

RESEARCH CENTER Grenoble - Rhône-Alpes

FIELD

Activity Report 2017

Section New Results

Edition: 2018-02-19

1. AGORA Team
2. AIRSEA Project-Team
3. ARIC Project-Team
4. AVALON Project-Team
5. BEAGLE Project-Team
6. BIPOP Project-Team
7. CHROMA Project-Team
8. CONVECS Project-Team
9. CORSE Project-Team
10. CTRL-A Project-Team
11. DANTE Project-Team
12. DATAMOVE Project-Team
13. DATASPHERE Team
14. DRACULA Project-Team
15. ERABLE Project-Team
16. IBIS Project-Team
17. IMAGINE Project-Team
18. MAVERICK Project-Team
19. MISTIS Project-Team 122
20. MOEX Project-Team
21. MORPHEO Project-Team
22. NANO-D Project-Team
23. NECS Project-Team
24. NUMED Project-Team (section vide)
25. PERCEPTION Project-Team
26. PERVASIVE INTERACTION Project-Team
27. POLARIS Team
28. PRIVATICS Project-Team
29. ROMA Project-Team
30. SOCRATE Project-Team
31. SPADES Project-Team
32. STEEP Project-Team
33. THOTH Project-Team
34. TYREX Project-Team

AGORA Team

7. New Results

7.1. Wireless network deployment

Participants: Ahmed Boubrima, Rodrigue Domga Komguem, Leo Le Taro, Jad Oueis, Walid Bechkit, Khaled Boussetta, Hervé Rivano, Razvan Stanica, Fabrice Valois.

7.1.1. Deployment of Wireless Sensor Networks for Pollution Monitoring

Air pollution has become a major issue of modern megalopolis because of industrial emissions and increasing urbanization along with traffic jams and heating/cooling of buildings. Monitoring urban air quality is therefore required by municipalities and by the civil society. Current monitoring systems rely on reference sensing stations that are precise but massive, costly and therefore seldom. In our work, we focus on an alternative or complementary approach, with a network of low cost and autonomic wireless sensors, aiming at a finer spatiotemporal granularity of sensing. Generic deployment models of the literature are not adapted to the stochastic nature of pollution sensing.

In this sense, in [2], our main contribution is to design integer linear programming models that compute sensor deployments capturing both the coverage of pollution under time-varying weather conditions and the connectivity of the infrastructure. We evaluate our deployment models on a real data set of Greater London. We analyze the performance of the proposed models and show that our joint coverage and connectivity formulation is tight and compact, with a reasonable enough execution time. We also conduct extensive simulations to derive engineering insights for effective deployments of air pollution sensors in an urban environment.

Unlike most of the existing methods, which rely on simple and generic detection models, our approach is based on the spatial analysis of pollution data, allowing to take into account the nature of the pollution phenomenon. As proof of concept, we apply our approach on real world data, namely the Paris pollution data, which was recorded in March 2014 [7]. In this paper, we consider citywide wireless sensor networks and tackle the minimum-cost node positioning issue for air pollution monitoring. We propose an efficient approach that aims to find optimal sensors and sinks locations while ensuring air pollution coverage and network connectivity.

Mobile wireless sensor networks can also be used for monitoring air pollution, where the aim is usually to generate accurate pollution maps in real time. The generation of pollution maps can be performed using either sensor measurements or physical models which simulate the phenomenon of pollution dispersion. The combination of these two information sources, known as data assimilation, makes it possible to better monitor air pollution by correcting the simulations of physical models while relying on sensor measurements. The quality of data assimilation mainly depends on the number of measurements and their locations. A careful deployment of nodes is therefore necessary in order to get better pollution maps. In an ongoing work [30], we tackle the placement problem of pollution sensors and design a mixed integer programming model allowing to maximize the assimilation quality while ensuring the connectivity of the network. We perform some simulations on a dataset of the city of Lyon in order to show the effectiveness of our model regarding the quality of pollution coverage.

For an air pollution monitoring system deployment to be relevant relative to urban air quality aspects, we are concerned with maintaining the system properties over time. Indeed, one of the major drawbacks of cheap sensors is their drift: chemical properties degrade over time and alter the measurement accuracy. We challenge this issue by designing distributed, online recalibration procedures. In [16], we present a simulation framework modelling a mobile wireless sensor network (WSN) and we assess the system's measurement confidence using trust propagation paradigms. As WSN calibrations translate to information exchange between sensors, we also study means of limiting the number of such transmissions by skipping the calibrations deemed least profitable to the system.

7.1.2. Wireless Sensor Networks with Linear Topology

In wireless sensor networks with linear topology, knowing the physical order in which nodes are deployed is useful not only for the target application, but also to some network services, like routing or data aggregation. Considering the limited resources of sensor nodes, the design of autonomous protocols to find this order is a challenging topic.

In [9], we propose a distributed and iterative centroid-based algorithm to address this problem. At each iteration, the algorithm selects two virtual anchors and finds the order of a subset of nodes, placed between these two anchors. The proposed algorithm requires local node connectivity knowledge and the identifier of the first sensor node of the network, which is the only manually configured parameter. This solution, scalable and lightweight from the deployment and maintenance point of view, is shown to be robust to connectivity degradation, correctly ordering more than 95% of the nodes, even under very low connectivity conditions.

7.1.3. Function Placement in Public Safety Networks

In response to the growing demand in the public safety community for broadband communication systems, LTE is currently being adopted as the base technology for next generation public safety networks. In parallel, notable efforts are being made by the 3GPP to enhance the LTE standard in order to offer public safety oriented services. In the recent Release 13, the Isolated E-UTRAN Operation for Public Safety (IOPS) concept was introduced. IOPS aims at maintaining a level of communication between public safety users, offering them local mission-critical services even when the backhaul connectivity to the core network is not fully functional. Isolated operation is usually needed in mission-critical situations, when the infrastructure is damaged or completely destroyed, and in out of coverage areas. In [6], we present a detailed technical overview on the IOPS specifications, and then identify several research prospects and development perspectives opened up by IOPS.

An isolated base station is a base station having no connection to a traditional core network. To provide services to users, an isolated base station is co-located with an entity providing the same functionalities as the traditional core network, referred to as Local EPC. In order to cover wider areas, several base stations are interconnected, forming a network that should be served by a single Local EPC. In [20], [24], we tackle the Local EPC placement problem in the network, to determine with which of the base stations the Local EPC must be co-located. We propose a novel centrality metric, flow centrality, which measures the capacity of a node to receive the total amount of flows in the network. We show that co-locating the Local EPC with the base station having the maximum flow centrality maximizes the total amount of traffic the Local EPC can receive from all base stations, under certain capacity and load distribution constraints. We compare the flow centrality to other state of the art centrality metrics, and emphasize its advantages.

7.1.4. User Association in Public Safety Oriented Mobile Networks

In many disaster scenarios, communication infrastructure fails to provide network services for both civilians and first responders. One solution is to have rapidly deployable mobile networks formed by interconnected base stations, that are easy to move, deploy, and configure. Such public safety-oriented networks are different from classical mobile networks in terms of scale, deployment, and architecture.

In this context, we revisit the user association problem [21], for two main reasons. First, the backhaul, formed by the links interconnecting the base stations, must be accounted for when deciding on the association, since it may present a bottleneck with its limited bandwidth. Second, the mission-critical nature of the traffic imposes strict guaranteed bit rate constraints, that must be respected when associating users. Therefore, we propose a network-aware optimal association that minimizes the bandwidth consumption on the backhaul, while still respecting the stringent performance requirements.

7.2. Wireless data collection

Participants: Yosra Bahri Zguira, Alexis Duque, Junaid Ahmed Khan, Abdoul-Aziz Mbacké, Romain Pujol, Hervé Rivano, Razvan Stanica, Fabrice Valois.

7.2.1. Smart Parking Systems

Considering the increase of urban population and traffic congestion, smart parking is always a strategic issue to work on, not only in the research field but also from economic interests. Thanks to information and communication technology evolution, drivers can more efficiently find satisfying parking spaces with smart parking services. The existing and ongoing works on smart parking are complicated and transdisciplinary. While deploying a smart parking system, cities, as well as urban engineers, need to spend a very long time to survey and inspect all the possibilities. Moreover, many varied works involve multiple disciplines, which are closely linked and inseparable.

To give a clear overview, we introduce a smart parking ecosystem and propose a comprehensive and thoughtful classification by identifying their functionalities and problematic focuses [5]. We go through the literature over the period of 2000-2016 on parking solutions as they were applied to smart parking development and evolution, and propose three macro-themes: information collection, system deployment, and service dissemination. In each macro-theme, we explain and synthesize the main methodologies used in the existing works and summarize their common goals and visions to solve current parking difficulties. Lastly, we give our engineering insights and show some challenges and open issues.

7.2.2. Data Offloading

Mobile users in an urban environment access content on the Internet from different locations. It is challenging for the current service providers to cope with the increasing content demand from a large number of collocated mobile users. In-network caching to offload content at nodes closer to users alleviates the issue, though efficient cache management is required to find out who should cache what, when and where in an urban environment, given nodes limited computing, communication and caching resources. To address this [14], we first define a novel relation between content popularity and availability in the network and investigate a node eligibility to cache content based on its urban reachability. We then allow nodes to self-organize into mobile fogs to increase the distributed cache and maximize content availability in a cost-effective manner. However, to cater rational nodes, we propose a coalition game for the nodes to offer a maximum virtual cache assuming a monetary reward is paid to them by the service/content provider. Nodes are allowed to merge into different spatio-temporal coalitions in order to increase the distributed cache size at the network edge. Results obtained through simulations using realistic urban mobility trace validate the performance of our caching system showing a ratio of 60 - 85% of cache hits compared to the 30 - 40% obtained by the existing schemes and 10% in case of no coalition.

Another option for data offloading is represented by vehicular traffic. With over 300 billion vehicle trips made in the USA and 64 billion in France per year, network operators have the opportunity to utilize the existing road and highway network as an alternative data network to offload large amounts of delay-tolerant traffic. To enable the road network as a large-capacity transmission system, we exploit the existing mobility of vehicles equipped with wireless and storage capacities together with a collection of offloading spots [1]. An offloading spot is a data storage equipment located where vehicles usually park. Data is transloaded from a conventional data network to the closest offloading spot and then shipped by vehicles along their line of travel. The subsequent offloading spots act as data relay boxes where vehicles can drop off data for later pickups by other vehicles, depending on their direction of travel. The main challenges of this offloading system are how to compute the road path matching the performance requirements of a data transfer and how to configure the sequence of offloading spots involved in the transfer. We propose a scalable and adaptive centralized architecture built on SDN that maximizes the utilization of the flow of vehicles connecting consecutive offloading spots. We simulate the performance of our system using real roads traffic counts for France. Results show that the centralized controlled offloading architecture can achieve an efficient and fair allocation of concurrent data transfers between major cities in France.

7.2.3. Hybrid Short/Long Range Networks

Despite the success of dedicated IoT networks, such as Sigfox or LoRa, several use cases can not be accommodated by these new technologies, mainly because of capacity constraints. For example, mobile sensing and proximity-based applications require smart devices to find other nodes in vicinity, though it is

challenging for a device to find neighbors in an energy efficient manner, while also running on low duty cycles.

Neighbor discovery schemes allow nodes to follow a schedule to become active and send beacons or listen for other active nodes in order to discover each other with a bounded latency. However, a trade-off exists between the energy consumption and the time a node takes to discover neighbors using a given activity schedule. Moreover, energy consumption is not the only bottleneck, as theoretically perfect schedules can result in discovery failures in a real environment. In [12], we provide an in-depth study on neighbor discovery, by first defining the relation between energy efficiency, discovery latency and the fraction of discovered neighbors. We evaluate existing mechanisms using extensive simulations for up to 100 nodes and testbed implementations for up to 15 nodes, with no synchronization between nodes and using duty cycles as low as 1% and 5%. Moreover, the literature assumes that multiple nodes active simultaneously always result in neighbor discovery, which is not true in practice as this can lead to collisions between the transmitted messages. Our findings reveal such scalability issues in existing schemes, where discovery fails because of collisions between beacons from multiple nodes active at the same time. Therefore, we show that energy efficient discovery schemes do not necessarily result in successful discovery of all neighbors, even when the activity schedules are computed in a deterministic manner.

A second use-case requiring a combination of long range and short range communications is related to intelligent transportation systems. As a matter of fact, communication is essential to the coordination of public transport systems. Nowadays, cities are facing an increasing number of bikes used by citizens therefore the need of monitoring and managing their traffic becomes crucial. Public bike sharing system has been introduced as an urban transportation system that can collect data from mobile devices. In this context, we introduce IoB-DTN [29], a protocol based on the Delay/Disruption Tolerant Network (DTN) paradigm adapted for an IoT-like applications running on bike sharing system based sensor network. We present simulation results obtained by evaluating the Binary Spray and Wait inspired variant of IoB-DTN with four buffer management policies and by comparing three variants of IoB-DTN by varying the number of packet copies sprayed in the network.

7.2.4. Visible Light Communications in IoT Networks

With the increasing consumer demand for smart objects, Visible Light Communications (VLC), and especially LED-to-Camera communication, appears as a low-cost alternative to radio to make any conventional device smart. Since LEDs are already on most electronics devices, that is achieved at the cost of negligible hardware modifications. However, as these LEDs are very different from the widely studied ceiling ones, several challenges need to be addressed to make this happen. In our work [31], we propose a line of sight bi-directional communication system between an ordinary LED and an off-the-shelf smartphone. We designed a cheap multi sensors device as a proof of concept of a near communication module for the IoT.

Among the issues we observed experimenting with this platform, we note the constrained physical layer data unit (PHY-SDU) length that complicates the use of coding strategies to cope with bits or packets erasure. To break this limitation, we present SeedLight [8], a coding scheme designed to face the inherent packet losses and enhance line-of-sight LED-to-Camera communication goodput. SeedLight leverages random linear coding to provide an efficient redundancy mechanism that works even on PHY-SDU of tens of bits. The key idea of SeedLight is to reduce the code overhead by replacing the usual coding coefficients by a seed. Since this work addresses IoT devices with low computational resources, SeedLight encoding algorithm complexity remains low. We develop an implementation of SeedLight on a low-cost MCU and a smartphone to evaluate both the communication and algorithmic performances. Experimental results show that SeedLight introduces a negligible overhead and can be implemented even on the cheapest MCU, such as the ones used in many IoT devices. The achievable goodput can be up to 2.5kbps, while the gain compared to a trivial retransmissions scheme is up to 100%.

To ease the evaluation of VLC systems, we present CamComSim [28], the first simulator for development and rapid prototyping of LED-to-Camera communication systems. Our event driven simulator relies on a standalone Java application that is easily extensible through a set of interfaces. A range of low and highlevel parameters, such as the camera characteristics, the PHY-SDU size, or the redundancy mechanism can be chosen. CamComSim uses empirically validated models for the LED-to-Camera channel and the broadcast protocols, configurable with a finely grained precision. To validate CamComSim implementation and accuracy, we use the previously discussed testbed, based on a color LED and a smartphone, and compare the performance reached by the testbed with the results given by our simulator. We illustrate with a real use case the full usage of CamComSim, tuning a broadcast protocol that implements the transmission of 1 kbyte of information. The results highlight that our simulator is very precise and predicts the performance of a real LED-to-Camera system with less than 10% of error in most cases.

7.2.5. Data Collection with RFID Devices

The popularization of Radio Frequency Identification (RFID) systems has conducted to large deployments of RFID solutions in various areas under different criteria. However, such deployments, specially in dense environments, can be subject to RFID collisions which in turn affect the quality of readings. In [17], [18], we propose two distributed and efficient solutions for dense mobile deployments of RFID systems. mDEFAR is an adaptation of a previous work highly performing in terms of collisions reduction, efficiency and fairness in dense static deployments. CORA is more of a locally mutual solution where each reader relies on its neighborhood to enable itself or not. Using a beaconing mechanism, each reader is able to identify potential (non-)colliding neighbors in a running frame and as such chooses to read or not. Performance evaluation shows high performance in terms of coverage delay for both proposals quickly achieving 100% coverage depending on the considered use case while always maintaining consistent efficiency levels above 70%. Compared to GDRA, our solutions proved to be better suited for highly dense and mobile environments, offering both higher throughput and efficiency. The results reveal that depending on the application considered, choosing either mDEFAR or CORA helps improve efficiency and coverage delay.

RFID solutions encounter two main issues: the first one is inherent to the technology itself which is readers collisions, the second one being the gathering of read data up to a base station, potentially in a multihop fashion. While the first one has been a main research subject in the late years, the second one has not been investigated for the sole purpose of RFID, but rather for wireless adhoc networks. This multihop tag information collection must be done in regards of the application requirements but it should also care for the deployment strategy of readers to take advantage of their relative positions, coverage, reading activity and deployment density to avoid interfering between tag reading and data forwarding. To the best of our knowledge, the issue for a joint scheduling between tag reading and forwarding has never been investigated so far in the literature, although important. In [17], we propose two new distributed, cross-layer solutions meant for the reduction of collisions and better efficiency of the RFID system, but also serving as a routing solution towards a base station. Simulations show high levels of throughput while not lowering on the fairness on medium access staying above 85% in the highest deployment density with up to 500 readers, also providing a 90% increase in data rate.

7.3. Network data exploitation

Participants: Panagiota Katsikouli, Elli Zavou, Stéphane D'Alu, Hervé Rivano, Razvan Stanica.

7.3.1. Spatio-temporal Characterization of Mobile Data Traffic

Mobile traffic data collected by network operators is a rich source of information about human habits, and its analysis provides insights relevant to many fields, including urbanism, transportation, sociology and networking. Urban landscapes present a variety of socio-topological environments that are associated to diverse human activities. As the latter affect the way individuals connect with each other, a bound exists between the urban tissue and the mobile communication demand. In [3], we investigate the heterogeneous patterns emerging in the mobile communication activity recorded within metropolitan regions. To that end, we introduce an original technique to identify classes of mobile traffic signatures that are distinctive of different urban fabrics. Our proposed technique outperforms previous approaches when confronted to ground-truth information, and allows characterizing the mobile demand in greater detail than that attained in the literature to date. We apply our technique to extensive real-world data collected by major mobile operators in ten cities. Results unveil the diversity of baseline communication activities across countries, but also evidence

the existence of a number of mobile traffic signatures that are common to all studied areas and specific to particular land uses.

Similarly to mobile phone data, GPS traces of vehicles convey information on transportation demand and human activities that can be related to the land use of the neighborhood where they take place. In [10], we investigate the land use patterns that emerge when studying simultaneously GPS traces of probe vehicles and mobile phone data collected by network providers. To this end, we extend previous definitions of mobile phone traffic signatures for land use detection, so as to incorporate additional information on human presence and mobility conveyed by GPS traces of vehicles. Leveraging these extended signatures, we exploit an unsupervised learning technique to identify classes of signatures that are distinctive of different land use. We apply our technique to real-world data collected in French and Italian cities. Results unveil the existence of signatures that are common to all studied areas and specific to particular land uses. The combined use of mobile phone data and GPS traces outperforms previous approaches when confronted to ground-truth information, and allows characterizing land use in greater detail than in the literature to date.

The spatial and temporal profiles of mobile phone traffic can be studied simultaneously. In [11], we present an original approach to infer both spatial and temporal structures hidden in the mobile demand, via a first-time tailoring of Exploratory Factor Analysis (EFA) techniques to the context of mobile traffic datasets. Casting our approach to the time or space dimensions of such datasets allows solving different problems in mobile traffic analysis, i.e., network activity profiling and land use detection, respectively. Tests with real-world mobile traffic datasets show that, in both its variants above, the proposed approach (i) yields results whose quality matches or exceeds that of state-of-the-art solutions, and (ii) provides additional joint spatiotemporal knowledge that is critical to result interpretation.

7.3.2. Using Mobile Phone Data in Cognitive Networking

In the next few years, mobile networks will undergo significant evolutions in order to accommodate the ever-growing load generated by increasingly pervasive smartphones and connected objects. Among those evolutions, cognitive networking upholds a more dynamic management of network resources that adapts to the significant spatiotemporal fluctuations of the mobile demand. Cognitive networking techniques root in the capability of mining large amounts of mobile traffic data collected in the network, so as to understand the current resource utilization in an automated manner. In [4], we take a first step towards cellular cognitive networks by proposing a framework that analyzes mobile operator data, builds profiles of the typical demand, and identifies unusual situations in network-wide usages. We evaluate our framework on two real-world mobile traffic datasets, and show how it extracts from these a limited number of meaningful mobile demand profiles. In addition, the proposed framework singles out a large number of outlying behaviors in both case studies, which are mapped to social events or technical issues in the network.

7.3.3. Study of Wi-Fi Localization from Crowdsourced Datasets

The wide adoption of mobile devices has created unprecedented opportunities to collect mobility traces and make them available for the research community to conduct interdisciplinary research. However, mobility traces available in the public domain are usually restricted to traces resulting from a single sensor (e.g., either GPS, GSM or WiFi). In [26], we present the PRIVA'MOV dataset, a novel dataset collected in the city of Lyon, France on which user mobility has been collected using multiple sensors. More precisely, this dataset contains mobility traces of about 100 persons including university students, staff and their family members over 15 months collected through the GPS, WiFi, GSM, and accelerometer sensors. We provide both a quantitative and a preliminary qualitative analysis of this dataset. Specifically, we report the number of visited points of interests, GSM antennas and WiFi hotspots and their distribution across the various users. We finally analyse the uniqueness of human mobility by considering the various sensors.

Thanks to this collected data, it is possible to combine information from several probes. A very common use case is the collection of network scans with location to help the localisation feature of these devices. Nevertheless, most users are not aware of this spying. The collected data might represent infringements of privacy. One possible solution to keep gathering these data while maintaining privacy would consist in device-to-device communications in order to break the links between data and users. In [22], we propose an approach

to test the feasibility of such a system. We collected data from mobile users to combine location and network scans data. With this data, we test the accuracy level we can reach while using Wi-Fi localisation. We analyse how a new measure should be pushed and how many scans should be realised to provide location-based Wi-Fi. We analyse the minimal dataset to cover the set of locations covered by users and prove that a multiuser gathering system can benefit the users.

AIRSEA Project-Team

7. New Results

7.1. Modeling for Oceanic and Atmospheric flows

7.1.1. Numerical Schemes for Ocean Modelling

Participants: Eric Blayo, Laurent Debreu, Florian Lemarié, Christopher Eldred, Farshid Nazari.

The increase of model resolution naturally leads to the representation of a wider energy spectrum. As a result, in recent years, the understanding of oceanic submesoscale dynamics has significantly improved. However, dissipation in submesoscale models remains dominated by numerical constraints rather than physical ones. Effective resolution is limited by the numerical dissipation range, which is a function of the model numerical filters (assuming that dispersive numerical modes are efficiently removed). A review paper on coastal ocean models has been written with German colleagues and will be published in Ocean Modelling early 2018 ([34]). Ocean models usually rely on a mode splitting procedure which separates the fast external gravity waves with the slower internal waves. A paper on the stability of the mode splitting has been submitted to Journal of Computational Physics ([21])

The team is involved in the HEAT (Highly Efficient ATmospheric Modelling) ANR project. This project aims at developing a new atmospheric dynamical core (DYNAMICO) discretized on an icosahedral grid. This project is in collaboration with Ecole Polytechnique, Meteo-France, LMD, LSCE and CERFACS. This year we worked on dispersion analysis of compatible Galerkin schemes for a 1D shallow water model ([8]).

7.1.2. Coupling Methods for Oceanic and Atmospheric Models

Participants: Eric Blayo, Laurent Debreu, Florian Lemarié, Charles Pelletier, Antoine Rousseau, Sophie Thery.

Coupling methods routinely used in regional and global climate models do not provide the exact solution to the ocean-atmosphere problem, but an approximation of one [61]. For the last few years we have been actively working on the analysis of ocean-atmosphere coupling both in terms of its continuous and numerical formulation. Our activities over the last few years can be divided into four general topics

- 1. *Stability and consistency analysis of existing coupling methods*: in [61] we showed that the usual methods used in the context of ocean-atmosphere coupling are prone to splitting errors because they correspond to only one iteration of an iterative process without reaching convergence. Moreover, those methods have an additional condition for the coupling to be stable even if unconditionally stable time stepping algorithms are used. This last remark was further studied recently in [1] and it turned out to be a major source of instability in atmosphere-snow coupling.
- 2. Study of physics-dynamics coupling: during the PhD-thesis of Charles Pelletier (funded by Inria) the scope is on including the formulation of physical parameterizations in the theoretical analysis of the coupling, in particular the parameterization schemes to compute air-sea fluxes [18]. To do so, a metamodel representative of the behavior of the full parameterization but with a continuous form easier to manipulate has been derived thanks to a sensitivity analysis based on Sobol' indexes. This metamodel has the advantage to be more adequate to conduct the mathematical analysis of the coupling while being physically satisfactory. This work is in revision for publication in Quarterly Journal of the Royal Meteorological Society and has been presented in various conferences [69], [24], [20], [17]. In parallel we have contributed to a general review gathering the main international specialists on the topic [53].

- 3. Design of a coupled single column model: in order to focus on specific problems of oceanatmosphere coupling, a work on simplified equation sets has been started. The aim is to implement a one-dimensional (in the vertical direction) coupled model with physical parameterizations representative of those used in realistic models. Thanks to this simplified coupled model the objective is to develop a benchmark suite for coupled models evaluation. Last year the single column oceanic and atmospheric components have been developed and coupled during the PhD-thesis of Rémi Pellerej and in the framework of the SIMBAD project. A publication describing this model and its interfacing with the OOPS software to allow the implementation of various data assimilation techniques is currently in preparation for the Geoscientific Model Development journal.
- 4. Analysis of air-sea interactions in realistic high-resolution realistic simulations: part of our activity has been in collaboration with atmosphericists and physical oceanographers to study the impact on some modeling assumptions (e.g. [62]) in large-scale realistic ocean-atmosphere coupled simulations [70], [66], [12].

These four topics are addressed through strong collaborations between the applied mathematicians and the climate community.

Moreover a PPR (*Projet à partenariat renforcé*) called SIMBAD (SIMplified Boundary Atmospheric layer moDel for ocean modeling purposes) is funded by Mercator-Ocean for the next three years (from march 2015 to march 2018). The aim of this project in collaboration with Meteo-France, Ifremer, LMD, and LOCEAN is to derive a metamodel to force high-resolution oceanic operational models for which the use of a full atmospheric model is not possible due to a prohibitive computational cost. First results have been presented during international conferences [22], [23] and a publication is currently in preparation. Another industrial contract named ALBATROS is also funded by (from June 2016 to June 2019) to couple SIMBAD with the NEMO global ocean model and a wave model called WW3.

An ANR project COCOA (COmprehensive Coupling approach for the Ocean and the Atmosphere, P.I.: E. Blayo) has been funded in 2016 and has officially start in January 2017.

7.1.3. Data assimilation for coupled models

In the context of operational meteorology and oceanography, forecast skills heavily rely on proper combination of model prediction and available observations via data assimilation techniques. Historically, numerical weather prediction is made separately for the ocean and the atmosphere in an uncoupled way. However, in recent years, fully coupled ocean-atmosphere models are increasingly used in operational centers to improve the reliability of seasonal forecasts and tropical cyclones predictions. For coupled problems, the use of separated data assimilation schemes in each medium is not satisfactory since the result of such assimilation process is generally inconsistent across the interface, thus leading to unacceptable artefacts. Hence, there is a strong need for adapting existing data assimilation techniques to the coupled framework. As part of our ERACLIM2 contribution, R. Pellerej started a PhD on that topic late 2014 and will defend it early 2018. So far, three general data assimilation algorithms, based on variational data assimilation techniques, have been developed and applied to a single column coupled model. The dynamical equations of the considered problem are coupled using an iterative Schwarz domain decomposition method. The aim is to properly take into account the coupling in the assimilation process in order to obtain a coupled solution close to the observations while satisfying the physical conditions across the air-sea interface. Preliminary results shows significant improvement compared to the usual approach on this simple system [68], [25]. The aforementioned system has been coded within the OOPS framework (Object Oriented Prediction System) in order to ease the transfer to more complex/realistic models.

The second contribution to ERACLIM2 was to investigate the importance of the quality of the data assimilation scheme in the ocean in the coupled system. It led to the proposition of cost effective approximations either in term of resolution reduction or equation simplifications, along with a metric to asses the quality of said approximations [35]

Finally, CASIS, a new collaborative project with Mercator Océan has started late 2017 in order to extend developments to iterative Kalman smoother data assimilation scheme, in the framework of a coupled oceanatmospheric boundary layer context.

7.1.4. Parameterizing subgrid scale eddy effects

Participant: Eugene Kazantsev.

Basing on the maximum entropy production principle, the influence of subgrid scales on the flow is presented as the harmonic dissipation accompanied by the backscattering of the dissipated energy. This parametrization is tested on the shallow water model in a square box. Two possible solutions of the closure problem are compared basing on the analysis of the energy dissipation-backscattering balance. Results of this model on the coarse resolution grid are compared with the reference simulation at four times higher resolution. It is shown that the mean flow is correctly recovered, as well as variability properties, such as eddy kinetic energy fields and its spectrum [33].

7.2. Model reduction / multiscale algorithms

7.2.1. Intrusive sensitivity analysis, reduced models

Participants: Maëlle Nodet, Clémentine Prieur.

Another point developed in the team for sensitivity analysis is model reduction. To be more precise regarding model reduction, the aim is to reduce the number of unknown variables (to be computed by the model), using a well chosen basis. Instead of discretizing the model over a huge grid (with millions of points), the state vector of the model is projected on the subspace spanned by this basis (of a far lesser dimension). The choice of the basis is of course crucial and implies the success or failure of the reduced model. Various model reduction methods offer various choices of basis functions. A well-known method is called "proper orthogonal decomposition" or "principal component analysis". More recent and sophisticated methods also exist and may be studied, depending on the needs raised by the theoretical study. Model reduction is a natural way to overcome difficulties due to huge computational times due to discretizations on fine grids. In [57], the authors present a reduced basis offline/online procedure for viscous Burgers initial boundary value problem, enabling efficient approximate computation of the solutions of this equation for parametrized viscosity and initial and boundary value data. This procedure comes with a fast-evaluated rigorous error bound certifying the approximation procedure. The numerical experiments in the paper show significant computational savings, as well as efficiency of the error bound.

When a metamodel is used (for example reduced basis metamodel, but also kriging, regression, ...) for estimating sensitivity indices by Monte Carlo type estimation, a twofold error appears: a sampling error and a metamodel error. Deriving confidence intervals taking into account these two sources of uncertainties is of great interest. We obtained results particularly well fitted for reduced basis metamodels [58]. In [55], the authors provide asymptotic confidence intervals in the double limit where the sample size goes to infinity and the metamodel converges to the true model. These results were also adapted to problems related to more general models such as Shallow-Water equations, in the context of the control of an open channel [59].

When considering parameter-dependent PDE, it happens that the quantity of interest is not the PDE's solution but a linear functional of it. In [56], we have proposed a probabilistic error bound for the reduced output of interest (goal-oriented error bound). By probabilistic we mean that this bound may be violated with small probability. The bound is efficiently and explicitly computable, and we show on different examples that this error bound is sharper than existing ones.

A collaboration has been started with Christophe Prieur (Gipsa-Lab) on the very challenging issue of sensitivity of a controlled system to its control parameters [59]. In [60], we propose a generalization of the probabilistic goal-oriented error estimation in [56] to parameter-dependent nonlinear problems. One aims at applying such results in the previous context of sensitivity of a controlled system.

7.3. Dealing with uncertainties

7.3.1. Sensitivity Analysis

Participants: Eric Blayo, Laurent Gilquin, François-Xavier Le Dimet, Elise Arnaud, Maëlle Nodet, Clémentine Prieur, Laurence Viry.

7.3.1.1. Scientific context

Forecasting geophysical systems require complex models, which sometimes need to be coupled, and which make use of data assimilation. The objective of this project is, for a given output of such a system, to identify the most influential parameters, and to evaluate the effect of uncertainty in input parameters on model output. Existing stochastic tools are not well suited for high dimension problems (in particular time-dependent problems), while deterministic tools are fully applicable but only provide limited information. So the challenge is to gather expertise on one hand on numerical approximation and control of Partial Differential Equations, and on the other hand on stochastic methods for sensitivity analysis, in order to develop and design innovative stochastic solutions to study high dimension models and to propose new hybrid approaches combining the stochastic and deterministic methods.

7.3.1.2. Sensitivity analysis with dependent inputs

An important challenge for stochastic sensitivity analysis is to develop methodologies which work for dependent inputs. For the moment, there does not exist conclusive results in that direction. Our aim is to define an analogue of Hoeffding decomposition [54] in the case where input parameters are correlated. Clémentine Prieur supervised Gaëlle Chastaing's PhD thesis on the topic (defended in September 2013) [41]. We obtained first results [42], deriving a general functional ANOVA for dependent inputs, allowing defining new variance based sensitivity indices for correlated inputs. We then adapted various algorithms for the estimation of these new indices. These algorithms make the assumption that among the potential interactions, only few are significant. Two papers have been recently accepted [40], [43]. We also considered the estimation of groups Sobol' indices, with a procedure based on replicated designs [52]. These indices provide information at the level of groups, and not at a finer level, but their interpretation is still rigorous.

Céline Helbert and Clémentine Prieur supervised the PhD thesis of Simon Nanty (funded by CEA Cadarache, and defended in October, 2015). The subject of the thesis is the analysis of uncertainties for numerical codes with temporal and spatio-temporal input variables, with application to safety and impact calculation studies. This study implied functional dependent inputs. A first step was the modeling of these inputs [64]. The whole methodology proposed during the PhD is presented in [65].

More recently, the Shapley value, from econometrics, was proposed as an alternative to quantify the importance of random input variables to a function. Owen [67] derived Shapley value importance for independent inputs and showed that it is bracketed between two different Sobol' indices. Song et al. [72] recently advocated the use of Shapley value for the case of dependent inputs. In a very recent work [13], in collaboration with Art Owen (Standford's University), we show that Shapley value removes the conceptual problems of functional ANOVA for dependent inputs. We do this with some simple examples where Shapley value leads to intuitively reasonable nearly closed form values. We also investigated further the properties of Shapley effects in [31].

7.3.2. Non-Parametric Estimation for Kinetic Diffusions

Participants: Clémentine Prieur, Jose Raphael Leon Ramos.

This research is the subject of a collaboration with Chile and Uruguay. More precisely, we started working with Venezuela. Due to the crisis in Venezuela, our main collaborator on that topic moved to Uruguay.

We are focusing our attention on models derived from the linear Fokker-Planck equation. From a probabilistic viewpoint, these models have received particular attention in recent years, since they are a basic example for hypercoercivity. In fact, even though completely degenerated, these models are hypoelliptic and still verify some properties of coercivity, in a broad sense of the word. Such models often appear in the fields of mechanics, finance and even biology. For such models we believe it appropriate to build statistical non-parametric estimation tools. Initial results have been obtained for the estimation of invariant density, in

conditions guaranteeing its existence and unicity [37] and when only partial observational data are available. A paper on the non parametric estimation of the drift has been accepted recently [38] (see Samson et al., 2012, for results for parametric models). As far as the estimation of the diffusion term is concerned, a paper has been accepted [38], in collaboration with J.R. Leon (Montevideo, Uruguay) and P. Cattiaux (Toulouse). Recursive estimators have been also proposed by the same authors in [39], also recently accepted. In a recent collaboration with Adeline Samson from the statistics department in the Lab, we considered adaptive estimation, that is we proposed a data-driven procedure for the choice of the bandwidth parameters.

In [5], we focused on damping Hamiltonian systems under the so-called fluctuation-dissipation condition. Idea in that paper were re-used with applications to neuroscience in [63].

Note that Professor Jose R. Leon (Caracas, Venezuela, Montevideo, Uruguay) was funded by an international Inria Chair, allowing to collaborate further on parameter estimation.

We recently proposed a paper on the use of the Euler scheme for inference purposes, considering reflected diffusions. This paper could be extended to the hypoelliptic framework.

We started a collaboration with Karine Bertin (Valparaiso, Chile) funded by a MATHAMSUD project. We are interested in new adaptive estimators for invariant densities on bounded domains, and would like to extend that results to hypo-elliptic diffusions.

7.3.3. Multivariate Risk Indicators

Participants: Clémentine Prieur, Patricia Tencaliec.

Studying risks in a spatio-temporal context is a very broad field of research and one that lies at the heart of current concerns at a number of levels (hydrological risk, nuclear risk, financial risk etc.). Stochastic tools for risk analysis must be able to provide a means of determining both the intensity and probability of occurrence of damaging events such as e.g. extreme floods, earthquakes or avalanches. It is important to be able to develop effective methodologies to prevent natural hazards, including e.g. the construction of barrages.

Different risk measures have been proposed in the one-dimensional framework. The most classical ones are the return level (equivalent to the Value at Risk in finance), or the mean excess function (equivalent to the Conditional Tail Expectation CTE). However, most of the time there are multiple risk factors, whose dependence structure has to be taken into account when designing suitable risk estimators. Relatively recent regulation (such as Basel II for banks or Solvency II for insurance) has been a strong driver for the development of realistic spatio-temporal dependence models, as well as for the development of multivariate risk measurements that effectively account for these dependencies.

We refer to [44] for a review of recent extensions of the notion of return level to the multivariate framework. In the context of environmental risk, [71] proposed a generalization of the concept of return period in dimension greater than or equal to two. Michele et al. proposed in a recent study [45] to take into account the duration and not only the intensity of an event for designing what they call the dynamic return period. However, few studies address the issues of statistical inference in the multivariate context. In [46], [48], we proposed non parametric estimators of a multivariate extension of the CTE. As might be expected, the properties of these estimators deteriorate when considering extreme risk levels. In collaboration with Elena Di Bernardino (CNAM, Paris), Clémentine Prieur is working on the extrapolation of the above results to extreme risk levels [29].

Elena Di Bernardino, Véronique Maume-Deschamps (Univ. Lyon 1) and Clémentine Prieur also derived an estimator for bivariate tail [47]. The study of tail behavior is of great importance to assess risk.

With Anne-Catherine Favre (LTHE, Grenoble), Clémentine Prieur supervised the PhD thesis of Patricia Tencaliec. We are working on risk assessment, concerning flood data for the Durance drainage basin (France). The PhD thesis started in October 2013 and was defended in February 2017. A first paper on data reconstruction has been accepted [73]. It was a necessary step as the initial series contained many missing data. A second paper is in revision, considering the modeling of precipitation amount with semi-parametric models, modeling both the bulk of the distribution and the tails, but avoiding the arbitrary choice of a threshold. We work in collaboration with Philippe Naveau (LSCE, Paris).

7.3.4. Extensions of the replication method for the estimation of Sobol' indices

Participants: Elise Arnaud, Laurent Gilquin, Clémentine Prieur.

Sensitivity analysis studies how the uncertainty on an output of a mathematical model can be attributed to sources of uncertainty among the inputs. Global sensitivity analysis of complex and expensive mathematical models is a common practice to identify influent inputs and detect the potential interactions between them. Among the large number of available approaches, the variance-based method introduced by Sobol' allows to calculate sensitivity indices called Sobol' indices. Each index gives an estimation of the influence of an individual input or a group of inputs. These indices give an estimation of how the output uncertainty can be apportioned to the uncertainty in the inputs. One can distinguish first-order indices that estimate the main effect from each input or group of inputs from higher-order indices that estimate the corresponding order of interactions between inputs. This estimation procedure requires a significant number of model runs, number that has a polynomial growth rate with respect to the input space dimension. This cost can be prohibitive for time consuming models and only a few number of runs is not enough to retrieve accurate informations about the model inputs.

The use of replicated designs to estimate first-order Sobol' indices has the major advantage of reducing drastically the estimation cost as the number of runs n becomes independent of the input space dimension. The generalization to closed second-order Sobol' indices relies on the replication of randomized orthogonal arrays. However, if the input space is not properly explored, that is if n is too small, the Sobol' indices estimates may not be accurate enough. Gaining in efficiency and assessing the estimate precision still remains an issue, all the more important when one is dealing with limited computational budget.

We designed approaches to render the replication method recursive, enabling the required number of evaluations to be controlled. With these approaches, more accurate Sobol' estimates are obtained while recycling previous sets of model evaluations. The estimation procedure is therefore stopped when the convergence of estimates is considered reached. One of these approaches corresponds to a recursive version of the replication method and is based on the iterative construction of stratified designs, latin hypercubes and orthogonal arrays [50]. A second approach combines the use of quasi-Monte Carlo sampling and the construction of a new stopping criterion [9] [32].

In [30] a new strategy to estimate the full set of first-order and second-order Sobol' indices with only two replicated designs based on orthogonal arrays of strength two. Such a procedure increases the precision of the estimation for a given computation budget. A bootstrap procedure for producing confidence intervals, that are compared to asymptotic ones in the case of first-order indices, is also proposed.

7.3.5. Parameter control in presence of uncertainties: robust estimation of bottom friction Participants: Victor Trappler, Elise Arnaud, Laurent Debreu, Arthur Vidard.

Many physical phenomena are modelled numerically in order to better understand and/or to predict their behaviour. However, some complex and small scale phenomena can not be fully represented in the models. The introduction of ad-hoc correcting terms, can represent these unresolved processes, but they need to be properly estimated.

A good example of this type of problem is the estimation of bottom friction parameters of the ocean floor. This is important because it affects the general circulation. This is particularly the case in coastal areas, especially for its influence on wave breaking. Because of its strong spatial disparity, it is impossible to estimate the bottom friction by direct observation, so it requires to do so indirectly by observing its effects on surface movement. This task is further complicated by the presence of uncertainty in certain other characteristics linking the bottom and the surface (eg boundary conditions). The techniques currently used to adjust these settings are very basic and do not take into account these uncertainties, thereby increasing the error in this estimate.

Classical methods of parameter estimation usually imply the minimisation of an objective function, that measures the error between some observations and the results obtained by a numerical model. In the presence of uncertainties, the minimisation is not straightforward, as the output of the model depends on those uncontrolled inputs and on the control parameter as well. That is why we will aim at minimising the objective

function, to get an estimation of the control parameter that is robust to the uncertainties. In this work, a toy model of a coastal area has been modelled and implemented. The control parameter is the bottom friction, upon which classical methods of estimation are applied in a simulation-reestimation experiment. The model is then modified to include uncertainties on the boundary conditions in order to apply robust control methods. First, a sensitivity analysis of the objective function has been performed to assess the influence of each considered variable. Then, a study on the meaning of different concepts of robustness have been carried on. Typically, one then seeks an optimal parameter set that would minimise the variance or the mean of the original objective function. Various associated algorithms from the literature have been implemented. They all rely on surrogate models and black-box optimisation techniques to solve this estimation problem.

7.3.6. Sensitivity of a floating offshore wind turbine to uncertain parameters

Participants: Adrien Hirvoas, Elise Arnaud, Clémentine Prieur, Arthur Vidard.

In a fast-changing energy context, marine renewable energies in general and floating offshore wind energy in particular are a promising source of energy in France and aboard. The design of these structures is made in a specific regulated framework related to their environment. Floating offshore wind turbines are submitted to various continuous environmental loadings (wind, current, swell), which generate solicitations and fatigue in some components. Fatigue lifetime is estimated with a dedicated software that allows performing coupled multi-physics simulations of the system (hydrodynamics, aerodynamics, mechanics and controls). The inputs of these simulations necessarily include uncertainties regarding the environmental loadings and the physical parameters of the models as well. These uncertainties can have an influence on the simulated behaviour of the system. The core of this work consists in conducting a sensitivity analysis to assess, how the uncertainty on an output of a model can be attributed to sources of uncertainty among the inputs. The approach that is considered, is based on the calculation of Sobol indices with the FAST method, and a meta-model using Kriging. These indices are used to evaluate in what extend an input or group of inputs is responsible for the output variance. The perspectives of this study is to understand what king of measurements could be of interest to properly estimate the sensible parameters, and where these measurements should be monitored on the structure. Such an estimation will be performed with data assimilation approaches, which optimally combine numerical models and physical observations. This work in done in collaboration with IFPEN.

7.3.7. Uncertainty and robustess analysis for models with functional input/output.

Participants: Mohammed Reda El Amri, Clémentine Prieur.

Numerical models are commonly used to study physical phenomena. They imply many inputs parameters, and potentially provide a large number of quantities of interest as outputs. Practitioners are not only interested in the response of their model for a given set of inputs (forward problem) but also in recovering the set of inputs values leading to a prescribed value or range for the quantity of interest (inversion problem). In collaboration with IFP Energies nouvelles, we develop data-driven strategies for robust inversion under functional uncertainties. Reda El Amri's PhD thesis aim at developping such tools with application to pollutant emission control.

7.4. Assimilation of Images

Participants: Elise Arnaud, François-Xavier Le Dimet, Maëlle Nodet, Arthur Vidard, Nelson Feyeux.

7.4.1. Direct assimilation of image sequences

At the present time the observation of Earth from space is done by more than thirty satellites. These platforms provide two kinds of observational information:

- Eulerian information as radiance measurements: the radiative properties of the earth and its fluid envelops. These data can be plugged into numerical models by solving some inverse problems.
- Lagrangian information: the movement of fronts and vortices give information on the dynamics of the fluid. Presently this information is scarcely used in meteorology by following small cumulus clouds and using them as Lagrangian tracers, but the selection of these clouds must be done by hand and the altitude of the selected clouds must be known. This is done by using the temperature of the top of the cloud.

Our current developments are targeted at the use of « Level Sets » methods to describe the evolution of the images. The advantage of this approach is that it permits, thanks to the level sets function, to consider the images as a state variable of the problem. We have derived an Optimality System including the level sets of the images. This approach is being applied to the tracking of oceanic oil spills [10]

A collaborative project started with C. Lauvernet (IRSTEA) in order to make use of our image assimilation strategies on the control of pesticide transfer.

7.4.2. Optimal transport for image assimilation

We investigate the use of optimal transport based distances for data assimilation, and in particular for assimilating dense data such as images. The PhD thesis of N. Feyeux studied the impact of using the Wasserstein distance in place of the classical Euclidean distance (pixel to pixel comparison). In a simplified one dimensional framework, we showed that the Wasserstein distance is indeed promising. Data assimilation experiments with the Shallow Water model have been performed and confirm the interest of the Wasserstein distance. Results have been presented at conferences and seminars and a paper is under minor revision at NPG [49].

7.5. Land Use and Transport Models Calibration

Participants: Thomas Capelle, Laurent Gilquin, Clémentine Prieur, Arthur Vidard, Peter Sturm, Elise Arnaud.

Given the complexity of modern urban areas, designing sustainable policies calls for more than sheer expert knowledge. This is especially true of transport or land use policies, because of the strong interplay between the land use and the transportation systems. Land use and transport integrated (LUTI) modelling offers invaluable analysis tools for planners working on transportation and urban projects. Yet, very few local authorities in charge of planning make use of these strategic models. The explanation lies first in the difficulty to calibrate these models, second in the lack of confidence in their results, which itself stems from the absence of any well-defined validation procedure. Our expertise in such matters will probably be valuable for improving the reliability of these models. To that purpose we participated to the building up of the ANR project CITIES led by the STEEP EPI. This project started early 2013 and two PhD about sensitivity analysis and calibration were launched late 2013. Laurent Gilquin defended his PhD in October 2016 [51] and Thomas Capelle defended his in April 2017 and published his latest results in [4].

ARIC Project-Team

7. New Results

7.1. Efficient approximation methods

7.1.1. Automatic generation of hardware FIR filters from a frequency domain specification

In [53], we present an open-source tool for the automatic design of reliable finite impulse response (FIR) filters, targeting FPGAs. It shows that user intervention can be limited to a very small number of relevant input parameters: a high-level frequency-domain specification, and input/output formats. All the other design parameters are computed automatically, using novel approaches to filter coefficient quantization and direct-form architecture implementation. Our tool guarantees a priori that the resulting architecture respects the specification while attempting to minimize its cost. Our approach is evaluated on a range of examples and shown to produce designs that are very competitive with the state of the art, with very little design effort.

7.1.2. Exponential sums and correctly-rounded functions

The 2008 revision of the IEEE-754 standard, which governs floating-point arithmetic, recommends that a certain set of elementary functions should be correctly rounded. Successful attempts for solving the Table Maker's Dilemma in binary64 made it possible to design CRlibm, a library which offers correctly rounded evaluation in binary64 of some functions of the usual libm. It evaluates functions using a two step strategy, which relies on a folklore heuristic that is well spread in the community of mathematical functions designers. Under this heuristic, one can compute the distribution of the lengths of runs of zeros/ones after the rounding bit of the value of the function at a given floating-point number. The goal of [13] was to change, whenever possible, this heuristic into a rigorous statement. The underlying mathematical problem amounts to counting integer points in the neighborhood of a curve, which we tackle using so-called exponential sums techniques, a tool from analytic number theory.

7.1.3. Continued fractions in power series fields

In [5], we explicitly describe a noteworthy transcendental continued fraction in the field of power series over \mathbb{Q} , having irrationality measure equal to 3. This continued fraction is a generating function of a particular sequence in the set $\{1, 2\}$. The origin of this sequence, whose study was initiated in a recent paper, is to be found in another continued fraction, in the field of power series over \mathbb{F}_3 , which satisfies a simple algebraic equation of degree 4, introduced thirty years ago by D. Robbins.

7.1.4. Validated and numerically efficient Chebyshev spectral methods for linear ordinary differential equations

In [51], we develop a validated numerics method for the solution of linear ordinary differential equations (LODEs). A wide range of algorithms (i.e., Runge-Kutta, collocation, spectral methods) exist for numerically computing approximations of the solutions. Most of these come with proofs of asymptotic convergence, but usually, provided error bounds are non-constructive. However, in some domains like critical systems and computer-aided mathematical proofs, one needs validated effective error bounds. We focus on both the theoretical and practical complexity analysis of a so-called *a posteriori* quasi-Newton validation method, which mainly relies on a fixed-point argument of a contracting map. Specifically, given a polynomial approximation, obtained by some numerical algorithm and expressed in Chebyshev basis, our algorithm efficiently computes an accurate and rigorous error bound. For this, we study theoretical properties like compactness, convergence, invertibility of associated linear integral operators and their truncations in a suitable coefficient space of Chebyshev series. Then, we analyze the almost-banded matrix structure of these operators, which allows for very efficient numerical algorithms for both numerical solutions of LODEs and rigorous computation of the approximation error. Finally, several representative examples show the advantages of our algorithms as well as their theoretical and practical limits.

7.1.5. Validated semi-analytical transition matrices for linearized relative spacecraft dynamics via Chebyshev series appproximations

In [47], we provide an efficient generic algorithm to compute validated approximations of transition matrices of linear time-variant systems using Chebyshev expansions, and apply it to two different examples of relative motion of satellites (spacecraft rendezvous with Tschauner-Hempel equations and geostationary station keeping with J2 perturbation in the linearized Orange model).

7.2. Lattices: algorithms and cryptology

7.2.1. All-But-Many Lossy Trapdoor Functions and Selective Opening Chosen-Ciphertext Security from LWE

In cryptography, selective opening (SO) security refers to adversaries that receive a number of ciphertexts and, after having corrupted a subset of the senders (thus obtaining the plaintexts and the senders' random coins), aim at breaking the security of remaining ciphertexts. So far, very few public-key encryption schemes are known to provide simulation-based selective opening (SIM-SO-CCA2) security under chosen-ciphertext attacks and most of them encrypt messages bit-wise. The only exceptions to date rely on all-but-many lossy trapdoor functions (as introduced by Hofheinz; Eurocrypt'12) and the Composite Residuosity assumption. In a paper [43] published at Crypto 2017, the team describes the first all-but-many lossy trapdoor function with security relying on the presumed hardness of the Learning-With-Errors problem (LWE) with standard parameters. The new construction exploits homomorphic computations on lattice trapdoors for lossy LWE matrices. By carefully embedding a lattice trapdoor in lossy public keys, the paper is able to prove SIM-SO-CCA2 security under the LWE assumption. As a result of independent interest, the paper describes a variant of our scheme whose multi-challenge CCA2 security tightly relates to the hardness of LWE and the security of a pseudo-random function.

7.2.2. Zero-Knowledge Arguments for Lattice-Based PRFs and Applications to E-Cash

This paper [41] deals with cryptographic pseudorandom functions from lattice assumptions and their use in e-cash systems. Beyond their security guarantees under well-studied assumptions, algebraic pseudo-random functions are motivated by their compatibility with efficient zero-knowledge proof systems, which is useful in a number of privacy applications like digital cash. The paper considers the problem of proving the correct evaluation of lattice-based PRFs based on the Learning-With-Rounding (LWR) problem introduced by Banerjee et al. (Eurocrypt'12). Namely, the paper provides zero-knowledge arguments of knowledge of triples (y, k, x) such that $y = F_k(x)$ is the correct evaluation of a PRF for a secret input x and a committed key k. While analogous statements admit efficient zero-knowledge protocols in the discrete logarithm setting, they have never been addressed in lattices so far. The paper provides such arguments for the key homomorphic PRF of Boneh et al. (Crypto'13) and the generic PRF implied by the LWR-based pseudo-random generator. As an application, the paper describes the first compact e-cash system based on lattice assumptions.

7.2.3. Adaptive Oblivious Transfer with Access Control from Lattice Assumptions

Adaptive oblivious transfer (OT) is a cryptographic protocol where a sender initially commits to a database $\{M_i\}_{i=1}^N$. Then, a receiver can query the sender up to k times with private indexes $\rho_1, ..., \rho_k$ so as to obtain $M_{\rho_1}, ..., M_{\rho_k}$ and nothing else. Moreover, for each $i \in [k]$, the receiver's choice ρ_i may depend on previously obtained messages. Oblivious transfer with access control (OT-AC) is a flavor of adaptive OT where database records are protected by distinct access control policies that specify which credentials a receiver should obtain in order to access each M_i . So far, all known OT-AC protocols only support access policies made of conjunctions or rely on *ad hoc* assumptions in pairing-friendly groups (or both). The paper [40] provides an OT-AC protocol where access policies may consist of any branching program of polynomial length, which is sufficient to realize any access policy in NC1. The security of the protocol is proved under the Learning-with-Errors (LWE) and Short-Integer-Solution (SIS) assumptions. As a result of independent interest, the paper provides protocols for proving the correct evaluation of a committed branching program on a committed input.

7.2.4. Encoding-Free ElGamal-Type Encryption Schemes on Elliptic Curves

At PKC 2006, Chevallier-Mames, Paillier, and Pointcheval proposed a very elegant technique over cyclic subgroups of \mathbb{F}_p eliminating the need to encode the message as a group element in the ElGamal encryption scheme. Unfortunately, it is unclear how to adapt their scheme over elliptic curves. In a previous attempt, Virat suggested an adaptation of ElGamal to elliptic curves over the ring of dual numbers as a way to address the message encoding issue. Advantageously the resulting cryptosystem does not require encoding messages as points on an elliptic curve prior to their encryption. Unfortunately, it only provides one-wayness and, in particular, it is not (and was not claimed to be) semantically secure. The paper revisits Virat's cryptosystem and extends the Chevallier-Mames et al.'s technique to the elliptic curve setting. The paper [35] considers elliptic curves over the ring $\mathbb{Z}/(p^2\mathbb{Z})$ and defines the underlying class function. This yields complexity assumptions whereupon new ElGamal-type encryption schemes are built. The so-obtained schemes are proved semantically secure and make use of a very simple message encoding: messages being encrypted are viewed as elements in the range [0, p-1]. Further, the new schemes come equipped with a partial ring-homomorphism property: anyone can add a constant to an encrypted message -or- multiply an encrypted message by a constant. This can prove helpful as a blinding method in a number of applications. Finally, in addition to practicability, the proposed schemes also offer better performance in terms of speed, memory, and bandwidth.

7.2.5. Structure-Preserving Chosen-Ciphertext Security with Shorter Verifiable Ciphertexts

Structure-preserving cryptography is a world where messages, signatures, ciphertexts and public keys are entirely made of elements of a group over which a bilinear map is efficiently computable. This property makes the primitives compatible with the Groth-Sahai non-interactive proof systems in the design of higherlevel privacy-preserving protocols. While structure-preserving signatures have received much attention the last 6 years, structure-preserving encryption schemes have undergone slower development. In particular, the best known structure-preserving cryptosystems with chosen-ciphertext (IND-CCA2) security either rely on symmetric pairings or require long ciphertexts comprised of hundreds of group elements or do not provide publicly verifiable ciphertexts. The paper [42] provides a publicly verifiable construction based on the SXDH assumption in asymmetric bilinear groups $e: G_1 \times G_2 \to G_T$, which features relatively short ciphertexts. For typical parameters, the ciphertext size amounts to less than 40 elements of G. As a second contribution, the paper provides a structure-preserving encryption scheme with perfectly randomizable ciphertexts and replayable chosen-ciphertext security. The new RCCA-secure system significantly improves upon the best known system featuring similar properties in terms of ciphertext size.

7.2.6. Tightly Secure IBE under Constant-size Master Public Key

This paper is about identity-based encryption (IBE). Chen and Wee (Crypto 2013) proposed the first almost tightly and adaptively secure IBE in the standard model and left two open problems which called for a tightly secure IBE with (1) constant-size master public key and/or (2) constant security loss. This paper proposes an IBE scheme with constant-size master public key and tighter security reduction. This (partially) solves Chen and Wee's first open problem and makes progress on the second one. Technically, the new IBE scheme is built based on Wee's petit IBE scheme (TCC 2016) in composite-order bilinear groups whose order is product of four primes. The sizes of master public key, ciphertexts, and secret keys are not only constant but also nearly optimal as Wee's petit IBE. The paper [33] proves its adaptive security in the multi-instance, multi-ciphertext setting (PKC 2015) based on the decisional subgroup assumption and a subgroup variant of DBDH assumption. The security loss is $O(\log q)$ where q is the upper bound of the total number of secret keys and challenge ciphertexts revealed to adversary in each single IBE instance. It is much smaller than those for all known adaptively secure IBE schemes in a concrete sense.

7.2.7. ABE with Tag Made Easy: Concise Framework and New Instantiations in Prime-order Groups

Among all existing identity-based encryption (IBE) schemes in bilinear groups, Wat-IBE proposed by Waters (CRYPTO 2009) and JR-IBE proposed by Jutla and Roy (Asiacrypt 2013) are quite special. A secret key and/or ciphertext in these two schemes consists of several group elements and an integer which is usually called tag.

A series of prior work was devoted to extending them towards more advanced attribute-based encryption (ABE) including inner-product encryption (IPE), hierarchical IBE (HIBE). Recently, Kim et al. (SCN 2016) introduced the notion of tag-based encoding and presented a generic framework for extending Wat-IBE. We may call these ABE schemes ABE with tag or tag-based ABE. Typically, a tag-based ABE construction is more efficient than its counterpart without tag. However, the research on tag-based ABE severely lags: we do not know how to extend JR-IBE in a systematic way and there is no tag-based ABE for Boolean span program even with Kim et al.'s generic framework.

This paper [32] proposes a generic framework for tag-based ABE which is based on JR-IBE and compatible with Chen et al.'s (attribute-hiding) predicate encoding (Eurocrypt 2015). The adaptive security in the standard model relies on the *k*-linear assumption in asymmetric prime-order bilinear groups. This is the first framework showing how to extend JR-IBE systematically. In fact, the framework and its simple extension are able to cover most concrete tag-based ABE constructions in previous literature. Furthermore, since Chen et al.'s predicate encoding supports a large number of predicates including boolean span program, the paper can give the first (both key-policy and ciphertext-policy) tag-based ABE for boolean span program in the standard model. Technically, the new framework is based on a simplified version of JR-IBE. Both the description and its proof are quite similar to the prime-order IBE derived from Chen et al.'s framework. This not only allows us to work with Chen et al.'s predicate encoding but also provides a clear explanation of the JR-IBE scheme and its proof technique.

7.2.8. Hardness of k-LWE and Applications in Traitor Tracing

The paper introduces the k-LWE problem, a Learning With Errors variant of the k-SIS problem. The Boneh-Freeman reduction from SIS to k-SIS suffers from an exponential loss in k. The paper [24] improves and extend it to an LWE to k-LWE reduction with a polynomial loss in k, by relying on a new technique involving trapdoors for random integer kernel lattices. Based on this hardness result, the paper presents the first algebraic construction of a traitor tracing scheme whose security relies on the worstcase hardness of standard lattice problems. The proposed LWE traitor tracing is almost as efficient as the LWE encryption. Further, it achieves public traceability, i.e., allows the authority to delegate the tracing capability to trusted parties. To this aim, the paper introduces the notion of projective sampling family in which each sampling function is keyed and, with a projection of the key on a well chosen space, one can simulate the sampling function in a computationally indistinguishable way. The construction of a projective sampling family from k-LWE allows us to achieve public traceability, by publishing the projected keys of the users.

7.2.9. Middle-Product Learning With Errors

The paper [45] introduces a new variant MPLWE of the Learning With Errors problem (LWE) making use of the Middle Product between polynomials modulo an integer q. It exhibits a reduction from the Polynomial-LWE problem (PLWE) parametrized by a polynomial f, to MPLWE which is defined independently of any such f. The reduction only requires f to be monic with constant coefficient coprime with q. It incurs a noise growth proportional to the so-called expansion factor of f. The paper also describes a public-key encryption scheme with quasi-optimal asymptotic efficiency (the bit-sizes of the keys and the run-times of all involved algorithms are quasi-linear in the security parameter), which is secure against chosen plaintext attacks under the MPLWE hardness assumption. The scheme is hence secure under the assumption that PLWE is hard for at least one polynomial f of degree n among a family of f's which is exponential in n.

7.2.10. Efficient Public Trace and Revoke from Standard Assumptions

The paper [27] provides efficient constructions for trace-and-revoke systems with public traceability in the black-box confirmation model. The constructions achieve adaptive security, are based on standard assumptions and achieve significant efficiency gains compared to previous constructions. The constructions rely on a generic transformation from inner product functional encryption (IPFE) schemes to trace-and-revoke systems. The proposed transformation requires the underlying IPFE scheme to only satisfy a very weak notion of security the attacker may only request a bounded number of random keys in contrast to the standard notion of security where she may request an unbounded number of arbitrarily chosen keys. The paper exploits the

much weaker security model to provide a new construction for bounded collusion and random key IPFE from the learning with errors assumption (LWE), which enjoys improved efficiency compared to the scheme of Agrawal et al. [CRYPTO'16]. Together with IPFE schemes from Agrawal et al., the paper obtains trace and revoke from LWE, Decision Diffie Hellman and Decision Quadratic Residuosity.

7.2.11. New Techniques for Structural Batch Verification in Bilinear Groups with Applications to Groth-Sahai Proofs

Bilinear groups form the algebraic setting for a multitude of important cryptographic protocols including anonymous credentials, e-cash, e-voting, e-coupon, and loyalty systems. It is typical of such crypto protocols that participating parties need to repeatedly verify that certain equations over bilinear groups are satisfied, e.g., to check that computed signatures are valid, commitments can be opened, or non-interactive zero-knowledge proofs verify correctly. Depending on the form and number of equations this part can quickly become a performance bottleneck due to the costly evaluation of the bilinear map.

To ease this burden on the verifier, batch verification techniques have been proposed that allow to combine and check multiple equations probabilistically using less operations than checking each equation individually. The paper [34] revisits the batch verification problem and existing standard techniques. It introduces a new technique which, in contrast to previous work, allows to fully exploit the structure of certain systems of equations. Equations of the appropriate form naturally appear in many protocols, e.g., due to the use of Groth–Sahai proofs.

The beauty of the new technique is that the underlying idea is pretty simple: the paper observes that many systems of equations can alternatively be viewed as a single equation of products of polynomials for which probabilistic polynomial identity testing following Schwartz–Zippel can be applied. Comparisons show that the new approach can lead to significant improvements in terms of the number of pairing evaluations. Indeed, for the BeleniosRF voting system presented at CCS 2016, it is possible to reduce the number of pairings (required for ballot verification) from 4k + 140, as originally reported by Chaidos et al., to k + 7. As the implementation and benchmarks demonstrate, this may reduce the verification runtime to only 5% to 13% of the original runtime.

7.2.12. Encryption Switching Protocols Revisited: Switching Modulo p

At CRYPTO 2016, Couteau, Peters and Pointcheval introduced a new primitive called Encryption Switching Protocols (ESP), allowing to switch ciphertexts between two encryption schemes. If such an ESP is built with two schemes that are respectively additively and multiplicatively homomorphic, it naturally gives rise to a secure 2-party computation protocol. It is thus perfectly suited for evaluating functions, such as multivariate polynomials, given as arithmetic circuits. Couteau et al. built an ESP to switch between Elgamal and Paillier encryptions which do not naturally fit well together. Consequently, they had to design a clever variant of Elgamal over $\mathbf{Z}/n\mathbf{Z}$ with a costly shared decryption.

In [31], we first present a conceptually simple generic construction for encryption switching protocols. We then give an efficient instantiation of our generic approach that uses two well-suited protocols, namely a variant of Elgamal in $\mathbf{Z}/p\mathbf{Z}$ and the Castagnos-Laguillaumie encryption which is additively homomorphic over $\mathbf{Z}/p\mathbf{Z}$. Among other advantages, this allows to perform all computations modulo a prime p instead of an RSA modulus. Overall, our solution leads to significant reductions in the number of rounds as well as the number of bits exchanged by the parties during the interactive protocols. We also show how to extend its security to the malicious setting.

7.3. Algebraic computing and high-performance kernels

7.3.1. Multiple binomial sums

Multiple binomial sums form a large class of multi-indexed sequences, closed under partial summation, which contains most of the sequences obtained by multiple summation of binomial coefficients and also all the sequences with algebraic generating function. We study the representation of the generating functions of

binomial sums by integrals of rational functions. The outcome is twofold. Firstly, we show that a univariate sequence is a multiple binomial sum if and only if its generating function is the diagonal of a rational function. Secondly we propose algorithms that decide the equality of multiple binomial sums and that compute recurrence relations for them. In conjunction with geometric simplifications of the integral representations, this approach behaves well in practice. The process avoids the computation of certificates and the problem of accurate summation that afflicts discrete creative telescoping, both in theory and in practice [12].

7.3.2. Algebraic Diagonals and Walks: Algorithms, Bounds, Complexity

The diagonal of a multivariate power series F is the univariate power series Diag(F) generated by the diagonal terms of F. Diagonals form an important class of power series; they occur frequently in number theory, theoretical physics and enumerative combinatorics. We study algorithmic questions related to diagonals in the case where F is the Taylor expansion of a bivariate rational function. It is classical that in this case Diag(F) is an algebraic function. We propose an algorithm that computes an annihilating polynomial for Diag(F). We give a precise bound on the size of this polynomial and show that generically, this polynomial is the minimal polynomial and that its size reaches the bound. The algorithm runs in time quasi-linear in this bound, which grows exponentially with the degree of the input rational function. We then address the related problem of enumerating directed lattice walks. The insight given by our study leads to a new method for expanding the generating power series of bridges, excursions and meanders. We show that their first N terms can be computed in quasi-linear complexity in N, without first computing a very large polynomial equation [10].

7.3.3. Computing minimal interpolation bases

In [20] we consider the problem of computing univariate polynomial matrices over a field that represent minimal solution bases for a general interpolation problem, some forms of which are the vector M-Padé approximation problem in [Van Barel and Bultheel, Numerical Algorithms 3, 1992] and the rational interpolation problem in [Beckermann and Labahn, SIAM J. Matrix Anal. Appl. 22, 2000]. Particular instances of this problem include the bivariate interpolation steps of Guruswami-Sudan hard-decision and Kötter-Vardy soft-decision decodings of Reed-Solomon codes, the multivariate interpolation step of list-decoding of folded Reed-Solomon codes, and Hermite-Padé approximation. In the mentioned references, the problem is solved using iterative algorithms based on recurrence relations. Here, we discuss a fast, divide-and-conquer version of this recurrence, taking advantage of fast matrix computations over the scalars and over the polynomials. This new algorithm is deterministic, and for computing shifted minimal bases of relations between m vectors of size σ it uses $\tilde{O}(m^{\omega-1}(\sigma + |s|))$ field operations, where ω is the exponent of matrix multiplication, |s| is the sum of the entries of the input shift s with min(s) = 0, and the soft-O notation indicates that logarithmic factors in the big-O are omitted. This complexity bound improves in particular on earlier algorithms in the case of bivariate interpolation for soft decoding, while matching fastest existing algorithms for simultaneous Hermite-Padé approximation.

7.3.4. Fast and deterministic computation of the Hermite normal form and determinant of a polynomial matrix

Given a nonsingular $n \times n$ matrix of univariate polynomials over a field, we present in [22] fast and deterministic algorithms to compute its determinant and its Hermite normal form. The proposed algorithms use $\tilde{O}(n^{\omega} \lceil s \rceil)$ field operations, where s is bounded from above by both the average of the degrees of the rows and that of the columns of the matrix, and ω is the exponent of matrix multiplication. The ceiling function indicates that the cost is $\tilde{O}(n^{\omega})$ when s = o(1). Our algorithms are based on a fast and deterministic triangularization method for computing the diagonal entries of the Hermite form of a nonsingular matrix.

7.3.5. Computing canonical bases of modules of univariate relations

We study in [44] the computation of canonical bases of sets of univariate relations $(p_1, ..., p_m) \in K[x]^m$ such that $p_1f_1 + \cdots + p_mf_m = 0$; here, the input elements $f_1, ..., f_m$ are from a quotient $K[x]^n/\mathcal{M}$, where \mathcal{M} is a K[x]-module of rank n given by a basis $M \in K[x]^{n \times n}$ in Hermite form. We exploit the triangular shape of M to generalize a divide-and-conquer approach which originates from fast minimal approximant basis algorithms. Besides recent techniques for this approach, we rely on high-order lifting to perform fast modular

products of polynomial matrices of the form $PF \mod M$. Our algorithm uses $\widetilde{O}(m^{\omega-1}D + n^{\omega}D/m)$ operations in K, where $D = \deg(\det(M))$ is the K-vector space dimension of $K[x]^n/\mathcal{M}$, $\widetilde{O}(\cdot)$ indicates that logarithmic factors are omitted, and ω is the exponent of matrix multiplication. This had previously only been achieved for a diagonal matrix M. Furthermore, our algorithm can be used to compute the shifted Popov form of a nonsingular matrix within the same cost bound, up to logarithmic factors, as the previously fastest known algorithm, which is randomized.

7.3.6. Matrices with displacement structure: generalized operators and faster algorithms

For matrices with displacement structure, basic operations like multiplication, inversion, and linear system solving can be expressed in terms of the following task: evaluate the product AB, where A is a structured $n \times n$ matrix of displacement rank α , and B is an arbitrary $n \times \alpha$ matrix. In [11], we first generalize classical displacement operators, based on block diagonal matrices with companion diagonal blocks, and then design fast algorithms to perform the task above for this extended class of structured matrices. The arithmetic cost of these algorithms ranges from $O(\alpha^{\omega-1}M(n))$ to $O(\alpha^{\omega-1}M(n)\log(n))$, with ω such that two $n \times n$ matrices over a field can be multiplied using $O(n^{\omega})$ field operations, and where M is a cost function for polynomial multiplication. By combining this result with classical randomized regularization techniques, we obtain faster Las Vegas algorithms for structured inversion and linear system solving.

7.3.7. Absolute real root separation

While the separation (the minimal nonzero distance) between roots of a polynomial is a classical topic, its absolute counterpart (the minimal nonzero distance between their absolute values) does not seem to have been studied much. We present the general context and give tight bounds for the case of real roots [14].

7.3.8. Weighted Lattice Walks and Universality Classes

In this work we consider two different aspects of weighted walks in cones. To begin we examine a particular weighted model, known as the Gouyou-Beauchamps model. Using the theory of analytic combinatorics in several variables we obtain the asymptotic expansion of the total number of Gouyou-Beauchamps walks confined to the quarter plane. Our formulas are parametrized by weights and starting point, and we identify six different asymptotic regimes (called universality classes) which arise according to the values of the weights. The weights allowed in this model satisfy natural algebraic identities permitting an expression of the weighted generating function in terms of the generating function of unweighted walks on the same steps. The second part of this article explains these identities combinatorially for walks in arbitrary cones and dimensions, and provides a characterization of universality classes for general weighted walks. Furthermore, we describe an infinite set of models with non-D-finite generating function [15].

7.3.9. Introduction to the IEEE 1788-2015 Standard for Interval Arithmetic

Interval arithmetic is a tool of choice for numerical software verification, as every result computed using this arithmetic is self-verified: every result is an interval that is guaranteed to contain the exact numerical values, regardless of uncertainty or roundoff errors. From 2008 to 2015, interval arithmetic underwent a standardization effort, resulting in the IEEE 1788-2015 standard. The main features of this standard are developed in [26]: the structure into levels, from the mathematic model to the implementation on computers; the possibility to accomodate different mathematical models, called flavors; the decoration system that keeps track of relevant events during the course of a calculation; the exact dot product for point (as opposed to interval) vectors.

