RESEARCH CENTER
Bordeaux - Sud-Ouest

FIELD

Activity Report 2018

Section New Results

Edition: 2019-03-07

AUCTUS Team

7. New Results

7.1. Posture and motion capture by smart textile

The objective of the work is to design a jacket made of smart textile, without the use of built-in sensors, to determine the posture of the operator.

We propose an innovative solution based on the electrical properties of a stretchable conductive tissue which is used in the manufacture of a smart garment. We use the Electrical Impedance Tomography (EIT) to reconstruct the resistance change of the conductive tissue during tissue extension/deformation caused by human movement. The conductive tissue is placed at strategic points of the jacket (e.g., elbow, shoulder). The model that describes the correlation between the operator's posture/motion and tissue deformation is difficult to obtain analytically. Neural networks are being used to associate the different postures and movements measured by the reference device with the electric field measured in the smart textile. After the learning phase, the neural network is able to predict articular angle with an accuracy of + -5 degrees from tissue extension/deformation only.

Following the successful validation on the first prototype, a request of the patent was drafted and submitted on November 6, 2018 under the number FR1860192 (Smart textile adapted for motion and/ or deformation detection). At the same time, we submitted an experiment project to COERLE. The experiments are planned for next year. This study will allow us to acquire a big database for the learning of artificial neural networks in order to try to propose a unique and stable solution of human posture capture by the smart textile, whatever the anthropometric parameters.

7.2. Appropriate design of kinematic chains

The goal of this research is to develop efficient and reliable tools based on the appropriate design framework using interval analysis that are capable of handling variations and uncertainties for the analysis and synthesis of serial kinematics chains. A primary application for this tool is to accurately model the true workspaces of the redundant human arm by imposing realistic joint constraints that may be obtained experimentally. The appropriate design framework makes it possible to model variations and uncertainties in the kinematics chains to describe families of mechanisms (e.g., sets of arms) and to understand the performance of the family. Through studying a person's usage of their available workspace on a given task, it is theorized that a task expert will make greater use of their available workspace to minimize the risk of fatigue, while a task amateur will confine themselves to a smaller region of their available workspace which will result in expedited fatigue. By understanding the range of motions of a family of task experts, collaborative robotics can be effectively incorporated to assist with the task. A C++ software library, titled the Kinematic Chain Appropriate Design Library, is being developed to efficiently model serial kinematics chains, where the main difficulty is to properly formulate the kinematic equations and incorporate additional constraints so that the problem can be quickly solved using interval analysis methods. The library will be capable of completely solving the forward and inverse kinematics problems, generating certified descriptions of various workspaces, and synthesizing appropriate design solutions.

7.3. Filtering method for human motion analysis

We have developed a series of filters to estimate the states of a dynamic system from a series of incomplete or noisy measurements for the analysis of human motion. They are also used for data fusion or for filtering noisy data from a model, especially for a Kinect and Orbbec sensor. In our case, we first developed an extended Kalman filter [13] that we improved to take into account the singularities of representations of the human kinematic module, the estimation of users' physiological parameters as well as the calibration of measurement systems. In addition, different strategies have been implemented to ensure the real-time operation of the filter, and the addition of joint constraints to improve the accuracy of the results.

In a second step, we implemented an interesting alternative technique for filtering time series. It consists of performing singular spectrum analysis. Due to the multidimensional nature of the type of data we use a specific version of this technique called Multivariate or Multidimensional Singular Spectrum Analysis (MSSA) [19].

This technique is based on a method called *decomposition into main components* which aims to compress the data both on their temporal and physical dimensions. Excellent results have been obtained.

7.4. A software architecture for the analysis of human movement and the prevention of musculoskeletal risk

Robot Operating System (ROS) is used to build the architecture of an in situ system for analyzing the movement of industrial operators. The system, presented in [5], allows us to manage data processing and modules for evaluating and recognizing a human's actions.

The ROS architecture has been chosen to guarantee a certain modularity in our system. More specifically, our objectives are to receive and merge any type of data. We want to set up an agile system that can be used in real time or in remote calculation. We also plan to use our architecture for human-robot interaction

7.5. Hamiltonian Monte Carlo with boundary reflections, and application to polytope volume calculations

In this work [7], we studied HMC with reflections on the boundary of a domain, providing an enhanced alternative to Hit-and-run (HAR) to sample a target distribution in a bounded domain. We make three contributions. First, we provide a convergence bound, paving the way to more precise mixing time analysis. Second, we present a robust implementation based on multi-precision arithmetic – a mandatory ingredient to guarantee exact predicates and robust constructions. Third, we use our HMC random walk to perform polytope volume calculations, using it as an alternative to HAR within the volume algorithm by Cousins and Vempala. The tests, conducted up to dimension 50, show that the HMC RW outperforms HAR.

This work is a collaboration with Frédéric Cazals and Augustin Chevallier from the ABS team at Inria Sophia-Antipolis. Augustin Chevallier visited our team on May 17-18, 2018. Volume calculation is a topic of interest for AUCTUS in light of the volume of configuration spaces.

7.6. Classification of cobotic systems

A new classification of cobotic systems has been proposed [1]. As there are many different ways to classify robots (robotic architecture, size, autonomy, moving ability, adaptability, etc.) and to classify human work or human roles, classifying cobotic systems (the teams formed by a robot and a human operator) is a complex problem. We proposed to focus on information exchanges and interactions among the robot, the human operator and objects of the environment. The graph describing these interactions provides interesting clues to classify cobotic systems. For example, in the surgical robotics and drone domains, the human operator is typically teleoperating (no direct contact with the environment) with constant information exchanges between him and the robot. For that reason, the graph describing these interactions called "scheme of interactions" is very specific. Further on, the description with a scheme of interactions seems particularly appropriate for cobotic systems classification. Several schemes present discriminant features that allow the qualification and naming of the cobotic systems. It is thus possible to identify the symbiotic system, with a constant information exchange and an efficient work sharing (drone), the augmented human case (work with exoskeleton), the subcontracting case, the assistance to effort case and the intelligent assistance case.

7.7. Use of Bayesian networks for situation awareness risks prediction

In all domains involving complex human systems interactions, such as the robotic domain, human errors may have dramatic impacts. These errors are often linked to situation awareness issues. We recently proposed a new method to predict situation awareness errors in training simulations [2]. It is based on Endsley's model and the 8 "situation awareness demons" that she described. The predictions are determined thanks to a Bayesian network and Noisy-Or nodes. A maturity model is introduced to come up with the initialization problem. The NASA behavioral competency model is also used to take individual differences into account.

7.8. Classification of human actions

It is important for the decomposition of human industrial activities to recognize and classify elementary gestures (a possible decomposition for measuring difficulty is described in section 8.3 or classical methods in industry such as MTM Methods Time Measurement). Due to the temporal nature of the signals, it is necessary to use a type of deep networks that manage this type of data. Recursive networks are therefore used where past observations influence the current prediction. Among recent deep network research, the so-called *long-short term memory* (LSTM) cells, represented here, seem well adapted. Unlike a simple recursive network where only data from the previous time is used for a new prediction, an LSTM cell can store data over a much longer period of time. With each prediction, the *forget gate* can decide to authorize the use or forget a previously observed data. We tested our algorithms on a classic benchmark (NTU RGB+D). In order to obtain interesting recognition rates, we showed that it was necessary to use the filters explained in section 7.3 to determinate the number of learning movements. Other less data-intensive methods are to be tested.

CAGIRE Project-Team

7. New Results

7.1. A density-based flux scheme scheme for simulating low Mach flows

Participants: Pascal Bruel, Jonathan Jung, Vincent Perrier.

The topic dealt with concerns acoustic computations in low Mach number flows with density based solvers. For ensuring a good resolution of the low Mach number base flow, a scheme able to deal with stationary low Mach number flows is necessary. Previously proposed low Mach number fixes have been tested with acoustic computations. Numerical results prove that they are not accurate for acoustic computations. The issues raised with acoustic computations with low Mach number fixes were studied and a new scheme has been developed, in order to be accurate not only for steady low Mach number flows, but also for acoustic computations. Numerical tests evidenced the improvement of the proposed scheme with respect to the state of the art [9].

7.2. A parameter free pressure based approach for simulating flows at all Mach

Participant: Pascal Bruel.

A pressure-correction algorithm developed in close partnership with Prof. E. Dick (Ghent University, Belgium) and Dr. Y. Moguen (UPPA, France) has been developed and extensively tested for a wide range of compressible fluid flow regimes. It proved to be well-suited to simulate flows at all levels of Mach number with smooth and discontinuous flow field changes, by providing a precise representation of convective transport and acoustic propagation. A co-located finite volume space discretization is used with the AUSM flux splitting. It is demonstrated that two ingredients are essential for obtaining good quality solutions: the presence of an inertia term in the transporting velocity expression; a velocity difference diffusive term in the face pressure expression, with a correct Mach number scaling to recover the hydrodynamic and acoustic low Mach number limits. To meet these two requirements, a new flux scheme, named MIAU, for Momentum Interpolation with Advection Upstream splitting has been proposed (one journal paper submitted in 2018).

7.3. New models for conjugate heat transfer

Participant: Rémi Manceau.

New models valid in the near-wall region have been proposed for both the turbulent heat flux and the dissipation rate of the temperature variance in the framework of the EDF CIFRE PhD thesis of G. Mangeon. The purpose is to extend the Elliptic Blending approaches developed in the team to all possible boundary conditions for the temperature: imposed wall-temperature, imposed heat flux or conjugate heat transfer, which is of primary importance for applications in the nuclear industry. The new full model (which associates the two above-mentioned models) is the first one to satisfy all the near-wall budgets and, consequently, the asymptotic behavior of all the quantities. These results have been presented at two international symposia [16], [15].

CARDAMOM Project-Team

7. New Results

7.1. Modelling of free surface flows

- Participants: Umberto Bosi, Mathieu Colin, Maria Kazolea, and Mario Ricchiuto
- Corresponding member: Mario Ricchiuto

This year we have continued our work on Boussinesq-type models. We have focused on the enhanced equations of Nwogu [101], and on a frequency enhanced version of the Green-Naghdi system as proposed in [55], [77]. These models allow to account for weak dispersive effects which become relevant in the near shore region. Two papers on the topic are published. The first one [9] compares two popular wave breaking closures. We perform a study of the behaviour of the two closures for different mesh sizes, with attention to the possibility of obtaining grid independent results. Based on a classical shallow water theory, we also suggest some monitors to quantify the different contributions to the dissipation mechanism, differentiating those associated to the scheme from those of the partial differential equation. Our main results show that numerical dissipation contributes very little to the the results obtained when using eddy viscosity method. This closure shows little sensitivity to the grid, and may lend itself to the development and use of non-dissipative/energy conserving numerical methods. The opposite is observed for the hybrid approach, for which numerical dissipation plays a key role, and unfortunately is sensitive to the size of the mesh. The second paper [8] presents the application and validation, with respect to the transformation, breaking and run-up of irregular waves, of an unstructured high-resolution finite volume (FV) numerical solver for the 2D extended BT equations of [101].

The extension of these techniques to also account for the presence of floating structures is ongoing [34].
The methods and models developed have also led to physical studies and applications. In particular, in collaboration with the EPOC laboratory in Bordeaux, we are conducting a parametric study of bore propagation in estuaries. In particular, Three types of bores are observed in nature long wavelength undulating, short wavelength undulating, and breaking. The first kind is invisible to the eye, but measurable. The other two can be seen during mascarets. Understanding the mechanisms ruling the transition from one to the other has tremendous impact on human activities in estuarine areas. We introduced a new set of dimensionless parameters to characterize the transition to breaking bores. We have shown that they allow to determine if the transition is dominated by friction or nonlinearity (wave amplitude). The work discussed in [7] somewhat represents an accomplishment of our activity, combining simulations using fully non-linear unstructured grid dispersive models, an exploration of parameter space based on adaptive sampling, locally enriched using a smoothness indicator related to the onset of wave breaking. Further extensions of this work are ongoing, and aim at proposing some mechanism for the first transition (see the preprint [35]).

7.2. Modelling of icing and de-icing of aircrafts

- Participants: Heloise Beaugendre, Mathieu Colin and Francois Morency
- Corresponding member: Heloise Beaugendre

Flying debris is generated in several situations: when a roof is exposed to a storm, when ice accretes on rotating wind turbines, or during inflight aircraft deicing. Four dimensionless parameters play a role in the motion of flying debris. The goal of our work was to investigate the relative importance of four dimensionless parameters: the Reynolds number, the Froude number, the Tachikawa number, and the mass moment of inertia parameters. Flying debris trajectories have been computed with a

fluid-solid interaction model formulated for an incompressible 2D laminar flow. The rigid moving solid effects are modelled in the Navier-Stokes equations using penalization. A VIC scheme was used to solve the flow equations. The aerodynamic forces and moments are used to compute the acceleration and the velocity of the solid. A database of 64 trajectories has been built using a two-level full factorial design for the four factors. The dispersion of the plate position at a given horizontal position decreases with the Froude number. Moreover, the Tachikawa number has a significant effect on the median plate position.
Ice release is of concern to aircraft manufacturers due to the potential damage that the ice debris can cause on aircraft components. This raises the need for accurate ice trajectory simulation tools to support pre-design, design and certification phases while improving cost efficiency. High-fidelity models involve fully coupled time-accurate aerodynamic and flight mechanics simulations and thus require the use of emerging simulation tools, such as approaches based on immersed boundary methods or chimera grids. The developments of current simulations tools for ice block trajectories performed in the scope of the recently completed research project STORM have been described and validated against a STORM experimental data base of trajectories created by the German Aerospace Center (DLR) in collaboration with the German-Dutch Wind Tunnel foundation.
Immersed boundary methods (IBM) are alternative methods to simulate fluid flows around complex geometries. The grid generation is fast as it does not need to conform to the fluid-solid interface. However, special treatments are needed in the flow equations to properly take into account the wall proximity. The penalization method is a particular case of the IBM in which the wall boundary conditions are imposed via continuous forcing terms into the governing equations. Reynolds Averaged NavierâStokes (RANS) equations completed with a turbulence model are still the most common way to model turbulence in engineering applications. However, RANS turbulence model implementation with penalization into a vortex formulation is not straight forward, in part because of the variable turbulent viscosity and partly because of the boundary conditions. Our work extends the penalization technique to turbulent flows. The objective is to validate the use of the SpalartâAllmaras turbulence model in the context of penalization and vortex formulation. Details of the resolution using a Vortex In Cell (VIC) numerical scheme are given. The proposed scheme is based on the advection of particles of vorticity and particles of turbulent viscosity. A Lagrangian framework is chosen to solve the advection part. The remaining parts of the system of equations are solved with an Eulerian framework using a Cartesian uniform grid. To avoid fine meshes near the wall, a wall function compatible with the penalization method and the vortex formulation is proposed. The formulation and the coding are validated against the wellknown periodic channel flow. Velocity profiles are computed without and with the wall function. Results agree with analytic law of the wall solutions, showing that RANS simulations can be conducted with VIC schemes and penalization.
Ice formation can reduce the efficiency of aircraft lifting surfaces. Experiments proved that even the onset of icing (increased roughness) could cause an increase of 63% on the minimum drag coefficient of a NACA0015 airfoil when compared to a smooth airfoil. To investigate stall behavior due to ice formation, a roughness wall extension for the Spalart-Allmaras turbulence model was implemented in the open-source code SU2 so that the onset of ice formation could also be evaluated.

7.3. High order embedded and immersed boundary methods

- Participants: Heloise Beaugendre and Mario Ricchiuto
- Corresponding member: Heloise Beaugendre

Immersed and embedded methods are a flexible and efficient approach to handle complex, and moving geometries. In these techniques, domain boundaries are not meshed exactly, but are embedded/immersed in the grid. Ad-hoc techniques are then required to account for the effect of the coupling/boundary conditions defined on the domain boundaries on the nodes in the computational domain. This year we have made some progress on these methods on two aspects.
Preliminary results on the use of higher order schemes with penalization have been obtained. As in [100], the main idea is to combine a simple penalization technique to impose the no-slip condition

with mesh adaptation to control the error. In the work done this year we have looked into the possibility of exploiting curved meshes and higher order schemes to further improve the accuracy in laminar flow computations [20], [21].

In parallel, we have worked on a different strategy to achieve higher order in the context of an embedded approach: the shifted boundary method, initially proposed in [92], [93]. We have worked on an extension of this method to hyperbolic problems, which are difficult to treat with penalization. The extension proposed for linear waves, as well as for the nonlinear shallow water equations show that second order results can be easily achieved in this context [15] (see also [27], [26]).

7.4. Adaptation techniques

- Participants: Heloise Beaugendre, Cecile Dobrzynski and Mario Ricchiuto
- Corresponding member: Cecile Dobrzynski
  Previous work on interpolation free adaptation for compressible flows [107] has been further extended to account for non-ideal gas effects. Simulations are carried out to assess grid adaptation criteria in presence of these effects [12]. We have found that for some cases, especially in presence of a strong non-ideal dependence of the speed of sound on the density and the temperature, Mach number estimators prove to be more effective.
  We have also completed the study of adaptive mesh deformation for shallow water flows, and proposed a comprehensive study of the efficiency of different techniques based on a adapt-project-evolve paradigm, or on a fully coupled ALE approach. The issue of marrying mass conservation and well-balancedness, which is peculiar to shallow water models, has been treated by means of a high order projection of the topography [3]. Preliminary results on the extension of these techniques have also been obtained [16].

7.5. Composites Materials

- Participants: Mathieu Colin, Cecile Dobrzynski and Mario Ricchiuto
- Corresponding member: Mario Ricchuito

  We have continued our study on composite materials. We have finalized the phd of Xi Lin. We have finished to write the fluid models describing the propagation of an oxyde in the cracks appearing in self-healing composite materials. More precisely, we obtained two families of models, governed by the boundaries conditions. The first one looks like Shallow Water Equations and the second one comes from Lubrication theory. In order to evaluate the relevance of our models, we investigate the well-posedness of our model. In this direction, in [58] we propose in the full generality a link between the BD entropy introduced by D. Bresch and B. Desjardins for the viscous shallow-water equations and the Bernis-Friedman (called BF) dissipative entropy introduced to study the lubrications equations. Different dissipative entropies are obtained playing with the drag terms on the viscous shallow water equations. It helps for instance to prove global existence of nonnegative weak solutions for the lubrication equations starting from the global existence of nonnegative weak solutions for appropriate viscous shallow-water equations. Two articles are in preparation.

CARMEN Project-Team

6. New Results

6.1. Exponential Adams Bashforth integrators for cardiac electrophysiology simulations

Models in cardiac electrophysiology are coupled systems of reaction-diffusion PDEs and ODEs. The ODE system displays a very stiff behavior. It is nonlinear and its upgrade at each time step has a high computational cost. A solution that we propose is to develop high-order explicit and stable integration methods. In an article published in 2018 [17] we investigated the use of exponential Adams Bashforth (EAB) integrators in cardiac electrophysiology. The paper demonstrates stability under perturbation (or 0-stability) and provides a new approach for the convergence analysis of the method. The Dahlquist stability properties of the method were also tested in a new framework that incorporates the discrepancy between the stabilizer and the system Jacobian matrix. Provided this discrepancy is small enough, the method was shown to be A(alpha)-stable. This result is interesting for an explicit time-stepping method. Numerical experiments were performed for two classes of stiff models in cardiac electrophysiology. The EAB method was observed to be as stable as implicit solvers and cheaper at the same level of accuracy.

6.2. Inverse models for identification of cellular ionic parameters from macroscopic signals

Traditional inverse models in cardiac electrophysiology have aimed at identifying the activation (and relaxation) order of the cardiac muscle from body surface potentials. However, many cardiac anomalies cannot be reduced to such simple parameters. Underlying the activation and “repolarization” of the cells is a complex interplay of different ion channels, each with its own dynamics. Genetic and other abnormalities express themselves in one or more of these channels. We have therefore taken several initiatives to identify the properties of the channels themselves.

Measurements with micro-electrodes can capture currents generated by a handful of cultured cells. Together with the REO team at the Inria center in Paris we have developed methods to identify properties of individual ionic currents from such measurements [28], [10].

On the level of the whole heart, two studies have attempted to identify cellular properties from surface electrocardiogram (ECG) signals [29], [11]. Ravon et al. have previously developed a method that identifies activation times as well as repolarization properties of a simplified ionic model (the Mitchell-Schaeffer model) [66], but considering only the outer surface of the heart muscle. We now evaluated the capability of this method to distinguish properties of the inner and outer surfaces, which are expected to differ importantly [29]. This turned out to be extremely challenging. Abidi et al. [11], on the other hand, demonstrated that the solution to such problems is unique even for intermediately complex ionic models such as the Beeler-Reuter [46] and Luo-Rudy I [61] models.

Thus, we found that a unique solution exists but is very hard to find in practical situations. This result suggests that further work should aim at removing confounding factors such as limitations in the volume conductor models.

6.3. Cellular and sub-cellular models

The electrical activation of the heart relies on rapid propagation of activation impulses through intercellular connections. Cardiac arrhythmia are often due to damage to these intercellular connections. In various pathologies, loss of individual connections can lead to the formation of a complicated maze in which very slow propagation is possible, leading to reentrant arrhythmias. To investigate such phenomena, the PhD thesis work of P.-E. Bécue was dedicated to the development of a three-dimensional model of the cardiac tissue with sub-cellular resolution [31], [50], [51], [52], [49]. This work builds on our CEPS software, which was specifically extended for this purpose.

Aouadi et al. have specifically developed a model for the connections between the network of cardiac Purkinje fibers and the working myocardium [13] [45]. It handles the myocardium with a standard bidomain model and the Purkinje network, which consists of discrete bundles, with a one-dimensional monodomain model.

On an even smaller scale, the team also worked on a three-dimensional model of calcium release from intracellular organelles and its diffusion inside the cell [42].

6.4. Highly scalable ECG simulation with electrocardiographic lead fields

Currently a monodomain reaction-diffusion model is a well-established method to simulate the electrical activity of the heart [62], [63], even more so because it can be adapted to approximate a bidomain model very closely [48], [54]. Computing the electrocardiogram (ECG) from the results of such models is harder because it requires large linear systems to be solved, and does not scale well to large numbers of processors. A possible solution is to use so-called lead fields, the electrocardiographic term for a linear combination of fundamental solutions that express the ECG potential as an integral over a field of electric current dipoles. A paper published in 2018 has shown that this method gives a huge scaling advantage on highly parallel computers [27]. This result is also of practical importance for our applied work.

CQFD Project-Team

6. New Results

6.1. A new characterization of the jump rate for piecewise-deterministic Markov processes with discrete transitions

Piecewise-deterministic Markov processes form a general class of non-diffusion stochastic models that involve both deterministic trajectories and random jumps at random times. In this paper, we state a new characterization of the jump rate of such a process with discrete transitions. We deduce from this result a nonparametric technique for estimating this feature of interest. We state the uniform convergence in probability of the estimator. The methodology is illustrated on a numerical example.

Authors: Alexandre Genadot (Inria CQFD) and Romain Azaïs.

6.2. Estimation of the average number of continuous crossings for non-stationary non-diffusion processes

Assume that you observe trajectories of a non-diffusive non-stationary process and that you are interested in the average number of times where the process crosses some threshold (in dimension d = 1) or hypersurface (in dimension d $\geq$ 2). Of course, you can actually estimate this quantity by its empirical version counting the number of observed crossings. But is there a better way? In this paper, for a wide class of piecewise smooth processes, we propose estimators of the average number of continuous crossings of an hypersurface based on Kac-Rice formulae. We revisit these formulae in the uni-and multivariate framework in order to be able to handle non-stationary processes. Our statistical method is tested on both simulated and real data.

Authors: Alexandre Genadot (Inria CQFD) and Romain Azaïs.

6.3. ClustGeo: an R package for hierarchical clustering with spatial constraints

In this paper, we propose a Ward-like hierarchical clustering algorithm including spatial/geographical constraints. Two dissimilarity matrices D_0 and D_1 are inputted, along with a mixing parameter $\alpha \in [0, 1]$. The dissimilarities can be non-Euclidean and the weights of the observations can be non-uniform. The first matrix gives the dissimilarities in the "feature space" and the second matrix gives the dissimilarities in the "constraint space". The criterion minimized at each stage is a convex combination of the homogeneity criterion calculated with D_0 and the homogeneity criterion calculated with D_1. The idea is then to determine a value of α which increases the spatial contiguity without deteriorating too much the quality of the solution based on the variables of interest i.e. those of the feature space. This procedure is illustrated on a real dataset using the R package ClustGeo.

Authors: Marie Chavent (Inria CQFD), Vanessa Kuentz-Simonet, Amaury Labenne and Jérôme Saracco (Inria CQFD).

6.4. Change-point detection for Piecewise Deterministic Markov Processes

We consider a change-point detection problem for a simple class of Piecewise Deterministic Markov Processes (PDMPs). A continuous-time PDMP is observed in discrete time and through noise, and the aim is to propose a numerical method to accurately detect both the date of the change of dynamics and the new regime after the change. To do so, we state the problem as an optimal stopping problem for a partially observed discrete-time Markov decision process taking values in a continuous state space and provide a discretization of the state space based on quantization to approximate the value function and build a tractable stopping policy. We provide error bounds for the approximation of the value function and numerical simulations to assess the performance of our candidate policy.

Authors: Alice Cleynen and Benoîte de Saporta (Inria CQFD).

6.5. A sharp first order analysis of Feynman–Kac particle models, Part I: Propagation of chaos

This article provides a new theory for the analysis of forward and backward particle approximations of Feynman–Kac models. Such formulae are found in a wide variety of applications and their numerical (particle) approximation is required due to their intractability. Under mild assumptions, we provide sharp and non-asymptotic first order expansions of these particle methods, potentially on path space and for possibly unbounded functions. These expansions allow one to consider upper and lower bound bias type estimates for a given time horizon n and particle number N; these non-asymptotic estimates are $O(n/N)$. Our approach is extended to tensor products of particle density profiles, leading to new sharp and non-asymptotic propagation of chaos estimates. The resulting upper and lower bound propagations of chaos estimates seem to be the first result of this kind for mean field particle models.

Authors: Pierre Del Moral (Inria CQFD) and Ajay Jasrab.

6.6. A sharp first order analysis of Feynman–Kac particle models, Part II: Particle Gibbs samplers

This article provides a new theory for the analysis of the particle Gibbs (PG) sampler (Andrieu et al., 2010). Following the work of Del Moral and Jasra (2017) we provide some analysis of the particle Gibbs sampler, giving first order expansions of the kernel and minorization estimates. In addition, first order propagation of chaos estimates are derived for empirical measures of the dual particle model with a frozen path, also known as the conditional sequential Monte Carlo (SMC) update of the PG sampler. Backward and forward PG samplers are discussed, including a first comparison of the contraction estimates obtained by first order estimates. We illustrate our results with an example of fixed parameter estimation arising in hidden Markov models.

Authors: Pierre Del Moral (Inria CQFD) and Ajay Jasrab.

6.7. Exponential mixing properties for time inhomogeneous diffusion processes with killing

We consider an elliptic and time-inhomogeneous diffusion process with time-periodic coefficients evolving in a bounded domain of RdRd with a smooth boundary. The process is killed when it hits the boundary of the domain (hard killing) or after an exponential time (soft killing) associated with some bounded rate function. The branching particle interpretation of the non absorbed diffusion again behaves as a set of interacting particles evolving in an absorbing medium. Between absorption times, the particles evolve independently one from each other according to the diffusion semigroup; when a particle is absorbed, another selected particle splits into two offsprings. This article is concerned with the stability properties of these non absorbed processes. Under some classical ellipticity properties on the diffusion process and some mild regularity properties of the hard obstacle boundaries, we prove an uniform exponential strong mixing property of the process conditioned to not be killed. We also provide uniform estimates w.r.t. the time horizon for the interacting particle interpretation of these non-absorbed processes, yielding what seems to be the first result of this type for this class of diffusion processes evolving in soft and hard obstacles, both in homogeneous and non-homogeneous time settings.

Authors: Pierre Del Moral (Inria CQFD) and Denis Villemonais.

6.8. Investigation of asymmetry in E. coli growth rate

The data we analyze derives from the observation of numerous cells of the bacterium Escherichia coli (E. coli) growing and dividing. Single cells grow and divide to give birth to two daughter cells, that in turn grow and divide. Thus, a colony of cells from a single ancestor is structured as a binary genealogical tree. At each node

the measured data is the growth rate of the bacterium. In this paper, we study two different data sets. One set corresponds to small complete trees, whereas the other one corresponds to long specific sub-trees. Our aim is to compare both sets. This paper is accessible to post graduate students and readers with advanced knowledge in statistics.

Authors: Bernard Delyon, Benoîte De Saporta (Inria CQFD), Nathalie Krell, Lydia Robert.

6.9. Design of estimators for restoration of images degraded by haze using genetic programming

Restoring hazy images is challenging since it must account for several physical factors that are related to the image formation process. Existing analytical methods can only provide partial solutions because they rely on assumptions that may not be valid in practice. This research presents an effective method for restoring hazy images based on genetic programming. Using basic mathematical operators several computer programs that estimate the medium transmission function of hazy scenes are automatically evolved. Afterwards, image restoration is performed using the estimated transmission function in a physics-based restoration model. The proposed estimators are optimized with respect to the mean-absolute-error. Thus, the effects of haze are effectively removed while minimizing overprocessing artifacts. The performance of the evolved GP estimators given in terms of objective metrics and a subjective visual criterion, is evaluated on synthetic and real-life hazy images. Comparisons are carried out with state-of-the-art methods, showing that the evolved estimators can outperform these methods without incurring a loss in efficiency, and in most scenarios achieving improved performance that is statistically significant.

Authors: Jose Enrique Hernandez-Beltran, Victor H.Diaz-Ramirez, Leonardo Trujillo and Pierrick Legrand (Inria CQFD).

6.10. Controlling IL-7 injections in HIV-infected patients

Immune interventions consisting in repeated injection are broadly used as they are thought to improve the quantity and the quality of the immune response. However, they also raised several questions that remains unanswered, in particular the number of injections to make or the delay to respect between different injections to acheive this goal. Practical and financial considerations add constraints to these questions, especially in the framework of human studies. We specifically focus here on the use of interleukine-7 (IL-7) injections in HIV-infected patients under antiretroviral treatment, but still unable to restore normal levels of CD4+ T lymphocytes. Clinical trials have already shown that repeated cycles of injections of IL-7 could help maintaining CD4+ T lymphocytes levels over the limit of 500 cells per microL, by affecting proliferation and survival of CD4+ T cells. We then aim at answering the question : how to maintain a patient's level of CD4+ T lymphocytes by using a minimum number of injections (ie optimizing the strategy of injections) ? Based on mechanistic models that were previously developed for the dynamics of CD4+ T lymphocytes in this context, we model the process by a piecewise deterministic Markov model. We then address the question by using some recently established theory on impulse control problem in order to develop a numerical tool determining the optimal strategy. Results are obtained on a reduced model, as a proof of concept : the method allows to defined an optimal strategy for a given patient. This method could applied to optimize injections schedules in clinical trials.

Authors: Chloé Pasin, François Dufour (Inria CQFD), Laura Villain, Huilong Zhang (Inria CQFD), Rodolphe Thiébaut.

6.11. Stochastic Control of Observer Trajectories in Passive Tracking with Acoustic Signal Propagation Optimization

The authors present in this study a numerical method which computes the optimal trajectory of a underwater vehicle subject to some mission objectives. The method is applied to a submarine whose goal is to best detect one or several targets, or/and to minimise its own detection range perceived by the other targets. The signal

considered is acoustic propagation attenuation. This approach is based on dynamic programming of a finite horizon Markov decision process. A quantisation method is applied to fully discretise the problem and allows a numerically tractable solution. Different scenarios are considered. The authors suppose at first that the position and the velocity of the targets are known and in the second they suppose that they are unknown and estimated by a Kalman type filter in a context of passive tracking.

Authors: Huilong Zhang (Inria CQFD), Benoite de Saporta (Inria CQFD), Francois Dufour (Inria CQFD), Dann Laneuville and Adrien Nègre.

6.12. Computable approximations for average Markov decision processes in continuous time

In this paper we study the numerical approximation of the optimal long-run average cost of a continuous-time Markov decision process, with Borel state and action spaces, and with bounded transition and reward rates. Our approach uses a suitable discretization of the state and action spaces to approximate the original control model. The approximation error for the optimal average reward is then bounded by a linear combination of coefficients related to the discretization of the state and action spaces, namely, the Wasserstein distance between an underlying probability measure μ and a measure with finite support, and the Hausdorff distance between the original and the discretized actions sets. When approximating μ with its empirical probability measure we obtain convergence in probability at an exponential rate. An application to a queueing system is presented.

Authors: Jonatha Anselmi (Inria CQFD), François Dufour (Inria CQFD) and Tomás Prieto-Rumeau.

6.13. Zero-Sum Discounted Reward Criterion Games for Piecewise Deterministic Markov Processes

This papers deals with the zero-sum game with a discounted reward criterion for piecewise deterministic Markov process (PDMPs) in general Borel spaces. The two players can act on the jump rate and transition measure of the process, with the decisions being taken just after a jump of the process. The goal of this paper is to derive conditions for the existence of min?max strategies for the infinite horizon total expected discounted reward function, which is composed of running and boundary parts. The basic idea is, by using the special features of the PDMPs, to re-write the problem via an embedded discrete-time Markov chain associated to the PDMP and re-formulate the problem as a discrete-stage zero sum game problem.

Authors: Oswaldo Costa and François Dufour (Inria CQFD) and Tomás Prieto-Rumeau.

6.14. Approximation of discounted minimax Markov control problems and zero-sum Markov games using Hausdorff and Wasserstein distances

This paper is concerned with a minimax control problem (also known as a robust Markov decision process (MDP) or a game against nature) with general state and action spaces under the discounted cost optimality criterion. We are interested in approximating numerically the value function and an optimal strategy of this general discounted minimax control problem. To this end, we derive structural Lipschitz continuity properties of the solution of this robust MDP by imposing suitable conditions on the model, including Lipschitz continuity of the elements of the model and absolute continuity of the Markov transition kernel with respect to some probability measure. Then, we are able to provide an approximating minimax control model with finite state and action spaces, and hence computationally tractable, by combining these structural properties with a suitable discretization procedure of the state space (related to a probabilistic criterion) and the action spaces (associated to a geometric criterion). Finally, it is shown that the corresponding approximation errors for the value function and the optimal strategy can be controlled in terms of the discretization parameters. These results are also extended to a two-player zero-sum Markov game.

Authors: François Dufour (Inria CQFD) and Tomás Prieto-Rumeau.

6.15. On the expected total cost with unbounded returns for Markov decision processes

We consider a discrete-time Markov decision process with Borel state and action spaces. The performance criterion is to maximize a total expected utility determined by unbounded return function. It is shown the existence of optimal strategies under general conditions allowing the reward function to be unbounded both from above and below and the action sets available at each step to the decision maker to be not necessarily compact. To deal with unbounded reward functions, a new characterization for the weak convergence of probability measures is derived. Our results are illustrated by examples.

Authors: François Dufour (Inria CQFD) and Alexandre Genadot (Inria CQFD).

6.16. Applying Genetic Improvement to a Genetic Programming library in C++

A young subfield of Evolutionary Computing that has gained the attention of many researchers in recent years is Genetic Improvement. It uses an automated search method that directly modifies the source code or binaries of a software system tofind improved versions based on some given criteria. Genetic Improvement has achieved notable results and the acceptance of several research communities, namely software engineering and evolutionary computation. Over the past 10 years there has been core publications on the subject, however, we have identified, to the best of our knowledge, that there is no work on applying Genetic Improvement to a meta-heuristic system. In this work we apply the GI framework called GISMO to the Beagle Puppy library version 0.1 in C++, a Genetic Programming system configured to perform symbolic regression on several benchmark and real-world problems. The objective is to improve the processing time while maintaining a similar or better test-fitness of the best individual produced by the unmodified Genetic Programming search. Results show that GISMO can generate individuals that present an improvement on those two key aspects over some problems, while also reducing the effects of bloat, one of the main issues in Genetic Programming.

Authors: Victor R. López-López, Leonardo Trujillo, Pierrick Legrand (Inria CQFD).

FLOWERS Project-Team

7. New Results

7.1. Computational Models Of Human Learning and Development

7.1.1. Computational Models Of Information-Seeking and Curiosity-Driven Learning in Humans and Animals

Participants: Pierre-Yves Oudeyer [correspondant], William Schueller, Sébastien Forestier, Alexandr Ten.

This project involves a collaboration between the Flowers team and the Cognitive Neuroscience Lab of J. Gottlieb at Columbia Univ. (NY, US), on the understanding and computational modeling of mechanisms of curiosity, attention and active intrinsically motivated exploration in humans.

It is organized around the study of the hypothesis that subjective meta-cognitive evaluation of information gain (or control gain or learning progress) could generate intrinsic reward in the brain (living or artificial), driving attention and exploration independently from material rewards, and allowing for autonomous lifelong acquisition of open repertoires of skills. The project combines expertise about attention and exploration in the brain and a strong methodological framework for conducting experimentations with monkeys, human adults and children together with computational modeling of curiosity/intrinsic motivation and learning.

Such a collaboration paves the way towards a central objective, which is now a central strategic objective of the Flowers team: designing and conducting experiments in animals and humans informed by computational/mathematical theories of information seeking, and allowing to test the predictions of these computational theories.

7.1.1.1. Context

Curiosity can be understood as a family of mechanisms that evolved to allow agents to maximize their knowledge (or their control) of the useful properties of the world - i.e., the regularities that exist in the world - using active, targeted investigations. In other words, we view curiosity as a decision process that maximizes learning/competence progress (rather than minimizing uncertainty) and assigns value ("interest") to competing tasks based on their epistemic qualities - i.e., their estimated potential allow discovery and learning about the structure of the world.

Because a curiosity-based system acts in conditions of extreme uncertainty (when the distributions of events may be entirely unknown) there is in general no optimal solution to the question of which exploratory action to take [100], [125], [135]. Therefore we hypothesize that, rather than using a single optimization process as it has been the case in most previous theoretical work [82], curiosity is comprised of a family of mechanisms that include simple heuristics related to novelty/surprise and measures of learning progress over longer time scales [123] [54], [111]. These different components are related to the subject's epistemic state (knowledge and beliefs) and may be integrated with fluctuating weights that vary according to the task context. Our aim is to quantitatively characterize this dynamic, multi-dimensional system in a computational framework based on models of intrinsically motivated exploration and learning.

Because of its reliance on epistemic currencies, curiosity is also very likely to be sensitive to individual differences in personality and cognitive functions. Humans show well-documented individual differences in curiosity and exploratory drives [98], [134], and rats show individual variation in learning styles and novelty seeking behaviors [74], but the basis of these differences is not understood. We postulate that an important component of this variation is related to differences in working memory capacity and executive control which, by affecting the encoding and retention of information, will impact the individual's assessment of learning, novelty and surprise and ultimately, the value they place on these factors [130], [146], [48], [150]. To start understanding these relationships, about which nothing is known, we will search for correlations between curiosity and measures of working memory and executive control in the population of children we test in our tasks, analyzed from the point of view of a computational models of the underlying mechanisms.

A final premise guiding our research is that essential elements of curiosity are shared by humans and non-human primates. Human beings have a superior capacity for abstract reasoning and building causal models, which is a prerequisite for sophisticated forms of curiosity such as scientific research. However, if the task is adequately simplified, essential elements of curiosity are also found in monkeys [98], [93] and, with adequate characterization, this species can become a useful model system for understanding the neurophysiological mechanisms.

7.1.1.2. Objectives

Our studies have several highly innovative aspects, both with respect to curiosity and to the traditional research field of each member team.

- Linking curiosity with quantitative theories of learning and decision making: While existing investigations examined curiosity in qualitative, descriptive terms, here we propose a novel approach that integrates quantitative behavioral and neuronal measures with computationally defined theories of learning and decision making.
- Linking curiosity in children and monkeys: While existing investigations examined curiosity in humans, here we propose a novel line of research that coordinates its study in humans and non-human primates. This will address key open questions about differences in curiosity between species, and allow access to its cellular mechanisms.
- Neurophysiology of intrinsic motivation: Whereas virtually all the animal studies of learning and decision making focus on operant tasks (where behavior is shaped by experimenter-determined primary rewards) our studies are among the very first to examine behaviors that are intrinsically motivated by the animals' own learning, beliefs or expectations.
- Neurophysiology of learning and attention: While multiple experiments have explored the single-neuron basis of visual attention in monkeys, all of these studies focused on vision and eye movement control. Our studies are the first to examine the links between attention and learning, which are recognized in psychophysical studies but have been neglected in physiological investigations.
- Computer science: biological basis for artificial exploration: While computer science has proposed and tested many algorithms that can guide intrinsically motivated exploration, our studies are the first to test the biological plausibility of these algorithms.
- Developmental psychology: linking curiosity with development: While it has long been appreciated that children learn selectively from some sources but not others, there has been no systematic investigation of the factors that engender curiosity, or how they depend on cognitive traits.

7.1.1.3. Current results: experiments in Active Categorization

In 2018, we have been occupied by analyzing data of the human adult experiment conducted in 2017. In this experiment we asked whether humans possess, and use, metacognitive abilities to guide performance-based or LP-based exploration in two contexts in which they could freely choose to learn about 4 competing tasks. Participants (n = 505, recruited via Amazon Mechanical Turk) were tested on a paradigm in which they could freely choose to engage with one of four different classification tasks. The experiment yielded a rich but complex set of data. The data includes records of participants' classification responses, task choices, reaction times, and post-task self-reports about various subjective evaluations of the competing tasks (e.g. subjective interest, progress, learning potential, etc.). We are currently analyzing the results and working on a computational models of the underlying cognitive and motivational mechanisms.

