Activity Report 2016

Section New Results

<p style="text-align: center; color: red;">**CAGIRE Project-Team**</p>

7. New Results

7.1. Heat transfer for effusion flows

The conjugate heat-transfer problem of a flow around a multi- perforated plate under realistic conditions has been addressed by the coupling of the LES-AVBP solver for the flow and the AVTP for solving the heat equation in the solid. A description of the topology of the heat exchange has been realized for the aspiration and injection sides of the walls as well as in the inner side of the holes. This work highlights the potential of such a fluid-solid coupling strategy in the description of the heat exchange distribution for combustor liners. Different analytical expressions have been assessed for each category of exchange surface.[15]

7.2. All-Mach numerical fluxes

This study was split into three self-consistent parts. In the first one, the low Mach number problem through a linear analysis of a perturbed linear wave equation was defined and analyzed. Then, we show how to modify Godunov type schemes applied to the linear wave equation to make this scheme accurate at any Mach number. This allows to define an all Mach correction and to propose a linear all Mach Godunov scheme for the linear wave equation. In the second one, we apply the all Mach correction proposed previously to the case of the non-linear barotropic Euler system when the Godunov type scheme is a Roe scheme. A linear stability result is proposed and a formal asymptotic analysis justifies the construction in this non-linear case by showing how this construction is related with the linear analysis. At last, we apply the all Mach correction to the case of the full Euler compressible system. Numerous numerical results justify the theoretical results and show that the obtained all Mach Godunov type schemes are both accurate and stable for all Mach numbers. We also underline that the proposed approach can be applied to other schemes and allows to justify other existing all Mach schemes.

7.3. Extension and validation of the EB-RSM model

The EB-RSM RANS turbulence model, an innovative model based on second moment closure, has been developed for almost 15 years and is now gradually deployed in the industrial practise. It is already implemented in several industrial codes (Code_Saturne, StarCCM+, EZNSS), as well as the open-source code OpenFOAM. In collaboration with industrial partners, the model is now being confronted to more and more complex industrial configurations: Wall-cooling using impinging jets; Measurement/control of head losses in pipes or injectors via local restrictions of the section (diaphragms); Turbine blade cooling by pin matrices; Control of boundary layer separation by local blowing to exploit the Coanda effect; Wing-tip vortices around airfoils representative of the spoilers of Formula One racing cars; Open-water propeller. All these results confirm the interest of the model compared to well-established models, and its numerical robustness.

7.4. Creation of a database of a direct numerical simulation of a jet in cross flow with and without gyration and with non isothermal flows

This year, we performed the direct numerical simulation of a jet in cross flow without gyration and with a 90° skidding with respect to the cross flow, and with a cross flow 800 Kelvin hotter than the jet. The Sutherland law was implemented for accounting for viscous effects in the non isothermal case and was validated. Then direct numerical simulations have been performed, by using synthetic eddy methods for inlet boundary conditions. Third order discretization was used. A limiter on the density was also used for damping oscillations in strong shear layers, which in this case include both large density and velocity gradients.

The database contains the mean flow at all points, the Reynolds tensor at the degrees of freedom, and the time dependent data at some probes.

CARDAMOM Project-Team

6. New Results

6.1. High order discretizations on unstructured meshes

- Participants: Héloise Beaugendre, Cécile Dobrzynski, Léo Nouveau, Mario Ricchiuto, Quentin Viville
- Corresponding member: Héloise Beaugendre

Our work on high order unstructured discretizations this year has pursued three main avenues:

- We have extended the team's previous work on the consistent residual based approximation of viscous flow equations to the framework of Immersed Boundary Methods (IBM). This is an increasingly popular approach in Computational Fluid Dynamics as it simplifies the mesh generation problem. In our work, we consider a technique based on the addition of a penalty term to the Navier-Stokes equations to account for the wall boundary conditions. To adapt the residual distribution method method to the IBM, we developed a new formulation based on a Strang splitting appproach in time. This approach, couples in a fully consistent manner an implicit asymptoticly exact integration procedure of the penalization ODE, with the explicit residual distribution discretization for the Navier-Stokes equations, based on the method proposed in [122]. The ODE integrator provides an operator which is exact up to orders η^2, with η the penalty parameter assuming values of the order of 10^{-10}. To preserve the accuracy of the spatial discretization in the Navier-Stokes step, we have introduced, in vicinity of the penalised region, a modification of the solution gradient reconstruction required for the evaluation of the viscous fluxes. We have shown formally and numerically that the approach proposed is second order accurate for smooth solutions, and shown its potential when combined with unstructured mesh adaptation strategies w.r.t. the (implicitly described) solid walls [16]. This approach has been combined with $r-$adaptation techniques to account for moving bodies and validated on simulations involving flapping wings, and computations of ices block trajectories in the framework of the STORM project [56], [46];

- Another research axis consists in proposing a novel approach that allows to use p-adaptation with continuous finite elements. Under certain conditions, primarily the use of a residual distribution scheme, it is possible to avoid the continuity constraint imposed to the approximate solution, while still retaining the advantages of a method using continuous finite elements. The theoretical material, the complete numerical method and practical results show as a proof of concept that p-adaptation is possible with continuous finite elements. This year, we extended the p-adaptation method to Navier-Stokes equations and coupled it with immersed boundary method.

- We have studied the high order approximation of problems with dispersion and suggested a route allowing to construct high order methods (up to order 4) allowing to obtain the same accuracy for the solution, and for its first and second order derivatives. Initial validation for the approach proposed has been shown for the time dependent KdV equations [14], [49].

6.2. High order mesh generation and mesh adaptation

- Participants: Luca Arpaia , Cécile Dobrzynski, Ghina El Jannoun, Léo Nouveau, Mario Ricchiuto
- Corresponding member: Cécile Dobrzynski

This year several new algorithmic improvements have been obtained which will allow to enhance our meshing tools:

- We have enhanced our work on r-adaptation techniques for time dependent equations. These techniques are based on mesh deformations obtained by solving continuous differential equations for the local displacements. These equations are controlled by an error monitor. Several improvements have been made. We have studied in depth the formulation of the coupling of the mesh movement with the flow solver. We have found that for both finite volume and residual distribution methods, a coupling of mesh and solution evolution (by means of an ALE method) provides accuracy enhancements, and is to be preferred to a simpler adapt-project-evolve approach. The method has been fully tested in two space dimensions and preliminary results have been performed in three dimensions. We have applied this technic to immersed boundary methods to compressible simulations. For problems with source terms, and in particular problems admitting some important physical invariants as the shallow water equations, we have solved the conflict between the conservation of either mass or the invariant, allowing for the conservation of both quantities up to machine accuracy. In parallel we have proposed a modified formulation of an elasticity equation allowing to reduce the nonlinearity of the mesh PDE to the force imposed in the right hand side. Initial validation has been shown in [56] and in the PhD of L. Nouveau ;