7.3.10. Influence of the Condition Number on Interval Computations: Some Examples

The condition number is a quantity that is well-known in "classical" numerical analysis, that is, where numerical computations are performed using floating-point numbers. This quantity appears much less frequently in interval numerical analysis, that is, where the computations are performed on intervals. In [56], two aspects are developed. On the one hand, it is stressed that the notion of condition number already appears in the literature on interval analysis, even if it does not bear that name. On the other hand, three small examples are used to illustrate experimentally the impact of the condition number on interval computations. As expected, problems with a larger condition number are more difficult to solve: this means either that the solution is not very accurate (for moderate condition numbers) or that the method fails to solve the problem, even inaccurately (for larger condition numbers). Different strategies to counteract the impact of the condition number are discussed and experimented: use of a higher precision, iterative refinement, bisection of the input.

7.3.11. Error bounds on complex floating-point multiplication with an FMA

The accuracy analysis of complex floating-point multiplication done by Brent, Percival, and Zimmermann is extended to the case where a fused multiply-add (FMA) operation is available. Considering floating-point arithmetic with rounding to nearest and unit roundoff u, we show that their bound $\sqrt{5} u$ on the normwise relative error $|\hat{z}/z - 1|$ of a complex product z can be decreased further to 2u when using the FMA in the most naive way. Furthermore, we prove that the term 2u is asymptotically optimal not only for this naive FMA-based algorithm, but also for two other algorithms, which use the FMA operation as an efficient way of implementing rounding error compensation. Thus, although highly accurate in the componentwise sense, these two compensated algorithms bring no improvement to the normwise accuracy 2u already achieved using the FMA naively. Asymptotic optimality is established for each algorithm thanks to the explicit construction of floating-point inputs for which we prove that the normwise relative error then generated satisfies $|\hat{z}/z - 1| \rightarrow 2u$ as $u \rightarrow 0$. All our results hold for IEEE floating-point arithmetic, with radix β , precision p, and rounding to nearest; it is only assumed that underflows and overflows do not occur and that $\beta^{p-1} \ge 24$ [19].

7.3.12. Automatic source-to-source error compensation of floating-point programs

Numerical programs with IEEE 754 floating-point computations may suffer from inaccuracies, since finite precision arithmetic is an approximation of real arithmetic. Solutions that reduce the loss of accuracy are available, such as, compensated algorithms or double-double precision floating-point arithmetic. Our goal is to automatically improve the numerical quality of a numerical program with the smallest impact on its performance. In [25] we define and implement source code transformations in order to derive automatically compensated programs. We present several experimental results to compare the transformed programs and existing solutions. The transformed programs are as accurate and efficient as the implementations of compensated algorithms when the latter exist. Furthermore, we propose some transformation strategies allowing us to improve partially the accuracy of programs and to tune the impact on execution time. Trade-offs between accuracy and performance are assured by code synthesis. Experimental results show that, with the help of the tools presented here, user-defined trade-offs are achievable in a reasonable amount of time.

7.3.13. Formal correctness of comparison algorithms between binary64 and decimal64 floating-point numbers

We present a full Coq formalisation of the correctness of some comparison algorithms between binary64 and decimal64 floating-point numbers [28].

7.3.14. Implementation and performance evaluation of an extended precision floating-point arithmetic library for high-accuracy semidefinite programming

Semidefinite programming (SDP) is widely used in optimization problems with many applications, however, certain SDP instances are ill-posed and need more precision than the standard double-precision available. Moreover, these problems are large-scale and could benefit from parallelization on specialized architectures such as GPUs. In this article, we implement and evaluate the performance of a floating-point expansion-based arithmetic library (newFPLib) in the context of such numerically highly accurate SDP solvers. We plugged-in the newFPLib with the state-of-the-art SDPA solver for both CPU and GPU-tuned implementations. We compare and contrast both the numerical accuracy and performance of SDPA-GMP,-QD and-DD, which employ other multiple-precision arithmetic libraries against SDPA-newFPLib. We show that our newFPLib is a very good trade-off for accuracy and speed when solving ill-conditioned SDP problems [38].

7.3.15. The classical relative error bounds for computing $\sqrt{a^2 + b^2}$ and $c/\sqrt{a^2 + b^2}$ in binary floating-point arithmetic are asymptotically optimal

We study the accuracy of classical algorithms for evaluating expressions of the form $\sqrt{a^2 + b^2}$ and $c/\sqrt{a^2 + b^2}$ in radix-2, precision-*p* floating-point arithmetic, assuming that the elementary arithmetic operations \pm , ×, /, $\sqrt{a^2 + b^2}$ are rounded to nearest, and assuming an unbounded exponent range. Classical analyses show that the relative error is bounded by $2u + O(u^2)$ for $\sqrt{a^2 + b^2}$, and by $3u + O(u^2)$ for $c/\sqrt{a^2 + b^2}$, where $u = 2^{-p}$ is the unit roundoff. Recently, it was observed that for $\sqrt{a^2 + b^2}$ the $O(u^2)$ term is in fact not needed. We show here that it is not needed either for $c/\sqrt{a^2 + b^2}$. Furthermore, we show that these error bounds are asymptotically optimal. Finally, we show that the possible availability of an FMA instruction does not change the bounds, nor their asymptotic optimality [37].

7.3.16. On the relative error of computing complex square roots in floating-point arithmetic

We study the accuracy of a classical approach to computing complex square-roots in floating-point arithmetic. Our analyses are done in binary floating-point arithmetic in precision p, and we assume that the (real) arithmetic operations $+, -, \times, \div, \sqrt{}$ are rounded to nearest, so the unit roundoff is $u = 2^{-p}$. We show that in the absence of underflow and overflow, the componentwise and normwise relative errors of this approach are at most $\frac{7}{2}u$ and $\frac{\sqrt{37}}{2}u$, respectively, and this without having to neglect terms of higher order in u. We then provide some input examples showing that these bounds are reasonably sharp for the three basic binary interchange formats (binary32, binary64, and binary128) of the IEEE 754 standard for floating-point arithmetic.

7.3.17. More accurate complex multiplication for embedded processors

In [36] we present some work in progress on the development of fast and accurate support for complex floatingpoint arithmetic on embedded processors. Focusing on the case of multiplication, we describe algorithms and implementations for computing both the real and imaginary parts with high relative accuracy. We show that, in practice, such accuracy guarantees can be achieved with reasonable overhead compared with conventional algorithms (which are those offered by current implementations and for which the real or imaginary part of a product can have no correct digit at all). For example, the average execution-time overheads when computing an FFT on the ARM Cortex-A53 and -A57 processors range from 1.04x to 1.17x only, while arithmetic costs suggest overheads from 1.5x to 1.8x.

7.3.18. Tight and rigorous error bounds for basic building blocks of double-word arithmetic

We analyze several classical basic building blocks of double-word arithmetic (frequently called "doubledouble arithmetic" in the literature): the addition of a double-word number and a floating-point number, the addition of two double-word numbers, the multiplication of a double-word number by a floating-point number, the multiplication of two double-word numbers, the division of a double-word number by a floatingpoint number, and the division of two double-word numbers. For multiplication and division we get better relative error bounds than the ones previously published. For addition of two double-word numbers, we show that the previously published bound was incorrect, and we provide a new relative error bound. We introduce new algorithms for division. We also give examples that illustrate the tightness of our bounds [21].

7.3.19. On the robustness of the 2Sum and Fast2Sum algorithms

The 2Sum and Fast2Sum algorithms are important building blocks in numerical computing. They are used (implicitely or explicitely) in many compensated algorithms (such as compensated summation or compensated polynomial evaluation). They are also used for manipulating floating-point expansions. We show that these algorithms are much more robust than it is usually believed: The returned result makes sense even when the rounding function is not round-to-nearest, and they are almost immune to overflow [9].

7.3.20. Formal verification of a floating-point expansion renormalization algorithm

Many numerical problems require a higher computing precision than the one offered by standard floatingpoint formats. A common way of extending the precision is to use floating-point expansions. As the problems may be critical and as the algorithms used have very complex proofs (many sub-cases), a formal guarantee of correctness is a wish that can now be fulfilled, using interactive theorem proving. In this article we give a formal proof in Coq for one of the algorithms used as a basic brick when computing with floating-point expansions, the renormalization, which is usually applied after each operation. It is a critical step needed to ensure that the resulted expansion has the same property as the input one, and is more "compressed". The formal proof uncovered several gaps in the pen-and-paper proof and gives the algorithm a very high level of guarantee [30].

7.3.21. Interactive proof protocols

We present in [46] an interactive probabilistic proof protocol that certifies in $(\log N)^{O(1)}$ arithmetic and Boolean operations for the verifier for example the determinant of an $N \times N$ matrix over a field whose entries are given by a single $(\log N)^{O(1)}$ -depth arithmetic circuit, which contains $(\log N)^{O(1)}$ field constants and which is polynomial time uniform. The prover can produce the interactive certificate within a $(\log N)^{O(1)}$ factor of the cost of computing the determinant. Our protocol is a version of the proofs for muggles protocol by Goldwasser, Kalai and Rothblum [STOC 2008, J. ACM 2015]. More generally, our verifier checks a computation on a family of circuits of size $N^{O(1)}$, or even $2^{(\log N)^{O(1)}}$, for $g_N(f_N(0), ..., f_N(N-1))$ in $(\log N)^{O(1)}$ bit communication and bit-operation complexity. Here g_N is a family of $(\log N)^{O(1)}$ -depth circuits, and f_N is a family of $(\log N)^{O(1)}$ -depth circuits for the scalars (such as hypergeometric terms); f_N can contain $(\log N)^{O(1)}$ input field constants. If the circuits f_N for the scalars are of size $(\log N)^{O(1)}$, they are input for the verifier. The circuit g_N and in the general case f_N are $N^{O(1)}$ -sized and cannot be built by the verifier with poly-log complexity. The verifier rather accesses the circuits via algorithms that probe the circuit structures, which are called uniformity properties.

7.3.22. New development on GNU MPFR

Work on the new fast, low-level algorithm to compute the correctly rounded summation of several floatingpoint numbers in arbitrary precision in radix 2 (each number having its own precision), and its implementation in GNU MPFR (new mpfr_sum function), has been completed [23].

The basic operations of GNU MPFR have also been optimized in small precision, and faithful rounding (mainly for internal use) is now partly supported [39].

These improvements, among many other ones, will be available in GNU MPFR 4.0.0; a release candidate is distributed in December 2017.

AVALON Project-Team

6. New Results

6.1. Energy Efficiency in HPC and Large Scale Distributed Systems

Participants: Mathilde Boutigny, Radu Carpa, Marcos Dias de Assunção, Thierry Gautier, Olivier Glück, Laurent Lefèvre, Jean-Christophe Mignot, Issam Rais.

6.1.1. Combining Shutdown Policies with Multiple Constraints

Large scale distributed systems (high performance computing centers, networks, data centers) are expected to consume huge amounts of energy. In order to address this issue, shutdown policies constitute an appealing approach able to dynamically adapt the resource set to the actual workload. However, multiple constraints have to be taken into account for such policies to be applied on real infrastructures: the time and energy cost of switching on and off, the power and energy consumption bounds caused by the electricity grid or the cooling system, and the availability of renewable energy. We propose models translating these various constraints into different shutdown policies that can be combined for a multi-constraint purpose. Our models and their combinations are validated through simulations on a real workload trace [4], [13]. This work is done through the PhD of Issam Rais in the FSN ELCI Project with the collaboration of Anne Benoit (Roma team) and Anne-Cécile Orgerie (Myriads team).

6.1.2. Evaluating the Impact of SDN-Induced Frequent Route Changes on TCP Flows

Traffic engineering technologies such as MPLS have been proposed to adjust the paths of data flows according to network availability. Although the time interval between traffic optimisations is often on the scale of hours or minutes, modern SDN techniques enable reconfiguring the network more frequently. It is argued, however, that changing the paths of TCP flows too often could severely impact their performance by incurring packet loss and reordering. This work analyses and evaluates the impact of frequent route changes on the performance of TCP flows. Experiments carried out on a network testbed show that rerouting a flow can affect its throughput when reassigning it a path either longer or shorter than the original path. Packet reordering has a negligible impact when compared to the increase of RTT. Moreover, constant rerouting influences the performance of the congestion control algorithm. Designed to assess the limits on SDN-induced reconfiguration, a scenario where the traffic is rerouted every 0.1s demonstrates that the throughput can be as low as 35% of that achieved without rerouting.[7], [14].

6.1.3. Evaluating Energy Consumption of OpenMP Runtime

In a joint-work with J.V. Lima from UFSM, Santa Maria, Brazil [26], we analyse performance and energy consumption of four OpenMP runtime systems over a NUMA platform. We present an experimental study to characterize OpenMP runtime systems on the three main kernels in dense linear algebra algorithms (Cholesky, LU and QR) in terms of performance and energy consumption. Our experimental results suggest that OpenMP runtime systems can be considered as a new energy leverage. For instance, a LU factorization with concurrent write extension from libKOMP achieved up to 1.75 of performance gain and 1.56 of energy decrease.

6.2. Modeling and Simulation of Parallel Applications and Distributed Infrastructures

Participant: Frédéric Suter.

6.2.1. Simulating MPI Applications: the SMPI Approach

Predicting the behavior of distributed algorithms has always been a challenge, and the scale of next-generation High Performance Computing (HPC) systems will only make the situation more difficult. Performance modeling and software engineering for these systems increasingly require a simulation-based approach, and this need will only become more apparent with the arrival of Exascale computing by the end of the decade. In [6] we summarized our recent work and developments on SMPI, a flexible simulator of MPI applications. In this tool, we took a particular care to ensure our simulator could be used to produce fast and accurate predictions in a wide variety of situations. Although we did build SMPI on SimGrid whose speed and accuracy had already been assessed in other contexts, moving such techniques to a HPC workload required significant additional effort. Obviously, an accurate modeling of communications and network topology was one of the key to such achievements. Another less obvious key was the choice to combine in a single tool the possibility to do both offline and online simulation.

6.2.2. Modeling Distributed Platforms from Application Traces

Simulation is a fast, controlled, and reproducible way to evaluate new algorithms for distributed computing platforms in a variety of conditions. However, the realism of simulations is rarely assessed, which critically questions the applicability of a whole range of findings.

In [15], we present our efforts to build platform models from application traces, to allow for the accurate simulation of file transfers across a distributed infrastructure. File transfers are key to performance, as the variability of file transfer times has important consequences on the dataflow of the application. We present a methodology to build realistic platform models from application traces and provide a quantitative evaluation of the accuracy of the derived simulations. Results show that the proposed models are able to correctly capture real-life variability and significantly outperform the state-of-the-art model.

6.3. Data Stream Processing and Edge Computing

Participants: Eddy Caron, Marcos Dias de Assunção, Alexandre Da Silva Veith, Laurent Lefèvre, Felipe Rodrigo de Souza.

6.3.1. Resource Elasticity and Edge Computing for Data Stream Processing

We carried out an extensive survey on techniques for enabling resource elasticity for data stream processing applications. Moreover we have been investigating algorithms for placing stream processing tasks onto environments that comprise both cloud and edge computing resources [29].

We are currently working on modelling the placement scenario as a constraint programming problem as well as measuring the energy consumption of constrained devices, such as Raspberry Pi's. The power consumption information is being used for creating a model on power consumption model.

6.4. Large-Scale Cloud Resource Management

Participants: Yves Caniou, Eddy Caron, Marcos Dias de Assunção, Christian Perez, Pedro de Souza Bento Da Silva.

6.4.1. An Efficient Communication Aware Heuristic for Multiple Cloud Application Placement

To deploy a distributed application on the cloud, cost, resource and communication constraints have to be considered to select the most suitable Virtual Machines (VMs), from private and public cloud providers. This process becomes very complex in large scale scenarios and, as this problem is NP-Hard, its automation must take scalability into consideration. In this work [21], we propose a heuristic able to calculate initial placements for distributed component-based applications on possibly multiple clouds with the objective of minimizing VM renting costs while satisfying applications' resource and communication constraints. We evaluate the heuristic performance and determine its limitations by comparing it to other placement approaches, namely exact algorithms and meta-heuristics. We show that the proposed heuristic is able to compute a good solution much faster than them.

6.4.2. Production Deployment Tools for IaaSes: an Overall Model and Survey

Emerging applications for the Internet of Things (IoT) are complex programs which are composed of multiple modules (or services). For scalability, reliability and performance, modular applications are distributed on infrastructures that support utility computing (*e.g.*, Cloud, Fog). In order to simply operate such infrastructures, an Infrastructure-as-a-Service (IaaS) manager is required. OpenStack is the de-facto open-source solution to address the IaaS level of the Cloud paradigm. However, OpenStack is itself a large modular application composed of more than 150 modules that make it hard to deploy manually. To fully understand how IaaSes are deployed today, we propose in [16] an overall model of the application deployment process which describes each step with their interactions. This model then serves as the basis to analyse five different deployment tools used to deploy OpenStack in production: Kolla, Enos, Juju, Kubernetes, and TripleO. Finally, a comparison is provided and the results are discussed to extend this analysis.

6.4.3. Communication Aware Task Placement for Workflow Scheduling on DaaS-based Cloud

We proposed a framework for building an autonomous workflow manager and developped the different components that are required for this design to work. We believe that this design will help solve current issues with workflow deployment and scalling in the context of shared IaaS Cloud platforms. In that regard, our first contribution is the modelization of network topology [24], which is a key factor in predicting communication patterns and should therefore be considered by clustering algorithms. By designing a generic network model, we managed to improve the results of static scheduling in the context of DaaS-based Cloud platforms. In fact, the resulting clusters are both more efficient in terms of makespan (primary objective) and in terms of deployment cost compared to previous non-network-aware clustering algorithms.

6.4.4. Communication Aware Stochastic Tasks Scheduling Composing Scientific Workflows on a Cloud

In order to study the scheduling of workflows composed of stochastic tasks on a set of resources managed as a cloud, we firstly proposed a new execution model taking into account data transfers, heterogeneity, billing of used resources as close to reality based to a great extend on the offers of three big cloud providers: Google Cloud, Amazon EC2 and OVH [25]. We then studied new scheduling heuristics on a set of worflows taken from the Pegasus benchmark suite [23]. During the mapping process, the budget-aware algorithms make conservative assumptions to avoid exceeding the initial budget; we further improve our results with refined versions that aim at re-scheduling some tasks onto faster virtual machines, thereby spending any budget fraction leftover by the first allocation. These refined variants are much more time-consuming than the former algorithms, so there is a trade-off to find in terms of scalability. We report an extensive set of simulations. Most of the time our budget-aware algorithms succeed in achieving efficient makespans while enforcing the given budget, and despite the uncertainty in task weights.

6.5. HPC Component Models and Domain Specific Languages

Participants: Thierry Gautier, Christian Perez, Jérôme Richard.

6.5.1. Combining Both a Component Model and a Task-based Model for HPC Applications: a Feasibility Study on GYSELA

In [12], we studied the feasibility of efficiently combining both a software component model and a taskbased model. Task based models are known to enable efficient executions on recent HPC computing nodes while component models ease the separation of concerns of application and thus improve their modularity and adaptability. This paper describes a prototype version of the COMET programming model combining concepts of task-based and component models, and a preliminary version of the COMET runtime built on top of StarPU and L2C. Evaluations of the approach have been conducted on a real-world use-case analysis of a subpart of the production application GYSELA. Results show that the approach is feasible and that it enables easy composition of independent software codes without introducing overheads. Performance results are equivalent to those obtained with a plain OpenMP based implementation.

6.5.2. Extensibility and Composability of a Multi-Stencil Domain Specific Framework

As the computation power of modern high performance architectures increases, their heterogeneity and complexity also become more important. One of the big challenges of exascale is to reach programming models that give access to high performance computing (HPC) to many scientists and not only to a few HPC specialists. One relevant solution to ease parallel programming for scientists is domain specific language (DSL). However, one problem to avoid with DSLs is to mutualize existing codes and libraries instead of implementing each solution from scratch. For example, this phenomenon occurs for stencil-based numerical simulations, for which a large number of languages has been proposed without code reuse between them. The Multi-Stencil Framework (MSF) presented in this paper [5] combines a new DSL to component-based programming models to enhance code reuse and separation of concerns in the specific case of stencils. MSF can easily choose one parallelization technique or another, one optimization or another, as well as one backend implementation or another. It is shown that MSF can reach same performances than a non component-based MPI implementation over 16,384 cores. Finally, the performance model of the framework for hybrid parallelization is validated by evaluations.

BEAGLE Project-Team

6. New Results

6.1. The fossil record of microbes augmented by an order of magnitude

Participants: Eric Tannier

Biodiversity has always been predominantly microbial and the scarcity of fossils from bacteria, archaea and microbial eukaryotes has prevented a comprehensive dating of the tree of life. We have shown that patterns of lateral gene transfer deduced from the analysis of modern genomes encode a novel and abundant source of information about the temporal coexistence of lineages throughout the history of life. We constructed and used new phylogenetic methods to reconstruct the history of thousands of gene families and demonstrate that dates implied by gene transfers are consistent with estimates from relaxed molecular clocks in bacteria, archaea and eukaryotes. An inspection of discrepancies between transfers and clocks and a comparison with mammal fossils show that gene transfer in microbes is potentially as informative for dating the tree of life as the geological record in macroorganisms.

Main publications: [30]

6.2. Phylogenetics of dependence, and dependence of phylogenies

Participants: Wandrille Duchemin, Eric Tannier

Standard phylogenetics use DNA or protein sequences, along with probabilistic models of substitutions, which are Markov processes on trees. The big default of this methodology is to assume a common evolution of all sites inside a gene, and a total independence with other genes. This model does not capture the essence of living things, which is made of dependencies and interactions. We made several methodological developments to take into account these dependencies, by improving the gene tree species tree reconciliation methods and reconstructing phylogenies of relations between genes.

Publications: [16], [32], [19]

6.3. Beware batch culture: Seasonality and niche construction predicted to favor bacterial adaptive diversification

Participants: Charles Rocabert, Carole Knibbe, Guillaume Beslon

The evolution of stable bacterial cross-feeding interactions is often considered as the first step toward bacterial speciation in sympatry. It is thus important to study the conditions favoring the emergence and the stabilization of cross-feeding interactions in well-mixed environments. Experimental evolution in laboratory, where fast organisms are replicated for thousands of generations in controlled conditions, provides important insights on this question. Indeed cross-feeding is commonly observed in batch cultures or in chemostat. However, the reasons why cross-feeding interactions become stable and lead to monophyletic ecotypes remain unclear. Because laboratory experiments are a long and costly process, we explored this question by evolving digital organisms in artificial systems mimicking the conditions of wet experiments.

Models of digital evolution helped a lot to decipher the evolution of cross-feeding interactions. However, the evolution of real microorganisms implies the interaction of a wide range of biological structures and levels, while those models often include only two or three levels, limiting their ability to mimic real experiments. In this work, we developed a new multi-scale model of digital evolution, integrating a complex and realistic genotype-to-phenotype mapping (including a metabolic network) and a complex environment (that links the organism's metabolic networks together, opening the possibility for cross-feeding). This model has been developed under the European project EvoEvo, and has been inspired by previous models developed by the Beagle team, and in the Theoretical Biology and Bioinformatics group at Utrecht University. By mimicking laboratory experiment setups and running simulations for tens of thousands of generations, we were able to recover ecological dynamics similar to those found in real experiments.

In batch culture, like in the Long Term Evolution Experiment (LTEE), it is accepted that the seasonality generated by the serial transfers triggers the maintenance of cross-feeding interactions on the long-term by favoring niche construction and specialization. In chemostat, cross-feeding interactions are observed and seem stable for a few hundreds of generations but the reasons of their stability remain unclear. Thanks to our model, we were able to observe stable cross-feeding interactions reproducing the same properties as those observed in the LTEE. We then showed that seasonal conditions found in batch cultures are essential for the maintenance of stable ecotypes on the long-term, since it produces conditions for niche construction and stable cross-feeding. In chemostat conditions, the absence of seasonality and competitive exclusion precludes any stabilization of emerging cross-feeding interactions. Finally, we proposed to consider a cross-feeding interaction to be stable only if interacting ecotypes undergo independent periodic selection events on the long-term. Stable cross-feeding interactions could then be considered as premises to speciation in sympatry.

This work is the result of an enriching collaboration between the Beagle team (Charles Rocabert, Carole Knibbe, Guillaume Beslon), and microbiologists from the TIMC-IMAG in Grenoble (Jessika Consuegra, Dominique Schneider). It has been published in the renowned journal PLoS Computational Biology in January 2017. This work is of interest for the fields of evolutionary biology, microbiology but also for computer science. Indeed our findings also suggest that digital evolution is a useful tool to study bacterial evolution, and that the use of models integrating a complex genotype-to-phenotype mapping and complex interactions between digital organisms and their environment is important to accurately study real biological systems thus appealing for further fruitful transdisciplinary collaborations. Publication: [21].

6.4. Evolution of phenotypic noise

Participants: Charles Rocabert, Guillaume Beslon, Carole Knibbe

The phenotype of an organism is a complex non-linear cascade of developmental, physiological and regulatory processes, formalized by the concept of genotype-to-phenotype map. An increasing number of experimental studies demonstrate the existence of phenotypic noise, which can be finely tuned by the genotype-to-phenotype map, and that phenotypic noise can be adaptive.

In stabilizing selection, when the population is at a fitness optimum, phenotypic noise is deleterious and minimized by evolution. Nevertheless, phenotypic noise can be positively selected when the population is exposed to stressful conditions. It was thus suggested that during an adaptation event, phenotypic noise would increase in directional evolution, and then be reduced when the selection becomes stabilizing. In 1930, R.A. Fisher suggested with its so-called Fisher's Geometric Model (FGM) that organisms adapting to a new environment experience a "cost of complexity", where beneficial mutations become increasingly harder to fix when the number of phenotypic characters increases. Predictions made on the evolution of phenotypic noise are mostly based on single trait observations. Is there also a cost of complexity on the phenotypic noise?

To address this question, we extended the FGM by adding an evolvable phenotypic noise. First, using a simple form of noise, affecting similarly every phenotypic character, we show that a cost of complexity indeed makes phenotypic noise deleterious in directional evolution. Second, we extended the FGM with a fully evolvable noise, allowing evolution on noise amplitudes on each character, as well as on noise correlations between characters. In directional evolution, we show that phenotypic noise evolves towards a flattened shape, with elevated noise in the direction of the optimum, and minimized noise in all other directions. In this case, the noise becomes advantageous again, even with many characters. Non-isotropic phenotypic noise thus facilitates evolution towards the fitness optimum, and significantly reduces the cost of complexity. Our results show that such non-isotropic phenotypic noise could be exploited by evolution, and suggest further experiments to assess the functional nature of phenotypic noise.

This result is currently under review for the Evolution journal. It is the result of an enriching collaboration between the Beagle team (Charles Rocabert, Guillaume Beslon, Carole Knibbe), and the Dracula team (Samuel Bernard). Although the results are grounded in theory and mathematical modeling, they provide stringent conditions for noise to be beneficial, which are experimentally testable. We believe the results to be of wide interest for researchers working on phenotypic evolution. By deciphering the conditions in which phenotypic noise evolves towards specific patterns, our work may also contribute to a better understanding of drug resistance and cancer cells proliferation, and also to the growing field of predictive biology.

6.5. Impact of group size and social composition on group vocal activity and acoustic network in a social songbird

Participants: Marie Fernandez, Hédi Soula

In social species individuals living in the same group may synchronize activities such as movements, foraging or antipredator vigilance. Vocalizations are behaviors that can be coordinated between individuals, but simultaneous vocalizations in groups have mostly been considered as noise that does not bear any information. Indeed, little is known about the structure and function of vocal communications involving a network of individuals. Zebra finches, Taeniopygia guttata, are social, monogamous songbirds that form lifelong pair bonds. In the wild, they are typically found in small groups and they gather in 'social' trees where they produce vocalizations. Here we investigated in the laboratory the influence of group size and composition on general vocal activity and synchrony, as well as the influence of pair bond and spatial location on the finer characteristics of dyads' vocal interactions. We used a set-up that locked the birds at fixed spatial positions of our choosing to control the proximity network and allowed us to match most of the vocalizations with specific individuals. We used an in-house software suite that automatically detects vocalizations from hours of passive recording. We found that zebra finch groups synchronized their general vocal activity with waves of collective vocalizations, which depended on both the size and the composition of the group. The acoustic network was shaped by pair bonds at different timescales.

Publication: [17].

6.6. Dopamine-endocannabinoid interactions mediate spike-timing dependent potentiation in the striatum

Participants: Hugues Berry, Alexandre Foncelle, Ilya Prokin

Synaptic long-term plasticity underlies multiple forms of learning and memory in the brain. In most systems, its bidirectionality - depression (LTD) and potentiation (LTP) allows adaptive adjustment of the synaptic weight depending on the activity. Endocannabinoids (eCBs), one of the most widespread neurotransmitter systems, are very well established as depressing neuronal communication but recent experimental evidence challenges this depression-only vision. Our previous work in collaboration with L. Venance's lab at CIRB, College de France, Paris (experimental neuroscience) has combined experimental and mathematical modeling approaches to identify in the basal ganglia the existence of an eCB-mediated spike-timing dependent LTP (eCB-tLTP) induced by a low number of paired stimulations [50], [49]. However, the regulation and control mechanisms of eCB-tLTP remained unknown. Using the same combination of experimental and modelling approaches, we have now discovered that dopamine controls eCB-tLTP. The dopamine system is a key actor of associative learning in the basal ganglia and a pivotal system in several pathologies including Parkinson's disease. We identified that eCB-tLTP depends on dopamine and involves the activation of D2R dopamine receptors located presynaptically in corticostriatal glutamatergic afferents. We moreover show that dopamine control of eCB-tLTP is of pathological significance since it is impaired in a rodent model of Parkinson's disease and rescued by chronic L-DOPA treatment in those animals. Combining our experimental finding with a realistic mathematical model of the underlying signaling pathway, we could describe the mechanisms accounting for this endocannabinoid-dopamine regulation of eCB-tLTP.

This result is currently under review in the journal Nature Communications.

6.7. Subspace clustering based on medians using evolutionary algorithms

Participants: Sergio Peignier, Christophe Rigotti, Jonas Abernot, Guillaume Beslon

Subspace clustering is a data mining task that searches for objects sharing similar features, and at the same time looks for the subspaces where these similarities appear. For this reason subspace clustering is recognized as more general and complicated than standard clustering, since it needs to detect these relevant subspaces. Taking advantage of the expertise of the team in evolution in silico, we previously showed that evolutionary algorithms are promising approaches to address this problem. Another important clustering task is the K-medians one, where objects are grouped around medians, leading to cluster centers more robust to noise and outliers. In order to take advantage of these benefits within the subspace clustering process itself, we developed a new evolutionary algorithm, KymeroClust, that builds cluster centers that are medians in subspaces. This algorithm takes advantage of an evolvable representation of the genotypes to adapt the numbers of clusters produced and the subspace dimensionalities. It is based on new bio-inspired mutation operators to evolve the cluster centers as medians and is able to handle streaming data. KymeroClust has been compared to the main subspace clustering methods and turns out to be very competitive both in terms of cluster quality and runtime, while requiring an easier parameter setting.

Publications: [12], [24]
BIPOP Project-Team

7. New Results

7.1. The contact complementarity problem, and Painlevé paradoxes

Participants: Bernard Brogliato, Florence Bertails-Descoubes, Alejandro Blumentals.

The contact linear complementarity problem is an set of equalities and complementarity conditions whose unknowns are the acceleration and the contact forces. It has been studied in a frictionless context with possibly singular mass matrix and redundant constraints, using results on well-posedness of variational inequalities obtained earlier by the authors. This is also the topic of the first part of the Ph.D. thesis of Alejandro Blumentals where the frictional case is treated as a perturbation of the frictionless case. With R. Kikuuwe from Kyushu University, we have also proposed a new formulation of the Baumgarte's stabilisation method, for unilateral constraints and Coulomb's friction, which sheds new light on Painlevé paradoxes [24]. It relies on a particular limiting process of normal cones.

7.2. Discrete-time sliding mode control

Participants: Vincent Acary, Bernard Brogliato.

This topic concerns the study of time-discretized sliding-mode controllers. Inspired by the discretization of nonsmooth mechanical systems, we propose implicit discretizations of discontinuous, set-valued controllers [3]. This is shown to result in preservation of essential properties like simplicity of the parameters tuning, suppression of numerical chattering, reachability of the sliding surface after a finite number of steps, and disturbance attenuation by a factor h or h^2 . This work was part of the ANR project CHASLIM. Within the framework of CHASLIM we have performed many experimental validations on the electropneumatic setup of IRCCyN (Nantes), which nicely confirm our theoretical and numerical predictions: the implicit implementation of sliding mode control, drastically improves the input and output chattering behaviours, both for the classical order-one ECB-SMC and the twisting algorithms. In particular the high frequency bangbang controllers which are observed with explicit discretizations, are completely suppressed. The implicit discretization has been applied to the classical equivalent-based-control SMC, and also to the twisting sliding-mode controller [48]. The previous results deal with disturbances which are matched and uniformly upperbounded. In [26] the SMC of Lagrange systems is studied. In [35] this is extended to the case of parametric uncertainties, which are more difficult to handle because they may yield unmatched equivalent disturbances, and these disturbances are not uniformly upperbounded by a constant. A nested controller is proposed in [36], using a backstepping-like approach for the controller synthesis.

7.3. Linear Complementarity Systems

Participants: Bernard Brogliato, Christophe Prieur, Alexandre Vieira.

The quadratic optimal control of Linear complementarity systems (LCS) is analysed in [37]. The major difficulty comes from the fact that complementarity conditions introduce non convex constraints. Suitable algorithms have to be used to solve the MPEC problems for solving the direct method. The indirect (Pontryagin's) approach is quite delicate and is currently analysed in the PhD thesis of A. Vieira.

7.4. Numerical analysis of multibody mechanical systems with constraints

This scientific theme concerns the numerical analysis of mechanical systems with bilateral and unilateral constraints, with or without friction [2]. They form a particular class of dynamical systems whose simulation requires the development of specific simulators.

7.4.1. Numerical time-integration methods for event-detecting schemes.

Participants: Vincent Acary, Bernard Brogliato.

The CIFRE thesis of M. Haddouni concerns the numerical simulation of mechanical systems subject to holonomic bilateral constraints, unilateral constraints and impacts. This work is performed in collaboration with ANSYS and the main goal is to improve the numerical time–integration in the framework of event-detecting schemes. Between nonsmooth events, time integration amounts to numerically solving a differential algebraic equations (DAE) of index 3. We have compared dedicated solvers (Explicit RK schemes, Half-explicit schemes, generalizes α -schemes) that solve reduced index formulations of these systems. Since the drift of the constraints is crucial for the robustness of the simulation through the evaluation of the index sets of active contacts, we have proposed some recommendations on the use of the solvers of dedicated to index-2 DAE. The results are reported in [23].

7.4.2. Multibody systems with clearances (dynamic backlash)

Participants: Vincent Acary, Bernard Brogliato.

The PhD thesis of N. Akadkhar under contract with Schneider Electric has concerned the numerical simulation of mechanical systems with unilateral constraints and friction, where the presence of clearances in imperfect joints plays a crucial role. A first work deals with four-bar planar mechanisms with clearances at the joints, which induce unilateral constraints and impacts, rendering the dynamics nonsmooth. The objective is to determine sets of parameters (clearance value, restitution coefficients, friction coefficients) such that the system's trajectories stay in a neighborhood of the ideal mechanism (*i.e.* without clearance) trajectories. The analysis is based on numerical simulations obtained with the projected Moreau-Jean time-stepping scheme. Circuits breakers with 3D joint clearances have been studied in [20], [31] where it is demonstrated that the nonsmooth dynamics approach as coded in our software SICONOS, allows a very good prediction of the system's dynamics, with experimental validation. An overview of various approaches for the feedback control of multibody systems with joint clearances is proposed in [21].

7.5. Nonlinear waves in dissipative granular chains

Participants: Guillaume James, Bernard Brogliato, Kirill Vorotnikov.

Granular chains made of aligned beads interacting by contact (e.g. Newton's cradle) are widely studied in the context of impact dynamics and acoustic metamaterials. In order to describe the response of such systems to impacts or vibrations, it is important to analyze different wave effects such as the propagation of localized compression pulses (solitary waves) or the scattering of vibrations through the chain. Such phenomena are strongly influenced by contact nonlinearities (Hertz force), spatial inhomogeneities and dissipation. In this work, we analyze the Kuwabara-Kono (KK) model for contact damping, and we develop new approximations of this model which are efficient for the simulation of multiple impacts. The KK model is a simplified viscoelastic contact model derived from continuum mechanics, which allows for simpler calibration (using material parameters instead of phenomenological ones), but its numerical simulation requires a careful treatment due to its non-Lipschitzian character. Using different dissipative time-discretizations of the conservative Hertz model, we show that numerical dissipation can be tuned properly in order to reproduce the physical dissipation of the KK model and associated wave effects. This result is obtained analytically in the limit of small time steps (using methods from backward analysis) and is numerically validated for larger time steps. The resulting schemes turn out to provide good approximations of impact propagation even for relatively large time steps.

7.6. Periodic motions of coupled impact oscillators

Participants: Guillaume James, Vincent Acary, Franck Pérignon.

In the work [40], we study the existence and stability of time-periodic oscillations in an infinite chain of linearly coupled impact oscillators, for rigid impacts without energy dissipation. We reformulate the search of periodic solutions as a boundary value problem incorporating unilateral constraints. This formulation, together with an appropriate notion of nondegenerate modes, allows us to construct nonsmooth modes of oscillations (spatially localized or extended) when the oscillators are weakly coupled (this approach is an adaptation of the idea of "anticontinuum" limit to the nonsmooth setting). In this framework, we show the existence of exact solutions (in particular, we check the condition of non-penetration of the obstacle) for an arbitrary number of impacting particles. Different solution branches corresponding to stable or unstable breathers, multibreathers and nonsmooth normal modes are found. We provide a formula for the monodromy matrix that determines spectral stability of nonsmooth modes in the presence of simple impacts. These results are completed by a numerical computation of the time-periodic solutions at larger coupling, and the Siconos software is used to simulate the system and explore dynamical instabilities. The above approach is much more effective than numerical continuation of periodic solutions based on stiff compliant models, which leads to stiff ODEs and costly numerical continuation.

7.7. Stability analysis for rogue waves

Participant: Guillaume James.

The study of rogue waves (large amplitude waves localized both in space and time) has gained importance in various fields, such as the mathematical modeling of water waves and nonlinear optics. The analysis of their stability is delicate because of their transient nature. In the work [46], we introduce a new method to tackle this problem. Our approach relies on the approximation of rogue waves by large amplitude breathers (localized in space and time-periodic) having a large period, and the use of Floquet theory to analyze breather stability. This problem is examined for the nonlinear Schrödinger equation, which describes the envelope of nonlinear waves in a large class of systems, for example granular chains [15]. This model admits a family of breather solutions (Kuznetsov-Ma breathers) which converge to a rogue-wave profile (Peregrine soliton) when their period tends to infinity. We show numerically that the Floquet exponents of the breathers approach a finite limit for large periods, and observe that a motion of the localized wave can be induced by a dynamical instability. This work suggests an analytical way to define the spectral stability of the (transient) Peregrine soliton, but this remains an open problem to prove analytically the convergence of Floquet exponents in the limit of infinite period.

7.8. Travelling waves in a spring-block chain sliding down a slope

Participants: Guillaume James, Jose Eduardo Morales Morales, Arnaud Tonnelier.

Spatially discrete systems (lattice differential equations) have a wide range of applications in natural sciences, engineering and social sciences. They frequently occur in physics as mass-spring systems with nearest-neighbors coupling and they have been used extensively to describe the dynamics of microscopic structures such as crystals or micromechanical systems, or to model fragmentation phenomena.

In this work, we consider a spring-block system that slides down a slope due to gravity. Each block is subjected to a nonlinear friction force. This system differs from the Burridge-Knopoff model considered for the modeling of earthquakes, which incorporates local potentials. We perform numerical simulations of the coupled system and show that the bistability property induces traveling patterns, as fronts and pulses. For a piecewise-linear spinodal friction law, a closed-form expression of front waves is derived. Pulse waves are obtained as the matching of two travelling fronts with identical wave speeds. Explicit formulas are obtained for the wavespeed and the wave form in the anti-continuum limit. The link with propagating phenomena in the Burridge-Knopoff model is briefly discussed. These results have been published in [27].

7.9. Solitary waves in the excitable Burridge-Knopoff model

Participants: Guillaume James, Jose Eduardo Morales Morales, Arnaud Tonnelier.

The Burridge-Knopoff model is a lattice differential equation describing a chain of blocks connected by springs and pulled over a surface. This model was originally introduced to investigate nonlinear effects arising in the dynamics of earthquake faults. One of the main ingredients of the model is a nonlinear velocity-dependent friction force between the blocks and the fixed surface. We introduce a simplified piecewise linear friction law (reminiscent of the McKean nonlinearity for excitable cells) which allows us to obtain analytical expression of solitary waves and study some of their qualitative properties, such as wavespeed and propagation failure. These results have been reported in [28].

We have obtained an existence theorem for solitary waves in the Burridge-Knopoff model. Our approach uses a piecewise-linear friction force combined with a weak coupling strength. Using asymptotic arguments, we show that trial solutions, obtained semi-analytically, satisfy, for some parameter set, the inequality constraints associated with the threshold conditions. An approximation of the wave profile is obtained and a minimal wave speed is derived.

7.10. Propagation in space-discrete excitable systems

Participant: Arnaud Tonnelier.

We introduce a simplified model of excitable media where the response of an isolated element to an incoming signal is given by a fixed pulse-shape function. When the total activity of one element reaches a given threshold, a signal is sent to its N nearest neighbors. We show that an excitable chain supports the propagation of a set of simple traveling waves where the interval between the emitting time of two successive elements remains constant. We propose a classification of travelling waves that depends on the number of signals that are received by an element. Results on stability of travelling signals are derived. We also discussed the global shape of the speed curve (velocity of the wave with respect to the global coupling strength). In particular, we show that for a given network connectivity, different wave velocities can be obtained, i.e., depending on initial conditions, the network may propagate different signals. A comprehensive study is done for a transmission line with N = 2 and N = 3. Some necessary conditions for multistationarity are derived for an arbitrary N and for different network connectivities.

7.11. Inverse design of a suspended elastic rod

Participants: Florence Bertails-Descoubes, Victor Romero.

In collaboration with Alexandre Derouet-Jourdan (OLM Digital, Japan) and Arnaud Lazarus (UPMC, Laboratoire Jean le Rond d'Alembert), we have investigated the inverse design problem of a suspended elastic subject to gravity. We have proved that given an arbitrary space curve, there exists a unique solution for the natural configuration of the rod, which is independent of the initial framing of the input curve. Moreover, this natural configuration can be easily computed by solving three linear ODEs in sequence, starting from any input framing. This work has just been submitted for publication in physics.

7.12. Simulation of cloth contact with exact Coulomb friction

Participant: Florence Bertails-Descoubes.

In collaboration with Gilles Daviet (Weta Digital, New Zealand), Rahul Narain and Jie Lie (University of Minnesota), we have developed a new implicit solver for taking into account contact in cloth with Coulomb friction. Our key idea stems from the observation that for a nodal system like cloth, and in the case where each node is subject to at most one contacting constraint (either an external or self-contact), the frictional contact problem may be formulated based on velocities as primary variables, without having to compute the costly Delassus operator; then, by reversing the roles classically played by the velocities and the contact impulses, conical complementarity solvers of the literature may be leveraged to solve for compatible velocities at nodes. To handle the full complexity of cloth dynamics scenarios, we have extended this base algorithm in two ways: first, towards the accurate treatment of frictional contact at any location of the cloth, through an adaptive node refinement strategy; second, towards the handling of multiple constraints at each node, through the duplication of constrained nodes and the adding of pin constraints between duplicata. Our method proves to be both fast

and robust, allowing us to simulate full-size garments with more realistic body-cloth interactions compared to former methods, while maintaining similar computational timings. Our work will be submitted for publication to Siggraph 2018.

7.13. Model Predictive Control for biped walking motion generation

Participants: Pierre-Brice Wieber, Nestor Alonso Bohorquez Dorante, Nahuel Villa, Matteo Ciocca, Stanislas Brossette, Alexander Sherikov.

We proposed last year a feasible Newton scheme for nonlinear MPC by combining ideas from robust control and trust regions. This year, we applied this approach to nonlinearities that appear when adapting step durations [32]. We also investigated more thoroughly the strong recursive feasibility of our scheme [34] and how it could be adapted to situations with bounded uncertainty [29]. Finally, we proposed a new approach to collision avoidance based on separating planes [38].

CHROMA Project-Team

7. New Results

7.1. Bayesian Perception

Participants: Christian Laugier, Jean-Alix David, Thomas Genevois, Blanche Baudouin, Jerome Lussereau, Lukas Rummelhard [IRT Nanoelec], Tiana Rakotovao [CEA since October 2017], Nicolas Turro [SED], Jean-François Cuniberto [SED].

7.1.1. Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT)

Participants: Christian Laugier, Jean-Alix David, Thomas Genevois, Blanche Baudouin, Jerome Lussereau, Lukas Rummelhard [IRT Nanoelec], Amaury Nègre [Gipsa Lab], Nicolas Turro [SED].

The research work on Bayesian Perception has been done as a continuation and an extension of our previous research activity and results on this approach 0 . This work exploits the initial BOF⁰ paradigm (see section 3.2.1) and its recent new formulation and framework called $CMCDOT^0$ [76]. More details about these developments can be found in the Chroma Activity Report 2016.

The objective of the research work performed in 2017 on this topic, was to further refine the perception models and algorithms. Refinements have been made at three different levels:

The occupancy grid generation process has been improved in order to appropriately process 2D and 3D lidars. In particular, the new sensor model for 3D lidar (called GEOG-Estimator), which was initially designed by the team in 2016, has been improved by adding a temporal filtering step for both obtaining a more accurate estimation of the elevation of the point cloud elements and a better occupancy grid generation with respect to the sensor calibration. This work has been published and patented in 2017 [24] [23]. The results are illustrated by Figure 6 and Figure 7.

The sensor model for 2D lidar has also been modified in order to adapt it to the new paradigm introduce with GEOG-Estimator, with the objective to obtain a better fusion of the occupancy grids.



Figure 6. Example of Occupancy Grid generated using the classified point cloud and the Ground Estimator model.

- A new approach for fusing several occupancy grids has been developed to appropriately merge the • outputs of different sensors (2D/3D lidars, stereo cameras, ...),
- The CMCDOT framework, which takes as input the merged occupancy grid, has been improved by adding new filtering equations whose objective is mainly to provide better results. In addition, a new output format has been added in order to make it easier the connection with the decision & control related components (see below).

⁰This research activity has been started in the scope of the former Inria team-project e-Motion and it is now conducted (since 2015) in the scope of the Inria Chroma team ⁰Bayesian Occupancy Filter

⁰Conditional Monte Carlo Dense Occupancy Tracker



Figure 7. (a) Typical 3D point cloud generated by Velodyne LiDAR. (b) Point cloud segmentation between ground (green points) and non-ground (purple points), and estimated average elevation of the terrain (red grid). (c) Point Cloud Segmentation on 4-Ibeo Lux LiDAR data and estimated elevation of terrain.

All the above mentioned software modules are highly parallelizable. This is why a *GPU* implementation has been made and is continually optimized in order to obtain very efficient processing time and results. Now, the whole perception framework is able to run on low energy consumption embedded boards (Nvidia Jetson TX2). Thanks to this efficient embedded implementation, an industrial proof of concept on a commercial autonomous shuttle (from the EasyMile company) has successfully been done in a few weeks, Figure 8 illustrates.



Figure 8. CMCDOT on a commercial autonomous shuttle EZ10 from EasyMile. a. classified point cloud b. detected dynamic objects and their velocities c. portable-demonstrator.

These approach allow us to also develop a compact and portable demonstrator for conference or exhibitions, Figure 8 (fourth picture) illustrates this technology.

7.1.2. Simulation based validation

Participants: Thomas Genevois, Nicolas Turro [SED], Christian Laugier, Tiana Rakotovao [CEA since October 2017], Blanche Baudouin.

In 2017, we have started to address the concept of *simulation based validation* in the scope of the EU Enable-S3 project, with the objective of searching for novel approaches, methods, tools and experimental methodology for validating BOF-based algorithms. For that purpose, we have collaborated with the Inria Tamis team (Rennes) and with Renault for developing the simulation platform that is used in the test platform. The simulation of both the sensors and the driving environment are based on the Gazebo simulator. A simulation of the prototype car and its sensors has also been realized. In the simulator, the virtual Lidars generates the same format of data as the real ones, meaning that the same implementation of *CMCDOT* can handle both real data and simulated data. The test management component that generates random simulated scenarios has also been developed. Output of *CMCDOT* computed from the simulated scenarios are recorded by *ROS* and analyzed through the Statistical Model Checker developed by the Inria Tamis team.

Within the project Perfect, it has also been decided to use Gazebo to build our simulator. Gazebo is a simulation framework which gives many tools to simulate physics, sensors and actuators. Moreover, as Gazebo is fully

compatible with *ROS*, it becomes easy for us to connect the simulator with our own perception and control tools (which are all using *ROS*).

We have developed a Gazebo model of the Inria Renault Zoe demonstrator, including physical properties (friction, inertia, sliding), sensors models (lidars and odometry) and real-like actuators (steering, acceleration and brake), Figure 9 (first picture) illustrates. All parameters have been tuned to match with the actual vehicle and its equipment. In addition, several plug-in programs allow us to drive the virtual vehicle with the same commands as the actual vehicle; similarly, the simulated sensors provide data in the same format as the actual sensors. Therefore the simulated vehicle behaves much like the actual car, and any program running on the actual car can be directly used on the simulated model without any adaptation. This simulation is almost real time which is very convenient for development purposes.

Thanks to this simulated model, we have safely developed control and navigation software (emergency braking, path following and local navigation with obstacle avoidance). During this work, the simulation helped us to detect all issues and fully debug the programs before executing actual experiments. It showed that our model is well designed to behave like the real Zoe. Figure 9 illustrates the graphic output of the system.

Future work aims at improving these models for obtaining faster execution, better low level physical simulation (engine, brake, suspension), and new types of simulation sensors (cameras, IMU, radar, etc.). An industrial partner has already shown direct interest to our simulation model and asked for an adaptation to his own vehicle (no details can be given since this work is confidential).





Figure 9. First picture: Display of the Zoe virtual model. Second picture: CMCDOT output computed by the simulator (using simulated sensors data). Third picture: Simulated scenario on Gazebo.

7.1.3. Control and navigation

Participants: Thomas Genevois, Christian Laugier, Nicolas Turro [SED].

In January 2017, within a partnership with Ecole Centrale de Nantes, we realized a wired based control kit on the Renault Zoe experimental vehicle of the team.

We have first developed an *emergency braking system* to avoid collision during manual driving. This system triggers when the risk estimation provided by the *CMCDOT* reaches a threshold. Significant engineering work was necessary for being able to control the brake with the software, while letting the driver to drive normally. This has been successfully tested with both the simulator and the real vehicles in the PTL experimental platform of IRT-Nanoelec, Figure 10 illustrates.

We plan to improve this with different level of risk (long term, middle term and short term) which would trigger different actions (warning, progressive braking, emergency braking). To do so we will need to improve long term and middle term risk estimation reliability.

In a second step, we have started working on a *safe local navigation* component. For that purpose, we have implemented a Dynamic Window Approach (DWA) local planner based on occupancy grid. The DWA approach consists in computing online a set of feasible trajectories for the vehicle (in terms of vehicle control and no-collision in the near future). Then, a score is associated to each trajectory considering its collision

risk, its heading with respect to the goal and its speed. Finally the trajectory with the best score is chosen and provided as a reference for the low level controller. The DWA technique affords a simple and flexible architecture and we took advantage of it to introduce the notion of time-to-collision within the score function. At the moment, the time-to-collision is computed from the occupancy grid provided by *CMCDOT*.

This approach has been successfully tested with in simulation and with the real vehicle on the PTL experimental platform of IRT-Nanoelec. It showed a good behaviour at slow speeds (<20km) on simple slalom and going through automated barriers (stops in front of the barriers and restarts when the barrier opens), Figure 10 illustrates.



Figure 10. Slalom-like obstacle course experiment in fully autonomous navigation mode

Current and future work aims at improving the approach at two levels: (1) developing a parallel implementation in Cuda on GPU for drastically improving the computation time, allowing in this way to compute in realtime several shapes of trajectories (i.e. for obtaining a better global motion) and to consider higher velocities and dynamic objects (i.e. anticipating the future motion of the moving objects); (2) improving the low level controller and the dynamic model of the vehicle (e.g. for computing the future positions of the vehicle from its odometry and the applied control commands), for obtaining a more precise control and better performances at higher speeds.

7.2. Situation Awareness

Participants: Christian Laugier, Olivier Simonin, Jilles Dibangoye, David Sierra-Gonzalez, Mathieu Barbier, Victor Romero-Cano, Ozgur Erkent, Christian Wolf.

7.2.1. Dense & Robust outdoor perception for autonomous vehicles

Participants: Christian Laugier, Victor Romero-Cano, Ozgur Erkent, Christian Wolf.

Robust perception plays a crucial role in the development of autonomous vehicles. While perception in normal and constant environmental conditions has reached a plateau, robustly perceiving changing and challenging environments has become an active research topic, particularly due to the safety concerns raised by the introduction of autonomous vehicles to public streets. Solving the robustness issue in road and urban perception applications is the first challenge. Then, it is also mandatory to develop an appropriate framework for extracting relevant semantic information. Our approach is to reason about vision-based data and the output of our grid-based multi-sensors perception approach (see previous section).

The work presented in this section has partly been done in 2016 and completed in 2017, in the scope of our collaboration with Toyota Motor Europe (TME). The main objective was to develop a framework for low-level multi-sensor data fusion in the pixel space. This framework is independent of the scene coverage of any of the sensors. It outputs a new image-like data representation where each pixel contains not only color but also other low level features such as depth and object IDs.



Figure 11. The XDvision framework.

The system depicted by Figure 11 addresses the problem of extending color images with sparse range data at the pixel level. To this end, we developed a framework for densifying sparse range data in the image space. Our framework provides a methodology for creating extended images independently of the density of the range sensor at hand. We adapted and combined two powerful segmentation techniques such as SLIC and Graph Cuts into a hierarchical methodology for depth densification. The experimental results show the advantages of our new low level data representation over using color only in an image classification task. Our framework achieves, from a camera and sparse/cheap range sensors, recognition results that are equivalent to those obtained from a camera and a dense/expensive 3D Velodyne. This work was patented [75] and published in [22].

Novel approach: Towards Semantic Occupancy Grids and robust *DATMO*⁰**.** Current and future work in the scope of our collaboration with TME, aims at constructing *Semantic Occupancy Grids* and at developing robust *DATMO*. The first objective is to develop occupancy grid mapping systems that exploit the precision of Lidar sensors and the dense and high level semantic information that can be obtained from camera sensors, in particular semantic segmentation systems. We are currently developing technologies that allow extending 2D occupancy grids with semantic and appearance information provided by semantic segmentations systems (**deep learning based**). These technologies allow us to probabilistically fuse occupancy maps and projected 2D segmentation maps both spatially and temporally.

7.2.2. Towards Human-Like Motion Prediction and Decision-Making in Highway Scenarios

Participants: David Sierra-Gonzalez, Christian Laugier, Jilles Dibangoye, Victor Romero-Cano.

The objective of this project is to develop a collision risk estimation system capable of reliably finding the risk of collision associated to the different feasible trajectories of an autonomous or semi-autonomous vehicle. This research work is done in the scope of the Inria-Toyota long-term cooperation and of the PhD thesis work of David Sierra González. The practical output expected for the project is an Advanced Driver Assistance System (ADAS) for the Inria-TME Lexus experimental platform capable of suggesting to the driver in real-time the safest maneuver to perform while driving in the highway.

The research work done in the scope of this project in 2017 has focused on improving the state-of-the-art on short-term prediction of driver behavior in highways. The key idea is to model the behavior of highway drivers from demonstrated highway driving data using Inverse Reinforcement Learning and to exploit the resulting models for prediction. This has led to significant theoretical and experimental results on driver maneuver estimation [25], [26]. The proposed maneuver estimation framework has been validated with real-world highway data collected with an experimental vehicle (see Fig. 12 .a). The results show that the system

⁰Detection and Tracking of Mobile Objects



Figure 12.

a) D. Sierra-Gonzalez PhD. : In this highway traffic scene a driver cuts in front of the ego-vehicle. The red bar above the target shows the estimated marginal probability associated to the lane change maneuver. Lane change maneuvers are in most cases identified before the target touches the lane markings.

b) M. Barbier PhD. : Bottom left shows the most likely direction the driver intend to follow. The top left corner shows the probability over the possible longitudinal intention(stop, yield, pass) for the subject vehicle.

is characterized by short detection times and low false positive rates. This opens the door to the application of this work as part of a collision warning system. In 2017, this work has derived on an additional Inria-TME joint patent [79]. Current efforts are directed towards exploiting the uncertain predictions for decision-making. The PhD thesis of David Sierra-Gonzalez will be defended in June 2018.

7.2.3. Decision-making for safe road intersection crossing

Participants: Mathieu Barbier, Christian Laugier, Olivier Simonin.

In continuation of the work done in 2016 about the functional space representation [39], we proposed a framework to observe the intention of intruder vehicle approaching an intersection. Random forest classifiers are used in conjunction with our functional discretization to analyze the trajectories of cars approaching an intersection. However each intersection can be different and the approach would scale poorly.

To address this problem, a hybrid data set was constructed. It is built in a simulated environment and completed with real data after a car drove multiple times across the intersection. We compared our approach against other classifiers and space discretization. In addition, we demonstrated the impact and the usefulness of the mixture between simulated and real data. An improvement of 30% accuracy is obtained with the hybrid data set, and 5% using our functional discretization with respect to baseline approach. It was also been implemented in our experimental platform as shown in see Fig. 12 .b. This work has been published at IEEE IV 2017 [15]. The PhD thesis of Mathieu Barbier will be defended in June 2018.

The ongoing work use the bayesian network of [56] and extend it to a Partially observable decision process (POMDP). The Observation model uses the performances of the classification described previously to take into account the uncertainty over the intention ⁰. The performances of this decision making system is investigated in the scope of the ENABLES-S3 project. With the use of simulation tools, multiple configurations and scenarios are tested. For each test, a set of "Key performances indicator (KPY)" is computed and used to judge the system. Such validation system has yet to be investigated and represent a major interest for both industry and research.

⁰This work will be the subject of a paper to be submitted to IEEE IV 2018

7.3. Robust state estimation (Sensor fusion)

This research is the follow up of Agostino Martinelli's investigations carried out during the last four years, which are in the framework of the visual and inertial sensor fusion problem and the unknown input observability problem.

7.3.1. Visual-inertial structure from motion

Participants: Agostino Martinelli, Alexander Oliva, Alessandro Renzaglia.

During this year we achieved the following two objectives:

- 1. (Theoretical) Extension of the closed form solution introduced in [64] to the cooperative case.
- 2. (Experimental) Improvement of one order of magnitude on the precision on the absolute scale determined by our closed form solution introduced in [64];

Regarding the first objective, we obtained a new theoretical and basic result in the framework of cooperative visual-inertial sensor fusion. Specifically, the case of two agents was investigated. First, the entire observable state was analytically derived. This state includes the relative position between the two aerial vehicles (which includes the absolute scale), the relative velocity and the three Euler angles that express the rotation between the two vehicle frames. Then, the basic equations that describe this system were analytically obtained. These results have been presented at the first international symposium on multi robot and multi agent systems [69]. Finally, we extended the closed form solution introduced in [64] to the cooperative case. Specifically, the observable state was expressed in closed form in terms of the measurements provided by monocular vision and inertial sensors, during a short time interval. We believe that this is a fundamental theoretical result, since it allows us from one side to automatically retrieve the absolute scale in closed form (and consequently without prior knowledge) even without observing external point features in the environment and, on the other side, to carry out a theoretical investigation that will allow us to detect all the system singularities. Extensive simulations and real experiments clearly show that the proposed solution is successful.

Regarding the second objective, our former experimental implementation provided a precision on the absolute scale in the range 10%-20% (all the details about the experimental setup are available in [55]). These former results were obtained in collaboration with the Robotics and Perception Group at the university of Zurich, in the framework of the ANR-VIMAD project. Specifically, the experiments were carried out in Zurich. This year, by an extensive use of the platform KINOVIS available at Inria, we investigated the impact of several sources of systematic error (imperfect extrinsic camera calibration, time delay and imperfect time alignment between sensors, etc.). We used simple methods to remove these error sources and we achieved the precision on the absolute scale in the range 1%-5%.

7.3.2. Unknown Input Observability

Participant: Agostino Martinelli.

During this year I achieved the following two objectives:

- 1. (Theoretical) Extension of the analytic solution presented in [66] to the driftless multiple unknown inputs case.
- 2. (Theoretical) Application of the solution in [66] to several problems, in the framework of computer vision, neuroscience and robotics.

Regarding the former objective, I obtained the general analytic solution of the nonlinear unknown input observability problem. As for the observability rank condition, the analytic criterion in presence of unknown inputs is based on the computation of the observable codistribution by a recursive and convergent algorithm. The algorithm is unexpectedly simple and can be easily and automatically applied to nonlinear systems driven by both known and unknown inputs, independently of their complexity and type of nonlinearity. Very surprisingly, the complexity of the overall analytic criterion is comparable to the complexity of the standard method to check the state observability in the case without unknown inputs (i.e., the observability rank condition). Given any nonlinear system characterized by any type of nonlinearity, driven by both known and unknown inputs, is obtained automatically, i.e., by following a systematic procedure

(e.g., by the usage of a very simple code that uses symbolic computation). This is a fundamental practical (and unexpected) advantage. On the other hand, the analytic derivations and all the proofs necessary to analytically derive the algorithm and its convergence properties and to prove their general validity are very complex and they are extensively based on an ingenious analogy with the theory of General Relativity. In practice, these derivations largely use Ricci calculus with tensors (in particular, I largely adopt the Einstein notation to achieve notational brevity). All the results are fully described in a book, which is supposed to be published during the next year. A first draft of the book is now available on ArXiv (arXiv:1704.03252).

Regarding the second objective, the solution that holds in the driftless case and in presence of a single unknown input ([66]) has been used to investigate several problems. These problems include:

- 1. The unicycle in presence of a single disturbance (which has been presented at the SIAM on Control and Application 2017, [68]).
- 2. Vehicle moving in 3D in presence of a disturbance (which has been presented at the IROS 2017, [67])

Finally, the visual and inertial sensor fusion problem, when some of the inputs are unknown, has been investigated both in 2D and in 3D. All the results are described in chapter 5 of the book available on ArXiv (arXiv:1704.03252).

7.4. Motion-planning in human-populated environment

We explore motion planning algorithms to allow robots/vehicles to navigate in human populated environment, and to predict human motions. Since 2016, our work focuses on two directions, which are prediction of pedestrian behaviors in urban environments and mapping of human flows. We also started to investigate the navigation of a telepresence robot in collaboration with the GIPSA Lab. These works are presented here after.





Figure 13. Illustration of (a) enclosure of areas predicted by the framework where pedestrian movements could occur. (b) Growing HMM for observed pedestrian trajectories at a randomly chosen time in the dataset

7.4.1. Urban Behavioral Modeling

Participants: Pavan Vasishta, Anne Spalanzani, Dominique Vaufreydaz.

The objective of modeling urban behavior is to predict the trajectories of pedestrians in towns and around car or platoons (PhD work of P. Vasishta). In 2017 we proposed to model pedestrian behaviour in urban scenes by combining the principles of urban planning and the sociological concept of Natural Vision. This model assumes that the environment perceived by pedestrians is composed of multiple potential fields that influence their behaviour. These fields are derived from static scene elements like side-walks, cross-walks, buildings, shops entrances and dynamic obstacles like cars and buses for instance. This work was published in [30], [27]. Next year will be dedicated to combine this model with GHMM (Growing HMM) [86] to infer probable pedestrian paths in the scene to predict, for example, legal and illegal crossings, see. Fig. 13.

7.4.2. Learning task-based motion planning

Participants: Christian Wolf, Jilles Dibangoye, Laetitia Matignon, Olivier Simonin.

Our goal is the automatic learning of robot navigation in human populated environments based on specific tasks and from visual input. The robot automatically navigates in the environment in order to solve a specific problem, which can be posed explicitly and be encoded in the algorithm (e.g. recognize the current activities of all the actors in this environment) or which can be given in an encoded form as additional input. Addressing these problems requires competences in computer vision, machine learning, and robotics (navigation and paths planning).

We started this work in the end of 2017, following the arrival of C. Wolf, through combinations of reinforcement learning and deep learning. The underlying scientific challenge here is to automatic learn representations which allow the agent to solve multiple sub problems require for the task. In particular, the robot needs to learn a metric representation (a map) of its environment based from a sequence of ego-centric observations. Secondly, to solve the problem, it needs to create a representation which encodes the history of ego-centric observations which are relevant to the recognition problem. Both representations need to be connected, in order for the robot to learn to navigate to solve the problem. Learning these representations from limited information is a challenging goal.



Figure 14. (a) Flow-grid mapping in a cross-corridor where 200 moving pedestrians turns (b) A* path-planning computed with different cost functions (c) implementation and results with a real robot and environment

7.4.3. Modeling human-flows from robot(s) perception

Participants: Jacques Saraydaryan, Fabrice Jumel, Olivier Simonin.

In order to deal with robot navigation in dense human populated environments, eg. in flows of humans, we investigate the problem of mapping these flows. The challenge is to build such a map from robot perceptions while robots move autonomously to perform some tasks in the environment. We developed statistical learning

approach (ie. a counting-based grid model) which computes in each cell the likelihoods of crossing a human in each possible direction, see red vectors in Fig. 14 .a We extended the flow grid model with a human motion predictive model based on the Von Misses motion pattern, allowing to "accelerate" the flow grid mapping, see blue vectors in Fig. 14 .a.

Then we examined how path-planning can benefit of such a flow grid, that is taking into account the risk for a robot to encounter humans in opposite direction. We first implemented the Flow-Grid model in a simulator built upon PedSim and ROS tools, allowing to simulate mobile robots and crowd of pedestrians. We compared three A*-based path-planning models using different levels of information about human presence: non-informed, a grid of human presence likelihood proposed by Tipaldi [83] and our grid of human motion likelihood (see 14 .b). Experiments in simulations and with real robots allowed to show the efficiency of the flow-grid to build efficient paths through human flows (see 14 .c). These results have been published in ECRM [16].

This work will allow us to develop new solutions to the patrolling of moving people, that we called the waiters problem two years ago (see our article in RIA revue, 2017 [11]). Indeed, if robots build a flow grid of people they cross and have to serve, they will be able to optimize along the time their strategy to deploy and revisit people regularly.

7.4.4. Navigation of telepresence robots

Participants: Rémi Cambuzat, Olivier Simonin, Anne Spalanzani, Gerard Bailly [GIPSA, CNRS, Grenoble], Frederic Elisei [GIPSA, CNRS, Grenoble].

In 2016 we obtained with the team of Gérard Bailly, from GIPSA/CNRS Grenoble, a regional support for the TENSIVE project. It funds the PhD thesis of Remi Cambuzat on immersive teleoperation of telepresence robots for verbal interaction and social navigation, started in October 2016. We have 2 mains objectives: i) to design a new generation of immersive control platforms for telepresence robots and ii) to teach multimodal behaviors to social robots by demonstration. In both cases, a human pilot interacts with remote interlocutors via the mediation of a robotic embodiment that should faithfully reproduce the body movements of the pilots while providing rich sensory and proprioceptive feedback. During social interactions, people's eyes convey a wealth of information about their direction of attention and their emotional and mental states. Endowing telepresence robots with the ability to mimic the pilot's gaze direction as well as autonomous social robots with the possibility to generate gaze cues is necessary for enabling them to seamlessly interact with humans. During the first year of the PhD thesis, we focused on the immersive teleoperation of the Nina Robot Gaze 15 . Figure 15 abstracts the proposed methodology.



Figure 15. immersive teleoperation for social interaction

7.5. Decision Making in Multi-Robot Systems

7.5.1. Multi-robot planning in dynamic environments

7.5.1.1. Multi-Robot Routing (MRR) for evolving missions

Participants: Mihai Popescu, Olivier Simonin, Anne Spalanzani, Fabrice Valois [Inria, Agora team].

After considering Multi-Robot Patrolling of known targets in 2016 [73], we generalized our work to Dynamic Multi Robot-Routing (DMRR), an instance of continuously adapting the multi-robot target allocation process (MRTA). Target allocation problems have been frequently treated in contexts such as multi-robot rescue operations, exploration, or patrolling, being often formalized as multi-robot routing problems. There are few works addressing dynamic target allocation, such as allocation of previously unknown targets. However, existent solutions do not regard the continuous adaptation of the ongoing robot missions to new targets. These techniques are neither adapted to handle the missions growth in time (nor a possible saturation bound for the mission cost). We proposed a framework for dynamically adapting the existent robot missions to new discovered targets. Dynamic saturation-based auctioning (DSAT) is proposed for adapting the execution of robots to the new targets. Comparison was made with algorithms ranging from greedy to auction-based methods with provable sub-optimality. We tested the algorithms on exhaustive sets of inputs. The results for DSAT shows it outperforms state-of-the-art methods, like standard SSI or SSI with regret clearing, especially in optimizing the target allocation w.r.t. the target coverage in time and the robot resource usage (e.g. minimizing the worst mission cost). Results have been submitted to ICAPS 2018.

This work is developped in the PhD. work of M. Popescu, but also through the collaboration (PHC ⁰ 'DRONEM' project) started in 2017 with the team of Gabriela Czibula from Babes-Bolyai University in Cluj-Napoca (Romania). The project focuses on optimization and online adaptation of the multi-cycle patrolling with machine learning (RL) techniques in order to deal with the arrival of new targets in the environment.

7.5.1.2. Global-local optimization in autonomous multi-vehicles systems

Participants: Guillaume Bono, Jilles Dibangoye, Laetitia Matignon, Olivier Simonin, Florian Peyreron [VOLVO Group, Lyon].

This work is part of the PhD. thesis in progress of Guillaume Bono, with the VOLVO Group, in the context of the INSA-VOLVO Chair. The goal of this project is to plan and learn at both global and local levels how to act when facing a vehicle routing problem (VRP). We started with a state-of-the-art paper on vehicle routing problems as it currently stands in the literature [28]. We were surprise to notice that few attention has been devoted to deep reinforcement learning approaches to solving VRP instances. Hence, we investigated our own deep reinforcement learning approaches to solving VRP instances. Hence, we investigated our own deep reinforcement learning approach that can help one vehicle to learn how to generalize strategies from solved instances of tralleving salesman problems (an instance of VRPs) to unsolved ones. The difficulty of this problem lies in the fact that its Markov decision process' formulation is intractable, i.e., the number of states grows doubly exponentially with the number of cities to be visited by the salesman. To gain in scalability, we build inspiration on a recent work by DeepMind, which suggests using pointernet, i.e., a novel deep neural network architecture, to address learning problems in which entries are sequences (here cities to be visited) and output are also sequences (here order in which cities should be visited). Preliminary results are encouraging, and we plan to extend this work in the multi-agent setting during the coming year.

7.5.2. Multi-robot coverage and mapping

7.5.2.1. Human scenes observation

Participants: Laetitia Matignon, Olivier Simonin, Stephane d'Alu, Christian Wolf.

Solving complex tasks with a fleet of robots requires to develop generic strategies that can decide in real time (or time-bounded) efficient and cooperative actions. This is particularly challenging in complex real environments. To this end, we explore anytime algorithms and adaptive/learning techniques.

⁰Hubert Curien Partnership



Figure 16. (a) Concentric navigation model. (b) Experimental setup and multi-robot mapping with Turtlebot 2.

The "CROME" and "COMODYS" ⁰ projects ⁰ are motivated by the exploration of the joint-observation of complex (dynamic) scenes by a fleet of mobile robots. In our current work, the considered scenes are defined as a sequence of activities, performed by a person in a same place. Then, mobile robots have to cooperate to find a spatial configuration around the scene that maximizes the joint observation of the human pose skeleton. It is assumed that the robots can communicate but have no map of the environment and no external localisation.

To attack the problem, we proposed an original concentric navigation model allowing to keep easily each robot camera towards the scene (see fig. 16 .a). This model is combined with an incremental mapping of the environment and exploration guided by meta-heuristics in order to limit the complexity of the exploration state space. Results have been submitted to AAMAS'2018⁰.