The central question going into the study was, how active learners become interested in specific learning exercises: how do they decide which task to be interested in – i.e., allocate "study time" - given that the underlying rewards or patterns are sparse and unknown? Using a family of statistical (multinomial logit), subjective-utility-based models of discrete choice behavior [109] we performed an exploratory all-subsets model selection exercise [61] to see if we can identify important and/or interesting variables that could reliably predict task choices. The initial set of variables included, among other things, various performance-based competence heuristics (e.g. current hit rate, likelihood of current hit rate). Model selection and multimodel inference pointed to a handful of variables that had *relatively* high influence on task choices (including the

likelihood of current hit rate and relative amount of time spent on a task), but their absolute effects were small, leaving most of the variation in task choices unexplained. This exercise also pointed out the potential limitations of our approach, either in operationalization of competence as a purely performance-based statistic, or in the potential lack of behavioral constraints in design of the experiment (participants may have been basing their choices on unanticipated variables). This latter limitation is tricky, since we are interested in exploratory behavior in unconstrained settings. What could have alleviated this challenge is a more diverse set of measurements that could include, for example, online records of participants' subjective feelings of interest, competence, liking, or learning potential. At this point, results concerning the LP hypothesis still have not revealed themselves, but we have gained valuable clues on how to find them. The next important step is to use cognitive models with transparent knowledge representations (e.g. Bayesian classifiers or neural networks) as an alternative way to operationalize subjective feelings of competence. The cognitive modeling approach emphasizes the idealistic assumptions made about the mind and examines their implicated behavioral outcomes. By doing that, cognitive models of learning and subjective competence can show whether our assumptions about the cognitive processes involved lead to the same behavioral patterns as the ones humans actually produce.

Although, the results of the Active Categorization study are still inconclusive, we found some interesting interim behavioral trends that are worth replicating and investigating. Participants showed preference for tasks of what we intended to be extreme complexity (i.e. too easy or too difficult) by spending more time on them (see figure 8). The group that was instructed to explore freely allocated their time more evenly, but showed a slight preference towards the easiest task where classification was based on a single dimension. The group that was instructed to try to maximize their learning during the experiment and expected a test at the end spent the majority of their time on the hardest (in fact, impossible) task to learn, where class assignment was independent of the two dimensions of variability. This suggests that active sampling strategies are subjected to extraneous constraints, and specifically, that some constraints may lead to inefficient exploration. It also potentially challenges the LP hypothesis, but it is to early to come to any strong conclusions about that, since we do not know how difficulty of the tasks was ranked subjectively by the participants.

Another puzzling observation comes from self-reported meta-cognitive judgments about the tasks. Figure 9 shows the average (min-max normalized) ratings of future learning potential and sensing the existence of a rule of each task. It is not clear why the learning potential for the hardest task (R) was reported to be high, despite the fact that it was believed to have no rule for classification. On the one hand, it is possible that while participants had not discovered the rule yet, they might have still believed there was a rule to be discovered. On the other hand, participants could really believe tha there was no rule to be discovered, but were not confident in that judgment, so high learning potential relates not to classification per se, but to discovering an interesting aspect of the task itself. There are other competing interpretations. Again, these observations compel us to better understand the contents of knowledge and knowledge-dependent processes used in the task, which we hope to achieve by applying and examining computational cognitive models of learning and meta-cognition.

7.1.2. Computational Models Of Tool Use and Speech Development: the Roles of Active Learning, Curiosity and Self-Organization

Participants: Pierre-Yves Oudeyer [correspondant], Sébastien Forestier, Rémy Portelas.

7.1.2.1. Modeling Speech and Tool Use Development in Infants

A scientific challenge in developmental and social robotics is to model how autonomous organisms can develop and learn open repertoires of skills in high-dimensional sensorimotor spaces, given limited resources of time and energy. This challenge is important both from the fundamental and application perspectives. First, recent work in robotic modeling of development has shown that it could make decisive contributions to improve our understanding of development in human children, within cognitive sciences [82]. Second, these models are key for enabling future robots to learn new skills through lifelong natural interaction with human users, for example in assistive robotics [127].

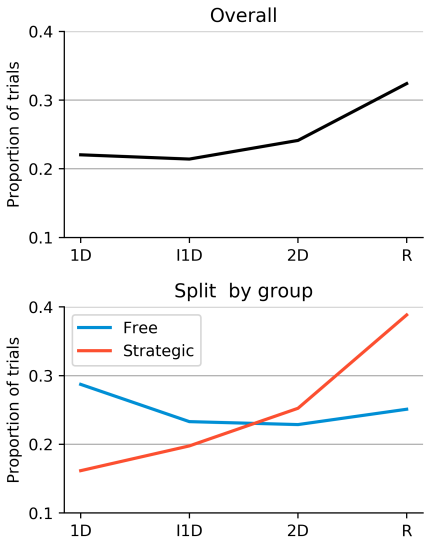


Figure 8. Proportion of trials on each task (1D, I1D, 2D, and R). 1D was the task were categorization was determined by a single variable dimension. In I1D (ignore 1D), the stimuli varied across 2 dimensions, but only one determined the stimulus category. In 2D, there were 2 variable dimensions and both jointly determined the category. Finally in R, there were 2 variable dimensions, but none of them could reliably predict the stimulus class. The top plot shows data aggregated across experimental groups, shown separately in the bottom plot.

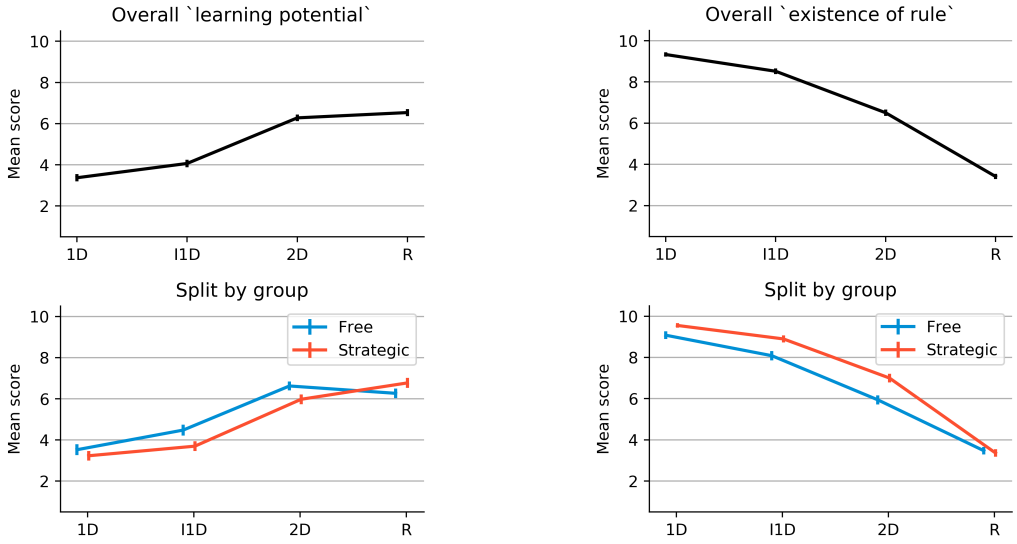


Figure 9. Average self-reported ratings of learning potential and existence of a rule for each task (1D, I1D, 2D, R; see figure 8 for disambiguation). The learning potential ratings ("Rate each monster family based on how much more you think you could learn if you had more time to play with it") were reported on a 10-point scale ([1] Definitely Could Not Learn More – [10] Definitely Could Learn More). The existence of rule ratings ("Rate each monster family based on how likely you think it had a rule for food preferences") were similarly reported ([1] Definitely No Rule – [10] Definitely a Rule. The error bars represent standard errors. The top plots show data aggregated across experimental groups, shown separately in the bottom plots.

In recent years, two strands of work have shown significant advances in the scientific community. On the one hand, algorithmic models of active learning and imitation learning combined with adequately designed properties of robotic bodies have allowed robots to learn how to control an initially unknown high-dimensional body (for example locomotion with a soft material body [53]). On the other hand, other algorithmic models have shown how several social learning mechanisms could allow robots to acquire elements of speech and language [62], allowing them to interact with humans. Yet, these two strands of models have so far mostly remained disconnected, where models of sensorimotor learning were too "low-level" to reach capabilities for language, and models of language acquisition assumed strong language specific machinery limiting their flexibility. Preliminary work has been showing that strong connections are underlying mechanisms of hierarchical sensorimotor learning, artificial curiosity, and language acquisition [128].

Recent robotic modeling work in this direction has shown how mechanisms of active curiosity-driven learning could progressively self-organize developmental stages of increasing complexity in vocal skills sharing many properties with the vocal development of infants [112]. Interestingly, these mechanisms were shown to be exactly the same as those that can allow a robot to discover other parts of its body, and how to interact with external physical objects [122].

In such current models, the vocal agents do not associate sounds to meaning, and do not link vocal production to other forms of action. In other models of language acquisition, one assumes that vocal production is mastered, and hand code the meta-knowledge that sounds should be associated to referents or actions [62]. But understanding what kind of algorithmic mechanisms can explain the smooth transition between the learning of vocal sound production and their use as tools to affect the world is still largely an open question.

The goal of this work is to elaborate and study computational models of curiosity-driven learning that allow flexible learning of skill hierarchies, in particular for learning how to use tools and how to engage in social interaction, following those presented in [122], [53], [117], [112]. The aim is to make steps towards addressing the fundamental question of how speech communication is acquired through embodied interaction, and how it is linked to tool discovery and learning.

We take two approaches to study those questions. One approach is to develop robotic models of infant development by looking at the developmental psychology literature about tool use and speech and trying to implement and test the psychologists' hypotheses about the learning mechanisms underlying infant development. Our second approach is to directly collaborate with developmental psychologists to analyze together the data of their experiments and develop other experimental setup that are well suited to answering modeling questions about the underlying exploration and learning mechanisms. We thus started last year a collaboration with Lauriane Rat-Fischer, a developmental psychologist working on the emergence of tool use in the first years of human life (now in Université Paris-Nanterre). We are currently analyzing together the behaviour of 22 month old infants in a tool use task where the infants have to retrieve a toy put in the middle of a tube by inserting sticks into the tube and pushing the toy out. We are looking at the different actions of the infant with tools and toys but also its looking behaviour, towards the tool, toys or the experimenter, and we are studying the multiple goals and exploration strategies of the babies other than the salient goal that the experimenter is pushing the baby to solve (retrieving a toy inside a tube).

In our recent robotic modeling work, we showed that the Model Babbling learning architecture allows the development of tool use in a robotic setup, through several fundamental ideas. First, goal babbling is a powerful form of exploration to produce a diversity of effects by self-generating goals in a task space. Second, the possible movements of each object define a task space in which to choose goals, and the different task spaces form an object-based representation that facilitates prediction and generalization. Also, cross-learning between tasks updates all skills while exploring one in particular. A novel insight was that early development of tool use could happen without a combinatorial action planning mechanism: modular goal babbling in itself allowed the emergence of nested tool use behaviors.

Last year we extended this architecture so that the agent can imitate caregiver's sounds in addition to exploring autonomously [78]. We hypothesized that these same algorithmic ingredients could allow a joint unified development of speech and tool use. Our learning agent is situated in a simulated environment where a vocal tract and a robotic arm are to be explored with the help of a caregiver. The environment is composed of three

toys, one stick that can be used as a tool to move toys, and a caregiver moving around. The caregiver helps in two ways. If the agent touches a toy, the caregiver produces this toy's name, but otherwise produces a distractor word as if it was talking to another adult. If the agent produces a sound close to a toy's name, the caregiver moves this toy within agent reach (see Fig. 10).

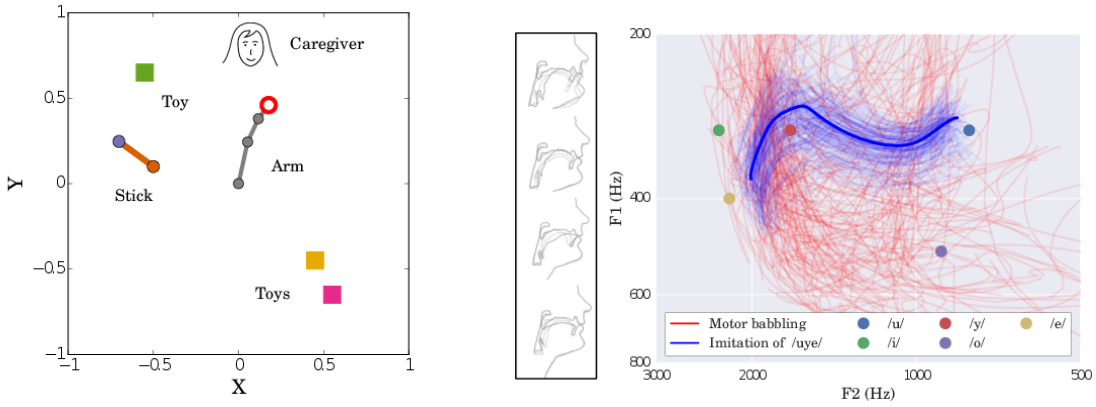


Figure 10. Agent's robotic and vocal environment. Left: Agent's 3 DOF arm, controlled with 21 parameters, grabs toys with its hand, or uses the stick to reach toys. Caregiver brings a toy within reach if the agent says its name. Right: Agent's vocal environment representing sounds as trajectories in the two first formants space. Agent's simulated vocal tract produces sounds given 28 parameters. When agent touches a toy, caregiver says toy's name. Some sounds corresponding to random parameters are plotted in red, and some sounds produced when imitating caregiver's /uye/ word in blue (best imitation in bold, error 0.3*).*

We showed that our learning architecture based on Model Babbling allows agents to learn how to 1) use the robotic arm to grab a toy or a stick, 2) use the stick as a tool to get a toy, 3) learn to produce toy names with the vocal tract, 4) use these vocal skills to get the caregiver to bring a specific toy within reach, and 5) choose the most relevant of those strategies to retrieve a toy that can be out-of-reach. Also, the grounded exploration of toys accelerates the learning of the production of accurate sounds for toy names once the caregiver is able to recognize them and react by bringing them within reach, with respect to distractor sounds without any meaning in the environment. Our model is the first to allow the study of the early development of tool use and speech in a unified framework. It predicts that infants learn to vocalize the name of toys in a natural play scenario faster than learning other words because they often choose goals related to those toys and engage caregiver's help by trying to vocalize those toys' names.

This year, we extended that model and we are currently studying on the one hand the impact of a partially contingent caregiver on agent's learning, and on the other hand the impact of attentional windows in agent's sensory perception, to see if and how an attentional window that do not match the time structure of the interaction with the caregiver could impair learning.

We also transposed this experiment to a real robotic setting during a 6-months research internship dedicated to study how IMGEP approaches scale to a real-world robotic setup. This work is related to ongoing research on simulating human babies' curiosity-driven learning mechanisms, which objectives are to test psychologists' hypotheses on human learning and to leverage these models to increase efficiency in reinforcement learning applied to robotics. Previous experiments [78] showed in simulation that intrinsically motivated reinforcement learning could be successfully applied to the early developments of speech and tool-use. The main goal of this internship was to extend this work by designing a real-world robotic experiment using a Poppy-Torso robot and a Baxter. The contributions made during this internship were 1) The design of the Poppy-Baxter

robotic playground (see figure [11] including the implementation of the communication architecture using ROS and the modeling of a 3D-printed toy, 2) Tuning of the experiment's parameters and learning process and 3) Analysis of the results in terms of exploration. Using this setup, we showed that the intrinsically motivated approach to model the early developments of speech and tool use developed in simulation can successfully scale to such a real-world experiment. Our curiosity-driven agents efficiently learned to vocalize the toy's name and to handle it in various and complex ways.

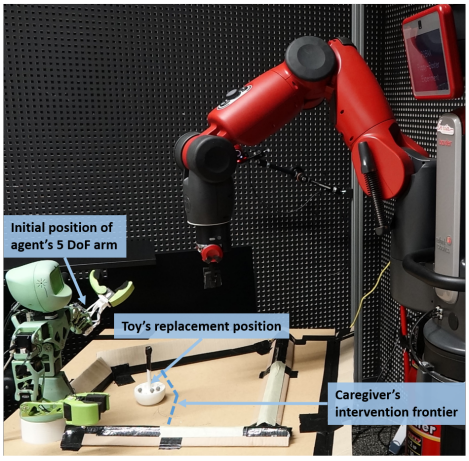


Figure 11. The POBAX Playground, in which the learning agent (Poppy-Torso) is able to interact with its arm and a simulated vocal tract. For each episode, if the agent touches the toy, the Caregiver (Baxter robot) says its name, otherwise the caregiver gives one of 3 distractor names. If the agent says the toys' name, the caregiver replaces it within the agent's arm-reach.

7.1.3. Models of Self-organization of lexical conventions: the role of Active Learning and Active Teaching in Naming Games

Participants: William Schueller [correspondant], Pierre-Yves Oudeyer.

How does language emerge, evolve and gets transmitted between individuals? What mechanisms underly the formation and evolution of linguistic conventions, and what are their dynamics? Computational linguistic studies have shown that local interactions within groups of individuals (e.g. humans or robots) can lead to self-organization of lexica associating semantic categories to words [143]. However, it still doesn't scale well to complex meaning spaces and a large number of possible word-meaning associations (or lexical conventions), suggesting high competition among those conventions.

In statistical machine learning and in developmental sciences, it has been argued that an active control of the complexity of learning situations can have a significant impact on the global dynamics of the learning process [82], [92], [101]. This approach has been mostly studied for single robotic agents learning sensorimotor affordances [123], [113]. However active learning might represent an evolutionary advantage for language formation at the population level as well [128], [145].

Naming Games are a computational framework, elaborated to simulate the self-organization of lexical conventions in the form of a multi-agent model [144]. Through repeated local interactions between random couples of agents (designated *speaker* and *hearer*), shared conventions emerge. Interactions consist of uttering a word – or an abstract signal – referring to a topic, and evaluating the success or failure of communication.

However, in existing works processes involved in these interactions are typically random choices, especially the choice of a communication topic.

The introduction of active learning algorithms in these models produces significant improvement of the convergence process towards a shared vocabulary, with the speaker [121], [140], [67] or the hearer [141] actively controlling vocabulary growth.

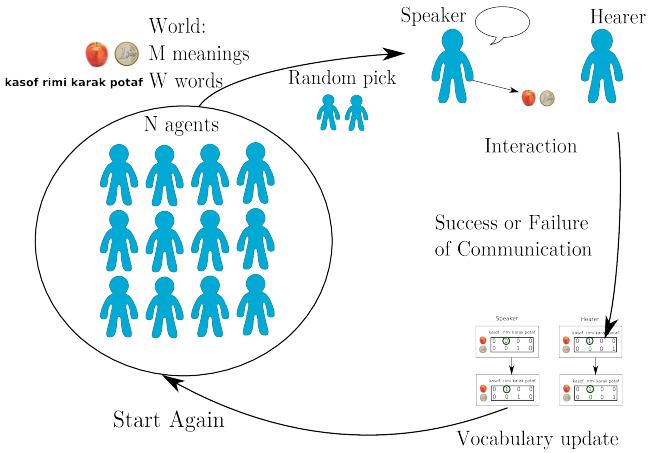


Figure 12. Illustration of the Naming Game model. Through repeated pairwise interactions, a population of individuals agrees on a shared lexicon mapping words to meanings.

7.1.3.1. Active topic choice strategies

Usually, the topic used in an interaction of the Naming Game is picked randomly. A first way of introducing active control of complexity growth is through the mechanism of topic choice: choosing it according to past memory. It allows each agent to balance reinforcement of known associations and invention of new ones, which can be seen as an exploitation vs. exploration problem. This can speed up convergence processes, and even lower significantly local and global complexity: for example in [140], [141], where heuristics based on the number of past successful interactions were used.

We defined new strategies in [31], [14] based on a maximization of an internal measure called LAPS, or Local Approximated Probability of Success. The derived strategies are called LAPSmax (exact measure but heuristical optimization algorithm) and Coherence (simplified measure but exact optimization).

Those strategies can speed up convergence the convergence process, but also diminish significantly the local complexity – i.e. the maximum number of distinct word-meaning association present in the population. See figure 13 .

7.1.3.2. Statistical lower bounds and performance measures

We showed that the time needed to converge to a shared lexicon admits a statistical lower bound [14]:

$$t_{conv} \geq N \cdot M \cdot \ln M \tag{1}$$

Where M is the number of meanings and N the population size.

Using this lower bound, we can define performance measures (between 0 and 1, best value being 1) to classify strategies and compare behavior for different values of the parameters (like population size). We distinguish in particular performance measures for convergence time, convergence speed, and maximum lexicon size. Using this, we can show that LAPSmax and Coherence yield good performance measures, which are stable with population size (cf fig. 14), and significantly better than previous strategies.

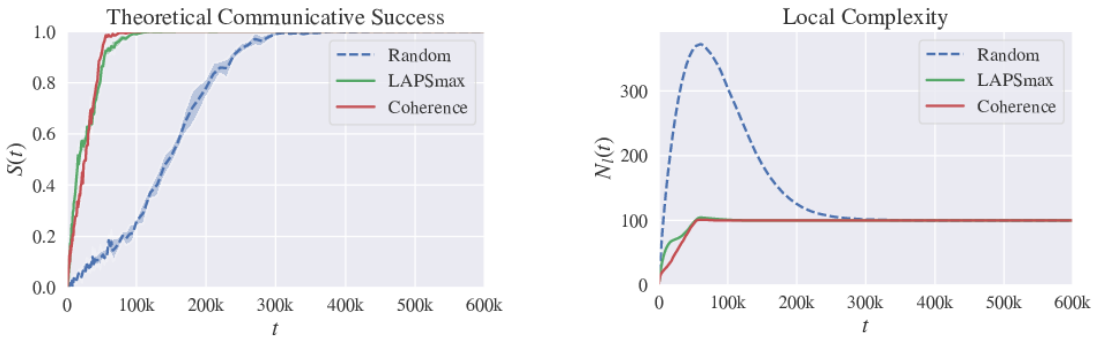


Figure 13. Evolution through time (number of interactions) of the measures of convergence (global probability of success) and global complexity (number of distinct word-meaning association present in the population) for simulations using Random Topic Choice vs. LAPSmax and Coherence Topic Choice strategies. The active topic choice strategy yields faster convergence, with less complexity. $N = 100$, $M = 100$, W *is unbounded.*

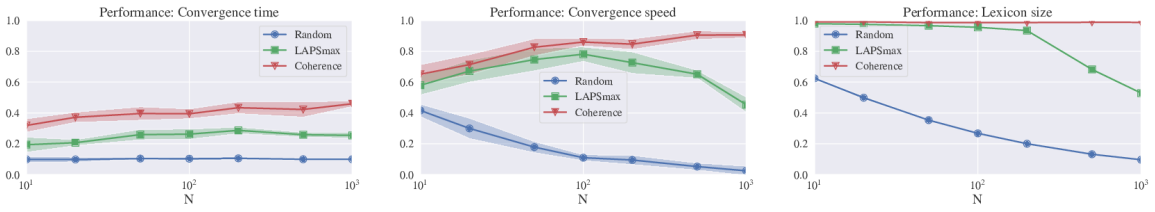


Figure 14. Performance measures for LAPSmax and Coherence strategies, compared with Random Topic Choice, for different values of N *(population size).* $M = 100$, W *is unbounded.*

7.1.3.3. Interactive application for collaborative creation of a language: Experimenting how humans actively negotiate new linguistic conventions

How do humans agree and negotiate linguistic conventions? This question is at the root of the domain of experimental semiotics [80], which is the context of our experiment/application. Typically, the experiments of this field consist in making human subjects play a game where they have to learn how to interact/collaborate through a new unknown communication medium (such as abstract symbols). In recent years, such experiments allowed to see how new conventions could be formed and evolve in population of individuals, shading light on the origins and evolution of languages [94], [79].

We consider a version of the Naming Game [152], [102], focusing on the influence of active learning/teaching mechanisms on the global dynamics. In particular, agreement is reached sooner when agents actively choose the topic of each interaction [121], [140], [141].

Through this experiment, we confront existing topic choice algorithms to actual human behavior. Participants interact through the mediation of a controlled communication system – a web application – by choosing words to refer to objects. Similar experiments have been conducted in previous work to study the agreement dynamics on a name for a single picture [63]. Here, we make several pictures or interaction topics available, and quantify the extent to which participants actively choose topics in their interactions.

Global description: Each user interacts for about 3-4 min (<30 interactions) with a brand new population of 4 simulated agents. They take the role of one designated agent, and play the Naming Game as this agent. Each time they interact as speakers, they can select the topics of conversation from a set of 5 objects, and are offered 6 possible words to refer to them. Their choices influence the global emergence of a common lexical convention, reached when communications are successful. The goal is to maximize a score based on the number of successful interactions (among the 50 in total for each run). They can see a list of the past interactions, with chosen topic, chosen word, and whether the interaction was successful or not. This experiment allows us to directly measure if there is a bias in the choice of topics, compared to random choice, based on memory of past interactions. Performance can then be compared to existing topic choice algorithms [121], [140], [141] and [31].

First version: A first version was developed for the Kreyon Conference in Rome, in September 2017. The experiment was however too close to the theoretical model, and users were not motivated to play and finish the experiment. Provided feedback was often perceived as frustrating.

Second version: A second version was developed with the help of Sandy Manolios. This second version is more entertaining, includes a motivating context, a backstory, more adapted feedback, and a more user-friendly visual interface.

Results: Users in both experiments seem to actively control their rate of invention of new conventions, by selecting more often (than random) objects that they already have a word for. See figure 16 .

7.2. Autonomous Learning in Artificial Intelligence

7.2.1. Intrinsically Motivated Goal Exploration and Multi-Task Reinforcement Learning

Participants: Sébastien Forestier, Pierre-Yves Oudeyer [correspondant], Alexandre Péré, Olivier Sigaud, Cédric Colas, Adrien Laversanne-Finot, Rémy Portelas.

7.2.1.1. Intrinsically Motivated Exploration of Spaces of Parameterized Goals and Application to Robot Tool Learning

A major challenge in robotics is to learn goal-parametrized policies to solve multi-task reinforcement learning problems in high-dimensional continuous action and effect spaces. Of particular interest is the acquisition of inverse models which map a space of sensorimotor goals to a space of motor programs that solve them. For example, this could be a robot learning which movements of the arm and hand can push or throw an object in each of several target locations, or which arm movements allow to produce which displacements of several objects potentially interacting with each other, e.g. in the case of tool use. Specifically, acquiring such repertoires of skills through incremental exploration of the environment has been argued to be a key target for life-long developmental learning [52].

Figure 15. Examples with the interface of respectively the first and the second version. Play the game here: http://naming-game.space

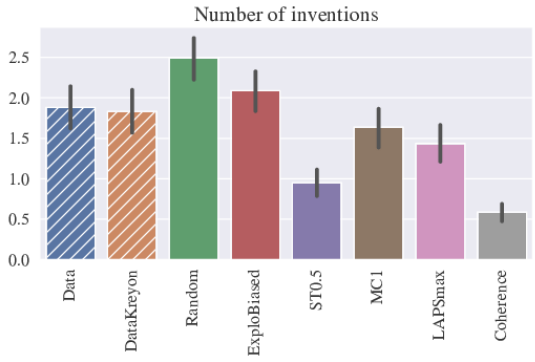


Figure 16. Average number of inventions for both experiments. DataKreyon: data collected at the Kreyon Conference in 2017 (first version); Data: data collected during summer 2018 (second version version); Other: simulated data for various strategies.

Last year we developed a formal framework called “Unsupervised Multi-Goal Reinforcement Learning”, as well as a formalization of intrinsically motivated goal exploration processes (IMGEPs), that is both more compact and more general than our previous models [76]. We experimented several implementations of these processes in a complex robotic setup with multiple objects (see Fig. 17), associated to multiple spaces of parameterized reinforcement learning problems, and where the robot can learn how to use certain objects as tools to manipulate other objects. We analyzed how curriculum learning is automated in this unsupervised multi-goal exploration process, and compared the trajectory of exploration and learning of these spaces of problems with the one generated by other mechanisms such as hand-designed learning curriculum, or exploration targeting a single space of problems, and random motor exploration. We showed that learning several spaces of diverse problems can be more efficient for learning complex skills than only trying to directly learn these complex skills. We illustrated the computational efficiency of IMGEPs as these robotic experiments use a simple memory-based low-level policy representations and search algorithm, enabling the whole system to learn online and incrementally on a Raspberry Pi 3.

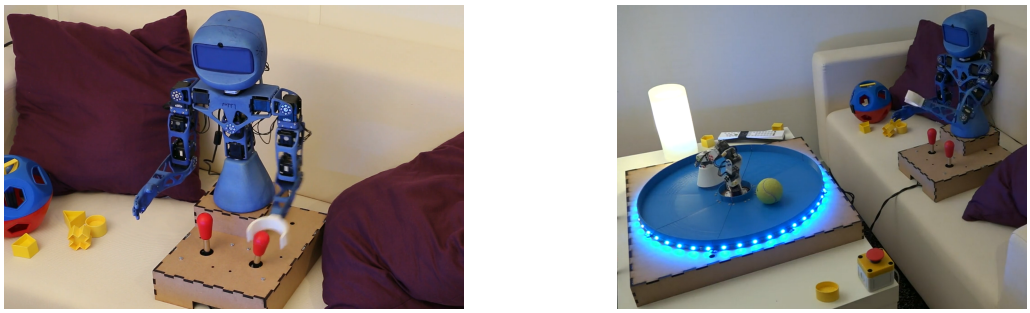


Figure 17. Robotic setup. Left: a Poppy Torso robot (the learning agent) is mounted in front of two joysticks. Right: full setup: a Poppy Ergo robot (seen as a robotic toy) is controlled by the right joystick and can hit a tennis ball in the arena which changes some lights and sounds.

In order to run many systematic scientific experiments in a shorter time, we scaled up this experimental setup to a platform of 6 identical Poppy Torso robots, each of them having the same environment to interact with. Every robot can run a different task with a specific algorithm and parameters each (see Fig. 18). Moreover, each Poppy Torso can also perceives the motion of a second Poppy Ergo robot, than can be used, this time, as a distractor performing random motions to complicate the learning problem. 12 top cameras and 6 head cameras can dump video streams during experiments, in order to record video datasets. This setup is now used to perform more experiments to compare different variants of curiosity-driven learning algorithms.

7.2.1.2. Leveraging the Malmo Minecraft platform to study IMGEP in rich simulations

In 2018 we started to leverage the Malmo platform to study curiosity-driven learning applied to multi-goal reinforcement learning tasks (https://github.com/Microsoft/malmo). The first step was to implement an environment called Malmo Mountain Cart (MMC), designed to be well suited to study multi-goal reinforcement learning (see figure [19]). We then showed that IMGEP methods could efficiently explore the MMC environment without any extrinsic rewards. We further showed that, even in the presence of distractors in the goal space, IMGEP methods still managed to discover complex behaviors such as reaching and swinging the cart, especially Active Model Babbling which ignored distractors by monitoring learning progress.

7.2.1.3. Unsupervised Deep Learning of Goal Spaces for Intrinsically Motivated Goal Exploration

Intrinsically motivated goal exploration algorithms enable machines to discover repertoires of policies that produce a diversity of effects in complex environments. These exploration algorithms have been shown to allow real world robots to acquire skills such as tool use in high-dimensional continuous state and action

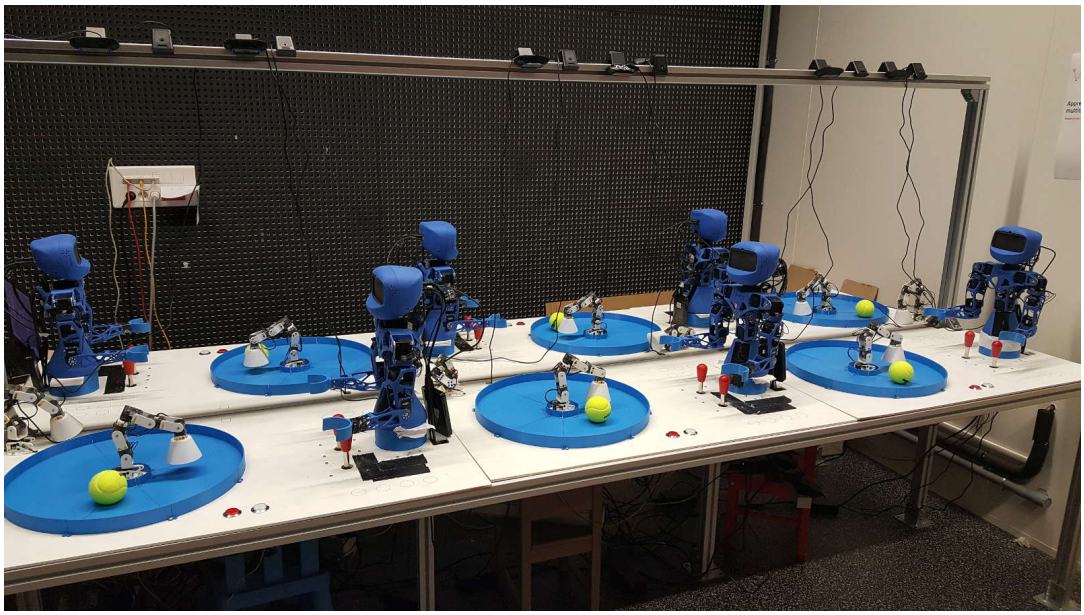


Figure 18. Platform of 6 robots with identical environment: joysticks, Poppy Ergo, ball in an arena, and a distractor. The central bar supports the 12 top cameras.

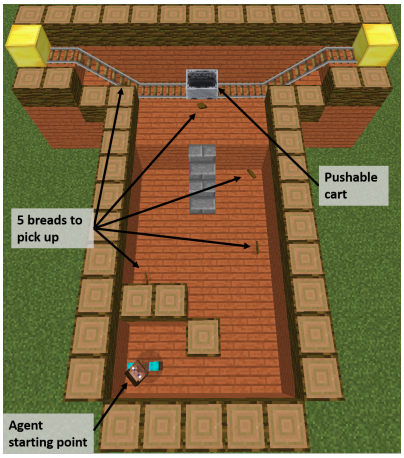


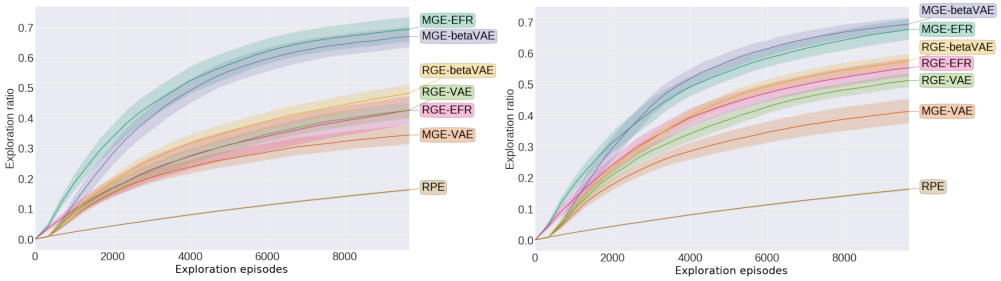
Figure 19. Malmo Mountain Cart. In this episodic environment the agent starts at the bottom left corner of the arena and is able to act on the environment using 2 continuous action commands: move and strafe. If the agent manages to get out of its starting area it may be able to collect items dispatched within the arena. If the agent manages to climb the stairs it may get close enough to the cart to move it along its railroad.

spaces. However, they have so far assumed that self-generated goals are sampled in a specifically engineered feature space, limiting their autonomy. We have proposed an approach using deep representation learning algorithms to learn an adequate goal space. This is a developmental 2-stage approach: first, in a perceptual learning stage, deep learning algorithms use passive raw sensor observations of world changes to learn a corresponding latent space; then goal exploration happens in a second stage by sampling goals in this latent space. We made experiments with a simulated robot arm interacting with an object, and we show that exploration algorithms using such learned representations can closely match, and even sometimes improve, the performance obtained using engineered representations. This work was presented at ICLR 2018 [38].

7.2.1.4. Curiosity Driven Exploration of Learned Disentangled Goal Spaces

As described in the previous paragraph, we have shown in [38] that one can use deep representation learning algorithms to learn an adequate goal space in simple environments. However, in the case of more complex environments containing multiple objects or distractors, an efficient exploration requires that the structure of the goal space reflects the one of the environment. We studied how the structure of the learned goal space using a representation learning algorithm impacts the exploration phase. In particular, we studied how disentangled representations compare to their entangled counterparts [28].

Those ideas were evaluated on a simple benchmark where a seven joints robotic arm evolves in an environment containing two balls. One of the ball can be grasped by the arm and moved around whereas the second one acts as a distractor: it cannot be grasped by the robotic arm and moves randomly across the environment.



(a) Small exploration noise ($\sigma = 0.05$) (b) Large exploration noise ($\sigma = 0.1$)

Figure 20. Exploration ratio during exploration for different exploration noises. Architectures with disentangled representations as a goal space (βVAE) explore more than those with entangled representations (VAE). Similarly modular architectures (MGE) explore more than flat architecture (RGE).

Our results showed that using a disentangled goal space leads to better exploration performances than an entangled goal space: the goal exploration algorithm discovers a wider variety of outcomes in less exploration steps (see Figure 20). We further showed that when the representation is disentangled, one can leverage it by sampling goals that maximize learning progress in a modular manner. Lastly, we have shown that the measure of learning progress, used to drive curiosity-driven exploration, can be used simultaneously to discover abstract independently controllable features of the environment.

7.2.1.5. Combining deep reinforcement learning and curiosity-driven exploration

A major challenge of autonomous robot learning is to design efficient algorithms to learn sensorimotor skills in complex and high-dimensional continuous spaces. Deep reinforcement learning (RL) algorithms are natural candidates in this context, because they can be adapted to the problem of learning continuous control policies with low sample complexity. However, these algorithms, such as DDPG [97] suffer from exploration issues in the context of sparse or deceptive reward signals.

In this project, we investigate how to integrate deep reinforcement learning algorithms with curiosity-driven exploration methods. A key idea consists in decorrelating the exploration stage from the policy learning stage by using a memory structure used in deep RL called a replay buffer. Curiosity-driven exploration algorithms, also called Goal Exploration Processes (GEPs) are used in a first stage to efficiently explore the state and action space of the problem, and the corresponding data is stored into a replay buffer. Then a DDPG learns a control policy from the content of this replay buffer.

Last year, an internship has been dedicated to trying this methodology in practice but did not reach conclusions because of instability issues related to DDPG. The project was restarted this year, which led to an ICML publication [25]. We used an open-source implementations [72], and compared the combination GEP-PG (GEP + DDPG) to the traditional DDPG [97] as well as the former state-of-the art DDPG implementation endowed with parameter-based exploration [131]. Results were presented on the popular OpenAI Gym benchmarks Continuous Mountain Car (CMC) and Half-Cheetah (HC) [58], where state-of-the-art results were demonstrated.

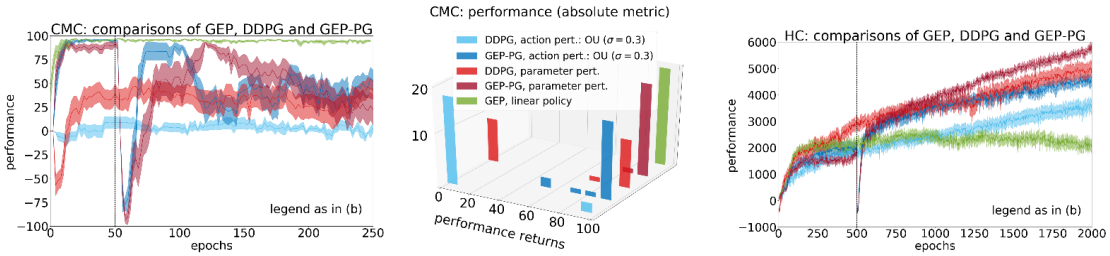


Figure 21. Left: learning curves on Continuous Mountain Car (20 seeds, mean +/- sem). Middle, best performances reached across learning on CMC. Right, learning curves on Half-Cheetah (20 seeds, mean +/- sem).

7.2.1.6. Monolithic Intrinsically Motivated Multi-Goal and Multi-Task Reinforcement Learning

In this project we merged two families of algorithms. The first family is the Intrinsically Motivated Goal Exploration Processes (IMGEP) developped in the team (see [77] for a description of the framework). In this family, autonomous learning agents sets their own goals and learn to reach them. Intrinsic motivation under the form of absolute learning progress is used to guide the selection of goals to target. In some variations of this framework, goals can be represented as coming from different *modules* or *tasks*. Intrinsic motivations are then used to guide the choice of the next task to target.

The second family encompasses goal-parameterized reinforcement learning algorithms. The first algorithm of this category used an architecture called Universal Value Function Approximators (UVFA), and enabled to train a single policy on an infinite number of goals (continuous goal spaces) [137] by appending the current goal to the input of the neural network used to approximate the value function and the policy. Using a single network allows to share weights among the different goals, which results in faster learning (shared representations). Later, HER [49] introduced a goal replay policy: the actual goal aimed at, could be replaced by a fictive goal when learning. This could be thought of as if the agent were pretending it wanted to reach a goal that it actually reached later on in the trajectory, in place of the true goal. This enables cross-goal learning and speeds up training. Finally, UNICORN [105] proposed to use UVFA to achieve multi-task learning with a discrete task-set.

In this project, we developed CURIOUS [43], an intrinsically motivated reinforcement learning algorithm able to achieve both multiple tasks and multiple goals with a single neural policy. It was tested on a custom multi-task, multi-goal environment adapted from the OpenAI Gym Fetch environments [58], see Figure 22 . CURIOUS is inspired from the second family as it proposes an extension of the UVFA architecture. Here,

the current task is encoded by a one-hot code corresponding to the task id. The goal is of size $\sum_{i=1}^{N} dim(\mathcal{G}_i)$ where $\mathcal{G}_i$ is the goal space corresponding to task T_i. All components are zeroed except the ones corresponding to the current goal g_i of the current task T_i, see Figure 23 .

