B6.4. - Internet of things
B6.6. - Embedded systems
B8.2. - Connected city
B9.2.3. - Video games
B9.4.1. - Computer science

1. Personnel

Research Scientists

Mesaac Makpangou [Inria, Researcher, HDR]
Marc Shapiro [Inria, Senior Researcher, HDR]

Faculty Members

Luciana Bezerra Arantes [Univ Pierre et Marie Curie, Associate Professor]
Philippe Darche [Univ René Descartes, Associate Professor]
Swan Dubois [Univ Pierre et Marie Curie, Associate Professor]
Jonathan Lejeune [Univ Pierre et Marie Curie, Associate Professor]
Franck Petit [Univ Pierre et Marie Curie, Professor, HDR]
Pierre Sens [Team leader, Univ Pierre et Marie Curie, Professor, HDR]
Julien Sopena [Univ Pierre et Marie Curie, Associate Professor]

External Collaborator

Sébastien Monnet [Université Savoie Mont-Blanc]

Technical Staff

Sreeja Nair [Univ Pierre et Marie Curie]

PhD Students

Sébastien Bouchard [Inria]
Marjorie Bournat [Univ Pierre et Marie Curie]
Damien Carver [Magency]
João Paulo de Araujo [Univ Pierre et Marie Curie]
Guillaume Fraysse [IO Lab, Orange Lab]
Lyes Hamidouche [Magency]
Denis Jeanneau [Univ Pierre et Marie Curie]
Francis Laniel [Univ Pierre et Marie Curie]
Vinh Tao Thanh [Scality, until Dec 2017]
Alejandro Tomsic [Inria]
Ilyas Toumlilt [Univ Pierre et Marie Curie]
Dimitrios Vasilas [Scality]
Gauthier Voron [Univ Pierre et Marie Curie]

Post-Doctoral Fellow

Paolo Viotti [Univ Pierre et Marie Curie]

Administrative Assistant

Nelly Maloisel [Inria]

2. Overall Objectives

2.1. Overall Objectives

The research of the Regal team addresses the theory and practice of *Computer Systems,* including multicore computers, clusters, networks, peer-to-peer systems, cloud computing systems, and other communicating entities such as swarms of robots. It addresses the challenges of communicating, sharing information, and computing correctly 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.

Regal is a joint research team between LIP6 (UPMC/CNRS) and Inria Paris.

3. Research Program

3.1. Research rationale

The research of Regal addresses both theoretical and practical issues of *Computer Systems*, i.e., its goal is a dual expertise in theoretical and experimental research. Our approach is a "virtuous cycle" of algorithm design triggered by issues with real systems, which we prove correct and evaluate theoretically, and then eventually implement and test experimentally.

Regal's major challenges comprise communication, sharing of information, and correct execution in large-scale and/or highly dynamic computer systems. While Regal's historically focused in static distributed systems, since some years ago we have covered a larger spectrum of distributed computer systems: multicore computers, clusters, mobile networks, peer-to-peer systems, cloud computing systems, and other communicating entities such as swarms of robots. This holistic approach allows the handling of related problems at different levels. Among such problems we can highlight communication between cores, consensus, fault detection, scalability, search and diffusion of information, allocation resource, replication and consistency of shared data, dynamic content distribution, and multi-core concurrent algorithms.

Computer Systems is a rapidly evolving domain, with strong interactions with industry and modern computer systems, which are increasingly distributed. Ensuring persistence, availability, and consistency of data in a distributed setting is a major requirement: the system must remain correct despite slow networks, disconnection, crashes, failures, churn, and attacks. Easiness of use, performance, and efficiency are equally fundamental. However, these requirements are somewhat conflicting, and there are many algorithmic and engineering trade-offs, which often depend on specific workloads or usage scenarios. At the same time, years of research in distributed systems are now coming to fruition, and are being used by millions of users of web systems, peer-to-peer systems, gaming and social applications, or cloud computing. These new usages bring new challenges of extreme scalability and adaptation to dynamically-changing conditions, where knowledge of the system state might only be partial and incomplete. Therefore, the scientific challenges of the distributed computing systems listed above are subject to additional trade-offs which include scalability, fault tolerance, dynamics, and virtualization of physical infrastructure. 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.

In in particular, Regal focuses on three key challenges:

- the adaptation of algorithms to the new dynamics of distributed systems;
- data management on extreme large configurations;
- the adaptation of execution support to new multi-core architectures.

We should emphasize that these challenges are complementary: the two first challenges aim at building new distributed algorithms and strategies for large and dynamic distributed configurations whereas the last one focusses on the scalability of internal OS mechanisms.

4. New Software and Platforms

4.1. Antidote

KEYWORDS: Distributed computing - Distributed Data Management - Cloud storage - Large scale
FUNCTIONAL DESCRIPTION: Antidote is the flexible cloud database platform currently under development in the SyncFree and LightKone European projects. Antidote aims to be both a research platform for studying replication and consistency at the large scale, and an instrument for exploiting research results. The platform supports replication of CRDTs, in and between sharded (partitioned) data centres (DCs). The current stable version supports strong transactional consistency inside a DC, and causal transactional consistency between DCs. Ongoing research includes support for explicit consistency , for elastic version management, for adaptive replication, for partial replication, and for reconfigurable sharding.

- Participants: Marc Shapiro, Paolo Viotti, Alejandro Tomsic, Ilyas Toumlilt and Dimitrios Vasilas
- Partners: Université Catholique de Louvain (UCL), Louvain-la-Neuve,Belgium - Universidade NOVA de Lisboa, Portugal - Technische Universität Kaiserslautern (UniKL), Allemagne
- Contact: Marc Shapiro
- Publications: Bringing the cloud closer to users - Write Fast, Read in the Past: Causal Consistency for Client-side Applications - Extending Eventually Consistent Cloud Databases for Enforcing Numeric Invariants - Designing a causally consistent protocol for geo-distributed partial replication - Towards Fast Invariant Preservation in Geo-replicated Systems - Putting Consistency back into Eventual Consistency - The Case for Fast and Invariant-Preserving Geo-Replication - Improving the scalability of geo-replication with reservations - Conflict-free Replicated Data Types - An encounter with Marc Shapiro and his SyncFree European project - PhysiCS-NMSI: efficient consistent snapshots for scalable snapshot isolation - Geo-Replication: Fast If Possible, Consistent If Necessary - Cure: Strong semantics meets high availability and low latency - Cure: Strong semantics meets high availability and low latency

4.2. CISE Tool

KEYWORDS: Distributed Applications - Program verification
FUNCTIONAL DESCRIPTION: Static analysis of the model of a distributed application, to prove (under the assumption of causal consistency) whether the invariants of the application are always satisfied, and to provide a counterexample if not.

- Participants: Sreeja Nair and Marc Shapiro
- Contact: Marc Shapiro
- Publications: Evaluation of the CEC (Correct Eventual Consistency) Tool - The CISE Tool: Proving Weakly-Consistent Applications Correct - The CISE Tool: Proving Weakly-Consistent Applications Correct - CISE Safety Tool - 'Cause I'm Strong Enough: Reasoning about Consistency Choices in Distributed Systems - Putting Consistency back into Eventual Consistency

4.3. PUMA

Puma: pooling unused memory in virtual machines
KEYWORDS: Virtualization - Operating system - Distributed systems - Linux kernel
FUNCTIONAL DESCRIPTION: PUMA is a system that is based on a kernel-level remote caching mechanism that provides the ability to pool VMs memory at the scale of a data center. An important property while lending memory to another VM, is the ability to quickly retrieve memory in case of need. Our approach aims at lending memory only for clean cache pages: in case of need, the VM which lent the memory can retrieve it easily. We use the system page cache to store remote pages such that: (i) if local processes allocate memory the borrowed memory can be retrieved immediately, and (ii) if they need cache the remote pages have a lower priority than the local ones.

- Participants: Maxime Lorrillere, Julien Sopena and Pierre Sens
- Partner: LIP6
- Contact: Julien Sopena
- Publications: Conception et évaluation d'un système de cache réparti adapté aux environnements virtualisés - Puma: pooling unused memory in virtual machines for I/O intensive applications
- URL: https://github.com/mlorrillere/puma

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, 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 2017, we obtained the following results.

5.1.1. Algorithms for Dynamic and Large Systems

In [32] we propose VCube-PS, a new topic-based Publish/Subscribe system built on the top of a virtual hypercube like topology. Membership information and published messages to subscribers (members) of a topic group are broadcast over dynamically built spanning trees rooted at the message's source. For a given topic, delivery of published messages respects causal order. Performance results of experiments conducted on the PeerSim simulator confirm the efficiency of VCube-PS in terms of scalability, latency, number, and size of messages when compared to a single rooted, not dynamically, tree built approach.

We also explore in [20] scheduling challenges in providing probabilistic Byzantine fault tolerance in a hybrid cloud environment, consisting of nodes with varying reliability levels, compute power, and monetary cost. In this context, the probabilistic Byzantine fault tolerance guarantee refers to the confidence level that the result of a given computation is correct despite potential Byzantine failures. We formally define a family of such scheduling problems distinguished by whether they insist on meeting a given latency limit and trying to optimize the monetary budget or vice versa. For the case where the latency bound is a restriction and the budget should be optimized, we propose several heuristic protocols and compare between them using extensive simulations.

In [27], we a new resource reservation protocol in the context of delay-sensitive rescue mobile networks. The search for service providers (e.g., ambulance, fire truck, etc.) after a disaster, must take place within a short time. Therefore, service discovery protocol which looks for providers that can attend victims, respecting time constraints, is crucial. In such a situation, a commonly solution for ensuring network connectivity between victims and providers is ad hoc networks (MANET), composed by battery-operated mobile nodes of persons (victims or not). Using message aggregations techniques, we propose an new reservation protocol aiming at reducing the number of messages over the network and, therefore, node's battery consumption

5.1.2. Self-Stabilization

Self-stabilization is a generic paradigm to tolerate transient faults (*i.e.*, faults of finite duration) in distributed systems. In [14], we propose a silent self-stabilizing leader election algorithm for bidirectional arbitrary connected identified networks. This algorithm is written in the locally shared memory model under the distributed unfair daemon. It requires no global knowledge on the network. Its stabilization time is in $\Theta(n^3)$ steps in the worst case, where n is the number of processes. Its memory requirement is asymptotically optimal, *i.e.*, $\Theta(\log n)$ bits per processes. Its round complexity is of the same order of magnitude — *i.e.*, $\Theta(n)$ rounds — as the best existing algorithms designed with similar settings. To the best of our knowledge, this is the first self-stabilizing leader election algorithm for arbitrary identified networks that is proven to achieve a stabilization time polynomial in steps. By contrast, we show that the previous best existing algorithms designed with similar settings stabilize in a non polynomial number of steps in the worst case.

5.1.3. Mobile Agents

In [21], we consider systems made of autonomous mobile robots evolving in highly dynamic discrete environment *i.e.*, graphs where edges may appear and disappear unpredictably without any recurrence, stability, nor periodicity assumption. Robots are uniform (they execute the same algorithm), they are anonymous (they are devoid of any observable ID), they have no means allowing them to communicate together, they share no common sense of direction, and they have no global knowledge related to the size of the environment. However, each of them is endowed with persistent memory and is able to detect whether it stands alone at its current location. A highly dynamic environment is modeled by a graph such that its topology keeps continuously changing over time. In this paper, we consider only dynamic graphs in which nodes are anonymous, each of them is infinitely often reachable from any other one, and such that its underlying graph (*i.e.*, the static graph made of the same set of nodes and that includes all edges that are present at least once over time) forms a ring of arbitrary size.

In this context, we consider the fundamental problem of *perpetual exploration*: each node is required to be infinitely often visited by a robot. This paper analyzes the computability of this problem in (fully) synchronous settings, *i.e.*, we study the deterministic solvability of the problem with respect to the number of robots. We

provide three algorithms and two impossibility results that characterize, for any ring size, the necessary and sufficient number of robots to perform perpetual exploration of highly dynamic rings.

5.1.4. Approach in the Plane

In [35] we study the task of *approach* of two mobile agents having the same limited range of vision and moving asynchronously in the plane. This task consists in getting them in finite time within each other's range of vision. The agents execute the same deterministic algorithm and are assumed to have a compass showing the cardinal directions as well as a unit measure. On the other hand, they do not share any global coordinates system (like GPS), cannot communicate and have distinct labels. Each agent knows its label but does not know the label of the other agent or the initial position of the other agent relative to its own. The route of an agent is a sequence of segments that are subsequently traversed in order to achieve approach. For each agent, the computation of its route depends only on its algorithm and its label. An adversary chooses the initial positions of both agents in the plane and controls the way each of them moves along every segment of the routes, in particular by arbitrarily varying the speeds of the agents. Roughly speaking, the goal of the adversary is to prevent the agents from solving the task, or at least to ensure that the agents have covered as much distance as possible before seeing each other. A deterministic approach algorithm is a deterministic algorithm that always allows two agents with any distinct labels to solve the task of approach regardless of the choices and the behavior of the adversary. The cost of a complete execution of an approach algorithm is the length of both parts of route travelled by the agents until approach is completed.

Let Δ and l be the initial distance separating the agents and the length of (the binary representation of) the shortest label, respectively. *Assuming that Δ and l are unknown to both agents, does there exist a deterministic approach algorithm whose cost is polynomial in Δ and l?*

Actually the problem of approach in the plane reduces to the network problem of rendezvous in an infinite oriented grid, which consists in ensuring that both agents end up meeting at the same time at a node or on an edge of the grid. By designing such a rendezvous algorithm with appropriate properties, as we do in this paper, we provide a positive answer to the above question.

Our result turns out to be an important step forward from a computational point of view, as the other algorithms allowing to solve the same problem either have an exponential cost in the initial separating distance and in the labels of the agents, or require each agent to know its starting position in a global system of coordinates, or only work under a much less powerful adversary.

5.2. Large scale data distribution

Participants: Mesaac Makpangou, Sébastien Monnet, Pierre Sens, Marc Shapiro, Paolo Viotti, Sreeja Nair, Ilyas Toumlilt, Alejandro Tomsic, Dimitrios Vasilas.

5.2.1. Data placement and searches over large distributed storage

Distributed storage systems such as Hadoop File System or Google File System (GFS) ensure data availability and durability using replication. Persistence is achieved by replicating the same data block on several nodes, and ensuring that a minimum number of copies are available on the system at any time. Whenever the contents of a node are lost, for instance due to a hard disk crash, the system regenerates the data blocks stored before the failure by transferring them from the remaining replicas. In [33] we focused on the analysis of the efficiency of replication mechanism that determines the location of the copies of a given file at some server. The variability of the loads of the nodes of the network is investigated for several policies. Three replication mechanisms are tested against simulations in the context of a real implementation of a such a system: Random, Least Loaded and Power of Choice. The simulations show that some of these policies may lead to quite unbalanced situations. It is shown in this paper that a simple variant of a power of choice type algorithm has a striking effect on the loads of the nodes. Mathematical models are introduced and investigated to explain this interesting phenomenon. The analysis of these systems turns out to be quite complicated mainly because of the large dimensionality of the state spaces involved. Our study relies on probabilistic methods, mean-field analysis, to analyze the asymptotic behavior of an arbitrary node of the network when the total number of nodes gets large.

In the summary prefix tree (SPT), a trie data structure that supports efficient superset searches over DHT. Each document is summarized by a Bloom filter which is then used by SPT to index this document. SPT implements an hybrid lookup procedure that is well-adapted to sparse indexing keys such as Bloom filters. It also proposes a mapping function that permits to mitigate the impact of the skewness of SPT due to the sparsity of Bloom filters, especially when they contain only few words. To perform efficient superset searches, SPT maintains on each node a local view of the global tree. The main contributions are the following. First, the approximation of the superset relationship among keyword-sets by the descendance relationship among Bloom filters. Second, the use of a summary prefix tree (SPT), a trie indexing data structure, for keyword-based search over DHT. Third, an hybrid lookup procedure which exploits the sparsity of Bloom filters to offer good performances. Finally, an algorithm that exploits SPT to efficiently find descriptions that are supersets of query keywords.

5.2.2. Just-Right Consistency

Consistency is a major concern in the design of distributed applications, but the topic is still not well understood. It is clear that no single consistency model is appropriate for all applications, but how do developers find their way in the maze of models and the inherent trade-offs between correctness and availability? The Just-Right Consistency approach presented here offers some guidance. First, we classify the safety patterns that are of interest to maintain application correctness. Second, we show how two of these patterns are "AP-compatible" and can be guaranteed without impacting availability, thanks to an appropriate data model and consistency model. Then we address the last, "CAP-sensitive" pattern. In a restricted but common case it can be maintained efficiently in a mostly-available way. In the general case, we exhibit a static analysis logic and tool which ensures just enough synchronisation to maintain the invariant, and availability otherwise.

In summary, instead of pre-defining a consistency model and shoe-horning the application to fit it, and instead of making the application developer compensate for the imperfections of the data store in an *ad-hoc* way, we have a provably correct approach to tailoring consistency to the specific application requirements. This approach is supported by several artefacts developed by Regal and collaborators: Conflict-Free Replicated Data Types (CRDTs), the Antidote cloud database, and the CISE verification tool.

This paper is under submission.

5.3. Memory management in system software

Participants: Damien Carver, Jonathan Lejeune, Pierre Sens, Julien Sopena [correspondent], Gauthier Voron.

Recent years have seen the increasingly widespread use of **multicore** architectures and **virtualized environments**. This development has an impact on all parts of the system software. Virtual machine (VM) technology offers both isolation and flexibility but has side effects such as fragmentation of the physical resources, including memory. This fragmentation reduces the amount of available memory a VM can use. Many recent works study that a NUMA (Non Uniform Memory Access) architecture, common in large multi-core processors, highly impacts application performance. We focus on improving the memory and cache management in various virtualized environments such as Xen hypervisor or linux-containers targeting big data applications on multicore architectures.

While virtualization only introduces a small overhead on machines with few cores, this is not the case on larger ones. Most of the overhead on the latter machines is caused by the NUMA architecture they are using. In order to reduce this overhead, in [34] we show how NUMA placement heuristics can be implemented inside Xen. With an evaluation of 29 applications on a 48-core machine, we show that the NUMA placement heuristics can multiply the performance of 9 applications by more than 2.

We also study the memory arbitration between containers. In the Damien Carver's PhD thesis, we are designing ACDC [23] (Advanced Consolidation for Dynamic Containers), a kernel-level mechanisms that automatically provides more memory to the most active containers.

In the Francis Laniel's PhD thesis, we study a new architecture using Non Volatile RAM NVRAM. Although NVRAM are slower than classical RAM, they have better energetic features. We investigate solutions where RAM and NVRAM coexist in order to balance the energy consumption and performance according to the needs of the system.

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral Contracts with Industry

Regal has two CIFRE contracts with Scality SA:

- Vinh Tao is advised by Marc Shapiro and Vianney Rancurel. He works on highly available geo-replicated file systems, building on CRDT technology. He defended his thesis in December 2017.
- Dimitrios Vasilas is advised by Marc Shapiro and Brad King. He works on secondary indexing in large-scale storage systems under weak consistency.

Regal 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.

Regal has one contract 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.

7. Partnerships and Cooperations

7.1. National Initiatives

7.1.1. Labex SMART - (2012–2019)

Members: ISIR (UPMC/CNRS), LIP6 (UPMC/CNRS), LIB (UPMC/INSERM), LJLL (UPMC/CNRS), LTCI (Institut Mines-Télécom/CNRS), CHArt-LUTIN (Univ. Paris 8/EPHE), L2E (UPMC), 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 Regal is responsible of the axe "Autonomic Distributed Environments for Mobility."

The project involves a PhD grant of 100 000 euros over 3 years.

7.1.2. ESTATE - (2016–2020)

Members: LIP6 (Regal, 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 Regal.

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.

7.1.3. RainbowFS - (2016–2020)

Members: LIP6 (Regal, 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 Regal.

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.

7.2. 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. CNRS-Inria-FAP's

Title: Autonomic and Scalable Algorithms for Building Resilient Distributed Systems

International Partner (Institution - Laboratory - Researcher):

Universida de Federal do Paraná (UFPR), Brazil, Prof. Elias Duarte

Duration: 2015–2017

In the context of autonomic computing systems that detect and diagnose problems, self-adapting themselves, the VCube (Virtual Cube), proposed by Prof. Elias Duarte , is a distributed diagnosis algorithm that organizes the system nodes on a virtual hypercube topology. VCube has logarithmic properties: when all nodes are fault-free, processes are virtually connected to form a perfect hypercube; as soon as one or more failures are detected, links are automatically reconnected to remove the faulty nodes and the resulting topology, connecting only fault-free nodes, keeps the logarithmic properties. The goal of this project is to exploit the autonomic and logarithmic properties of the VCube by proposing self-adapting and self-configurable services.

7.3.1.3. 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.

7.3.1.4. 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.

7.4. International Research Visitors

7.4.1. Visits of International Scientists

7.4.1.1. Internships

Ajay Singh of Indian Institute Of Technology Hyderabad, India, was invited for a six-month internship, on data structures for concurrency and persistent memory. This work is published at the HiPC SRS 2017 workshop [43].

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, Co-organizer (with Arnaud Casteigts, University of Bordeaux) of the Second Workshop on Computing in Dynamic Networks, in conjunction with DISC'17, Vienna, Austria, October 20th, 2017.

8.1.2. Scientific Events Selection

8.1.2.1. Member of the Conference Program Committees

Pierre Sens, 28th International Symposium on Software Reliability Engineering (ISSRE'2017), IEEE 46th International Conference on Parallel Processing (ICPP'2017), 16th IEEE International Symposium on Network Computing and Applications (NCA 2017).

Marc Shapiro, Steering Committee of Int. Conf. on Principles of Distributed Systems (OPODIS).

Marc Shapiro, Steering Committed of Workshop on Principles and Practice of Consistency for Distr. Data (PaPoC).

Swan Dubois, 19th Workshop on Advances in Parallel and Distributed Computational Models (APDCM'2017), 5th International Symposium on Computing and Networking (CANDAR'2017).

Luciana Arantes, 16th IEEE International Symposium on Network Computing and Applications (NCA 2017), 13th European Dependable Computing Conference (EDCC 2017), 17th IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS 2017)

8.1.2.2. Reviewer

- Marc Shapiro, reviewer for Int. Conf. on Middleware (MIDDLEWARE 2017).
- Swan Dubois, 31st International Symposium on Distributed Computing (DISC'2017), 24th International Colloquium on Structural Information and Communication Complexity (SIROCCO'2017).
- Luciana Arantes, ACM Symposium on Principles of Distributed Computing (PODC 2017), 23rd International European Conference on Parallel and Distributed Computing (Euro-Par 2017).

8.1.3. Journal

8.1.3.1. Member of the Editorial Boards

Pierre Sens, Associate editor of International Journal of High Performance Computing and Networking (IJHPCN)

Lucian Arantes, Special Issue of Concurrency and Computation: Practice and Experience, Volume 29, january 2017

8.1.3.2. Reviewer - Reviewing Activities

- Marc Shapiro, reviewer IEEE Transactions on Software Engineering.
- Swan Dubois, Theoretical Computer Science (TCS), Theory of Computing Systems (TOCS), International Journal of Networking and Computing (IJNC).
- Luciana Arantes, reviewer Journal of Parallel and Distributed Systems (JPDC).

8.1.4. Invited Talks

Pierre Sens, *Failure detection in large and dynamic distributed systems*. Univ. Curitiba, Brazil. October 2017.

Pierre Sens, *Failure detection in large and dynamic distributed systems*. Keynote speaker, GDR RSD and ASF Winter School on Distributed Systems, Pleynet, Sept Laux, Mars, 2017

Marc Shapiro, *The Antidote Cloud Database*. Comité de pilotage Groupe de Travail Logiciels Libres, Paris, Jan. 2017.

Marc Shapiro, *Just-Right Consistency*. Invited talk, joint session of PaPoC and LADIS workshops, EuroSys April 2017. Belgrade, Serbia.

Marc Shapiro, *AntidoteDB: a developer-friendly, open-source cloud database*. Invited talk, Datageeks Paris, Vente Privée la Plaine-Saint-Denis, May 2017.

Marc Shapiro, *AntidoteDB: a developer-friendly, open-source cloud database*. Invited talk, Open Source Innovation Spring, Paris May 2017.

Marc Shapiro, *Just-Right Consistency*. *La demie-heure de science*, Inria Paris, Oct. 2017.

Marc Shapiro, *AntidoteDB : Une base de données nuage pour la juste cohérence.* Paris Open Source Summit, Saint-Denis, Nov. 2017.

Marc Shapiro, *Semantics and proof of geo-replicated file system*. Invited talk, ENS Ulm, Nov. 2017.

Marc Shapiro, Invited speaker, workshop of Verification of Distributed Systems, Essaouira, May 2018.

8.1.5. Scientific Expertise

Pierre Sens, Project in Indo-French Centre for the Promotion of Advancded Research

Marc Shapiro, reviewer for ERC Starting Grant panel PE-6.

Marc Shapiro, invited for CACM Technical Perspective [18].

Marc Shapiro, invited for several entries in Springer Encyclopedia of Database Systems [36], [37], [38] and Encyclopedia of Big Data Technologies (to appear).

Marc Shapiro, reviewer for Irish Research Council.

Marc shapiro, reviewer for Indo French Centre for the Promotion of Advanced Research (CEFIPRA).

8.1.6. Research Administration

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.

Pierre Sens, since 2015, officer at scientific research vice presidency UPMC

Pierre Sens, since 2014: Member of Steering Committee of International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD).

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Julien Sopena is Member of "Directoire des formations et de l'insertion professionnelle" of UPMC Sorbonne Universités, France

Master: Julien Sopena is responsible of Computer Science Master's degree in Distributed systems and applications (in French, SAR), UPMC Sorbonne Universités, France

Master: Luciana Arantes, Swan Dubois, Jonathan Lejeune, Franck Petit, Pierre Sens, Julien Sopena, Advanced distributed algorithms, M2, UPMC Sorbonne Universités, France

Master: Jonathan Lejeune, Designing Large-Scale Distributed Applications, M2, UPMC Sorbonne Universités, France

Master: Maxime Lorrillere, Julien Sopena, Linux Kernel Programming, M1, UPMC Sorbonne Universités, France

Master: Luciana Arantes, Swan Dubois, Jonathan Lejeune, Pierre Sens, Julien Sopena, Operating systems kernel, M1, UPMC Sorbonne Universités, France

Master: Luciana Arantes, System distributed Programming, M1, UPMC Sorbonne Universités, France

Master: Luciana Arantes, Swan Dubois, Franck Petit, Distributed Algorithms, M1, UPMC Sorbonne Universités, France

Master: Jonathan Lejeune, Julien Sopena, Client-server distributed systems, M1, UPMC Sorbonne Universités, France

Licence: Pierre Sens, Luciana Arantes, Julien Sopena, Principles of operating systems, L3, UPMC Sorbonne Universités, France

Licence: Swan Dubois, Initiation to operating systems, L3, UPMC Sorbonne Universités, France

Licence: Jonathan Lejeune, Oriented-Object Programming, L3, UPMC Sorbonne Universités, France

Licence: Swan Dubois, Franck Petit, Advanced C Programming, L2, UPMC Sorbonne Universités, France

Licence: Swan Dubois, Sébastien Monnet, Introduction to operating systems, L2, UPMC Sorbonne Universités, France

Licence: Mesaac Makpangou, C Programming Language, 27 h, L2, UPMC Sorbonne Universités, France

Ingénieur 4ème année : Marc Shapiro, Introduction aux systèmes d'exploitation, 26 h, M1, Polytech UPMC Sorbonne Universités, France.

8.2.2. Supervision

PhD: Antoine Blin, "Execution of real-time applications on a small multicore embedded system", 30 January 2017, Gilles Muller (Whisper) and Julien Sopena, CIFRE Renault

PhD: Tao Thanh Vinh, "Ensuring Availability and Managing Consistency in Geo-Replicated File Systems", UPMC, CIFRE, 8 December 2017, Marc Shapiro, Vianney Rancurel (Scality).

PhD: Rudyar Cortes,"Un Environnement à grande échelle pour le traitement de flots massifs de données," UPMC, funded by Chile government, 6 April 2017, Olivier Marin, Luciana Arantes, 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", UPMC, since Oct. 2016, Swan Dubois, Franck Petit, Yoann Dieudonné (University of Picardy Jules Verne)

PhD in progress: Marjorie Bournat, “Exploration with robots in dynamic networks”, UPMC, 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.”, UPMC, CIFRE, since Jan. 2015, Sébastien Monnet, Pierre Sens, Julien Sopena, Dimitri Refauvelet (Magency).

PhD in Progress: Florent Coriat, “Géolocalisation et routage en situation de crise” since Sept 2014, UPMC, Anne Fladenmuller (NPA-LIP6) and Luciana Arantes.

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: Lyes Hamidouche, “Data replication and data sharing in mobile networks”, UPMC, CIFRE, since Nov. 2014, Sébastien Monnet, Pierre Sens, Dimitri Refauvelet (Magency).

PhD in progress: Denis Jeanneau,“Problèmes d’accord et détecteurs de défaillances dans les réseaux dynamique,” UPMC, funded by Labex Smart, since Oct. 2015, Luciana Arantes, Pierre Sens.

PhD in progress: Francis Laniel, UPMC, 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: Ilyas Toumlilt, UPMC, since Sept. 2017, advised by Marc Shapiro. “Bridging the CAP gap all the way to the edge.”

PhD in progress: Alejandro Z. Tomsic, UPMC, since Feb. 2014, Marc Shapiro. “Computing over widely-replicated data in a hybrid cloud.”

CIFRE PhD in progress: Dimitrios Vasilas, UPMC, “Indexing in large-scale storage systems.” Since Sept. 2016, advised by Marc Shapiro.

PhD in progress: Gauthier Voron, “Big-Os : un OS pour les grands volumes de données,”, UPMC, since Sep. 2014, Gaël Thomas, Pierre Sens.

8.2.3. Juries

Pierre Sens was the reviewer of:

- Jalil Boukhobza, HDR, UBO, Brest
- Adrien Lebre, HDR, EMN, Nantes
- Cédric Tedeschi, HDR, IRISA, Rennes
- Ismael Cuadrado-Cordero, PhD, IRSIA, Rennes
- Hana Teyeb, PhD, Telecom SudParis, Evry

Pierre Sens was Chair of

- Georgios Bouloukakis, PhD, UPMC-Inria, Paris, (Advisor: V. Issarny)
- José Manuel Rubio-Hernan, PhD, Telecom SudParis, Evry (Advisor: J. Garcia-Alfaro)
- Luis Eduardo Pineda Morales, PhD, Irisa, Rennes (Advisors: F. Desprez, A. Lebre)

Marc Shapiro was a member of the PhD defense committee of Paolo Viotti, EURECOM, April 2017.

8.3. Popularization

Jonathan Lejeune and Julien Sopena animated an activity during the Science Festival 2017 at UPMC

9. Bibliography

Major publications by the team in recent years

[1] V. BALEGAS, S. DUARTE, C. FERREIRA, R. RODRIGUES, N. PREGUIÇA, M. NAJAFZADEH, M. SHAPIRO.*Putting Consistency back into Eventual Consistency*, in "Euro. Conf. on Comp. Sys. (EuroSys)", Bordeaux, France, ACM, 2015, p. 6:1–6:16 [*DOI :* 10.1145/2741948.2741972], https://hal.inria.fr/hal-01248191.

[2] G. BOSILCA, A. BOUTEILLER, A. GUERMOUCHE, T. HÉRAULT, Y. ROBERT, P. SENS, J. DONGARRA.*Failure Detection and Propagation in HPC systems*, in "SC 2016 - The International Conference for High Performance Computing, Networking, Storage and Analysis", Salt Lake City, United States, November 2016, https://hal.inria.fr/hal-01352109.

[3] S. DUBOIS, R. GUERRAOUI, P. KUZNETSOV, F. PETIT, P. SENS.*The Weakest Failure Detector for Eventual Consistency*, in "34th Annual ACM Symposium on Principles of Distributed Computing (PODC-2015), Donostia-San Sebastián, Spain", Donostia-San Sebastián, Spain, July 2015, p. 375-384 [*DOI :* 10.1145/2767386.2767404], https://hal.archives-ouvertes.fr/hal-01213330.

[4] L. GIDRA, G. THOMAS, J. SOPENA, M. SHAPIRO, N. NGUYEN.*NumaGiC: a Garbage Collector for Big Data on Big NUMA Machines*, in "20th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)", Istanbul, Turkey, Architectural Support for Programming Languages and Operating Systems (ASPLOS), ACM, March 2015, p. 661-673 [*DOI :* 10.1145/2694344.2694361], https://hal.archives-ouvertes.fr/hal-01178790.

[5] A. GOTSMAN, H. YANG, C. FERREIRA, M. NAJAFZADEH, M. SHAPIRO.*'Cause I'm Strong Enough: Reasoning about Consistency Choices in Distributed Systems*, in "Symposium on Principles of Programming Languages", Saint Petersburg, FL, United States, January 2016, p. 371–384 [*DOI :* 10.1145/2837614.2837625], https://hal.inria.fr/hal-01243192.

[6] M. SAEIDA ARDEKANI, P. SUTRA, M. SHAPIRO.*Non-Monotonic Snapshot Isolation: scalable and strong consistency for geo-replicated transactional systems*, in "Symp. on Reliable Dist. Sys. (SRDS)", Braga, Portugal, IEEE Comp. Society, Oct. 2013, p. 163–172 [*DOI :* 10.1109/SRDS.2013.25], http://lip6.fr/Marc.Shapiro/papers/NMSI-SRDS-2013.pdf.

[7] M. SAEIDA ARDEKANI, P. SUTRA, M. SHAPIRO.*G-DUR: A Middleware for Assembling, Analyzing, and Improving Transactional Protocols*, in "Middleware", Bordeaux, France, IEEE, December 2014, 12 [*DOI :* 10.1145/2663165.2663336], https://hal.inria.fr/hal-01109114.

[8] Y. SAITO, M. SHAPIRO.*Optimistic Replication*, in "ACM Computing Surveys", March 2005, vol. 37, n^{o} 1, p. 42–81, http://lip6.fr/Marc.Shapiro/papers/Optimistic_Replication_Computing_Surveys_2005-03_cameraready.pdf.

[9] M. SHAPIRO, N. PREGUIÇA, C. BAQUERO, M. ZAWIRSKI.*Conflict-free Replicated Data Types*, in "Int. Symp. on Stabilization, Safety, and Security of Distributed Systems (SSS)", Grenoble, France, X. DÉFAGO, F. PETIT, V. VILLAIN (editors), Lecture Notes in Comp. Sc., Springer-Verlag, Oct. 2011, vol. 6976, p. 386–400.

[10] V. VAFEIADIS, M. HERLIHY, T. HOARE, M. SHAPIRO.*Proving Correctness of Highly-Concurrent Linearisable Objects*, in "Symp. on Principles and Practice of Parallel Prog. (PPoPP)", New York, USA, March 2006, p. 129–136.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] A. BLIN.*Towards an efficient use of multi-core processors in mixed criticality embedded systems*, Université Pierre et Marie Curie - Paris VI, January 2017, https://tel.archives-ouvertes.fr/tel-01624259.

[12] R. CORTÉS.*Scalable Location-Temporal Range Query Processing for Structured Peer-to-Peer Networks*, Pierre et Marie Curie, Paris VI ; LIP6 UMR 7606 UPMC Sorbonne Universités, France ; équipe REGAL, April 2017, https://tel.archives-ouvertes.fr/tel-01552377.

[13] V. TAO THANH.*Ensuring Availability and Managing Consistency in Geo-Replicated File Systems*, Pierre et Marie Curie, Paris VI ; Inria Paris ; REGAL ; Scality, December 2017, https://hal.inria.fr/tel-01673030.

Articles in International Peer-Reviewed Journal

[14] K. ALTISEN, A. COURNIER, S. DEVISMES, A. DURAND, F. PETIT.*Self-Stabilizing Leader Election in Polynomial Steps*, in "Information and Computation", 2017 [*DOI :* 10.1016/J.IC.2016.09.002], http://hal.upmc.fr/hal-01347471.

[15] X. BONNAIRE, R. CORTÉS, F. KORDON, O. MARIN.*ASCENT: a Provably-Terminating Decentralized Logging Service*, in "The Computer Journal", December 2017, vol. 60, n° 12, p. 1889–1911 [*DOI :* 10.1093/COMJNL/BXX076], http://hal.upmc.fr/hal-01547514.

[16] G. BOSILCA, A. BOUTEILLER, A. GUERMOUCHE, T. HÉRAULT, Y. ROBERT, P. SENS, J. DONGARRA.*A Failure Detector for HPC Platforms*, in "International Journal of High Performance Computing Applications", 2017, https://hal.inria.fr/hal-01531522.

[17] T.-M.-T. NGUYEN, L. HAMIDOUCHE, F. MATHIEU, S. MONNET, S. ISKOUNEN.*SDN-based Wi-Fi Direct Clustering for Cloud Access in Campus Networks*, in "Annals of Telecommunications, Springer", 2017 [*DOI :* 10.1007/S12243-017-0598-Z], http://hal.upmc.fr/hal-01567735.

[18] M. SHAPIRO.*Technical Perspective: Unexpected Connections*, in "Communications- ACM", July 2017, vol. 60, n° 8, p. 82–82 [*DOI :* 10.1145/3068768], https://hal.inria.fr/hal-01570845.

[19] P. VIOTTI, D. DOBRE, M. VUKOLIĆ.*Hybris: Robust Hybrid Cloud Storage*, in "Transactions on Storage", September 2017, vol. 13, n° 3, p. 1 - 32 [*DOI :* 10.1145/3119896], https://hal.inria.fr/hal-01610463.

International Conferences with Proceedings

[20] L. ARANTES, R. FRIEDMAN, O. MARIN, P. SENS.*Probabilistic Byzantine Tolerance Scheduling in Hybrid Cloud Environments*, in "18th International Conference on Distributed Computing and Networking (ICDCN 2017)", Hyderabad, India, January 2017 [*DOI :* 10.1145/1235], https://hal.inria.fr/hal-01399026.

[21] M. BOURNAT, S. DUBOIS, F. PETIT.*Computability of Perpetual Exploration in Highly Dynamic Rings*, in "The 37th IEEE International Conference on Distributed Computing Systems (ICDCS 2017)", Atlanta, United States, June 2017, https://hal.inria.fr/hal-01548109.

[22] M. BOURNAT, S. DUBOIS, F. PETIT.*Quel est le nombre optimal de robots pour explorer un anneau hautement dynamique ?* , in "ALGOTEL 2017 - 19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Quiberon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01516182.

[23] D. CARVER, J. SOPENA, S. MONNET.*ACDC : Advanced Consolidation for Dynamic Containers*, in "NCA", Cambridge, MA, United States, October 2017, https://hal.inria.fr/hal-01673304.

[24] L. HAMIDOUCHE, S. MONNET, F. BARDOLLE, P. SENS, D. REFAUVELET.*EDWiN : leveraging device-to-device communications for Efficient data Dissemination over Wi-Fi Networks*, in "The 31st IEEE International Conference on. Advanced Information Networking and Applications (AINA-2017)", Taipei, Taiwan, March 2017, https://hal.inria.fr/hal-01515372.

[25] L. HAMIDOUCHE, S. MONNET, P. SENS, D. REFAUVELET.*Toward heterogeneity-aware device-to-device data dissemination over Wi-Fi networks*, in "ICPADS 2017 - International Conference on Parallel and Distributed Systems", Shenzhen, China, December 2017, http://hal.univ-smb.fr/hal-01619216.

[26] D. JEANNEAU, T. RIEUTORD, L. ARANTES, P. SENS.*Détecteur de fautes pour le k-accord dans les systèmes inconnus et dynamiques*, in "ALGOTEL 2017 - 19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Quiberon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01511559.

[27] J. KNIESS, L. ARANTES, P. SENS, C. V. N. ALBUQUERQUE.*Saving Resources in Discovery Protocol on Delay-Sensitive Rescue Mobile Networks*, in "The 31st IEEE International Conference on. Advanced Information Networking and Applications (AINA-2017)", Tapei, Taiwan, March 2017, https://hal.inria.fr/hal-01515369.

[28] L. LE FRIOUX, S. BAARIR, J. SOPENA, F. KORDON.*PaInleSS: a Framework for Parallel SAT Solving*, in "The 20th International Conference on Theory and Applications of Satisfiability Testing", Melbourne, Australia, Lecture Notes in Computer Science, Springer, August 2017, vol. 10491, p. 233-250 [*DOI :* 10.1007/978-3-319-66263-3_15], https://hal.archives-ouvertes.fr/hal-01540785.

[29] J. LEJEUNE, F. ALVARES, T. LEDOUX.*Towards a generic autonomic model to manage Cloud Services*, in "The 7th International Conference on Cloud Computing and Services Science (CLOSER 2017)", Porto, Portugal, April 2017, https://hal.archives-ouvertes.fr/hal-01511360.

[30] 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.

[31] B. NGOM, M. MAKPANGOU.*Summary Prefix Tree: An over DHT Indexing Data Structure for Efficient Superset Search*, in "NCA 2017 : 16th IEEE International Symposium on Network Computing and Applications", Cambridge, MA, United States, October 2017, https://hal.inria.fr/hal-01672052.

[32] J. PAULO DE ARAUJO, L. ARANTES, E. P. DUARTE, L. A. RODRIGUES, P. SENS.*A Publish/Subscribe System Using Causal Broadcast Over Dynamically Built Spanning Trees*, in "SBAC-PAD 2017 - 29th International Symposium on Computer Architecture and High Performance Computing", Campinas, Brazil, IEEE, October 2017, p. 161-168 [*DOI :* 10.1109/SBAC-PAD.2017.28], https://hal.inria.fr/hal-01644469.

[33] W. SUN, V. SIMON, S. MONNET, P. ROBERT, P. SENS.*Analysis of a Stochastic Model of Replication in Large Distributed Storage Systems: A Mean-Field Approach*, in "ACM Sigmetrics 2017- International Conference on Measurement and Modeling of Computer Systems", Urbana-Champaign, Illinois, United States, ACM, June 2017, p. 51–51, https://arxiv.org/abs/1701.00335 [*DOI :* 10.1145/3078505.3078531], https://hal.inria.fr/hal-01494235.

[34] G. VORON, G. THOMAS, V. QUEMA, P. SENS.*An interface to implement NUMA policies in the Xen hypervisor*, in "Twelfth European Conference on Computer Systems, EuroSys 2017", Belgrade, Serbia, April 2017, 15, https://hal.inria.fr/hal-01515359.

Conferences without Proceedings

[35] S. BOUCHARD, M. BOURNAT, Y. DIEUDONNÉ, S. DUBOIS, F. PETIT.*Asynchronous Approach in the Plane: A Deterministic Polynomial Algorithm*, in "31st International Symposium on Distributed Computing, DISC 2017", Vienna, Austria, October 2017, http://hal.upmc.fr/hal-01672916.

Scientific Books (or Scientific Book chapters)

[36] M. SHAPIRO, B. KEMME.*Eventual Consistency*, in "Encyclopedia of Database Systems", L. LIU, M. T. ÖZSU (editors), Springer, June 2017, 2 [*DOI :* 10.1007/978-1-4899-7993-3_1366-2], https://hal.inria.fr/hal-01547451.

[37] M. SHAPIRO.*Optimistic Replication and Resolution*, in "Encyclopedia Of Database Systems", L. LIU, M. T. ÖZSU (editors), Springer-Verlag, April 2017, vol. Optimistic Replication and Resolution, p. 1–8 [*DOI :* 10.1007/978-1-4899-7993-3_258-4], https://hal.inria.fr/hal-01576333.

[38] M. SHAPIRO.*Replicated Data Types*, in "Encyclopedia Of Database Systems", L. LIU, M. T. ÖZSU (editors), Springer-Verlag, July 2017, vol. Replicated Data Types, p. 1–5 [*DOI :* 10.1007/978-1-4899-7993-3_80813-1], https://hal.archives-ouvertes.fr/hal-01578910.

Research Reports

[39] G. BOSILCA, A. BOUTEILLER, A. GUERMOUCHE, T. HÉRAULT, Y. ROBERT, P. SENS, J. DONGARRA.*A Failure Detector for HPC Platforms*, Inria, February 2017, n° RR-9024, https://hal.inria.fr/hal-01453086.

[40] E. MAUFFRET, D. JEANNEAU, L. ARANTES, P. SENS.*The Weakest Failure Detector to Solve the Fault Tolerant Mutual Exclusion Problem in an Unknown Dynamic Environment*, LISTIC ; Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, December 2017, https://hal.archives-ouvertes.fr/hal-01661127.

[41] S. S. NAIR.*Evaluation of the CEC (Correct Eventual Consistency) Tool*, Inria Paris ; LIP6 UMR 7606, UPMC Sorbonne Universités, France, November 2017, n° RR-9111, p. 1-27, https://hal.inria.fr/hal-01628719.

Other Publications

[42] P. DARCHE.*Évolution des mémoires à semi-conducteurs à accès aléatoire*, February 2017, Article E2491 des Techniques de l'Ingénieur-article de référence sur le domaine, http://hal.upmc.fr/hal-01341972.

[43] A. SINGH, M. SHAPIRO, G. THOMAS.*Persistent Memory Programming Abstractions in Context of Concurrent Applications*, December 2017, https://arxiv.org/abs/1712.04989 - Accepted in HiPC SRS 2017, https://hal.inria.fr/hal-01667772.

Project-Team REO

Numerical simulation of biological flows

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

IN PARTNERSHIP WITH:
CNRS

Université Pierre et Marie Curie (Paris 6)

RESEARCH CENTER
Paris

THEME
Modeling and Control for Life Sciences

Table of contents

Project-Team REO

Creation of the Project-Team: 2005 April 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
A6.3.1. - 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. Personnel

Research Scientists

Jean-Frédéric Gerbeau [Team leader, Inria, Senior Researcher, HDR]
Miguel Ángel Fernández Varela [Inria, Senior Researcher, HDR]
Céline Grandmont [Inria, Senior Researcher, HDR]
Damiano Lombardi [Inria, Researcher]
Marc Thiriet [CNRS, Researcher, HDR]
Marina Vidrascu [Inria, Senior Researcher]
Irene Vignon Clementel [Inria, Senior Researcher, HDR]

Faculty Members

Laurent Boudin [Univ Pierre et Marie Curie, Associate Professor, HDR]
Muriel Boulakia [Univ Pierre et Marie Curie, Associate Professor, HDR]

Technical Staff

Gautier Bureau [Inria]
Axel Fourmont [Inria, until Sep 2017]
Fabien Raphel [Inria, granted by NOTOCORD]

PhD Students

Matteo Aletti [Inria, until May 2017]
Chloé Audebert [Inria, until Jul 2017]
Ludovic Boilevin-Kayl [Inria]
Chen-Yu Chiang [Univ Pierre et Marie Curie]
Felipe Galarce Marin [Inria, from Dec 2017]
Nicolas Golse [MD, part-time project-team MIMESIS]
Nicolas Pozin [ALSI (Air Liquide Sante International) until Feb 2017, Inria until Aug 2017]
Alexandre This [Philips, PhD Student, granted by CIFRE]
Eliott Tixier [Univ Pierre et Marie Curie]

Post-Doctoral Fellows

Dena Kazerani [Inria, until Sep 2017]
Lara Trussardi [Inria, until Sep 2017]
Florian Joly [Inria]

Visiting Scientists

Gonzalo Castineira Veiga [Universidade da Coruña, from Apr 2017 until Jun 2017]
Marthe(prisca) Combari [Hopital Paul Brousse, from Nov 2017]

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 Universités, 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 [57] in the frame of combustion. It was later used to develop the Kiva code [45] at the Los Alamos National Laboratory, or the Hesione code [52], for example. It has a wide range of applications, besides the nuclear setting: diesel engines, rocket engines [49], 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 [53]. 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 [56], and Biot [46] 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 [47]), 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.* [50], [55], [54] 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

Mikel Landajuela Larma was awarded the 2017 SMAI-GAMNI PhD thesis prize by the French Society of Industrial and Applied Mathematics for his thesis supervised by Miguel Fernández.

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, Faisal Amlani, Irène Vignon-Clementel, Jean-Frédéric Gerbeau, Marina Vidrascu, Matteo Aletti, Miguel Angel Fernandez Varela, Mikel Landajuela Larma, Philippe Moireau and Sébastien Gilles
- Contact: Jean-Frédéric Gerbeau
- URL: http://felisce.gforge.inria.fr

6.2. MODULEF

FUNCTIONAL DESCRIPTION: The numerical method to approximate the constitutive laws for rubber elasticity derived from polymer physics are implemented in the Inria software Modulef.

It is based on : - algorithms from stochastic geometry to generate suitable polymer networks, - Delaunay tessellation algorithms to deal with steric effects (courtesy of the Inria project-team GAMMA2), - the introduction of 1-dimensional finite elements for the polymer-chains in Modulef.

- Participants: Antoine Gloria and Marina Vidrascu
- Contact: Marina Vidrascu
- URL: https://www.rocq.inria.fr/modulef/

6.3. 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/

7. New Results

7.1. Mathematical and numerical analysis of fluid-structure interaction problems

Participants: Matteo Aletti, Ludovic Boilevin-Kayl, Chen-Yu Chiang, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Céline Grandmont, Damiano Lombardi, Marc Thiriet, Marina Vidrascu.

In [15] a reduced order modeling method is developed to simulate multi-domain multi-physics problems. In particular we considered the case in which one problem of interest, described by a generic non-linear partial differential equation is coupled to one or several problems described by a set of linear partial differential equations. In order to speed up the resolution of the coupled system, a low-rank representation of the Poincaré-Steklov operator is built by a reduced-basis approach. A database for the secondary problems is built when the interface condition is set to be equal to a subset of the Laplace-Beltrami eigenfunctions on the surface. The convergence of the method is analysed and several 3D fluid-fluid and fluid-structure couplings are presented as numerical experiments.

In [43] we study an unsteady nonlinear fluid–structure interaction problem. We consider a Newtonian incompressible two-dimensional flow described by the Navier-Stokes equations set in an unknown domain depending on the displacement of a structure, which itself satisfies a linear wave equation or a linear beam equation. We prove existence of a unique local-in-time strong solution. In the case of the wave equation or a beam equation with inertia of rotation, this is, to our knowledge the first result of existence of strong solutions for which no viscosity is added. One key point, is to use the fluid dissipation to control, in appropriate function spaces, the structure velocity.

In [26] a fluid-structure interaction solver based on 3D Eulerian monolithic formulation for an incompressible Newtonian fluid coupled with a hyperelastic incompressible solid has been implemented, verified, and validated. It is based on a Eulerian formulation of the full system. After a fully implicit discretization in time, displacement is eliminated and the variational equation is solved for the velocity and pressure. Its main application in medicine is venous flow in inferior limbs.

7.2. Numerical methods for biological flows

Participants: Chloé Audebert, Ludovic Boilevin-Kayl, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Florian Joly, Alexandre This, Marc Thiriet, Irene Vignon Clementel.

Peripheral pulmonary artery stenosis (PPS) is a congenital abnormality resulting in pulmonary blood flow disparity and right ventricular hypertension, for which optimal surgical strategies remain unclear. In [35], we conduct a pilot study to use recently refined computational simulation in the setting of multiple surgical strategies and to examine the influence of pulmonary artery reconstruction on hemodynamics in this population. Obstruction relief along with pulmonary artery vasodilation determines postoperative pulmonary flow distribution and newer methods can incorporate these physiologic changes.