In 2017, we also proposed an hybrid metric-topological mapping for multi-robot observation of a human scene. Robots are individually building a map that is updated cooperatively by exchanging only high-level data between robots, thereby reducing the communication payload. We combined an on-line distributed multi-robot decision with this hybrid mapping. These modules have been implemented and evaluated on our platform composed of several Turtlebots2, see fig. 16 .b. Results have been published in 2017 in [21] (ECMR).

7.5.2.2. Multi-UAV Visual Coverage of Partially Known 3D Surfaces

Participants: Alessandro Renzaglia, Jilles Dibangoye, Olivier Simonin.

It has been largely proved that the use of Unmanned Aerial Vehicles (UAVs) is an efficient and safe way to deploy visual sensor networks in complex environments. In this context, a widely studied problem is the cooperative coverage of a given environment. In a typical scenario, a team of UAVs is called to achieve the mission without a perfect knowledge on the environment and needs to generate the trajectories on-line, based only on the information acquired during the mission through noisy measurements. For this reason, guaranteeing a global optimal solution of the problem is usually impossible. Furthermore, the presence of several constraints on the motion (collision avoidance, dynamics, etc.) as well as from limited energy and computational capabilities, makes this problem particularly challenging.

Depending on the sensing capabilities of the team (number of UAVs, range of on-board sensor, etc.) and the dimension of the environment to cover, different formulations of this problem can be considered. We firstly approached the deployment problem, where the goal is to find the optimal static UAVs configuration from which the visibility of a given region is maximized. A suitable way to tackle this problem is to adopt derivative-free optimization methods based on numerical approximations of the objective function. In 2012, Renzaglia et al. [74] proposed an approach based on a stochastic optimization algorithm to obtain a solution

⁰COoperative Multi-robot Observation of DYnamic human poSes

⁰Funded by a LIRIS transversal project in 2016-2017 and a FIL project in 2017-2019 (led by L. Matignon)

⁰Multi-Robot Simultaneous Coverage and Mapping of Complex Scene - Comparison of Different Strategies

for arbitrary, initially unknown 3D terrains. However, adopting this kind of approaches, the final configuration can be strongly dependent on the initial positions and the system can get stuck in local optima very far from the global solution. We identified that a way to overcome this problem can be found in initializing the optimization with a suitable starting configuration. An a priori partial knowledge on the environment is a fundamental source of information to exploit to this end. The main contribution of our work is thus to add another layer to the optimization scheme in order to exploit this information. This step, based on the concept of Centroidal Voronoi Tessellation, will then play the role of initialization for the on-line, measurement-based local optimizer. The resulting method, taking advantages of the complementary properties of geometric and stochastic optimization, significantly improves the result of the previous approach and notably reduces the probability of a far-to-optimal final configuration. Moreover, the number of iterations necessary for the convergence of the on-line algorithm is also reduced. This work led to a paper submitted to ICRA 2018⁰, currently under review. The development of a realistic simulation environment based on Gazebo is an important on-going activity in Chroma and will allow us to further test the approach and to prepare the implementation of this algorithm on the real robotic platform available in the team.

We are currently also investigating the dynamic version of this problem, where the information is collected along the trajectories and the environment reconstruction is obtained from the fusion of the total visual data.

7.5.2.3. Middleware for open multi-robot systems

Participants: Stefan Chitic, Julien Ponge [CITI, Dynamid], Olivier Simonin.

Multi-robots systems (MRS) require dedicated software tools and models to face the complexity of their design and deployment. In the context of the PhD work of Stefan Chitic, we address service self-discovery and property proofs in an ad-hoc network formed by a fleet of robots. This led us to propose a robotic middleware, SDfR, that is able to provide service discovery, see [44]. In 2017, we defined a tool-chain based on timed automata, called ROSMDB, that offers a framework to formalize and implement multi-robot behaviors and to check some (temporal) properties (both offline and online). S. Chtic will defend his Phd thesis on March 2018.

7.5.3. Sequential decision-making under uncertainty

This research is the follow up of team led by Jilles S. Dibangoye carried out during the last three years, which include foundations of sequential decision making by a group of cooperative or competitive robots or more generally agents.

7.5.3.1. Optimally solving cooperative and competitive games as continuous Markov decision processes

Participants: Jilles S. Dibangoye, Olivier Buffet [Inria Nancy], Vincent Thomas [Inria Nancy], Christopher Amato [Univ. New Hampshire], François Charpillet [Inria Nancy, Larsen team].

Our major findings this year include:

- 1. (Theoretical) As an extension of [47] in the cooperative case, we characterize the optimal solution of partially observable stochastic games.
- 2. (Theoretical) We further exhibit new underlying structures of the optimal solution for both cooperative and non-cooperative settings.
- 3. (Algorithmic) We extend a non-trivial procedure for computing such optimal solutions when only an incomplete knowledge about the model is available.

This work proposes a novel theory and algorithms to optimally solving a two-person zero-sum POSGs (zs-POSGs). That is, a general framework for modeling and solving two-person zero-sum games (zs-Games) with imperfect information. Our theory builds upon a proof that the original problem is reducible to a zs-Game—but now with perfect information. In this form, we show that the dynamic programming theory applies. In particular, we extended Bellman equations [40] for zs-POSGs, and coined them maximin (resp. minimax) equations. Even more importantly, we demonstrated Von Neumann & Morgenstern's minimax theorem [87] [88] holds in zs-POSGs. We further proved that value functions—solutions of maximin (resp. minimax)

⁰A. Renzaglia, J. Dibangoye and O. Simonin, "Multi-UAV Visual Coverage of Partially Known 3D Surfaces: Voronoi-based Initialization for Stochastic Optimization", IEEE International Conference on Robotics and Automation (ICRA), 2018, *under review*.

equations—yield special structures. More specifically, the maximin value functions are convex whereas the minimax value functions are concave. Even more surprisingly, we prove that for a fixed strategy, the optimal value function is linear. Together these findings allow us to extend planning and learning techniques from simpler settings to zs-POSGs. To cope with high-dimensional settings, we also investigated low-dimensional (possibly non-convex) representations of the approximations of the optimal value function. In that direction, we extended algorithms that apply for convex value functions to lipschitz value functions [43].

7.5.3.2. Learning to act in continuous decentralized partially observable Markov decision process

Participants: Jilles S. Dibangoye, Olivier Buffet [Inria Nancy], Laëtitia Matignon, Christian Wolf, Guillaume Bono, Jacques Saradaryan, Olivier Simonin, Florian Peyreron.

During the last year, we investigated deep and standard reinforcement learning for solving decentralized partially observable Markov decision processes. Our preliminary results include:

- 1. (Theoretical) Proofs that the optimal value function is linear in the occupancy-state space, the set of all possible distributions over hidden states and histories.
- 2. (Algorithmic) Value-based and policy-based (deep) reinforcement learning for common-payoff partially observable stochastic games.

This work addresses a long-standing open problem of Multi-Agent Reinforcement Learning (MARL) in decentralized stochastic control. MARL previously applied to finite decentralized decision making with a focus on team reinforcement learning methods, which at best lead to local optima. In this research, we build on our recent approach [47], which converts the original problem into a continuous-state Markov decision process, allowing knowledge transfer from one setting to the other. In particular, we introduce the first optimal reinforcement learning method for finite cooperative, decentralized stochastic control domains. We achieve significant scalability gains by allowing the latter to feed deep neural networks. Experiments show our approach can learn to act optimally in many finite decentralized stochastic control problems from the literature.

CONVECS Project-Team

6. New Results

6.1. New Formal Languages and their Implementations

LNT is a next generation formal description language for asynchronous concurrent systems, which attempts to combine the best features of imperative programming languages and value-passing process algebras. LNT is increasingly used by CONVECS for industrial case studies and applications (see § 6.5) and serves also in university courses on concurrency, in particular at ENSIMAG (Grenoble) and at Saarland University.

6.1.1. Translation from LNT to LOTOS

Participants: Hubert Garavel, Frédéric Lang, Wendelin Serwe.

The move towards "safer" LNT exceptions initiated in 2016 has been completed in 2017: the two concepts of gates and exceptions have been unified in both LNT processes and LNT functions. The static semantics of LNT no longer requires that variables and exceptions share the same name space.

LNT now permits simple loops (of the form "loop ... end loop", without loop label nor "while" condition) in LNT functions, as well as in LNT processes.

The pragma names "comparedby", "external", "implementedby", "iteratedby", "printedby", and "representedby" are no longer reserved LNT keywords, meaning that it is now permitted to declare LNT identifiers having these names. Two new type pragmas "!card" and "!bits" have been added to specify the maximum number of values and the number of bits to be used when storing the values of a given type in "hash-consing" tables.

The LPP preprocessor and the LNT2LOTOS translator, which implement the LNT language, have been enhanced in many ways. In addition to 9 bug fixes, the following enhancements have been made:

- LPP now implements LNT character strings more concisely.
- LPP automatically adds the ".lnt" extension to input and output files if this extension is missing.
- The algorithm that computes which LNT gates are used in each function or process has been made more precise, and LNT2LOTOS now warns about gates that are declared but never used.
- LNT2LOTOS performs stricter compile-time checks that produce dedicated error messages, rather than generating invalid LOTOS code that was subsequently rejected by CAESAR and/or CAE-SAR.ADT. Also, several error messages displayed by LNT2LOTOS during its static-analysis phases have been enhanced.
- The translation from LNT functions to LOTOS operations has been significantly improved by eliminating unreachable or redundant LOTOS equations, removing unused auxiliary LOTOS operations, simplifying the premises of certain LOTOS equations, factorizing identical assignments in "if-thenelse" instructions, and optimizing long sequences of assignments intertwined with assertions. Thus, LNT2LOTOS is now faster, uses less memory, generates more compact LOTOS code, and can compile larger LNT specifications that could not be handled before.

The LNT2LOTOS Reference Manual, which contains the definition of the LNT language, has been revised, enriched, and simplified in many ways. A paper presenting the historical background and motivation behind the definition of LNT was published in an international conference [18].

6.1.2. NUPN

Participant: Hubert Garavel.

Nested-Unit Petri Nets (NUPNs) is an upward-compatible extension of P/T nets, which are enriched with structural information on their concurrent structure. Such additional information can easily be produced when NUPNs are generated from higher-level specifications (e.g., process calculi); quite often, such information allows logarithmic reductions in the number of bits required to represent states, thus enabling verification tools to perform better. The principles of NUPNs are exposed in [29] and its PNML representation is described here ⁰.

In 2017, we studied an abstraction called *place fusion*, which takes advantage of the compositional, hierarchical structure of NUPNs. We formulated key theorems stating which properties are preserved or not under this abstraction. On the practical side, our collection of NUPN models grew to more than 8 000 benchmarks. Statistical studies were done on this collection to estimate the compression factor permitted by the NUPN model. A journal article providing an overview of NUPNs was written.

The NUPN model has been adopted by the Model Checking Contest and implemented in ten different tools developed in four countries. In 2017, the NUPN model was also adopted for the parallel problems of the RERS'2017 (*Rigorous Examination of Reactive Systems*) challenge ⁰.

6.1.3. Analysis of BPMN via Translation to LNT

Participants: Ajay Muroor Nadumane, Gwen Salaün.

Business process modeling is an important concern in companies and organizations. Formal analysis techniques are crucial to detect semantic issues in the corresponding models, or to help with their refactoring and evolution. However, business process development frameworks often fall short when it comes to go beyond simulation or syntactic checking of the models. To ensure a more robust development of business processes, we developed the VBPMN verification framework. It features several techniques for the automatic analysis of business processes modeled using BPMN, the de facto standard for business process modeling.

The business processes, described using a Web application compliant with BPMN 2.0, are transformed into an intermediate format called PIF (*Process Intermediate Format*). Then, from the PIF descriptions, models in LNT and model-specific verification scripts in SVL are generated. In the end, CADP is used to check either for functional properties of a given business process, or for the correctness of the evolution of a business process into another one. This latter kind of verification supported by VBPMN is particularly helpful in order to improve a process w.r.t. certain optimization criteria. A paper presenting these results was published in an international conference [16].

6.1.4. Translation of Term Rewrite Systems

Participants: Hubert Garavel, Lina Marsso.

We pursued the development undertaken in 2015 of a software platform for systematically comparing the performance of rewrite engines and pattern-matching implementations in algebraic specification and functional programming languages. Our platform reuses the benchmarks of the three Rewrite Engine Competitions (2006, 2009, and 2010). Such benchmarks are term-rewrite systems expressed in a simple formalism named REC, for which we developed automated translators that convert REC benchmarks into many languages, among which AProVE, Clean, Haskell, LNT, LOTOS, Maude, mCRL, MLTON, OCAML, Opal, Rascal, Scala, SML-NJ, Stratego/XT, and Tom.

In 2017, we revised and enhanced the largest REC benchmark, the MAA (*Message Authenticator Algorithm*), a Message Authentication Code used for financial transactions (ISO 8731-2) between 1987 and 2002. This model (13 sorts, 18 constructors, 644 non-constructors, and 684 rewrite rules) was proven to be confluent, and terminating. Implementations in thirteen different languages have been automatically derived from this model and used to validate 200 official test vectors for the MAA. These results led to a publication in an international conference [14].

⁰http://mcc.lip6.fr/nupn.php

⁰http://www.rers-challenge.org

We also corrected and/or enhanced several of the existing REC translators (e.g., Clean) and added support of CafeOBJ and compiled OCAML. A scientific paper on this study has been prepared.

6.1.5. Other Language Developments

Participants: Hubert Garavel, Frédéric Lang, Radu Mateescu, Wendelin Serwe.

The ability to compile and verify formal specifications with complex, user-defined operations and data structures is a key feature of the CADP toolbox since its very origins.

In 2017, we brought various enhancements to several compilers handling formal specification languages (LOTOS, MCL, XTL, and GRL):

- A buffer overflow and two out-of-bound array accesses have been corrected in both CAESAR and CAESAR.ADT. Two memory allocation bugs have also been corrected in CAESAR.ADT. The latter tool now generates C code that gives better diagnostic when the evaluation of a constant fails at run time (e.g., when it triggers an exception signal, or exhausts the stack or heap memory).
- In addition to two bug fixes, the warning and error messages displayed by MCL_EXPAND and XTL_EXPAND have been made more precise and stringent. The XTL libraries "walk" and "walk_nice" have been modified not to trigger the extra warnings recently introduced.
- The GRL2LNT translator takes as input a formal description in GRL of a GALS system and generates an equivalent LNT specification. A new version 1.1 of GRL2LNT has been released, which corrects a bug concerning the LNT code generated by the "-merge" option.

H. Garavel pursued the study of the most suitable axiomatization of signed integers undertaken in 2016. He reviewed a tenth of such Peano-like axiom systems, which he classified and evaluated according to complexity and efficiency criteria. These results have been published in an international conference [13].

6.2. Parallel and Distributed Verification

6.2.1. Distributed State Space Manipulation

Participants: Hubert Garavel, Wendelin Serwe.

For distributed verification, CADP provides the PBG format, which implements the theoretical concept of *Partitioned LTS* [34] and provides a unified access to an LTS distributed over a set of remote machines.

In 2017, many changes have been done to simplify the code of the CAESAR_NETWORK_1 communication library, which is the backbone of the distributed verification tools of CADP, as well as the code of other tools such as BCG_MIN, but most of these changes are not directly observable by end users. In addition to two bug fixes in CAESAR_NETWORK_1 and two other bug fixes in the BES_SOLVE tool, the error messages displayed by the various tools and the statistical information produced by the "-stat" option of BES_SOLVE have been made more concise and more informative.

6.2.2. Debugging of Concurrent Systems

Participants: Gianluca Barbon, Gwen Salaün.

Model checking is an established technique for automatically verifying that a model satisfies a given temporal property. When the model violates the property, the model checker returns a counterexample, which is a sequence of actions leading to a state where the property is not satisfied. Understanding this counterexample for debugging the specification is a complicated task for several reasons: (i) the counterexample can contain hundreds of actions, (ii) the debugging task is mostly achieved manually, and (iii) the counterexample does not explicitly highlight the source of the bug that is hidden in the model.

We proposed an approach that improves the usability of model checking by simplifying the comprehension of counterexamples. Our solution aims at keeping only actions in counterexamples that are relevant for debugging purposes. To do so, we first extract in the model all the counterexamples. Second, we define an analysis algorithm that identifies actions that make the model skip from incorrect to correct behaviours, making these actions relevant from a debugging perspective. Our approach is fully automated by a tool we implemented and applied on real-world case studies from various application areas for evaluation purposes. This work led to a publication in an international conference [11].

In 2017, we focused on extending our approach following three directions: (a) we introduced new notions to identify new types of relevant actions; (b) we developed a set of heuristics to extract these actions from counterexamples; (c) we proposed an alternative approach to focus on a broader range of properties (i.e., liveness properties). These new extensions have been integrated into our tool. A paper was submitted to an international journal.

6.3. Timed, Probabilistic, and Stochastic Extensions

6.3.1. Tools for Probabilistic and Stochastic Systems

Participants: Hubert Garavel, Jean-Philippe Gros, Frédéric Lang, Julie Parreaux, Wendelin Serwe.

Formal models and tools dealing with quantititative aspects (such as time, probabilities, and other continuous physical quantities) have become unavoidable for a proper study and computer-aided verification of functional and non-functional properties of cyberphysical systems. The wealth of such formal models is sometimes referred to as a quantitative "zoo" [39].

The CADP toolbox already implements some of these probabilistic/stochastic models, namely DTMCs and CTMCs (*Discrete-Time* and *Continuous-Time Markov Chains*), and IMCs (*Interactive Markov Chains*) [41]. Our long-term goal is to increase the capability and flexibility of the CADP tools, so as to support other quantitative models more easily.

In 2017, we undertook a systematic review of the existing theoretical models and built a comprehensive list of more than 70 software tools implementing these models [37]. The results of this study have been made widely available as a Web catalog 0 .

In parallel, we also undertook a systematic review [50] of the benchmarks made available for these tools. We downloaded more than 21 000 files from the web and developed triage scripts to analyze these files and classify them automatically, separating various kinds of automata-based models (e.g., Markov chains, Markov automata, hybrid automata, etc.) from temporal-logic formulas. One finding of this "big data" study is the present lack of diversity, as four tools (PRISM, MRMC, STORM, and SiSAT) provide nearly 60% of the models.

To address this issue, we started investigating the probabilistic and stochastic models of complex industrial systems produced by former PhD students of the VASY and CONVECS teams. We analyzed these models (written in BCG, EXP, LOTOS, LNT, SVL, and/or Makefiles) to separate functional aspects from performance ones, leading to a collection of DTMCs, CTMCs, IMCs, and IPCs (*Interactive Probabilistic Chains*). We updated these models to ensure compatibility with the latest versions of CADP and C compilers, and we started enhancing EXP.OPEN with new features that simplify the parallel composition of IPCs (see § 6.4.1).

6.3.2. On-the-fly Model Checking for Extended Regular Probabilistic Operators

Participant: Radu Mateescu.

⁰http://cadp.inria.fr/resources/zoo

Specifying and verifying quantitative properties of concurrent systems requires expressive and user-friendly property languages combining temporal, data-handling, and quantitative aspects. In collaboration with José Ignacio Requeno (Univ. Zaragoza, Spain), we undertook the quantitative analysis of concurrent systems modeled as PTSs (*Probabilistic Transition Systems*), whose actions contain data values and probabilities. We proposed a new regular probabilistic operator that extends naturally the Until operators of PCTL (*Probabilistic Computation Tree Logic*) [38], by specifying the probability measure of a path characterized by a generalized regular formula involving arbitrary computations on data values. We integrated the regular probabilistic operator into MCL, we devised an associated on-the-fly model checking method based on a combined local resolution of linear and Boolean equation systems, and we implemented the method in a prototype extension of the EVALUATOR model checker.

In 2017, we continued experimenting the extended model checker on further examples of protocols (Bounded Retransmission Protocol, randomized philosophers, self-stabilization) and observed that it exhibits a performance comparable with the explicit-state algorithms of the PRISM model checker ⁰. A paper was submitted to an international journal.

6.4. Component-Based Architectures for On-the-Fly Verification

6.4.1. Compositional Verification

Participants: Hubert Garavel, Frédéric Lang.

The CADP toolbox contains various tools dedicated to compositional verification, among which EXP.OPEN, BCG_MIN, BCG_CMP, and SVL play a central role. EXP.OPEN explores on the fly the graph corresponding to a network of communicating automata (represented as a set of BCG files). BCG_MIN and BCG_CMP respectively minimize and compare behavior graphs modulo strong or branching bisimulation and their stochastic extensions. SVL (*Script Verification Language*) is both a high-level language for expressing complex verification scenarios and a compiler dedicated to this language.

In 2017, two bugs have been solved in SVL and one bug has been solved in EXP.OPEN. Several improvements have been brought to both tools. In particular:

• EXP.OPEN now has two new options "-prob" and "-rate" for handling probabilistic and stochastic transitions, respectively; without these options, probabilistic and stochastic transitions are considered as ordinary transitions (this enables EXP.OPEN to be used for implementing alternative semantics, such as *Interactive Probabilistic Chains* [27] where probabilistic transitions are synchronized using a global clock). Consequently, the former "-ratebranching" option has been replaced by "- rate -branching".

Also, error messages about synchronization vectors have been made more precise and EXP.OPEN performs tighter checks about labels containing only blanks and unexpected synchronization of probabilistic or stochastic transitions. Two bugs have been fixed in EXP.OPEN and style files have been added to bring support for the EXP format by mainstream text editors.

• A new option "-v" has been added to set SVL variables from the command line (similar to "awk" or "make"). Debugging SVL scripts has been made easier: the "-debug" option of SVL now stops the execution as soon as a shell command (e.g., a CADP tool or a Unix command) terminates with a non-zero exit status, so that problems are detected as soon as they occur.

Also, SVL now performs tighter semantic checks, making sure that all partial-order reduction options passed to EXP.OPEN (namely, options explicitly set by the user and options automatically computed by SVL from the context of the EXP composition expression) are not contradictory.

6.4.2. On-the-Fly Test Generation

Participants: Hubert Garavel, Lina Marsso, Radu Mateescu, Wendelin Serwe.

⁰http://www.prismmodelchecker.org/

The CADP toolbox provides support for conformance test case generation by means of the TGV tool. Given a formal specification of a system and a test purpose described as an input-output LTS (IOLTS), TGV automatically generates test cases, which assess using black box testing techniques the conformance of a system under test w.r.t. the formal specification. A test purpose describes the goal states to be reached by the test and enables one to indicate parts of the specification that should be ignored during the testing process. TGV does not generate test cases completely on the fly (i.e., *online*), because it first generates the complete test graph (CTG) and then traverses it backwards to produce controllable test cases.

In 2017, we carried out the following activities:

- We developed the prototype tool TESTOR to extract test cases completely on the fly. Compared to TGV, the new tool TESTOR presents several advantages: (i) it has a more modular architecture, based on generic graph transformation components taken from the OPEN/CAESAR libraries (τ-compression, τ-confluence, τ-closure, determinization, resolution of Boolean equation systems); (ii) it is capable of extracting a test case completely on the fly, by exploiting the diagnostic generation features of the Boolean equation system resolution algorithms; (iii) it enables a more flexible expression of test purposes, taking advantage of the multiway rendezvous, a primitive to express communication and synchronization among a set of distributed processes [15]. We evaluated TESTOR on three published case studies and more than 10 000 examples taken from the non-regression test suites of CADP. A paper describing this work was accepted for publication in an international conference.
- We also revised TGV, which is now by default much less verbose and only displays the most important information, but the former behaviour can still be retained using option "-verbose". A new option "-monitor" allows to follow in real time how the test case generation progresses. Many warning and error messages have been enhanced, various bugs (especially buffer overflows) have been fixed, and memory allocation results are now strictly controlled.

6.4.3. Other Component Developments

Participants: Lian Apostol, Soren Enevoldsen, Hubert Garavel, Frédéric Lang, Radu Mateescu, Wendelin Serwe.

The CAESAR_STANDARD library was enriched with the new CAESAR_TYPE_FORMAT type and its associated primitives, and with two new functions CAESAR_SET_SIGNALS() and CAE-SAR_RESET_SIGNALS() for handling POSIX signals (including SIGSEGV, i.e., segmentation violation). The CAESAR_GRAPH interface, which remained stable for two decades, has been modified: its two functions CAESAR_FORMAT_STATE() and CAESAR_FORMAT_LABEL() became more powerful, while its two functions CAESAR_MAX_FORMAT_STATE() and CAESAR_MAX_FORMAT_LABEL() have been removed from the interface. The same changes apply as well to all the other similar functions of the OPEN/CAESAR libraries. All the OPEN/CAESAR compilers, application tools, and demo examples have been modified to reflect these changes.

Sustained effort has been made to ensure that CADP works properly on mainstream computing platforms. In particular, the RFL and TST scripts and the documentation have been continuously updated. Changes were brought to CADP to cope with recent C compilers (such as GCC 6 and Clang) and to work around problems with the "indent" command available on Solaris and macOS/Xcode. On Linux, CADP was ported to the latest versions of Centos, Debian 9, and Ubuntu 17.04. The support for the various desktop environments (Gnome, KDE, Mate, etc.) available in Linux distributions has improved. On macOS, support of obsolete versions (from Mac OS X 10.6 "Snow Leopard" to OS X 10.9 "Mavericks" included) was withdrawn and support of macOS 10.13 "High Sierra" was added. Preliminary steps have been made to prepare a 64-bit version of CADP on macOS. On Windows, support of obsolete versions (Windows XP and Vista) was dropped. CADP was also adapted to follow the changes in the Cygwin software regarding pipe management. Many changes were made to CADP so as to support the case where Cygwin is not installed in "C:/" but in a different folder. Finally, preliminary steps have been made to wards a 64-bit version of CADP for Windows.

In collaboration with Soren Enevoldsen (Aalborg University, Denmark), we studied the applicability of CADP tools for analyzing concurrent systems described using weighted CCS (WCCS) [43], an extension of CCS with an action prefix operator carrying a weight represented as a natural number. We developed a prototype OPEN/CAESAR-compliant compiler for WCCS, which enables to produce, in conjunction with the GENERATOR tool of CADP, the corresponding LTS model in which transitions are labeled with actions and weights. For specifying temporal properties of WCCS systems, we developed a prototype MCL library defining the operators of weighted CTL (WCTL) [43] using MCL fixed point operators parameterized by natural numbers. This library, used in conjunction with the EVALUATOR tool, provides an on-the-fly model checker for WCTL equipped with diagnostic capabilities (counterexamples and witnesses).

6.5. Real-Life Applications and Case Studies

6.5.1. Autonomous Resilience of Distributed IoT Applications in a Fog Environment

Participants: Umar Ozeer, Gwen Salaün.

The first year of the PhD thesis started with a state of the art on the resilience mechanisms, broadly in distributed systems and then more specifically in distributed IoT (*Internet of Things*) applications. This resulted, firstly in defining the scope of the thesis and, secondly, in identifying the steps to manage failures, namely state saving, failure detection, fault isolation, and repairing in a consistent state.

A study of the mechanisms for saving the execution state of processes in distributed systems was done. This enabled us to identify the specificities of our environment and to adapt existing snapshot and message logging mechanisms to fit the context of state saving and manipulation in distributed IoT applications in view of repairing failures and re-establishing consistency. We devised a first failure management protocol, which is being tested on an instance of an IoT application test bed at Orange Labs. Next steps include formally verifying the failure management protocol, as well as carrying out further tests on larger scaled applications for the purpose of performance evaluation.

6.5.2. Activity Detection in a Smart Home

Participants: Waqas Imtiaz, Frédéric Lang, Radu Mateescu, Wendelin Serwe.

Ambient intelligence is an active research field, whose aim is to design and analyze smart environments that are capable of automated interaction with users and the physical world, through sensors, actuators, displays, and computational elements, embedded in everyday objects, and connected through a network. In the Grenoble area, the Equipex Amiqual4Home⁰ provides among others access to a Smart Home, which is a fully functional two-stage 90 meters square flat with 4 rooms including an open to kitchen living room, a bedroom, a bathroom and a small office. All the rooms are equipped with cameras, microphones, sensors and actuators to remote control various appliances like rollershutter, lights or multimedia devices. The software architecture of the Smart Home is based on the open source home automation software OpenHAB⁰. It allows a complete control of the flat devices with a single system, despite the various protocols used. Using the rule engine, it also enables the definition of rules expressing how the Smart Home should react to physical (human action, sensors, etc.) or external (weather prediction service, calendar, etc.) events. A difficult question is how to make sure that smart environments are programmed correctly, and will not lead to unexpected or even harmful behaviour.

Smart environments are concurrent and asynchronous by nature. To address the question above, we started, in collaboration with Nicolas Bonnefond (PERVASIVE INTERACTION team and Amiqual4Home), to study how existing tools for the formal design and verification of concurrent asynchronous systems present in the CADP toolbox can be used to verify a smart environment. Firstly, we proposed a translation from OpenHAB rules into a formal LNT model on which properties can be verified [42]. Secondly, in collaboration with Paula Lago and Claudia Roncancio (SIGMA team of LIG), we exploited the dataset ContextAct@A4H of daily living activities collected and annotated within Amiqual4Home for the purpose of activity recognition. Each activity was described as an MCL temporal logic formula that is checked repeatedly on the log of sensor

⁰http://amiqual4home.inria.fr

⁰http://www.openhab.org

measurements until all occurrences of the activity have been found. This approach has the ability to recognize the start and end points of activities (thus not requiring to segment sensor data) and also expresses the temporal order of events, thus palliating a limitation of existing ontology based activity recognition techniques. This led to a publication in an international conference [17].

6.5.3. Other Case Studies

Participants: Hubert Garavel, Frédéric Lang, Lina Marsso, Wendelin Serwe.

The demo examples of CADP, which have been progressively accumulated since the origins of the toolbox, are a showcase for the multiple capabilities of CADP, as well as a test bed to assess the new features of the toolbox. In 2017, the effort to maintain and enhance these demos has been pursued. The demo 05 (Airplane-ground communication protocol) has been modified to use the new syntax of exceptions in the LNT language. The LOTOS and LNT specifications of demo 12 (Message Authenticator Algorithm) have been entirely revised, based upon the fine knowledge acquired by modelling this cryptographic function as a term rewrite system [14]. The LNT specification has also been extended to incorporate the test vectors given in the International Standards ISO 8730 and 8731-2. The resulting specification, which was initially too large to be compiled, is now successfully handled after the enhancements brought to the LNT2LOTOS translator. Demo 19 (Production Cell) has been simplified and is now fully documented in a publication [15].

In the framework of the SECURIOT-2 project (see § 8.2.2.1), a Memory Protection Unit has been formally specified in LNT and verified at Tiempo using CADP. A paper has been submitted to an international conference.

CORSE Project-Team

6. New Results

6.1. Simplification and Run-time Resolution of Data Dependence Constraints for Loop Transformations

Participants: Diogo Nunes Sampaio, Alain Ketterlin [Inria CAMUS], Louis-Noël Pouchet [CSU, USA], Fabrice Rastello.

Loop optimizations such as tiling, thread-level parallelization or vectorization are essential transformations to improve performance. Their use rely on the ability to compute dependence information at compile-time to assess their validity, but in many real situations, dependence analysis fails to provide precise enough information. Typical examples where this happens are when working over compilers IR (e.g., LLVM IR) or with legacy source code, with pointers and linearized arrays (e.g., packed symmetric matrices in BLAS LAPACK). In this scenario, the compiler will often be unable to apply aggressive transformations due to lack of conclusive static dependence analysis.

This work makes a fundamental leap towards enabling complex loop transformations in real-life scenarios, by using a hybrid static+dynamic analysis to disambiguate may-dependencies. Similarly to GCC's autovectorization, our approach consists in adding a lightweight run-time test to check whether ambiguous maydependencies do exist at execution time, to determine whether the optimized or unmodified code version should be called. The main contribution of our work is to generalize this pragmatic approach to a large class of loop-nest transformations, including tiling, loop invariant code motion, parallelization, etc. In particular, we design a quantifier elimination scheme on integer multivariate-polynomials, which can aid application of off-the-shelf polyhedral transformations on a larger class of programs, that holds polynomial memory access and affine loop bounds.

The preciseness of the presented scheme and the low run-time overhead of the test are key to make this approach realistic. We experimentally validate our technique on 25 benchmarks using complex loop transformations, achieving negligible overhead. Preciseness is assessed by the observed success of generated test in practical cases.

IPFME tool 5.5 has been developped in this context. This work is the fruit of the collaboration 8.4.1.1 with OSU. It has been presented at the ACM/SIGARCH International Conference on Supercomputing, ICS 2017 [25].

6.2. Optimizing the Four-Index Integral Transform Using Data Movement Lower Bounds Analysis

Participants: Samyam Rajbhandari [Microsoft, USA], Fabrice Rastello, Karol Kowalski [PNNL, USA], Sriram Krishnamoorthy [PNNL, USA], P. Sadayappan [OSU, USA].

The four-index integral transform is a fundamental and computationally demanding calculation used in many computational chemistry suites such as NWChem. It transforms a four-dimensional tensor from an atomic basis to a molecular basis. This transformation is most efficiently implemented as a sequence of four tensor contractions that each contract a four-dimensional tensor with a two-dimensional transformation matrix. Differing degrees of permutation symmetry in the intermediate and final tensors in the sequence of contractions cause intermediate tensors to be much larger than the final tensor and limit the number of electronic states in the modeled systems.

Loop fusion, in conjunction with tiling, can be very effective in reducing the total space requirement, as well as data movement. However, the large number of possible choices for loop fusion and tiling, and data/computation distribution across a parallel system, make it challenging to develop an optimized parallel implementation for the four-index integral transform. We develop a novel approach to address this problem, using lower bounds modeling of data movement complexity. We establish relationships between available aggregate physical memory in a parallel computer system and ineffective fusion configurations, enabling their pruning and consequent identification of effective choices and a characterization of optimality criteria. This work has resulted in the development of a significantly improved implementation of the four-index transform that enables higher performance and the ability to model larger electronic systems than the current implementation in the NWChem quantum chemistry software suite.

This work is the fruit of the collaboration 8.4.1.1 with OSU. It has been presented at the ACM/SIGPLAN Symposium on Principles and Practice of Parallel Programming, PPoPP 2017 [21].

6.3. Register Optimizations for Stencils on GPUs

Participants: Aravind Sukumaran-Rajam [OSU, USA], Atanas Rountev [OSU, USA], Fabrice Rastello, Louis-Noël Pouchet [CSU, USA], P. Sadayappan [OSU, USA].

The recent advent of compute-intensive GPU architecture has allowed application developers to explore highorder 3D stencils for better computational accuracy. A common optimization strategy for such stencils is to expose sufficient data reuse by means such as loop unrolling, with the hope of register-level reuse. However, the resulting code is often highly constrained by register pressure. While the current state-of-the-art register allocators are satisfactory for most applications, they are unable to effectively manage register pressure for such complex high-order stencils, resulting in a sub-optimal code with a large number of register spills. In this paper, we develop a statement reordering framework that models stencil computations as DAG of trees with shared leaves, and adapts an optimal scheduling algorithm for minimizing register usage for expression trees. The effectiveness of the approach is demonstrated through experimental results on a range of stencils extracted from application codes.

This work is the fruit of the collaboration 8.4.1.1 with OSU. It will be presented at the ACM/SIGPLAN Symposium on Principles and Practice of Parallel Programming, PPoPP 2018.

6.4. Data-Flow/Dependence Profiling for Structured Transformations

Participants: Diogo Nunes Sampaio, Fabian Gruber, Christophe Guillon [STMicroelectronics], Antoine Moynault [STMicroelectronics], Louis-Noël Pouchet [CSU, USA], Fabrice Rastello.

Profiling feedback is an important technique used by developers for performance debugging, where it is usually used to pinpoint performance bottlenecks and also to find optimization opportunities. Assessing the validity and potential benefit of a program transformation requires accurate knowledge of the data flow and data dependencies, which can be uncovered by profiling a particular execution of the program.

In this work we develop MICKEY 5.4, an end-to-end infrastructure for dynamic binary analysis, which produces feedback about the potential to apply structured transformations to uncover non-trivial parallelism and data locality via complex program re-scheduling. Our tool can handle both inter- and intra-procedural aspects of the program in a unified way, thus enabling inter-procedural structured transformations. It is based on QEMU and uses dynamic binary translation to instrument arbitrary programs at run-time. The design of this tool was driven by the goal of achieving portability, both in terms of targeted CPU architectures, but also in terms of programming environment and the use of third-party libraries for which no source code is available.

This work is the fruit of the collaboration 8.4.1.1 with CSU and the contract 7.2 with STMicroelectronics.

6.5. Dynamic Load Balancing of Monte Carlo Particle Transport Applications

Participants: Thomas Gonçalves, Marc Pérache [CEA DAM, Bruyères le Châtel], Frederic Desprez, Jean-Francois Mehaut.

Monte Carlo particle transport applications consist in studying the behavior of particles moving about a simulation domain. Particle distribution among simulation domains is not uniform and changes dynamically during simulation. The parallelization of this kind of applications on massively parallel architectures leads to solve complex issues of workloads and data balancing among numerous compute cores.

This research work started by identifying parallelization pitfalls of Monte Carlo particle transport applications using theoretical and experimental analysis of reference parallelization methods. A semi-dynamic loadbalancing based on partitioning techniques has then been proposed. Finally, we designed a dynamic approach which is able to redistribute workloads and data while keeping a low communication volume. Compared to the perfectly balanced domain replication method using strong scaling measurement, the dynamic approach leads both to speedups and reduction of memory footprint.

This work is part of the Thomas Gonçalves's PhD thesis defended in September 2017 at TERATEC (Bruyères le Châtel). The main contributions of this work were also presented in the ParCo conference [14].

6.6. BOAST: A Meta-Programming Framework to Produce Portable and Efficient Computing Kernels for HPC Applications

Participants: Brice Videau, Kevin Pouget, Luigi Genovese [CEA Inac, Grenoble], Thierry Deutsch [CEA Inac, Grenoble], Dimitri Komatitsch [LMA, CNRS, Marseille], Frédéric Desprez, Jean-Francois Mehaut.

Application portability is an important issue that should be solved efficiently, especially given the large number of different processors now available for today's supercomputers. The work needed to get performance portability is a tedious task, even for experienced programmers. The availability of semi-automatic tools is therefore mandatory for the development of large simulation applications. Computing kernels' identification and optimization has to be carefully performed as they usually consume most of the computing resources.

BOAST is a framework (DSL and run-time) that aims at describing kernels in a high-level language and it allows the comparison of the performance of different versions of the code in a simple and seamless way. We described its application to three use cases from the Mont-Blanc project. Results are encouraging as BOAST proved to be a powerful and flexible tool that allowed gains in performance compared to hand-tuned codes. Performance portability of those codes is also improved.

This work was mainly developed by Brice Videau in the context of the Mont-Blanc FP7 European projects 8.3.1.1. It lead to a publication in the International Journal of High Performance Computing Applications (IJHPCA) [11]. A paper will also be published in 2018 describing the BOAST usage for the Gysela Application (see http://gyseladoc.gforge.inria.fr/).

6.7. Auto-tuning at Run-time with Multiple Implementations of OpenMP Tasks

Participants: Luis Felipe Garlet Millani, Lucas Mello Schnorr [UFRGS, Brazil], Jean-François Mehaut.

OpenMP established itself as the de facto standard for parallel programming in shared memory environments. It received many additions over the years enabling OpenMP to be used with heterogeneous systems. We propose an extension to the task pragma of OpenMP allowing it to provide multiple ways to compute the desired result. The run-time can thus be provided with implementations with different trade-offs.

With the use of the BOAST 5.7 auto-tuning framework, these implementations can be generated automatically before the execution. But within this framework, the auto-tuned kernel is selected in an environment different from that of an actual execution of the application. As a consequence, it may be the case that no interactions occur between different tasks during the auto-tuning, while, in the actual execution, tasks do affect each other due to shared resources like cache or memory bandwidth: Kernel selection done in isolation during the auto-tuning process is probably not the best choice for the embedded execution as part of the full application.

We propose dealing with this limitation by having the auto-tuning phase select not a single but a set of implementations, to be later further selected during execution. Our approach also permits the tuning of different parameters (such as memory accesses and number of operations), and allows to use whichever implementation is more adequate for the thread based on monitored load.

Our extension is implemented within the LLVM framework and Clang compiler front-end. Furthermore we extend the LLVM OpenMP Run-time to be aware of the multiple task implementations. We verify the efficacy of our proposal with the Ondes3D seismic wave simulator and a sparse matrix multiplication application.

6.8. Improving Characterization of NUMA Architectures through Applications' Kernels

Participants: Philippe Virouleau, Francois Broquedis, Thierry Gautier, Julien Langou [UCD, USA], Fabrice Rastello.

Programmers need tools to be able to study their applications. When targeting NUMA architectures, many existing tools allow to observe and identify the critical parts of the application. However there is a need for tools that enable programmers to clearly understand how critical parts of their applications behave, and how they could be improved on a given architecture.

In the context of data-flow applications each part - task - of the application is clearly identified in the data-flow graph. All manipulated data are also clearly available as, within such framework, they constitute what links tasks with one another.

On NUMA architectures, a task's execution time depends, among others, on both the core which executes the task and the NUMA node on which has been allocated its data. Assume one can characterize a task behavior (with regard to its execution context) as follow: run it in isolation from the overall application, and change various of its properties (such as the size of input or the placement of data). Then the scheduler of a run-time system can use this characteristic to improve the overall performance: It would have full information about what is running on the machine (e.g.: on the same NUMA node as the idle thread), and could sort the tasks ready for execution according to how good their behavior would be on the idle thread, given the current state.

We designed a tool which goal is to help the user execute a given *scenario* on the architecture. This scenario describes:

- What are the data and where to allocate them on the architecture
- What are the tasks to execute, where to execute them on the architecture, and with which data
- What are the characteristics to observe for each task (execution time, performance counters, ...)

The tool guarantees that the scenario will be executed correctly on the architecture, letting him focus on understanding on his application rather than taking care of the low level implementation details.

We applied this approach to a dense linear algebra algorithm: the Cholesky factorization. It has enabled us to profile the four kernels of the application by running them in various configurations of data placements, sizes, and concurrent workload. We believe we've tested enough configurations to reliably find the best and worst cases for all the kernels Assuming the behavior of the kernel stays the same within the application, we've been able to estimate upper-bound and lower-bound execution time for the overall application given those best and worst cases.

6.9. Workload-aware Loop Scheduling of Irregular Loops

Participants: Pedro Henrique de Mello Morado Penna, Marcio Castro [UFSC, Universidade Federal de Santa Catarina, Brazil], Henrique Cota de Freitas [Pontifical Catholic University of Minas Gerais, Brazil], Francois Broquedis, Jean-Francois Mehaut.

The input workload of an irregular application must be evenly distributed among its threads to enable cuttingedge performance. To address this need in OpenMP, several loop scheduling strategies were proposed. While having this ever-increasing number of strategies at disposal is helpful, it has become a non-trivial task to select the best one for a particular application. Nevertheless, this challenge becomes easier to be tackled when existing scheduling strategies are extensively evaluated. Therefore, we present a performance and scalability evaluation of the recently proposed loop scheduling strategy named Smart Round-Robin (SRR). To deliver a comprehensive analysis, we coupled a synthetic kernel benchmarking technique with several rigorous statistical tools, and considered OpenMP's Static and Dynamic loop schedulers as our baselines. Our results unveiled that SRR performs better on irregular applications with symmetric workloads and coarsegrained parallelization, achieving up to 1.9x and 1.5x speedup over OpenMP's Static and Dynamic schedulers, respectively.

This work laid the foundations of a collaboration between CORSE, the UFSC, and PUC Minas, that led to the publication of two conference papers (ICCS'17 [20], WSCAD'17 [27]) and one international journal paper (CCPE'17 [5]). The WSCAD'17 paper has also been selected to be extended for a special issue of the CCPE journal. This extension will be based on recent works with the MHM (Multiscale Hybrid-Mixed Methods) simulator developed at LNCC for the H2020 HPC4e project 8.3.1.3 between Europe and Brazil.

6.10. Error-Rate Prediction and Radiation Experiments on a 28nm Many-Core Processor

Participants: Vanessa Vargas [TIMA Labs Grenoble & ESPE, Ecuador], Pablo Ramos [TIMA Labs Grenoble & ESPE, Ecuador], Vincent Ray [kalray, Montbonnot Saint-Martin], Camille Jalier [kalray, Montbonnot Saint-Martin], Renaud Stevens [kalray, Montbonnot Saint-Martin], Benoît Dupont de Dinechin [kalray, Montbonnot Saint-Martin], Maud Baylac [LSPC Labs, Université Grenoble Alpes, CNRS/IN2P3], Francesca Villa [LSPC Labs, Université Grenoble Alpes, CNRS/IN2P3], Nacer-Eddine Zergainoh [TIMA Labs, Université Grenoble Alpes & CNRS], Jean-Francois Mehaut, Raoul Velazco [TIMA Labs, Université Grenoble Alpes & CNRS].

This work analyses the 14 MeV neutron sensitivity of the MPPA-256 many-core processor. Analysis results suggest that ECC and interleaving implemented in the shared memories of clusters are very effective to mitigate SEUs as all detected events of this type were corrected.

The evaluation of the device dynamic response shows that by enabling the cache memories, it is possible to gain in performance of the application without compromising reliability, as all the detected errors produced in data and instruction cache memories were corrected by the parity protection. The non-correctable errors that occurred in the different dynamic tests were produced by bit-flips in general purpose registers since registers do not implement any protection mechanism.

Results show that the predicted application error-rate is reasonably close to the measured one. Consequently, despite the complexity of the many-core processor, this work supports the relevance of the use of the CEU approach to predict the error-rate of applications implemented in such devices.

This work is part of the Vanessa Vargas's PhD thesis defended in April 2017. This PhD thesis was advised by Raoul Velazco (TIMA) and Jean-François Méhaut. Four of the authors worked for the Kalray company. The experiments under radiation were performed on the GENEPI2 platform of the LSPC laboratory. This work was also published in the IEEE Transactions on Nuclear Sciences [10]. It was partially funded by the STIC-Amsud EnergySFE project 8.4.2.

6.11. CAP Bench: A Benchmark Suite for Low-Power Many-Core Processors

Participants: Matheus Souza [PUC Minas], Pedro Henrique de Mello Morado Penna, Matheus Queiroz [PUC Minas], Alyson Pereira [UFSC], Luis Góes [PUC Minas], Henrique Cota de Freitas [PUC Minas], Márcio Castro [UFSC], Philippe Navaux [UFRGS], Jean-Francois Mehaut.

CAP Bench is an open source benchmark suite that includes parallel applications suitable to evaluate emerging low-power many-core processors such as MPPA-256. The benchmark contains a diverse set of applications that evaluated key aspects of MPPA-256, namely the use of its compute clusters, I/O subsystem, NoC and energy consumption. We expose development difficulties and potential bottlenecks that can stem from the shift in development paradigm when programming for low-power many-core architectures. The results showed us that different applications can have different performance bottlenecks, which is why a solid knowledge about the low-power many-core architecture is necessary for the development of efficient programs.

Our analysis shows that CAP Bench is prepared for the analysis of low-power many-core processors such as the MPPA-256, being scalable and concerned with new trends on this type of architectures. To achieve good performance and scalability, we developed applications considering aspects such as parallel patterns, load balance and architecture limitations. This allowed us to evaluate several aspects of the MPPA-256.

Our benchmark explores the hybrid programming model, which is a trend in low-power many- core processors, following parallel patterns. This enables us to verify that, in the case of MPPA-256, communication time may surpass computation time, which would ideally never occur. This behavior was highlighted by the LU application available in CAP Bench, which may indicate that the NoC should be improved to achieve better performance on NoC-bound applications. In this manner, CAP Bench comes up with the proposal to identify such bottlenecks, revealing potential improvements that might be done in future many-core architectures.

Application development challenges are still out there, and have to be solved to enable the evaluation of next generation many-core processors. As future work, we intend to incorporate other applications to the benchmark, to make it more diverse and allow for a better characterization of the architecture and its aspects. We also intend to extend the benchmark use to other many-core architectures, to achieve a broader understanding of them and the differences between many- core processors

This work was developed in the context of the EnergySFE STIC Amsud project 8.4.2. A description of CAP Bench has been published in the CCPE (Concurrency Computation: Practice and Experience) international journal [9]. CAP Bench will be used and extended during Pedro Henrique Penna's doctoral thesis.

6.12. Social Network Analysis on Multi-Core Architectures

Participants: Thomas Messi Nguele, Maurice Tchuente [Univ Yaoundé 1, LIRIMA], Jean-Francois Mehaut.

One of social graph properties is the community structure, that is, subsets where nodes belonging to the same subset have a higher link density between themselves and a low link density with nodes belonging to external subsets. Furthermore, most social network mining algorithms comprise a local exploration of the underlying graph, which consists in referencing nodes in the neighborhood of a particular node.

The main contribution of this work is to use the community structure during the storage of large graphs that arise in social network mining. The goal is to reduce cache misses and consequently, execution time. After formalizing the problem of social network ordering as a problem of optimal linear arrangement which is known as NP-Complete, we propose NumBaCo, a heuristic based on the community structure. We present, for Katz score and Pagerank, simulations that compare classic data structures Bloc and Yale to their corresponding versions that use NumBaCo. Results on a 32-cores NUMA machine using real datasets (amazon, dblp and web-google) show that NumBaCo allows to reduce from 62% to 80% of cache misses and from 15% to 50% of execution time.

This work was initiated inside the LIRIMA Inria International Laboratory with the University of Yaoundé and Maurice Tchuenté. Those results are part of Thomas Messi Nguélé' PhD which is prepared with a Cotutelle agreement. Those results have been presented at the ParCO international conference [19] and published in the ARIMA (Revue Africaine de Recherche en Informatique et Mathématiques Appliquées) journal [3].

6.13. Run-Time Enforcement Using Büchi Games

Participants: Matthieu Renard [LaBRI], Antoine Rollet [LaBRI], Yliès Falcone.

In this work, we leverage Büchi games for the run-time enforcement of regular properties with uncontrollable events. Run-Time enforcement consists in modifying the execution of a running system to have it satisfy a given regular property, modeled by an automaton. We revisit run-time enforcement with uncontrollable events and propose a framework where we model the run-time enforcement problem as a Büchi game and synthesize sound, compliant, and optimal enforcement mechanisms as strategies. We present algorithms and a tool implementing enforcement mechanisms. We reduce the complexity of the computations performed by enforcement mechanisms at run-time by pre-computing the decisions of enforcement mechanisms ahead of time.

This work has been presented at the 24th ACM/SIGSOFT International SPIN Symposium on Model Checking of Software, SPIN 2017 [23].

6.14. GREP: Games for the Run-Time Enforcement of Properties

Participants: Matthieu Renard [LaBRI], Antoine Rollet [LaBRI], Yliès Falcone.

In this work, we developed GREP, a tool for the run-time enforcement of (timed) properties. GREP takes an execution sequence as input (stdin), and modifies it (stdout) as necessary to enforce the desired property, when possible. GREP can enforce any regular timed property described by a deterministic and complete Timed Automaton. The main novelties of GREP are twofold: It uses game theory to improve the synthesis of enforcement mechanisms, and it accounts for uncontrollable events, i.e. events that cannot be controlled by the enforcement mechanisms and thus have to be released immediately. The usability of GREP has been validated with a performance evaluation.

The associated work has been presented at the IFIP International Conference on Testing Software and Systems, ICTSS 2017 [22]

6.15. Verifying Policy Enforcers

Participants: Oliviero Riganelli [University of Milano Bicocca], Daniela Micucci [University of Milano Bicocca], Leonardo Mariani [University of Milano Bicocca], Yliès Falcone.

Policy enforcers are sophisticated run-time components that can prevent failures by enforcing the correct behavior of the software. While a single enforcer can be easily designed focusing only on the behavior of the application that must be monitored, the effect of multiple enforcers that enforce different policies might be hard to predict. So far, mechanisms to resolve interferences between enforcers have been based on priority mechanisms and heuristics. Although these methods provide a mechanism to take decisions when multiple enforcers try to affect the execution at a same time, they do not guarantee the lack of interference on the global behavior of the system. In this work we propose a verification strategy that can be exploited to discover interferences between sets of enforcers and thus safely identify a priori the enforcers that can co-exist at run-time. In our evaluation, we experimented our verification method with several policy enforcers for Android and discovered some incompatibilities.

This work has been presented at the 17-th International Conference on Run-Time Verification, RV 2017 [24].

6.16. Monitoring Decentralized Specifications

Participants: Antoine El-Hokayem, Yliès Falcone.

In this work, we define two complementary approaches to monitor decentralized systems. The first approach relies on those with a centralized specification, i.e, when the specification is written for the behavior of the entire system. To do so, our approach introduces a data-structure that i) keeps track of the execution of an automaton, ii) has predictable parameters and size, and iii) guarantees strong eventual consistency. The second approach defines decentralized specifications wherein multiple specifications are provided for separate parts of the system. We study decentralized monitorability, and present a general algorithm for monitoring decentralized specifications. We map three existing algorithms to our approaches and provide a framework for analyzing their behavior.

The associate tool THEMIS 5.1 is a framework for designing such decentralized algorithms, and simulating their behavior. This work has been presented at the 26th ACM/SIGSOFT International Symposium on Software Testing and Analysis, ISSTA 2017 [12].

6.17. Interactive Run-Time Verification - When Interactive Debugging Meets Run-Time Verification

Participants: Raphael Jakse, Yliès Falcone, Kevin Pouget, Jean-Francois Mehaut.

Run-Time Verification consists in studying a system at run-time, looking for input and output events to discover, check or enforce behavioral properties. Interactive debugging consists in studying a system at runtime in order to discover and understand its bugs and fix them, inspecting interactively its internal state. Interactive Run-Time Verification (i-RV) combines run-time verification and interactive debugging. We define an efficient and convenient way to check behavioral properties automatically on a program using a debugger. We aim at helping bug discovery and understanding by guiding classical interactive debugging techniques using run-time verification.

This work has been presented at the IEEE 28th International Symposium on Software Reliability Engineering, ISSRE 2017 [15]. It is also a part of the Nano2017 DEMA project 7.2 with STMicroelectronics.

6.18. Predictive Run-Time Verification of Timed Properties

Participants: Srinivas Pinisetty [Aalto University], Thierry Jéron [Inria Rennes], Stravos Tripakis [Aalto University], Yliès Falcone, Hervé Marchand [Inria Rennes], Viorel Preoteasa [Aalto University].

Run-Time verification (RV) techniques are used to continuously check whether the (un-trustworthy) output of a black-box system satisfies or violates a desired property. When we consider run-time verification of timed properties, physical time elapsing between actions influences the satisfiability of the property. This work introduces predictive run-time verification of timed properties where the system is not entirely a blackbox but something about its behavior is known a priori. A priori knowledge about the behavior of the system allows the verification monitor to foresee the satisfaction (or violation) of the monitored property. In addition to providing a conclusive verdict earlier, the verification monitor also provides additional information such as the minimum (maximum) time when the property can be violated (satisfied) in the future. The feasibility of the proposed approach is demonstrated by a prototype implementation, which is able to synthesize predictive run-time verification monitors from timed automata.

This work has been published in the Journal of Systems and Software 2017 [6].

6.19. Concurrency-Preserving and Sound Monitoring of Multi-Threaded Component-based Systems: Theory, Algorithms, Implementation, and Evaluation

Participants: Hosein Nazarpour [Verimag], Yliès Falcone, Saddek Bensalem [Verimag], Marius Bozga [Verimag].

This work addresses the monitoring of logic-independent linear-time user-provided properties in multithreaded component-based systems. We consider intrinsically independent components that can be executed concurrently with a centralized coordination for multiparty interactions. In this context, the problem that arises is that a global state of the system is not available to the monitor. A naive solution to this problem would be to plug in a monitor which would force the system to synchronize in order to obtain the sequence of global states at run-time. Such a solution would defeat the whole purpose of having concurrent components. Instead, we reconstruct on-the-fly the global states by accumulating the partial states traversed by the system at run-time. We define transformations of components that preserve their semantics and concurrency and, at the same time, allow to monitor global-state properties. Moreover, we present RVMT-BIP, a prototype tool implementing the transformations for monitoring multi-threaded systems described in the Behavior, Interaction, Priority (BIP) framework, an expressive framework for the formal construction of heterogeneous systems. Our experiments on several multi-threaded BIP systems show that RVMT-BIP induces a cheap run-time overhead. This work has been published in the journal Formal Aspects of Computing 2017 [4].

6.20. Formal Analysis and Offline Monitoring of Electronic Exams

Participants: Ali Kassem [Inria Grenoble], Yliès Falcone, Pascal Lafourcade [University of Auvergne].

More and more universities are moving toward electronic exams (in short e-exams). This migration exposes exams to additional threats, which may come from the use of the information and communication technology. In this work, we identify and define several security properties for e-exam systems. Then, we show how to use these properties in two complementary approaches: model-checking and monitoring. We illustrate the validity of our definitions by analyzing a real e-exam used at the pharmacy faculty of University Grenoble Alpes (UGA) to assess students. On the one hand, we instantiate our properties as queries for ProVerif, an automatic verifier of cryptographic protocols, and we use it to check our modeling of UGA exam specifications. ProVerif found some attacks. On the other hand, we express our properties as Quantified Event Automata (QEAs), and we synthesize them into monitors using MarQ, a Java tool designed to implement QEAs. Then, we use these monitors to verify real exam executions conducted by UGA. Our monitors found fraudulent students and discrepancies between the specifications of UGA exam and its implementation.

This work has been published in the journal of Formal Methods in System Design 2017 [2].

6.21. Teaching Algorithms using Problem and Challenge Based Learning

Participant: Florent Bouchez - Tichadou.

Teaching algorithms is always a challenge at any level of the CS curriculum, as it is often viewed as a theoretical field. While many exercises revolve around classical examples that illustrate interesting algorithmic points, they are often disconnected from reality, which is a major drawback for students trying to learn. During the last four years, we have been trying to reconnect the teaching of algorithms with their applicability in the real world to M1 and L2 students, by giving them actual problems that could arise in their life of future software engineers, challenging enough to force them to use particular algorithmic techniques or data structures—e.g., linked lists, binary trees, dynamic programming or approximation algorithms.

By assigning students in groups of 5 to 6 members, we wanted to create an environment where they function as a team trying to work together to solve a problem. This allowed them to help each other in their respective comprehension, and made them more autonomous in their learning. The effective materials was provided as online pdf files so they had to read and learn from them by themselves, while the class sessions with a tutor (teacher) where used for the problem-solving part, with guidance from the tutor (who is there to make sure the learning takes place).

After four years of experimentation with M1 students, we found that the student's grades were stable, in particular there was no decrease in exams' performances compared to the classical course that was taught in the previous years. However, the students progressed in trans-disciplinary skills such a communication and the writing of essays. More importantly, students show a strong adhesion to the teaching method, 50% of them rating it as "excellent" (6) and 25% as "good" (resp. 6 and 5 on a scale from 1 (terrible) to 6 (excellent)). No student rated the course below average.

This work is still ongoing, and our plan now is to use our knowledge of the internals of compilers and run-time systems to: First, extract real-life algorithmic problems that have concrete applications; Second, create a tool that exposes the working mechanics of a running program, hence helping students to better understand how algorithms work.
CTRL-A Project-Team

6. New Results

6.1. Programming support for Autonomic Computing

6.1.1. Reactive languages

Participants: Gwenaël Delaval, Eric Rutten.

Our work in reactive programming for autonomic computing systems is focused on the specification and compilation of declarative control objectives, under the form of contracts, enforced upon classical mode automata as defined in synchronous languages. The compilation involves a phase of Discrete Controller Synthesis in order to obtain an imperative executable code. The programming language Heptagon / BZR (see Section Software and Platforms) integrates our research results [7].

Recent work concerns exploring new possibilities offered by logics-numeric control. We target the problem of the safe control of reconfigurations in component-based software systems (see also Section 6.1.2 for the component-based aspects), where strategies of adaptation to variations in both their environment and internal resource demands need to be enforced. In this context, the computing system involves software components that are subject to control decisions. We approach this problem under the angle of Discrete Event Systems (DES), involving properties on events observed during the execution (e.g., requests of computing tasks, work overload), and a state space representing different configurations such as activity or assemblies of components. We consider in particular the potential of applying novel logico-numerical control techniques to extend the expressivity of control models and objectives, thereby extending the application of DES in component-based software systems. We elaborate methodological guidelines for the application of logico-numerical control based on a case-study, and validate the result experimentally.

This work is in cooperation with the Sumo team at Inria Rennes and University of Liverpool, and is published in the CCTA 2017 conference [15].

6.1.2. Component-based approaches

Participants: Gwenaël Delaval, Eric Rutten.

Our work in component-based programming for autonomic computing systems as examplified by e.g., FRAC-TAL, considers essentially the problem of specifying the control of components assembly reconfiguration, with an approach based on the integration within such a component-based framework of a reactive language as in Section 6.1.1 [6].

Dynamic reconfiguration is a key capability of Component-based Software Systems to achieve self-adaptation as it provides means to cope with environment changes at runtime. The space of configurations is defined by the possible assemblies of components, and navigating this space while achieving goals and maintaining structural properties is managed in an autonomic loop. The natural architectural structure of component-based systems calls for hierarchy and modularity in the design and implementation of composites and their managers, and requires support for coordinated multiple autonomic loops. [1] [12].

In recent work, we leverage the modularity capability to strengthen the Domain-Specific Language (DSL) Ctrl-F, targeted at the design of autonomic managers in component-based systems. Its original definition involved discrete control-theoretical management of reconfigurations, providing assurances on the automated behaviors. The objective of modularity is two-fold: from the design perspective, it allows designers to seamlessly decompose a complex system into smaller pieces of reusable architectural elements and adaptive behaviours. From the compilation point of view, we provide a systematical and generative approach to decompose control problems described in the architectural level while relying on mechanisms of modular Discrete Control Synthesis (DCS), which allows us to cope with the combinatorial complexity that is inherent to DCS problems. We show the applicability of our approach by applying it to the self-adaptive case study of the existing RUBiS/Brownout eBay-like web auction system. This work is done in cooperation with Inria teams ASCOLA in Nantes and SPIRALS in Lille, and is published in the Journal of Systems and Software [12] and the SeAC 2017 - 2nd Workshop on Self-Aware Computing, a satellite of the ICAC'17 conference [14].

We are also considering integration at the DSL level of expressivity extensions, for which the compilation and controller synthesis is relying on the ReaX tool developed at Inria Rennes, in the Sumo team as mentioned in Section 6.1.1 [15].

6.1.3. Rule-based systems

Participants: Adja Sylla, Gwenaël Delaval, Eric Rutten.

This work concerns a high-level language for safe rule-based programming in the LINC transactional rulebased platform developed at CEA [17]. Rule based middlewares such as LINC enable high level programming of distributed adaptive systems behaviours. LINC also provides the systems with transactional guarantees and hence ensures their reliability at runtime. However, the set of rules may contain design errors (e.g. conflicts, violations of constraints) that can bring the system in unsafe safe or undesirables states, despite the guarantees provided by LINC. On the other hand, automata based languages such as Heptagon/BZR enable formal verification and especially synthesis of discrete controllers to deal with design errors. Our work studies these two languages and combines their execution mechanisms, from a technical perspective. We target applications to the domain of Internet of Things and more particularly smart building, office or home (see Section 6.2.2.1).

This work is in cooperation with CEA LETI/DACLE, it is the topic of the PhD of Adja Sylla at CEA, coadvised with M. Louvel [11], and aspects on Software Engineering and Software Architecture for Multiple Autonomic Loops are published in the ICCAC 2017 conference [18].

6.1.4. A Language for the Smart Home

Participant: Gwenaël Delaval.

This work is about the design of the CCBL programming language (Cascading Contexts Based Language), an end-user programming language dedicated to Smart Home. CCBL has been proposed to avoid the problems encountered by end-users programming with ECA (Event Conditions Actions), which is the dominant approach in the field. This language has been evaluated by means of a user-based experiment where 21 adults (11 experimented programmers and 10 non-programmers) have been asked to express four increasingly complex behaviors using both CCBL and ECA. It has been shown that significantly less errors were made using CCBL than using ECA. From this experiment, some categorization and explanation of the errors made when using ECA have been proposed, with explanations about why users avoid these errors when programming with CCBL. Finally, error reporting for CCBL have been explored by identifying two specific errors and by developing a solution based on Heptagon and ReaX to detect them in CCBL programs.

This work is done in cooperation with the IIHM team of LIG (Alexandre Demeure), in the framework of a LIG « projet émergence » and was the topic of the MSc internship of Lénaïg Terrier [22].

6.2. Design methods for reconfiguration controller design in computing systems

We apply the results of the previous axes of the team's activity to a range of infrastructures of different natures, but sharing a transversal problem of reconfiguration control design. From this very diversity of validations and experiences, we draw a synthesis of the whole approach, towards a general view of Feedback Control as MAPE-K loop in Autonomic Computing [20] [19].

6.2.1. High-Performance Computing

Participants: Soguy Mak Kare Gueye, Gwenaël Delaval, Stéphane Mocanu, Bogdan Robu, Eric Rutten.

6.2.1.1. Towards a Control-Theory based approach for cluster overload avoidance

This work is addressing the problem of automated resource management in an HPC infrastructure, using techniques from Control Theory to design a controller that maximizes cluster utilization while avoiding overload. We put in place a mechanism for feedback (Proportional Integral, PI) to system software, through a maximum number of jobs to be sent to the cluster, in response to system information about the current number jobs processed.

This work is done in cooperation with the Datamove team of Inria/LIG, and Gipsa-lab. It was the topic of the internship of Emmanuel Stahl for the Grenoble INP ENSE³ engineering school, [21].

6.2.1.2. Reconfiguration control in DPR FPGA

6.2.1.2.1. DPR FPGA and discrete control for reconfiguration

Implementing self-adaptive embedded systems, such as UAV drones, involves an offline provisioning of the several implementations of the embedded functionalities with different characteristics in resource usage and performance in order for the system to dynamically adapt itself under uncertainties. FPGA-based architectures offer for support for high flexibility with dynamic partial reconfiguration (DPR) features. We propose an autonomic control architecture for self-adaptive and self-reconfigurable FPGA-based embedded systems. The control architecture is structured in three layers: a mission manager, a reconfiguration manager and a scheduling manager. In this work we focus on the design of the reconfiguration manager. We propose a design approach using automata-based discrete control. It involves reactive programming that provides formal semantics, and discrete controller synthesis from declarative objectives.

This work is in the framework of the ANR project HPeC (see Section 8.2.1), and is published in the AHS 2017 conference [16].

6.2.1.2.2. Mission management and stochastic control

In the Mission Management workpackage of the ANR project HPeC, a concurrent control methodology is constructed for the optimal mission planning of a U.A.V. in stochastic environnement. The control approach is based on parallel ressource sharing Partially Observable Markov Decision Processes modeling of the mission. The parallel POMDP are reduced to discrete Markov Decision Models using Bayesian Networks evidence for state identification. The control synthesis is an iterative two step procedure : first MDP are solved for the optimisation of a finite horizon cost problem ; then the possible ressource conflicts between parallel actions are solved either by a priority policy or by a QoS degradation of actions, e.g., like using a lower resolution version of the image processing task if the ressource availability is critical.

6.2.2. IoT

Participants: Neïl Ayeb, Adja Sylla, Gwenaël Delaval, Stéphane Mocanu, Eric Rutten.

6.2.2.1. Control of smart buildings

A smart environment is equipped with numerous devices (i.e., sensors, actuators) that are possibly distributed over different locations (e.g., rooms of a smart building). These devices are automatically controlled to achieve different objectives related, for instance, to comfort, security and energy savings. Controlling smart environment devices is not an easy task. This is due to: the heterogeneity of devices, the inconsistencies that can result from communication errors or devices failure, and the conflicting decisions including those caused by environment dependencies.

Our work proposes a design framework for the reliable and environment aware management of smart environment devices. The framework is based on the combination of the rule based middleware LINC and the automata based language Heptagon/BZR (H/BZR). It consists of: an abstraction layer for the heterogeneity of devices, a transactional execution mechanism to avoid inconsistencies and a controller that, based on a generic model of the environment, makes appropriate decisions and avoids conflicts. A case study with concrete devices, in the field of building automation, is presented to illustrate the framework.

This work is in the framework of the cooperation with CEA (see Section 7.1), and is published in the Springer Journal of Internet Services and Applications, with recognized editors from the Middleware community [13]

75

6.2.2.2. Device management

The research topic is targeting an adaptative and decentralized management for the IoT. It will contribute design methods for processes in virtualized gateways in order to enhance IoT infrastructures.

More precisely, it concerns Device Management in the case of large numbers of connected sensors and actuators, as can be found in Smart Home and Building, Smart Electricity grids, and industrial frameworks as in Industry 4.0.

In contrast with a centralized management of such large sets of devices, for the autonomic management of their adaptations, upgrades and other commands, the objective its to target a distributed management, enabling local decisions, by proposing an appropriate middleware framework. These local adjustments will be processed using context data. The context is a synchronized (i.e., always up-to-date with reality) description of concepts and relations. Technically, the context data information are extracted from multiple sources such as IT environment, user environment and physical environment.

This work is in the framework of the Inria/Orange labs joint laboratory (see Section 7.2.1), and supported by the CIFRE PhD thesis grant of Neïl Ayeb, starting dec. 2017.

6.2.2.3. Security in SCADA industrial systems

We focus mainly on vulnerability search, automatic attack vectors synthesis and intrusion detection. Model checking techniques are used for vulnerability search and automatic attack vectors construction. Intrusion detection is mainly based on process-oriented detection with a technical approach from run-time monitoring. The LTL formalism is used to express safety properties which are mined on an attack-free dataset. The resulting monitors are used for fast intrusion detections.

A demonstrator of attack/defense scenarion in SCADA systems will be built on the existing G-ICS lab (hosted by ENSE3/Grenoble-INP).

This work is in the framework of the ANR project Sacade on cybersecurity of industrial systems (see Section 8.2.2),

DANTE Project-Team

7. New Results

7.1. Graph & Signal Processing

Participants: Paulo Gonçalves, Éric Fleury, Sarra Ben Alaya, Esteban Bautista Ruiz, Gaëtan Frusque, Sarah de Nigris, Mikhail Tsitsvero.

7.1.1. Fractional Semi-Supervised Machine Learning

Graph-based semi-supervised learning for classification endorses a nice interpretation in terms of diffusive random walks, where the regularisation factor in the original optimisation formulation plays the role of a restarting probability. Recently, a new type of biased random walks for characterising certain dynamics on networks have been defined and rely on the γ -th power of the standard Laplacian matrix \mathbf{L}^{γ} , with $\gamma > 0$. In particular, these processes embed long range transitions, the Lévy flights, that are capable of one-step jumps between far-distant states (nodes) of the graph. In a series of two articles [28] and [29], we envisioned to build upon these volatile random walks to propose two new versions of graph based semi-supervised learning algorithms: one called fractional SSL corresponds to the case where $0 < \gamma < 1$ whose classification outcome could benefit from the dynamics induced by the fractional transition matrix, and the other less straightforwardly connected to random walks, derives from $\gamma > 1$.

7.1.2. Design of graph filters and filterbanks

Basic operations in graph signal processing consist in processing signals indexed on graphs either by filtering them or by changing their domain of representation, in order to better extract or analyze the important information they contain. The aim of our chapter [58] is to review general concepts underlying such filters and representations of graph signals. We first recall the different Graph Fourier Transforms that have been developed in the literature, and show how to introduce a notion of frequency analysis for graph signals by looking at their variations. Then, we move to the introduction of graph filters, that are defined like the classical equivalent for 1D signals or 2D images, as linear systems which operate on each frequency of a signal. Some examples of filters and of their implementations are given. Finally, as alternate representations of graph signals, we focus on multiscale transforms that are defined from filters. Continuous multiscale transforms such as spectral wavelets on graphs are reviewed, as well as the versatile approaches of filterbanks on graphs. Several variants of graph filterbanks are discussed, for structured as well as arbitrary graphs, with a focus on the central point of the choice of the decimation or aggregation operators.

7.1.3. GraSP: A Matlab Toolbox for Graph Signal Processing

In [30], we publicised the recent developments and new functionalities of our Graph Signal Processing Toolbox (GraSP).

7.2. Performance analysis and networks protocols

Participants: Mohammed Amer, Thomas Begin, Anthony Busson, Éric Fleury, Yannick Leo, Isabelle Guerin Lassous, Philippe Nain, Huu Nghi Nguyen, Laurent Reynaud.