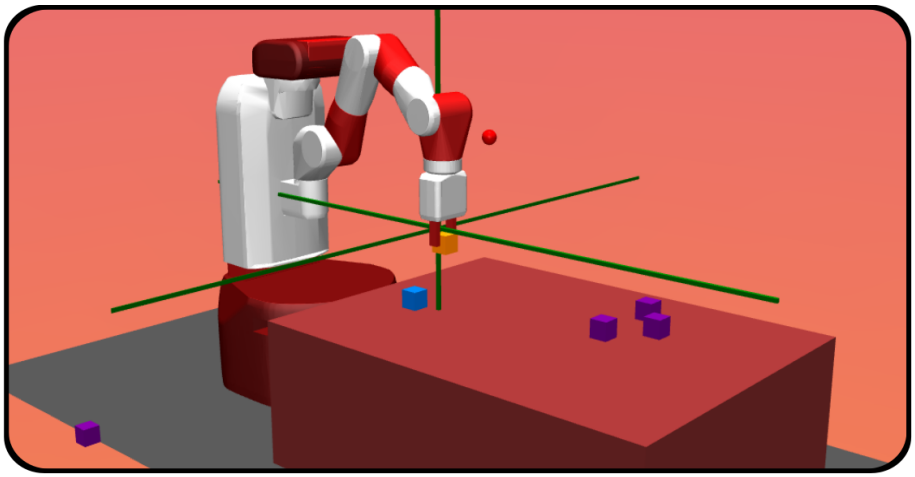


Figure 22. Custom multi-task and multi-goal environment to test the CURIOUS algorithm.

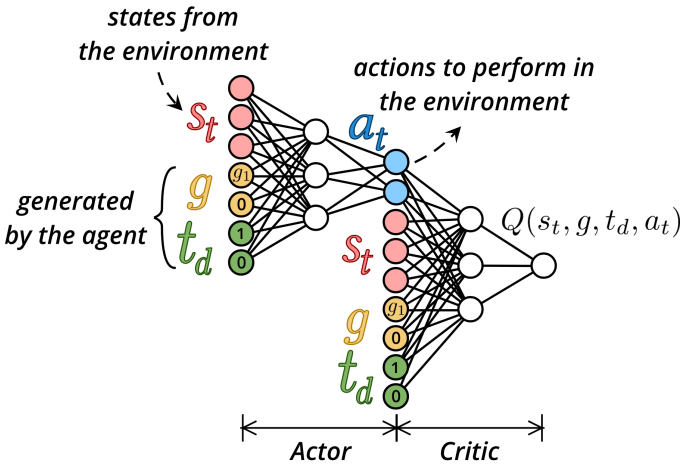


Figure 23. Architecture extended from Universal Value Function Approximators. In this example, the agent is targeting task T_1 among two tasks, each corresponding to a 1 dimension goal.

CURIOUS is also inspired from the first family, as it self-generates its own tasks and goals and uses a measure of learning progress to decide which task to target at any given moment. The learning progress is computed as the absolute value of the difference of non-overlapping window average of the successes or failures

$$LP_i(t) = \frac{|\sum_{\tau=t-2l}^{t-l} S_\tau - \sum_{\tau=t-l}^{t} S_\tau|}{2l},$$

where S_τ describes a success (1) or a failure (0) and l is a time window length. The learning progress is then used in two ways: it guides the selection of the next task to attempt, and it guides the selection of the task to replay. Cross-goal and cross-task learning are achieved by replacing the goal and/or task in the transition by another. When training on one combination of task and goal, the agent can therefore use this sample to learn about other tasks and goals. Here, we decide to replay and learn more on tasks for which the absolute learning progress is high. This helps for several reasons: 1) the agent does not focus on already learned tasks, as the corresponding learning progress is null, 2) the agent does not focus on impossible tasks for the same reason. The agent focuses more on tasks that are being learned (therefore maximizing learning progress), and on tasks that are being forgotten (therefore fighting the problem of forgetting). Indeed, when many tasks are learned in a same network, chances are tasks that are not being attempted often will be forgotten after a while.

In this project, we compare CURIOUS to two baselines: 1) a flat representation algorithm where goals are set from a multi dimensional space including all tasks (equivalent to HER); 2) a task-expert algorithm where a multi-goal UVFA expert policy is trained for each task. The results are shown in Figure 24 .

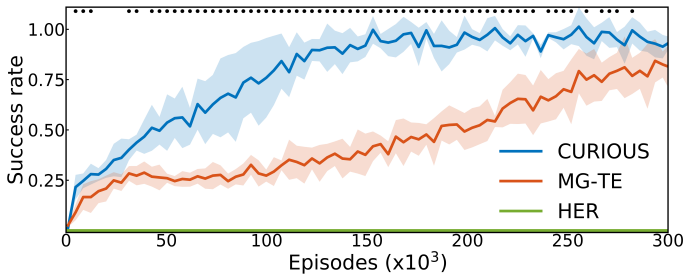


Figure 24. Comparison of CURIOUS to alternative algorithms.

7.2.2. Transfer Learning from Simulated to Real World Robotic Setups with Neural-Augmented Simulators

Participants: Florian Golemo [correspondant], Pierre-Yves Oudeyer.

This work was made in collaboration with Adrien Ali Taiga and Aaron Courville, and presented at CoRL 2018 [27]. Reinforcement learning with function approximation has demonstrated remarkable performance in recent years. Prominent examples include playing Atari games from raw pixels, learning complex policies for continuous control, or surpassing human performance on the game of Go. However most of these successes were achieved in non-physical environments (simulations, video games, etc.). Learning complex policies on physical systems remains an open challenge. Typical reinforcement learning methods require a lot of data which makes it unsuitable to learn a policy on a physical system like a robot, especially for dynamic tasks like throwing or catching a ball. One approach to this problem is to use simulation to learn control policies before applying them in the real world. This raises new problems as the discrepancies between simulation and real world environments ("reality gap") prevent policies trained in simulation from performing well when transfered to the real world. This is an instance of *domain adaption* where the input distribution of a model changes between training (in simulation) and testing (in real environment). The focus of this work is in settings where resetting the environment frequently in order to learn a policy directly in the real environment is highly impractical. In these settings the policy has to be learned entirely in simulation but is evaluated in the real environment, as *zero-shot transfer*.

In simulation there are differences in physical properties (like torques, link weights, noise, or friction) and in control of the agent, specifically joint control in robots. We propose to compensate for both of these source of issues with a generative model to bridge the gap between the source and target domain. By using data collected in the target domain through task-independent exploration we train our model to map state transitions from the source domain to state transition in the target domain. This allows us to improve the quality of our simulated robot by grounding its trajectories in realistic ones. With this learned transformation of simulated trajectories we are able to run an arbitrary RL algorithm on this augmented simulator and transfer the learned policy directly to the target task. We evaluated our approach in several OpenAI gym environments that were modified to allow for drastic torque and link length differences.

7.2.3. Curiosity-driven Learning for Automated Discovery of Physico-Chemical Structures

Participants: Chris Reinke [correspondant], Pierre-Yves Oudeyer.

Intrinsically motivated goal exploration algorithms enable machines to discover repertoires of action policies that produce a diversity of effects in complex environments. In robotics, these exploration algorithms have been shown to allow real world robots to acquire skills such as tool use in high-dimensional continuous state and action spaces (e.g. [75], [53]). In other domains such as chemistry and physics, they open the possibility to automate the discovery of novel chemical or physical structures produced by complex dynamical systems (e.g. [132]). However, they have so far assumed that self-generated goals are sampled in a specifically engineered feature space, limiting their autonomy. Recent work has shown how unsupervised deep learning approaches could be used to learn goal space representations (e.g. [38]), but they have focused on goals represented as static target configurations of the environment in robotics sensorimotor spaces. This project studies how these new families of machine learning algorithms can be extended and used for automated discovery of behaviours of dynamical systems in physics/chemistry.

The work on the project started in November 2018 and is currently in the state of exploring potential systems and algorithms.

7.2.4. Statistical Comparison of RL Algorithms.

In this project [42], we review the statistical tests recommended by [87] to compare RL algorithms (T-test, bootstrap test, Kolmogorov-Smirnov). Kolmogorov-test is discarded as it only allows to compare distributions and not mean or median performance. We wrote a tutorial in the form of an arxiv paper [42] to present the use of the Welch's t-test and the bootstrap test to compare algorithm performances. In the last section of that paper, initial assumptions of the test are described. In particular, two limiting points are discussed. First, the computation of the required sample size to satisfy requirements in type-II error (false negative) is highly sensitive to the estimations of the empirical standard deviations of the algorithms performance distributions. It is showed that for small sample sizes (< 20) the empirical standard deviation of a Gaussian distribution is biased negatively in average. Using an empirical standard deviation smaller than the true one results in under-estimations of the type-II error and therefore of the required sample size to meet requirement on that error. Second we propose empirical estimations of the type-I error (false positive) as a function of the sample size for the Welch's t-test and the bootstrap test for a particular example (Half-Cheetah environment from OpenAI Gym [58] using DDPG algorithm [97]). It is showed that the type-I error is largely underestimated by the bootstrap test for small sample size (x6 for $n = 2$, x2 for $n = 5$, x1.5 for $n = 20$). The Welch's t-test offers a satisfying estimation of the type-I error for all sample size. In conclusion, the bootstrap test should not be used. The Welch's t-test should be used with a sufficient number of seeds to obtain a reasonable estimation of the standard deviation so as to get an accurate measure of type-II error (N>20).

7.3. Representation Learning

7.3.1. State Representation Learning in the Context of Robotics

Participants: David Filliat [correspondant], Natalia Diaz Rodriguez, Timothee Lesort, Antonin Raffin, René Traoré, Ashley Hill.

During the DREAM project, we participated in the development of a conceptual framework of open-ended lifelong learning [20] based on the idea of representational re-description that can discover and adapt the states, actions and skills across unbounded sequences of tasks.

In this context, State Representation Learning (SRL) is the process of learning without explicit supervision a representation that is sufficient to support policy learning for a robot. We have finalized and published a large state-of-the-art survey analyzing the existing strategies in robotics control [23], and we have developed unsupervised methods to build representations with the objective to be minimal, sufficient, and that encode the relevant information to solve the task. More concretely, we have developed and open sourced[0] the S-RL toolbox [39] containing baseline algorithms, data generating environments, metrics and visualization tools for assessing SRL methods. The framework has been published at the NIPS workshop on Deep Reinforcement Learning 2018.

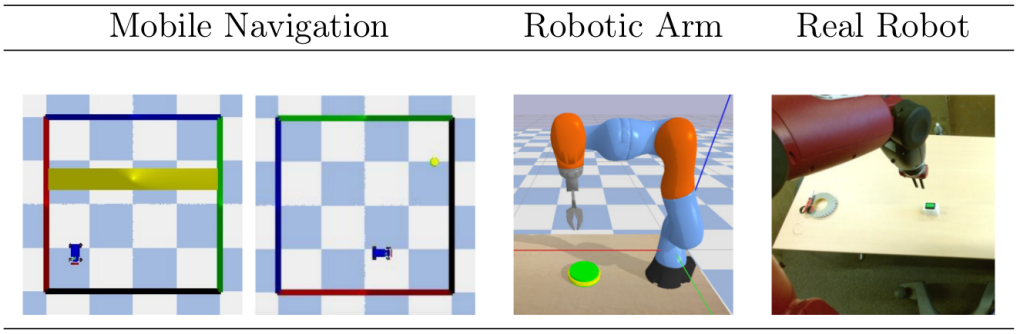


Figure 25. Environments and datasets for state representation learning.

The environments proposed in Fig. 25 are variations of two environments: a 2D environment with a mobile robot and a 3D environment with a robotic arm. In all settings, there is a controlled robot and one or more targets (that can be static, randomly initialized or moving). Each environment can either have a continuous or discrete action space, and the reward can be sparse or shaped, allowing us to cover many different situations.

The evaluation and visualization tools are presented in Fig. 26 and make it possible to qualitatively verify the learned state space behavior (e.g., the state representation of the robotic arm dataset is expected to have a continuous and correlated change with respect to the arm tip position).

We also proposed a new approach that consists in learning a state representation that is split into several parts where each optimizes a fraction of the objectives. In order to encode both target and robot positions, auto-encoders, reward and inverse model losses are used.

Because combining objectives into a single embedding is not the only option to have features that are *sufficient* to solve the tasks, by stacking representations, we favor *disentanglement* of the representation and prevent objectives that can be opposed from cancelling out. This allows a more stable optimization. Fig. 27 shows the split model where each loss is only applied to part of the state representation.

As using the learned state representations in a Reinforcement Learning setting is the most relevant approach to evaluate the SRL methods, we use the developed S-RL framework integrated algorithms (A2C, ACKTR, ACER, DQN, DDPG, PPO1, PPO2, TRPO) from Stable-Baselines [72], Augmented Random Search (ARS), Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) and Soft Actor Critic (SAC). Due to its stability, we perform extensive experiments on the proposed datasets using PPO and states learned with the approaches described in [39] along with ground truth (GT).

[0] https://github.com/araffin/robotics-rl-srl

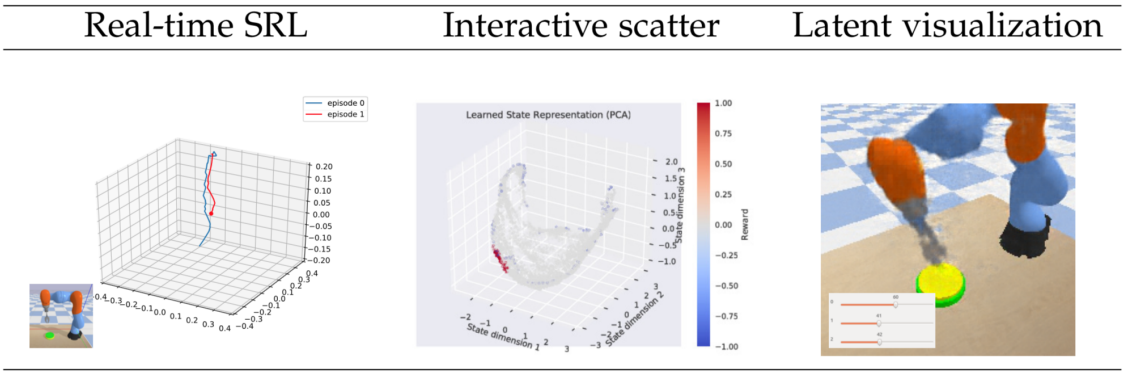


Figure 26. Visual tools for analysing SRL; Left: Live trajectory of the robot in the state space. Center: 3D scatter plot of a state space; clicking on any point displays the corresponding observation. Right: reconstruction of the point in the state space defined by the sliders.

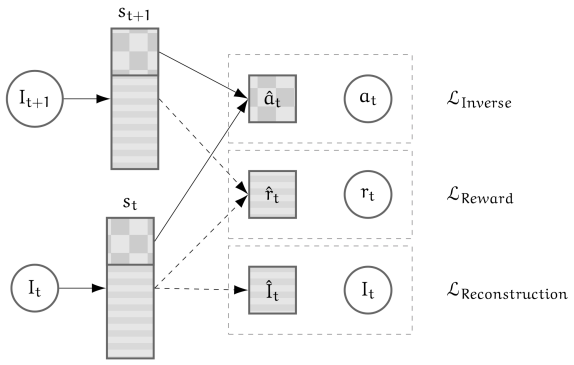


Figure 27. SRL Splits model: combines a reconstruction of an image I, a reward (r) prediction and an inverse dynamic models losses, using two splits of the state representation s. Arrows represent model learning and inference, dashed frames represent losses computation, rectangles are state representations, circles are real observed data, and squares are model predictions.

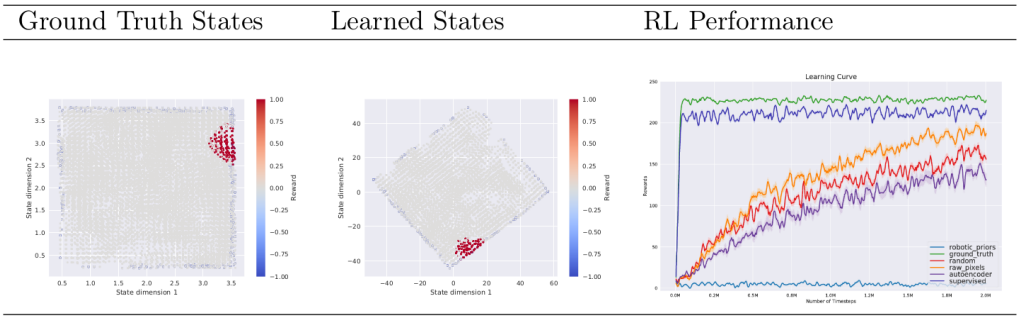


Figure 28. Ground truth states (left), states learned (Inverse and Forward) (center), and RL performance evaluation (PPO) (right) for different baselines in the mobile robot environment. Colour denotes the reward, red for positive, blue for negative and grey for null reward (left and center).

Table 28 illustrates the qualitative evaluation of a state space learned by combining forward and inverse models on the mobile robot environment. It also shows the performance of PPO algorithm based on the states learned by several baseline approaches [39].

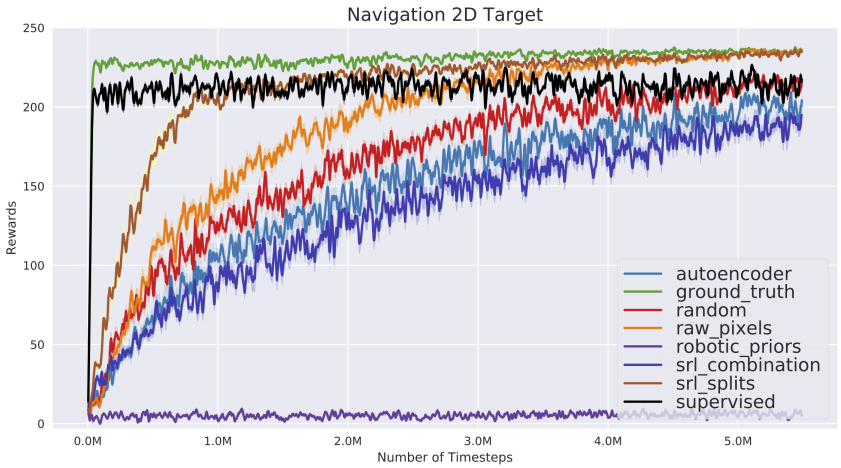


Figure 29. Performance (mean and standard error for 10 runs) for PPO algorithm for different state representations learned in Navigation 2D random target environment.

We verified that our new approach (described in Task 2.1) makes it possible for reinforcement learning to converge faster towards the optimal performance in both environments with the same amount of budget timesteps. Learning curve in Fig. 29 shows that our unsupervised state representation learned with the split model even improves on the supervised case.

7.3.2. Continual learning

Participants: David Filliat [correspondant], Natalia Díaz Rodríguez, Timothee Lesort, Hugo Caselles-Dupré.

Continual Learning (CL) algorithms learn from a stream of data/tasks continuously and adaptively through time to better enable the incremental development of ever more complex knowledge and skills. The main problem that CL aims at tackling is catastrophic forgetting [108], i.e., the well-known phenomenon of a neural network experiencing a rapid overriding of previously learned knowledge when trained sequentially on new data. This is an important objective quantified for assessing the quality of CL approaches, however, the almost exclusive focus on catastrophic forgetting by continual learning strategies, lead us to propose a set of comprehensive, implementation independent metrics accounting for factors we believe have practical implications worth considering with respect to the deployment of real AI systems that learn continually, and in "Non-static" machine learning settings. In this context we developed a framework and a set of comprehensive metrics [34] to tame the lack of consensus in evaluating CL algorithms. They measure Accuracy (A), Forward and Backward (*/remembering*) knowledge transfer (FWT, BWT, REM), Memory Size (MS) efficiency, Samples Storage Size (SSS), and Computational Efficiency (CE). Results on iCIFAR-100 classification sequential class learning is in Table 30 .

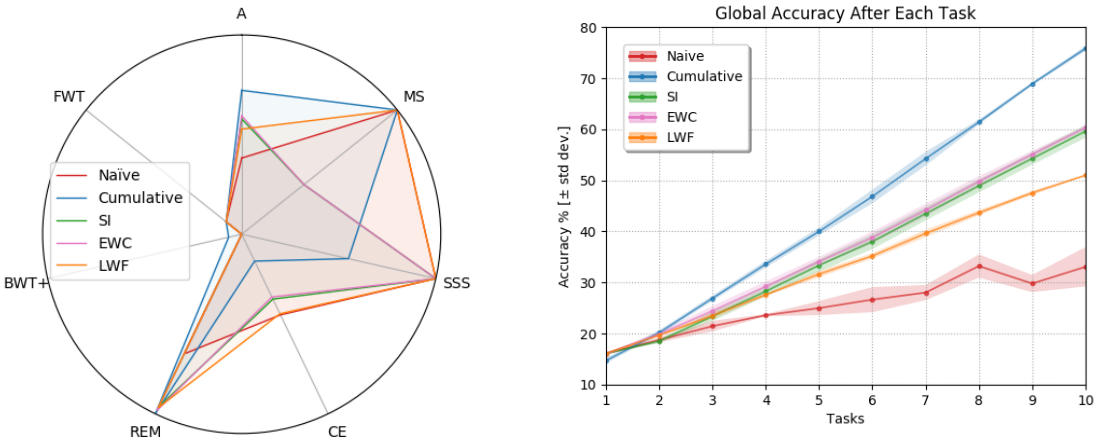


Figure 30. (left) Spider chart: CL metrics per strategy (larger area is better) and (right) Accuracy per CL strategy computed over the fixed test set.

Generative models can also be evaluated from the perspective of Continual learning.This work aims at evaluating and comparing generative models on disjoint sequential image generation tasks. We study the ability of Generative Adversarial Networks (GANS) and Variational Auto-Encoders (VAEs) and many of their variants to learn sequentially in continual learning tasks. We investigate how these models learn and forget, considering various strategies: rehearsal, regularization, generative replay and fine-tuning. We used two quantitative metrics to estimate the generation quality and memory ability. We experiment with sequential tasks on three commonly used benchmarks for Continual Learning (MNIST, Fashion MNIST and CIFAR10). We found (see Figure 32) that among all models, the original GAN performs best and among Continual Learning strategies, generative replay outperforms all other methods. Even if we found satisfactory combinations on MNIST and Fashion MNIST, training generative models sequentially on CIFAR10 is particularly instable, and remains a challenge. This work has been published at the NIPS workshop on Continual Learning 2018.

Another extension of previous section on state representation learning (SRL) to the continual learning setting is in our paper [33]. This work proposes a method to avoid catastrophic forgetting when the environment changes using generative replay, i.e., using generated samples to maintain past knowledge. State representations are learned with variational autoencoders and automatic environment change is detected through VAE

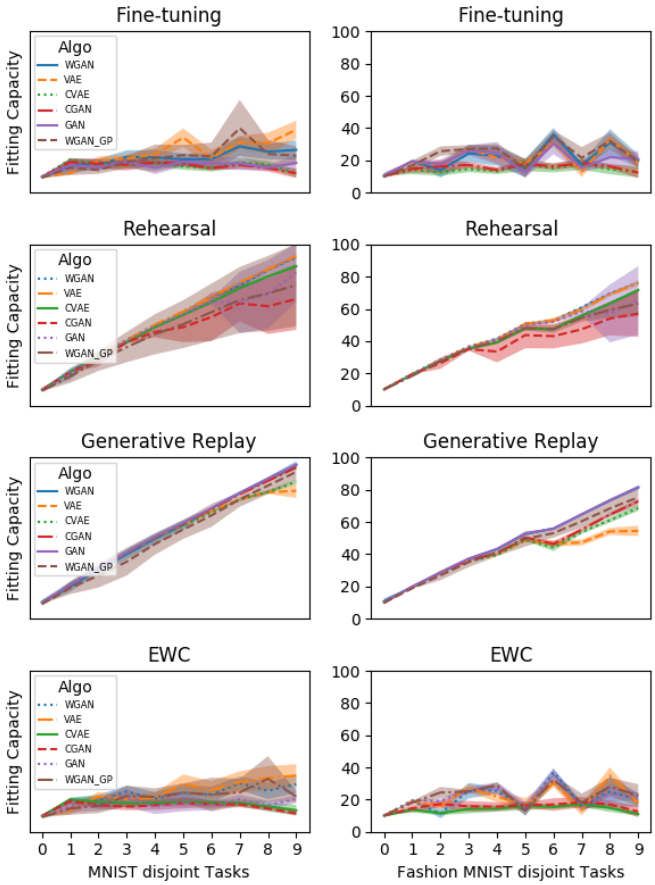


Figure 31. Means and standard deviations over 8 seeds of Fitting Capacity metric evaluation of VAE, CVAE, GAN, CGAN and WGAN. The four considered CL strategies are: Fine Tuning, Generative Replay, Rehearsal and EWC. The setting is 10 disjoint tasks on MNIST and Fashion MNIST.

reconstruction error. Results show that using a state representation model learned continually for RL experiments is beneficial in terms of sample efficiency and final performance, as seen in Figure 32 . This work has been published at the NIPS workshop on Continual Learning 2018 and is currently being extended.

The experiments were conducted in an environment built in the lab, called Flatland [32]. This is a lightweight first-person 2-D environment for Reinforcement Learning (RL), designed especially to be convenient for Continual Learning experiments. Agents perceive the world through 1D images, act with 3 discrete actions, and the goal is to learn to collect edible items with RL. This work has been published at the ICDL-Epirob workshop on Continual Unsupervised Sensorimotor Learning 2018, and was accepted as oral presentation.

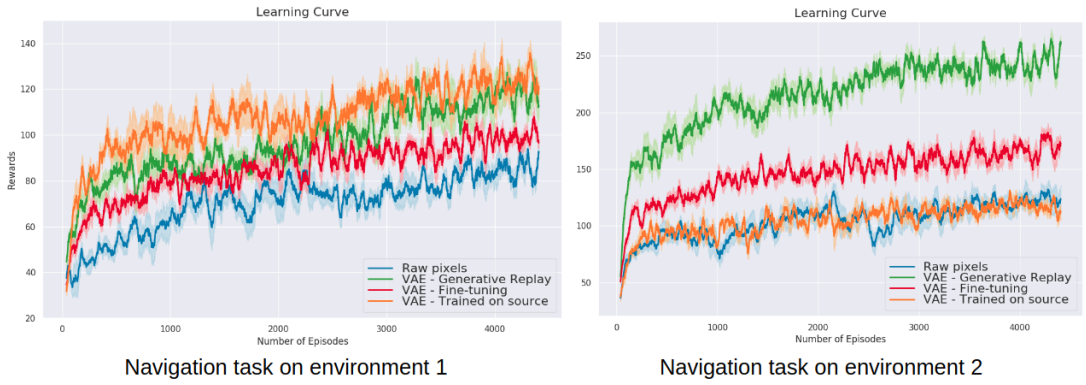


Figure 32. Mean reward and standard error over 5 runs of RL evaluation using PPO with different types of inputs. Fine-tuning and Generative Replay models are trained sequentially on the first and second environment, and then used to train a policy for both tasks. Generative Replay outperforms all other methods. It shows the need for continually learning features in State Representation Learning in settings where the environment changes.

Real life examples of applications envisioned for continual learning include learning on the edge, real time embedded systems, and applications such as the project proposal at the NeurIPS workshop on AI for Good on intelligent drone swarms for search and rescue operations at sea [36].

7.3.3. Knowledge engineering tools for neural-symbolic learning

Participant: Natalia Díaz Rodríguez [correspondant].

This section includes diverse partners distributed world wide and is result of former stablished collaborations and includes work in the context of knowledge engineering for neural-symbolic learning and reasoning systems. In [35] we presented *Datil*, a tool for learning fuzzy ontology datatypes based on clustering techniques and fuzzyDL reasoner. Ontologies for modelling healthcare data aggregation as well as knowledge graphs for modelling influence in the fashion domain are concrete ontological proposals for knowledge modelling. The former looks at wearables data interoperability for Ambient Assisted Living application development, including concepts such as height, weight, locations, activities, activity levels, activity energy expenditure, heart rate, or stress levels, among others [41]. The second proposal, considers the intrinsic subjectivity needed to effectively model subjective domains such as fashion in recommendations systems. Subjective influence networks are proposed to better quantify influence and novelty in networks. A set of use cases this approach is intended to address is discussed, as well as possible classes of prediction questions and machine learning experiments that could be executed to validate or refute the model [40].

7.4. Applications in Robotic myoelectric prostheses

Participant: Pierre-Yves Oudeyer [correspondant].

Together with the Hybrid team at INCIA, CNRS (Sébastien Mick, Daniel Cattaert, Florent Paclet, Aymar de Rugy) and Pollen Robotics (Matthieu Lapeyre, Pierre Rouanet), the Flowers team continued to work on a project related to the design and study of myoelectric robotic prosthesis. The ultimate goal of this project is to enable an amputee to produce natural movements with a robotic prosthetic arm (open-source, cheap, easily reconfigurable, and that can learn the particularities/preferences of each user). This will be achieved by 1) using the natural mapping between neural (muscle) activity and limb movements in healthy users, 2) developing a low-cost, modular robotic prosthetic arm and 3) enabling the user and the prosthesis to co-adapt to each other, using machine learning and error signals from the brain, with incremental learning algorithms inspired from the field of developmental and human-robot interaction.

7.4.1. Reachy, a 3D-printed Human-like Robotic Arm as a Test Bed for Prosthesis Control Strategies

To this day, despite the increasing motor capability of robotic prostheses, elaborating efficient control strategies is still a key challenge for their design. To provide an amputee with efficient ways to drive a prosthesis, this task requires thorough testing prior to integration into finished products. To preserve consistency with prosthetic applications, employing an actual robot for such testing requires it to show human-like features. To fulfill this need for a biomimetic test platform, we developed the Reachy robotic platform, a seven-joint human-like robotic arm that can emulate a prosthesis. Although it does not include an articulated hand and is therefore more suitable for studying reaching than manipulation, a robotic hand from available research prototypes could be integrated to Reachy. Its 3D-printed structure and off-the-shelf actuators make it inexpensive relatively to the price of a genuine prosthesis. Using an open-source architecture, its design makes it broadly connectable and customizable, so it can be integrated into many applications. To illustrate how Reachy can connect to external devices, we developed several proofs of concept where it is operated with various control strategies, such as tele- operation or vision-driven control. In this way, Reachy can help researchers to develop and test innovative control strategies on a human-like robot.

7.5. Applications in Educational Technologies

7.5.1. Machine Learning for Adaptive Personalization in Intelligent Tutoring Systems

Participants: Benjamin Clement, Alexandra Delmas, Pierre-Yves Oudeyer [correspondant], Didier Roy, Helene Sauzeon.

7.5.1.1. The Kidlearn project

Kidlearn is a research project studying how machine learning can be applied to intelligent tutoring systems. It aims at developing methodologies and software which adaptively personalize sequences of learning activities to the particularities of each individual student. Our systems aim at proposing to the student the right activity at the right time, maximizing concurrently his learning progress and its motivation. In addition to contributing to the efficiency of learning and motivation, the approach is also made to reduce the time needed to design ITS systems.

We continued to develop an approach to Intelligent Tutoring Systems which adaptively personalizes sequences of learning activities to maximize skills acquired by students, taking into account the limited time and motivational resources. At a given point in time, the system proposes to the students the activity which makes them progress faster. We introduced two algorithms that rely on the empirical estimation of the learning progress, **RiARiT** that uses information about the difficulty of each exercise and **ZPDES** that uses much less knowledge about the problem.

The system is based on the combination of three approaches. First, it leverages recent models of intrinsically motivated learning by transposing them to active teaching, relying on empirical estimation of learning progress provided by specific activities to particular students. Second, it uses state-of-the-art Multi-Arm Bandit (MAB) techniques to efficiently manage the exploration/exploitation challenge of this optimization process. Third, it leverages expert knowledge to constrain and bootstrap initial exploration of the MAB, while requiring only coarse guidance information of the expert and allowing the system to deal with didactic gaps in its

knowledge. The system was evaluated in several large-scale experiments relying on a scenario where 7-8 year old schoolchildren learn how to decompose numbers while manipulating money [65]. Systematic experiments were also presented with simulated students.

7.5.1.2. Kidlearn Experiments in 2018: Evaluating the impact of ZPDES and choice on learning efficiency and motivation

An experiment was held between mars 2018 and July 2018 in order to test the Kidlearn framework in classrooms in Bordeaux Metropole. 600 students from Bordeaux Metropole participated in the experiment. This study had several goals. The first goal was to evaluate the impact of the Kidlearn framework on motivation and learning compared to an Expert Sequence without machine learning. The second goal was to observe the impact of using learning progress to select exercise types within the ZPDES algorithm compared to a random policy. The third goal was to observe the impact of combining ZPDES with the ability to let children make different kinds of choices during the use of the ITS. The last goal was to use the psychological and contextual data measures to see if correlation can be observed between the students psychological state evolution, their profile, their motivation and their learning. The different observations showed that generally, algorithms based on ZPDES provided a better learning experience than an expert sequence. In particular, they provide a better motivating and enriching experience to self-determined students. The details of these new results, as well as the overall results of this project, were presented during the PhD defense of Benjamin Clement in decembre 2018.

7.5.1.3. Fostering Health Education With a Serious Game in Children With Asthma: Pilot Studies for Assessing Learning Efficacy and Automatized Learning Personalization

Coupled with Health Education programs, an e-learning platform—KidBreath—was participatory designed [19] and assessed in situ (Study 1) and was augmented and tested with an Intelligent Tutoring System (ITS) based on Multi-Armed Bandit Methods (Study 2). For each study, the impact of KidBreath practice was assessed in children with asthma in terms of pedagogical efficacy (knowledge of the illness), pedagogical efficiency (usability, type of motivation and level of interest elicited), and therapeutic effect (illness perception, system's expectation and judgement in disease self-management, child's implication in study). For the Study 1, asthma children aged 8 to 11 years used the tool at home without time pressure for 2 months according to a predefined learning sequence defined by the research team. Results supported pedagogical efficacy of KidBreath, with a significant increase of general knowledge about asthma after use. It also featured a greater learning gain for children knowing the least about the illness before use. Results on pedagogical efficiency revealed a great intrinsic motivation elicited by KidBreath showing a deep level of interest in the edutainment activities. Study 2 explored an augmented version of KidBreath with learning optimization algorithm (called ZPDES) after its use during 1 month. Pedagogical efficacy was less conclusive than Study 1 because less content was displayed due to algorithm parameters. However, the ITS-augmented KidBreath use showed a strong impact in pedagogical efficiency and therapeutic adherence features. Even if implementation improvements must be done in future works, this preliminary study highlighted the viability of our methods to design an ITS as serious game in health education context for all chronic diseases.

- Journée EdTech, Talence, mai 2018

7.5.2. Poppy Education: Designing and Evaluating Educational Robotics Kits

Participants: Pierre-Yves Oudeyer [correspondant], Didier Roy, Thibault Desprez, Théo Segonds, Stéphanie Noirpoudre.

The Poppy Education project aims to create, evaluate and disseminate all-inclusive pedagogical kits, open-source and low cost, for teaching computer science and robotics in secondary education and higher education, scientific literacy centers and Fablabs.

It is designed to help young people to take ownership with concepts and technologies of the digital world, and provide the tools they need to allow them to become actors of this world, with a considerable socio-economic potential. It is carried out in collaboration with teachers and several official french structures (French National Education, High schools, engineering schools, ...).

Poppy Education is based on the robotic platform poppy (open-source platform for the creation, use and sharing of interactive 3D printed robots), including:

- web interface connection (see figure 33)

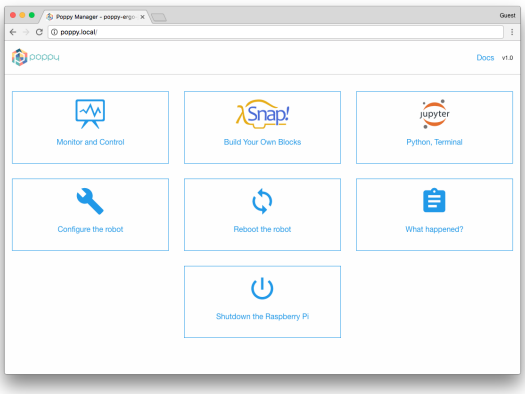


Figure 33. Home page on http://poppy.local

- Poppy Humanoid, a robust and complete robotics platform designed for genuine experiments in the real world and that can be adapted to specific user needs.
- Poppy Torso, a variant of Poppy Humanoid that can be easily installed on any flat support.
- Ergo Jr, a robotic arm. Durable and inexpensive, it is perfect to be used in class. It can be programmed in Python, directly from a web browser, using Ipython notebooks (an interactive terminal, in a web interface for the Python Programming Language).
- Snap. The visual programming system Snap (see figure 34), which is a variant of Scratch. Its features allow a thorough introduction of information technology. Several specific "blocks" have been developed for this.

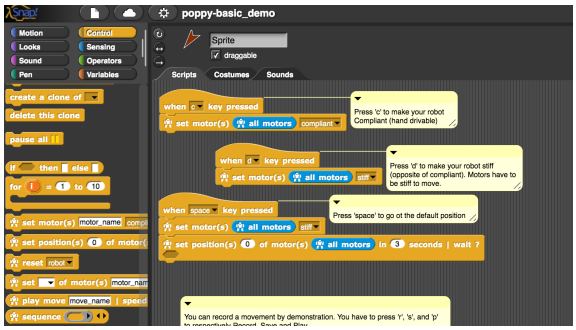


Figure 34. The visual programming system Snap

- C++, Java, Matlab, Ruby, Javascript, etc. thanks to a REST API that allows you to send commands and receive information from the robot with simple HTTP requests.

- Virtual robots (Poppy Humanoid, Torso and Ergo) can be simulated with the free simulator V-REP (see figure 35). It is possible in the classroom to work on the simulated model and then allow students to run their program on the physical robot.
- Virtual robots (Poppy Ergo) can also be simulated with a 3D web viewer (see figure 36).

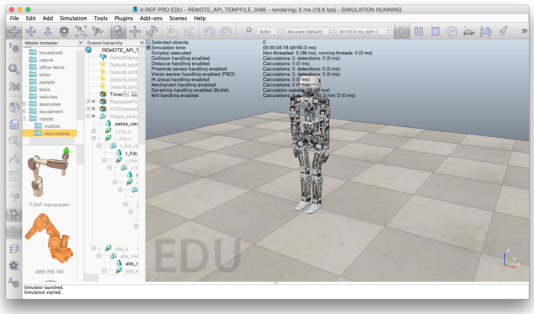


Figure 35. V-rep

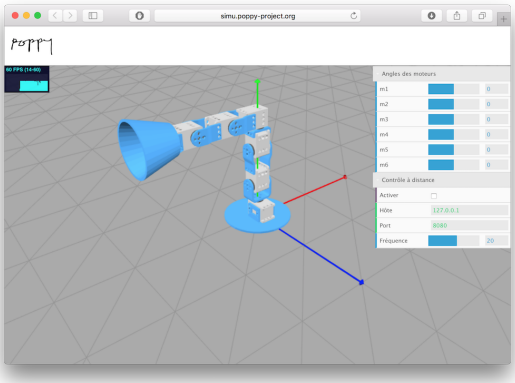


Figure 36. 3D viewer

7.5.2.1. Pedagogical experimentations : Design and experiment robots and the pedagogical activities in classroom.

The robots are designed with the final users in mind. The pedagogical tools of the project (robots and resources) are being created directly with the users and evaluated in real life by experiments. So teachers and researchers co-create activities, test them with students in class-room, share their experience and develop the platform as needed [120].

The activities were designed mainly with Snap! and Python. Most activities use Poppy Ergo Jr, but some use Poppy Torso (mostly in higher school due to its cost).

The pedagogical experiments in classroom carried out during the first year of the project notably allowed to create and experiment many robotic activities. These activities are designed as pedagogical resources introducing robotics. The main objective of the second year was to make all the activities and resources reusable (with description, documentation and illustration) easily and accessible while continuing the experiments and the diffusion of the robotic kits.

Figure 37. Experiment robots and pedagogical activities in classroom

- Pedagogical working group : the teacher partners continued to use the robots in the classroom and to create and test new classroom activities. We organized some training to help them to discover and learn how to use the robotics platform. Also, an engineer of the Poppy Education team went to visit the teachers in their school to see and to evaluate the pedagogical tools (robots and activities) in a real context of use.
  Five meetings have been organized during the year including all teachers part of the project as well as the Poppy Education team in order to exchange about their experience using the robots as a pedagogical tool, to understand their need and to get some feedback from them. This is helping us to understand better the educational needs, to create and improve the pedagogical tools.
  You can see the videos of pedagogical robotics activities here:
  https://www.youtube.com/playlist?list=PLdX8RO6QsgB7hM_7SQNLvyp2QjDAkkzLn

7.5.2.2. Pedagogical documents and resources

- We continued to improve the documentation of the robotic platform Poppy (https://docs.poppy-project.org/en/) and the documentation has been translated into French (https://docs.poppy-project.org/fr/.
  We configured a professional platform to manage the translation of the documentation (https://crowdin.com/project/poppy-docs. This platforms allows anybody to participate in the translation of the documentation to the language of their choice.
- To complete the pedagogical booklet [119] that provides guided activities and small challenges to become familiar with Poppy Ergo Jr robot and the Programming language Snap! (https://hal.inria.fr/hal-01384649/document) we provided a list of Education projects. Educational projects have been written for each activity carried out and tested in class. Each project has its own web page including resources allowing any teacher to carry out the activity (description, pedagogical sheet, photos / videos, pupil's sheet, teacher's sheet with correction etc.).
  The activities are available here:

https://www.poppy-education.org/activites/activites-lycee
The pedagogical activities are also available on the Poppy project forum where everyone is invited to comment and create new ones:
https://forum.poppy-project.org/t/liste-dactivites-pedagogiques-avec-les-robots-poppy/2305

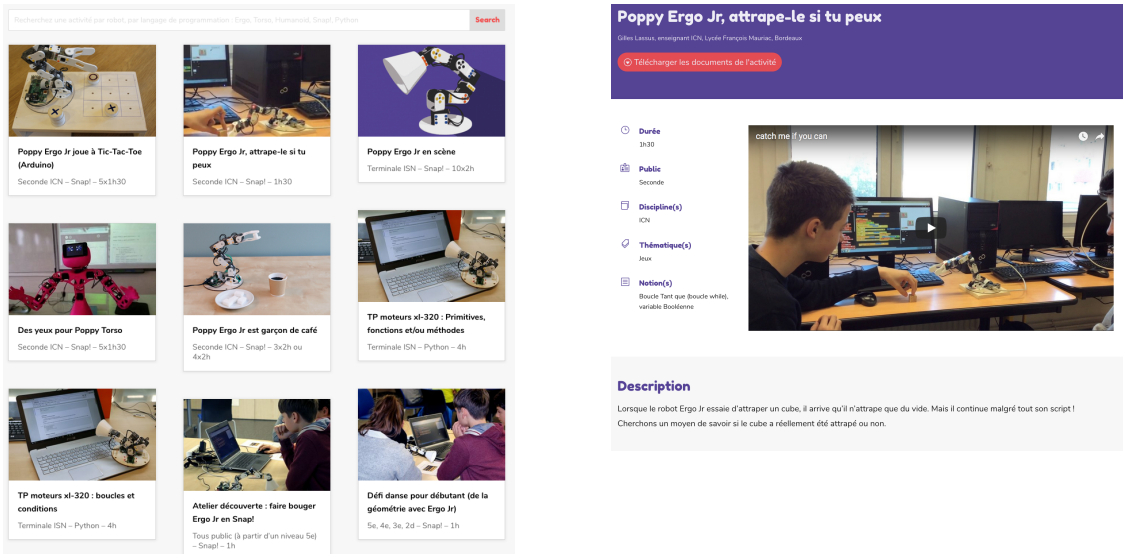


Figure 38. Open-source educational activities with Poppy robots are available on Poppy-Education.org

- A FAQ have been written with the most frequents questions to help the users: https://www.poppy-education.org/aide/
- A website has been created to present the project and to share all resources and activities. https://www.poppy-education.org/

7.5.2.3. Evaluation of the pedagogical kits

The impact of educational tools created in the lab and experimented in class had to be evaluated qualitatively and quantitatively. First, the usability, efficiency and user satisfaction must be evaluated. We must therefore assess, at first, if these tools offer good usability (i.e. effectiveness, efficiency, satisfaction). Then, in a second step, select items that can be influenced by the use of these tools. For example, students' representations of robotics, their motivation to perform this type of activity, or the evolution of their skills in these areas. In 2017 we conducted experiments to evaluate the usability of kits. We also collected data on students' perceptions of robotics.