6.3. Uncertainty Quantification and robust design optimization

- Participants: Andrea Cortesi , Pietro Marco Congedo, Nassim Razaaly, Sanson Francois
- Corresponding member: Pietro Marco Congedo

We have developed an efficient sparse polynomial decomposition for sensitivity analysis and for building a surrogate in a problems featuring a large number of parameters. The Polynomial Dimensional Decomposition (PDD) is employed in this work for the global sensitivity analysis and uncertainty quantification (UQ) of stochastic systems subject to a moderate to large number of input random variables. Due to the intimate connection between the PDD and the Analysis of Variance (ANOVA) approaches, PDD is able to provide a simpler and more direct evaluation of the Sobol sensitivity indices, when compared to the Polynomial Chaos expansion (PC). Unfortunately, the number of PDD terms grows exponentially with respect to the size of the input random vector, which makes the computational cost of standard methods unaffordable for real engineering applications. In order to address the problem of the curse of dimensionality, this work proposes essentially variance-based adaptive strategies aiming to build a cheap meta- model (i.e. surrogate model) by employing the sparse PDD approach with its coefficients computed by regression. Three levels of adaptivity are carried out : 1) the truncated dimensionality for ANOVA component functions, 2) the active dimension technique especially for second- and higher-order parameter interactions, and 3) the stepwise regression approach designed to retain only the most influential polynomials in the PDD expansion. During this adaptive procedure featuring stepwise regressions, the surrogate model representation keeps containing few terms, so that the cost to resolve repeatedly the linear systems of the least-squares regression problem is negligible. The size of the finally obtained sparse PDD representation is much smaller than the one of the full expansion, since only significant terms are eventually retained. Consequently, a much smaller number of calls to the deterministic model is required to compute the final PDD coefficients.

Concerning sensitivity analysis, we illustrate how third and fourth-order moments, i.e. skewness and kurtosis, respectively, can be decomposed mimicking the ANOVA approach. It is also shown how this decomposition is correlated to a Polynomial Chaos (PC) expansion leading to a simple strategy to compute each term. New sensitivity indices, based on the contribution to the skewness and kurtosis, are proposed. The outcome of the proposed analysis is depicted by considering several test functions. Moreover, the ranking of the sensitivity indices is shown to vary according to their statistics order. Furthermore, the problem of formulating a truncated polynomial representation of the original function is treated. Both the reduction of the number of dimensions and the reduction of the order of interaction between parameters are considered. In both cases, the impact on the reduction is assessed in terms of statistics, namely the probability density function. Feasibility of the proposed analysis in a real-case is then demonstrated by presenting the sensitivity analysis of the performances of a turbine cascade in an Organic Rankine Cycles (ORCs), in the presence of complex thermodynamic models and multiple sources of uncertainty.

Moreover, we have developed a new framework for performing robust design optimization. a strategy is developed to deal with the error affecting the objective functions in uncertainty-based optimization. We refer to the problems where the objective functions are the statistics of a quantity of interest computed by an uncertainty quantification technique that propagates some uncertainties of the input variables through the system under consideration. In real problems, the statistics are computed by a numerical method and therefore they are affected by a certain level of error, depending on the chosen accuracy. The errors on the objective function can be interpreted with the abstraction of a bounding box around the nominal estimation in the objective functions space. In addition, in some cases the uncertainty quantification methods providing the objective functions also supply the possibility of adaptive refinement to reduce the error bounding box. The novel method relies on the exchange of information between the outer loop based on the optimization algorithm and the inner uncertainty quantification loop. In particular, in the inner uncertainty quantification loop, a control is performed to decide whether a refinement of the bounding box for the current design is appropriate or not. In single-objective problems, the current bounding box is compared to the current optimal design. In multi-objective problems, the decision is based on the comparison of the error bounding box of the current design and the current Pareto front. With this strategy, fewer computations are made for clearly dominated solutions and an accurate estimate of the objective function is provided for the interesting, non-dominated solutions. The results presented in this work prove that the proposed method improves the efficiency of the global loop, while preserving the accuracy of the final Pareto front.

Concerning semi-intrusive methods, a novel multiresolution framework, namely the Truncate and Encode (TE) approach is generalized and extended for taking into account uncertainty in partial differential equations (PDEs). Innovative ingredients are given by an algorithm permitting to recover the multiresolution representation without requir- ing the fully resolved solution, the possibility to treat a whatever form of pdf and the use of high-order (even non-linear, i.e. data-dependent) reconstruction in the stochastic space. Moreover, the spatial-TE method is introduced, which is a weakly intrusive scheme for uncertainty quantification (UQ), that couples the physical and stochastic spaces by minimizing the computational cost for PDEs. The proposed scheme is particularly attractive when treating moving discontinuities (such as shock waves in compressible flows), even if they appear during the simulations as it is common in unsteady aerodynamics applications. The proposed method is very flexible since it can easily coupled with different deterministic schemes, even with high-resolution features. Flexibility and performances of the present method are demonstrated on various numerical test cases (algebraic functions and ordinary differential equations), including partial differential equations, both linear and non-linear, in presence of randomness.

We applied a part of this method to a problem associated to the atmospheric reentry. In fact, an accurate determination of the catalytic property of thermal protection materials is crucial to design reusable atmospheric entry vehicles. This property is determined by combining experimental measurements and simulations of the reactive boundary layer near the material surface. The inductively-driven Plasmatron facility at the von Karman Institute for Fluid Dynamics provides a test environment to analyze gas-surface interactions under effective hypersonic conditions. In this study, we develop an uncertainty quantification methodology to rebuild values of the gas enthalpy and material catalytic property from Plasmatron experiments. A non-intrusive spectral projection method is coupled with an in-house boundary-layer solver, to propagate uncertainties and provide error bars on the rebuilt gas enthalpy and material catalytic property, as well as to determine which uncertainties have the largest contribution to the outputs of the experiments. We show that the uncertainties computed with the methodology developed are significantly reduced compared to those determined using a more conservative engineering approach adopted in the analysis of previous experimental campaigns.

6.4. Modelling of free surface flows

- Participants: Luca Arpaia , Stevan Bellec , Mathieu Collin , Sebastien De Brye , Andrea Filippini , Maria Kazolea, Luc Mieussens, and Mario Ricchiuto
- Corresponding member: Mario Ricchiuto

We have introduced a new systematic method to obtain discrete numerical models for incompressible free-surface flows. our approach allows to recover discrete asymptotic equations from a semi-discretized form

(keeping the vertical z variable and time continuous) of the incompressible Euler equations with free surface. In particular, starting from a (continuous) Galerkin finite element discretization in the horizontal direction, we perform an asymptotic analysis of the resulting semi-discrete system. This has allowed to obtain new discrete equivalents of the Peregrine equations [5], as well as enhanced variants in the spirit of [115]. This has been done in the PhD of S. Bellec. We have demonstrated that the resulting discrete equations present dispersion characteristics much improved w.r.t. those obtained by directly discretizing the asymptotic Boussinesq equations with continuous finite elements. This has been confirmed by numerical experiments on long wave propagation benchmarks. Concerning more classical continuous Boussinesq models, additional work has been done to characterize some of their exact solutions. This has provided some improved solutions to benchmark our codes, as well as some additional knownledge on these models [4].