7.2.1. Network Softwarization

We have developed a modelling framework to analytically evaluate the performance of DPDK-based virtual switches in the context of NFV (Network Function Virtualisation) networks. In [34], we extended our previous work [82] to enable non-null switch-over times that account for a delay overhead whenever a CPU starts polling a different queue. More recently, in [35], we refined our framework to let it deal with batches of packets (i.e. several packets on the same queue are processed together) that tends to speed up the performance of the virtual switches. These works were partly funded by the French ANR REFLEXION under the "ANR-14-CE28-0019" project.

7.2.2. Wi-Fi optimization

Densification of Wi-Fi networks has led to the possibility for a station to choose between several access points (APs). On the other hand, the densification of APs generates interference, contention and decreases the global throughput as APs have to share a limited number of channels. Optimizing the association step between APs and stations can alleviate this problem and increase the overall throughput and fairness between stations. We have proposed an original solution to this optimization problem based on a mathematical model and introduce a local search algorithm to solve this problem through a suitable neighborhood structure. Our evaluation, based on simulations, shows that the proposed solution improves the overall throughput and the fairness of the network. We are currently working on variant of this problem where the traffic to the stations is taken into account in the model and the optimization formulation.

7.2.3. Caching

In [72] we focus on the LRU cache where requests for distinct contents are described by independent stationary and ergodic processes. We extend a TTL-based approximation of the cache hit probability first proposed by R. Fagin in 1977 for the independence reference model to this more general workload model. We show that under very general conditions this approximation is exact as the cache size and the number of contents go to infinity. Moreover, we establish this not only for the aggregate cache hit probability but also for every individual content. Last, we obtain a rate of convergence.

In [70] we consider the problem of allocating cache resources among multiple content providers. The cache can be partitioned into slices and each partition can be dedicated to a particular content provider, or shared among a number of them. It is assumed that each partition employs the LRU policy for managing content. We propose utility-driven partitioning, where we associate with each content provider a utility that is a function of the hit rate observed by the content provider. We consider two scenarios: i) content providers serve disjoint sets of files, ii) there is some overlap in the content served by multiple content providers. In the first case, we prove that cache partitioning outperforms cache sharing as cache size and numbers of contents served by providers go to infinity. In the second case, It can be beneficial to have separate partitions for overlapped content. In the case of two providers it is usually always beneficial to allocate a cache partition to serve all overlapped content and separate partitions to serve the non-overlapped contents of both providers. We establish conditions when this is true asymptotically but also present an example where it is not true asymptotically. We develop online algorithms that dynamically adjust partition sizes in order to maximize the overall utility and prove that they converge to optimal solutions, and through numerical evaluations we show they are effective.

7.2.4. Mobile networks

The development of analytical models to analyze the behavior of vehicular ad hoc networks (VANETs) is a challenging aim. Adaptive methods are suitable for many algorithms (*e.g.* choice of forwarding paths, dynamic resource allocation, channel control congestion) and services (*e.g.* provision of multimedia services, message dissemination). These adaptive algorithms help the network to maintain a desired performance level. However, this is a difficult goal to achieve, especially in VANETs due to fast position changes of the VANET nodes. Adaptive decisions should be taken according to the current conditions of the VANET. Therefore, evaluation of transient measures is required for the characterization of VANETs. In the literature, different works address the characterization and measurement of the idle (or busy) time to be used in different proposals to attain a more efficient usage of wireless network. We have developed an analytical model based on a straightforward Markov reward chain (MRC) to obtain transient measurements of the idle time of the link between two VANET nodes. We have shown that numerical results from the analytical model fit well with simulation results [20].

In another study, we have investigated the application of an adapted controlled mobility strategy on selfpropelling nodes, which could efficiently provide network resource to users scattered on a designated area. We have designed a virtual force-based controlled mobility scheme (called VFPc) and evaluated its ability to be jointly used with a dual packet-forwarding and epidemic routing protocol. In particular, we have studied the possibility for end-users to achieve synchronous communications at given times of the considered scenarios. On this basis, we have studied the delay distribution for such user traffic and show the advantages of our solution compared to other packet-forwarding and packet-replication schemes, and highlighted that VFPcenabled applications could take benefit of both schemes to yield a better user experience, despite challenging network conditions [21].

7.3. Modeling of Dynamics of Complex Networks

Participants: Jean Pierre Chevrot, Christophe Crespelle, Sicheng Dai, Éric Fleury, Eric, Philippe Guichard, Márton Karsai, Yannick Leo, Sebastien Lerique, Jacob Levy Abitbol, Jean-Philippe Magué, Matteo Morini, Samuel Unicomb, Samuel Unicomb.

7.3.1. Multilayer networks

In [67] we introduce a new class of stochastic multilayer networks. A stochastic multilayer network is the aggregation of M networks (one per layer) where each is a subgraph of a foundational network G. Each layer network is the result of probabilistically removing links and nodes from G. The resulting network includes any link that appears in at least K layers. This model is an instance of a non-standard site-bond percolation model. Two sets of results are obtained: first, we derive the probability distribution that the M-layer network is in a given configuration for some particular graph structures (explicit results are provided for a line, an algorithm is provided for a tree), where a configuration is the collective state of all links (each either active or inactive). Next, we show that for appropriate scalings of the node and link selection processes in a layer, links are asymptotically independent as the number of layers goes to infinity, and follow a Poisson distribution. Numerical results are provided to highlight the impact of having several layers on some metrics of interest (including expected size of the cluster a node belongs to in the case of the line). This model finds applications in wireless communication networks, and, more generally, in any scenario where a common set of nodes can be linked via co-existing means of connectivity.

7.3.2. Models of time varying networks

In terms of modelling temporal networks we had the following main contributions in 2017.

A book on Bursty Human Dynamics, written by M. Karsai as a the leading author. Bursty dynamics is a common temporal property of various complex systems in Nature but it also characterises the dynamics of human actions and interactions. At the phenomenological level it is a feature of all systems that evolve heterogeneously over time by alternating between periods of low and high event frequencies. In such systems, bursts are identified as periods in which the events occur with a rapid pace within a short time-interval while these periods are separated by long periods of time with low frequency of events. As such dynamical patterns occur in a wide range of natural phenomena, their observation, characterisation, and modelling have been a long standing challenge in several fields of research. However, due to some recent developments in communication and data collection techniques it has become possible to follow digital traces of actions and interactions of humans from the individual up to the societal level. This led to several new observations of bursty phenomena in the new but largely unexplored area of human dynamics, which called for the renaissance to study these systems using research concepts and methodologies, including data analytics and modelling. As a result, large amount of new insight and knowledge as well as innovations have been accumulated in the field, which provided the timely opportunity to write a monograph book [56] to make an up-to-date review and summary of the observations, appropriate measures, modelling, and applications of heterogeneous bursty patterns occurring in the dynamics of human behaviour.

In another contribution M. Karsai and collaborators introduced a new representation of temporal networks [73]. The dynamics of diffusion-like processes on temporal networks are influenced by correlations in the times of contacts. This influence is particularly strong for processes where the spreading agent has a limited lifetime at nodes: disease spreading (recovery time), diffusion of rumors (lifetime of information), and passenger routing (maximum acceptable time between transfers). We introduce weighted event graphs as a powerful and fast framework for studying connectivity determined by time-respecting paths where the allowed waiting times between contacts have an upper limit. We study percolation on the weighted

event graphs and in the underlying temporal networks, with simulated and real-world networks. We show that this type of temporal-network percolation is analogous to directed percolation, and that it can be characterized by multiple order parameters.

M. Karsai also contributed to a new definition to better quantify attention distributed in dynamical egocentric social networks [64]. Granovetter's weak tie theory of social networks is built around two central hypotheses. The first states that strong social ties carry the large majority of interaction events; the second maintains that weak social ties, although less active, are often relevant for the exchange of especially important information (e.g., about potential new jobs in Granovetter's work). While several empirical studies have provided support for the first hypothesis, the second has been the object of far less scrutiny. A possible reason is that it involves notions relative to the nature and importance of the information of both Granovetter's hypotheses. We find clear empirical support for the first. We also provide empirical evidence and a quantitative interpretation for the second. We show that attention, measured as the fraction of interactions devoted to a particular social connection, is high on weak ties — possibly reflecting the postulated informational purposes of such ties — but also on very strong ties. Data from online social media and mobile communication reveal network-dependent mixtures of these two effects on the basis of a platform's typical usage. Our results establish a clear relationships between attention, importance, and strength of social links, and could lead to improved algorithms to prioritize social media content.

7.3.3. Dynamical processes on networks

Another field which has been intensively studied during the last year addresses dynamical processes on temporal and static networks.

In a book chapter M. Karsai summarised his recent findings on temporal network immunisation [57]. The vast majority of strategies aimed at controlling contagion processes on networks consider a timescale separation between the evolution of the system and the unfolding of the process. However, in the real world, many networks are highly dynamical and evolve, in time, concurrently to the contagion phenomena. Here, we review the most commonly used immunization strategies on networks. In the first part of the chapter, we focus on controlling strategies in the limit of timescale separation. In the second part instead, we introduce results and methods that relax this approximation. In doing so, we summarize the main findings considering both numerical and analytically approaches in real as well as synthetic time-varying networks.

With the PhD student S. Unicomb, M. Karsai and a collaborator developed a new formalism, which is capable to precisely capture and predict the non-monotonous dependence of threshold driven dynamics on weight heterogeneities in networks [76]. Weighted networks capture the structure of complex systems where interaction strength is meaningful. This information is essential to a large number of processes, such as threshold dynamics, where link weights reflect the amount of influence that neighbours have in determining a node's behaviour. Despite describing numerous cascading phenomena, such as neural firing or social contagion, the modelling of threshold dynamics on weighted networks has been largely overlooked. We fill this gap by studying a dynamical threshold model over synthetic and real weighted networks with numerical and analytical tools. We show that the time of cascade emergence depends non-monotonously on weight heterogeneities, which accelerate or decelerate the dynamics, and lead to non-trivial parameter spaces for various networks and weight distributions. Our methodology applies to arbitrary binary state processes and link properties, and may prove instrumental in understanding the role of edge heterogeneities in various natural and social phenomena.

With other co-authors, M. Karsai published another book chapter about his recent findings on the modelling threshold driven dynamics on networks [55]. The collective behaviour of people adopting an innovation, product or online service is commonly interpreted as a spreading phenomenon throughout the fabric of society. This process is arguably driven by social influence, social learning and by external effects like media. Observations of such processes date back to the seminal studies by Rogers and Bass, and their mathematical modelling has taken two directions: One paradigm, called simple contagion, identifies adoption spreading with an epidemic process. The other one, named complex contagion, is concerned with behavioural thresholds

and successfully explains the emergence of large cascades of adoption resulting in a rapid spreading often seen in empirical data. The observation of real world adoption processes has become easier lately due to the availability of large digital social network and behavioural datasets. This has allowed simultaneous study of network structures and dynamics of online service adoption, shedding light on the mechanisms and external effects that influence the temporal evolution of behavioural or innovation adoption. These advancements have induced the development of more realistic models of social spreading phenomena, which in turn have provided remarkably good predictions of various empirical adoption processes. In this chapter we review recent datadriven studies addressing real-world service adoption processes. Our studies provide the first detailed empirical evidence of a heterogeneous threshold distribution in adoption. We also describe the modelling of such phenomena with formal methods and data-driven simulations. Our objective is to understand the effects of identified social mechanisms on service adoption spreading, and to provide potential new directions and open questions for future research.

Y. Leo, E. Fleury and M. Karsai is in the final stage to publish a study on a unique mobile call/banking dataset on the dynamics of purchasing patterns. We analyse a coupled dataset collecting the mobile phone communications and bank transactions history of a large number of individuals living in a Latin American country. After mapping the social structure and introducing indicators of socioeconomic status, demographic features, and purchasing habits of individuals we show that typical consumption patterns are strongly correlated with identified socioeconomic classes leading to patterns of stratification in the social structure. In addition we measure correlations between merchant categories and introduce a correlation network, which emerges with a meaningful community structure. We detect multivariate relations between merchant categories and show correlations in purchasing habits of individuals. Finally, by analysing individual consumption histories, we detect dynamical patterns in purchase behaviour and their correlations with the socioeconomic status, demographic characters and the egocentric social network of individuals. Our work provides novel and detailed insight into the relations between social and consuming behaviour with potential applications in resource allocation, marketing, and recommendation system design.

7.3.4. SoSweet

The SoSweet project focuses on the synchronic variation and the diachronic evolution of the variety of French language used on Twitter.

In one paper accepted to WWW'18 we addressed some of the main questions of the project using a unique dataset combining the largest French Twitter dataset and demographic data coming from INSEE [31]. Our usage of language is not solely reliant on cognition but is arguably determined by myriad external factors leading to a global variability of linguistic patterns. This issue, which lies at the core of sociolinguistics and is backed by many small-scale studies on face-to-face communication, is addressed here by constructing a dataset combining the largest French Twitter corpus to date with detailed socioeconomic maps obtained from national census in France. We show how key linguistic variables measured in individual Twitter streams depend on factors like socioeconomic status, location, time, and the social network of individuals. We found that (*i*) people of higher socioeconomic status, active to a greater degree during the daytime, use a more standard language; (*ii*) the southern part of the country is more prone to use more standard language than the northern one, while locally the used variety or dialect is determined by the spatial distribution of socioeconomic status; and (*iii*) individuals connected in the social network are closer linguistically than disconnected ones, even after the effects of status homophily have been removed. Our results inform sociolinguistic theory and may inspire novel learning methods enabling the inference of socioeconomic status of people from the way they tweet.

7.3.5. Relational methods for media studies

A very relevant application of the research that DANTE carries out on networks structures and networks dynamics concerns the field of journalism and media study. Relational analysis may be helpful in these fields in two different way.

On the one hand, the advent of digital media has challenged the established vertical structure of information distribution typical of broadcasting media with a decentralised organisation that facilitates the spreading of contents through all sort of horizontal channels (in the Web and in Social Media). This new type of circulation

is still insufficiently studied and require both quantitative and qualitative investigation. We tried to provide the first in our Field Guide to Fake News already introduced in the highlights of this document [68] and in a forthcoming chapter on the heterogeneous clustering of French Media system for the The Routledge Handbook to Developments in Digital Journalism Studies [60]. As for the qualitative study of the structure of the media system, we published and analysis of the strategies employed by Facebook to steer the evolution of the technology of Live Video Streaming [23].

On the other hand, network analysis can be a powerful tool to investigate and narrate journalistic stories, but its techniques need adapted to the language used by journalists and understood by their audiences. We tried to provide such a translation in a paper for the journal Digital Journalism [13] and in a chapter for the Datafied Society book [59].

The use of network analysis to study vast societal phenomena has also profound implications for the theory of social sciences, which we tried to explore in a paper for the journal Big Data & Society [25] and in a chapter of a book on the Frontiers of Social Science [63], and for their practice [62].

7.3.6. Philosophy of technologies revisited by Internet

The Internet, as a technology of writing, helps us to understand that a technology is not always a mean to reach a goal, nor an application of science. In fact, the Internet does not appear as a revolution, but as a revealer. We understand that a technology can be reflexive (it invites us to think it) and that it cannot be clearly separated from human activities (writing, etc.). For instance, 50 years ago, we imagined we could think with our own mind (and perhaps with a paper and pencil). Now, we know that we cannot think without material stuff (a computer, the internet, etc.). A very few philosophers knew this fact (Leibniz, Boole, etc.). But this evidence transforms completely the philosophy of technologies. Another important point is the effect of technology is not there (eg: social cartography or statistics). This fact invites us to insert technologies and methods in the traditional diptych of theory and experience. In synthesis, we also discover strong links between technology and culture; hence the role of engineers in the construction of culture [53], [18], [52], [54], [71].

DATAMOVE Project-Team

7. New Results

7.1. Integration of High Performance Computing and Data Analytics

New results on the topic *Integration of High Performance Computing and Data Analytics* are related to compression [15], automatic data extraction for in situ processing [14], in transit sensitivity analysis [16] and management of heterogeneous HPC and BigData workloads [21]. We detail the two last here.

- Large Scale In Transit Sensitivity Analysis Avoiding Intermediate Files [16]. Global sensitivity analysis is an important step for analyzing and validating numerical simulations. One classical approach consists in computing statistics on the outputs from well-chosen multiple simulation runs. Simulation results are stored to disk and statistics are computed postmortem. Even if supercomputers enable to run large studies, scientists are constrained to run low resolution simulations with a limited number of probes to keep the amount of intermediate storage manageable. In this paper we propose a file avoiding, adaptive, fault tolerant and elastic framework that enables high resolution global sensitivity analysis at large scale. Our approach combines iterative statistics are updated on-the-fly as soon as the in transit parallel server receives results from one of the running simulations. For one experiment, we computed the Sobol' indices on 10M hexahedra and 100 timesteps, running 8000 parallel simulations executed in 1h27 on up to 28672 cores, avoiding 48TB of file storage. Based on this work we open sourced the associated framework called Melissa (https://melissa-sa.github.io).
- Big Data and HPC collocation: Using HPC idle resources for Big Data Analytics [21]. Executing • Big Data workloads upon High Performance Computing (HPC) infrastractures has become an attractive way to improve their performances. However, the collocation of HPC and Big Data workloads is not an easy task, mainly because of their core concepts' differences. This paper focuses on the challenges related to the scheduling of both Big Data and HPC workloads on the same computing platform. In classic HPC workloads, the rigidity of jobs tends to create holes in the schedule: we can use those idle resources as a dynamic pool for Big Data workloads. We propose a new idea based on Resource and Job Management System's (RJMS) configuration, that makes HPC and Big Data systems to communicate through a simple prolog/epilog mechanism. It leverages the built-in resilience of Big Data frameworks, while minimizing the disturbance on HPC workloads. We present the first study of this approach, using the production RJMS middleware OAR and Hadoop YARN from the HPC and Big Data ecosystems respectively. Our new technique is evaluated with real experiments upon the Grid5000 platform. Our experiments validate our assumptions and show promising results. The system is capable of running an HPC workload with 70% cluster utilization, with a Big Data workload that fills the schedule holes to reach a full 100% utilization. We observe a penalty on the mean waiting time for HPC jobs of less than 17% and a Big Data effectiveness of more than 68% in average.

7.2. Data Aware Batch Scheduling

New results on the topic *Data Aware Batch Scheduling* are related to graph algorithm for dense k-subset detection [8], scheduling heuristic for multi-CPU multi-GPU computing platform with performance guarantee [9], machine learning for designing scheduling policies [11] and multi-objective scheduling heuristic [13]. We detail the two last here.

• Obtaining Dynamic Scheduling Policies with Simulation and Machine Learning [11]. Dynamic scheduling of tasks in large-scale HPC platforms is normally accomplished using ad-hoc heuristics, based on task characteristics, combined with some backfilling strategy. Defining heuristics that work efficiently in different scenarios is a difficult task, specially when considering the large variety of

task types and platform architectures. In this work, we present a methodology based on simulation and machine learning to obtain dynamic scheduling policies. Using simulations and a workload generation model, we can determine the characteristics of tasks that lead to a reduction in the mean slowdown of tasks in an execution queue. Modeling these characteristics using a nonlinear function and applying this function to select the next task to execute in a queue dramatically improved the mean task slowdown in synthetic workloads. When applied to real workload traces from highly different machines, these functions still resulted in important performance improvements, attesting the generalization capability of the obtained heuristics.

• A new on-line method for scheduling independent tasks [13]. We present a new method for scheduling independent tasks on a parallel machine composed of identical processors. This problem has been studied extensively for a long time with many variants. We are interested here in designing a generic algorithm in the on-line non-preemptive setting whose performance is good for various objectives. The basic idea of this algorithm is to detect some problematic tasks that are responsible for the delay of other shorter tasks. Then the former tasks are redirected to be executed in a dedicated part of the machine. We show through an extensive experimental campaign that this method is effective and in most cases is closer to some standard lower bounds than the base-line method for the problem.

DATASPHERE Team

7. New Results

7.1. Political economy

We pursued our work on digital platforms and their impact on the structure of socio-economic systems, which results from the capacity to separate data or information from the actors of the physical world. In [14], we showed how the movement above ground of the intermediation activity transforms territories. We developed further this idea in [5] to suggest that a new political grammar is necessary to understand the relationships between the actors.

7.2. Anthropocene

In [6], we investigated the possible similarities between biological systems and social systems facing shortage of resources, suggesting that the digital revolution might have something to do with the Anthropocene. The timing of the digital revolution was further investigated in [7], to further analyse the relationships between the two transformation affecting the contemporary period. An investigation of the world of images and photography in the time of algorithms was conducted in [4].

7.3. Laws and digital

The emergence of digital services affects the legal system. The law is always associated to a territory, while digital systems act remotely over large regions crossing borders to reach the population, imposing new norms. In [3], [2], we suggest that a new framework is necessary to apprehend new phenomena, such as the those resulting from the conflicts between global search engines and local rules with respect to the Right to be forgotten for instance.

7.4. Network data analytics

In collaboration with the Chinese Academy of Sciences, we worked on packet processing algorithmic for high speed network measurements. In [9] a packet capture archive system is developed and described. In [8] a theoretical analysis of the TCAM updates delay that is the main shortcoming of TCAM usage in high speed packet processors is presented. Quality of service for network functions were considered in [13].

7.5. Data protection

We developed new mechanisms to process aggregation while preserving the confidentiality of the initial data in the framework of Robert Riemann's thesis [1]. The benefits of distributed protocols for online voting was considered [12]. A distributed aggregation mechanisms preserving confidentiality of data based on Kademlia was proposed in [11]. Applications of the previous algorithms to lotteries was considered in [10].

DRACULA Project-Team

5. New Results

5.1. Cancer

5.1.1. Long-term treatment effects in chronic myeloid leukemia

We propose and analyze in [12] a simplified version of a partial differential equation (PDE) model for chronic myeloid leukemia (CML) derived from an agent-based model proposed by Roeder *et al.* in [38]. This model describes the proliferation and differentiation of leukemic stem cells in the bone marrow and the effect of the drug Imatinib on these cells. We first simplify the PDE model by noting that most of the dynamics occurs in a subspace of the original 2D state space. Then we determine the dominant eigenvalue of the corresponding linearized system that controls the long-term behavior of solutions. We mathematically show a non-monotonous dependence of the dominant eigenvalue with respect to treatment dose, with the existence of a unique minimal negative eigen-value. In terms of CML treatment, this shows that there is a unique dose that maximizes the docay rate of the CML tumor load decay. Numerical simulations of the full model confirm that this phenomenon is not an artifact of the simplification. Therefore, while optimal asymptotic dosage might not be the best one at short time scales, our results raise interesting perspectives in terms of strategies for achieving and improving long-term deep response.

5.1.2. A model of interaction between the immune system and cancer cells in chronic myelogenous leukemia

We describe in [11] a simple model for the interaction between leukemic cells and the autologous immune response in chronic phase chronic myelogenous leukemia (CML). This model is a simplified version of the model we proposed in 2015 ([34]). Our simplification is based on the observation that certain key characteristics of the dynamics of CML can be captured with a three compartments model: two for the leukemic cells (stem cells and mature cells) and one for the immune response. We characterize the existence of steady states and their stability for generic forms of immunosuppres-sive effects of leukemic cells. We provide a complete co-dimension one bifurcation analysis. Our results show how clinical response to tyrosine kinase inhibitors treatment is compatible with the existence of a stable low-disease, treatment-free steady state.

5.1.3. A hybrid computation model to describe the progression of multiple myeloma and its intra-clonal heterogeneity

Multiple myeloma (MM) is a genetically complex hematological cancer that is characterized by proliferation of malignant plasma cells in the bone marrow. MM evolves from the clonal premalignant disorder monoclonal gammopathy of unknown significance (MGUS) by sequential genetic changes involving many different genes, resulting in dysregulated growth of multiple clones of plasma cells. The migration, survival, and proliferation of these clones require the direct and indirect interactions with the non-hematopoietic cells of the bone marrow. We develop in [14], a hybrid discrete-continuous model of MM development from the MGUS stage. The discrete aspect of the model is observed at the cellular level: cells are represented as individual objects which move, interact, divide, and die by apoptosis. Each of these actions is regulated by intracellular and extracellular processes as described by continuous models. The hybrid model consists of the following submodels that have been simplified from the much more complex state of evolving MM: cell motion due to chemotaxis, intracellular regulation of plasma cells, extracellular regulation in the bone marrow, and acquisition of mutations upon cell division. By extending a previous, simpler model in which the extracellular matrix was considered to be uniformly distributed, the new hybrid model provides a more accurate description in which cytokines are produced by the marrow microenvironment and consumed by the

myeloma cells. The complex multiple genetic changes in MM cells and the numerous cell-cell and cytokinemediated interactions between myeloma cells and their marrow microenviroment are simplified in the model such that four related but evolving MM clones can be studied as they compete for dominance in the setting of intraclonal heterogeneity.

5.1.4. Fast and binary assay for predicting radiosensitivity based on the nucleoshuttling of ATM protein: development, validation and performances

The societal and clinical impact of post-radiotherapy adverse tissue events (AE) has highlighted the need of molecular parameters to predict individual radiosensitivity. Recent studies have stressed the role of the phosphorylated forms of the ATM protein (pATM) and its nucleoshuttling in response to radiation. The statistical performance of the pATM immunofluorescence assay to predict AE is promising. However, immunofluorescence requires a time-consuming amplification of cells. The purpose of the study in [24] was to develop a predictive assay based on the ELISA technique that renders faster the previous approach. Materials and methods This study was performed on 30 skin fibroblasts from 9 radioresistant and 21 AE patients. Patients were divided in 2 groups, radioresistant (toxicity grade < 2) and radiosensitive (toxicity grade ≥ 2). The quantity of nuclear pATM molecules was assessed by ELISA method at 10 min and 1 h after 2 Gy and compared to pATM immunofluorescence data. Results The pATM ELISA data were found in quantitative agreement with the immunofluorescence ones. A ROC analysis was applied first to two data sets (a training (n = 14) and a validating (n = 16) one) and thereafter to the whole data with a 2-fold cross-validation method. The assay showed an AUC value higher than 0.8, a sensitivity of 0.8 and a specificity ranging from 0.75 and 1, which strongly document the predictive power of the pATM ELISA assay. Conclusion This study showed that the assessment of nuclear pATM quantity after 2 Gy via ELISA technique can be the basis of a predictive assay with the highest statistical performance among the available predictive approaches.

5.2. Immune response

5.2.1. Identification of nascent memory CD8 T cells and modeling of their ontogeny

Primary immune responses generate short-term effectors and long-term protective memory cells. The delineation of the genealogy linking naive, effector and memory cells has been complicated by the lack of phenotypes discriminating effector from memory differentiation stages. Using transcriptomics and phenotypic analyses, we identify (see [17]) a novel marker combination that allows us to track nascent memory cells within the effector phase. We then use a formal approach based on mathematical models describing the dynamics of population-size evolutions to test potential progeny links and demonstrate that most cells follow a linear naive-early effector-late effector-memory pathway. Moreover, our mathematical model allows long-term prediction of memory cell numbers from a few early experimental measurements. Our work thus provides a phenotypic means to identify effector and memory cells, as well as a mathematical framework to investigate the ontology of their generation and to predict the outcome of immunization regimens in terms of memory cell numbers generated.

5.2.2. Modelling the dynamics of virus infection and immune response in space and time

Spreading of viral infection in the tissues such as lymph nodes or spleen depends on virus multiplication in the host cells, their transport and on the immune response. Reaction–diffusion systems of equations with delays in proliferation and death terms of the immune cells represent an appropriate model to study this process. The properties of the immune response and the initial viral load determine the regimes of infection spreading. In the proposed model [13], the proliferation rate of the immune cells is represented by a bell-shaped function of the virus concentration which increases for small concentrations and decreases if the concentration is sufficiently high. We use such a model system to show that an infection can be completely eliminated or it can remain present together with a decreased concentration of immune cells. Finally, immune cells can be completely exhausted leading to a high virus concentration in the tissue. In addition, we predicted two novel regimes of infection dynamics not observed before. Infection propagation in the tissue can occur as a superposition of two travelling waves: first wave propagates as a low level infection front followed by a high level infection

front with a smaller speed of propagation. Both of the travelling waves can have a positive or a negative speed corresponding to infection advancement or retreat. These regimes can be accompanied by instabilities and the emergence of complex spatiotemporal patterns.

5.2.3. Estimates and impact of lymphocyte division parameters from CFSE data using mathematical modeling

Carboxyfluorescein diacetate succinimidyl ester (CFSE) labelling has been widely used to track and study cell proliferation. In [23], we use mathematical modeling to describe the kinetics of immune cell proliferation after an in vitro polyclonal stimulation tracked with CFSE. This approach allows us to estimate a set of key parameters, including ones related to cell death and proliferation. We develop a three-phase model that distinguishes a latency phase, accounting for non-divided cell behaviour, a resting phase and the active phase of the division process. Parameter estimates are derived from model results, and numerical simulations are then compared to the dynamics of in vitro experiments, with different biological assumptions tested. Our model allows us to compare the dynamics of CD4+ and CD8+ cells, and to highlight their kinetic differences. Finally we perform a sensitivity analysis to quantify the impact of each parameter on proliferation kinetics. Interestingly, we find that parameter sensitivity varies with time and with cell generation. Our approach can help biologists to understand cell proliferation mechanisms and to identify potential pathological division processes.

5.3. Erythropoiesis

5.3.1. Investigating the role of the experimental protocol in phenylhydrazine-induced anemia on mice recovery

Production of red blood cells involves growth-factor mediated regulation of erythroid progenitor apoptosis and self-renewal. During severe anemia, characterized by a strong fall of the hematocrit followed by a recovery phase, these controls allow a fast recovery of the hematocrit and survival of the organism. Using a mathematical model of stress erythropoiesis and an ad hoc numerical method, we investigate in [9] the respective roles of anemia-inducing phenylhydrazine injections and physiological regulation on the organism's recovery. By explicitly modeling the experimental protocol, we show that it mostly characterizes the fall of the hematocrit following the anemia and its severeness, while physiological process regulation mainly controls the recovery. We confront our model and our conclusions to similar experiments inducing anemia and show the model's ability to reproduce several protocols of phenylhydrazine-induced anemia. In particular, we establish a link between phenylhydrazine effect and the severeness of the anemia.

5.3.2. Numerical integration of an erythropoiesis model with explicit growth factor dynamics

Erythropoiesis, the red blood cell production process, involves interactions between cell populations with different differentiation states, mainly immature progenitor cells and mature erythrocytes, and growth factors such as erythropoietin and glucocorticoids, known to respectively inhibit cell apoptosis, stimulate proliferation and differentiation, and stimulate self-renewal. The feedback regulation of this process allows a very fast and efficient recovery in the case of a severe anemia. We consider in [8] an age-structured model of red blood cell production accounting for these feedback regulations and the dynamics of growth factors. We theoretically show the existence of a unique positive steady state for the model and we propose a numerical method to obtain an approximation to its solution. Experiments are reported to show numerically, on one hand, the optimal convergence order of the numerical scheme and, on the other hand, a fine approximation to real experimental data, with a suitable selection of the parameters involved.

5.4. Methodological developments

5.4.1. Traveling waves for a model of hematopoiesis

The formation and development of blood cells (hematopoiesis) is a very complex process. This process involves a small population of cells called hematopoietic stem cells (HSCs). The HSCs are undifferentiated

cells, located in the bone marrow before they become mature blood cells and enter the blood stream. They have a unique ability to produce either similar cells (self-renewal), or cells engaged in one of different lineages of blood cells: red blood cells, white cells and platelets (differentiation). The HSCs can be either in a proliferating or in a quiescent phase. In [6], we distinguish between dividing cells that enter directly to the quiescent phase and dividing cells that return to the proliferating phase to divide again. We propose a mathematical model describing the dynamics of HSC population, taking into account their spatial distribution. The resulting model is an age-structured reaction-diffusion system. The method of characteristics reduces this model to a coupled reaction-diffusion equation and difference equation with delay. We study the existence of traveling wave fronts connecting the zero steady state with the unique positive uniform one. We use a monotone iteration technique coupled with the upper and lower solutions method.

5.4.2. A hybrid finite volume method for advection equations and its applications in population dynamics

We present in [30] a very adapted finite volume numerical scheme for transport type-equation. The scheme is an hybrid one combining an anti-dissipative method with down-winding approach for the flux ([35]; [36]) and an high accurate method as the WENO5 one ([37]). The main goal is to construct a scheme able to capture in exact way the numerical solution of transport type-equation without artifact like numerical diffusion or without "stairs" like oscillations and this for any regular or discontinuous initial distribution. This kind of numerical hybrid scheme is very suitable when properties on the long term asymptotic behavior of the solution are of central importance in the modeling what is often the case in context of population dynamics where the final distribution of the considered population and its mass preservation relation are required for prediction.

5.4.3. Inferring gene regulatory networks from single-cell data: a mechanistic approach

The recent development of single-cell transcriptomics has enabled gene expression to be measured in individual cells instead of being population-averaged. Despite this considerable precision improvement, inferring regulatory networks remains challenging because stochasticity now proves to play a fundamental role in gene expression. In particular, mRNA synthesis is now acknowledged to occur in a highly bursty manner. We propose in [21] to view the inference problem as a fitting procedure for a mechanistic gene network model that is inherently stochastic and takes not only protein, but also mRNA levels into account. We first explain how to build and simulate this network model based upon the coupling of genes that are described as piecewise-deterministic Markov processes. Our model is modular and can be used to implement various biochemical hypotheses including causal interactions between genes. However, a naive fitting procedure would be intractable. By performing a relevant approximation of the stationary distribution, we derive a tractable procedure that corresponds to a statistical hidden Markov model with interpretable parameters. This approximation turns out to be extremely close to the theoretical distribution in the case of a simple toggleswitch, and we show that it can indeed fit real single-cell data. As a first step toward inference, our approach was applied to a number of simple two-gene networks simulated in silico from the mechanistic model and satisfactorily recovered the original networks. Our results demonstrate that functional interactions between genes can be inferred from the distribution of a mechanistic, dynamical stochastic model that is able to describe gene expression in individual cells. This approach seems promising in relation to the current explosion of single-cell expression data.

5.5. Physiology

5.5.1. A multiscale modeling approach for the regulation of the cell cycle by the circadian clock

We present in [18] a multiscale mathematical model for the regulation of the cell cycle by the circadian clock. Biologically, the model describes the proliferation of a population of heterogeneous cells connected to each other. The model consists of a high dimensional transport equation structured by molecular contents of the cell cycle-circadian clock coupled oscillator. We propose a computational method for resolution adapted from the concept of particle methods. We study the impact of molecular dynamics on cell proliferation and show an example where discordance of division rhythms between population and single cell levels is observed. This highlights the importance of multiscale modeling where such results cannot be inferred from considering solely one biological level.

5.5.2. The lifespan and turnover of microglia in the human brain

The hematopoietic system seeds the CNS with microglial progenitor cells during the fetal period, but the subsequent cell generation dynamics and maintenance of this population have been poorly understood. We report in [25] that microglia, unlike most other hematopoietic lineages, renew slowly at a median rate of 28% per year, and some microglia last for more than two decades. Furthermore, we find no evidence for the existence of a substantial population of quiescent long-lived cells, meaning that the microglia population in the human brain is sustained by continuous slow turnover throughout adult life.

5.5.3. Impact of fat mass and distribution on lipid turnover in human adipose tissue

Differences in white adipose tissue (WAT) lipid turnover between the visceral (vWAT) and subcutaneous (sWAT) depots may cause metabolic complications in obesity. In [26], we compare triglyceride age and, thereby, triglyceride turnover in vWAT and sWAT biopsies from 346 individuals and find that subcutaneous triglyceride age and storage capacity are increased in overweight or obese individuals. Visceral triglyceride age is only increased in excessively obese individuals and associated with a lower lipid removal capacity. Thus, although triglyceride storage capacity in sWAT is higher than in vWAT, the former plateaus at substantially lower levels of excess WAT mass than vWAT. In individuals with central or visceral obesity, lipid turnover is selectively increased in vWAT. Obese individuals classified as 'metabolically unhealthy' (according to ATPIII criteria) who have small subcutaneous adipocytes exhibit reduced triglyceride turnover. We conclude that excess WAT results in depot-specific differences in lipid turnover and increased turnover in vWAT and/or decreased turnover in sWAT may result in metabolic complications of overweight or obesity.

ERABLE Project-Team

6. New Results

6.1. General comments

We present in this section the main results obtained in 2017.

We tried to organise these following the six main axes of research of the team. Clearly, in some cases, a result obtained overlaps more than one axis. In such case, we chose the one that could be seen as the main one concerned by such results.

We did not indicate here the results on more theoretical aspects of computer science if it did not seem for now that they could be relevant in contexts related to computational biology. Actually, we do believe those on scheduling [25], [24], and on text [37], graph [4], [32], [34], [5], [36], [35] or general algorithmic problems notably related to performance issues [23], [28] could in the future become more specifically relevant for life sciences (biology or ecology). We did not indicate either work that was done a few years ago by members who were in ERABLE but whose associated publication appeared only this year [27].

Notice that the theoretical results related to problems closely resembling questions that have already been addressed by us in computational biology and that we present below concern not only cross-fertilising issues among different computational approaches, and we therefore extended the title of this axis for the purpose of presenting such results, for now purely theoretical.

A few other results of 2017 are not mentioned in this report, not because the corresponding work is not important, but because it was likewise more specialised, or the work represented a survey.

6.2. Identifying the molecular elements

Motif tries for pattern discovery. In [14], the motif trie data structure was introduced to improve the extraction of recurring patterns in sequences. Such extraction concerned maximal patterns with at most k don't care symbols and at least q occurrences, according to a given maximality notion. The motif trie was applied to this problem, also showing how to build it efficiently. This led to the first algorithm that attains a stronger notion of output-sensitivity, where the cost for an input sequence of n symbols is proportional to the actual number of occurrences of each pattern, which is at most n (much smaller in practice). This avoids the best-known cost of O(nc)O(nc) per pattern, for a constant c > 1, which is otherwise impractical for massive sequences with a large value of n.

Identification of genome and alternative splicing variants in RNA-seq data. The team's work on identifying alternative splicing and other genome variants such as SNPs (Single Nucleotide Polymorphism), indels, etc., started around 2010. This has concerned mostly RNA-seq data also for the variants investigated.

Both DNA and RNA-seq data analysis using so-called NGS (Next Generation Sequencing) is a domain of research that has been active for decades now, with many open questions remaining despite such long and intense activity. One is the case of non-model organisms, but actually there is another major problem that has not been solved, at least in any really satisfying way since the premises of genome sequencing. This is the problem of repeats. Notice however that repeats are not just "problems to be avoided", but have a strong biological interest in themselves, notably those related to transposable elements. Various papers of the team in 2017, notably [13], [19], [22], [1], were concerned with the study of such elements.

As concerns non-model organisms, the team extended a method it had previously developed, called KISS-PLICE, to identify, quantify and annotate SNPs without any reference genome, using RNA-seq data only. The paper (Lopez-Maestre *et al.*, *Nucleic Acids Research*, 44(19):e148, 2016) appeared at the end of 2016. There we showed that individuals can be pooled prior to sequencing if not enough material is available from one individual. Using pooled human RNA-seq data, we clarified the precision and the recall of our method and discussed them with respect to others which use a reference genome or an assembled transcriptome. We then validated experimentally the predictions of our method using RNA-seq data from two non-model species. The method can be used for any species to annotate SNPs and predict their impact on the protein sequence. It enables to test for the association of the identified SNPs with a phenotype of interest. One of the phenotypes explored was related to the dependence of the insect *Asobara tabida* on its endosymbiont *Wolbachia*.

The methodological part of the work above relied in part on a number of more theoretical results, related to algorithmics and more specifically focused on the problem of repeats [21]. The most theoretical recent work of the team, accepted at the 43rd International Workshop on Graph-Theoretic Concepts in Computer Science (WG) in 2017 [30], proposed the notion of a bubble generator set, *i.e.* of a polynomial-sized subset of bubbles from which all the others can be obtained, also in polynomial time, through the application of a specific symmetric difference operator. This is further described in the last axis (Axis 6).

Genome and haplotype assembly. Fully assembling the genome sequence of an organism remains an important and challenging task. Genome scaffolding (*i.e.* the process of ordering and orientating contigs) of *de novo* assemblies usually represents the first step in most genome finishing pipelines. The team started by developing an algorithm (called MEDUSA) for such task (Bosi *et al.*, *Bioinformatics*, 31(15):2443-2451, 2015). It exploited information obtained from a set of (draft or closed) genomes from related organisms to determine the correct order and orientation of the contigs. It formalised the scaffolding problem by means of a combinatorial optimisation formulation on graphs and implements an efficient constant factor approximation algorithm to solve it. In contrast to the majority of the scaffolders, it did not require either prior knowledge on the input dataset (usually of micro-organisms) or the availability of paired-end read libraries. MEDUSA however presented limitations both in the construction of the scaffolding graph for large genomes, and in the subsequent assembly. The first aspect has been recently greatly improved by a method developed in collaboration with researchers (among which Alex di Genova) from Chile. This work led to the software FAST-SG already publicly available, and to a first publication that is in revision.

6.3. Inferring and analysing the networks of molecular elements

Metabolic impact of a change of conditions. The increasing availability of metabolomics data enables to better understand the metabolic processes involved in the immediate response of an organism to environmental changes, where the latter can be related to the presence of other species. The data usually come in the form of a list of metabolites whose concentrations significantly changed under some conditions, and are thus not easy to interpret without being able to precisely infer how such metabolites are interconnected. The team introduced a method that enables to organise the data from any metabolomics experiment into what we initially called *metabolic stories* when we were working with a simpler, graph representation of metabolism, and which have now become *metabolic hyperstories* as more accurate directed hypergraphs representations are considered. Each (hyper)story corresponds to a possible scenario explaining the flow of matter between the metabolites of interest. The initial work on a graph representation led to the GOBBOLINO + TOUCHÉ software (Milre *et al., Bioinformatics*, 30(1):61-70, 2014). Two newer works working with directed hypergraphs were presented in the PhD of Alice Julien-Laferrière (defended in 2016). Two papers are currently in preparation. They led to the software TOTORO (which uses a qualitative measurement of concentrations in two steady-states) and KOTOURA (which infers quantitative changes of the reactions) which are both already publicly available.

Metabolic network reconstruction and comparison for understanding virulence. The respiratory tract of swines is colonised by several bacteria among which are three Mycoplasma species: *Mycoplasma flocculare*, *Mycoplasma hyopneumoniae* and *Mycoplasma hyorhinis*. While colonisation by *M. flocculare* is virtually asymptomatic, *M. hyopneumoniae* is the causative agent of enzootic pneumonia and *M. hyorhinis* is present in cases of pneumonia, polyserositis and arthritis. The genomic resemblance among these three Mycoplasma species combined with their different levels of pathogenicity is an indication that they have unknown mechanisms of virulence and differential expression, as for most mycoplasmas. We performed whole-genome metabolic network reconstructions for the three mycoplasmas, as well as cultivation tests and metabolomic experiments through nuclear magnetic resonance spectroscopy (NMR) (Ferrarini *et al., BMC Genomics*, 17(1):353, 2016). We were able to infer from such reconstructed networks that the lack of pathogenicity of *M. flocculare* if compared to the highly pathogenic *M. hyopneumoniae* may be related to its incapacity to produce cytotoxic hydrogen peroxide. A second, more experimentally oriented-paper is currently under revision.

6.4. Modelling and analysing a network of individuals, or a network of individuals' networks

On unrooted and root-uncertain variants of several well-known phylogenetic network problems Genetic hybridisation is the process individuals from genetically distinct populations that are able to interbreed and this produce a hybrid.

The hybridisation number problem refers to finding the minimum number of hybridisation events necessary to explain conflicts among several evolutionary trees. It requires to embed a set of binary rooted phylogenetic trees into a binary rooted phylogenetic network such that the number of nodes with in-degree two is minimised. However, from a biological point of view accurately inferring the root location in a phylogenetic tree is notoriously difficult and poor root placement can artificially inflate the hybridisation number. To this end, a number of relaxed variants of this problem were studied in [29]. We started by showing that the fundamental problem of determining whether an unrooted phylogenetic network displays (*i.e.* embeds) an unrooted phylogenetic tree, is NP-hard. On the positive side, we showed that this problem is FPT in reticulation number. In the rooted case, the corresponding FPT result is trivial, but here a more subtle argumentation was required. Next, we showed that the hybridisation number problem for unrooted networks (when given two unrooted trees) is equivalent to the problem of computing the tree bisection and reconnect distance of the two unrooted trees. We then considered the "root uncertain" variant of the hybridisation number. Here we are free to choose the root location in each of a set of unrooted input trees such that the hybridisation number of the resulting rooted trees is minimised. On the negative side, we showed that this problem is APX-hard. On the positive side, we showed that it is FPT in the hybridisation number, via kernelisation, for any number of input trees.

Phylogenetic tree reconciliation. Phylogenetic tree reconciliation consists in a mapping of one tree (usually the symbiont tree) to the other (the host tree) using event-based maximum parsimony. Given a cost model for the events, many optimal reconciliations are however possible. Any further biological interpretation of them must therefore take this into account, making the capacity to enumerate all optimal solutions a crucial point. Indeed, the problem is not just that if we proposed a single solution, there is a good chance we would miss the "true" answer, but also that we would lose the capacity to verify whether there exist some characteristics that are common to enough of the solutions to increase our confidence in the "story" such reconciliation tells of the past.

When the ERABLE team started addressing this issue, only two algorithms existed that attempted such enumeration; in one case (software CORE-PA) not all possible solutions were produced while in the other (software NOTUNG) not all cost vectors were handled. We then introduced a polynomial-delay algorithm, called EUCALYPT, for enumerating all optimal reconciliations, and showed that in general many solutions exist (Donati *et al., Algorithms for Molecular Biology*, 10(1):11, 2015). Some might not be time-feasible. However, we further showed that, among the many solutions that are usually found, in the majority of the cases, at least some will be time-feasible, and we provided a polynomial algorithm to test for time-feasibility.

We also considered a restricted version of the model where host switches are allowed to happen only between species that are within some fixed distance along the host tree. This restriction allows to reduce the number of time-feasible solutions while preserving the same optimal cost, as well as to find time-feasible solutions with a cost close to the optimal in the cases where no time-feasible optimal solution is found.

More recently, we defined two equivalence relations that enable to identify many reconciliations with a single one, thereby reducing their number. These results were published in a paper which was accepted at CIBB 2017 and will appear in the *LNCI-LNCS* proceedings of the conference (published after CIBB). Extensive experiments indicated that the number of output solutions greatly decreases in general. By how much clearly depends on the constraints that are given as input. An extended journal version of this work that includes its theoretical part will be submitted at the beginning of 2018. Other forms of grouping (or clustering) solutions are also being explored that rely instead on defining a distance between two different reconciliations. Two approaches are being investigated, one in collaboration with a researcher in Italy (paper in preparation), and the other with researchers in the UK (one paper submitted and one in preparation).

Improving the biological realism of coevolutionary models. The host-symbiont coevolutionary models developed so far needed also to be improved. The realism we wished to add to such models was for now the possibility to handle the case of multiple associations of a symbiont. Among the few previous works that allowed for this, all presented some limitation either in terms of the model or of the algorithm developed. Handling such multiple associations requires to introduce an event that was little or not formally considered in the literature. This is the event of spread, which precisely corresponds to the invasion of different hosts by a same symbiont. In this case, as when spreads are not considered, the optimal reconciliations obtained will depend on the choice made for the costs of the events. The need to develop statistical methods to assign the most appropriate ones therefore remained also of actuality. This is one of the problems we addressed in the PhD of Laura Urbini that was defended in October 2017. Two types of spread were in fact introduced: vertical and horizontal. The first corresponds to the case where the evolution of the symbiont "freezes" while the symbiont continues to be associated with a host and with the new species that descend from this host. The second includes both an invasion, of the symbiont which remains with the initial host but at the same time gets associated with ("invades") another one incomparable with the first, and a double freeze (in relation to the evolution of the host with which it was initially associated and in relation to the evolution of the second one it "invaded"). Two papers addressing distinct aspects related to the spread problem with different approaches are in preparation and will be submitted before the end of 2017 or beginning of 2018.

Estimating the frequency and expansion process of an infection We addressed the question of how often an infection occurs and of whether its expansion reached an equilibrium using as model *Wolbachia*. *Wolbachia* is a bacterial genus that infects about half of all arthropods, with diverse and extreme consequences ranging from sex-ratio distortion and mating incompatibilities to protection against viruses. These phenotypic effects, combined with efficient vertical transmission from mothers to offspring, satisfactorily explain the invasion dynamics of *Wolbachia* within species. However, beyond the species level, the lack of congruence between the host and symbiont phylogenetic trees indicates that *Wolbachia* horizontal transfers and extinctions do happen and underlie its global distribution.

In [3], we inferred recent acquisition/loss events from the distribution of *Wolbachia* lineages across the mitochondrial DNA tree of 3600 arthropod specimens, spanning 1100 species from Tahiti and the surrounding islands. We showed that most events occurred within the last million years, but are likely attributable to individual level variation (*e.g.*, imperfect maternal transmission) rather than to population level variation (*e.g.*, *Wolbachia* extinction). At the population level, we estimated that mitochondria typically accumulate 4.7% substitutions per site during an infected episode, and 7.1% substitutions per site during the uninfected phase. Using a Bayesian time calibration of the mitochondrial tree, these numbers translate into infected and uninfected phases of approximately 7 and 9 million years. Infected species thus lose *Wolbachia* slightly more often than uninfected species acquire it, supporting the view that its present incidence, estimated here slightly below 0.5, represents an epidemiological equilibrium.

6.5. Going towards control

Quantitative synthetic biology. Synthetic biology has boomed since the early 2000s when it started being shown that it was possible to efficiently synthetise compounds of interest in a much more rapid and effective way by using other organisms than those naturally producing them. However, to thus engineer a single organism, often a microbe, to optimise one or a collection of metabolic tasks may lead to difficulties when attempting to obtain a production system that is efficient, or to avoid toxic effects for the recruited microorganism. The idea of using instead a microbial consortium has thus started being developed in the last decade. Establishing which consortium is best for the production of a given compound or set thereof remains however a great challenge. The team introduced an initial model and a method, called MULTIPUS, that enable to propose a consortium to synthetically produce compounds that are either exogenous to it, or are endogenous but where interaction among the species in the consortium could improve the production line (Julien-Laferrière *et al., Scientific Reports*, 6, 2016).

Since the work on MULTIPUS, the team has been considering quantitative approaches for synthetic biology. We thus explored the concept of multi-objective optimisation in the field of metabolic engineering when both continuous and integer decision variables are involved in the model. In particular, we proposed multi-objective models, initially for a single species, to suggest reaction deletion strategies, and also to deal with situations where several functions must be optimised simultaneously, such as the maximisation of bioproducts while minimising toxicity (Hartmann *et al.*, *BMC Systems Biology*, see https://www.ncbi.nlm. nih.gov/pubmed/29268790, just accepted and not yet visible in Hal-Inria). We compared our results with those obtained by using the well-known bi-level optimisation model of OPTKNOCK, and studied two multi-objective optimisation problems arising from the metabolic engineering of microorganisms. One of them, using Yeast, has been validated experimentally. The work is submitted. The team has then started expanding it to communities (Master Thesis of Irene Ziska who is continuing into a PhD).

6.6. Health

Rare Diseases. Splicing is an essential step in the process leading to gene expression because it not only removes the introns from the primary transcripts, but also generates a combination of mature transcripts through the differential inclusion/exclusion of exons and sometimes retention of introns. Some pathologies are associated to such abnormal splicing. This is the case of the Taybi-Linder Syndrome (TALS), a very rare malformative syndrome with autosomal recessive transmission, belonging to the group of microcephalic dwarfism and responsible for death usually before the age of 2 years. This pathology was recently found to be caused by mutations in RNU4ATAC, a small nuclear RNA, which is an essential component of the minor spliceosome. We started a collaboration with the group of Pr. P. Edery (who first identified this alteration in 2012) with the objective to establish a comprehensive catalog of splice alterations in several cohorts of TALS patients. In this work, we take advantage of our reference-free assembly approach of transcripts (KISSPLICE) in order to detect new splicing alterations and to identify the associated deregulated signalling genes and pathways.

Cancer. Alain Viari has continued to develop a strong interaction with clinicians concerned with cancer, notably of the breast and in the early human embryo. A number of papers have appeared in 2017 that describe this work [7], [6], [10], [11], [12], [16], [17], [18]. We highlight here just two.

The first [7] refers to breast cancer. Mismatch repair (MMR)-deficient cancers have been discovered to be highly responsive to immune therapies such as PD-1 checkpoint blockade, making their definition in patients, where they may be relatively rare, paramount for treatment decisions. In the study published in [7], we utilised patterns of mutagenesis known as mutational signatures, which are imprints of the mutagenic processes associated with MMR deficiency, to identify MMR-deficient breast tumours from a whole-genome sequencing dataset comprising a cohort of 640 patients. We identified 11 of 640 tumours as MMR deficient, but only 2 of 11 exhibited germline mutations in MMR genes or Lynch Syndrome. Two additional tumours had a substantially reduced proportion of mutations attributed to MMR deficiency, where the predominant mutational signatures were related to APOBEC enzymatic activity. Overall, 6 of 11 of the MMR-deficient

cases in this cohort were confirmed genetically or epigenetically as having abrogation of MMR genes. However, IHC analysis of MMR-related proteins revealed all but one of 10 samples available for testing as MMR deficient. Thus, the mutational signatures more faithfully reported MMR deficiency than sequencing of MMR genes, because they represent a direct pathophysiologic readout of repair pathway abnormalities. As whole-genome sequencing continues to become more affordable, it could be used to expose individually abnormal tumours in tissue types where MMR deficiency has been rarely detected, but also rarely sought.

The second [18] concerns early human embryo. Somatic cells acquire mutations throughout the course of an individual's life. Mutations occurring early in embryogenesis are often present in a substantial proportion of, but not all, cells in postnatal humans and thus have particular characteristics and effects. Depending on their location in the genome and the proportion of cells they are present in, these mosaic mutations can cause a wide range of genetic disease syndromes and predispose carriers to cancer. They have a high chance of being transmitted to offspring as de novo germline mutations and, in principle, can provide insights into early human embryonic cell lineages and their contributions to adult tissues. Although it is known that gross chromosomal abnormalities are remarkably common in early human embryos, our understanding of early embryonic somatic mutations is very limited. In this work, whole-genome sequences of normal blood from 241 adults was used to identify 163 early embryonic mutations. It was estimated that approximately three base substitution mutations occur per cell per cell-doubling event in early human embryogenesis and these are mainly attributable to two known mutational signatures. The mutations were then used to reconstruct developmental lineages of adult cells and demonstrate that the two daughter cells of many early embryonic cell-doubling events contribute asymmetrically to adult blood at an approximately 2:1 ratio. This study provided insights into the mutation rates, mutational processes and developmental outcomes of cell dynamics that operate during early human embryogenesis.

6.7. Cross-fertilising different computational approaches and other theoretical results

Bubble generator.

As mentioned earlier, a theoretical recent work of the team related to NGS analysis was accepted at the 43rd International Workshop on Graph-Theoretic Concepts in Computer Science (WG) in 2017 [30]. It introduced what was called a bubble generator.

Bubbles are pairs of internally vertex-disjoint (s, t)-paths with applications in the processing of DNA and RNA data. For example, enumerating alternative splicing events in a reference-free context can be done by enumerating all bubbles in a de Bruijn graph built from RNA-seq reads. However, listing and analysing all bubbles in a given graph is usually unfeasible in practice, due to the exponential number of bubbles present in real data graphs. In [30], we proposed a notion of a bubble generator set, *i.e.* a polynomial-sized subset of bubbles from which all the others can be obtained through the application of a specific symmetric difference operator. This set provides a compact representation of the bubbles can be more conveniently extracted from this compact set. Furthermore, we provide a polynomial-time algorithm to decompose any bubble of a graph into the bubbles of such a generator in a tree-like fashion.

IBIS Project-Team

6. New Results

6.1. Qualitative modeling of gene regulatory networks in food-borne pathogens

Bacteria are able to respond to a variety of environmental stresses, which poses food safety problems when these bacteria are food-borne pathogens. Addition of salt, one of the most ancient and common way of preserving food, subjects the bacteria to an osmotic stress to which some may survive. However, the molecular mechanisms of adaptation in food-born pathogens are largely unknown. As a first step towards better understanding these adaptation processes on the molecular level, Delphine Ropers and Aline Métris from the Institute for Food Research in Norwich (UK), invited researcher in IBIS last year, developed a qualitative model of the osmotic stress response in the model bacterium *Escherichia coli*. The qualitative dynamics of the network has been analyzed using the tool GENETIC NETWORK ANALYZER (GNA). The model has allowed to reproduce the behavior of *E. coli* cells adapting to an osmotic stress by including the regulatory mechanisms involved in the process. This work has been published in the *International Journal of Food Microbiology* [21]. It paves the way to modelling stress responses of other foodborne pathogens like *Salmonella* to stresses relevant for the food industry, for which much less is known.

6.2. Analysis of fluorescent reporter gene data

The use of fluorescent and luminescent reporter genes allows real-time monitoring of gene expression, both at the level of individual cells and cell populations (Section 3.2). In order to fully exploit this technology, we need methods to rapidly construct reporter genes, both on plasmids and on the chromosome, mathematical models to infer biologically relevant quantities from the primary data, and computer tools to achieve this in an efficient and user-friendly manner. For instance, in a typical microplate experiment, 96 cultures are followed in parallel, over several hours, resulting in 10,000-100,000 measurements of absorbance and fluorescence and luminescence intensities.

Valentin Zulkower, former PhD student in IBIS, developed novel methods for the analysis of reporter gene data obtained in microplate experiments, based on the use of regularized linear inversion. This allows a range of estimation problems in the analysis of reporter gene data, notably the inference of growth rate, promoter activity, and protein concentration profiles, to be solved in a mathematically sound and practical manner. The linear inversion methods, published in *Bioinformatics* in 2015 [12], in have been implemented in the Python package WELLFARE and integrated in the web application WELLINVERTER. Funded by a grant from the Institut Français de Bioinformatique (IFB), Yannick Martin has improved WellInverter by developing a parallel computational architecture with a load balancer to distribute the analysis queries over several backend servers, a new graphical user interface, and a plug-in system for defining high-level routines for parsing data files produced by microplate readers from different manufacturers. This has resulted in a scalable and user-friendly web service providing a guaranteed quality of service, in terms of availability and response time. This web service has been deployed on the IFB cloud and on an Inria server, accompanied by extensive user documentation, online help, and a tutorial. A paper on WELLINVERTER is in preparation.

While the use of microplate readers results in population-level measurements of gene expression, for many applications it is mandatory to monitor gene expression over time on the level of individual cells. Several developments in the past decade have enormously extended the capabilities to achieve this, in particular the combination of fluorescence time-lapse microscopy for precisely quantifying gene expression in single cells and microfluidics technology for cultivating bacteria in confined spatial compartments and under well-controlled experimental conditions. One of the most wide-spread microfluidics devices is the so-called mother machine shown in Figure 5 . A major problem is that functional software for image analysis (segmentation, tracking, lineage reconstruction, ...) adapted to the requirements of mother machine applications are still rare. IBIS has therefore collaborated with the BEAGLE project-team for the adaptation of their tool FLUOBACTRACKER

to the analysis of time-lapse movies of fluorescent reporter expression and bacterial growth in microfluidics devices. This collaboration, which has also involved the SERPICO project-team, was supported by the Technology Transfer and Innovation department of Inria, in the framework of the Inria Hub program, and has allowed the hiring of Cyril Dutrieux as a software engineer in IBIS.

6.3. Analysis of dynamic metabolomics data

An important step in the study of intracellular metabolism is the quantification of growth rates as well as uptake and excretion rates of metabolites in growing cellular populations. Traditional approaches are based on steady-state experiments, where time-invariant growth rates and exchange fluxes are measured in different experimental conditions. Technological advances in metabolomics have made it possible to monitor the concentration of extracellular metabolites over time, thus paving the way for the study of metabolism in transient conditions. Recovering time-varying exchange and growth rates from time-lapse metabolimics data is a key aspect of this challenge.

We have investigated the reconstruction of exchange reaction and growth rates from time-lapse measurements of external metabolite concentrations and population growth. In particular we have focused on the case of exhaustion of specific substrates, entailing sudden metabolic reorganization of the cell such as diauxie shifts in *E. coli*. Such discontinuities in the metabolic dynamics make data analysis and rate reconstruction particularly challenging but also information-rich. We have developed a Bayesian method that explicitly accounts for these sudden changes and the correlated adaptation of growth in order to accurately estimate time-varying exchange reaction and growth rates, and tested the method on real data from batch and fed-batch cultures of *E. coli* and *L. lactis* obtained at INRA/INSA Toulouse. The method is based on a time-inhomogeneous Gaussian process characterization of the rate dynamics, and Kalman smoothing techniques for the solution of the regularized estimation problem. Method and results were presented at the joint 2017 ISMB-ECCB conference, and published in the corresponding special issue of *Bioinformatics* [17]. The software implementing the method in Matlab is available at https://team.inria.fr/ibis/rate-estimation-software/, and has also been used for the data analysis in another joint publication with INRA/INSA Toulouse [20]. Further developments of the method are under consideration.

6.4. Models of carbon metabolism in bacteria

Adaptation of bacterial growth to changes in environmental conditions, such as the availability of specific carbon sources, is triggered at the molecular level by the reorganization of metabolism and gene expression: the concentration of metabolites is adjusted, as well as the concentration and activities of enzymes, the rate of metabolic reactions, the transcription and translation rates, and the stability of proteins and RNAs. This reprogramming of the bacterial cell is carried out by i) specific interactions involving regulatory proteins or RNAs that specifically respond to the change of environmental conditions and ii) global regulation involving changes in the concentration of RNA polymerase, ribosomes, and metabolite pools that globally affect the rates of transcription, translation, and degradation of all RNAs and proteins. While these phenomena have been well studied in steady-state growth conditions, recent works by IBIS members and collaborators support the view that regulatory mechanisms of growth adaptation are best observed in dynamical conditions.

A first study concerns the second messenger cAMP in *E. coli* and its role in carbon catabolite repression, the mechanism by which bacterial cells select their preferred carbon source for growth. Studies performed in steady-state conditions have questioned the importance of cAMP, leading to a controversy on its physiological role, more than fifty years after its discovery. In a recently submitted journal paper, reporting work started during the PhD thesis of Valentin Zulkower and continued over the past two years, we argue that in order to properly assess the role of cAMP one should shift the focus from steady-state to dynamical conditions. We show, by a combination of fluorescent reporter gene assays and quantitative modeling, that a transient peak in the expression of cAMP-dependent genes leads to the accumulation of proteins necessary for growth on a variety of alternative carbon sources. In the long run, the expression of genes cognate to the alternative carbon source present in the environment is maintained by dedicated positive feedback circuits. Our results thus demonstrate that carbon catabolite repression and diauxic growth need to be understood from a dynamical perspective within the context of a hierarchical regulatory network.

A quantitative description and understanding of this complex network, cutting across metabolism, gene expression, and signalling, can be accessed through mathematical modelling only. In collaboration with Andreas Kremling, professor at TU München and former visiting scientist in the IBIS project-team, Hans Geiselmann, Delphine Ropers and Hidde de Jong developed an ensemble of variants of a simple core model of carbon catabolite repression. The model variants, with two substrate assimilation pathways and four intracellular metabolites only, differ from one another in only a single aspect, each breaking the symmetry between the two pathways in a different manner. Interestingly, all model variants are able to reproduce the data from a reference diauxic growth experiment. For each of the model variants, we predicted the behaviour in two new experimental conditions. When qualitatively comparing these predictions with experimental data, a number of models could be excluded while other model variants are still not discriminable. The best-performing model variants are based on inducer inclusion and activation of enzymatic genes by a global transcription factor, but the other proposed factors may complement these well-known regulatory mechanisms. The model ensemble, which was described in a journal paper recently submitted for publication, offers a better understanding of the variety of mechanisms that have been proposed to play a role in carbon catabolite repression, but is also useful as an educational resource for systems biology.

The same focus on the dynamics of physiological processes has shaped a project on the post-transcriptional control of carbon central metabolism in *E. coli*. In the framework of the PhD thesis of Manon Morin, supported by a Contrat Jeune Scientifique INRA-Inria, the collaboration of Delphine Ropers with Muriel Cocaign-Bousquet and Brice Enjalbert at INRA/INSA Toulouse has demonstrated the key role played by the post-transcriptional regulatory system CSR in growth transitions. In a multi-scale analysis of several wild-type and mutant strains of the CSR system, a variety of experimental data have been acquired in relevant conditions, including growth parameters, gene expression levels and metabolite pools. Data integration through the estimation of fermentation fluxes and flux balance analysis, using the method described above (Section 6.3), have elucidated the role of post-transcriptional regulation in the dynamics of glycogen storage and consumption, as well as the key role of the latter compound for bacterial fitness, through the regulation of intracellular energy levels. A paper summarizing the work has been published in *mBio* [20].

The collaboration with INRA/INSA de Toulouse is continued in the context of the PhD thesis of Thibault Etienne, funded by an INRA-Inria PhD grant, with the objective of developing models able to explain how cells coordinate their physiology and the functioning of the transcription, translation, and degradation machineries following changes in the availability of carbon sources in the environment.

6.5. Stochastic modeling and identification of gene regulatory networks in bacteria

At the single-cell level, the processes that govern single-cell dynamics in general and gene expression in particular are better described by stochastic models. Modern techniques for the real-time monitoring of gene expression in single cells enable one to apply stochastic modelling to study the origins and consequences of random noise in response to various environmental stresses, and the emergence of phenotypic variability. The potential impact of single-cell stochastic analysis and modelling ranges from a better comprehension of the biochemical regulatory mechanisms underlying cellular phenotypes to the development of new strategies for the (computer assisted or genetically engineered) control of cell populations and even of single cells.

Work in IBIS on gene expression and interaction dynamics at the level of individual cells is addressed in terms of identification of intrinsic noise models from population snapshot data, on the one hand, and the inference of models focusing on cellular variability within isogenic populations from fluorescence microscopy gene expression profiles, on the other hand. Along with modelling and inference comes analysis of the inferred models in various respects, notably in terms of identifiability, single-cell state estimation and control. Other problems related with single-cell modelling and extracellular variability are considered in eukaryotic cells through external collaborations.

Concerning identification of intrinsic noise dynamics in single cells, previous results on the contribution of stochasticity to parameter identifiability have been revisited in the context of reconstruction of unknown networks. For the case of population snapshot meaurements, where the dynamics of the population statistics are observed by simple time-lapse experiments, we performed an analytical study of the additional information provided by variance measurements for the reconstruction of unknown first-order kinetics. Based on simulated example, we showed that a tremendous improvement in network reconstruction is achieved relative to the utilization of population-average statistics alone, as addressed by deterministic modelling. These exciting yet preliminary results were published in the form of a paper in the proceedings of the *IFAC World Congress* [22] and will be further developed.

Reconstruction of promoter activity statistics from reporter gene population snapshot data has been further investigated, leading to a full-blown spectral analysis and reconstruction method for reporter gene systems. Building upon reults in previous conference papers, in the context of the ANR project MEMIP (Section 8.2), we have characterized reporter systems as noisy linear systems operating on a stochastic input (promoter activity), and developed an inversion method for nonparametric estimation of promoter dynamics, namely the autocovariance function, from the considered readouts. These theoretical and simulation results have been submitted for journal publication and are also available as an arXiv pre-print. The method will be further developed and applied to real data and case studies.

Modelling of heterogeneity in isogenic cell populations is also an active research direction. Still in the context of MEMIP, in collaboration with the INBIO team, we are considering generalizations of our achievements on Mixed-Effects modelling and inference on yeast, in order to account for different sources of noise and lineage effects. As an offspring of this work, a study of inter-individual variability of *E. coli* gene expression and growth rate in growth arrest-and-restart experiments has been carried out with BIOCORE. Results obtained so far are part of the PhD thesis of Stefano Casagrande.

6.6. Modelling bacterial growth

Various mathematical approaches have been used in the literature to describe the networks of biochemical reactions involved in microbial growth. With various levels of detail, the resulting models provide an integrated view of these reaction networks, including the transport of nutrients from the environment and the metabolism and gene expression allowing the conversion of these nutrients into biomass. The models hence bridge the scale between individual reactions to the growth of cell populations. In a review article published in the *Journal of the Royal Society Interface* [18], several IBIS members as well as colleagues from the BIOCORE project-team, discuss various models of microbial growth that are, at first sight, quite diverse. They have a different scope and granularity, make different simplifications, use different approaches to obtain predictions from the model structure and have their origin in different fields. In the review we derive a general framework for the kinetic modelling of microbial growth from a few basic hypotheses on the systems of biochemical reactions underlying microbial growth found in the literature, including self-replicator models of bacterial growth developed by Nils Giordano in his PhD thesis and published in *PLoS Computational Biology* last year [5]. This reveals how the models are related on a deeper level and provides a sound basis for further modelling studies.

Analysing the dynamics of some of the network models mentioned above becomes quickly intractable, when mathematical functions are for instance given by complex algebraic expressions resulting from the mass balance of biochemical reactions. In a paper published in the *Bulletin of Mathematical Biology* [16], Edith Grac, former post-doc in Ibis, Delphine Ropers, and Stefano Casagranda and Jean-Luc Gouzé from the BIOCORE project-team, have studied how monotone system theory and time-scale arguments can be used to reduce high-dimension models based on the mass-action law. Applying the approach to an important positive feedback loop regulating the expression of RNA polymerase in *E. coli*, made it possible to study the stability of the system steady states and relate the dynamical behaviour of the system to observations on the physiology of the bacterium *E. coli*.

6.7. Growth control in bacteria and biotechnological applications

The ability to experimentally control the growth rate is crucial for studying bacterial physiology. It is also of central importance for applications in biotechnology, where often the goal is to limit or even arrest growth. Growth-arrested cells with a functional metabolism open the possibility to channel resources into the production of a desired metabolite, instead of wasting nutrients on biomass production. The objective of the RESET project, supported in the framework of the Programme d'Investissements d'Avenir (Section 8.2), is to develop novel strategies to limit or completely stop microbial growth and to explore biotechnological applications of these approaches.

A foundation result for growth control in bacteria was published in the journal *Molecular Systems Biology* at the end of 2015 [6]. In that publication, we described an engineered *E. coli* strain where the transcription of a key component of the gene expression machinery, RNA polymerase, is under the control of an inducible promoter. By changing the inducer concentration in the medium, we can adjust the RNA polymerase concentration and thereby switch bacterial growth between zero and the maximal growth rate supported by the medium. The publication also presented a biotechnological application of the synthetic growth switch in which both the wild-type *E. coli* strain and our modified strain were endowed with the capacity to produce glycerol when growing on glucose. Cells in which growth has been switched off continue to be metabolically active and harness the energy gain to produce glycerol at a twofold higher yield than in cells with natural control of RNA polymerase expression. Remarkably, without any further optimization, the improved yield is close to the theoretical maximum computed from a flux balance model of *E. coli* metabolism. This work has been continued in several directions in the context of the RESET project by Célia Boyat. Moreover, extending work on self-replicator models of bacterial growth, we have studied the production of metabolites by means of the growth switch from an optimal control perspective, in a paper that is currently being prepared for publication.

In a review published in *Trends in Microbiology* this year [19], we have put the scientific results mentioned above in a broader context. As illustrated by the synthetic growth switch, reengineering the gene expression machinery allows modifying naturally evolved regulatory networks and thereby profoundly reorganizing the manner in which bacteria allocate resources to different cellular functions. This opens new opportunities for our fundamental understanding of microbial physiology and for a variety of applications. We describe how recent breakthroughs in genome engineering and the miniaturization and automation of culturing methods have offered new perspectives for the reengineering of the transcription and translation machinery in bacteria as well as the development of novel *in vitro* and *in vivo* gene expression systems. In our paper, we review different examples from the unifying perspective of resource reallocation, and discuss the impact of these approaches for microbial systems biology and biotechnological applications.

IMAGINE Project-Team

6. New Results

6.1. User-centered Models for Shapes and Shape Assemblies

- Scientist in charge: Stefanie Hahmann.
- Other permanent researchers: Marie-Paule Cani, Frédéric Devernay, Jean-Claude Léon, Damien Rohmer.

Our goal, is to develop responsive shape models, i.e. 3D models that respond in the expected way under any user action, by maintaining specific application-dependent constraints (such as a volumetric objects keeping their volume when bent, or cloth-like surfaces remaining developable during deformation, etc). We are extending this approach to composite objects made of distributions and/or combination of sub-shapes of various dimensions.

6.1.1. Deformation Grammars: Hierarchical Constraint Preservation Under Deformation



Figure 1. Deformation grammars [24] allow to freely deform complex objects or object assemblies, while preserving their consistency. Top row: Original hierarchical objects (tree, house, bird flock, scene with mixed elements). The tree and the bird flock are made of parts of the same type, while the other objects are hetegogeneous hierarchies. Bottom row: Deformed objects, where the interpretation of user-controled deformations through deformation grammars is used to automatically maintain consistency constraints.