- Population

  Our sample is made up of 28 teachers and 146 students from the region Nouvelle Aquitaine. Each subject completed an online survey in June 2017. Here, we study several groups of individuals: teachers and students. Among the students we are interested in those who practiced classroom activities with the Ergo Jr kit during the school year 2016 - 2017 (N = 68) (age = 16, std = 2.44). Among these students, 37 where High School students following the "Computer Science and Digital Sciences" stream (BAC S option ISN), 12 followed the stream "Computer and Digital Creation" (BAC S option ICN) and 18 where in Middle School.

  Among the 68 students, 13 declared having used the educational booklet provided in the kit and 16 declared having used other robotic kits. Concerning the time resource dedicated to activities with the robot, 30 students declared having spent less than 6 hours, 22 declared between 6 and 25 hours, and 16 declared having spent more than 25 hours.

have practiced less than 6 hours of activity with the robot (N = 30), between 6 and 25 hours (N = 22) or more than 25 hours (N = 16); having built the robot (N = 12); have used the visual programming language Snap! (N = 46), the language of Python textual programming (N = 21), both (N = 8) or none (N = 9), it should be noted that these two languages are directly accessible via the main interface of the robot.

- Evaluation of the tool

  We have selected two standardized surveys dealing with this issue: SUS (The Systeme Usability Scales) [59] and The AttrakDiff [96]. These two surveys are complementary and allow to identify the design problems and to account for the perception of the user during the activities. The results of these surveys are available in the article (in French) [26] published at the conference Didapro (Lausanne Feb, 2018). Figures 39 and 40 show the averages of the 96 respondents (68 students + 28 teachers) for each of the 10 statements from the SUS and 28 pairs of antonyms to be scored on a scale of 1 to 5 and a 7-point scale, respectively.

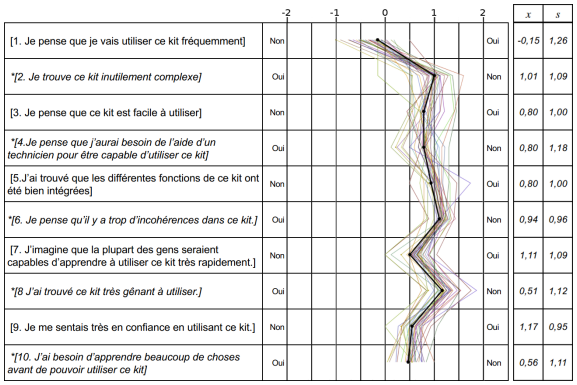


Figure 39. Result of SUS survey

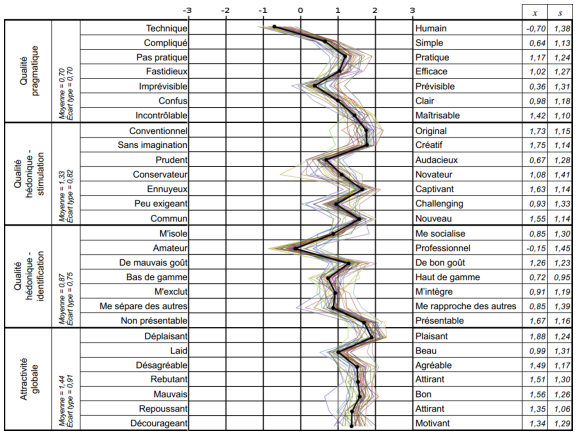


Figure 40. Result of AttrakDiff survey

- Evaluation of impact on learner

One of the objectives of the integration of digital sciences in school is to allow students to have a better understanding of the technological tools that surround them daily (i.e. web, data, algorithm, connected object, etc.). So, we wanted to measure how the practice of activities with ErgoJr robot had changed this apprehension; especially towards robots. For that, we used a standardized survey: "attitude towards robot" *EuroBarometer 382* originally distributed in 2012 to more than 1000 people in each country of the European Union. On the one hand, we sought to establish whether there had been a change in response between 2012 and 2017, and secondly whether there was an impact on the responses of 2017 according to the participation, or not, in educational activities with ErgoJr robot. The analysis of the results is in progress and will be published in 2019.

- Web page for the experimentations

  To facilitate the storage of documents, their availability, and to highlight some information and news, a page dedicated to the experimentations is now available on the website. https://www.poppy-education.org/evaluation/

7.5.2.4. Partnership on education projects

- Ensam

  The Arts and Métiers campus at Bordeaux-Talence in partnership with Inria wishes to contribute to its educational and scientific expertise to the development of new teaching methods and tools. The objective is to develop teaching sequences based on a project approach, relying on an attractive multidisciplinary technological system: the humanoid Inria Poppy robot.

  The humanoid Inria Poppy robot offers an open platform capable of providing an unifying thread for the different subjects covered during the 3-years of the Bachelor training: mechanics, manufacturing (3D printing), electrical, mecha-tronics, computer sciences, design.

- Poppy entre dans la danse (Poppy enters the dance)

  The project "Poppy enters the dance" (Canope 33) took place for the second year. It uses the humanoid robot Poppy. This robot is able to move and experience the dance. The purpose of this project is to allow children to understand the interactions between science and choreography, to play with the random and programmable, to experience movement in dialogue with the machine. At the beginning of the project they attended two days of training on the humanoid robot (Inria - Poppy Education). During the project, they met the choreographer Eric Minh Cuong Castaing and the engineer Segonds Theo (Inria - Poppy Education).

  You can see a description and an overview of the project here:
  https://www.youtube.com/watch?v=XfxXaq899kY

- DANE

  The Academic Delegation for Digital Educational is in charge of supporting the development of digital uses for pedagogy. It implements the educational digital policy of the academy in partnership with local authorities. She accompanies institutions daily, encourages innovations and participates in their dissemination.

- RobotCup Junior

  RoboCupJunior OnStage invites teams to develop a creative stage performance using autonomous robots that they have designed, built and programmed. The objective is to create a robotic performance between 1 to 2 minutes that uses technology to engage an audience. The challenge is intended to be open-ended. This includes a whole range of possible performances, for example dance, storytelling, theatre or an art installation. The performance may involve music but this is optional. Teams are encouraged to be as creative, innovative and entertaining, in both the design of the robots and in the design of the overall performance.

7.5.3. IniRobot: Educational Robotics in Primary Schools

Participants: Didier Roy [correspondant], Pierre-Yves Oudeyer.

Reminder : IniRobot (a project done in collaboration with EPFL/Mobsya) aims to create, evaluate and disseminate a pedagogical kit which uses Thymio robot, an open-source and low cost robot, for teaching computer science and robotics.

IniRobot Project aims to produce and diffuse a pedagogical kit for teachers and animators, to help them and to train them directly or by the way of external structures. The aim of the kit is to initiate children to computer science and robotics. The kit provides a micro-world for learning, and takes an inquiry-based educational approach, where kids are led to construct their understanding through practicing an active investigation methodology within teams. See https://dm1r.inria.fr/c/kits-pedagogiques/inirobot or http://www.inirobot.fr.

Deployment: After 4 years of activity, IniRobot is used by more than 3000 adults, 30 000 children in France. Inirobot is also used in higher education, for example in Master 2 "Neurosciences, human and animal cognition" at the Paul Sabatier University in Toulouse. Inirobot is additionally used to train the management and elected officials of the Bordeaux metropolitan area (20 people). The digital mediators of the 8 Inria centers are trained to Inirobot and use it in their activities.

7.5.3.1. Partnership

The project continues to be carried out in main collaboration with the LSRO Laboratory from EPFL (Lausanne) and others collaborations such as the French National Education/Rectorat d'Aquitaine, the Canopé Educational Network, the ESPE (teacher's school) Aquitaine, the ESPE Martinique, the ESPE Poitiers and the National Directorate of Digital Education.

7.5.3.2. Created pedagogical documents and resources

- The inirobot pedagogical kit [83]: This pedagogical booklet provides activities scenarized as missions to do. An updated version of the Inirobot pedagogical kit is available at: https://dm1r.inria.fr/uploads/default/original/1X/70037bdd5c290e48c7ec4cb4f26f0e426a4b4cf6.pdf. Another pedagogical booklet has been also created by three pedagogical advisers for primary school, with pedagogical instructions and aims, under our supervision. The new pedagogical kit,"Inirobot Scolaire, Langages et robotique", which extends Inirobot to a full primary school approach is available at http://tice33.ac-bordeaux.fr/Ecolien/ASTEP/tabid/5953/language/fr-FR/Default.aspx
- Inirobot website and forum: https://dm1r.inria.fr/c/kits-pedagogiques/inirobot or http://www.inirobot.fr On this website, teachers, animators and general public can download documents, exchange about their use of inirobot's kit.

7.5.3.3. Scientific mediation

Inirobot is very popular and often presented in events (conferences, workshops, ...) by us and others.

7.5.3.4. Spread of Inirobot activities

Inirobot activities are used by several projects: Dossier 123 codez from Main à la Pâte Fundation, Classcode project, ...

7.5.3.5. MOOC Thymio

The MOOC Thymio, released in october 2018, in collaboration with Inria Learning Lab and EPFL (Lausanne, Switzerland), on FUN platform and edX EPFL Platform), use Inirobot activities to teach how to use Thymio robot in education.

GEOSTAT Project-Team

7. New Results

7.1. Excitable systems

Participants: G. Attuel, E. Gerasimova-Chechkina, F. Argoul, H. Yahia, A. Arnéodo.

In a companion paper (I. Multifractal analysis of clinical data), we used a wavelet-based multiscale analysis to reveal and quantify the multifractal intermittent nature of the cardiac impulse energy in the low frequency range (2 Hz during atrial fibrillation (AF). It demarcated two distinct areas within the coronary sinus (CS) with regionally stable multifractal spectra likely corresponding to different anatomical substrates. The electrical activity also showed no sign of the kind of temporal correlations typical of cascading processes across scales, thereby indicating that the multifractal scaling is carried by variations in the large amplitude oscillations of the recorded bipolar electric potential. In the present study, to account for these observations, we explore the role of the kinetics of gap junction channels (GJCs), in dynamically creating a new kind of imbalance between depolarizing and repolarizing currents. We propose a one-dimensional (1D) spatial model of a denervated myocardium, where the coupling of cardiac cells fails to synchronize the network of cardiac cells because of abnormal transjunctional capacitive charging of GJCs. We show that this non-ohmic nonlinear conduction 1D modeling accounts quantitatively well for the "multifractal white noise" dynamics of the electrical activity experimentally recorded in the left atrial posterior wall area. We further demonstrate that the multifractal properties of the numerical impulse energy are robust to changes in the model parameters.

Publication: Frontiers in Physiology (in review forum).

7.2. Multiscale description in terms of multiplicative cascade, application to Earth observation signals

Participants: I. Hernandez-Carrasco, V. Garçon, J. Sudre, C. Garbe, H. Yahia.

A new methodology has been developed in order to improve the description of the spatial and temporal variability of not well-resolved oceanic variables from other well-observed high-resolution oceanic variables. The method is based on the cross-scale inference of information, incorporating the common features of different multifractal high-resolution variables into a coarser one. An exercise of validation of the methodology has been performed based on the outputs of coupled physical-biogeochemical Regional Ocean Modeling System adapted to the eastern boundary upwelling systems at two spatial resolutions. Once the algorithm has been proved to be effective in increasing the spatial resolution of modeled partial pressure of CO_2 at the surface ocean (pCO_2), we have investigated the capability of our methodology when it is applied to remote sensing data, focusing on the improvement of the temporal description. In this regard, we have inferred daily pCO_2 maps at high resolution (4 kms) fusing monthly pCO_2 data at low resolution (100 kms) with the small-scale features contained in daily high-resolution maps of satellite sea surface temperature and Chlorophyll-a. The algorithm has been applied to the South Eastern Atlantic Ocean opening the possibility to obtain an accurate quantification of the CO_2 fluxes in relevant coastal regions, such as the eastern boundary upwelling systems. Outputs of our algorithm have been compared with in situ measurements, showing that daily maps inferred from monthly products are in average 6 μatm closer to the in situ values than the original coarser monthly maps. Furthermore, values of pCO_2 have been improved in points close to the coast with respect to the original input data.

Publication: IEEE Transactions on Geoscience and Remote Sensing, HAL.

7.3. Multiscale description in terms of multiplicative cascade, application to Earth observation signals

Participants: H. Yahia, V. Garçon, J. Sudre, C. Maes.

We evidence and study the differences in turbulence statistics in ocean dynamics carried by wind forcing at the air-sea interface. Surface currents at the air-sea interaction are of crucial importance because they transport heat from low to high latitudes. At first order, oceanic currents are generated by the balance of the Coriolis and pressure gradient forces (geostrophic current) and the balance of the Coriolis and the frictional forces dominated by wind stress (Ekman current) in the surface ocean layers. The study was conducted by computing statistical moments on the shapes of spectra computed within the framework of microcanonical multi-fractal formalism. Remotely sensed daily datasets derived from one year of altimetry and wind data were used in this study, allowing for the computation of two kinds of vector fields: geostrophy with and geostrophy without wind stress forcing. We explore the statistical properties of singularity spectra computed from velocity norms and vorticity data, notably in relation with kurtosis information to underline the differences in the turbulent regimes associated with both kinds of velocity fields.

Publication: Frontiers of Information Technology & Electronic Engineering , Springer, 2018, 19 (8), HAL.

7.4. Multiscale description in terms of multiplicative cascade, application to Earth observation signals

Participants: A. El Aouni, K. Daoudi, H. Yahia, K. Minaoui.

We study coherent vortex detection from particles trajectories analysis and surface mixing and biological activity in the north african upwelling.

Publications: SIAM Conference on Nonlinear Waves and Coherent Structures HAL and AGU Ocean Sciences Meeting 2018 HAL.

7.5. Data-based identification of characteristic scales and automated modeling

Participants: N. Brodu, G. S. Phartiyal, D. Singh, H. Yahia.

Low-rankness transfer for denoising Sentinel-1 SAR images. Published in the 9th International Symposium on Signal, Image, Video and Communications ISIVC, Rabat, 2018, HAL.

A mixed spectral and spatial Convolutional Neural Network for Land Cover Classification using SAR and Optical data. Published in EGU General Assembly, Vienna, 2018, HAL.

Inferrence of causal states from time series for empirical modeling at prescribed scales. The goal of this research is to recover physical systems internal states from data and build a model of their evolution. Clustering together data with the same causal effets leads to consistent internal states: each measured data inferred to match the same state has by definition the same consequence, hence the same functional role. The theory behind this is well established, with major steps in the 80's by Jim Crutchfield. This leads to computational mechanics and epsilon-machines in the discrete case. The theory has however always suffers from computationability issues and it is very hard to apply in practice on large systems and real data. N. Brodu has made (unpublished) progress in 2018 in this theory, showing links between epsilon-machines and stochastic processes in the continuous case. The goal is to form a new class of algorithms drawing on the continuous representation, which would not suffer from the explicit discretization steps needed by current algorithms. N. Brodu has initiated a collaboration with Jim Crutchfield in 2017 and hope to further enhance that collaboration in 2019. This plan was presented to the reviewers during the team evaluation and deemed to be of high priority.

7.6. Speech analysis

Participants: G. Li, K. Daoudi, J. Klempir, J. Rusz, B. Das.

In the early stage of disease, the symptoms of Parkinson's disease (PD) are similar to atypical Parkinsonian sydromes (APS). The early differential diagnosis between PD and APS is thus a very challenging task. It trurns out that speech disorder is an early and common symptom to PD and APS. The goal of reserach is to develop a digital marker based on speech analysis in order to assist the neurologists in their diagnosis.

Publication: IEEE-ICASSP - 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, Apr 2018, HAL.

7.7. InnovationLab with I2S, sparse signals & optimisation

Participants: M. Martin, A. Zebadua, S. Sakka, N. Brodu, K. Daoudi, A. Cherif [I2S], J. L. Vallancogne [I2S], A. Cailly [I2S].

During 2018 A. Zebadua was involved for the first time in the use of non-convex optimization methods and proximal operators, to solve inverse problems in image processing. Within the framework of the joint project with i2s, his main task was to work together with the research engineers to understand, implement, adapt and reduce the execution time of the algorithms developed by H. Badri in his doctoral thesis. These algorithms were developed in a scientific context, it was, therefore, necessary to adapt them to an industrial context with practical constraints.

S. Sakka has been implementating of the demosaicing algorithm to reconstruct a full-color image from raw images acquired by scanning. Colors are generated by a mask of Bayer. This algorithm uses the Hamilton Method and the Edge aware smoothing algorithm. He has been performing benchmarks for performance and quality test, and has written the technical report about the demosaicing algorithm. S. Sakka has been implementing iterative algorithms for linear system resolution useful for inpaitnig image processing algorithm.

S. Sakka has been implementating of the demosaicing algorithm to reconstruct a full-color image from raw images acquired by scanning. Colors are generated by a mask of Bayer. This algorithm uses the Hamilton Method and the Edge aware smoothing algorithm. He has been performing benchmarks for performance and quality test, and has written the technical report about the demosaicing algorithm. S. Sakka has been implementing iterative algorithms for linear system resolution useful for inpainting image processing algorithm.

C. Sakka has been participating in the GPU Technology Conference organized by Nvidia in Munich and attending the NVIDIA Deep Learning School, and A. Zebadua has been participating to the winter school itwist18 in optimization from November 19th to 20th in Marseille.

M. Martin studied and made comparison between two methods : EAS (Edge Aware Smoothing) algorithm and LRT (low rank transfer) for denoising. She has been writing technical report to choose with I2S the best method : more efficient, less time. M. Martin has also been implementing denoising in C++ with librairy opencv and gpu. She has been setting up code sharing with I2S on gitlab Inria

HIEPACS Project-Team

7. New Results

7.1. High-performance computing on next generation architectures

7.1.1. Evaluation of dataflow programming models for electronic structure theory

Dataflow programming models have been growing in popularity as a means to de- liver a good balance between performance and portability in the post-petascale era. In this paper we evaluate different dataflow programming models for electronic struc- ture methods and compare them in terms of programmability, resource utilization, and scalability. In particular, we evaluate two programming paradigms for expressing scientific applications in a dataflow form: (1) explicit dataflow, where the dataflow is specified explicitly by the developer, and (2) implicit dataflow, where a task schedul- ing runtime derives the dataflow using per-task data-access information embedded in a serial program. We discuss our findings and present a thorough experimental analysis using methods from the NWChem quantum chemistry application as our case study, and OpenMP, StarpU and ParSEC as the task-based runtimes that enable the different forms of dataflow execution. Furthermore, we derive an abstract model to explore the limits of the different dataflow programming paradigms.

More information on these results can be found in [8]

7.1.2. On soft errors in the Conjugate Gradient method: sensitivity and robust numerical detection

The conjugate gradient (CG) method is the most widely used iterative scheme for the solution of large sparse systems of linear equations when the matrix is symmetric positive definite. Although more than sixty year old, it is still a serious candidate for extreme-scale computation on large computing platforms. On the technological side, the continuous shrinking of transistor geometry and the increasing complexity of these devices affect dramatically their sensitivity to natural radiation, and thus diminish their reliability. One of the most common effects produced by natural radiation is the single event upset which consists in a bit-flip in a memory cell producing unexpected results at application level. Consequently, the future computing facilities at extreme scale might be more prone to errors of any kind including bit-flip during calculation. These numerical and technological observations are the main motivations for this work, where we first investigate through extensive numerical experiments the sensitivity of CG to bit-flips in its main computationally intensive kernels, namely the matrix-vector product and the preconditioner application. We further propose numerical criteria to detect the occurrence of such faults; we assess their robustness through extensive numerical experiments.

More information on these results can be found in [16].

7.1.3. Energy analysis of a solver stack for frequency-domain electromagnetics

High-performance computing (HPC) aims at developing models and simulations for applications in numerous scientific fields. Yet, the energy consumption of these HPC facilities currently limits their size and performance, and consequently the size of the tackled problems. The complexity of the HPC software stacks and their various optimizations makes it difficult to finely understand the energy consumption of scientific applications. To highlight this difficulty on a concrete use-case, we perform an energy and power analysis of a software stack for the simulation of frequency-domain electromagnetic wave propagation. This solver stack combines a high order finite element discretization framework of the system of three-dimensional frequency-domain Maxwell equations with an algebraic hybrid iterative-direct sparse linear solver. This analysis is conducted on the KNL-based PRACE-PCP system. Our results illustrate the difficulty in predicting how to trade energy and runtime.

More information on these results can be found in [18].

7.1.4. A compiler front-end for OpenMP's variants

OpenMP 5.0 introduced the concept of *variant*: a directive which can be used to indicate that a function is a variant of another existing *base function*, in a specific context (eg: `foo_gpu_nvidia` could be declared as a variant of `foo`, but only when executing on specific NVIDIA hardware).

In the context of PRACE-5IP, in collaboration with the Inria STORM team, we want to leverage this construct to be able to take advantage of the StarPU heterogeneous scheduler through the interoperability layer between OpenMP and StarPU. We started this work by implementing the necessary changes in the Clang front-end to support OpenMP's *variant*. We have assessed this support in the `Chameleon` dense linear algebra package. Indeed, `Chameleon` relies on sequential task-based algorithms where sub-tasks of the overall algorithms are submitted to a runtime system. Additionally to the `quark`, `PaRSEC` and `StarPU` support, we have implemented an OpenMP support in `Chameleon`. The originality of the proposed approach is that this OpenMP support can either rely on a native OpenMP runtime system or indirectly use the above mentioned OpenMP-StarPU back-end. We are currently assessing the approach on multicore homogeneous machines, the next step being heterogeneous architectures.

7.2. High performance solvers for large linear algebra problems

7.2.1. Partitioning and communication strategies for sparse non-negative matrix factorization

Non-negative matrix factorization (NMF), the problem of finding two non-negative low-rank factors whose product approximates an input matrix, is a useful tool for many data mining and scientific applications such as topic modeling in text mining and blind source separation in microscopy. In this paper, we focus on scaling algorithms for NMF to very large sparse datasets and massively parallel machines by employing effective algorithms, communication patterns, and partitioning schemes that leverage the sparsity of the input matrix. In the case of machine learning workflow, the computations after SpMM must deal with dense matrices, as Sparse-Dense matrix multiplication will result in a dense matrix. Hence, the partitioning strategy considering only SpMM will result in a huge imbalance in the overall workflow especially on computations after SpMM and in this specific case of NMF on non-negative least squares computations. Towards this, we consider two previous works developed for related problems, one that uses a fine-grained partitioning strategy using a point-to-point communication pattern and on that uses a checkerboard partitioning strategy using a collective-based communication pattern. We show that a combination of the previous approaches balances the demands of the various computations within NMF algorithms and achieves high efficiency and scalability. From the experiments, we could see that our proposed algorithm communicates atleast 4x less than the collective and achieves upto 100x speed up over the baseline FAUN on real world datasets. Our algorithm was experimented in two different super computing platforms and we could scale up to 32000 processors on Bluegene/Q.

More information on these results can be found in [21].

7.2.2. Low-rank factorizations in data sparse hierarchical algorithms for preconditioning Symmetric positive definite matrices

We consider the problem of choosing low-rank factorizations in data sparse ma- trix approximations for preconditioning large scale symmetric positive definite matrices. These approximations are memory efficient schemes that rely on hierarchical matrix partitioning and compression of certain sub-blocks of the matrix. Typically, these matrix approximations can be constructed very fast, and their matrix product can be applied rapidly as well. The common prac- tice is to express the compressed sub-blocks by low-rank factorizations, and the main contribution of this work is the numerical and spectral analysis of SPD preconditioning schemes represented by $2\times$ 2 block matrices, whose off-diagonal sub-blocks are low-rank approximations of the original matrix off-diagonal sub-blocks. We propose an optimal choice of low-rank approximations which minimizes the condition number of the preconditioned system, and demonstrate that the analysis can be applied to the class of hierarchically off-diagonal low-rank matrix approximations. Spec- tral estimates that take into account the error propagation through levels of the hierarchy which quantify the impact of the choice of low-rank compression on the global condition number are provided. The numerical results indicate that

the properties of the preconditioning scheme using proper low-rank compression are superior to employing standard choices for low-rank compression. A major goal of this work is to provide an insight into how proper reweighted prior to low-rank compression influences the condition number for a simple case, which would lead to an extended analysis for more general and more efficient hierarchical matrix approximation techniques.

More information on these results can be found in [5].

7.2.3. Analyzing the effect of local rounding error propagation on the maximal attainable accuracy of the pipelined Conjugate Gradient method

Pipelined Krylov subspace methods typically offer improved strong scaling on par- allel HPC hardware compared to standard Krylov subspace methods for large and sparse linear systems. In pipelined methods the traditional synchronization bottleneck is mitigated by overlap- ping time-consuming global communications with useful computations. However, to achieve this communication-hiding strategy, pipelined methods introduce additional recurrence relations for a number of auxiliary variables that are required to update the approximate solution. This paper aims at studying the influence of local rounding errors that are introduced by the additional recurrences in the pipelined Conjugate Gradient (CG) method. Specifically, we analyze the impact of local round-off effects on the attainable accuracy of the pipelined CG algorithm and compare it to the traditional CG method. Furthermore, we estimate the gap between the true residual and the recursively computed residual used in the algorithm. Based on this estimate we suggest an automated residual replacement strategy to reduce the loss of attainable accuracy on the final iterative solution. The resulting pipelined CG method with residual replacement improves the maximal attainable accuracy of pipelined CG while maintaining the efficient parallel performance of the pipelined method. This conclusion is substantiated by numerical results for a variety of benchmark problems.

More information on these results can be found in [7].

7.2.4. Sparse supernodal solver using block low-rank compression: Design, performance and analysis

We propose two approaches using a Block Low-Rank (BLR) compression technique to reduce the memory footprint and/or the time-to-solution of the sparse supernodal solver PaStiX. This flat, non-hierarchical, compression method allows to take advantage of the low-rank property of the blocks appearing during the factorization of sparse linear systems, which come from the discretization of partial differential equations. The proposed solver can be used either as a direct solver at a lower precision or as a very robust preconditioner. The first approach, called *Minimal Memory*, illustrates the maximum memory gain that can be obtained with the BLR compression method, while the second approach, called *Just-In-Time*, mainly focuses on reducing the computational complexity and thus the time-to-solution. Singular Value Decomposition (SVD) and Rank-Revealing QR (RRQR), as compression kernels, are both compared in terms of factorization time, memory consumption, as well as numerical properties. Experiments on a shared memory node with 24 threads and 128 GB of memory are performed to evaluate the potential of both strategies. On a set of matrices from real-life problems, we demonstrate a memory footprint reduction of up to 4 times using the *Minimal Memory* strategy and a computational time speedup of up to 3.5 times with the *Just-In-Time* strategy. Then, we study the impact of configuration parameters of the BLR solver that allowed us to solve a 3D laplacian of 36 million unknowns a single node, while the full-rank solver stopped at 8 million due to memory limitation.

These contributions have been published in International Journal of Computational Science and Engineering (JoCS) [9].

7.2.5. Supernodes ordering to enhance Block Low-Rank compression in a sparse direct solver

Solving sparse linear systems appears in many scientific applications, and sparse direct linear solvers are widely used for their robustness. Still, both time and memory complexities limit the use of direct methods to solve larger problems. In order to tackle this problem, low-rank compression techniques have been introduced in direct solvers to compress large dense blocks appearing in the symbolic factorization. In this paper, we consider the Block Low-Rank compression (BLR) format and adress the problem of clustering unknowns

that come from separators issued from the nested dissection process. We show that methods considering only intra-separators connectivity (i.e., k-way or recursive bisection) as well as methods managing only interaction between separators have some limitations. We propose a new strategy that considers interactions between a separator and its children to pre-select some interactions while reducing the number of off-diagonal blocks in the symbolic structure. We demonstrate how this new method enhances the BLR strategies in the sparse direct supernodal solver PaStiX.

These contributions have been submitted in SIAM Journal on Matrix Analysis and Applications (SIMAX) [22].

7.3. Parallel Low-Rank Linear System and Eigenvalue Solvers Using Tensor Decompositions

At the core of numerical simulations for scientific computing applications, one typically needs to solve an equation either in the form of a linear system (Ax = b) or an eigenvalue problem (Ax = x) to determine the course of the simulation. A major breakthrough in this solution step is exploiting the inherent low-rank structure in the problem; an idea stemming from the observation that particles in the same spatial locality exhibit similar interactions with others in a distant cluster/region. This property has been exploited in many contexts such as fast multipole methods (FMM) and hierarchical matrices (H-matrices) in applications ranging from n-body simulations to electromagnetics, which amount to numerically compressing the matrix in order to reduce computational and memory costs. Recent theory along this direction involves employing tensor decomposition to quantize the matrix in the form of a tensor (through logical restructuring/reshaping) and use tensor decomposition to approximate it with a controllable global error. Once the matrix and vectors are compressed this way, one can similarly use the compressed tensor to carry out matrix-vector operations with significantly better compression rate than the H-matrix approach.

Despite these major recent breakthroughs in the theory and application of tensor-based methods, addressing large-scale real-world problems with these methods requires immense computational power, which necessitates highly optimized parallel algorithms and implementations. To this end, we have initiated the development of a tensor-based linear system and eigenvalue solver library called Celeste++(C++ library for Efficient low-rank Linear and Eigenvalue Solvers using Tensor decomposition) providing a complete framework for expressing a problem in tensor form, then effectuating all matrix-vector operations under this compressed form with tremendous computational and memory efficiency. The fruits of our preliminary studies led two project submissions at the national scale (ANR JCJC and CNRS PEPS JCJC, currently under evaluation) and one Severo Ochoa Mobility Grant for a collaboration visit to Barcelona Supercomputing Center (BSC). We also supervised an internship on the application of tensor solvers in the context of electromagnetic applications with very promising results for future work.

7.4. Efficient algorithmic for load balancing and code coupling in complex simulations

7.4.1. StarPart Redesign

In the context of the french ICARUS project (FUI), which focuses the development of high-fidelity calculation tools for the design of hot engine parts (aeronautics & automotive), we are looking to develop new load-balancing algorithms to optimize the complex numerical simulations of our industrial and academic partners (Turbomeca, Siemens, Cerfacs, Onera, ...). Indeed, the efficient execution of large-scale coupled simulations on powerful computers is a real challenge, which requires revisiting traditional load-balancing algorithms based on graph partitioning. A thesis on this subject has already been conducted in the Inria HiePACS team in 2016 by Maria Predari, which has successfully developed a co-partitioning algorithm that balances the load of two coupled codes by taking into account the coupling interactions between these codes.

This work was initially integrated into the StarPart platform. The necessary extension of our algorithms to parallel & distributed (increasingly dynamic) versions has led to a complete redesign of StarPart, which has been the focus of our efforts this year. The StarPart framework provides the necessary building blocks to develop new graph algorithms in the context of HPC, such as those we are targeting. The strength of StarPart lies in the fact that it is a light runtime system applied to the issue of "graph computing". It provides a unified data model and a uniform programming interface that allows easy access to a dozen partitioning libraries, including Metis, Scotch, Zoltan, etc. Thus, it is possible, for example, to load a mesh from an industrial test case provided by our partners (or an academic graph collection as DIMACS'10) and to easily compare the results for the different partitioners integrated in StarPart.

7.5. Application Domains

7.5.1. Material physics

7.5.1.1. EigenSolver

The adaptive vibrational configuration interaction algorithm has been introduced as a new eigennvalues method for large dimension problem. It is based on the construction of nested bases for the discretization of the Hamiltonian operator according to a theoretical criterion that ensures the convergence of the method. It efficiently reduce the dimension of the set of basis functions used and then we are able solve vibrationnal eigenvalue problem up to the dimension 15 (7 atoms). This year we have worked on three main areas. First, we extend our shared memory parallelization to distributed memory using the message exchange paradigm. This new version should allow us to process larger systems quickly. To target the eigenvalues relevant for chemists, i. eigenvalues with an intensity. This requires calculating the scalar product between the smallest eigenvalues and the dipole moment applied to an eigenvector to evaluate its intensity. In addition, to get closer to the experimental values, we introduced the Coriolis operator into the Hamiltonian. A paper is being written on these last two points.

7.5.1.2. Dislocation

We have focused on the improvements of the parallel collision detection and the data structure in the OPTIDIS code [11].

- The new collision detection algorithm to reliably handle junction formation for Dislocation Dynamics using hybrid OpenMP + MPI parallelism has been developed. The enhanced precision and reliability of this new algorithm allows the use of larger time-steps for faster simulations. Hierarchical methods for collision detection, as well as hybrid parallelism are also used to improve performance;
- A new distributed data structure has been developed to enhance the reliability and modularity of OPTIDIS. The new data structure provides an interface to modify safely and reliably the distributed dislocation mesh in order to enforce data consistency across all computation nodes. This interface also improves code modularity allowing the study of data layout performance without modifying the algorithms.

7.5.2. Co-design for scalable numerical algorithms in scientific applications

7.5.2.1. A geometric view of biodiversity: scaling to metagenomics

We have designed a new efficient dimensionality reduction algorithm in order to investigate new ways of accurately characterizing the biodiversity, namely from a geometric point of view, scaling with large environmental sets produced by NGS ($\sim$105 sequences). The approach is based on Multidimensional Scaling (MDS) that allows for mapping items on a set of n points into a low dimensional euclidean space given the set of pairwise distances. We compute all pairwise distances between reads in a given sample, run MDS on the distance matrix, and analyze the projection on first axis, by visualization tools. We have circumvented the quadratic complexity of computing pairwise distances by implementing it on a hyperparallel computer (Turing, a Blue Gene Q), and the cubic complexity of the spectral decomposition by implementing a dense random projection based algorithm. We have applied this data analysis scheme on a set of 10^5 reads, which are amplicons of a diatom environmental sample from Lake Geneva. Analyzing the shape of the point cloud paves the way for a geometric analysis of biodiversity, and for accurately building OTUs (Operational Taxonomic Units), when the data set is too large for implementing unsupervised, hierarchical, high-dimensional clustering.

More information on these results can be found in [19].

7.5.2.2. High performance simulation for ITER tokamak

Concerning the GYSELA global non-linear electrostatic code, a critical problem is the design of a more efficient parallel gyro-average operator for the deployment of very large (future) GYSELA runs. The main unknown of the computation is a distribution function that represents either the density of the guiding centers, either the density of the particles in a tokamak. The switch between these two representations is done thanks to the gyro-average operator. In the previous version of GYSELA, the computation of this operator was achieved thanks to a Padé approximation. In order to improve the precision of the gyro-averaging, a new parallel version based on an Hermite interpolation has been done (in collaboration with the Inria TONUS project-team and IPP Garching). The integration of this new implementation of the gyro-average operator has been done in GYSELA and the parallel benchmarks have been successful. This work is carried on in the framework of the PhD of Nicolas Bouzat (funded by IPL C2S@Exa) co-advised with Michel Mehrenberger from TONUS project-team and in collaboration with Guillaume Latu from CEA-IRFM. The scientific objectives of this work are first to consolidate the parallel version of this gyro-average operator, in particular by designing a scalable MPI+OpenMP parallel version and by using a new communication scheme, and second to design new numerical methods for the gyro-average, source and collision operators to deal with new physics in GYSELA. The objective is to tackle kinetic electron configurations for more realistic complex large simulations. This has been done by using a new data distribution for a new irregular mesh in order to take into account the complex geometries of modern tokamak reactors. All these contributions have been validated on a new object-oriented proptotype of GYSELA which uses a task based programming model. The PhD thesis of Nicolas Bouzat has been defended on December 17, 2018.

In the context of the EoCoE project, we have collaborations with CEA-IRFM. First, with G. Latu, we have investigated the potential of using the last release of the PaStiX solver (version 6.0) on Intel KNL architecture, and more especially on the MARCONI machine (one of the PRACE supercomputers at Cineca, Italia). The results obtained on this architecture are really promising since we are able to reach more than 1 Tflops using a single node. Secondly, we also have a collaboration with P. Tamain and G. Giorgani on the TOKAM3X code to analyze the performance of using PaStiX as a preconditioner. Since a distributed memory is required during the simulation, the previous release of PaStiX is then used. Some difficulties regarding the Fortran wrapper and some memory issues should be fixed when we will have reimplemented the MPI interface in the current release.

7.5.2.3. Numerical and parallel scalable hybrid solvers in large scale calculations

Numerically scalable hybrid solvers based on a fully algebraic coarse space correction have been theoretically studied within the PhD thesis of Louis Poirel defended on November 28, 2018. Some of the proposed numerical schemes have been integrated in the MaPHyS parallel package. In particular, multiple parallel strategies have been designed and their parallel efficiencies were assessed in two large application codes. The first one is Alya developed at BSC, that is a high performance computational mechanics code to solve coupled multi-physics / multi-scale problems, which are mostly coming from engineering applications. This activity was carried out in the framework of the EoCoE project. Thye second large code is AVIP jointly developed by CERFACS and Laboratoire de Physique des Plasmas at E´cole Polytechnique for the calculation of plasma propulsion. For this latter code, part of the parallel experiments were conduced on a PRACE Tier-0 machine within a PRACE Project Access.

LFANT Project-Team

6. New Results

6.1. Cryptographic Protocols

Participants: Guilhem Castagnos, Ida Tucker.

In [24], G. Castagnos, F. Laguillaumie and I. Tucker revisit a recent cryptographic primitive called *Functional encryption for inner products* (FE4IP).

Functional encryption (FE) is an advanced cryptographic primitive which allows, for a single encrypted message, to finely control how much information on the encrypted data each receiver can recover. To this end many functional secret keys are derived from a master secret key. Each functional secret key allows, for a ciphertext encrypted under the associated public key, to recover a specific function of the underlying plaintext.

Since constructions for general FE that appear in the past five years are far from practical, the problem arose of building efficient FE schemes for restricted classes of functions; and in particular for linear functions, (i.e. the inner product functionality). Such constructions yield many practical applications, while developing our understanding of FE.

Though such schemes had already been conceived in the past three years (Abdalla *et al.* 2015, Agrawal *et al.* 2016), they all suffered of practical drawbacks. Namely the computation of inner products modulo a prime are restricted, in that they require that the resulting inner product be small for decryption to be efficient. The only existing scheme that overcame this constraint suffered of poor efficiency due in part to very large ciphertexts. This work overcomes these limitations and we build the first FE schemes for inner products modulo a prime that are both efficient and recover the result whatever its size.

To this end, Castagnos *et al.* introduce two new cryptographic assumptions. These are variants of the assumptions used for the Castagnos-Laguillaumie encryption of 2015. This supposes the existence of a cyclic group G where the decision Diffie-Hellman assumption holds together with a subgroup F of G where the discrete logarithm problem is easy. This setting allows to encode information in the exponent of the subgroup F, which can be efficiently recovered whatever its size.

From these assumptions Castagnos *et al.* construct generic, linearly homomorphic encryption schemes over a field of prime order which are semantically secure under chosen plaintext attacks. They then use the homomorphic properties of the above schemes to construct generic inner product FE schemes over the integers and over fields of prime order. They thereby provide constructions for inner product FE modulo a prime p that do not restrict the size of the inputs or of the resulting inner product, which are the most efficient such schemes to date.

This paper was presented at the ASIACRYPT Conference 2018, and is part of the ALAMBIC project.

6.2. Computation of Euclidean minima in totally definite quaternion fields

Participant: Jean-Paul Cerri.

In collaboration with Pierre Lezowski, Jean-Paul Cerri has studied norm-Euclidean properties of totally definite quaternion fields over number fields. Building on their previous work about number fields, they have proved that the Euclidean minimum and the inhomogeneous minimum of orders in such quaternion fields are always equal. Besides, they are rational under the hypothesis that the base number field is not quadratic. This single remaning open case corresponds to the similar open case remaining for real number fields.

They also have extended Cerri's algorithm for the computation of the upper part of the norm-Euclidean spectrum of a number field to this noncommutative context. This algorithm has allowed to compute the exact value of the norm-Euclidean minimum of orders in totally definite quaternion fields over a quadratic number field. This has provided the first known values of this minimum when the base number field has degree strictly greater than 1.

Consequently, both theoretical and practical milestones set in the previous quadrennial report were reached. These results are presented in [19], due to appear in *International Journal of Number Theory*.

6.3. Can you hear the homology of 3-dimensional drums?

Participant: Aurel Page.

In [16], A. Bartel and A. Page describe all possible actions of groups of automorphisms on the homology of 3-manifolds, and prove that for every prime p, there are 3-dimensional drums that sound the same but have different p-torsion in their homology. This completes previous work [42] by proving that the behaviour observed by computer experimentation was indeed a general phenomenon.

