RESEARCH CENTER
Paris

FIELD

Activity Report 2017

Section New Results

Edition: 2018-02-19

ALMANACH Team

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]

7.2. Digital Humanities and Cultural Heritage

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

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

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.

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

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

(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.

7.9. Quantitative and computational morphology

Participant: Benoît Sagot.

[0] fang2014entity

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).

ALPINES Project-Team

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.

ANGE Project-Team

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.

ANTIQUE Project-Team

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/.

AOSTE2 Team

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.

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.

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

ARAMIS Project-Team

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 multi-layer 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]

CASCADE Project-Team

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

COML Team

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.

DYOGENE Project-Team

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.

EVA Project-Team

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).

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.*

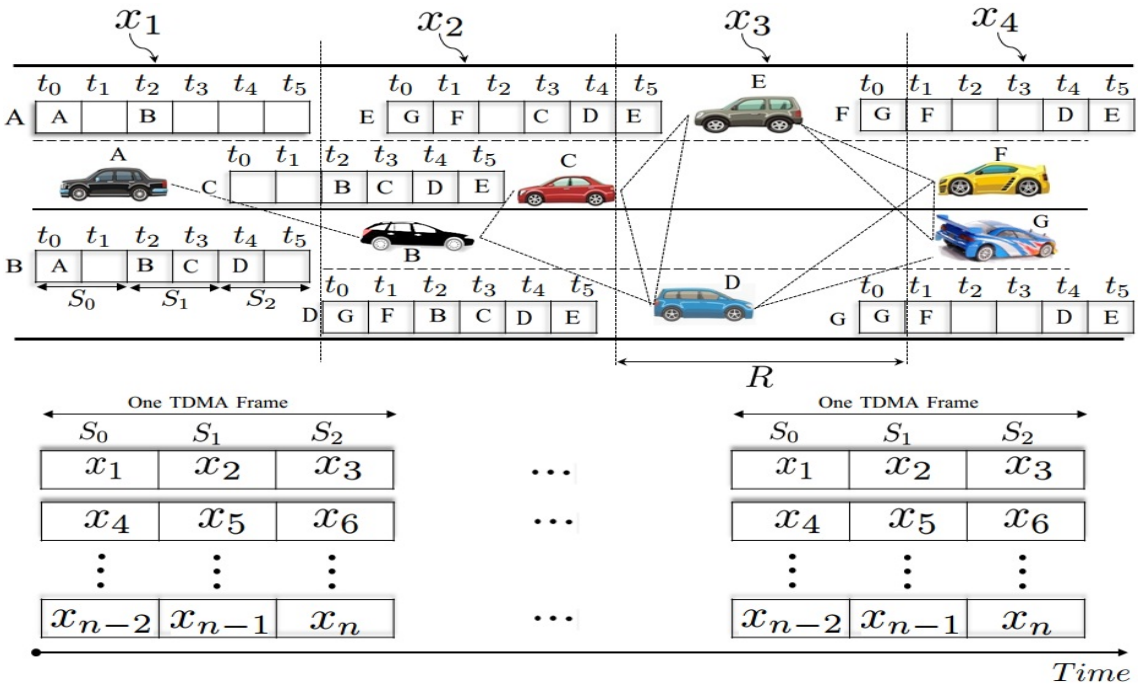

Figure 3. VANET network using DTMAC scheduling scheme.

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.

7.6.1.3. A Flooding-Based Location Service in VANETs

Participants: Selma Boumerdassi, Paul Muhlethaler.

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)

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.

GALLIUM Project-Team

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+.

GANG Project-Team

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.

MAMBA Project-Team

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].

MATHERIALS Project-Team

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.

MATHRISK Project-Team

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.

MIMOVE Team

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).

MOKAPLAN Project-Team

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.

MYCENAE Project-Team

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.

PARKAS Project-Team

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).

PI.R2 Project-Team

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.

POLSYS Project-Team

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.

PROSECCO Project-Team

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].

QUANTIC Project-Team

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.

RAP2 Team

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.

REGAL Project-Team

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.

REO Project-Team

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.

RITS Project-Team

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.

SECRET Project-Team

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.

SERENA Project-Team

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.