This year we extended our work on fully non-linear weakly dispersive wave models in two dimensional horizontal coordinates. The proposed framework in [11], to approximate the so-called Green-Naghdi equations is followed. The method proposed, while remaining unsplit in time, is based on a separation of the elliptic and hyperbolic components of the equations. This separation is designed to recover the standard shallow water equations in the hyperbolic step, so that the method can be written as an *algebraic* correction to an existing shallow water code. More precisely, we re-write the standard form of the equations by splitting the original system in its elliptic and hyperbolic parts, through the definition of a new variable, accounting for the dispersive effects and having the role of a non-hydrostatic pressure gradient in the shallow water equations. We consider a two-step solution procedure. In the first step we compute a source term by inverting the elliptic coercive operator associated to the dispersive effects; then in a hyperbolic step we evolve the flow variables by using the non-linear shallow water equations, with all nonhydrostatic effects accounted by the source computed in the elliptic phase. The advantages of this procedure are firstly that the GN equations are used for propagation and shoaling, while locally reverting to the non-linear shallow water equations to model energy dissipation in breaking regions. Secondly and from the numerical point of view, this strategy allows each step to be solved with an appropriate numerical method on arbitrary unstructured meshes. We propose a hybrid finite element (FE) finite volume (FV) scheme, where the elliptic part of the system is discretized by means of the continuous Galerkin FE method and the hyperbolic part is discretized using a third-order node-centered finite volume (FV) technique. This work was a part of Andrea Filippini's PhD and a research paper is under preparation.

We also continue our study on wave breaking techniques on unstructured meshes [55]. In particular, we evaluate the coupling of both a weakly and a fully non-linear Boussinesq system with a turbulence model. We reformulate an evolution model for the turbulent kinetic energy, initially proposed by Nwogu [115], and evaluate its capabilities to provide sufficient dissipation in breaking regions. We also compare this dissipation to the one introduced by the numerical discretization. A research paper on the topic, is under preparation. Further more we studied and tested the application and validation of TUCWave code on the transformation breaking and run-up of irregular waves. Its is the first time that an unstructured high-resolution FV numerical solver for the 2D extended BT equations of Nwogu is tested on the generation and propagation of irregular waves. A research paper is under preparation.

The tools developed have been also used intensively in funded research programs. Within the TANDEM project, several benchmarks relevant to tsunami modelling have been performed and several common publications with the project partners are submitted and/or in preparation [54], [45]. We also our code SLOWS, to study the conditions for tidal bore formation in convergent alluvial estuaries [7]. A new set of dimensionless parameters has been introduced to describe the problem, and the code SLOWS has been used to explore the space of these parameters allowing to determine a critical curve allowing to characterize an estuary as "bore forming" or not. Surprising physical behaviours, in terms of dissipation and nonlinearity of the tides, have been highlighted.

Finally, in collaboration with F. Veron (University of Delaware at Newark, USA), L. Mieussens has developed a model to describe the effect of rain falling on water waves [20]. This model is based on a kinetic description of rain droplets that is used to compute the induced pression on a water wave. This allows to estimate the dissipation (or amplification) of the wave due to rainy conditions.

6.5. Wave energy conversion hydrodynamics

- Participants: Umberto Bosi, Mario Ricchiuto

- Corresponding member: Mario Ricchiuto

We have developed a prototype spectral element solver four a coupled set of differential equations modelling wave propagation (so-called outer domain), and the submerged flow under a floating body (inner domain). Both systems of equations are depth-averaged (Boussinesq type) systems involving some dispersive terms. They are further coupled to a force balance providing a (system of) ODE(s) for the floater. This model constitutes an intermediate fidelity approximation for the hydrodynamics of a wave energy converter. Differently from all industrial state of the art, it is a (fully) nonlinear model. However, its cost is extremely low when compared to full three-dimensional CFD analyses, due to the dimensional reduction brought from the depth averaged modelling. Last year we have shown the potential of this approach to predict the hydrodynamics of a floater in a simplified case [90], [91] (journal version to appear on *J. Ocean Eng. and Marine Energy*). This year we have further studied the issue of the coupling between domains with different PDE models (as in our case the inner and outer domains), and suggested an approach (based on a first order reformulation) allowing to coupled domains with different equations and with or without dispersive effects on either side. This work is done in the framework of the MIDWEST project funded by the EU OCEANEranet call.

CARMEN Project-Team

7. New Results

7.1. Convergence analysis of a bidomain-bath model

M. Bendahmane and N. Chamakuri performed a convergence analysis for optimal control of a bidomain-bath model by using a finite-element scheme. The bidomain-bath model represents a commonly used experimental setup where a small piece of cardiac tissue is kept alive and studied for some time in a nutrient bath. The bidomain-bath model equations describe the cardiac bioelectric activity in the tissue and bath volumes where the control acts at the boundary of the tissue. The existence of the finite element scheme and convergence to a unique weak solution of the direct problem were established. The convergence proof was based on deriving a series of a-priori estimates and using a general L2-compactness criterion. Moreover, the well-posedness of the adjoint problem and the first order necessary optimality conditions were shown. Comparing to the direct problem, the convergence proof of the adjoint problem is based on using a general L1-compactness criterion. The model was used for a simulation of low-energy defibrillation.

7.2. An exponential Adams–Bashforth ODE solver for stiff problems

C. Douanla Lontsi, together with Y. Coudière and C. Pierre, obtained an important result on time integration of stiff differential problems. They considered Adams exponential integrators with general varying stabilizers. General stabilization brings flexibility and facilitates the integration of ODE systems and semilinear evolution PDEs coupled with ODE systems. They were able to prove the stability and convergence of this type of integrator by introducing a new framework that extends multistep linear methods. Dahlquist stability was numerically investigated. A(α)-stability was observed under a condition on the stabilizer, which is a singular property for explicit schemes. The method was numerically studied for two stiff models in electrophysiology. Its performance was compared with several classical methods. The authors concluded that for stiff ODE systems, it provides a cheaper way to compute accurate solutions at large time steps than implicit solvers.