Incoming velocity at open boundaries, or backflow, often yields to unphysical instabilities already for moderate Reynolds numbers. Several treatments to overcome these backflow instabilities have been proposed in the literature. In [17], we present a set of benchmark problems in order to compare different methods in different backflow regimes (with a full reversal flow and with propagating vortices after a stenosis). The examples are implemented in FreeFem++ and the source code is openly available.

The simulation of cardiac blood flow using patient-specific geometries can help for the diagnosis and treatment of cardiac diseases. Current patient-specific cardiac flow simulations requires a significant amount of human expertise and time to pre-process image data and obtain a case ready for simulations. In [38] a new procedure is proposed to alleviate this pre-processing by registering a unique generic mesh on patient-specific cardiac segmentations and transferring appropriately the spatiotemporal dynamics of the ventricle. The method is applied on real patient data acquired from 3D ultrasound imaging. Both a healthy and a pathological conditions are simulated. The resulting simulations exhibited physiological flow behavior in cardiac cavities and the experiments confirm a significant reduction in pre-processing work.

In order to reduce the complexity of heart hemodynamics simulations, one-way coupling approaches are often considered as an alternative to 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 [39] we propose an enhanced resistive immersed surface (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.

In [51], a computational model of unsteady blood flow in the cerebral venous circuit inside the skull reconstructed from medical images has been carried out. This venous network runs separately from the arterial bed perfusing the brain. The major aspects are boundary conditions and flow governing parameters.

7.3. Numerical methods for cardiac electrophysiology

Participants: Muriel Boulakia, Jean-Frédéric Gerbeau, Damiano Lombardi, Fabien Raphel, Eliott Tixier.

In [32], we propose a model to represent the electrical potential of cardiomyocytes derived from stem cells in Multi Electrodes Arrays (MEA). This model based on the bidomain equations and a model for the MEA electrodes is used to analyze experimental signals. Our numerical algorithm is able to provide for different drugs dose-response curves which are in very good agreement with known values.

In [14], we are interested in the electrical activity of cardiomyocytes under the action of drugs in MEA devices. We present numerical simulations based on the same model as in [32] enriched with a pore block model to assay the action of drugs. The simulation results show that the model properly reflects the main effects of several drugs on the electrical potential.

In [33] the variability of phenomena in cardiac electro-physiology is investigated by using a moment matching approach. The cells activity is described by parametric systems of Ordinary Differential Equations. Given the population statistics on a system observables (which is the action potential of the cells), the probability density distribution of the parameters is sought such that the statistics of the model outputs match the observed ones. An uncertainty quantification step is solved once for all by using a non-instrusive approach, and then the inverse problem is solved by introducing an entropy regularisation. Several numerical experiments are considered to validate the approach on realistic datasets.

In [34] a realistic application on the classification of the drugs effect on cardiac cells is investigated. In particular, the electrical activity of the cells is recorder by Micro Electrode Arrays in normal conditions and under drugs, at different concentrations. In order to perform a classification of a drug in terms of promoting or inhibit the activity of certain ion channels a machine learning approach is used (support vector machine). Since the data amount is not big and the variability and alea sources have a large impact on the signals recorded, the data set is augmented by in silico experiments. Several tests on realistic data are performed.

7.4. Lung and respiration modeling

Participants: Céline Grandmont, Dena Kazerani, Nicolas Pozin, Marina Vidrascu, Marc Thiriet, Irene Vignon Clementel.

In [30] we use the coupled model tree-parenchyma model introduced in [31] to study the impact of asthma on effort and ventilation distribution along with the effect of Heliox compared to air. Indeed, in spite of numerous clinical studies, there is no consensus on the benefit Heliox mixtures can bring to asthmatic patients in terms of work of breathing and ventilation distribution. For this study, lung surface displacement fields extracted from computed tomography medical images are used to prescribe realistic boundary conditions to the system. Asthma is simulated by imposing bronchoconstrictions to some airways of the tracheo-bronchial tree based on statistical laws deduced from the literature. This study illuminates potential mechanisms for patient responsiveness to Heliox when affected by obstructive pulmonary diseases. Responsiveness appears to be function of the pathology severity, as well as its distal position in the tracheo-bronchial tree and geometrical position within the lung. Moreover, as already stated, in asthma and COPD, some airways of the tracheo-bronchial tree can be constricted, from moderate narrowing up to closure. These pathological patterns affect the lung ventilation distribution. While some imaging techniques enable visualization and quantification of constrictions in proximal generations, no non-invasive technique provides precise insights on what happens in more distal areas. In [44] we propose a process that exploits dynamical lung ventilation measurements to access positions of airways closures in the tree. This identification approach combines our lung ventilation model along with a machine learning approach. Based on synthetic data generated with typical temporal and spatial resolutions as well as reconstruction errors, we obtain encouraging results with a detection rate higher than 90%.

The human tracheobronchial tree surface is covered with mucus that ensures clearance of foreign material. An alteration of mucus or its environment such as in cystic fibrosis dramatically impacts the mucociliary clearance. In [48] the numerical method is able to manage variations of more than 5 orders of magnitude in the shear rate and viscosity. It leads to a cartography that enables to discuss major issues on defective mucociliary clearance in cystic fibrosis. In addition, cystic fibrosis is associated with a shear-thinning mucus that tends to aggregate in regions of lower clearance. However, a rarefaction of periciliary fluid has a greater impact than the mucus shear-thinning.

7.5. Miscellaneous

Participants: Damiano Lombardi, Irene Vignon Clementel.

In [27] an adaptive tensor method is developed to build a parsimonious discretization for the kinetic equations, starting from separated, arbitrary and a priori chosen discretizations for the space and the velocity variables. The method automatically adapts the rank of the decomposition in order to ensure that a criterion on the residual of the equations is satisfied, and the proof of the convergence is provided. The method is tested on the Vlasov-Poisson equation but can be extended to other kinetic equations and to systems in which the domain is the cartesian product of separated domains.

In [42] an a posteriori error estimator for hermitian positive eigenvalue problem is proposed. This estimator, which is based on a residual formulation, is constructed by shifting the operators in such a way that the error between the exact eigenvalues and the approximated ones can be estimated efficiently. It is conditionally certified and sharp.

Diffusion-weighted magnetic resonance imaging (DWI) is a key non-invasive imaging technique for cancer diagnosis and tumor treatment assessment; yet its relation to the underlying tissue structure is not clear. In [36], in order to link low-resolution but non-invasive DWI data with high resolution (invasive) histological information, we developed an image processing and analysis chain, which was used to study the correlation between the DWI diffusion coefficient and tumor cellularity from serial histological slides of a resected non-small cell lung cancer tumor.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Air Liquide Santé International

Participants: Céline Grandmont, Nicolas Pozin, Irene Vignon Clementel.

CIFRE convention and contract with Air Liquide Santé International (March 2014 - December 2017) in the context of the ANRT on “Multiscale lung ventilation modeling in health and disease”, for the PhD thesis of Nicolas Pozin.

8.1.2. 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.

8.1.3. 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.

8.1.4. Instem/NOTOCORD

Participants: Muriel Boulakia, Damiano Lombardi, Jean-Frédéric Gerbeau, Fabien Raphel, Eliott Tixier.

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).

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

9.1.1.1. ANR Project "iFLOW"

Participants: Chloé Audebert, Jean-Frédéric Gerbeau, Florian Joly, Irene Vignon Clementel [co-Principal Investigator].

Period: 2013-2017.
This ANR-TecSan, co-managed by Eric Vibert (Paul Brousse Hospital) and Irene Vignon Clementel, aims at developing an Intraoperative Fluorescent Liver Optimization Workflow to better understand the relationship between architecture, perfusion and function in hepatectomy.

Other partners: DHU Hepatinov - Hôpital Paul Brousse, Inria Mamba, Fluoptics, IfADo, MID.

9.1.1.2. 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.

9.1.1.3. 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
- Marina Vidrascu is a member of the ANR ARAMIS
- 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).

9.1.2. Inria initiatives

9.1.2.1. ADT Project "PARASOL"

Participants: Miguel Ángel Fernández Varela [Principal Investigator], Axel Fourmont, Marina Vidrascu.

Period: 2016-2017
The aim of this project, coordinated by Miguel Ángel Fernández Varela, is to implement in the FELiScE library several balancing domain decomposition methods (BDD) for solid-mechanics.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. REVAMMAD

Title: "Retinal Vascular Modeling, Measurement and Diagnosis"

Programm: FP7

Duration: April 2013 - March 2017

Coordinator: University of Lincoln

Partners: : See the web site http://revammad.blogs.lincoln.ac.uk/partners/

Inria contact: J-F Gerbeau

REVAMMAD is a European Union project aimed at combatting some of the EU's most prevalent chronic medical conditions using retinal imaging. The project aims to train a new generation of interdisciplinary scientists for the academic, clinical and industrial sectors, and to trigger a new wave of biomedical interventions. The role of REO team within this consortium is to propose a mathematical model and a simulation tool for the retina hemodynamics. See http://revammad.blogs.lincoln.ac.uk for more details.

9.2.2. Collaborations in European Programs, Except FP7 & H2020

9.2.2.1. 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. Internships

- Gonzalo Castineira Veiga, Visiting PhD student, Universidade da Coruña, Apr 2017–Jun 2017

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- Céline Grandmont
  - Co-organizer of Inria-LJLL meeting in scientific computing
  - Co-organizer of the conference in the honor of Y. Maday for his 60th birthday, may 2017, Roscoff
  - Co-organizer of the conference "Analysis and control of fluid-structure interaction systems", october 2017, IMB, Bordeaux University
- Irene Vignon Clementel
  - Organized a minisymposium at the ICCB conference, September 2017, Compiègne
  - Organized a minisymposium with A. Marsden (Stanford U.) at the 5th International Conference on Computational & Mathematical Biomedical Engineering, April 2017, Pittsburg, USA.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Laurent Boudin
  - Coordinator of the 6th "Forum Emploi Maths 2017" organizing committee, with Bertrand Michel (École Centrale Nantes), December 2017, Paris, France

10.1.2.2. Member of the Conference Program Committees

- Jean-Frédéric Gerbeau
  - 5th International conference on Computational & Mathematical Biomedical Engineering (CMBE 2017), Pittsburgh, USA.
  - ENUMATH 2017 conference. Voss, Norway.
  - 9th International Conference on Functional Imaging and Modeling of the Heart (FIMH 2017). Toronto, Canada.
- Céline Grandmont
  - Member of the scientific committee of the "EDP Normandie 2017" conference
- Irene Vignon Clementel
  - Programme committee member, Computational and Mathematical Biomedical Engineering Conference, Pittsburgh 2017
  - Conference steering committee, International Conference on Engineering Frontiers in Pediatric and Congenital Heart Disease, 2015-present

10.1.2.3. Reviewer

- Jean-Frédéric Gerbeau
  - Member of the Scientific Program Committee of the Millennium Science Initiative, a program of the Ministry of Economy of Chile.
- Irene Vignon Clementel
  - Reviewer for the Netherlands Organisation for Scientific Research (NWO)

10.1.3. Journal

10.1.3.1. 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.
- Marc Thiriet
  - Member of the editorial board of Digital Medicine

10.1.4. 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.*)

10.1.5. Research Administration

- Muriel Boulakia
  - Supervisor of the teaching of mathematics at the engineer school Polytech Paris-UPMC
- Miguel Ángel Fernández Varela
  - Deputy Head of Science, Inria Paris (from Sept. 2017)
  - Member of the Inria Evaluation Committee (from Sept. 2017)
  - Co-president of the Scientific Positions Commission, Inria Paris
- Jean-Frédéric Gerbeau
  - Head of science, Inria Paris (until Sept. 2017)
  - Member of the Inria Evaluation Committee (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
  - Committee member for PhD students at Inria "Commission consultative des doctorants", since July 2016.
  - Mediator between PhD students and their supervisors for Inria Paris

10.1.6. Conferences

- Matteo Aletti
  - Minisymposium talk, International conference on Finite elements in flow problems (FEF), Rome (Italy), 2017
- Chloé Audebert
  - Invited speaker, GDR Mamovi, Sep 27-29, Lyon, France.
  - Minisymposium talk, International Conference on Computational Bioengineering, Sep 6-8, Compiègne, France.
  - Talk, Colloquium 595 Biomechanics and computer assisted surgery meets medical reality, Aug 29-31, Lille, France.
  - Weekend de l'innovation chirurgicale (WIC), Jun 23-25, Cabourg, France.
  - Talk, Congrès SMAI 2017, 8e Biennale Française des Mathématiques Appliquées et Industrielles, Jun 5-9, Ronce-les-bains, France.
  - Minisymposium talk, Computational and mathematical Biomedical Engineering (CMBE) , Apr 10-12, Pittsburgh, USA.
  - Seminar, bioMMeda group, Mar 8th, Ghent University, Ghent, Belgium.
  - Seminar, Laboratoire de mathématique MAP5, Mar 3rd, Université Paris Descartes, Paris, France.
- Ludovic Boilevin-Kayl
  - Minisymposium talk, 5th International Conference on Computational and Mathematical Biomedical Engineering (CMBE), April 10th-12th, Pittsburgh, PA, United States.
- Laurent Boudin

  - Contributed talk, Workshop "Franco-Italian meeting on kinetic theory and singular parabolic equations", Mar 16-17, Paris, France
  - Invited talk, "Aerosolstorming" meeting of the "Société de Pneumologie en Langue Française", May 2017, Paris, France
  - Invited talk, Summer School on Computational Social and Behavioral Sciences, Sep 2017, Paris, France
  - Seminar, ENS Paris-Saclay Starters scientific talks in mathematics and computer science, Sep 2017, Cachan, France
  - Invited talk, Workshop "Modelling, simulation and study of social behaviour", Oct 2017, Nancy, France
  - Invited talk, Conference on Partial differential equations and semi-groups, Dec 2017, Besançon, France
- Muriel Boulakia
  - Seminar Analysis, Ceremade, Univ. Paris Dauphine, December 2017
  - Seminar, Univ. Paris Descartes, November 2017
  - Seminar PDE, Univ. Versailles, May 2017
- Miguel Ángel Fernández Varela
  - Invited semi-plenary lecture, International conference on Finite elements in flow problems (FEF), Rome (Italy), 2017
- Jean-Frédéric Gerbeau
  - Invited lectures (6 hours), Maxwell Institute Graduate School on Evolution Equations, ICMS, Edinburgh, UK, 2017
  - Invited lecture, Conference "Quiet 2017", Quantification of Uncertainty: Improving Efficiency and Technology, SISSA, Trieste, Italy, July 2017
  - Invited lecture, Workshop "In Silico Human drug safety and efficacy", Oxford, UK, September 2017
  - Invited lecture, Workshop GDR Mamovi.
  - Seminar, Weierstrass Institute for Applied Analysis and Stochastics (WIAS), Berlin, Germany, November 2017
  - Minisymposium talk, ICCB conference, Compiègne, France, September 2017
- Céline Grandmont
  - ULB Maths Colloquium, Bruxelles, april 2017
  - Colloquium Univ. Lille, december 2017
- Florian Joly
  - Weekend de l'innovation chirurgicale (WIC), Jun 23-25, Cabourg, France.
  - Talk, GRIC Journées Françaises de Radiologie, Oct 12th, Paris, France.
- Damiano Lombardi
  - CMBE 2017, Pittsburgh (US), invited to the minisymposium on *Adaptation, growth and remodelling*
  - CMBE 2017, Pittsburgh (US), organiser of the minisymposium on *Reliable predictions in biomedical applications*
- Dena Kazerani
  - Talk, Groupe de travail de biologie, Tours-Orléans, October 2017, Orléans, France.
  - Poster, Congrès SMAI 2017, 8e Biennale Française des Mathématiques Appliquées et Industrielles, June 2017, Ronce-les-bains, France.

  - Talk, Ecole EGRIN: Ecoulements Gravitaires et Risques Naturels, June 2017, Cargèse, France.
  - Talk, Séminaire du Laboratoire Hydraulique Saint-Venant, January 2017, Chatou, France
  - Poster, vingt-neuème séminaire sur "la mécanique des fluides numériques" organised by GAMNI-CEA, Institut Herni Poincaré (IHP), January 2017, Paris, France.
- Alexandre This
  - Poster, 9th international conference on Functional Imaging and Modeling of the Heart, FIMH, Toronto, June 2017
- Eliott Tixier
  - Invited talk, Workshop on Mathematical Methods in Cardiac Electrophysiology, Nov 4-6, Ottawa, Canada
  - Invited talk, GdR Mamovi meeting, Sep 27-29, Lyon, France
  - Poster presentation, QUIET 2017 Workshop, Jul 18-21, Trieste, Italy
- Marina Vidrascu
  - Invited talk ,The Sixth International Conference on Scientific Computing and Partial Differential Equations, Hong Kong Baptist University, June 5-8, 2017
- Irene Vignon Clementel
  - Seminar, Universittsklinikum Aachen, Dec 4th 2017, Aachen, Germany
  - Seminar, LiSyM consortium, Nov. 28th 2017, Paris, France
  - Seminar, U. de Caen, Dpt of Mathematics, Nov. 13th 2017, Caen, France
  - Invited, SimInhale workshop, Oct. 4-5th 2017, Athens, Greece
  - Seminar, Ecole Polytechnique, Oct 16th 2017, Palaiseau, France
  - Seminar, Marie-Lannelongue Hospital, Sept 18th 2017, Plessis-Robinson, France
  - Keynote (invited), Colloquium 595 Biomechanics and computer assisted surgery meets medical reality, Aug 29-31, Lille, France.
  - Invited talk, Weekend de l'innovation chirurgicale (WIC), Jun 23-25, Cabourg, France.
  - Minisymposium talk, Computational and mathematical Biomedical Engineering (CMBE), Apr 10-12, Pittsburgh, USA.
  - Seminar, Eindhoven U., Dpt of Bioengineering, March 24th 2017, Eindhoven, The Netherlands
  - Seminar, Organox - Paul Brousse Hospital, Jan 24th 2017, Villejuif, France

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence:

- Ludovic Boilevin-Kayl
  - Calculus, 80h, L1, UPMC
  - Matrix computations, 18h, L1, UPMC
- Laurent Boudin
  - Introduction to series for signal theory, 18h, L2, UPMC
  - Shared studies supervision in mathematics licence for approximately 500 students, 24h, L2, UPMC (until Aug 31, 2017)
  - Calculus, 38.5h, L1, UPMC

  - Student advising for orientation and professional insertion, 22h, L2, UPMC
  - Numerical methods for ODE, 26.5h, L3, UPMC
- Muriel Boulakia
  - Scilab, 35h, L2, UPMC
  - Nonlinear systems and optimization, 35h, L3, Polytech'Paris
  - Oral tests in in topology and differential calculus, 20h, L3
- Miguel Ángel Fernández Varela
  - Analysis and Scientific Computing, 30h, L3, ENPC
- Jean-Frédéric Gerbeau
  - Control of dynamical systems, 32h, L3, Ecole Polytechnique.
- Céline Grandmont
  - Ordinary differential equations, 24h, L3, UPMC
- Damiano Lombardi
  - Analysis and Scientific Computing, 32h, L3, ENPC
  - Numerical Methods, 48h, L3, Polytech'Paris
- Marc Thiriet
  - Modeling and Simulation for Computer-Aided Medicine and Surgery, 16h, L3-M2, National Taiwan University
- Irene Vignon Clementel
  - Numerical Methods for Ordinary Differential Equations, 24h ETD, L3, UPMC
  - Numerical simulations of blood flow, 1h30, as part of the undergraduate "continuum mechanics", AgroParisTech

Master:

- Laurent Boudin
  - Basics for numerical methods, 27h, M1, UPMC
  - Student advising for orientation and professional insertion, 20h, M1, UPMC
- Muriel Boulakia
  - Preparatory course for teaching admission examination "Agrégation", 25h, M2, UPMC
- Miguel Ángel Fernández Varela
  - Modeling and numerical methods for hemodynamics, 30h, M2, UPMC
- Jean-Frédéric Gerbeau
  - Seminar for M2 students of the master "Math SV" (1h), M2, Univ Paris-Sud, December
  - Seminar for M2 students at Ecole des Mines (3h), Paris, February
- Irene Vignon Clementel
  - Modélisation hémodynamique & simulation numérique comme outil pour la chirurgie, 1h, M2, Université Paris Sud

10.2.2. Supervision

PhD: Chloé Audebert, Modeling of liver hemodynamics, defended on February 24, 2017. Supervisors: J.-F. Gerbeau & I. Vignon Clementel.

PhD: Matteo Aletti, Multiscale retinal vascular modeling, defended on May 30, 2017. Supervisors: J.-F. Gerbeau & D. Lombardi.

PhD: Eliott Tixier, Variability modeling and numerical biomarkers design in cardiac electrophysiology, defended on December 18, 2017. Supervisors: J-F. Gerbeau & D. Lombardi.
PhD: Nicolas Pozin, Multiscale lung ventilation modeling in health and disease, defended on October 6, 2017. Supervisors: C. Grandmont & I. Vignon Clementel.
PhD in progress: Andrea Bondesan, Kinetic and fluid models, numerical and asymptotic analysis, since October 2015. Supervisors: L. Boudin, B. Grec & S. Martin.
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.

10.2.3. Juries

- Muriel Boulakia
  - PhD committee: Charlie Douanla Lontsi, Inria Bordeaux Sud-Ouest
- Jean-Frédéric Gerbeau
  - PhD committees: Ivan Fumagalli, Politecnico di Milano (referee); Rajnesh Lal, Univ Montpellier (referee).
  - HDR committees: Edmond Vigmond, Univ Bordeaux; Sébastien Boyaval, Univ Paris-Est.
  - Hiring committee: Inria Paris (CR2).
- Céline Grandmont
  - Member of the “agrégation” jury in mathematics.
  - Hiring committees: Inria Nancy (CR2), Inria DR2, Professor position Paris-Sud University.
  - PHD committee: M. Deville, Bordeaux University (Referee), Nicolas Pozin, UPMC (co-advisor).
- Marina Vidrascu
  - Hiring committee: IR CNRS
- Irene Vignon Clementel
  - Hiring committees: Starting Research Positions and Senior Research Positions at Inria
  - PhD comittee: Roch Mollero, University of Nice-Sophia Antipolis (Referee), Nicolas Pozin, UPMC, (co-advisor), Arthur Ghigo, UPMC, Chloé Audebert, UPMC, (co-advisor), Andie de Villiers, U. of Cape Town, (Referee), Petru Bucur, Inserm.

10.3. Popularization

- Céline Grandmont
  - Conference “Métier”: Master Maths students, UPMC, Oct 2017
- Irene Vignon Clementel
  - Presentation to celebrate the renewal of UPMC-Inria Paris partnership, June 22nd 2017, Paris, 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^{o} 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^{o} 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^{o} 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^{o} 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^{o} 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^{o} 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^{o} 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

Doctoral Dissertations and Habilitation Theses

[12] M. C. M. ALETTI.*Mathematical Modelling and Simulations of the Hemodynamics in the eye*, Université Pierre et Marie Curie (UPMC Paris 6), May 2017, https://tel.archives-ouvertes.fr/tel-01538557.

[13] C. AUDEBERT.*Mathematical liver modeling: hemodynamics and function in hepatectomy*, Université Pierre & Marie Curie - Paris 6, February 2017, https://tel.archives-ouvertes.fr/tel-01512620.

Articles in International Peer-Reviewed Journal

[14] 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", July 2017, https://hal.inria.fr/hal-01562673.

[15] M. ALETTI, D. LOMBARDI.*A Reduced Order representation of the Poincaré-Steklov operator: an application to coupled multi-physics problems*, in "International Journal for Numerical Methods in Engineering", 2017, https://hal-auf.archives-ouvertes.fr/hal-01396286.

[16] C. AUDEBERT, P. BUCUR, M. BEKHEIT, E. VIBERT, I. VIGNON-CLEMENTEL, J.-F. GERBEAU.*Kinetic scheme for arterial and venous blood flow, and application to partial hepatectomy modeling*, in "Computer Methods in Applied Mechanics and Engineering", February 2017, vol. 314, p. 102-125 [*DOI :* 10.1016/J.CMA.2016.07.009], https://hal.archives-ouvertes.fr/hal-01347500.

[17] C. BERTOGLIO, A. CAIAZZO, Y. BAZILEVS, M. BRAACK, M. ESMAILY, V. GRAVEMEIER, A. MARSDEN, O. PIRONNEAU, I. VIGNON-CLEMENTEL, W. A. WALL.*Benchmark problems for numerical treatment of backflow at open boundaries*, in "International Journal for Numerical Methods in Biomedical Engineering", 2017 [*DOI :* 10.1002/CNM.2918], https://hal.inria.fr/hal-01569770.

[18] F. BONALDI, G. GEYMONAT, F. KRASUCKI, M. VIDRASCU.*Mathematical and numerical modeling of plate dynamics with rotational inertia*, in "Journal of Numerical Mathematics", 2017, p. 1 - 20 [*DOI :* 10.1515/JNMA-2016-1020], https://hal.inria.fr/hal-01413037.

[19] L. BOUDIN, C. GRANDMONT, A. MOUSSA.*Global existence of solutions to the incompressible Navier-Stokes-Vlasov equations in a time-dependent domain*, in "Journal of Differential Equations", February 2017 [*DOI :* 10.1016/J.JDE.2016.10.012], https://hal.inria.fr/hal-01312262.

[20] L. BOUDIN, B. GREC, V. PAVAN.*The Maxwell-Stefan diffusion limit for a kinetic model of mixtures with general cross sections*, in "Nonlinear Analysis: Theory, Methods and Applications", February 2017, vol. 159, p. 40-61 [*DOI :* 10.1016/J.NA.2017.01.010], https://hal.inria.fr/hal-01303312.

[21] M. BOULAKIA, S. GUERRERO.*On the interaction problem between a compressible fluid and a Saint-Venant Kirchhoff elastic structure*, in "Advances in Differential Equations", January 2017, vol. 22, n^{o} 1-2, https://hal.inria.fr/hal-01229577.

[22] M. BOULAKIA, E. SCHENONE.*Stability estimates for some parameters of a reaction-diffusion equation coupled with an ODE*, in "Applicable Analysis", 2017, vol. 96, n^{o} 7, https://hal.inria.fr/hal-01227427.

[23] 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.

[24] P. BUCUR, M. BEKHEIT, C. AUDEBERT, I. E. VIGNON-CLEMENTEL, E. VIBERT.*Simplified technique for 75% and 90% hepatic resection with hemodynamic monitoring in large white swine model*, in "Journal of Surgical Research", March 2017, vol. 209, p. 122-130 [*DOI :* 10.1016/J.JSS.2016.09.018], https://hal.archives-ouvertes.fr/hal-01405171.

[25] E. BURMAN, A. ERN, M. A. FERNÁNDEZ.*Fractional-step methods and finite elements with symmetric stabilization for the transient Oseen problem*, in "ESAIM: Mathematical Modelling and Numerical Analysis", March 2017, vol. 51, n^{o} 2, p. 487-507 [*DOI :* 10.1051/M2AN/2016028], https://hal.inria.fr/hal-01218328.

[26] C.-Y. CHIANG, O. PIRONNEAU, T. W. H. SHEU, M. THIRIET.*Numerical Study of a 3D Eulerian Monolithic Formulation for Incompressible Fluid-Structures Systems*, in "Fluids", June 2017, vol. 2, n^{o} 2, p. 34-53 [*DOI :* 10.3390/FLUIDS2020034], http://hal.upmc.fr/hal-01558747.

[27] V. EHRLACHER, D. LOMBARDI.*A dynamical adaptive tensor method for the Vlasov-Poisson system*, in "Journal of Computational Physics", 2017, vol. 319, p. 285-306, https://hal.archives-ouvertes.fr/hal-01335507.

[28] M. LANDAJUELA, M. VIDRASCU, D. CHAPELLE, M. A. FERNÁNDEZ.*Coupling schemes for the FSI forward prediction challenge: comparative study and validation*, in "International Journal for Numerical Methods in Biomedical Engineering", 2017, vol. 33, n^{o} 4, e02813 [*DOI :* 10.1002/CNM.2813], https://hal.inria.fr/hal-01239931.

[29] S. PANT, C. CORSINI, C. BAKER, T.-Y. HSIA, G. PENNATI, I. VIGNON-CLEMENTEL.*Inverse problems in reduced order models of cardiovascular haemodynamics: aspects of data-assimilation and heart-rate variability*, in "Journal of the Royal Society Interface", 2017, https://hal.inria.fr/hal-01413446.

[30] N. POZIN, S. MONTESANTOS, I. KATZ, M. PICHELIN, C. GRANDMONT, I. VIGNON-CLEMENTEL.*Calculated Ventilation and Effort Distribution as a Measure Of Respiratory Disease and Heliox Effectiveness*, in "Journal of Biomechanics", June 2017, n^{o} 60C, p. 100-109 [*DOI :* 10.1016/J.JBIOMECH.2017.06.009], https://hal.archives-ouvertes.fr/hal-01468861.

[31] N. POZIN, S. MONTESANTOS, I. KATZ, M. PICHELIN, I. E. VIGNON-CLEMENTEL, C. GRANDMONT.*A tree-parenchyma coupled model for lung ventilation simulation*, in "International Journal for Numerical Methods in Biomedical Engineering", February 2017 [*DOI :* 10.1002/CNM.2873], https://hal.archives-ouvertes.fr/hal-01407055.

[32] 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", 2017, In press [*DOI :* 10.1109/TBME.2017.2748798], https://hal.archives-ouvertes.fr/hal-01570341.

[33] E. TIXIER, D. LOMBARDI, B. RODRIGUEZ, J.-F. GERBEAU.*Modeling Variability in Cardiac Electrophysiology: A Moment Matching Approach*, in "Journal of the Royal Society Interface", August 2017, vol. 14, n^{o} 133 [*DOI :* 10.1098/RSIF.2017.0238], https://hal.archives-ouvertes.fr/hal-01570828.

[34] 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.

[35] W. YANG, F. L. HANLEY, F. P. CHAN, A. L. MARSDEN, I. E. VIGNON-CLEMENTEL, J. A. FEINSTEIN.*Computational simulation of postoperative pulmonary flow distribution in Alagille patients with peripheral pulmonary artery stenosis*, in "Congenital Heart Disease", 2017 [*DOI :* 10.1111/CHD.12556], https://hal.inria.fr/hal-01657634.

[36] 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.

International Conferences with Proceedings

[37] E. LLUCH, R. DOSTE, S. GIFFARD-ROISIN, A. THIS, M. SERMESANT, O. CAMARA, M. DE CRAENE, H. G. MORALES.*Smoothed Particle Hydrodynamics for Electrophysiological Modeling: An Alternative to Finite Element Methods*, in "Functional imaging and modelling of the heart 2017", Toronto, Canada, Functional imaging and modelling of the heart 2017 Proceedings, Springer International Publishing, June 2017, vol. 141, p. 333-343 [*DOI :* 10.1007/978-3-319-59448-4_32], https://hal.inria.fr/hal-01533371.

[38] A. A. THIS, L. BOILEVIN-KAYL, H. G. MORALES, O. BONNEFOUS, P. A. ALLAIN, M. A. FERNÁNDEZ, J.-F. GERBEAU.*One Mesh To Rule Them All: Registration-Based Personalized Cardiac Flow Simulations*, in "FIMH 2017 - 9th international conference on Functional Imaging and Modeling of the Heart", Toronto, Canada, LNCS - Lecture Notes in Computer Science, Springer, June 2017, vol. 10263, https://hal.inria.fr/hal-01512309.

Conferences without Proceedings

[39] L. BOILEVIN-KAYL, A. THIS, M. A. FERNÁNDEZ, J.-F. GERBEAU.*An efficient valve model based on resistive immersed surfaces enhanced with physiological data*, in "5th International Conference on Computational & Mathematical Biomedical Engineering", Pittsburgh, United States, Etienne Boileau and David Nordsletten, April 2017, https://hal.inria.fr/hal-01519602.

Scientific Books (or Scientific Book chapters)

[40] D. ANNE-ARCHARD, R. CHATELIN, M. MURRIS-ESPIN, D. SANCHEZ, M. THIRIET, A. DIDIER, P. PONCET.*Modeling Cystic Fibrosis and Mucociliary Clearance*, in "Modeling of microscale transport in biological processes", S. M. BECKER (editor), Academic Press, 2017, p. 113-154 [*DOI :* 10.1016/B978-0-12-804595-4.00005-5], https://hal.inria.fr/hal-01476216.

Research Reports

[41] A. NICOLOPOULOS, N. RIANE, A. SAINT-DIZIER, L. TRUSSARDI.*Analyse d'images de pales d'éoliennes. Semaine d'étude maths-entreprises, Paris, 11-15 septembre 2017*, AMIES, November 2017, p. 1-18, https://hal.archives-ouvertes.fr/hal-01651796.

Other Publications

[42] A. BAKHTA, D. LOMBARDI.*An a posteriori error estimator based on shifts for positive hermitian eigenvalue problems*, September 2017, working paper or preprint, https://hal.inria.fr/hal-01584180.

[43] C. GRANDMONT, M. HILLAIRET, J. LEQUEURRE.*Existence of local strong solutions to fluid-beam and fluid-rod interaction systems* , July 2017, working paper or preprint, https://hal.inria.fr/hal-01567661.

[44] N. POZIN, S. MONTESANTOS, I. KATZ, M. PICHELIN, I. VIGNON-CLEMENTEL, C. GRANDMONT.*From dynamical lung ventilation data to plugs distribution in asthma – A numerical diagnosis tool*, July 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01568065.

References in notes

[45] 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.

[46] 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.

[47] 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].

[48] R. CHATELIN, D. ANNE-ARCHARD, M. MURRIS-ESPIN, M. THIRIET, P. PONCET.*Numerical and experimental investigation of mucociliary clearance breakdown in cystic fibrosis*, in "Journal of Biomechanics", 2017, 8 [*DOI :* 10.1016/J.JBIOMECH.2016.12.026], https://hal.inria.fr/hal-01476172.

[49] 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.

[50] 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.

[51] O. MIRAUCOURT, S. SALMON, M. SZOPOS, M. THIRIET.*Blood flow in the cerebral venous system: modeling and simulation*, in "Computer Methods in Biomechanics and Biomedical Engineering", 2016 [*DOI :* 10.1080/10255842.2016.1247833], http://hal.upmc.fr/hal-01384285.

[52] R. MOTTE.*A numerical method for solving particle-fluid equations*, in "Trends in numerical and physical modeling for industrial multiphase flows", Cargèse, France, 2000.

[53] 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.

[54] F. SACHSE.*Computational Cardiology: Modeling of Anatomy, Electrophysiology, and Mechanics*, Springer-Verlag, 2004.

[55] J. SUNDNES, G. LINES, X. CAI, B. NIELSEN, K.-A. MARDAL, A. TVEITO.*Computing the electrical activity in the heart*, Springer-Verlag, 2006.

[56] K. TERZAGHI.*Theoretical Soil Mechanics*, John Wiley and Sons, New-York, 1943.

[57] 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. - 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
B9.4.5. - Data science

1. Personnel

Research Scientists

Fawzi Nashashibi [Team leader, Inria, Senior Researcher, HDR]
Guy Fayolle [Inria, Senior Researcher Emeritus]
Jean-Marc Lasgouttes [Inria, Researcher]
Gérard Le Lann [Inria, Senior Researcher Emeritus]
Anne Verroust-Blondet [Inria, Researcher, HDR]

External Collaborators

Oyunchimeg Shagdar [VEDECOM, until Aug 2017]
Itheri Yahiaoui [Univ de Reims Champagne-Ardennes]

Technical Staff

Rafael Colmenares [Inria, until May 2017]
Azary Abboud [Inria, until Jan 2017]
Mohammad Abualhoul [Inria]
Younes Bouchaala [Inria, from Oct 2017]
Raoul de Charette [Inria]
Mohamed Elhadad [Inria, from Oct 2017]
Ahmed Soua [Inria, until Jun 2017]
Thomas Streubel [Inria, until Aug 2017]
Ilias Xydias [Inria, from Aug 2017]
Armand Yvet [Inria]

PhD Students

Zayed Alsayed [VEDECOM]
Pierre de Beaucorps [Inria]
Carlos Flores [Inria]
Fernando Garrido [VEDECOM]
Farouk Ghallabi [CIFRE Renault]
David González Bautista [Inria, until Mar 2017]
Maximilian Jaritz [CIFRE Valeo]
Imane Mahtout [CIFRE Renault, from Dec 2017]
Kaouther Messaoud [Inria, from Apr 2017]
Francisco Navas [Inria]
Dinh-Van Nguyen [Vietnamese grant]
Danut-Ovidiu Pop [Inria]
Luis Roldao Jimenez [CIFRE AKKA, from Oct 2017]

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, Raoul de Charette, Rafael Colmenares, Aitor Gomez, Fernando Garrido, Farouk Ghallabi, Aitor Gomez, David González Bautista, Kaouther Messaoud, Francisco Navas, Fawzi Nashashibi, Carlos Flores, Dinh-Van Nguyen, Danut-Ovidiu Pop, Luis Roldao Jimenez, Oyunchimeg Shagdar, Thomas Streubel, 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, Rafael Colmenares, Farouk Ghallabi, Aitor Gomez, Fawzi Nashashibi, Dinh-Van Nguyen, Danut-Ovidiu Pop, Luis Roldao Jimenez, 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.

3.1.2. Cooperative Multi-sensor data fusion

Participants: Fawzi Nashashibi, Oyunchimeg Shagdar.

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: Fernando Garrido, David González Bautista, Imane Mahtout, Fawzi Nashashibi, Francisco Navas, Carlos Flores.

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. V2V and V2I Communications for ITS

Participants: Oyunchimeg Shagdar, Gérard Le Lann, Mohammad Abualhoul, Younes Bouchaala, Fawzi Nashashibi.

Wireless communications are expected to play an important role for road safety, road efficiency, and comfort of road users. Road safety applications often require highly responsive and reliable information exchange between neighboring vehicles in any road density condition. Because the performance of the existing radio

communications technology largely degrades with the increase of the node density, the challenge of designing wireless communications for safety applications is enabling reliable communications in highly dense scenarios. Targeting this issue, RITS has been working on medium access control design and visible light communications, especially for highly dense scenarios. The works have been carried out considering the vehicle behavior such as vehicle merging and vehicle platooning.

Unlike many of the road safety applications, the applications regarding road efficiency and comfort of road users, on the other hand, often require connectivity to the Internet. Based on our expertise in both Internet-based communications in the mobility context and in ITS, we are now investigating the use of IPv6 (Internet Protocol version 6 which is going to replace the current version, IPv4, in a few years from now) for vehicular communications, in a combined architecture allowing both V2V and V2I.

The wireless channel and the topology dynamics need to be studied when understanding the dynamics and designing efficient communications mechanisms. Targeting this issue, we 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 geocast message delivery, and access point selection.

Below follows a more detailed description of the related research issues.

3.2.1. Geographic multicast addressing and routing

Participant: Oyunchimeg Shagdar.

Many ITS applications such as fleet management require multicast data delivery. Existing work on this subject tackles mainly the problems of IP multicasting inside the Internet or geocasting in the VANETs. To enable Internet-based multicast services for VANETs, we introduced a framework that:
i) defines a distributed and efficient geographic multicast auto-addressing mechanism to ensure vehicular multicast group reachability through the infrastructure network,
ii) introduces a simplified approach that locally manages the group membership and distributes the packets among them to allow simple and efficient data delivery.

3.2.2. Platooning control using visible light communications

Participants: Mohammad Abualhoul, Oyunchimeg Shagdar, Fawzi Nashashibi.

The main purpose of our research is to propose and test new successful supportive communication technology, which can provide stable and reliable communication between vehicles, especially for the platooning scenario. Although VLC technology has a short history in comparison with other communication technologies, the infrastructure availability and the presence of the congestion in wireless communication channels lead to propose VLC technology as a reliable and supportive technology which can takeoff some loads of the wireless radio communication. The first objective of this work is to develop an analytical model of VLC to understand its characteristics and limitations. The second objective is to design vehicle platooning control using VLC. In platooning control, a cooperation between control and communication is strongly required in order to guarantee the platoon's stability (e.g. string stability problem). For this purpose we work on VLC model platooning scenario, to permit for each vehicle the trajectory tracking of the vehicle ahead, altogether with a prescribed inter-vehicle distance and considering all the VLC channel model limitations. The integrated channel model of the main Simulink platooning model will be responsible for deciding the availability of the Line-of-Sight for different trajectory's curvatures, which means the capability of using light communication between each couple of vehicles in the platooning queue. At the same time the model will compute all the required parameters acquired from each vehicle controller.

3.2.3. V2X radio communications for road safety applications

Participants: Mohammad Abualhoul, Oyunchimeg Shagdar, Fawzi Nashashibi.

While 5.9 GHz radio frequency band is dedicated to ITS applications, the channel and network behaviors in mobile scenarios are not very well known. In this work we theoretically and experimentally study the radio channel characteristics in vehicular networks, especially the radio quality and bandwidth availability. Based on our study, we develop mechanisms for efficient and reliable V2X communications, channel allocation, congestion control, and access point selection, 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: λ for channel access delays, and Δ 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 λ and Δ 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 [48]. 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 Elhadad, 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 [44], 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 [50];
- build macroscopic variables that measure the overall state of the underlying graph, in order to improve the local propagation of information [45];
- make the underlying model as sparse as possible, in order to improve BP convergence and quality [49].

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 [42] 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 [39]):

$$Q(x,y)\pi(x,y) = q(x,y)\pi(x) + \widetilde{q}(x,y)\widetilde{\pi}(y) + \pi_0(x,y). \tag{1}$$

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 [43], 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. New Software and Platforms

5.1. 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

5.2. 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

5.3. 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

6. New Results

6.1. Scene Understanding with Computer Vision

Participants: Maximilian Jaritz, Raoul de Charette, Rafael Colmenares, Ziyang Hong, Fawzi Nashashibi.

This axis is in the continuation of previous year axis on scene understanding. It is crucial for autonomous driving. While last year we focused more on road estimation and ego velocity estimation (research report [51]), this year we focused on object recognition either from a single RGB camera or from a fusion of sensors (PhD of Maximilian Jaritz). Road estimation was also extended using graph energy minimization techniques and lead to interesting results in the scope of Rafael Colmenares internship. For object recognition a number of popular deep learning techniques were evaluated and the outcome of this evaluation study is that existing approaches suffers either from performance issues or processing time issues. In the scope of Maximilian Jaritz thesis a multi-modal approach is being developed were RGB and LiDAR are used to detect objects in the direct vicinity of the autonomous car. Preliminary results using state-of-the-art network architecture leads to satisfactory performances in terms of precision but a non-optimal localization of their spatial position (especially when rotated).

6.2. Computer Vision in Bad Weather

Participants: Raoul de Charette, Aitor Gomez, Sule Kahraman.

Common assumption of any perception system is to consider the atmosphere transparent so that the light rays travel directly from a point in the scene to the camera. While this assumption is true in clear weather, in fog/rain/hail or snow conditions this assumption isn't valid and all perception system will struggle. This can have a dramatic impact in autonomous driving. Following some of his previous works in former labs, Raoul de Charette lead several works to investigate and quantify the influence of rain and fog on computer vision for autonomous driving. Two internships were conducted in that axis (Aitor Gomez, Sule Kahraman) and there are on going results and research to be output. More detail can be found in [41].

6.3. Perception for Cooperative Driving

Participants: Raoul de Charette, Carlos Flores, Francisco Navas, Fawzi Nashashibi.

In cooperation with the control/planning group the computer vision group has worked on practical and applied research for 2D processing of LiDAR sensors in the context of cooperative driving. Practical failures cases were addressed such as the case of occluded vulnerables. In a dense urban environment were buildings may occlude pedestrian we proposed for example a perception system fusing both LiDAR and communication data retrieved from pedestrian communication streaming there GPS position. This allows us to detect and predict possible collisions of car and pedestrians. Experiments were conducted in the site of Rocquencourt and the results lead to a submit journal publication in cooperation with Vicente Milanés from Renault. Another practical case of failures in cooperative driving occurs are the cut-in or cut-out cars in platoon scenarios. When cars travel in a platoon, a car leaving or entering may disrupt the whole platoon. In collaboration with the control group, the detection and prediction of such behavior was addressed using 2D LiDAR data and tested on Cycabs. A journal was submitted in cooperation with Vicente Milanés from Renault.

6.4. Recognizing Pedestrians using Cross-Modal Convolutional Networks

Participants: Danut-Ovidiu Pop, Fawzi Nashashibi.

Pedestrian detection and recognition is of great importance for autonomous vehicles. A pedestrian detection system depends on: 1) the sensors utilized to capture the visual data, 2) the features extracted from the acquired images and 3) the classification process. Considering existing data-sets of images (Daimler, Caltech and KITTI) we have focused only on the last two points. Our question is whether one modality can be used exclusively (standpoint one) for training the classification model used to recognize pedestrians in another modality or only partially (standpoint two) for improving the training of the classification model in another modality. If it is trained on multi-modal data, can the system still work when the data from one of the domains is missing? How much information is redundant across the domains (can we regenerate data in one domain on the basis of the observation from the other domain)? How could a multi-modal system be trained, when data in one of the modalities is scarce (e.g. many more images in the visual spectrum than depth). To our knowledge, these questions have not yet been answered for the pedestrian recognition task. Our work proposes to solve this brain-teaser through various experiments based on the Daimler stereo vision data set. This year, we perform the following experimental studies (More detail can be found in [32], [33], [34]):

1. Three different image modalities (Intensity, Depth, Optical Flow) for improving the classification component are considered. The Classical Training and the Cross Training methods are analyzed. On the Cross Training method, the CNN is trained and validated on different images modalities, in contrast to classical training method in which the training and validation of each CNN is on same images modality.
2. In [33], [34] we study how learning representations from one modality would enable prediction for other modalities, which one terms as cross modality. Several approaches are proposed:

   a) A correlated model where a unique CNN is trained with Intensity, Depth and Flow images for each frame,

b) An incremental model where a CNN is trained with the first modality images frames, then a second CNN, initialized by transfer learning on the first one is trained on the second modality images frames, and finally a third CNN initialized on the second one, is trained on the last modality images frames.

c) A particular cross-modality model, where each CNN is trained on one modality, but tested on a different one.

3. In [32] two different fusion schemes are studied:

   a) The early fusion model is built by concatenating three image modalities (intensity, depth and optical flow) to feed a unique CNN.

   b) The late fusion model consists in fusing the outputs scores (the class probability estimate) of three independent CNNs, trained on intensity, depth and optical flow images, by a classifier system.

6.5. A Fusion method of WiFi and Laser-SLAM for Vehicle Localization

Participants: Dinh-Van Nguyen, Fawzi Nashashibi.

Precise positioning plays a key role in successful navigation of autonomous vehicles. A fusion architecture of Global Positioning System (GPS) and Laser-SLAM (Simultaneous Localization and Mapping) is widely adopted. While Laser-SLAM is known for its highly accurate localization, GPS is still required to overcome accumulated error and give SLAM a required reference coordinate. However, there are multiple cases where GPS signal quality is too low or not available such as in multi-story parking, tunnel or urban area due to multipath propagation issue etc. [30] proposes an alternative approach for these areas with WiFi Fingerprinting technique to replace GPS. Result obtained from WiFi Fingerprinting are then fused with LaserSLAM to maintain the general architecture, allow seamless adaptation of vehicle to the environment (cf. [29]).

6.6. SLAM failure scenario detection for laser-based SLAM methods

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 [18]. This year, we worked on the detection of SLAM failure and non-failure scenarios and a technique detecting *a priori* potential failure scenarios for 2D laser-based SLAM methods has been introduced. Our approach is independent of the underlying SLAM implementation as it uses raw sensor data to extract a relevant scene descriptor, which is used in a decision-making process to detect failure scenarios. Experimental evaluations on three realistic experiments show the relevance of our approach. See [22] for more detail.

6.7. Motion planning techniques

Participants: Fernando Garrido, David González Bautista, Fawzi Nashashibi.

Overtaking and lane change maneuvers represent some of the major causes of fatalities in road transport. The role of the path planning in these maneuvers is essential, not only for designing collision-free trajectories, but also to provide comfort to the occupants of the vehicle.

Having this in mind, a novel two-phase dynamic local planning algorithm to deal with these dynamic scenarios has been proposed, based on previous work. In the first phase (pre-planning) [47], a multi-objective trajectory optimization considering static information (i.e. digital maps) is carried out, using quartic Bézier curves as the path generation, which let us consider the constraints of both vehicle and road, generating continuous paths in the next phase. In the second phase (real-time planning) [46], time-horizon based trajectory generation is provided on a real-time using the pre-planned information. A human-like driving style is provided evaluating the sharpness of the road bends and the available space among them, smoothing the path. There, the paths are generated by joining the already optimized quartic Bézier curves ensuring continuity in the transitions among bends and straights.

Based on this architecture, a dynamic path planning approach has been introduced to safely avoid the possible obstacles in the path. A grid based solution has been developed to discretize the space and process the obstacles. It computes a virtual lane that re-plans the local path to be tracked by modifying the global itinerary using a geometric approach considering dynamics of both overtaking and overtaken vehicles to find smooth lane changes. That way, the dynamic problem can be addressed with the described real-time static local planner. Then, the overtaking path is built by joining two curves for each lane change, minimizing the slopes, according to the virtual lane configuration, loading these curves from the pre-planning stage.

The proposed architecture has been validated both on simulation (with Pro-Sivic and RTMaps) and on the Inria Rocquencourt terrain (with Cybercars and a Citroen C1) for the static scenario, and on simulation for the dynamic scenario. The results showed a smoother tracking of the curves, reduction on the execution times and reduced global accelerations increasing comfort. Future works will improve the capacity to deal with unexpected circumstances while making the overtaking maneuvers, testing with different car types as obstacles.

6.8. Decision-making for automated vehicles adapting human-like behavior

Participants: Pierre de Beaucorps, Thomas Streubel, Anne Verroust-Blondet, Fawzi Nashashibi.

Learning from human driver's strategies for solving complex and potentially dangerous situations including interaction with other road users has the potential to improve decision-making methods for automated vehicles. In [37], we focus on simple unsignalized intersections and roundabouts in presence of another vehicle. We propose a human-like decision-making algorithm for these scenarios built up from human drivers recordings. The algorithm includes a risk assessment to avoid collisions in the intersection area. Three road topologies with different interaction scenarios were presented to human participants on a previously developed simulation tool. The same scenarios have been used to validate our decision-making process. We obtained promising results with no collisions in all setups and the ability to successfully determine to go before or after another vehicle. A further study was conducted in [36] to assess the acceptability of the approach by human drivers.

6.9. Deep Reinforcement Learning for end-to-end driving

Participants: Maximilian Jaritz, Raoul de Charette, Fawzi Nashashibi.

We conducted works on a very new research field that is end-to-end driving, where an artificial intelligence learns to drive directly from RGB images, without the use of any mediated perception (object recognition, scene understanding). Using a recent rally game with realistic physics and graphics we have trained a car in a simulator to drive. Several approaches were attempted. The most successful one uses an Asynchronous Actor Critic (A3C) trained in an end-to-end fashion and propose new strategies that improve training and generalization. The network was trained simultaneously on tracks with various road structures (sharp turns, etc.), graphics (snow, mountain, and coast) and physics (road adherence). As for other problems, we have shown that learning in a simulated environment (here a racing car game) can be transposed to other tracks and even real driving. Despite complex and varying dynamics of the car and road the trained agent learns to drive in challenging scenarios using only RGB image and vehicle speed. To prove its generalization the algorithm is also tested in unseen tracks, under legal speed limit and with real images. Initial work was published in [31] and recent works were submitted. The work was conducted in cooperation with Etienne Perot and Marin Toromanoff from Valeo.