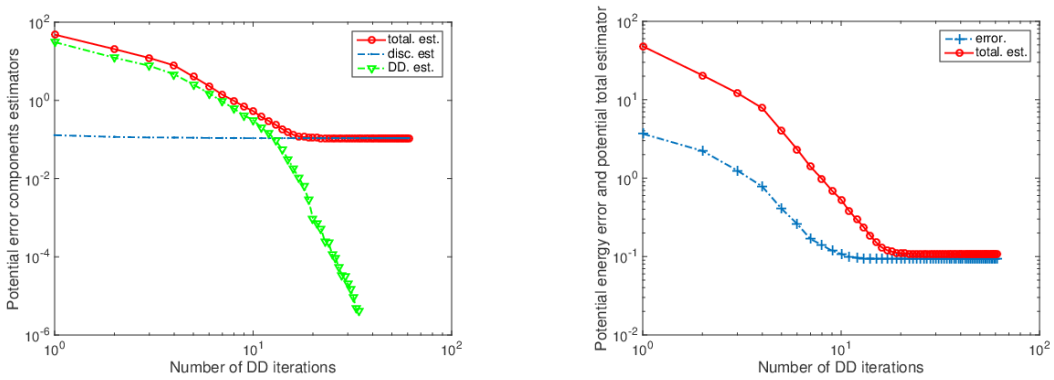

Figure 1. Error component estimates (left) and total energy error and its estimate (right), DD with GMRES solver

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.

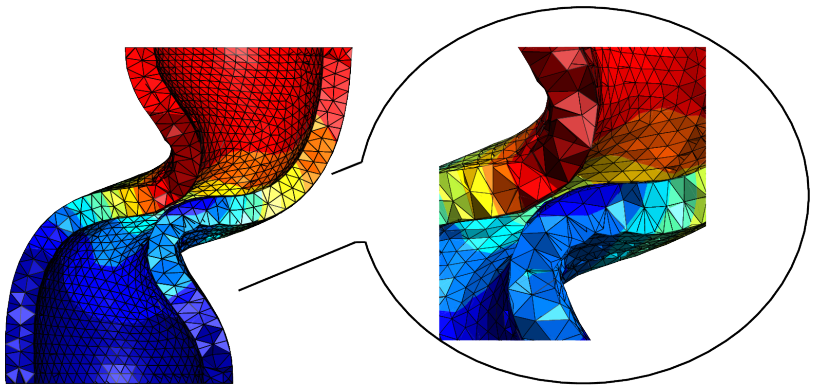

Figure 2. Euclidean displacement norm on the deformed configuration for the shear and compressed cylinder, and a zoom where the deformations are the most important. The color scale goes from 0.0 *(blue) to* 1.8 *(red).*

7.4. A nonlinear consistent penalty method for positivity preservation

Participant: Alexandre Ern.

Publication: [16]

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 .

7.5. A simple a posteriori estimate on general polytopal meshes

Participant: Martin Vohralík.

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* $4 \cdot 10^{-3}\%$.

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 .

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.

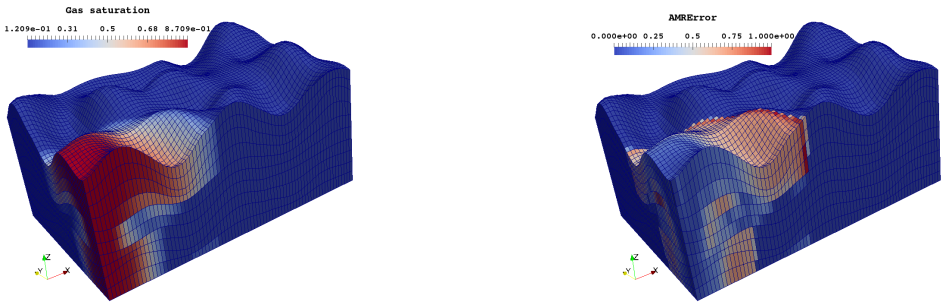

Figure 4. Simulated gas saturation after 1000 days (left) and corresponding a posteriori error estimate (right)