7.3. Homogeneous Neumann condition on the torso for solving inverse problems

The electrical activity of the heart creates an electrical field in the body. This phenomenon is classically modelled in a quasistatic manner by Laplace's equation. The non-invasive electrocardiographic imaging (ECGI) problem consists in retrieving the best electrical map on the heart from given torso measurements. Classically, the solution is found as the best fit between data generated by a forward problem and the actual torso measurements, and it needs a regularization. Hence the inverse solution depends on the matrix of the forward problem, called the transfer matrix, and the choice of the regularization procedure. In 2006, a meshless method based on the method of fundamental solutions (MFS) was adapted by Y. Wang and Y. Rudy [54] to directly solve the inverse problem, combined with a 0-th order Tikhonov regularization. The MFS method is notably more robust than previous methods (e.g. BEM) to the uncertainties introduced by the segmentation of the geometries. In the MFS, the potential is expressed as summation of the fundamental solution of the Laplace equation over a discrete set of virtual point sources placed outside of the domain of interest. The inverse solution is searched as the set of sources that best fit the boundary conditions on the torso, up to the regularization term. This formulation yields a linear system, which matrix depends on the torso and heart geometries, and the boundary conditions at the torso surface. The regularization parameter also heavily depends on the properties of the transfer matrix. The boundary conditions considered in [54] are: i) the Dirichlet conditions, meaning that the potentials at the torso surface are fitted to the recorded ones, ii) homogeneous Neumann conditions (HNC) meaning that the normal flux of current is minimized.

Numerically, the HNC requires to build accurate directions at each measurement location of the body surface, which is a first difficulty. In addition, the body is cut at the top and the bottom where no-flux conditions are probably not relevant. Lastly, the matrix coefficients related to the HNC appears to be much smaller than the ones from the Dirichlet condition, due to the distance between the torso and the actual electrical source (the heart).

J. Chamorro-Servent, Y. Coudière and R. Dubois studied the effect of the HNC on the matrix. They showed that enforcing the Neumann condition has a negligible effect on the solution of the inverse problem. Reconstructed potentials and activation time maps were built for in-silico data. No major differences were found between the standard MFS and the MFS removing the HNC in terms of potentials and activation times. In addition, removing the HNC reduces the ill-conditioning of the problem and the computational burden: the normal at the torso surface is not required anymore, and the problem size is divided by 2. The results of this work were presented as a poster in CinC 2016, and collected in a proceeding for the same conference by J. Chamorro-Servent et al. [18].

7.4. Adaptive placement of the pseudo-boundaries improves the conditioning of the inverse problem

In order to complete the investigaton concerning the MFS technique from [54], J. Chamorro-Servent, Y. Coudière and R. Dubois also studied the effect of the location of the virtual sources of the MFS method on the solutions of the inverse problem. Specifically, the regularization term spoils the biophysical content of the solution, and the regularization parameter must be chosen as small as possible. But the problem must be regularized enough to overcome its sensitivity to: i) noise on the measured potentials, ii) uncertainty in the location of measurement sites with respect to the surface on which the sources are distributed, iii) errors of segmentation of the geometries, iv) influence of cardiac motion, etc.

The regularization parameter can be studied in view of the singular values of the matrix, or for given measurments, the discrete Picart condition as defined by Hansen [47].

In the MFS problem, explained in section 7.3 , the virtual sources are placed by inflating and deflating the heart and torso surfaces with respect to the heart's geometric center. However, for some heart-torso geometries, this geometrical center is a poor reference. Furthermore, it has been proved in other fields that the placement of the virtual sources influences the ill-posedness of the MFS problem. However, this has not been tested for the ECGI problem.

J. Chamorro-Servent, R. Dubois and Y. Coudière proposed a new method of placement of these virtual sources based on the minimal distance of each point considered on the heart surface to the torso electrodes. The singular value analysis and the discrete Picard condition were used to optimize the location of these sources. The new distribution of sources was compared with the standard one for a set of experimental data. These data consist of simultaneous acquisition of the cardiac (on a Langendorff perfusion of the heart) and body surface potentials, in a controlled experimental environment.

The results presented by J. Chamorro-Servent et al. at CinC2016 [24] showed that the new distribution of sources made the inverse problem less ill-posed and therefore, less sensitive to the regularization parameter chosen. This improved the reconstructed potentials on the heart surface, especially when artefact (as for example the baseline) or noise were present.

Further results from the combination of the works described here and in section 7.3 were presented in a poster in the Liryc workshop of October 2016 [33] by J. Chamorro-Servent et al. A journal manuscript is currently under preparation (to submit in 2017).

7.5. Reduced sodium current in the lateral ventricular wall induces J waves in the ECG

"J waves," a particular abnormal waveform in electrocardiogram (ECG) leads, are associated with a higher risk for ventricular fibrillation. M. Potse has performed a series of simulations to investigate three possible

mechanisms that could explain such waves and the associated arrhythmia risk. Out of these, a reduced sodium current in the lateral area of the left ventricular wall turned out to be the most powerful to cause J waves. The lateral area is particular because it is normally late activated, and a further delay due to regionally reduced sodium current can lead to J waves in the ECG. If the same occurs elsewhere in the heart, the resulting J waves would be masked by other ECG peaks. The simulations were supported, as far as possible, by experiments performed at the University of Amsterdam. The results have been published in the journal Frontiers in Physiology, and further refinements have recently been shown in a poster at the Annual workshop of IHU Liryc [14], [43].

7.6. Atrial fibrillation due to complex geometry

Atrial fibrillation (AF), a situation in which the electrical activation of the atria proceeds chaotically, is believed to be due to abnormal tissue structure (for example fibrosis), which slows propagation, and abnormalities in ionic currents, which make the action potential shorter. In collaboration with the Center for Computational Medicine in Cardiology in Lugano, Switzerland, we performed series of simulations in which we tried to reproduce these effects [20]. Rapid stimulation of the atria caused AF in some of the simulations, with a likelihood related to the severity of fibrosis. However, we also observed a 30 % likelihood of AF initiation in a model with no fibrosis at all. In these cases, the complex structure of our highly realistic models alone in combination with the rapid-pacing protocol sufficed to create situations of conditional propagation block, which led to a reentrant arrhythmia. These results may shed a new light on the course of new-onset AF. A manuscript on this topic is under preparation.

<div style="text-align:center; color:red;">**CQFD Project-Team**</div>

7. New Results

7.1. Computable approximations for continuous-time Markov decision processes on Borel spaces based on empirical measures

The following result has been obtained by J. Anselmi (Inria CQFD), F. Dufour (Inria CQFD) and T. Prieto-Rumeau.

We propose an approach for approximating the value function and computing an ε-optimal policy of a continuous-time Markov decision processes with Borel state and action spaces, with possibly unbounded cost and transition rates, under the total expected discounted cost optimality criterion. Under the assumptions that the controlled process satisfies a Lyapunov type condition and the transition rate has a density function with respect to a reference measure, together with piecewise Lipschitz continuity of the elements of the control model, one can approximate the original controlled process by a sequence of models that are computationally solvable. Convergence of the approximations takes place at an exponential rate in probability.

7.2. Decentralized Proportional Load Balancing

The following result has been obtained by J. Anselmi (Inria CQFD) and N. Walton.

Load balancing is a powerful technique commonly used in communication and computer networks to improve system performance, robustness and fairness. In this paper, we consider a general model capturing the performance of communication and computer networks, and on top of it we propose a decentralized algorithm for balancing load among multiple network paths. The proposed algorithm is inspired by the modus operandi of the processor-sharing queue and on each network entry point operates as follows: every time a unit of load completes its service on a path, it increases by one unit the load of that path and decreases by one unit the load of a path selected at random with probability proportional to the amount of load on each of the available paths. We develop a dynamical system to argue that our load-balancer achieves a desirable network-wide utility optimization.