6.10. A Time Gap-Based Spacing Policy for Full-Range Car-Following

Participants: Carlos Flores, Fawzi Nashashibi.

Car-Following techniques are a promising solution to reduce traffic jams, while increasing driver comfort and safety. The first version of such systems, Adaptive Cruise Control (ACC), proposes the employment of throttle/brake automation with ranging sensors to regulate the spacing gap with respect to the vehicle in front. Afterwards, the addition of Vehicle to Vehicle (V2V) communication links permits tighter string formations allowing Cooperative-ACC (CACC). The reaction time towards speed changes from forward vehicles can be significantly reduced, given that the ego-vehicle reacts before an spacing error is detected in feedback, employing preceding or leader vehicles' information.

To take further advantage of car-following benefits, a spacing policy is introduced in the control structure in function of the application requirements. In the state-of-the-art approaches, several works have proposed different policies to address performance metrics as: safety, traffic flow increase, stability, string stability, among others. A more complete spacing policy is studied to target all of these criteria for the full speed range and adaptable for both ACC and CACC techniques.

Towards achieving these goals, it is proposed to divide the speed range in low/high speeds and employ a variable time gap setting. A time gap transition from the minimum value for which string stability is ensured to the targeted value in high speeds is suggested. The minimal distance required in case of an unexpected braking on the preceding vehicle is also evaluated to determine the distance to keep at standstill. Both the time gaps and standstill distance are in function of the employed technique–i.e. ACC or CACC–. Among the research lines to be followed, one can mention:

- Development of a robust controller based on fractional-order calculus to achieve a more performing car-following, fulfilling more requirements.
- Further investigation on the effects of communication delays and latency in the V2V links, as well as study different control structures that react not with the preceding vehicle's behavior but also other string members.
- Consider strings which vehicles may account with different dynamics, which introduces perturbations to the car-following control structure.

More detail can be found in [23].

6.11. Plug&Play control for highly non-linear systems: Stability analysis of autonomous vehicles

Participants: Francisco Navas, Fawzi Nashashibi.

The final stage for automating a vehicle relies on the control algorithms. They are in charge of providing the proper behavior and performance to the vehicle, leading to provide fully automated capabilities. Controllability and stability of dynamic complex systems are the key aspects when it comes to design intelligent control algorithms for vehicles.

Nowadays, the problem is that control systems are "monolithic". That means that a minor change in the system could require the entire redesign of the control system. It addresses a major challenge, a system able to adapt the control structure automatically when a change occurred.

An autonomous vehicle is built by combining a set-of-sensors and actuators together with sophisticated algorithms. Since sensors and actuators are prone to intermittent faults, the use of different sensors is better and more cost effective than duplicating the same sensor type. The problem is to deal with the different availability of each sensor/actuator and how the vehicle should react to these changes. Another possible modification is the change in vehicle dynamics over time; or difference in dynamics from one vehicle to another.

A methodology that improves the security of autonomous driving systems by providing a framework managing different dynamics and sensor/actuator setups should be carried out. New trends are proposing intelligent algorithms able to handle any unexpected circumstances as unpredicted uncertainties or even fully outages from sensors. This is the case of Plug & Play control, which is able to provide stability responses for autonomous vehicles under uncontrolled circumstances.

Here, the basis of Plug & Play control, Youla-Kucera parameterization, has been used to develop different applications within the autonomous driving field.

- Stable controller reconfiguration when some change occurs. Last year, the already commercially available Adaptive Cruise Controller (ACC) system, and its evolution by adding vehicle-to-vehicle communication (CACC) were examined. The Youla-Kucera parameterization was used for providing stable transitions between both controllers when the vehicle-to-vehicle communication link is changing from available to disable or vice-versa. More details can be found in [52]. This year,

this work has been extended in what is called Youla-Kucera-based Advanced Cooperative Adaptive Cruise Control (ACACC). In the literature, CACC degrades to ACC when communication when the preceding vehicle is no longer available. This degradation occurs even if information from another V2V-equipped vehicle ahead (different from the preceding vehicle) is still available. ACACC benefits from the existing communication with this vehicle ahead in the string, reducing the inter-vehicle distance whereas keeping string stability. The proposed structure uses YK parameterization to obtain a hybrid behavior between two CACC controllers with different time gaps. Stable transition between both controllers is also ensured. This work has been submitted to IEEE Transactions on Vehicular Technology. Finally, Youla-Kucera has been also employed to assure stable transitions when other CACC-equipped vehicles are joining/leaving a CACC string of vehicles.

- Online closed loop identification. Youla-Kucera has a dual formulation that allows recasting closed-loop identification into open-loop-like identification. [28] deals with the identification of longitudinal dynamics of a cycab for subsequent control performance's improvement. Here, the dual Youla-Kucera formulation is used to transform a closed-loop identification problem in an open-loop-like. The algorithm is tested in a string of two cycabs equipped with a proportional-derivative-based CACC, showing how the resulting model is improved in comparison with a classical open-loop identification algorithm. Closed-loop identification results have been also obtained for a production vehicle when connected to a lane following control system. Thanks to that, lateral dynamics are known for velocities between 8 and $20m/s$.
- A final step that integrates both stable controller reconfiguration and closed-loop identification: Automatic control reconfiguration to achieve optimal performance based on the identification of the new situation. This idea has been used to obtain an adaptive approach able to ensure string stability when different dynamics are involved in the same string of vehicles (a heterogeneous string of vehicles). A supervisor is able to provide the closest model in a predefined set, activating the controller that ensures string stability. The closest model in the set can be known without using identification algorithms, thanks to Youla-Kucera properties, with the consequent computational saving.

6.12. Large scale simulation interfacing

Participants: Ahmed Soua, Jean-Marc Lasgouttes, Oyunchimeg Shagdar.

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 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.

6.13. Belief propagation inference for traffic prediction

Participant: Jean-Marc Lasgouttes.

This work [50], [49], in collaboration with Cyril Furtlehner (TAU, Inria), deals with real-time prediction of traffic conditions in a setting where the only available information is floating car data (FCD) sent by probe vehicles. The main focus is on finding a good way to encode some coarse information (typically whether traffic on a segment is fluid or congested), and to decode it in the form of real-time traffic reconstruction and prediction. Our approach relies in particular on the belief propagation algorithm.

This year, following an agreement signed with the company SISTeMA ITS (Italy), we obtained access to large amounts of data from the cities of Vienna and Turin. We are now working on assessing the performance of our techniques in real-world city networks, and to compare it to the sate of the art techniques.

6.14. Platoons Formation for autonomous vehicles redistribution

Participants: Mohamed Elhadad, Jean-Marc Lasgouttes, Ilias Xydias.

As part of the VALET ANR project, we aim to optimize platoon formation for vehicle retrieval, where parked vehicles are collected and guided by a fleet manager in a given area. Each platoon follows an optimized route to collect and guide the parked vehicles to their final destinations. The Multi-Platoons Parked Vehicles Collection consists in minimizing the total travel duration, total travel distance, the number of platoons, under constraints of battery level. After a linear formal definition of the problem, we show how to use a multi-objective version of genetic algorithms, more precisely the NSGA-II algorithm, to solve this multi-criteria optimization problem.

This is a work in progress.

6.15. Random Walks in Orthants

Participant: Guy Fayolle.

The Second Edition of the Book [39] *Random walks in the Quarter Plane*, prepared in collaboration with R. Iasnogorodski (St-Petersburg, Russia) and V. Malyshev (MGU, Moscow), has been published by Springer in the collection *Probability Theory and Stochastic Processes*.

Part II of this second edition borrows specific case-studies from queueing theory and enumerative combinatorics. Five chapters have been added, including examples and applications of the general theory to enumerative combinatorics. Among them:

- Explicit criteria for the finiteness of the group, both in the genus 0 and genus 1 cases.
- Chapter *Coupled-Queues* shows the first example of a queueing system analyzed by reduction to a BVP in the complex plane.
- Chapter *Joining the shorter-queue* analyzes a famous model, where maximal homogeneity conditions do not hold, hence leading to a system of functional equations.
- Chapter *Counting Lattice Walks* concerns the so-called *enumerative combinatorics*. When counting random walks with small steps, the nature (rational, algebraic or holonomic) of the generating functions can be found and a precise classification is given for the basic (up to symmetries) 79 possible walks.

6.16. Lattice path combinatorics

Participant: Guy Fayolle.

In the second edition of the book [39], 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, 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 which case the underlying algebraic curve is of genus 0 or 1. 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 Galois theory for genus greater than 1 (i.e. when some jumps are ≥ 2), etc.

6.17. Facing ADAS validation complexity with usage oriented testing

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 maybe 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.

The COVADEC project (type FUI/FEDER 15), which was aiming to provide methods and techniques to deal with these problems, was actually successfully completed in May 2017. The test cases automatic generation relies on a *Model Based Testing (MBT)* approach. The tool used for MBT is the software MaTeLo (Markov Test Logic), developed by the company All4Tec. MaTeLo is an MBT tool, which makes it possible to build a model of the expected behavior of the system under test and then to generate, from this model, a set of test cases suitable for particular needs. MaTeLo is based on Markov chains, and, for non-deterministic generation of test cases, uses the Monte Carlo methods. To cope with the inherent combinatorial explosion, we couple the graph generated by MaTeLo to an ad hoc *random scan Gibbs sampler (RSGS)*, which converges at geometric speed to the target distribution. Thanks to these test acceleration techniques, MaTeLo also makes it possible to obtain a maximal coverage of system validation by using a minimum number of test cases. As a consequence, the number of driving kilometers needed to validate an ADAS is substantially reduced, see [53] and [54]. These methods do interest the French manufacturer *Groupe PSA*, who wishes to establish a contractual collaboration involving Armines-MINES ParisTech.

6.18. Safety, Privacy, Trust, and Immunity to Cyberthreats

Participant: Gérard Le Lann.

Safety (significant reductions of severe accident figures) and traffic efficiency (smaller safe inter-vehicular gaps, higher occupancy of asphalt resources) are dual and antagonistic goals targeted with autonomous vehicles. On-board robotics and inter-vehicular communications (IVCs) are essential for achieving proactive and reactive safety (ability to influence behaviors and moves of nearby vehicles).

Existing US standards (WAVE) and European standards (ETSI ITS-G5) for IVCs based on omnidirectional radio technologies have been shown to be inadequate in this respect. Numerous publications demonstrate that they induce channel access delays which are unacceptably high in average and worst-case load or contention conditions. Periodic beaconing (the broadcasting of messages carrying identifiers, UTC time and GNSS positions) at frequencies ranging from 1 Hz to 10 Hz is mistakenly believed to provide every vehicle with a correct local dynamic map (LDM) giving the accurate geo-localizations of surrounding vehicles. Radio broadcasts are unreliable. Therefore, the LDMs of any two vehicles arbitrarily close to each other may differ. Safe coordination implies exact agreements (a.k.a. consensus), i.e. strictly identical LDMs. This has been shown to be impossible in asynchronous systems (WAVE/G5 networks) and in synchronous systems (deterministic MAC protocols) in the presence of message losses.

Periodic beaconing may lead to radio channel saturation. Furthermore, since GNSS coordinates are unencrypted, periodic beaconing atop WAVE/G5 favors eavesdropping and tracking, as well as cyberattacks from unknown distant entities (malicious vehicles or terrestrial nodes). Pseudonymous authentication based on asymmetric key pairs and certificates delivered by Public Key Infrastructures shall thwart such threats. Unfortunately, numerous problems are yet unsolved. Tracking and cyberattacks are feasible with the set of aforementioned solutions (referred to as WAVE 1.0).

In 2017, we have contributed to the work conducted by scientists and engineers in various countries, aimed at demonstrating that it is possible to achieve safety, privacy, trust, and immunity to cyberthreats altogether (no mitigation), following approaches that differ from WAVE 1.0. We are also working with experts who have expressed concerns regarding the risks of cyber-surveillance induced by WAVE 1.0 solutions when better solutions are available. Two essential observations are in order.

Firstly, networks of connected autonomous vehicles are instances of life-critical systems. Inevitably, future on-board (OB) systems will have to be designed in accordance with the segregation principle (a fundamental design rule in the domain of safety/life-critical systems). A critical sub-system must be isolated from a non-critical sub-system. In a vehicle, a critical sub-system hosts critical robotics and critical IVCs (novel IVC protocols and distributed algorithms for time-bounded decision-making and IV coordination). WAVE 1.0 solutions are implemented in the non-critical sub-system.

Secondly, only vehicles very close to each other may be involved in an accident. It follows that short-range and directional IVCs are necessary and sufficient for safety. In [25] and [27], we present IVC protocols and agreement algorithms that achieve small worst-case time bounds for longitudinal and lateral message dissemination within and across cohorts (spontaneous linear vehicular networks). These bounds are such that no vehicle moves by more than 1 asphalt slot while messages are being disseminated and agreements are reached, in the presence of message losses. A brief summary can be found in [38]. Similar IVC protocols and agreement algorithms can be devised for upcoming technologies, namely 5G radio communications (MIMO antennas) and optical communications ignored in WAVE 1.0 solutions.

These solutions (referred to as WAVE 2.0) have additional merits regarding cyberthreats. Remote cyberattacks cannot jeopardize safety (contrary to WAVE 1.0), given that OB critical sub-systems are isolated from *the outside world*. This is discussed in [24] and in [26]. In [26], we introduce an OB system architecture consistent with the segregation principle, which includes a tamper-proof device (for non-repudiation and accountability), and novel protocols for IVCs. In addition to pseudonymous authentication, sources and destinations of safety messages are fully anonymous, and certified pseudonyms can be used ad infinitum, thus circumventing the deficiencies of WAVE 1.0 solutions. With WAVE 2.0 solutions, proximate eavesdropping and tracking are unfeasible and vain. Also, we show that proximate cyberattacks (e.g., masquerading, injection of bogus data, falsification, Sybil attack) are immediately detected, and how to stop a malicious or misbehaving vehicle safely.

Our on-going research targets crossings of un-signaled intersections, roundabouts, and spontaneous formations of heterogeneous vehicular networks (SAE automation levels from 0 to 5), where properties of safety, efficiency, privacy and immunity to cyberattacks shall hold.

7. Bilateral Contracts and Grants with Industry

7.1. 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 and the post-doc of Thomas Streubel under the framework of VALEO project “Daring”
- *SMART* project: on the *Design and development of multisensor fusion system for road vehicles detection and tracking*. This project funds the internship of Alfredo Valle.
- 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 begun on November 2017 (Imane MATHOUT).

AKKA Technologies: Collaboration with AKKA since 2012 (for the Link & Go prototype).

- Inria and AKKA Technologies are partners in the COCOVEA and the VALET projects (ANR projects).
- A new CIFRE PhD thesis (Luis ROLDAO JIMENEZ) dealing with 3D-environment modeling for autonomous vehicles begun in October 2017.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

8.1.1.1. COCOVEA

Title: Coopération Conducteur-Véhicule Automatisé

Instrument: ANR

Duration: November 2013 - April 2017

Coordinator: Jean-Christophe Popieul (LAMIH - University of Valenciennes)

Partners: LAMIH, IFSTTAR, Inria, University of Caen, COMETE, PSA, CONTINENTAL, VALEO, AKKA Technologies, SPIROPS

Inria contact: Fawzi Nashashibi

Abstract: CoCoVeA project aims at demonstrating the need to integrate from the design of the system, the problem of interaction with the driver in resolving the problems of sharing the driving process and the degree of freedom, authority, level of automation, prioritizing information and managing the operation of the various systems. This approach requires the ability to know at any moment the state of the driver, the driving situation in which he finds himself, the operating limits of the various assistance systems and from these data, a decision regarding activation or not the arbitration system and the level of response.

8.1.1.2. 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.

8.1.2. FUI

8.1.2.1. Sinetic

Title: Système Intégré Numérique pour les Transports Intelligents Coopératifs

Instrument: FUI

Duration: December 2014 - May 2017

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.

8.1.2.2. 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.

8.1.3. 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 is financing the PhD theses of Mr. Fernando Garrido and Mr. Zayed Alsayed.

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

8.2.1.1. 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 - December 2018

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.

8.2.2. Collaborations with Major European Organizations

RITS is member of the **euRobotics AISBL** 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/

8.3. International Initiatives

8.3.1. Participation in Other International Programs

8.3.1.1. ICT-Asia

SIM-Cities

Title: "Sustainable and Intelligent Mobility for Smart Cities"

International Partner (Institution - Laboratory - Researcher):

- Nanyang Technical University (NTU), School of Electrical and Electronic Engineering – Singapore. Prof. Dan Wei Wang

- National University of Singapore (NUS), Department of Mechanical Engineering – Singapore. Dr. Marcelo Ang

- Kumamotoo University - Japan. Intelligent Transportation Systems Lab, Graduate School of Science and Technology, Prof. James Hu / Prof. Ogata

- Shanghai Jiao-Tong University (SJTU), Department of Automation – China. Prof. Ming Yang

- Hanoi University of Science and Technology, International Center MICA Institute – Vietnam. Prof. Eric Castelli

- Inria, RITS Project-Team – France. Dr. Fawzi Nashashibi

- Inria, e-Motion/CHROMA Project-Team – France. Dr. Christian Laugier

- Ecole Centrale de Nantes, IRCCyN – France. Prof. Philippe Martinet

Duration: Jan. 2015 - May 2017

Start year: 2015

This project aims at conducting common research and development activities in the field of sustainable transportation and advanced mobility of people and goods in order to move in the direction of smart, clean and sustainable cities.

RITS and MICA lab have obtained from the Vietnamese Program 911 the financing of the joint PhD thesis of Dinh-Van Nguyen (co-directed by Eric Castelli from MICA lab and Fawzi Nashashibi).

8.3.1.2. ECOS Nord – Venezuela

ECOS Nord

Title: "Les Techniques de l'Information et de la Communication pour la Conception de Systèmes Avancés de Mobilité durable en Milieu Urbain."

International Partner (Institution - Laboratory - Researcher):

- Simon Bolivar University, Department of Mecatronics – Venezuela. Dr. Gerardo Fernandez

- Inria, RITS Project-Team – France. Dr. Fawzi Nashashibi

Duration: Jan. 2014 - Dec. 2017

Start year: 2014

The main objective of this project is to contribute scientifically and technically to the design of advanced sustainable mobility systems in urban areas, particularly in dense cities where mobility, comfort and safety needs are more important than in other types of cities. In this project, we will focus on the contribution of advanced systems of perception, communication and control for the realization of intelligent transport systems capable of gradually integrating into the urban landscape. These systems require the development of advanced dedicated urban infrastructures as well as the development and integration of on-board intelligence in individual vehicles or mass transport.

This year, a session of courses has been organized at University Simon Bolivar, Caracas (Venezuela). Following several PhDs and interns recruitments from this university, prof G. Fernandez and J. Capeletto invited Raoul de Charette to organize a 32Hr Computer Vision Master Class in December 2017. PhDs Carlos Flores and Luis Roldao were also part of the master class and teached control (10Hr) and point cloud processing (7Hr), respectively.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

8.4.1.1. Internships

Julio Blanco Deniz, Nievsabel Molina from Simon Bolivar University, Venezuela.

They both worked on a cascade control architecture based on PID controllers for a Citroen C1: the longitudinal control was developed by Julio Blanco Deniz , under the supervision of Carlos Flores and the lateral control (for the action on the steering wheel) was done by Nievsabel Molina, under the supervision of Francisco Navas. Using this architecture, a reference trajectory can be smoothly followed by the vehicle at different speeds.

Aitor Gomez, Alfredo Valle, Edgar Talavera Munoz from Universidad Politécnica de Madrid, Spain.

Ziyang Hong from Université de Bourgogne, Dijon, France.

Maradona Rodrigues from University of Warwick, United Kingdom.

Sule Kahraman from MIT, USA.

Arthur Lecert from ESIEE Paris, France. He was supervised by Pierre de Beaucorps.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Organizing Committees

Mohammad Abualhoul and Fawzi Nashashibi organized a workshop within the framework of the European project AUTOCITS "Regulation Study for Interoperability in the Adoption of Autonomous Driving in European Urban Nodes". Date: 05/04/2017 - Inria Paris. https://project.inria.fr/autocits/autoc-its-workshop-paris/

9.1.2. Scientific Events Selection

9.1.2.1. Member of the Conference Program Committees

Fawzi Nashashibi: IEEE Intelligent Vehicles Symposium - IV 2017, IEEE 20th International Conference on Intelligent Transportation Systems - ITSC 2017.

9.1.2.2. Reviewer

Carlos Flores: IEEE 20th International Conference on Intelligent Transportation Systems - ITSC 2017.

Francisco Navas: IEEE Intelligent Vehicles Symposium - IV 2017, IEEE 20th International Conference on Intelligent Transportation Systems - ITSC 2017.

Anne Verroust-Blondet: IEEE Intelligent Vehicles Symposium - IV 2017, IEEE 20th International Conference on Intelligent Transportation Systems - ITSC 2017.

9.1.3. Journal

9.1.3.1. 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* and of *IEEE Transactions on Intelligent Transportation Systems*.

Fawzi Nashashibi: guest editor with Erwin Schoitsch (AIT) of *ERCIM News* N°109 on "Autonomous Vehicles".

9.1.3.2. Reviewer - Reviewing Activities

Guy Fayolle: *AAP, MPRF, PTRF, QUESTA, European Journal of Combinatorics, JSP, Physica A, Springer Science*.

Jean-Marc Lasgouttes: *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Intelligent Transportation Systems*.

Anne Verroust-Blondet: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *The Visual Computer*.

9.1.4. Invited Talks

Raoul de Charette: Talk on "Computer Vision" at Adomik (artificial intelligence software company), Paris, November 29th.

Guy Fayolle was a keynote speaker at the conference at the ACMPT-2017 Conference http://acmpt.moscow/, *(Analytical and Computational Methods in Probability Theory and its Applications)*, held in Moscow State University and in RUDN University, Moscow, 23-27th October 2017. His talk presented the substance of the article [21].

Fawzi Nashashibi was a keynote speaker at IEEE International Conference on Intelligent Vehicles Symposium - IV 2017, *Latest Advancements in Intelligent Vehicles Research in Europe*, held in June 2017.

Fawzi Nashashibi was a keynote speaker at IEEE International Conference on Vehicular Electronics and Safety (ICVES), *(Automated vehicles in urban environments: challenges and technical solutions)*, June 2017.

Fawzi Nashashibi was a keynote speaker at Convergence Technology Symposium 2017 (ConTech 2017) *(Autonomous Driving, Future Mobility, Future Smart City with Autonomous Driving)*, held in June 2017, at the AICT in Suwon, Korea.

Fawzi Nashashibi was a keynote speaker at IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), *(A functional architecture for the navigation of an autonomous vehicle)*, held in September 2017.

9.1.5. Scientific Expertise

Guy Fayolle is scientific advisor and associate researcher at the *Robotics Laboratory of Mines ParisTech*.

Jean-Marc Lasgouttes is member of the *Conseil Académique* of Université Paris-Saclay.

Anne Verroust-Blondet is member of the COST-GTRI committee at Inria and of the "emploi scientifique" committee of Inria Paris.

9.1.6. 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 relations of Inria Paris.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence: Fawzi Nashashibi, "Programmation avancée", 84h, L1, Université Paris-8 Saint-Denis, France.

Master: Raoul de Charette, "Computer Vision for Autonomous Driving", 32Hr, Master 2, University of Simon Bolivar, Venezuela, December 2017.

Master: Carlos Flores, "Control for Autonomous Driving", 10Hr, Master 2, University of Simon Bolivar, Venezuela, December 2017.

Master: Jean-Marc Lasgouttes, "Analyse de données", 54h, second year of Magistère de Finance (M1), University Paris 1 Panthéon Sorbonne, France.

Master: Luis Roldao Jimenez, "Point Cloud Processing for Autonomous Driving", 7Hr, Master 2, University of Simon Bolivar, Venezuela, December 2017.

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 2017.

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.

Doctorat: Raoul de Charette, "Introduction, Data Visualization and Signal Processing in Python", 21Hr, class organized with Paris Science et Lettres, France, January 2017.

Doctorat: Jean-Marc Lasgouttes, "Analyse de données fonctionnelles", 31.5h, Mastère Spécialisé "Expert en sciences des données", INSA-Rouen, France

9.2.2. Supervision

PhD: David González Bautista, "Architecture fonctionnelle pour la planification des trajectoires des véhicules automatisés dans des environnements complexes", Mines ParisTech, April 2017, supervisor: Fawzi Nashashibi.

PhD in progress: Zayed Alsayed, “Système de localisation redondant en environnement extérieur ouvert pour véhicule urbain automatique”, Télécom ParisTech, October 2014, supervisor: Anne Verroust-Blondet, co-supervisor: Guillaume Bresson.

PhD in progress: Pierre de Beaucorps, “Autonomous vehicle: behavior prediction and Interaction with road users”, UPMC Paris, January 2016, supervisor: Anne Verroust-Blondet, co-supervisor: Fawzi Nashashibi.

PhD in progress: Carlos Flores, “Analysis and design of cooperative systems for trains of green cars”, Mines ParisTech, December 2015, supervisor: Fawzi Nashashibi, co-supervisor: Vicente Milanés.

PhD in progress: Fernando Garrido, “Optimal trajectory generation for autonomous vehicles in urban environments” , Mines ParisTech, November 2014, supervisor: Fawzi Nashashibi, co-supervisors: Vicente Milanés, Joshué Pérez.

PhD in progress: Farouk Ghallabi, “Environment modeling and simultaneous localization of a mobile vehicle on motorways: a multi-sensor approach”, Mines ParisTech, October 2016, supervisor: Fawzi Nashashibi.

PhD in progress: Maximilian Jaritz, “Perception multi-capteur pour la conduite autonome grâce à l’apprentissage profond”, Mines ParisTech, January 2017, supervisor: Fawzi Nashashibi, co-supervisor: Raoul de Charette.

PhD in progress: Francisco Navas, “Plug&Play control for highly non-linear systems: Stability analysis of autonomous vehicles”, Mines ParisTech, October 2015, supervisor: Fawzi Nashashibi, co-supervisor: Vicente Milanés.

PhD in progress: Dinh-Van Nguyen, "Wireless sensor networks for indoor mapping and accurate localization for low speed navigation in smart cities", Mines ParisTech, December 2015, supervisor: Fawzi Nashashibi, co-supervisor: Eric Castelli.

PhD in progress: Danut-Ovidiu Pop, "Deep learning techniques for intelligent vehicles", INSA Rouen, May 2016, supervisor: Abdelaziz Bensrhair, co-supervisor: Fawzi Nashashibi.

Starting PhD: Imane Matout, “Estimation de l’intention des véhicules pour la prise de décision et le contrôle sans faille en navigation autonome”, Mines ParisTech, October 2017, supervisor: Fawzi Nashashibi, co-supervisor: Vicente Milanés.

Starting PhD: Kaouther Messaoud, “Détermination des manoeuvres et des intentions des véhicules avoisinant un véhicule autonome”, UPMC Paris, October 2017, supervisor: Anne Verroust-Blondet, co-supervisor: Fawzi Nashashibi, Itheri Yahiaoui.

Starting PhD: Luis Roldao Jimenez, “Modélisation 3D de l’environnement et de la manoeuvrabilité d’un véhicule”, UPMC Paris, October 2017, supervisor: Anne Verroust-Blondet, co-supervisor: Raoul de Charette.

9.2.3. Juries

Guy Fayolle was a jury member of PhD thesis of Younes Bouchaala - *Handling Safety Messages in Vehicular Ad-Hoc Networks (VANETs)*, University of Versailles Saint-Quentin-en-Yvelines, 21 October 2017.

Fawzi Nashashibi was a reviewer of the PhD thesis of Ange Nizard - *Planification et commande pour véhicules à deux trains directeurs en milieu encombré* Université Blaise Pascal, Clermont-Ferrand. 31 March 2017.

Fawzi Nashashibi was an examiner of the PhD thesis of Bastien Béchadergue - *Mesure de distance et transmission de données intervéhicules par phares à LED*. University of Paris-Saclay, prepared at the University of Versailles Saint-Quentin, 10 November 2017.

Fawzi Nashashibi was the Jury President for the thesis of Viet-Cuong Ta - *Smartphone-based indoor positioning using WIFI, inertial sensor and Bluetooth*, under the co-supervision of Université Grenoble Alpes and Hanoi University of Science & Technology, 15 December 2017.

Anne Verroust-Blondet was a reviewer of the PhD thesis of Mathias Paget - *Optimisation discrète et indices de stabilité appliqués à la stéréoscopie en contexte routier*, Université Paris-Est, 13 December 2017.

9.3. Popularization

RITS team: press article on the research axes of Inria RITS team entitled "RITS Team at Inria", with participation of Fawzi Nashashibi, Anne Verroust-Blondet, Jean-Marc Lasgouttes and Raoul de Charette. IEEE Intelligent Transportation Systems Magazine, Vol 9, Issue 2. April 19th. (http://ieeexplore.ieee.org/document/7904770/).

Raoul de Charette: participation to the Hackathon round-table in collaboration with approximately 10 journalists and a handful of Inria researchers. Exchange on the implementation of future autonomous transportation in cities. It lead to several press article. June 7th, Paris.

Raoul de Charette: public conference "Voiture Autonomes: Où en est-on ?" at Futur en Seine, Paris, June 8th.

Raoul de Charette: exchange on upcoming challenges for autonomous driving for the web article "Véhicules Autonomes : Où En Est-On en France ?", journalist Mathilde Ragot. T.O.M. June 14th. (http://www.tom.travel/2017/06/14/vehicules-autonomes-ou-en-est-on-en-france/).

Raoul de Charette: exchange on upcoming challenges for autonomous driving for the web article "Véhicules autonomes : où en est-on réellement ?", journalist Benoit Fleuret. Microsoft RSLN. June 22th. (https://rslnmag.fr/cite/vehicules-autonomes-ou-en-est-on-reellement/).

Raoul de Charette: interview on current progress of autonomous driving and real upcoming challenges for the web article "Pourquoi la voiture 100 % autonome n'est pas près de rouler", journalist Reynald Fléchaux. Silicon. June 26th. (https://www.silicon.fr/voiture-autonome-pas-prete-rouler-178787.html?inf_by=5a0c4d80681db858068b471e).

Raoul de Charette: interview on the progress of computer vision and danger of artificial intelligence for the press article "Robots tueurs soudés pour déssouder", journalist Erwan Cario. Libération. August 24th. (http://www.liberation.fr/futurs/2017/08/24/robots-tueurs-soudes-pour-dessouder_1591778).

Gérard Le Lann: Interview published in *Journal of Internet Histories*, vol. 1, issue 1-2, pp. 188-196, (http://tandfonline.com/doi/full/10.1080/24701475.2017.1301132).

Gérard Le Lann: interview for Inria's 50 years "Networks, distributed algorithms and critical cyber-physical systems", (https://50ans.inria.fr/en/gerard-le-lann-networks-distributed-algorithms-and-critical-cyber-physical-systems/]).

Fawzi Nashashibi: participation to the round-table on "L'Intelligence Artificielle pour le transport autonome" during the workshop "Journée de l'IA / Evenement Public ♯FranceIA" organized by SystemX at Nano-INNOV (Palaiseau), March 3rd.

Fawzi Nashashibi: participation to the round-table on "Transports et mobilité" during the workshop "Les Mystères du XXIème siècle" (http://www.mysteres21.org/edition-2017/), November 25th.

Fawzi Nashashibi: participation to the round-table on "les questions éthiques et juridiques en robotique" during the workshop on "Ethique de la recherche en numérique", organized by the CERNA, Paris (Institut Mines-Télécom), June 15th.

Fawzi Nashashibi: interview for "L'Esprit Sorcier" number 29 on "La voiture autonome", journalist: Melvin Martineau (https://www.lespritsorcier.org/dossier-semaine/la-voiture-autonome/).

Fawzi Nashashibi: interviewed by Tiffany Blandin for the book *Un Monde sans travail ?*, Paris: Seuil, 24 August 2017, (Chapter 7: la recherche sur les voitures autonomes).

Fawzi Nashashibi: interviewed by Coralie Baumard for the article "L'Intelligence Artificielle, Nouveau Moteur De l'Industrie Automobile", FORBES France, February 11th. (https://www.forbes.fr/business/lintelligence-artificielle-nouveau-moteur-de-lindustrie-automobile/).

Fawzi Nashashibi: interviewed by Jérôme Bonaldi (Science & Vie TV) for "Le Mag de la Science" on September 2017 (to be broadcasted in 2018).

10. Bibliography

Major publications by the team in recent years

[1] M. ABUALHOUL, O. SHAGDAR, F. NASHASHIBI.*Visible Light Inter-Vehicle Communication for Platooning of Autonomous Vehicles*, in "2016 IEEE Intelligent Vehicles Symposium IV2016", Gothenburg, Sweden, June 2016, https://hal.inria.fr/hal-01308430.

[2] G. FAYOLLE, R. IASNOGORODSKI, V. A. MALYSHEV.*Random walks in the Quarter Plane*, Applications of Mathematics, Springer-Verlag, 1999, n^{o} 40.

[3] C. FLORES, V. MILANÉS, F. NASHASHIBI.*Using Fractional Calculus for Cooperative Car-Following Control*, in "Intelligent Transportation Systems Conference 2016", Rio de Janeiro, Brazil, IEEE, November 2016, https://hal.inria.fr/hal-01382821.

[4] 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.

[5] D. GONZÁLEZ BAUTISTA, J. PÉREZ RASTELLI, R. LATTARULO, V. MILANÉS, F. NASHASHIBI.*Continuous curvature planning with obstacle avoidance capabilities in urban scenarios*, in "2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)", Qingdao, China, October 2014, https://hal.inria.fr/hal-01086888.

[6] G. LE LANN.*Cohorts and groups for safe and efficient autonomous driving on highways*, in "Vehicular Networking Conference (VNC)", IEEE, 2011, p. 1-8.

[7] H. LI, F. NASHASHIBI.*Multi-vehicle cooperative localization using indirect vehicle-to-vehicle relative pose estimation*, in "ICVES 2012 - IEEE International Conference on Vehicular Electronics and Safety", Istanbul, Turkey, IEEE, July 2012, p. 267 - 272 [*DOI :* 10.1109/ICVES.2012.6294256], https://hal.inria.fr/hal-00763825.

[8] H. LI, F. NASHASHIBI.*Cooperative Multi-Vehicle Localization Using Split Covariance Intersection Filter*, in "IEEE Intelligent Transportation Systems Magazine", April 2013, vol. 5, n^{o} 2, p. 33-44, https://hal.inria.fr/hal-00833707.

[9] M. MAROUF, E. POLLARD, F. NASHASHIBI.*Automatic parking and platooning for electric vehicles redistribution in a car-sharing application*, in "IOSR Journal of Electrical and Electronics Engineering", 2015, vol. 10, n^{o} 1, 9, https://hal.inria.fr/hal-01254336.

[10] 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.

[11] 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° 1, p. 123-154 [*DOI :* 10.1007/s10472-015-9470-x], https://hal.inria.fr/hal-01186220.

[12] P. MERDRIGNAC, O. SHAGDAR, F. NASHASHIBI.*Fusion of Perception and V2P Communication Systems for Safety of Vulnerable Road Users*, in "IEEE Transactions on Intelligent Transportation Systems", 2016, https://hal.inria.fr/hal-01399150.

[13] P. MORIGNOT, J. PÉREZ RASTELLI, F. NASHASHBI.*Arbitration for balancing control between the driver and ADAS systems in an automated vehicle: Survey and approach*, in "2014 IEEE Intelligent Vehicles Symposium (IV)", Dearborn, United States, June 2014, p. 575 - 580 [*DOI :* 10.1109/IVS.2014.6856577], https://hal.inria.fr/hal-01081302.

[14] F. NAVAS, V. MILANÉS, F. NASHASHIBI.*Using Plug&Play Control for stable ACC-CACC system transitions*, in "Intelligent Vehicles Symposium 2016", Gothemburg, Sweden, June 2016 [*DOI :* 10.1109/IVS.2016.7535464], https://hal.inria.fr/hal-01304542.

[15] P. PETROV, F. NASHASHIBI.*Modeling and Nonlinear Adaptive Control for Autonomous Vehicle Overtaking*, in "IEEE Transactions Intelligent Transportation Systems", August 2014, vol. 15, n° 4, p. 1643–1656.

[16] G. TREHARD, E. POLLARD, B. BRADAI, F. NASHASHIBI.*On line Mapping and Global Positioning for autonomous driving in urban environment based on Evidential SLAM*, in "Intelligent Vehicles Symposium - IV 2015", Seoul, South Korea, June 2015, https://hal.inria.fr/hal-01149504.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[17] D. GONZA´LEZ BAUTISTA.*Functional architecture for automated vehicles trajectory planning in complex environments*, PSL Research University, April 2017, https://pastel.archives-ouvertes.fr/tel-01568505.

Articles in International Peer-Reviewed Journal

[18] G. BRESSON, Z. ALSAYED, L. YU, S. GLASER.*Simultaneous Localization And Mapping: A Survey of Current Trends in Autonomous Driving*, in "IEEE Transactions on Intelligent Vehicles", 2017, vol. XX, 1 [*DOI :* 10.1109/TIV.2017.2749181], https://hal.archives-ouvertes.fr/hal-01615897.

[19] D. GONZALEZ, J. PÉREZ, V. MILANÉS.*Parametric-based path generation for automated vehicles at roundabouts*, in "Expert Systems with Applications", April 2017, vol. 71, p. 332 - 341 [*DOI :* 10.1016/j.eswa.2016.11.023], https://hal.inria.fr/hal-01674532.

[20] Z. YASSEEN, A. VERROUST-BLONDET, A. NASRI.*View selection for sketch-based 3D model retrieval using visual part shape description*, in "Visual Computer", May 2017, vol. 33, n° 5, p. 565-583 [*DOI :* 10.1007/s00371-016-1328-7], https://hal.inria.fr/hal-01396333.

Invited Conferences

[21] G. FAYOLLE.*Functional equations as an important analytic method in stochastic modelling and in combinatorics*, in "ACMPT 2017 - Analytical and Computational Methods in Probability Theory and its Applications", Moscou, Russia, Analytic and Computational Methods in Probability Theory and its Applications, October

2017, p. 1-25, https://arxiv.org/abs/1712.02271 - To appear in MPRF (Markov Processes and Related Fields), https://hal.inria.fr/hal-01657154.

International Conferences with Proceedings

[22] Z. ALSAYED, G. BRESSON, A. VERROUST-BLONDET, F. NASHASHIBI.*Failure Detection for Laser-based SLAM in Urban and Peri-Urban Environments*, in "ITSC 2017 - IEEE 20th International Conference on Intelligent Transportation Systems", Yokohama, Japan, October 2017, p. 1-7, https://hal.inria.fr/hal-01623394.

[23] C. FLORES, V. MILANÉS, F. NASHASHIBI.*A Time Gap-Based Spacing Policy for Full-Range Car-Following*, in "Intelligent Transportation Systems Conference 2017", Yokohama, Japan, October 2017, https://hal.inria.fr/hal-01634494.

[24] G. LE LANN.*Anonymat, non-traçabilité et sécurité-innocuité dans les réseaux de véhicules autonomes connectés*, in "8ème Atelier sur la Protection de la Vie Privée (APVP'17)", Autrans, France, Equipe Privatics du laboratoire CITI d'Inria / INSA-Lyon, June 2017, https://hal.archives-ouvertes.fr/hal-01556192.

[25] G. LE LANN.*Fast Distributed Agreements and Safety-Critical Scenarios in VANETs*, in "2017 IEEE International Conference on Computing, Networking and Communications", Santa Clara, CA, United States, 2017 IEEE International Conference on Computing, Networking and Communications, IEEE ComSoc, January 2017, 7, https://hal.inria.fr/hal-01402159.

[26] G. LE LANN.*Protection de la vie privée, innocuité et immunité envers les cybermenaces dans les futurs réseaux de véhicules autonomes connectés*, in "C&ESAR 2017 - Protection des données face à la menace cyber", Rennes, France, November 2017, p. 1-22, https://hal.archives-ouvertes.fr/hal-01621500.

[27] G. LE LANN.*Safe Automated Driving on Highways – Beyond Today's Connected Autonomous Vehicles*, in "8th Complex Systems Design & Management Conference "Towards smarter and more autonomous systems"", Paris, France, December 2017, https://hal.archives-ouvertes.fr/hal-01610957.

[28] F. NAVAS, V. MILANÉS, F. NASHASHIBI.*Youla-Kucera Based Online Closed-Loop Identification For Longitudinal Vehicle Dynamics*, in "21st International Conference on System Theory, Control and Computing", Sinaia, Romania, October 2017, https://hal.inria.fr/hal-01591705.

[29] D. V. N. NGUYEN, F. NASHASHIBI, T.-K. DAO, E. CASTELLI.*Improving Poor GPS Area Localization for Intelligent Vehicles*, in "MFI 2017 - IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems", Daegu, South Korea, November 2017, p. 1-5, https://hal.inria.fr/hal-01613132.

[30] D.-V. NGUYEN, F. NASHASHIBI, T.-H. NGUYEN, E. CASTELLI.*Indoor Intelligent Vehicle localization using WiFi Received Signal Strength Indicator*, in "3rd IEEE MTT-S International Conference on Microwaves for Intelligent Mobility 2017", Nagoya, Japan, March 2017, https://hal.inria.fr/hal-01433785.

[31] E. PEROT, M. JARITZ, M. TOROMANOFF, R. DE CHARETTE.*End-to-End Driving in a Realistic Racing Game with Deep Reinforcement Learning*, in "International conference on Computer Vision and Pattern Recognition - Workshop", Honolulu, United States, July 2017, https://hal.inria.fr/hal-01620595.

[32] D. O. POP, A. ROGOZAN, F. NASHASHIBI, A. BENSRHAIR.*Fusion of Stereo Vision for Pedestrian Recognition using Convolutional Neural Networks*, in "ESANN 2017 - 25th European Symposium on Artificial Neural

Networks, Computational Intelligence and Machine Learning", Bruges, Belgium, April 2017, https://hal.inria.fr/hal-01501735.

[33] D. O. POP, A. ROGOZAN, F. NASHASHIBI, A. BENSRHAIR.*Incremental Cross-Modality Deep Learning for Pedestrian Recognition*, in "IV'17 - IEEE Intelligent Vehicles Symposium", Redondo Beach, CA, United States, June 2017, https://hal.inria.fr/hal-01501711.

[34] D. O. POP, A. ROGOZAN, F. NASHASHIBI, A. BENSRHAIR.*Pedestrian Recognition through Different Cross-Modality Deep Learning Methods*, in "IEEE International Conference on Vehicular Electronics and Safety", Vienna, Austria, June 2017, https://hal.inria.fr/hal-01588441.

[35] O. SHAGDAR, F. NASHASHIBI, S. TOHMÉ.*Performance study of CAM over IEEE 802.11p for cooperative adaptive cruise control*, in "Wireless Days, 2017", Porto, Portugal, March 2017 [*DOI :* 10.1109/WD.2017.7918118], https://hal.inria.fr/hal-01675460.

[36] T. STREUBEL, P. DE BEAUCORPS, F. NASHASHIBI.*Evaluation of automated vehicle behavior in intersection scenarios*, in "RSS2017 - Road Safety & Simulation International Conference", The Hague, Netherlands, October 2017, https://hal.inria.fr/hal-01632434.

[37] P. DE BEAUCORPS, T. STREUBEL, A. VERROUST-BLONDET, F. NASHASHIBI, B. BRADAI, P. RESENDE.*Decision-making for automated vehicles at intersections adapting human-like behavior*, in "IV'17 - IEEE Intelligent Vehicles Symposium", Redondo Beach, United States, IEEE, June 2017, https://hal.inria.fr/hal-01531516.

Conferences without Proceedings

[38] G. LE LANN.*On safety in ad hoc networks of autonomous and communicating vehicles: A rationale for time-bounded deterministic solutions*, in "CoRes 2017- 2ème Rencontres Francophones sur la Conception de Protocoles, l'Évaluation de Performance et l'Expérimentation des Réseaux de Communication", Quiberon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01517823.

Scientific Books (or Scientific Book chapters)

[39] G. FAYOLLE, R. IASNOGORODSKI, V. A. MALYSHEV. , S. ASMUSSEN, P. W. GLYNN, Y. L. 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 [*DOI :* 10.1007/978-3-319-50930-3], https://hal.inria.fr/hal-01651919.

[40] D. GONZÁLEZ, J. PÉREZ, V. MILANÉS, F. NASHASHIBI, M. SAEZ TORT, A. CUEVAS.*Arbitration and Sharing Control Strategies in the Driving Process*, in "Towards a Common Software/Hardware Methodology for Future Advanced Driver Assistance Systems The DESERVE Approach", River Publishers Series in Transport Technology , April 2017, 24, https://hal.inria.fr/hal-01676355.

Research Reports

[41] S. KAHRAMAN, R. DE CHARETTE.*Influence of Fog on Computer Vision Algorithms*, Inria Paris, September 2017, p. 1-3, https://hal.inria.fr/hal-01620602.

References in notes

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] F. GARRIDO, D. GONZALEZ BAUTISTA, V. MILANÉS, J. PÉREZ, F. NASHASHIBI.*Optimized trajectory planning for Cybernetic Transportation Systems*, in "9th IFAC Symposium on Intelligent Autonomous Vehicles IAV 2016", Leipzig, Germany, June 2016, vol. 49, n^{o} 15, p. 1-6, https://hal.inria.fr/hal-01356691.

[47] F. GARRIDO, D. GONZALEZ BAUTISTA, V. MILANÉS, J. PÉREZ, F. NASHASHIBI.*Real-time planning for adjacent consecutive intersections*, in "19th International IEEE Conference on Intelligent Transportation Systems - ITSC 2016", Rio de Janeiro, Brazil, November 2016, https://hal.inria.fr/hal-01356706.

[48] 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.

[49] 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.

[50] 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.

[51] A. MEYER, R. DE CHARETTE.*Computing Ego Velocity from Scene Flow estimation*, Inria Paris, December 2016, https://hal.inria.fr/hal-01620608.

[52] F. NAVAS, V. MILANÉS, F. NASHASHIBI.*Using Plug&Play Control for stable ACC-CACC system transitions*, in "Intelligent Vehicles Symposium 2016", Gothemburg, Sweden, June 2016, https://hal.inria.fr/hal-01304542.

[53] L. RAFFAELLI, G. FAYOLLE, F. VALLÉE.*ADAS Reliability and Safety*, in "20ème Congrès de maîtrise des risques et de sûreté de fonctionnement ", Saint-Malo, France, E. LARDEUX, A. BRACQUEMOND (editors), Congrès Lambda MU 20, Institut pour la Maîtrise des Risques (IMdR), October 2016, 10, https://hal.inria.fr/hal-01398428.

[54] L. RAFFAELLI, F. VALLÉE, G. FAYOLLE, P. DE SOUZA, X. ROUAH, M. PFEIFFER, S. GÉRONIMI, F. PÉTROT, S. AHIAD.*Facing ADAS validation complexity with usage oriented testing*, in "ERTS 2016", Toulouse, France, January 2016, 13, https://hal.inria.fr/hal-01277494.

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
B9.4.1. - Computer science
B9.4.2. - Mathematics
B9.8. - Privacy

1. Personnel

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]
Anthony Leverrier [Inria, Researcher, HDR]
María Naya Plasencia [Inria, 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, from Sep 2017, en délégation]

PhD Students

Xavier Bonnetain [Univ Pierre et Marie Curie, AMX]
Rémi Bricout [Univ Pierre et Marie Curie, from Sep 2017, AMN]
Rodolfo Canto Torres [Inria]
Kevin Carrier [Ministère de la Défense]
Kaushik Chakraborty [Inria, until Oct 2017]

Julia Chaulet [Thales, until Mar 2017]
Thomas Debris [Univ Pierre et Marie Curie]
Sébastien Duval [Univ Pierre et Marie Curie]
Shouvik Ghorai [Univ Pierre et Marie Curie, from Oct 2017]
Antoine Grospellier [Univ Pierre et Marie Curie, AMN]
Adrien Hauteville [Univ de Limoges, until Sep 2017]
Matthieu Lequesne [Univ Pierre et Marie Curie, from Sep 2017, AMX]
Vivien Londe [Univ de Bordeaux, AMX]
Andrea Olivo [Inria, from Nov 2017]
Yann Rotella [Inria]
Ferdinand Sibleyras [Inria, from Oct 2017, DGA-Inria]
Valentin Vasseur [Univ René Descartes, from Oct 2017]

Post-Doctoral Fellow

Leo Perrin [Inria, from Sep 2017, Fondation Sciences Mathématiques de Paris]

Visiting Scientists

Christof Beierle [Univ. Bochum, Germany, from Apr 2017 until Jun 2017]
Özgül Küçük [Istanbul Bilgi Univ., Turkey, from Jul 2017 until Aug 2017, bourse SSHN]
Thomas Peyrin [NTU, Singapore, May 2017 and July 2017]

Administrative Assistants

Laurence Bourcier [Inria]
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.

[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/

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:

(i) 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;

(ii) 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

5.1.1. NIST post-quantum cryptography standardisation

The end of this year was the deadline to submit proposals to the NIST competition [0], 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.

We have contributed to three proposals to the NIST call. In two of them, "BIKE" [67] and "Big Quake" [69], our action is central and we also have a marginal participation in another, "Classic McEliece". Those projects are of great importance for us because they are a means to demonstrate our long lasting expertise in code-based cryptography. They are the product of numerous research works, including several PhD theses, on the design, the implementation, and the cryptanalysis of code-based cryptographic primitives. There are 69 projects in that call, which will be evaluated by the NIST and the academic cryptographic community in the next three to five years and whose outcome will certainly influence cryptographic applications for one or several decades.

5.1.2. Quantum symmetric cryptanalysis and collision search

The resistance of symmetric primitives to quantum computers is a topic that has received recently a lot of attention from our community. The ERC starting grant QUASYModo on this subject, awarded to M. Naya-Plasencia, has started in September 2017. We have continued the work started last year obtaining new results, as cryptanalysis of concrete proposals [44], or analysis on attacks considering modular additions (preliminary described in [14]). In particular, we have proposed in [47] a new quantum algorithm for finding collisions. This new algorithm, based on BHT, exploits distinguished points as well as an improved optimization of the parameters, and allows to find for the first time, collisions on n bits with a better time complexity than $2^{n/2}$. Its time and query complexity are of about $2^{2n/5}$, needing $2^{n/5}$ classical memory and a polynomial amount of quantum memory. As collision search is a tool widely used in symmetric cryptanalysis, this algorithm, that also can be applied to multiple preimage search, considerably improves the best known previous attacks when having a relatively small quantum computer available.

5.1.3. Émergences grant on quantum money

André Chailloux was awarded an Émergences grant from the city of Paris for a project on quantum money. This project aims at providing a comprehensive theoretical and experimental study of unforgeable quantum money, one of the most powerful protocols in quantum information science, and historically the first. A quantum money scheme enables a secure transaction between a client, a vendor and a bank via the use of a credit card or banknotes, with maximal security guarantees, unreachable with classical technologies. This application is central in the context of the emerging quantum network infrastructures guaranteeing the long-term security of data and communications against all-powerful adversaries.