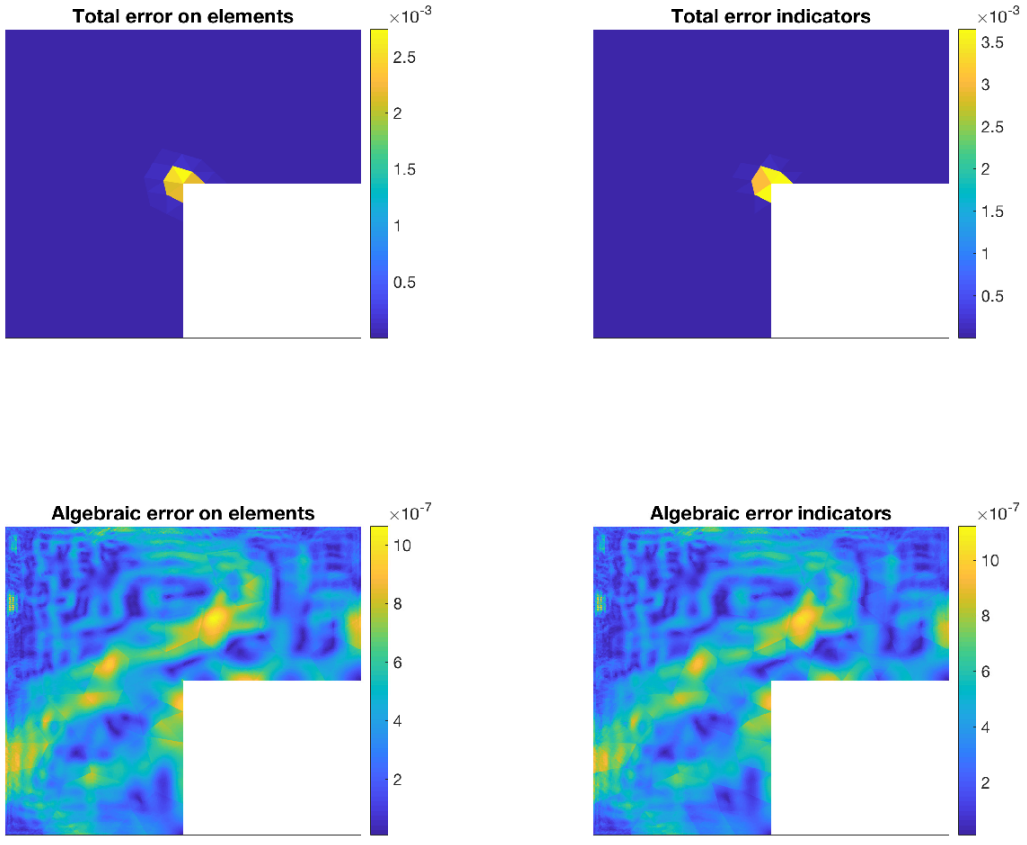

Figure 5. Actual total error (top left) and its a posteriori error estimate (top right). Actual algebraic error (bottom left) and its a posteriori error estimate (bottom right).

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.

SIERRA Project-Team

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.

TAPDANCE Team (section vide)

Valda Team

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.

WHISPER Project-Team

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.

WILLOW Project-Team

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].

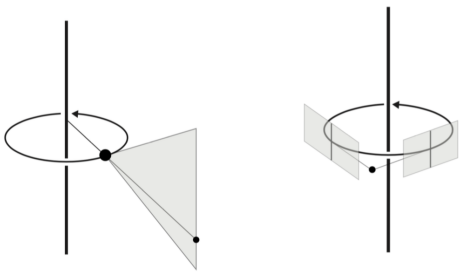

Figure 1. Non-central panoramic (left) and stereo panoramic cameras (right) are examples of non-linear cameras that can be modeled using line congruences.

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].

7.1.3. Changing Views on Curves and Surfaces

Participants: Kathlén Kohn, Bernd Sturmfels, Matthew Trager.

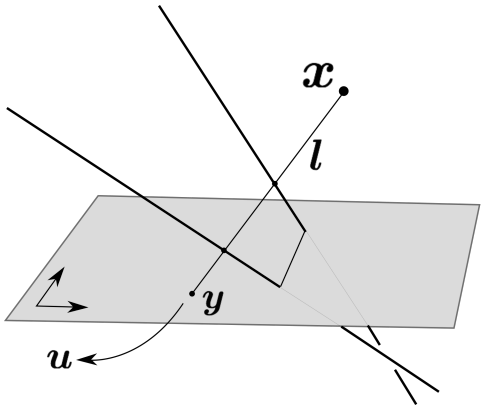

Figure 2. A general camera associates a scene point x with a visual ray l, then maps the ray l to its intersection y with some retinal plane π, and finally uses a projective coordinate system on π to express y as a point u in $\mathbb{P}^2$.

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].