7.3. Constrained and Unconstrained Optimal Discounted Control of Piecewise Deterministic Markov Processes

The following result has been obtained by O. Costa, F. Dufour (Inria CQFD), and A. B. Piunovskiy.

The main goal of this paper is to study the infinite-horizon expected discounted continuous-time optimal control problem of piecewise deterministic Markov processes with the control acting continuously on the jump intensity λ and on the transition measure Q of the process but not on the deterministic flow ϕ. The contributions of the paper are for the unconstrained as well as the constrained cases. The set of admissible control strategies is assumed to be formed by policies, possibly randomized and depending on the history of the process, taking values in a set valued action space. For the unconstrained case we provide sufficient conditions based on the three local characteristics of the process ϕ, λ, Q and the semicontinuity properties of the set valued action space, to guarantee the existence and uniqueness of the integro-differential optimality equation (the so-called Bellman–Hamilton–Jacobi equation) as well as the existence of an optimal (and δ-optimal, as well) deterministic stationary control strategy for the problem. For the constrained case we show that the values of the constrained control problem and an associated infinite dimensional linear programming (LP) problem are the same, and moreover we provide sufficient conditions for the solvability of the LP problem as well as for the existence of an optimal feasible randomized stationary control strategy for the constrained problem.

7.4. Approximate Kalman-Bucy filter for continuous-time semi-Markov jump linear systems

The following result has been obtained by B. Saporta and E. F. Costa.

The aim of this paper is to propose a new numerical approximation of the Kalman-Bucy filter for semi-Markov jump linear systems. This approximation is based on the selection of typical trajectories of the driving semi-Markov chain of the process by using an optimal quantization technique. The main advantage of this approach is that it makes pre-computations possible. We derive a Lipschitz property for the solution of the Riccati equation and a general result on the convergence of perturbed solutions of semi-Markov switching Riccati equations when the perturbation comes from the driving semi-Markov chain. Based on these results, we prove the convergence of our approximation scheme in a general infinite countable state space framework and derive an error bound in terms of the quantization error and time discretization step. We employ the proposed filter in a magnetic levitation example with markovian failures and compare its performance with both the Kalman-Bucy filter and the Markovian linear minimum mean squares estimator.

7.5. Investigation of asymmetry in E. coli growth rate

The following result has been obtained by B. Saporta in collaboration with B. Delyon, N. Krell and Lydia Robert.

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.

7.6. Impulsive Control for Continuous-Time Markov Decision Processes: A Linear Programming Approach

The following result has been obtained by F. Dufour (Inria CQFD) and A. B. Piunovskiy.

In this paper, we investigate an optimization problem for continuous-time Markov decision processes with both impulsive and continuous controls. We consider the so-called constrained problem where the objective of the controller is to minimize a total expected discounted optimality criterion associated with a cost rate function while keeping other performance criteria of the same form, but associated with different cost rate functions, below some given bounds. Our model allows multiple impulses at the same time moment. The main objective of this work is to study the associated linear program defined on a space of measures including the occupation measures of the controlled process and to provide sufficient conditions to ensure the existence of an optimal control.

7.7. Conditions for the Solvability of the Linear Programming Formulation for Constrained Discounted Markov Decision Processes

The following result has been obtained by F. Dufour (Inria CQFD) and T. Prieto-Rumeau.

We consider a discrete-time constrained discounted Markov decision process (MDP) with Borel state and action spaces, compact action sets, and lower semi-continuous cost functions. We introduce a set of hypotheses related to a positive weight function which allow us to consider cost functions that might not be bounded below by a constant, and which imply the solvability of the linear programming formulation of the constrained MDP. In particular, we establish the existence of a constrained optimal stationary policy. Our results are illustrated with an application to a fishery management problem.

7.8. Spatio-temporal averaging for a class of hybrid systems and application to conductance-based neuron models

The following result has been obtained by A. Genadot (Inria CQFD).

We obtain a limit theorem endowed with quantitative estimates for a general class of infinite dimensional hybrid processes with intrinsically two different time scales and including a population. As an application, we consider a large class of conductance-based neuron models describing the nerve impulse propagation along a neural cell at the scales of ion channels.

7.9. A comparison of fitness-case sampling methods for genetic programming

The following result has been obtained by Pierrick Legrand (Inria CQFD) in collaboration with Y. Martinez, E. Naredo, L. Trujillo, U. Lopez.

The canonical approach towards fitness evaluation in Genetic Programming (GP) is to use a static training set to determine fitness, based on a cost function averaged over all fitness-cases. However, motivated by different goals, researchers have recently proposed several techniques that focus selective pressure on a subset of fitness-cases at each generation. These approaches can be described as fitness-case sampling techniques, where the training set is sampled, in some way, to determine fitness. This paper shows a comprehensive evaluation of some of the most recent sampling methods, using benchmark and real-world problems for symbolic regression. The algorithms considered here are Interleaved Sampling, Random Interleaved Sampling, Lexicase Selection and a new sampling technique is proposed called Keep-Worst Interleaved Sampling (KW-IS). The algorithms are extensively evaluated based on test performance, overfitting and bloat. Results suggest that sampling techniques can improve performance compared with standard GP. While on synthetic benchmarks the difference is slight or none at all, on real-world problems the differences are substantial. Some of the best results were achieved by Lexicase Selection and KeepWorse-Interleaved Sampling. Results also show that on real-world problems overfitting correlates strongly with bloating. Furthermore, the sampling techniques provide efficiency, since they reduce the number of fitness-case evaluations required over an entire run.

7.10. Prediction of Expected Performance for a Genetic Programming Classifier

The following result has been obtained by Pierrick Legrand (Inria CQFD) in collaboration with Y. Martínez, L. Trujillo and E. Galván-López.

The estimation of problem difficulty is an open issue in genetic programming (GP). The goal of this work is to generate models that predict the expected performance of a GP-based classifier when it is applied to an unseen task. Classification problems are described using domain-specific features, some of which are proposed in this work, and these features are given as input to the predictive models. These models are referred to as predictors of expected performance. We extend this approach by using an ensemble of specialized predictors (SPEP), dividing classification problems into groups and choosing the corresponding SPEP. The proposed predictors are trained using 2D synthetic classification problems with balanced datasets. The models are then used to predict the performance of the GP classifier on unseen real-world datasets that are multidimensional and imbalanced. This work is the first to provide a performance prediction of a GP system on test data, while previous works focused on predicting training performance. Accurate predictive models are generated by posing a symbolic regression task and solving it with GP. These results are achieved by using highly descriptive features and including a dimensionality reduction stage that simplifies the learning and testing process. The proposed approach could be extended to other classification algorithms and used as the basis of an expert system for algorithm selection.