Quantum money has been largely considered difficult to bring to the experimental realm, but a demonstration became more accessible recently, thanks to the conception of new practical schemes. The goal of our project will be to perform a theoretical analysis of such schemes, both in the discrete and continuous-variable frameworks, to adapt them to realistic conditions, and to implement them using state-of-the-art photonic quantum technologies. The project, centered around Inria, is interdisciplinary at its core, bringing together young partners with world leading expertise in all aspects of the proposed work, including theoretical and experimental quantum cryptography.

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

[0] https://csrc.nist.gov/projects/post-quantum-cryptography/post-quantum-cryptography-standardization

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, Sébastien Duval, Gaëtan Leurent, María Naya Plasencia, Yann Rotella, Ferdinand Sibleyras, Tim Beyne, Mathilde de La Morinerie, André Schrottenloher.

7.1.1. Primitives: block ciphers, stream ciphers, ...

Our recent results mainly concern either the analysis and design of lightweight block ciphers.
Recent results:

- Analysis of linear invariant attacks [41], [54], [28], [29]: C. Beierle, A. Canteaut, G. Leander and Y. Rotella have studied SPN ciphers with a very simple key schedule, such as PRINCE. They introduce properties of the linear layer and of the round constants than can be used to prove that there are no nonlinear invariants.
- Analysis of the probability of differential characteristics for unkeyed constructions [19]: This work shows that the probabilities of some fixed-key differential characteristics are higher than expected when assuming independent S-Boxes. This leads to improved attacks against ROADRUNNER and Minalpher.
- Design and study of a new construction for low-latency block ciphers, named *reflection ciphers*, which generalizes the so-called α-reflection property exploited in PRINCE. This construction aims at reducing the implementation overhead of decryption on top of encryption [15].
- Modular construction of primitives with code-hardness, time-hardness or memory-hardness [42]. A. Biryukov and L. Perrin have introduced new definitions to formalize hardness, and constructions that are hard to compute for common users, but easy for users knowing a secret.
- 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].

7.1.2. 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:

- Boolean functions with restricted input: Y. Rotella, together with C. Carlet and P. Méaux, has introduced some new criteria on filtering Boolean functions, which measure the security of the recent stream cipher proposal FLIP. Indeed, in this context, the inputs of the filtering function are not uniformly distributed but have a fixed Hamming weight. Then, the main properties of filtering functions (e.g. nonlinearity, algebraic immunity...) have been revisited [20].
- 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 [45]. 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.
- A. Canteaut, S. Duval and L. Perrin proposed a construction of a new family of permutations over binary fields of dimension $(4k+2)$ with good cryptographic properties. An interesting property is that this family includes as a specific case the only known APN permutation of an even number of variables [55], [18].
- Construction of cryptographic permutations over finite fields with a sparse representation: P. Charpin, together with N. Cepak and E. Pasalic, exhibited permutations which are derived from sparse functions via linear translators [21].
- New methods for determining the differential spectrum of an Sbox: P. Charpin and G. Kyureghyan have proved that the whole differential spectrum of an Sbox can be determined without examining all derivatives of the mapping, but only the derivatives with respect to an element within a hyperplane [23]. Also, they have proved that, for mappings of a special shape, it is enough to consider the derivatives with respect to all elements within a suitable multiplicative subgroup of $\mathbb{F}_{2^n}$.

7.1.3. Side-channel attacks

Physical attacks must be taken into account in the evaluation of the security of lightweight primitives. Indeed, these primitives are often dedicated to IoT devices in pervasive environments, where an attacker has an easy access to the devices where the primitive is implemented.

Recent results:

- Differential fault attack against LS-designs and SCREAM [52]: this attack generalized previous work on PRIDE to the class of LS-Designs.

7.1.4. 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 security, and we now that their use is secure as long as the underlying primitive are secure, and we respect some limits on the amount of data processed. The analysis of generic attack helps us understand what happens when the hypothesis of the security proofs do not hold, or the corresponding limits are not respected. Comparing proofs and attack also shows gaps where our analysis is incomplete, and improved proof or attacks are required.

Recent results:

- Use of block ciphers operating on small blocks with the CBC mode [31]: it is well-known that CBC is not secure if the same key is used for encrypting $2^{n/2}$ blocks of plaintext, but this threat has traditionally been dismissed as impractical, even for 64-bit blocks. K. Bhargavan and G. Leurent demonstrated concrete attacks that exploit such short block ciphers in CBC mode.
- Use of block ciphers operating on small blocks with the CTR mode [77]: the security proof of the CTR mode also 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 even less problematic than on CBC. During his internship with G. Leurent, F. Sibleyras has studied concrete attacks against the CTR mode when processing close to $2^{n/2}$ blocks of data, and has shown that an attacker can actually extract as much information as in the case of CBC encryption.
- Improved generic attacks against hash-based MAC [25].
- Modes of operation for full disk encryption [51]: L. Khati, N. Mouha and D. Vergnaud have classified various FDE modes of operation according to their security in a setting where there is no space to store additional data, like an IV or a MAC value. They also introduce the notion of a diversifier, which does not require additional storage, but allows the plaintext of a particular sector to be encrypted into different ciphertexts.

7.2. Code-based cryptography

Participants: Rodolfo Canto Torres, Julia Chaulet, André Chailloux, Thomas Debris, Adrien Hauteville, Nicolas Sendrier, Jean-Pierre Tillich, Matthieu Lequesne, Valentin Vasseur, Matthieu Vieira.

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.

As mentioned in Section 5.1.1, the NIST is currently running a standardization effort for quantum-safe cryptography, where code based cryptography is a promising technique.

Our work in this area can be decomposed as follows:

- suggesting code-based solutions to the NIST competition;
- cryptanalyzing code-based schemes;
- fundamental work on code-based cryptography.

7.2.1. Code-based solutions to the NIST competition

We have proposed two key-exchange protocols to the NIST competition:

- the first one [67] is based on quasi-cyclic MDPC codes and the work [40];
- the second one [69] is based on quasi-cyclic Goppa codes.

Both of them are able to reduce significantly the keysizes by relying on quasi-cyclic codes.

7.2.2. Cryptanalysis of code-based cryptography

Here our work can be summarized as follows:

- cryptanalysis of McEliece schemes based on wild Goppa codes over quadratic extension fields [24];
- improving generic attacks on rank metric codes [68];
- side-channel attacks on quasi-cyclic MDPC bit flipping decoder [74].

7.2.3. Fundamental work on code-based cryptography

- studying precisely the complexity of statistical decoding techniques [71], [48];
- suggesting the first code-based identity-based encryption by using rank metric codes [49];
- suggesting a code-based signature scheme [43];
- analysing and improving the decoding of quasi-cyclic MDPC codes [12], [78];
- studying families of codes that might be used in a cryptographic setting [53].
- improving the complexity of quantum decoding algorithms [50];
- studying [70], [56], [30] whether security reductions for signature schemes are quantum safe when considering the quantum random oracle model (QROM). We were particularly interested in code-based Full Domain Hash constructions. We show that if the underlying correcting code we use has good pseudo random properties then it is possible to perform a quantum security reduction in the QROM.

7.3. Quantum Information

Participants: Xavier Bonnetain, Rémi Bricout, Kaushik Chakraborty, André Chailloux, Shouvik Ghorai, Antoine Grospellier, Anirudh Krishna, Gaëtan Leurent, Anthony Leverrier, Vivien Londe, María Naya Plasencia, Andrea Olivo, Jean-Pierre Tillich, Sristy Agrawal, 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.

Recent results:

- Decoding algorithm for quantum expander codes [72], [57], [58], [59], [73], [35]. In this work, A. Grospellier, A. Leverrier and O. Fawzi analyze an efficient decoding algorithm for quantum expander codes and prove that it suppresses errors exponentially in the local stochastic noise model. 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.
- Construction of quantum LDPC codes from regular tessellations of hyperbolic 4-space [64], [62]. In this work, V. Londe proposes a variant of a construction of Guth and Lubotzky that yields a family of constant rate codes with a polynomial minimum distance. The main interest of this construction is that is is based on a regular tessellation of hyperbolic 4-space by hypercubes. This nice local structure is exploited to design and analyze an efficient decoding algorithm that corrects arbitrary errors of weight logarithmic in the code length.

- Construction of quantum codes based on the real projective space [63]. In this work, V. Londe studies a family of almost LDPC codes with a large minimum distance and another efficient decoding algorithm.
- We were also awarded a European Quantera project "QCDA" to investigate and develop better quantum error-correcting codes and schemes for fault-tolerance.

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. Another primitive is quantum money and was in fact the first proposed idea of quantum cryptography in the 70s. However, this primitive hasn't received much attention because its implementation requires quantum memories, which weren't available until now.

Recent results:

- Full security proof for BB84 [27]. In this work A. Leverrier, with M. Tomamichel, give a detailed and self-contained security proof for BB84, the most studied quantum key distribution protocol. Many simplified proofs appear in the literature, but are usually incomplete and fail to address the whole protocol.
- Security proof of continuous-variable quantum key distribution [26], [36], [37]. In this work, A. Leverrier establishes for the first time a security reduction from general attacks to a class of simple attacks called "collective Gaussian" attacks. This result exploits in a crucial way a recent Gaussian de Finetti theorem that applies to quantum systems of infinite dimension [75], [61], [34].
- In [22], A. Chailloux and I. Kerenidis present an extended version on results for optimal quantum bit commitment and coin flipping. Those results show what is the best way to quantumly perform those protocols in the information-theoretic setting. In the extended version, we also show that the bound for quantum bit commitment cannot be achieved classically, even with an access to an ideal coin flipping primitive.
- We were also awarded an ANR project quBIC and an "Émergence" project from Ville de Paris to study quantum money schemes in collaboration with UPMC, LKB and IRIF.

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 zero-knowledge: In [46], A. Chailloux and A. Leverrier construct 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. While this technique is applied to the relativistic setting, it also has implications for more standard quantum cryptography.
- In [16], R. Bricout and A. Chailloux study relativistic multi-round bit commitment schemes. They show optimal classical cheating strategies for the canonical F_Q commitment scheme. This shows that the security proof derived last year on the relativistic F_Q commitment scheme is essentially optimal against classical adversaries.

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, that has started on september 2017.

Recent results:

- In a result published in Asiacrypt 2017 [47] and done during the internship of André Schrottenloher [76] a new quantum algorithm for finding collisions is proposed. The algorithm is based on BHT and exploits distinguished points as well as an improved optimization of the parameters, and allows to find, for the first time, collisions on n bits with a better time complexity than $2^{n/2}$ while needing a polynomial amount of quantum memory.
- Two of the most popular symmetric cryptanalysis families are differential and linear cryptanalysis. In [60] (also presented in [33]), G. Leurent, M. Kaplan, A. Leverrier and M. Naya-Plasencia have proposed efficient ways of quantizing these attacks in different models, obtaining some non-intuitive results: just quantizing the best classical attack does not always provide the best quantum attack.
- X. Bonnetain and M. Naya-Plasencia have obtained some new results, preliminarily described in [14] and presented at [38], that consider the tweak proposed at Eurocrypt this year of using modular additions to counter Simon's attacks. They have studied the best attacks on these constructions, that use Kuperberg's algorithm. They have also simulated the cost of such attacks, improved the algorithm, applied this to a widely-used construction and to some slide attacks, and finally dimensionated the symmetric construction in order to stay secure to these attacks. They have concluded that the proposed tweak does not seam realistic.
- In [44], an attack on the superposition model of the CAESAR cadidate AEZ is proposed, showing that this construction would be completely broken in that scenario.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

- **Thales** ($02/14 \rightarrow 01/17$)
  Funding for the supervision of Julia Chaulet's PhD.
  30 kEuros.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. 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.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. 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.

9.2.1.2. 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.

9.2.1.3. 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.

9.2.2. Collaborations in European Programs, Except FP7 & H2020

9.2.2.1. COST Action IC1306

Program: COST

Project acronym: ICT COST Action IC1306

Project title: Cryptography for Secure Digital Interaction

Duration: January 2014 - November 2017

Coordinator: Claudio Orlandi, Aarhus University, Denmark

Other partners: see http://www.cost.eu/domains_actions/ict/Actions/IC1306

Abstract: 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.

Anne Canteaut is co-leader of the working group on cryptographic primitives. She co-organized a 2-day workshop for PhD students and early-career researchers in symmetric cryptography, DISC 2016 (Bochum, Germany, March 23-24 2016) and a winter school dedicated to Symmetric Cryptography and Blockchain (Torremolinos, Spain, February 19-23, 2018). She also serves on the program committee of the CryptoAction Symposium organized every year.

9.2.2.2. 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)

Abstract: 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.

9.3. International Initiatives

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

9.3.1.1. 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, a real pair of colliding messages was only published recently by a team from CWI and Google, because the estimated attack complexity is around 2^{63} SHA-1 computations (this represents about 70000 years of computation on a normal PC).

While this SHA-1 collision clearly demonstrates 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.

9.3.2. Inria International Partners

9.3.2.1. 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.

9.3.2.2. Informal International Partners

- Nanyang Technological University (Singapore): cryptanalysis of symmetric primitives.
- Ruhr-Universität Bochum (Germany): design and cryptanalysis of symmetric primitives.

9.3.3. Participation in Other International Programs

Anirudh Krishna, PhD student at Sherbroke University (Canada) spends six months in our team within the MITACS program.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Giannicola Scarpa, Universidad Complutense de Madrid, Spain, April 2017.
- Thomas Peyrin, NTU Singapore, May 2017, July 2017 and January 2018.
- Kaisa Nyberg, University of Helsinki, Finlande, May 2017.
- Adi Shamir, The Weizmann Institute of Science, Rehovot, Israel, May 2017.
- Christof Beierle, Bochum University, Germany, visiting PhD student, April-June 2017.
- Özgül Küçük, Bilgi University, Turkey, July-August 2017 (Bourse SSHN du Gouvernement Français).

9.4.1.1. Internships

- Sristy Agrawal, Kolkata, India, June-Aug. 2017
- Tim Beyne, Univ. Leuven, Belgium, Aug.-Sept. 2017
- Mathilde De La Morinerie, École Polytechnique, April-July 2017

- Matthieu Lequesne, MPRI, March-Aug. 2017
- André Schrottenloher, MPRI and Telecom ParisTech, March-Aug. 2017
- Ferdinand Sibleyras, MPRI, March-Aug. 2017
- Valentin Vasseur, Univ. Grenoble, March-Aug. 2017
- Matthieu Vieira, ENS Lyon, May-July 2017

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

- NTU, Singapore, October 16 - November 3, joint work within the CHOCOLAT Associate Team (G. Leurent).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- EuroS&P 2017: April 26-28, 2015, Paris (France): G. Leurent (poster chair);
- TQC 2017 (Theory of Quantum Computation, Communication and Cryptography): June 20-22, 2017, Paris (France): A. Chailloux, A. Leverrier.
- Dagstuhl Seminar 17401, "Quantum Cryptanalysis": October 1-6, 2017, Dagstuhl (Germany): N. Sendrier (co-organizer)
- Training School on Symmetric Cryptography and Blockchain: February 19-23, 2018, Torremolinos (Spain): A. Canteaut (co-organizer).

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

As a co-editor-in-chief of the journal *IACR Transactions on Symmetric Cryptology*, María Naya-Plasencia serves as a program chair of the conference *Fast Software Encryption (FSE)*, hold in Tokyo March 2017, and to be held in Bruges March 2018.

10.1.2.2. Member of the Conference Program Committees

- QIP 2017: January 16-20 2017, Seattle, USA (A. Chailloux, A. Leverrier);
- CT-RSA 2017: February 14-17, 2017, San Francisco, US (M. Naya-Plasencia);
- FSE 2017: March 5-8, 2017, Tokyo, Japan (A. Canteaut, G. Leurent, M. Naya-Plasencia);
- CryptoAction Symposium 2017: March 27-28, Amsterdam, the Netherlands (A. Canteaut);
- Financial Crypto 2017: April 3-7, 2017, Sliema, Malta (G. Leurent);
- Journées Codage et Cryptographie - C2 2017: April 23-28, La Bresse, France (G. Leurent);
- Eurocrypt 2017: 30 April- 4 May, 2017, Paris, France (M. Naya-Plasencia);
- Fq13: June 4-9, 2017, Gaeta, Italy (A. Canteaut);
- CEWQO 2017: June 26-30 2017, Lyngby, Denmark (A. Leverrier);
- PQCrypto 2017: 26-28 June, 2017, Utrecht, the Netherlands (M. Naya-Plasencia, N. Sendrier, J.P. Tillich)
- SAC 2017: August 16-18, 2017, Ottawa, Canada (G. Leurent, M. Naya-Plasencia);
- Crypto 2017: August 20-24, 2017, Santa Barbara, CA, USA (G. Leurent);
- AQIS 2017: September 4-8, 2017, Singapore (A. Chailloux);

- SCN 2018: September 5-7, 2018, Amalfi, Italy (G. Leurent);
- QCrypt 2017: 2017, September 18-22 2017, Cambridge, UK (A. Leverrier);
- WCC 2017: September 18-22, Saint-Petersburg, Russia (P. Charpin, J.-P. Tillich);
- 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, (N. Sendrier, J.P. Tillich);

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- *Designs, Codes and Cryptography*, associate editor: P. Charpin.
- *Finite Fields and Applications*, associate editors: 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, L. Perrin, co-editor-in-chief: M. Naya-Plasencia.
- *Annals of telecommunications*, associate editor: J.P. Tillich.
- *Advances in Mathematics for Communications*, associated editor : J.P. Tillich

10.1.3.2. Editor for books or special issues

- Special Issue on Coding and Cryptography, *Designs, Codes and Cryptography* : P. Charpin, T. Johansson, G. Kyureghyan, N. Sendrier and J.-P. Tillich, Eds., Volume 82, Issue 1-2, January 2017

10.1.3.3. Reviewer - Reviewing Activities

- Reviewer for Mathematical Reviews: P. Charpin.
- Reviewer for ERC proposals: G. Leurent

10.1.4. Invited Talks

- A. Leverrier, *Security of continuous-variable quantum key distribution via a Gaussian de Finetti reduction*, Trustworthy Quantum Information TyQi 2017, Paris, France, 19-21 June 2017.
- A. Leverrier, *Challenges in continuous-variable quantum cryptography*, QCRYPT 2017, Cambridge, UK, 18-22 September 2017.
- N. Sendrier, *Quantum Safe Cryptography from Codes: Present and Future*, 16th IMA International Conference on Cryptography and Coding, Oxford, UK, December 13, 2017.

The members of the project-team have also been invited to give talks to some workshops or international seminars, including:

- A. Canteaut, *Proving Resistance against Invariant Attacks: Properties of the Linear Layer* , Early Symmetric Crypto - ESC 2017, Canach, Luxembourg, January 2017
- A. Canteaut, *Proving resistance of a block cipher against invariant attacks*, BFA 2017 - Boolean Functions and their Applications, Os, Norway, July 2017.
- A. Chailloux, *A tight security reduction in the quantum random oracle model for code-based signature schemes*, IRIF Algocomp seminar, Paris, France, November 2017
- G. Leurent, *On the Practical (In-)Security of 64-bit Block Ciphers*, Early Symmetric Crypto - ESC 2017, Canach, Luxembourg, January 2017
- G. Leurent, *Breaking Symmetric Cryptosystems Using Quantum Algorithms*, Frontiers of Quantum Safe Cryptography - FOQUS, April 2017, Paris, France.
- G. Leurent, *Bad Symmetric Crypto in the Real World*, Journées Nationales 2017 Pré-GDR Sécurité Informatique, Paris, France, May 2017.

- A. Leverrier, *A Gaussian de Finetti theorem and application to truncations of random Haar matrices*, Workshop on "Probabilistic techniques and Quantum Information Theory", IHP, Paris, France, 23-27 October 2017.
- A. Leverrier, *Efficient decoding of random errors for quantum expander codes*, Conference on "Quantum Information Theory", IHP, Paris, France, 11-15 December 2017.
- M. Naya-Plasencia, *New Results on Quantum Symmetric Cryptanalysis* Dagstuhl seminar "Quantum Cryptanalysis", Dagstuhl, Germany, October 2017.
- N. Sendrier,*Code-based Cryptography*, PQCRYPTO Summer School on Post-Quantum Cryptography 2017, TU Eindhoven, June 2017. 5 hours.
- N. Sendrier,*Code-based Cryptography*, Executive School on Post-Quantum Cryptography 2017, TU Eindhoven, June 2017. 1 1/2 hours.
- J.P. Tillich *Décodage de codes LDPC quantiques*, Journées C2 La Bresse, April 27, 2017.
- J.P. Tillich *Code based cryptography and quantum attacks*, Dagstuhl seminar "Quantum Cryptanalysis", Dagstuhl, Germany, October 2017.
- J.P. Tillich *Recent advances in decoding quantum LDPC codes*, Recent advances in Quantum Computing, CEA Saclay, December 13, 2017.

10.1.5. 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)*.
- N. Sendrier serves on the steering committee of the WCC conference series.
- N. Sendrier is a member of the “Comité de pilotage” of the ANR (défi 9).
- A. Leverrier serves on the steering committee of *DIM SIRTEQ* (réseau francilien pour les technologies quantiques).

10.1.6. Research Administration

- A. Canteaut serves as Head of Science of the Inria Paris research center since September 2017. She was deputy Head of Science from January to August 2017.
- A. Canteaut serves on the *Evaluation Committee* since September 2017.
- A. Canteaut was a member of the steering committee of the Fondation Sciences Mathématiques de Paris until June 2017.
- P. Charpin serves on the *Comité Parité* at Inria.
- M. Naya-Plasencia is a member of *Inria Paris CES 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 on the jury for PhD scholarships from EDITE.
- M. Naya-Plasencia serves on the *Comité des usagers du projet "rue Barrault"*.
- M. Naya-Plasencia serves on the *commission bureaux*.

10.1.7. Committees for the selection of professors, assistant professors and researchers

- Inria Paris Chargés de recherche: M. Naya-Plasencia
- Inria Directeurs de recherche: A. Canteaut
- Université Pierre-et-Marie-Curie, professor: A. Canteaut

[0] https://competitions.cr.yp.to/caesar.html

- Université de Rouen, assistant professor: C. Boura, M. Naya-Plasencia
- Université de Limoges, assistant professor: C. Boura, M. Naya-Plasencia
- ENSEA, assistant Professor: M. Naya-Plasencia
- DTU Denmark, associate professor: A. Canteaut.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master: A. Canteaut, *Error-correcting codes and applications to cryptology*, 12 hours, M2, University Paris-Diderot (MPRI), France;

Master: A. Canteaut, *Symmetric crypography*, 6 hours M2, Ecole des Mines de Saint-Etienne, campus de Gardanne (ingénieurs Spécialité Microélectronique et Informatique), France, 2017.

Master: A. Chailloux, *Quantum Computing*, 9 hours, M2, University Paris-Diderot (MPRI), France;

Master: G. Leurent *Algorithmique et programmation*, 25 hours, M1, UVSQ, France;

Corps des Mines: G. Leurent *Cryptographie symétrique*, 9 hours, Telecom ParisTech, France;

Master: J.-P. Tillich, *Introduction to Information Theory*, 32 hours, M2, Ecole Polytechnique, France;

10.2.2. Supervision

HdR : María Naya Plasencia, *Symmetric Cryptography for Long-Term Security*, University Pierre-et-Marie-Curie, May 5, 2017.

HdR : Anthony Leverrier, *Protecting information in a quantum world: from cryptography to error correction*, University Pierre-et-Marie-Curie, September 27, 2017.

PhD : Julia Chaulet, *Study of public-key cryptosystems based on MDPC quasi-cyclic codes*, University Pierre-et-Marie-Curie, March 20, 2017.

PhD : Kaushik Chakraborty, *Cryptography with spacetime constraints*, Université Pierre-et-Marie Curie, October 12, 2017.

PhD : Adrien Hauteville, *Nouveaux protocoles et nouvelles attaques pour la cryptologie basée sur les codes en métrique rang* , University of Limoges, December 4, 2017.

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: Sébastien Duval, *Constructions for lightweight cryptography*, since October 2015, supervisor: A. Canteaut and G. Leurent

PhD in progress: Yann Rotella, *Finite fields and symmetric cryptography*, since October 2015, supervisor: A. Canteaut

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).

10.2.3. Juries

- C. Mavromati, *Cryptanalyse des algorithmes de type Even-Mansour*, University Paris-Saclay, January 24, 2017, committee: A. Canteaut (reviewer).
- J. Chaulet, *Study of public-key cryptosystems based on MDPC quasi-cyclic codes*, University Pierre-et-Marie-Curie, March 20, 2017, committee: N. Sendrier (supervisor), JP Tillich.
- R. do Canto de Loura, *Quantum measures, noise and measurement errors in a quantum bit commitment protocol*, Universidade de Lisboa, Instituto Superior Tecnico, March 31, 2017, committee: A. Leverrier (reviewer).
- M. Naya-Plasencia, *Symmetric Cryptography for Long-Term Security*, Habilitation, University Pierre-et-Marie-Curie, May 5, 2017, committee: A. Canteaut.
- H. Kalachi, *Sécurité de Protocoles Cryptographiques Fondés sur la Théorie des Codes Correcteurs d'Erreurs*, July 5, 2017, University of Rouen, committee: J.P. Tillich;
- B. Dravie, *Synchronisation et systèmes dynamiques, application à la cryptographie*, University of Lorraine, July 6, 2017, committee: A. Canteaut (reviewer).
- V. Dragoi, *Approche algébrique pour l'étude et la résolution de problèmes algorithmiques issus de la cryptographie et de la théorie des codes*, July 6, 2017, University of Rouen, committee: N. Sendrier (reviewer), J.P. Tillich;
- V. Migliore, *Cybersécurité matérielle et conception de composants dédiés au calcul homomorphe*, Université de Bretagne Sud, September 26, 2017. committee: N. Sendrier (reviewer);
- A. Leverrier, *Protecting information in a quantum world: from cryptography to error correction*, Habilitation, University Pierre-et-Marie-Curie, September 27, 2017, committee: J.P. Tillich;
- A. Bannier, *Combinatorial analysis of block ciphers with trapdoor*, Arts et Métiers ParisTech, September 29, 2017, committee: A. Canteaut;
- K. Chakraborty, *Cryptography with spacetime constraints*, Université Pierre-et-Marie Curie, October 12, 2017, committee: A. Leverrier (supervisor), J.P. Tillich (supervisor);
- A. Hauteville, *Nouveaux protocoles et nouvelles attaques pour la cryptologie basée sur les codes en métrique rang* , University of Limoges, December 4, 2017, committee: N. Sendrier, J.P. Tillich (supervisor);
- D. Mirandola, *On products of linear error correcting codes*, Leiden University, the Netherlands, and Univ. de Bordeaux, December 6, 2017, committee: A. Canteaut (reviewer, chair).
- G. Spini, *Unconditionally secure cryptographic protocols from coding-theoretic primitives*, Leiden University, the Netherlands, and Univ. de Bordeaux, December 6, 2017, committee: A. Canteaut (chair).

- P. Méaux, *Chiffrement complètement homomorphe hybride*, Research University PSL, December 8, 2017, committee: A. Canteaut (reviewer);
- M. Saad Taha. *Algebraic Approach for Code Equivalence*, Université de Rouen, December 18, 2017. committee: N. Sendrier (reviewer), J.P. Tillich;

10.3. Popularization

- **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/. Matthieu Lequesne organized the challenge and created the scientific content of the competition. He also gave a talk during the final of the cipher challege Alkindi on May 17 at the "Cité des Sciences" in Paris. The 2018 edition of the competition has been launched in December 2017 at Lycée de la Vallée de Chevreuse, Gif-sur-Yvette. Matthieu Lequesne, Sébastien Duval and Yann Rotella gave talks on cryptography during the opening ceremony. The best teams from Académies de Dijon and Orléans-Tours have been visiting the SECRET project-team in June 2017 https://www.youtube.com/watch?v=EVLHEOWAORc.
- N. Sendrier, *Code-Based Cryptography: State of the Art and Perspectives*, IEEE Security & Privacy, Special Issue on Post-quantum Cryptography. July/August 2017.
- A. Chailloux *Cryptographie Quantique en théorie* - Journée Maths en Mouvement sur l'ordinateur quantique organized by the FSMP, Paris, France, May 2017
- Matthieu Lequesne co-organized the final of the French Tournament of Young Mathematicians at École polytechnique on May 26-28 and was chaired the jury sessions. He also participated to the elaboration of the problems for the 8th French Tournament of Young Mathematicians (TFJM^2) in December 2017.
- Matthieu Lequesne co-organized the International Tournament of Young Mathematicians (ITYM) in Iasi, Romania in July 2017 and was part of the international jury.
- Matthieu Lequesne taught for one week during a mathematical summer camp for high school students in Bethlehem, Palestine, organized by the Al Khwarizmi Noether Institute in August 2017.
- Matthieu Lequesne co-organized a weekend for female high-school students interested in mathematics (Rendez-vous des Jeunes Mathématiciennes) at ENS Ulm, November 25-26.
- Yann Rotella gave a talk on cryptography at Lycée Théophile Gautier, Tarbes, January 31, 2017.
- Yann Rotella gave a presentation for *Raconte-moi ta thèse !* during Fete de la Science, at IHP, Paris, October 2017.
- Several members of the team (C. Boura, A. Canteaut, M. Lequesne, A. Leverrier, Y. Rotella) have been involved in the *Cinquante ans d'Inria*, November 2017. They hold a stand to present a serious game on cryptography. A. Canteaut has participated on a panel on Cyber-security. A. Leverrier gave a short talk (pitch de science) on quantum computing.
- Matthieu Lequesne was auditioned by the committee in charge of proposing a reform of mathematical education (Mission Maths Villani-Torossian) on November 29.

11. Bibliography

Major publications by the team in recent years

[1] 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.

[2] A. CANTEAUT, B. CHEVALLIER-MAMES, A. GOUGET, P. PAILLIER, T. PORNIN, E. BRESSON, C. CLAVIER, T. FUHR, T. ICART, J.-F. MISARSKY, M. NAYA-PLASENCIA, J.-R. REINHARD, C. THUILLET, M. VIDEAU.*Shabal, a Submission to NIST's Cryptographic Hash Algorithm Competition*, October 2008, Submission to NIST.

[3] A. CANTEAUT, M. NAYA-PLASENCIA, B. VAYSSIÈRE.*Sieve-in-the-Middle: Improved MITM Attacks*, in "Advances in Cryptology - CRYPTO 2013, Part I", Lecture Notes in Computer Science, Springer, 2013, vol. 8042, p. 222–240.

[4] 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.

[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] J.-C. FAUGÈRE, A. OTMANI, L. PERRET, J.-P. TILLICH.*Algebraic Cryptanalysis of McEliece Variants with Compact Keys*, in "Advances in Cryptology - EUROCRYPT 2010", LNCS, Springer, 2010, n^{o} 6110, p. 279-298, http://dx.doi.org/10.1007/978-3-642-13190-5_14.

[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] K. CHAKRABORTY.*Cryptography with Spacetime Constraints*, Université Pierre et Marie Curie - Paris VI, October 2017, https://hal.inria.fr/tel-01637818.

[12] J. CHAULET.*Study of public key cryptosystems based on quasi-cyclic MDPC codes*, Université Pierre et Marie Curie - Paris VI, March 2017, https://tel.archives-ouvertes.fr/tel-01599347.

[13] A. LEVERRIER.*Protecting information in a quantum world: from cryptography to error correction*, Université Pierre et Marie Curie - Paris VI, September 2017, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01636624.

[14] M. NAYA-PLASENCIA.*Symmetric Cryptography for Long-Term Security*, Université Pierre et Marie Curie - Paris VI, May 2017, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01656036.

Articles in International Peer-Reviewed Journal

[15] C. BOURA, A. CANTEAUT, L. R. KNUDSEN, G. LEANDER.*Reflection ciphers*, in "Designs, Codes and Cryptography", January 2017, vol. 82, n° 1–2, p. 3–25 [*DOI :* 10.1007/s10623-015-0143-x], https://hal.inria.fr/hal-01237135.

[16] R. BRICOUT, A. CHAILLOUX.*Recursive cheating strategies for the relativistic $\mathbb{F}_Q$ bit commitment protocol*, in "MDPI - Cryptography", August 2017, https://arxiv.org/abs/1608.03820 [*DOI :* 10.3390/CRYPTOGRAPHY1020014], https://hal.inria.fr/hal-01409563.

[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", 2017, https://hal.inria.fr/hal-01650012.

[18] A. CANTEAUT, S. DUVAL, L. PERRIN.*A generalisation of Dillon's APN permutation with the best known differential and nonlinear properties for all fields of size 2^{4k+2}*, in "IEEE Transactions on Information Theory", 2017, vol. 63, n° 11, p. 7575–7591 [*DOI :* 10.1109/TIT.2017.2676807], https://hal.inria.fr/hal-01589131.

[19] A. CANTEAUT, E. LAMBOOIJ, S. NEVES, S. RASOOLZADEH, Y. SASAKI, M. STEVENS.*Refined Probability of Differential Characteristics Including Dependency Between Multiple Rounds*, in "IACR Transactions on Symmetric Cryptology", May 2017, vol. 2017, n° 2, p. 203–227 [*DOI :* 10.13154/TOSC.V2017.I2.203-227], https://hal.inria.fr/hal-01649954.

[20] C. CARLET, P. MÉAUX, Y. ROTELLA.*Boolean functions with restricted input and their robustness; application to the FLIP cipher*, in "IACR Transactions on Symmetric Cryptology", 2017, vol. 2017, n° 3, p. 192–227 [*DOI :* 10.13154/TOSC.V2017.I3.192-227], https://hal.inria.fr/hal-01633506.

[21] N. CEPAK, P. CHARPIN, E. PASALIC.*Permutations via linear translators*, in "Finite Fields and Their Applications", 2017, vol. 45, p. 19–42, https://arxiv.org/abs/1609.09291 [*DOI :* 10.1016/J.FFA.2016.11.009], https://hal.inria.fr/hal-01412487.

[22] A. CHAILLOUX, I. KERENIDIS.*Physical Limitations of Quantum Cryptographic Primitives or Optimal Bounds for Quantum Coin Flipping and Bit Commitment*, in "SIAM Journal on Computing", January 2017, vol. 46, n° 5, p. 1647–1677 [*DOI :* 10.1137/15M1010853], https://hal.inria.fr/hal-01650970.

[23] P. CHARPIN, G. M. KYUREGHYAN.*On sets determining the differential spectrum of mappings*, in "International journal of information and Coding Theory", 2017, vol. 4, n° 2/3, p. 170–184, Special issue on the honor of Gerard Cohen [*DOI :* 10.1504/IJICOT.2017.083844], https://hal.inria.fr/hal-01406589.

[24] 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.

[25] I. DINUR, G. LEURENT.*Improved Generic Attacks Against Hash-Based MACs and HAIFA*, in "Algorithmica", December 2017, vol. 79, n^{o} 4, p. 1161–1195 [*DOI :* 10.1007/s00453-016-0236-6], https://hal.inria.fr/hal-01407953.

[26] A. LEVERRIER.*Security of Continuous-Variable Quantum Key Distribution via a Gaussian de Finetti Reduction*, in "Physical Review Letters", May 2017, vol. 118, n^{o} 20, p. 1–24, https://arxiv.org/abs/1701.03393 [*DOI :* 10.1103/PHYSREVLETT.118.200501], https://hal.inria.fr/hal-01652082.

[27] M. TOMAMICHEL, A. LEVERRIER.*A largely self-contained and complete security proof for quantum key distribution*, in "Quantum", 2017, vol. 1, 14, https://arxiv.org/abs/1506.08458 [*DOI :* 10.22331/Q-2017-07-14-14], https://hal.inria.fr/hal-01237240.

Invited Conferences

[28] C. BEIERLE, A. CANTEAUT, G. LEANDER, Y. ROTELLA.*Proving Resistance against Invariant Attacks: Properties of the Linear Layer* , in "ESC 2017 - Early Symmetric Crypto", Canach, Luxembourg, January 2017, https://hal.inria.fr/hal-01649994.

[29] C. BEIERLE, A. CANTEAUT, G. LEANDER, Y. ROTELLA.*Proving resistance of a block cipher against invariant attacks*, in "BFA 2017 - Boolean Functions and their Applications", Os, Norway, July 2017, https://hal.inria.fr/hal-01649990.

[30] A. CHAILLOUX.*A tight security reduction in the quantum random oracle model for code-based signature schemes*, in "2017 - IRIF Algocomp seminar", Paris, France, November 2017, https://hal.inria.fr/hal-01660701.

[31] G. LEURENT, K. BHARGAVAN.*On the Practical (In-)Security of 64-bit Block Ciphers*, in "ESC 2017 - Early Symmetric Crypto", Canach, Luxembourg, January 2017, https://hal.inria.fr/hal-01105128.

[32] G. LEURENT.*Bad Symmetric Crypto in the Real World*, in "Journées Nationales 2017 Pré-GDR Sécurité Informatique", Paris, France, May 2017, https://hal.inria.fr/hal-01652853.

[33] G. LEURENT.*Breaking Symmetric Cryptosystems Using Quantum Algorithms*, in "FOQUS - Frontiers of Quantum Safe Cryptography", Paris, France, April 2017, https://hal.inria.fr/hal-01652852.

[34] A. LEVERRIER.*A Gaussian de Finetti theorem and application to truncations of random Haar matrices*, in "Workshop on ”Probabilistic techniques and Quantum Information Theory””, Paris, France, October 2017, p. 1-60, https://hal.inria.fr/hal-01656425.

[35] A. LEVERRIER.*Efficient decoding of random errors for quantum expander codes*, in "Conference on ”Quantum Information Theory””, Paris, France, December 2017, p. 1-33, https://hal.inria.fr/hal-01656427.

[36] A. LEVERRIER.*Security of continuous-variable quantum key distribution via a Gaussian de Finetti reduction*, in "TyQi 2017 - Trustworthy Quantum Information", Paris, France, June 2017, https://hal.inria.fr/hal-01656418.

[37] A. LEVERRIER.*Theoretical challenges in continuous-variable quantum cryptography*, in "QCrypt 2017 - 7th International Conference on Quantum Cryptography", Cambridge, United Kingdom, September 2017, p. 1-26, https://hal.inria.fr/hal-01656419.

[38] M. NAYA-PLASENCIA.*New results on symmetric quantum cryptanalysis*, in "Dagstuhl Seminar 17401 - Quantum Cryptanalysis", Dagstuhl, Germany, October 2017, https://hal.inria.fr/hal-01671913.

[39] J.-P. TILLICH.*Code based cryptography and quantum attacks*, in "Dagstuhl Seminar 17401 - Quantum cryptanalysis", Dagstuhl, Germany, October 2017, https://hal.archives-ouvertes.fr/hal-01671921.

International Conferences with Proceedings

[40] P. S. L. M. BARRETO, S. GUERON, T. GUNEYSU, R. MISOCZKI, E. PERSICHETTI, N. SENDRIER, J.-P. TILLICH.*CAKE: Code-based Algorithm for Key Encapsulation*, in "IMACC 2017 - 16th IMA International Conference on Cryptography and Coding", Oxford, United Kingdom, M. O'NEILL (editor), LNCS - Lecture Notes in Computer Science, Springer, December 2017, vol. 10655, p. 207–226 [*DOI :* 10.1007/978-3-319-71045-7_11], https://hal.inria.fr/hal-01661949.

[41] 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.

[42] A. BIRYUKOV, L. PERRIN.*Symmetrically and Asymmetrically Hard 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. 10626, p. 417–445 [*DOI :* 10.1007/978-3-319-70700-6_15], https://hal.inria.fr/hal-01650044.

[43] O. BLAZY, P. GABORIT, J. SCHREK, N. SENDRIER.*A code-based blind signature*, in "ISIT 2017 - IEEE International Symposium on Information Theory", Aachen, Germany, IEEE, June 2017, p. 2718–2722 [*DOI :* 10.1109/ISIT.2017.8007023], https://hal.archives-ouvertes.fr/hal-01610410.

[44] X. BONNETAIN.*Quantum Key-Recovery on full AEZ*, in "SAC 2017 - Selected Areas in Cryptography", Ottawa, Canada, August 2017, https://hal.inria.fr/hal-01650026.

[45] C. BOURA, A. CANTEAUT, J. JEAN, V. SUDER.*Two Notions of Differential Equivalence on Sboxes*, in "WCC 2017 - Workshop on Coding and Cryptography", Saint Petersburg, Russia, September 2017, https://hal.inria.fr/hal-01650010.

[46] A. CHAILLOUX, A. LEVERRIER.*Relativistic (or 2-Prover 1-Round) Zero-Knowledge Protocol for NP Secure Against Quantum Adversaries*, in "Eurocrypt 2017 - Advances in Cryptology", Paris, France, J.-S. CORON, J. B. NIELSEN (editors), LNCS - Lecture Notes in Computer Science, Springer, April 2017, vol. 10212, p. 369–396 [*DOI :* 10.1007/978-3-319-56617-7_13], https://hal.inria.fr/hal-01650985.

[47] 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.

[48] T. DEBRIS-ALAZARD, J.-P. TILLICH.*Statistical Decoding*, in "ISIT 2017 - IEEE International Symposium on Information Theory", Aachen, Germany, IEEE, June 2017, p. 1789–1802 [*DOI :* 10.1109/ISIT.2017.8006839], https://hal.inria.fr/hal-01661749.

[49] P. GABORIT, A. HAUTEVILLE, D. H. PHAN, J.-P. TILLICH.*Identity-based Encryption from Codes with Rank Metric*, 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. 10403, p. 194–224 [*DOI :* 10.1007/978-3-319-63697-9_7], https://hal.inria.fr/hal-01589463.

[50] G. KACHIGAR, J.-P. TILLICH.*Quantum Information Set Decoding Algorithms*, in "PQCrypto 2017 - The Eighth International Conference on Post-Quantum Cryptography", Utrecht, Netherlands, T. LANGE, T. TAKAGI (editors), LNCS - Lecture Notes in Computer Science, Springer, June 2017, vol. 10346, p. 69-89 [*DOI :* 10.1007/978-3-319-59879-6_5], https://hal.inria.fr/hal-01661905.

[51] L. KHATI, N. MOUHA, D. VERGNAUD.*Full Disk Encryption: Bridging Theory and Practice*, in "CT-RSA 2017 - RSA Conference Cryptographers' Track", San Francisco, United States, H. HANDSCHUH (editor), Lecture Notes in Computer Science, Springer, February 2017, vol. 10159, p. 241–257 [*DOI :* 10.1007/978-3-319-52153-4_14], https://hal.inria.fr/hal-01403418.

[52] B. LAC, A. CANTEAUT, J. J. A. FOURNIER, R. SIRDEY.*DFA on LS-Designs with a Practical Implementation on SCREAM*, in "COSADE 2017 - Constructive Side-Channel Analysis and Secure Design", Paris, France, S. GUILLEY (editor), LNCS - Lecture Notes in Computer Science, Springer, April 2017, vol. 10348, p. 223–247 [*DOI :* 10.1007/978-3-319-64647-3_14], https://hal.inria.fr/hal-01649974.

[53] I. MARQUEZ-CORBELLA, J.-P. TILLICH.*Attaining Capacity with iterated $(U|U+V)$ codes based on AG codes and Koetter-Vardy soft decoding*, in "ISIT 2017 - IEEE International Symposium on Information Theory", Aachen, Germany, IEEE, June 2017, p. 6–10 [*DOI :* 10.1109/ISIT.2017.8006479], https://hal.inria.fr/hal-01661977.

Conferences without Proceedings

[54] C. BEIERLE, A. CANTEAUT, G. LEANDER, Y. ROTELLA.*Attaques par invariant : comment s'en protéger?*, in "Journées codage et cryptographie 2017", La Bresse, France, April 2017, 1, https://hal.inria.fr/hal-01633519.

[55] A. CANTEAUT, S. DUVAL, L. PERRIN.*On a generalisation of Dillon's APN permutation*, in "Fq13 - Finite Fields and Applications", Gaeta, Italy, June 2017, https://hal.inria.fr/hal-01650001.

[56] A. CHAILLOUX.*A tight security reduction in the quantum random oracle model for code-based signature schemes*, in "Code based crypto seminar", Paris, France, October 2017, p. 1-22, https://hal.inria.fr/hal-01660693.

[57] 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.

[58] A. GROSPELLIER, A. LEVERRIER, O. FAWZI.*Efficient decoding of random errors for quantum expander codes*, in "Journées Informatique Quantique 2017", Bordeaux, France, November 2017, https://hal.archives-ouvertes.fr/hal-01671491.

[59] A. GROSPELLIER, A. LEVERRIER, O. FAWZI.*Quantum expander codes*, in "Journées codage et cryptographie 2017", La Bresse, France, April 2017, https://hal.archives-ouvertes.fr/hal-01671485.

[60] G. LEURENT, M. KAPLAN, A. LEVERRIER, M. NAYA-PLASENCIA.*Quantum differential and linear cryptanalysis*, in "FSE 2017 - Fast Software Encryption", Tokyo, Japan, March 2017, https://hal.inria.fr/hal-01652807.

[61] A. LEVERRIER.*SU(p,q) coherent states and Gaussian de Finetti theorems*, in "QIP 2017 - 20th Annual Conference on Quantum Information Processing", Seattle, United States, January 2017, p. 1-24, https://hal.inria.fr/hal-01656414.

[62] V. LONDE.*Golden codes: 4D hyperbolic regular quantum codes*, in "8th colloquium of the GDR IQFA - Ingénierie Quantique, des Aspects Fondamentaux aux Applications", Nice, France, November 2017, https://hal.inria.fr/hal-01671528.

[63] V. LONDE.*Homological quantum error correcting codes and real projective space*, in "Journées Codage et Cryptographie 2017", La Bresse, France, April 2017, https://hal.inria.fr/hal-01671444.

[64] V. LONDE.*4D hyperbolic regular quantum codes*, in "Journées Informatique Quantique 2017", Bordeaux, France, November 2017, https://hal.inria.fr/hal-01671456.

[65] N. SENDRIER.*Quantum Safe Cryptography from Codes: Present and Future*, in "16th IMA International Conference on Cryptography and Coding", Oxford, United Kingdom, December 2017, https://hal.archives-ouvertes.fr/hal-01671452.

Scientific Popularization

[66] A. CHAILLOUX.*Cryptographie Quantique en théorie*, in "2017 - 9ème Journée Mathématiques en Mouvement sur l'ordinateur quantique", Paris, France, FSMP, May 2017, https://hal.inria.fr/hal-01660726.

Other Publications

[67] N. ARAGON, P. S. L. M. BARRETO, S. BETTAIEB, L. BIDOUX, O. BLAZY, J.-C. DENEUVILLE, P. GABORIT, S. GUERON, T. GUNEYSU, C. AGUILAR MELCHOR, R. MISOCZKI, E. PERSICHETTI, N. SENDRIER, J.-P. TILLICH, G. ZÉMOR.*BIKE: Bit Flipping Key Encapsulation*, December 2017, Submission to the NIST post quantum standardization process, https://hal.archives-ouvertes.fr/hal-01671903.

[68] N. ARAGON, P. GABORIT, A. HAUTEVILLE, J.-P. TILLICH.*Improvement of Generic Attacks on the Rank Syndrome Decoding Problem*, October 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01618464.

[69] M. BARDET, E. BARELLI, O. BLAZY, R. CANTO TORRES, A. COUVREUR, P. GABORIT, A. OTMANI, N. SENDRIER, J.-P. TILLICH.*BIG QUAKE BInary Goppa QUAsi–cyclic Key Encapsulation*, December 2017, submission to the NIST post quantum cryptography standardization process, https://hal.archives-ouvertes.fr/hal-01671866.

[70] A. CHAILLOUX, T. DEBRIS-ALAZARD.*A tight security reduction in the quantum random oracle model for code-based signature schemes* , December 2017, working paper or preprint, https://hal.inria.fr/hal-01671870.

[71] T. DEBRIS-ALAZARD, J.-P. TILLICH.*Statistical Decoding*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01661745.

[72] O. FAWZI, A. GROSPELLIER, A. LEVERRIER.*Efficient decoding of random errors for quantum expander codes*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01671348.

[73] O. FAWZI, A. GROSPELLIER, A. LEVERRIER.*Efficient decoding of random errors for quantum expander codes*, November 2017, 8th colloquium of the GDR IQFA - Ingénierie Quantique, des Aspects Fondamentaux aux Applications, Poster, https://hal.archives-ouvertes.fr/hal-01671496.

[74] M. LEQUESNE.*Side Channel Key Recovery Attacks on QC-MDPC Codes*, MPRI, September 2017, p. 1-22, https://hal.inria.fr/hal-01658381.

[75] A. LEVERRIER.$SU(p,q)$ *coherent states and a Gaussian de Finetti theorem*, November 2017, working paper or preprint, https://hal.inria.fr/hal-01652084.

[76] A. SCHROTTENLOHER.*Collision search and quantum symmetric cryptanalysis*, Université Paris-Saclay, September 2017, p. 1-25, https://hal.inria.fr/hal-01654190.

[77] F. SIBLEYRAS.*Cryptanalysis of the Counter mode of operation*, Paris 7, September 2017, https://hal.inria.fr/hal-01662040.

[78] V. VASSEUR.*Cryptographie post-quantique : étude du décodage des codes QC-MDPC*, Université Grenoble-Alpes, September 2017, https://hal.inria.fr/hal-01664082.

[79] M. DE LA MORINERIE.*Implémentation à seuil de boîtes S*, Ecole Polytechnique, July 2017, https://hal.inria.fr/hal-01672270.

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:

A2.1.2. - Object-oriented programming
A2.1.3. - 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. Personnel

Research Scientists

Martin Vohralík [Team leader, Inria, Senior Researcher, HDR]
François Clément [Inria, Researcher]
Alexandre Ern [Ecole Nationale des Ponts et Chaussées, Professor, HDR]
Michel Kern [Inria, Researcher]
Laurent Monasse [Ecole Nationale des Ponts et Chaussées, Associate Professor, until August 2017]
Géraldine Pichot [Inria, Researcher]
Iain Smears [Inria, Starting Research Position, until August 2017]
Pierre Weis [Inria, Senior Researcher]

External Collaborators

Hend Ben Ameur [IPEST and ENIT-Lamsin (Tunisia), Professor, HDR]
Erik Burman [University College London, Professor, HDR]
Guy Chavent [Univ Paris-Dauphine, Professor (retired), HDR]
Jean-Luc Guermond [Texas A&M University, Professor, HDR]
Jérôme Jaffré [Inria, Senior Researcher (retired), HDR]
Caroline Japhet [Univ Paris-Nord, Associate Professor]

Vincent Martin [Univ de technologie de Compiègne, Associate Professor]
Jean-Elizabeth Roberts [Inria, Senior Researcher (retired), HDR]

Technical Staff

Sébastien Furic [Inria, from June 2017]

PhD Students

Sarah Ali Hassan [Inria, until June 2017]
Amina Benaceur [Univ Paris-Est, CIFRE EDF]
Karol Cascavita [Univ Paris-Est, granted by Labex MMCD]
Jad Dabaghi [Inria, CORDI-C]
Patrik Daniel [Inria, CORDI-C granted by ERC GATIPOR]
Frédéric Marazzato [Univ Paris-Est, contract CEA]
Riccardo Milani [Univ Paris-Est, CIFRE EDF, from October 2017]
Ani Miraci [Inria, CORDI-C granted by ERC GATIPOR, from October 2017]
Nicolas Pignet [Univ Paris-Est, CIFRE EDF]
Rita Riedlbeck [University Montpellier 2, CIFRE EDF, until November 2017]

Post-Doctoral Fellows

Elyes Ahmed [Univ Paris-Nord, until April 2017]
Thomas Boiveau [Univ Paris-Est, until September 2017]
Matteo Cicuttin [Univ Paris-Est]
Seyed Mohammad Zakerzadeh [Inria, from October 2017]

Administrative Assistants

Virginie Collette [Inria, until June 2017, and from December 2017]
Azzeddine Saidani [Inria, from June to August 2017]
Kévin Bonny [Inria, from August to December 2017]

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). Our requirement is to derive *structure-preserving* methods, i.e., methods that mimic 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* for the lowest-order schemes for the Navier–Stokes equations and to the recently-devised theory of gradient schemes for discrete element methods applied to elasto-plasticity. For the schemes we develop, we study existence, uniqueness, and stability questions, and derive a priori convergence estimates. Our special interest is in higher-order methods like the hybrid high-order method, which have recently begun to receive significant attention. Even though their use in practice may not be immediate, we believe that they belong to the future generation of numerical methods for industrial simulations.