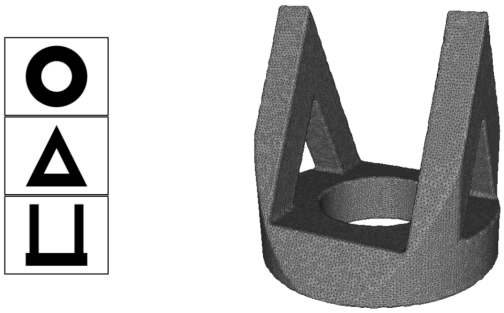

Figure 3. Changing views of a curve correspond to Reidemeister moves. The viewpoint z crosses the tangential surface (left), edge surface (middle), or trisecant surface (right).

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.

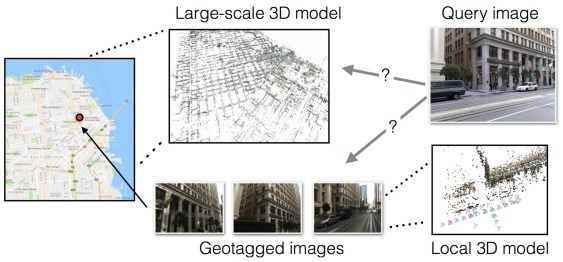

Figure 4. ***Large-scale visual localization.*** *2D image-based methods (bottom) use image retrieval and return the pose of the most relevant database image. 3D structure-based methods (top) use 2D-3D matches against a 3D model for camera pose estimation. Both approaches have been developed largely 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].

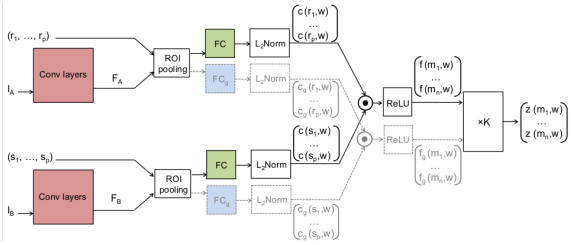

Figure 5. The SCNet architectures. Three variants are proposed: SCNet-AG, SCNet-A, and SCNet-AG+. The basic architecture, SCNet-AG, is drawn in solid lines. Colored boxes represent layers with learning parameters and the boxes with the same color share the same parameters. "$\times K$" denotes the voting 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.

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].

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

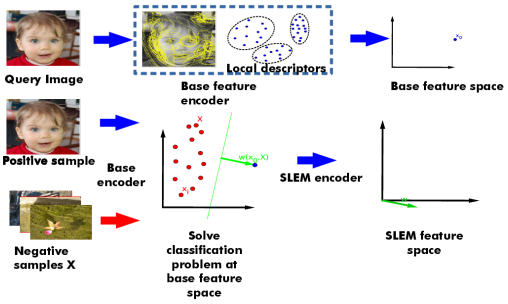

Figure 6. Pipeline of SLEM. First row encapsulates the construction of a base feature for a query image, which usually consists of extracting, embedding and aggregating local descriptors into a vector, here written as x_0. After repeating the process of base feature calculation a database of sample images 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.

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 .

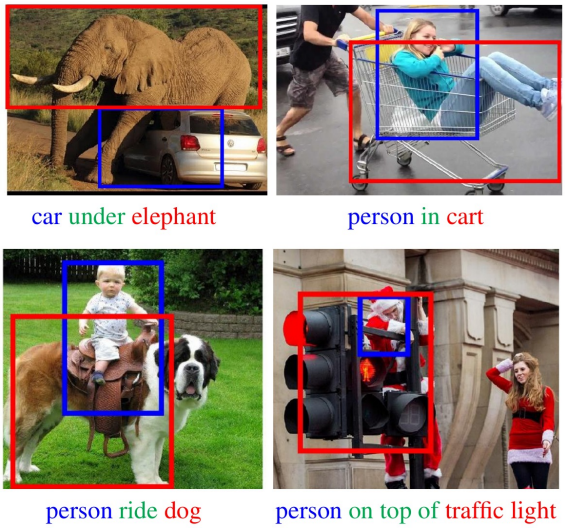

Figure 7. Examples of top retrieved pairs of boxes in UnRel dataset for unusual queries with our weakly supervised model

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].