7.11. Evolving Genetic Programming Classifiers with Novelty Search

The following result has been obtained by Pierrick Legrand (Inria CQFD) in collaboration with E. Naredo, L. Trujillo, S. Silvac andLuis Muñoza.

Novelty Search (NS) is a unique approach towards search and optimization, where an explicit objective function is replaced by a measure of solution novelty. However, NS has been mostly used in evolutionary robotics while its usefulness in classic machine learning problems has not been explored. This work presents a NS-based genetic programming (GP) algorithm for supervised classification. Results show that NS can solve real-world classification tasks, the algorithm is validated on real-world benchmarks for binary and multiclass problems. These results are made possible by using a domain-specific behavior descriptor. Moreover, two new versions of the NS algorithm are proposed, Probabilistic NS (PNS) and a variant of Minimal Criteria NS (MCNS). The former models the behavior of each solution as a random vector and eliminates all of the original NS parameters while reducing the computational overhead of the NS algorithm. The latter uses a standard objective function to constrain and bias the search towards high performance solutions. The paper also discusses the effects of NS on GP search dynamics and code growth. Results show that NS can be used as a realistic alternative for supervised classification, and specifically for binary problems the NS algorithm exhibits an implicit bloat control ability.

7.12. Regularity and Matching Pursuit Feature Extraction for the Detection of Epileptic Seizures

The following result has been obtained by Pierrick Legrand (Inria CQFD) in collaboration with E. Z-Floresa, L. Trujillo, A. Soteloa and L. N. Coriaa.

The neurological disorder known as epilepsy is characterized by involuntary recurrent seizures that diminish a patient's quality of life. Automatic seizure detection can help improve a patient's interaction with her/his environment, and while many approaches have been proposed the problem is still not trivially solved.

In this work, we present a novel methodology for feature extraction on EEG signals that allows us to perform a highly accurate classification of epileptic states. Specifically, Hölderian regularity and the Matching Pursuit algorithm are used as the main feature extraction techniques, and are combined with basic statistical features to construct the final feature sets. These sets are then delivered to a Random Forests classification algorithm to differentiate between epileptic and non-epileptic readings.

Several versions of the basic problem are tested and statistically validated producing perfect accuracy in most problems and 97.6% accuracy on the most difficult case. Comparison with existing methods: A comparison with recent literature, using a well known database, reveals that our proposal achieves state-of-the-art performance. The experimental results show that epileptic states can be accurately detected by combining features extracted through regularity analysis, the Matching Pursuit algorithm and simple time-domain statistical analysis. Therefore, the proposed method should be considered as a promising approach for automatic EEG analysis.

7.13. Probabilistic safety analysis of the collision between a space debris and a satellite with an island particle algorithm

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with C. Verge, J. Morio and J.C Dolado Perez.

Collision between satellites and space debris seldom happens, but the loss of a satellite by collision may have catastrophic consequences both for the satellite mission and for the space environment. To support the decision to trigger off a collision avoidance manoeuver, an adapted tool is the determination of the collision probability between debris and satellite. This probability estimation can be performed with rare event simulation techniques when Monte Carlo techniques are not enough accurate. In this chapter, we focus on analyzing the influence of different simulation parameters (such as the drag coefficient) that are set for to simplify the simulation, on the collision probability estimation. A bad estimation of these simulation parameters can strongly modify rare event probability estimations. We design here a new island particle Markov chain Monte Carlo algorithm to determine the parameters that, in case of bad estimation, tend to increase the collision probability value. This algorithm also gives an estimate of the collision probability

maximum taking into account the likelihood of the parameters. The principles of this statistical technique are described throughout this chapter.

7.14. Particle association measures and multiple target tracking

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with J. Houssineau.

In the last decade, the area of multiple target tracking has witnessed the introduction of important concepts and methods, aiming at establishing principled approaches for dealing with the estimation of multiple objects in an efficient way. One of the most successful classes of multi-object filters that have been derived out of these new grounds includes all the variants of the Probability Hypothesis Density (phd) filter. In spite of the attention that these methods have attracted, their theoretical performances are still not fully understood. In this chapter, we first focus on the different ways of establishing the equations of the phd filter, using a consistent set of notations. The objective is then to introduce the idea of observation path, upon which association measures are defined. We will see how these concepts highlight the structure of the first moment of the multi-object distributions in time, and how they allow for devising solutions to practical estimation problems.

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

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with D. Villemonais.

We consider an elliptic and time-inhomogeneous diffusion process with time-periodic coefficients evolving in a bounded domain of Rd 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.

7.16. On particle Gibbs Markov chain Monte Carlo models

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with R. Kohn and F. Patras.

This result analyses a new class of advanced particle Markov chain Monte Carlo algorithms recently introduced by Andrieu, Doucet, and Holenstein (2010). We present a natural interpretation of these methods in terms of well known unbiasedness properties of Feynman-Kac particle measures, and a new duality with Feynman-Kac models. This perspective sheds new light on the foundations and the mathematical analysis of this class of methods. A key consequence is their equivalence with the Gibbs sampling of a (many-body) Feynman-Kac target distribution. Our approach also presents a new stochastic differential calculus based on geometric combinatorial techniques to derive non-asymptotic Taylor type series for the semigroup of a class of particle Markov chain Monte Carlo models around their invariant measures with respect to the population size of the auxiliary particle sampler. These results provide sharp quantitative estimates of the convergence rate of the models with respect to the time horizon and the size of the systems. We illustrate the direct implication of these results with sharp estimates of the contraction coefficient and the Lyapunov exponent of the corresponding samplers, and explicit and non-asymptotic L p -mean error decompositions of the law of the random states around the limiting invariant measure. The abstract framework developed in the article also allows the design of natural extensions to island (also called SMC^2) type particle methodologies. We illustrate

this general framework and results in the context of nonlinear filtering, hidden Markov chain problems with fixed unknown parameters, and Feynman-Kac path- integration models arising in computational physics and chemistry.

7.17. Sequential Monte Carlo with Highly Informative Observations

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with L. Murray.

We propose sequential Monte Carlo (SMC) methods for sampling the posterior distribution of state-space models under highly informative observation regimes, a situation in which standard SMC methods can perform poorly. A special case is simulating bridges between given initial and final values. The basic idea is to introduce a schedule of intermediate weighting and resampling times between observation times, which guide particles towards the final state. This can always be done for continuous-time models, and may be done for discrete-time models under sparse observation regimes; our main focus is on continuous-time diffusion processes. The methods are broadly applicable in that they support multivariate models with partial observation, do not require simulation of the backward transition (which is often unavailable), and, where possible, avoid pointwise evaluation of the forward transition. When simulating bridges, the last cannot be avoided entirely without concessions, and we suggest an epsilon-ball approach (reminiscent of Approximate Bayesian Computation) as a workaround. Compared to the bootstrap particle filter, the new methods deliver substantially reduced mean squared error in normalising constant estimates, even after accounting for execution time. The methods are demonstrated for state estimation with two toy examples, and for parameter estimation (within a particle marginal Metropolis–Hastings sampler) with three applied examples in econometrics, epidemiology and marine biogeochemistry.