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 shall then address theoretically the question of convergence of the new algorithms and prove their numerical quasi-optimality, meaning that they need, up to a generic constant, the smallest possible number of degrees of freedom to achieve the given accuracy. We in particular seek to prove a guaranteed error reduction in terms of the number of degrees of freedom.

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
- geological sequestration of CO2
- production of oil and gas

4.2. Complex Stokes and Navier–Stokes flows

- industrial risks in energy production (fission)

4.3. Energy production, sustainable use of resources

- simulation of shock waves impinging on deformable or fragmentable structures
- use of nets of rods for sustainable construction

4.4. 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

The most important results of the ERC GATIPOR are now centralized in the ERC GATIPOR Gallery.

5.1.1. Awards

Laurent Monasse was awarded an ANR JCJC (young researcher) grant.

6. New Software and Platforms

6.1. GEOFRAC

GEOFRACFLOW
KEYWORDS: Hydrogeology - Numerical simulations - 3D

SCIENTIFIC DESCRIPTION: GEOFRACFLOW 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 GEOFRACFLOW 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.

RELEASE FUNCTIONAL DESCRIPTION: The last version includes optimisations of the code, mainly with an efficient upload of the mesh data generated with BLSURF and vectorization of the operations.

- 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

6.2. Ref-indic

Refinement indicators
KEYWORD: Inverse problem
SCIENTIFIC DESCRIPTION: The refinement indicator algorithm is suited for the estimation of a distributed parameter occurring in a mathematical simulation model, typically a set of partial differential equations. When the numerical simulation model must be solved on a fine grid, the refinement indicator algorithm provides an adaptive parameterization of the sought parameter that avoids overparameterization difficulties. In each grid cell, the estimated parameter may be of dimension greater than one, i.e. the algorithm is able to estimate several scalar distributed parameters.

Ref-indic implements a generic version of the refinement indicator algorithm that can dock specific programs provided they conform to the generic algorithm API.

The API of Ref-indic requires four main functionalities (called tasks) for the user specific program, it must be able: * to initialize, i.e. to open all necessary data files, to perform all necessary preliminary computation, and to return an initial coarse parameterization (giving a zone number between 0 and the initial number of zones minus one for each cell of the fine grid), * to compute the gradient on the fine grid for a given fine parameterization, * to optimize the problem for a given coarse parameterization, * and to finalize, i.e. to store the resulting coarse parameterization.

Given any such user specific program, the inversion platform automatically provides a program that solves the corresponding user inverse problem using the refinement indicator algorithm.
FUNCTIONAL DESCRIPTION: Ref-indic is an adaptive parameterization platform using refinement indicators. Slogan is “invert details only where they are worth it”. Ref-indic implements a generic version of the refinement indicator algorithm that can dock specific programs provided they conform to the generic algorithm API.
NEWS OF THE YEAR: In its current implementation, the inversion platform can only build coarse parameterizations for a distributed parameter defined on a fine rectangular grid. From version 1.5+pl0, the user has the possibility to specify masked cells in the fine rectangular grid that will be ignored by the algorithm (with the use of the specific zone number -1 in the initial coarse parameterization). This allows for the treatment of inverse problems defined on unstructured meshes. The handling of both-way interpolations must be taken care of by the gradient computation and optimization tasks. The masked cells must be the same for all components of the parameter.

- Contact: François Clément
- Publications: Image Segmentation with Multidimensional Refinement Indicators - The Multi-Dimensional Refinement Indicators Algorithm for Optimal Parameterization
- URL: http://refinement.inria.fr/ref-indic/

6.3. Sklml

The OCaml parallel skeleton system
KEYWORDS: Parallel programming - Functional programming
SCIENTIFIC DESCRIPTION: Writing parallel programs is not easy, and debugging them is usually a nightmare. To cope with these difficulties, the skeleton programming approach uses a set of predefined patterns for parallel computations. The skeletons are higher order functional templates that describe the program underlying parallelism.

Sklml is a new framework for parallel programming that embeds an innovative compositional skeleton algebra into the OCaml language. Thanks to its skeleton algebra, Sklml provides two evaluation regimes to programs: a regular sequential evaluation (merely used for prototyping and debugging) and a parallel evaluation obtained via a recompilation of the same source program in parallel mode.

Sklml was specifically designed to prove that the sequential and parallel evaluation regimes coincide.
FUNCTIONAL DESCRIPTION: Sklml is a functional parallel skeleton compiler and programming system for OCaml programs. Slogan is “easy coarse grain parallelization”.
NEWS OF THE YEAR: Caml preprocessors are no longer needed.

- Participants: François Clément and Pierre Weis
- Contact: François Clément
- URL: http://sklml.inria.fr

6.4. GENFIELD

KEYWORDS: Hydrogeology - Algorithm - Heterogeneity
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.

RELEASE FUNCTIONAL DESCRIPTION: The new version includes: - The use of the FFTW3-mpi library for discrete Fourier transform - Non regression tests (and continuous integration through gitlab.inria.fr) - Debugging of the parallel algorithm

- Participants: Géraldine Pichot, Simon Legrand, Grégoire Lecourt, Jean-Raynald De Dreuzy and Jocelyne Erhel
- Contact: Géraldine Pichot
- Publication: Algorithms for Gaussian random field generation
- URL: https://gitlab.inria.fr/slegrand/Genfield_dev

6.5. 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/datafl4sh/diskpp

6.6. 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/

6.7. 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/

7. New Results

7.1. A posteriori stopping criteria for domain decomposition methods

Participants: Sarah Ali Hassan, Michel Kern, Martin Vohralík.

Publication: [45]

In [45] we propose a new method for stopping iterations in a domain decomposition (DD) algorithm. The approach is based on a posteriori error estimates, and builds estimators that distinguish between the (space and time) discretization errors and that caused by the DD iterations. This enables stopping the iterations as soon as the DD error is smaller than the discretization error. In practice, numerous unnecessary iterations can be avoided, as illustrated in Figure 1 (here we stop at iteration 17 in place of the usual 61, economizing 72 % iterations). The method has been extended to global-in-time domain decomposition and to nonlinear problems. This was the topic of the Ph.D. thesis of Sarah Ali Hassan.

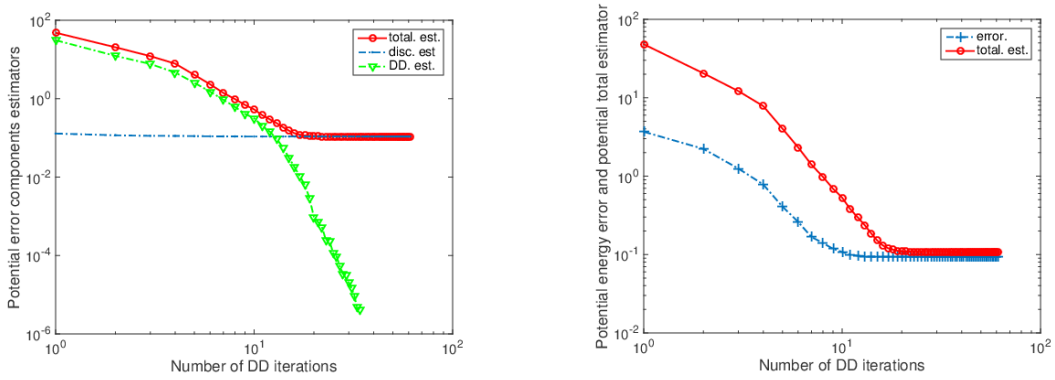


Figure 1. Error component estimates (left) and total energy error and its estimate (right), DD with GMRES solver

7.2. Finite element quasi-interpolation and best-approximation

Participant: Alexandre Ern.

Publication: [21]

In [21], we introduce a quasi-interpolation operator for scalar- and vector-valued finite element spaces constructed on affine, shape-regular meshes with some continuity across mesh interfaces. This operator gives optimal estimates of the best approximation error in any L^p-norm assuming regularity in the fractional Sobolev spaces $W^{r,p}$, where $p \in [1, \infty]$ and the smoothness index r can be arbitrarily close to zero. The operator is stable in L^1, leaves the corresponding finite element space point-wise invariant, and can be modified to handle homogeneous boundary conditions. The theory is illustrated on H^1-, $\mathbb{H}(\mathrm{curl})$-, and $\mathbb{H}(\mathrm{div})$-conforming spaces.

7.3. Hybrid High-Order methods for hyperelasticity

Participants: Alexandre Ern, Nicolas Pignet.

Publication: [13]

In [13], we devise and evaluate numerically Hybrid High-Order (HHO) methods for hyperelastic materials undergoing finite deformations. The HHO methods use as discrete unknowns piecewise polynomials of order $k \geq 1$ on the mesh skeleton, together with cell-based polynomials that can be eliminated locally by static condensation. The discrete problem is written as the minimization of a broken nonlinear elastic energy where a local reconstruction of the displacement gradient is used. Two HHO methods are considered: a stabilized method where the gradient is reconstructed as a tensor-valued polynomial of order k and a stabilization is added to the discrete energy functional, and an unstabilized method which reconstructs a stable higher-order gradient and circumvents the need for stabilization. Both methods satisfy the principle of virtual work locally with equilibrated tractions. We present a numerical study of the two HHO methods on test cases with known solution and on more challenging three-dimensional test cases including finite deformations with strong shear layers and cavitating voids. We assess the computational efficiency of both methods, and we compare our results to those obtained with an industrial software using conforming finite elements and to results from the literature. The two HHO methods exhibit robust behavior in the quasi-incompressible regime. In Figure 2, we present some results for a hollow cylinder under shear and compression.

7.4. A nonlinear consistent penalty method for positivity preservation

Participant: Alexandre Ern.

Publication: [16]

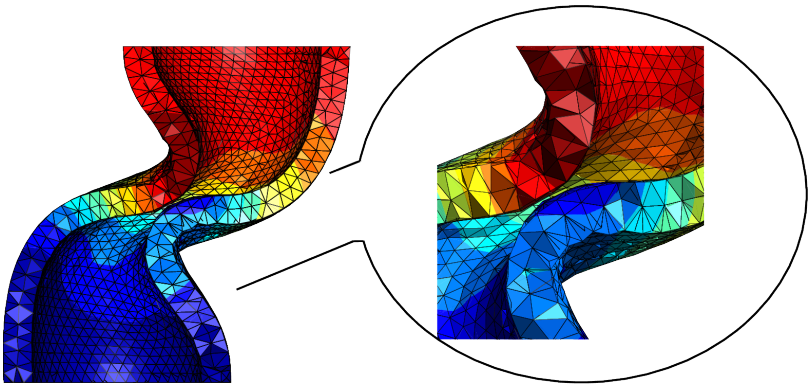


Figure 2. Euclidean displacement norm on the deformed configuration for the shear and compressed cylinder, and a zoom where the deformations are the most important. The color scale goes from 0.0 *(blue) to* 1.8 *(red).*

In [16], we devise and analyze a new stabilized finite element method to solve the first-order transport (or advection-reaction) equation. The method combines the usual Galerkin/Least-Squares approach to achieve stability with a nonlinear consistent penalty term inspired by recent discretizations of contact problems to weakly enforce a positivity condition on the discrete solution. We prove the existence and the uniqueness of the discrete solution. Then we establish quasioptimal error estimates for smooth solutions bounding the usual error terms in the Galerkin/Least-Squares error analysis together with the violation of the maximum principle by the discrete solution. A numerical example is presented in Figure 3.

Figure 3. Elevations of solutions using piecewise quadratic elements. Left: standard method, the nodal discrete maximum principle violation is 21%*. Right: consistent penalty method, violation is less than* $4 \cdot 10^{-3}\%$*.*

7.5. A simple a posteriori estimate on general polytopal meshes

Participant: Martin Vohralík.

Publication: [30]

The recent publication [30] develops an a posteriori error estimate for lowest-order locally conservative methods on meshes consisting of general polytopal elements. We focus here on the ease of implementation and evaluation cost of the methodology based on H^1-conforming potential reconstructions and $\mathbf{H}(\mathrm{div})$-conforming flux reconstructions that we develop in the SERENA project-team. In particular, the evaluation of our estimates for steady linear diffusion equations merely consists in some local matrix-vector multiplications, where, on each mesh element, the matrices are either directly inherited from the given numerical method, or easily constructed from the element geometry, while the vectors are the flux and potential values on the given element. This is probably the smallest computational price that one can imagine. We next extend our approach to steady nonlinear problems. We obtain a guaranteed upper bound on the total error in the fluxes that is still obtained by local matrix-vector multiplications, with the same element matrices as above. Moreover, the estimate holds true on any linearization and algebraic solver step and allows to distinguish the different error components. Finally, we apply this methodology to unsteady nonlinear coupled degenerate problems describing complex multiphase flows in porous media. It leads to an easy-to-implement and fast-to-run adaptive algorithm with guaranteed overall precision, adaptive stopping criteria, and adaptive space and time mesh refinements. An example of its application to a complex porous media flow (three-phases/three-components black-oil problem) can be found in Figure 4.

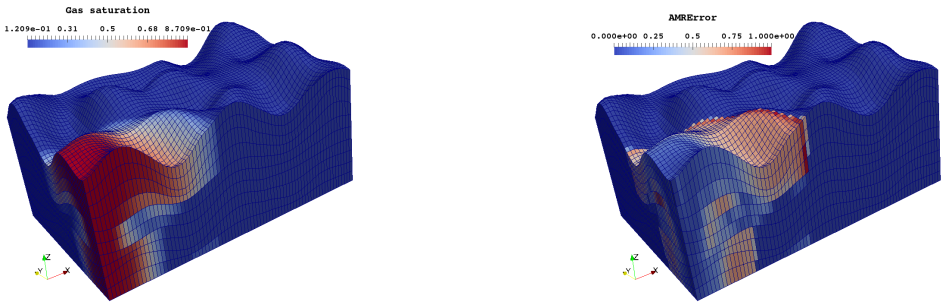


Figure 4. Simulated gas saturation after 1000 days (left) and corresponding a posteriori error estimate (right)

7.6. Sharp algebraic and total a posteriori error bounds

Participant: Martin Vohralík.

Publication: [66]

In [66], we derive guaranteed, fully computable, constant-free, and sharp upper and lower a posteriori estimates on the algebraic, total, and discretization errors of finite element approximations of the Poisson equation obtained by an arbitrary iterative solver. Though guaranteed bounds on the discretization error, when the associated algebraic system is solved exactly, are now well-known and available, this is definitely not the case for the error from the linear algebraic solver (algebraic error), and a beautiful problem arises when these two error components interact. We try to analyze it here while identifying a decomposition of the algebraic error over a hierarchy of meshes, with a global residual solve on the coarsest mesh. Mathematically, we prove equivalence of our computable total estimate with the unknown total error, up to a generic polynomial-degree-independent constant. Numerical experiments illustrate sharp control of all error components and accurate prediction of their spatial distribution in several test problems, as we illustrate it in Figure 5 for the higher-order conforming finite element method and the conjugate gradient algebraic solver.

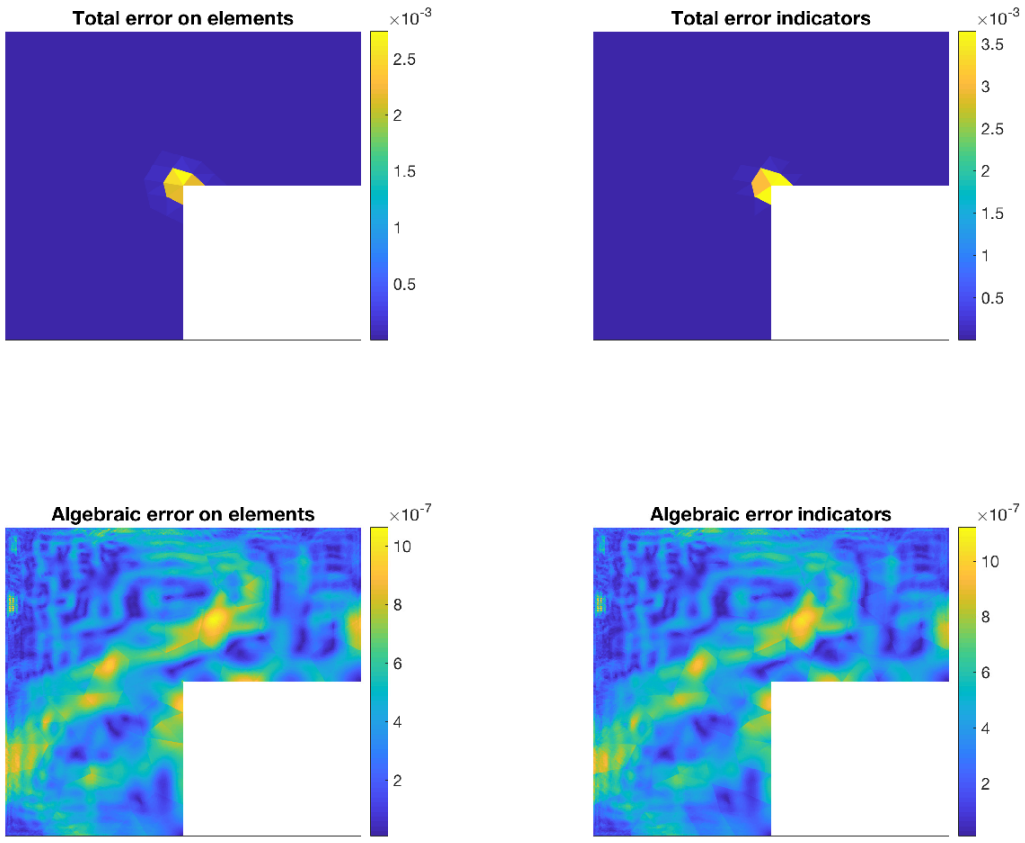


Figure 5. Actual total error (top left) and its a posteriori error estimate (top right). Actual algebraic error (bottom left) and its a posteriori error estimate (bottom right).

7.7. Analytic expressions of the solutions of advection-diffusion problems in 1D with discontinuous coefficients

Participant: Géraldine Pichot.

Publication: [64]

Grants: H2MN04 3

In [64], we provide a general methodology to compute the resolvent kernel as well as the density when available for a one-dimensional second-order differential operators with discontinuous coefficients. In a sequel, the computed resolvent kernel will be used to set-up an efficient and accurate simulation scheme.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

Three contracts with EDF accompanying the PhD theses of Amina Benaceur, Nicolas Pignet, and Riccardo Milani.

One contract with CEA accompanying the PhD thesis of Frédéric Marazzato.

One contract with ANDRA accompanying the PhD thesis of Sarah Ali Hassan (ended, Ph.D. defended in June 2017).

One contract with IFP Energies Nouvelles, in the framework of the Inria–IFP Energies Nouvelles "contrat cadre".

Three-parts contract Inria–EDF–Sciworks Technologies (from April 2017) on "Form-L for the formalization of constraints of complex systems".

9. Partnerships and Cooperations

9.1. Regional Initiatives

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 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 project had its intermediate review in December 2016, and received excellent marks from the expert panel.

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 2014–2017.

ANR GEOPOR: "Geometrical approach for porous media flows: theory and numerics". A new approach to numerical methods for multiphase simulations based on the concept of gradient flows is investigated. With Laboratoire Jacques-Louis Lions, University Pierre and Marie Curie. SERENA representant is M. Vohralík, period 2013–2017.

ANR H2MNO4: "Original optimized object-oriented numerical model for heterogeneous hydrogeology". The project H2MNO4 develops numerical models for reactive transport in heterogeneous media. The objective is to design both Eulerian and Lagrangian models. Three applications are concerned: freshwater supply, remediation of mine drainage, and waste geological disposal. The project relies on a consortium of six partners, involving four public research laboratories (Inria, Geosciences Rennes, University of Lyon 1, University of Poitiers, Pprime Institute), one public institution (ANDRA), and one enterprise (ITASCA). International collaborations are pursued with University of San Diego (USA) and UPC (Spain). SERENA representant is G. Pichot, period 2012–2016.

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.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

ERC GATIPOR: "Guaranteed fully adaptive algorithms with tailored inexact solvers for complex porous media flows". The subject of this project 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.

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.

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.

MoRe

Program: Research, Development and Innovation Council of the Czech Republic

Project acronym: MoRe

Project title: Implicitly constituted material models: from theory through model reduction to efficient numerical methods

Duration: September 2012 – September 2017

Coordinator: Josef MÁLEK, Charles University in Prague. SERENA representant is M. Vohralík.

Other partners: Institute of Mathematics, Czech Academy of Sciences; University of Oxford

Abstract: A multidisciplinary project on nonlinear Navier–Stokes flows with implicit constitutive laws. It focuses on development of accurate, efficient, and robust numerical methods for simulations of the new class of implicit models.

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 medial applications.

Barbara Wohlmuth, Professor at Technical University of München, Germany, mixed finite element methods.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Lars Diening, Professor at University of Bielefeld, Germany, February 17–23, 2017.

Christian Kreuzer, Professor at University of Dortmund, Germany, February 19–25, 2017.

Joscha Gedicke, post-doc at University Vienna, Austria, May 29–June 2, 2017.

Martin Eigel, post-doc at Weierstrass Institute Berlin, Germany, May 29–June 2, 2017.

Carsten Carstensen, Professor at Humboldt University Berlin, Germany, August 15–September 15, 2017.

Peter Minev, Professor at the University of Alberta, Canada, September 15–October 15, 2017.

Hend Ben Ameur, Professor at IPEST and member of ENIT-Lamsin, Tunis, Tunisia, October 23–November 3, 2017.

9.5.1.1. Internships

K. Talali, université de Fez, Morocco, April 1–August 31 (Master degree).

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

Alexandre Ern participated as Invited Professor to the HIM Program on Multiscale Problems: Algorithms, Numerical Analysis and Computation, in Bonn, Germany, January 2017.

Martin Vohralík was invited for two weeks stay to Charles University in Prague collaboration with J. Málek, April 2017.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Alexandre Ern, Iain Smears, and Martin Vohralík have organized a 2-day workshop on **A posteriori error estimates, adaptivity, and advanced applications**, in the framework of the ERC GATIPOR project. 40 participants from the whole world.

Soleiman Yousef 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

Géraldine Pichot and Michel Kern are members of the local organizing committee of the Computational Methods in Water Resources 2018 conference.

Martin Vohralík was a member of the organizing committee of the *Finite Volumes for Complex Applications* conference.

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

Alexandre Ern and Martin Vohralík were members of the scientific committee of the ENUMATH 2017 conference.

Martin Vohralík was a member of the scientific committee of the *Finite Volumes for Complex Applications* conference.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

François Clément is a member of the editorial board of Matapli from September 2017.

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, Computational Methods in Applied Mathematics, and Journal de l'Ecole polytechnique, Mathématiques.

M. Vohralík is a member of the editorial boards of SIAM Journal on Numerical Analysis and of Acta Polytechnica.

10.1.3.2. Reviewer - Reviewing Activities

Thomas Boiveau was a reviewer for the Journal of Scientific Computing and the journal Geometrically Unfitted Finite Element Method and Applications (Proceedings of the UCL Workshop 2016).

Matteo Cicuttin was a reviewer for the journal Journal of Computational and Applied Mathematics.

Alexandre Ern served as reviewer for tens of papers in different journals.

Frédéric Marazzato was a reviewer for the journal Modelling and Simulation in Materials Science and Engineering.

Laurent Monasse was a reviewer for the Journal of Computational Physics, the European Journal of Mechanics B Fluids, and the International Journal for Numerical Methods in Engineering.

Michel Kern was a reviewer for Computers and Geosciences, Advances in Computational Mathematics, ESAIM: proceeedings, ARIMA, and Applicable Analysis.

Martin Vohralík served as reviewer for tens of papers in different journals.

10.1.4. Invited Talks

Alexandre Ern, plenary speaker, SIAM Conference on Mathematical and Computational Issues in the Geosciences 2017, Erlangen, Germany.

Alexandre Ern, invited speaker, POEMS 2017, Milano, Italy.

Alexandre Ern, invited speaker, Inauguration workshop of the French-German-Italian LIA COPDESC on Applied Analysis.

Michel Kern, plenary speaker, 6th Workshop on Parallel-in-Time methods, Locarno, Switzerland.

Martin Vohralík, plenary speaker, TamTam 2017, Hammamet, Tunisia, May 2017.

Martin Vohralík, plenary speaker, International Conference on Approximation Methods and Numerical Modelling in Environment and Natural Resources, Oujda, Morocco, May 2017.

Martin Vohralík, plenary speaker, Czech workshop on complex systems, Prague, Czech Republic, September 2017.

10.1.5. Leadership within the Scientific Community

M. Kern is a reviewer for the German Supercomputing Center JARA program.

M. Kern is a member of the Scientific Committee of Orap (ORganisation Associative du Parallélisme), 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.

10.1.6. Research Administration

F. Clément is a member of the *Comité local d'hygiène, de sécurité et des conditions de travail* of the Inria Research Center of Paris.

F. Clément is the *AMIES facilitator* of the Inria Research Center of Paris.

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.

G. Pichot is a member of the *Comité local d'hygiène, de sécurité et des conditions de travail* of the Inria center of Paris.

G. Pichot is member of the Conseil de département MAM of Polytech Lyon.

G. Pichot is member of the Commission de developpement technologique (CDT) of the Inria center of Paris.

G. Pichot is a member of the CES commission of the Inria center of Paris.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence : Amina Benaceur, Eléments d'arithmétique, 19h15, L2, University Pierre et Marie Curie, France.

Licence : Thomas Boiveau, Equations aux dérivées partielles et éléments finis 15h, L3, Ecole des ponts ParisTech, France

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.

Licence : Frédéric Marazzato, Analyse et Calcul Scientifique, 15h, L3, Ecole Nationale des Ponts et Chaussées, France.

Master : Alexandre Ern, Discontinuous Galerkin methods, 20h, M2, University Pierre et Marie Curie, France.

Master : Alexandre Ern is the leader of the Master Mathématiques et applications of ENPC.
Master : Frédéric Marazzato, Remise à niveau en Mécanique des Milieux Continus, 9h, M2, Ecole Nationale des Ponts et Chaussées, France.
Master : Frédéric Marazzato, Reliability of Materials and Structures, 15h, M1, Ecole Nationale des Ponts et Chaussées, France.
Master : Frédéric Marazzato, Projet de Département Génie Mécanique et Matériaux, 20h, M1, Ecole Nationale des Ponts et Chaussées, 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, 30 h (with E. Mouche), M2, Université Paris Saclay, France
Martin Vohralík, CIMPA Summer School on Multiscale Computational Methods and Error Control, Kanpur, India, July 2017, 10h.

10.2.2. Supervision

PhD: Sarah Ali Hassan, A posteriori error estimates and stopping criteria for solvers using the domain decomposition method and with local time stepping, University Pierre and Marie Curie, 26 June 2017, Martin Vohralík, Caroline Japhet, and Michel Kern.
PhD: Yannick Masson, Tchebyshev nets and application to Grid Shells, University Paris-Est, 09 June 2017, Alexandre Ern and Olivier Baverel.
PhD: Rita Riedlbeck, Adaptive algorithms for poromechanics and poroplasticity, University of Montpellier, 27 November 2017, Daniele di Pietro and Alexandre Ern.
PhD in progress: Amina Benaceur, Model reduction in thermo-mechanics, 01 January 2016, Alexandre Ern.
PhD in progress: Karol Cascavita, Discontinuous skeletal methods for yield fluids, 01 October 2015, Alexandre Ern and Xavier Chateau.
PhD in progress: Jad Dabaghi, Adaptive modeling via complementarity of phase appearance and disappearance in fractured and porous media, 01 November 2015, Martin Vohralík Vincent Martin.
PhD in progress: Patrik Daniel, Adaptive multilevel solvers with a posteriori error control for porous media flows, 01 October 2015, Martin Vohralík and Alexandre Ern.
PhD in progress: Frédéric Marazzato, Fracture and fragmentation simulated by the discrete element method, 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, Examiner, HDR C. Le Potier, University Paris Est, 15 November 2017.
Laurent Monasse, Examiner, PhD J. Ridoux, University Pierre and Marie Curie, 4 October 2017.
Martin Vohralík, President of the committee, PhD R. Riedelbeck, University of Montpellier, 27 November 2017.
Martin Vohralík, Examiner, HDR P. Gosselet, ENS Cachan, 10 February 2017.

10.3. Popularization

F. Clément was the coordinator of the Maths-Enterprises booth for AMIES and Labex mathématiques Hadamard at the *18e Salon Culture & Jeux Mathématiques*, held in Paris, 27–30 May, 2017.
F. Clément coordinated an article for AMIES in Interstices ([42]).

Major publications by the team in recent years: [1], [2], [3], [4], [5], [6], [7], [8], [9], [10].

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] A. ERN, M. VOHRALÍK.*Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs*, in "SIAM J. Sci. Comput.", 2013, vol. 35, n^{o} 4, p. A1761–A1791, http://dx.doi.org/10.1137/120896918.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] I. SMEARS.*Robust and efficient preconditioners for the discontinuous Galerkin time-stepping method*, in "IMA Journal of Numerical Analysis", October 2016 [*DOI :* 10.1093/IMANUM/DRW050], https://hal.archives-ouvertes.fr/hal-01357497.

[10] M. VOHRALÍK, B. I. WOHLMUTH.*Mixed finite element methods: implementation with one unknown per element, local flux expressions, positivity, polygonal meshes, and relations to other methods*, in "Math. Models Methods Appl. Sci.", 2013, vol. 23, n^{o} 5, p. 803–838, http://www.worldscientific.com/doi/abs/10.1142/S0218202512500613.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] S. ALI HASSAN.*A posteriori error estimates and stopping criteria for solvers using the domain decomposition method and with local time stepping*, Paris 6 ; Inria Paris, June 2017, https://hal.inria.fr/tel-01672497.

[12] R. RIEDLBECK.*Adaptive algorithms for poro-mechanics and poro-plasticity*, Université de Montpellier, November 2017, https://hal.inria.fr/tel-01676709.

Articles in International Peer-Reviewed Journal

[13] M. ABBAS, A. ERN, N. PIGNET.*Hybrid High-Order methods for finite deformations of hyperelastic materials*, in "Computational Mechanics", 2018, p. 1-29, https://hal.archives-ouvertes.fr/hal-01575370.

[14] 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", March 2017, 31, https://arxiv.org/abs/1602.08304 [*DOI :* 10.1080/17415977.2017.1290087], https://hal.inria.fr/hal-01279503.

[15] S. BLIUDZE, S. FURIC, J. SIFAKIS, A. VIEL.*Rigorous Design of Cyber-Physical Systems: Linking Physicality and Computation*, in "Software and Systems Modeling", 2017, p. 1-24 [*DOI :* 10.1007/s10270-017-0642-5], https://hal.inria.fr/hal-01636392.

[16] E. BURMAN, A. ERN.*A nonlinear consistent penalty method weakly enforcing positivity in the finite element approximation of the transport equation*, in "Computer Methods in Applied Mechanics and Engineering", 2017, vol. 320, p. 122-132 [*DOI :* 10.1016/J.CMA.2017.03.019], https://hal.archives-ouvertes.fr/hal-01383295.

[17] E. BURMAN, A. ERN, M. A. FERNÁNDEZ.*Fractional-step methods and finite elements with symmetric stabilization for the transient Oseen problem*, in "ESAIM: Mathematical Modelling and Numerical Analysis", March 2017, vol. 51, n^{o} 2, p. 487-507 [*DOI :* 10.1051/M2AN/2016028], https://hal.inria.fr/hal-01218328.

[18] 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 Journal on Numerical Analysis", September 2017, vol. 55, n^{o} 5, p. 2228-2254 [*DOI :* 10.1137/15M1038633], https://hal.inria.fr/hal-01194364.

[19] P. CANTIN, A. ERN.*An edge-based scheme on polyhedral meshes for vector advection-reaction equations*, in "ESAIM: Mathematical Modelling and Numerical Analysis", 2017 [*DOI :* 10.1051/M2AN/2016075], https://hal.archives-ouvertes.fr/hal-01324545.

[20] D. A. DI PIETRO, A. ERN.*Arbitrary-order mixed methods for heterogeneous anisotropic diffusion on general meshes*, in "IMA Journal of Numerical Analysis", January 2017, vol. 37, n^{o} 1, p. 40-63, https://hal.archives-ouvertes.fr/hal-00918482.

[21] A. ERN, J.-L. GUERMOND.*Finite element quasi-interpolation and best approximation* , in "ESAIM: Mathematical Modelling and Numerical Analysis", 2017 [*DOI :* 10.1051/M2AN/2016066], https://hal.archives-ouvertes.fr/hal-01155412.

[22] A. ERN, I. SMEARS, M. VOHRALÍK.*Discrete p-robust H(div)-liftings and a posteriori estimates for elliptic problems with H^{-1} source terms*, in "Calcolo", January 2017, vol. 54, n^{o} 3, p. 1009-1025 [*DOI :* 10.1007/s10092-017-0217-4], https://hal.inria.fr/hal-01377007.

[23] A. ERN, I. SMEARS, M. VOHRALÍK.*Guaranteed, locally space-time efficient, and polynomial-degree robust a posteriori error estimates for high-order discretizations of parabolic problems*, in "SIAM Journal on Numerical Analysis", November 2017, vol. 55, n^{o} 6, p. 2811-2834 [*DOI :* 10.1137/16M1097626], https://hal.archives-ouvertes.fr/hal-01377086.

[24] C. LUSSO, F. BOUCHUT, A. ERN, A. MANGENEY.*A free interface model for static/flowing dynamics in thin-layer flows of granular materials with yield: simple shear simulations and comparison with experiments*, in "Applied Sciences", April 2017, vol. 7, n^{o} 4, 386 [*DOI :* 10.3390/APP7040386], https://hal-upec-upem.archives-ouvertes.fr/hal-00992309.

[25] C. LUSSO, A. ERN, F. BOUCHUT, A. MANGENEY, M. FARIN, O. ROCHE.*Two-dimensional simulation by regularization of free surface viscoplastic flows with Drucker-Prager yield stress and application to granular collapse*, in "Journal of Computational Physics", March 2017, vol. 333, p. 387-408 [*DOI :* 10.1016/J.JCP.2016.12.036], https://hal-upec-upem.archives-ouvertes.fr/hal-01133786.

[26] J. PAPEŽ, Z. STRAKOŠ, M. VOHRALÍK.*Estimating and localizing the algebraic and total numerical errors using flux reconstructions*, in "Numerische Mathematik", September 2017 [*DOI :* 10.1007/S00211-017-0915-5], https://hal.inria.fr/hal-01312430.

[27] J. RIDOUX, N. LARDJANE, L. MONASSE, F. COULOUVRAT.*Comparison of Geometrical Shock Dynamics and Kinematic models for shock wave propagation*, in "Shock Waves", 2017, https://hal.archives-ouvertes.fr/hal-01511489.

[28] R. RIEDLBECK, D. DI PIETRO, A. ERN, S. GRANET, K. KAZYMYRENKO.*Stress and flux reconstruction in Biot's poro-elasticity problem with application to a posteriori error analysis*, in "Computers and Mathematics with Applications", April 2017, vol. 73, n^{o} 7, p. 1593–1610 [*DOI :* 10.1016/J.CAMWA.2017.02.005], https://hal.archives-ouvertes.fr/hal-01366646.

[29] I. SMEARS.*Nonoverlapping domain decomposition preconditioners for discontinuous Galerkin approximations of Hamilton–Jacobi–Bellman equations*, in "Journal of Scientific Computing", 2017 [*DOI :* 10.1007/S10915-017-0428-5], https://hal.inria.fr/hal-01428790.

[30] 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", December 2017 [*DOI :* 10.1016/J.CMA.2017.11.027], https://hal.archives-ouvertes.fr/hal-01532195.

[31] 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", November 2017 [*DOI :* 10.1007/S00211-017-0925-3], https://hal.inria.fr/hal-01097662.

International Conferences with Proceedings

[32] S. BOLDO, F. CLÉMENT, F. FAISSOLE, V. MARTIN, M. MAYERO.*A Coq Formal Proof of the Lax–Milgram theorem*, in "6th ACM SIGPLAN Conference on Certified Programs and Proofs", Paris, France, January 2017 [*DOI :* 10.1145/3018610.3018625], https://hal.inria.fr/hal-01391578.

[33] P. LAUG, G. PICHOT, J.-R. D. DREUZY.*Realistic geometric modeling of fracture networks*, in "8th International Conference on Adaptive Modeling and Simulation (ADMOS 2017) - Symposium "Mesh generation and mesh adaptativity: methods and applications"", Verbania, Italy, June 2017, https://hal.inria.fr/hal-01591579.

National Conferences with Proceeding

[34] S. BOLDO, F. CLÉMENT, F. FAISSOLE, V. MARTIN, M. MAYERO.*Preuve formelle du théorème de Lax–Milgram*, in "16èmes journées Approches Formelles dans l'Assistance au Développement de Logiciels", Montpellier, France, June 2017, https://hal.archives-ouvertes.fr/hal-01581807.

Conferences without Proceedings

[35] H. BARUCQ, H. CALANDRA, G. CHAVENT, M. V. DE HOOP, F. FAUCHER.*Stability and convergence analysis for seismic depth imaging using FWI*, in "Computational Inverse Problems for Partial Differential Equations Workshop", Oberwolfach, Germany, May 2017, https://hal.archives-ouvertes.fr/hal-01623952.

[36] H. BARUCQ, H. CALANDRA, G. CHAVENT, M. V. DE HOOP, F. FAUCHER.*Stability and convergence for seismic reconstruction using full waveform inversion*, in "Group Seminar", Linz, Austria, July 2017, https://hal.archives-ouvertes.fr/hal-01623950.

[37] H. BARUCQ, H. CALANDRA, G. CHAVENT, F. FAUCHER.*Convergence Analysis for Seismic Full Waveform Inversion*, in "MATHIAS – TOTAL Symposium on Mathematics", Paris, France, October 2017, https://hal.archives-ouvertes.fr/hal-01623953.

[38] H. BARUCQ, H. CALANDRA, G. CHAVENT, F. FAUCHER.*Convergence of seismic full waveform inversion and extension to Cauchy data*, in "Inverse Days 2017", Oulu, Finland, December 2017, https://hal.archives-ouvertes.fr/hal-01662677.

[39] M. CHAVENT, G. CHAVENT.*Approche bloc en ACP group-sparse: le package sparsePCA*, in "AngletR 2017 - 6ème Rencontres R", Anglet, France, June 2017, 1, https://hal.archives-ouvertes.fr/hal-01662652.

[40] M. CICUTTIN, A. ERN, S. LEMAIRE.*On the implementation of the multiscale Hybrid High-Order method*, in "ENUMATH 2017 - European Conference on Numerical Mathematics and Advanced Applications", Voss, Norway, September 2017, p. 1-8, https://hal.archives-ouvertes.fr/hal-01661925.

Scientific Books (or Scientific Book chapters)

[41] R. RIEDLBECK, D. A. DI PIETRO, A. ERN.*Equilibrated stress reconstructions for linear elasticity problems with application to a posteriori error analysis*, in "Finite Volumes for Complex Applications VIII – Methods and Theoretical Aspects", June 2017 [*DOI :* 10.1007/978-3-319-57397-7], https://hal.archives-ouvertes.fr/hal-01433841.

Scientific Popularization

[42] F. CLÉMENT.*Les mathématiques s'appliquent aussi à l'industrie*, in "Interstices", January 2017, https://hal.inria.fr/hal-01466798.

Other Publications

[43] E. AHMED, S. ALI HASSAN, C. JAPHET, M. KERN, M. VOHRALÍK.*A posteriori error estimates and stopping criteria for space-time domain decomposition for two-phase flow between different rock types*, June 2017, working paper or preprint, https://hal.inria.fr/hal-01540956.

[44] E. AHMED, A. BEN ABDA.*The sub-Cauchy Stokes Problem: Solvability Issues and Lagrange Multiplier Methods with Artificial Boundary Conditions*, October 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01467425.

[45] S. ALI HASSAN, C. JAPHET, M. KERN, M. VOHRALÍK.*A posteriori stopping criteria for optimized Schwarz domain decomposition algorithms in mixed formulations*, May 2017, working paper or preprint, https://hal.inria.fr/hal-01529532.

[46] S. ALI HASSAN, C. JAPHET, M. VOHRALÍK.*A posteriori stopping criteria for space-time domain decomposition for the heat equation in mixed formulations* , September 2017, working paper or preprint, https://hal.inria.fr/hal-01586862.

[47] L. AMIR, M. KERN.*Preconditioning a coupled model for reactive transport in porous media*, September 2017, https://arxiv.org/abs/1710.01483 - working paper or preprint, https://hal.inria.fr/hal-01327307.

[48] A. BENACEUR, V. EHRLACHER, A. ERN, S. MEUNIER.*A progressive reduced basis/empirical interpolation method for nonlinear parabolic problems* , November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01599304.

[49] T. BOIVEAU, V. EHRLACHER, A. ERN, A. NOUY.*Low-rank approximation of linear parabolic equations by space-time tensor Galerkin methods*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01668316.

[50] E. BURMAN, A. ERN.*A cut-cell Hybrid High-Order method for elliptic problems with curved boundaries*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01653685.

[51] E. BURMAN, A. ERN.*An unfitted Hybrid High-Order method for elliptic interface problems* *, October 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01625421.

[52] V. M. CALO, M. CICUTTIN, Q. DENG, A. ERN.*Spectral approximation of elliptic operators by the Hybrid High-Order method*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01628698.

[53] E. CANCÈS, G. DUSSON, Y. MADAY, B. STAMM, M. VOHRALÍK.*Guaranteed and robust a posteriori bounds for Laplace eigenvalues and eigenvectors: a unified framework*, March 2017, working paper or preprint, https://hal.inria.fr/hal-01483461.

[54] M. CHAVENT, G. CHAVENT.*Group-sparse block PCA and explained variance*, December 2017, https://arxiv.org/abs/1705.00461 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01662605.

[55] P. CIARLET, M. VOHRALÍK.*Localization of global norms and robust a posteriori error control for transmission problems with sign-changing coefficients*, August 2017, working paper or preprint, https://hal.inria.fr/hal-01148476.

[56] M. CICUTTIN, D. A. DI PIETRO, A. ERN.*Implementation of Discontinuous Skeletal methods on arbitrary-dimensional, polytopal meshes using generic programming*, September 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01429292.

[57] M. CICUTTIN, A. ERN, S. LEMAIRE.*A Multiscale Hybrid High-Order method*, February 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01467434.

[58] J. DABAGHI, V. MARTIN, M. VOHRALÍK.*Adaptive inexact semismooth Newton methods for the contact problem between two membranes*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01666845.

[59] P. DANIEL, A. ERN, I. SMEARS, M. VOHRALÍK.*An adaptive hp-refinement strategy with computable guaranteed bound on the error reduction factor*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01666763.

[60] A. ERN, J.-L. GUERMOND.*Abstract nonconforming error estimates and application to boundary penalty methods for diffusion equations and time-harmonic Maxwell's equations*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01563594.

[61] A. ERN, J.-L. GUERMOND.*Analysis of the edge finite element approximation of the Maxwell equations with low regularity solutions*, June 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01531940.

[62] 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*, March 2017, working paper or preprint, https://hal.inria.fr/hal-01489721.

[63] M. JENSEN, I. SMEARS.*On the notion of boundary conditions in comparison principles for viscosity solutions*, March 2017, working paper or preprint, https://hal.inria.fr/hal-01493586.

[64] A. LEJAY, L. LENÔTRE, G. PICHOT.*Analytic expressions of the solutions of advection-diffusion problems in 1D with discontinuous coefficients*, November 2017, working paper or preprint, https://hal.inria.fr/hal-01644270.

[65] F. MARAZZATO, A. ERN, C. MARIOTTI, L. MONASSE.*An explicit energy-momentum conserving time-integration scheme for Hamiltonian dynamics*, December 2017, working paper or preprint, https://hal-enpc.archives-ouvertes.fr/hal-01661608.

[66] J. PAPEŽ, U. RÜDE, M. VOHRALÍK, B. WOHLMUTH.*Sharp algebraic and total a posteriori error bounds for h and p finite elements via a multilevel approach*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01662944.

[67] 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.

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:

B9.4.5. - Data science

1. Personnel

Research Scientists

Francis Bach [Team leader, Inria, Senior Researcher, HDR]
Alexandre d'Aspremont [CNRS, Senior Researcher]
Pierre Gaillard [Inria, Researcher]
Alessandro Rudi [Inria, Starting Research Position, from Sep 2017]

Faculty Member

Marco Cuturi Cameto [ENSEA, Associate Professor]

External Collaborators

Christophe Dupuy [from Jul 2017]
Senanayak Karri [until Sep 2017]
Simon Lacoste-Julien
Guillaume Obozinski [Ecole Nationale des Ponts et Chaussées, until Jun 2017]
Balamurugan Palaniappan [École Nationale Supérieure de Techniques Avancées]
Fabian Pedregosa [from Apr 2017 until Aug 2017]
Federico Vaggi [Ecole Normale Supérieure Paris, from May 2017 until Aug 2017]

Technical Staff

Fabian Pedregosa [until Apr 2017]

PhD Students

Jean-Baptiste Alayrac [Ecole polytechnique, until Sep 2017]
Dmitry Babichev [Inria]
Anaël Bonneton [Ecole Normale Supérieure Paris]
Margaux Bregere [EDF, from Oct 2017]
Alexandre Defossez [Facebook]
Aymeric Dieuleveut [Ecole Normale Supérieure Paris, until Sep 2017]
Christophe Dupuy [Technicolor, until Jun 2017]
Nicolas Flammarion [Ecole Normale Supérieure Lyon, until Aug 2017]
Damien Garreau [Inria, until Aug 2017]
Thomas Kerdreux [Ecole Normale Supérieure Paris, from Oct 2017]
Remi Leblond [Inria]

Loucas Pillaud Vivien [Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer, from Sep 2017]
Antoine Recanati [CNRS]
Vincent Roulet [Ecole polytechnique]
Damien Scieur [Inria]
Tatiana Shpakova [Inria]

Post-Doctoral Fellows

Lenaic Chizat [Inria, from Dec 2017]
Igor Colin [Institut Telecom ex GET Groupe des Ecoles des Télécommunications]
Pascal Germain [Inria, until Oct 2017]
Robert Gower [Inria, until Aug 2017]
Adrien Taylor [Inria, from Oct 2017]
Federico Vaggi [Ecole Normale Supérieure Paris, until Apr 2017]

Visiting Scientists

Marwa El Halabi [Ecole polytechnique, until Apr 2017]
Gauthier Gidel [from Aug 2017]
Lucas Rencker [from Mar 2017 until Sep 2017]
Jonathan Weed [from Mar 2017 until May 2017]
Alfredo Zermini [from Mar 2017 until Jun 2017]

Administrative Assistants

Anja Plos [Inria]
Lindsay Polienor [Inria]
Sandrine Verges [Inria]

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. Application Domains

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. New Software and Platforms

5.1. 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

5.2. 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.

- Contact: Jean-Baptiste Alayrac

6. New Results

6.1. On Structured Prediction Theory with Calibrated Convex Surrogate Losses

In [16], we provide novel theoretical insights on structured prediction in the context of efficient convex surrogate loss minimization with consistency guarantees. For any task loss, we construct a convex surrogate that can be optimized via stochastic gradient descent and we prove tight bounds on the so-called "calibration function" relating the excess surrogate risk to the actual risk. In contrast to prior related work, we carefully monitor the effect of the exponential number of classes in the learning guarantees as well as on the optimization complexity. As an interesting consequence, we formalize the intuition that some task losses make learning harder than others, and that the classical 0-1 loss is ill-suited for general structured prediction.

6.2. Domain-Adversarial Training of Neural Networks

In [18], we introduce a new representation learning approach for domain adaptation, in which data at training and test time come from similar but different distributions. Our approach is directly inspired by the theory on domain adaptation suggesting that, for effective domain transfer to be achieved, predictions must be made based on features that cannot discriminate between the training (source) and test (target) domains. The approach implements this idea in the context of neural network architectures that are trained on labeled data from the source domain and unlabeled data from the target domain (no labeled target-domain data is necessary). As the training progresses, the approach promotes the emergence of features that are (i) discriminative for the main learning task on the source domain and (ii) indiscriminate with respect to the shift between the domains. We show that this adaptation behaviour can be achieved in almost any feed-forward model by augmenting it with few standard layers and a new gradient reversal layer. The resulting augmented architecture can be trained using standard backpropagation and stochastic gradient descent, and can thus be implemented with little effort using any of the deep learning packages. We demonstrate the success of our approach for two distinct classification problems (document sentiment analysis and image classification), where state-of-the-art domain adaptation performance on standard benchmarks is achieved. We also validate the approach for descriptor learning task in the context of person re-identification application.

6.3. Linearly Convergent Randomized Iterative Methods for Computing the Pseudoinverse

In [25], we develop the first stochastic incremental method for calculating the Moore-Penrose pseudoinverse of a real matrix. By leveraging three alternative characterizations of pseudoinverse matrices, we design three methods for calculating the pseudoinverse: two general purpose methods and one specialized to symmetric matrices. The two general purpose methods are proven to converge linearly to the pseudoinverse of any given matrix. For calculating the pseudoinverse of full rank matrices we present two additional specialized methods which enjoy a faster convergence rate than the general purpose methods. We also indicate how to develop randomized methods for calculating approximate range space projections, a much needed tool in inexact Newton type methods or quadratic solvers when linear constraints are present. Finally, we present numerical experiments of our general purpose methods for calculating pseudoinverses and show that our methods greatly outperform the Newton-Schulz method on large dimensional matrices.

6.4. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance

The Wasserstein distance between two probability measures on a metric space is a measure of closeness with applications in statistics, probability, and machine learning. In [39], we consider the fundamental question of how quickly the empirical measure obtained from n independent samples from μ approaches μ in the Wasserstein distance of any order. We prove sharp asymptotic and finite-sample results for this rate of convergence for general measures on general compact metric spaces. Our finite-sample results show the existence of multi-scale behavior, where measures can exhibit radically different rates of convergence as n grows. Collaboration with Jonathan Weed, Francis Bach)

6.5. Efficient Algorithms for Non-convex Isotonic Regression through Submodular Optimization

In [19], we consider the minimization of submodular functions subject to ordering constraints. We show that this optimization problem can be cast as a convex optimization problem on a space of uni-dimensional measures, with ordering constraints corresponding to first-order stochastic dominance. We propose new discretization schemes that lead to simple and efficient algorithms based on zero-th, first, or higher order oracles; these algorithms also lead to improvements without isotonic constraints. Finally, our experiments show that non-convex loss functions can be much more robust to outliers for isotonic regression, while still leading to an efficient optimization problem.