More precisely: if M is a manifold with an action of a group G, then the homology group $H_1(M, \mathbb{Q})$ is naturally a $\mathbb{Q}[G]$-module, where $\mathbb{Q}[G]$ denotes the rational group ring. Bartel and Page prove that for every finite group G, and for every $\mathbb{Q}[G]$-module V, there exists a closed hyperbolic 3-manifold M with a free G-action such that the $\mathbb{Q}[G]$-module $H_1(M, \mathbb{Q})$ is isomorphic to V. They give an application to spectral geometry: for every finite set P of prime numbers, there exist hyperbolic 3-manifolds N and N' that are strongly isospectral such that for all $p \in P$, the p-power torsion subgroups of $H_1(N, \mathbb{Z})$ and of $H_1(N', \mathbb{Z})$ have different orders. They also show that, in a certain precise sense, the rational homology of oriented Riemannian 3-manifolds with a G-action "knows" nothing about the fixed point structure under G, in contrast to the 2-dimensional case. The main geometric techniques are Dehn surgery and, for the spectral application, the Cheeger-Müller formula, but they also make use of tools from different branches of algebra, most notably of regulator constants, a representation theoretic tool that was originally developed in the context of elliptic curves.

6.4. Error-correcting codes based on non-commutative algebras

Participant: Aurel Page.

In [36], C. Maire and A. Page revisit a construction due to Lenstra and Guruswami by generalising them to unit groups of division algebras.

Lenstra and Guruswami described number field analogues of the algebraic geometry codes of Goppa. Recently, Maire and Oggier generalised these constructions to other arithmetic groups: unit groups in number fields and orders in division algebras; they suggested to use unit groups in quaternion algebras but could not completely analyse the resulting codes. Maire and Page prove that the noncommutative unit group construction yields asymptotically good families of codes for division algebras of any degree, and estimate the smallest possible size of the alphabet in terms of the degree of the algebra.

6.5. Towards practical key exchange from ordinary isogeny graphs

Participant: Jean Kieffer.

In [25], L. De Feo, J. Kieffer and B. Smith revisit the ordinary isogeny-graph based cryptosystems of Couveignes and Rostovtsev–Stolbunov, long dismissed as impractical.

De Feo, Kieffer and Smith give algorithmic improvements that accelerate key exchange in this framework, and explore the problem of generating suitable system parameters for contemporary pre-and post-quantum security that take advantage of these new algorithms. They prove the session-key security of this key exchange in the Canetti-Krawczyk model, and the IND-CPA security of the related public-key encryption scheme, under reasonable assumptions on the hardness of computing isogeny walks. This system admits efficient key-validation techniques that yield CCA-secure encryption, thus providing an important step towards efficient post-quantum non-interactive key exchange (NIKE).

6.6. Optimal addition sequences for theta functions

Participants: Andreas Enge, Fredrik Johansson.

In [20], A. Enge, F. Johansson and their coauthor W. Hart consider the problem of numerically evaluating one-dimensional θ-functions and the elliptic η-function. They construct short addition sequences reaching an optimal number of $N + o(N)$ multiplications for evaluating the function as a sparse series with N terms. The proof relies on the representability of specific quadratic progressions of integers as sums of smaller numbers of the same kind. For example, they show that every generalised pentagonal number $c > 5$ can be written as $c = 2a + b$, where a, b are smaller generalised pentagonal numbers. They then give a baby-step giant-step algorithm that breaks through the theoretical barrier achievable with addition sequences, and which uses only $O(N/(logN)^r)$ multiplications for any $r > 0$. These theoretical improvements also lead to an interesting speed-up in practice, and they have been integrated into the CM and the ARB software.

6.7. Reed–Solomon-Gabidulin Codes

Participant: Xavier Caruso.

In [31], X. Caruso and A. Durand define a new family of linear codes which is a common generalization of Reed–Solomon codes on the one hand and Gabidulin codes on the other hand. Their construction works over an arbitrary field (not necessarily finite) equipped with an automorphism of finite order and a twisted derivation whose subfield of constants is sufficiently large. This setting allows for example the base field to be $\mathbb{F}_q(t)$ equipped with its natural derivation and then provides a new large family of interesting codes. Caruso and Durand then compute the minimal distance of their codes and design an efficient algorithm for decoding up to the half of the minimal distance.

6.8. Computing Stieltjes constants using complex integration

Participant: Fredrik Johansson.

In [32], F. Johansson and I. Blagouchine devise an efficient algorithm to compute the generalized Stieltjes constants $\gamma_n(a)$ to arbitrary precision with rigorous error bounds, for the first time achieving this with low complexity with respect to the order n. The algorithm consists of locating an approximate steepest descent contour and then evaluating the integral numerically in ball arithmetic using the Petras algorithm with a Taylor expansion for bounds near the saddle point. An implementation is provided in the Arb library.

6.9. Numerical Evaluation of Elliptic Functions, Elliptic Integrals and Modular Forms

Participant: Fredrik Johansson.

In [33], F. Johansson describes algorithms to compute elliptic functions and their relatives (Jacobi theta functions, modular forms, elliptic integrals, and the arithmetic-geometric mean) numerically to arbitrary precision with rigorous error bounds for arbitrary complex variables. Implementations in ball arithmetic are available in the Arb library. This overview article discusses the standard algorithms from a concrete implementation point of view, and also presents some improvements.

6.10. Numerical integration in arbitrary-precision ball arithmetic

Participant: Fredrik Johansson.

In [26], F. Johansson describes an implementation of arbitrary-precision numerical integration with rigorous error bounds in the Arb library. Rapid convergence is ensured for piecewise complex analytic integrals by use of the Petras algorithm, which combines adaptive bisection with adaptive Gaussian quadrature where error bounds are determined via complex magnitudes without evaluating derivatives. The code is general, easy to use, and efficient, often outperforming existing non-rigorous software.

6.11. Fast and rigorous arbitrary-precision computation of Gauss-Legendre quadrature nodes and weights

Participant: Fredrik Johansson.

In [26], F. Johansson and M. Mezzarobba describe a strategy for rigorous arbitrary-precision evaluation of Legendre polynomials on the unit interval and its application in the generation of Gauss-Legendre quadrature rules. The focus is on making the evaluation practical for a wide range of realistic parameters, corresponding to the requirements of numerical integration to an accuracy of about 100 to 100 000 bits. The algorithm combines the summation by rectangular splitting of several types of expansions in terms of hypergeometric series with a fixed-point implementation of Bonnet's three-term recurrence relation. Rigorous enclosures of the Gauss-Legendre nodes and weights are then computed using the interval Newton method. The work provides rigorous error bounds for all steps of the algorithm. The approach is validated by an implementation in the Arb library, which achieves order-of-magnitude speedups over previous code for computing Gauss-Legendre rules with simultaneous high degree and precision.

6.12. On a two-valued sequence and related continued fractions in power series fields

Participant: Bill Allombert.

In [15], Bill Allombert with Nicolas Brisebarre and Alain Lasjaunias describe a noteworthy transcendental continued fraction in the field of power series over $\mathbb{Q}$, having irrationality measure equal to 3. This article has been published in The Ramanujan Journal.

6.13. Moduli space

Participant: Nicolas Mascot.

The article [22] by Nicolas Mascot, on the Certification of modular Galois representations has been published in Mathematics of Computation.

6.14. Modular forms

Participants: Karim Belabas, Henri Cohen, Bill Allombert.

In [18], K. Belabas and H. Cohen give theoretical and practical information on the Pari/GP modular forms package, using the formalism of trace formulas. This huge package (about 70 exported public functions) handles standard operations on classical modular forms in $M_k(\Gamma_0(N), \chi)$, also in weight 1 and non-integral weight (which are not cohomological, hence not directly handled by trace formulas). It is the first publicly available package which can compute Fourier expansions at any cusps, evaluate modular forms near the real axis, evaluate L-functions of non-eigenforms, and compute general Petersson scalar products.

In [39], H. Cohen explained how to compute Fourier expansions at all cusps of any modular form of integral or half-integral weight.

A complementary package using modular symbols is used in [17] by Karim Belabas, Dominique Bernardi and Bernadette Perrin-Riou to compute Manin's constant and the modular degree of elliptic curves defined over $\mathbb{Q}$.

6.15. L-functions

Participant: Henri Cohen.

In [29], H. Cohen gives an overview of Computational Number Theory in Relation with L-Functions, both in the local case (counting points on varieties over finite fields, involving in particular a detailed study of Gauss and Jacobi sums), and in the global case (for instance Dirichlet L-functions, involving in particular the study of inverse Mellin transforms). He also gives a number of little-known but very useful numerical methods, usually but not always related to the computation of L-functions.

6.16. Number fields

Participant: Henri Cohen.

In https://hal.inria.fr/hal-01379473/, H. Cohen and F. Thorne give explicit formulas for the Dirichlet series generating function of D_ℓ-extensions of odd prime degree ℓ with given quadratic resolvent.

MAGIQUE-3D Project-Team

6. New Results

6.1. Seismic Imaging and Inverse Problems

6.1.1. Shape-reconstruction and parameter identification of an elastic object immersed in a fluid

Participants: Izar Azpiroz Iragorri, Hélène Barucq, Julien Diaz.

We have developed a procedure to reconstruct the shape and material parameters of an elastic obstacle immersed in a fluid medium from some external measurements given by the so called far-field pattern. It is a nonlinear and ill-posed problem which is solved by applying a Newton-like iterative method involving the Fréchet derivatives of the scattered field. These derivatives express the sensitivity of the scattered field with respect to the parameters of interest. They are defined as the solution of boundary value problems which differ from the direct one only at the right-hand sides level. We have been able to establish the well-posedness of each problem in the case of a regular obstacle and it would be interesting in the near future to extend those results to the case of scatterers with polygonal boundaries. It requires to work with less regular Sobolev spaces for which the definition of traces is not obvious. We have also provided an analytical representation of the Fréchet derivatives in the case of a circle.

Next, we have introduced a series of numerical experiments that have been performed by applying two algorithms which propose two strategies of full reconstruction regarding the material parameters are retrieved simultaneously with the shape or not. It turns out that both work similarly delivering the same level of accuracy but the simultaneous reconstruction requires less iterations. We have thus opted for retrieving all the parameters simultaneously. Since realistic configurations include noisy data, we have performed some simulations for the reconstruction of the shape along with the Lamé coefficients for different noise levels. Other interesting experiments have been carried out using a multistage procedure where the parameters of interest are the density of the solid interior, the shape of the obstacle and its position. We have considered the case of Limited Aperture Data in back-scattering configurations, using multiple incident plane waves, mimicing a physical disposal of non-destructive testing.

We extended the solution methodology to the case of anisotropic media. Since the impact of some of the anisotropic parameters on the FFP is even weaker than the Lamé coefficients, the reconstruction of these parameters together with the shape parameters requires several frequencies and carefully adapted regularization parameters. It is in particular difficult to retrieve the Thomsen parameters ϵ and δ because their reconstruction requires to have an accurate adjustment on the rest of material and shape parameters. The recovery process is thus computationally intensive and some efforts should be done in the near future to decrease the computational costs. We were able to recover all the anisotropic parameters when the shape were assumed to be known. However, when trying to recover both shape and material parameters, we could only recover the shape and some of the physical parameters (namely the three most important ones : the density and the two velocities V_p and V_s).

These results have been obtained in collaboration with Rabia Djellouli (California State University at Northridge, USA) and are presented in Izar Azpiroz Ph.D thesis [1]

6.1.2. Time-harmonic seismic inverse problem with dual-sensors data

Participants: Hélène Barucq, Florian Faucher.

We study the inverse problem for the time-harmonic acoustic wave equation. The seismic context implies restrictive set of measurements: it consists of reflection data (resulting from an artificial source) acquired from the near surface area only. The inverse problem aims at recovering the subsurface medium parameters and we use the Full Waveform Inversion (FWI) method, which defines an iterative minimization algorithm of the difference between the measurement and simulation.

We investigate the use of new devices that have been introduced in the acoustic setting. They are able to capture both the pressure field and the vertical velocity of the waves and are called *dual-sensors*. For solving the inverse problem of interest, we define a new cost function, adapted to these two-components data. We first note that the stability of the problem can be shown to be Lipschitz, assuming the parameters to be piecewise linear.

The usefulness of the cost function is to allow a separation between the observational and numerical sources. Therefore, the numerical sources do not have to coincide with the observational ones, offering new possibilities to create adapted computational acquisitions, and possibilities to reduce the numerical burden. We illustrate our approach with three-dimensional medium reconstructions, where we start with minimal information on the target models.
This work is a collaboration with Giovanni Alessandrini (Università di Trieste), Maarten V. de Hoop (Rice University), Romina Gaburro (University of Limerick) and Eva Sincich (Università di Trieste). It has been presented in the GDR-Meca Wave conference [31].

6.1.3. Stability and convergence analysis for seismic depth imaging using Full Waveform Inversion

Participants: Hélène Barucq, Florian Faucher.

We study the convergence of the inverse problem associated with the time-harmonic wave equations. In the context of seismic, the inverse problem uses reflection data which can only be obtained from the near surface area. We consider the propagation of waves in a domain Ω and the forward problem is defined from the displacement vector field u, solution to

$$-\rho\omega^2 u - \nabla \cdot \sigma = g, \quad \text{in } \Omega, \tag{2}$$

where g stands for the source, ρ is the density and σ the stress tensor. The inverse problem aims at the recovery of the medium parameters (contained in the stress tensor) and can be solved with an iterative minimization algorithm: this is the Full Waveform Inversion (FWI) method. We study the convergence of the minimization by introducing the framework of *Finite Curvature/Limited Deflection* (FC/LD) problems. The idea is to obtain the FC/LD properties by restricting the model space to guarantee *strictly quasiconvex* attainable set. It allows us to numerically estimate the size of the basin of attraction depending on characteristics of the inverse problem such as the frequency or the geometry of the target. In particular, it allows a quantitative comprehension of frequency progression during the iterative scheme, which is an aspect that appeared mostly intuitive. It also allows a comparison of methods from a convergence point of view. This analysis is to relate with stability estimates in order to provide a consistent scheme where frequency progression is justified from the quantitative estimates. Eventually, we illustrate our approach with elastic medium reconstructions, starting from minimal information on the initial models; this also serves to illustrate the numerical requirement of the large scale optimization seismic experiments.
This work is a collaboration with Guy Chavent (Inria Rocquencourt) and Henri Calandra (TOTAL). The results have been presented in the conference "Reconstruction Methods for Inverse Problems" [15].

6.1.4. Quantitative localization of small obstacles with single-layer potential fast solvers

Participants: Hélène Barucq, Florian Faucher, Ha Howard Faucher.

In this work, we numerically study the inverse problem of locating small circular obstacles in a homogeneous medium using noisy backscattered data collected at several frequencies. The main novelty of our work is the implementation of a single-layer potential based fast solver (called FSSL) in a Full-Waveform inversion procedure, to give high quality reconstruction with low-time cost. The efficiency of FSSL was studied in our previous works. We show reconstruction results with up to 12 obstacles in structured or random configurations with several initial guesses, all allowed to be far and different in nature from the target. This last assumption is not expected in results using nonlinear optimization schemes in general. For results with 6 obstacles, we also investigate several optimization methods, comparing between nonlinear gradient descent and quasi-Newton, as well as their convergence with different line search algorithms.

The work is published in Journal of Computational Physics [9]. This work has been presented at GDR-Meca Wave conference in Fréjus *cf.* [21].

6.1.5. Time Domain Full Waveform Inversion Adjoint Studies

Participants: Hélène Barucq, Julien Diaz, Pierre Jacquet.

Full Waveform Inversion (FWI) allows retrieving the physical parameters (e.g. the velocity, the density) from an iterative procedure underlying a global optimization technique. The recovering of the medium corresponds to the minimum of a cost function quantifying the difference between experimental and numerical data. In this study we have considered the adjoint state method to compute the gradient of this cost function. At each iteration the parameters are updated with the solution of an adjoint equation which can be defined either as the adjoint of the continuous equation or the discrete problem. Some studies have addressed the question of establishing what the best strategy is. The answer is still unclear and turns out to be strongly dependent on the problem under study.

The purpose of this study was to investigate several computations of the adjoint state as a preamble of a FWI method applied to the time-dependent acoustic wave approximated in a Discontinuous Galerkin framework involving Bernstein elements. We have considered different time schemes to feature the inherited properties of the computed adjoint state. By comparing the different discrete adjoint operators both from a mathematical and numerical point of view, we aim at defining the best option for computing the adjoint state with accuracy at least cost.

This work is a collaboration with Henri Calandra (TOTAL). It was presented at Total MATHIAS conference in Paris [28].

6.1.6. Seismic imaging of remote targets buried in an unknown media

Participant: Yder Masson.

Box Tomography:first application to the imaging of upper-mantle shear velocity and radial anisotropy structure beneath the North American continent: The EarthScope Transpotable Array (TA) deployment provides dense array coverage through- out the continental United States and with it, the opportunity for high-resolution 3-D seismic velocity imaging of the stable part of the North American (NA) upper mantle. Building upon our previous long-period waveform tomographic modeling, we present a higher resolution 3-D isotropic and radially anisotropic shear wave velocity model of the NA lithosphere and asthenosphere. The model is constructed using a combination of teleseismic and regional waveforms down to 40 s period and wavefield computations are performed using the spectral element method both for regional and teleseismic data. Our study is the first tomographic ap- plication of ‘Box Tomography’, which allows us to include teleseismic events in our inversion, while computing the teleseismic wavefield only once, thus significantly reducing the numerical computational cost of several iterations of the regional inversion. We confirm the presence of high-velocity roots beneath the Archean part of the continent, reaching 200–250 km in some areas, however the thickness of these roots is not everywhere correlated to the crustal age of the corresponding cratonic province. In particular, the lithosphere is thick (250 km) in the western part of the Superior craton, while it is much thinner (150 km) in its eastern part. This may be related to a thermomechanical erosion of the cratonic root due to the passage of the NA plate over the Great Meteor hotspot during the opening of the Atlantic ocean 200–110 Ma. Below the lithosphere, an upper-mantle low-velocity zone (LVZ) is present everywhere under the NA continent, even under the thickest parts of the craton, although it is less developed there. The depth of the minimum in shear velocity has strong lateral variations, whereas the bottom of the LVZ is everywhere relatively flat around 270–300 km depth, with minor undulations of maximum 30 km that show upwarping under the thickest lithosphere and downwarping under tectonic regions, likely reflecting residual temperature anomalies. The radial anisotropy structure is less well resolved, but shows distinct signatures in highly deformed regions of the lithosphere.

This is the first application to a real case study of a novel imaging method called "Box Tomography". These results were obtained through collaborations with Barbara Romanowicz (Berkeley Seimological Laboratory, UC Berkeley; Collège de France) and Pierre Clouzet (Institut de Physique du Globe de Paris). The results have been published in the Geophysical Journal International [11].

Additional developments are conducted in collaboration with Sevan Adourian and Barbara Romanowicz at the Berkeley Seismological Laboratory, UC Berkeley, in particular, to efficiently account for receivers located outside the imaged region. These new results have been presented in different international conferences [23], [19].

To strengthen existing collaborations, a proposal has been submitted to the France-Berkeley Fund (13000$ for travelling and living expenses). We propose a joint effort to further develop and apply a novel seismic tomographic approach, Box-Tomography, to image and characterize small scale structures of interest in the deep Earth, such as the roots of mantle plumes, ultra-low velocity zones, or the edges of large low shear velocity provinces. Our objective is to forge a long-term collaboration between applied mathematicians at Magique3D developing wave propagation modeling methods and the seismologists at the Berkeley Seismological Laboratory (UC Berkeley) using these methods to investigate the Earth's internal structure.

6.2. Mathematical modeling of multi-physics involving wave equations

6.2.1. Hybrid space discretization to solve elasto-acoustic coupling

Participants: Hélène Barucq, Julien Diaz, Aurélien Citrain.

Accurate wave propagation simulations require selecting numerical schemes capable of taking features of the medium into account. In case of complex topography, unstructured meshes are the most adapted and in that case, Discontinuous Galerkin Methods (DGM) have demonstrated great performance. Off-shore exploration involves propagation media which can be well represented by hybrid meshes combining unstructured meshes with structured grids that are best for representing homogeneous media like water layers. Then it has been shown that Spectral Element Methods (SEM) deliver very accurate simulations on structured grids with much lower computational costs than DGMs.

We have developed a SEM-DG numerical method for solving time-dependent elasto- acoustic wave problems. We consider the first-order coupled formulation for which we propose a SEM-DG formulation which turns out to be stable. In the 2D case, the coupling is quite straightforward due to the natural way of mixing triangles with quadrangles. 3D coupling is much more difficult and the interface between tetrahedra and hexahedra deserves a particular attention.

These results have been obtained in collaboration with Henri Calandra(TOTAL) and Christian Gout (INSA Rouen) and have been presented at the Fifth International congress on multiphysics, multiscale and optimization problems in Bilbao, the 13th World Congress on Computational Mecanics in New-York and MATHIAS conference in Paris [16], [17], [24].

6.2.2. Signal and noise in helioseismic holography

Participant: Hélène Barucq.

Helioseismic holography is an imaging technique used to study heterogeneities and flows in the solar interior from observations of solar oscillations at the surface. Holograms contain noise due to the stochastic nature of solar oscillations. Aims. We provide a theoretical framework for modeling signal and noise in Porter-Bojarski helioseismic holography. Methods. The wave equation may be recast into a Helmholtz-like equation, so as to connect with the acoustics literature and define the holography Green's function in a meaningful way. Sources of wave excitation are assumed to be stationary, horizontally homogeneous, and spatially uncorrelated. Using the first Born approximation we calculate holograms in the presence of perturbations in sound-speed, density, flows, and source covariance, as well as the noise level as a function of position. This work is a direct extension of the methods used in time-distance helioseismology to model signal and noise. Results. To illustrate the theory, we compute the hologram intensity numerically for a buried sound-speed perturbation at different depths in the solar interior. The reference Green's function is obtained for a spherically-symmetric solar model using a finite-element solver in the frequency domain. Below the pupil area on the surface, we find that the spatial resolution of the hologram intensity is very close to half the local wavelength. For a sound-speed perturbation of size comparable to the local spatial resolution, the signalto-noise ratio is approximately constant with depth. Averaging the hologram intensity over a number N of frequencies above 3 mHz increases

the signal-to-noise ratio by a factor nearly equal to the square root of N. This may not be the case at lower frequencies, where large variations in the holographic signal are due to the individual contributions of the long-lived modes of oscillation. This work has been done in collaboration with Laurent Gizon, Damien Fournier, Dan Yang and Aaron C. Birch of the Max-Planck-Institut für Sonnensystemforschung at Göttingen (Germany) and published in Astronomy and Astrophysics [14]

6.2.3. *Sensitivity kernels for time-distance helioseismology. Efficient computation for spherically symmetric solar models*

Participant: Hélène Barucq.

The interpretation of helioseismic measurements, such as wave travel-time, is based on the computation of kernels that give the sensitivity of the measurements to localized changes in the solar interior. These kernels are computed using the ray or the Born approximation. The Born approximation is preferable as it takes finite-wavelength effects into account, although it can be computationally expensive. Aims.We propose a fast algorithm to compute travel-time sensitivity kernels under the assumption that the background solar medium is spherically symmetric. Methods. Kernels are typically expressed as products of Green's functions that depend upon depth, latitude, and longitude. Here, we compute the spherical harmonic decomposition of the kernels and show that the integrals in latitude and longitude can be performed analytically. In particular, the integrals of the product of three associated Legendre polynomials can be computed. Results. The computations are fast and accurate and only require the knowledge of the Green's function where the source is at the pole. The computation time is reduced by two orders of magnitude compared to other recent computational frameworks. Conclusions. This new method allows flexible and computationally efficient calculations of a large number of kernels, required in addressing key helioseismic problems. For example, the computation of all the kernels required for meridional flow inversion takes less than two hours on 100 cores. This work has been done in collaboration with Damien Fournier, Chris S. Hanson and Laurent Gizon of the Max-Planck-Institut für Sonnensystemforschung at Göttingen (Germany) and published in Astronomy and Astrophysics [13]

6.2.4. *Characterization of partial derivatives with respect to material parameters in a fluid-solid interaction problem.*

Participants: Izar Azpiroz Iragorri, Hélène Barucq, Ha Howard Faucher.

For a fluid-solid interaction problem with Lipschitz interface, we investigate the partial Fréchet differentiability of the solutions and the approximate far-field-pattern with respect to solid material parameters. Differentiability is shown in standard Sobolev framework, and the derivatives are characterized as solutions to inhomogeneous fluid-solid transmission problems. To validate the accuracy of the characterization, we compare analytical values with numerical ones given by Interior Penalty Discontinuous Galerkin (IPDG) in a setting with circular obstacles. Our comparisons also show that IPDG gives results with high precision and incurs almost no effect of discretization error accumulation. This work has been done in colllaboration with Rabia Djellouli (california State University at Northridge, USA). It has been published in [3].

6.2.5. *Asymptotic modeling of multiple electromagnetic scattering problems by small obstacles*

Participants: Justine Labat, Victor Péron, Sébastien Tordeux.

The detection of small conductive heterogeneities in three dimensional domains by non-destructive electromagnetic imaging is a real challenge. Basic finite element-based methods are very expensive in terms of computation time and memory burden, since they involve a huge number of degrees of freedom when the obstacles are very small compared to the testing wavelength. Using the matched asymptotic expansions method, we have developed a meshless reduced model, which consists of replacing the scatterers by equivalent point sources. This method has been numerically implemented in Matlab and its accuracy validated with analytical solutions in spherical geometries. The details of the results are given in [35] and were presented at the fifth International Congress on Multiphysics, Multiscale and Optimization Problems in Bilbao [18] and at ECCOMAS conferences in Glasgow [22]. Following the Born and Foldy-Lax models, we can extend the results for one obstacle to the multiple scattering problem, thus provide meshless methods in this case. Numerical simulations with thousands of small scatterers, up to 10000, were presented at the seminar of RWTH Aachen University [40].

6.2.6. Discontinuous Galerkin Trefftz type method for solving the Maxwell equations

Participant: Sébastien Tordeux.

Trefftz type methods have been developed in Magique 3D to solve Helmholtz equation. Theese methods reduce the numerical dispersion and the condition number of the linear system. This work aims in pursuing this development for electromagnetic scattering. We have adapted and tested the method for an academical 2D configuration. This work has been achieved in the context of the Master trainee of Hakon Fure in collaboration with Sébastien Pernet of ONERA Toulouse.

6.2.7. Comparison between Galbrun and linearized Euler models in the context of helioseismology

Participants: Hélène Barucq, Juliette Chabassier, Marc Duruflé, Nathan Rouxelin.

Helioseismology aims to probe the Sun's internal structure thanks to surface observations and our knowledge of acoustic wave propagation. In this work we focus on modeling and simulating the propagation of waves below the surface of the Sun.

In the first part, we establish the equations for acoustic wave propagation by linearizing the Euler equations describing the fluid flow. We then compare two linearization processes based on the eulerian and lagrangian description of fluid dynamics.

In the second part, we solve those equations in time-harmonic domain using high order Discontinuous Galerkin methods. Those numerical methods seem to lack consistency and stability when applied to our problems. Specifically, we notice the presence of spurious modes in our numerical solutions.

To fully understand those results further investigations are needed. In particular, two questions seem to stand out : Is the acoustic wave propagation problem in time-harmonic domain well posed for a recirculating background flow ? Is this approach valid ? Can we really assume that the solar plasma solves the Euler equations ?

6.2.8. Asymptotic Models for the Electric Potential across a Highly Conductive Casing

Participant: Victor Péron.

We analyze a configuration that involves a steel-cased borehole, where the casing that covers the borehole is considered as a highly conductive thin layer. We develop an asymptotic method for deriving reduced problems capable of efficiently dealing with the numerical difficulties caused by the casing when applying traditional numerical methods. We derive several reduced models by employing two different approaches, each of them leading to different classes of models. We prove stability and convergence results for these models. The theoretical orders of convergence are supported by numerical results obtained with the finite element method. These results have been obtained with D. Pardo (UPV/EHU, BCAM, Ikerbasque) and Aralar Erdozain. It was published in Computers and Mathematics with Applications [12].

6.2.9. Boundary Element Method for 3D Conductive Thin Layer in Eddy Current Problems

Participant: Victor Péron.

Thin conducting sheets are used in many electric and electronic devices. Solving numerically the eddy current problems in presence of these thin conductive sheets requires a very fine mesh which leads to a large system of equations, and becoming more problematic in case of high frequencies. In this work we show the numerical pertinence of asymptotic models for 3D eddy current problems with a conductive thin layer of small thickness based on the replacement of the thin layer by its mid-surface with impedance transmission conditions that satisfy the shielding purpose, and by using an efficient discretization with the Boundary Element Method in order to reduce the computational cost. These results have been obtained in collaboration with M. Issa, R. Perrussel and J-R. Poirier (LAPLACE, CNRS/INPT/UPS, Univ. de Toulouse) and O. Chadebec (G2Elab, CNRS/INPG/UJF, Institut Polytechnique de Grenoble). This work has been accepted for publication in COMPEL - The international journal for computation and mathematics in electrical and electronic engineering. This work has been presented in the symposium IABEM 2018.

6.2.10. Model-based digital pianos: from physics to sound synthesis

Participant: Juliette Chabassier.

A review article has been published in IEEE Signal Processing Magazine on model-based digital pianos in collaboration with Balasz Bank [4] .

6.2.11. The virtual workshop : towards versatile optimal design of musical wind instruments for the makers

Participants: Juliette Chabassier, Robin Tournemenne.

Our project aims at proposing optimization solutions for wind instrument making. Our approach is based on a strong interaction with makers and players, aiming at defining interesting criteria to optimize from their point of view. After having quantified those criteria under the form of a cost function and a design parameters space, we wish to implement state-of-the-art numerical methods (finite elements, full waveform inversion, neuronal networks, diverse optimization techniques...) that are versatile (in terms of models, formulations, couplings...) in order to solve the optimization problem. More precisely, we wish to take advantage of the fact that sound waves in musical instruments satisfy the laws of acoustics in pipes (PDE), which gives us access to the full waveform inversion technique, usable in harmonic or temporal regime. The methods that we want to use are attractive because the weekly depend on the chosen criterion, and they are easily adaptable to various physical situations (multimodal decomposition in the pipe, coupling with the embouchure, ...), which can therefore be modified a posteriori. The goal is to proceed iteratively between instrument making and optimal design (the virtual workshop) in order to get close to tone quality related and playability criteria. In 2018 we have implemented a python 3 toolbox named OpenWind that includes the first simulation module. Next modules will be implemented next year. This work has been presented to the Congrès Français d'Acoustique [26].

6.2.12. Collaboration with Augustin Humeau, wind instrument maker

Participants: Juliette Chabassier, Robin Tournemenne.

We have initiated a strong collaboration with Augustin Humeau, bassoon maker in Dordogne, France. The goal is to develop practical tools for instrument design, in the realistic context of an artisanal workshop. Until now, an input impedance measurement setup has been developed in collaboration with Samuel Rodriguez, I2M Univ. Bordeaux. It is based on the use of five microphones and the need of one calibration. It has been specifically adapted to the small entrances of some wind instruments (bassoon, oboes). We have attended the JFIS (journées facture instrumentale et science) in November 2018, Le Mans, where the approach has been presented and demonstrated. Given the great interest showed by other instrument makers attendanting the conference, the future of this collaboration is in discussion and may integrate an Inria startup process.

6.2.13. Optimization of brass wind instruments based on sound simulations

Participant: Robin Tournemenne.

We exploited a new optimization method of the inner shape of brass instruments using sound simulations to derive objective functions. The novelties are the obtention of optimal bores for objective functions representative of the intonation but also of the spectrum of the instrument, and the possibility to include constraints in the optimization problem. A complete physics-based model, taking into account the instrument and the musician's embouchure, is used, in order to simulate sounds' permanent regimes using the harmonic balance technique. The instrument is modeled by its input impedance computed with the transfer matrix method under plane wave propagation and visco-thermal losses. Some embouchure's parameters remain variable during the optimization procedure in order to get the average behavior of the instrument. The design variables are the geometrical dimensions of the resonator. Given the computationally expensive function evaluation and the unavailability of gradients, a surrogate-assisted optimization framework is implemented using the mesh adaptive direct search algorithm (MADS). Two optimization examples of a Bb trumpet's bore (with two and ten design optimization variables) demonstrate the effectiveness of the approach. Results show that solvers can deal flawlessly with high dimensional problems, under constraints, improving significantly the value of the objective functions.

6.2.14. Energy based model and simulation in the time domain of linear acoustic waves in a radiating pipe

Participants: Juliette Chabassier, Robin Tournemenne.

We model in the time domain linear acoustic waves in a radiating pipe without damping. The acoustic equations system in formulated in flow and pressure, which leads to a first order space time equations system. The radiation condition is also written as a first order in time equation, and is parametrized by two real coefficients. Moreover, an auxiliary variable is introduced at the radiating boundary. The choice of this variable is adapted to the considered source type in order to ensure the model stability by energy techniques, under some conditions on the radiating condition. We then propose a stable space time explicit discretization, which ensures the dissipation of a discrete energy. The novelty of the discretization lies, on the one hand, in the variational nature of the space approximation (which leads to arbitrary order finite elements with no required matrix inversion), and on the other hand, on the definition of the auxiliary variable for any acoustic source type (which leads to the decay of a well defined energy). Finally, we quantify the frequential domain of validity of the used radiation condition by comparison with theoretical and experimental models of the literature. This is a collaboration with Morgane Bergot (Université Claude Bernard, Lyon 1). An article has been written and will be submitted soon. This work has been presented to the Congrès Français d'Acoustique [25].

6.2.15. Computation of the entry impedance of a dissipative radiating pipe

Participants: Juliette Chabassier, Robin Tournemenne.

Modeling the entry impedance of wind instruments pipes is essential for sound synthesis or instrument qualification. We study this modeling with the finite elements method in one dimension (FEM1D) and with the more classically used transfer matrix method (TMM). The TMM gives an analytical formula of the entry impedance depending on the bore (intern geometry of the instrument) defined as a concatenation of simple elements (cylinders, cones, etc). The FEM1D gives the entry impedance for any instrument geometry. The main goals of this work are to assess the viability of the FEM1D and to study and analyse the approximations necessary for the TMM in dissipative pipes. First, lossless Webster's equation in one dimension is studied with arbitrary radiation conditions. In this context and for cylinders or cones, the TMM is exact. We verify that the error made with FEM1D for fine enough elements is as small as desired. When we consider viscothermal losses, the TMM does not solve the classical Kirchhoff model because two terms are supposed constant. In order to overcome this model approximation, simple segments, on which are based the TMM, are decomposed into much smaller segments. We show that using the TMM actually amounts to solving a different equation than the original one, on each small segment. The FEM1D does not necessitate any model approximation, and it is possible to show that it solves the dissipative equation with any arbitrarily small error. With this in hand, we can quantify the TMM model approximation error. The methods are compared in terms of accuracy and computational burden. On realistic cases as the case of a trumpet, the FEM show a better efficiency. Moreover, unusual phenomena as a non constant air temperature can easily be tackle with the FEM. An article has been written and will be submitted soon. This work has been presented to the Congrès Français d'Acoustique [27].

6.2.16. Seismic wave propagation in carbonate rocks at the core scale

Participants: Julien Diaz, Florian Faucher, Chengyi Shen.

Reproduction of large-scale seismic exploration at lab-scale with controllable sources is a promising approach that could not only be applied to study small-scale physical properties of the medium, but also contribute to significant progress in wave-propagation understanding and complex media imaging at exploration scale via upscaling methods. We propose to apply a laser-generated seismic source for lab-scale new geophysical experiments. This consists in generating seismic waves in various media by well-calibrated pulsed-laser impacts and measuring precisely the wavefield (displacement) by Laser Doppler Vibrometer. Parallel 2D/3D simulations featuring the Discontinuous Galerkin discretization method with Interior Penalties (IPDG) are done to match the experimental data. The IPDG method is of particular interest when it comes to solve wave propagation problems in highly heterogeneous media, such as the limestone cores that we are studying.

Current seismic data allowed us to retrieve V_p tomography slices. Further more, qualitative/quantitative comparisons between simulations and experimental data validated the experiment protocol and vice-versa the numerical schemes, opening the possibility of performing FWI on these high resolution data.

This work is in collaboration with Clarisse Bordes, Daniel Brito and Deyuan Zhang (LFCR, UPPA) and with Stéphane Garambois (ISTerre). It was presented at conference AGU [42].

6.3. Supercomputing for Helmholtz problems

6.3.1. Numerical libraries for hybrid meshes in a discontinuous Galerkin context

Participants: Hélène Barucq, Aurélien Citrain, Julien Diaz.

Elasticus team code has been designed for triangles and tetrahedra mesh cell types. The first part of this work was dedicated to add quadrangle libraries and then to extend them to hybrid triangles-quadrangles (so in 2D). This implied to work on polynomials to form functions basis for the (discontinuous) finite element method, to finally be able to construct reference matrices (mass, stiffness, ...).

A complementary work has been done on mesh generation. The goal was to encircle an unstructured triangle mesh, obtained by third-party softwares, with a quadrangle mesh layer. At first, we built scripts to generate structured triangle meshes, quadrangle meshes and hybrid meshes (triangles surrounded by quadrangles). In 2018, we have implemented the coupling between Discontinuous Galerkin methods (using the triangles/tetrahedra) and Spectral Element methods (using quadrangles/hexahedra). We have also implemented the PML in the SEM part, and we are now working on the local time-stepping feature.

6.4. Hybrid time discretizations of high-order

6.4.1. Space-Time Discretization of Elasto-Acoustic Wave Equation in Polynomial Trefftz-DG Bases

Participants: Elvira Shishenina, Hélène Barucq, Julien Diaz.

In the context of the strategic action "Depth Imaging Partnership" between Inria and Total we have investigated to the development of an explicit Trefftz-DG formulation for elasto-acoustic problem, solving the global sparse matrix by constructing an approximate inverse obtained from the decomposition of the global matrix into a block-diagonal one. The inversion is then justified under a CFL-type condition. This idea allows for reducing the computational costs but its accuracy is limited to small computational domains. According to the limitations of the method, we have investigated the potential of Tent Pitcher algorithms following the recent works of Gopalakrishnan et al. It consists in constructing a space-time mesh made of patches that can be solved independently under a causality constraint. We have obtained very promising numerical results illustrating the potential of Tent Pitcher in particular when coupled with a Trefftz-DG method involving only surface terms. In this way, the space-time mesh is composed of elements which are 3D objects at most. It is also worth noting that this framework naturally allows for local time-stepping which is a plus to increase the accuracy while decreasing the computational burden. The results of this work have been published in the Applicable Analysis Journal [6], in the book of proceedings for European Conference on Numerical Mathematics and Advanced Applications (ENUMATH 2017) (due date April 27, 2019), and in the PhD thesis [2], as well as presented during the International Conference on Spectral and High-Order Methods (ICOSAHOM 2018, London - UK), the 13$^{\text{th}}$ World Congress on Computational Mechanics (WCCM 2018, New-York - USA), and during the annual seminar on Computational Science Engineering and Data Science organized by TOTAL (MATHIAS 2018, Serris - France).

6.4.2. Performance analysis of local time-stepping schemes for wave propagation

Participants: Julien Diaz, Rose-Cloé Meyer.

The efficiency of numerical simulation of wave propagation is highly dependent of the quality of the mesh. For complex simulations, the size of the cells in the mesh can strongly vary, either because of the geometry or because of the different propagation celerity of the waves. To ensure stability, explicit numerical schemes must match with the CFL conditions of every cells of the mesh. When significant disparities appear in the domain, the time step used on big cells is not optimal, which can cause heavy calculation cost and result in a loss of efficiency. To improve the performance of the programs, local time-stepping methods based on a spatial Discontinuous Galerkin discretization have been implemented. In this work, we compared three local time-stepping methods: a conservative method, a recursive method, and an asynchron method. The two first methods use local time steps that are fractions of the global time step, while the third method can use independent time steps on each cell of the mesh. The accuracy of the solution, the computation cost and the speedup of local-time stepping are presented on cases in two and three dimensions on configurations as fine slot or domains with geometric singularities. The results are presented in Rose-Cloé Meyer Master thesis [41]. This work has been achieved in collaboration with Guillaume Dufour and Xavier Ferrières (Onera)

6.4.3. *Construction and analysis of a fourth order, energy preserving, explicit time discretization for dissipative linear wave equations.*

Participants: Juliette Chabassier, Julien Diaz.

We submitted a paper to M2AN [37]. This paper deals with the construction of a fourth order, energy preserving, explicit time discretization for dissipative linear wave equations. This family of schemes is obtained by replacing the inversion of a matrix, that comes naturally after using the technique of the Modified Equation on the second order Leap Frog scheme applied to dissipative linear wave equations, by an explicit approximation of its inverse. The series can be truncated at different orders, which leads to several schemes. The stability of the schemes is studied. Numerical results in 1D illustrate the good behavior regarding space/time convergence and the efficiency of the newly derived scheme compared to more classical time discretizations. A loss of accuracy is observed for non smooth profiles of dissipation, and we propose an extension of the method that fixes this issue. Finally, we assess the good performance of the scheme for a realistic dissipation phenomenon in Lorentz's materials. This work has been done in collaboration with Sébastien Imperiale (Inria Project-Team M3DISIM).

6.4.4. *Construction and convergence analysis of conservative second order local time discretisation for wave equations*

Participant: Juliette Chabassier.

In this work we present and analyse a time discretisation strategy for linear wave propagation that aims at using locally in space the most adapted time discretisation among a family of implicit or explicit centered second order schemes. The domain of interest being decomposed into several regions, different time discretisations can be chosen depending on the local properties of the spatial discretisations (mesh size and quality) or the physical parameters (high wave speed, low density). We show that, under some conditions on the time step, the family of time discretisations obtained combined with standard finite elements methods in space ensures a second order space-time convergence. This work has been done in collaboration with Sébastien Imperiale (Inria Project-Team M3DISIM). It has been submitted to Numerische Mathematik.