7.18. A duality formula for Feynman-Kac path particle models

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with R. Kohn and F. Patras.

This result presents a new duality formula between genetic type genealogical tree based particle models and Feynman–Kac measures on path spaces. Among others, this formula allows us to design reversible Gibbs–Glauber Markov chains for Feynman–Kac integration on path spaces. Our approach yields new Taylor series expansions of the particle Gibbs–Glauber semigroup around its equilibrium measure w.r.t. the size of the particle system, generalizing the recent work of Andrieu, Doucet, and Holenstein [1]. We analyze the rate of convergence to equilibrium in terms of the ratio of the length of the trajectories to the number of particles. The analysis relies on a tree-based functional and combinatorial representation of a class of Feynman–Kac particle models with a frozen ancestral line. We illustrate the impact of these results in the context of Quantum and Diffusion Monte Carlo methods.

7.19. Non-Asymptotic Analysis of Adaptive and Annealed Feynman-Kac Particle Models

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with F. Giraud.

Sequential and quantum Monte Carlo methods, as well as genetic type search algorithms can be interpreted as a mean field and interacting particle approximations of Feynman-Kac models in distribution spaces. The performance of these population Monte Carlo algorithms is strongly related to the stability properties of nonlinear Feynman-Kac semigroups. In this paper, we analyze these models in terms of Dobrushin ergodic coefficients of the reference Markov transitions and the oscillations of the potential functions. Sufficient conditions for uniform concentration inequalities w.r.t. time are expressed explicitly in terms of these two quantities. We provide an original perturbation analysis that applies to annealed and adaptive Feynman-Kac models, yielding what seems to be the first results of this kind for these types of models. Special attention is devoted to the particular case of Boltzmann-Gibbs measures' sampling. In this context, we design an explicit way of tuning the number of Markov chain Monte Carlo iterations with temperature schedule. We also design

an alternative interacting particle method based on an adaptive strategy to define the temperature increments. The theoretical analysis of the performance of this adaptive model is much more involved as both the potential functions and the reference Markov transitions now depend on the random evolution on the particle model. The nonasymptotic analysis of these complex adaptive models is an open research problem. We initiate this study with the concentration analysis of a simplified adaptive models based on reference Markov transitions that coincide with the limiting quantities, as the number of particles tends to infinity.

7.20. Uniform stability of a particle approximation of the optimal filter derivative

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with A. Doucet and S.S. Singh.

Particle methods, also known as Sequential Monte Carlo methods, are a principled set of algorithms used to approximate numerically the optimal filter in nonlinear non-Gaussian state-space models. However, when performing maximum likelihood parameter inference in state-space models, it is also necessary to approximate the derivative of the optimal filter with respect to the parameter of the model. References [G. Poyiadjis, A. Doucet, and S. S. Singh, Particle methods for optimal filter derivative: Application to parameter estimation, in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 5, Philadelphia, 2005, pp. 925–928 and G. Poyiadjis, A. Doucet, and S. S. Singh, Biometrika, 98 (2011), pp. 65–80] present an original particle method to approximate this derivative, and it was shown in numerical examples to be numerically stable in the sense that it did not deteriorate over time. In this paper we theoretically substantiate this claim. $\mathbb{L}_p$ bounds and a central limit theorem for this particle approximation are presented. Under mixing conditions these $\mathbb{L}_p$ bounds and the asymptotic variance are uniformly bounded with respect to the time index.

7.21. Combining clustering of variables and feature selection using random forests: the CoV/VSURF procedure

The following result has been obtained by M. Chavent, and J. Saracco in collaboration with R. Genuer.

High-dimensional data classification is a challenging problem. A standard approach to tackle this problem is to perform variables selection, e.g. using stepwise procedures or LASSO approches. Another standard way is to perform dimension reduction, e.g. by Principal Component Analysis (PCA) or Partial Least Square (PLS) procedures. The approach proposed in this paper combines both dimension reduction and variables selection. First, a procedure of clustering of variables (CoV) is used to built groups of correlated variables in order to reduce the redundancy of information. This dimension reduction step relies on the R package ClustOfVar which can deal with both numerical and categorical variables. Secondly, the most relevant synthetic variables (which are numerical variables summarizing the groups obtained in the first step) are selected with a procedure of variable selection using random forests (VSURF), implemented in the R package VSURF. Numerical performances of the proposed methodology called CoV/VSURF are compared with direct applications of VSURF or random forests (RF) on the original p variables. Improvements obtained with the CoV/VSURF procedure are illustrated on two simulated mixed datasets (cases $n > p$ and $n << p$) and on a real proteomic dataset.

7.22. An innovative approach combining animal performances, nutritional value and sensory quality of meat

The following result has been obtained by M. Chavent and J. Saracco in collaboration with M.P. Ellies.

This work sets out a methodological approach to assess how to simultaneously control together animal performances, nutritional value, sensory quality of meat. Seventy-one young bulls were characterized by 97 variables. Variables of each element were arranged into either 5 homogeneous Intermediate Scores (IS) or 2 Global Indices (GI) via a clustering of variables and analysed together by Principal Component Analysis (PCA). These 3 pools of 5 IS (or 2 GI) were analysed together by PCA to established the links existing among the triptych. Classification on IS showed no opposition between animal performances and nutritional value of meat, as it seemed possible to identify animals with a high butcher value and intramuscular fat relatively rich in polyunsaturated fatty acids. Concerning GI, the classification indicated that animal performances were negatively correlated with sensory quality. This method appeared to be a useful contribution to the management of animal breeding for an optimal trade-off between the three elements of the triptych.

7.23. BIG-SIR: a Sliced Inverse Regression approach for massive data

The following result has been obtained by J. Saracco (Inria CQFD) in collaboration with B. Liquet.

In a massive data setting, we focus on a semiparametric regression model involving a real dependent variable Y and a p-dimensional covariable X. This model includes a dimension reduction of X via an index $X'\beta$. The Effective Dimension Reduction (EDR) direction β cannot be directly estimated by the Sliced Inverse Regression (SIR) method due to the large volume of the data. To deal with the main challenges of analysing massive datasets which are the storage and computational efficiency, we propose a new SIR estimator of the EDR direction by following the "divide and conquer" strategy. The data is divided into subsets. EDR directions are estimated in each subset which is a small dataset. The recombination step is based on the optimisation of a criterion which assesses the proximity between the EDR directions of each subset. Computations are run in parallel with no communication among them. The consistency of our estimator is established and its asymptotic distribution is given. Exten- sions to multiple indices models, q-dimensional response variable and/or SIR α-based methods are also discussed. A simulation study using our edrGraphicalTools R package shows that our approach enables us to reduce the computation time and conquer the memory constraint problem posed by massive datasets. A combination of foreach and bigmemory R packages are exploited to offer efficiency of execution in both speed and memory. Finally, results are visualised using the bin-summarise-smooth approach through the bigvis R package

<p style="text-align:center; color:red; font-weight:bold">FLOWERS Project-Team</p>

7. New Results

7.1. Robotic And Computational Models Of Human Development and Cognition

7.1.1. *Computational Models Of Information-Seeking, Curiosity And Attention in Humans and Animals*

Participants: Manuel Lopes, Pierre-Yves Oudeyer [correspondant], Jacqueline Gottlieb, Celeste Kidd, Alvaro Ovalle, William Schueller, Sebastien Forestier, Nabil Daddaouda, Nicholas Foley.