6.6. Bridging the Gap between Constant Step Size Stochastic Gradient Descent and Markov Chains

In [21], we consider the minimization of an objective function given access to unbiased estimates of its gradient through stochastic gradient descent (SGD) with constant step-size. While the detailed analysis was only performed for quadratic functions, we provide an explicit asymptotic expansion of the moments of the averaged SGD iterates that outlines the dependence on initial conditions, the effect of noise and the step-size, as well as the lack of convergence in the general (non-quadratic) case. For this analysis, we bring tools from Markov chain theory into the analysis of stochastic gradient and create new ones (similar but different from stochastic MCMC methods). We then show that Richardson-Romberg extrapolation may be used to get closer to the global optimum and we show empirical improvements of the new extrapolation scheme.

6.7. AdaBatch: Efficient Gradient Aggregation Rules for Sequential and Parallel Stochastic Gradient Methods

In [22], we study a new aggregation operator for gradients coming from a mini-batch for stochastic gradient (SG) methods that allows a significant speed-up in the case of sparse optimization problems. We call this method AdaBatch and it only requires a few lines of code change compared to regular mini-batch SGD algorithms. We provide a theoretical insight to understand how this new class of algorithms is performing and show that it is equivalent to an implicit per-coordinate rescaling of the gradients, similarly to what Adagrad methods can do. In theory and in practice, this new aggregation allows to keep the same sample efficiency of SG methods while increasing the batch size. Experimentally, we also show that in the case of smooth convex optimization, our procedure can even obtain a better loss when increasing the batch size for a fixed number of samples. We then apply this new algorithm to obtain a parallelizable stochastic gradient method that is synchronous but allows speed-up on par with Hogwild! methods as convergence does not deteriorate with the increase of the batch size. The same approach can be used to make mini-batch provably efficient for variance-reduced SG methods such as SVRG.

6.8. Structure-Adaptive, Variance-Reduced, and Accelerated Stochastic Optimization

In [38], we explore the fundamental structure-adaptiveness of state of the art randomized first order algorithms on regularized empirical risk minimization tasks, where the solution has intrinsic low-dimensional structure (such as sparsity and low-rank). Such structure is often enforced by non-smooth regularization or constraints. We start by establishing the fast linear convergence rate of the SAGA algorithm on non-strongly-convex objectives with convex constraints, via an argument of cone-restricted strong convexity. Then for the composite minimization task with a coordinate-wise separable convex regularization term, we propose and analyse a two stage accelerated coordinate descend algorithm (Two-Stage APCG). We provide the convergence analysis showing that the proposed method has a global convergence in general and enjoys a local accelerated linear convergence rate with respect to the low-dimensional structure of the solution. Then based on this convergence result, we proposed an adaptive variant of the two-stage APCG method which does not need to foreknow the restricted strong convexity beforehand, but estimate it on the fly. In numerical experiments we compare the adaptive two-stage APCG with various state of the art variance-reduced stochastic gradient methods on sparse regression tasks, and demonstrate the effectiveness of our approach.

6.9. Exponential convergence of testing error for stochastic gradient methods

In [31], 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.

6.10. Optimal algorithms for smooth and strongly convex distributed optimization in networks

In [35], we determine the optimal convergence rates for strongly convex and smooth distributed optimization in two settings: centralized and decentralized communications over a network. For centralized (i.e. *master/slave*) algorithms, we show that distributing Nesterov's accelerated gradient descent is optimal and achieves a precision $\varepsilon > 0$ in time $O(\sqrt{\kappa_g}(1+\Delta\tau)\ln(1/\varepsilon))$, where κ_g is the condition number of the (global) function to optimize, Δ is the diameter of the network, and τ (resp. 1) is the time needed to communicate values between two neighbors (resp. perform local computations). For decentralized algorithms based on gossip, we provide the first optimal algorithm, called the *multi-step dual accelerated* (MSDA) method, that achieves a precision $\varepsilon > 0$ in time $O(\sqrt{\kappa_l}(1+\frac{\tau}{\sqrt{\gamma}})\ln(1/\varepsilon))$, where κ_l is the condition number of the local functions and γ is the (normalized) eigengap of the gossip matrix used for communication between nodes. We then verify the efficiency of MSDA against state-of-the-art methods for two problems: least-squares regression and classification by logistic regression.

6.11. Stochastic Composite Least-Squares Regression with convergence rate O(1/n)

In [23], we consider the minimization of composite objective functions composed of the expectation of quadratic functions and an arbitrary convex function. We study the stochastic dual averaging algorithm with a constant step-size, showing that it leads to a convergence rate of $O(1/n)$ without strong convexity assumptions. This thus extends earlier results on least-squares regression with the Euclidean geometry to (a) all convex regularizers and constraints, and (b) all geome-tries represented by a Bregman divergence. This is achieved by a new proof technique that relates stochastic and deterministic recursions.

6.12. Sharpness, Restart and Acceleration

The Łojasiewicz inequality shows that sharpness bounds on the minimum of convex optimization problems hold almost generically. Sharpness directly controls the performance of restart schemes. The constants quantifying error bounds are of course unobservable, but we show in [33] that optimal restart strategies are robust, and searching for the best scheme only increases the complexity by a logarithmic factor compared to the optimal bound. Overall then, restart schemes generically accelerate accelerated methods.

6.13. PAC-Bayes and Domain Adaptation

In [24], we provide two main contributions in PAC-Bayesian theory for domain adaptation where the objective is to learn, from a source distribution, a well-performing majority vote on a different, but related, target distribution. Firstly, we propose an improvement of the previous approach we proposed in Germain et al. (2013), which relies on a novel distribution pseudodistance based on a disagreement averaging, allowing us to derive a new tighter domain adaptation bound for the target risk. While this bound stands in the spirit of common domain adaptation works, we derive a second bound (recently introduced in Germain et al., 2016) that brings a new perspective on domain adaptation by deriving an upper bound on the target risk where the distributions' divergence—expressed as a ratio—controls the trade-off between a source error measure and the target voters' disagreement. We discuss and compare both results, from which we obtain PAC-Bayesian generalization bounds. Furthermore, from the PAC-Bayesian specialization to linear classifiers, we infer two learning algorithms, and we evaluate them on real data.

6.14. Kernel Square-Loss Exemplar Machines for Image Retrieval

Zepeda and Pérez have recently demonstrated the promise of the exemplar SVM (ESVM) as a feature encoder for image retrieval. The paper [6] extends this approach in several directions: We first show that replacing the hinge loss by the square loss in the ESVM cost function significantly reduces encoding time with negligible effect on accuracy. We call this model square-loss exemplar machine, or SLEM. We then introduce a kernelized

SLEM which can be implemented efficiently through low-rank matrix decomposition , and displays improved performance. Both SLEM variants exploit the fact that the negative examples are fixed, so most of the SLEM computational complexity is relegated to an offline process independent of the positive examples. Our experiments establish the performance and computational advantages of our approach using a large array of base features and standard image retrieval datasets.

6.15. Breaking the Nonsmooth Barrier: A Scalable Parallel Method for Composite Optimization

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 [10], 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.

6.16. PAC-Bayesian Analysis for a two-step Hierarchical Multiview Learning Approach

In [15], we study a two-level multiview learning with more than two views under the PAC-Bayesian framework. This approach, sometimes referred as late fusion, consists in learning sequentially multiple view-specific classifiers at the first level, and then combining these view-specific classifiers at the second level. Our main theoretical result is a generalization bound on the risk of the majority vote which exhibits a term of diversity in the predictions of the view-specific classifiers. From this result it comes out that controlling the trade-off between diversity and accuracy is a key element for multiview learning, which complements other results in multiview learning.

6.17. Integration Methods and Accelerated Optimization Algorithms

In [37], we show that accelerated optimization methods can be seen as particular instances of multi-step integration schemes from numerical analysis, applied to the gradient flow equation. In comparison with recent advances in this vein, the differential equation considered here is the basic gradient flow and we show that multi-step schemes allow integration of this differential equation using larger step sizes, thus intuitively explaining acceleration results.

6.18. GANs for Biological Image Synthesis

In [17], we propose a novel application of Generative Adversarial Networks (GAN) to the synthesis of cells imaged by fluorescence microscopy. Compared to natural images, cells tend to have a simpler and more geometric global structure that facilitates image generation. However, the correlation between the spatial pattern of different fluorescent proteins reflects important biological functions, and synthesized images have to capture these relationships to be relevant for biological applications. We adapt GANs to the task at hand and propose new models with casual dependencies between image channels that can generate multi-channel images, which would be impossible to obtain experimentally. We evaluate our approach using two independent techniques and compare it against sensible baselines. Finally, we demonstrate that by interpolating across the latent space we can mimic the known changes in protein localization that occur through time during the cell cycle, allowing us to predict temporal evolution from static images.

6.19. Nonlinear Acceleration of Stochastic Algorithms

Extrapolation methods use the last few iterates of an optimization algorithm to produce a better estimate of the optimum. They were shown to achieve optimal convergence rates in a deterministic setting using simple gradient iterates. In [36], we study extrapolation methods in a stochastic setting, where the iterates are produced by either a simple or an accelerated stochastic gradient algorithm. We first derive convergence bounds for arbitrary, potentially biased perturbations, then produce asymptotic bounds using the ratio between the variance of the noise and the accuracy of the current point. Finally, we apply this acceleration technique to stochastic algorithms such as SGD, SAGA, SVRG and Katyusha in different settings, and show significant performance gains.

6.20. Algorithmic Chaining and the Role of Partial Feedback in Online Nonparametric Learning

In [20], we investigate contextual online learning with nonparametric (Lipschitz) comparison classes under different assumptions on losses and feedback information. For full information feedback and Lipschitz losses, we design the first explicit algorithm achieving the minimax regret rate (up to log factors). In a partial feedback model motivated by second-price auctions, we obtain algorithms for Lipschitz and semi-Lipschitz losses with regret bounds improving on the known bounds for standard bandit feedback. Our analysis combines novel results for contextual second-price auctions with a novel algorithmic approach based on chaining. When the context space is Euclidean, our chaining approach is efficient and delivers an even better regret bound.

6.21. Frank-Wolfe Algorithms for Saddle Point Problems

In [14], we extend the Frank-Wolfe (FW) optimization algorithm to solve constrained smooth convex-concave saddle point (SP) problems. Remarkably, the method only requires access to linear minimization oracles. Leveraging recent advances in FW optimization, we provide the first proof of convergence of a FW-type saddle point solver over polytopes, thereby partially answering a 30 year-old conjecture. We also survey other convergence results and highlight gaps in the theoretical underpinnings of FW-style algorithms. Motivating applications without known efficient alternatives are explored through structured prediction with combinatorial penalties as well as games over matching polytopes involving an exponential number of constraints.

6.22. Convex optimization over intersection of simple sets: improved convergence rate guarantees via an exact penalty approach

In [29], We consider the problem of minimizing a convex function over the intersection of finitely many simple sets which are easy to project onto. This is an important problem arising in various domains such as machine learning. The main difficulty lies in finding the projection of a point in the intersection of many sets. Existing approaches yield an infeasible point with an iteration-complexity of $O(1/\varepsilon^2)$ for nonsmooth problems with no guarantees on the in-feasibility. By reformulating the problem through exact penalty functions, we derive first-order algorithms which not only guarantees that the distance to the intersection is small but also improve the complexity to $O(1/\varepsilon)$ and $O(1/\sqrt{\varepsilon})$ for smooth functions. For composite and smooth problems, this is achieved through a saddle-point reformulation where the proximal operators required by the primal-dual algorithms can be computed in closed form. We illustrate the benefits of our approach on a graph transduction problem and on graph matching. (Collaboration with Achintya Kundu, Francis Bach, Chiranjib Bhattacharyya)

6.23. A Generic Approach for Escaping Saddle points

A central challenge to using first-order methods for optimizing nonconvex problems is the presence of saddle points. First-order methods often get stuck at saddle points, greatly deteriorating their performance. Typically, to escape from saddles one has to use second-order methods. However, most works on second-order methods rely extensively on expensive Hessian-based computations, making them impractical in large-scale settings. To tackle this challenge, we introduce in [32] a generic framework that minimizes Hessian based computations

while at the same time provably converging to second-order critical points. Our framework carefully alternates between a first-order and a second-order subroutine, using the latter only close to saddle points, and yields convergence results competitive to the state-of-the-art. Empirical results suggest that our strategy also enjoys a good practical performance. (Collaboration with Sashank Reddi, Manzil Zaheer, Suvrit Sra, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander Smola)

6.24. Tracking the gradients using the Hessian: A new look at variance reducing stochastic methods

The goal of [26] is to improve variance reducing stochastic methods through better control variates. We first propose a modification of SVRG which uses the Hessian to track gradients over time, rather than to recondition, increasing the correlation of the control variates and leading to faster theoretical convergence close to the optimum. We then propose accurate and computationally efficient approximations to the Hessian, both using a diagonal and a low-rank matrix. Finally, we demonstrate the effectiveness of our method on a wide range of problems.

6.25. Combinatorial Penalties: Which structures are preserved by convex relaxations?

In [28] we consider the homogeneous and the non-homogeneous convex relaxations for combinatorial penalty functions defined on support sets. Our study identifies key differences in the tightness of the resulting relaxations through the notion of the lower combinatorial envelope of a set-function along with new necessary conditions for support identification. We then propose a general adaptive estimator for convex monotone regularizers, and derive new sufficient conditions for support recovery in the asymptotic setting. (Collaboration with Marwa El Halabi, Francis Bach, Volkan Cevher)

6.26. On the Consistency of Ordinal Regression Methods

Many of the ordinal regression models that have been proposed in the literature can be seen as methods that minimize a convex surrogate of the zero-one, absolute, or squared loss functions. A key property that allows to study the statistical implications of such approximations is that of Fisher consistency. Fisher consistency is a desirable property for surrogate loss functions and implies that in the population setting, i.e., if the probability distribution that generates the data were available, then optimization of the surrogate would yield the best possible model. In [3] we will characterize the Fisher consistency of a rich family of surrogate loss functions used in the context of ordinal regression, including support vector ordinal regression, ORBoosting and least absolute deviation. We will see that, for a family of surrogate loss functions that subsumes support vector ordinal regression and ORBoosting, consistency can be fully characterized by the derivative of a real-valued function at zero, as happens for convex margin-based surrogates in binary classification. We also derive excess risk bounds for a surrogate of the absolute error that generalize existing risk bounds for binary classification. Finally, our analysis suggests a novel surrogate of the squared error loss. We compare this novel surrogate with competing approaches on 9 different datasets. Our method shows to be highly competitive in practice, outperforming the least squares loss on 7 out of 9 datasets.

6.27. Iterative hard clustering of features

In [34], we seek to group features in supervised learning problems by constraining the prediction vector coefficients to take only a small number of values. This problem includes non-convex constraints and is solved using projected gradient descent. We prove exact recovery results using restricted eigenvalue conditions. We then extend these results to combine sparsity and grouping constraints, and develop an efficient projection algorithm on the set of grouped and sparse vectors. Numerical experiments illustrate the performance of our algorithms on both synthetic and real data sets.

6.28. Asaga: Asynchronous Parallel Saga

In [9], we describe Asaga, an asynchronous parallel version of the incremental gradient algorithm Saga that enjoys fast linear convergence rates. We highlight 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 proposed "perturbed iterate" framework that resolves it. We thereby prove that Asaga 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 speedup as well as the hardware overhead.

6.29. Sparse Accelerated Exponential Weights

In [8], we consider the stochastic optimization problem where a convex function is minimized observing recursively the gradients. We introduce SAEW, a new procedure that accelerates exponential weights procedures with the slow rate $1/\sqrt{T}$ to procedures achieving the fast rate $1/T$. Under the strong convexity of the risk, we achieve the optimal rate of convergence for approximating sparse parameters in $\mathbb{R}^d$. The acceleration is achieved by using successive averaging steps in an online fashion. The procedure also produces sparse estimators thanks to additional hard threshold steps.

7. Bilateral Contracts and Grants with Industry

7.1. 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/.

7.2. Bilateral Grants with Industry

- A. d'Aspremont: AXA, "mécénat scientifique, chaire Havas-Dauphine", machine learning.
- F. Bach: Gift from Facebook AI Research.

8. Partnerships and Cooperations

8.1. National Initiatives

- A. d'Aspremont: IRIS, PSL "Science des données, données de la science".

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

- **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.

8.3. International Initiatives

8.3.1. 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: http://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).

8.4. International Research Visitors

8.4.1. Internships

- Marwa El Halabi, from Jan. until Apr. 2017, EPFL, Lausanne, Switzerland
- Jonathan Weed, from Mar. 2017 until May 2017, MIT, US
- Alfredo Zermini, from Mar 2017 until June 2017, University of Surrey, UK
- Billy Tang, visited from Sept. 2017 until Dec. 2017, University of Edimburgh, UK

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

- P. Germain and F. Bach: co-organization of NIPS workshop: "(Almost) 50 Shades of Bayesian Learning: PAC-Bayesian trends and insights" https://bguedj.github.io/nips2017/50shadesbayesian.html
- A. d'Aspremont: co-organization of the workshop: "Optimization and Statistical Learning", Les Houches, France

9.1.1.1. Member of the Organizing Committees

- F. Bach: Senior Area chair for NIPS 2017

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

- F. Bach: Action Editor, Journal of Machine Learning Research.
- 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.

9.1.3. Invited Talks

- F. Bach: Workshop on Shape, Images and Optimization, Muenster, Germany invited talk, February 2017
- F. Bach: SIAM conference on Optimization, Vancouver, Canada, invited tutorial, May 2017
- F. Bach: LCCC workshop on Large-Scale and Distributed Optimization, Lund, Sweden, invited talk, June 2017
- F. Bach: Summer school on Structured Regularization for High-Dimensional Data Analysis, Paris, invited talk, June 2017
- F. Bach: FOCM Barcelona, two invited talks in special sessions, July 2017
- F. Bach: European Signal Processing conference (EUSIPCO), Kos, Greece, keynote speaker, August 2017
- F. Bach: StatMathAppli 2017, Frejus, mini-course on optimization, September 2017
- F. Bach: 2017 ERNSI Workshop on System Identification, Lyon, invited plenary talk, September 2017
- F. Bach: New-York University, Data science seminar, October 2017
- F. Bach: Workshop on Generative models, parameter learning and sparsity, Cambridge, UK, invited talk, November 2017
- F. Bach: NIPS workshops, two invited talks, Long Beach, CA, December 2017
- A. d'Aspremont: "Regularized Nonlinear Acceleration"
  - GdR MOA, Bordeaux.
  - GdR MEGA, Paris.
  - SIAM OPtimization conference
  - Oxford computational math seminar
  - Alan Turing institute
- A. d'Aspremont: "Sharpness, Restart and Acceleration". Foundations of Computational Mathematics, Barcelona.
- P. Germain: "Generalization of the PAC-Bayesian Theory, and Applications to Semi-Supervised Learning", Modal Seminars, Lille, France, January 2017

- P. Germain: "Theory Driven Domain Adaptation Algorithm", Google Brain TechTalk, Mountain View (CA), USA, April 2017
- P. Gaillard: "Sparse acceleration of exponential weights"
  - Seminar of the SEQUEL project team, Lilles, February 2017
  - 49e Journées Françaises de Statistique, Avignon, Juin 2017
- P. Gaillard: "Obtaining sparse and fast convergence rates online under Bernstein condition", CWI-Inria Workshop, September 2017
- P. Gaillard: "Online nonparametric learning"
  - Cambridge Statistics Seminar, October 2017
  - Statistics Seminar of the University Aix-Marseille, December 2017

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master: A. d'Aspremont, "Optimization", 21h, M1, Ecole Normale Supérieure, France

Master: A. d'Aspremont, "Optimization", 21h, M2 (MVA), ENS Cachan, France

Master: F. Bach and P. Gaillard, "Apprentissage statistique", 35h, M1, Ecole Normale Supérieure, France.

Master: F. Bach (together with G. Obozinski), "Graphical models", 30h, M2 (MVA), ENS Cachan, France.

Master: F. Bach, "Optimisation et apprentissage statistique", 20h, M2 (Mathématiques de l'aléatoire), Université Paris-Sud, France.

Master: F. Pedregosa (together with Fajwel Fogel), "Introduction to scikit-learn", M2 (MASH), Université Paris-Dauphine, France.

9.2.2. Supervision

- PhD: Nicolas Flammarion, July 2017, co-directed by Alexandre d'Aspremont and Francis Bach.
- PhD: Aymeric Dieuleveut, September 2017, directed by Francis Bach.
- PhD: Christophe Dupuy, June 2017, directed by Francis Bach.
- PhD: Rafael Rezende, December 2017, Francis Bach, co-advised with Jean Ponce.
- PhD: Vincent Roulet, December 2017, directed by Alexandre d'Aspremont.
- PhD in progress: Damien Scieur, started September 2015, co-directed with Alexandre d'Aspremont and Francis Bach
- PhD in progress: Antoine Recanati, started September 2015, directed by Alexandre d'Aspremont
- PhD in progress: Anaël Bonneton, started December 2014, co-advised by Francis Bach, located in Agence nationale de la sécurité des systèmes d'information (ANSSI).
- PhD in progress: Dmitry Babichev, started September 2015, co-advised by Francis Bach and Anatoly Judistky (Univ. Grenoble).
- PhD in progress: Tatiana Shpakova, started September 2015, advised by Francis Bach.
- PhD in progress: Loucas Pillaud-Vivie, started September 2017, co-directed by Alessandro Rudi and Francis Bach
- PhD in progress: Margaux Brégère, started September 2017, co-advised by Pierre Gaillard, Gilles Stoltz and Yannig Goude (EDF R&D)

9.3. Popularization

- A. d'Aspremont: Paris Science et Data, PSL & Inria.

- A. d'Aspremont: Journée innovation défense
- P. Gaillard: testimony for EDF fellows day

10. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] J.-B. ALAYRAC, P. BOJANOWSKI, N. AGRAWAL, J. SIVIC, I. LAPTEV, S. LACOSTE-JULIEN.*Learning from narrated instruction videos*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", September 2017, vol. XX, https://hal.archives-ouvertes.fr/hal-01580630.

[2] F. BACH.*On the Equivalence between Kernel Quadrature Rules and Random Feature Expansions*, in "Journal of Machine Learning Research", 2017, vol. 18, n^{o} 21, p. 1-38, https://arxiv.org/abs/1502.06800 , https://hal.archives-ouvertes.fr/hal-01118276.

[3] F. PEDREGOSA, F. BACH, A. GRAMFORT.*On the Consistency of Ordinal Regression Methods*, in "Journal of Machine Learning Research", 2017, vol. 18, p. 1 - 35, https://arxiv.org/abs/1408.2327 , https://hal.inria.fr/hal-01054942.

[4] N. P. ROUGIER, K. HINSEN, F. ALEXANDRE, T. ARILDSEN, L. BARBA, F. C. Y. BENUREAU, C. T. BROWN, P. DE BUYL, O. CAGLAYAN, A. P. DAVISON, M. A. DELSUC, G. DETORAKIS, A. K. DIEM, D. DRIX, P. ENEL, B. GIRARD, O. GUEST, M. G. HALL, R. N. HENRIQUES, X. HINAUT, K. S. JARON, M. KHAMASSI, A. KLEIN, T. MANNINEN, P. MARCHESI, D. MCGLINN, C. METZNER, O. L. PETCHEY, H. E. PLESSER, T. POISOT, K. RAM, Y. RAM, E. ROESCH, C. ROSSANT, V. ROSTAMI, A. SHIFMAN, J. STACHELEK, M. STIMBERG, F. STOLLMEIER, F. VAGGI, G. VIEJO, J. VITAY, A. VOSTINAR, R. YURCHAK, T. ZITO.*Sustainable computational science: the ReScience initiative*, in "PeerJ Computer Science", 2017, https://arxiv.org/abs/1707.04393 - 8 pages, 1 figure, https://hal.inria.fr/hal-01592078.

Articles in Non Peer-Reviewed Journal

[5] A. MEURER, C. SMITH, M. PAPROCKI, O. ČERTÍK, S. KIRPICHEV, M. ROCKLIN, A. KUMAR, S. IVANOV, J. MOORE, S. SINGH, T. RATHNAYAKE, S. VIG, B. GRANGER, R. MULLER, F. BONAZZI, H. GUPTA, S. VATS, F. JOHANSSON, F. PEDREGOSA, M. CURRY, A. TERREL, Š. ROUČKA, A. SABOO, I. FERNANDO, S. KULAL, R. CIMRMAN, A. SCOPATZ.*SymPy: symbolic computing in Python*, in "PeerJ Comput.Sci.", 2017, vol. 3, e103 [*DOI :* 10.7717/PEERJ-CS.103], https://hal.archives-ouvertes.fr/hal-01645958.

Invited Conferences

[6] R. S. REZENDE, J. ZEPEDA, J. S. PONCE, F. S. BACH, P. PEREZ.*Kernel Square-Loss Exemplar Machines for Image Retrieval*, in "Computer Vision and Pattern Recognition 2017", Honolulu, United States, Computer vision and pattern recognition 2017, July 2017, https://hal.inria.fr/hal-01515224.

International Conferences with Proceedings

[7] A. BEAUGNON, P. CHIFFLIER, F. BACH.*ILAB: An Interactive Labelling Strategy for Intrusion Detection*, in "RAID 2017: Research in Attacks, Intrusions and Defenses", Atlanta, United States, September 2017, https://hal.archives-ouvertes.fr/hal-01636299.

[8] P. GAILLARD, O. WINTENBERGER.*Sparse Accelerated Exponential Weights*, in "20th International Conference on Artificial Intelligence and Statistics (AISTATS)", Fort Lauderdale, United States, April 2017, https://hal.archives-ouvertes.fr/hal-01376808.

[9] R. LEBLOND, F. PEDREGOSA, S. LACOSTE-JULIEN.*Asaga: Asynchronous Parallel Saga*, in "20th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017", Fort Lauderdale, Florida, United States, April 2017, https://hal.inria.fr/hal-01665255.

[10] F. PEDREGOSA, R. LEBLOND, S. LACOSTE-JULIEN.*Breaking the Nonsmooth Barrier: A Scalable Parallel Method for Composite Optimization*, in "NIPS 2017 - Thirty-First Annual Conference on Neural Information Processing Systems", Long Beach, United States, December 2017, p. 1-28, https://arxiv.org/abs/1707.06468 - Appears in Advances in Neural Information Processing Systems 30 (NIPS 2017), 28 pages, https://hal.inria.fr/hal-01638058.

[11] F. PEDREGOSA, R. LEBLOND, S. LACOSTE-JULIEN.*Breaking the Nonsmooth Barrier: A Scalable Parallel Method for Composite Optimization*, in "31st Conference on Neural Information Processing Systems (NIPS 2017)", Long Beach, California, United States, December 2017, https://hal.inria.fr/hal-01665260.

National Conferences with Proceeding

[12] A. GOYAL, E. MORVANT, P. GERMAIN.*Une borne PAC-Bayésienne en espérance et son extension à l'apprentissage multivues*, in "Conférence Francophone sur l'Apprentissage Automatique (CAp)", Grenoble, France, June 2017, https://hal.archives-ouvertes.fr/hal-01529219.

Conferences without Proceedings

[13] A. BEAUGNON, A. HUSSON.*Le Machine Learning confronté aux contraintes opérationnelles des systèmes de détection*, in "SSTIC 2017: Symposium sur la sécurité des technologies de l'information et des communications", Rennes, France, June 2017, p. 317-346, https://hal.archives-ouvertes.fr/hal-01636303.

[14] G. GIDEL, T. JEBARA, S. LACOSTE-JULIEN.*Frank-Wolfe Algorithms for Saddle Point Problems*, in "The 20th International Conference on Artificial Intelligence and Statistics", Fort Lauderdale, Florida, United States, April 2017, https://arxiv.org/abs/1610.07797 , https://hal.archives-ouvertes.fr/hal-01403348.

[15] A. GOYAL, E. MORVANT, P. GERMAIN, M.-R. AMINI.*PAC-Bayesian Analysis for a two-step Hierarchical Multiview Learning Approach*, in "European Conference on Machine Learning & Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD)", Skopje, Macedonia, September 2017, Long version : https://arxiv.org/abs/1606.07240, https://hal.archives-ouvertes.fr/hal-01546109.

[16] A. OSOKIN, F. BACH, S. LACOSTE-JULIEN.*On Structured Prediction Theory with Calibrated Convex Surrogate Losses*, in "The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)", Long Beach, United States, December 2017, https://arxiv.org/abs/1703.02403 , https://hal.archives-ouvertes.fr/hal-01611691.

[17] A. OSOKIN, A. CHESSEL, R. E. C. SALAS, F. VAGGI.*GANs for Biological Image Synthesis*, in "ICCV 2017 - IEEE International Conference on Computer Vision", Venice, Italy, October 2017, https://arxiv.org/abs/1708.04692 , https://hal.archives-ouvertes.fr/hal-01611692.

Scientific Books (or Scientific Book chapters)

[18] Y. GANIN, E. USTINOVA, H. AJAKAN, P. GERMAIN, H. LAROCHELLE, F. LAVIOLETTE, M. MARCHAND, V. LEMPITSKY.*Domain-Adversarial Training of Neural Networks*, in "Domain Adaptation in Computer Vision Applications", G. CSURKA (editor), Advances in Computer Vision and Pattern Recognition, Springer, September 2017 [*DOI :* 10.1007/978-3-319-58347-1], https://hal.archives-ouvertes.fr/hal-01624607.

Other Publications

[19] F. BACH.*Efficient Algorithms for Non-convex Isotonic Regression through Submodular Optimization*, July 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01569934.

[20] N. CESA-BIANCHI, P. GAILLARD, C. GENTILE, S. GERCHINOVITZ.*Algorithmic Chaining and the Role of Partial Feedback in Online Nonparametric Learning*, June 2017, https://arxiv.org/abs/1702.08211 - This document is the full version of an extended abstract accepted for presentation at COLT 2017., https://hal.archives-ouvertes.fr/hal-01476771.

[21] A. DIEULEVEUT, A. DURMUS, F. BACH.*Bridging the Gap between Constant Step Size Stochastic Gradient Descent and Markov Chains*, July 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01565514.

[22] A. DÉFOSSEZ, F. BACH.*AdaBatch: Efficient Gradient Aggregation Rules for Sequential and Parallel Stochastic Gradient Methods*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01620513.

[23] N. FLAMMARION, F. BACH.*Stochastic Composite Least-Squares Regression with convergence rate O(1/n)*, February 2017, working paper or preprint, https://hal.inria.fr/hal-01472867.

[24] P. GERMAIN, A. HABRARD, F. LAVIOLETTE, E. MORVANT.*PAC-Bayes and Domain Adaptation*, July 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01563152.

[25] R. M. GOWER, P. RICHTÁRIK.*Linearly Convergent Randomized Iterative Methods for Computing the Pseudoinverse*, January 2017, 28 pages, 10 figures, https://hal.inria.fr/hal-01430489.

[26] R. M. GOWER, N. L. ROUX, F. BACH.*Tracking the gradients using the Hessian: A new look at variance reducing stochastic methods*, November 2017, https://arxiv.org/abs/1710.07462 - 17 pages, 2 figures, 1 table [*DOI :* 10.07462], https://hal.archives-ouvertes.fr/hal-01652152.

[27] A. GOYAL, E. MORVANT, P. GERMAIN, M.-R. AMINI.*PAC-Bayesian Analysis for a two-step Hierarchical Multiview Learning Approach*, July 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01336260.

[28] M. E. HALABI, F. BACH, V. CEVHER.*Combinatorial Penalties: Which structures are preserved by convex relaxations?*, November 2017, working paper or preprint [*DOI :* 10.06273], https://hal.archives-ouvertes.fr/hal-01652151.

[29] A. KUNDU, F. BACH, C. BHATTACHARYYA.*Convex optimization over intersection of simple sets: improved convergence rate guarantees via an exact penalty approach*, November 2017, working paper or preprint [*DOI :* 10.06465], https://hal.archives-ouvertes.fr/hal-01652149.

[30] R. LEBLOND, J.-B. ALAYRAC, A. OSOKIN, S. LACOSTE-JULIEN.*SEARNN: Training RNNs with global-local losses*, December 2017, https://arxiv.org/abs/1706.04499 - 12 pages, https://hal.inria.fr/hal-01665263.

[31] L. PILLAUD-VIVIEN, A. RUDI, F. BACH.*Exponential convergence of testing error for stochastic gradient methods*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01662278.

[32] S. J. REDDI, M. ZAHEER, S. SRA, B. POCZOS, F. BACH, R. SALAKHUTDINOV, A. J. SMOLA.*A Generic Approach for Escaping Saddle points*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01652150.

[33] V. ROULET, A. D'ASPREMONT.*Sharpness, Restart and Acceleration*, February 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01474362.

[34] V. ROULET, F. FOGEL, A. D'ASPREMONT, F. BACH.*Iterative hard clustering of features*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01664964.

[35] K. SCAMAN, F. BACH, S. BUBECK, Y. T. LEE, L. MASSOULIÉ.*Optimal algorithms for smooth and strongly convex distributed optimization in networks*, February 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01478317.

[36] D. SCIEUR, A. D'ASPREMONT, F. BACH.*Nonlinear Acceleration of Stochastic Algorithms*, October 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01618379.

[37] D. SCIEUR, V. ROULET, F. BACH, A. D'ASPREMONT.*Integration Methods and Accelerated Optimization Algorithms*, February 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01474045.

[38] J. TANG, F. BACH, M. GOLBABAEE, M. E. DAVIES.*Structure-Adaptive, Variance-Reduced, and Accelerated Stochastic Optimization*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01658487.

[39] J. WEED, F. BACH.*Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance*, July 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01555307.

Team TAPDANCE

Theory and Practice of Nanoscale Computing Engines

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
Computational Biology

Table of contents

Team TAPDANCE

Creation of the Team: 2016 June 03

Keywords:

Computer Science and Digital Science:

A1.1.12. - Non-conventional architectures
A1.3. - Distributed Systems
A2.2. - Compilation
A8.1. - Discrete mathematics, combinatorics

Other Research Topics and Application Domains:

B5.3. - Nanotechnology
B5.6. - Robotic systems

1. Personnel

Research Scientists

Pierre Etienne Meunier [Inria, Starting Research Position]
Damien Woods [Inria, Advanced Research Position, Team Leader]

External Collaborator

David Doty [UC Davis, from Sep 2017; Also visitor from Jun 2017 until Jul 2017]

Administrative Assistant

Helene Milome [Inria]

2. Overall Objectives

2.1. Overall Objectives

In biological systems we see extraordinarily sophisticated growth processes, where molecular self-assembly is combined with active molecular components. Indeed, biological systems consume energy (e.g. ATP) and exhibit phenomena such as rapid growth in cell size and numbers, reconfiguration of internal components, molecular motors that push and pull large structures around, as well as molecular complexes, cells and whole organs that actively respond to the environment. Computer science gives us tools and methodologies to think about and design systems with large number of interacting components. Our goal is to bring these ideas together to design computational molecular systems.

The work of the TAPDANCE team is concerned with the theory and practice of active DNA nanostructures that build structures and compute, all at the nanoscale.

We focus on:

1. Proposing and analysing models of computation for nanoscale biomolecular systems. This includes finding new models for the systems we wish to build, proving theorems (e.g. about their computational power), as well as developing the theory of existing models.
2. Implementing these models in the wet-lab, primarily using DNA.
3. Software to design these kinds of systems (e.g. DNA sequence design) as well as coarse-grained molecular models for system analysis. Software tools are one of the main ways we bridge the gap between theory and experiments.

3. Research Program

3.1. Recent results and ongoing work

Theoretical work by Meunier and Woods "The non-cooperative tile assembly model is not intrinsically universal or capable of bounded Turing machine simulation" was published in 2017 at the conference STOC, Montreal, Canada, and later presented as a poster and chosen for a short oral talk at DNA23, UT Austin, TX, USA, (presentations by Meunier). In this model, called the noncooperative (or temperature 1) abstract Tile Assembly Model, square tiles assemble into structures in the discrete plane ($\mathbb{Z}^2$) where each tile binds to a growing structure if one of its 4 coloured edges matches the colour of some available edge on the growing structure. It has been conjectured since 2000 that this model is not capable of computation or other sophisticated forms of growth. We show two results. One of our results states that time-bounded Turing machine computation is impossible in this model if we require the simulation to occur in a bounded rectangle in the plane. This result has a short proof that essentially follows from our other main result (with a more involved prof) which states that this model is not "intrinsically universal". This latter result means that there is no single tileset in this model that can simulate any instance of the model, answering a question from [Meunier, Patitz, Summers, Theyssier, Winslow, Woods; SODA 2014] and contrasting a result [Doty, Lutz, Patitz, Schweller, Summers, Woods; FOCS 2012] for the more general cooperative (temperature 2) model.

A number of popular models of computation for molecular computing are kinetic in nature: the rules of the systems describe how system evolves, step-by-step over time. However, such models allow one to program structures and behaviours that contradict how molecular systems would behave on long enough (e.g. infinite) time-scales. Recently, Doty, Rogers, Soloveichik, Thachuk and Woods proposed a thermodynamic based model called Thermodynamic Binding Networks. One programs the model by specifying a multiset of molecules, and then the 'output' is defined as the (or a, if many) multiset of polymers deemed most 'thermodynamically stable' out of all possible multisets of polymers definable in the model. In order to clarify the roles of fundamental thermodynamic concepts in molecular computing, the model makes a number of simplifying assumptions including a lack of geometry and (essentially) an infinite biasing of enthalpy over entropy. This model was published at the DNA23 conference at the University of Texas at Austin in 2017, along with some results: The authors show how to program the model to evaluate Boolean AND/OR formulas, and how to efficiently self-assemble simple structures (a binary counter), Also, limitations (upper bounds) on the size of objects self-assemblable are shown. It is hoped this work will lead to new ways of thinking about computation with molecules.

In experimental work, Thubagere, Li, Johnson, Chen, Doroudi, Lee, Izatt, Wittman, Srinivas, Woods, Winfree and Qian, implemented a molecular walker made out of DNA. The walker randomly around on a 2D nanoscale testing ground, picking up one of two types of cargos (also DNA molecules) and dropping them off at specific goal locations. The work shows the power of simple randomised algorithms in molecular systems. This work was led by Thubagary and Qian at Caltech; Woods' contribution occurred while at Caltech (before joining Inria). The work was published in Science.

Experimental work by Woods and collaborators Doty, Myhrvold, Hui, Zhou, Yin and Winfree has focused on experimentally implementing a wide class of Boolean circuits of a certain form. Experiments were mostly carried out at Caltech, with data analysis and paper writeup being carried out a Inria, UC Davis and Caltech. A publication is in preparation.

ENS student Tristan Stérin is leading theoretical work on analysing the computational power of the previously mentioned iterated Boolean circuit model of Woods, Doty, Myhrvold, Hui, Zhou, Yin, Winfree. Preliminary results were presented as a poster at DNA23 (UT Austin, TX, USA) where Tristan won "DNA23 best poster award". The work is in preparation for a conference publication.

There are a number of ongoing projects along the lines of topics above in Overall Objectives.

4. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Awards

Tristan Stérin won a best poster award at the conference DNA 23.

5. New Software and Platforms

5.1. Sanakirja

KEYWORD: Databases
FUNCTIONAL DESCRIPTION: Sanakirja is a fully transactional (all operations are atomic) key-value dictionary stored in a file (usable a the low-level layer of a more full-featured database engine), with a zero-copy fork operation (fork is in time and space O(log n), where n is the number of keys in the file). This project is written in Rust.

- Contact: Pierre-Etienne Meunier
- URL: https://nest.pijul.com/pijul_org/sanakirja

5.2. Thrussh

KEYWORD: Security
FUNCTIONAL DESCRIPTION: Purely asynchronous SSH library in Rust.

- Contact: Pierre-Etienne Meunier

5.3. Pijul

KEYWORDS: Data structures - Distributed systems
FUNCTIONAL DESCRIPTION: Patch-based distributed version control system using category theory. This solves a number of problems in other systems (such as Git), like:

- Scaling up to giant repositories (as used by Facebook, Google and Mozilla). - Being easy to understand and use, because based on a solid theory.

- Contact: Pierre-Etienne Meunier
- URL: https://pijul.org

5.4. SeqDesign

KEYWORDS: Chemistry - Molecular simulation
FUNCTIONAL DESCRIPTION: SeqDesign is a free energy calculation tool for DNA secondary structures. We use it to design sequences capable of self-assembling in a designed way.

- Contact: Pierre-Etienne Meunier

6. Partnerships and Cooperations

6.1. European Initiatives

6.1.1. FP7 & H2020 Projects

Woods applied for an ERC Consolidator award. The application was successful and begins in 2018.

6.2. International Research Visitors

6.2.1. Visits of International Scientists

David Doty (UC Davis) visited the team several times in 2017.

7. Dissemination

7.1. Promoting Scientific Activities

7.1.1. Scientific Events Selection

7.1.1.1. Reviewer

Both Damien Woods and Pierre-Étienne Meunier blind-reviewed papers for a number of computer science theory conferences (and do not wish to disclose further information related to blind peer review).

7.1.2. Journal

7.1.2.1. Guest Editor for Journal Special Issue

Damien Woods was guest editor (along with Yannick Rondelez, CNRS & ESPCI) of a special issue of the journal *Natural Computing* dedicated to the conference DNA22 (2016). The special issue will appear in 2018.

7.1.2.2. Reviewer - Reviewing Activities

Both Damien Woods and Pierre-Étienne Meunier reviewed papers for a variety of scientific journals (and do not wish to disclose further information related to blind peer review).

7.1.3. Invited Talks

- Woods. Centre for Research and Interdisciplinarity (CRI), Masterclass, Paris, France, 2017.
- Woods. Paris Sciences et Lettres (PSL institute), group on Origins and Conditions of Appearance of Life, Observatoire de Paris, France. 2017.
- Meunier, RustFest 2017 in Kiev, Ukraine, April 2017.

7.1.4. Contributed Talks (with no associated publication)

- Woods (speaker). FNANO 2017: 14th Annual Conference on Foundations on Nanoscience: Self-Assembled Architectures and Devices. Snowbird, UT. [Woods, Doty, Myhrvold, Hui, Zhou, Yin and Winfree. *Iterated circuit computation by DNA self-assembly in a field-programmable gate array architecture*]
- Woods. DNA23: The 23rd International Conference on DNA Computing and Molecular Programming, 2017. UT Ausin, Texas [Woods, Doty, Myhrvold, Hui, Zhou, Yin and Winfree. *Iterated circuit computation by DNA self-assembly in a field-programmable gate array architecture*]
- Meunier (speaker). DNA 23: The 23rd International Conference on DNA Computing and Molecular Programming. UT Austin Texas, USA. [Meunier, Woods. *The non-cooperative tile assembly model is not intrinsically universal or capable of bounded Turing machine simulation*]

7.1.5. Conference Programme Committees

- Woods. CiE 2017. Computability in Europe. Turku, Finland.
- Woods. DNA23: The 23rd International Conference on DNA Computing and Molecular Programming, 2017. UT Ausin, Texas
- Woods. UCNC 2017. Unconventional Computation and Natural Computation. Fayetteville, AR.

7.2. Teaching - Supervision - Juries

7.2.1. Teaching

Woods taught a 1-week course on molecular computing to Masters students at ENS Lyon (with N. Schabanel, Y. Rondelez, C. Moskalenko, L. Bellon). Course title: ER02: Molecular programming: Theory & Wet Lab Nano-Scale Computation. https://www.irif.fr/~nschaban/2017-ER02/

8. Bibliography

Major publications by the team in recent years

[1] A. THUBAGERE, W. LI, R. JOHNSON, Z. CHEN, S. DOROUDI, Y. LEE, S. W. G. IZATT, N. SRINIVAS, D. WOODS, E. WINFREE, L. QIAN.*A cargo-sorting DNA robot*, in "Science", 2017, vol. 357, n^{o} 6356, eaan6558.

Publications of the year

Articles in International Peer-Reviewed Journal

[2] L. KARI, S. KOPECKI, P.-É. MEUNIER, M. J. PATITZ, S. SEKI.*Binary Pattern Tile Set Synthesis Is NP-Hard*, in "Algorithmica", May 2017, vol. 78, n^{o} 1, p. 1 - 46 [*DOI :* 10.1007/s00453-016-0154-7], https://hal.inria.fr/hal-01659779.

International Conferences with Proceedings

[3] D. WOODS, D. DOTY, T. A. ROGERS, D. SOLOVEICHIK, C. THACHUK.*Thermodynamic Binding Networks*, in "DNA 2017 - The 23rd International Conference on DNA Computing and Molecular Programming", Austin, United States, DNA 2017: DNA Computing and Molecular Programming, September 2017, vol. 10467, p. 249-266, https://arxiv.org/abs/1709.07922 [*DOI :* 10.1007/978-3-319-66799-7_16], https://hal.inria.fr/hal-01662285.

[4] D. WOODS, P.-É. MEUNIER.*The non-cooperative tile assembly model is not intrinsically universal or capable of bounded Turing machine simulation*, in "STOC 2017 - Theory Fest 49th Annual ACM Symposium on the Theory of Computing", Montréal, Canada, STOC 2017 Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, ACM, June 2017, p. 328-341, https://arxiv.org/abs/1702.00353 - This submission is an extended (arxiv) version of a STOC 2017 conference paper, with arxiv identifier arxiv:1702.00353 (or version arxiv:1702.00353v2 at the time of submission on HAL) [*DOI :* 10.1145/3055399.3055446], https://hal.inria.fr/hal-01659730.

Other Publications

[5] C. GEARY, P.-É. MEUNIER, N. SCHABANEL, S. SEKI.*Folding Turing is hard but feasible*, December 2017, https://arxiv.org/abs/1508.00510 - working paper or preprint, https://hal.inria.fr/hal-01659777.

Team Valda

Value from Data

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
Data and Knowledge Representation and Processing

Table of contents

Team Valda

Creation of the Team: 2016 December 01, updated into Project-Team: 2018 January 01

VALDA has integrated members of the Inria DAHU project-team in 2017. Their relevant activity has been integrated in this activity report.

Keywords:

Computer Science and Digital Science:

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
A3.1.9. - Database
A3.2.2. - Knowledge extraction, cleaning
A3.2.3. - Inference
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
A9.1. - Knowledge

Other Research Topics and Application Domains:

B6.3.1. - Web
B6.3.4. - Social Networks
B6.5. - Information systems
B9.5.5. - Sociology
B9.5.10. - Digital humanities
B9.7.2. - Open data
B9.8. - Privacy
B9.10. - Ethics

1. Personnel

Research Scientists

Serge Abiteboul [Inria, Senior Researcher, HDR]
Luc Segoufin [Inria, Senior Researcher, from Sep 2017, HDR]

Faculty Member

Pierre Senellart [Team leader, École normale supérieure, Professor, HDR]

External Collaborator

Yann Ramusat [Student at École normale supérieure on a long-term project, from Sep 2017]

PhD Students

Julien Grange [École normale supérieure, from Sep 2017]

Miyoung Han [Institut Telecom ex GET Groupe des Ecoles des Télécommunications]
Quentin Lobbe [Institut Telecom ex GET Groupe des Ecoles des Télécommunications]
Mikael Monet [Institut Telecom ex GET Groupe des Ecoles des Télécommunications]
David Montoya [ENGIE, until March 2017]
Karima Rafes [BorderCloud]
Alexandre Vigny [Univ Denis Diderot, from Sep 2017]
Su Yang [Ecole Nationale Supérieure des Mines de Paris, until Apr 2017]

Visiting Scientist

Victor Vianu [UCSD & École normale supérieure, from Jul 2017]

Administrative Assistants

Linday Polienor [until June 2017]
Sandrine Vergès [from September 2017]

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 [2], [5] 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 [33][3], and possibly distributed [78] 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.

[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*.

These problems have already been studied individually and have led to techniques such as *query rewriting* [59] or *distributed query optimization* [65].

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 [50] or aggregated search [77]. 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 [75] (train an agent to accomplish a task in a partially known world based on rewards obtained) and of active learning [72] (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 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 [2]. Since these early developments, logic has also been successfully used to capture a large variety of query modes, such as data aggregation [64], recursive queries (Datalog), or querying of XML databases [5]. 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 [76] 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 [35], queries, expressed in temporal logics, can often by compiled to automata; in graph databases [41], 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 [68], or for more complex languages to data tree automata[7]. Another reason of the importance of automata theory, and tree automata in particular, comes from Courcelle's results [48] 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* [2]. 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 [46].

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 [32].

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 [2]. 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* [61] 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 [57], and to a probabilistic formalism by assigning independent probabilities to the Boolean events [58]. These two extensions form the

basis of modern provenance and probability management, subsuming in a large way previous works [47], [42]. 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 [74] [63], [38].

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 [70], crowdsourcing [40], focused crawling [56], or automatic database tuning [43] critically rely on machine learning techniques, such as classification [60], probabilistic models [55], or reinforcement learning [75].

Machine learning is also a rich source of complex data management problems: thus, the probabilities produced by a conditional random field [66] 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* [72], 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 [71] consider more realistic costs. Also, oracles are usually assumed to be perfect with only a few exceptions [51]. 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, 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.

[0] http://db.ucsd.edu/WAVE/default.html
[0] http://researcher.watson.ibm.com/researcher/view_group.php?id=2501

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 [46], [62],[6], 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 [46], or query rewriting using views [6].

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 [69]. We have contributed a lot to this fields by providing several such indexes [62] 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.

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; 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 [63], [36][12], provenance management [4], incomplete information [37], and uncertainty analysis and propagation in conflicting datasets [53], [34]. 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 [57], extensions to queries with negation [54], aggregation [39], or recursion [67]?

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 [49] on probabilistic query evaluation with the instance-based one of [4] [38].

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, [1] 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) [1].

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 [70].

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 [73] 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 [3] in a broad sense: semi-structured, structured, or unstructured content extracted from Web databases [70]; knowledge bases from the Semantic Web [73]; social networks [9]; Web archives and Web crawls [52]; Web applications and deep Web databases [45]; crowdsourcing platforms [40]. We intend to continue using Web data as a natural application domain for the research within Valda when relevant. For instance [44], 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: ProvSQL becomes usable for a large range of queries. Support for semirings and m-semirings is present, support for probability computation has been added through a variety of techniques, including knowledge compilation, support for where-provenance is currently being implemented.

- Participants: Pierre Senellart and Yann Ramusat
- Contact: Pierre Senellart
- Publication: Provenance and Probabilities in Relational Databases: From Theory to Practice
- URL: https://github.com/PierreSenellart/provsql

5.2. Thymeflow

KEYWORD: Personal information
FUNCTIONAL DESCRIPTION: ThymeFlow allows in particular the development of plugins for both interacting with existing Web sources and presenting users with rich interfaces and query facilities over their personal information. A preliminary version of ThymeFlow tools has also been deployed on the Cozy Cloud personal cloud system. The model allows the open-source community to contribute individual plugins while we focus on providing users with useful ways to exploit their personal information.
NEWS OF THE YEAR: Minor maintenance.

- Participants: David Montoya, Pierre Senellart, Serge Abiteboul and Su Yang
- Partner: ENGIE
- Contact: Pierre Senellart
- Publication: Personal Knowledge Base Systems
- URL: https://github.com/thymeflow/thymeflow/

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: Ability to specify a sectioning counter, Compilation fix of proofs-ketch in inline mode
NEWS OF THE YEAR: Overall software maintenance. Support for more document classes. Some new features.