Deformation grammars are a novel procedural framework enabling to sculpt hierarchical 3D models in an object-dependent manner [24]. They process object deformations as symbols thanks to user-defined interpretation rules. We use them to define hierarchical deformation behaviors tailored for each model, and enabling any sculpting gesture to be interpreted as some adapted constraint-preserving deformation. This is illustrated in Figure 1. A variety of object-specific constraints can be enforced using this framework, such as maintaining distributions of sub-parts, avoiding self-penetrations, or meeting semantic-based user-defined rules. The operations used to maintain constraints are kept transparent to the user, enabling them to focus on their design. We demonstrate the feasibility and the versatility of this approach on a variety of examples, implemented within an interactive sculpting system.

6.1.2. Patterns from Photograph: Reverse-Engineering Developable Products



Figure 2. Patterns from Photograph: Reverse-Engineering Developable Products [17]. Overview of our approach.
(1) Our system takes as input a photograph where the user has traced the object silhouette and the surface patch boundaries. We also ask users to annotate a global symmetry plane and to indicate symmetric curves. (2) We analyze these annotations to register symmetric curves and to propagate rulings from the silhouettes towards the interior of the surface patches. The detected symmetric points and rulings provide us with geometric constraints on the 3D curves, which we express as linear terms in an optimization. (3) Solving this optimization produces a 3D curve network, which we subsequently surface with developable patches. (4,5) These curves are used as boundaries to generate a piecewise developable mesh of the object.

Developable materials are ubiquitous in design and manufacturing. Unfortunately, general-purpose modeling tools are not suited to modeling 3D objects composed of developable parts. We propose an interactive tool to model such objects from a photograph [17]. This is illustrated in Figure 2. Users of our system load a single picture of the object they wish to model, which they annotate to indicate silhouettes and part boundaries. Assuming that the object is symmetric, we also ask users to provide a few annotations of symmetric correspondences. The object is then automatically reconstructed in 3D. At the core of our method is an algorithm to infer the 2D projection of rulings of a developable surface from the traced silhouettes and boundaries. We impose that the surface normal is constant along each ruling, which is a necessary property for the surface to be developable. We complement these developability constraints with symmetry constraints to lift the curve network in 3D. In addition to a 3D model, we output 2D patterns enabling to fabricate real prototypes of the object on the photo. This makes our method well suited for reverse engineering products made of leather, bent cardboard or metal sheets.

6.1.3. Defining the Pose of any 3D Rigid Object and an Associated Distance

The pose of a rigid object is usually regarded as a rigid transformation, described by a translation and a rotation. However, equating the pose space with the space of rigid transformations is in general abusive, as it does not account for objects with proper symmetries – which are common among man-made objects. In our recent work [13], we define pose as a distinguishable static state of an object, and equate a pose with a set of rigid transformations. This is illustrated in Figure 3. Based solely on geometric considerations, we propose a frame-invariant metric on the space of possible poses, valid for any physical rigid object, and requiring no arbitrary tuning. This distance can be evaluated efficiently using a representation of poses within an Euclidean space of at most 12 dimensions depending on the object's symmetries. This makes it possible to efficiently perform neighborhood queries such as radius searches or k-nearest neighbor searches within a large set of poses using off-the-shelf methods. Pose averaging considering this metric can similarly be performed easily, using a projection function from the Euclidean space onto the pose space. The practical value of those theoretical developments is illustrated with an application of pose estimation of instances of a 3D rigid object given an input depth map, via a Mean Shift procedure.



Figure 3. Defining the Pose of any 3D Rigid Object and an Associated Distance [13]. Illustration of our proposed distance for a 2D object with a rotation symmetry of 2p=3. (a) The distance between two poses consists in the minimum distance between two poses of an equivalent object without proper symmetry – here there are 3 possible poses of the equivalent object for each pose of the original object. The distance between poses of an object without proper symmetry corresponds to the RMS distance between corresponding object points (dashed segments). (b) Equivalently, the proposed distance can be considered as a measure of the smallest displacement from one pose to an other – here there are actually only 3 different displacements between those two poses (solid, dotted and dashed boxes).

6.2. Motion & Sound Synthesis

- Scientist in charge: Damien Rohmer.
- Other permanent researchers: Marie-Paule Cani, Frédéric Devernay, Stéfanie Hahmann, Rémi Ronfard.

Animating objects in real-time is mandatory to enable user interaction during motion design. Physicallybased models, an excellent paradigm for generating motions that a human user would expect, tend to lack efficiency for complex shapes due to their use of low-level geometry (such as fine meshes). Our goal is therefore two-folds: first, develop efficient physically-based models and collisions processing methods for arbitrary passive objects, by decoupling deformations from the possibly complex, geometric representation; second, study the combination of animation models with geometric responsive shapes, enabling the animation of complex constrained shapes in real-time. The last goal is to start developing coarse to fine animation models for virtual creatures, towards easier authoring of character animation for our work on narrative design.

6.2.1. Interactive paper tearing

In this work, we proposed an efficient method to model paper tearing in the context of interactive modeling [23]. This is illustrated in Figure 4. The method uses geometrical information to automatically detect potential starting points of tears. We further introduce a new hybrid geometrical and physical-based method to compute the trajectory of tears while procedurally synthesizing high resolution details of the tearing path using a texture based approach. The results obtained are compared with real paper and with previous studies on the expected geometric paths of paper that tears.

6.2.2. A Generative Audio-Visual Prosodic Model for Virtual Actors

In this new work [9], we proposed a method for generating natural speech and facial animation in various attitudes using neutral speech and animation as input. This is illustrated in Figure 5. Given a neutral sentence, we use the phonotactic information to predict prosodic feature contours. The predicted rhythm is used to compute phoneme durations. The expressive speech is synthesized with a vocoder that uses the neutral utterance, predicted rhythm, energy, and voice pitch, and the facial animation parameters are obtained by adding the warped neutral motion to the reconstructed and warped predicted motion contours.



Figure 4. Interactive paper tearing [23]. *The path of a tear follows a geometrical curve but also presents stochastic details.*



Figure 5. A Generative Audio-Visual Prosodic Model for Virtual Actors [9]. The rows present corresponding frames extracted from (a) the video, (b) ground-truth animation, and (c) synthetic animation. From left to right, the images correspond to comforting, fascinated, thinking (male actor), fascinated, ironic, and scandalized (female actor) attitudes.

6.2.3. Which prosodic features contribute to the recognition of dramatic attitudes?

In this new work [10], we explored the capability of audiovisual prosodic features (such as fundamental frequency, head motion or facial expressions) to discriminate among different dramatic attitudes. We extracted the audiovisual parameters from an acted corpus of attitudes and structured them as frame, syllable and sentence-level features. Using Linear Discriminant Analysis classifiers, we showed that prosodic features present a higher discriminating rate at sentence-level. This finding is confirmed by the perceptual evaluation results of audio and/or visual stimuli obtained from the recorded attitudes.

6.3. Knowledge-based Models for Narrative Design

- Scientist in charge: Rémi Ronfard.
- Other permanent researchers: Marie-Paule Cani, Frédéric Devernay, Jean-Claude Léon, Olivier Palombi.

Our long term goal is to develop high-level models helping users to express and convey their own narrative content (from fiction stories to more practical educational or demonstrative scenarios). Before being able to specify the narration, a first step is to define models able to express some a priori knowledge on the background scene and on the object(s) or character(s) of interest. Our first goal is to develop 3D ontologies able to express such knowledge. The second goal is to define a representation for narration, to be used in future storyboarding frameworks and virtual direction tools. Our last goal is to develop high-level models for virtual cinematography such as rule-based cameras able to automatically follow the ongoing action and semi-automatic editing tools enabling to easily convey the narration via a movie.

6.3.1. Zooming On All Actors: Automatic Focus+Context Split Screen Video Generation



Figure 6. Zooming On All Actors: Automatic Focus+Context Split Screen Video Generation [20]. Illustration of our approach showing a frame from the input video (a) and its corresponding split screen composition (b). The split screen view shows the emotions and expressions of the performers, which are hardly visible in the original view (master shot).

Recordings of stage performances are easy to capture with a high-resolution camera, but are difficult to watch because the actors' faces are too small. We present an approach to automatically create a split screen video that transforms these recordings to show both the context of the scene as well as close-up details of the actors [20]. This is illustrated in Figure 6. Given a static recording of a stage performance and tracking information about the actors positions, our system generates videos showing a focus+context view based on computed close-up camera motions using crop-and zoom. The key to our approach is to compute these camera motions such that they are cinematically valid close-ups and to ensure that the set of views of the different actors are properly coordinated and presented. We pose the computation of camera motions as convex optimization that creates

detailed views and smooth movements, subject to cinematic constraints such as not cutting faces with the edge of the frame. Additional constraints link the close up views of each actor, causing them to merge seamlessly when actors are close. Generated views are placed in a resulting layout that preserves the spatial relationships between actors. We demonstrate our results on a variety of staged theater and dance performances.

6.3.2. Make Gestures to Learn: Reproducing Gestures Improves the Learning of Anatomical Knowledge More than Just Seeing Gestures

Manual gestures can facilitate problem solving but also language or conceptual learning. Both seeing and making the gestures during learning seem to be beneficial. However, the stronger activation of the motor system in the second case should provide supplementary cues to consolidate and re-enact the mental traces created during learning. In this work [14], we tested this hypothesis in the context of anatomy learning by naïve adult participants. Anatomy is a challenging topic to learn and is of specific interest for research on embodied learning, as the learning content can be directly linked to learners' body. Two groups of participants were asked to look at a video lecture on the forearm anatomy. The video included a model making gestures related to the content of the lecture. Both groups see the gestures but only one also imitate the model. Tests of knowledge were run just after learning and few days later. The results revealed that imitating gestures improves the recall of structures names and their localization on a diagram. This effect was however significant only in long-term assessments. This suggests that: (1) the integration of motor actions and knowledge may require sleep; (2) a specific activation of the motor system during learning may improve the consolidation and/or the retrieval of memories.

6.4. Creating and Interacting with Virtual Prototypes

- Scientist in charge: Jean-Claude Léon.
- Other permanent researchers: Marie-Paule Cani, Frédéric Devernay, Olivier Palombi, Damien Rohmer, Rémi Ronfard.

The challenge is to develop more effective ways to put the user in the loop during content authoring. We generally rely on sketching techniques for quickly drafting new content, and on sculpting methods (in the sense of gesture-driven, continuous distortion) for further 3D content refinement and editing. The objective is to extend these expressive modeling techniques to general content, from complex shapes and assemblies to animated content. As a complement, we are exploring the use of various 2D or 3D input devices to ease interactive 3D content creation.

6.4.1. EcoBrush: Interactive Control of Visually Consistent Large-Scale Ecosystems



Figure 7. EcoBrush: Interactive Control of Visually Consistent Large-Scale Ecosystems [18]. Terrain conditions (a), such as temperature, soil moisture, and sunlight exposure, are used to index a database of plant distributions (b) and synthesize an initial complete ecosystem (c), which can then be modified with semantic brushes, to adjust age, density and variability (d).

One challenge in portraying large-scale natural scenes in virtual environments is specifying the attributes of plants, such as species, size and placement, in a way that respects the features of natural ecosystems, while remaining computationally tractable and allowing user design. To address this, we combine ecosystem simulation with a distribution analysis of the resulting plant attributes to create biome-specific databases, indexed by terrain conditions, such as temperature, rainfall, sunlight and slope [18]. This is illustrated in Figure 7.

For a specific terrain, interpolated entries are drawn from this database and used to interactively synthesize a full ecosystem, while retaining the fidelity of the original simulations. A painting interface supplies users with semantic brushes for locally adjusting ecosystem age, plant density and variability, as well as optionally picking from a palette of precomputed distributions. Since these brushes are keyed to the underlying terrain properties a balance between user control and real-world consistency is maintained. Our system can be used to interactively design ecosystems up to 5 km \times 5 km in extent, or to automatically generate even larger ecosystems in a fraction of the time of a full simulation, while demonstrating known properties from plant ecology such as succession, self-thinning, and underbrush, across a variety of biomes.

6.4.2. Authoring Landscapes by Combining Ecosystem and Terrain Erosion Simulation



Figure 8. Authoring Landscapes by Combining Ecosystem and Terrain Erosion Simulation [16]. Results from different initial heightfield inputs produced using our simulation after 300 years of evolution. From left to right: an Alpine landscape from the U.S. Rockies, a portion of the Grand Canyon, a Mediterranean landscape, and a forested valley.

In this new paper [16], we introduced a novel framework for interactive landscape authoring that supports bi-directional feedback between erosion and vegetation simulation. This is illustrated in Figure 8. Vegetation and terrain erosion have strong mutual impact and their interplay influences the overall realism of virtual scenes. Despite their importance, these complex interactions have been neglected in computer graphics. Our framework overcomes this by simulating the effect of a variety of geomorphological agents and the mutual interaction between different material and vegetation layers, including rock, sand, humus, grass, shrubs, and trees. Users are able to exploit these interactions with an authoring interface that consistently shapes the terrain and populates it with details. Our method, validated through side-by-side comparison with real terrains, can be used not only to generate realistic static landscapes, but also to follow the temporal evolution of a landscape over a few centuries.

6.4.3. Shape from sensors: Curve networks on surfaces from 3D orientations

We presented a novel framework for acquisition and reconstruction of 3D curves using orientations provided by inertial sensors. This is illustrated in Figure 9. While the idea of sensor shape reconstruction is not new, we present the first method for creating well-connected networks with cell complex topology using only orientation and distance measurements and a set of user- defined constraints. By working directly with orientations, our method robustly resolves problems arising from data inconsistency and sensor noise.


Figure 9. Shape from sensors: Curve networks on surfaces from 3D orientations. Guitar and bathtub scanned using the Morphorider. Left to right, reconstructed networks, surfaces computed from the networks, the scanned objects. The guitar shape was created by extruding the surfaced network.

MAVERICK Project-Team

7. New Results

7.1. Expressive rendering

7.1.1. Edge- and substrate-based effects for watercolor stylization

Participants: Santiago Montesdeoca, Hock Soon Seah, Pierre Bénard, Romain Vergne, Joëlle Thollot, Hans-Martin Rall, Davide Benvenuti.



Figure 2. Our methods allow new and improved edge- and substrate-based effects for watercolor stylization: edge darkening (red), overlaps (green) and dry-brush (yellow). Still life, model by Dylan Sisson ©Pixar Animation Studios.

We investigate characteristic edge-and substrate-based effects for watercolor stylization. These two fundamental elements of painted art play a significant role in traditional watercolors and highly influence the pigment's behavior and application. Yet a detailed consideration of these specific elements for the stylization of 3D scenes has not been attempted before. Through this investigation, we contribute to the field by presenting ways to emulate two novel effects: dry-brush and gaps & overlaps. By doing so, we also found ways to improve upon well-studied watercolor effects such as edge-darkening and substrate granulation. Finally, we integrated controllable external lighting influences over the watercolorized result, together with other previously researched watercolor effects. These effects are combined through a direct stylization pipeline to produce sophisticated watercolor imagery (see Figure 2), which retains spatial coherence in object-space and is locally controllable in real-time. This work has been published in Expressive'2017 [9].

7.1.2. Direct 3D stylization pipelines

Participants: Santiago Montesdeoca, Hock Soon Seah, Pierre Bénard, Romain Vergne, Joëlle Thollot, Hans-Martin Rall, Davide Benvenuti.

Using 3D computer graphics to emulate watercolor presents a special challenge. Complex stylizations are commonly processed offline, by combining multiple passes in compositing, where art directing is slow and non-intuitive because the stylized result is not immediate. This direct 3D stylization pipeline allows art direction to happen in real time (Figure 3). Using the framework, artists can assign their desired local and global effects directly in the 3D scene, see the stylized results immediately, and intuitively adapt them to fit their stylized vision. The technique can be used and applied in 3D animations, games, VR, visualizations, illustrations, and interactive art. This work was presented at the SIGGRAPH real-time live session [13].



Figure 3. This rendering is stylized and animated in real-time using our tool.

7.2. Computer-aided image manipulation

7.2.1. Local texture-based color transfer and colorization

Participants: Benoit Arbelot, Romain Vergne, Thomas Hurtut, Joëlle Thollot.

This work targets two related color manipulation problems: Color transfer for modifying an image's colors and colorization for adding colors to a grayscale image. Automatic methods for these two applications propose to modify the input image using a reference that contains the desired colors. Previous approaches usually do not target both applications and suffer from two main limitations: possible misleading associations between input and reference regions and poor spatial coherence around image structures. In this paper, we propose a unified framework that uses the textural content of the images to guide the color transfer and colorization. Our method introduces an edge-aware texture descriptor based on region covariance, allowing for local color transformations. We show that our approach is able to produce results comparable or better than state-of-the-art methods in both applications (see Figure 4). This work is an extended version of an Expressive' 2016 paper and was published in the C&G journal [1].



Color transfer Colorization Figure 4. Our framework allows for automatic local color transfer (left) and colorization (right).

7.2.2. Texture Transfer Based on Texture Descriptor Variations

Participants: Benoit Arbelot, Romain Vergne, Thomas Hurtut, Joëlle Thollot.



Figure 5. The bark texture of the input tree (a) is replaced by the bark texture of the reference (b) in the final result (c). Alpha masks (bottom right) are used to define the input and reference textures in the images.

We tackle the problem of image-space texture transfer which aims to modify an object or surface material by replacing its input texture by another reference texture. The main challenge of texture transfer is to successfully reproduce the reference texture patterns while preserving the input texture variations due to its environment such as illumination or shape variations. We propose to use a texture descriptor composed of local luminance and local gradients orientation and magnitude to characterize the input texture variations. We then introduce a guided texture synthesis algorithm to synthesize a texture resembling the reference texture with the input texture variations. The main contribution of our algorithm is its ability to locally deform the reference texture according to local texture descriptors in order to better reproduce the input texture variations. We show that our approach is able to produce results comparable with current state-of-the-art approaches but with fewer user inputs. Preliminary results of this work are shown in a research report [14].

7.3. Illumination Simulation and Materials

7.3.1. Point-Based Rendering for Homogeneous Participating Media with Refractive Boundaries

Participants: Beibei Wang, Nicolas Holzschuch.

Illumination effects in translucent materials are a combination of several physical phenomena: refraction at the surface, absorption and scattering inside the material. Because refraction can focus light deep inside the material, where it will be scattered, practical illumination simulation inside translucent materials is difficult. We present an a Point-Based Global Illumination method for light transport on homogeneous translucent materials with refractive boundaries. We start by placing light samples inside the translucent material and organizing them into a spatial hierarchy. At rendering, we gather light from these samples for each camera ray. We compute separately the sample contributions for single, double and multiple scattering, and add them. We present two implementations of our algorithm: an offline version for high-quality rendering and an interactive GPU implementation. The offline version provides significant speed-ups and reduced memory footprints compared to state-of-the-art algorithms, with no visible impact on quality (Figure 6). The GPU version yields interactive frame rates: 30 fps when moving the viewpoint, 25 fps when editing the light position

or the material parameters. This work has been published in IEEE Transactions on Visualization and Computer Graphics [7].



Figure 6. Individual component validation on the Bumpy Sphere scene, high-quality offline rendering.

7.3.2. A Two-Scale Microfacet Reflectance Model Combining Reflection and Diffraction Participants: Nicolas Holzschuch, Romain Pacanowski.

Adequate reflectance models are essential for the production of photorealistic images. Microfacet reflectance models predict the appearance of a material at the macroscopic level based on microscopic surface details. They provide a good match with measured reflectance in some cases, but not always. This discrepancy between the behavior predicted by microfacet models and the observed behavior has puzzled researchers for a long time. In this work, we show that diffraction effects in the micro-geometry provide a plausible explanation. We describe a two-scale reflectance model, separating between geometry details much larger than wavelength and those of size comparable to wavelength. The former model results in the standard Cook-Torrance model. The latter model is responsible for diffraction effects. Diffraction effects at the smaller scale are convolved by the micro-geometry normal distribution. The resulting two-scale model provides a very good approximation to measured reflectances (Figure 7). This work has been published in TOG [2] and presented at SIGGRAPH 2017. It was also presented at the "tout sur les BRDF" day in Poitier (France) [17].

7.3.3. Precomputed Multiple Scattering for Light Simulation in Participating Medium Participanty, Paikai Wang, Nicolas Holgschuch

Participants: Beibei Wang, Nicolas Holzschuch.

Illumination simulation involving participating media is computationally intensive. The overall aspect of the material depends on simulating a large number of scattering events inside the material. Combined, the contributions of these scattering events are a smooth illumination. Computing them using ray-tracing or photon-mapping algorithms is expensive: convergence time is high, and pictures before convergence are low quality. In this work, we precompute the result of multiple scattering events, assuming an infinite medium, and store it in two 4D tables. These precomputed tables can be used with many rendering algorithms, such as Virtual Ray Lights (VRL), Unified Point Beams and Paths (UPBP) or Manifold Exploration Metropolis Light Transport (MEMLT), greatly reducing the convergence time (Figure 8). The original algorithm takes care of low order scattering (single and double scattering), while our precomputations are used for multiple scattering (more than two scattering events). This work was presented at SIGGRAPH [12].

7.3.4. The Effects of Digital Cameras Optics and Electronics for Material Acquisition Participants: Nicolas Holzschuch, Romain Pacanowski.



Figure 7. Material reflectance properties are caused by small variations in surface geometry. We separate these surface variations into micro-geometry, of size larger than the wavelength of visible light, and nano-geometry, of size comparable to the wavelength. The latter produces diffraction effects, with wavelength-dependent effects. The former corresponds to the classical Cook-Torrance lobe. We explain how these two levels interact and show that combined together, they reproduce measured materials faithfully, including subtle color shifts.



Figure 8. Our precomputed multiple scattering accelerates the convergence of existing algorithms for participating media.

For material acquisition, we use digital cameras and process the pictures. We usually treat the cameras as perfect pinhole cameras, with each pixel providing a point sample of the incoming signal. In this work, we study the impact of camera optical and electronic systems. Optical system effects are modelled by the Modulation Transfer Function (MTF). Electronic System effects are modelled by the Pixel Response Function (PRF). The former is convolved with the incoming signal, the latter is multiplied with it. We provide a model for both effects, and study their impact on the measured signal. For high frequency incoming signals, the convolution results in a significant decrease in measured intensity, especially at grazing angles. We show this model explains the strange behaviour observed in the MERL BRDF database at grazing angles. This work has been presented at the Workshop on Material Appearance Modeling [8].

7.3.5. A Versatile Parameterization of Measured Material Manifolds

Participants: Cyril Soler, Kartic Subr, Derek Nowrouzezahrai.

A popular approach for computing photorealistic images of virtual objects requires applying reflectance profiles measured from real surfaces, introducing several challenges: the memory needed to faithfully capture realistic material reflectance is large, the choice of materials is limited to the set of measurements, and image synthesis using the measured data is costly. Typically, this data is either compressed by projecting it onto a subset of its linear principal components or by applying non-linear methods. The former requires prohibitively large numbers of components to faithfully represent the input reflectance, whereas the latter necessitates costly algorithms to extrapolate reflectance data. We learn an underlying, low-dimensional non-linear reflectance manifold amenable to the rapid exploration and rendering of real-world materials. We show that interpolated materials can be expressed as linear combinations of the measured data, despite lying on an inherently non-linear manifold. This allows us to efficiently interpolate and extrapolate measured BRDFs, and to render directly from the manifold representation. To do so, we rely on a Gaussian process latent variable model of reflectance. We demonstrate the utility of our representation in the context of both high-performance and offline rendering with materials that interpolated from real-world captured BRDFs. This work has been accepted for publication at the Eurographics 2018 conference.

7.4. Complex Scenes

In order to render both efficiently and accurately ultra-detailed large scenes, this approach consists in developing representations and algorithms able to account compactly for the quantitative visual appearance of a regions of space projecting on screen at the size of a pixel.

7.4.1. Appearance pre-filtering

Participants: Guillaume Loubet, Neyret Fabrice.

We address the problem of constructing appearance-preserving level of details (LoDs) of complex 3D models such as trees and propose a hybrid method that combines the strength of mesh and volume representations. Our main idea is to separate macroscopic (i.e. larger than the target spatial resolution) and microscopic (sub-resolution) surfaces at each scale and to treat them differently, because meshes are very efficient at representing macroscopic surfaces while sub-resolution geometry benefit from volumetric approximations. We introduce a new algorithm based on mesh analysis that detects the macroscopic surfaces of a 3D model at a given resolution. We simplify these surfaces with edge collapses and provide a method for pre-filtering their BRDFs parameters. To approximate microscopic details, we use a heterogeneous microflake participating medium and provide a new artifact-free voxelization algorithm that preserves local occlusion. Thanks to our macroscopic surface analysis, our algorithm is fully automatic and can generate seamless LoDs at arbitrarily coarse resolutions for a wide range of 3D models. We validated our method on highly complex geometry and show that appearance is consistent across scales while memory usage and loading times are drastically reduced (see Figure 10). This work has been accepted at EG2017 [4].

7.4.2. Appearance pre-filtering of self-shadowing and anisotropic occlusion Participants: Guillaume Loubet, Neyret Fabrice.



Figure 9. Four of the images above (Number 2,4,6 and 12 in reading order) are rendered with measured BRDFs from the MERL dataset [MPBM03a], the remaining 11 being rendered with BRDFs randomly picked from our parameterization of the non-linear manifold containing MERL materials. We explore this manifold interactively to produce high-quality BRDFs which retain the physical properties and perceptual aspect of real materials.



Figure 10. A weeping willow 3D model pre-filtered with our method. Our LoDs use meshes for representing macroscopic surfaces and a volumetric representation to approximate sub-resolution geometry. This approach allows for accurate preservation of the appearance of complex geometry across scales while memory usage is drastic reduced. These images have been rendered with 256spp and a thin lense camera model in Mitsuba (http://www.mitsuba-renderer.org/).

This year, we addressed the problem of representing the effect of internal self-shadowing in elements about to be filtered out at a given LOD, in the scope of volume of voxels containing density and phase-function (represented by a microflakes).

Naïve linear methods for downsampling high resolution microflake volumes often produce inaccurate results, especially when input voxels are very opaque. Preserving correct appearance at all resolutions requires taking into account inter- and intravoxel self-shadowing effects (see Figure 11). We introduce a new microflake model whose parameters characterize self-shadowing effects at the microscopic scale. We provide an anisotropic self-shadowing function and a microflake distribution for which scattering coefficients and phase functions of our model have closed-form expressions. We use this model in a new downsampling approach in which scattering parameters are computed from local estimations of self-shadowing in the input volume. Unlike previous work, our method handles datasets with spatially varying scattering parameters, semi-transparent volumes and datasets with intricate silhouettes. We show that our method generates LoDs with correct transparency and consistent appearance through scales for a wide range of challenging datasets, allowing for huge memory savings and efficient distant rendering without loss of quality. This work has been accepted at EG2018.



Figure 11. Comparison between naïve downsampling of microflake volumes and our method ("Aniso"). Naïve dowsampling of volumes with dense voxels often lead to inaccurate results due to the loss of inter- and intra-voxel self-shadowing effects. Our method is based on a new participating medium model and on local estimations of self-shadowing. It generates LoDs with correct transparency and consistent appearance through scales. Rendered with volume path tracing in Mitsuba (http://www.mitsuba-renderer.org/): the trunk of the cedar is a mesh.

7.5. Texture Synthesis

7.5.1. Programmable 2D Arrangements for Element Texture Design

Participants: Hugo Loi, Thomas Hurtut, Romain Vergne, Joëlle Thollot.

This work introduces a programmable method for designing stationary 2D arrangements for element textures, namely textures made of small geometric elements. These textures are ubiquitous in numerous applications of computer-aided illustration. Previous methods, whether they be example-based or layout-based, lack control and can produce a limited range of possible arrangements. Our approach targets technical artists who will design an arrangement by writing a script. These scripts are using three types of operators: partitioning operators for defining the broad-scale organization of the arrangement, mapping operators for controlling the local organization of elements, and merging operators for mixing different arrangements. These operators are designed so as to guarantee a stationary result meaning that the produced arrangements will always be repetitive. We show that this simple set of operators is sufficient to reach a much broader variety of arrangements than previous methods. Editing the script leads to predictable changes in the synthesized



Figure 12. Some textures obtained using our operators and mapped onto 3D surfaces.

arrangement, which allows an easy iterative design of complex structures. Finally, our operator set is extensible and can be adapted to application-dependent needs such as 3D texturing, as shown in Figure 12. This work has been published in TOG [3] and presented at SIGGRAPH 2017.

7.6. Visualization and Geometric Design

7.6.1. Activelec: an Interaction-Based Visualization System to Analyze Household Electricity Consumption

Participants: Jérémy Wambecke, Georges-Pierre Bonneau, Romain Vergne, Renaud Blanch.

Everyone can now record and explore the evolution over time of his/her personal household electricity consumption. However understanding what links this data to our behavior remains a challenge. In this work, we present a visualization tool based on the direct manipulation, by the users, of their behavior (see Figure 13). Users can select and modify their actions over time, evaluating the results on the data with the visualization. We also conduct a user study, showing that our method allows users to understand the links between actions and data, and to use this knowledge in order to test and evaluate changes their behavior. This work has been published in Visualization In Practice (VIP) [11].

7.6.2. Computing Contour Trees for 2D Piecewise Polynomial Functions

Participants: Georges-Pierre Bonneau, Stefanie Hahmann, Girijanandan Nucha, Vijay Natarajan.

This work is a result from a collaboration with Vijay Natarajan from the Indian Institute of SCience (IISc), Bangalore, and team-project IMAGINE (Stefanie Hahmann). Contour trees are extensively used in scalar field analysis. The contour tree is a data structure that tracks the evolution of level set topology in a scalar field. Scalar fields are typically available as samples at vertices of a mesh and are linearly interpolated within each cell of the mesh. A more suitable way of representing scalar fields, especially when a smoother function needs to be modeled, is via higher order interpolants. We propose an algorithm to compute the contour tree for such functions. The algorithm computes a local structure by connecting critical points using a numerically stable monotone path tracing procedure. Such structures are computed for each cell and are stitched together to obtain the contour tree of the function. The algorithm is scalable to higher degree interpolants whereas previous methods were restricted to quadratic or linear interpolants. The algorithm is intrinsically parallelizable and has potential applications to isosurface extraction. Figure 14 shows examples of contour trees for quadratic and cubic scalar functions defined on 3D meshes. The results have been published in Computer Graphics Forum [5].



Figure 13. Interface of our visualization tool.



Figure 14. From left to right: quadratic (top) and cubic (bottom) scalar function visualized by a colormap, critical points of the scalar function, exact contour tree, simplified contour tree.

7.6.3. Shape from sensors: Curve networks on surfaces from 3D orientations

Participants: Tibor Stanko, Stefanie Hahmann, Georges-Pierre Bonneau, Nathalie Saguin-Sprynski.



Figure 15. From left to right: curves are scanned on an object with a device measuring orientations, simple Euler integration of tangents, reconstruction of curve network with incoherent orientations, reconstruction of curve network with smooth and coherent orientations, surface reconstruction

This is a joint work with team-project IMAGINE (Tibor Stanko and Stefanie Hahmann) at Inria-Grenoble and CEA-Leti (Nathalie Saguin-Sprynski). This work introduces a novel framework for acquisition and reconstruction of 3D curves using orientations provided by inertial sensors. While the idea of sensor shape reconstruction is not new, we present the first method for creating well-connected networks with cell complex topology using only orientation and distance measurements and a set of user- defined constraints. By working directly with orientations, our method robustly resolves problems arising from data inconsistency and sensor noise. Although originally designed for reconstruction of physical shapes, the framework can be used for "sketching" new shapes directly in 3D space. We test the performance of the method using two types of acquisition devices: a standard smartphone, and a custom-made device. The results have been published in C&G [6]. This paper has been awarded "Best Paper" at the conference Shape Modeling International 2017.

7.6.4. Morphorider: Acquisition and Reconstruction of 3D Curves with Mobile Sensors Participants: Tibor Stanko, Stefanie Hahmann, Georges-Pierre Bonneau, Nathalie Saguin-Sprynski.



Figure 16. Left: the Morphorider and the wireless keypad used for marking nodes during acquisition. Right: screenshots of our acquisition and reconstruction applications.

This is a joint work with team-project IMAGINE (Tibor Stanko and Stefanie Hahmann) at Inria-Grenoble and CEA-Leti (Nathalie Saguin-Sprynski). In this work we introduce a new mobile device called the Morphorider, which is equipped with a 3A3M-sensor node and an odometer for distance tracking. Using this single inertial

measurement unit (IMU), we propose a method to scan physical objects and to reconstruct digital 3D models. By moving the IMU along the surface, a network of local orientation data is acquired together with traveled distances and network topology. We then reconstruct a consistent network of curves and fit these curves by a globally smooth surface. This work has been published in IEEE Sensors [10].

MISTIS Project-Team

7. New Results

7.1. Mixture models

7.1.1. Robust and collaborative extensions of Sliced Inverse Regression.

Participants: Stéphane Girard, Florence Forbes.

This research theme was supported by a LabEx PERSYVAL-Lab project-team grant.

Joint work with: A. Chiancone and J. Chanussot (Gipsa-lab and Grenoble-INP).

Sliced Inverse Regression (SIR) has been extensively used to reduce the dimension of the predictor space before performing regression. Recently it has been shown that this technique is, not surprisingly, sensitive to noise. Different approaches have thus been proposed to robustify SIR. In [16], we investigate the properties of an inverse problem proposed by R.D. Cook and we show that the framework can be extended to take into account a non-Gaussian noise. Generalized Student distributions are considered and all parameters are estimated via an EM algorithm. The algorithm is outlined and tested comparing the results with different approaches on simulated data. Results on a real dataset show the interest of this technique in presence of outliers.

For further improvement of SIR, in his PhD thesis work, Alessandro Chiancone studied the extension of the SIR method to different sub-populations. The idea is to assume that the dimension reduction subspace is not the same for different clusters of the data [17]. One of the difficulties is that standard Sliced Inverse Regression (SIR) has requirements on the distribution of the predictors that are hard to check since they depend on unobserved variables. It has been shown that, if the distribution of the predictors is elliptical, then these requirements are satisfied. In case of mixture models, the ellipticity is violated and in addition there is no assurance of a single underlying regression model among the different components. Our approach clusters the predictors space to force the condition to hold on each cluster and includes a merging technique to look for different underlying models in the data. A study on simulated data as well as two real applications are provided. It appears that SIR, unsurprisingly, is not able to deal with a mixture of Gaussians involving different underlying models whereas our approach is able to correctly investigate the mixture.

7.1.2. Structured mixture of linear mappings in high dimension

Participant: Florence Forbes.

Joint work with: Benjamin Lemasson from Grenoble Institute of Neuroscience, Naisyin Wang and Chun-Chen Tu from University of Michigan, Ann Arbor, USA.

Regression is a widely used statistical tool. A large number of applications consists of learning the association between responses and predictors. From such an association, different tasks, including prediction, can then be conducted. To go beyond simple linear models while maintaining tractability, non-linear mappings can be handled through exploration of local linearity. The non-linear relationship can be captured by a mixture of locally linear regression models as proposed in the so-called Gaussian Locally Linear Mapping (GLLiM) model [6] that assumes Gaussian noise models. GLLiM is based on a joint modeling of both the responses and covariates, observed or latent. This joint modeling allows for the use of an inverse regression strategy to handle the high dimensionality of the data. Mixtures are used to approximate non-linear associations. GLLiM groups data with similar linear association together. Within the same cluster, the association can be considered as locally linear, which can then be resolved under the classical linear regression setting (see Figure 2 (a)). However, when the covariate dimension is much higher than the response dimension, GLLiM may result in erroneous clusters at the low dimension (eg Figure 2 (b)), leading to potentially inaccurate

predictions. Specifically, when the clustering is conducted at a high joint dimension, the distance at low dimension between two members of the same cluster (component) could remain large. As a result, a mixture component might contain several sub-clusters and/or outliers, violating the model Gaussian assumption. This results in a model misspecification effect that can seriously impact prediction performance. A natural way to lessen this effect is to increase the number of components in the mixture making each linear mapping even more local. But this also increases the number of parameters to estimate and therefore requires to be done in a parsimonious manner to avoid over-parameterization. In this work, we propose a parsimonious approach which we refer to as Structured Mixture of Gaussian Locally Linear Mapping (SMoGLLiM) to solve the aforementioned problems. It follows a two-layer hierarchical clustering structure where local components are grouped into global components sharing the same high-dimensional noise covariance structure, which effectively reduces the number of parameters of the model. SMoGLLiM also includes a pruning algorithm for eliminating outliers as well as determining an appropriate number of clusters. Moreover, the number of clusters and training outliers determined by SMoGLLiM can be further used by GLLiM for improving prediction performance. As an extension, a subsetting and parallelization techniques are discussed for the efficiency concern. A preliminary version of this work was presented at the American Statistical Association Joint Statistical Meeting in Baltimore USA in July 2017, [35].



Figure 2. illustration of the GLLiM model: (Top) Non linear relationship approximated as a mixture of locally linear mappings; (Bottom) problematic clustering with a non Gaussian component (orange region) due to unbalanced weights between the high dimensional responses and low dimensional covariates.

7.1.3. Dictionary-free MR fingerprinting parameter estimation via inverse regression

Participants: Florence Forbes, Fabien Boux, Julyan Arbel.

Joint work with: Emmanuel Barbier from Grenoble Institute of Neuroscience.

Magnetic resonance imaging (MRI) can map a wide range of tissue properties but is often limited to observe a single parameter at a time. In order to overcome this problem, Ma et al. introduced magnetic resonance fingerprinting (MRF), a procedure based on a dictionary of simulated couples of signals and parameters. Acquired signals called fingerprints are then matched to the closest signal in the dictionary in order to estimate parameters. This requires an exhaustive search in the dictionary, which even for moderately sized problems, becomes costly and possibly intractable . We propose an alternative approach to estimate more parameters at a time. Instead of an exhaustive search for every signal, we use the dictionary to learn the functional relationship between signals and parameters. This allows the direct estimation of parameters without the need of searching through the dictionary. We investigated the use of GLLiM that bypasses the problems associated with high-to-low regression. The experimental validation of our method is performed in the context of vascular fingerprinting could benefit from a regression approach to limit dictionary size and fasten computation time. Preliminary tests and results have been submitted to ISMRM 2018, International Society for Magnetic Resonance in Medicine.

7.1.4. Semiparametric copula-based clustering

Participants: Florence Forbes, Gildas Mazo, Yaroslav Averyanov.

Modeling of distributions mixtures has rested on Gaussian distributions and/or a conditional independence hypothesis for a long time. Only recently have researchers begun to construct and study broader generic models without appealing to such hypotheses. Some of these extensions use copulas as a tool to build flexible models, as they permit to model the dependence and the marginal distributions separately. Recently [70], a semiparametric copula-based mixture model has been proposed to cluster continuous data. This semiparametric feature allows for more flexibility and reduces the modelling effort for the practitioner. Nonetheless, these advantages come at the cost of assuming that the clusters do not differ in scale. The aim of the internship of Y. Averyanov was to get rid of this assumption by building a nonparametric estimator which have to satisfy certain moment constraints. The performance of the estimator was tested on simulations and then embedded into an EM-like algorithm framework.

7.1.5. Fully automatic lesion localization and characterization: application to brain tumors using multiparametric quantitative MRI data

Participants: Florence Forbes, Alexis Arnaud.

Joint work with: Emmanuel Barbier, Nora Collomb and Benjamin Lemasson from Grenoble Institute of Neuroscience.

When analyzing brain tumors, two tasks are intrinsically linked, spatial localization and physiological characterization of the lesioned tissues. Automated data-driven solutions exist, based on image segmentation techniques or physiological parameters analysis, but for each task separately, the other being performed manually or with user tuning operations. In this work, the availability of quantitative magnetic resonance (MR) parameters is combined with advanced multivariate statistical tools to design a fully automated method that jointly performs both localization and characterization. Non trivial interactions between relevant physiological parameters are captured thanks to recent generalized Student distributions that provide a larger variety of distributional shapes compared to the more standard Gaussian distributions. Probabilistic mixtures of the former distributions are then considered to account for the different tissue types and potential heterogeneity of lesions. Discriminative multivariate features are subsequently pooled together to build a statistical fingerprint model of the different lesion types that captures lesion characteristics while accounting for inter-subject variability. The potential of this generic procedure is demonstrated on a data set of 53 rats, with 36 rats bearing 4 different brain tumors, for which 5 quantitative MR parameters were acquired. This study has been submitted for publication [15].

Analyzing brain tumor tissue composition can then improve the handling of tumor growth and resistance to therapies. We showed on a 6 time point dataset of 8 rats that multiparametric MRI could be exploited via statistical clustering to quantify intra-lesional heterogeneity in space and time. More specifically, MRI can be used to map structural, eg diffusion, as well as functional, eg volume (BVf), vessel size (VSI), oxygen saturation of the tissue (StO2), characteristics. In previous work, these parameters have been analyzed to show the great potential of multiparametric MRI (mpMRI) to monitor combined radio- and chemo-therapies. However, to exploit all the information contained in mpMRI while preserving information about tumor heterogeneity, new methods need to be developed. We demonstrated the ability of clustering analysis applied to longitudinal mpMRI to summarize and quantify intra-lesional heterogeneity during tumor growth. This study showed the interest of a clustering analysis on mpMRI data to monitor the evolution of brain tumor heterogeneity. It highlighted the type of tissue that mostly contributes to tumor development and could be used to refine the evaluation of therapies and to improve tumor prognosis. This work has been presented at ISMRM 2017, International Society for Magnetic Resonance in Medicine [42].

7.1.6. Signature extraction in MR scans for de novo Parkinsonian patients

Participants: Florence Forbes, Veronica Munoz Ramirez, Julyan Arbel.

Joint work with: Michel Dojat from Grenoble Institute of Neuroscience.

This work is part of the cross-disciplinary project NeuroCoG. Parkinson's disease (PD) is characterized by the degeneration of dopaminergic neurons located in the substantia nigra pars compacta (SNc). This leads to well-known motor symptoms associated to Parkinson's disease, rigidity, akinesia and tremor. However, non-motor symptoms also appear. It is of primordial interest to understand these symptoms in order to optimize treatments and diagnose at an early stage the pathology's occurrence. The goal of the PhD work of Veronica Munoz Ramirez is the extraction of specific signatures from MR data of de novo PD patients. We investigated the possibility to use multivariate non-supervised clustering techniques as developed in the PhD thesis of Alexis Arnaud to cluster voxels taking into account interactions between various parameters.

7.1.7. Object-based classification from high resolution satellite image time series with Gaussian mean map kernels

Participant: Stéphane Girard.

Joint work with: C. Bouveyron (Univ. Paris 5), M. Fauvel and M. Lopes (ENSAT Toulouse)

In the PhD work of Charles Bouveyron [65], we proposed new Gaussian models of high dimensional data for classification purposes. We assume that the data live in several groups located in subspaces of lower dimensions. Two different strategies arise:

- the introduction in the model of a dimension reduction constraint for each group
- the use of parsimonious models obtained by imposing to different groups to share the same values of some parameters.

This modelling yielded a supervised classification method called High Dimensional Discriminant Analysis (HDDA)[4]. Some versions of this method have been tested on the supervised classification of objects in images. This approach has been adapted to the unsupervised classification framework, and the related method is named High Dimensional Data Clustering (HDDC)[3]. In the framework of Mailys Lopes PhD, our recent work [22], [23], consists in adapting this work to the classification of grassland management practices using satellite image time series with high spatial resolution. The study area is located in southern France where 52 parcels with three management types were selected. The spectral variability inside the grasslands was taken into account considering that the pixels signal can be modeled by a Gaussian distribution. To measure the similarity between two grasslands, a new kernel is proposed as a second contribution: the a-Gaussian mean kernel. It allows to weight the influence of the covariance matrix when comparing two Gaussian distributions. This kernel is introduced in Support Vector Machine for the supervised classification of grasslands from southwest France. A dense intra-annual multispectral time series of Formosat-2 satellite is used for the classification

of grasslands management practices, while an inter-annual NDVI time series of Formosat-2 is used for permanent and temporary grasslands discrimination. Results are compared to other existing pixel- and objectbased approaches in terms of classification accuracy and processing time. The proposed method shows to be a good compromise between processing speed and classification accuracy. It can adapt to the classification constraints and it encompasses several similarity measures known in the literature. It is appropriate for the classification of small and heterogeneous objects such as grasslands.

7.2. Semi and non-parametric methods

7.2.1. Robust estimation for extremes

Participants: Clément Albert, Stéphane Girard.

Joint work with: M. Stehlik (Johannes Kepler Universitat Linz, Austria and Universidad de Valparaiso, Chile) and A. Dutfoy (EDF R&D).

In the PhD thesis of Clément Albert (funded by EDF), we study the sensitivity of extreme-value methods to small changes in the data and we investigate their extrapolation ability [36], [37]. To reduce this sensitivity, robust methods are needed and we proposed a novel method of heavy tails estimation based on a transformed score (the t-score). Based on a new score moment method, we derive the t-Hill estimator, which estimates the extreme value index of a distribution function with regularly varying tail. t-Hill estimator is distribution sensitive, thus it differs in e.g. Pareto and log-gamma case. Here, we study both forms of the estimator, i.e. t-Hill and t-lgHill. For both estimators we prove weak consistency in moving average settings as well as the asymptotic normality of t-lgHill estimator in the i.i.d. setting. In cases of contamination with heavier tails than the tail of original sample, t-Hill outperforms several robust tail estimators, especially in small sample situations. A simulation study emphasizes the fact that the level of contamination is playing a crucial role. We illustrate the developed methodology on a small sample data set of stake measurements from Guanaco glacier in Chile. This methodology is adapted to bounded distribution tails in [26] with an application to extreme snow loads in Slovakia.

7.2.2. Conditional extremal events

Participant: Stéphane Girard.

Joint work with: L. Gardes (Univ. Strasbourg) and J. Elmethni (Univ. Paris 5)

The goal of the PhD theses of Alexandre Lekina and Jonathan El Methni was to contribute to the development of theoretical and algorithmic models to tackle conditional extreme value analysis, *ie* the situation where some covariate information X is recorded simultaneously with a quantity of interest Y. In such a case, the tail heaviness of Y depends on X, and thus the tail index as well as the extreme quantiles are also functions of the covariate. We combine nonparametric smoothing techniques [67] with extreme-value methods in order to obtain efficient estimators of the conditional tail index and conditional extreme quantiles [61].

7.2.3. Estimation of extreme risk measures

Participant: Stéphane Girard.

Joint work with: A. Daouia (Univ. Toulouse), L. Gardes (Univ. Strasbourg), J. Elmethni (Univ. Paris 5) and G. Stupfler (Univ. Nottingham, UK).

One of the most popular risk measures is the Value-at-Risk (VaR) introduced in the 1990's. In statistical terms, the VaR at level $\alpha \in (0, 1)$ corresponds to the upper α -quantile of the loss distribution. The Value-at-Risk however suffers from several weaknesses. First, it provides us only with a pointwise information: VaR(α) does not take into consideration what the loss will be beyond this quantile. Second, random loss variables with light-tailed distributions or heavy-tailed distributions may have the same Value-at-Risk. Finally, Value-at-Risk is not a coherent risk measure since it is not subadditive in general. A first coherent alternative risk measure is the Conditional Tail Expectation (CTE), also known as Tail-Value-at-Risk, Tail Conditional Expectation or Expected Shortfall in case of a continuous loss distribution. The CTE is defined as the expected loss given that the loss lies above the upper α -quantile of the loss distribution. In [61], we investigate the extreme properties of a new risk measure (called the Conditional Tail Moment) which encompasses various risk measures, such as the CTE, as particular cases. We study the situation where some covariate information is available under some general conditions on the distribution tail. We thus has to deal with conditional extremes (see paragraph 7.2.2).

A second possible coherent alternative risk measure is based on expectiles [18]. Compared to quantiles, the family of expectiles is based on squared rather than absolute error loss minimization. The flexibility and virtues of these least squares analogues of quantiles are now well established in actuarial science, econometrics and statistical finance. Both quantiles and expectiles were embedded in the more general class of M-quantiles [19] as the minimizers of a generic asymmetric convex loss function. It has been proved very recently that the only M-quantiles that are coherent risk measures are the expectiles.

7.2.4. Level sets estimation

Participant: Stéphane Girard.

Joint work with: G. Stupfler (Univ. Nottingham, UK).

The boundary bounding the set of points is viewed as the larger level set of the points distribution. This is then an extreme quantile curve estimation problem. We proposed estimators based on projection as well as on kernel regression methods applied on the extreme values set, for particular set of points [10]. We also investigate the asymptotic properties of existing estimators when used in extreme situations. For instance, we have established that the so-called geometric quantiles have very counter-intuitive properties in such situations [21] and thus should not be used to detect outliers.

7.2.5. Approximate Bayesian inference

Participant: Julyan Arbel.

Joint work with: Igor Prünster, Stefano Favaro.

Approximate Bayesian inference was tackled from two perspectives.

First, from a computational viewpoint, we have proposed an algorithm which allows for controlling the approximation error in Bayesian nonparametric posterior sampling. In [14], we show that completely random measures (CRM) represent the key building block of a wide variety of popular stochastic models and play a pivotal role in modern Bayesian Nonparametrics. The popular Ferguson & Klass representation of CRMs as a random series with decreasing jumps can immediately be turned into an algorithm for sampling realizations of CRMs or more elaborate models involving transformedCRMs. However, concrete implementation requires to truncate the random series at some threshold resulting in an approximation error. The goal of this work is to quantify the quality of the approximation by a moment-matching criterion, which consists in evaluating a measure of discrepancy between actual moments and moments based on the simulation output. Seen as a function of the truncation level, the methodology can be used to determine the truncation level needed to reach a certain level of precision. The resulting moment-matching Ferguson & Klass algorithm is then implemented and illustrated on several popular Bayesian nonparametric models.

In [57], we focus on the truncation error of a superposed gamma process in a decreasing order representation. As in [14], we utilize the constructive representation due to Ferguson and Klass which provides the jumps of the series in decreasing order. This feature is of primary interest when it comes to sampling since it minimizes the truncation error for a fixed truncation level of the series. We quantify the quality of the approximation in two ways. First, we derive a bound in probability for the truncation error. Second, we study a momentmatching criterion which consists in evaluating a measure of discrepancy between actual moments of the CRM and moments based on the simulation output. This work focuses on a general class of CRMs, namely the superposed gamma process, which suitably transformed have already been successfully implemented in Bayesian Nonparametrics. To this end, we show that the moments of this class of processes can be obtained analytically.

Second, we have proposed an approximation of Gibbs-type random probability measures at the level of the predictive probabilities. Gibbs-type random probability measures are arguably the most "natural" generalization of the Dirichlet process. Among them the two parameter Poisson–Dirichlet process certainly stands out for the mathematical tractability and interpretability of its predictive probabilistic models for linguistic applications, e.g., machine learning theory, probabilistic models for linguistic applications, Bayesian nonparametric statistics, excursion theory, measure-valued diffusions in population genetics, combinatorics and statistical physics. Given a sample of size n, in this work we show that the predictive probabilities of any Gibbs-type prior admit a large n approximation, with an error term vanishing as o(1/n), which maintains the same desirable features as the predictive probabilities of the two parameter Poisson–Dirichlet prior.

7.2.6. Bayesian nonparametric posterior asymptotic behavior

Participant: Julyan Arbel.

Joint work with: Olivier Marchal, Stefano Favaro, Bernardo Nipoti, Yee Whye Teh.

In [24], we obtain the optimal proxy variance for the sub-Gaussianity of Beta distribution, thus proving upper bounds recently conjectured by Elder (2016). We provide different proof techniques for the symmetrical (around its mean) case and the non-symmetrical case. The technique in the latter case relies on studying the ordinary differential equation satisfied by the Beta moment-generating function known as the confluent hypergeometric function. As a consequence, we derive the optimal proxy variance for the Dirichlet distribution, which is apparently a novel result. We also provide a new proof of the optimal proxy variance for the Bernoulli distribution, and discuss in this context the proxy variance relation to log-Sobolev inequalities and transport inequalities.

The article [13] deals with a Bayesian nonparametric inference for discovery probabilities: credible intervals and large sample asymptotics. Given a sample of size n from a population of individual belonging to different species with unknown proportions, a popular problem of practical interest consists in making inference on the probability $D_n(l)$ that the (n + 1)-th draw coincides with a species with frequency l in the sample, for any l = 0, 1, ..., n. We explore in this work a Bayesian nonparametric viewpoint for inference of $D_n(l)$. Specifically, under the general framework of Gibbs-type priors we show how to derive credible intervals for the Bayesian nonparametric estimator of $D_n(l)$, and we investigate the large n asymptotic behavior of such an estimator. We also compare this estimator to the classical Good–Turing estimator.

7.2.7. A Bayesian nonparametric approach to ecological risk assessment Participant: Julyan Arbel.

Joint work with Guillaume Kon Kam King, Igor Prünster.

We revisit a classical method for ecological risk assessment, the Species Sensitivity Distribution (SSD) approach, in a Bayesian nonparametric framework. SSD is a mandatory diagnostic required by environmental regulatory bodies from the European Union, the United States, Australia, China etc. Yet, it is subject to much scientific criticism, notably concerning a historically debated parametric assumption for modelling species variability. Tackling the problem using nonparametric mixture models, it is possible to shed this

parametric assumption and build a statistically sounder basis for SSD. We use Normalized Random Measures with Independent Increments (NRMI) as the mixing measure because they offer a greater flexibility than the Dirichlet process. Indeed, NRMI can induce a prior on the number of components in the mixture model that is less informative than the Dirichlet process. This feature is consistent with the fact that SSD practitioners do not usually have a strong prior belief on the number of components. In this work, we illustrate the advantage of the nonparametric SSD over the classical normal SSD and a kernel density estimate SSD on several real datasets[59].

7.2.8. Machine learning methods for the inversion of hyperspectral images

Participant: Stéphane Girard.

Joint work with: S. Douté (IPAG, Grenoble), M. Fauvel (INRA, Toulouse) and L. Gardes (Univ. Strasbourg).

We address the physical analysis of planetary hyperspectral images by massive inversion [58]. A direct radiative transfer model that relates a given combination of atmospheric or surface parameters to a spectrum is used to build a training set of synthetic observables. The inversion is based on the statistical estimation of the functional relationship between parameters and spectra. To deal with high dimensionality (image cubes typically present hundreds of bands), a two step method is proposed, namely K-GRSIR. It consists of a dimension reduction step followed by a regression with a non-linear least-squares algorithm. The dimension reduction is performed with the Gaussian Regularized Sliced Inverse Regression algorithm, which finds the most relevant directions in the space of synthetic spectra for the regression. The method is compared to several algorithms: a regularized version of k-nearest neighbors, partial least-squares, linear and non-linear support vector machines. Experimental results on simulated data sets have shown that non-linear support vector machines is the most accurate method followed by K-GRSIR. However, when dealing with real data sets, K-GRSIR gives the most interpretable results and is easier to train.

7.2.9. Multi sensor fusion for acoustic surveillance and monitoring

Participants: Florence Forbes, Jean-Michel Bécu.

Joint work with: Pascal Vouagner and Christophe Thirard from ACOEM company.

In the context of the DGA-rapid WIFUZ project, we addressed the issue of determining the localization of shots from multiple measurements coming from multiple sensors. The WIFUZ project is a collaborative work between various partners: DGA, ACOEM and HIKOB companies and Inria. This project is at the intersection of data fusion, statistics, machine learning and acoustic signal processing. The general context is the surveillance and monitoring of a zone acoustic state from data acquired at a continuous rate by a set of sensors that are potentially mobile and of different nature. The overall objective is to develop a prototype for surveillance and monitoring that is able to combine multi sensor data coming from acoustic sensors (microphones and antennas) and optical sensors (infrared cameras) and to distribute the processing to multiple algorithmic blocs. As an illustration, the MISTIS contribution is to develop technical and scientific solutions as part of a collaborative protection approach, ideally used to guide the best coordinated response between the different vehicles of a military convoy. Indeed, in the case of an attack on a convoy, identifying the threatened vehicles and the origin of the threat is necessary to organize the best response from all members on the convoy. Thus it will be possible to react to the first contact (emergency detection) to provide the best answer for threatened vehicles (escape, lure) and for those not threatened (suppression fire, riposte fire). We developed statistical tools that make it possible to analyze this information (characterization of the threat) using fusion of acoustic and image data from a set of sensors located on various vehicles. We used Bayesian inversion and simulation techniques to recover multiple sources mimicking collaborative interaction between several vehicles.

7.2.10. Extraction and data analysis toward "industry of the future"

Participants: Florence Forbes, Hongliang Lu, Fatima Fofana, Gildas Mazo, Jaime Eduardo Arias Almeida.

Joint work with: J. F. Cuccaro and J. C Trochet from Vi-Technology company.

Industry as we know it today will soon disappear. In the future, the machines which constitute the manufacturing process will communicate automatically as to optimize its performance as whole. Transmitted information essentially will be of statistical nature. In the context of VISION 4.0 project with Vi-Technology, the role of MISTIS is to identify what statistical methods might be useful for the printed circuits boards assembly industry. The topic of F. Fofana's internship was to extract and analyze data from two inspection machines of a industrial process making electronic cards. After a first extraction step in the SQL database, the goal was to enlighten the statistical links between these machines. Preliminary experiments and results on the Solder Paste Inspection (SPI) step, at the beginning of the line, helped identifying potentially relevant variables and measurements (eg related to stencil offsets) to identify future defects and discriminate between them. More generally, we have access to two databases at both ends (SPI and Component Inspection) of the assembly process. The goal is to improve our understanding of interactions in the assembly process, find out correlations between defects and physical measures, generate proactive alarms so as to detect departures from normality.

7.3. Graphical and Markov models

7.3.1. Fast Bayesian network structure learning using quasi-determinism screening

Participants: Thibaud Rahier, Stéphane Girard, Florence Forbes.

Joint work with: Sylvain Marié, Schneider Electric.

Learning the structure of Bayesian networks from data is a NP-Hard problem that involves an optimization task on a super-exponential sized space. In this work, we show that in most real life datasets, a number of the arcs contained in the final structure can be prescreened at low computational cost with a limited impact on the global graph score. We formalize the identification of these arcs via the notion of quasi-determinism, and propose an associated algorithm that reduces the structure learning to a subset of the original variables. We show, on diverse benchmark datasets, that this algorithm exhibits a significant decrease in computational time and complexity for only a little decrease in performance score.

7.3.2. Robust graph estimation

Participants: Karina Ashurbekova, Florence Forbes.

Joint work with: Sophie Achard, CNRS, Gipsa-lab.

Graphs are an intuitive way of representing and visualizing the relationships between many variables. A graphical model is a probabilistic model whose conditional independence or other measures of relationship between random variables is given by a graph. Learning graphical models using their observed samples is an important task, and involves both structure and parameter estimation. Generally, graph estimation consists of several steps. First of all, we do not know the distribution of the real data. But we can do an assumption about this distribution. Then the measure of relationship between variables we are interested in can be chosen based on this assumption. All these measures of relationship are related with elements of the covariance or precision matrices. After estimating the covariance/precision matrix the

final graph can be constructed based on elements of this matrix. A lot of graph estimation methods rely on the Gaussian graphical model, in which the random vector Y is assumed to be Gaussian. Under this assumption, the most popular method is the graphical lasso (glasso). In practice, data may deviate from the Gaussian model in various ways. Outliers and heavy tails frequently occur. Contamination of a handful of variables in a few experiments can lead to a drastically wrong graph. So one of our objective is to deal with heavy tailed data using a new family of multivariate heavy-tailed distributions [8] and infer a graph robust to outliers without having to remove them.

7.3.3. Spatial mixtures of multiple scaled t-distributions

Participants: Florence Forbes, Alexis Arnaud.

Joint work with: Steven Quinito Masnada, Inria Grenoble Rhone-Alpes

The goal is to implement an hidden Markov model version of our recently introduced mixtures of non standard multiple scaled *t*-distributions. The motivation for doing that is the application to multiparametric MRI data for lesion analysis. When dealing with MRI human data, spatial information is of primary importance. For our preliminary study on rat data [15], the results without spatial information were already quite smooth. The main anatomical structures can be identified. We suspect the reason is that the measured parameters already contain a lot of information about the underlying tissues. However, introducing spatial information is always useful and is our ongoing work. In the statistical framework we have developed (mixture models and EM algorithm), it is conceptually straightforward to introduce an additional Markov random field. In addition, when using a Markov random field it is easy to incorporate additional atlas information.

7.3.4. Spectral CT reconstruction with an explicit photon-counting detector model: a "one-step" approach

Participants: Florence Forbes, Pierre-Antoine Rodesch.

Joint work with: Veronique Rebuffel and Clarisse Fournier from CEA-LETI Grenoble.

In the context of Pierre-Antoine Rodesh's PhD thesis, we investigate new statistical and optimization methods for tomographic reconstruction from non standard detectors providing multiple energy signals. Recent developments in energy-discriminating Photon-Counting Detector (PCD) enable new horizons for spectral CT. With PCDs, new reconstruction methods take advantage of the spectral information measured through energy measurement bins. However PCDs have serious spectral distortion issues due to charge-sharing, fluorescence escape, pileup effect Spectral CT with PCDs can be decomposed into two problems: a noisy geometric inversion problem (as in standard CT) and an additional PCD spectral degradation problem. The aim of this study is to introduce a reconstruction method which solves both problems simultaneously: a one-step approach. An explicit linear detector model is used and characterized by a Detector Response Matrix (DRM). The algorithm reconstructs two basis material maps from energy-window transmission data. For comparison, we also perform a standard two-step approach: an advanced polynomial decomposition of measured sinograms combined with a filtered-back projection reconstruction. The results demonstrate the potential uses of this method for medical imaging or for non-destructive control in industry. Preliminary results will be presented at the SPIE medical imaging 2018 conference in Houston, USA [44].

7.3.5. Non parametric Bayesian priors for hidden Markov random fields

Participants: Florence Forbes, Julyan Arbel, Hongliang Lu.

Hidden Markov random field (HMRF) models are widely used for image segmentation or more generally for clustering data under spatial constraints. They can be seen as spatial extensions of independent mixture models. As for standard mixtures, one concern is the automatic selection of the proper number of components in the mixture, or equivalently the number of states in the hidden Markov field. A number of criteria exist to select this number automatically based on penalized likelihood (eg. AIC, BIC, ICL etc.) but they usually require to run several models for different number of classes to choose the best one. Other techniques (eg. reversible jump) use a fully Bayesian setting including a prior on the class number but at the cost of prohibitive computational times. In this work, we investigate alternatives based on the more recent field of Bayesian nonparametrics. In particular, Dirichlet process mixture models (DPMM) have emerged as promising candidates for clustering applications where the number of clusters is unknown. Most applications of DPMM involve observations which are supposed to be independent. For more complex tasks such as unsupervised image segmentation with spatial relationships or dependencies between the observations, DPMM are not satisfying.

7.3.6. Automated ischemic stroke lesion MRI quantification

Participant: Florence Forbes.

Joint work with: Senan Doyle (Pixyl), Assia Jaillard (CHUGA), Olivier Heck (CHUGA), Olivier Detante (CHUGA) and Michel Dojat (GIN)

Manual delineation by an expert is currently the gold standard for lesion quantification, but is resourceintensive, suffers from inter-rater and intra-rater variability, and does not scale well to large population cohorts. We develop an automated lesion quantification method to assess the efficacy of cell therapy in patients after ischemic stroke. A high-quality sub-acute and chronic stroke dataset was supplied by HERMES. T1-w and 3D-Flair MRIs were acquired from 20 ischemic stroke patients with MCA infarct at 2 and 6 months postevent. Manual delineation was performed by an expert using the Flair image. We propose an unsupervised method employing a hidden Markov random field, with innovations to address the challenges posed by stroke MR scans. We introduce a probabilistic vascular territory atlas, adapted to the patient-specific data in a joint segmentation and registration framework, to model the potential progression and delimitation of vascular accidents. After segmentation, a good correlation is observed between manual and automated lesion volume delineation for the two time points. We therefore propose an unsupervised method with the hypothesis that such a class of methods is more robust to the diversity of images obtained with different sequence parameters and scanners; a particularly sensitive point for multi-center studies. Interestingly, this approach will be used in the European **RESSTORE** cohort.

7.3.7. PyHRF: A python library for the analysis of fMRI data based on local estimation of hemodynamic response function

Participants: Florence Forbes, Jaime Eduardo Arias Almeida, Aina Frau Pascual.

Joint work with: Michel Dojat and Jan Warnking from Grenoble Institute of Neuroscience.

Functional Magnetic Resonance Imaging (fMRI) is a neuroimaging technique that allows the non-invasive study of brain function. It is based on the hemodynamic changes induced by cerebral activity following sensory or cognitive stimulation. The measured signal depends on the variation of blood oxygenation level (BOLD signal) which is related to brain activity: a decrease in deoxyhemoglobin induces an increase in BOLD signal. In fact, the signal is convoluted by the Hemodynamic Response Function (HRF) whose exact form is unknown and fluctuates with various parameters such as age, brain region or physiological conditions. In this work we focused on PyHRF, a software to analyze fMRI data using a joint detection-estimation (JDE) approach. It jointly detects cortical activation and estimates the HRF. In contrast to existing tools, PyHRF estimates the HRF instead of considering it as constant in the entire brain, improving thus the reliability of the results. We investigated a number of real data case to demonstrate that PyHRF was a suitable tool for clinical applications. This implied the definition of guidelines to set some of the parameters required to run the software. We investigated a calibration method by comparing results with the standard SPM software in the case of a fixed HRF. An overview of the package and its performance was presented at the 16th Python in Science Conference (SciPy 2017) in Austin, TX, United States [38].

7.3.8. Hidden Markov models for the analysis of eye movements

Participants: Jean-Baptiste Durand, Brice Olivier.

This research theme is supported by a LabEx PERSYVAL-Lab project-team grant.

Joint work with: Anne Guérin-Dugué (GIPSA-lab) and Benoit Lemaire (Laboratoire de Psychologie et Neurocognition)

In the last years, GIPSA-lab has developed computational models of information search in web-like materials, using data from both eye-tracking and electroencephalograms (EEGs). These data were obtained from experiments, in which subjects had to decide whether a text was related or not to a target topic presented to them beforehand. In such tasks, reading process and decision making are closely related. Statistical analysis of such data aims at deciphering underlying dependency structures in these processes. Hidden Markov models (HMMs) have been used on eye movement series to infer phases in the reading process that can be interpreted as steps in the cognitive processes leading to decision. In HMMs, each phase is associated with a state of the Markov chain. The states are observed indirectly though eye-movements. Our approach was inspired by Simola et al. (2008), but we used hidden semi-Markov models for better characterization of phase length distributions [55]. The estimated HMM highlighted contrasted reading strategies (ie, state transitions), with

both individual and document-related variability. However, the characteristics of eye movements within each phase tended to be poorly discriminated. As a result, high uncertainty in the phase changes arose, and it could be difficult to relate phases to known patterns in EEGs.

This is why, as part of Brice Olivier's PhD thesis, we are developing integrated models coupling EEG and eye movements within one single HMM for better identification of the phases. Here, the coupling should incorporate some delay between the transitions in both (EEG and eye-movement) chains, since EEG patterns associated to cognitive processes occur lately with respect to eye-movement phases. Moreover, EEGs and scanpaths were recorded with different time resolutions, so that some resampling scheme must be added into the model, for the sake of synchronizing both processes.

New results were obtained in the standalone analysis of the eye-movements. A comparison between the effects of three types of texts was performed, considering texts either closely related, moderately related or unrelated to the target topic.

Our goal for this coming year is to develop and implement a model for jointly analyzing eye-movements and EEGs in order to improve the discrimination of the reading strategies.

7.3.9. Markov models for the analysis of the alternation of flowering in apple tree progenies **Participant:** Jean-Baptiste Durand.

This research theme is supported by a Franco-German ANR grant (AlternApp project).

Joint work with: Evelyne Costes (INRA AGAP, AFEF team)

A first study was published to characterize genetic determinisms of the alternation of flowering in apple tree progenies. Data were collected at two scales: at whole tree scale (with annual time step) and a local scale (annual shoots, which correspond to portions of stems that were grown during the same year). One or several replications of each genotype were available.

Three families of indices were proposed for early detection of alternation during the juvenile phase. The first family was based on a trend model and a quantification of the deviation amplitudes and dependency, with respect to the trend. The second family was based on a 2nd-order Markov chain with fixed and random effect in transition probabilities. The third family was based on entropy indices, in which flowering probabilities were corrected from fixed effects using Generalized Linear Models.

This allowed early quantification of alternation from the yearly numbers of inflorescences at tree scale. Some quantitative trait loci (QTL) were found in relation with these indices [40], [20].

New data sets where collected in other F1 progenies. Ancestral relationships between parents of different progenies were taken into account to enhance the power of QTL detection using Bayesian methods. Other QTLs are expected to be found using these new indices and genetic material. However, the amount of replicate per genotype and of data per replicate is quite reduced compared to those of our previous work. This is why we will investigate the loss of power in QTL detection due to a degraded amount of data, by simulating data deletion in our reference results.

MOEX Project-Team

4. New Results

4.1. Cultural knowledge evolution

Agents may use ontology alignments to communicate when they represent knowledge with different ontologies: alignments help reclassifying objects from one ontology to the other. Such alignments may be provided by dedicated algorithms [9], but their accuracy is far from satisfying. Yet agents have to proceed. They can take advantage of their experience in order to evolve alignments: upon communication failure, they will adapt the alignments to avoid reproducing the same mistake.

Such repair experiments have been performed [3] and revealed that, by playing simple interaction games, agents can effectively repair random networks of ontologies.

4.1.1. Expansion and relaxation modalities for cultural alignment repair

Participant: Jérôme Euzenat [Correspondent].

We repeated these experiments and, using new measures, showed that the quality of previous results was underestimated. We introduced new adaptation operators that improve those previously considered. We also allowed agents to go beyond the initial operators in two ways [8]: they can generate new correspondences when they discard incorrect ones, and they can provide less precise answers. The combination of these modalities satisfy the following properties: (1) agents still converge to a state in which no mistake occurs, (2) they achieve results far closer to the correct alignments than previously found, (3) they reach again 100% precision and coherent alignments.

4.1.2. Starting with empty alignments in cultural alignment repair

Participant: Jérôme Euzenat [Correspondent].

The results of §4.1.1 suggest that, with the expansion modality, agents could develop alignments from scratch. We explored the use of expanding repair operators for that purpose. When starting from empty alignments, agents fail to create them as they have nothing to repair. Hence, we introduced the capability for agents to risk adding new correspondences when no existing one is useful [7]. We compared and discussed the results provided by this modality and showed that, due to this generative capability, agents reach better results than without it in terms of the accuracy of their alignments. When starting with empty alignments, alignments reach the same quality level as when starting with random alignments, thus providing a reliable way for agents to build alignment from scratch through communication. The evolution curves of both approaches (random and empty alignments), passed a starting phase in which figures correspond to the initial conditions, superimpose nearly exactly. This comfort a posteriori the experiments with random initialisation.

4.2. Link keys

Link keys ($\S3.3$) are explored following two directions:

- Extracting link keys;
- Reasoning with link keys.

4.2.1. Link key extraction with relational concept analysis

Participants: Manuel Atencia, Jérôme David [Correspondent], Jérôme Euzenat, Jérémy Vizzini.

A first method has been designed to extract and select link keys from two classes which deals with multiple values but not object values [1]. Moreover, the extraction step has been rephrased in formal concept analysis (FCA) allowing to generate link keys across relational tables [2].

We have extended this latter work so that it can deal with multiple object values when the data set is cycle free. This encoding does not necessarily generate the optimal link keys. Hence, we use relational concept analysis (RCA), an extension of FCA taking relations between concepts into account. We show that a new expression of this problem is able to extract the optimal link keys even in the presence of cyclic dependencies. Moreover, the proposed process does not require information about the alignments of the ontologies to find out from which pairs of classes to extract link keys.

We implemented these methods and evaluated them by reproducing the experiments made in previous studies [14]. This shows that the method extracts the expected results as well as (also expected) scalability issues.

4.2.2. Tableau method for ALC+Link key reasoning

Participants: Manuel Atencia [Correspondent], Jérôme David, Jérôme Euzenat.

Link keys can also be thought of as axioms in a description logic. As such, they can contribute to infer ABox axioms, such as links, or terminological axioms, and other link keys. We have extended the tableau method designed for the ALC description logic to support reasoning with link keys in ALC [5]. This year, we have proven that this method is sound, complete and that it always terminates.