6.4.5. *High-order locally implicit time schemes for linear ODEs*

Participants: Hélène Barucq, Marc Duruflé, Mamadou N'Diaye.

In this work we have proposed a method that combines optimized explicit schemes and implicit schemes to form locally implicit schemes for linear ODEs, including in particular ODEs coming from the space discretization of wave propagation phenomena. This method can be applied to the following ODE

$$M_h \frac{dU}{dt} = K_h U + F(t)$$

Like in the local time-stepping developed by Grote and co-workers, the computational domain is split into a fine region and a coarse region. The matrix A_h is given as

$$A_h = M_h^{-1} K_h = A_h P + A_h (I - P)$$

where P is the projector on the fine region of the computational domain. Then the proposed locally implicit method is obtained from the combination of the A-stable implicit schemes we have developed in 2016 (Padé schemes or Linear-SDIRK schemes detailed in [8]) on the fine region and explicit schemes with optimal CFL number in the coarse region. The developed method has been used to solve the acoustic wave equation and we have checked the convergence in time of these schemes for order 4, 6 and 8.

This work is a chapter of the thesis defended by Mamadou N'diaye under the joint supervision of Hélène Barucq and Marc Duruflé. In 2018, the implemented method has been parallelized in Montjoie and 3-D numerical results have been obtained. An article is in preparation.

MANAO Project-Team

7. New Results

7.1. Analysis and Simulation

7.1.1. Visual Features in the Perception of Liquids

Perceptual constancy—identifying surfaces and objects across large image changes—remains an important challenge for visual neuroscience. Liquids are particularly challenging because they respond to external forces in complex, highly variable ways, presenting an enormous range of images to the visual system. To achieve constancy, the brain must perform a causal inference that disentangles the liquid's viscosity from external factors—like gravity and object interactions—that also affect the liquid's behavior. Here, we tested whether the visual system estimates viscosity using "midlevel" features that respond more to viscosity than other factors. Our findings demonstrate that the visual system achieves constancy by representing stimuli in a multidimensional feature space—based on complementary, midlevel features—which successfully cluster very different stimuli together and tease similar stimuli apart, so that viscosity can be read out easily.

7.1.2. Teaching Spatial Augmented Reality: a Practical Assignment for Large Audiences

We conceived a new methodology to teach spatial augmented reality in a practical assignment to large audiences. Our approach does not require specific equipment such as video projectors while teaching the principal topics and difficulties involved in spatial augmented reality applications, and especially calibration and tracking. The key idea is to set up a scene graph consisting of a 3D scene with a simulated projector that "projects" content onto a virtual representation of the real-world object. For illustrating the calibration, we simplify the intrinsic parameters to using the field of view, both for the camera and the projector. For illustrating the tracking, instead of relying on specific hardware or software, we exploit the relative transformations in the scene graph.

7.2. From Acquisition to Display

7.2.1. Comparison of Plenoptic Imaging Systems

Plenoptic cameras provide single-shot 3D imaging capabilities, based on the acquisition of the Light-Field, which corresponds to a spatial and directional sampling of all the rays of a scene reaching a detector. Specific algorithms applied on raw Light-Field data allow for the reconstruction of an object at different depths of the scene. Two different plenoptic imaging geometries have been reported, associated with two reconstruction algorithms: the traditional or unfocused plenoptic camera, also known as plenoptic camera 1.0, and the focused plenoptic camera, also called plenoptic camera 2.0. Both systems use the same optical elements, but placed at different locations: a main lens, a microlens array and a detector. These plenoptic systems have been presented as independent. We have demonstrated the continuity between them, by simply moving the position of an object. We have also compared the two reconstruction methods. We have finally theoretically shown that the two algorithms are intrinsically based on the same principle and could be applied to any Light-Field data. However, the resulting images resolution and quality depend on the chosen algorithm.

7.2.2. Capturing Illumination for Augmented Reality using RGB-D Images

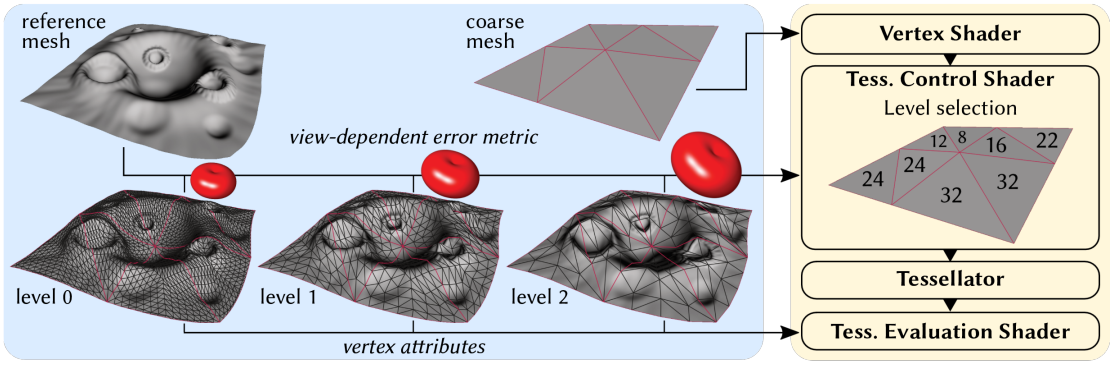
RGB-D sensors is becoming more and more available. We have proposed an automatic framework to recover the illumination (from light sources both in and out of the camera's view) of indoor scenes based on a single RGB-D image. Unlike previous works, our method can recover spatially varying illumination without using any lighting capturing devices or HDR information. The recovered illumination can produce realistic rendering results. Using the estimated light sources and geometry model, environment maps at different points in the scene are generated that can model the spatial variance of illumination. The experimental results have demonstrated the validity of our approach and the possibilities offered to Augemented Reality by the use of more dedicated hardware.

7.2.3. Diffraction Removal in an Image-based BRDF Measurement Setup

Material appearance is traditionally represented through its Bidirectional Reflectance Distribution Function (BRDF), quantifying how incident light is scattered from a surface over the hemisphere. To speed up the measurement process of the BRDF for a given material, which can necessitate millions of measurement directions, image-based setups are often used for their ability to parallelize the acquisition process: each pixel of the camera gives one unique configuration of measurement. With highly specular materials, the High Dynamic Range (HDR) imaging techniques are used to acquire the whole BRDF dynamic range, which can reach more than 10 orders of magnitude. Unfortunately, HDR can introduce star-burst patterns around highlights arising from the diffraction by the camera aperture. Therefore, while trying to keep track on uncertainties throughout the measurement process, one has to be careful to include this underlying diffraction convolution kernel. A purposely developed algorithm is used to remove most part of the pixels polluted by diffraction, which increase the measurement quality of specular materials, at the cost of discarding an important amount of BRDF configurations (up to 90% with specular materials). Finally, our setup succeed to reach a 1.5 degree median accuracy (considering all the possible geometrical configurations), with a repeatability from 1.6% or the most diffuse materials to 5.5% for the most specular ones. Our new database, with their quantified uncertainties, will be helpful for comparing the quality and accuracy of the different experimental setups and for designing new image-based BRDF measurement devices.

7.3. Rendering, Visualization and Illustration

7.3.1. A View-Dependent Metric for Patch-Based LOD Generation & Selection



(a) LOD generation **(b)** GPU evaluation

*Figure 8. Full processing pipeline — **(a)** as a pre-process, the LOD is generated from the reference mesh by decimation and, for each patch, at each level, its approximation error with respect to the reference surface is summarized into a compact view-dependent metric, **(b)** these are then used during hardware tessellation to select the most appropriate patch level according to the current viewing distance and direction.*

With hardware tessellation, highly detailed geometric models are decomposed into patches whose tessellation factor can be specified dynamically and independently at render time to control polygon resolution. Yet, to achieve maximum efficiency, an appropriate factor needs to be selected for each patch according to its content (geometry and appearance) and the current viewpoint distance and orientation. We proposed [4] a novel patch-based error metric that addresses this problem (Fig. 8). It summarizes both the geometrical error and the texture parametrization deviation of a simplified patch compared to the corresponding detailed surface. This metric is compact and can be efficiently evaluated on the GPU along any view direction. Furthermore, based

on this metric, we devise an easy-to-implement refitting optimization that further reduces the simplification error of any decimation algorithm, and propose a new placement strategy and cost function for edge-collapses to reach the best quality/performances trade-off.

7.3.2. MNPR: A Framework for Real-Time Expressive Non-Photorealistic Rendering of 3D Computer Graphics

Figure 9. A 3D scene rendered through MNPR in watercolor, oil and charcoal styles. Baba Yaga's hut model ©Inuciian.

We developed [12] a framework for expressive non-photorealistic rendering of 3D computer graphics: MNPR. Our work focuses on enabling stylization pipelines with a wide range of control, thereby covering the interaction spectrum with real-time feedback. In addition, we introduce control semantics that allow cross-stylistic art-direction, which is demonstrated through our implemented watercolor, oil and charcoal stylizations (Fig. 9). Our generalized control semantics and their style-specific mappings are designed to be extrapolated to other styles, by adhering to the same control scheme. We then share our implementation details by breaking down our framework and elaborating on its inner workings. Finally, we evaluate the usefulness of each level of control through a user study involving 20 experienced artists and engineers in the industry, who have collectively spent over 245 hours using our system. MNPR is implemented in Autodesk Maya and open-sourced through this publication, to facilitate adoption by artists and further development by the expressive research and development community.

7.4. Editing and Modeling

7.4.1. Interactive optimal transport solver

Optimal transport is a fundamental tool that appeared in various forms in numerous application domains. We developed a novel and extremely fast algorithm to compute continuous transport maps between 2D probability densities discretized on uniform grids. It follows the Monge-Ampère formulation, and it converges in a few cheap iterations thanks to the novel derivative-free non-linear solver we developed along this work. We achieve interactive performance in various applications such as blue noise sampling, feature sensitive remeshing, and caustic design (Fig. 10).

7.4.2. A Composite BRDF Model for Hazy Gloss

A new bidirectional reflectance distribution function (BRDF) model is introduced for the rendering of materials that exhibit hazy reflections, whereby the specular reflections appear to be flanked by a surrounding halo. The focus of this work is on artistic control and ease of implementation for real-time and off-line rendering. The material model is based on a pair of arbitrary BRDF models; however, instead of controlling their physical parameters, we expose perceptual parameters inspired by visual experiments. The

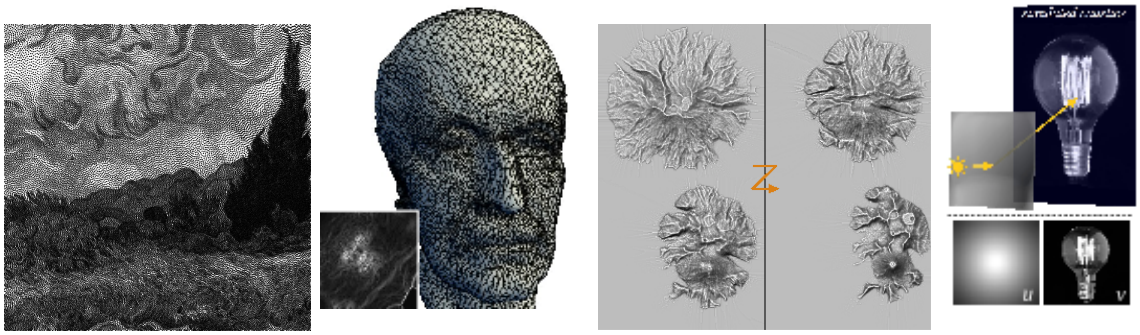


Figure 10. Our fast mass-transport solver enables many applications such as adaptive sampling, surface remeshing, heightfield morphing and caustic design with interactive performance.

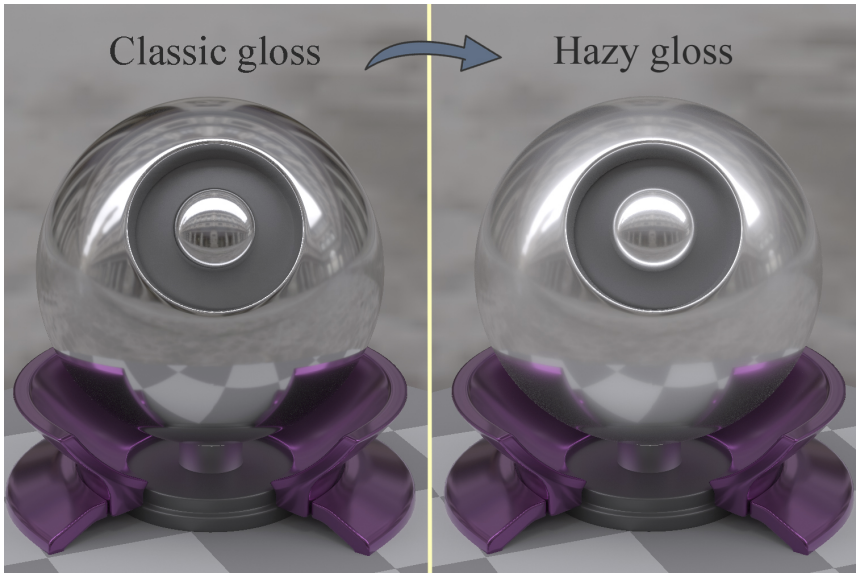


Figure 11. An object rendered with a classic glossy material (left), and with our hazy gloss material model (right), exhibiting specular reflections flanked by a halo.

main contribution then consists in a mapping from perceptual to physical parameters that ensures the resulting composite BRDF is valid in terms of reciprocity, positivity and energy conservation. The immediate benefit of this approach is to provide direct artistic control over both the intensity and extent of the haze effect (Fig. 11), which is not only necessary for editing purposes, but also essential to vary haziness spatially over an object surface.

MEMPHIS Project-Team

7. New Results

7.1. Hybrid FOM/ROM simulations for turbulent flows

We present below results concerning the application of the hybrid FOM/ROM method to a realistic problem in CFD. The purpose of the study is to investigate the behavior of the flow past a car for several front bumper configurations. We here resort to Free-Form Deformation (FFD) based on two parameters to determine a satisfactory parametrization of all possible configurations, and we consider a steady RANS solver at $\mathrm{Re} = 4.87 \cdot 10^6$ with Spalart-Allmaras turbulence model to simulate the flow. Simulations are performed in collaboration with OPTIMAD (http://www.optimad.it/) exploiting the methodology proposed in [15].

Figure 6 (left) shows the domain decomposition; in the blue region we solve the Full-Order model, while in the outer region we rely on a POD-Galerkin Reduced Order model. The partitioning is obtained adaptively using the algorithm described in [15]. Figure 6 (right) shows the flow prediction error relative to dynamic pressure for the worst-case parameter: the proposed method leads to 2% accurate results with a speed-up compared to the full-order model of 8.

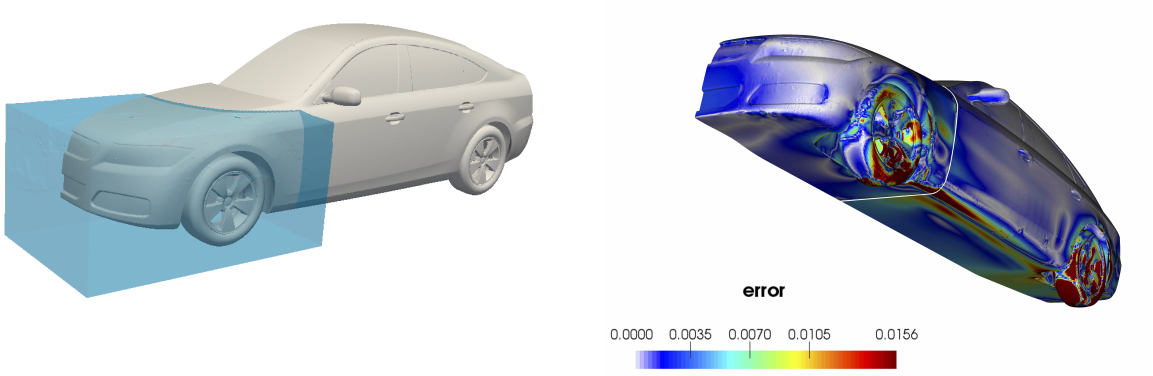


Figure 6. hybrid FOM/ROM approach; application to RANS modelling of the flow around a car.

7.2. All-speed multi-material schemes

We are interested in the development of numerical models for phenomena involving fluid flows and elastic material deformations. We pursue a monolithic approach, which describes the behavior of each material (gas, liquid or solid) through a system of conservation laws and appropriate constitutive relationships. Our method is designed to handle both high-Mach and low-Mach regimes.

It is well-known that Godunov-type schemes are inadequate for low-Mach problems: first, they introduce an eccessive amount of numerical artificial viscosity; second, they require the enforcement of a CFL stability condition which leads to unpractical time steps. For this reason, we resort to the relaxation method proposed in [36], to derive a novel discretization scheme which can be applied to problems characterized by a broad range of Mach numbers. As opposed to [36], we propose in [1] to treat the advective term implicitly.

Figure 7 shows results for a quasi 1D de Laval nozzle problem in water: the flow is low-Mach and almost incompressible. In the present simulation, we impose at the inlet the total pressure $P_{\rm tot} = 10{\rm Pa}$ and the absolute temperature $T = 280K$ and at the outlet the pressure $p_{\rm out} = 1{\rm Pa}$. Figure 7 (center) shows results for the explicit scheme proposed in [36], while Figure 7 (right) shows results of our implicit scheme; for the explicit schemes, we impose the acoustic CFL $\nu_{\rm ac} = 0.4$, while for the implicit scheme, we consider $\nu_{\rm ac} = 100$. We observe that our method outperforms the method in [36].

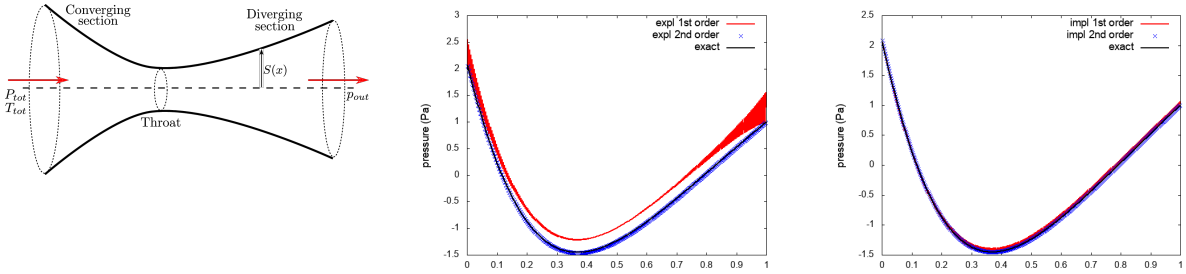


Figure 7. all-speed relaxation scheme. Left: de Laval nozzle. Center: pressure distribution predicted by the explicit method in [36] ($\nu_{\rm ac} = 0.4$). Right: pressure distribution predicted by the implicit method proposed in [1] ($\nu_{\rm ac} = 100$).

7.3. Thermal convection on a hemisphere

Hamid Kellay (LOMA) performs a physical experiment using a half soap bubble heated at the equator. This device allows to study thermal convection and the movement of large scale structures on the surface of the bubble. The results show strong similarities with atmospheric flows on the earth. In particular large vortical structures on the half bubble and tropical cyclones in the atmosphere have the same dynamics.

Using a stereographic transform we solve Navier-Stokes equations on the half bubble and get very good agreement with the experiment. In addition we find that the Nusselt and Reynolds numbers varify scaling laws quite close to the scaling law given in the literature for Rayleigh-Bénard convection: ${\rm Nu} \propto {\rm Ra}^{0.31}$ and ${\rm Re} \propto {\rm Ra}^{1/2}$. Finally a Bolgieno regime is found with scaling as $Ra^{-1/4}$.

Adding the rotation like on the earth we show that the rotation changes the nature of turbulent fluctuations and a new scaling regime is obtained for the temperature field.

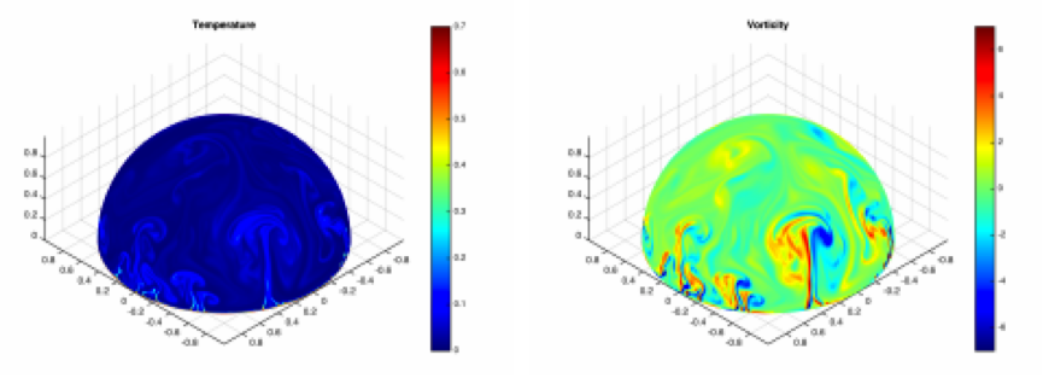


Figure 8. Temperature (left) and vorticity (right) fields at the stationary state for $Ra = 3.10^8$ and $Pr = 7$. Plumes move upward and interact with each other.

MNEMOSYNE Project-Team

7. New Results

7.1. Overview

This year we have explored the main cortico-basal loops of the cerebral architecture and their associated memory mechanisms. The limbic loop (*cf.* § 7.2) concerns the taking into account of the emotional and motivational aspects by the respondant and operant conditioning and their relations with the semantic and episodic memories. The associative loop (*cf.* § 7.3) is about mechanisms of working memory. Concerning the motor loop (*cf.* § 7.4), we have studied mechanisms of song acquisition and production in birds.

We have also worked on the systemic integration of our models (*cf.* § 7.5), raising the question of the conditions of autonomous learning and certain global characteristics such as the transition from goal-directed to habitual behaviors.

Finally, we study the links between our bio-inspired modeling work and Machine Learning (*cf.* § 7.6), revisiting this latter domain in the light of the principles highlighted by our models.

7.2. The limbic loop

We have continued to explore the mnemonic characteristics of the main information flows and cerebral structures interacting with the limbic loop. This is the case with the main learning paradigms associated to the involvement of the amygdala in respondent conditioning [16]. We have also continued our work on the hippocampus and particularly its function of pattern separation [7].

7.3. The associative loop

The prefrontal cortex is known to be involved in many high-level cognitive functions, in particular working memory. In [23], we studied to what extent a group of randomly connected units (namely an Echo State Network, ESN) can store and maintain (as output) an arbitrary real value from a streamed input, i.e. can act as a sustained working memory unit. Furthermore, we explored to what extent such an architecture can take advantage of the stored value in order to produce non-linear computations. Comparison between different architectures (with and without feedback, with and without a working memory unit) shows that an explicit memory improves the performances. This property of the model can be considered as a gated memory: a value enters the memory at the moment of the (input) trigger and is kept constant in face of incoming distractors (the continuous streamed input).

7.4. The motor loop

Sensorimotor learning represents a challenging problem for artificial and natural systems. Several computational models try to explain the neural mechanisms at play in the brain to implement such learning. These models have several common components: a motor control model, a sensory system and a learning architecture. In S. Pagliarini's PhD, our challenge is to build a biologically plausible model for song learning in birds including neuro-anatomical and developmental constraints.

We made a review on a specific type of sensorimotor learning referred to as imitative vocal learning and exemplified by song learning in birds or human complex vocalizations. We aim to compare the various approaches used in existent sensorimotor models relevant for our purpose and to place them in a common framework. We propose a bio-inspired model for an imitative sensorimotor learning, which aims at building a map between the sensory representations of gestures (sensory targets) and their underlying motor pattern through random exploration of the motor space. An example of such learning process occurs during vocal learning in humans or birds, when young subjects babble and learn to copy previously heard adult vocalizations. Previous work (by Hahnloser and Ganguli) has suggested that a simple Hebbian learning rule allows perfect imitation when sensory feedback is a purely linear function of the motor pattern underlying movement production. We aim at generalizing this model to the more realistic case where sensory responses are sparse and non-linear. To this end, we explore the performance of various learning rules and normalizations. Importantly, the proposed model is robust whatever normalization is chosen. We showed that both the imitation quality and the convergence time are highly dependent on the sensory selectivity and dimension of the motor representation.

On this topic, X. Hinaut is also collaborating with Catherine del Negro's team (CNRS, University of Paris-Sud, Orsay) on the representation of syntax in songbird brains. In particular, the project aims at (1) linking the neural activity of a sensori-motor area (HVC) to syntax elements in the songs of domestic canaries ; (2) analysing the audio files and transcripts of canary songs in order to find syntax cues and higher order representations (graph properties of songs, evaluate Markovian forward and backward transition probabilities of various orders). The song analyses part has been performed with the help of a L3 intern, Pierre Marcus.

7.5. Systemic integration

This year, we have considered characteristics of interactions of cortico-basal loops [32], firstly to continue the development of a software environment based on the Minecraft videogame allowing for the survival behavior of an autonomous agent [25], [24] and secondly to revisit the principles of habits formation.

The dorsal pallium (a.k.a. the cortex in the mammals) makes a large loop circuit with the basal ganglia and the thalamus known to control and adapt behavior but the who's who of the functional roles of these structures is still debated. Influenced by the Triune brain theory that was proposed in the early sixties, many current theories propose a hierarchical organization on the top of which stands the cortex to which the subcortical structures are subordinated. In particular, habits formation has been proposed to reflect a switch from conscious on-line control of behavior by the cortex, to a fully automated subcortical control. We have proposed in [3] instead to revalue the function of the network in light of the current experimental evidence concerning the anatomy and physiology of the basal ganglia-cortical circuits in vertebrates.

This theory is supported by a model [11] that includes interactions between the cortex, the basal ganglia and the thalamus based on a dual competition. We hypothesize that the striatum, the subthalamic nucleus, the internal pallidum (GPi), the thalamus, and the cortex are involved in closed feedback loops through the hyperdirect and direct pathways. These loops support a competition process that results in the ability of basal ganglia to make a cognitive decision followed by a motor decision. Considering lateral cortical interactions, another competition takes place inside the cortex allowing the latter to make a cognitive and a motor decision. We show how this dual competition endows the model with two regimes. One is driven by reinforcement learning, the other by Hebbian learning. The final decision is made according to a combination of these two mechanisms with a gradual transfer from the former to the latter. We confirmed these theoretical results on primates (Macaca mulata) using a novel paradigm predicted by the model.

7.6. Machine Learning

In this section, we report on some neuronal adaptive mechanisms, that we develop at the frontier between Machine Learning and Computational Neuroscience. Our goal is to consider and adapt models in Machine Learning for their integration in a bio-inspired framework.

Concerning the manipulation of temporal sequences, we have proposed an original algorithm for the extraction of sequences from LSTM, a major class of recurrent neural networks [1]. These sequences are then interpreted as rules representing implicit knowledge within electrical diagrams (*cf.* § 8.1).

Concerning our work on reservoir computing, X. Hinaut is collaborating with Michael Spranger (Sony Lab, Tokyo, Japan) on grounding of language, adapting Hinaut's previous Reservoir Language Model (RLM) with the representational system of Spranger: IRL (Incremental Recruitment Language). He is also collaborating with Hamburg on the use of reservoir models for robotic tasks (*cf.* § 9.3). In this work, we have shown that the RLM can successfully learn to parse sentences related to home scenarios in fifteen languages [6]. This demonstrates that (1) the learning principle of our model is not limited to a particular language (or particular sentence structures), and (2) it can deal with various kinds of representations (not only predicates), which enable users to adapt it to their own needs.

Regarding the extraction of characteristics from and the use of hierarchical networks, as in the case of deep networks, we have been able to consider how to deal with not-so-big data sets, and target the notion of interpretability of the obtained results which is a key issue: since deep learning applications are increasingly present in the society, it is important that the underlying processes be accessible and understandable to every one. In order to target these challenges, we have analyzed how considering prototypes in a rather generalized sense (with respect to the state of the art) allows to reasonably work with small data sets while providing an interpretable view of the obtained results. Some mathematical interpretation of this proposal have also been discussed. Sensitivity to hyperparameters is a key issue for reproducible deep learning results, and has been carefully considered in our methodology. Performances and (even more interesting, in a sense) limitations of the proposed setup have been explored in details, under different hyperparameters sets, in an analogous way as biological experiments are conducted. We obtain a rather simple architecture, easy to explain, and which allows, combined with a standard method, to target both performances and interpretability [4].

MONC Project-Team

7. New Results

7.1. Mathematical Modeling of the Proliferation Gradient in MultiCellular Tumor Spheroids

Authors: *Thomas Michel*, J. Fehrenbach, V. Lobjois, J. Laurent, A. Gomes, *Thierry Colin, Clair Poignard*. Paper published in the Journal of Theoretical Biology. https://hal.inria.fr/hal-01883189

MultiCellular Tumor Spheroids are 3D cell cultures that can accurately reproduce the behavior of solid tumors. It has been experimentally observed that large spheroids exhibit a decreasing gradient of proliferation from the periphery to the center of these multicellular 3D models: the proportion of proliferating cells is higher in the periphery while the non-proliferating quiescent cells increase in depth. In this paper, we propose to investigate the key mechanisms involved in the establishment of this gradient with a Partial Differential Equations model that mimics the experimental setup of growing spheroids under different nutrients supply conditions. The model consists of mass balance equations on the two cell populations observed in the data: the proliferating cells and the quiescent cells. The spherical symmetry is used to rewrite the model in radial and relative coordinates. Thanks to a rigorous data postprocessing the model is then fit and compared quantitatively with the experimental quantification of the percentage of proliferating cells from EdU immunodetection on 2D spheroid cryosection images. The results of this calibration show that the proliferation gradient observed in spheroids can be quantitatively reproduced by our model.

7.2. Viscoelastic modeling of the fusion of multicellular tumor spheroids in growth phase

Authors: *Guillaume Dechristé*, Jérôme Fehrenbach, Elena Griseti, Valérie Lobjois, *Clair Poignard*. Paper published in the Journal of Theoretical Biology. https://hal.inria.fr/hal-01786027

Background. Since several decades, the experiments have highlighted the analogy of fusing cell aggregates with liquid droplets. The physical macroscopic models have been derived under incompressible assumptions. The aim of this paper is to provide a 3D model of growing spheroids, which is more relevant regarding embryo cell aggregates or tumor cell spheroids. Methods. We extend the past approach to a compressible 3D framework in order to account for the tumor spheroid growth. We exhibit the crucial importance of the effective surface tension, and of the inner pressure of the spheroid to describe precisely the fusion. The experimental data were obtained on spheroids of colon carcinoma human cells (HCT116 cell line). After 3 or 6 days of culture, two identical spheroids were transferred in one well and their fusion was monitored by live videomicroscopy acquisition each 2hours during 72h. From these images the neck radius and the diameter of the assembly of the fusing spheroids are extracted. Results. The numerical model is fitted with the experiments. It is worth noting that the time evolution of both neck radius and spheroid diameter are quantitatively obtained. The interesting feature lies in the fact that such measurements characterise the macroscopic rheological properties of the tumor spheroids. Conclusions. The experimental determination of the kinetics of neck radius and overall diameter during spheroids fusion characterises the rheological properties of the spheroids. The consistency of the model is shown by fitting the model with two different experiments, enhancing the importance of both surface tension and cell proliferation. General Significance. The paper sheds new light on the macroscopic rheological properties of tumor spheroids. It emphasizes the role of the surface tension and the inner pressure in the fusion of growing spheroid. Under geometrical assumptions, the model reduces to a 2–parameter differential equation fit with experimental measurements. The 3–D partial differential system makes it possible to study the fusion of spheroids in non-symmetrical or more general frameworks.

7.3. Mathematical analysis and 2-scale convergence of a heterogeneous microscopic bidomain model

Authors: *Annabelle Collin*, Sébastien Imperiale. Paper published in Mathematical Models and Methods in Applied Sciences. https://hal.inria.fr/hal-01759914

The aim of this paper is to provide a complete mathematical analysis of the periodic homogenization procedure that leads to the macroscopic bidomain model in cardiac elec-trophysiology. We consider space-dependent and tensorial electric conductivities as well as space-dependent physiological and phenomenological non-linear ionic models. We provide the nondimensionalization of the bidomain equations and derive uniform estimates of the solutions. The homogenization procedure is done using 2-scale convergence theory which enables us to study the behavior of the non-linear ionic models in the homogenization process.

7.4. Pre-treatment magnetic resonance-based texture features as potential imaging biomarkers for predicting event free survival in anal cancer treated by chemoradiotherapy

Authors: Arnaud Hocquelet, Thibaut Auriac, *Cynthia Perier*, Clarisse Dromain, Marie Meyer, Jean-Baptiste Pinaquy, Alban Denys, Hervé Trillaud, *Baudouin Denis de Senneville*, Véronique Vendrely. Paper published in European Radiology. https://hal.archives-ouvertes.fr/hal-01962472

AIM: To assess regular MRI findings and tumour texture features on pre-CRT imaging as potential predictive factors of event-free survival (disease progression or death) after chemoradiotherapy (CRT) for anal squamous cell carcinoma (ASCC) without metastasis.
MATERIALS AND METHODS: We retrospectively included 28 patients treated by CRT for pathologically proven ASCC with a pre-CRT MRI. Texture analysis was carried out with axial T2W images by delineating a 3D region of interest around the entire tumour volume. First-order analysis by quantification of the histogram was carried out. Second-order statistical texture features were derived from the calculation of the grey-level co-occurrence matrix using a distance of 1 (d1), 2 (d2) and 5 (d5) pixels. Prognostic factors were assessed by Cox regression and performance of the model by the Harrell C-index.
RESULTS: Eight tumour progressions led to six tumour-specific deaths. After adjusting for age, gender and tumour grade, skewness (HR = 0.131, 95% CI = 0-0.447, p = 0.005) and cluster shade_d1 (HR = 0.601, 95% CI = 0-0.861, p = 0.027) were associated with event occurrence. The corresponding Harrell C-indices were 0.846, 95% CI = 0.697-0.993, and 0.851, 95% CI = 0.708-0.994.
CONCLUSION: ASCC MR texture analysis provides prognostic factors of event occurrence and requires additional studies to assess its potential in an "individual dose" strategy for ASCC chemoradiation therapy.
KEY POINTS: MR texture features help to identify tumours with high progression risk. Texture feature maps help to identify intra-tumoral heterogeneity. Texture features are a better prognostic factor than regular MR findings.
KEYWORDS: Anal squamous cell carcinoma; Definitive chemoradiotherapy; Imaging biomarkers; Magnetic resonance imaging; Texture analysis

7.5. T2-based MRI Delta-radiomics improve response prediction in soft-tissue sarcomas treated by neoadjuvant chemotherapy

Authors: *Amandine Crombé, MS Cynthia Périer*, Michèle Kind, *Baudouin Denis De Senneville*, François Le Loarer, Antoine Italiano, Xavier Buy, *Olivier Saut*. Paper published in the Journal of Magnetic Resonance Imaging. https://hal.inria.fr/hal-01929807

Background: Standard of care for patients with high-grade soft-tissue sarcoma (STS) are being redefined since neoadjuvant chemotherapy (NAC) has demonstrated a positive effect on patients' outcome. Yet response evaluation in clinical trials still relies on RECIST criteria.

Purpose: To investigate the added value of a Delta-radiomics approach for early response prediction in patients with STS undergoing NAC.
Study Type: Retrospective.
Population: Sixty-five adult patients with newly-diagnosed, locally-advanced, histologically proven high-grade STS of trunk and extremities. All were treated by anthracycline-based NAC followed by surgery and had available MRI at baseline and after two chemotherapy cycles.
Field Strength/Sequence: Pre- and postcontrast enhanced T1-weighted imaging (T1-WI), turbo spin echo T2-WI at 1.5 T.
Assessment: A threshold of $< 10\%$ viable cells on surgical specimens defined good response (Good-HR). Two senior radiologists performed a semantic analysis of the MRI. After 3D manual segmentation of tumors at baseline and early evaluation, and standardization of voxel-sizes and intensities, absolute changes in 33 texture and shape features were calculated.
Statistical Tests: Classification models based on logistic regression, support vector machine, k-nearest neighbors, and random forests were elaborated using crossvalidation (training and validation) on 50 patients ("training cohort") and was validated on 15 other patients ("test cohort").
Results: Sixteen patients were good-HR. Neither RECIST status (P = 0.112) nor semantic radiological variables were associated with response (range of P-values: 0.134–0.490) except an edema decrease (P = 0.003), although 14 shape and texture features were (range of P-values: 0.002–0.037). On the training cohort, the highest diagnostic performances were obtained with random forests built on three features: Δ_Histogram_Entropy, Δ_Elongation, Δ_Surrounding_Edema, which provided: area under the curve the receiver operating characteristic = 0.86, accuracy = 88.1%, sensitivity = 94.1%, and specificity = 66.3%. On the test cohort, this model provided an accuracy of 74.6% but 3/5 good-HR were systematically ill-classified.
Data Conclusion: A T2-based Delta-radiomics approach might improve early response assessment in STS patients with a limited number of features.

7.6. Revisiting bevacizumab + cytotoxics scheduling using mathematical modeling: proof of concept study in experimental non-small cell lung carcinoma

Authors: Diane-Charlotte Imbs, Raouf El Cheikh, Arnaud Boyer, Joseph Ciccolini, Celine Mascaux, Bruno Lacarelle, Fabrice Barlesi, Dominique Barbolosi, *Sébastien Benzekry*. Paper published in Clinical Pharmacology and Therapeutics: Pharmacometrics and Systems Pharmacology. https://hal.inria.fr/hal-01624423v2

Concomitant administration of bevacizumab and pemetrexed-cisplatin is a common treatment for advanced non-squamous non-small cell lung cancer (NSCLC). Vascular normalization following bevacizumab administration may transiently enhance drug delivery, suggesting improved efficacy with sequential administration. To investigate optimal scheduling, we conducted a study in NSCLC-bearing mice using. First, experiments demonstrated improved efficacy when using sequential versus concomitant scheduling of bevacizumab and chemotherapy. Using a mathematical model of tumor growth under therapy accounting for the normalization effect, we predicted an optimal delay of 2.8 days between bevacizumab and chemotherapy. This prediction was confirmed experimentally, with reduced tumor growth of 38% as compared to concomitant scheduling, and prolonged survival (70 vs. 74 days). Alternate sequencing of 8 days failed in achieving similar increase in efficacy, thus emphasizing the utility of modeling support to identify optimal scheduling. The model could also be a useful tool in the clinic to personally tailor regimen sequences.

PHOENIX-POST Team

6. New Results

6.1. Towards context-aware assistive applications for aging in place via real-life-proof activity detection

Assisted living applications can support aging in place efficiently when their context-awareness is based on a real-life-proof approach to activity detection. Recently, Caroux et al. proposed a new approach to monitoring activities dedicated to older adults, named "activity verification". This approach uses a knowledge-driven framework that draws from the literature on older adults. The purpose of the present study is to address the limitations of this approach by scaling it up and by demonstrating that it is applicable to context-aware assistive applications for aging in place. First, an experimental study was conducted in which this approach was used to monitor a large range of daily activities, for a long period (8 weeks of experimentation) and involving several participants (7 participants). Second, this approach was used to validate two examples of context-aware assisted living applications, via simulation, based on real-life sensor log data. Results showed that the applicability of the "activity verification" approach scales up to a large range of daily activities by extending this approach (with accuracy values ranging between 0.82 and 1.00 depending on the activity of interest). Its inter-participant and intra-participant consistencies were demonstrated. Its limitations were addressed and the applicability to context-aware assistive applications for aging in place running on a dedicated platform was demonstrated.

6.2. Are visual cues helpful for virtual spatial navigation and spatial memory in patients with mild cognitive impairment or Alzheimer's disease?

Objective: To evaluate whether visual cues are helpful for virtual spatial navigation and memory in Alzheimer's disease (AD) and patients with mild cognitive impairment (MCI). Method: 20 patients with AD, 18 patients with MCI and 20 age-matched healthy controls (HC) were included. Participants had to actively reproduce a path that included 5 intersections with one landmark at each intersection that they had seen previously during a learning phase. Three cueing conditions for navigation were offered: salient landmarks, directional arrows and a map. A path without additional visual stimuli served as control condition. Navigation time and number of trajectory mistakes were recorded. Results: With the presence of directional arrows, no significant difference was found between groups concerning the number of trajectory mistakes and navigation time. The number of trajectory mistakes did not differ significantly between patients with AD and patients with MCI on the path with arrows, the path with salient landmarks and the path with a map. There were significant correlations between the number of trajectory mistakes under the arrow condition and executive tests, and between the number of trajectory mistakes under the salient landmark condition and memory tests. Conclusion: Visual cueing such as directional arrows and salient landmarks appears helpful for spatial navigation and memory tasks in patients with AD and patients with MCI. This study opens new research avenues for neuro-rehabilitation, such as the use of augmented reality in real-life settings to support the navigational capabilities of patients with MCI and patients with AD.

6.3. Early detection of mild cognitive impairment with in-home monitoring technologies using functional measures: A systematic review

Introduction: The aging of the world population is accompanied by a substantial increase in neurodegenerative disorders such as dementia. Early detection of dementia, i.e. at the mild cognitive impairment (MCI) stage, could be an essential condition for slowing down the loss of autonomy and quality of life caused by the disease, as it would provide a critical window for the implementation of early pharmacological and non-pharmacological interventions. However, the current assessments for MCI have several limitations. In this

context, approaches involving smart home technologies offer many attractive advantages, including the continuous measurement of functional abilities in ecological environments. Objective: This systematic review aims to investigate the current state of knowledge on the effectiveness of smart home technologies for the early detection of MCI through the monitoring of everyday life activities. Methods: A systematic search of publications in Medline, EMBASE, CINAHL was conducted. Results: Sixteen studies were included in this review. Twelve studies were based on real-life monitoring, with several sensors installed in participants' actual homes, and four studies included scenario-based evaluations in which the participants had to complete various tasks in a research lab apartment. In real-life monitoring, the most used indicators of MCI were walking speed and activity/motion in the house. In scenario-based evaluation, time of completion, quality of activity completion, number of errors, amount of assistance needed, and task-irrelevant behaviors during the performance of everyday activities predicted MCI in participants. Discussion: Despite technological limitations and the novelty of the field, smart home technologies represent a promising potential for the early screening of MCI and could support clinicians in geriatric care.