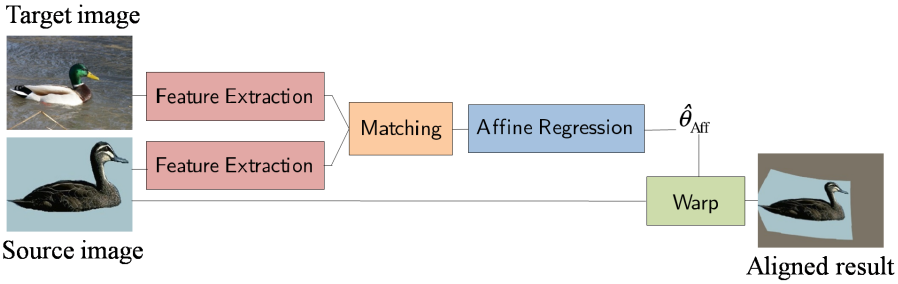

Figure 8. Proposed CNN architecture for geometric matching. Source and target images are passed through feature extraction networks which have tied parameters, followed by a matching network which matches the descriptors. The output of the matching network is passed through a regression network which 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.

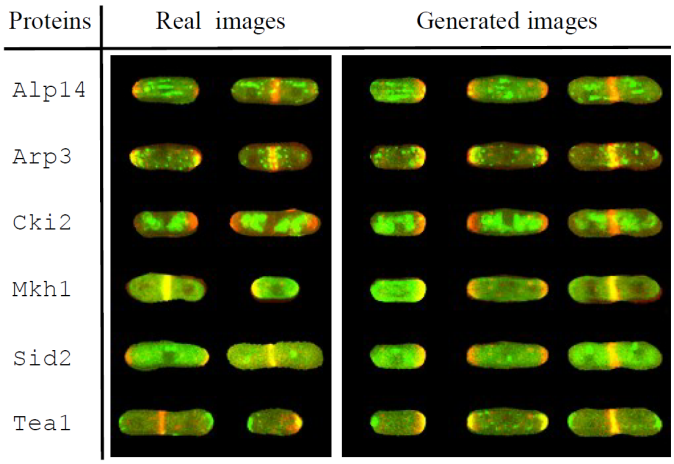

Figure 9. Real (left) and generated (right) images of fission yeast cells with protein bgs4 depicted in the red channel and 6 other proteins depicted in the green channel. The synthetic images were generated with our star-shaped GAN. The star-shaped model can generate multiple green channels aligned with the same red channel whereas the training images have only one green channel.

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 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].

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]

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

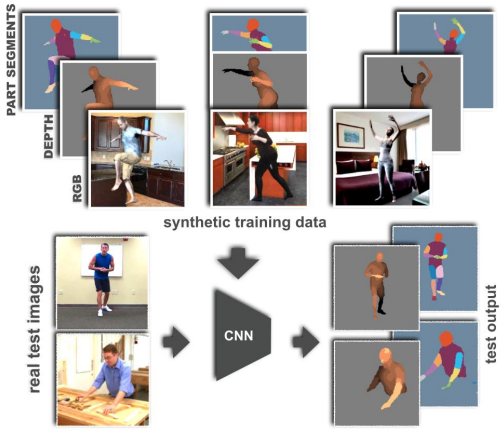

Figure 10. We generate photo-realistic synthetic images and their corresponding ground truth for learning pixel-wise classification problems: human parts segmentation and depth estimation. The convolutional neural network trained only on synthetic data generalizes on real images sufficiently for both tasks.

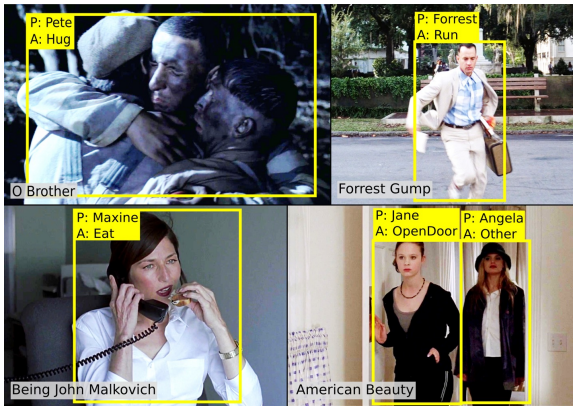

Figure 11. We automatically recognize actors and their actions in a of dataset of 66 movies with scripts as weak supervision

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 .