MORPHEO Project-Team

7. New Results

7.1. Multi-View Dynamic Shape Refinement Using Local Temporal Integration



Figure 4. 3D Shape Model Refinement

We consider 4D shape reconstructions in multi-view environments and investigate how to exploit temporal redundancy for precision refinement (see Figure 4). In addition to being beneficial to many dynamic multi-view scenarios this also enables larger scenes where such increased precision can compensate for the reduced spatial resolution per image frame. With precision and scalability in mind, we propose a symmetric (non-causal) local time-window geometric integration scheme over temporal sequences, where shape reconstructions are refined framewise by warping local and reliable geometric regions of neighboring frames to them. This is in contrast to recent comparable approaches targeting a different context with more compact scenes and real-time applications. These usually use a single dense volumetric update space or geometric template, which they causally track and update globally frame by frame, with limitations in scalability for larger scenes and in topology and precision with a template based strategy. Our templateless and local approach is a first step towards temporal shape super-resolution. We show that it improves reconstruction accuracy by considering multiple frames. To this purpose, and in addition to real data examples, we introduce a multi-camera synthetic dataset that provides ground-truth data for mid-scale dynamic scenes.

This work was presented at the International Conference on Computer Vision [8].

7.2. Controllable Variation Synthesis for Surface Motion Capture

We address the problem of generating variations of captured 4D models automatically (see Figure 5), and we particularly focus on dynamic human shapes as observed from multi-view videos. Variation is an essential component of motion realism, however recent mesh animation datasets and tools lack such richness. Given a few 4D models representing movements of the same type, our method builds a probabilistic low dimensional embedding of shape poses using Gaussian Process Dynamical Models, and novel variants of motions are



Figure 5. Animation Synthesis with Variability

obtained by sampling trajectories from this manifold using Monte Carlo Markov Chain. We can synthesize an unlimited number of variations of any of the input movements, and also any blended version of them, without costly non-linear interpolation of input movement variations in mesh domain. The output variations are statistically similar to the input movements but yet slightly different in poses and timings. As we show through our results, the generated mesh sequences match the training examples in realism, which facilitates 4D model dataset augmentation.

This work was presented at the International Conference on 3D Vision [6].

7.3. CT from Motion: Volumetric Capture of Moving Shapes with X-rays and Videos



Figure 6. Volumetric Model of a Moving Arm

We consider the capture of dense volumetric X-ray attenuation models of non-rigidly moving samples (see Figure 6). Traditional 3D medical imaging apparatus, e.g. CT or MRI, do not easily adapt to shapes that deform significantly such as a moving hand. We propose an approach that simultaneously recovers dense volumetric shape and motion information by combining video and X-ray modalities. Multiple colour images are captured to track shape surfaces while a single X-ray device is used to infer inner attenuations. The approach does not assume prior models which makes it versatile and easy to generalise over different shapes. Results on synthetic and real-life data are presented that demonstrate the approach feasibility with a limited number of X-ray views. The resulting dense 4D attenuation data provides unprecedented insights for motion analysis.

This work was presented at the British Machine Vision Conference [9].

7.4. Surface Motion Capture Transfer with Gaussian Process Regression



Figure 7. Motion Transfer between Captured 4D Models

We address the problem of transferring motion between captured 4D models (see Figure 7). We particularly focus on human subjects for which the ability to automatically augment 4D datasets, by propagating movements between subjects, is of interest in a great deal of recent vision applications that builds on human visual corpus. Given 4D training sets for two subjects for which a sparse set of corresponding key-poses are known, our method is able to transfer a newly captured motion from one subject to the other. With the aim to generalize transfers to input motions possibly very diverse with respect to the training sets, the method contributes with a new transfer model based on non-linear pose interpolation. Building on Gaussian process regression, this model intends to capture and preserve individual motion properties, and thereby realism, by accounting for pose inter-dependencies during motion transfers. Our experiments show visually qualitative, and quantitative, improvements over existing pose-mapping methods and confirm the generalization capabilities of our method compared to state of the art.

This work was presented at the Conference on Computer Vision and Pattern Recognition [7].

7.5. Dynamic Filters in Graph Convolutional Networks



Figure 8. Shape Registration (left: reference shape, right: matched shapes) with Graph Convolutional Networks

Convolutional neural networks (CNNs) have massively impacted visual recognition in 2D images, and are now ubiquitous in state-of-the-art approaches. While CNNs naturally extend to other domains, such as audio and video, where data is also organized in rectangular grids, they do not easily generalize to other types of data such as 3D shape meshes, social network graphs or molecular graphs. To handle such data, we propose a novel graph-convolutional network architecture that builds on a generic formulation that relaxes the 1-to-1 correspondence between filter weights and data elements around the center of the convolution. The main novelty of our architecture is that the shape of the filter is a function of the features in the previous network layer, which is learned as an integral part of the neural network. Experimental evaluations on digit recognition, semi-supervised document classification, and 3D shape correspondence (see Figure 8) yield state-of-the-art results, significantly improving over previous work for shape correspondence.

This work was published as research report [11].

NANO-D Project-Team

5. New Results

5.1. Incremental methods for long range interactions

Participants: Semeho Edorh, Stephane Redon.

Adaptively Restrained Molecular Dynamics (**ARMD**) were recently proposed with the purpose of speeding up molecular simulations. The main idea is to modify the Hamiltonian such that the kinetic energy is set to zero for low velocities, which allows to save computational time since particles do not move and forces need not be updated.

We continued our work on developing an extension of **ARMD** to electrostatic simulations. Therefore, we developed a fast method dedicated to the computation of the electrostatic potential in adaptively restrained systems. The proposed algorithm is derived from a multigrid-based alternative to the popular particle mesh methods. Our algorithm ,labeled as Incremental Mesh Continuum Method (IMCM),was implemented inside LAMMPS, a popular molecular dynamics simulation package. During ARMD simulations, IMCM scales with the number of active particles.

The performance of the new algorithm was accessed on various molecular systems. 1 showed that IMCM is able to outperform the well-established Particle Particle Particle Mesh (P3M) for adaptively restrained simulations. For an aqueous solution of sodium chloride, water molecules can be adaptively restrained. On this system, ARMD was able to reproduce static properties of sodium chloride (2). When a functionalized nanopore is placed at the center of the system, ARMD and IMCM were able to reproduce the ion selectivity property (3). For this benchmark, this positively charged nanopore acts like a sieve that blocks the flux of Sodium atoms, while promoting the crossing of Chlorine particles(4).

5.2. Adaptive Algorithms for Orbital-Free Density Functional Theory

Participants: Francois Rousse, Stephane Redon.

Our Orbital-Free Density Functional Theory (OF-DFT) program has been enriched and improved. Its accuracy has been demonstrated just like in he work of [79] : by comparison with the PROFESS software [47] of energies and relaxed geometries of several aluminium clusters. An incremental version has been developed and tested: it can be tuned smoothly from fast and approximative to slow and precise. We have shown that for cases where few particles positions have changed, like in an adaptively restrained dynamical [23], the update of electronic density is faster with the adaptive version.

A SAMSON App computing electronic energies through Orbital-Free DFT has been released for SAMSON 0.6.0. It is a light version, with only one optimization algorithm and no adaptive version. The parallel implementation is available for most operating systems. Another SAMSON App is being developed to generate the input files of PROFESS, our reference software for OF-DFT.

5.3. As-Rigid-As-Possible molecular interpolation paths

Participants: Minh Khoa Nguyen, Jaillet Leonard, Stephane Redon.

We proposed a new method to generate interpolation paths between two given molecular conformations [11]. It relies on the As-Rigid-As- Possible (ARAP) paradigm used in Computer Graphics to manipulate complex meshes while preserving their essential structural characteristics. Experiments conducted on a large set of benchmarks show how such a strategy can efficiently compute relevant interpolation paths with large conformational rearrangements.



Figure 1. Performance depending on the percentage of active particles for different number of processes. Performance of LAMMPS P3M is shown as a reference (dotted lines, pentagram marker) — it does not depend on the percentage of active particles. In all cases electrostatics were computed at similar accuracy ($\sim 10^{-5}$).



Figure 2. Ion-water pair distribution functions using armd with the NaCl/ ϵ force field at 298 K the rigid water model SPC/ ϵ and an ionic concentration of 10.0 molal. Different restraining parameters (ε^r , ε^f) were tested on water molecules. Na and Cl are always active. Black line corresponds to a standard molecular dynamics simulation of the system.



Figure 3. Number density of chlorine (red dotted line) and sodium (blue dashed line) ions along z-axis using standard ARMD (Left) and MD (Right). Both methods show the ion selectivity of the nanopore which is located at z = 0.

5.4. ART-RRT: As-Rigid-As-Possible Exploration of Ligand Unbinding Pathways

Participants: Minh Khoa Nguyen, Jaillet Leonard, Stephane Redon.

We proposed a method to efficiently generate approximate ligand unbinding pathways (to appear in the Journal of Computational Chemistry). It combines an efficient tree-based exploration method with a morphing technique from Computer Graphics for dimensionality reduction. This method is computationally cheap and, unlike many existing approaches, does not require a reaction coordinate to guide the search. It can be used for finding pathways with known or unknown directions beforehand. The approach is evaluated on several benchmarks and the obtained solutions are compared with the results from other state-of-the-art approaches. We show that the method is time-efficient and produces pathways in good agreement with other state-of-the-art solutions. These paths can serve as first approximations that can be used, analyzed or improved with more specialized methods.

5.5. IM-UFF: extending the Universal Force Field for interactive molecular modeling

Participants: Jaillet Leonard, Svetlana Artemova, Stephane Redon.

IM-UFF, the extension of UFF to interactive modeling was completed. It led to an analysis demonstrating that IM-UFF allows to obtain statistical measures that are normally only accessible to reactive force fields (cf. Figure 7). It resulted in the paper "IM-UFF: extending the Universal Force Field for interactive molecular modeling" published at the Journal of Molecular Graphics and Modelling [7]. IM-UFF will be proposed as a module available to all public in the 0.7.0 version of SAMSON that will be released soon.

5.6. Exploring Chemical Reaction Paths

Participant: Jaillet Leonard.



Figure 4. Nonuniform distributions of number density of chlorine (Top) and sodium (Bottom) ions driven by an external electric field (black arrow) E = 1 V/A using standard MD (Left) and ARMD (Right). The gray rectangles at z = 0 mark the graphene sheet. Both ions form concentration polarization layers.


Figure 5. Motion of Adenylate Kinase from "open" to "close" state generated with the ARAP method.



Figure 6. Paths (in colored sticks) obtained by ART-RRT for the unbinding of retinoic acid hormone from its receptor. The protein is represented by ribbons and the ligand by orange balls. Two different views are shown for clarity. The left picture shows pathways I in red, II in blue, III in green and Other in black. The right picture shows pathways IV in yellow, V in purple and VI in cyan. These main pathways are also reported by other studies by the SMD and RAMD methods for nuclear hormone receptors.



Figure 7. The interaction between 10 methane molecules restrained in a fixed volume is simulated through Monte Carlo simulation (left). The radial distribution functions (RDF) obtained with Brenner (blue curves) are qualitatively the same as those obtained with IM-UFF (red curves), in case of a 500 K temperature setup (top) as well as a 7000 K temperature setup (bottom). It means that, it might be possible to use IM-UFF to obtain statistical measures that are normally only accessible to reactive force fields.

5.6.1. Context

In the past, we have developed a methodology to explore chemical reaction paths based on stochastic trees. One difficulty was the assessment of the quality of the paths found, and the comparison with existing state of the art methods.

To address these limitations, we have developed several new modules in SAMSON that propose state of the art methods and helpful tools to find and manipulate the paths and the important states of the considered systems. One can classify these modules into three categories: interpolation methods, minima and saddle point search methods and supporting tools. More details regarding these modules are provided below.

5.6.2. Interpolation methods

We have implemented the artificial force induced reaction (AFIR) method [59], that helps to find a transition path from a given initial state made of two compounds A and B, towards a goal compound X. In the futur, we would like to combine AFIR with our exploration methods.

The Linear Synchronous Transit (LST) and the Quadratic Synchronous Transit (QST) methods [42] have also been integrated in SAMSON. These methods generate paths such that each atom-pair distance in an intermediate structure is the interpolated value between those in the initial and target structures. The QST variant differs of the LST one from that the interpolated path also passes through a third intermediate point. Moreover, we have implemented the MINIMAX method as in [50] that alternates phases of minimization and QST interpolation to search for a transition path. These three methods appear to show a better behavior than a simple linear interpolation approach. In the future, we are also planning to combine them with our exploration methods.

5.6.3. Search for minima and saddle points

We have developed a SAMSON module to describe the energy basins associated to the various conformations of a given set. For this, conformations are first minimized and then clustered. This tool is also convenient when analyzing conformations along a given path to search for the states the closest to the saddle points.

In collaboration with the post doc Clément Betone, we have implemented two modules. First, the Dimer method as proposed in [45]. This method allows to find from a given point the closest saddle points in high dimensional potential surfaces, using only first derivatives. Some works remains to be done to solve the sensibility problem related to the initial orientation of the dimmer. Second, the freezing string method as proposed in [24]. This method searches efficiently for a saddle point between two given end states through a combination of interpolatoin and optimization.

To assess the various methods and obtain a visual feedback of their behavior, we have developed several 2D force fields where the energetical landscape can easily be apprehended. Such an example of landscape is illustrated in Figure 8.



Figure 8. Example of 2D force field used to test the various methods performing force field searches. One can see at the bottom right, the current state represented by a red cross and at a local minimum

5.6.4. Supporting tools

We have proposed in SAMSON a set of tools to manipulate and perform some measures associated to set of conformations, that could potentially represent molecular paths.

A first tool was proposed to perform given measures for a given set of conformations, such as bond lengths, angle bends, or torsion angles. A second tool allows to align various conformations onto a reference on. Finally, another tool allows to compute the RMSD distance between pairs of conformations.

5.7. Combination of force fields

We have proposed in SAMSON several modules related to force fields, that can be used as soft constraints applied to a system. We have proposed a force field that sums up the contributions of two given force fields for a given system. Another force field was developed based on three force fields, where the first one concerns one part of the system, the second one an another part and the last one the interaction between these two parts. We have also added elementary force fields for the standard interactions that may be represented classical force fields: bond springs, angle bend springs, torsion springs, Steric clash and Van der Waals interactions.

5.8. Simulating nanomaterials

We have initiated an informal collaboration with Cyril Guedj, a permanent researcher of the Leti at CEA. This researcher is an expert in nanomaterials and the goal is to develop in SAMSON tools to manipulate, simulate and measure these nanomaterials. We focused in particular in crystals that appears in many new materials such as semiconductors.

So far, we have developed three versions of Keating for simulating crystals: an harmonic Keating force field for elements of type IV based on [55], an harmonic Keating force field extended to elements of type III and V based on [30] and a non-harmonic Keating force field based on [74]. An example of crystal simulated with Keating force field in SAMSON is shown in Figure 9 Currently, we are in a phase of validation of these force fields. Cyril Guedj as experts in nanomaterials is focusing on the calibration of the force field parameters and the comparison with state of the art results. In the future, we plan to extend the functionality of these force fields to address more complex scenarii.

In parallel, we have developed a set of tools to manipulate and measure these crystal: a module to rescale them, one to simplify their visualization, one to measure a set of characteristics, and finally, one to simulate constraints applied to some part of the system.



Figure 9. Example of crystal simulated in SAMSON, using one variant of the Keating force field.

5.9. Parallel algorithms for adaptive molecular dynamics simulations

Participants: Dmitriy Marin, Stephane Redon.

We worked on the development and improvement of a parallel implementation of the Adaptively Restrained Molecular Dynamics (ARMD) method in the LAMMPS molecular dynamics simulator. The parallelization was done in application to multi-core CPU and hybrid CPU/GPU systems thanks to the Kokkos package provided by LAMMPS. The ARMD can be used for decreasing computational complexity by restraining degrees of freedom for some particles in the simulated system [23], therefore allowing to gain speed-up by either decreasing precision or focusing on select subsystems. The developed parallel implementation allows us to run LAMMPS with an ARMD integrator on central processing units (CPU), graphics processing units (GPU), and many integrated core architecture (MIC). We developed a new algorithm for processing particles that switches their state from a restrained state to a full-dynamics state and vice versa. The new algorithm is modified for efficient usage of GPU and many-core CPUs (computations are performed on a computational device, communications between host and device are decreased). The results on performance and speed-up for ARMD in comparison with the non-modified LAMMPS for a standard Lennard–Jones liquid benchmark are shown in Figures 10 and 11. We showed that starting from some number of atoms in the system and from some percentage of restrained atoms, ARMD provides better performance over classical MD. The results are published in [14].

5.10. Development of Convex-PL, a scoring function for protein-ligand interactions

Participants: Sergei Grudinin, Maria Kadukova.



Figure 10. Performance for a system with ARMD parameters: $\varepsilon^r = 4.5$, $\varepsilon^f = 6.5$



Figure 11. Speed-up for a system with 1 372 000 atoms



Figure 12.

We have continued developing Convex-PL, which is a knowledge-based scoring function for protein-ligand interactions. It is based on the assumption that protein-ligand interactions linearly depend on radial distributions of pairs of atoms of various types. The corresponding coefficients are deduced with a convex optimization problem from the structural data. We augmented Convex-PL with a term standing for steric clashes, ran a leave-one-out cross-validation procedure, and additionally validated it on the D3R Grand Challenge 2 usersubmitted poses. The corresponding paper [8] was accepted to the Journal of Computer-Aided Molecular Design. We have also presented Convex-PL at two scientific conferences [53], [52].

5.11. Participation in the D3R Grand Challenge 2 with the Convex-PL scoring function

Participants: Sergei Grudinin, Maria Kadukova.



Figure 13. Structural heterogeneity of the target FXR proteins.

We have participated in the pose prediction stage of the D3R Grand Challenge 2 using Convex-PL to re-score the ligand poses obtained with Autodock Vina. The target protein of this challenge was a farnesoid X receptor (FXR). After the correct co-crystal poses were released, we carefully repeated the experiments and compared them with several other docking protocols. For these protocols we have used a modified version of Autodock Vina with Convex-PL as a built-in scoring function used in sampling. The protocols we have tried include comparison of docking to various co-crystal and mutated co-crystal FXR structures, as well as docking to the correct protein structures to evaluate the influence of receptor flexibility and docking of ligand structures generated with two different algorithms, as well as the co-crystal ones. This study was published in the Journal of Computer-Aided Molecular Design. [9]

5.12. Development of a Normal Modes Analysis SAMSON Element

Participants: Yassine Naimi, Alexandre Hoffmann, Sergei Grudinin.

We developed a SAMSON Element based on the method proposed by Alexandre Hoffman and Sergei Grudinin on Linux and Mac operating systems. This SAMSON Element was implemented in two versions, a Lite version and an Advanced version.



Figure 14. Normal Modes Analysis element Lite version.

The first one (14) computes the nonlinear normal modes of a molecular system (protein, RNA, DNA) very quickly using the NOLB algorithm developed by Alexandre Hoffmann and Sergei Grudinin (J. Chem. Theory Comput., 2017, 13 (5), pp 2123-2134, DOI: 10.1021 / acs.jctc.7b00197.). The user indicates the desired number of modes, the interactions cutoff distance and the potential function. For now, the elastic network model potential is the one that is available but more potential functions, like the Gaussian network model, will be added in the future. In the output, each mode is represented by a slider. The user can visualize the motion of each mode independently by moving its corresponding slider manually or by checking its checkbox and then pressing on the play button. Also, the user can visualize the motion of a combination of modes selecting them before playing the motion.

The transformations used in this motion can be set to linear or nonlinear and the amplitude of the motion can be increased/decreased by changing the scaling factor. During this motion, the user can activate a real time minimization using one of the provided algorithms (steepest descent, conjugated gradient or LBGF) and defined values of minimization steps and minimization tolerance. Finally, the user can either save/export a given conformation of the structure or the entire displayed trajectory by going into the "Save Frames" tabulation of the SAMSON element.



Figure 15. Normal Modes Analysis element Advanced version (A)

The second one (15) is an advanced version of the Nonlinear Normal Modes Analysis module. It computes the nonlinear normal modes of a molecular system (protein, RNA, DNA) in the same time given that it uses the same algorithm developed by Alexandre Hoffmann and Sergei Grudinin (J. Chem. Theory Comput., 2017, 13 (5), pp 2123-2134, DOI: 10.1021 / acs.jctc.7b00197.). It has the same functionalities than the Lite version of the Normal Modes Analysis element. In addition, the element has an additional tabulation called "Structure Definition" (16). In this section, users can define a pocket and ask for a combination of modes that contribute the most to the opening and closing of this pocket. Also, in the near future, another functionality will be added in the "Structure Definition" tabulation called "reference structure definition". Using this functionality, users can define the conformation of structure as a reference and the element will provide a combination of modes that will lead this structure to present this conformation.

5.13. Development of a Hex SAMSON Element

Participants: Yassine Naimi, Sergei Grudinin.

We developed a SAMSON Element to wrap Hex (17), an interactive protein docking software, written by Dave Ritchie (LORIA/Inria Nancy). With this element, users can define receptor/ligand structures displayed



Figure 16. Normal Modes Analysis element Advanced version (B)



Figure 17. Hex element.

in the SAMSON viewport or import their PDB files. All the options available on the Hex software have been implemented in this SAMSON element.



Figure 18. Hex results.

Then, by clicking on the Run button, docking solutions will be computed and clustered in a table (18) using the Hex algorithm. Users can show the resulted docking solutions by clicking on the play button, or the solution line in the table or the next/previous buttons.

5.14. Development of RDKit Smiles Manager SAMSON Elements

Participant: Yassine Naimi.

We integrated RDKit, an open-source collection of cheminformatics and machine-learning software written in C++ and Python, in SAMSON. One of RDKit's features is the conversion of molecules from their SMILES code to a 2D and 3D structures. Therefore, it is now possible to use these features in the SAMSON platform. SMILES code files (.smi) or text files (.txt) containing several SMILES codes can be read using the import button (19).

Users can manage the imported data with several manners (modify the SMILES code, add manually a new SMILES code, assign names to the molecules...). Also, 2D depictions of the SMILES code are generated on the fly (20). When a SMILES code is invalid an error image is automatically generated. By right-clicking on these images, users can open or generate the 3D structure in SAMSON or save the image as png or svg. The main feature of this element is to generate 3D structures from imported/written SMILES codes. After selecting the molecules, users can click on the Generate 3D structures button. Few seconds later, the 3D structures of the molecules (presenting a valid SMILES code) are added to the SAMSON document view. Finally, RDKit provides a feature to filter the selected molecules using a substructure pattern (SMILES or SMARTS). By default in RDKit, information about stereochemistry is not used in substructure searches but this can be changed by using the chirality. For information, the name of the molecules that did not include the given pattern are displayed in a pop-up.

5.15. Symmetry mate generator for SAMSON

Participants: Guillaume Pagès, Sergei Grudinin.



Figure 19. RDkit Smiles Manager element.

•	RDKi	it-SMILES Manager		■ 🔒 ? 🗙
Smiles Code Import From File: Cr/Users/naimi/Desktop/smileFile.smi				Import
	Sel	ect All		New line Clear table
		Name	Code	2D Conformation
	1	☑ c1cccn(=O)c1	c1cccn(=O)c1	<pre></pre>
	2	☑ C==	C==	Syntax error in SMILES string
Substructure Filtering				
Generate 30 structures				

Figure 20. RDkit Smiles Manager element.



Figure 21. The Symmetry Mate element.

Many biological systems are composed of several identical units structurally organized in a symmetric manner. To observe the atomistic contacts in a system, one needs to replicate the asymmetric unit and apply specific rigid-body operations to it. We developed a SAMSON element that is able to read these operations from a file, and which provides a way to replicate the subunits in a user-friendly fashion.

5.16. Development of a symmetry detection software AnAnaS

Participants: Guillaume Pagès, Sergei Grudinin.

Macromolecules are generally not rigid bodies at physiological temperature and they adopt different conformational states. Thus, if one considers a macromolecular assembly made of N subunits, do we expect that all the units will be structurally identical to each other? Most probably not, since at any given moment of time, each unit may be sampling a different conformational state. For example, there are plenty of X-ray structures of homo-dimers, where the individual monomers are not structurally identical.

In order to quantitatively assess these differences, we developed a method for Analytical Analysis of Symmetries (AnAnaS) in protein complexes. The method is extremely fast, robust and accurate. Two manuscripts describing the method are currently submitted for publication. This method is available on the website of the team (https://team.inria.fr/nano-d/software/ananas/).

5.17. Integration of AnAnaS to SAMSON

Participants: Guillaume Pagès, Sergei Grudinin.

We created a SAMSON element to make the symmetry detection and symmetry axes' visualization as easy and intuitive as possible.

5.18. Deep Learning for Symmetry detection

Participants: Guillaume Pagès, Sergei Grudinin.

We are working on a fully-structural method for detecting symmetries in molecular structures. This will allow us to detect tandem repeats, or even symmetry in density maps. We created a method based on neural network and deep learning, inspired by the advances in computer vision in the past decade. According to our first tests on simulated examples, our method is able to detect the order of a cyclic symmetry (which can be 1 for asymmetric structure) with a 92% accuracy, and guesses the direction of the axis of symmetry with an average error of 3° . We are still working on improving it and doing tests on more realistic examples.



Figure 22. System with an octahedral symmetry, with the axes displayed in SAMSON.

5.19. Pepsi-SAXS calculator of small-angle X-ray scattering profiles

Participants: Sergei Grudinin, Maria Garkavenko.

We have continued the development of a new method called Pepsi-SAXS that calculates small angle X-ray scattering profiles from atomistic models [2]. The method is based on the multipole expansion scheme and is significantly faster with a comparable precision than other methods.

Our method has been highlighted in a recent SAXS-related review [49] and was one of the best performers in the recent CASP12 data-assisted protein structure prediction experiment [81].

Pepsi-SAXS is available at http://team.inria.fr/nano-d/software/pepsi-saxs. A SAMSON module will be made available at https://www.samson-connect.net.

5.20. NOLB nonlinear normal modes

Participants: Alexandre Hoffmann, Sergei Grudinin.

We developed a new conceptually simple and computationally efficient method for nonlinear normal mode analysis called NOLB [4]. This is a logical evolution of the RTB-subspace method developed by Y.-H. Sanejouand and colleagues [38], [80]. We demonstrated how to physically interpret the eigenvalues computed in the RTB basis in terms of angular and linear velocities applied to the rigid blocks and how to construct a nonlinear extrapolation of motion out of these velocities. The key observation of our method is that the angular velocity of a rigid block can be interpreted as the result of an implicit force, such that the motion of the rigid block can be considered as a pure rotation about a certain center.

Overall, our method produces better structures compared to the standard approach, especially at large deformation amplitudes, as we demonstrate by visual inspection, energy and topology analyses, and also by the MolProbity service validation. Also, our method is scalable and can be applied to very large molecular systems, such as ribosomes.

Standalone executables of the NOLB normal mode analysis method are available at https://team.inria.fr/nanod/software/nolb-normal-modes/. A graphical user interface created for the SAMSON software platform will be made available at https://www.samson-connect.net.

5.21. Applications of the NOLB NMA method to structural biology

Participant: Sergei Grudinin.



Figure 23.



Figure 24.

Participants: Sergei Grudinin

Using the created nonlinear normal mode analysis NMA tool, we have successfully predicted structural transitions in several protein complexes, whose structures were solved by our collaborators [6], [3].

5.22. Off-grid fitting method

Participants: Alexandre Hoffmann, Sergei Grudinin.





We developed a novel Fast Fourier Transform (FFT)-based exhaustive search method extended to off-grid translational and rotational degrees of freedom [5]. The method combines the advantages of the FFT-based exhaustive search, which samples all the conformations of a system under study on a grid, with a local optimization technique that guarantees to find the nearest optimal off-grid conformation. The method is demonstrated on a fitting problem and can be readily applied to a docking problem.

The algorithm first samples a scoring function on a six-dimensional grid of size N^6 using the FFT. This operation has the asymptotic complexity of $O(N^6 \log N)$. Then, the method performs the off-grid search using a local quadratic approximation of the cost function and the trust region optimization algorithm. The computation of the quadratic approximation is also accelerated by FFT at the same additional asymptotic cost of $O(N^6 \log N)$. We demonstrate our method on fitting atomic protein models into several simulated and experimental maps from cryo-electron microscopy. The method is available at https://team.inria.fr/nano-d/software/offgridfit.

5.23. RapidRMSD library

Participants: Emilie Neveu, Petr Popov, Sergei Grudinin.

The root mean square deviation (RMSD) is one of the most used similarity criteria in structural biology and bioinformatics. Standard computation of the RMSD has a linear complexity with respect to the number of atoms in a molecule, making RMSD calculations time-consuming for the large-scale modeling applications, such as assessment of molecular docking predictions or clustering of spatially proximate molecular conformations. Previously we introduced the *RigidRMSD* algorithm to compute the RMSD corresponding to the



Figure 26.

rigid-body motion of a molecule [62]. Recently, we went beyond the limits of the rigid-body approximation by taking into account conformational flexibility of the molecule. We model the flexibility with a reduced set of collective motions computed with e.g. normal modes or principal component analysis.

The initialization of our algorithm is linear in the number of atoms and all the subsequent evaluations of RMSD values between flexible molecular conformations depend only on the number of collective motions that are selected to model the flexibility. Therefore, our algorithm is much faster compared to the standard RMSD computation for large-scale modeling applications. The method can be applied e.g. to cluster flexible docking or to generate pseudo-random constant-RMSD structural molecular ensembles.

The algorithm is written in C++ as the open-source *RapidRMSD* library governed by the BSD-compatible license, which is available at http://team.inria.fr/nano-d/software/RapidRMSD/. The constant-RMSD structural ensemble application is available at http://team.inria.fr/nano-d/software/nolb-normal-modes/.

5.24. SBROD protein quality assessment method

Participants: Mikhail Karasikov, Sergei Grudinin.

Protein quality assessment (QA) is a crucial element of protein structure prediction, a fundamental but yet open problem in structural bioinformatics. QA aims at ranking predicted protein models from a set of proposed candidates. Although consensus-model QA methods often outperform single-model QA methods, their performance substantially depends on the pool of available candidates. This makes single-model QA methods a particularly important research target since these usually assist in the sampling of candidates.

We developed a novel single-model QA method called SBROD. The SBROD (Smooth Backbone-Reliant Orientation-Dependent) method uses only the conformation of the protein backbone, and hence it can be applied to scoring the coarse-grained protein models. The proposed method deduces the scoring function from a training set of protein models. This function is composed of four terms related to different structural features, residue-residue orientations, contacts between the backbone atoms, hydrogen bonding, and solvent-solvate interactions. The SBROD scoring function is smooth with respect to atomic coordinates and thus is applicable to continuous gradient-based optimization of protein conformations. Furthermore, it can also be used for coarse-grained protein modeling and computational protein design. Computational experiments conducted on diverse datasets (Stage1 and Stage2 from CASP11, and MOULDER) proved SBROD to achieve the state-of-the-art performance among single-model QA methods including meta algorithms.

The standalone application implemented in C++ and Python is freely available at https://team.inria.fr/nanod/software/SBROD and supported on Linux, MacOS, and Windows.



Figure 27.

5.25. Symmetry detection methods

Participants: Etienne Bamas, Sergei Grudinin.

We have developed a novel framework for the computational detection of point-group symmetries in electron density maps. The method is based on the symmetry-based reduced representation of the density using polynomial expansions.

5.26. SAXS-assisted protein docking

Participants: Gaurav Dhar, Sergei Grudinin.

We have developed an extension of the Pepsi-SAXS method [2] applicable to rescoring of rigid-body proteinprotein docking predictions.

5.27. Methods for the estimation of collective motions

Participants: Robin Gullo, Sergei Grudinin.

We have studied novel ways to predict structural conformational transitions in macromolecules.

5.28. Smoothed-force energy optimization

Participants: Clement Beitone, Stephane Redon.

Many approaches have been developed during the last decades to improve the speed of convergence of optimization methods used to find minima of potential energy surfaces. We proposed to spatially and temporally smooth the force vector given by the force field. Our approach alters the deformation of the structure being minimized and makes it behave as if it was locally more rigid. We apply this filtering method to two well-known optimization methods, steepest descent and FIRE, and evaluate its efficiency on several benchmarks, including nanomaterials and biomolecules. We demonstrated that the smoothed force variants may significantly speed up energy minimization.

5.29. Adaptively Restrained Molecular Dynamics in LAMMPS

Participants: Krishna Kant Singh, Stephane Redon.

Adaptively Restrained Molecular Dynamics (ARMD) is a recently introduced particles simulation method that switches positional degrees of freedom on and off during simulation in order to speed up calculations. In the NVE ensemble, ARMD allows users to trade between precision and speed while, in the NVT ensemble, it makes it possible to compute statistical averages faster. Despite the conceptual simplicity of the approach, however, integrating it in existing molecular dynamics packages is non-trivial, in particular since implemented potentials should a priori be rewritten to take advantage of frozen particles and achieve a speed-up. We proposed novel algorithms for integrating ARMD in LAMMPS, a popular multipurpose molecular simulation package [12]. In particular, we demonstrated how to enable ARMD in LAMMPS without having to reimplement all available force fields. The proposed algorithms were assessed on four different benchmarks, and showed how they allowed us to speed up simulations up to one order of magnitude.

5.30. Single-pass Incremental Force Updates for Adaptively Restrained Molecular Dynamics

Participants: Krishna Kant Singh, Stephane Redon.

We proposed new, single-pass incremental force updates algorithms to efficiently simulate a system using ARMD [13]. We assessed different algorithms for speedup measurements and implemented them in the LAMMPS MD package. We validated the single-pass incremental force update algorithm on four different benchmarks using diverse pair potentials. The proposed algorithm allows us to perform simulation of a system faster than traditional MD in both NVE and NVT ensembles. Moreover, ARMD using the new single-pass algorithm speeds up the convergence of observables in wall-clock time.

5.31. Auto update process for SAMSON & SAMSON-SDK

Participant: Jocelyn Gate.

Since SAMSON 0.6.0, instead of manually installing the latest SAMSON updates, the existing SAMSON can keep itself up-to-date automatically. If an internet connection is established and as soon as we add a new version of SAMSON on SAMSON Connect, all users that launch SAMSON will be notified that a new version is available. The previous version will remain running until SAMSON is closed, but the updated version will be launched automatically the next time you start SAMSON. It is a one click process.

5.31.1. The SAMSON auto update

When a new SAMSON is available users have a notification and a summary if all elements they use are still available in the updated version.

5.31.2. The SAMSON-SDK auto update

When a new SAMSON-SDK is available users have notification and they can install it in one click when SAMSON starts.

5.32. SAMSON Elements policies

Participants: Jocelyn Gate, Stephane Redon.

Some developers of SAMSON wanted to restrain access to their developed SAMSON Elements, hence we defined four different access policies:

- Private : only the developer and the collaborators can see and use the corresponding published element
- Public : everyone can see and use the corresponding published element
- Hidden : only users that get the corresponding hidden link can see and use the corresponding published element
- Shared : only the users that have been added to the list of shared users can see and use the corresponding published element



Figure 28. The update notification



Figure 29. The available element summary



Figure 30. The update progress



Figure 31. The policy configuration

5.33. Improvements to our software development pipeline

Participant: Jocelyn Gate.

In order to fully automate the deployment process we increased the number of jenkins features.

- SAMSON Element packaging and deployment for every NANO-D users to samson-connect.
- SAMSON & SAMSON-SDK documentation build.
- SAMSON & SAMSON-SDK documentation upload via FTP to the new documentation website. (https://documentation.samson-connect.net/)

5.34. SAMSON Connect Forum

Participants: Jocelyn Gate, Stephane Redon.

To help the community to use SAMSON and develop elements with the SAMSON SDK, we setup a forum thanks to Jean-Francois Scariot from Inria (https://forum.samson-connect.net/). Despite the fact that the samson-connect.net and forum.samson-connect.net websites are separate, the login functionality is shared, since the user logs in samson-connect.net to access the forum.

NECS Project-Team

7. New Results

7.1. Networks: modeling, analysis and estimation

7.1.1. Cyber-Physical Systems: a control-theoretic approach to privacy and security

Participants: A. Kibangou [Contact person], F. Garin, S. Gracy, H. Nouasse.

Cyber-physical systems are composed of many simple components (agents) with interconnections giving rise to a global complex behaviour. Interesting recent research has been exploring how the graph describing interactions affects control-theoretic properties such as controllability or observability, namely answering the question whether a small group of agents would be able to drive the whole system to a desired state, or to retrieve the state of all agents from the observed local states only.

A related problem is observability in the presence of an unknown input, where the input can represent a failure or a malicious attack, aiming at disrupting the normal system functioning while staying undetected. We study linear network systems, and we aim at characterizing input and state observability (ISO), namely the conditions under which both the whole network state and the unknown input can be reconstructed from some measured local states. We complement the classical algebraic characterizations with novel structural results, which depend only on the graph of interactions (equivalently, on the zero pattern of the system matrices). More precisely, there are two kinds of results: structural results, true for almost all interaction weights, and strongly structural results, true for all non-zero interaction weights.

In [32], we consider linear time-invariant (LTI) systems, for which we provide a full characterization of structural ISO. The characterization of strongly structural ISO is on-going work.

In [33], instead, we consider linear time-varying (LTV) systems, under some assumptions on the input and output matrices, namely that each attack input and each output measurement concerns a single local state, and that there is no direct feedthrough of the input to the output. Under these assumptions, we characterize strongly structural ISO; in [23] we also give the characterization of structural ISO under the same assumptions.

We are currently working on analogous characterizations for the more general case, removing these assumptions.

Observability is also related to privacy issues. In the ProCyPhyS project, started in October 2016, we are studying privacy-preserving properties of cyber-physical systems, by analyzing observability properties of such systems, in order to derive privacy-preserving policies for applications related to smart mobility. Precisely, by assuming scenarios where nodes compute an average of their initial condition in a finite number of steps with have state privacy-preserving conditions and devise a simple policy that guarantee privacy in case of observable networks.

7.1.2. Sensor networks: multisensor data fusion for navigation

Participants: H. Fourati [Contact person], T. Michel.

Attitude estimation consists in the determination of rigid body orientation in 3D space (principally in terms of Euler angles, rotation matrix, or quaternion). In [27], we solved the attitude determination problem based on a single sensor observation. The rotation equation is transformed into a quadratic quaternion form and is then derived to a linear matrix equation with pseudoinverse matrices. The analytic solutions to the equation are computed via elementary row operations. The solutions show that the attitude determination from a single sensor observation has infinite solutions and the general one is governed by two limiting quaternions. Accordingly, the variance analysis is given in view of probabilistic characters. The authors explore the experimental results via the accelerometer attitude determination system. The properties of the two limiting quaternions are investigated in the experiment. The results show that the gravity-determination abilities of the two limiting quaternions are quite different. Using the rotation vector and eigenvalue decomposition of the

attitude matrix, the authors prove that one limiting quaternion is better than another one geometrically. The singularity analysis is also performed revealing the non-existence of singularities for limiting quaternions. The above findings are novel, which are quite different from the conclusions made in a previously published study. In [26], we presents a novel linear approach to solve this problem. We name the proposed method the Fast Linear Attitude Estimator (FLAE) because it is faster than known representative algorithms. The original Wahba's problem is extracted to several 1-dimensional equations based on quaternions. They are then investigated with pseudo-inverse matrices establishing a linear solution to n-dimensional equations, which are equivalent to the conventional Wahba's problem. To obtain the attitude quaternion in a robust manner, an eigenvalue-based solution is proposed. Symbolic solutions to the corresponding characteristic polynomial is derived showing higher computation speed. Simulations are designed and conducted using test cases evaluated by several classical methods e.g. M. D. Shuster's QUaternion ESTimator (QUEST), F. L. Markley's SVD method, D. Mortari's Second Estimator of the Optimal Quaternion (ESOQ2) and some recent representative methods e.g. Y. Yang's analytical method and Riemannian manifold method. The results show that FLAE generates attitude estimates as accurate as that of several existing methods but consumes much less computation time (about 50% of the known fastest algorithm). Also, to verify the feasibility in embedded application, an experiment on the accelerometer-magnetometer combination is carried out where the algorithms are compared via C++ programming language. An extreme case is finally studied, revealing a minor improvement that adds robustness to FLAE. We have been interested in other work [28] to some critical issues on Kalman filter observed in navigation solutions of Global Navigation Satellite System (GNSS). The Kalman filtering (KF) is optimal under the assumption that both process and observation noises are independent white Gaussian noise. However, this assumption is not always satisfed in real-world navigation campaigns. In this paper, two types of KF methods are investigated, i.e. augmented KF (AKF) and the second moment information based KF (SMIKF) with colored system noises, including process and observation noises. As a popular noise-whitening method, the principle of AKF is briefly reviewed for dealing with the colored system noises. The SMIKF method is developed for the colored and correlated system noises, which directly compensates for the covariance through stochastic model in the sense of minimum mean square error. To accurately implement the SMIKF, a refned SMIKF is further derived regarding the continuous-time dynamic model rather than the discrete one. The computational burdens of the proposed SMIKF along with representative methods are analyzed and compared. The simulation results demonstrate the performances of proposed methods.

7.1.3. Network reduction towards a scale-free structure preserving physical properties

Participants: N. Martin, P. Frasca, C. Canudas de Wit [Contact person].

In the context of the ERC project, we are addressing a problem of graph reduction, where a given arbitrary weighted graph is reduced to a (smaller) scale-free graph while preserving a consistency with the initial graph and some physical properties. This problem can be formulated as a minimization problem. We give specifications to this general problem to treat a particular case: to this end we define a metric to measure the scale-freeness of a graph and another metric to measure the similarity between two graphs with different dimensions, based on a notion of spectral centrality. Moreover, through the reduction we also preserve a property of mass conservation (essentially, Kirchoff's first law). We study the optimization problem and, based on the gained insights, we derive an algorithm allowing to find an approximate solution. Finally, we have simulated the algorithm both on synthetic networks and on real-world examples of traffic networks that represent the city of Grenoble.

7.1.4. The Observability Radius of Networks

Participants: G. Bianchin, P. Frasca [Contact person], A. Gasparri, F. Pasqualetti.

Our group is undergoing an effort to understand the system-theoretic properties of networks, namely in terms of controllability and observability. In this context, we have studied the observability radius of network systems, which measures the robustness of a network to perturbations of the edges. We consider linear networks, where the dynamics are described by a weighted adjacency matrix and dedicated sensors are positioned at a subset of nodes. We allow for perturbations of certain edge weights with the objective of

preventing observability of some modes of the network dynamics. To comply with the network setting, our work considers perturbations with a desired sparsity structure, thus extending the classic literature on the observability radius of linear systems. The paper [14] proposes two sets of results. First, we propose an optimization framework to determine a perturbation with smallest Frobenius norm that renders a desired mode unobservable from the existing sensor nodes. Second, we study the expected observability radius of networks with given structure and random edge weights. We provide fundamental robustness bounds dependent on the connectivity properties of the network and we analytically characterize optimal perturbations of line and star networks, showing that line networks are inherently more robust than star networks.

7.1.5. Distributed Estimation from Relative and Absolute Measurements

Participants: P. Frasca [Contact person], W.s. Rossi, F. Fagnani.

Important applications in machine learning, in robotic coordination and in sensor networks require distributed algorithms to solve the so-called relative localization problem: a node-indexed vector has to be reconstructed from measurements of differences between neighbor nodes. In [22] we define the problem of least-squares distributed estimation from relative and absolute measurements, by encoding the set of measurements in a weighted undirected graph. The role of its topology is studied by an electrical interpretation, which easily allows distinguishing between topologies that lead to "small" or "large" estimation errors. The least-squares problem is solved by a distributed gradient algorithm, which we have studied in detail. Remarkably, we have observed that the computed solution is approximately optimal after a number of steps that does not depend on the size of the problem or on the graph-theoretic properties of its encoding. This fact indicates that only a limited cooperation between the sensors is necessary to solve this problem.

7.2. Multi-agent systems and network games

7.2.1. Distributed control and game theory: self-optimizing systems

Participants: F. Garin [Contact person], B. Gaujal [POLARIS], S. Durand.

The design of distributed algorithms for a networked control system composed of multiple interacting agents, in order to drive the global system towards a desired optimal functioning, can benefit from tools and algorithms from game theory. This is the motivation of the Ph.D. thesis of Stéphane Durand, a collaboration between POLARIS and NECS teams.

The first results of this thesis concerned the complexity of the best response algorithm under round-robin revision sequence, a classical centralized iterative algorithm to find a Nash Equilibrium. In a more recent work, submitted for publication, and described in the report [40], we focus on distributed versions of the same algorithm. We compute the average complexity over all potential games of best response dynamics under a random i.i.d. revision sequence, since it can be implemented in a distributed way using Poisson clocks. We obtain a distributed algorithm whose execution time is within a constant factor of the optimal centralized one. We then show how to take advantage of the structure of the interactions between players in a network game: noninteracting players can play simultaneously. This improves best response algorithm, both in the centralized and in the distributed case.

7.2.2. Using a linear gain to accelerate average consensus over unreliable networks

Participants: F. Acciani, P. Frasca [Contact person], G. Heijenk, A. Stoorvogel.

Packet loss is a serious issue in wireless consensus networks, as even few failures might prevent a network to converge to the desired consensus value. In some recent work, we have devised a possible way to compensate for the errors caused by packet collisions, by modifying the updating weights. Such a modification compensates for the loss of information in an unreliable network, but results in a reduced convergence speed. In [30], we propose a faster method - based on a suitable gain in the consensus dynamics - to solve the unreliable average consensus problem. We find a sufficient condition for the gain to preserve stability of the network. Simulations are used to discuss the choice of the gain, and to compare our method with the literature.

7.2.3. Mean-field analysis of the convergence time of message-passing computation of harmonic influence in social networks

Participants: W. S. Rossi, P. Frasca [Contact person].

In the study of networks, identifying the most important nodes is of capital importance. The concept of Harmonic Influence has been recently proposed as a metric for the importance of nodes in a social network. This metric evaluates the ability for one node to sway the 'opinions' of the other nodes in the network, under the assumption of a linear diffusion of opinions in the network. A distributed message passing algorithm for its computation has been proposed by Vassio et al., 2014, and proved to converge on general graphs by Rossi and Frasca, 2016. In [36], we presented an want to evaluate the convergence time of this algorithm by using a mean-field approach. The mean-field dynamics is first introduced in a "homogeneous" setting, where it is exact, then heuristically extended to a non-homogeneous setting. The rigorous analysis of the mean-field dynamics is complemented by numerical examples and simulations that demonstrate the validity of the approach.

7.2.4. Modeling birds on wires

Participants: A. Aydogdu, P. Frasca [Contact person], C. d'Apice, R. Manzo, J. M. Thornton, B. Gachomo, T. Wilson, B. Cheung, U. Tariq, W.m. Saidel, B. Piccoli.

The paper [13] introduces a mathematical model to study the group dynamics of birds resting on wires. The model is agent-based and postulates attraction-repulsion forces between the interacting birds: the interactions are "topological", in the sense that they involve a given number of neighbors irrespective of their distance. The main properties of the model are investigated by combining rigorous mathematical analysis and simulations. This analysis gives indications about the total length of a group and the inter-animal spacings within it: in particular, the model predicts birds to be more widely spaced near the borders of each group. We compare these insights from the model with new experimental data, derived from the analysis of pictures of pigeons and starlings taken by the team in New Jersey. We have used two different image elaboration protocols to derive the data for the statistical analysis, which allowed us to establish a good agreement with the model and to quantify its main parameters. Our data also seem to indicate potential handedness of the birds: we investigated this issue by analyzing the group organization features and the group dynamics at the arrival of new birds. However, data are still insufficient to draw a definite conclusion on this matter. Finally, arrivals and departures of birds from the group are included in a refined version of the model, by means of suitable stochastic processes

7.2.5. Network Games: Condensation of the Graph as a Hierarchical interpretation of the Game

Participants: G. Casadei, C. Canudas de Wit [Contact person].

Control and optimization over large population networks have become a popular topic within the control community. The main reason is that modern applications re- quire multiple systems to communicate and interact with each other to fulfill the desired task. For instance power networks, sensor networks and social networks are solid examples in which is fundamental to control different parts of the network to achieve a global desired behavior. In the recent years, the control community has largely focused on cooperative approaches to networks. In this framework the agents in the network are willing to collaborate and find an agreement between each other in such a way that they coordinate their motion.

However, not in all the frameworks and not in all the situations, it is possible to consider a cooperative approach. In several scenarios, the nodes are selfish and in competition with the others to pursue their goal. This leads to a non-cooperative interaction between the agents and thus to games played over networks. When the number of nodes in the network is large, it becomes analytically impossible to use conventional game theoretic tools to find a solution to the problem. This motivated researchers to define a new type of games, named aggregative, where the response of an agent depends, rather than on each other players decision, on the aggregation of all the other agents action.

We considered a refined typology of networks games in which the aggregate information is depending on a directed communication graph and showed that under a certain number of conditions the players reach a Nash Equilibrium. Then we study the influence of this graph topology on the structure of the game and show that the condensation of the graph leads to a hierarchical interpretation of the game and thus to a quasi-sequential architecture of optimization. Then, we introduce the concept of physical graph and control graph in flow networks, and show that the condensation of the control graph helps in determining the equilibrium the agents will reach.

7.3. Transportation networks and vehicular systems

7.3.1. Travel time prediction

Participants: A. Kibangou [Contact person], H. Fourati, C. Canudas de Wit, A. Ladino, M Rodriguez.

One of the regular performance metrics for qualifying the level of congestion in traffic networks is the travel time. In [24], we addressed the problem of dynamic travel time (DTT) forecasting within highway traffic networks using speed measurements. Definitions, computational details and properties in the construction of DTT are provided. DTT is dynamically clustered using a K-means algorithm and then information on the level and the trend of the centroid of the clusters is used to devise a predictor computationally simple to be implemented. To take into account the lack of information in the cluster assignment for the new predicted values, a weighted average fusion based on a similarity measurement is proposed to combine the predictions of each model. The algorithm is deployed in a real time application and the performance is evaluated using real traffic data from the South Ring of the Grenoble city in France. We consider in a recent paper submitted to European Control Conference 2018 the problem of joint reconstruction of flow and density in a urban traffic network using heterogeneous sources of information. The traffic network is modeled within the framework of macroscopic traffic models, where we adopt Lighthill-Whitham-Richards model (LWR) conservation equation and a piecewise linear fundamental diagram. The estimation problem considers three key principles. First, the principle governing traffic models where flow is maximized in a junction. Second, the error minimization between the measured and reconstructed flows and velocities, and finally the equilibrium state of the network which establishes flow propagation within the network. All principles are integrated and the problem is casted as a constrained quadratic optimization with inequality and equality constraints in order to shrink the feasible region of estimated variables. Some simulation scenarios based on synthetic data for a Manhattan grid network are provided in order to validate the performance of the proposed algorithm.

7.3.2. Urban traffic control

Participants: C. Canudas de Wit [Contact person], F. Garin, P. Grandinetti.

The PhD thesis of Pietro Grandinetti deals with optimal or near-optimal operation of traffic lights in an urban area, e.g., a town or a neighborhood. The goal is on-line optimization of traffic lights schedule in real time, so as to take into account variable traffic demands, with the objective of obtaining a better use of the road infrastructure. More precisely, we aim at maximizing total travel distance within the network, together with balancing densities across the network. The complexity of optimization over a large area is addressed both in the formulation of the optimization problem, with a suitable choice of the traffic model, and in a distributed solution, which not only parallelizes computations, but also respects the geometry of the town, i.e., it is suitable for an implementation in a smart infrastructure where each intersection can compute its optimal traffic lights by local computations combined with exchanges of information with neighbor intersections. A modified version of the algorithm uses simplified optimization (purely local, instead of distributed) but takes into account the real constraints in Grenoble downtown traffic lights network, such as priority to public transportation, and imposed minimal and maximal green duration, leading to a fully realistic implementation, tested using Aimsun microscopic simulator.

7.3.3. Traffic Regulation Via Controlled Speed Limit

Participants: M. L. Delle Monache [Contact person], B. Piccoli, F. Rossi.

The work [21] address the speed limit problem on a single road. The control variable is the maximal allowed velocity, which may vary in time but we assume to be of bounded total variation, and we aim at tracking a given target outgoing flow. More precisely, the main goal is to minimize the quadratic difference between the achieved outflow and the given target outflow. Mathematically the problem is very hard, because of the delays in the effect of the control variable (speed limit). In fact, the link entering time, which represents the entering time of the car exiting the road at time t, depends on the given inflow and the control policy on the whole time interval. Moreover, the input-output map is defined in terms of the Link Entering Time, thus the achieved outflow at time t depends on the control variable on the whole time interval. After formulating the optimal control problem, we consider needle-like variations for the control policy as used in the classical Pontryagin maximum principle. We are able to derive an analytical expression of the one-sided variation of the cost, corresponding to needle-like variations of the control policy, using

fine properties of functions with bounded variation. In particular the one-sided variations depend on the sign of the control variation and involve integrals w.r.t. the distributional derivative of the solution as a measure. This allows us to prove Lipschitz continuity of the cost functional in the space of a bounded variation function and prove existence of a solution. Afterwards, we define three different techniques to numerically solve this problem and we compare the three approaches on two test cases.

7.3.4. Scalar conservation laws with moving flux constraints

Participants: M. L. Delle Monache [Contact person], P. Goatin [Acumes, Inria], C. Chalons.

This problem is motivated by the modeling of a moving bottleneck in traffic flow, which can be caused by a large, slow moving vehicle. A slow moving large vehicle, like a bus or a truck, reduces the road capacity and thus generates a moving bottleneck for the surrounding traffic flow. This situation can be modeled by a PDE–ODE strongly coupled system consisting of a scalar conservation law with moving flux constraint accounting for traffic evolution and an ODE describing the slower vehicle motion. In [18], we introduce a novel approach to solve numerically this problem. The main point here is related to the presence of non-classical shocks in the solutions of the model under consideration. It is well-known that, in this context, standard conservative finite volume methods cannot be applied and fail in producing good numerical results. Glimm's scheme can be used but it is not strictly conservative. In order to propose a numerical scheme which is conservative on fixed meshes and able to compute non-classical solutions, we propose to adapt a reconstruction strategy approach, which allows to precisely capture moving non-classical discontinuities on fixed meshes still guaranteeing conservation, unlike Glimm's scheme. An important feature of the proposed method is to be exact for isolated classical and non-classical shocks, which means in particular only one point of numerical diffusion (on each cell the approximate value corresponds to the value of the average of the exact solution). In the general case, shocks are still computed without numerical diffusion and convergence is proved numerically.

In [19] we study well-posedness of a scalar conservation laws with moving flux constraints. In this work we assume that the constraint trajectory is given and it does not depend on the solution of the PDE. In this setting we then show Lipschitz continuous dependence of bounded variation solutions with respect to the initial data ant the constraint trajectory.

7.3.5. Priority-based Riemann solver for traffic flow on networks

Participants: M. L. Delle Monache [Contact person], P. Goatin [Acumes, Inria], B. Piccoli.

In [20] we introduce a novel solver for traffic intersection which considers priorities among the incoming roads as the first criterion and maximization of flux as the second. The main idea is that the road with the highest priority will use the maximal flow taking into account also outgoing roads constraints. If some room is left for additional flow then the road with the second highest priority will use the left space and so on. A precise definition of the new Riemann solver, called Priority Riemann Solver, is based on a traffic distribution matrix, a priority vector and requires a recursion method. The general existence theorem for Riemann solvers on junctions can not be applied in the present case. Therefore, we achieve existence via a new set of general properties.

7.3.6. Discrete-time system optimal dynamic traffic assignment (SO-DTA) with partial control for horizontal queuing networks

Participants: S. Samaranayake, J. Reilly, W. Krichene, M. L. Delle Monache [Contact person], P. Goatin [Acumes, Inria], A. Bayen.

Dynamic traffic assignment (DTA) is the process of allocating time-varying origin-destination (OD) based traffic demand to a set of paths on a road network. There are two types of traffic assignment that are generally considered, the user equilibrium or Wardrop equilibrium allocation (UE-DTA), in which users minimize individual travel-time in a selfish manner, and the system optimal allocation (SODTA) where a central authority picks the route for each user and seeks to minimize the aggregate total travel-time over all users. It can be shown that the price of anarchy (PoA), the worst-case ratio of the system delay caused by the selfish behavior over the system optimal solution, may be arbitrarily large even in simple networks. System optimal (SO) traffic assignment on the other hand leads to optimal utilization of the network resources, but is hard to achieve in practice since the overriding objective for individual drivers in a road network is to minimize their own travel-time. It is well known that setting a toll on each road segment corresponding to the marginal delay of the demand moves the user equilibrium towards a SO allocation. In [25], we formulate the system optimal dynamic traffic assignment problem with partial control (SO-DTAPC), using a Godunov discretization of the Lighthill-Williams-Richards (LWR) partial differential equation (PDE) with a triangular flux function. We propose solving the SO-DTA-PC problem with the non-convex traffic dynamics and limited OD data with complete split ratios as a non-linear optimal control problem. This formulation generalizes to multiple sources and multiple destinations. We show that the structure of our dynamical system allows for very efficient computation of the gradient via the discrete adjoint method.

7.3.7. Measuring trajectories and fuel consumption in oscillatory traffic: experimental results

Participants: F. Wu, R. Stern, M. Churchill, M. L. Delle Monache [Contact person], K. Han, B. Piccoli.

In [37] we present data collected through a set of experiments with nine to 10 vehicles driving on a ring road constructed on a closed track. Vehicle trajectory data is extracted via a series of vision processing algorithms (for background subtraction, vehicle identification, and trajectory extraction) from a 360-degree panoramic camera placed at the center of the ring. The resulting trajectory data is smoothed via a two-step algorithm which applies a combination of RLOESS smoothing and regularized differentiation to produce consistent position, velocity, and acceleration data that does not exhibit unrealistic accelerations common in raw trajectory data extracted from video. A subset of the vehicles also record real-time fuel consumption data of the vehicles using OBD-II scanners. The tests include both smooth and oscillatory traffic conditions, which are useful for constructing and calibrating microscopic models, as well as fuel consumption estimates from these models. The results show a an increase in fuel consumption in the experiments in which traffic oscillations are observed as compared to experiments where vehicles maintain a smooth ow. However, this is partially due to the higher average speed at which vehicles travel in the experiments in which oscillatory traffic is observed. The article contains a complete, publicly available dataset including the video data, the extracted trajectories, the smoothed trajectories, and the OBD-II logs from each equipped vehicle. In addition to the dataset, this article also contains a complete source code for each step of the data processing. It is the first of several experiments planned to collect detailed trajectory data and fuel consumption data with smooth and unsteady traffic flow in a controlled experimental environment.

7.3.8. Large Scale Traffic Networks and Aggregation

Participants: G. Casadei, V. Bertrand, B. Gouin, C. Canudas de Wit [Contact person].

Large scale traffic networks are a popular topic nowadays due to the impact traffic has in our everyday life, both economically and health-wise. City management are interested in understanding the evolution of traffic and its patterns over the city in order to take decision on potential changes and to design new and more functional infrastructure. However, monitoring the current state of a large scale traffic network is a demanding task. The heterogeneity of available measures poses several question on how to merge different sources of information coming from private and public sources. Furthermore, sparsity is an intrinsic issues related to

large scale systems: independently from the source we choose to rely on, we cannot expect the measurements to be sufficiently dense to cover the full network in detail.

For large scale urban network, managing real-time traffic information from thousands of links simultaneously is an overwhelming task and extracting interesting and meaningful insights from these tangle of data can be even a more challenging aim. In recent years more and more data are becoming available from new sources, such as smart phones, GPS navigators, and their technological penetration nowadays allows to have an impressive amount of real-time traffic information, not requiring the placement of physical sensors over the network and thus reducing incredibly costs due to installation and maintenance: in other words, each user becomes a moving sensor inside the network.

One way to deal with this huge amount of data over a urban traffic network is to look at the graph describing the network with a clusterization approach: this would reduce the number of nodes, thus the computational cost, proportionally to the clusterization rate and potentially would help with sparsity by merging areas in which no data are available with areas with sufficient penetration of information. In this work we presented an aggregation-based technique to analyze GPS velocity data from a private source (TomTom) and to calculate multi-origin multi-destination travel time. The technique we propose allows to perform the aggregation and the necessary computation in such a way that its application in a real time framework is feasible. The information and results we obtain are of great interest to understand the macroscopic evolution of the traffic from a large-scale point of view and to evaluate the average time that users spend in transiting between different areas along the day. In practice, we show that reducing the *complexity* of the network by 95% thanks to aggregation, we introduce an error in the calculation of the traveling times that in the average is below 25%.

7.3.9. Two dimensional models for traffic

Participants: S. Mollier, M. L. Delle Monache, C. Canudas de Wit [Contact person].

The work deals with the problem of modeling traffic flow in urban area, e. g. a town. More precisely, the goal is to design a two-dimensional macroscopic traffic flow model suitable to model large network as the one of a city. Macroscopic traffic models are inspired from fluid dynamic. They represent vehicles on the road by a density and describe their evolution with partial differential equations. Usually, these models are one dimensional models and, for instance, give a good representation of the evolution of traffic states in highway. The extension of these 1D models to a network is possible thanks to models of junction but can be tedious according to the number of parameters to fit. In the last few years, the idea of models based on a two dimensional conservation laws arose in order to represent traffic flow in large and dense networks. This study aims to develop such models with new designs especially including the network topology, and validation with simulation.

NUMED Project-Team (section vide)

PERCEPTION Project-Team

6. New Results

6.1. Audio-Source Localization

In previous years we have developed several *supervised* sound-source localization algorithms. The general principle of these algorithms was based on the learning of a mapping (regression) between binaural feature vectors and source locations [6], [8]. While fixed-length wide-spectrum sounds (white noise) are used for training to reliably estimate the model parameters, we showed that the testing (localization) can be extended to variable-length sparse-spectrum sounds (such as speech), thus enabling a wide range of realistic applications. Indeed, we demonstrated that the method could be used for audio-visual fusion, namely to map speech signals onto images and hence to spatially align the audio and visual modalities, thus enabling to discriminate between speaking and non-speaking faces. This year we released a novel corpus of real-room recordings that allow quantitative evaluation of the co-localization method in the presence of one or two sound sources. Experiments demonstrate increased accuracy and speed relative to several state-of-the-art methods. During the period 2015-2016 we extended this method to an arbitrary number of microphones based on the *relative transfer function* – *RTF* (between any channel and a reference channel). In the period 2016-2017 we extended this work and developed a novel transfer function that contains the direct path between the source and the microphone array, namely the *direct-path relative transfer function* [23], [35].

Websites:

https://team.inria.fr/perception/research/acoustic-learning/ https://team.inria.fr/perception/research/binaural-ssl/ https://team.inria.fr/perception/research/ssl-rtf/

6.2. Audio-Source Separation

We addressed the problem of separating audio sources from both static and time-varying convolutive mixtures. We proposed an unsupervised probabilistic framework based on the local complex-Gaussian model combined with non-negative matrix factorization [22]. The time-varying mixing filters are modeled by a continuous temporal stochastic process. This model extended the case of static filters which corresponds to static audio sources. While static filters can be learnt in advance, e.g. [6], time-varying filters cannot and therefore the problem is more complex. We developed a variational expectation-maximization (VEM) algorithm that employs a Kalman smoother to estimate the time-varying mixing matrix, and that jointly estimates the source parameters. In 2017 we extended this method to incorporate the concept of diarization. Indeed, audio sources such as speaking persons do not emit continuously, bet merely take "turns". We formally modeled speech turn-taking within a combined separation and diarization formulation [45], [44]. We also started to investigate the use of the convolutive transfer function for audio-source separation [49], [48], [54]. Websites:

https://team.inria.fr/perception/research/vemove/ https://team.inria.fr/perception/research/nmfig/ https://team.inria.fr/perception/research/dnd/

6.3. Speech Dereverberation and Noise Reduction

We address the problems of blind multichannel identification and equalization for *joint speech dereverberation and noise reduction*. The standard time-domain cross-relation methods are hardly applicable for blind room impulse response identification due to the near-common zeros of the long impulse responses. We extend the cross-relation formulation to the short-time Fourier transform (STFT) domain, in which the time-domain impulse response is approximately represented by the convolutive transfer function (CTF) with much less coefficients. For the oversampled STFT, CTFs suffer from the common zeros caused by the non-flat-top STFT window. To overcome this, we propose to identify CTFs using the STFT framework with oversampled signals and critically sampled CTFs, which is a good trade-off between the frequency aliasing of the signals and the common zeros problem of CTFs. The phases of the identified CTFs are inaccurate due to the frequency aliasing of the CTFs, and thus only their magnitudes are used. This leads to a non-negative multichannel equalization method based on a non-negative convolution model between the STFT magnitude of the source signal and the CTF magnitude. To recover the STFT magnitude of the source signal and to reduce the additive noise, the ℓ_2 norm fitting error between the STFT magnitude of the microphone signals and the non-negative convolution is constrained to be less than a noise power related tolerance. Meanwhile, the ℓ_1 -norm of the STFT magnitude of the source signal is minimized to impose the sparsity [53].

Website: https://team.inria.fr/perception/research/ctf-dereverberation/

6.4. Acoustic-Articulatory Mapping

In this series of studies, we tackle the problem of adapting an acoustic-articulatory inversion model of a reference speaker to the voice of another source speaker. We exploited the framework of Gaussian mixture regressors (GMR) with missing data. To address speaker adaptation, we previously proposed a general framework called Cascaded-GMR (C-GMR) which decomposes the adaptation process into two consecutive steps: spectral conversion between source and reference speaker and acoustic-articulatory inversion of converted spectral trajectories. In particular, we proposed the Integrated C-GMR technique (IC-GMR) in which both steps are tied together in the same probabilistic model. In [34], [43], we extend the C-GMR framework with another model called Joint-GMR (J-GMR). Contrary to the IC-GMR, this model aims at exploiting all potential acoustic-articulatory relationships, including those between the source speaker's acoustics and the reference speaker's articulation. We present the full derivation of the exact Expectation-Maximization (EM) training algorithm for the J-GMR. It exploits the missing data methodology of machine learning to deal with limited adaptation data. We provide an extensive evaluation of the J-GMR on both synthetic acoustic-articulatory data and on the multi-speaker MOCHA EMA database. We compare the J-GMR performance to other models of the C-GMR framework, notably the IC-GMR, and discuss their respective merits. We also exploited the IC-GMR framework with visual data to provide visual biofeedback [32]. Visual biofeedback is the process of gaining awareness of physiological functions through the display of visual information. As speech is concerned, visual biofeedback usually consists in showing a speaker his/her own articulatory movements, which has proven useful in applications such as speech therapy or second language learning. We automatically animate an articulatory tongue model from ultrasound images. We benchmarked several GMR-based techniques on a multispeaker database. The IC-GMR approach is able (i) to maintain good mapping performance while minimizing the amount of adaptation data (and thus limiting the duration of the enrollment session), and (ii) to generalize to articulatory configurations not seen during enrollment better than the plain GMR approach. As a result, the GMR appears to be a good mapping technique for non-linear regression tasks, and in particular for those requiring adaptation (either using J-GMR or IC-GMR).

6.5. Visual Tracking of Multiple Persons

Object tracking is an ubiquitous problem in computer vision with many applications in human-machine and human-robot interaction, augmented reality, driving assistance, surveillance, etc. Although thoroughly investigated, tracking multiple persons remains a challenging and an open problem. In this work, an online variational Bayesian model for multiple-person tracking is proposed. This yields a variational expectationmaximization (VEM) algorithm. The computational efficiency of the proposed method is made possible thanks to closed-form expressions for both the posterior distributions of the latent variables and for the estimation of the model parameters. A stochastic process that handles person birth and person death enables the tracker to handle a varying number of persons over long periods of time [4]. The method was combined with visual servoing and implemented on our robot platform (Fig. 3) [38]. Websites:

https://team.inria.fr/perception/research/ovbt/ https://team.inria.fr/perception/research/mot-servoing/



Figure 3. The multi-person tracking method is combined with a visual servoing module. The latter estimates the optimal robot commands and the expected impact of the tracked person locations. The multi-person tracking module refines the locations of the persons with the new observations and the information provided by the visual servoing.

6.6. Audio-Visual Speaker Tracking and Diarization

We are particularly interested in modeling the interaction between an intelligent device and a group of people. For that purpose we develop audio-visual person tracking methods [33], [41], [52], [39]. As the observed persons are supposed to carry out a conversation, we also include speaker diarization into our tracking methodology. We cast the diarization problem into a tracking formulation whereby the active speaker is detected and tracked over time. A probabilistic tracker exploits the spatial coincidence of visual and auditory observations and infers a single latent variable which represents the identity of the active speaker. Visual and auditory observations are fused using our recently developed weighted-data mixture model [10], while several options for the speaking turns dynamics are fulfilled by a multi-case transition model. The modules that translate raw audio and visual data into image observations are also described in detail. The performance of the proposed method are tested on challenging datasets that are available from recent contributions which are used as baselines for comparison [33].

Websites:

https://team.inria.fr/perception/research/wdgmm/ https://team.inria.fr/perception/research/speakerloc/ https://team.inria.fr/perception/research/speechturndet/ https://team.inria.fr/perception/research/avdiarization/



Figure 4. This figure illustrates the audiovisual tracking and diarization method that we have recently developed. First row: A number is associated with each tracked person. Second row: diarization result. Third row: the ground truth diarization. Fourth row: acoustic signal recorded by one of the two microphones.

6.7. Head Pose Estimation and Tracking

Head pose estimation is an important task, because it provides information about cognitive interactions that are likely to occur. Estimating the head pose is intimately linked to face detection. We addressed the problem of head pose estimation with three degrees of freedom (pitch, yaw, roll) from a single image and in the presence of face detection errors. Pose estimation is formulated as a high-dimensional to low-dimensional mixture of linear regression problem [7]. We propose a method that maps HOG-based descriptors, extracted from face bounding boxes, to corresponding head poses. To account for errors in the observed bounding-box position, we learn regression parameters such that a HOG descriptor is mapped onto the union of a head pose and an offset, such that the latter optimally shifts the bounding box towards the actual position of the face in the image. The performance of the proposed method is assessed on publicly available datasets. The experiments that we carried out show that a relatively small number of locally-linear regression functions is sufficient to deal with the non-linear mapping problem at hand. Comparisons with state-of-the-art methods show that our method outperforms several other techniques [30]. This work is part of the PhD of Vincent Drouard [28] that received the best student paper award (second place) at the IEEE ICIP'15.

In 2017 we extended this work and we proposed a head-pose tracker based on a switching Kalman filter (SKF) formalism. The SKF governs the temporal predictive distribution of the pose parameters (modeled as continuous latent variables) conditioned by the discrete variables associated with the mixture of linear inverse-regression formulation of [7]. We formally derived the equations of the proposed switching linear regression model, we proposed an approximation that is both identifiable and computationally tractable, we designed an EM procedure to estimate the SKF parameters in closed-form, and we carried out experiments and comparisons with other methods using recently released datasets [40]. Websites:

https://team.inria.fr/perception/research/head-pose/ https://team.inria.fr/perception/research/head-pose-tracking/

6.8. Tracking Eye Gaze and of Visual Focus of Attention

The visual focus of attention (VFOA) has been recognized as a prominent conversational cue. We are interested in estimating and tracking the VFOAs associated with multi-party social interactions. We note that in this type of situations the participants either look at each other or at an object of interest; therefore their eyes are not always visible. Consequently both gaze and VFOA estimation cannot be based on eye detection and tracking. We propose a method that exploits the correlation between eye gaze and head movements. Both VFOA and gaze are modeled as latent variables in a Bayesian switching state-space model (also named switching Kalman filter). The proposed formulation leads to a tractable learning method and to an efficient online inference procedure that simultaneously tracks gaze and visual focus. The method is tested and benchmarked using two publicly available datasets, Vernissage and LAEO, that contain typical multi-party human-robot and humanhuman interactions [36].

Website:

https://team.inria.fr/perception/research/eye-gaze/.



Figure 5. This figure shows some results obtained with the LAEO dataset. The top row shows results obtained with coarse head orientation and the bottom row shows results obtained with fine head orientation. Head orientations are shown with red arrows. The algorithm infers gaze directions (green arrows) and VFOAs (blue circles). People looking at each others are shown with a dashed blue line.

6.9. Attention-Gated Conditional Random Fields

Recent works have shown that exploiting multi-scale representations deeply learned via convolutional neural networks (CNN) is of tremendous importance for accurate contour detection. We present [51] a novel approach for predicting contours which advances the state of the art in two fundamental aspects, i.e. multi-scale feature generation and fusion. Different from previous works directly considering multi-scale feature maps obtained from the inner layers of a primary CNN architecture, we introduce a hierarchical deep model which produces more rich and complementary representations. Furthermore, to refine and robustly fuse the representations learned at different scales, the novel Attention-Gated Conditional Random Fields (AG-CRFs) are proposed. The experiments ran on two publicly available datasets (BSDS500 and NYUDv2) demonstrate the effectiveness of the latent AG-CRF model and of the overall hierarchical framework.