6.4. A Language for Online State Processing of Binary Sensors, Applied to Ambient Assisted Living

There is a large variety of binary sensors in use today, and useful context-aware services can be defined using such binary sensors. However, the currently available approaches for programming context-aware services do not conveniently support binary sensors. Indeed, no existing approach simultaneously supports a notion of state, central to binary sensors, offers a complete set of operators to compose states, allows to define reusable abstractions by means of such compositions, and implements efficient online processing of these operators. This paper proposes a new language for event processing specifically targeted to binary sensors. The central contributions of this language are a native notion of state and semi-causal operators for temporal state composition including: Allen's interval relations generalized for handling multiple intervals, and temporal filters for handling delays. Compared to other approaches such as CEP (complex event processing), our language provides less discontinued information, allows less restricted compositions, and supports reusable abstractions. We implemented an interpreter for our language and applied it to successfully rewrite a full set of real Ambient Assisted Living services. The performance of our prototype interpreter is shown to compete well with a commercial CEP engine when expressing the same services.

6.5. Implementing a semi-causal domain-specific language for context detection over binary sensors

In spite of the fact that many sensors in use today are binary (i.e. produce only values of 0 and 1), and that useful context-aware applications are built exclusively on top of them, there is currently no development approach specifically targeted to binary sensors. Dealing with notions of state and state combinators, central to binary sensors, is tedious and error-prone in current approaches. For instance, developing such applications in a general programming language requires writing code to process events, maintain state and perform state transitions on events, manage timers and/or event histories. In another paper, we introduced a domain specific language (DSL) called Allen, specifically targeted to binary sensors. Allen natively expresses states and state combinations, and detects contexts on line, on incoming streams of binary events. Expressing state combinations in Allen is natural and intuitive due to a key ingredient: semi-causal operators. That paper focused on the concept of the language and its main operators, but did not address its implementation challenges. Indeed, online evaluation of expressions containing semi-causal operators is difficult, because semi-causal sub-expressions may block waiting for future events, thus generating unknown values, besides 0 and 1. These unknown values may or may not propagate to the containing expressions, depending on the current value of the other arguments. This paper presents a compiler and runtime for the Allen language, and shows how they implement its state combining operators, based on reducing complex expressions to a core subset of operators, which are implemented natively. We define several assisted living applications both in Allen and in a general scripting language. We show that the former are much more concise in Allen, achieve more effective code reuse, and ease the checking of some domain properties.

6.6. Towards Truly Accessible MOOCs for Persons with Cognitive Disabilities: Design and Field Assessment

MOOCs are playing an increasingly important role in education systems. Unfortunately, MOOCs are not fully accessible. In this paper, we propose design principles to enhance the accessibility of MOOC players, especially for persons with cognitive disabilities. These principles result from a participatory design process gathering 7 persons with disabilities and 13 expert professionals. They are also inspired by various design approaches (Universal Design for Learning, Instructional Design, Environmental Support). We also detail the creation of a MOOC player offering a set of accessibility features that users can alter according to their needs and capabilities. We used it to teach a MOOC on digital accessibility. Finally, we conducted a field study to assess learning and usability outcomes for persons with cognitive and non-cognitive impairments. Results support the effectiveness of our player for increasing accessibility.

6.7. Assistive Computing: a Human-Centered Approach to Developing Computing Support for Cognition

The growing population of cognitively impaired individuals calls for the emergence of a research area dedicated to developing computing systems that address their needs. The nature of this research area requires to bridge the many disciplines needed to develop human-centered, assistive computing systems. Such bridging may seem unattainable considering the conceptual and practical gaps between the related disciplines and the challenges of propagating human-related concerns throughout the many stages of the development process of assistive technologies. As a consequence, existing assistive technologies lack a proper needs analysis; their development is often driven by technology concerns, resulting in ill-designed and stereotype-biased systems; and, most of them are not tested for their effectiveness in assisting users. In this paper, we propose a systematic exploration of this vast challenge. First, we decline Assistive Computing as a research area and propose key principles to drive its study. Then, we introduce a tool-based methodology dedicated to developing assistive computing support, integrating a range of disciplines from human-related sciences to computer science. This methodology is purposefully pragmatic in that it leverages, aggregates and revisits numerous research results, concretizing it with a range of examples. More generally, our goal is i) to provide a framework to conduct research in the area of Assistive Computing and ii) to identify the necessary bridges between disciplines to account for all the dimensions of such systems.

PLEIADE Team

7. New Results

7.1. Introgressions as a source of diversity

Several prominent mechanisms of genomic evolution have been described for the yeasts, among them inter-specific hybridization, reticulated evolution, aneuploidization, recent or ancient poly-ploidization events, large chromosomal duplication or more limited gene duplication, and horizontal transfer. These mechanisms are usually so closely intertwined that it is difficult to determine which ones are causes or consequences. Regardless of mechanisms the result has been a drastic reshaping of yeasts genome along evolution. Understanding these mechanisms is important, not only for strain construction in biotechnology, but also more fundamentally for insight into the causes and effects of genome reshaping on much shorter time scales.

Introgression, the transfer of large or more limited genetic information from one species to another, is an evolutionary mechanism of particular interest in industrial applications such as wine making where large vat cultures are used. Introgression results in mosaic genomes, and can be the result of interspecific hybridization fol- lowed by the extensive loss of one parental genome, either through repeated backcross with one parental species or through missegregation of the hybrid at meiosis.

In collaboration with the Institut des Science de la Vigne et du Vin and Bordeaux Sciences Agro, PLEIADE developed tools to rapidly assess the presence of introgressed regions in a large population of *Saccharomyces uvarum* isolates (104 strains), focusing on Holarctic isolates from natural, cider and wine environments since introgressed regions are absent in Southern hemisphere isolates. The overall number of introgressed regions is significantly higher in cider-associated strains compared to wild strains, and is higher in wine isolates. However, only a subset of the introgressed regions were found to be overrepresented in anthropic activities and their number and quality varied between cider- and wine-making processes.

Paradoxically, the low Holarctic genetic diversity observed in [1] contrasts with the relative high phenotypic diversity found for technological traits. This contradiction suggests that interspecific introgressions found among Holarctic *S. uvarum* strains could be the most important source of genetic, and by extention of phenotypic, diversity.

7.2. Geometric view of biodiversity and molecular taxonomy

For the geometric view on biodiversity and molecular based taxonomy (metabarcoding), 2018 has been characterized by significant progress in the connections between the questions posed by metabarcoding, and the implementaton of solutions by intensive computing in collaboration with HiePACS team. We are now able to run a Multidimensional Scaling (MDS) to build a point cloud where each point is a read (a short sequence from an environmental sample) on a sample of nearly 2×10^5 reads on one node of a cluster with shared memory. PLEIADE is member of two projects lead by Inria BSO to extend these possibilities to work with distributed memory (connection with the Chameleon dense linear algebra solver) with the possibility to reach results with one million sequences at hand. To our knowledge, such a connection between scientific questions related to biodiversity and HPC-HTC has up to now had no equivalent and is new.

In collaboration with HiePACS, the SED of Inria BSO, and the GRICAD (Mésocentre of Grenoble), PLEIADE has developed a scientific library of tools for dimension reductions in the framework of Big Data, specifically linear algebra of full matrices. A prototype was developed in numpy, then reimplemented in C++ with an external library for Random Projection, and made available in Julia. This library will be connected with the workflow developed in Regional project and ADT Gordon.

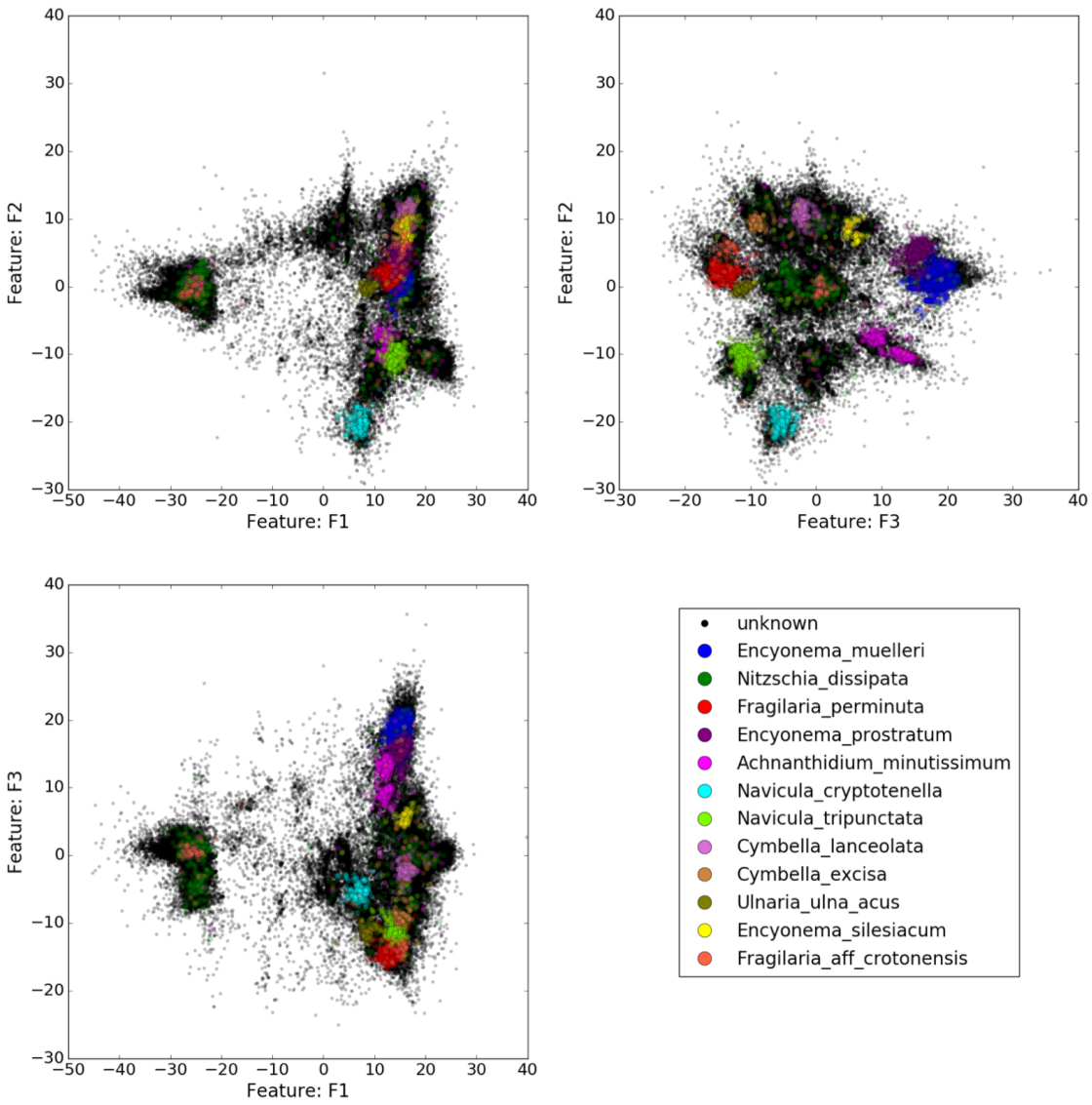


Figure 4. Validation of high density islands using supervised classification. Metagenomic reads from diatoms in Lake Geneva [26] were analyzed by the method from [7] and colored by species according to a reference database.

7.3. Tangible exploration of protein families

Protein families[3] are an effective way to compare the complete genomes of fungal species. In general these comparisons are very challanging due to the large evolutionary distances involved, the wide range of GC compositions observed from one species to the next, and the extensive map reshuffling that is characteristic of the yeasts in particular. Protein families are a classification of protein-coding gene sequences into phylogenetic groups, using clustering methods and semi-supervised classification. Members of a family are homologous and in many cases this homology is suggestive of functional similarity.

An intriguing feature of protein families is that the weighted graph constructed from their pairwise distance matrices has a structure that reflects the evolutionary history of the family. We developed software (family-3d, see above) that uses truncated distances to construct a weighted graph and to lay it out using an adaptation of the three-dimensional extension of the Kamada-Kawai force-directed layout. The resulting shapes for a set of protein families are then clustered manually by similarity (figure 5). Similarity in shape is highly suggestive of similarity in evolutionary scenarios, leading to hypotheses about the histories of individual protein families and the mechanisms by which functional diversity is obtained.

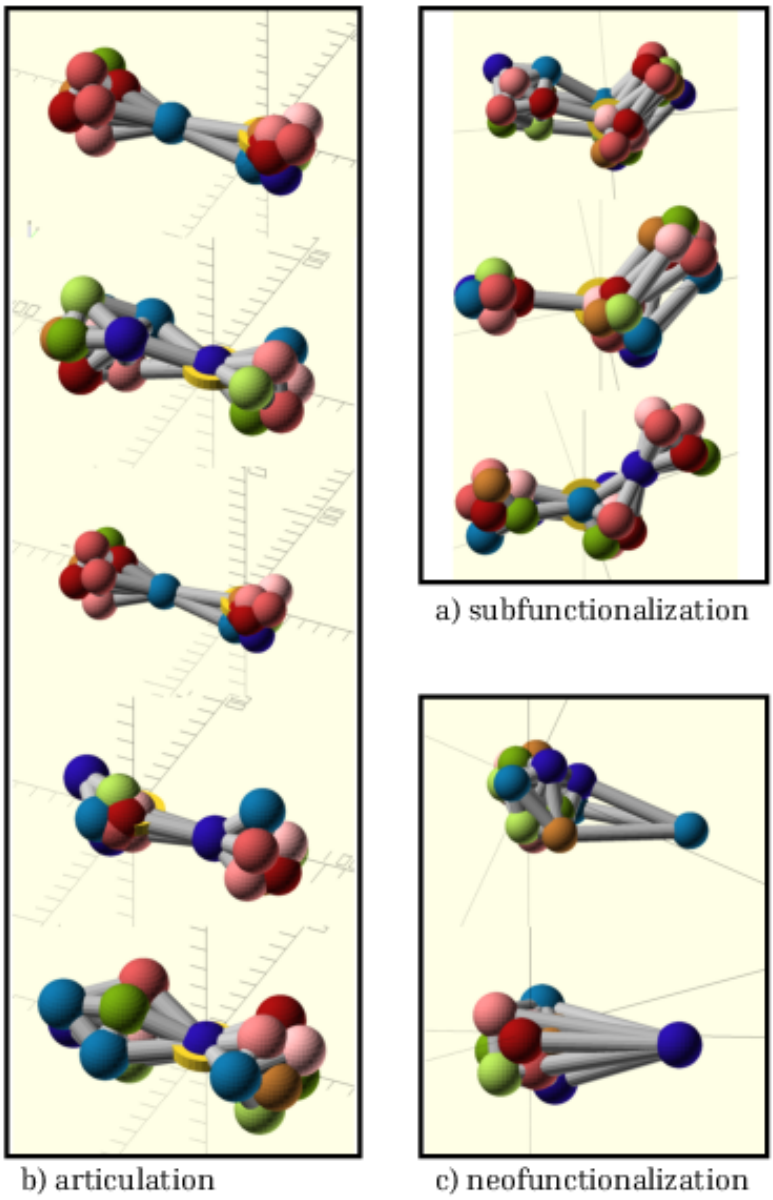


Figure 5. Three examples of classes of protein families with similar shapes: a) articulation, b) subfunctionalization, c) neofunctionalization

POTIOC Project-Team

7. New Results

7.1. Transition between AR and VR spaces

Participants: Joan Sol Roo, Martin Hachet, Pierre-Antoine Cinquin

Mixed Reality systems combine physical and digital worlds, with great potential for the future of HCI. It is possible to design systems that support flexible degrees of virtuality by combining complementary technologies. In order for such systems to succeed, users must be able to create unified mental models out of heterogeneous representations. We conducted two studies focusing on the users' accuracy on heterogeneous systems using Spatial Augmented Reality (SAR) and immersive Virtual Reality (VR) displays (see Figure 4), and combining viewpoints (egocentric and exocentric). The results show robust estimation capabilities across conditions and viewpoints [31].

Figure 4. A user experiencing transitions between spatial augmented reality and virtual reality spaces.

7.2. Tangible and augmented interfaces for Schoolchildren

Participants: Philippe Giraudeau, Théo Segonds, Martin Hachet

In 2018, we have continued working on the exploration of tangible and augmented interfaces for Schoolchildren. We have notably evaluated the pedagogical potential of Teegi in a user study conducted at school [24].

We have also pursued our work on collaborative learning at school, part of the e-Tac project. In particular, based on focus group with children and practitioners, we have refined our interactive pedagogical environment, and we have implemented a new version (see Figure 5) [53]

7.3. Ambient interfaces dedicated to the awareness of energy consumption

Participants: Pierre-Antoine Cinquin, Philippe Giraudeau

Figure 5. Tangible and augmented objects to foster collaborative learning at school.

Inspired by studies in data physicalization, we explore the use of tangible and ambient interfaces to raise people's awareness of energy consumption. As a first approach, we are developing an interactive and collaborative environment named Erlen. This year, we have designed a first prototype taking the form of an Erlenmeyer flask with fluid simulation. Through manipulation, users can visualize information about their electricity consumption. This prototype was demonstrated at IHM 2018 [21]. Based on the feedback we obtained, we are actually developing a new set of individuals and shared interfaces along with new interactions.

7.4. Drones for Human interaction

Participants: Rajkumar Darbar, Anke Brock, Martin Hachet.

We have also continued working with drones. In particular, we have proposed FlyMap as a novel user experience for interactive maps projected from a drone. We iteratively designed three interaction techniques for FlyMap's usage scenarios. In a comprehensive indoor study (N = 16), we show the strengths and weaknesses of the techniques on users' cognition, task load and satisfaction. We then pilot tested FlyMap outdoors in real world conditions with four groups of participants. We show that its interactivity is exciting to users, opening the space for more direct interactions with drones [20].
We are currently exploring the use of drones to bring passive haptic feedback in immersive VR scenario. Concretely, we are building a system where drones, equipped with flat panels, co-locate themselves with virtual objects to provide physical feedbacks to VR users.

7.5. Mixed reality based interfaces for visual impaired persons

Participant: Lauren Thévin, Anke Brock

Current low-tech Orientation & Mobility (O&M) tools for visually impaired people, e.g. tactile maps, possess limitations. Interactive accessible maps have been developed to overcome these. However, most of them are limited to exploration of existing maps, and have remained in laboratories. Using a participatory design approach, we have worked closely with 15 visually impaired students and 3 O&M instructors over 6 months. We iteratively designed and developed an augmented reality map destined at use in O&M classes in special education centers. This prototype combines projection, audio output and use of tactile tokens, and thus allows both map exploration and construction by low vision and blind people. Our user study demonstrated that all students were able to successfully use the prototype, and showed a high user satisfaction. A second phase with 22 international special education teachers allowed us to gain more qualitative insights. This work shows that augmented reality has potential for improving the access to education for visually impaired people [18].

We have pursued this work to make the visual and audio augmentation of real objects easy and convenient. In a user study, six teachers created their own audio-augmentation of objects, such as a botanical atlas (Figure 6 , within 30 minutes or less. Teachers found the tool easy to use and were confident about re-using it. Participants found the resulting interactive graphics exciting to use independently of their mental imagery skills [32].

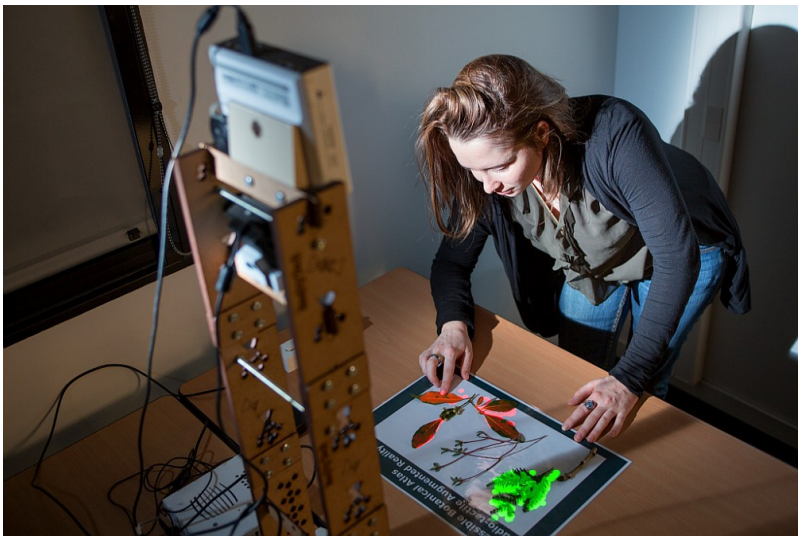


Figure 6. Multimodal Augmented Reality for visual impaired students.

7.6. Accessibility of e-learning systems

Participants: Pierre-Antoine Cinquin, Damien Caselli and Pascal Guitton

In 2018, we continued to work on new digital teaching systems such as MOOCs. Unfortunately, accessibility for people with disabilities is often forgotten, which excludes them, particularly those with cognitive impairments for whom accessibility standards are fare from being established. We have shown in [11] that very few research activities deal with this issue.

In past years, we have proposed new design principles based on knowledge in the areas of accessibility (Ability-based Design and Universal Design), digital pedagogy (Instruction Design with functionalities that reduce the cognitive load : navigation by concept, slowing of the flow...), specialized pedagogy (Universal Design for Learning, eg, automatic note-taking, and Self Determination Theory, e.g., configuration of the interface according to users needs and preferences) and psychoducational interventions (eg, support the joint teacher-learner attention), but also through a participatory design approach involving students with disabilities and experts in the field of disability. From these framework, we have designed interaction features which have been implemented in a specific MOOC player called Aïana. Moreover, we have produced a MOOC on digital accessibility which is published on the national MOOC platform (FUN) using Aïana (4 sessions since 2016 with more than 9000 registered participants). https://mooc-francophone.com/cours/mooc-accessibilite-numerique/. Our first field studies demonstrate the benefits of using Aïana for disabled participants [22].

7.7. Improving EEG Signal Processing for Brain-Computer Interfaces

Participants: Aurélien Appriou, Satyam Kumar, Fabien Lotte

Figure 7. The Aïana MOOC player.

A review of classification algorithms for BCI: Most current Electroencephalography (EEG)-based Brain-Computer Interfaces (BCIs) are based on machine learning algorithms. We surveyed the BCI and machine learning literature to identify the classification approaches that have been investigated to design BCIs. We found that the recently designed classification algorithms for EEG-based BCIs can be divided into four main categories: adaptive classifiers, matrix and tensor classifiers, transfer learning and deep learning, plus a few other miscellaneous classifiers. Among these, adaptive classifiers were demonstrated to be generally superior to static ones, even with unsupervised adaptation. Transfer learning can also prove useful although the benefits of transfer learning remain unpredictable. Riemannian geometry-based methods have reached state-of-the-art performances on multiple BCI problems and deserve to be explored more thoroughly, along with tensor-based methods. Shrinkage linear discriminant analysis and random forests also appear particularly useful for small training samples settings. On the other hand, deep learning methods have not yet shown convincing and consistent improvement over state-of-the-art BCI methods. This survey was published in Journal of Neural Engineering in [14].

Exploring Modern Machine Learning Methods to Estimate Mental Workload From EEG Signals: Estimating mental workload from brain signals such as EEG has proven very promising in multiple HCI applications, e.g., to design games or educational applications with adaptive difficulty. However, currently obtained workload classification accuracies are relatively low, making the resulting estimations not fully trustable. We thus studied promising modern machine learning algorithms, including Riemannian geometry-based methods and Convolutional Neural Networks, to estimate workload from EEG signals. We studied them with both user-specific and user-independent calibration, to go towards calibration-free systems. Our results suggested that a shallow Convolutional Neural Network obtained the best performance in both conditions, outperforming state-of-the-art methods on the used data sets. This work was published as a work-in-progress in the CHI conference [19].

BCPy, an open-source python platform for offline EEG signals decoding and analysis: Although promising, BCIs are still barely used outside laboratories due to their poor robustness. Moreover, they are sensitive to noise, outliers and the non-stationarity of EEG signals. Many algorithms have been developed for EEG signals processing and classification, in order to improve BCIs robustness. We proposed BCPy, an open-source, easy-to-use python BCI platform for offline EEG signal analysis. Python is free and contains good scalable libraries for scientific computing. Moreover, Python is the major language used to implement recent advances

in ML and Deep Learning, thus making them easily available for BCI research. This work was published in the International BCI meeting [48].

Adaptive Riemannian classification methods: The omnipresence of non-stationarity and noise in EEG signals restricts the ubiquitous use of BCIs. One of the possible ways to tackle this problem is to adapt the computational model used to detect and classify different mental states. Adapting the model will possibly help us to track the changes and thus reducing the effect of non-stationarities. In this paper, we present different adaptation strategies for state of the art Riemannian geometry based classifiers. The offline evaluation of our proposed methods on two different datasets showed a statistically significant improvement over baseline non-adaptive classifiers. Moreover, we also demonstrate that combining different (hybrid) adaptation strategies generally increased the performance over individual adaptation schemes. Also, the improvement in average classification accuracy for a 3-class mental imagery BCI with hybrid adaption is as high as around 17% above the baseline non-adaptive classifier. This was published in [26].

Regularized spatial filters for EEG regression problems: In collaboration with University Freiburg, we reported on novel supervised algorithms for single-trial brain state decoding. When brain activity is assessed by multichannel recordings, spatial filters computed by the source power comodulation (SPoC) algorithm allow identifying oscillatory subspaces. In small dataset scenarios, this supervised method tends to overfit to its training data. To improve upon this, we proposed and characterize three types of regularization techniques for SPoC. Evaluating all methods on real-world data, we observed an improved regression performance mainly for datasets from subjects with initially poor performance. This was published in the Neuroinformatics journal [16].

SEREEGA: a toolbox to Simulate EEG activity: EEG is a popular method to monitor brain activity, but it is difficult to evaluate EEG-based analysis methods because no ground-truth brain activity is available for comparison. Therefore, to test and evaluate such methods, in collaboration with TU Berlin, we proposed SEREEGA, a free and open-source matlab toolbox for Simulating Event-Related EEG Activity. The toolbox is available at https://github.com/lrkrol/SEREEGA. SEREEGA unifies the majority of past simulation methods reported in the literature into one toolbox. This toolbox and its use were published in journal of neuroscience methods [13].

7.8. Understanding Brain-Computer Interfaces user Training

Participants: Léa Pillette, Camille Benaroch, Fabien Lotte

Computational models of BCI performance: Mental-Imagery based BCIs (MI-BCIs) use signals produced during mental imagery tasks to control the system. Current MI-BCIs are rather unreliable, which is due at least in part to the use of inappropriate user-training procedures. Understanding the processes underlying user-training by modelling it computationally could enable us to improve MI-BCI training protocols and adapt the latter to the profile of each user. Indeed, we developed theoretical and conceptual models of BCI performances suggesting that the users' profiles does impact their performances [12]. Our objective is to create a statistical/probabilistic model of training that could explain, if not predict, the learning rate and the performances of a BCI user over training time using user's personality, skills, state and timing of the experiment. Preliminary analyses on previous data revealed positive correlations between MI-BCI performances and mental rotation scores among two of three different studies based on the same protocol [49]. This suggests that spatial abilities play a major role in MI-BCI users' abilities to learn to perform MI tasks, which is consistent with the literature.

Modeling and measuring users' skills at MI-BCI control: Studying and improving the reliability issue of BCI requires the use of appropriate reliability metrics to quantify both the classification algorithm and the BCI user's performances. So far, Classification Accuracy (CA) is the typical metric used for both aspects. However, we argued that CA is a poor metric to study BCI users' skills. Thus, we proposed a definition and new metrics to quantify such BCI skills for MI-BCIs, independently of any classification algorithm. By re-analyzing EEG data sets with such new metrics, we indeed confirmed that CA may hide some increase in MI-BCI skills or hide the user inability to self-modulate a given EEG pattern. On the other hand, our new metrics could reveal such skill improvements as well as identify when a mental task performed by a user was no different than rest EEG. This work was published in Journal of Neural Engineering [15].

Towards measuring the impact of attention: "Attention" is a generic word encompasses alertness and sustained attentions, referring to the intensity of attention (i.e., strength), as well as selective and divided attentions, referring to its selectivity (i.e., amount of monitored information). BCI literature indicates an influence of both users' attention traits and states (i.e., respectively stable and unstable attentional characteristics) on the ability to control a BCI. Though the types of attention involved remain unclear. Therefore, assessing which types of attention are involved during BCI use might provide information to improve BCI usability. Before testing this hypothesis, we first needed to assess if the different types of attention are recognizable using EEG. Our first results suggested that indeed, using machine learning, we can discriminate attention types for each other in EEG, at least when comparing them two by two [59].

The Influence of the experimenter: Through out the research and development process of MI-BCI, human supervision (e.g., experimenters or caregivers) plays a central role. People need to present the technology to users and ensure the smooth progress of the BCI learning and use. Though, very little is known about the influence they might have on their results. Such influence is to be expected as social and emotional feedback were shown to influence MI-BCI performances and user experience. Furthermore, literature from different fields indicate an effect of experimenters, and specifically their gender, on experiment outcome. Therefore, we assessed the impact of gender on MI-BCI performances, progress and user experience. An interaction of the runs, subjects gender and experimenters gender was found to have an impact on the performances of the subjects, suggesting users learn better with female experiments [30] (see Fig. 8).

7.9. Improving BCI user performance and training

Participants: Jelena Mladenovic, Léa Pillette, Thibaut Monseigne, Fabien Lotte

The potential of learning companions: As mentioned before, current BCI training protocols do not enable every user to acquire the skills required to use BCIs. We showed that learning companions were promising tools to increase BCI user experience during training, as well as to increase the performances of users who are more inclined to work in a group. Encouraged by these first results we investigated all the other potential benefits learning companions could bring to BCI training by improving the feedback, i.e., the information provided to the user, which is primordial to the learning process and yet have proven both theoretically and practically inadequate in BCI. From these considerations, some guidelines were drawn, open challenges identified and potential solutions were suggested to design and use learning companions for BCIs [29].

Figure 8. An EEG cap is being placed on the head of a subject by an experimenter on the right while another experimenter on the left is setting up the necessary software on the computer.

Active Inference for P300 speller: Brain Computer Interface (BCI) mostly relies, on one hand, on the stability of a person's mental commands, and on the other, on the machine's capacity to interpret those commands. As a person is naturally changing and adapting all the time, the machine becomes less successful in interpreting user's commands. In turn, the machine should be able to predict and minimize undesired user fluctuations. Moreover, it should build bottom-up information about the user through physiological input (EEG observations), and influence the user by providing optimal task (action) to minimize prediction error. A novel neuroscience approach, Active (Bayesian) Inference, is a very generic and flexible computational framework that can predict user intentions through a series of optimal actions and observations. On simulated data, we have shown that Active Inference has great potential to enable the machine to co-adapt with the user, and increase performance levels in a P300 speller BCI. We further tested Active Inference on real data, and show that active inference surpasses the standard algorithms while permitting the implementation of various cases of p300 speller BCI within one single framework [57].

Towards Congruent Feedback for BCI: Congruent visual environment in MI BCI has been researched in virtual reality, giving a sense of body ownership illusion, and showed to be more robust and improve performance. On the other hand, the effects of a congruent, purely audio environment, have not yet been explicitly explored in BCI. This inspired us to explore the benefits of a task-related (congruent) and synchronised audio feedback which would comply with the user's imagined movements. We investigate the potential of such an audio feedback congruent to the task, tackling the sensory illusion of presence by providing realistic audio feedback using natural sounds. Our preliminary results show the benefits of a congruent, audio MI feedback of feet (sound of footsteps in gravel) as opposed to no congruent feedback using abstract sound [50].

Neurofeedback of daytime alertness: Neurofeedback consists in providing a subject with information about his own EEG by means of a sensory feedback (visual, auditory ...) in real-time, in order to enable cognitive learning. In collaboration with SANPSY (Pellegrin Hospital/Univ. Bordeaux), we implemented a complete Neurofeedback solution as a proof of concept that aims to determine the level of effectiveness of Neurofeedback on daytime alertness ability. Indeed, excessive daytime sleepiness (EDS) is a common complaint associated with increased accidental risk. The usual countermeasures such as blue light, caffeine or nap have been shown to be effective but have limitations. With a test on five subjects, preliminary data showed that it was possible to learn how to regulate our own EEG activity with a short number of sessions (8 sessions of 40 min). Clinical trials to confirm these results should be initiated in the course of 2019.

7.10. Physiological computing

Participants: Jelena Mladenovic, Fabien Lotte

ElectroGastoGraphy: Recent research in the enteric nervous system, sometimes called the second brain, has revealed potential of the digestive system in predicting emotions. Even though people regularly experience changes in their gastrointestinal (GI) tract which influence their mood and behavior multiple times per day, robust measurements and wearable devices are not quite developed for such phenomena. However, other manifestations of the autonomic nervous system such as electrodermal activity, heart rate, and facial muscle movement have been extensively used as measures of emotions or in biofeedback applications, while neglecting the gut. In [28], we exposed electrogastrography (EGG), i.e., recordings of the myoelectric activity of the GI tract, as a possible measure for inferring human emotions.

EEG-based neuroergonomics: In collaboration with ISAE Toulouse, we explored the use of EEG to monitor cognitive processes in real flight situation, using dry EEG sensors. We showed that doing so is possible, however with low performances, given the strong noise in signals occurring in this challenging context [23]. In general, we presented in [17] and [40] how BCIs could be useful for neuroergonomics, i.e., to estimate user interfaces ergonomics quality from neurophysiological measures. Finally, in collaboration with RIKEN BSI, Japan, we showed that emotions could be monitored to some extent in EEG signals from multiple users watching the same emotional video clips at the same time. Interestingly, emotions decoding performances were increased by using the EEG data from several users compared to using EEG from each individual user [33].

REALOPT Project-Team

7. New Results

7.1. Improving Branch-and-Price Methods

We have made progress on stabilization techniques and math-heuristics that are essential components for generic Branch-and-Price methods.

The convergence of a column generation algorithm can be improved in practice by using stabilization techniques. Smoothing and proximal methods based on penalizing the deviation from the incumbent dual solution have become standards of the domain. Interpreting column generation as cutting plane strategies in the dual problem, we have analyzed [6] the mechanisms on which stabilization relies. In particular, the link is established between smoothing and in-out separation strategies to derive generic convergence properties. For penalty function methods as well as for smoothing, we describe proposals for parameter self-adjusting schemes. Such schemes make initial parameter tuning less of an issue as corrections are made dynamically. Such adjustments also allow to adapt the parameters to the phase of the algorithm. Extensive test reports validate our self-adjusting parameter scheme and highlight their performances. Our results also show that using smoothing in combination with penalty function yields a cumulative effect on convergence speed-ups.

Math heuristics have become an essential component in mixed integer programming (MIP) solvers. Extending MIP based heuristics, we have studied [8] generic procedures to build primal solutions in the context of a branch-and-price approach. As the Dantzig-Wolfe reformulation of a problem is typically tighter than that of the original compact formulation, heuristics based on rounding its linear programing (LP) solution can be more competitive. We focus on the so-called diving methods that used re-optimization after each LP rounding. We explore combination with diversification- intensification paradigms such as Limited Discrepancy Search, sub-MIPing, relaxation induced neighbourhood search, local branching, and strong branching. The dynamic generation of variables inherent to a column generation approach requires specific adaptation of heuristic paradigms. We manage to use simple strategies to get around these technical issues. Our numerical results on generalized assignment, cutting stock, and vertex coloring problems sets new benchmarks, highlighting the performance of diving heuristics as generic procedures in a column generation context and producing better solutions than state-of-the-art specialized heuristics in some cases.

7.2. Routing Problems

In [7] we deal with the Minimum Latency Problem (MLP), another variant of the well-known Traveling Salesman Problem in which the objective is to minimize the sum of waiting times of customers. This problem arises in many applications where customer satisfaction is more important than the total time spent by the server. This paper presents a novel branch-and-price algorithm for MLP that strongly relies on new features for the *ng*-path relaxation, namely: (1) a new labeling algorithm with an enhanced dominance rule named multiple partial label dominance; (2) a generalized definition of *ng*-sets in terms of arcs, instead of nodes; and (3) a strategy for decreasing *ng*-set sizes when those sets are being dynamically chosen. Also, other elements of efficient exact algorithms for vehicle routing problems are incorporated into our method, such as reduced cost fixing, dual stabilization, route enumeration and strong branching. Computational experiments over TSPLIB instances are reported, showing that several instances not solved by the current state-of-the-art method can now be solved.

We also considered a family of Vehicle Routing Problem (VRP) variants that generalize the classical Capacitated VRP by taking into account the possibility that vehicles differ by capacity, costs, depot allocation, or even by the subset of customers that they can visit. In [5] we propose a branch-cut-and-price algorithm that adapts advanced features found in the best performing exact algorithms for homogeneous fleet VRPs. The original contributions include: (i) the use of Extended Capacity Cuts, defined over a pseudo-polynomially

large extended formulation, together with Rank-1 Cuts, defined over the Set Partitioning Formulation; (ii) the concept of vehicle-type dependent memory for Rank-1 Cuts; and (iii) a new family of lifted Extended Capacity Cuts that takes advantage of the vehicle-type dependent route enumeration. The algorithm was extensively tested in instances of the literature and was shown to be significantly better than previous exact algorithms, finding optimal solutions for many instances with up to 200 customers and also for some larger instances. Several new best solutions were found too.

We examined the robust counterpart of the classical Capacitated Vehicle Routing Problem (CVRP) in [13], [20]. We considered two types of uncertainty sets for the customer demands: the classical budget polytope introduced by Bertsimas and Sim (2003), and a partitioned budget polytope proposed by Gounaris et al. (2013). We showed that using the set-partitioning formulation it is possible to reformulate our problem as a deterministic heterogeneous vehicle routing problem. Thus, many state-of-the-art techniques for exactly solving deterministic VRPs can be applied for the robust counterpart, and a modern branch-and-cut-and-price algorithm can be adapted to our setting by keeping the number of pricing subproblems strictly polynomial. More importantly, we introduced new techniques to significantly improve the efficiency of the algorithm. We present analytical conditions under which a pricing subproblem is infeasible. This result is general and can be applied to other combinatorial optimization problems with knapsack uncertainty. We also introduced robust capacity cuts which are provably stronger than the ones known in the literature. Finally, a fast iterated local search algorithm was proposed to obtain heuristic solutions for the problem. Using our branch-and-cut-and-price algorithm incorporating existing and new techniques, we were able to solve to optimality all but one open instances from the literature.

In [14], we have generalized our Branch-Cut-and-Price algorithm to solve other Vehicle Routing and related combinatorial optimization problems, as Generalized Assignment, Bin Packing, and Vector Packing. Our generic approach outperformed several problem specific algorithms.

7.3. Scheduling and Clustering Problems

In [19] we consider the unrelated parallel machine scheduling problem with setup times to minimize a general objective function. In this work we present a novel exact algorithm that is capable of solving this problem $R|r_j, s_{ij}^k| \sum f_j(C_j)$ and the large class of problems that can be derived as particular cases from it. The proposed algorithm consists of a branch-cut-and-price approach that combines several features such as non-robust cuts, strong branching, reduced cost fixing and dual stabilization. To our knowledge, this is the first exact algorithm for unrelated machines with earliness and/or tardiness criteria that can solve consistently instances with more than 20 jobs. We report improved bounds for instances of problems $R|r_j, s_{ij}^k| \sum w'_j E_j + w_j T_j$ and $R|| \sum w'_j E_j + w_j T_j$ with up to 80 and 120 jobs, respectively.

A cross-docking terminal is a transshipment facility in supply chains, where products transported by inbound trucks are unloaded at inbound doors, sorted, and reloaded on outbound trucks at outbound doors. In [16], we address the truck-to-door scheduling problem at a multi-door cross-docking terminal where temporary storage is considered. We propose two types of time-indexed formulation for the problem to assign trucks to dock doors and determine their arrival and departure times so that tadiness and earliness as well as unsatisfied demand are minimized. We examine the effectiveness of the proposed formulations by numerical experiment.

7.4. Scheduling and placement for HPC

In High Performance Computing, heterogeneity is now the norm with specialized accelerators like GPUs providing efficient computational power. Resulting complexity led to the development of task-based runtime systems, where complex computations are described as task graphs, and scheduling decisions are made at run-time to perform load balancing between all resources of the platforms. In [2], we consider the problem of developing good scheduling strategies, even at the scale of a single node, and analyzing them both theoretically and in practice is expected to have a very high impact on the performance of current HPC systems. The special case of two kinds of resources, typically CPUs and GPUs is already of great practical interest. The scheduling policy Hetero-Prio has been proposed in the context of fast multipole computations (FMM), and has been

extended to general task graphs with very promising results. In this paper, we provide a theoretical study of the performance of HeteroPrio, by proving approximation bounds compared to the optimal schedule, both in the case of independent tasks and in the case of general task graphs. Interestingly, our results establish that spoliation (a technique that enables resources to restart uncompleted tasks on another resource) is enough to prove bounded approximation ratios for a list scheduling algorithm on two unrelated resources, which is known to be impossible otherwise. This result holds true both for independent and dependent tasks graphs. Additionally, we provide an experimental evaluation of HeteroPrio on real task graphs from dense linear algebra computation, that establishes its strong performance in practice.

In [1], we consider the problem of partitioning a matrix into a set of sub-matrices, that has received increased attention recently and is crucial when considering dense linear algebra and kernels with similar communication patterns on heterogeneous platforms. The problem of load balancing and minimizing communication is traditionally reducible to an optimization problem that involves partitioning a square into rectangles. This problem has been proven to be NP-Complete for an arbitrary number of partitions. In this paper, we present recent approaches that relax the restriction that all partitions be rectangles. The first approach uses an original mathematical technique to find the exact optimal partitioning. Due to the complexity of the technique, it has been developed for a small number of partitions only. However, even at a small scale, the optimal partitions found by this approach are often non-rectangular and sometimes non-intuitive. The second approach is the study of approximate partitioning methods by recursive partitioning algorithms. In particular we use the work on optimal partitioning to improve preexisting algorithms. In this paper we discuss the different perspectives it opens and present two algorithms, SNRPP which is a $\sqrt{3/2}$ approximation, and NRPP which is a $2/\sqrt{3}$ approximation. While sub-optimal, this approach works for an arbitrary number of partitions. We use the first exact approach to analyze how close to the known optimal solutions the NRRP algorithm is for small numbers of partitions.