- Participant: Pierre Senellart
- Contact: Pierre Senellart
- URL: https://github.com/PierreSenellart/apxproof

6. New Results

6.1. Enumeration of Query Results

In many applications the output of a query may have a huge size and computing all the answers may already consume too many of the allowed resources. In this case it may be appropriate to first output a small subset of the answers and then, on demand, output a subsequent small numbers of answers and so on until all possible answers have been exhausted. To make this even more attractive it is preferable to be able to minimize the time necessary to output the first answers and, from a given set of answers, also minimize the time necessary to output the next set of answers - this second time interval is known as the *delay*. We have shown that this was doable with a almost linear preprocessing time and constant enumeration delay for first-order queries over structures having local bounded expansion [22].

6.2. Ethical Data Management

Issues of responsible data analysis and use are coming to the forefront of the discourse in data science research and practice [14]. The research has been focused on analyzing the fairness, accountability and transparency (FAT) properties of specific algorithms and their outputs. Although these issues are most apparent in the social sciences where fairness is interpreted in terms of the distribution of resources across protected groups, management of bias in source data affects a variety of fields. Consider climate change studies that require representative data from geographically diverse regions, or supply chain analyses that require data that represents the diversity of products and customers. In a paper [23], we argue that FAT properties must be considered as database system issues, further upstream in the data science lifecycle: bias in source data goes unnoticed, and bias may be introduced during pre-processing (fairness), spurious correlations lead to reproducibility problems (accountability), and assumptions made during pre-processing have invisible but significant effects on decisions (transparency). As machine learning methods continue to be applied broadly by non-experts, the potential for misuse increases. There is a need for a data sharing and collaborative analytics platform with features to encourage (and in some cases, enforce) best practices at all stages of the data science lifecycle. We describe features of such a platform, which we term Fides, in the context of of urban analytics, outlining a systems research agenda in responsible data science.

6.3. Structure and Tractability of Uncertain Data

A major part of the work conducted in Valda has been to study the connections between tractability and structure in databases, in particular uncertain databases.

In a first line of work, we have investigated incompleteness related to order. In [18], we have introduced a query language for order-incomplete data, based on the positive relational algebra with order-aware accumulation. We have used partial orders to represent order-incomplete data, and studied possible and certain answers for queries in this context, showing these problems are respectively NP-complete and coNP-complete, but identifying tractable cases depending on query operators and the structure of input partial orders. In [16], we consider a different setting where some partial order is known, but actual values are unknown. Our work is the first to propose a principled scheme to derive the value distributions and expected values of unknown items in this setting, with the goal of computing estimated top-k results by interpolating the unknown values from the known ones. We have studied the complexity of this general task, and show tight complexity bounds, proving that the problem is intractable, but can be tractably approximated. We have also isolated structure-based restrictions that allow for a PTIME solution.

In [17], we have investigated parameterizations of both database instances and queries that make query evaluation fixed-parameter tractable in combined complexity, first in a setting without uncertainty. For this, we have introduced a new Datalog fragment with stratified negation, intensional-clique-guarded Datalog (ICG-Datalog), with linear-time evaluation on structures of bounded treewidth for programs of bounded rule size. Our result is shown by compiling to alternating two-way automata, whose semantics is defined via cyclic provenance circuits (cycluits) that can be tractably evaluated. Finally, we move to the probabilistic setting and have shown that probabilistic query evaluation remains intractable in combined complexity under this parameterization.

Finally, a last line of work concerns efficient queries over probabilistic graphs. In a first theoretical work [19], we have studied the combined complexity of conjunctive query evaluation on probabilistic graphs, which can be alternatively phrased as a probabilistic version of the graph homomorphism problem. We have shown that the complexity landscape is surprisingly rich, using a variety of technical tools. In a more practical work [12], we have proposed indexing techniques and algorithms to evaluate source-to-target queries in probabilistic graphs, by exploiting their structure. We have shown that these significantly enhance the accuracy and efficiency of existing query evaluation approaches on probabilistic graphs.

7. Partnerships and Cooperations

7.1. Regional Initiatives

Valda has obtained a 10k€ budget from ENS in 2017, as a start-up grant from the team (*Action Concertée Incitative*).

Inria established a bilateral contract with the Centre – Val de Loire region, for the expertise and audit of a research project by Pierre Senellart. Because of delays due to the company being audited, the expertise is still in progress.

7.2. National Initiatives

7.2.1. ANR

Valda has been part of one ANR project in 2017 (Headwork, budget managed by Inria), together with IRISA (DRUID team, coordinator), Inria Lille (LINKS & SPIRAL), and Inria Rennes (SUMO), and two application partners: MNHN (Cesco) and FouleFactory. The topic is workflows for crowdsourcing. See http://headwork.gforge.inria.fr/.

In addition, another project (BioQOP, budget managed by ENS) will start in January 2018, with Morpho and GREYC, on the optimization of queries for privacy-aware biometric data management

7.3. International Initiatives

7.3.1. Informal International Partners

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, and Georg Gottlob

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 6 months within Valda: three months employed by Inria and three months as an ENS invited professor.

7.4.1.1. Internships

Deabrota Basu, PhD student at National University of Singapore, stayed 2.5 months within Valda, to work with Pierre Senellart.

7.4.2. Visits to International Teams

7.4.2.1. Research Stays Abroad

- Pierre Senellart has spent around two months at the University of Edinburgh, collaborating with Peter Buneman and Leonid Libkin.
- Pierre Senellart has spent a cumulated time of more than one month at National University of Singapore, co-advising Debabrota Basu, PhD student working under the co-supervision of Stéphane Bressan.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events Organisation

8.1.1.1. Member of the Organizing Committees

- Serge Abiteboul, organization of Personal Analytics & Privacy workshop, joint with ECML-PKDD 2017, Skopje, Macedonia
- Serge Abiteboul, scientific organization of colloquium on *La communauté scientifique face au renseignement*, École militaire, Parsi, France
- Serge Abiteboul, organization of colloquium on *Les enjeux scientifiques de l'éthique du numérique*, Académie des sciences
- Pierre Senellart, organization of ParisBD 2017, Télécom ParisTech, Paris, France
- Pierre Senellart, co-organizer of ACM-ICPC Southwestern Europe 2017 competition

8.1.2. Scientific Events Selection

8.1.2.1. Chair of Conference Program Committees

- Pierre Senellart, BDA 2017 (French conference on data management)
- Pierre Senellart, WebDB workshop, joint with SIGMOD 2017

8.1.2.2. Member of the Conference Program Committees

- Pierre Senellart, *Gems of PODS 2017* committee
- Pierre Senellart, SIGMOD 2017 (*distinguished PC member*), ICDT 2017, EDBT 2018

8.1.3. Journal

8.1.3.1. Reviewer - Reviewing Activities

- Pierre Senellart, *Journal of the ACM*, *VLDB Journal*, *Artificial Intelligence*

8.1.4. Invited Talks

- Serge Abiteboul, keynote at PPDP-LOPSTR, Namur, Belgium
- Serge Abiteboul, keynote at ETAPS, Uppsala, Sweden
- Serge Abiteboul, keynote at Law & Big Data Conference, Paris, 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.

8.1.6. Scientific Expertise

- Pierre Senellart, ANR, NSF

8.1.7. Research Administration

- Serge Abiteboul was the president of the Dune jury (*Développement d'universités numériques expérimentales*)
- Serge Abiteboul participated in the NCU jury (*nouveaux cursus à l'université*)
- Serge Abiteboul contributed to the report on *Éthique de la recherche en apprentissage machine* of Cerna-Allistene
- Serge Abiteboul is co-chair of the "Committee on Gender Equality and Equal Opportunities" of Inria.
- 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 vice-director of the DI ENS laboratory, joint between ENS, CNRS, and Inria.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Licence: Pierre Senellart, *Databases*, 54 heqTD, L3, École normale supérieure

Licence: Pierre Senellart, *Algorithms*, 18 heqTD, L3, École normale supérieure

Master: Serge Abiteboul & Pierre Senellart, *Web data management*, 36 heqTD, M2, MPRI

Master: Luc Segoufin, *Logic, descriptive complexity and database theory*, 36 heqTD, M2, MPRI

Pierre Senellart has various teaching responsibilities (L3 internships, M2 internships, M2 administration) at ENS.

Serge Abiteboul proposed with Benjamin Nguyen and Philippe Rigaux a second session of the Mooc "Bases de données relationnelles: comprendre pour maîtriser" (FUN). He proposed with Julia Stoyanovich a course on "Ethical data management" at the EDBT Summer School, Genova, 2017.

8.2.2. Supervision

PhD : David Montoya, *Une base de connaissance personnelle intégrant les données d'un utilisateur et une chronologie de ses activités*, Université Paris-Saclay, 6 March 2017, Serge Abiteboul & Pierre Senellart

PhD in progress: Debabrota Basu, *Reinforcement learning applications to data management problems*, started in 2015, Stéphane Bressan & Pierre Senellart

PhD in progress: Julien Grange, *Graph properties: order and arithmetic in predicate logics*, started in 2017, Luc Segoufin

PhD in progress: Miyoung Han, *Learning approaches to dynamic data management*, started in 2015, Pierre Senellart

PhD in progress: Quentin Lobbé, *Diachronic analysis of diaspora communities through web archives enrichment*, started in 2015, Pierre Senellart & Dana Diminescu

PhD in progress: Mikaël Monet, *Efficient querying of large uncertain graphs by exploiting their structure*, started in 2015, Pierre Senellart & Antoine Amarilli

PhD in progress: Karima Rafes, *Security and management of personal data in the Web of things*, started in 2015, Serge Abiteboul & Sarah Cohen-Boulakia

PhD in progress: Alexandre Vigny, *Query enumeration on nowhere-dense graphs*, started in 2015, Luc Segoufin & Arnaud Durand

8.2.3. Juries

- PhD Paul Lagrée, October 2017, Université Paris-Saclay, Pierre Senellart

8.3. Popularization

Serge Abiteboul is involved in several popular science activities. He founded and animates the blog http://binaire.blog.lemonde.fr/ on computer science. He was the scientific curator (*commissaire scientifique*) of the exhibition "Terra Data" at the Cité des Sciences. He published two scientific popularization books in 2017, "Le temps des algorithmes" [26], with Gilles Dowek, and "Terra data" [27], with Valéeie Peugeot.

Serge Abiteboul is the president of the strategic committee of the Blaise Pascal foundation for scientific mediation.

9. Bibliography

Major publications by the team in recent years

[1] 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.

[2] S. ABITEBOUL, R. HULL, V. VIANU.*Foundations of Databases*, Addison-Wesley, 1995, http://webdam.inria.fr/Alice/.

[3] S. ABITEBOUL, I. MANOLESCU, P. RIGAUX, M. ROUSSET, P. SENELLART.*Web Data Management*, Cambridge University Press, 2011, http://webdam.inria.fr/Jorge.

[4] 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.

[5] 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.

[6] 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.

[7] 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^{o} 2, https://doi.org/10.2168/LMCS-12(2:3)2016.

[8] W. KAZANA, L. SEGOUFIN.*Enumeration of monadic second-order queries on trees*, in "ACM Trans. Comput. Log.", 2013, vol. 14, n^{o} 4, p. 25:1–25:12, http://doi.acm.org/10.1145/2528928.

[9] 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.

[10] 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.

Publications of the year

Articles in International Peer-Reviewed Journal

[11] D. FIGUEIRA, L. SEGOUFIN.*Bottom-up automata on data trees and vertical XPath*, in "Logical Methods in Computer Science", 2017, p. 1-40, https://arxiv.org/abs/1710.08748 [*DOI :* 10.08748], https://hal.inria.fr/hal-01631219.

[12] 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.

[13] P. SENELLART.*Provenance and Probabilities in Relational Databases: From Theory to Practice*, in "SIGMOD record", December 2017, p. 1-11, https://hal.inria.fr/hal-01672566.

Invited Conferences

[14] S. ABITEBOUL.*Issues in Ethical Data Management - Extended Abstract*, in "PPDP 2017 - 19th International Symposium on Principles and Practice of Declarative Programming", Namur, Belgium, October 2017, https://hal.inria.fr/hal-01621687.

[15] S. ABITEBOUL, D. MONTOYA.*Personal Knowledge Base Systems*, in "PAP 2017, Personal analytics and privacy", Skopje, Macedonia, September 2017, https://hal.inria.fr/hal-01592601.

International Conferences with Proceedings

[16] A. AMARILLI, Y. AMSTERDAMER, T. MILO, P. SENELLART.*Top-k Querying of Unknown Values under Order Constraints*, in "ICDT 2017 - International Conference on Database Theory", Venice, Italy, March 2017 [*DOI :* 10.4230/LIPICS.ICDT.2017.5], https://hal.inria.fr/hal-01439295.

[17] A. AMARILLI, P. BOURHIS, M. MONET, P. SENELLART.*Combined Tractability of Query Evaluation via Tree Automata and Cycluits*, in "ICDT 2017 - International Conference on Database Theory", Venice, Italy, March 2017 [*DOI :* 10.4230/LIPICS.ICDT.2017.6], https://hal.inria.fr/hal-01439294.

[18] A. AMARILLI, M. LAMINE BA, D. DEUTCH, P. SENELLART.*Possible and Certain Answers for Queries over Order-Incomplete Data*, in "24th International Symposium on Temporal Representation and Reasoning (TIME 2017)", Mons, Belgium, S. SCHEWE, T. SCHNEIDER, J. WIJSEN (editors), Schloss Dagstuhl, October 2017, vol. 90, p. 4:1-4:19, https://arxiv.org/abs/1707.07222 [*DOI :* 10.4230/LIPICS.TIME.2017.4], https://hal.inria.fr/hal-01570603.

[19] A. AMARILLI, M. MONET, P. SENELLART.*Conjunctive Queries on Probabilistic Graphs: Combined Complexity*, in "Principles of Database Systems (PODS)", Chicago, United States, May 2017, https://arxiv.org/abs/1703.03201 [*DOI :* 10.1145/3034786.3056121], https://hal.inria.fr/hal-01486634.

[20] M. CROCHEMORE, A. HELIOU, G. KUCHEROV, L. MOUCHARD, S. P. PISSIS, Y. RAMUSAT.*Minimal absent words in a sliding window & applications to on-line pattern matching*, in "FCT 2017", Bordeaux, France, Lecture Notes in Computer Science, Springer, September 2017, https://hal.archives-ouvertes.fr/hal-01569264.

[21] O. SAVKOVIĆ, E. KHARLAMOV, W. NUTT, P. SENELLART.*Towards Approximating Incomplete Queries over Partially Complete Databases (Extended Abstract)*, in "AMW", Montevideo, Uruguay, AMW 2017 - 11th Alberto Mendelzon International Workshop on Foundations of Data Management Montevideo, Uruguay June 5 – 9, 2017, June 2017, https://hal.inria.fr/hal-01586884.

[22] L. SEGOUFIN, A. VIGNY.*Constant Delay Enumeration for FO Queries over Databases with Local Bounded Expansion*, in "ICDT", Venise, Italy, March 2017, https://hal.inria.fr/hal-01589303.

[23] J. STOYANOVICH, B. HOWE, S. ABITEBOUL, G. MIKLAU, A. SAHUGUET, G. WEIKUM.*Fides: Towards a Platform for Responsible Data Science*, in "SSDBM'17 - 29th International Conference on Scientific and Statistical Database Management", Chicago, United States, June 2017 [*DOI :* 10.1145/3085504.3085530], https://hal.inria.fr/hal-01522418.

National Conferences with Proceeding

[24] K. RAFES, S. COHEN-BOULAKIA, S. ABITEBOUL.*Une autocomplétion générique de SPARQL dans un contexte multi-services*, in "BDA 2017 - 33ème conférence sur la «Gestion de Données — Principes, Technologies et Applications»", Nancy, France, November 2017, https://hal.inria.fr/hal-01627760.

Books or Proceedings Editing

[25] A. MELIOU, P. SENELLART (editors). *Proceedings of the 20th International Workshop on the Web and Databases, WebDB 2017*, May 2017, https://hal.inria.fr/hal-01523772.

Scientific Popularization

[26] S. ABITEBOUL, G. DOWEK.*Le temps des algorithmes*, Editions Le Pommier, 2017, 192, https://hal.inria.fr/hal-01502505.

[27] S. ABITEBOUL, V. PEUGEOT.*Terra Data : Qu'allons-nous faire des données numériques ?*, Editions Le Pommier, 2017, 320, https://hal.inria.fr/hal-01502512.

[28] P. SENELLART.*Archivage du Web*, in "Les Big Data à découvert", CNRS Éditions, March 2017, https://hal.inria.fr/hal-01497800.

Other Publications

[29] A. AMARILLI, Y. AMSTERDAMER, T. MILO, P. SENELLART.*Top-k Querying of Unknown Values under Order Constraints (Extended Version)*, January 2017, https://arxiv.org/abs/1701.02634 - 32 pages, 1 figure, 1 algorithm, 51 references. Extended version of paper at ICDT'17, https://hal.inria.fr/hal-01439310.

[30] A. AMARILLI, M. L. BA, D. DEUTCH, P. SENELLART.*Possible and Certain Answers for Queries over Order-Incomplete Data*, October 2017, https://arxiv.org/abs/1707.07222 - 55 pages, 5 figures, 1 table, 44 references. Accepted at TIME'17. This paper is the full version with appendices of the article in the TIME proceedings. The main text of this full version is the same as the TIME proceedings version, except some superficial changes (to fit the proceedings version to 15 pages, and to obey LIPIcs-specific formatting requirements) [*DOI :* 10.4230/LIPICS.TIME.2017.4], https://hal.inria.fr/hal-01614571.

[31] P. SENELLART, A. AMARILLI, M. MONET.*Connecting Width and Structure in Knowledge Compilation*, October 2017, https://arxiv.org/abs/1709.06188 - 32 pages, no figures, 39 references. Submitted, https://hal.inria.fr/hal-01614551.

References in notes

[32] S. ABITEBOUL, P. BOURHIS, V. VIANU.*Comparing workflow specification languages: A matter of views*, in "ACM Trans. Database Syst.", 2012, vol. 37, n° 2, p. 10:1–10:59, http://doi.acm.org/10.1145/2188349.2188352.

[33] S. ABITEBOUL, P. BUNEMAN, D. SUCIU.*Data on the Web: From Relations to Semistructured Data and XML*, Morgan Kaufmann, 1999.

[34] 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.

[35] S. ABITEBOUL, L. HERR, J. V. 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.

[36] S. ABITEBOUL, B. KIMELFELD, Y. SAGIV, P. SENELLART.*On the expressiveness of probabilistic XML models*, in "VLDB J.", 2009, vol. 18, n° 5, p. 1041–1064, https://doi.org/10.1007/s00778-009-0146-1.

[37] S. ABITEBOUL, L. SEGOUFIN, V. VIANU.*Representing and querying XML with incomplete information*, in "ACM Trans. Database Syst.", 2006, vol. 31, n° 1, p. 208–254, http://doi.acm.org/10.1145/1132863.1132869.

[38] 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.

[39] 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.

[40] Y. AMSTERDAMER, Y. GROSSMAN, T. MILO, P. SENELLART.*CrowdMiner: Mining association rules from the crowd*, in "PVLDB", 2013, vol. 6, n° 12, p. 1250–1253, http://www.vldb.org/pvldb/vol6/p1250-amsterdamer.pdf.

[41] 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.

[42] D. BARBARÁ, H. GARCIA-MOLINA, D. PORTER.*The Management of Probabilistic Data*, in "IEEE Trans. Knowl. Data Eng.", 1992, vol. 4, n° 5, p. 487–502, https://doi.org/10.1109/69.166990.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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. V. 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] F. GEERTS, A. POGGI.*On database query languages for K-relations*, in "J. Applied Logic", 2010, vol. 8, n° 2, p. 173–185, https://doi.org/10.1016/j.jal.2009.09.001.

[55] L. GETOOR.*Introduction to statistical relational learning*, MIT Press, 2007.

[56] 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.

[57] 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.

[58] T. J. GREEN, V. TANNEN.*Models for Incomplete and Probabilistic Information*, in "IEEE Data Eng. Bull.", 2006, vol. 29, n° 1, p. 17–24, http://sites.computer.org/debull/A06mar/green.ps.

[59] A. Y. HALEVY.*Answering queries using views: A survey*, in "VLDB J.", 2001, vol. 10, n° 4, p. 270–294, https://doi.org/10.1007/s007780100054.

[60] M. A. HEARST, S. T. DUMAIS, E. OSUNA, J. PLATT, B. SCHOLKOPF.*Support vector machines*, in "IEEE Intelligent Systems", 1998, vol. 13, n° 4, p. 18–28, https://doi.org/10.1109/5254.708428.

[61] T. IMIELINSKI, W. L. JR..*Incomplete Information in Relational Databases*, in "J. ACM", 1984, vol. 31, n° 4, p. 761–791, http://doi.acm.org/10.1145/1634.1886.

[62] 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.

[63] 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.

[64] A. C. KLUG.*Equivalence of Relational Algebra and Relational Calculus Query Languages Having Aggregate Functions*, in "J. ACM", 1982, vol. 29, n° 3, p. 699–717, http://doi.acm.org/10.1145/322326.322332.

[65] D. KOSSMANN.*The State of the art in distributed query processing*, in "ACM Comput. Surv.", 2000, vol. 32, n° 4, p. 422–469, http://doi.acm.org/10.1145/371578.371598.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] B. SETTLES.*Active Learning*, Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Publishers, 2012, https://doi.org/10.2200/S00429ED1V01Y201207AIM018.

[73] 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.

[74] 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.

[75] R. S. SUTTON, A. G. BARTO.*Reinforcement learning - an introduction*, Adaptive computation and machine learning, MIT Press, 1998, http://www.worldcat.org/oclc/37293240.

[76] 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.

[77] 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.

[78] 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

Université Pierre et Marie Curie (Paris 6)

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. - Software
A2.1.6. - Concurrent programming
A2.1.10. - Domain-specific languages
A2.1.11. - Proof languages
A2.2.1. - Static analysis
A2.2.3. - Run-time systems
A2.3.1. - Embedded systems
A2.3.3. - Real-time systems
A2.4. - Verification, reliability, certification
A2.4.3. - Proofs
A2.5. - Software engineering
A2.6. - Infrastructure software
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. Personnel

Research Scientists

Gilles Muller [Team leader, Inria, Senior Researcher, HDR]
Pierre-Évariste Dagand [CNRS, Researcher]
Julia Lawall [Inria, Senior Researcher]

Faculty Member

Bertil Folliot [Univ Pierre et Marie Curie, Professor, HDR]

Technical Staff
Antoine Blin [Inria, granted by ORANGE SA]
Lucas Serrano [Inria, from Sep 2017 until Oct 2017]

PhD Students
Cédric Courtaud [Thales]
Redha Gouicem [Univ Pierre et Marie Curie]
Lucas Serrano [Univ Pierre et Marie Curie, from Nov 2017]
Darius Mercadier [Univ Pierre et Marie Curie, from Nov 2017]

Post-Doctoral Fellow
Van-Anh Nguyen [Univ Pierre et Marie Curie, financed by ANR ITrans]

Visiting Scientist
Gregory Kroah-Hartman [Linux Foundation, until Jul 2017]

Administrative Assistants
Nelly Maloisel [Inria, Assistant]
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 [45], [51], [90]. 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 [35], [37], [39] or dynamic analysis [59], [62], [69]. 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* [37]. 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* [50] 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 [77]. More recently, more general forms of symbolic execution [35] 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 [31], [33].

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. [42] 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 [36], 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 [23]. 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) [55], 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 [54].

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] [63].

3.1.2.1. Traditional approach.

Using little languages to aid in software development is a tried-and-trusted technique [80] 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, [24] 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 [44] and its descendant Pharo [28] – or multi-paradigm languages – such as the Scala programming

language [66] – 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 [25].

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 [34], [29] 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, [30] 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 [26], [27], [67], [70], [74], [76], [79], [60],[16], 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 [82], 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. Dagand's recent work 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 [29], [34], [64]. 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 [51], 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. [41], numerous research tools have been applied to the Linux kernel, typically for finding bugs [39], [58], [71], [81] or for computing software metrics [47], [87]. In our work, we have studied generic C bugs in Linux code [9], bugs in function protocol usage [52], [53], issues related to the processing of bug reports [75] and crash dumps [46], and the problem of backporting [70], [82], 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, [33] a template-based approach for porting device drivers from one OS to another. Ryzhyk *et al.* propose Termite, [72], [73] 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 [49] 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 [23], Coverity [41], CP-Miner, [57] PR-Miner [58], 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

As part of a collaborative effort with Timothy Bourke, Lélio Brun, Marc Pouzet (Parkas team), Xavier Leroy (Gallium team), Lionel Rieg (Collège de France) and Pierre-Évariste Dagand, our work on a certified Lustre compiler was accepted at PLDI [13].

Julia Lawall was invited to present a talk as part of the Colloquium Jacques Morgenstern at Inria - Sophia Antipolis. The talk was entitled "Coccinelle: synergy between programming language research and the Linux kernel". A video of the presentation is available. [0]

The work of Julia Lawall on the Linux kernel was featured in the Linux Foundation's 2017 Linux Kernel Development Report. [0]

[0] https://www.canal-u.tv/video/inria/coccinelle_synergy_between_programming_language_research_and_the_linux_kernel.38185
[0] https://www.linuxfoundation.org/2017-linux-kernel-report-landing-page

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/

7. New Results

7.1. Software engineering for infrastructure software

Work in 2017 on the Linux kernel has focused on the problem of kernel device driver porting and on kernel compilation as a validation mechanism in the presence of variability. We have also completed a study with researchers at Singapore Management University on the relationship between the code coverage of test cases and the number of post-release defects, focusing on a range of popular open-source projects. Finally, we have worked with researchers at the University of Frankfurt on the design of a transformation language targeting data representation changes.

Porting Linux device drivers to target more recent and older Linux kernel versions to compensate for the ever-changing kernel interface is a continual problem for Linux device driver developers. Acquiring information about interface changes is a necessary, but tedious and error prone, part of this task. To address these problems, we have proposed two tools, *Prequel* and *gcc-reduce*, to help the developer collect the needed information. Prequel provides language support for querying git commit histories, while gcc-reduce translates error messages produced by compiling a driver with a target kernel into appropriate Prequel queries. We have used our approach in porting 33 device driver files over up to 3 years of Linux kernel history, amounting to hundreds of thousands of commits. In these experiments, for 3/4 of the porting issues, our approach highlighted commits that enabled solving the porting task. For many porting issues, our approach retrieves relevant commits in 30 seconds or less. This work was published at USENIX ATC [16] and a related talk was presented at Linuxcon Europe. The Prequel tool and some of our experimental results are available at http://prequel-pql.gforge.inria.fr/. The complete tool suite is available at http://select-new.gforge.inria.fr/.

The Linux kernel is highly configurable, and thus, in principle, any line of code can be included or excluded from the compiled kernel based on configuration operations. Configurability complicates the task of a *kernel janitor*, who cleans up faults across the code base. A janitor may not be familiar with the configuration options that trigger compilation of a particular code line, leading him to believe that a fix has been compile-checked when this is not the case. We have proposed JMake, a mutation-based tool for signaling changed lines that are not subjected to the compiler. JMake shows that for most of the 12,000 file-modifying commits between Linux v4.3 and v4.4 the configuration chosen by the kernel allyesconfig option is sufficient, once the janitor chooses the correct architecture. For most commits, this check requires only 30 seconds or less. We furthermore characterize the situations in which changed code is not subjected to compilation in practice. This work was published at DSN [15] and a related talk was presented at Linuxcon Europe. JMake is available at http://jmake-release.gforge.inria.fr/.

Testing is a pivotal activity in ensuring the quality of software. Code coverage is a common metric used as a yardstick to measure the efficacy and adequacy of testing. However, does higher coverage actually lead to a decline in post-release bugs? Do files that have higher test coverage actually have fewer bug reports? The direct relationship between code coverage and actual bug reports has not yet been analysed via a comprehensive empirical study on real bugs. In an empirical study, we have examined these questions in the context of 100 large open-source Java software projects based on their actual reported bugs. Our results show that coverage has an insignificant correlation with the number of bugs that are found after the release of the software at the project level, and no such correlation at the file level. This work was done in collaboration with researchers at Singapore Management University and has been published in the IEEE Transactions on Reliability [12].

Data representation migration is a program transformation that involves changing the type of a particular data structure, and then updating all of the operations that somehow depend on that data structure according to the new type. Changing the data representation can provide benefits such as improving efficiency and improving the quality of the computed results. Performing such a transformation is challenging, because it requires applying data-type specific changes to code fragments that may be widely scattered throughout the source code, connected by dataflow dependencies. Refactoring systems are typically sensitive to dataflow dependencies, but are not programmable with respect to the features of particular data types. Existing program transformation languages provide the needed flexibility, but do not concisely support reasoning about dataflow dependencies.

To address the needs of data representation migration, we have proposed a new approach to program transformation that relies on a notion of semantic dependency: every transformation step propagates the transformation process onward to code that somehow depends on the transformed code. Our approach provides a declarative transformation-specification language, for expressing type-specific transformation rules. Our approach further provides scoped rules, a mechanism for guiding rule application, and tags, a device for simple program analysis within our framework, to enable more powerful program transformations. Evaluation of our prototype based on our approach, targeting C and C++ software, shows that it can improve program performance and the precision of the computed results, and that it scales to programs of up to 3700 lines. This work was done in collaboration with researchers at the University of Frankfurt and was published at PEPM [18].

7.2. Trustworthy domain-specific compilers

This year, we concluded the correctness proof of the compiler back-end of the Lustre [32] synchronous dataflow language. Synchronous dataflow languages are widely used for the design of embedded systems: they allow a high-level description of the system and naturally lend themselves to a hierarchical design. Developed in collaboration with members of the Parkas team of Inria Paris (Tim Bourke, Lélio Brun, Marc Pouzet), the Gallium team of Inria Paris (Xavier Leroy) and Collège de France (Lionel Rieg), this work formalizes the compilation of a synchronous data-flow language into an imperative sequential language, which is eventually translated to Cminor [56], one of CompCert's intermediate languages. The proof has been developed and verified in the Coq theorem prover. This project illustrates perfectly our methodology: the design of synchronous dataflow languages is first governed by semantic considerations (Kahn process

networks and the synchrony hypothesis) that are then reifed into syntactic artefacts. The implementation of a certified compiler highlights this dependency on semantics, forcing us to give as crisp a semantics as possible for the proof effort to be manageable. This work was published in a national conference [19] as well as in an international conference [13], both on the topic of language design and implementation.

Expanding upon these ideas, Darius Mercadier started his PhD with us in October. We are currently developing a synchronous dataflow language targeting verified and high-performance implementations of bitsliced algorithms, with application to cryptographical algorithms [40]. Our preliminary results [22] are encouraging.

7.3. Algebra of programming

We have pursued our study of the algebraic structures of programming languages, from a syntactic as well as semantics perspective. Tackling the semantics aspect, Pierre-Évariste Dagand published a journal article introducing the theory of ornaments [11] to a general audience of functional programmers. Ornaments amount to a domain-specific language, usually described in type theory, for describing structure-preserving changes in algebraic datatypes. Such descriptions can be used to improve code reuse as well as ease of refactoring in functional languages. This work is part of a wider effort by our community to foster the adoption of ornaments when programming with algebraic datatypes, be it in type theory [48] or general-purpose functional programming languages [65], [89]. Tackling the syntactic aspect and in collaboration with researchers at the University of Utrecht (Victor Miraldo, Wouter Swierstra), Pierre-Évariste Dagand has worked on a formalization of `diffs` for structured data [20]. This preliminary and foundational work aims at providing a typed specification to the problem of computing the difference of two pieces of structured data. Unlike previous approaches [43], following a type-theoretical approach allowed us to formalize the difference of two structure as a typed object. The task of computing the difference of two structured objects is then able to exploit this typing information to control the search space (which is otherwise gigantic). Having a typed difference also ensures that applying such a `diff` to a well-structured data results in either a failure (the difference is in conflict with the given file) or another well-structured data.

7.4. Developing infrastructure software using Domain Specific Languages

In terms of DSL design for domains where correctness is critical, our current focus is first on process scheduling for multicore architecture, and second on selfishness in distributed systems. Ten years ago, we developed Bossa, targeting process scheduling on unicore processors, and primarily focusing on the correctness of a scheduling policy with respect to the requirements of the target kernel. At that time, the main use cases were soft real-time applications, such as video playback. Bossa was and still continues to be used in teaching, because the associated verifications allow a student to develop a kernel-level process scheduling policy without the risk of a kernel crash. Today, however, there is again a need for the development of new scheduling policies, now targeting multicore architectures. As identified by Lozi *et al.* [61], large-scale server applications, having specific resource access properties, can exhibit pathological properties when run with the Linux kernel's various load balancing heuristics. We are working on a new domain-specific language, Ipanema, to enable verification of critical scheduling properties such as liveness and work-conservation; for the latter, we are exploring the use of the Leon theorem prover from EPFL [17]. A first version of the language has been designed and we expect to release a prototype of Ipanema working next year. The work around Ipanema is the subject of a very active collaboration between researchers at four institutions (Inria, University of Nice, University of Grenoble, and EPFL (groups of V. Kuncak and W. Zwaenepoel)). Baptiste Lepers (EPFL) is supported in 2017 as a postdoc as part of the Inria-EPFL joint laboratory.

Selfishness is one of the key problems that confronts developers of cooperative distributed systems (e.g., file-sharing networks, voluntary computing). It has the potential to severely degrade system performance and to lead to instability and failures. Current techniques for understanding the impact of selfish behaviours and designing effective countermeasures remain manual and time-consuming, requiring multi-domain expertise. To overcome these difficulties, we have proposed SEINE, a simulation framework for rapid modelling and evaluation of selfish behaviours in a cooperative system. SEINE relies on a domain-specific language (SEINE-L) for specifying selfishness scenarios, and provides semi-automatic support for their implementation and

study in a state-of-the-art simulator. We show in a paper published at DSN 2017 [14] that (1) SEINE-L is expressive enough to specify fifteen selfishness scenarios taken from the literature, (2) SEINE is accurate in predicting the impact of selfishness compared to real experiments, and (3) SEINE substantially reduces the development effort compared to traditional manual approaches.

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. More precisely, our goal is to develop optimizations for a software TDMA hypervisor developed by Thales that provides full time-isolation of tasks. The PhD of Cédric Courtaud is supported by a CIFRE fellowship with Thales Research.
- OSADL, 2016-2017, development of the Prequel patch query language, 20 000 euros. OSADL is an organization headquartered in Germany that promotes and supports the use of open source software in the automation and machine industry. The project is in the context of the OSADL project SIL2LinuxMP bringing together various companies in automotive and embedded sytems with the goal of developing methodologies for certifying the basic components of a GNU/Linux-based RTOS.

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.

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.

9.3. International Initiatives

9.3.1. Inria International Labs

- EPFL-Inria Lab Our work on the Ipanema DSL [17] is done as part of the EPFL-Inria Lab. Baptiste Lepers (EPFL) is supported in 2017 as a joint postdoc between the Whisper and the groups of V. Kuncak and W. Zwaenepoel.

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 [12], [68], [78], [83], [84], [85], [88], [86], at conferences including ASE and ICSME. The ITrans ANR is a joint project with them.
- We collaborate with Christoph Reichenbach of the University of Lund and Krishna Narasimhan of Itemis (Germany) on program transformation [18] and the design of tools for code clone management.
- We collaborate with Wouter Swierstra of the University of Utrecht (Netherlands) on type-directed structured differences [20].
- We collaborate with Eric Tanter of the University of Chile (Chile) on the theoretical and practical aspects of dependent interoperability [38] in type theory.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

As part of the Invited Professor program of LIP6, we have hosted Prof. Éric Tanter (University of Chile) for two weeks (December 2017) who took this opportunity to give an introductory master class as well as a research seminar on the topic of gradual typing.

9.4.1.1. Internships

- Lukas Gnirke, Oberlin College, January 2017, evaluation of our methodology for searching for examples to guide driver porting [16].
- Adina Johnson, Oberlin College, May - August 2017, analysis of the differences between the Linux kernel and the Android kernel.
- Jonathan Carroll, Oberlin College, May - August 2017, use of machine learning to identify stable-kernel relevant patches.
- Bhumika Goyal, October - November 2017, constification of Linux kernel structures, supported by the Linux Foundation's Core Infrastructure Initiative.
- Peio Borthelle, École Normale Supérieure de Lyon, June - July 2017, solving the Oware on a single machine.
- Darius Mercadier, Université Pierre et Marie Curie, January - August 2017, designing and implementing Usuba, a bitslicing compiler.

9.4.1.2. Research Stays Abroad

- Julia Lawall, visit to David Lo and Lingxiao Jiang at Singapore Management University (two weeks in May 2017).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- Gilles Muller: PLOS 2017

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Julia Lawall: PLOS 2017

10.1.2.2. Member of the Conference Program Committees

- Gilles Muller: Usenix ATC 2017, DSN 2017, Systor 2017,
- Julia Lawall: ICSE NIER, ML workshop, OCaml workshop, SPLASH workshops.
- Pierre-Évariste Dagand: HOPE workshop

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Julia Lawall: Editorial board of Science of Computer Programing (2008 - present).

10.1.3.2. Reviewer - Reviewing Activities

- Gilles Muller: IEEE Transactions on Computer Systems
- Julia Lawall: Computer Languages, Empirical Software Engineering, IEEE Transactions on Reliability
- Pierre-Évariste Dagand: Journal of Functional Programming, Journal of Logical and Algebraic Methods in Programming, Journée Francophones des Langages Applicatifs

10.1.4. Invited Talks

- Gilles Muller: University of North Carolina, Inria Grenoble, University of Bordeaux.
- Julia Lawall: Colloquium Jacques Morgenstern Sopha Antipolis, Vrije Universiteit Brussel, University of Copenhagen, 2017 FSD Meeting.
- Pierre-Évariste Dagand: École Normale Supérieure de Cachan

10.1.5. Research Administration

- Pierre-Évariste Dagand: Member of the steering committee for the Colloquium d'Informatique de L'UPMC Sorbonne Universités, organizer of the Colloquium of Philippa Gardner (March 2017) and Timothy Roscoe (November 2017).
- Julia Lawall: IFIP TC secretary (2012 - present). Elected member of IFIP WG 2.11.

  Hiring committees: UPMC (PR)

  Board member of Software Heritage (https://www.softwareheritage.org/).

  Organized the Colloquium d'Informatique de L'UPMC Sorbonne Universités of Simon Peyton Jones (May 2017)
- Gilles Muller: EuroSys steering committee (2013-2017), 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, member of the Paris committee for allocating post-docs, PhD stipends and sabbaticals.

  Hiring committees: University of Grenoble (MdC).
- 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

- Licence: Pierre-Évariste Dagand, Distributed cooperating objects, L3, 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

10.2.2. Supervision

- PhD in progress : Mariem Saeid, soutenance en 2018, Jens Gustedt (Camus), Gilles Muller.
- PhD in progress : Cédric Courtaud, CIFRE Thalès, 2016-2019, Gilles Muller, Julien Sopéna (Regal).
- PhD in progress : Redha Gouicem, 2016-2019, Gilles Muller, Julien Sopéna (Regal).
- 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

- Gilles Muller: Reporter of the HDR of A. Tchana (U. Toulouse), member of the HDR of S. Ben Mokhtar (U. of Lyon).
- Julia Lawall: Vinh Tao (PhD, UPMC, president), Marcelino Cancio Rodriguez (PhD, University of Rennes, reporter), Victor Allombert (PhD, Paris Est-Creteil, reporter), Reinout Stevens (PhD, VUB, reporter).

10.3. Popularization

- Julia Lawall: Coordinator of the Outreachy internship program for the Linux kernel. Outreachy provides remote 3-month internships twice a year for women and other underrepresented minorities on open source projects. Julia Lawall also mentored Bhumika Goyal as part of this program.
- Julia Lawall, “Fast and Precise Retrieval of Forward and Back Porting Information for Linux Device Drivers”, Open Source Summit Europe, October 2017.
- Julia Lawall, “JMake: Dependable Compilation for Kernel Janitors”, Open Source Summit Europe, October 2017.
- Julia Lawall, “Panel Discussion: Outreachy Kernel Internship Report” (moderator), Open Source Summit Europe, October 2017.
- Julia Lawall, “Overview of Coccinelle”, Linux Lund Conference, May 2018.
- Julia Lawall, “Constification of Linux kernel structures”, OSADL Networking Day, May 2017.

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

Doctoral Dissertations and Habilitation Theses

[10] A. BLIN.*Towards an efficient use of multi-core processors in mixed criticality embedded systems*, Université Pierre et Marie Curie - Paris VI, January 2017, https://tel.archives-ouvertes.fr/tel-01624259.

Articles in International Peer-Reviewed Journal

[11] P.-E. DAGAND.*The essence of ornaments*, in "Journal of Functional Programming", January 2017, vol. 27 [*DOI :* 10.1017/S0956796816000356], https://hal.archives-ouvertes.fr/hal-01461209.

[12] 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.

International Conferences with Proceedings

[13] 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.

[14] G. L. COTA, S. BEN MOKHTAR, G. GIANINI, E. DAMIANI, J. LAWALL, G. MULLER, L. BRUNIE.*Analysing Selfishness Flooding with SEINE*, in "The 47th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'17)", Denver, Colorado, United States, June 2017, p. 603 - 614 [*DOI :* 10.1109/DSN.2017.51], https://hal.archives-ouvertes.fr/hal-01581628.

[15] J. LAWALL, G. MULLER.*JMake: Dependable Compilation for Kernel Janitors*, in "The 47th IEEE/IFIP International Conference on Dependable Systems and Networks", Denver,Colorado, United States, IEEE/IFIP, June 2017 [*DOI :* 10.1109/DSN.2017.62], https://hal.inria.fr/hal-01555711.

[16] 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.

[17] 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.

[18] K. NARASIMHAN, C. REICHENBACH, J. LAWALL.*Interactive Data Representation Migration: Exploiting Program Dependence to Aid Program Transformation*, in "PEPM 2017 Workshop on Partial Evaluation and Program Manipulation", Paris, France, January 2017, https://hal.inria.fr/hal-01408266.

National Conferences with Proceeding

[19] T. BOURKE, P.-E. DAGAND, M. POUZET, L. RIEG.*Vérification de la génération modulaire du code impératif pour Lustre*, in "JFLA 2017 - Vingt-huitième Journées Francophones des Langages Applicatifs", Gourette, France, January 2017, https://hal.inria.fr/hal-01403830.

Conferences without Proceedings

[20] V. C. MIRALDO, P.-E. DAGAND, W. SWIERSTRA.*Type-directed diffing of structured data*, in "TyDe 2017 Proceedings of the 2nd ACM SIGPLAN International Workshop on Type-Driven Development", Oxford, United Kingdom, September 2017 [*DOI :* 10.1145/3122975.3122976], https://hal.archives-ouvertes.fr/hal-01673541.

Other Publications

[21] P.-E. DAGAND, N. TABAREAU, É. TANTER.*Foundations of Dependent Interoperability*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01629909.

[22] D. MERCADIER, P.-É. DAGAND, L. LACASSAGNE, G. MULLER.*Usuba, Optimizing & Trustworthy Bitslicing Compiler*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01657259.

References in notes

[23] 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.

[24] 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.

[25] 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.

[26] 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^{o} 2, p. 177-196, https://hal.archives-ouvertes.fr/hal-00844536.

[27] 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.

[28] A. P. BLACK, S. DUCASSE, O. NIERSTRASZ, D. POLLET.*Pharo by Example*, Square Bracket Associates, 2010.

[29] 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.

[30] 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.

[31] 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.

[32] P. CASPI, N. HALBWACHS, D. PILAUD, J. PLAICE.*Lustre: a declarative language for programming synchronous systems*, in "14th ACM Symposium on Principles of Programming Languages", ACM, 1987.

[33] V. CHIPOUNOV, G. CANDEA.*Reverse Engineering of Binary Device Drivers with RevNIC*, in "EuroSys", 2010, p. 167–180.

[34] A. CHLIPALA.*The Bedrock Structured Programming System: Combining Generative Metaprogramming and Hoare Logic in an Extensible Program Verifier*, in "ICFP", 2013, p. 391–402.

[35] 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.

[36] 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.

[37] P. COUSOT, R. COUSOT.*Abstract Interpretation: Past, Present and Future*, in "CSL-LICS", 2014, p. 2:1–2:10.

[38] P.-E. DAGAND, N. TABAREAU, É. TANTER.*Partial Type Equivalences for Verified Dependent Interoperability*, in "ICFP 2016 - 21st ACM SIGPLAN International Conference on Functional Programming", Nara, Japan, September 2016, p. 298-310, http://dx.doi.org/10.1145/2951913.2951933.

[39] I. DILLIG, T. DILLIG, A. AIKEN.*Sound, complete and scalable path-sensitive analysis*, in "PLDI", June 2008, p. 270–280.

[40] 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.

[41] 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.

[42] 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.

[43] J. FALLERI, F. MORANDAT, X. BLANC, M. MARTINEZ, M. MONPERRUS.*Fine-grained and accurate source code differencing*, in "ACM/IEEE International Conference on Automated Software Engineering, ASE '14, Vasteras, Sweden - September 15 - 19, 2014", 2014, p. 313–324, http://dx.doi.org/10.1145/2642937.2642982.

[44] A. GOLDBERG, D. ROBSON.*Smalltalk-80: The Language and Its Implementation*, Addison-Wesley, 1983.

[45] 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.

[46] 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.

[47] 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.

[48] H.-S. KO, J. GIBBONS.*Programming with ornaments*, in "Journal of Functional Programming", 2017, vol. 27, http://dx.doi.org/10.1017/S0956796816000307.

[49] A. KADAV, M. M. SWIFT.*Understanding modern device drivers*, in "ASPLOS", 2012, p. 87–98.

[50] G. A. KILDALL.*A Unified Approach to Global Program Optimization*, in "POPL", 1973, p. 194–206.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] X. LEROY.*Formal verification of a realistic compiler*, in "Communications of the ACM", 2009, vol. 52, n^{o} 7, p. 107–115.

[57] 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.

[58] 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.

[59] D. LO, S. KHOO.*SMArTIC: towards building an accurate, robust and scalable specification miner*, in "FSE", 2006, p. 265–275.

[60] 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.

[61] J. LOZI, B. LEPERS, J. R. FUNSTON, F. GAUD, V. QUÉMA, A. FEDOROVA.*The Linux scheduler: a decade of wasted cores*, in "Proceedings of the Eleventh European Conference on Computer Systems, EuroSys 2016, London, United Kingdom, April 18-21, 2016", C. CADAR, P. PIETZUCH, K. KEETON, R. RODRIGUES (editors), ACM, 2016, p. 1:1–1:16, http://doi.acm.org/10.1145/2901318.2901326.

[62] 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.

[63] 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.

[64] 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.

[65] S. NAJD, S. P. JONES.*Trees that Grow*, in "Journal of Universal Computer Science", jan 2017, vol. 23, n^{o} 1, p. 42–62.

[66] M. ODERSKY, T. ROMPF.*Unifying functional and object-oriented programming with Scala*, in "Commun. ACM", 2014, vol. 57, n^{o} 4, p. 76–86.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] L. RYZHYK, P. CHUBB, I. KUZ, E. LE SUEUR, G. HEISER.*Automatic device driver synthesis with Termite*, in "SOSP", 2009, p. 73–86.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] D. A. SCHMIDT.*Data Flow Analysis is Model Checking of Abstract Interpretations*, in "POPL", 1998, p. 38–48.

[78] P. SENNA, 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.

[79] 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.

[80] M. SHAPIRO.*Purpose-built languages*, in "Commun. ACM", 2009, vol. 52, n^{o} 4, p. 36–41.

[81] 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.

[82] 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.

[83] 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.

[84] 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.

[85] 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.

[86] 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.

[87] W. WANG, M. GODFREY.*A Study of Cloning in the Linux SCSI Drivers*, in "Source Code Analysis and Manipulation (SCAM)", IEEE, 2011.

[88] 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.

[89] T. WILLIAMS, P. DAGAND, D. RÉMY.*Ornaments in practice*, in "Proceedings of the 10th ACM SIGPLAN workshop on Generic programming, WGP 2014, Gothenburg, Sweden, August 31, 2014", 2014, p. 15–24, http://dx.doi.org/10.1145/2633628.2633631.

[90] 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:

B9.4.1. - Computer science
B9.4.5. - Data science

1. Personnel

Research Scientists

Ivan Laptev [Inria, Senior Researcher, HDR]
Josef Sivic [Inria, Senior Researcher, HDR]

Faculty Member

Jean Ponce [Team leader, ENS Paris, Professor]

Technical Staff

Igor Kalevatykh [Inria]
Mauricio Diaz [Inria]

PhD Students

Guilhem Cheron [Inria]
Theophile Dalens [Inria]
Thomas Eboli [ENS]
Yana Hasson [Inria]
Vadim Kantorov [Inria]
Zongmian Li [Inria]
Antoine Miech [Inria]
Maxime Oquab [Inria]
Julia Peyre [Inria]
Ronan Riochet [Inria]
Ignacio Rocco Spremolla [Inria]
Rafael Sampaio de Rezende [Inria]
Matthew Trager [Inria]
Gül Varol [Inria]
Tuan-Hung Vu [Inria]
Dmitry Zhukov [Inria]

Post-Doctoral Fellow

Anton Osokin [Inria]

Visiting Scientists
Hildegard Kuehne [Universtiy of Bonn, Apr 2017]
Jason Corso [University of Michigan, Apr 2017]
Alexei Efros [UC Berkeley, June 2017]

Administrative Assistants
Sabrine Boumizy [Inria]
Sarah Le [Inria]

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 four new Phd students: Thomas Eboli (ENS), Yana Hasson (Inria), Zongmian Li (Inria) and Dmitry Zhukov (Inria). Alexei Efros (Professor, UC Berkeley, USA) visited Willow during June. Hildegard Kuehne (Universtiy of Bonn) and Jason Corso (University of Michigan) visited Willow during April.

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.

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.

[0]The patent: "Match, Expand, and Filter Technique for Multi-View Stereopsis" was issued December 11, 2012 and assigned patent number 8,331,615.

3.4. 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. Awards

- J. Sivic (together with J. Philbin, O. Chum, M. Isard, and A. Zisserman) received the Longuet-Higgins Prize for "Fundamental contributions in Computer Vision", awarded at the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- J. Sivic (together with A. Zisserman) received the Helmholtz Prize for "fundamental contributions to computer vision", awarded at the International Conference on Computer Vision, 2017.
- J. Sivic (together with B. Russell, A. Efros, B. Freeman and A. Zisserman) received the Helmholtz Prize for "fundamental contributions to computer vision", awarded at the International Conference on Computer Vision, 2017.
- I. Laptev (together with T. Lindeberg) received the Helmholtz Prize for "fundamental contributions to computer vision", awarded at the International Conference on Computer Vision, 2017.

6. New Software and Platforms

6.1. LOUPE

Learnable mOdUle for Pooling fEatures
KEYWORDS: Video analysis - Computer vision
FUNCTIONAL DESCRIPTION: LOUPE (Learnable mOdUle for Pooling fEatures) is a Tensorflow toolbox that implements several modules for pooling features such as NetVLAD, NetRVLAD, NetFV and Soft-DBoW. It also allows to use their Gated version. This toolbox was mainly use in the winning approach of the Youtube 8M Large Scale Video Understanding challenge

- Participants: Antoine Miech, Ivan Laptev and Josef Sivic
- Contact: Antoine Miech
- Publication: Learning from Video and Text via Large-Scale Discriminative Clustering
- URL: https://github.com/antoine77340/LOUPE

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

6.3. SURREAL

Learning from Synthetic Humans
KEYWORDS: Synthetic human - Segmentation - Neural networks
FUNCTIONAL DESCRIPTION: The SURREAL dataset consisting of synthetic videos of humans, and models trained on this dataset are released in this package. The code for rendering synthetic images of people and for training models is also included in the release.