6.10. Pooling Local Virality

In our overly-connected world, the automatic recognition of virality - the quality of an image or video to be rapidly and widely spread in social networks - is of crucial importance, and has recently awaken the interest of the computer vision community. Concurrently, recent progress in deep learning architectures showed that global pooling strategies allow the extraction of activation maps, which highlight the parts of the image most
likely to contain instances of a certain class. We extended this concept by introducing a pooling layer that learns the size of the support area to be averaged: the learned top-N average (LENA) pooling [37]. We hypothesize that the latent concepts (feature maps) describing virality may require such a rich pooling strategy. We assess the effectiveness of the LENA layer by appending it on top of a convolutional siamese architecture and evaluate its performance on the task of predicting and localizing virality. We report experiments on two publicly available datasets annotated for virality and show that our method outperforms state-of-the-art approaches.

6.11. Registration of Multiple Point Sets

We have also addressed the rigid registration problem of multiple 3D point sets. While the vast majority of state-of-the-art techniques build on pairwise registration, we proposed a generative model that explains jointly registered multiple sets: back-transformed points are considered realizations of a single Gaussian mixture model (GMM) whose means play the role of the (unknown) scene points. Under this assumption, the joint registration problem is cast into a probabilistic clustering framework. We formally derive an expectation-maximization procedure that robustly estimates both the GMM parameters and the rigid transformations that map each individual cloud onto an under-construction reference set, that is, the GMM means. GMM variances carry rich information as well, thus leading to a noise- and outlier-free scene model as a by-product. A second version of the algorithm is also proposed whereby newly captured sets can be registered online. A thorough discussion and validation on challenging data-sets against several state-of-the-art methods confirm the potential of the proposed model for jointly registering real depth data [31]. Website:

https://team.inria.fr/perception/research/jrmpc/



Figure 6. Integrated point clouds from the joint registration of 10 TOF images that record a static scene (EXBI data-set). Top: color images that roughly show the scene content of each range image (occlusions due to cameras baseline may cause texture artefacts). Bottom: front-view and top-view of integrated sets after joint registration. The results obtained with the proposed method (JRMPC-B) are compared with several other methods.

PERVASIVE INTERACTION Project-Team

6. New Results

6.1. Simulating Haptic Sensations

Participants: Jingtao Chen, Sabine Coquillart, Partners: Inria GRA, LIG, GIPSA, G-SCOP

Pseudo-haptic feedback is a technique aiming to simulate haptic sensations without active haptic feedback devices. Peudo-haptic techniques have been used to simulate various haptic feedbacks such as stiffness, torques, and mass. In the framework of the Persyval project, a novel pseudo-haptic experiment has been set up. The aim of this experiment is to study the force and EMG signals during a pseudo-haptic task. A stiffness discrimination task similar to the one published in Lecuyer's PhD thesis has been chosen. The experimental set-up has been developed, as well as the software controlling the experiment. Pre-tests have been conducted. They have been followed by formal tests with subjects.

6.2. Observing and Modeling Awareness and Expertise During Problem Solving

Participants: Thomas Guntz, Dominique Vaufreydaz, James Crowley, Philippe Dessus, Raffaella Balzarini

We have constructed an instrument for the capture and interpretation of multi-modal signals of humans engaged in solving challenging problems. Our instrument captures eye gaze, fixation, body posture and facial expressions and other physiological signals from humans engaged in interactive tasks on a touch screen.

As a pilot study, we have observed the eye gaze, fixation, body posture and facial expressions for 21 chess experts engaged in solving chess problems. We are currently exploring how such recordings can be used to estimate a subjects awareness of the current situation, and ability to respond effectively to challenging tasks.

6.3. Learning Routine Patterns of Activity in the Home

Participants: Julien Cumin, James Crowley Other Partners: Fano Ramparany, Greg Lefevre (Orange Labs)

During the month of February 2017, we have collected 4 weeks of data on daily activities within the Amiqual4Home Smart Home Living lab apartment. This dataset was presented at the international Conference on Ubiquitous Computing and Ambient Intelligence, UCAmI 2017, at Bethlehem PA, in Nov 2017 and is currently available for download from the Amiqual4Home web server (http://amiqual4home.inria.fr/en/orange4home/)

The objective of this research action is to develop a scalable approach to learning routine patterns of activity in a home using situation models. Information about user actions is used to construct situation models in which key elements are semantic time, place, social role, and actions. Activities are encoded as sequences of situations. Recurrent activities are detected as sequences of activities that occur at a specific time and place each day. Recurrent activities provide routines that can be used to predict future actions and anticipate needs and services. An early demonstration has been to construct an intelligent assistant that can respond to and filter inter-personal communications.

POLARIS Team

7. New Results

7.1. Simgrid for MPI (SMPI)

Several new results on the usage of Simgrid to assess MPI performance have been published in 2017. The general framework introducing the methodology for a proper use of SimGrid to simulate MPI applications was presented in [5]. One more specific line of work concerns the prediction of the performance and the energy consumption of MPI applications using SimGrid [39], [19]. Other applications have also been simulated using this approach. General capacity planning of supercomputers is analyzed in [35] using simulation. More specifically, we have shown that the HPL benchmark (high Performance Linpack), used to establish the top 500 ranking of the most powerful supercomputers in the world, can be emulated faithfully on a commodity server, at the scale of a supercomputer [36]. We have also shown that SimGrid is reliable and fast enough to evaluate and tune the performance of dynamic load balancing in seismic simulations [22].

7.2. Visualisation for Performance Analysis of Task-Based Applications

The performance of task-based application heavily depends on the runtime scheduling and on its ability to exploit computing and communication resources. Unfortunately, the traditional performance analysis strategies are unfit to fully understand task-based runtime systems and applications: they expect a regular behavior with communication and computation phases, while task-based applications demonstrate no clear phases. Moreover, the finer granularity of task-based applications typically induces a stochastic behavior that leads to irregular structures that are difficult to analyze. We have introduced a flexible framework combining visualization panels to understand and pinpoint performance problems incurred by bad scheduling decisions in task-based applications. Three case-studies using StarPU-MPI, a task-based multi-node runtime system, have been investigated in more details to show how our framework is used to study the performance of the well-known Cholesky factorization. Performance improvements include a better task partitioning among the multi-(GPU,core) to get closer to theoretical lower bounds, improved MPI pipelining in multi-(node,core,GPU) to reduce the slow start, and changes in the runtime system to increase MPI bandwidth, with gains of up to 13% in the total makespan [38].

7.3. Convergence of game dynamics

The study of game dynamics is crucial in understanding the long-run behavior of optimizing agents in an environment that changes dynamically over time, whether endogenously (i.e. via the agents' interactions) or exogenously (i.e. due to factors beyond the agents' influence). Starting with the observation that oblivious agents should seek to at least minimize their regret, we showed in [9] that players that "follow the regularized leader" in continuous time achieve no regret at an optimal rate. The multi-agent implications of this property where subsequently explored in [3], [24] (for games with finite and continuous action sets respectively), where we established a wide range of conditions guaranteeing convergence to Nash equilibrium, even when the players' payoff observations are subject to noise and/or other stochastic disturbances.

7.4. Multi-agent learning

In contrast to [9], [3], [24], the above works focus squarely on multi-agent interactions that occur in discrete time (as is typically the case in practical applications). In the case of games with finite action spaces, we showed in [16] that no-regret learning based on "following the regularized leader" converges to Nash equilibrium in potential games, thus complementing the analysis of [15] where it was shown that this family of learning methods eliminates dominated strategies and converges locally to strict Nash equilibria. The former result was extended to mixed-strategy learning in games with continuous action spaces in [11], while [42], [28] established the convergence of no-regret regularized learning to variationally stable equilibria in continuous games, even with imperfect and/or delayed/asynchronous feedback.

7.5. Selfishness vs efficiency in traffic networks

Empirical studies in real-world networks show that the efficiency ratio between selfishly and socially optimal states (the so-called price of anarchy) is close to 1 in both light and heavy traffic conditions, thus raising the question: can these observations be justified theoretically? In [17] we showed that this is not always the case: the price of anarchy may remain bounded away from 1 for all values of the traffic inflow, even in simple three-link networks with a single O/D pair and smooth, convex costs. On the other hand, for a large class of cost functions (including all polynomials), the price of anarchy does converge to 1 in both heavy and light traffic conditions, and irrespective of the network topology and the number of O/D pairs in the network.

7.6. Online Energy Optimization in Embedded Systems

We have used a Markov Decision Process (MDP) approach to compute the optimal on-line speed scaling policy to minimize the energy consumption of a single processor executing a finite or infinite set of jobs with real-time constraints. We provide several qualitative properties of the optimal policy: monotonicity with respect to the jobs parameters, comparison with on-line deterministic algorithms. Numerical experiments in several scenarios show that our proposition performs well when compared with off-line optimal solutions and out-performs on-line solutions oblivious to statistical information on the jobs [33]. Several extension to online learning (Q-learning) as well as hidden Markov chain theory for offline computation of the statistical parameters of the system are currently being investigated.

7.7. Asymptotic Models

• Mean field approximation is a popular means to approximate large and complex stochastic models that can be represented as N interacting objects. The idea of mean field approximation to study the limit of this system as N goes to infinity.

In [18], we study how accurate is mean field approximation as N goes to infinity. We show that under very general conditions the expectation of any performance indicator converges at rate O(1/N) to its mean field approximation. In [7] we continue this analysis and establish a result that expresses the constant associated with this 1/N term. This allows us to propose what we call a *refined mean field approximation*. By considering a variety of applications, we illustrate that the proposed refined mean field approximation is significantly more accurate that the classic mean field approximation for small and moderate values of N: the relative errors of this refined approximation is often below 1% for systems with N = 10.

- Computer system and network performance can be significantly improved by caching frequently used information. When the cache size is limited, the cache replacement algorithm has an important impact on the effectiveness of caching. In [8] we introduce time-to-live (TTL) approximations to determine the cache hit probability of two classes of cache replacement algorithms: h-LRU and LRU(m). Using a mean field approach, we provide both numerical and theoretical support for the claim that the proposed TTL approximations are asymptotically exact. We use this approximation and trace-based simulation to compare the performance of h-LRU and LRU(m). First, we show that they perform alike, while the latter requires less work when a hit/miss occurs. Second, we show that as opposed to LRU, h-LRU and LRU(m) are sensitive to the correlation between consecutive interrequest times. Last, we study cache partitioning. In all tested cases, the hit probability improved by partitioning the cache into different parts—each being dedicated to a particular content provider. However, the gain is limited and the optimal partition sizes are very sensitive to the problem's parameters.
- Mean field approximation is often use to characterize the transient or steady state performance of a stochastic system. In [6], we use this approach to compute absorbing times. We use mean field approximation to provide an asymptotic expansion of this absorbing time that uses the spectral decomposition of the kernel of the original chains. Our results rely on extreme values theory. We show the applicability of this approach with three different problems: the coupon collector, the erasure channel lifetime and the coupling times of random walks in high dimensional spaces.

7.8. Secret Key Generation

Secret key generation (SKG) from shared randomness at two remote locations has been shown to be vulnerable to denial of service attacks in the form of jamming. In [13], [2], we develop as a novel counterjamming approach by using energy harvesting. The idea is that part of the jamming signal can potentially be harvested and converted into useful communication power. In [14], we investigate the use of frequency hopping/spreading in Rayleigh block fading additive white Gaussian noise (BF-AWGN) channels to counteract attacks from jamming. In both cases, we formulate the problems as a zero-sum game and characterize the unique Nash equilibrium of the game in closed form. Through numerical evaluations, we show that energy harvesting is an efficient counter-jamming approach that offers substantial gains in terms of relative SKG rates. In the case of BF-AWGN channels, we also use numerical results to show that frequency hopping/spreading is an effective technique for combating jamming attacks in SKG systems; a modest increase of the system bandwidth can substantially increase the SKG rates.

7.9. Power control in wireless systems

Channel state information (CSI) is essential for efficient power and spectrum allocation policies. In cognitive radio (CR) channels, although perfect CSI of the direct link (between the secondary transmitter and the secondary receiver) is a reasonable assumption at the secondary transmitter (ST), however, perfect knowledge of its interfering links to the primary receivers (PRs) is not. Power allocation and scheduling algorithms are often based on perfect global CSI at the secondary transmitter (ST). In [31], we analyze the impact of channel estimation errors on both the secondary and primary users. On the one hand, the robustness of water-filling type of algorithms allowing the secondary user (SU) to minimize its power consumption under QoS and CR interference power constraints to channel estimation errors in the SU interfering links is analyzed. On the other hand, the impact of these estimation errors on the PU interference constraints is also analyzed. To this aim, we consider the worst case with respect to these estimation errors. Our analysis shows that the water-filling algorithm provides robustness in terms of power consumption and scheduling of the SU given the realistic estimation error models especially when the SU is overestimating the interfering power gains. We also provide possible solutions to ensure that the created interference is below the tolerated thresholds.

7.10. Routing in SDN-based networks

A new adaptive multi-flow routing algorithm to select end-to-end paths in packet-switched networks has been proposed. This algorithm provides provable optimality guarantees in the following game theoretic sense: The network configuration converges to a configuration arbitrarily close to a pure Nash equilibrium. This algorithm has several robustness properties making it suitable for real-life usage: it is robust to measurement errors, outdated information, and clocks desynchronization. Furthermore, it is only based on local information and only takes local decisions, making it suitable for a distributed implementation. Our SDN-based proof-of-concept is built as an Openflow controller. We also set up an emulation platform based on Mininet to test the behavior of our proof-of- concept implementation in several scenarios. Although real-world conditions do not conform exactly to the theoretical model, all experiments exhibit satisfying behavior, in accordance with the theoretical predictions [20], [21].

7.11. Distributed Best Response

We have designed and analyzed distributed algorithms to compute a Nash equilibrium in random potential games. Our algorithms are based on best-response dynamics, with suitable revision sequences (orders of play). We compute the average complexity over all potential games of best response dynamics under a random i.i.d. revision sequence, since it can be implemented in a distributed way using Poisson clocks. We obtain a distributed algorithm whose execution time is within a constant factor of the optimal centralized one. We also showed how to take advantage of the structure of the interactions between players in a network game: non-interacting players can play simultaneously. This improves best response algorithm, both in the centralized and in the distributed case [32].

PRIVATICS Project-Team

6. New Results

6.1. A refinement approach for the reuse of privacy risk analysis results

Participants: Daniel Le Métayer, Sourya joyee de.

With the adoption of the EU General Data Protection Regulation (GDPR), conducting a data protection impact assessment will become mandatory for certain categories of personal data processing. A large body of literature has been devoted to data protection impact assessment and privacy impact assessment. However, most of these papers focus on legal and organizational aspects and do not provide many details on the technical aspects of the impact assessment, which may be challenging and time consuming in practice. The general objective of [10] was to fill this gap and to propose a methodology which can be applied to conduct a privacy risk analysis in a systematic way, to use its results in the architecture selection process (following the privacy by design approach and to re-use its generic part for different products or deployment contexts. The proposed analysis proceeds in three broad phases: (1) a generic privacy risk analysis phase which depends only on the specifications of the system and yields generic harm trees; (2) an architecture-based privacy risk analysis which takes into account the definitions of the possible architectures of the system and refines the generic harm trees into architecture-specific harm trees. (3) a context-based privacy risk analysis which takes into account the definitions of the system (e.g., a casino, an office cafeteria, a school) and further refines the architecture-specific harm trees into context-specific harm trees.

6.2. Interdisciplinarity in practice: Challenges and benefits for privacy research

Participant: Daniel Le Métayer.

The objective of this work was to draw the lessons learned from a project that involved security systems engineers, computer scientists, lawyers and social scientists. Since one of the goals of the project was to propose actual solutions following the privacy by design approach, its aim was to go beyond multidisciplinarity and build on the variety of expertise available in the consortium to follow a true interdisciplinary approach. We have described the challenges before analyzing the solutions adopted by the project to meet them and the outcomes and benefits of the approach. We have concluded the study with some lessons to be drawn from this experience and recommendations for future interdisciplinary projects.

6.3. Capacity: an abstract model of control over personal data

Participant: Daniel Le Métayer.

While the control of individuals over their personal data is increasingly seen as an essential component of their privacy, the word "control" is usually used in a very vague way, both by lawyers and by computer scientists. This lack of precision may lead to misunderstandings and makes it difficult to check compliance. To address this issue, we have proposed in [17] a formal framework based on capacities to specify the notion of control over personal data and to reason about control properties. We have illustrated our framework with social network systems and shown that it makes it possible to characterize the types of control over personal data that they provide to their users and to compare them in a rigorous way. This work will be presented at CODASPY 2018.

6.4. Privacy Risk Analysis to Enable Informed Privacy Settings

Participants: Daniel Le Métayer, Sourya joyee de.

We have proposed in [16] a method to enable better informed choices of privacy preferences or privacy settings by individuals. The method relies on a privacy risk analysis framework parameterized with privacy settings. The user can express his choices, visualize their impact on the privacy risks through a user-friendly interface, and decide to revise them as necessary to reduce risks to an acceptable level.

6.5. Secure electronic documents: is the centralisation of biometric data really inevitable? Inria Analysis Note

Participants: Claude Castelluccia, Daniel Le Métayer.

The decree of 28 October 2016 authorising the creation of a centralised file of "secure electronic documents" (TES) has raised a certain number of questions and concerns. The main aim put forward by the French government is the fight against identity fraud. However, the text of the decree also authorises certain accesses to the database by officers of the national police, national Gendarmerie and intelligence. Many voices have been raised to highlight the risks that such a centralised file could represent with regard to individual freedom, and particularly the invasion of citizens' privacy. The strengthening of the means to fight fraud (and, more generally, criminality) and the requirement to protect privacy are not necessarily in contradiction. However, in order to be able to reach a decision on the advantages and disadvantages of a management system for electronic documents, it seemed necessary to:

- Clearly define the desired functionalities and the advantages that can be expected from them, in particular with respect to the current situation and other solutions.
- Describe the technical solution chosen in a sufficiently precise way to enable its analysis.
- Rigorously analyse the risks of an invasion of privacy with regard to the expected benefits.

As a contribution to this debate, we have analyzed several architectures and alternative solutions which are described in an Inria Analysis Note [15].

6.6. Biometric Systems Private by Design: Reasoning about privacy properties of biometric system architectures

Participant: Daniel Le Métayer.

The goal of this was to show the applicability of the privacy by design approach to biometric systems and the benefit of using formal methods to this end. Starting from a general framework to define privacy architectures and to formally reason about their properties, we have described its adaptation to biometrics. The choice of particular techniques and the role of the components (central server, secure module, biometric terminal, smart card, etc.) in the architecture have a strong impact on the privacy guarantees provided by a biometric system. In the literature, some architectures have already been analysed in some way. However, the existing proposals were made on a case by case basis, which makes it difficult to compare them and to provide a rationale for the choice of specific options. In this work, we have described, on different architectures providing different levels of protection, how a general framework for the definition of privacy architectures can be used to specify the design options of a biometric systems and to reason about them in a formal way.

6.7. Wi-Fi and privacy

Participants: Mathieu Cunche, Célestin Matte.

As communications-enabled devices are becoming more and more ubiquitous, it becomes easier to track the movements of individuals through the radio signals broadcasted by their devices. While there is a strong interest for physical analytics platforms to leverage this information for many purposes, this tracking also threatens the privacy of individuals. To solve this issue, we propose a privacy-preserving solution for collecting aggregate mobility patterns while at the same time satisfying the strong guarantee of ε -differential privacy. More precisely, we introduce a sanitization mechanism for efficient, privacy-preserving and non-interactive approximate distinct counting for physical analytics based on perturbed Bloom filters. We also extend and generalize previous approaches for estimating distinct count of events and joint events (i.e., intersection, and more generally tout of -n cardinalities). Finally, we experimentally evaluate our approach and compare it to previous ones on a real dataset.

Wi-Fi signals emitted by mobile smartphones can be exploited to passively track users' mobility. Turning off the Wi-Fi interface of the device is often presented as a mean to evade those tracking systems. As a matter of fact this method is sometime suggested by the actors of the Wi-Fi tracking industry as a way to opt-out from those systems. The Android system features an option to enable or disable Wi-Fi on the device. However, disabling Wi-Fi through this option is not sufficient to prevent all Wi-Fi activity of the device. Based on measurements on a range of Android devices, we show in [18] that another option, called "Always allow scanning", when activated, makes a device send Wi-Fi frames which can be used to track this device, even if the Wi-Fi switch is off. This option is not clearly described in all Android versions, and sometimes even not deactivatable. Besides, the Google Maps application prompts the user to activate this option.

6.8. Towards Privacy-preserving Wi-Fi Analytics

Participants: Mathieu Cunche, Mohammad Alaggan.

A new technique enabling non-interactive (t, n)-incidence count estimation for indicator vectors ensuring Differential Privacy has been introduced. Given one or two differentially private indicator vectors, estimating the distinct count of elements in each and their intersection cardinality (equivalently, their inner product) have been studied in the literature, along with other extensions for estimating the cardinality set intersection in case the elements are hashed prior to insertion. The core contribution behind all these studies was to address the problem of estimating the Hamming weight (the number of bits set to one) of a bit vector from its differentially private version, and in the case of inner product and set intersection, estimating the number of positions which are jointly set to one in both bit vectors. We develop in [13] the most general case of estimating the number of positions which are set to one in exactly t out of n bit vectors (this quantity is denoted the (t, n)-incidence count), given access only to the differentially private version of those bit vectors. This means that if each bit vector belongs to a different owner, each can locally sanitize their bit vector prior to sharing it, hence the non-interactive nature of our algorithm. The newly introduced algorithm simultaneously estimates the (t, n)incidence counts for all $t \in \{0, ..., n\}$. Upper and lower bounds to the estimation error have been derived. The lower bound is achieved by generalizing the limit of two-party differential privacy into n-party differential privacy, which is a contribution of independent interest. We prove that a lower bound on the additive error that must be incurred by any *n*-wise inner product of *n* mutually differentially-private bit vectors. Those results are very general and are not limited to differentially private bit vectors. They should apply to a large class of sanitization mechanism of bit vectors which depend on flipping the bits with a constant probability. Some potential applications for this technique include physical mobility analytics, call-detail-record analysis, and similarity metrics computation.

6.9. Towards Implicit Visual Memory-Based Authentication

Participant: Claude Castelluccia.

Selecting and remembering secure passwords puts a high cognitive burden on the user, which has adverse effects on usability and security. Authentication schemes based on implicit memory can relieve the user of the burden of actively remembering a secure password. In [8], we propose a new authentication scheme (MooneyAuth) that relies on implicitly remembering the content of previously seen Mooney images. These images are thresholded two-tone images derived from images containing single objects. Our scheme has two phases: In the enrollment phase, a user is presented with Mooney images, their corresponding original images, and labels. This creates an implicit link between the Mooney image and the object in the user's memory that serves as the authentication secret. In the authentication phase, the user has to label a set of Mooney images, a task that gets performed with substantially fewer mistakes if the images have been seen in the enrollment phase. We applied an information-theoretical approach to compute the eligibility of the user, based on which images were labeled correctly. This new dynamic scoring is substantially better than previously proposed static scoring by considering the surprisal of the observed events. We built a prototype and performed three experiments with 230 and 70 participants over the course of 264 and 21 days, respectively. We show that MooneyAuth outperforms current implicit memory-based schemes, and demonstrates a promising new approach for fallback authentication procedures on the Web.

6.10. MyAdChoices: Bringing transparency and control to online advertising

Participant: Claude Castelluccia.

The intrusiveness and the increasing invasiveness of online advertising have, in the last few years, raised serious concerns regarding user privacy and Web usability. As a reaction to these concerns, we have witnessed the emergence of a myriad of ad-blocking and antitracking tools, whose aim is to return control to users over advertising. The problem with these technologies, however, is that they are extremely limited and radical in their approach: users can only choose either to block or allow all ads. With around 200 million people regularly using these tools, the economic model of the Web—in which users get content free in return for allowing advertisers to show them ads—is at serious peril. In [3], we propose a smart Web technology that aims at bringing transparency to online advertising, so that users can make an informed and equitable decision regarding ad blocking. The proposed technology is implemented as a Web-browser extension and enables users to exert fine-grained control over advertising, thus providing them with certain guarantees in terms of privacy and browsing experience, while preserving the Internet economic model. Experimental results in a real environment demonstrate the suitability and feasibility of our approach, and provide preliminary findings on behavioral targeting from real user browsing profiles.

6.11. Differentially Private Mixture of Generative Neural Networks

Participant: Claude Castelluccia.

Generative models are used in a wide range of applications building on large amounts of contextually rich information. Due to possible privacy violations of the individuals whose data is used to train these models, however, publishing or sharing generative models is not always viable. In [4], we develop a novel technique for privately releasing generative models and entire high-dimensional datasets produced by these models. We model the generator distribution of the training data with a mixture of k generative neural networks. These are trained together and collectively learn the generator distribution of a dataset. Data is divided into k clusters, using a novel differentially private kernel k-means, then each cluster is given to separate generative neural networks, such as Restricted Boltzmann Machines or Variational Autoencoders, which are trained only on their own cluster using differentially private gradient descent. We evaluate our approach using the MNIST dataset, as well as call detail records and transit datasets, showing that it produces realistic synthetic samples, which can also be used to accurately compute arbitrary number of counting queries.

6.12. Revisiting Private Web Search using Intel SGX

Participant: Antoine Boutet.

The leakage of user search queries by search engines, which is at the heart of their economic model, makes private Web search an essential functionality to offer to those users that care about their privacy. Nowadays, there exists no satisfactory approach to enable users to access search engines in a privacy-preserving way. Existing solutions are either too costly due to the heavy use of cryptographic mechanisms (e.g., private information retrieval protocols), subject to attacks (e.g., Tor, TrackMeNot, GooPIR) or rely on weak adversarial models (e.g., PEAS). This work [6] introduces X-Search, a novel private Web search mechanism building on the disruptive software guard extensions (SGX) proposed by Intel. We compare X-Search to its closest competitors, Tor and PEAS using a dataset of real web search queries. Our evaluation shows that: (1) X-Search offers stronger privacy guarantees than its competitors as it operates under a stronger adversarial model; (2) it better resists state-of-the-art re-identification attacks; (3) from the performance perspective, X-Search outperforms its competitors both in terms of latency and throughput by orders of magnitude.

6.13. PULP: Achieving Privacy and Utility Trade-off in User Mobility Data

Participant: Antoine Boutet.

Leveraging location information in location-based services leads to improving service utility through geocontextualization. However, this raises privacy concerns as new knowledge can be inferred from location records, such as user's home and work places, or personal habits. Although Location Privacy Protection Mechanisms (LPPMs) provide a means to tackle this problem, they often require manual configuration posing significant challenges to service providers and users. Moreover, their impact on data privacy and utility is seldom assessed. In [9], we present PULP, a model-driven system which automatically provides user-specific privacy protection and contributes to service utility via choosing adequate LPPM and configuring it. At the heart of PULP is nonlinear models that can capture the complex dependency of data privacy and utility for each individual user under given LPPM considered, i.e., Geo-Indistinguishability and Promesse. According to users' preferences on privacy and utility, PULP efficiently recommends suitable LPPM and corresponding configuration. We evaluate the accuracy of PULP's models and its effectiveness to achieve the privacy-utility trade-off per user, using four real-world mobility traces of 770 users in total. Our extensive experimentation shows that PULP ensures the contribution to location service while adhering to privacy constraints for a great percentage of users, and is orders of magnitude faster than non-model based alternatives.

6.14. The Pitfalls of Hashing for Privacy

Participants: Cédric Lauradoux, Mathieu Cunche, Levent Demir.

Boosted by recent legislations, data anonymization is fast becoming a norm. However, as of yet no generic solution has been found to safely release data. As a consequence, data custodians often resort to ad-hoc means to anonymize datasets. Both past and current practices indicate that hashing is often believed to be an effective way to anonymize data. Unfortunately, in practice it is only rarely effective. In [2], we expose the limits of cryptographic hash functions as an anonymization technique. Anonymity set is the best privacy model that can be achieved by hash functions. However, this model has several shortcomings. We provide three case studies to illustrate how hashing only yields a weakly anonymized data. The case studies include MAC and email address anonymization as well as the analysis of Google Safe Browsing.

6.15. Duck Attack on Accountable Distributed Systems

Participant: Cédric Lauradoux.

Accountability plays a key role in dependable distributed systems. It allows to detect, isolate and churn malicious/selfish nodes that deviate from a prescribed protocol. To achieve these properties, several accountable systems use at their core cryptographic primitives that produce non-repudiable evidence of inconsistent or incorrect behavior. In [11], we show how selfish and colluding nodes can exploit the use of cryptographic digests in accountability protocols to mount what we call a duck attack. In a duck attack, selfish and colluding nodes exploit the use of cryptographic digests to alter the transmission of messages while masquerading as honest entities. The end result is that their selfish behavior remains undetected. This undermines the security guarantees of the accountability protocols. We first discover the duck attack while analyzing PAG – a custom cryptographic protocol to build accountable systems presented at ICDCS 2016. We later discover that accountable distributed systems based on a secure log (essentially a hash-based data structure) are also vulnerable to the duck attack and apply it on AcTinG – a protocol presented at SRDS 2014. To defeat our attack, we modify the underlying secure log to have high-order dependency on the messages stored in it.

6.16. Less Latency and Better Protection with AL-FEC Sliding Window Codes: a Robust Multimedia CBR Broadcast Case Study

Participants: Vincent Roca, Belkacem Teibi.

Application-Level Forward Erasure Correction (AL-FEC) codes have become a key component of communication systems in order to recover from packet losses. This work analyzes the benefits of the AL-FEC codes based on a sliding encoding window (A.K.A. convolutional codes) for the reliable broadcast of real-time flows to a potentially large number of receivers over a constant bit rate channel. It first details the initialization of both sliding window codes and traditional block codes in order to keep the maximum AL-FEC decoding latency below a target latency budget. Then it presents detailed performance analyzes using official 3GPP mobility traces, representative of our use case which involves mobile receivers. This work highlights the major benefits of RLC codes, representative of sliding window codes, that outperform any block code, from Raptor codes (that are part of 3GPP MBMS standard) up to ideal MDS codes, both in terms of reduced added latency and improved robustness. It also demonstrates that our RLC codec features decoding speeds that are an order of magnitude higher than that of Raptor codes.

6.17. Coding for efficient Network Communications Research Group (NWCRG)

Participants: Vincent Roca, Belkacem Teibi.

In the context of the "Coding for efficient Network Communications IRTF Research Group (NWCRG) » (https://datatracker.ietf.org/rg/nwcrg/), several activities have been carried out. First of all, a recommended terminology for Network Coding concepts and constructs has been elaborated. It provides a comprehensive set of terms in order to avoid ambiguities in future Network Coding IRTF and IETF documents.

Then, in order to facilitate the use of Sliding Window Codes, such as RLC (see the above FEC Scheme) and RLNC codes (i.e., the well known codes for network coding applications, with potential re-encoding within the network), a work started that introduces a generic Application Programming Interface (API) for window-based FEC codes. This API is meant to be usable by any sliding window FEC code, independently of the FEC Scheme or network coding protocol that may rely on it. This API defines the core procedures and functions meant to control the codec (i.e., implementation of the FEC code), but leaves out all upper layer aspects (e.g., signalling) that are the responsibility of the application making use of the codec. A goal of this document is to pave the way for a future open-source implementation of such codes.

Finally, we started to work on the motivation and requirements for the use of Network Level Packet Erasure Coding to improve the performance of the QUIC protocol that is proposed a new transport protocol. The goal at this level is not specify a specific code but to list the salient features that a code should have in order to deal with know loss patterns on QUIC paths.

ROMA Project-Team

7. New Results

7.1. Acyclic partitioning of large directed acyclic graphs

Participant: Bora Uçar.

Finding a good partition of a computational directed acyclic graph associated with an algorithm can help find an execution pattern improving data locality, conduct an analysis of data movement, and expose parallel steps. The partition is required to be acyclic, i.e., the inter-part edges between the vertices from different parts should preserve an acyclic dependency structure among the parts. In this work [26], we adopt the multilevel approach with coarsening, initial partitioning, and refinement phases for acyclic partitioning of directed acyclic graphs and develop a direct k-way partitioning scheme. To the best of our knowledge, no such scheme exists in the literature. To ensure the acyclicity of the partition at all times, we propose novel and efficient coarsening and refinement heuristics. The quality of the computed acyclic partitions is assessed by computing the edge cut, the total volume of communication between the parts, and the critical path latencies. We use the solution returned by well-known undirected graph partitioners as a baseline to evaluate our acyclic partitioner, knowing that the space of solution is more restricted in our problem. The experiments are run on large graphs arising from linear algebra applications.

7.2. Further notes on Birkhoff-von Neumann decomposition of doubly stochastic matrices

Participants: Ioannis Panagiotas, Bora Uçar.

The well-known Birkhoff-von Neumann (BvN) decomposition expresses a doubly stochastic matrix as a convex combination of a number of permutation matrices. For a given doubly stochastic matrix, there are many BvN decompositions, and finding the one with the minimum number of permutation matrices is NP-hard. There are heuristics to obtain BvN decompositions for a given doubly stochastic matrix. A family of heuristics are based on the original proof of Birkhoff and proceed step by step by subtracting a scalar multiple of a permutation matrix at each step from the current matrix, starting from the given matrix. At every step, the subtracted matrix contains nonzeros at the positions of some nonzero entries of the current matrix and annihilates at least one entry, while keeping the current matrix nonnegative. Our first result shows that this family of heuristics can miss optimal decompositions. We also investigate the performance of two heuristics from this family theoretically [46].

7.3. Low-Cost Approximation Algorithms for Scheduling Independent Tasks on Hybrid Platforms

Participants: Louis-Claude Canon, Loris Marchal, Frédéric Vivien.

Hybrid platforms embedding accelerators such as GPUs or Xeon Phis are increasingly used in computing. When scheduling tasks on such platforms, one has to take into account that a task execution time depends on the type of core used to execute it. We focus on the problem of minimizing the total completion time (or makespan) when scheduling independent tasks on two processor types, also known as the $(Pm, Pk)||C_{\max}$ problem. We propose BalanceEstimate and BalanceMakespan, two novel 2-approximation algorithms with low complexity. Their approximation ratio is both on par with the best approximation algorithms using dual approximation techniques (which are, thus, of high complexity) and significantly smaller than the approximation ratio of existing low-cost approximation algorithms. We compared both algorithms by simulations to existing strategies in different scenarios. These simulations showed that their performance is among the best ones in all cases.

This work has been presented at the EuroPar 2017 conference [22].

7.4. Memory-aware tree partitioning on homogeneous platforms

Participants: Anne Benoit, Changjiang Gou, Loris Marchal.

Scientific applications are commonly modeled as the processing of directed acyclic graphs of tasks, and for some of them, the graph takes the special form of a rooted tree. This tree expresses both the computational dependencies between tasks and their storage requirements. The problem of scheduling/traversing such a tree on a single processor to minimize its memory footprint has already been widely studied. Hence, we move to parallel processing and study how to partition the tree for a homogeneous multiprocessor platform, where each processor is equipped with its own memory. We formally state the problem of partitioning the tree into subtrees such that each subtree can be processed on a single processor and the total resulting processing time is minimized. We prove that the problem is NP-complete, and we design polynomial-time heuristics to address it. An extensive set of simulations demonstrates the usefulness of these heuristics.

This work has been accepted as a short paper in the PDP 2018 conference [50].

7.5. Parallel scheduling of DAGs under memory constraints.

Participants: Loris Marchal, Bertrand Simon, Frédéric Vivien.

Scientific workflows are frequently modeled as Directed Acyclic Graphs (DAG) of tasks, which represent computational modules and their dependencies, in the form of data produced by a task and used by another one. This formulation allows the use of runtime sys- tems which dynamically allocate tasks onto the resources of increasingly complex and hetero- geneous computing platforms. However, for some workflows, such a dynamic schedule may run out of memory by exposing too much parallelism. This work focuses on the problem of transforming such a DAG to prevent memory shortage, and concentrates on shared memory platforms. We first propose a simple model of DAG which is expressive enough to emulate com- plex memory behaviors. We then exhibit a polynomial-time algorithm that computes the max- imum peak memory of a DAG, that is, the maximum memory needed by any parallel schedule. We consider the problem of reducing this maximum peak memory to make it smaller than a given bound by adding new fictitious edges, while trying to minimize the critical path of the graph. After proving this problem NP-complete, we provide an ILP solution as well as several heuristic strategies that are thoroughly compared by simulation on synthetic DAGs modeling actual computational workflows. We show that on most instances, we are able to decrease the maximum peak memory at the cost of a small increase in the critical path, thus with little im- pact on quality of the final parallel schedule.

This work has been accepted for presentation at the IPDPS 2018 conference [56].

7.6. On the Complexity of the Block Low-Rank Multifrontal Factorization

Participants: Patrick Amestoy [INP-IRIT], Alfredo Buttari [CNRS-IRIT], Jean-Yves L'Excellent, Théo Mary [UPS-IRIT].

Matrices coming from elliptic Partial Differential Equations have been shown to have a low-rank property: well defined off-diagonal blocks of their Schur complements can be approximated by low-rank products and this property can be efficiently exploited in multifrontal solvers to provide a substantial reduction of their complexity. Among the possible low-rank formats, the Block Low-Rank format (BLR) is easy to use in a general purpose multifrontal solver and has been shown to provide significant gains compared to full-rank on practical applications. However, unlike hierarchical formats, such as \mathcal{H} and HSS, its theoretical complexity was unknown. We extended the theoretical work done on hierarchical matrices in order to compute the theoretical complexity of the BLR multifrontal factorization. We then studied several variants of the BLR multifrontal factorization. We then studied several variants of the BLR multifrontal pivoting is handled. We showed that these variants can further reduce the complexity of the factorization. In the best case (3D, constant ranks), we obtain a complexity of the order of $O(n^{4/3})$. We provide an experimental study with numerical results to support our complexity bounds.

This work has been published in the SIAM Journal on Scientific Computing [6].

7.7. Large-scale 3-D EM modelling with a Block Low-Rank multifrontal direct solver

Participants: Daniil Shantsev [EMGS-Univ. Oslo], Piyoosh Jaysaval [Univ. Oslo], Sébastien de La Kethulle de Ryhove [EMGS], Patrick Amestoy [INP-IRIT], Alfredo Buttari [CNRS-IRIT], Jean-Yves L'Excellent, Théo Mary [UPS-IRIT].

We put forward the idea of using a Block Low-Rank (BLR) multifrontal direct solver to efficiently solve the linear systems of equations arising from a finite-difference discretization of the frequency-domain Maxwell equations for 3-D electromagnetic (EM) problems. The solver uses a low-rank representation for the offdiagonal blocks of the intermediate dense matrices arising in the multifrontal method to reduce the computational load. A numerical threshold, the so-called BLR threshold, controlling the accuracy of low-rank representations was optimized by balancing errors in the computed EM fields against savings in floating point operations (flops). Simulations were carried out over large-scale 3-D resistivity models representing typical scenarios for marine controlled-source EM surveys, and in particular the SEG SEAM model which contains an irregular salt body. The flop count, size of factor matrices and elapsed run time for matrix factorization are reduced dramatically by using BLR representations and can go down to, respectively, 10, 30 and 40 per cent of their full-rank values for our largest system with N = 20.6 million unknowns. The reductions are almost independent of the number of MPI tasks and threads at least up to $90 \times 10 = 900$ cores. The BLR savings increase for larger systems, which reduces the factorization flop complexity from $O(N^2)$ for the full-rank solver to $O(N^m)$ with m = 1.4-1.6. The BLR savings are significantly larger for deep-water environments that exclude the highly resistive air layer from the computational domain. A study in a scenario where simulations are required at multiple source locations shows that the BLR solver can become competitive in comparison to iterative solvers as an engine for 3-D controlled-source electromagnetic Gauss-Newton inversion that requires forward modelling for a few thousand right-hand sides.

This work has been published in the Geophysical Journal International [16].

7.8. On exploiting sparsity of multiple right-hand sides in sparse direct solvers

Participants: Patrick Amestoy [INP-IRIT], Jean-Yves L'Excellent, Gilles Moreau.

The cost of the solution phase in sparse direct methods is sometimes critical. It can be larger than the one of the factorization in applications where systems of linear equations with thousands of right-hand sides (RHS) must be solved. In this work, we focus on the case of multiple sparse RHS with different nonzero structures in each column. Given a factorization A = LU of a sparse matrix A and the system AX = B (or LY = B when focusing on the forward elimination), the sparsity of B can be exploited in two ways. First, vertical sparsity is exploited by pruning unnecessary nodes from the elimination tree, which represents the dependencies between computations in a direct method. Second, we explain how *horizontal* sparsity can be exploited by working on a subset of RHS columns at each node of the tree. A combinatorial problem must then be solved in order to permute the columns of B and minimize the number of operations. We propose a new algorithm to build such a permutation, based on the tree and on the sparsity structure of B. We then propose an original approach to split the columns of B into a minimal number of blocks (to preserve flexibility in the implementation or maintain high arithmetic intensity, for example), while reducing the number of operations down to a given threshold. Both algorithms are motivated by geometric intuitions and designed using an algebraic approach, and they can be applied to general systems of linear equations. We demonstrate the effectiveness of our algorithms on systems coming from real applications and compare them to other standard approaches. Finally, we give some perspectives and possible applications for this work.

This work is available as a research report [34] and has been submitted to a journal.

7.9. Revisiting temporal failure independence in large scale systems

Participants: Guillaume Aupy [Inria Tadaam], Leonardo Bautista Gomez [Barcelona Supercomputing Center, Spain], Yves Robert, Frédéric Vivien.

This work revisits the *failure temporal independence* hypothesis which is omnipresent in the analysis of resilience methods for HPC. We explain why a previous approach is incorrect, and we propose a new method to detect failure cascades, i.e., series of non-independent consecutive failures. We use this new method to assess whether public archive failure logs contain failure cascades. Then we design and compare several cascade-aware checkpointing algorithms to quantify the maximum gain that could be obtained, and we report extensive simulation results with archive and synthetic failure logs. Altogether, there are a few logs that contain cascades, but we show that the gain that can be achieved from this knowledge is not significant. The conclusion is that we can wrongly, but safely, assume failure independence!

This work is available as a research report and has been submitted to a journal. A preliminary version appears in the proceedings of the FTS'17 workshop.

7.10. Co-scheduling Amdahl applications on cache-partitioned systems

Participants: Guillaume Aupy [Inria Tadaam], Anne Benoit, Sicheng Dai [East China Normal University, China], Loïc Pottier, Padma Raghavan [Vanderbilt University, Nashville TN, USA], Yves Robert, Manu Shantharam [San Diego Supercomputer Center, San Diego CA, USA].

Cache-partitioned architectures allow subsections of the shared last-level cache (LLC) to be exclusively reserved for some applications. This technique dramatically limits interactions between applications that are concurrently executing on a multi-core machine. Consider n applications that execute concurrently, with the objective to minimize the makespan, defined as the maximum completion time of the n applications. Key scheduling questions are: (i) which proportion of cache and (ii) how many processors should be given to each application? In this work, we provide answers to (i) and (ii) for Amdahl applications. Even though the problem is shown to be NP-complete, we give key elements to determine the subset of applications that should share the LLC (while remaining ones only use their smaller private cache). Building upon these results, we design efficient heuristics for Amdahl applications. Extensive simulations demonstrate the usefulness of co-scheduling when our efficient cache partitioning strategies are deployed.

This work is available as a research report and has been accepted for publication in the IJHPCA journal.

7.11. Coping with silent and fail-stop errors at scale by combining replication and checkpointing

Participants: Anne Benoit, Franck Cappello [Argonne National Laboratory, USA], Aurélien Cavelan [University of Basel, Switzerland], Padma Raghavan [Vanderbilt University, Nashville TN, USA], Yves Robert, Hongyang Sun [Vanderbilt University, Nashville TN, USA].

This work provides a model and an analytical study of replication as a technique to detect and correct silent errors, as well as to cope with both silent and fail-stop errors on large-scale platforms. Fail-stop errors are immediately detected, unlike silent errors for which a detection mechanism is required. To detect silent errors, many application-specific techniques are available, either based on algorithms (ABFT), invariant preservation or data analytics, but replication remains the most transparent and least intrusive technique. We explore the right level (duplication, triplication or more) of replication for two frameworks: (i) when the platform is subject only to silent errors, and (ii) when the platform is subject to both silent and fail-stop errors. A higher level of replication is more expensive in terms of resource usage but enables to tolerate more errors and to correct some silent errors, hence there is a trade-off to be found. Replication is combined with checkpointing and comes with two flavors: process replication and group replication. Process replication applies to messagepassing applications with communicating processes. Each process is replicated, and the platform is composed of process pairs, or triplets. Group replication applies to black-box applications, whose parallel execution is replicated several times. The platform is partitioned into two halves (or three thirds). In both scenarios, results are compared before each checkpoint, which is taken only when both results (duplication) or two out of three results (triplication) coincide. If not, one or more silent errors have been detected, and the application rolls back to the last checkpoint, as well as when fail-stop errors have struck. We provide a detailed analytical study for all of these scenarios, with formulas to decide, for each scenario, the optimal parameters as a function of the error rate, checkpoint cost, and platform size. We also report a set of extensive simulation results that nicely corroborates the analytical model.

This work is available as a research report and has been submitted to a journal. A preliminary version appears in the proceedings of the FTXS'17 workshop.

7.12. Optimal checkpointing period with replicated execution on heterogeneous platforms

Participants: Anne Benoit, Aurélien Cavelan [University of Basel, Switzerland], Valentin Le Fèvre, Yves Robert.

In this work, we design and analyze strategies to replicate the execution of an application on two different platforms subject to failures, using checkpointing on a shared stable storage. We derive the optimal pattern size W for a periodic checkpointing strategy where both platforms concurrently try and execute W units of work before checkpointing. The first platform that completes its pattern takes a checkpoint, and the other platform interrupts its execution to synchronize from that checkpoint. We compare this strategy to a simpler onfailure checkpointing strategy, where a checkpoint is taken by one platform only whenever the other platform encounters a failure. We use first or second-order approximations to compute overheads and optimal pattern sizes, and show through extensive simulations that these models are very accurate. The simulations show the usefulness of a secondary platform to reduce execution time, even when the platforms have relatively different speeds: in average, over a wide range of scenarios, the overhead is reduced by 30%. The simulations also demonstrate that the periodic checkpointing strategy is globally more efficient, unless platform speeds are quite close.

This work is available as a research report. A preliminary version appears in the proceedings of the FTXS'17 workshop.

7.13. A Failure Detector for HPC Platforms

Participants: George Bosilca [ICL, University of Tennessee Knoxville, USA], Aurélien Bouteiller [ICL, University of Tennessee Knoxville, USA], Amina Guermouche [Telecom SudParis, France], Thomas Hérault [ICL, University of Tennessee Knoxville, USA], Yves Robert, Pierre Sens [LIP6, Université Paris 6, France].

Building an infrastructure for exascale applications requires, in addition to many other key components, a stable and efficient failure detector. This work describes the design and evaluation of a robust failure detector, that can maintain and distribute the correct list of alive resources within proven and scalable bounds. The detection and distribution of the fault information follow different overlay topologies that together guarantee minimal disturbance to the applications. A virtual observation ring minimizes the overhead by allowing each node to be observed by another single node, providing an unobtrusive behavior. The propagation stage is using a non uniform variant of a reliable broadcast over a circulant graph overlay network, and guarantees a logarithmic fault propagation. Extensive simulations, together with experiments on the Titan ORNL supercomputer, show that the algorithm performs extremely well and exhibits all the desired properties of an exascale-ready algorithm.

This work is available as a research report and has been accepted for publication in the IJHPCA journal. A preliminary version appears in the proceedings of the SC'16 conference.

7.14. Budget-aware scheduling algorithms for scientific workflows on IaaS cloud platforms

Participants: Yves Caniou [Inria Avalon], Eddy Caron [Inria Avalon], Aurélie Kong Win Chang, Yves Robert.

This work introduces several budget-aware algorithms to deploy scientific workflows on IaaS cloud platforms, where users can request Virtual Machines (VMs) of different types, each with specific cost and speed parameters. We use a realistic application/platform model with stochastic task weights, and VMs communicating through a datacenter. We extend two well-known algorithms, HEFT and MinMin, and make scheduling decisions based upon machine availability *and* available budget. During the mapping process, the budget-aware algorithms make conservative assumptions to avoid exceeding the initial budget; we further improve our results with refined versions that aim at re-scheduling some tasks onto faster VMs, thereby spending any budget fraction leftover by the first allocation. These refined variants are much more time-consuming than the former algorithms, so there is a trade-off to find in terms of scalability. We report an extensive set of simulations with workflows from the Pegasus benchmark suite. Budget-aware algorithms generally succeed in achieving efficient makespans while enforcing the given budget, and despite the uncertainty in task weights.

This work is available as a research report and has been submitted to a journal.

7.15. Resilience for stencil computations with latent errors

Participants: Aurélien Cavelan [University of Basel, Switzerland], Andrew Chien [University of Chicago, USA], Aiman Fang [University of Chicago, USA], Yves Robert.

Projections and measurements of error rates in near-exascale and exascale systems suggest a dramatic growth, due to extreme scale (10^9 cores) , concurrency, software complexity, and deep submicron transistor scaling. Such a growth makes resilience a critical concern, and may increase the incidence of errors that "escape", silently corrupting application state. Such errors can often be revealed by application software tests but with long latencies, and thus are known as *latent errors*. We explore how to efficiently recover from latent errors, with an approach called application-based focused recovery (ABFR). Specifically we present a case study of stencil computations, a widely useful computational structure, showing how ABFR focuses recovery effort where needed, using intelligent testing and pruning to reduce recovery effort, and enables recovery effort to be overlapped with application computation. We analyze and characterize the ABFR approach on stencils, creating a performance model parameterized by error rate and detection interval (latency). We compare projections from the model to experimental results with the Chombo stencil application, validating the model and showing that ABFR on stencil can achieve a significant reductions in error recovery cost (up to 400x) and recovery latency (up to 4x). Such reductions enable efficient execution at scale with high latent error rates.

This work is available as a research report . A short version appears in the proceedings of the ICPP'17 conference.

7.16. Checkpointing workflows for fail-stop errors

Participants: Louis-Claude Canon, Henri Casanova [University of Hawai'i at Manoa, USA], Li Han, Yves Robert, Frédéric Vivien.

We consider the problem of orchestrating the execution of workflow applications structured as Directed Acyclic Graphs (DAGs) on parallel computing platforms that are subject to fail-stop failures. The objective is to minimize expected overall execution time, or makespan. A solution to this problem consists of a schedule of the workflow tasks on the available processors and of a decision of which application data to checkpoint to stable storage, so as to mitigate the impact of processor failures. For general DAGs this problem is hopelessly intractable. In fact, given a solution, computing its expected makespan is still a difficult problem. To address this challenge, we consider a restricted class of graphs, Minimal Series-Parallel Graphs (GSPGs). It turns out that many real-world workflow applications are naturally structured as GSPGs. For this class of graphs, we propose a recursive list-scheduling algorithm that exploits the GSPG structure to assign sub-graphs to individual processors, and uses dynamic programming to decide which tasks in these sub-gaphs should be checkpointed. Furthermore, it is possible to efficiently compute the expected makespan for the solution produced by this algorithm, using a first-order approximation of task weights and existing evaluation algorithms for 2-state probabilistic DAGs. We assess the performance of our algorithm for production workflow configurations, comparing it to (i) an approach in which all application data is checkpointed, which

corresponds to the standard way in which most production workflows are executed today; and (ii) an approach in which no application data is checkpointed. Our results demonstrate that our algorithm strikes a good compromise between these two approaches, leading to lower checkpointing overhead than the former and to better resilience to failure than the latter. To the best of our knowledge, this is the first scheduling/checkpointing algorithm for workflow applications with fail-stop failures that considers workflow structures more general than mere linear chains of tasks.

This work is available as a research report and has been submitted to a journal. A short version appears in the proceedings of the IEEE Cluster'17 conference.

7.17. Optimal Cooperative Checkpointing for Shared High-Performance Computing Platforms

Participants: Dorian Arnold [Emory University, Atlanta, GA, USA], George Bosilca [ICL, University of Tennessee Knoxville, USA], Aurélien Bouteiller [ICL, University of Tennessee Knoxville, USA], Jack Dongarra [ICL, University of Tennessee Knoxville, USA], Kurt Ferreira [Center for Computing Research, Sandia National Laboratory, USA], Thomas Hérault [ICL, University of Tennessee Knoxville, USA], Yves Robert.

In high-performance computing environments, input/output (I/O) from various sources often contend for scarce available bandwidth. Adding to the I/O operations inherent to the failure-free execution of an application, I/O from checkpoint/restart (CR) operations (used to ensure progress in the presence of failures) places an additional burden as it increases I/O contention, leading to degraded performance. In this work, we consider a cooperative scheduling policy that optimizes the overall performance of concurrently executing CR-based applications which share valuable I/O resources. First, we provide a theoretical model and then derive a set of necessary constraints needed to minimize the global *waste* on the platform. Our results demonstrate that the optimal checkpoint interval as defined by Young/Daly, while providing a sensible metric for a single application, is not sufficient to optimally address resource contention at the platform scale. We therefore show that combining optimal checkpointing periods with I/O scheduling strategies can provide a significant improvement on the overall application performance, thereby maximizing platform throughput. Overall, these results provide critical analysis and direct guidance on checkpointing large-scale workloads in the presence of competing I/O while minimizing the impact on application performance.

This work is available as a research report and has been submitted to a conference.

7.18. Parallel Code Generation of Synchronous Programs for a Many-core Architecture

Participant: Matthieu Moy.

Embedded systems tend to require more and more computational power. Many-core architectures are good candidates since they offer power and are considered more time predictable than classical multi-cores.

Data-flow Synchronous languages such as Lustre or Scade are widely used for avionic critical software. Programs are described by networks of computational nodes. Implementation of such programs on a manycore architecture must ensure a bounded response time and preserve the functional behavior by taking interference into account.

We consider the top-level node of a Lustre application as a software architecture description where each subnode corresponds to a potential parallel task. Given a mapping (tasks to cores), we automatically generate code suitable for the targeted many-core architecture. This code uses hardware synchronization mechanisms and time-triggered execution. This minimizes memory interferences and allows usage of a framework to compute the Worst-Case Response Time.

This work was accepted for publication at the DATE 2018 conference [82].

7.19. Optimizing Affine Control with Semantic Factorizations

Participants: Christophe Alias, Alexandru Plesco.

Hardware accelerators generated by polyhedral synthesis techniques make an extensive use of affine expressions (affine functions and convex polyhedra) in control and steering logic. Since the control is pipelined, these affine objects must be evaluated at the same time for different values, which forbids aggressive reuse of operators.

In this work, we propose a method to factorize a collection of affine expressions without preventing pipelining. Our key contributions are (i) to use semantic factorizations exploiting arithmetic properties of addition and multiplication and (ii) to rely on a cost function whose minimization ensures a correct usage of FPGA resources. Our algorithm is totally parametrized by the cost function, which can be customized to fit a target FPGA. Experimental results on a large pool of linear algebra kernels show a significant improvement compared to traditional low-level RTL optimizations. In particular, we show how our method reduces resource consumption by revealing hidden strength reductions.

This work has been published in ACM TACO [5]

7.20. Improving Communication Patterns in Polyhedral Process Networks

Participant: Christophe Alias.

Process networks are a natural intermediate representation for HLS and more generally automatic parallelization. Compiler optimizations for parallelism and data locality restructure deeply the execution order of the processes, hence the read/write patterns in communication channels. This breaks most FIFO channels, which have to be implemented with addressable buffers. Expensive hardware is required to enforce synchronizations, which often results in dramatic performance loss.

In this work, we present an algorithm to partition the communications so that most FIFO channels can be recovered after a loop tiling, a key optimization for parallelism and data locality. Experimental results show a drastic improvement of FIFO detection for regular kernels at the cost of (few) additional storage. As a bonus, the storage can even be reduced in some cases.

This work will be presented at the HiP3ES'2018 workshop [32].

7.21. Static Analyses of pointers

Participants: Laure Gonnord, Maroua Maalej.

The design and implementation of static analyses that disambiguate pointers has been a focus of research since the early days of compiler construction. One of the challenges that arise in this context is the analysis of languages that support pointer arithmetics, such as C, C++ and assembly dialects. In 2017, we contributed to this research area with a conference paper and a journal paper.

The CGO'17 paper[27] contributes to solve this challenge. We start from an obvious, yet unexplored, observation: if a pointer is strictly less than another, they cannot alias. Motivated by this remark, we use the abstract interpretation framework to build strict less-than relations between pointers. To this end, we construct a program representation that bestows the Static Single Information (SSI) property onto our dataflow analysis. SSI gives us an efficient sparse algorithm, which, once seen as a form of abstract interpretation, is correct by construction. We have implemented our static analysis in LLVM. It runs in time linear on the number of program variables, and, depending on the benchmark, it can be as much as six times more precise than the pointer disambiguation techniques already in place in that compiler.

Pentagons is an abstract domain invented by Logozzo and Fahndrich to validate array accesses in low-level programming languages. This algebraic structure provides a cheap "less-than check", which builds a partial order between the integer variables used in a program. In the Science of Computer Programming journal paper[15], we show how we have used the ideas available in Pentagons to design and implement a novel alias analysis. With this new algorithm, we are able to disambiguate pointers with off- sets, that commonly occur in C programs, in a precise and efficient way. Together with this new abstract domain we describe several implementation decisions that let us produce a practical pointer disambiguation algorithm on top of the LLVM compiler. Our alias analysis is able to handle programs as large as SPEC CPU2006's gcc in a few minutes. Furthermore, it improves on LLVM's industrial quality analyses. As an extreme example, we have observed a 4x improvement when analyzing SPEC's lbm.

7.22. Dataflow static analyses and optimisations

Participants: Laure Gonnord, Lionel Morel, Szabolcs-Marton Bagoly, Romain Fontaine.

Nowadays, parallel computers have become ubiquitous and current processors contain several execution cores, anywhere from a couple to hundreds. This multi-core tendency is due to constraints preventing the increase of clock frequencies, such as heat generation and power consumption. A variety of low-level tools exist to program these chips efficiently, but they are considered hard to program, to maintain, and to debug, because they may exhibit non-deterministic behaviors. We explore the potentiality of the data flow programming, which allows the programmer to specify only the operations to perform and their dependencies, without actually scheduling them. The work is published in two research reports: [48] and [49].

In [48], we explore the combination of a dataflow paradigm language, SigmaC, with the Polyhedral Model, which allows automatic parallelization and optimization of loop nests, in order to make the programming easier by delegating work to the compilers and static analyzers, in various case studies.

In [49], we explore the expressivity of the horn clause format for static analyses of Lustre programs with arrays. We propose a translation from a Lustre core language to horn clauses, with or without array variables.

SOCRATE Project-Team

6. New Results

6.1. Flexible Radio Front-End

6.1.1. RFID tag-to-tag communication

RFID is a well-known technique for wireless authentication. Usually, such a system only consists on a reader communicating with one or several tags. The concept of passive RFID tag-to-tag communications has been recently introduced and opens new promising perspectives, especially in the field of Internet-of-Things. In this work, a simulation framework was proposed as a new tool allowing the performance evaluation of tag-to-tag radio links. The modeling takes into consideration the external source supplying the communication between tags, radiating characteristics of tag antennas, and reception system aspects. Performance results are expressed in terms of Bit Error Rate (BER) with respect to the distance between the tags and the position of the energy source relative to the position of the two tags [36], [35].

6.1.2. Optimization of waveforms for energy harvesting

We have studied the incidence of the modulation scheme as well as the input power on the RF to DC rectifier conversion efficiency for an energy harvesting system based on radiowaves. A commercial energy harvesting (EH) P21XXCSR-EVB evaluation board from Powercast Corporation is used as measurement target and several waveforms are employed to evaluate the rectification efficiency. With a continuous wave as reference, QPSK, QAM and OFDM waveforms usage demonstrates that digital modulated signals can lead to a better efficiency. Thus, by selecting a high peak to average power ratio (PAPR), and under certain conditions, the performance of the energy harvesting circuit is enhanced [30].

6.2. Multi-User Communications

6.2.1. Fundamental limits : contributions in Multi-User Information Theory (MU-IT)

6.2.1.1. Interference channel with feedback

In this work [29], [44], [43], the η -Nash equilibrium (η -NE) region of the two-user linear deterministic interference channel (IC) with noisy channel-output feedback is characterized for all $\eta > 0$. The η -NE region, a subset of the capacity region, contains the set of all achievable information rate pairs that are stable in the sense of an η -NE. More specifically, given an η -NE coding scheme, there does not exist an alternative coding scheme for either transmitter-receiver pair that increases the individual rate by more than η bits per channel use. Existing results such as the η -NE region of the linear deterministic IC without feedback and with perfect output feedback are obtained as particular cases of the result. We also characterized in [15] the price of anarchy (PoA) and the price of stability (PoS) of this η -NE. The price of anarchy is the ratio between the sum-rate capacity and the smallest sum-rate at an η -NE. The price of stability is the ratio between the sum-rate capacity and the biggest sum-rate at an η -NE. Some of the main conclusions of this work are the following: (a) When both transmitter-receiver pairs are in low interference regime, the PoA can be made arbitrarily close to one as η approaches zero, subject to a particular condition. More specifically, there are scenarios in which even the worst η -NE (in terms of sum-rate) is arbitrarily close to the Pareto boundary of the capacity region. (b) The use of feedback plays a fundamental role on increasing the PoA, in some interference regimes. This is basically because in these regimes, the use of feedback increases the sum-capacity, whereas the smallest sumrate at an η -NE remains the same. (c) The PoS is equal to one in all interference regimes. This implies that there always exists an η -NE in the Pareto boundary of the capacity region. The ensemble of conclusions of this work reveal the relevance of jointly using equilibrium selection methods and channel-output feedback for reducing the effect of anarchical behavior of the network components in the η -NE sum-rate of the interference channel.

6.2.1.2. Simultaneous information and energy transmission

In this work [42], [25], [48], the fundamental limits of simultaneous information and energy transmission in the two-user Gaussian interference channel (G-IC) with and without feedback are fully characterized. More specifically, an achievable and converse region in terms of information and energy transmission rates (in bits per channel use and energy-units per channel use, respectively) are identified. In both cases, with and without feedback, an achievability scheme based on power-splitting, common randomness, rate splitting, block-Markov superposition coding, and backward decoding is presented. Finally, converse regions for both cases are obtained using some of the existing outer bounds for information transmission rates, as well as a new outer bound for the energy transmission rate.

6.2.1.3. Ultra-dense wireless networks

Ultra-dense networks represent an interesting model for future IoT networks. The analysis of these metworks relies on the association of stochastic geometry models with information theory in the finite blocklength regime. Considering an isolated wireless cell containing a high density of nodes, the fundamental limit can be defined as the maximal number of nodes the associate base station can serve under some system level constraints including maximal rate, reliability, latency and transmission power. This limit can be investigated in the downlink, modeled as a spatial continuum broadast channel (SCBC) as well as in the uplink modeled as a spatial continuum multiple access channel (SCMAC). In this work, we define the different steps towards the characterization of this fundamental limit, considering four figures of merit: energy efficiency, spectral efficiency, latency, reliability [13]. To address this question in the uplink scenario [10], we uses a large scale Multiple Access Channel (MAC) to model IoT nodes randomly distributed over the coverage area of a unique base station. The traffic is represented by an information rate spatial density $\rho(x)$. This model, referred to as the Spatial Continuum Multiple Access Channel, is defined as the asymptotic limit of a sequence of discrete MACs. The access capacity region of this channel is defined as the set of achievable information rate spatial densities achievable with vanishing transmission errors and under a sum-power constraint. Simulation results validate the model and show that this fundamental limit theoretically achievable when all nodes transmit simultaneously over an infinite time, may be reached even with a relatively small number of simultaneous transmitters (typically around 20 nodes) which gives credibility to the model. The results also highlight the potential interest of non-orthogonal transmissions for IoT uplink transmissions when compared to an ideal time sharing strategy. We then developed a powerful analytical model of wireless network with Superposition Coding (SC), also referred to as Non Orthogonal Multiple Access (NOMA), taking into consideration a multi cell interference limited network. This model allows to establish a closed form expression of the minimum power a base station (BS) needs to transmit to its users and to achieve a given SINR (signal to interference plus noise ratio) whatever its location in the area covered by the base station. It moreover allows to establish a closed form expression of the minimum total transmit power of a base station. These closed form expressions allow to establish performance of wireless networks, by minimizing the base stations transmit powers. As an application, we show that these closed form expressions allow to quantify the energetic performance, spectral efficiency, total throughput and the coverage of a BS, in a simple and quick way.

6.2.1.4. Broadcast channel in the Finite Blocklength regime

In order to analyse wireless networks with short packets, theoretical results in information theory for the finite blocklength regime in multi-user scenarios were missin. In [34], [33], we analyzed the performance of superposition coding for Gaussian broadcast channels with finite blocklength. To this end, we adapted two different achievability bounds, the dependence testing and the $\kappa - \beta$ -bounds introduced by Polyanskiy et al. in 2010 to the broadcast setting. The distinction between these bounds lies in fixing either the input or the output distributions of the channel. For the first case of the dependence testing bound, an upper bound on the average error probability of the system is derived whereas for the latter, lower bounds on the maximal code sizes of each user are presented.

6.2.1.5. Capacity sensitivity

In this work [22], [40], a new framework based on the notion of *capacity sensitivity* is introduced to study the capacity of continuous memoryless point-to-point channels. The capacity sensitivity reflects how the capacity changes with small perturbations in any of the parameters describing the channel, even when the capacity is

not available in closed-form. This includes perturbations of the cost constraints on the input distribution as well as on the channel distribution. The framework is based on continuity of the capacity, which is shown for a class of perturbations in the cost constraint and the channel distribution. The continuity then forms the foundation for obtaining bounds on the capacity sensitivity. As an illustration, the capacity sensitivity bound is applied to obtain scaling laws when the support of additive α -stable noise is truncated.

6.2.2. Performance evaluation of large scale systems

6.2.2.1. UNB networks performance evaluation

UNB (Ultra Narrow Band) stands out as one promising PHY solution for low-power, low-throughput and longrange IoT. The dedicated MAC scheme is RFTMA (Random Frequency and Time Multiple Access), where nodes access the channel randomly both in frequency and in time domain, without prior channel sensing. This blind randomness sometimes introduces interference and packet losses. In order to quantify the system performance, we have derived and exploited a theoretical expression of the outage probability in a UNB based IoT network, when taking into account both interference due to the spectral randomness and path loss due to the propagation [14], [5]. Besides, we also proposed to use the well-known SIC (Successive Interference Cancellation) to cancel the interference in a recursive way. We provided a theoretical analysis of network performance, when considering jointly SIC and the specific spectral randomness of UNB. We analytically and numerically highlighted the SIC efficiency in enhancing UNB system performance [26].

6.2.2.2. Wireless networks on FIT/CorteXlab

In this work we study the FIT/CorteXlab platform where all radio nodes are confined to an electromagnetically (EM) shielded environment and have flexible radio-frequency (RF) front-end for experimenting on software defined radio (SDR) and cognitive radio (CR). A unique feature of this testbed is that it offers roughly 40 SDR nodes that can be accessed from anywhere in the world in a reproducible manner: the electromagnetic shield prevents from external interference and channel variability. In this work [16] we show why it is important to have such a reproducible radio experiment testbed and we highlight the reproducibility by the channel characteristics between the nodes of the platform. We back our claims with a large set of measurements done in the testbed, that also refines our knowledge on the propagation characteristics of the testbed.

One of the major goals of the 5G technology roadmap is to create disruptive innovation for the efficient use of the radio spectrum to enable rapid access to bandwidth-intensive multimedia services over wireless networks. The biggest challenge toward this goal lies in the difficulty in exploiting the multicast nature of the wireless channel in the presence of wireless users that rarely access the same content at the same time. Recently, the combined use of wireless edge caching and coded multicasting has been shown to be a promising approach to simultaneously serve multiple unicast demands via coded multicast transmissions, leading to order-of-magnitude bandwidth efficiency gains. However, a crucial open question is how these theoretically proven throughput gains translate in the context of a practical implementation that accounts for all the required coding and protocol overheads. In [3], in collaboration with Nokia Bell Labs, New Jersey, we first provide an overview of the emerging caching- aided coded multicast technique, including state-ofthe-art schemes and their theoretical performance. We then focus on the most competitive scheme proposed to date and describe a fully working prototype implementation in CorteXlab, one of the few experimental facilities where wireless multiuser communication scenarios can be evaluated in a reproducible environment. We use our prototype implementation to evaluate the experimental performance of state-of-the-art cachingaided coded multicast schemes compared to state-of-the-art uncoded schemes, with special focus on the impact of coding computation and communication overhead on the overall bandwidth efficiency performance. Our experimental results show that coding overhead does not significantly affect the promising performance gains of coded multicasting in small-scale realworld scenarios, practically validating its potential to become a key next generation 5G technology.

6.2.3. Cognitive networks

6.2.3.1. Game theory based approaches

In [4], a generalization of the satisfaction equilibrium (SE) for games in satisfaction form (SF) is presented. This new solution concept is referred to as the generalized satisfaction equilibrium (GSE). In games in SF, players choose their actions to satisfy an individual constraint that depends on the actions of all the others. At a GSE, players that are unsatisfied are unable to unilaterally deviate to be satisfied. The concept of GSE generalizes the SE in the sense that it allows mixed-strategy equilibria in which there exist players who are unable to satisfy their individual constraints. The pure-strategy GSE problem is closely related to the constraint satisfaction problem and finding a pure-strategy GSE is proven to be NP-hard. The existence of at least one GSE in mixed strategies is proven for the class of games in which the constraints are defined by a lower limit on the expected utility. A dynamics referred to as the satisfaction response is shown to converge to a GSE in certain classes of games. Finally, Bayesian games in SF and the corresponding Bayesian GSE are introduced. These results provide a theoretical framework for studying service-level provisioning problems in communications networks as shown by several examples.

Device-to-device (D2D) communications can enhance spectrum and energy efficiency due to direct proximity communication and frequency reuse. However, such performance enhancement is limited by mutual interference and energy availability, especially when the deployment of D2D links is ultra-dense. In this contribution [9], we present a distributed power control method for ultra-dense D2D communications underlaying cellular communications. In this power control method, in addition to the remaining battery energy of the D2D transmitter, we consider the effects of both the interference caused by the generic D2D transmitter to others and interference from all others' caused to the generic D2D receiver. We formulate a mean-field game (MFG) theoretic framework with the interference mean-field approximation. We design the cost function combining both the performance of the D2D communication and cost for transmit power at the D2D transmitter. Within the MFG framework, we derive the related Hamilton-Jacobi-Bellman (HJB) and Fokker-Planck-Kolmogorov (FPK) equations. Then, a novel energy and interference aware power control policy is proposed, which is based on the Lax-Friedrichs scheme and the Lagrange relaxation. The numerical results are presented to demonstrate the spectrum and energy efficiency performances of our proposed approach. Index Terms—Device-to-device communication, mean field game, spectrum efficiency, energy efficiency.

6.2.3.2. Learning approaches

Fast initialization of cognitive radio systems is a key problem in a variety of wireless communication systems, particularly for public safety organizations in emergency crises. In the initialization problem, the goal is to rapidly identify an unoccupied frequency band. In this contribution [21], we formalize the initialization problem within the framework of active hypothesis testing. We characterize the optimal scanning policy in the case of at most one free band and show that the policy is computationally challenging. Motivated by this challenge for the implementation of the optimal policy and the need to cope with an unknown number of interferers larger than one, we propose the constrained DGF algorithm. We show that for strict constraints on the maximum number of observations, the constrained DGF algorithm can outperform the error probability of the state-of-the-art C-SPRT algorithm by an order of magnitude, for comparable average delays.

6.2.3.3. Asynchronous transmissions in VLC

In a visible light communications system (VLC), light sources are responsible for both illumination, communications and positioning. These light sources inevitably interfere each others at the receiver. To retain the appealing advantage that VLC systems can reuse existing lighting infrastructure, using an extra network to control or synchronize the light sources should be avoided. This work [31] proposes an uncoordinated multiple access scheme for VLC systems with positioning capability. The proposed scheme does not require a central unit to coordinate the transmission of the transmitters. Transmitters can be asynchronous with one another and with the receiver. Each transmitter is allocated a unique codeword with L chips for a system with up to (L-1)/2 transmitters where L is prime. Due to the linear growth in complexity with respect to number of transmitters, our proposed scheme is feasible for systems with large numbers of transmitters. Our novel decoder can minimize the effect of additive Gaussian noise at the receiver side. Simulation results show that the proposed decoder outperforms zero-forcing decoder.