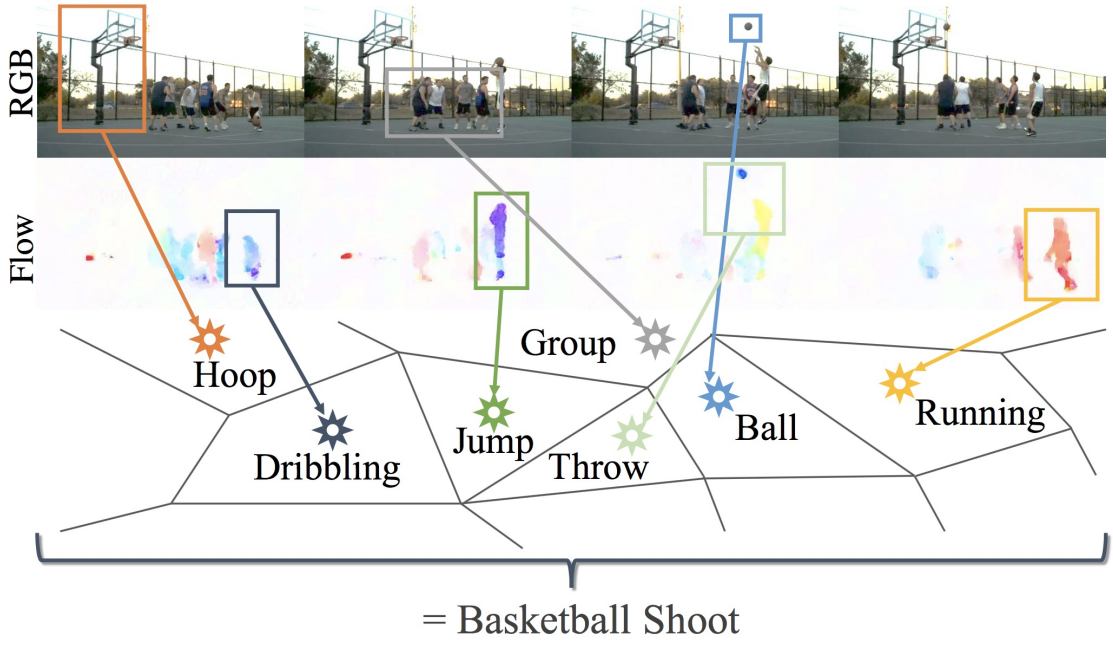

Figure 12. How do we represent actions in a video? We propose ActionVLAD, a spatio-temporal aggregation of a set of action primitives over the appearance and motion streams of a video. For example, a basketball shoot may be represented as an aggregation of appearance features corresponding to 'group of players', 'ball' and 'basketball hoop'; and motion features corresponding to 'run', 'jump', and 'shoot'.

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 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 .

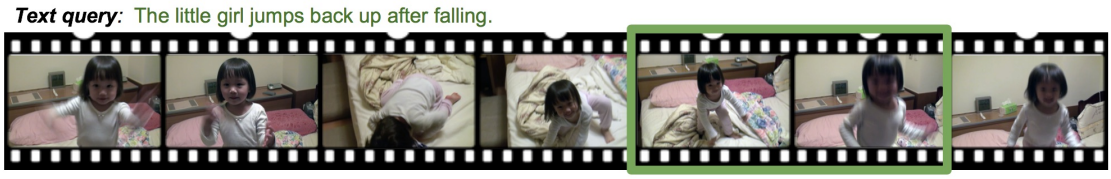

Figure 13. We consider localizing moments in video with natural language and demonstrate that incorporating local and global video features is important for this task. To train and evaluate our model, we collect the Distinct Describable Moments (DiDeMo) dataset which consists of over 40,000 pairs of 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].

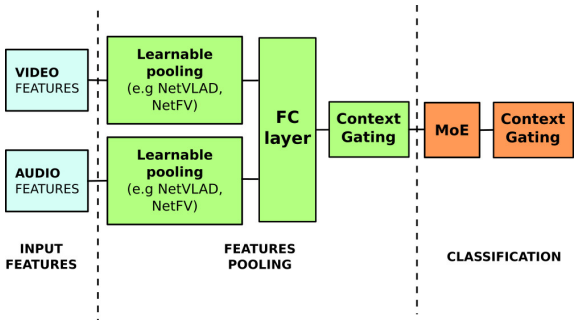

Figure 14. Overview of our network architecture for video classification