In [12], we consider the problem of data allocation when performing matrix multiplication on a heterogeneous node, with multicores and GPUs. Classical (cyclic) allocations designed for homogeneous settings are not appropriate, but the advent of task-based runtime systems makes it possible to use more general allocations. Previous theoretical work has proposed square and cube partitioning algorithms aimed at minimizing data movement for matrix multiplication. We propose techniques to adapt these continuous square partitionings to allocating discrete tiles of a matrix, and strategies to adapt the static allocation at run-time. We use these techniques in an implementation of Matrix Multiplication based on the StarPU runtime system, and we show through extensive experiments that this implementation allows to consistently obtain a lower communication volume while improving slightly the execution time, compared to standard state-of-the-art dynamic strategies.

7.5. Convergence between HPC and Data Science

In [11] paper we concentrate on a crucial parameter for efficiency in Big Data and HPC applications: data locality. We focus on the scheduling of a set of independent tasks, each depending on an input file. We assume that each of these input files has been replicated several times and placed in local storage of different nodes of a cluster, similarly of what we can find on HDFS system for example. We consider two optimization problems, related to the two natural metrics: makespan optimization (under the constraint that only local tasks are allowed) and communication optimization (under the constraint of never letting a processor idle in order to optimize makespan). For both problems we investigate the performance of dynamic schedulers, in particular the basic greedy algorithm we can for example find in the default MapReduce scheduler. First we theoretically study its performance, with probabilistic models, and provide a lower bound for communication metric and asymptotic behaviour for both metrics. Second we propose simulations based on traces from a Hadoop cluster to compare the different dynamic schedulers and assess the expected behaviour obtained with the theoretical study.

In [10], we consider the use of Burst-Buffers, that are high throughput, small size intermediate storage systems typically based on SSDs or NVRAM that are designed to be used as a potential buffer between the computing nodes of a supercomputer and its main storage system consisting of hard drives. Their purpose is to absorb the bursts of I/O that many HPC applications experience (for example for saving checkpoints or data from

intermediate results). In this paper, we propose a probabilistic model for evaluating the performance of Burst-Buffers. From a model of application and a data management strategy, we build a Markov chain based model of the system, that allows to quickly answer issues about dimensioning of the system: for a given set of applications, and for a given Burst-Buffer size and bandwidth, how often does the buffer overflow? We also provide extensive simulation results to validate our modeling approach.

7.6. Energy management

In [9], we consider energy management optimization problems in a future wherein an interaction with micro-grids has to be accounted for. We model this interaction through a set of contracts between the generation companies owning centralized assets and the micro-grids. We formulate a general stylized model that can, in principle, account for a variety of management questions such as unit-commitment. The resulting model, a bilevel stochastic mixed integer program will be numerically tackled through a novel preprocessing procedure. As a result the solution for the bilevel (or single leader multiple follower) problem will be neither " optimistic " nor " pessimistic ". We numerically evaluate the difference of the resulting solution with the " optimistic " solution. We also demonstrate the efficiency and potential of our methodology on a set of numerical instances.

7.7. Network Design Problems

The delivery of freight from manufacturing platforms to demand zones is often managed through one or more intermediate locations where storing, merging, transshipment and consolidation activities are performed. In [56], we design a Two-Echelon Distribution Network that helps synchronise different flows of product. Under demand uncertainty, our model integrates decisions on the locations and the size of second echelon facilities an decisions on the flows assignment between the echelons, and on delivery routes to serve the demand zones.

In [33], we study the k-edge-connected L-hop-constrained network design problem. Given a weighted graph $G = (V, E)$, a set D of pairs of nodes, two integers $L \geq 2$ and $k \geq 2$, the problem consists in finding a minimum weight subgraph of G containing at least k edge-disjoint paths of length at most L between every pair $\{s, t\} \in D$. We consider the problem in the case where $L = 2$, 3 and $|D| \geq 2$. We first discuss integer programming formulations introduced in the literature. Then, we introduce new integer programming formulations for the problem that are based on the transformation of the initial undirected graph into directed layered graphs. We present a theoretical comparison of these formulations in terms of LP-bound. Finally, these formulations are tested using CPLEX and compared in a computational study for $k = 3, 4, 5$.

In [72], we consider a multi-layer network design model arising from a real-life telecommunication application where traffic routing decisions imply the installation of expensive nodal equipment. Customer requests come in the form of bandwidth reservations for a given origin destination pair. Bandwidth demands are expressed as multiples of nominal granularities. Each request must be single-path routed. Grooming several requests on the same wavelength and multiplexing wavelengths in the same optical stream allow a more efficient use of network capacity. However, each addition or withdrawal of a request from a wavelength requires optical to electrical conversion and the use of cross-connect equipment with expensive ports of high densities. The objective is to minimize the number of required ports of the cross-connect equipment. We deal with backbone optical networks, therefore with networks with a moderate number of nodes (14 to 20) but thousands of requests. Further difficulties arise from the symmetries in wavelength assignment and traffic loading. Traditional multi-commodity network flow approaches are not suited for this problem. Instead, four alternative models relying on Dantzig-Wolfe and/or Benders' decomposition are introduced and compared. The formulations are strengthened using symmetry breaking restrictions, variable domain reduction, zero-one discretization of integer variables, and cutting planes. The resulting dual bounds are compared to the values of primal solutions obtained through hierarchical optimization and rounding procedures. For realistic size instances, our best approaches provide solutions with optimality gap of approximately 5% on average in around two hours of computing time.

7.8. Packing and Cutting Problems

The two-dimensional knapsack problem consists in packing a set of small rectangular items into a given large rectangle while maximizing the total reward associated with selected items. In [3], we restrict our attention to packings that emanate from a k-stage guillotine-cut process. We introduce a generic model where a knapsack solution is represented by a flow in a directed acyclic hypergraph. This hypergraph model derives from a forward labeling dynamic programming recursion that enumerates all non-dominated feasible cutting patterns. To reduce the hypergraph size, we make use of further dominance rules and a filtering procedure based on Lagrangian reduced costs fixing of hyperarcs. Our hypergraph model is (incrementally) extended to account for explicit bounds on the number of copies of each item. Our exact forward labeling algorithm is numerically compared to solving the max-cost flow model in the base hyper-graph with side constraints to model production bounds. Benchmarks are reported on instances from the literature and on datasets derived from a real-world application.

Also we consider a variant of two-dimensional guillotine cutting-stock problem that arises when different bills of order (or batches) are considered consecutively. The raw material leftover of the last cutting pattern is not counted as waste as it can be reused for cutting the next batch. The objective is thus to maximize the length of the leftover. In [21] we propose a diving heuristic based on a Dantzig-Wolfe reformulation solved by column generation in which the pricing problem is solved using dynamic programming (DP). This DP generates so-called non-proper columns, *i.e.* cutting patterns that cannot participate in a feasible integer solution of the problem. We show how to adapt the standard diving heuristic to this "non-proper" case while keeping its effectiveness. We also introduce the partial enumeration technique, which is designed to reduce the number of non-proper patterns in the solution space of the dynamic program. This technique helps to strengthen the lower bounds obtained by column generation and improve the quality of solutions found by the diving heuristic. Computational results are reported and compared on classical benchmarks from the literature as well as on new instances inspired from industrial data. According to these results, proposed diving algorithms outperform constructive and evolutionary heuristics.

SISTM Project-Team

7. New Results

7.1. Mecanistic modeling

7.1.1. Methodology

Estimation methods in mechanistic models can be seen as an inverse problem, in which we want to recover the individual parameters that have produced some observations. In collaboration with the Inria MONC & M3DISIM team, we propose a method for estimation in ODE with mixed effects on parameters based on Kalman based filter (also known as linear quadratic estimation (LQE)) that consist in correcting the original dynamic at each time by a feedback control.

7.1.2. Applications

7.1.2.1. Applied to IL-7 therapy

We have developed two approaches to optimize the injection of IL-7 in HIV-infected patients: one based on a statistical model and one based on Piecewise Deterministic Markov Processes (PDMP).

Villain L, Commenges D, Pasin C, Prague M, Thiébaut R. Adaptive protocols based on predictions from a mechanistic model of the effect of IL7 on CD4 counts. Statistics in Medicine. 2018;38:221-235.

Pasin C, Dufour F, Villain L, Zhang H, Thiébaut R. Controlling IL-7 Injections in HIV-Infected Patients. Bulletin of Mathematical Biology. 2018;80:2349-2377.

7.1.2.2. Applied to development of new HIV immunotherapies and vaccines

HIV infection can be treated but not cured with combination antiretroviral therapy, and new therapies that instead target the host immune response to infection are now being developed. Two recent studies of such immunotherapies, conducted in an animal model (SIV-infected rhesus macaques), have shown that agents which target the innate immune receptor TLR7 along with recombining viral-vector vaccines can prevent or control the rebound in viremia that usually accompanies the discontinuation of antiretroviral drugs. However, the mechanism of action of these therapies remains unknown. In collaboration with Harvard School of public health and Harvard program for evolutionary dynamics, we delineate the best model and the best procedure for treatment effect selection in order to delineate the each of each immunotherapies and design subsequent trials.

7.1.2.3. Applied to development of Ebola vaccines

We have analysed data from three clinical trials conducted under the EBOVAC1 consortium in 4 different countries (UK, Kenya and Uganda/Tanzania) and assessing the safety and immunogenicity of prime-boost vaccine regimens combining one adenovirus-based vector and one modified vaccine Ankara. In particular, we have modeled the dynamics of the humoral immune response following the boost immunization and fitted both linear mixed models and ODEs-based mechanistic models to the antibody concentrations data. This analysis allowed the estimation of the durability of the antibody response, as well as the identification of factors of variability of the response.

Pasin C, Balelli I, Van Effelterre T, Bockstal V, Soloforosi L, Prague M, Douoguih M, Thiébaut R. Dynamics of the humoral immune response to a prime-boost Ebola vaccine: quantification and sources of variation. Under revision in Journal of Virology.

7.2. Statistical learning methods for high-dimensional data

7.2.1. Automatic analysis of cell populations

We have developed two different approaches to classify the cell populations according to the high dimensional data obtained with flow cytometry assays. The first one is based on an interesting development of Dirichlet processes and the second one is based on a simple tree classification providing very high performances:

Hejblum BP, Alkhassim C, Gottardo R, Caron F, Thiébaut R. Sequential Dirichlet process mixtures of multivariate skew t-distributions for model-based clustering of flow cytometry data. Annals of Applied Statistics. In press.

Commenges D, Alkhassim C, Gottardo R, Hejblum B, Thiébaut R. cytometree: A binary tree algorithm for automatic gating in cytometry analysis. Cytometry A. 2018;93:1132-1140.

7.2.2. Missing Value Treatment in Longitudinal High Dimensional Supervised Problems

Poor blood sample quality introduces a large number of missing values in the context of sequencing data production. Furthermore, strong technical biases may force the analyst to remove the considered sequenced samples. Then entire day dependent data are then missing.

We have developed a regularized SVD based method using the temporal structure (through multi-block approach) of the missing values to estimate missing values with the objective of predicting uni-variate or multivariate regression responses but also classification problems. That regularizing method uses soft-thresholding on the co-variance matrices implying natural variable selection of covariate and response through a single hyper-parameter to be tuned.

7.2.3. Left-censored data treatment in High Dimensional Supervised Problems

Data could be censored either by the limit of detection or the limit of quantification. We have developed a regularized method for handling high-dimensional exposure data in the presence of censored values in the field of HIV that could be applied to other fields.

Soret P, Avalos M, Wittkop L, Commenges D, Thiébaut R. Lasso regularization for left-censored Gaussian outcome and high-dimensional predictors. BMC Med Res Methodol. 2018 Dec 4;18(1):159.

7.3. Analysis of results from Clinical trials and cohorts

7.3.1. In the HIV field

We have performed the statistical analyses of the immunogenicity endpoints, including high dimensional assays such as gene expression (RNA Seq), of two HIV vaccine clinical trials: 1) ANRS VRI01, a randomized phase I/II trial evaluating for different prime boost vaccine strategies in healthy volunteers; 2) ANRS 159 LIGHT, a randomized phase II trial comparing a prime-boost therapeutic HIV vaccine strategy to placebo in HIV-infected patients undergoing antiretroviral treatment interruption. The results of each of these two trials have been presented as an oral presentation at the HIV R4P conference in Madrid in October 2018 (Richert L et al. and Lacabaratz C et al). Integrative statistical analyses using sPLS methods (as developed by the team) are currently ongoing to relate markers from different high-dimensional immunogenicity assays or virological assays to each other.

7.3.2. In the Ebola field

We have performed a review of all existing clinical trials available to evaluate Ebola vaccines in macaques and humans.

Gross L, Lhomme E, Pasin C, Richert L, Thiébaut R. Ebola vaccine development: Systematic review of pre-clinical and clinical studies, and meta-analysis of determinants of antibody response variability after vaccination. International Journal of Infectious Diseases. 2018. pii: S1201-9712(18)34457-6.

7.3.3. In cluster randomized trials

Accounting for missing outcome is highly important to recover unbiased results of treatment effects. Weighting approached are less common compared to multilevel multiple imputation to analyse clustered data with missing outcome. >In collaboration with Duke University, we compared the two approaches and evaluated their performances to conclude that weighted appraoch should be considered a viable strategy to account for missing outcomes in cluster randomized trials

In cluster randomized trials, it is often desirable to improve the understanding of intervention effects in the presence of dissemination/spillover. In collaboration with Rhode Island university and Yale University, we aims at proposing innovative approach to analyze the TasP ANRS 12249 trial.We proposed innovative methods to individual, disseminated and overal effect of a clustered intervention based on counterfactuals averages in the presence of dissemination.

7.4. Conferences

Members of the team were involved in several talks during conferences and colloquium.

- Big Data and Information Analytics 2018 BigDIA Conference – 17-19th December 2018, Houston, Texas, USA – Random Forests for high-dimensional longitudinal data [invited talk]
- Big Data and Information Analytics 2018 BigDIA Conference – 17-19th December 2018, Houston, Texas, USA – Analysis of high-dimensional longitudinal data from the French health-administrative databases using machine learning methods: performance comparison between LSTM neural networks and Lasso for the analysis of the risk of road traffic crashes associated with medicinal drug consumption [invited talk]
- Journées Recherche et Santé (JRS) Inserm Phénotypage Clinique et biologie des systèmes – 22 Novembre 2018, Institut Imagine Paris, France - Visualisation des omics et des données cliniques [invited talk]
- Center for Modelling and Simulation in the Biosciences (BIOMS) Symposium 2018 – 1-2nd Octobre 2018, Heidelberg, Germany - Finding the cells in the middle of the data [invited talk]
- Mc Gill University, Departement of Epidemiology Seminar – 24th September 2018, Montreal, Canada - Big Data In Vaccine Clinical Trials: A Dive Into Data Science [invited talk]
- 4th Neurepiomics summer school - 17-20th September 2018, Arcachon, France - Systems biology approaches applied to omics data [invited talk]
- International Workshop in honor of Daniel Commenges' 70th birthday- June 4-5, 2018, Bordeaux, France - The mechanistic model point of view of causality [invited talk]
- Population Approach Group Europe (PAGE meeting) - May 29 - June 1, 2018, Montreux, Switzerland - Use of mathematical modeling for optimizing and adapting immunotherapy protocols in HIV-infected patients [oral contribution]
- International Symposium on HIV and Emerging Infectious Diseases (ISHEID) - May 16-18, 2018, Marseille, France - In silico clinical research, keep it dynamic! [invited talk]
- International Conference « Statistics and Health » - 11-12 January 2018, Toulouse Institute of Mathematics, Toulouse - Use of mathematical modeling for accelerating and personalizing clinical trials [invited talk]
- Workshop Developments in cluster randomised and stepped wedge designs, London, 21-22 Nov. 2018 Performance of weighting as an alternative to multilevel multiple imputation in cluster randomized trials with missing binary outcomes [oral contribution]
- International Biometrics Society, Barcelona, Spain, 09-13 July 2018. Optimizing the administration of IL7. [oral contribution]
- International Biometrics Society, Barcelona, Spain, 09-13 July 2018. Fitting pharmacokinetics data with a population-based Kalman filters [oral contribution]
- International Biometrics Society, Barcelona, Spain, 09-13 July 2018. Random Forests for high-dimensional longitudinal data. [oral contribution]
- ENBIS European Network for Business and Industrial Statistics conference, 2-6 Sept 2018. Mechanistic modeling for in silico trials [oral contribution]
- IMI 10th Anniversary Scientific Symposium, Brussels, Belgium, 22-23 October 2018. Modelling the humoral immune response to Ebola vaccine [oral contribution]
- Montreal University, Faculté de Pharmacie, Departement of Mathematical pharmacology Seminar – 21th December 2018, Montreal, Canada - Mechanistic modeling for in silico trials [invited talk]

STORM Project-Team

7. New Results

7.1. InKS Programming Model

Existing programming models tend to tightly interleave algorithm and optimization in HPC simulation codes. This requires scientists to become experts in both the simulated domain and the optimization process and makes the code difficult to maintain and port to new architectures. The InKS programming model, developed within the context of the PhD. Thesis of Ksander Ejjaaouani [9], decouples these two concerns with distinct languages for each. The simulation algorithm is expressed in the InKS pia language with no concern for machine-specific optimizations. Optimizations are expressed using both a family of dedicated optimizations DSLs (InKS O) and plain C++. InKS O relies on the InKS pia source to assist developers with common optimizations while C++ is used for less common ones. Our evaluation demonstrates the soundness of the approach by using it on synthetic benchmarks and the Vlasov-Poisson equation. It shows that InKS offers separation of concerns at no performance cost.

7.2. Porting Chameleon on top of OpenMP

Chameleon is a dense linear algebra software relying on sequential task-based algorithms where sub-tasks of the overall algorithms are submitted to a Runtime system. Algorithms were implemented on top of several task-based runtime systems: QUARK, PaRSEC, and StarPU (for which there is also an optional heterogeneous implementation). In the context of PRACE-5IP, we introduced OpenMP as an alternative backend for these linear algebra kernels.

7.3. StarPU in Julia

Julia is a modern language for parallelism and simulation that aims to ease the effort for developping high performance codes. In this context, we have started to develop a StarPU binding inside Julia. It is now possible to launch StarPU kernels inside Julia, either given as libraries, or described in Julia directly. Julia has the advantage to simplify significantly the syntax required to express the task and data management in StarPU (defining a new scope for StarPU, using automatic desallocation of buffers, ...).

Besides, using the introspection properties of Julia, the kernels written in Julia are automatically translated in both C codes and CUDA codes. Some preliminary experimental results show encouraging speedups on some limited codes. This is a work in progress, developped with A.Juven and M.Keryell.

7.4. Simulation and Validation of Error Correction Code Algorithms

The AFF3CT Error Correction Code (ECC) development and experimentation toolchain reached a major milestone with the release of the 2.x branch. It incorporates a hefty set of new modules and capabilities:

- New code families: Reed-Solomon, Turbo Product Code (TCP);
- New decoders: Maximum Likelihood (ML), Chase, LDPC Approximate Min-Star (AMS), LDPC Vertical Layered, LDPC Peeling, LDPC Bit Flipping;
- New channels: Optical, Binary Erasure Channel (BEC), Binary Symmetric Channel (BSC);
- New modem: On-Off Keying (OOK).

This new branch comes with extensive documentation of all available parameters at any point in the chain (https://aff3ct.readthedocs.io). In the process of the new release, the source code has also be reorganized in a rational and compartimentalized way, in terms of modules, tasks and sockets. This refactoring streamlines the use AFF3CT as a library, building on the concept of tasks, with well defined input and output sockets.

7.5. Speeding-Up Error Correction Code Processing using a Portable SIMD Wrapper

Error correction code (ECC) processing has so far been performed on dedicated hardware for previous generations of mobile communication standards, to meet latency and bandwidth constraints. As the 5G mobile standard, and its associated channel coding algorithms, are now being specified, modern CPUs are progressing to the point where software channel decoders can viably be contemplated. A key aspect in reaching this transition point is to get the most of CPUs SIMD units on the decoding algorithms being pondered for 5G mobile standards. The nature and diversity of such algorithms requires highly versatile programming tools. We proposed the virtues and versatility of our MIPP SIMD wrapper in implementing a high performance portfolio of key ECC decoding algorithms on AFF3CT [8].

7.6. Runtime System Interoperability with StarPU

Parallel HPC applications increasingly build on multiple parallel libraries, which results in interferences if the parallel entities in the application and in the libraries it uses access computing resources in an uncoordinated manner. A set of resource management APIs has therefore been designed within the context of H2020 project INTERTWinE (see http://www.intertwine-project.eu/developer-hub/resource-manager), and implemented in the StarPU task-based runtime system developed by Team STORM, as well as in the OmpSs/Nanos 6 task-based runtime system developed at the Barcelona Supercomputing Center (BSC). It enables StarPU and OmpSs to interoperate within an application, along multiple scenarios such as *nested interoperability*, with a host runtime system executing parallel tasks over a guest runtime system, or *concurrent interoperability*, with several runtime systems dynamically sharing computing resources over the application lifespan.

7.7. Hierarchicals Tasks

The programming model of StarPU, namely the sequential task flow model, was successfuly used in several applicative areas and was able to achieve high performance. However, the submission process needed either to be completely static, in the sense that the whole task graph is submitted at once, or to be stopped from time to time in order to control the execution. To overcome these limitations, we have introduced a new paradigm which we call *hierarchical tasks* where the so-called *control tasks* allow to submit at runtime a task subgraph. By allowing the submission of some parts of the task graph to be delayed until the execution of the corresponding control task, this feature allows to timely and dynamically choose the right version of the computation task subgraph (e.g. OpenMP, StarPU, cuda, or sequential, etc. implementations).

The graph of control tasks provide a high-level description of the computations which allow to use and design sophisticated scheduling algorithms. Furthermore, the cost of managing the control graph being much smaller than the one of the computation task graph, relying on the hierarchical tasks scheme enhances the scalability of the runtime system and allows to parallelize the submission process. Finally, this mechanism represents an elegant way of tackling the granularity issues which represent a key problem for achieving high performance in a heterogeneous context. The specificity of our implementation is to nicely combine hierarchical tasks with data partitioning, without needless synchronisation points.

7.8. Load Balancing Management in a Distributed Task-Based Programming Model

Distributed task-based programming models such as StarPU optimize the execution of applications based on an initial distribution of data. The resulting computational load on each node may however evolve over the course of the application, to the point where this initial distribution of data leads becomes suboptimal. It becomes necessary to correct the distribution the distribution of data to rebalance the load among nodes. Tools such as Zoltan or ParMetis do exist to perform this rebalancing job. However, they cannot be employed without breaking the application execution flow, and force synchronizing steps in fundamentally asynchronous task parallelism paradigms. Within the context of the internship of Loïc Jouans, we proposed a mechanism to enable the detection of load imbalance as well as the application of corrective measures to rebalance it while preserving the execution asynchrony.

7.9. Task-based Execution Visualization

One of the purpose of task-based programming is to let asynchronous execution achieve extreme pipelining of operations. This however make it a real challenge to determine why an execution performs poorly, since the execution trace shows the mixture of unrelated tasks. With the the University of Grenoble, we have designed a visualization framework which allows to easily visualize different metrics of the execution trace and apply different techniques to reveal the execution behavior. This allowed to determine and fix some erratic behaviors for instance in the StarPU runtime system and OpenMPI communication library [13], [4]

7.10. Interprocedural Collectives Verification

The advent to exascale requires more scalable and efficient techniques to help developers to locate, analyze and correct errors in parallel applications. PARallel COntrol flow Anomaly CHecker (PARCOACH) is a framework that detects the origin of collective errors in applications using MPI and/or OpenMP. In MPI, such errors include collective operations mismatches. In OpenMP, a collective error can be a barrier not called by all tasks in a team. We have developed an extension of PARCOACH which improves its collective errors detection [11]. The new analysis is more precise and accurate than the previous one on different benchmarks and real applications.

7.11. Profile-Guided Scope-Based Data Allocation Method

The complexity of High Performance Computing nodes memory system increases in order to challenge application growing memory usage and increasing gap between computation and memory access speeds. As these technologies are just being introduced in HPC supercomputers no one knows if it is better to manage them with hardware or software solutions. Thus both are being studied in parallel. For both solutions, the problem consists in choosing which data to store on which memory at any time.

In this context, we propose a linear formulation of the data allocation problem. Moreover, we propose a new profile- guided scope-based approach which reduces the data allocation problem complexity, thus enhancing the precision of state of the art analyzes. Finally we have implemented our method in a framework made of GCC plugins, dynamic libraries and python scripts, allowing to test the method on several benchmarks. We have evaluated our method on an INTEL Knight's Landing processor. To this aim we have run LULESH, HydroMM, two hydrodynamic codes, and MiniFE, a finite element mini application. We have compared our framework performance over these codes to several straight- forward solutions: MCDRAM as a cache, in hybrid mode, in flat mode using numactl command and existing AutoHBW dynamic library [7]

7.12. Lightweight Containerization of Computing Resources

SwLoc is a library for flexible and generic partitioning of computing resources (CPU, accelerators). It allows applications to create contexts (i.e. resource partitions) and run parallel codes inside such lightweight containers. Many libraries developed using OpenMP, Pthreads or Intel TBB can ben executed concurrently with little or no modification. SwLoc also features dynamic context resizing capabilities that enables parallel applications to perform resource negotiation.

7.13. Adaptive Partitioning for Iterated Sequences of Irregular OpenCL Kernels

OpenCL defines a common parallel programming language for CPU and GPU devices, although writing tasks adapted to the devices, managing communication and load-balancing issues are left to the programmer. We propose [10] a static/dynamic approach for the execution of an iterated sequence of data-dependent kernels on a multi-device heterogeneous architecture. The method allows to automatically distribute irregular kernels onto multiple devices and tackles, without training, both load balancing and data transfers issues coming from hardware heterogeneity, load imbalance within the application itself and load variations between repeated executions of the sequence. Our evaluation on some benchmarks and a complex N-body application, SOTL, simulating the electromagnetic Coulomb force applied on particles, show the interest of our approach.

7.14. A compiler front-end for OpenMP's variants

OpenMP 5.0 introduced the concept of *variant*: a directive which can be used to indicate that a function is a variant of another existing *base function*, in a specific context (eg: `foo_gpu_nvidia` could be declared as a variant of `foo`, but only when executing on specific NVidia hardware).

In the context of PRACE-5IP, we want to leverage this construct to be able to take advantage of the StarPU heterogeneous scheduler through the interoperability layer between OpenMP and StarPU.

We started this work by implementing the necessary changes in the Clang front-end to support OpenMP's *variant*.

7.15. Combining Task-based Parallelism and Adaptive Mesh Refinement Techniques in Molecular Dynamics Simulations

Modern parallel architectures require applications to generate massive parallelism so as to feed their large number of cores and their wide vector units. We have revisited the extensively studied classical Molecular Dynamics N-body problem in the light of these hardware constraints. We have introduced Adaptive Mesh Refinement techniques to store particles in memory, and to optimize the force computation loop using multi-threading and vectorization-friendly data structures [14]. Our design is guided by the need for load balancing and adaptivity raised by highly dynamic particle sets, as typically observed in simulations of strong shocks resulting in material micro-jetting. We have analyzed performance results on several simulation scenarios, over 512 nodes equipped by Intel Xeon Phi Knights Landing (KNL) processors. Performance obtained with our OpenMP implementation outperforms state-of-the-art implementations (LAMMPS) on both steady and micro-jetting particles simulations. In the latter case, our implementation is 1.38 times faster on KNL.

These results were obtained in the context of joint work between Inria and CEA/DAM.

TADAAM Project-Team

7. New Results

7.1. Checkpointing Strategies for Adjoint Computation on Hierarchical Platforms

The Adjoint Computation problem can be split in two phases: the forward phase where functions are successively evaluated on a particular input, and a backward phase computing the gradient descent. In the backward phase, the outputs of the forward phase are used* for the corresponding backward computation. On very large problems, all the forward outputs can not be kept in the memory at the same time, and one has to decide which output should be checkpointed and which output should be recomputed later on. The goal is to minimize the number of recomputation when reversing an Adjoint Computation Graph.

Griewank and Walther proved that, for a given number of available checkpoints with negligible writing and reading costs, the schedule that minimizes the amount of recomputation uses a binomial checkpointing strategy. We have designed an optimal algorithm to tackle the more general problem where we don't have only one level of memory with negligible access cost, but a hierarchical storage architecture. Each level of memory has its own size, writing and reading cost. The problem becomes more complex, since, not only we have to decide if an output should be checkpointed, be we have to decide in which level of the memory it should be kept. A trade-off must be found between the cost of memory accesses and that of recomputations.

We have designed an exact algorithm providing the optimal checkpointing strategy for a given Adjoint Computation Graph size and a description of the Hierarchical Platform; as well as heuristics. These algorithms can be found in the Software DISK-REVOLVE and a paper describing them is under writing process.

7.2. Modeling Non-Uniform Memory Access on Large Compute Nodes with the Cache-Aware Roofline Model

The trend of increasing the number of cores on-chip is enlarging the gap between compute power and memory performance. This issue leads to design systems with heterogeneous memories, creating new challenges for data locality. Before the release of those memory architectures, the Cache-Aware Roofline Model [33] (CARM) offered an insightful model and methodology to improve application performance with knowledge of the cache memory subsystem.

With the help of the HWLOC library, we are able to leverage the machine topology to extend the CARM for modeling NUMA and heterogeneous memory systems, by evaluating the memory bandwidths between all combinations of cores and NUMA nodes. The new Locality Aware Roofline Model [5] (LARM) scopes most contemporary types of large compute nodes and characterizes three bottlenecks typical of those systems, namely contention, congestion and remote access. We also designed a hybrid memory bandwidth model to better estimate the roof when heterogeneous memories are involved or when read and write bandwidths differ.

This work has been achieved in collaboration with the authors of the CARM from Universidade de Lisboa.

7.3. Cross Platform Classification for Detecting Locality Sensitivity and Selecting Data and Threads Placement Strategy

Individual nodes composing High Performance Computing (HPC) systems embed complex multicore and manycore processors. At this scale, compute tasks and data placement can double or halve execution times with respectively trivial are wise placements. While state of the art placement solutions can offer good performance improvements, they failed to set up as standards in supercomputers software stack. Current solutions are rather directed toward data or thread driven static policies. Among existing or promising future placement solutions a deep evaluation of applications response to these had yet to be done in order to wisely choose the best one.

With a set of 37 HPC representative applications, three different HPC processors, and 51 state of the art characterization metrics we built thousands models to evaluate applications response to data and threads placement policies. Thanks to a thorough methodology, our models were able to predict applications sensitivity to locality and their preferred placement policy both on new platforms and new applications. In the first case we were able to achieve more than 75% accuracy while preferred policy predictions approach optimal speedups in the second case.

This work was conducted using the PlaFRIM experimental testbed, in collaboration with Thomas Ropars from Laboratoire d'Informatique de Grenoble.

Several leads can be taken toward an extension of this work. For instance, predictions can be improved with benchmark directed learning. Models interpretation can also be furthered studied to refine the design of application characterization metrics.

7.4. Co-scheduling HPC workloads on cache-partitioned CMP platforms

Co-scheduling techniques are used to improve the throughput of applications on chip multiprocessors (CMP), but sharing resources often generates critical interferences.

In collaboration with ENS Lyon and Georgia Tech, we looked at the interferences in the last level of cache (LLC) and use the *Cache Allocation Technology* (CAT) recently provided by Intel to partition the LLC and give each co-scheduled application their own cache area. We considered m iterative HPC applications running concurrently and answer the following questions: (i) how to precisely model the behavior of these applications on the cache partitioned platform? and (ii) how many cores and cache fractions should be assigned to each application to maximize the platform efficiency? Here, platform efficiency is defined as maximizing the performance either globally, or as guaranteeing a fixed ratio of iterations per second for each application. Through extensive experiments using CAT, we demonstrated the impact of cache partitioning when multiple HPC application are co-scheduled onto CMP platforms. [13]

7.5. Memory Footprint of Locality Information on Many-Core Platforms

Exploiting the power of HPC platforms requires knowledge of their increasingly complex hardware topologies. Multiple components of the software stack, for instance MPI implementations or OpenMP runtimes, now perform their own topology discovery to find out the available cores and memory, and to better place tasks based on their affinities.

We studied the impact of this topology discovery in terms of memory footprint. Storing locality information wastes an amount of physical memory that is becoming an issue on many-core platforms on the road to exascale.

We demonstrated that this information may be factorized between processes by using a shared-memory region. Our analysis of the physical and virtual memories in supercomputing architectures showed that this shared region can be mapped at the same virtual address in all processes, hence dramatically simplifying the software implementation. [19]

Our implementation in HWLOC and Open MPI showed a memory footprint that does not increase with the number of MPI ranks per node anymore. Moreover the job launch time decreased by more than a factor of 2 on an Intel Knights Landing Xeon Phi and on a 96-core NUMA platform.

7.6. New abstraction to manage hardware topologies in MPI applications

Since the end of year 2016, we have been working on new abstractions and mechanisms that can allow the programmer to take advantage of the underlying hardware topology in their parallel applications developed in MPI. For instance, taking into account the intricate network/memory hierarchy can lead to substantial improvements in communication performance and reduce altogether the overall execution time of the application. However, it is important to find the relevant level of abstraction, as too much details are not usable practically because the programmer is not a a hardware specialist most of the time. Also, MPI being hardware-agnostic,

it is important to find means to use the hardware specifics without being tied to a particular architecture or hardware design.

With these goals in mind, we proposed the HSPLIT (see Section 6.1) libray that implements a solution based on a well-known MPI concept, the *communicators* (that can be seen as groups of communicating processes) [7], [19]. With HSPLIT, each level in the hardware hierarchy is accessible through a dedicated communicator. In this way, the programmer can leverage the underlying hierarchy in their application quite simply . The current implementation of HSPLIT is based on both HWLOC and NETLOC.

This work led to the creation of a new active working group within the MPI Forum, coordinated and led by Inria.

Also, this work has lead to the joint developement of the Hippo software with the CERFACS. Thanks to this piece of software, hybrid OpenMP/MPI applications can leverage the underlying physical hierarchy in order to better place MPI processes and OpenMP threads. This is particularly useful in a context where the application is composed of several kernels that use their own placement and mapping policy for processes and threads to achieve the best performance. Thanks to HSPLIT and HWLOC, CERFACS is now able to write codes in a more portable fashion without to solely rely on interactions of OpenMP and MPI runtimes for mapping and binding of processes/threads management.

7.7. Scheduling Parallel Tasks under Multiple Resources: List Scheduling vs. Pack Scheduling

Scheduling in High-Performance Computing (HPC) has been traditionally centered around computing resources (e.g., processors/cores). The ever-growing amount of data produced by modern scientific applications start to drive novel architectures and new computing frameworks to support more efficient data processing, transfer and storage for future HPC systems. This trend towards data-driven computing demands the scheduling solutions to also consider other resources (e.g., I/O, memory, cache) that can be shared amongst competing applications. In this paper, we study scheduling HPC applications while exploring the availability of multiple resources that could impact their performance. The goal is to minimize the overall execution time, or makespan, for a set of moldable tasks under multi-resource constraints. Two scheduling paradigms, namely, list scheduling and pack scheduling, are compared through both theoretical analyses and experimental evaluation. Theoretically, we prove, for several algorithms falling in the two scheduling paradigms, tight approximation ratios that increase linearly with the number of resource types. As the complexity of the direct solutions grows exponentially with the number of resource types, we also design a strategy to indirectly solve the problem via a transformation to a single-resource problem, which can significantly reduce the algorithms' running times without compromising their approximation ratios. Experiments conducted on Intel Knights Landing with two types of resources (processor cores and high-bandwidth memory) and simulations designed on more resource types confirm the benefit of the transformation strategy and show that pack-based scheduling, despite having a slightly worse theoretical bound, offers a practically promising and easy-to-implement solution, especially when managing a large number of resources. [20]

7.8. Sizing Burst-Buffers efficiently

Burst-Buffers are high throughput, small size intermediate storage systems typically based on SSDs or NVRAM that are designed to be used as a potential buffer between the computing nodes of a supercomputer and its main storage system consisting of hard drives. Their purpose is to absorb the bursts of I/O that many HPC applications experience (for example for saving checkpoints or data from intermediate results). In this paper, we propose a probabilistic model for evaluating the performance of Burst-Buffers. From a model of application and a data management strategy, we build a Markov-chain-based model of the system, that allows us to quickly answer issues about dimensioning of the system: for a given set of applications, and for a given Burst-Buffers size and bandwidth, how often does the buffer overflow? We also provide extensive simulation results to validate our modeling approach. [12], [25]

7.9. Scheduling for Neurosciences

In this project in collaboration with the Vanderbilt University, we are interested in scheduling stochastic jobs (originating from Neuroscience applications) on a reservation-based platform. Specifically, we consider jobs whose execution time follows a known probability distribution. The platform is reservation-based, meaning that the user has to request fixed-length time slots. The cost depends on both the request duration and the actual execution time of the job. A reservation strategy is a sequence of increasing-length reservations, which are paid for until one of them allows the job to successfully complete. The goal is to minimize the total expected cost of the strategy. We provide some properties of the optimal solution, which we characterize up to the length of the first reservation. We evaluate these heuristics using two different platform models and cost functions: The first one targets a cloud-oriented platform (e.g., Amazon AWS) using jobs that follow a large number of usual probability distributions (e.g., Uniform, Exponential, LogNormal, Weibull, Beta), and the second one is based on interpolating traces from a real neuroscience application executed on an HPC platform. [14], [27]

7.10. Process Affinity, Metrics and Impact on Performance

Process placement, also called topology mapping, is a well-known strategy to improve parallel program execution by reducing the communication cost between processes. It requires two inputs: the topology of the target machine and a measure of the affinity between processes. In the literature, the dominant affinity measure is the communication matrix that describes the amount of communication between processes. The goal of this work is to study the accuracy of the communication matrix as a measure of affinity. We have done an extensive set of tests with two fat-tree machines and a 3d-torus machine to evaluate several hypotheses that are often made in the literature and to discuss their validity. First, we have checked the correlation between algorithmic metrics and the performance of the application. Then, we have checked whether a good generic process placement algorithm never degrades performance. And finally, we have seen whether the structure of the communication matrix can be used to predict gain [16].

7.11. Scheduling bi-colored-chains

In high performance computing, platform are shared by concurrent applications, each able to work with immense amount of data. As the file system is shared, we need to tackle congestion problems. One way to avoid increased I/O duration is to schedule the tasks with regards to their requests.

We proposed a theoretical model, bi-colored chains, that models applications with two alternating phases on distinct resources. After showing that minimizing the makespan with this model is a NP-complete problem in most cases. We studied particular cases, especially periodic applications and periodic schedule and provided approximation algorithms. This model will be developed in a PhD that started this fall, and enrich with practical data from simulations.

An extended intership report is available here: [31].

7.12. Experimenting task-based runtimes on a legacy Computational Fluid Dynamics code with unstructured meshes

Advances in high performance computing hardware systems lead to higher levels of parallelism and optimizations in scientific applications and more specifically in computational fluid dynamics codes. To reduce the level of complexity that such architectures bring while attaining an acceptable amount of the parallelism offered by modern clusters, the task-based approach has gained a lot of popularity recently as it is expected to deliver portability and performance with a relatively simple programming model. In this work, we have reviewed and presented the process of adapting part of Code Saturne, a legacy code at EDF R&D into a task-based form using the PARSEC (Parallel Runtime Scheduling and Execution Control) framework. We have first shown show the adaptation of our prime algorithm to a simpler form to remove part of the complexity of our code and then present its task-based implementation. We then have compared performance of various forms of our code and discuss the perks of task-based runtimes in terms of scalability, ease of incremental deployment in a legacy CFD code, and maintainability [8].

7.13. Progress threads placement for overlapping MPI non-blocking collectives using simultaneous multi-threading

Non-blocking collectives have been proposed so as to allow communications to be overlapped with computation in order to amortize the cost of MPI collective operations. To obtain a good overlap ratio, communications and computation have to run in parallel. To achieve this, different hardware and software techniques exists. Using dedicated cores for progress threads is one of them. However, some CPUs provide Simultaneous Multi-Threading, which is the ability for a core to have multiple hardware threads running simultaneously, sharing the same arithmetic units. We propose [18], [3] to use SMT to run progress threads to avoid dedicated cores allocation. We have run benchmarks on Haswell processors, using its Hyper-Threading capability, and get good results for both performance and overlap for inter-node communications. However, we have shown that Simultaneous Multi-Threading for intra-communications leads to bad performances due to contention on cache.

7.14. Dynamic placement of progress thread for overlapping MPI non-blocking collectives on manycore processor

To amortize the cost of MPI collective operations, non-blocking collectives have been proposed so as to allow communications to be overlapped with computation. Unfortunately, collective communications are more CPU-hungry than point-to-point communications and running them in a communication thread on a single dedicated CPU core makes them slow. On the other hand, running collective communications on the application cores leads to no overlap. To address these issues, we proposed [28], [17], [21], [3] an algorithm for tree-based collective operations that splits the tree between communication cores and application cores. To get the best of both worlds, the algorithm runs the short but heavy part of the tree on application cores, and the long but narrow part of the tree on one or several communication cores, so as to get a trade-off between overlap and absolute performance. We provided a model to study and predict its behavior and to tune its parameters. We implemented it in the MPC framework, which is a thread-based MPI implementation. We have run benchmarks on manycore processors such as the KNL and Skylake and got good results both in terms of performance and overlap.

7.15. Multi-criteria graph partitioning

The inclusion of multi-constraint graph partitioning algorithms in SCOTCH resulted in the obtainment of balanced multi-constraint partitions for a simulation software used in an industrial context [15]. This protype version is being transferred into the trunk of the SCOTCH package. Much of this year's software development has been devoted to the refactoring of the multi-threading management of the sequential version of the SCOTCH library.