6.2.4. Contributions in other application fields

6.2.4.1. Smart Grids

The advanced operation of future electricity distribution systems is likely to require significant observability of the different parameters of interest (e.g., demand, voltages, currents, etc.). Ensuring completeness of data is, therefore, paramount. In this context, an algorithm for recovering missing state variable observations in electricity distribution systems is presented in [47]. The proposed method exploits the low rank structure of the state variables via a matrix completion approach while incorporating prior knowledge in the form of second order statistics. Specifically, the recovery method combines nuclear norm minimization with Bayesian estimation. The performance of the new algorithm is compared to the information-theoretic limits and tested trough simulations using real data of an urban low voltage distribution system. The impact of the prior knowledge is analyzed when a mismatched covariance is used and for a Markovian sampling that introduces structure in the observation pattern. Numerical results demonstrate that the proposed algorithm is robust and outperforms existing state of the art algorithms.

In addition, Gaussian random attacks that jointly minimize the amount of information obtained by the operator from the grid and the probability of attack detection are presented in [38]. The construction of the attack is posed as an optimization problem with a utility function that captures two effects: firstly, minimizing the mutual information between the measurements and the state variables; secondly, minimizing the probability of attack detection via the Kullback-Leibler (KL) divergence between the distribution of the measurements with an attack and the distribution of the measurements without an attack. Additionally, a lower bound on the utility function achieved by the attacks constructed with imperfect knowledge of the second order statistics of the state variables is obtained. The performance of the attack construction using the sample covariance matrix of the state variables is numerically evaluated. The above results are tested in the IEEE 30–Bus test system.

6.2.4.2. Molecular Communications

Molecular communications is emerging as a technique to support coordination in nanonetworking, particularly in biochemical systems. In complex biochemical systems such as in the human body, it is not always possible to view the molecular communication link in isolation as chemicals in the system may react with chemicals used for the purpose of communication. There are two consequences: either the performance of the molecular communication link is reduced; or the molecular link disrupts the function of the biochemical system. As such, it is important to establish conditions when the molecular communication link can coexist with a biochemical system. In this work [45], we develop a framework to establish coexistence conditions based on the theory of chemical reaction networks. We then specialize our framework in two settings: an enzyme-aided molecular communication system; and a low-rate molecular communication system near a general biochemical system. In each case, we prove sufficient conditions to ensure coexistence.

6.3. Software Radio Programming Model

6.3.1. Dataflow programming models

Parallel computers have become ubiquitous and current processors contain several execution cores. A variety of low-level tools exist to program these chips efficiently, but they are considered hard to program, to maintain, and to debug, because they may exhibit non-deterministic behaviors. A solution is to use the higher-level formalism of dataflow programming to specify only the operations to perform and their dependencies. This paradigm may then be combined with the Polyhedral Model, which allows automatic parallelization and optimization of loop nests. This makes programming easier by delegating the low-level work to compilers and static analyzers [41].

Existing dataflow runtime systems either focus on the efficient execution of a single data-flow application, or on scenarii where applications are known a priori. CalMAR is a Multi-Application Dataflow Runtime built on top of the RVC-Cal environment that addresses the problem of executing an a priori unknown number of dataflow applications concurrently on the same multi-core system. Its efficiency has been validated compared to the RVC-CAL traditional approach [27].

6.3.2. Environments for transiently powered devices

An important research initiative has been started in Socrate recently: the study of the new NVRAM technology and its use in ultra-low power context. NVRAM stands for Non-Volatile Radom Access Memory. Non-Volatile mémory has been extising for a while (Nand Flash for instance) but was not sufficiently fast to be used as main memory. Many emerging technologies are forseen for Non-Volatile RAM to replace current RAM [58].

Socrate has started a work on the applicability of NVRAM for *transiantly powered systems*, i.e. systems which may undergo power outage at any time. This study resulted in the Sytare software presented in a research report and at the IoENT conference [39], [37], [17] and also to the starting of an Inria Project Lab: ZEP.

The Sytare software introduces a checkpointing system that takes into account peripherals (ADC, leds, timer, radio communication, etc.) present on all embedded system. Checkpointing is the natural solution to power outage: regularly save the state of the system in NVRAM so as to restore it when power is on again. However, no work on checkpointing took into account the restoration of the states of peripherals, Sytare provides this possibility

6.3.3. Filter synthesis

[46] presents an open-source tool for the automatic design of reliable finite impulse response (FIR) filters, targeting FPGAs. It shows that user intervention can be limited to a very small number of relevant input parameters: a high-level frequency-domain specification, and input/output formats. All the other design parameters are computed automatically, using novel approaches to filter coefficient quantization and direct-form architecture implementation. Our tool guarantees a priori that the resulting architecture respects the specification, while attempting to minimize its cost. Our approach is evaluated on a range of examples and shown to produce designs that are very competitive with the state of the art, with very little design effort.

Linear Time Invariant (LTI) filters are often specified and simulated using high-precision software, before being implemented in low-precision fixed-point hardware. A problem is that the hardware does not behave exactly as the simulation due to quantization and rounding issues. The article [53] advocates the construction of LTI architectures that behave as if the computation was performed with infinite accuracy, then rounded only once to the low-precision output format. From this minimalist specification, it is possible to deduce the optimal values of many architectural parameters, including all the internal data formats. This requires a detailed error analysis that captures not only the rounding errors but also their infinite accuracy specification at the minimal hardware cost. This generic methodology is detailed for the case of low-precision LTI filters in the Direct Form I implemented in FPGA logic. The approach is demonstrated by a fully automated and open-source architecture generator tool, and validated on a range of Infinite Impulse Response filters.

6.3.4. Hardware computer arithmetic

In collaboration with researchers from Istanbul, Turkey, operators have been developed for division by a small positive constant [8]. The first problem studied is the Euclidean division of an unsigned integer by a constant, computing a quotient and a remainder. Several new solutions are proposed and compared against the state of the art. As the proposed solutions use small look-up tables, they match well the hardware resources of an FPGA. The article then studies whether the division by the product of two constants is better implemented as two successive dividers or as one atomic divider. It also considers the case when only a quotient or only a remainder are needed. Finally, it addresses the correct rounding of the division of a floating-point number by a small integer constant. All these solutions, and the previous state of the art, are compared in terms of timing, area, and area-timing product. In general, the relevance domains of the various techniques are very different on FPGA and on ASIC.

[23] presents the new framework for semi-automatic circuit pipelining that will be used in future releases of the FloPoCo generator. From a single description of an operator or datapath, optimized implementations are obtained automatically for a wide range of FPGA targets and a wide range of frequency/latency trade-offs. Compared to previous versions of FloPoCo, the level of abstraction has been raised, enabling easier development, shorter generator code, and better pipeline optimization. The proposed approach is also more flexible

than fully automatic pipelining approaches based on retiming: in the proposed technique, the incremental construction of the pipeline along with the circuit graph enables architectural design decisions that depend on the pipeline.

FPGAs are well known for their ability to perform non-standard computations not supported by classical microprocessors. Many libraries of highly customizable application-specific IPs have exploited this capablity. However, using such IPs usually requires handcrafted HDL, hence significant design efforts. High Level Synthesis (HLS) lowers the design effort thanks to the use of C/C++ dialects for programming FPGAs. However, high-level C language becomes a hindrance when one wants to express non-standard computations: this languages was designed for programming microprocessors and carries with it many restrictions due to this paradigm. This is especially true when computing with floating-point, whose data-types and evaluation semantics are defined by the IEEE-754 and C11 standards. If the high-level specification was a computation-specific efficiency and HLS ease of use. It specifically targets the ubiquitous floating-point summation-reduction pattern. A source-to-source compiler transforms selected floating-point additions into sequences of simpler operators using non-standard arithmetic formats. This improves performance and accuracy for several benchmarks, while keeping the ease of use of a high-level C description.

The previous uses a variation of Kulisch' proposal to use an internal accumulator large enough to cover the full exponent range of floating-point. With it, sums and dot products become exact operations with one single rounding at the end. This idea failed to materialize in general purpose processors, as it was considered too slow and/or too expensive in terms of resources. It may however be an interesting option in reconfigurable computing, where a designer may use use smaller, more resource-efficient floating-point formats, knowing that sums and dot products will be exact. Another motivation of this work is that these exact operations, contrary to classical floating point ones, are associative, which enables better compiler optimizations in a High-Level Synthesis context. Kulisch proposed several architectures for the large accumulator, all using a sign/magnitude representation: the internal accumulator always represents a positive significand. [52] introduces an architecture using a 2's complement representation instead, and demonstrates improvements over Kulisch' proposal in both area and speed.

Another alternative to floating point is the UNUM, a variable length floating-point format conceived to replace the formats defined in the IEEE 754 standard. [18] discusses the implementation of UNUM arithmetic and reports hardware implementation results for the main UNUM operators.

SPADES Project-Team

6. New Results

6.1. Components and contracts

Participants: Alain Girault, Christophe Prévot, Sophie Quinton, Jean-Bernard Stefani.

6.1.1. Contracts for the negotiation of embedded software updates

We address the issue of change during design and after deployment in safety-critical embedded system applications, in collaboration with Thales and also in the context of the CCC project (http://ccc-project.org/).

In collaboration with Thales, we mostly focus on timing aspects with the objective to anticipate, at design time, future software evolutions and identify potential schedulability bottlenecks. This year we have paved the way for an extension, to more complex systems, of the approach developed last year to quantify the flexibility of a system with respect to timing. Specifically, we have focused on systems with task chains, and have proposed new methods for computing upper and lower bounds on task chain latencies. This work will be submitted to a conference early 2018. Our methods are also being implemented in the Thales tool chain, in order to be used in industry.

6.1.2. Location graphs

The design of configurable systems can be streamlined and made more systematic by adopting a componentbased structure, as demonstrated with the FRACTAL component model [38]. However, the formal foundations for configurable component-based systems, featuring higher-order capabilities where components can be dynamically instantiated and passivated, and non-hierarchical structures where components can be contained in different composites at the same time, are still an open topic. We have recently introduced the location graph model [70], where components are understood as graphs of locations hosting higher-order processes, and where component structures can be arbitrary graphs.

We have continued the development of location graphs, revisiting the underlying structural model (hypergraphs instead of graphs), and simplifying its operational semantics while preserving the model expressivity. Towards the development of a behavioral theory of location graphs, we have defined different notions of bisimilarity for location graphs and shown them to be congruences, although a fully fledged co-inductive characterization of contextual equivalence for location graphs is still in the works. This work has not yet been published.

6.2. Real-Time multicore programming

Participants: Pascal Fradet, Alain Girault, Gregor Goessler, Xavier Nicollin, Sophie Quinton.

6.2.1. Dynamicity in dataflow models

Recent dataflow programming environments support applications whose behavior is characterized by dynamic variations in resource requirements. The high expressive power of the underlying models (*e.g.*, Kahn Process Networks or the CAL actor language) makes it challenging to ensure predictable behavior. In particular, checking *liveness* (*i.e.*, no part of the system will deadlock) and *boundedness* (*i.e.*, the system can be executed in finite memory) is known to be hard or even undecidable for such models. This situation is troublesome for the design of high-quality embedded systems. In the past few years, we have proposed several parametric dataflow models of computation.

We have written a survey that provides a comprehensive description of the existing parametric dataflow MoCs (constructs, constraints, properties, static analyses) and compares them using a common example [10]. The main objectives are to help designers of streaming applications to choose the most suitable model for their needs and to pave the way for the design of new parametric MoCs.

We have studied *symbolic* analyses of dataflow graphs [11]. Symbolic analyses express the system performance as a function of parameters (*i.e.*, input and output rates, execution times). Such functions can be quickly evaluated for each different configuration or checked *w.r.t.* different quality-of-service requirements. These analyses are useful for parametric MoCs, partially specified graphs, and even for completely static SDF graphs. Our analyses compute the maximal throughput of acyclic synchronous dataflow graphs, the minimum required buffers for which as soon as possible (asap) scheduling achieves this throughput, and finally the corresponding input-output latency of the graph.

We have proposed an original method to deal with lossy communication channels in dataflow graphs. Lossy channels intrinsically violate the dataflow model of computation. Yet, many real-life applications encounter some form of lossy channels, for instance IoT applications. The challenge that is raised is how to manage the retransmissions in case of lost or corrupted tokens. The solution that we have proposed involves decomposing the execution of the dataflow graph into three phases: (i) an upstream phase where all the actors before the lossy channel are executed as usual; (ii) a lossy phase where only the two actors linked by the lossy channel are executed, as many times as required until all the tokens are correctly transmitted; and (iii) a downstream phase where all the actors after the lossy channel are executed as usual. When a graph includes several lossy channels, things become more complex. We rely on the Boolean parameters of BPDF [32] to encode enabling conditions on channels so that the execution follows this upstream-lossy-downstream semantics [12].

We are now studying models allowing dynamic reconfigurations of the *topology* of the dataflow graphs. This would be of interest for C-RAN and 5G telecommunication applications. This is one of the research topic of Arash Shafiei's PhD in collaboration with Orange Labs.

6.2.2. Synthesis of switching controllers using approximately bisimilar multiscale abstractions

The use of discrete abstractions for continuous dynamics has become standard in hybrid systems design (see *e.g.*, [71] and the references therein). The main advantage of this approach is that it offers the possibility to leverage controller synthesis techniques developed in the areas of supervisory control of discrete-event systems [66]. The first attempts to compute discrete abstractions for hybrid systems were based on traditional systems behavioral relationships such as simulation or bisimulation, initially proposed for discrete systems most notably in the area of formal methods. These notions require inclusion or equivalence of observed behaviors which is often too restrictive when dealing with systems observed over metric spaces. For such systems, a more natural abstraction requirement is to ask for closeness of observed behaviors. This leads to the notions of approximate simulation and bisimulation introduced in [45].

These approaches are based on sampling of time and space where the sampling parameters must satisfy some relation in order to obtain abstractions of a prescribed precision. In particular, the smaller the time sampling parameter, the finer the lattice used for approximating the state-space; this may result in abstractions with a very large number of states when the sampling period is small. However, there are a number of applications where sampling has to be fast; though this is generally necessary only on a small part of the state-space. We have been exploring two approaches to overcome this state-space explosion [4].

We are currently investigating an approach using mode sequences of given length as symbolic states for our abstractions. By using mode sequences of variable length we are able to adapt the granularity of our abstraction to the dynamics of the system, so as to automatically trade off precision against controllability of the abstract states.

6.2.3. Schedulability of weakly-hard real-time systems

We focus on the problem of computing tight deadline miss models for real-time systems, which bound the number of potential deadline misses in a given sequence of activations of a task. In practical applications, such guarantees are often sufficient because many systems are in fact not hard real-time [3].

We have developed an extension of sensitivity analysis for budgeting in the design of weakly-hard real-time systems [18]. During design, it often happens that some parts of a task set are fully specified while other parameters, *e.g.*, regarding recovery or monitoring tasks, will be available only much later. In such cases, sensitivity analysis can help anticipate how these missing parameters can influence the behavior of the whole

system so that a resource budget can be allocated to them. We have developed an extension of sensitivity analysis for deriving task budgets for systems with hard and weakly-hard requirements. This approach has been validated on synthetic test cases and a realistic case study given by our partner Thales.

A second contribution in this area is the application of our method for computing deadline miss models, called Typical Worst-Case Analysis (TWCA), to systems with finite queue capacity [9]. Finite ready queues, implemented by buffers, are a system reality in embedded real-time computing systems and networks. The dimensioning of queues is subject to constraints in industrial practice, and often the queue capacity is sufficient for typical system behavior, but is not sufficient in peak overload conditions. This may lead to overflow and consequently to the discarding of jobs. In this paper, we explore whether finite queue capacity can also be used as a mean of design in order to reduce workload peaks and thus shorten a transient overload phase. We have proposed an analysis method which is to the best of our knowledge the first one able to give (a) worst-case response times guarantees as well as (b) weakly-hard guarantees for tasks which are executed on a computing system with finite queues. Experimental results show that finite queue capacity may only have weak overload limiting effect. This unexpected outcome can be explained by the system behavior in the worst-case corner cases. The analysis shows nevertheless that a trade-off between weakly-hard guarantees and queue sizes is possible.

Finally, in collaboration with TU Braunschweig and Daimler we have worked on the application of the Logical Execution Time (LET) paradigm, according to which data are read and written at predefined time instants, to the automotive industry. Specifically, we have bridged the gap between LET, as it was originally proposed [59], and its current use in the automotive industry. One interesting outcome of this research is that it can nicely be combined with the use of TWCA. This work has not been published yet.

6.2.4. A Markov Decision Process approach for energy minimization policies

In the context of independent real-time sporadic jobs running on a single-core processor equipped with Dynamic Voltage and Frequency Scaling (DVFS), we have proposed a Markov Decision Process approach (MDP) to compute the scheduling policy that dynamically chooses the voltage and frequency level of the processor such that each job meets its deadline and the total energy consumption is minimized. We distinguish two cases: the finite case (there is a fixed time horizon) and the infinite case. In the finite case, several offline solutions exist, which all use the complete knowledge of all the jobs that will arrive within the time horizon [74], i.e., their size and deadlines. But clearly this is unrealistic in the embedded context where the characteristics of the jobs are not known in advance. Then, an optimal offline policy called Optimal Available (OA) has been proposed in [30]. Our goal was to improve this result by taking into account the statistical characteristics of the upcoming jobs. When such information is available (for instance by profiling the jobs based on execution traces), we have proposed several speed policies that optimize the expected energy consumption. We have shown that this general constrained optimization problem can be modeled as an unconstrained MDP by choosing a proper state space that also encodes the constraints of the problem. In particular, this implies that the optimal speed at each time can be computed using a dynamic programming algorithm, and that the optimal speed at any time t will be a deterministic function of the current state at time t [21]. This is the topic of Stephan Plassart's PhD, funded by the CASERM Persyval project.

6.2.5. Formal proofs for schedulability analysis of real-time systems

We have started to lay the foundations for computer-assisted formal verification of schedulability analysis results. Specifically, we contribute to Prosa [26], a foundational Coq library of reusable concepts and proofs for real-time schedulability analysis. A key scientific challenge is to achieve a modular structure of proofs for response time analysis. We intend to use this library for:

- 1. a better understanding of the role played by some assumptions in existing proofs;
- 2. a formal comparison of different analysis techniques; and
- 3. the verification of proof certificates generated by instrumenting (existing and efficient) analysis tools.

Two schedulability analyses for uniprocessor systems have been formalized and mechanically verified in Coq for:

- sporadic task sets scheduled according to the Time Division Multiple Access (TDMA) policy.
- periodic task sets with offsets scheduled according to the Fixed Priority Preemptive (FPP) policy [15].

The analysis for TDMA has mainly served to familiarize ourselves with the Prosa library. Schedulability analysis in presence of offsets is a non-trivial problem with a high computational complexity. In contrast to the traditional (offset oblivious) analysis, many scenarios must be tested and compared to identify which one represents the worst-case scenario. We have formalized and proved in Coq the basic analysis presented by Tindell [72]. This has allowed us to: (1) underline implicit assumptions made in Tindell's informal analysis; (2) ease the generalization of the verified analysis; (3) generate a certifier and an analyzer. We are investigating these two tools in terms of computational complexity and implementation effort, in order to provide a good solution to guarantee schedulability of industrial systems.

In parallel, we have worked on a Coq formalization of Typical Worst Case Analysis (TWCA). We aim to provide certified generic results for weakly-hard real-time systems in the form of (m, k) guarantees (a task may miss at most m deadlines out of k consecutive activations). So far, we have adapted the initial TWCA for arbitrary schedulers. The proof relies on a practical definition of the concept of busy window which amounts to being able to perform a local response time analysis. We provide such an instantiation for Fixed Priority Preemptive (FPP) schedulers as in the original paper. Future work includes making the state of the art TWCA suitable for formal proofs, exploring more complex systems (*e.g.*, bounded buffers) and providing instantiations of our results for other scheduling policies.

6.3. Language Based Fault-Tolerance

Participants: Pascal Fradet, Alain Girault, Gregor Goessler, Jean-Bernard Stefani, Martin Vassor.

6.3.1. Fault Ascription in Concurrent Systems

The failure of one component may entail a cascade of failures in other components; several components may also fail independently. In such cases, elucidating the exact scenario that led to the failure is a complex and tedious task that requires significant expertise.

The notion of causality (did an event e cause an event e'?) has been studied in many disciplines, including philosophy, logic, statistics, and law. The definitions of causality studied in these disciplines usually amount to variants of the counterfactual test "e is a cause of e' if both e and e' have occurred, and in a world that is as close as possible to the actual world but where e does not occur, e' does not occur either". In computer science, almost all definitions of logical causality — including the landmark definition of [54] and its derivatives — rely on a causal model that. However, this model may not be known, for instance in presence of black-box components. For such systems, we have been developing a framework for blaming that helps us establish the causal relationship between component failures and system failures, given an observed system execution trace. The analysis is based on a formalization of counterfactual reasoning [6].

In [16] we have discussed several shortcomings of existing approaches to counterfactual causality from the computer science perspective, and sketched lines of work to try and overcome these issues. In particular, research on counterfactual causality analysis has been marked, since its early days, by a succession of definitions of causality that are informally (in)validated against human intuition on mostly simple examples, see *e.g.*, [54], [53]. We call this approach TEGAR, *textbook example guided analysis refinement*. As pointed out in [48], it suffers from its dependence on the tiny number and incompleteness of examples in the literature, and from the lack of stability of the intuitive judgments against which the definitions are validated. We have argued that we need a formalization of counterfactual causality based on *first principles*, in the sense that causality definitions should not be driven by individual examples but constructed from a set of precisely specified requirements. Example of such requirements are robustness of causation under equivalence of models, and well-defined behavior under abstraction and refinement. To the best of our knowledge, none of the existing causality analysis techniques provides sufficient guarantees in this regard.

We are currently working on a revised version of our general semantic framework for fault ascription in [50] that satisfies a set of formally stated requirements, and on its instantiation to acyclic models of computation, in order to compare our approach with the standard definition of *actual causality* proposed by Halpern and Pearl.

6.3.2. Tradeoff exploration between energy consumption and execution time

We have continued our work on multi-criteria scheduling, in two directions. First, in the context of dynamic applications that are launched and terminated on an embedded homogeneous multi-core chip, under execution time and energy consumption constraints, we have proposed a two layer adaptive scheduling method [14]. In the first layer, each application (represented as a DAG of tasks) is scheduled statically on subsets of cores: 2 cores, 3 cores, 4 cores, and so on. For each size of these sets (2, 3, 4, ...), there may be only one topology or several topologies. For instance, for 2 or 3 cores there is only one topology (a "line"), while for 4 cores there are three distinct topologies ("line", "square", and "T shape"). Moreover, for each topology, we generate statically several schedules, each one subject to a different total energy consumption constraint, and consequently with a different Worst-Case Reaction Time (WCRT). Coping with the energy consumption constraints is achieved thanks to Dynamic Frequency and Voltage Scaling (DVFS). In the second layer, we use these pre-generated static schedules to reconfigure dynamically the applications running on the multi-core each time a new application is launched or an existing one is stopped. The goal of the second layer is to perform a dynamic global optimization of the configuration, such that each running application meets a pre-defined quality-of-service constraint (translated into an upper bound on its WCRT) and such that the total energy consumption be minimized. For this, we (i) allocate a sufficient number of cores to each active application, (ii) allocate the unassigned cores to the applications yielding the largest gain in energy, and (iii) choose for each application the best topology for its subset of cores (*i.e.*, better than the by default "line" topology). This is a joint work with Ismail Assayad (U. Casablanca, Morocco) who visits the team regularly.

Second, we have proposed the first of its kind multi-criteria scheduling heuristics for a DAG of tasks onto an homogeneous multi-core chip, optimizing the execution time, the reliability, the power consumption, and the temperature. Specifically, we have worked on the static scheduling minimizing the execution time of the application under the multiple constraints that the reliability, the power consumption, and the temperature remain below some given thresholds. There are multiple difficulties: (i) the reliability is not an invariant measure w.r.t. time, which makes it impossible to use backtrack-free scheduling algorithms such as list scheduling [28]; to overcome this, we adopt instead the Global System Failure Rate (GSFR) as a measure of the system's reliability, which is invariant with time [46]; (ii) keeping the power consumption under a given threshold requires to lower the voltage and frequency, but this has a negative impact both on the execution time and on the GSFR; keeping the GSFR below a given threshold requires to replicate the tasks on multiple cores, but this has a negative impact both on the execution time, on the power consumption, and on the temperature; (iii) keeping the temperature below a given threshold is even more difficult because the temperature continues to increase even after the activity stops, so each scheduling decision must be assessed not based on the current state of the chip (*i.e.*, the temperature of each core) but on the state of the chip at the end of the candidate task, and cooling slacks must be inserted. We have proposed a multi-criteria scheduling heuristics to address these challenges. It produces a static schedule of the given application graph and the given architecture description, such that the GSFR, power, and temperature thresholds are satisfied, and such that the execution time is minimized. We then combine our heuristic with a variant of the ε -constraint method [52] in order to produce, for a given application graph and a given architecture description, its entire Pareto front in the 4D space (exec. time, GSFR, power, temp.). This is a joint work with Athena Abdi and Hamid Zarandi from Amirkabir U., Iran, who have visited the team in 2016.

6.3.3. Concurrent flexible reversibility

Reversible concurrent models of computation provide natively what appears to be very fine-grained checkpoint and recovery capabilities. We have made this intuition clear by formally comparing a distributed algorithm for checkpointing and recovery based on causal information, and the distributed backtracking algorithm that lies at the heart of our reversible higher-order pi-calculus. We have shown that (a variant of) the reversible higher-order calculus with explicit rollback can faithfully encode a distributed causal checkpoint and recovery algorithm. The reverse is also true but under precise conditions, which restrict the ability to rollback a computation to an identified checkpoint. This work has currently not been published.

STEEP Project-Team

7. New Results

7.1. Calculating spatial urban sprawl indices using open data

Urban sprawl has been related to numerous negative environmental and socioeconomic impacts. Meanwhile, urban areas have been growing at alarming rates, urging for assessing sprawl towards sustainable development. However, sprawl is an elusive term and different approaches to measure it have lead to heterogeneous results. Moreover, most studies rely on private/commercial data-sets and their software is rarely made public, impeding research reproducibility and comparability. Furthermore, many works give as result a unique value for a region of analysis, dismissing local spatial diversity that is vital for urban planners and policy makers.

Based on our last year's initial work [19], we have developed an extended open source framework for assessing urban sprawl using open data. Locations of residential and activity units are used to measure mixed use development and built-up dispersion, whereas the street network is used to measure accessibility between different land uses. Sprawl patterns are identified, and the resulting spatial information allows focusing on particular neighborhoods for a fine-grained analysis, as well as visualizing each sprawl dimension separately.

This work has been published in [10], [14], [15] and the associated implementation is available as open source software (see previous section).

7.2. A method for downscaling open population data

To extend our ongoing work on urban sprawl indicators (see above), we have developed a method to perform dissagregated population estimations at building level using open data. Our goal is to estimate the number of people living at the fine level of individual households by using open urban data and coarse-scaled population data. First, a fine scale description of residential land use per building is built using OpenStreetMap. Then, using coarse-scale gridded population data, we perform the down-scaling for each household given their containing area for residential usage. We rely solely on open data in order to ensure replicability, and to be able to apply our method to any city in the world, as long as sufficient data exists. The evaluation is carried out using fine-grained census block data for cities in France as ground-truth.

This work is published in [11] and the associated software implementation is made available as open source code at https://github.com/lgervasoni/urbansprawl.

7.3. Ecological accounting

The most important result obtained on this front bears on the quantification of the errors associated with the national road freight transport database (SITRAM). This database is informed year by year through a dedicated sampling campaign, but the errors associated with the various types of material goods transported have never been quantified. This was achieved by our team through the use of appropriate error estimators. This result is eagerly awaited by a number of scientific teams and public territorial agencies. Furthermore, the methodology that we have developed can easily be transposed to other countries. This result constitutes an important piece in the overall effort that the team has devoted to the question of the quantification of uncertainties in material flow analyses. This work is in the press in the leading journal in this field, the *Journal of Industrial Ecology*.

7.4. A computer framework for measuring urban land-use mix

The number of people living in cities has been increasing considerably since 1950, from 746 million to 3.9 billion in 2014, and more than 66% of the world's population are projected to live in urban areas by 2050. As this continuing population growth and urbanization are projected to add 2.5 billion people to the world's urban population in 30 years, this situation brings new challenges on how to conceive cities that host such amounts of population in a sustainable way. This sustainability question should address several aspects, ranging from economical to social and environmental matters among others. In this work, we focus on the formalization of a measure of mixed use development or land use mix in a city, i.e. how the structure of the city can help to provide a car-free sustainable living. Such type of land use mix has been largely proven to contain beneficial outcomes in terms of sustainability and to positively contribute to societal outcome, health, and public transportation among others. We developped a framework to compute mixed uses development index. A main caracteristic of our approach is to use only crowd-sourcing data (from OpenStreetMap) to extract the geo-localized land uses. Due to the universality of this data source, we are able to process any geographical area in the world, as long as sufficient data are available in OSM. A Kernel Density Estimation is performed for each of the land uses, outputing the spatial distribution of the different land uses. Based on this representation, a measure of land use mix is then calculated using the Entropy Index. The resulting GIS output shows enriched information for urban planners, supporting and aiding the decision-making procedure.

The framework, still in the phase of validation, was applied on the cities of London and Grenoble [19]. Future work includes integrating the LUM output for measuring the urban sprawl phenomenon and performing numerical interpretations of desirable mixed use values. We will also study the potential integration to transportation models, where land use mix correlation with the activities and residential uses can help to improve demand estimation. In addition, further investigation can be done by means of analyzing in detail the different types of activities. Finally, the estimation of LUM can be refined by taking into account, besides their location, the accessibility between different land uses, which is partly conditioned by the transportation infrastructure.

7.5. Calibration and sensitivity analysis for LUTI models

This year, we have consolidated our previous works on calibration of LUTI models, in particular of the Tranus model. The developed approaches are currently applied to instantiate a complete Tranus model for the Grenoble catchment area, in collaboration with AURG (Urban Planning Agency of the Grenoble area) and Brian Morton (U North Carolina); see [12], [13].

THOTH Project-Team

7. New Results

7.1. Visual recognition in images

7.1.1. Dynamic Filters in Graph Convolutional Networks

Participants: Nitika Verma, Edmond Boyer [MORPHEO, Inria Grenoble], Jakob Verbeek.

Convolutional neural networks (CNNs) have massively impacted visual recognition in 2D images, and are now ubiquitous in state-of-the-art approaches. While CNNs naturally extend to other domains, such as audio and video, where data is also organized in rectangular grids, they do not easily generalize to other types of data such as 3D shape meshes, social network graphs or molecular graphs. In our recent paper [39], we propose a novel graph-convolutional network architecture to handle such data. The architecture builds on a generic formulation that relaxes the 1-to-1 correspondence between filter weights and data elements around the center of the convolution, see Figure 1 for an illustration. The main novelty of our architecture is that the shape of the filter is a function of the features in the previous network layer, which is learned as an integral part of the neural network. Experimental evaluations on digit recognition, semi-supervised document classification, and 3D shape correspondence yield state-of-the-art results, significantly improving over previous work for shape correspondence.



Figure 1. Left: Illustration of a standard CNN, representing the parameters as a set of $M = w \times h$ weight matrices, each of size $D \times E$. Each weight matrix is associated with a single relative position in the input patch. Right: Our graph convolutional network, where each node in the input patch is associated in a soft manner to each of the M weight matrices based on its features using the weight $q_m(\mathbf{x}_i, \mathbf{x}_i)$.

7.1.2. LCR-Net: Localization-Classification-Regression for Human Pose

Participants: Grégory Rogez, Philippe Weinzaepfel, Cordelia Schmid.

In this paper [24], we propose an end-to-end architecture for joint 2D and 3D human pose estimation in natural images. Key to our approach is the generation and scoring of a number of pose proposals per image, which allows us to predict 2D and 3D pose of multiple people simultaneously. See example in Figure 2. Hence, our approach does not require an approximate localization of the humans for initialization. Our architecture, named LCR-Net, contains 3 main components: 1) the pose proposal generator that suggests potential poses at different locations in the image; 2) a classifier that scores the different pose proposals ; and 3) a regressor that refines pose proposals both in 2D and 3D. All three stages share the convolutional feature layers and are trained jointly. The final pose estimation is obtained by integrating over neighboring pose hypotheses , which is shown to improve over a standard non maximum suppression algorithm. Our approach significantly outperforms the state of the art in 3D pose estimation on Human3.6M, a controlled environment. Moreover, it shows promising results on real images for both single and multi-person subsets of the MPII 2D pose benchmark.


Figure 2. Examples of joint 2D-3D pose detections in a natural image. Even in case of occlusion or truncation, we estimate the joint locations by reasoning in term of full-body 2D-3D poses.

7.1.3. Incremental Learning of Object Detectors without Catastrophic Forgetting

Participants: Konstantin Shmelkov, Cordelia Schmid, Karteek Alahari.

In the paper [25] we introduce a framework for incremental learning of object detectors based on convolutional neural networks, i.e., adapting the original model trained on a set of classes to additionally detect objects of new classes, in the absence of the initial training data. They suffer from "catastrophic forgetting"—an abrupt degradation of performance on the original set of classes, when the training objective is adapted to the new classes. We present a method to address this issue, and learn object detectors incrementally, when neither the original training data nor annotations for the original classes in the new training set are available. The core of our proposed solution is a loss function to balance the interplay between predictions on the new classes and a new distillation loss which minimizes the discrepancy between responses for old classes from the original and the updated networks (see Figure 3). This incremental learning can be performed multiple times, for a new set of classes in each step, with a moderate drop in performance compared to the baseline network trained on the ensemble of data. We present object detection results on the PASCAL VOC 2007 and COCO datasets, along with a detailed empirical analysis of the approach.



Figure 3. Overview of our framework for learning object detectors incrementally. It is composed of a frozen copy of the detector (Network A) and the detector (Network B) adapted for the new class(es).

7.1.4. BlitzNet: A Real-Time Deep Network for Scene Understanding

Participants: Mikita Dvornik, Konstantin Shmelkov, Julien Mairal, Cordelia Schmid.

Real-time scene understanding has become crucial in many applications such as autonomous driving. In this work [16], we propose a deep architecture, called BlitzNet, that jointly performs object detection and semantic segmentation in one forward pass, allowing real-time computations. Besides the computational gain of having a single network to perform several tasks, we show that object detection and semantic segmentation benefit from each other in terms of accuracy. Experimental results for VOC and COCO datasets show state-of-the-art performance for object detection and segmentation among real time systems.

To achieve these goals we designed a novel architecture (see fig.4 that naturally suits well for each of the tasks by allowing embedding of precise local details and reach global semantical information in a single featurespace. This solution allows to better localize and segment small objects. The usage of common architecture for both tasks allows more efficient feature sharing and and a simple training procedure that introduces benefits for semantic segmentation by adding extra data with only bounding box annotations. To reduce the computational overhead introduced by the upscale stream we slightly modify the NMS procedure to speed up post-processing in test time without no effect on detection accuracy.



Figure 4. Atchirecture of the neutal network used in the project. The middle stream (blue and violet blocks) is shared between the task. The upper stream (in green) predicts categories of object proposals and their localization offsets to perform object detection. The bottom stream classifies each pixel to output a semantic segmentation mask.

7.1.5. SCNet: Learning semantic correspondence

Participants: Kai Han, Rafael Rezende, Bumsub Ham, Kwan-Yee Wong, Minsu Cho, Cordelia Schmid, Jean Ponce.

In this work [17], 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 by 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.

7.1.6. Auxiliary Guided Autoregressive Variational Autoencoders

Participants: Thomas Lucas, Jakob Verbeek.

Generative modeling of high-dimensional data is a key problem in machine learning. Successful approaches include latent variable models and autoregressive models. The complementary strengths of these approaches, to model global and local image statistics respectively, suggest hybrid models combining the strengths of both. Our contribution is to train such hybrid models using an auxiliary loss function that controls which information is captured by the latent variables and what is left to the autoregressive decoder, as illustrated in Figure 6. In contrast, prior work on such hybrid models needed to limit the capacity of the autoregressive modeling. Our approach results in models with meaningful latent variables and only rely on autoregressive modeling. Our approach results in models with meaningful latent variable representations, and which rely on powerful autoregressive decoders to model image details. Our model generates qualitatively convincing samples, and yields state-of-the-art quantitative results.

7.1.7. Areas of Attention for Image Captioning

Participants: Marco Pedersoli, Thomas Lucas, Jakob Verbeek.



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_a for geometry.



Figure 6. Schematic illustration of our auxiliary guided autoregressive variational autoencoder (AGAVE). An input image is encoded into a latent representation and decoded back into an image. This first reconstruction is guided by an auxiliary maximum likelihood loss and regularized with a Kullback-Liebler divergence. An autoregressive model is then conditionned on the auxiliary reconstruction and also trained with maximum likelihood.

We propose "Areas of Attention", a novel attention-based model for automatic image captioning. Our approach models the dependencies between image regions, caption words, and the state of an RNN language model, using three pairwise interactions. In contrast to previous attention-based approaches that associate image regions only to the RNN state, our method allows a direct association between caption words and image regions. During training these associations are inferred from image-level captions, akin to weakly-supervised object detector training. These associations help to improve captioning by localizing the corresponding regions during testing. We also propose and compare different ways of generating attention areas: CNN activation grids, object proposals, and spatial transformers nets applied in a convolutional fashion, as illustrated in Figure 7. Spatial transformers give the best results. They allow for image specific attention areas, and can be trained jointly with the rest of the network. Our attention mechanism and spatial transformer attention areas together yield state-of-the-art results on the MSCOCO dataset.



Figure 7. An attention mechanism jointly predicts the next caption word and the corresponding region at each time-step given the RNN state (top). Attention areas can be defined using CNN activation grids or object proposals (left and middle), as used in previous work. We also present a end-to-end trainable convolutional spatial transformer approach to compute image specific attention areas (bottom).

7.1.8. Enhancing Energy Minimization Framework for Scene Text Recognition with Top-Down Cues

Participants: Anand Mishra, Karteek Alahari, C. v. Jawahar.

Color and strokes are the salient features of text regions in an image. In this work, presented in [10], we use both these features as cues, and introduce a novel energy function to formulate the text binarization problem. The minimum of this energy function corresponds to the optimal binarization. We minimize the energy function with an iterative graph cut based algorithm. Our model is robust to variations in foreground and background as we learn Gaussian mixture models for color and strokes in each iteration of the graph cut. We show results on word images from the challenging ICDAR 2003/2011, born-digital image and street

view text datasets, as well as full scene images containing text from ICDAR 2013 datasets, such as the ones shown in Figure 8, and compare our performance with state-of-the-art methods. Our approach shows significant improvements in performance under a variety of performance measures commonly used to assess text binarization schemes. In addition, our method adapts to diverse document images, like text in videos, handwritten text images.



Figure 8. Sample images we consider in the work presented in [10]. Due to large variations in foreground and background colors, most of the popular binarization techniques in the literature tend to fail on such images.

7.1.9. Learning deep face representations using small data

Participants: Guosheng Hu, Xiaojiang Peng [Hengyang Normal University, China], Yongxin Yang [Queen Mary University of London, UK], Timothy Hospedales [University of Edinburgh, UK], Jakob Verbeek.

Deep convolutional neural networks have recently proven extremely effective for difficult face recognition problems in uncontrolled settings. To train such networks very large training sets are needed with millions of labeled images. For some applications, such as near-infrared (NIR) face recognition, such large training datasets are, however, not publicly available and very difficult to collect. In our recent paper [8] we propose a method to generate very large training datasets of synthetic images by compositing real face images in a given dataset. Our approach replaces facial parts (nose, mouth, eyes) from one face with those of another, see Figure 9 for several examples. We show that this method enables to learn models from as few as 10,000 training images, which perform on par with models trained from 500,000 images without using our data augmentation. Using our approach we also improve the state-of-the-art results on the CASIA NIR-VIS heterogeneous face recognition dataset.

7.1.10. Invariance and Stability of Deep Convolutional Representations

Participants: Alberto Bietti, Julien Mairal.

In [13] and [29], we study deep signal representations that are near-invariant to groups of transformations and stable to the action of diffeomorphisms without losing signal information. This is achieved by generalizing the multilayer kernel introduced in the context of convolutional kernel networks and by studying the geometry of the corresponding reproducing kernel Hilbert space. We show that the signal representation is stable, and that models from this functional space, such as a large class of convolutional neural networks, may enjoy the same stability.

7.1.11. Weakly-supervised learning of visual relations

Participants: Julia Peyre, Ivan Laptev, Cordelia Schmid, Josef Sivic.



Figure 9. Illustration of generated synthetic training images, mixing either images of the same person (left) or from different people (right).

This work [23] introduces a novel approach for modeling visual relations between pairs of objects. We call relation a triplet of the form (*subject*, *predicate*, *object*) where the predicate is typically a preposition (eg. 'under', 'in front of') or a verb ('hold', 'ride') that links a pair of objects (*subject*, *object*). Learning such relations is challenging as the objects have different spatial configurations and appearances depending on the relation in which they occur. Another major challenge comes from the difficulty to get annotations, especially at box-level, for all possible triplets, which makes both learning and evaluation difficult. The contributions of this paper are threefold. First, we design strong yet flexible visual features that encode the appearance and spatial configuration for pairs of objects. Second, we propose a weakly-supervised discriminative clustering model to learn relations from image-level labels only. Third we introduce a new challenging dataset of unusual relations (UnRel) together with an exhaustive annotation, that enables accurate evaluation of visual relationship dataset significantly improving performance on previously unseen relations (zero-shot learning), and confirm this observation on our newly introduced UnRel dataset. Example results are shown in Figure 10.

7.1.12. Learning from Synthetic Humans

Participants: Gül Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael Black, Ivan Laptev, Cordelia Schmid.

Estimating human pose, shape, and motion from images and video are fundamental challenges with many applications. Recent advances in 2D human pose estimation use large amounts of manually-labeled training data for learning convolutional neural networks (CNNs). Such data is time consuming to acquire and difficult to extend. Moreover, manual labeling of 3D pose, depth and motion is impractical. In [28], 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 11. 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 at CVPR 2017 [28].



Figure 10. Examples of top retrieved pairs of boxes in UnRel dataset for unusual queries with our weakly supervised model



Figure 11. We generate photo-realistic synthetic images and their corresponding ground truth for learning pixel-wise classification problems: human parts segmentation and depth estimation. The convolutional neural network trained only on synthetic data generalizes on real images sufficiently for both tasks.

7.2. Visual recognition in videos

7.2.1. Detecting Parts for Action Localization

Participants: Nicolas Chesneau, Grégory Rogez, Karteek Alahari, Cordelia Schmid.

we propose a new framework for action localization that tracks people in videos and extracts full-body human tubes, i.e., spatio-temporal regions localizing actions, even in the case of occlusions or truncations. This is achieved by training a novel human part detector that scores visible parts while regressing full-body bounding boxes. The core of our method is a convolutional neural network which learns part proposals specific to certain body parts. These are then combined to detect people robustly in each frame. Our tracking algorithm connects the image detections temporally to extract full-body human tubes. We apply our new tube extraction method on the problem of human action localization, on the popular JHMDB dataset, and a very recent challenging dataset DALY (Daily Action Localization in YouTube), showing state-of-the-art results. An overview of the method is show in Figure 12. More details are provided in [15].



Figure 12. Two example videos from the DALY dataset to illustrate the difference between our human tube extraction and previous methods.

7.2.2. Learning from Web Videos for Event Classification

Participants: Nicolas Chesneau, Karteek Alahari, Cordelia Schmid.

Traditional approaches for classifying event videos rely on a manually curated training dataset. While this paradigm has achieved excellent results on benchmarks such as TrecVid multimedia event detection (MED) challenge datasets, it is restricted by the effort involved in careful annotation. Recent approaches have attempted to address the need for annotation by automatically extracting images from the web, or generating queries to retrieve videos. In the former case, they fail to exploit additional cues provided by video data, while in the latter, they still require some manual annotation to generate relevant queries. We take an alternate approach in this paper, leveraging the synergy between visual video data and the associated textual metadata, to learn event classifiers without manually annotating any videos. Specifically, we first collect a video dataset with queries constructed automatically from textual description of events, prune irrelevant videos with text and video data, and then learn the corresponding event classifiers. We evaluate this approach in the challenging setting where no manually annotated training set is available, i.e., EK0 in the TrecVid challenge, and show state-of-the-art results on MED 2011 and 2013 datasets. An overview of the method is show in Figure 13. More details are provided in [4].

7.2.3. Learning Motion Patterns in Videos

Participants: Pavel Tokmakov, Karteek Alahari, Cordelia Schmid.



Changing a vehicle tire One or more people work to replace a tire on a vehicle A vehicle is any device, motorized or not, used to transport people and/or other items. Thes are ring-shaped inflated objects, usually made of rubber, that fit over the wheel of a vehicle. The process for replacing a tire includes removing the existing tire and installing the new tire onto the wheel of the vehicle. Tires typically are replaced because they are damaged or worn down. If a tire is damaged and loses air pressure as a result, it is called a "flat tire". Generally the driver of the vehicle with a flat tire will stop the vehicle as soon as possible and replace the affected tire with a temporary tire called a "spare tire", which may be stored elsewhere on/in the vehicle. In other cases, the tire may be changed not by the vehicle operator, but by a professional (e.g. a mechanic) who may use dedicated tools and work in a repair shop or similarsetting. Evidential description: scene: garage, outdoors, street, parking lot objects/people: tire, lug wrench, hubcap, vehicle (car, bike, lawnmower, etc), tire jack **activities:**removing hubcap, turning lugwrench, unscrewing bolts, pulling rim out of tire **audio**: narration of the process; sounds of tools being used; street/traffic noise; background noises from repair shop Text + visual concepts **Query expansion** Web videos Pruned training set pruning Visual concepts space Selected Videos List of queries x Changing vehicle tire Changing car tire Changing vehicle tire driver Changing vehicle tire hubcap Changing vehicle tire garage Changing bike tire 12 Visual concepts space 3 🗙 × Pruned Videos

Figure 13. Overview: Given the description of an event ("Event description"), relevant queries are automatically generated ("Query generation") to collect an initial training set ("Web videos"). Text metadata and visual concepts extracted from these videos are used to select the relevant ones automatically ("Text + visual concepts pruning"), and build a training set for event classification ("Pruned training set").

× 🗸

→X →X X

The problem of determining whether an object is in motion, irrespective of the camera motion, is far from being solved. We address this challenging task by learning motion patterns in videos [26]. The core of our approach is a fully convolutional network (see Figure 14), which is learnt entirely from synthetic video sequences, and their ground-truth optical flow and motion segmentation. This encoder-decoder style architecture first learns a coarse representation of the optical flow field features, and then refines it iteratively to produce motion labels at the original high-resolution. The output label of each pixel denotes whether it has undergone independent motion, i.e., irrespective of the camera motion. We demonstrate the benefits of this learning framework on the moving object segmentation task, where the goal is to segment all the objects in motion. To this end we integrate an objectness measure into the framework. Our approach outperforms the top method on the recently released DAVIS benchmark dataset, comprising real-world sequences, by 5.6%. We also evaluate on the Berkeley motion segmentation database, achieving state-of-the-art results.



Figure 14. Our motion pattern network: MP-Net. The blue arrows in the encoder part (a) denote convolutional layers, together with ReLU and max-pooling layers. The red arrows in the decoder part (b) are convolutional layers with ReLU, 'up' denotes 2 x 2 upsampling of the output of the previous unit. The unit shown in green represents bilinear interpolation of the output of the last decoder unit

7.2.4. Learning Video Object Segmentation with Visual Memory

Participants: Pavel Tokmakov, Karteek Alahari, Cordelia Schmid.

This paper [27] addresses the task of segmenting moving objects in unconstrained videos. We introduce a novel two-stream neural network with an explicit memory module shown in Figure 15 to achieve this. The two streams of the network encode spatial and temporal features in a video sequence respectively, while the memory module captures the evolution of objects over time. The module to build a "visual memory" in video, i.e., a joint representation of all the video frames, is realized with a convolutional recurrent unit learned from a small number of training video sequences. Given a video frame as input, our approach assigns each pixel an object or background label based on the learned spatio-temporal features as well as the "visual memory" specific to the video, acquired automatically without any manually-annotated frames. The visual memory is implemented with convolutional gated recurrent units, which allows to propagate spatial information over time. We evaluate our method extensively on two benchmarks, DAVIS and Freiburg-Berkeley motion segmentation datasets, and show state-of-the-art results. For example, our approach outperforms the top method on the DAVIS dataset by nearly 6%. We also provide an extensive ablative analysis to investigate the influence of each component in the proposed framework.

7.2.5. Learning to Segment Moving Objects

Participants: Pavel Tokmakov, Cordelia Schmid, Karteek Alahari.

We study the problem of segmenting moving objects in unconstrained videos [37]. Given a video, the task is to segment all the objects that exhibit independent motion in at least one frame. We formulate this as a learning problem and design our framework with three cues: (i) independent object motion between a pair of frames, which complements object recognition, (ii) object appearance, which helps to correct errors in motion



Figure 15. Overview of our segmentation approach. Each video frame is processed by the appearance (green) and the motion (yellow) networks to produce an intermediate two-stream representation. The ConvGRU module combines this with the learned visual memory to compute the final segmentation result.

estimation, and (iii) temporal consistency, which imposes additional constraints on the segmentation. The framework is a two-stream neural network with an explicit memory module, shown in Figure 15. The two streams encode appearance and motion cues in a video sequence respectively, while the memory module captures the evolution of objects over time, exploiting the temporal consistency. The motion stream is a convolutional neural network trained on synthetic videos to segment independently moving objects in the optical flow field. The module to build a "visual memory" in video, i.e., a joint representation of all the video frames, is realized with a convolutional recurrent unit learned from a small number of training video sequences. For every pixel in a frame of a test video, our approach assigns an object or background label based on the learned spatio-temporal features as well as the "visual memory" specific to the video. We evaluate our method extensively on three benchmarks, DAVIS, Freiburg-Berkeley motion segmentation dataset and SegTrack. In addition, we provide an extensive ablation study to investigate both the choice of the training data and the influence of each component in the proposed framework.

7.2.6. Joint learning of object and action detectors

Participants: Vasiliki Kalogeiton, Philippe Weinzaepfel, Vittorio Ferrari, Cordelia Schmid.

While most existing approaches for detection in videos focus on objects or human actions separately, we aim at jointly detecting objects performing actions, such as cat eating or dog jumping [19]. We introduce an end-to-end multitask objective that jointly learns object-action relationships, see Figure 16. We compare it with different training objectives, validate its effectiveness for detecting objects-actions in videos, and show that both tasks of object and action detection benefit from this joint learning. Moreover, the proposed architecture can be used for zero-shot learning of actions: our multitask objective leverages the commonalities of an action performed by different objects, e.g. dog and cat jumping , enabling to detect actions of an object without training with these object-actions pairs. In experiments on the A2D dataset, we obtain state-of-the-art results on segmentation of object-action pairs. We finally apply our multitask architecture to detect visual relationships between objects in images of the VRD dataset.

7.2.7. Action Tubelet Detector for Spatio-Temporal Action Localization

Participants: Vasiliki Kalogeiton, Philippe Weinzaepfel, Vittorio Ferrari, Cordelia Schmid.

Current state-of-the-art approaches for spatio-temporal action detection rely on detections at the frame level that are then linked or tracked across time. In this paper [18], we leverage the temporal continuity of videos instead of operating at the frame level. We propose the ACtion Tubelet detector (ACT-detector) that takes as input a sequence of frames and outputs tubelets, ie., sequences of bounding boxes with associated scores, see Figure 17. The same way state-of-the-art object detectors rely on anchor boxes, our ACT-detector is based on anchor cuboids. We build upon the state-of-the-art SSD framework. Convolutional features are extracted for each frame, while scores and regressions are based on the temporal stacking of these features, thus exploiting information from a sequence. Our experimental results show that leveraging sequences of frames



Figure 16. Overview of our end-to-end multitask network architecture for joint object-action detection in videos. Blue color represents convolutional layers while green represents fully connected layers. The end-to-end training is done by concatenating the fully connected layers from both streams. Here, pO and pA are the outputs of the two branches that predict the object and action labels, resulting in the final joint object-action loss.

significantly improves detection performance over using individual frames. The gain of our tubelet detector can be explained by both more relevant scores and more precise localization. Our ACT-detector outperforms the state of the art methods for frame-mAP and video-mAP on the J-HMDB and UCF-101 datasets, in particular at high overlap thresholds.



Figure 17. Overview of our ACT-detector. Given a sequence of frames, we extract convolutional features with weights shared between frames. We stack the features from subsequent frames to predict scores and regress coordinates for the anchor cuboids (middle figure, blue color). Depending on the size of the anchors, the features come from different convolutional layers (left figure, color coded: yellow, red, purple, green). As output, we obtain tubelets (right figure, yellow color).

7.3. Large-scale statistical learning

7.3.1. Stochastic Optimization with Variance Reduction for Infinite Datasets with Finite-Sum Structure

Participants: Alberto Bietti, Julien Mairal.

Stochastic optimization algorithms with variance reduction have proven successful for minimizing large finite sums of functions. Unfortunately, these techniques are unable to deal with stochastic perturbations of input data, induced for example by data augmentation. In such cases, the objective is no longer a finite sum, and the main candidate for optimization is the stochastic gradient descent method (SGD). In [14], we introduce a variance reduction approach for these settings when the objective is composite and strongly convex. The

convergence rate outperforms SGD with a typically much smaller constant factor, which depends on the variance of gradient estimates only due to perturbations on a single example.

7.3.2. Catalyst Acceleration for First-order Convex Optimization: from Theory to Practice

Participants: Hongzhou Lin, Julien Mairal, Zaid Harchaoui.

In this paper [35], we introduce a generic scheme for accelerating gradient-based optimization methods in the sense of Nesterov. The approach, called Catalyst, builds upon the inexact accelerated proximal point algorithm for minimizing a convex objective function, and consists of approximately solving a sequence of well-chosen auxiliary problems, leading to faster convergence. One of the key to achieve acceleration in theory and in practice is to solve these sub-problems with appropriate accuracy by using the right stopping criterion and the right warm-start strategy. In this paper, we give practical guidelines to use Catalyst and present a comprehensive theoretical analysis of its global complexity. We show that Catalyst applies to a large class of algorithms, including gradient descent, block coordinate descent, incremental algorithms such as SAG, SAGA, SDCA, SVRG, Finito/MISO, and their proximal variants. For all of these methods, we provide acceleration and explicit support for non-strongly convex objectives. We conclude with extensive experiments showing that acceleration is useful in practice, especially for ill-conditioned problems.

7.3.3. A Generic Quasi-Newton Algorithm for Faster Gradient-Based Optimization

Participants: Hongzhou Lin, Julien Mairal, Zaid Harchaoui.

In this paper [34], we propose a generic approach to accelerate gradient-based optimization algorithms with quasi-Newton principles. The proposed scheme, called QuickeNing, can be applied to incremental first-order methods such as stochastic variance-reduced gradient (SVRG) or incremental surrogate optimization (MISO). It is also compatible with composite objectives, meaning that it has the ability to provide exactly sparse solutions when the objective involves a sparsity-inducing regularization. QuickeNing relies on limited-memory BFGS rules, making it appropriate for solving high-dimensional optimization problems. Besides, it enjoys a worst-case linear convergence rate for strongly convex problems. We present experimental results where QuickeNing gives significant improvements over competing methods for solving large-scale high-dimensional machine learning problems, see Figure 18 for example.



Figure 18. An illustration of the minimization of logistic regression. Significant improvement is observed by applying QuickeNing.

7.3.4. Catalyst Acceleration for Gradient-Based Non-Convex Optimization

Participants: Courtney Paquette, Hongzhou Lin, Dmitriy Drusvyatskiy, Julien Mairal, Zaid Harchaoui.

In this paper [36], we introduce a generic scheme to solve nonconvex optimization problems using gradientbased algorithms originally designed for minimizing convex functions. When the objective is convex, the proposed approach enjoys the same properties as the Catalyst approach of Lin et al, 2015. When the objective is nonconvex, it achieves the best known convergence rate to stationary points for first-order methods. Specifically, the proposed algorithm does not require knowledge about the convexity of the objective; yet, it obtains an overall worst-case efficiency of $O(\epsilon^{-2})$ and, if the function is convex, the complexity reduces to the near-optimal rate $O(\epsilon^{-2/3})$. We conclude the paper by showing promising experimental results obtained by applying the proposed approach to SVRG and SAGA for sparse matrix factorization and for learning neural networks (see Figure 19).



Figure 19. Comparison of different algorithms for the minimization of a two-layer neural network. Applying our method provides a clear acceleration in terms of function value.

7.4. Machine learning and pluri-disciplinary research

7.4.1. Predicting Transcription Factor Binding Sites with Convolutional Kernel Networks

Participants: Dexiong Chen, Laurent Jacob, Julien Mairal.

The growing amount of biological sequences available makes it possible to learn genotype-phenotype relationships from data with increasingly high accuracy. By exploiting large sets of sequences with known phenotypes, machine learning methods can be used to build functions that predict the phenotype of new, unannotated sequences. In particular, deep neural networks have recently obtained good performances on such prediction tasks, but are notoriously difficult to analyze or interpret. In this work, we introduce a hybrid approach between kernel methods and convolutional neural networks for sequences, which retains the ability of neural networks to learn good representations for a learning problem at hand, while defining a well characterized Hilbert space to describe prediction functions. Our method (see Figure 20), dubbed CKN-seq, outperforms state-of-the-art convolutional neural networks on a transcription factor binding prediction task while being much faster to train and yielding more stable and interpretable results.

Source code is freely available at https://gitlab.inria.fr/dchen/CKN-seq.

7.4.2. Loter: Inferring local ancestry for a wide range of species Participants: Thomas Dias-Alves, Julien Mairal, Michael Blum [CNRS].



Figure 20. Construction scheme of CKN-seq

Admixture between populations provides opportunity to study biological adaptation and phenotypic variation. Admixture studies can rely on local ancestry inference for admixed individuals, which consists of computing at each locus the number of copies that originate from ancestral source populations. Existing software packages for local ancestry inference are tuned to provide accurate results on human data and recent admixture events. Here, we introduce Loter, an open-source software package that does not require any biological parameter besides haplotype data in order to make local ancestry inference available for a wide range of species. Using simulations, we compare the performance of Loter to HAPMIX, LAMP-LD, and RFMix. HAPMIX is the only software severely impacted by imperfect haplotype reconstruction. Loter is the less impacted software by increasing admixture time when considering simulated and admixed human genotypes. LAMP-LD and RFMIX are the most accurate method when admixture took place 20 generations ago or less; Loter accuracy is comparable or better than RFMix accuracy when admixture took place of 50 or more generations; and its accuracy is the largest when admixture is more ancient than 150 generations. For simulations of admixed Populus genotypes, Loter and LAMP-LD are robust to increasing admixture times by contrast to RFMix. When comparing length of reconstructed and true ancestry tracts, Loter and LAMP-LD provide results whose accuracy is again more robust than RFMix to increasing admixture times. We apply Loter to admixed Populus individuals and lengths of ancestry tracts indicate that admixture took place around 100 generations ago.

The Loter software package and its source code are available at https://github.com/bcm-uga/Loter.

7.4.3. High Dimensional Classification with combined Adaptive Sparse PLS and Logistic Regression

Participant: Ghislain Durif.

The high dimensionality of genomic data calls for the development of specific classification methodologies, especially to prevent over-optimistic predictions. This challenge can be tackled by compression and variable selection, which combined constitute a powerful framework for classification, as well as data visualization and interpretation. However, current proposed combinations lead to unstable and non convergent methods due to inappropriate computational frameworks. We hereby propose a computationally stable and convergent approach for classification in high dimensional based on sparse Partial Least Squares (sparse PLS). In this work [6], we start by proposing a new solution for the sparse PLS problem that is based on proximal operators for the case of univariate responses. Then we develop an adaptive version of the sparse PLS for classification, called logit-SPLS, which combines iterative optimization of logistic regression and sparse PLS to ensure computational convergence and stability. Our results are confirmed on synthetic and experimental data. In particular we show how crucial convergence and stability can be when cross-validation is involved for calibration purposes. Using gene expression data we explore the prediction of breast cancer relapse (c.f. figure 21 for an example of data visualization). We also propose a multi-categorical version of our method, used

to predict cell-types based on single-cell expression data. Our approach is implemented in the plsgenomics R-package.



Figure 21. Visualization of gene expression profiles sampled from breast cancer patients in a two-dimensional subspace, using our supervised approach logit-SPLS, compared to other SPLS methods for supervised classification and PCA (unsupervised). Data are separated in two groups of individuals, presenting a relapse or not.

7.4.4. Probabilistic Count Matrix Factorization for Single Cell Expression Data Analysis Participant: Ghislain Durif.

The development of high-throughput biology technologies now allows the investigation of the genome-wide diversity of transcription in single cells. This diversity has shown two faces: the expression dynamics (gene to gene variability) can be quantified more accurately, thanks to the measurement of lowly-expressed genes. Second, the cell-to-cell variability is high, with a low proportion of cells expressing the same gene at the same time/level. Those emerging patterns appear to be very challenging from the statistical point of view, especially to represent and to provide a summarized view of single-cell expression data. PCA is one of the most powerful framework to provide a suitable representation of high dimensional datasets, by searching for latent directions catching the most variability in the data. Unfortunately, classical PCA is based on Euclidean distances and projections that work poorly in presence of over-dispersed counts that show drop-out events (zero-inflation) like single-cell expression data. In this work [32], we propose a probabilistic Count Matrix Factorization (pCMF) approach for single-cell expression data analysis, that relies on a sparse Gamma-Poisson factor model. This hierarchical model is inferred using a variational EM algorithm. We show how this probabilistic framework induces a geometry that is suitable for single-cell data visualization, and produces a compression of the data that is very powerful for clustering purposes. Our method is competed to other standard representation methods like t-SNE, and we illustrate its performance for the representation of zero-inflated over-dispersed count data (c.f. figure 22). We also illustrate our work with results on a publicly available data set, being single-cell expression profile of neural stem cells. Our work is implemented in the pCMF R-package.



Figure 22. Visualization of synthetic zero-inflated over-dispersed count data in a two-dimensional subspace, using our approach pCMF, compared to PCA, Non-negative Matrix Factorization (NMF), Zero-Inflated Factor Analysis (ZIFA) and t-SNE. Data are generated with n = 100 indivuduals and p = 1000 recorded variables and 3 groups of individuals. t-SNE is applied with a preliminary dimension reduction step based on pCMF or PCA (default behavior).

TYREX Project-Team

6. New Results

6.1. Experimental evaluation of attitude estimation algorithms for smartphones

• **Context:** Pervasive applications on smartphones increasingly rely on techniques for estimating attitude. Attitude is the orientation of the smartphone with respect to Earth's local frame.

Modern smartphones embed sensors such as accelerometer, gyroscope, and magnetometer which make it possible to leverage existing attitude estimation algorithms.

• **Contribution:** we focused on smartphone attitude estimation. We proposed the first benchmark using a motion lab with a high precision (the Inria Kinovis platform) for the purpose of comparing and evaluating filters from the literature on a common basis. This allowed us to provide the first in-depth comparative analysis in this context. In particular, we focused on typical motions of smartphones when carried by pedestrians. Furthermore, we proposed a new parallel filtering technique for limiting the impact of magnetic perturbations with any attitude estimation algorithm used in this context. We showed how our technique compares and improves over previous works. We made our benchmark available (see Benchmarks Attitude Smartphones in Software section) and payed attention to the reproducibility of results. We analyzed and discussed the obtained results and reported on lessons learned [14], [9], [24].

6.2. The SPARQLGX System for Distributed Evaluation of SPARQL Queries

In this work [19], we propose SPARQLGX: an implementation of a distributed RDF datastore based on Apache Spark. SPARQLGX is designed to leverage existing Hadoop infrastructures for evaluating SPARQL queries efficiently. SPARQLGX relies on an automated translation of SPARQL queries into optimized executable Spark code. We show that SPARQLGX makes it possible to evaluate SPARQL queries on billions of triples distributed across multiple nodes, while providing attractive performance figures. We report on experiments which show how SPARQLGX compares to state-of-the-art implementations and we show that our approach scales better than other systems in terms of supported dataset size. With its simple design, SPARQLGX represents an interesting alternative in several scenarios.

6.3. HAP: Building Pipelines with Heterogeneous Data and Hive

The increasing number of available datasets gives opportunities to build large and complex applications which aggregate results coming from several sources. These emerging use cases require new systems where the combinations of heterogeneous sources are both allowed and efficient. To tackle these challenges, we built a system [17] offering a simple high-level set of primitives – called HAP – to easily describe processing chains. These descriptions are then compiled into optimized SQL queries executed on Hive.

6.4. Multi-Criteria Experimental Classification of Distributed SPARQL Evaluators

In this work [13], we provide a new perspective on distributed sparql evaluators, based on a multi-criteria ranking obtained through extensive experiments. Specifically, we propose a set of five principal features which we use to rank evaluators. Each system exhibits a particular combination of these features. Similarly, the various requirements of practical use cases can also be decomposed in terms of these features. Our suggested set of features provides a more comprehensive description of the behavior of a distributed evaluator when compared to traditional performance metrics. We show how it helps in more accurately evaluating to which extent a given system is appropriate for a given use case. For this purpose, we systematically benchmarked a panel of 10 state-of-the-art implementations. We ranked them using this reading grid to pinpoint the advantages and limitations of current sparql evaluation systems.

6.5. SPARUB: SPARQL UPDATE Benchmark

One aim of the RDF data model, as standardized by the W3C, is to facilitate the evolution of data over time without requiring all the data consumers to be changed. To this end, one of the latest addition to the SPARQL standard query language is an update language for RDF graphs. The research on efficient and scalable SPARQL evaluation methods increasingly relies on standardized methodologies for benchmarking and comparing systems. However, current RDF benchmarks do not support graphs updates. We propose SPARUB: a benchmark for the SPARQL UPDATE language on RDF graphs [18]. The aim of SPARUB is not to be yet another RDF benchmark. Instead it provides the mean to automatically extend and improve existing RDF benchmarks along a new dimension of data updates, while preserving their structure and query scenarios.

6.6. Optimizing SPARQL query evaluation with a worst-case cardinality estimation

SPARQL is the w3c standard query language for querying data expressed in the Resource Description Framework (RDF). There exists a variety of SPARQL evaluation schemes and, in many of them, estimating the cardinality of intermediate results is key for performance, especially when the computation is distributed and the datasets very large. For example it helps in choosing join orders that minimize the size of intermediate subquery results.

In this context [23], we propose a new cardinality estimation based on statistics about the data. Our cardinality estimation is a worst-case analysis tailored for SPARQL and capable of taking advantage of the implicit schema often present in RDF datasets (e.g. functional dependencies). This implicit schema is captured by statistics therefore our method does not need for the schema to be explicit or perfect (our system performs well even if there are a few "violations" of these implicit dependencies). We implemented our cardinality estimation and used it to optimize the evaluation of SPARQL queries: equipped with our cardinality estimation, the query evaluator performs better against most queries (sometimes by an order of magnitude) and is only ever slightly slower.

6.7. Extending the SPARQL Algebra for the optimization of Property Paths

In this work [22], [21], we propose a new algebra, μ -algebra, inspired by works on the relational algebra, SQL and NoSQL languages (especially SPARQL) along with a prototype implementation of a SPARQL optimizer based on this algebra1. Our algebra has the following properties: (1) It subsumes the SPARQL Algebra (under the set semantics) with a more general recursion. (2) SPARQL with Property Paths can be e ciently translated to this algebra. (3) We have a type system and rewriting rules for terms of this algebra that allow optimization, notably of terms involving recursion. We illustrate the differences and the benefits of our approach on recursive query optimization. While a generic approach often comes at the cost of performance, we experimentally show that this approach actually leads to more efficient evaluation of queries with Property Paths. We also show that our approach produces Query Execution Plans (QEP) that are not considered by other existing approaches.

6.8. SPARQL Query Containment with ShEx Constraints

Data structured in the Resource Description Framework (RDF) are increasingly available in large volumes. This leads to a major need and research interest in novel methods for query analysis and compilation for making the most of RDF data extraction. SPARQL is the widely used and well supported standard query language for RDF data. In parallel to query language evolutions, schema languages for expressing constraints on RDF datasets also evolve. Shape Expressions (ShEx) are increasingly used to validate RDF data, and to communicate expected graph patterns. Schemas in general are important for static analysis tasks such as query optimisation and containment.

In this work [10], [8], we investigate the means and methodologies for SPARQL query static analysis in the presence of ShEx schema constraints. Our contribution consists in considering the problem of SPARQL query containment in the presence of ShEx constraints. We propose a sound and complete procedure for the problem of containment with ShEx, considering several SPARQL fragments. Particularly our procedure considers OPTIONAL query patterns, that turns out to be an important feature to be studied with schemas. We provide complexity bounds for the containment problem with respect to the language fragments considered. We also propose alternative method for SPARQL query containment with ShEx by reduction into First Order Logic satisfiability, which allows for considering SPARQL fragment extension in comparison to the first method. This is the first work addressing SPARQL query containment in the presence of ShEx constraints.

6.9. Selectivity Estimation for SPARQL Triple Patterns with Shape Expressions for Optimising SPARQL Query Evaluation

ShEx (Shape Expressions) is a language for expressing constraints on RDF graphs. In this work [11], [8], [15], we optimize the evaluation of conjunctive SPARQL queries, on RDF graphs, by taking advantage of ShEx constraints. Our optimization is based on computing and assigning ranks to query triple patterns, dictating their order of execution. The presence of intermediate joins between the query triple patterns is the reason why ordering is important in increasing efficiency. We first define a set of well formed ShEx schemas, that possess interesting characteristics for SPARQL query optimization. We then define our optimization method by exploiting information extracted from a ShEx schema. We finally report on evaluation results performed showing the advantages of applying our optimization on the top of an existing state-of-the-art query evaluation system.

6.10. A Circuit-Based Approach to Efficient Enumeration

In this work [12], we study the problem of enumerating the satisfying valuations of a circuit while bounding the delay, i.e., the time needed to compute each successive valuation. We focus on the class of structured d-DNNF circuits originally introduced in knowledge compilation, a sub-area of artificial intelligence. We propose an algorithm for these circuits that enumerates valuations with linear preprocessing and delay linear in the Hamming weight of each valuation. Moreover, valuations of constant Hamming weight can be enumerated with linear preprocessing and constant delay. Our results yield a framework for efficient enumeration that applies to all problems whose solutions can be compiled to structured d-DNNFs. In particular, we use it to recapture classical results in database theory, for factorized database representations and for MSO evaluation. This gives an independent proof of constant-delay enumeration for MSO formulae with first-order free variables on bounded-treewidth structures.

6.11. XQuery Static Type-Checking

Although XQuery is a statically typed, functional query language for XML data, some of its features such as upward and horizontal XPath axes are typed imprecisely. The main reason is that while the XQuery data model allows us to navigate upwards and between siblings from a given XML node, the type model, e.g., regular tree types, can describe only the subtree structure of the given node. Recently, Castagna et al. (2015) and Genevès and Gesbert (2015) independently propose a precise forward type inference system for XQuery using an extended type language that can describe not only a given XML node but also its context. In this work [20], as a complementary method to forward type inference systems, we propose a novel backward type inference system for XQuery, using the type language proposed by Genevès and Gesbert (2015). Our backward type inference system provides an exact typing result for XPath axes and a sound typing result for XQuery expressions.

6.12. Predicting At-Risk Patient Profiles from Big Prescription Data

In this work [16], we show how the analysis of very large amounts of drug prescription data make it possible to detect, on the day of hospital admission, patients at risk of developing complications during their hospital stay. We explore, for the first time, to which extent volume and variety of big prescription data help in constructing predictive models for the automatic detection of at-risk profiles. Our methodology is designed to validate our claims that: (1) drug prescription data on the day of admission contain rich information about the patient's situation and perspectives of evolution, and (2) the various perspectives of big medical data (such as veracity, volume, variety) help in extracting this information. We build binary classification models to identify at-risk patient profiles. We use a distributed architecture to ensure scalability of model construction with large volumes of medical records and clinical data. We report on practical experiments with real data of millions of patients and hundreds of hospitals. We demonstrate how the fine-grained analysis of such big data can improve the detection of at-risk patients, making it possible to construct more accurate predictive models that significantly benefit from volume and variety, while satisfying important criteria to be deployed in hospitals.