- Participants: Gül Varol Simsekli, Xavier Martin, Ivan Laptev and Cordelia Schmid
- Contact: Gül Varol Simsekli
- Publication: Learning from Synthetic Humans
- URL: http://www.di.ens.fr/willow/research/surreal/

6.4. UNREL

Weakly-supervised learning of visual relations
KEYWORDS: Recognition - Computer vision
FUNCTIONAL DESCRIPTION: Open source release of the software package for the ICCV17 paper by Peyre et al. "Weakly-supervised learning of visual relations". The package provides a full implementation of the method (training and evaluation) and the release of the UnRel dataset. Links to all of these are available at the project page http://www.di.ens.fr/willow/research/unrel/

- Participants: Julia Peyre, Ivan Laptev, Cordelia Schmid and Josef Sivic
- Contact: Julia Peyre
- Publication: Weakly-supervised learning of visual relations
- URL: http://www.di.ens.fr/willow/research/unrel/

6.5. BIOGAN

GANs for Biological Image Synthesis
KEYWORDS: Computer vision - Biology
FUNCTIONAL DESCRIPTION: This software package implements the method in the ICCV 2017 paper by Osokin et al. "GANs for Biological Image Synthesis".

- Participants: Federico Vaggi, Anton Osokin and Anatole Chessel
- Contact: Anton Osokin
- Publication: GANs for Biological Image Synthesis

6.6. KernelImageRetrieval

Kernel square-loss exemplar machines for image retrieval
KEYWORD: Computer vision
FUNCTIONAL DESCRIPTION: This software package contains the code for the CVPR'17 paper by Rezende et al. "Kernel square-loss exemplar machines for image retrieval". It provides the implementation of all variants of the pipeline as well as the trained parameters for each of the tested base features.

- Participants: Jean Ponce, Francis Bach, Patrick Pérez and Rafael Sampaio De Rezende
- Contact: Rafael Sampaio De Rezende
- Publication: Kernel Square-Loss Exemplar Machines for Image Retrieval
- URL: https://github.com/rafarez/slem/

6.7. SCNet

SCNet: Learning semantic correspondence

KEYWORD: Computer vision
FUNCTIONAL DESCRIPTION: This software package implements the method for the ICCV'17 paper by Han et al. "SCNet: Learning Semantic Correspondence". The package provides the code, the training and testing subsets and the trainable architecture.

- Participants: Rafael Sampaio De Rezende, Bumsub Ham, Minsu Cho, Cordelia Schmid and Jean Ponce
- Contact: Rafael Sampaio De Rezende
- Publication: SCNet: Learning Semantic Correspondence
- URL: https://github.com/k-han/SCNet/

6.8. CNNGeometric

Convolutional neural network architecture for geometric matching
KEYWORD: Computer vision
FUNCTIONAL DESCRIPTION: Open source release of the software package for the CVPR'17 paper by Rocco et al. "Convolutional neural network architecture for geometric matching". 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: Convolutional neural network architecture for geometric matching
- URL: http://www.di.ens.fr/willow/research/cnngeometric/

6.9. LSDClustering

Large-Scale Discriminative Clustering
KEYWORDS: Video analysis - Computer vision
FUNCTIONAL DESCRIPTION: This software package implements the method in the ICCV'17 paper by Miech et al. "Learning from Video and Text via Large-Scale Discriminative Clustering".

- Participants: Antoine Miech, Jean-Baptiste Alayrac, Piotr Bojanowski, Ivan Laptev and Josef Sivic
- Contact: Antoine Miech
- Publication: Learning from Video and Text via Large-Scale Discriminative Clustering
- URL: http://www.di.ens.fr/willow/research/learningvideotext/

7. New Results

7.1. 3D object and scene modeling, analysis, and retrieval

7.1.1. Congruences and Concurrent Lines in Multi-View Geometry

Participants: Jean Ponce, Bernd Sturmfels, Matthew Trager.

We present a new framework for multi-view geometry in computer vision. A camera is a mapping between $\mathbb{P}^3$ and a line congruence. This model, which ignores image planes and measurements, is a natural abstraction of traditional pinhole cameras. It includes two-slit cameras, pushbroom cameras, catadioptric cameras, and many more (Figure 1). We study the concurrent lines variety, which consists of n-tuples of lines in $\mathbb{P}^3$ that intersect at a point. Combining its equations with those of various congruences, we derive constraints for corresponding images in multiple views. We also study photographic cameras which use image measurements and are modeled as rational maps from $\mathbb{P}^3$ to $\mathbb{P}^2$ or $\mathbb{P}^1 \times \mathbb{P}^1$. This work has been published in [7].

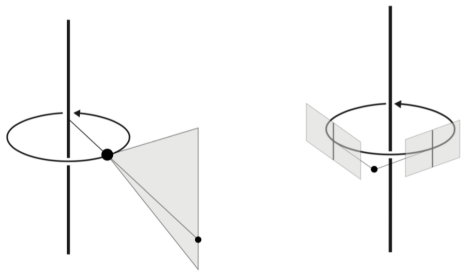


Figure 1. Non-central panoramic (left) and stereo panoramic cameras (right) are examples of non-linear cameras that can be modeled using line congruences.

7.1.2. General models for rational cameras and the case of two-slit projections

Participants: Matthew Trager, Bernd Sturmfels, John Canny, Martial Hebert, Jean Ponce.

The rational camera model provides a general methodology for studying abstract nonlinear imaging systems and their multi-view geometry. This paper builds on this framework to study "physical realizations" of rational cameras. More precisely, we give an explicit account of the mapping between between physical visual rays and image points, which allows us to give simple analytical expressions for direct and inverse projections (Figure 2). We also consider "primitive" camera models, that are orbits under the action of various projective transformations, and lead to a general notion of intrinsic parameters. The methodology is general, but it is illustrated concretely by an in-depth study of two-slit cameras, that we model using pairs of linear projections. This simple analytical form allows us to describe models for the corresponding primitive cameras, to introduce intrinsic parameters with a clear geometric meaning, and to define an epipolar tensor characterizing two-view correspondences. In turn, this leads to new algorithms for structure from motion and self-calibration. This work has been published in [22].

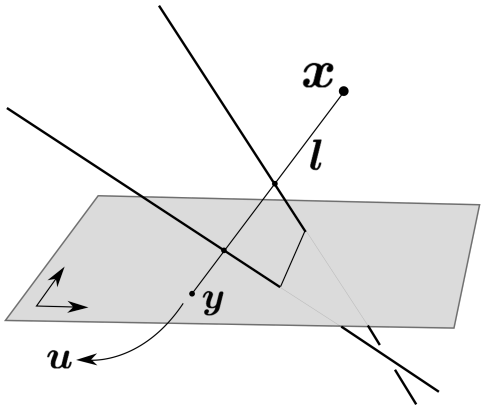


Figure 2. A general camera associates a scene point x with a visual ray l, then maps the ray l to its intersection y with some retinal plane π, and finally uses a projective coordinate system on π to express y as a point u in $\mathbb{P}^2$.

7.1.3. Changing Views on Curves and Surfaces

Participants: Kathlén Kohn, Bernd Sturmfels, Matthew Trager.

In this paper, 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 (Figure 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 was published in [6].

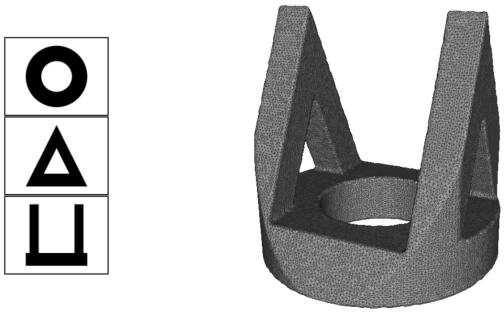


Figure 3. Changing views of a curve correspond to Reidemeister moves. The viewpoint z crosses the tangential surface (left), edge surface (middle), or trisecant surface (right).

7.1.4. On point configurations, Carlsson-Weinshall duality, and multi-view geometry

Participants: Matthew Trager, Martial Hebert, Jean Ponce.

We propose in this project projective point configurations as a natural setting for studying perspective projection in a geometric, coordinate-free manner. We show that classical results on the effect of permutations on point configurations give a purely synthetic formulation of the well known analytical Carlsson-Weinshall duality between camera pinholes and scene points. We further show that the natural parameterizations of configurations in terms of subsets of their points provides a new and simple analytical formulation of Carlsson-Weinshall duality in any scene and image coordinate systems, not just in the reduced coordinate frames used traditionally. When working in such reduced coordinate systems, we give a new and complete characterization of multi-view geometry in terms of a reduced joint image and its dual. We also introduce a new parametrization of trinocular geometry in terms of reduced trilinearities, and show that, unlike trifocal tensors, these are not subject to any nonlinear internal constraints. This leads to purely linear primal and dual structure-from-motion algorithms, that we demonstrate with a preliminary implementation on real data. This work has been submitted to CVPR'18 [27].

7.1.5. Are Large-Scale 3D Models Really Necessary for Accurate Visual Localization?

Participants: Torsten Sattler, Akihiko Torii, Josef Sivic, Marc Pollefeys, Hajime Taira, Masatoshi Okutomi, Tomas Pajdla.

Accurate visual localization is a key technology for autonomous navigation. 3D structure-based methods employ 3D models of the scene to estimate the full 6DOF pose of a camera very accurately. However, constructing (and extending) large-scale 3D models is still a significant challenge. In contrast, 2D image retrieval-based methods only require a database of geo-tagged images, which is trivial to construct and to maintain. They are often considered inaccurate since they only approximate the positions of the cameras. Yet,

the exact camera pose can theoretically be recovered when enough relevant database images are retrieved. In this paper, we demonstrate experimentally that large-scale 3D models are not strictly necessary for accurate visual localization. We create reference poses for a large and challenging urban dataset. Using these poses, we show that combining image-based methods with local reconstructions results in a pose accuracy similar to the state-of-the-art structure-based methods. Our results, published at [21] and illustrated in Figure 4, suggest that we might want to reconsider the current approach for accurate large-scale localization.

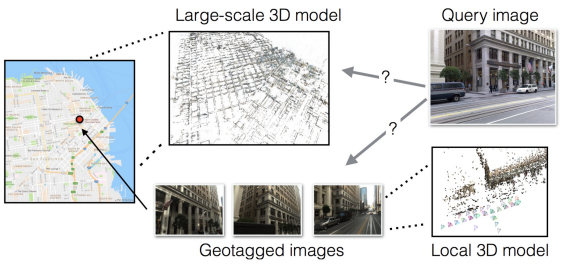


Figure 4. ***Large-scale visual localization.*** *2D image-based methods (bottom) use image retrieval and return the pose of the most relevant database image. 3D structure-based methods (top) use 2D-3D matches against a 3D model for camera pose estimation. Both approaches have been developed largely independently of each other and never compared properly before. We provide such comparison in this work.*

7.2. Category-level object and scene recognition

7.2.1. SCNet: Learning semantic correspondence

Participants: Kai Han, Rafael S. Rezende, Bumsub Ham, Kwan-Yee K. Wong, Minsu Cho, Cordelia Schmid, Jean Ponce.

In this work we propose a convolutional neural network architecture, called SCNet, for learning a geometrically plausible model for establishing semantic correspondence between images depicting different instances of the same object or scene category. SCNet uses region proposals as matching primitives, and explicitly incorporates geometric consistency in its loss function. An overview of the architecture can be seen in Figure 5. It is trained on image pairs obtained from the PASCAL VOC 2007 keypoint dataset, and a comparative evaluation on several standard benchmarks demonstrates that the proposed approach substantially outperforms both recent deep learning architectures and previous methods based on hand-crafted features. This work has been published in [13].

7.2.2. Kernel square-loss exemplar machines for image retrieval

Participants: Rafael S. Rezende, Joaquin Zepeda, Jean Ponce, Francis Bach, Patrick Pérez.

In this work we explore the promise of an exemplar classifier, such as exemplar SVM (ESVM), as a feature encoder for image retrieval and extends this approach in several directions: We first show that replacing the hinge loss by the square loss in the ESVM cost function significantly reduces encoding time with negligible effect on accuracy. We call this model square-loss exemplar machine, or SLEM. An overview of the pipeline can be seen in Figure 6. We then introduce a kernelized SLEM which can be implemented efficiently through low-rank matrix decomposition, and displays improved performance. Both SLEM variants exploit the fact that the negative examples are fixed, so most of the SLEM computational complexity is relegated to an offline process independent of the positive examples. Our experiments establish the performance and computational advantages of our approach using a large array of base features and standard image retrieval datasets. This work has been published in [19].

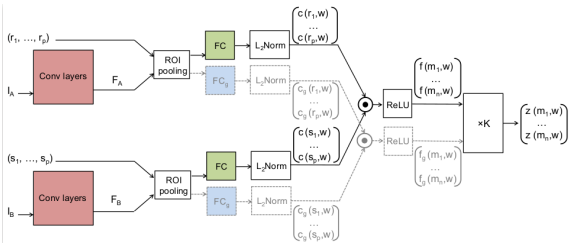


Figure 5. The SCNet architectures. Three variants are proposed: SCNet-AG, SCNet-A, and SCNet-AG+. The basic architecture, SCNet-AG, is drawn in solid lines. Colored boxes represent layers with learning parameters and the boxes with the same color share the same parameters. "$\times K$" denotes the voting layer for geometric scoring. A simplified variant, SCNet-A, learns appearance information only by making the voting layer an identity function. An extended variant, SCNet-AG+, contains an additional stream drawn in dashed lines. SCNet-AG learns a single embedding c for both appearance and geometry, whereas SCNet-AG+ learns an additional and separate embedding c_g for geometry.

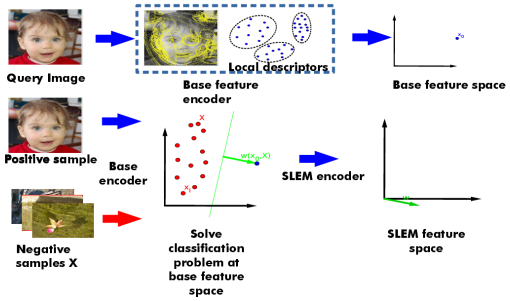


Figure 6. Pipeline of SLEM. First row encapsulates the construction of a base feature for a query image, which usually consists of extracting, embedding and aggregating local descriptors into a vector, here written as x_0. After repeating the process of base feature calculation a database of sample images and obtaining a matrix X of base features, we solve a exemplar classifier by labeling x_0 as the lonely positive example (called exemplar) and the columns of X as negatives. The solution ω to this classification problem, which is a function of x_0 and X, is our SLEM encoding of the query image.

7.2.3. Weakly-supervised learning of visual relations

Participants: Julia Peyre, Ivan Laptev, Cordelia Schmid, Josef Sivic.

This paper introduces a novel approach for modeling visual relations between pairs of objects. We call relation a triplet of the form $(subject, predicate, object)$ where the predicate is typically a preposition (eg. 'under', 'in front of') or a verb ('hold', 'ride') that links a pair of objects $(subject, object)$. Learning such relations is challenging as the objects have different spatial configurations and appearances depending on the relation in which they occur. Another major challenge comes from the difficulty to get annotations, especially at box-level, for all possible triplets, which makes both learning and evaluation difficult. The contributions of this paper are threefold. First, we design strong yet flexible visual features that encode the appearance and spatial configuration for pairs of objects. Second, we propose a weakly-supervised discriminative clustering model to learn relations from image-level labels only. Third we introduce a new challenging dataset of unusual relations (UnRel) together with an exhaustive annotation, that enables accurate evaluation of visual relation retrieval. We show experimentally that our model results in state-of-the-art results on the visual relationship dataset significantly improving performance on previously unseen relations (zero-shot learning), and confirm this observation on our newly introduced UnRel dataset. This work has been published in [18] and example results are shown in Figure 7.

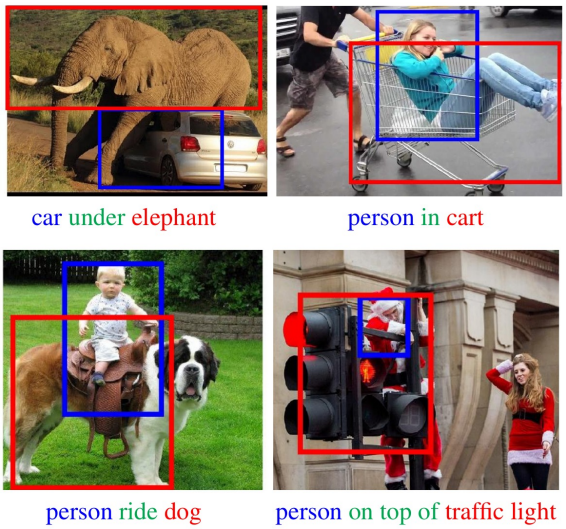


Figure 7. Examples of top retrieved pairs of boxes in UnRel dataset for unusual queries with our weakly supervised model

7.2.4. Convolutional neural network architecture for geometric matching

Participants: Ignacio Rocco, Relja Arandjelović, Josef Sivic.

We address the problem of determining correspondences between two images in agreement with a geometric model such as an affine or thin-plate spline transformation, and estimating its parameters. The contributions of this work are three-fold. First, we propose a convolutional neural network architecture for geometric matching, illustrated in Figure 8. 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 Proposal Flow dataset. This work has been published in [20].

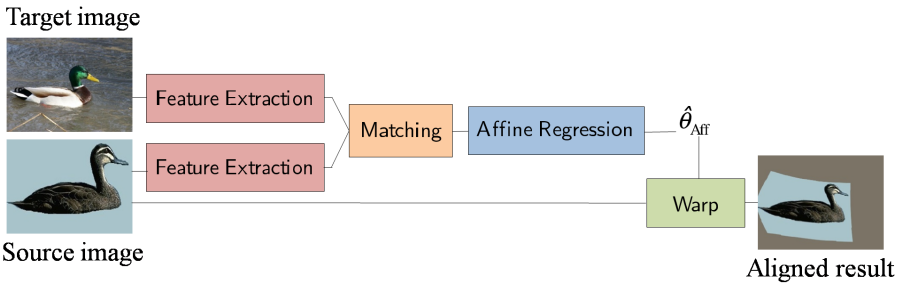


Figure 8. Proposed CNN architecture for geometric matching. Source and target images are passed through feature extraction networks which have tied parameters, followed by a matching network which matches the descriptors. The output of the matching network is passed through a regression network which outputs the parameters of the geometric transformation, which are used to produce the final alignment.

7.3. Image restoration, manipulation and enhancement

7.3.1. GANs for Biological Image Synthesis

Participants: Anton Osokin, Anatole Chessel, Rafael E. Carazo Salas, Federico Vaggi.

In this work we propose a novel application of Generative Adversarial Networks (GAN) to the synthesis of cells imaged by fluorescence microscopy. Compared to natural images, cells tend to have a simpler and more geometric global structure that facilitates image generation. However, the correlation between the spatial pattern of different fluorescent proteins reflects important biological functions, and synthesized images have to capture these relationships to be relevant for biological applications. We adapt GANs to the task at hand and propose new models with casual dependencies between image channels that can generate multi-channel images, which would be impossible to obtain experimentally (see Figure 9). We evaluate our approach using two independent techniques and compare it against sensible baselines. Finally, we demonstrate that by interpolating across the latent space we can mimic the known changes in protein localization that occur through time during the cell cycle, allowing us to predict temporal evolution from static images. This paper has been published in [17].

7.4. Human activity capture and classification

7.4.1. Learning from Synthetic Humans

Participants: Gül Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael Black, Ivan Laptev, Cordelia Schmid.

Estimating human pose, shape, and motion from images and video are fundamental challenges with many applications. Recent advances in 2D human pose estimation use large amounts of manually-labeled training data for learning convolutional neural networks (CNNs). Such data is time consuming to acquire and difficult to extend. Moreover, manual labeling of 3D pose, depth and motion is impractical. In [23], we present SURREAL: a new large-scale dataset with synthetically-generated but realistic images of people rendered from 3D sequences of human motion capture data. We generate more than 6 million frames together with ground truth pose, depth maps, and segmentation masks. We show that CNNs trained on our synthetic dataset

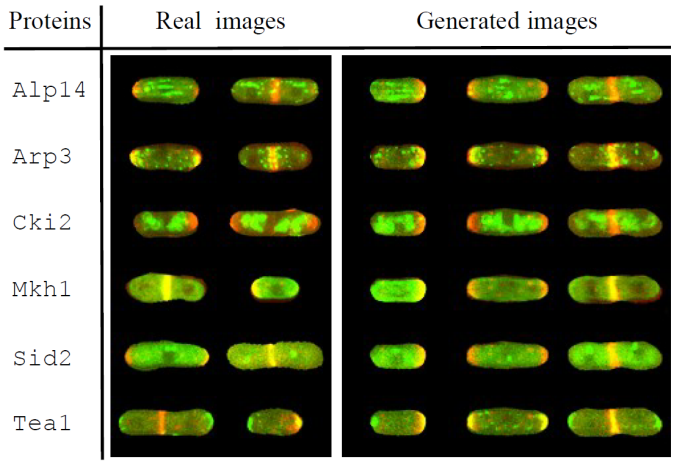


Figure 9. Real (left) and generated (right) images of fission yeast cells with protein bgs4 depicted in the red channel and 6 other proteins depicted in the green channel. The synthetic images were generated with our star-shaped GAN. The star-shaped model can generate multiple green channels aligned with the same red channel whereas the training images have only one green channel.

allow for accurate human depth estimation and human part segmentation in real RGB images, see Figure 10. Our results and the new dataset open up new possibilities for advancing person analysis using cheap and large-scale synthetic data. This work has been published in [23].

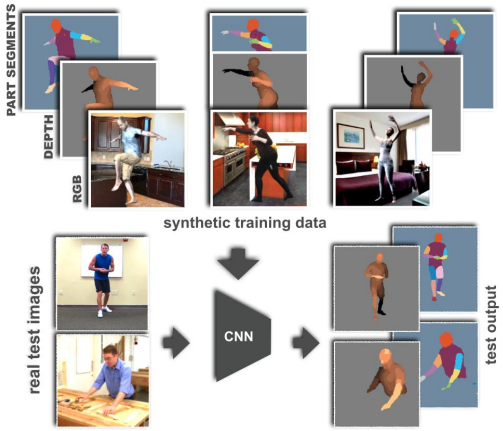


Figure 10. We generate photo-realistic synthetic images and their corresponding ground truth for learning pixel-wise classification problems: human parts segmentation and depth estimation. The convolutional neural network trained only on synthetic data generalizes on real images sufficiently for both tasks.

7.4.2. Learning from Video and Text via Large-Scale Discriminative Clustering

Participants: Miech Antoine, Alayrac Jean-Baptiste, Bojanowski Piotr, Laptev Ivan, Sivic Josef.

Discriminative clustering has been successfully applied to a number of weakly-supervised learning tasks. Such applications include person and action recognition, text-to-video alignment, object co-segmentation and colocalization in videos and images. One drawback of discriminative clustering, however, is its limited scalability. We address this issue and propose an online optimization algorithm based on the Block-Coordinate Frank-Wolfe algorithm. We apply the proposed method to the problem of weakly supervised learning of actions and actors from movies together with corresponding movie scripts. The scaling up of the learning problem to 66 feature length movies enables us to significantly improve weakly supervised action recognition. Figure 11 illustrates output of our method on movies. This work has been published in [15]

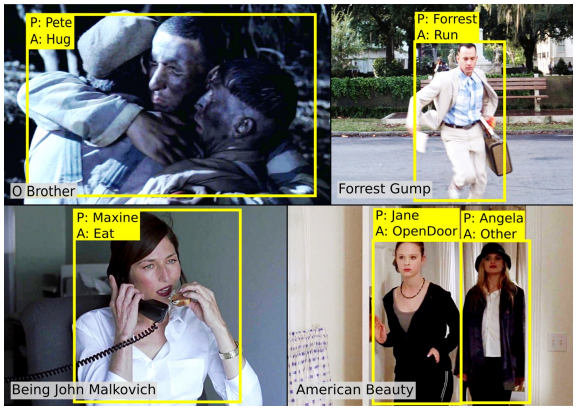


Figure 11. We automatically recognize actors and their actions in a of dataset of 66 movies with scripts as weak supervision

7.4.3. ActionVLAD: Learning spatio-temporal aggregation for action classification

Participants: Rohit Girdhar, Deva Ramanan, Abhinav Gupta, Josef Sivic, Bryan Russell.

In this work, we introduce a new video representation for action classification that aggregates local convolutional features across the entire spatio-temporal extent of the video. We do so by integrating state-of-the-art two-stream networks [42] with learnable spatio-temporal feature aggregation [6]. The resulting architecture is end-to-end trainable for whole-video classification. We investigate different strategies for pooling across space and time and combining signals from the different streams. We find that: (i) it is important to pool jointly across space and time, but (ii) appearance and motion streams are best aggregated into their own separate representations. Finally, we show that our representation outperforms the two-stream base architecture by a large margin (13out-performs other baselines with comparable base architec-tures on HMDB51, UCF101, and Charades video classification benchmarks. The work has been published at [12] and the method is illustrated in Figure 12.

7.4.4. Localizing Moments in Video with Natural Language

Participants: Lisa Hendricks, Oliver Wang, Eli Shechtman, Josef Sivic, Trevor Darrell, Bryan Russell.

We consider retrieving a specific temporal segment, or moment, from a video given a natural language text description. Methods designed to retrieve whole video clips with natural language determine what occurs in a video but not when. To address this issue, we propose the Moment Context Network (MCN) which effectively localizes natural language queries in videos by integrating local and global video features over time. A key obstacle to training our MCN model is that current video datasets do not include pairs of localized video segments and referring expressions, or text descriptions which uniquely identify a corresponding moment. Therefore, we collect the Distinct Describable Moments (DiDeMo) dataset which consists of over 10,000

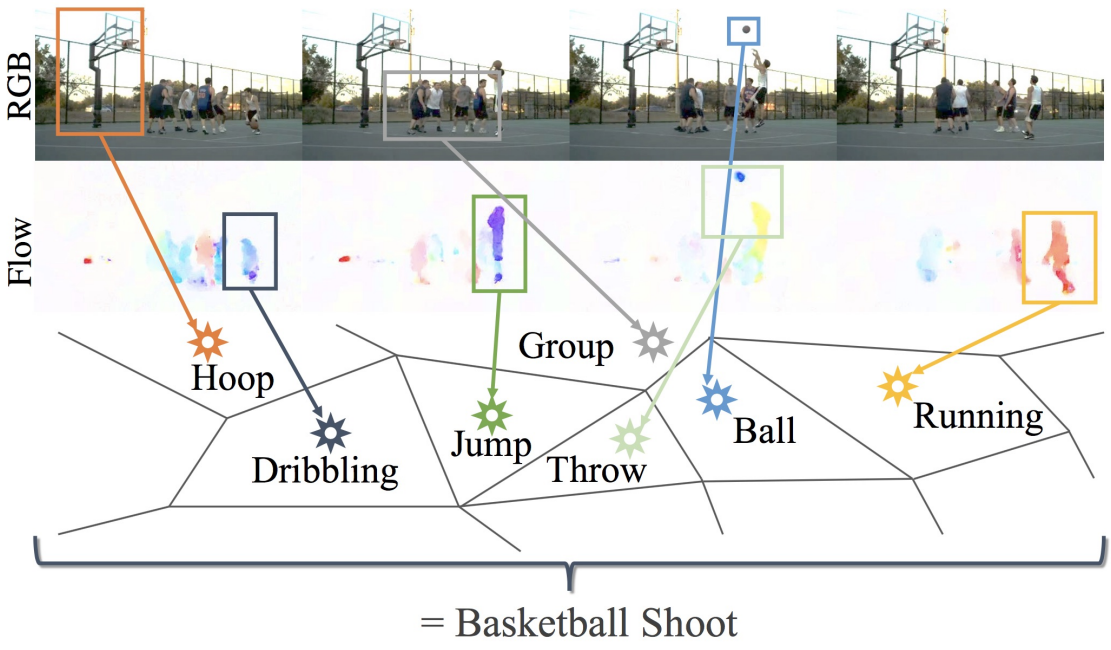


Figure 12. How do we represent actions in a video? We propose ActionVLAD, a spatio-temporal aggregation of a set of action primitives over the appearance and motion streams of a video. For example, a basketball shoot may be represented as an aggregation of appearance features corresponding to 'group of players', 'ball' and 'basketball hoop'; and motion features corresponding to 'run', 'jump', and 'shoot'.

unedited, personal videos in diverse visual settings with pairs of localized video segments and referring expressions. We demonstrate that MCN outperforms several baseline methods and believe that our initial results together with the release of DiDeMo will inspire further research on localizing video moments with natural language. The work has been published at [14] and results are illustrated in Figure 13.

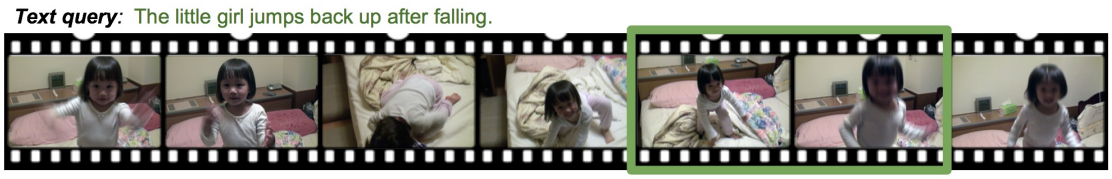


Figure 13. We consider localizing moments in video with natural language and demonstrate that incorporating local and global video features is important for this task. To train and evaluate our model, we collect the Distinct Describable Moments (DiDeMo) dataset which consists of over 40,000 pairs of localized video moments and corresponding natural language.

7.4.5. Learnable pooling with Context Gating for video classification

Participants: Miech Antoine, Laptev Ivan, Sivic Josef.

Common video representations often deploy an average or maximum pooling of pre-extracted frame features over time. Such an approach provides a simple means to encode feature distributions, but is likely to be suboptimal. As an alternative, in this work we explore combinations of learnable pooling techniques such as Soft Bag-of-words, Fisher Vectors, NetVLAD, GRU and LSTM to aggregate video features over time. We also introduce a learnable non-linear network unit, named Context Gating, aiming at modeling interdependencies between features. The overview of our network architecture is illustrated in Figure 14. We evaluate the method on the multi-modal Youtube-8M Large-Scale Video Understanding dataset using pre-extracted visual and audio features. We demonstrate improvements provided by the Context Gating as well as by the combination of learnable pooling methods. We finally show how this leads to the best performance, out of more than 600 teams, in the Kaggle Youtube-8M Large-Scale Video Understanding challenge. This work has been published in [26].

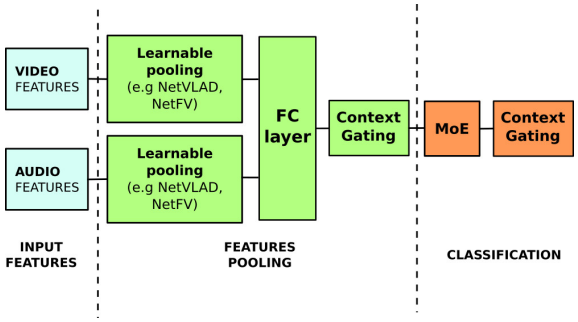


Figure 14. Overview of our network architecture for video classification

8. Bilateral Contracts and Grants with Industry

8.1. 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. 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.

8.3. 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.

8.4. 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 Science” 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.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. Agence Nationale de la Recherche (ANR): SEMAPOLIS

Participants: Mathieu Aubry, Josef Sivic.

The goal of the SEMAPOLIS project is to develop advanced large-scale image analysis and learning techniques to semantize city images and produce semantized 3D reconstructions of urban environments, including proper rendering. Geometric 3D models of existing cities have a wide range of applications, such as navigation in virtual environments and realistic sceneries for video games and movies. A number of players (Google, Microsoft, Apple) have started to produce such data. However, the models feature only plain surfaces, textured from available pictures. This limits their use in urban studies and in the construction industry, excluding in practice applications to diagnosis and simulation. Besides, geometry and texturing are often wrong when there are invisible or discontinuous parts, e.g., with occluding foreground objects such as trees, cars or lampposts, which are pervasive in urban scenes. This project will go beyond the plain geometric models by producing semantized 3D models, i.e., models which are not bare surfaces but which identify architectural elements such as windows, walls, roofs, doors, etc. Semantic information is useful in a larger number of scenarios, including diagnosis and simulation for building renovation projects, accurate shadow impact taking into account actual window location, and more general urban planning and studies such as solar cell deployment. Another line of applications concerns improved virtual cities for navigation, with object-specific rendering, e.g., specular surfaces for windows. Models can also be made more compact, encoding object repetition (e.g., windows) rather than instances and replacing actual textures with more generic ones according to semantics; it allows cheap and fast transmission over low- bandwidth mobile phone networks, and efficient storage in GPS navigation devices.

This is a collaborative effort with LIGM / ENPC (R. Marlet), University of Caen (F. Jurie), Inria Sophia Antipolis (G. Drettakis) and Acute3D (R. Keriven).

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 2017 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. Inria CityLab initiative

Participants: Josef Sivic, Jean Ponce, Ivan Laptev, Alexei Efros [UC Berkeley].

Willow participates in the ongoing CityLab@Inria initiative (co-ordinated by V. Issarny), which aims to leverage Inria research results towards developing “smart cities" by enabling radically new ways of living in, regulating, operating and managing cities. The activity of Willow focuses on urban-scale quantitative visual analysis and is pursued in collaboration with A. Efros (UC Berkeley).

Currently, map-based street-level imagery, such as Google Street-view provides a comprehensive visual record of many cities worldwide. Additional visual sensors are likely to be wide-spread in near future: cameras will be built in most manufactured cars and (some) people will continuously capture their daily visual experience using wearable mobile devices such as Google Glass. All this data will provide large-scale, comprehensive and dynamically updated visual record of urban environments.

The goal of this project is to develop automatic data analytic tools for large-scale quantitative analysis of such dynamic visual data. The aim is to provide quantitative answers to questions like: What are the typical architectural elements (e.g., different types of windows or balconies) characterizing a visual style of a city district? What is their geo-spatial distribution? How does the visual style of a geo-spatial area evolve over time? What are the boundaries between visually coherent areas in a city? Other types of interesting questions concern distribution of people and their activities: How do the number of people and their activities at particular places evolve during a day, over different seasons or years? Are there tourists sightseeing, urban dwellers shopping, elderly walking dogs, or children playing on the street? What are the major causes for bicycle accidents?

Break-through progress on these goals would open-up completely new ways smart cities are visualized, modeled, planned and simulated, taking into account large-scale dynamic visual input from a range of visual sensors (e.g., cameras on cars, visual data from citizens, or static surveillance cameras).

9.3.3. 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

Prof. Alexei Efros (UC Berkeley, USA) visited Willow during June. Hildegard Kuehne (Universtiy of Bonn) and Jason Corso (University of Michigan) visited Willow during April.

9.4.1.1. Internships

Kai Han has visited Willow from the University of Hong Kong.

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

Jean Ponce is visiting New York University since September 2017.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- I. Laptev will be program co-chair of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

10.1.1.2. Member of the Organizing Committees

- M. Trager is an organizer of “Minisymposium” on “Algebraic Vision” at the SIAM conference on Applied Algebraic Geometry (Atlanta, July 31 – August 4, 2017).
- G. Varol is an organizer of “Women in Computer Vision Workshop” at IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- G. Varol is an organizer of “Multiview Relationships in 3D Data Workshop” at International Conference on Computer Vision (ICCV), 2017.

10.1.2. Scientific Events Selection

10.1.2.1. Area chairs

- IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018 (J. Sivic).
- International Conference on Computer Vision (ICCV), 2017 (J. Sivic).
- European Conference on Computer Vision (ECCV), 2018 (I. Laptev, J. Sivic).

10.1.2.2. Member of the Conference Program Committees

- IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017 (G. Cheron, I. Laptev, A. Osokin, J. Sivic).
- IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018 (J.-B. Alayrac, M. Oquab, R. Rezende, I. Rocco, G. Varol).
- International Conference on Computer Vision (ICCV), 2017 (I. Laptev, A. Osokin).
- Neural Information Processing Systems (NIPS), 2017 (A. Osokin, J. Sivic).
- International Conference on Learning Representations (ICLR), 2017 (J. Sivic).
- International Conference on Machine Learning (ICML), 2017 (A. Osokin).
- IEEE International Conference on Robotics and Automation (ICRA), 2018 (A. Miech).

10.1.3. Journals

10.1.3.1. Member of the editorial board

- 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).
- I. Laptev co-edit a special issue on "Deep Learning for Computer Vision" in Computer Vision and Image Understanding.

10.1.3.2. Reviewer

- International Journal of Computer Vision (G. Cheron, M. Trager, G. Varol).
- IEEE Transactions on Pattern Analysis and Machine Intelligence (J.-B. Alayrac, G. Cheron, A. Osokin, M. Trager, G. Varol).
- IEEE Transactions on Circuits and Systems for Video Technology (G. Varol).

10.1.4. Others

- J. Sivic is senior fellow of the Neural Computation and Adaptive Perception program of the Canadian Institute of Advanced Research.
- A. Bursuc obtained the outstanding reviewer award at International Conference on Computer Vision (ICCV), 2017.

10.1.5. Invited Talks

- I. Laptev, Seminar, Inria Rennes, December, 2017.
- I. Laptev, Invited talk, Chalearn Workshop on Action, Gesture, and Emotion Recognition, Venice, October, 2017.
- I. Laptev, Invited talk, The Joint Video and Language Understanding Workshop, Venice, October, 2017.
- I. Laptev, Invited talk, Rentrée ENS, Paris, September, 2017
- I. Laptev, Invited talk, ML Day, Pré-GDR IA, Paris, September 2017.
- I. Laptev, Invited talk, Frontiers of Video Technology workshop, Adobe, July, 2017.
- I. Laptev, Invited talk, Workshop on YouTube-8M Large-Scale Video Understanding, Honolulu, July, 2017.
- I. Laptev, Invited talk, Workshop on Visual Understanding Across Modalities, Honolulu, July, 2017.
- I. Laptev, Plenary talk, Iberian Conference on Pattern Recognition and Image Analysis, Faro, June, 2017.
- I. Laptev, Invited talk, Paris ML Meetup Spatio-temporal Series Hackathon, Paris, February, 2017.
- A. Miech, Invited Talk, Facebook AI Research, Paris, September 2017.
- A. Miech, Invited Talk, DGA TIM2017 Seminar, Paris, July 2017.

- A. Miech, Invited Talk, Paris ML Meetup, Paris, June 2017.
- J. Ponce, Keynote speaker, Korean Conference on Computer Vision, Seoul, June 2017.
- J. Ponce, Invited talk, BioVision, Lyon, April 2017.
- J. Ponce, Invited talk, Dept. of computer science of the University of Central Florida, Orlando, February 2017.
- J. Ponce, Invited talk, New York University, Center for Data Science, September 2017.
- J. Ponce, Invited talk, Facebook AI Research, New York, October 2017.
- J. Ponce, Invited talk, Amazon, Seattle, November 2017.
- J. Ponce, Invited talk, AEF conference, Paris, 2017.
- J. Ponce, Invited talk, BIOVISION, The World Life Sciences Forum, Lyon, April 2017.
- J. Ponce, Invited talk, DGSI, 2017.
- J. Ponce, Invited talk, DRM 20th Anniversary, 2017.
- J. Sivic, Seminar, KTH Stockholm, January, 2017.
- J. Sivic, Invited talk, ParisTech Telecom, January, 2017.
- J. Sivic, Invited talk, University of Amsterdam, March, 2017.
- J. Sivic, Invited talk, ORASIS, journées francophones des jeunes chercheurs en vision par ordinateur, June, 2017.
- J. Sivic, Invited talk, Visual Understanding for Interaction workshop, CVPR 2017, July, 2017.
- J. Sivic, Invited talk, Frontiers of Video Technology workshop, Adobe, July, 2017.
- J. Sivic, Invited talk, Inria Rennes, September, 2017.
- J. Sivic, Seminar, UC Berkeley, December, 2017.
- J. Sivic, Invited talk, the CIFAR workshop, Long Beach, December 2017.

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. Sivic gave an overview of state-of-the-art in computer vision at the seminar on deep learning at Academie des Technologies, Paris, November 2017.

10.1.8. Research Administration

- Member, Bureau du comité des projets, Inria, Paris (J. Ponce)
- Director, Department of Computer Science, Ecole normale supérieure (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- (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.
- Master : I. Laptev and J. Sivic, Cours PSL-ITI - Informatique, mathematiques appliques pour le traitement du signal et l'imagerie, 20h.

10.2.2. Supervision

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 : Jean-Baptiste Alayrac, "Structured learning from video and natural language", started in 2014, I. Laptev, J. Sivic and S. Lacoste-Julien (Inria SIERRA / U. Montreal).

PhD : Rafael Sampaio de Rezende, "New methods for image classification, image retrieval and semantic correspondence", graduated in 2017, J. Ponce.

PhD in progress : Guilhem Cheron, "Structured modeling and recognition of human actions in video", started in 2014, I. Laptev and C. Schmid.

PhD in progress : Theophile Dalens, "Learning to analyze and reconstruct architectural scenes", starting in Jan 2015, M. Aubry and J. Sivic.

PhD in progress : Vadim Kantorov, "Large-scale video mining and recognition", started in 2012, I. Laptev.

PhD in progress : Maxime Oquab, "Learning to annotate dynamic scenes with convolutional neural networks", started in Jan 2014, L. Bottou (Facebook AI Research), I. Laptev and J. Sivic.

PhD in progress : Matthew Trager, "Projective geometric models in vision", started in 2014, J. Ponce and M. Hebert (CMU).

PhD in progress : Tuang Hung VU, "Learning functional description of dynamic scenes", started in 2013, I. Laptev.

10.2.3. Juries

PhD thesis committee:

- Ahmet Iscen, University of Rennes, France, 2017, (J. Sivic, rapporteur).
- Edouard Oyallon, ENS, France, 2017, (I. Laptev, examinatuer).
- Juan Manuel PÉREZ RÚA, Inria Rennes, France, 2017, (I. Laptev, examinatuer).
- Ali Razavian, KTH Stockholm, Sweden, 2017 (J. Sivic, reviewer).
- Francisco Suzano Massa, ENPC, France, 2017, (J. Sivic, examinatuer).
- Mattis Paulin, Universite de Grenoble, France, 2017 (J. Sivic, rapporteur).

10.3. Popularization

- Interview with Science et Vie, June 2017 (J. Ponce).
- Interview with Forbes Russia, June 2017 (J. Ponce).
- Interview "Science et Vie Junior", July 2017 (J. Ponce).
- Interview for "Le jaune et le rouge", Oct. 2017 (J. Ponce).
- Interview for the "les 100 français de l'IA" dossier of Usine Nouvelle, Dec. 2017 (J. Ponce).
- Two articles in "Binaires" for the newspaper Le Monde (J. Ponce).

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] R. SAMPAIO DE REZENDE.*New methods for image classification, image retrieval and semantic correspondence* , École normale supérieure de Paris, December 2017, https://hal.inria.fr/tel-01676893.

Articles in International Peer-Reviewed Journal

[2] J.-B. ALAYRAC, P. BOJANOWSKI, N. AGRAWAL, J. SIVIC, I. LAPTEV, S. LACOSTE-JULIEN.*Learning from narrated instruction videos*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", September 2017, vol. XX, https://hal.archives-ouvertes.fr/hal-01580630.

[3] R. ARANDJELOVIĆ, P. GRONAT, A. TORII, T. PAJDLA, J. SIVIC.*NetVLAD: CNN architecture for weakly supervised place recognition*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", June 2017, vol. XX, https://hal.inria.fr/hal-01557234.

[4] B. HAM, M. CHO, J. PONCE.*Robust Guided Image Filtering Using Nonconvex Potentials*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2017, Accepted pending minor revision, https://hal.archives-ouvertes.fr/hal-01279857.

[5] B. HAM, M. CHO, C. SCHMID, J. PONCE.*Proposal Flow: Semantic Correspondences from Object Proposals*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2018, https://hal.inria.fr/hal-01644132.

[6] K. KOHN, B. STURMFELS, M. TRAGER.*Changing Views on Curves and Surfaces*, in "Acta Mathematica Vietnamica", December 2017, https://arxiv.org/abs/1707.01877 - 31 pages [*DOI :* 10.1007/s40306-017-0240-1], https://hal.inria.fr/hal-01676208.

[7] J. PONCE, B. STURMFELS, M. TRAGER.*Congruences and Concurrent Lines in Multi-View Geometry*, in "Advances in Applied Mathematics", 2017, vol. 88, p. 62-91, https://arxiv.org/abs/1608.05924v2 , https://hal.inria.fr/hal-01423057.

[8] A. TORII, R. ARANDJELOVIC, J. SIVIC, M. OKUTOMI, T. PAJDLA.*24/7 place recognition by view synthesis*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", February 2017, 14 [*DOI :* 10.1109/TPAMI.2017.2667665], https://hal.inria.fr/hal-01616660.

[9] G. VAROL, I. LAPTEV, C. SCHMID.*Long-term Temporal Convolutions for Action Recognition*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2017, https://arxiv.org/abs/1604.04494 [*DOI :* 10.1109/TPAMI.2017.2712608], https://hal.inria.fr/hal-01241518.

[10] 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

[11] J.-B. ALAYRAC, J. SIVIC, I. LAPTEV, S. LACOSTE-JULIEN.*Joint Discovery of Object States and Manipulation Actions*, in "ICCV 2017 - IEEE International Conference on Computer Vision", Venice, Italy, October 2017, https://arxiv.org/abs/1702.02738 - Appears in: International Conference on Computer Vision 2017 (ICCV 2017). 15 pages, https://hal.archives-ouvertes.fr/hal-01676084.

[12] R. GIRDHAR, D. RAMANAN, A. GUPTA, J. SIVIC, B. RUSSELL.*ActionVLAD: Learning spatio-temporal aggregation for action classification*, in "IEEE Conference on Computer Vision and Pattern Recognition", Honolulu, United States, 2017, https://arxiv.org/abs/1704.02895 - Project page: https://rohitgirdhar.github.io/ActionVLAD/, https://hal.inria.fr/hal-01678686.

[13] K. K. HAN, R. S. REZENDE, B. HAM, K.-Y. K. WONG, M. CHO, C. S. SCHMID, J. S. PONCE.*SCNet: Learning Semantic Correspondence*, in "International Conference on Computer Vision", Venise, Italy, International conference on computer vision, October 2017, https://arxiv.org/abs/1705.04043 , https://hal.archives-ouvertes.fr/hal-01576117.

[14] L. A. HENDRICKS, O. WANG, E. SHECHTMAN, J. SIVIC, T. DARRELL, B. RUSSELL.*Localizing Moments in Video with Natural Language*, in "IEEE International Conference on Computer Vision - ICCV 2017", Venice, Italy, October 2017, https://arxiv.org/abs/1708.01641 , https://hal.inria.fr/hal-01678699.

[15] A. MIECH, J.-B. ALAYRAC, P. BOJANOWSKI, I. LAPTEV, J. SIVIC.*Learning from Video and Text via Large-Scale Discriminative Clustering*, in "ICCV 2017 - IEEE International Conference on Computer Vision", Venice, Italy, October 2017, https://arxiv.org/abs/1707.09074 , https://hal.inria.fr/hal-01569540.

[16] A. OSOKIN, F. BACH, S. LACOSTE-JULIEN.*On Structured Prediction Theory with Calibrated Convex Surrogate Losses*, in "The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)", Long Beach, United States, December 2017, https://arxiv.org/abs/1703.02403 , https://hal.archives-ouvertes.fr/hal-01611691.

[17] A. OSOKIN, A. CHESSEL, R. E. C. SALAS, F. VAGGI.*GANs for Biological Image Synthesis*, in "ICCV 2017 - IEEE International Conference on Computer Vision", Venice, Italy, October 2017, https://arxiv.org/abs/1708.04692 , https://hal.archives-ouvertes.fr/hal-01611692.

[18] J. PEYRE, I. LAPTEV, C. SCHMID, J. SIVIC.*Weakly-supervised learning of visual relations*, in "ICCV 2017-International Conference on Computer Vision 2017", Venice, Italy, October 2017, https://arxiv.org/abs/1707.09472 , https://hal.archives-ouvertes.fr/hal-01576035.

[19] R. S. REZENDE, J. ZEPEDA, J. S. PONCE, F. S. BACH, P. PÉREZ.*Kernel Square-Loss Exemplar Machines for Image Retrieval*, in "Computer Vision and Pattern Recognition 2017", Honolulu, United States, Computer vision and pattern recognition 2017, July 2017, https://hal.inria.fr/hal-01515224.

[20] I. ROCCO, R. ARANDJELOVIĆ, J. SIVIC.*Convolutional neural network architecture for geometric matching*, in "CVPR 2017 - IEEE Conference on Computer Vision and Pattern Recognition", Honolulu, United States, July 2017, https://arxiv.org/abs/1703.05593 , https://hal.inria.fr/hal-01513001.

[21] T. SATTLER, A. TORII, J. SIVIC, M. POLLEFEYS, H. TAIRA, M. OKUTOMI, T. PAJDLA.*Are Large-Scale 3D Models Really Necessary for Accurate Visual Localization?*, in "CVPR 2017 - IEEE Conference on Computer Vision and Pattern Recognition", Honolulu, United States, July 2017, 10, https://hal.inria.fr/hal-01513083.

[22] M. TRAGER, B. STURMFELS, J. CANNY, M. HEBERT, J. PONCE.*General models for rational cameras and the case of two-slit projections*, in "CVPR 2017 - IEEE Conference on Computer Vision and Pattern Recognition", Honolulu, United States, July 2017, https://arxiv.org/abs/1612.01160v4 , https://hal.archives-ouvertes.fr/hal-01506996.

[23] G. VAROL, J. J. ROMERO, X. MARTIN, N. MAHMOOD, M. J. BLACK, I. LAPTEV, C. SCHMID.*Learning from Synthetic Humans*, in "2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)", Honolulu, United States, July 2017, https://arxiv.org/abs/1701.01370 [*DOI :* 10.1109/CVPR.2017.492], https://hal.inria.fr/hal-01505711.

Scientific Books (or Scientific Book chapters)

[24] PARTHENOS (editor). *Digital 3D Objects in Art and Humanities: challenges of creation, interoperability and preservation. White paper: A result of the PARTHENOS Workshop held in Bordeaux at Maison des Sciences de l'Homme d'Aquitaine and at Archeovision Lab. (France), November 30th - December 2nd, 2016*, PARTHENOS, Bordeaux, France, May 2017, 71, https://hal.inria.fr/hal-01526713.

Other Publications

[25] R. LEBLOND, J.-B. ALAYRAC, A. OSOKIN, S. LACOSTE-JULIEN.*SEARNN: Training RNNs with global-local losses*, December 2017, https://arxiv.org/abs/1706.04499 - 12 pages, https://hal.inria.fr/hal-01665263.

[26] A. MIECH, I. LAPTEV, J. SIVIC.*Learnable pooling with Context Gating for video classification*, June 2017, https://arxiv.org/abs/1706.06905 - working paper or preprint, https://hal.inria.fr/hal-01547378.

[27] M. TRAGER, M. HEBERT, J. PONCE.*On point configurations, Carlsson-Weinshall duality, and multi-view geometry*, January 2018, working paper or preprint, https://hal.inria.fr/hal-01676732.