This project involves a collaboration between the Flowers team, the Cognitive Neuroscience Lab of J. Gottlieb at Columbia Univ. (NY, US), and the developmental psychology lab of Celeste Kidd at Univ. Rochester, US, on the understanding and modeling of mechanisms of curiosity, attention and active intrinsically motivated exploration that until now have been little explored in neuroscience, machine learning and cognitive robotics.

It is organized around the study of the hypothesis that information gain (or control gain) 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 (Gottlieb's lab) and children (Kidd's lab) together with computational modeling of curiosity/intrinsic motivation and learning in the Flowers team.

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 [29], [155], [162]. Therefore we hypothesize that, rather than using a single optimization process as it has been the case in most previous theoretical work [131], 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 [153] [110], [149]. 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. We will quantitatively characterize this dynamic, multi-dimensional system in the framework of Bayesian Reinforcement Learning, as described below.

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 [143], [161], and rats show individual variation in learning styles and novelty seeking behaviors [128], 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 [159], [171], [106], [175]. 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 model based on Bayesian reinforcement learning.

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 [143], [141] 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 Bayesian Reinforcement 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

In particular, new results in 2015 include:

7.1.1.4. Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates

Using a novel oculomotor paradigm, combined with reinforcement learning (RL) simulations, we show that monkeys are intrinsically motivated to search for and look at reward-predictive cues, and that their intrinsic motivation is shaped by a desire to reduce uncertainty, a desire to obtain conditioned reinforcement from

positive cues, and individual variations in decision strategy and the cognitive costs of acquiring information. The results suggest that free-viewing oculomotor behavior reveals cognitive and emotional factors underlying the curiosity driven sampling of information. These results were published in [66].

7.1.1.5. Experiments in Active Categorization

An ongoing effort to characterize curiosity and exploration in an experimental setting consists in evaluating the manner in which diverse tasks or goals are selected. This would include monitoring what does a test subject decide to learn, in what order and how is it done. This has been referred to as strategic learning [31]. Accordingly, it is of particular interest for the project to observe the type of learning dynamics in relation to their learning progress [153]. This principle tries to establish links between the selection and ordering of tasks and the speed or the rate of improvement a subject may achieve. This implies that during free exploration the subject would focus on tasks that are considered of certain complexity and where it makes consistent progress. At the same time the subject would avoid: (1) trivial tasks that do not offer much learning due to their simplicity or (2) very complicated tasks where little or no progress is achieved.

We have been working on prototyping an experiment where the subject is presented with different stimuli classification tasks of varying difficulty. The goal for each of the tasks is to correctly predict and differentiate between different classes of stimuli. Two main aspects of the task are under the control of the subject: (1) the task that he/she wants to learn and (2) once selected a task, what elements to explore in order to subsequently being able to predict future stimuli. Essentially the subject autonomously organizes which tasks to focus on and in what order. Therefore one of the objectives of this investigation is to analyze if the learning dynamics are guided by the amount of progress the subject achieves in the tasks.

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], Clement Moulin-Frier, Sébastien Forestier, Linda Smith.

7.1.2.1. Modeling Cognitive Development and Tool Use 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 [131]. 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 [157].

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 [3]). On the other hand, other algorithmic models have shown how several social learning mechanisms could allow robots to acquire elements of speech and language [118], 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 [49].

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 [37]. 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 [152].

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 [118]. 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 [152],[3], [43], [37]. 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.

A first question that we study in this work is the type of mechanisms that could be used for hierarchical skill learning allowing to manage new task spaces and new action spaces, where the action and task spaces initially given to the robot are continuous and high-dimensional and can be encapsulated as primitive actions to affect newly learnt task spaces.

We presented firsts results on that question at the 38th Annual Meeting of the Cognitive Science Society, Philadelphia, Pennsylvania, USA, August 10-13th [80]. In this work, we presented the HACOB (Hierarchical Active Curiosity-driven mOdel Babbling) architecture of algorithms that actively chooses which sensorimotor model to train in a hierarchy of models representing the environmental structure. We studied this architecture using a simulated robotic arm interacting with objects in a 2D environment (See Fig. 8). Studies of child development of tool use precursors showed successive but overlapping phases of qualitatively different types of behaviours [167]. We hypothesized that two mechanisms in particular play a role in the structuring of these phases: the intrinsic motivation to explore and the representation used to encode sensorimotor experience.
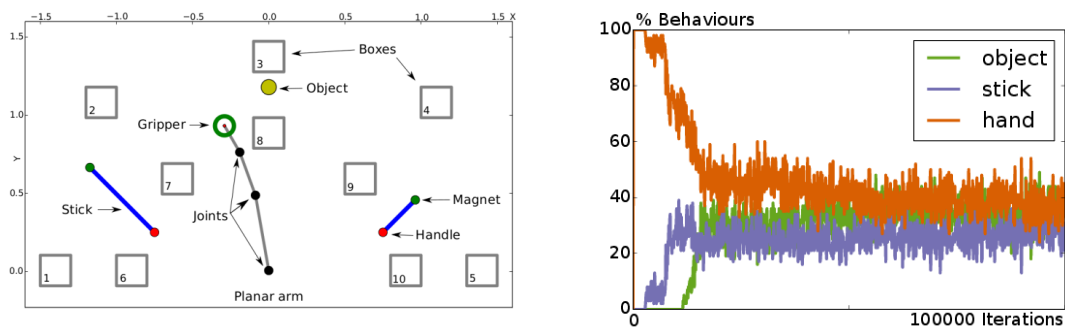

Figure 8. Left: simulated robotic environment with a 4 DOF robotic arm, 2 tools and a toy. Right: Observed behaviours of an agent: it first explores its arm to move its hand, then also explore to move the stick and the toy.

We showed that using a hierarchical structure of sensorimotor models and active model babbling as an intrinsic motivation to explore sensorimotor models that have a high learning progress, then overlapping phases of behaviours are autonomously emerging in the developmental trajectories of agents. To our knowledge, this is the first model of curiosity-driven development of simple tool use and of the self-organization of overlapping phases of behaviours. In particular, our model explains why and how intrinsically motivated exploration of non-optimal methods to solve certain sensorimotor problems can be useful to discover how to solve other sensorimotor problems, in accordance with Siegler's overlapping waves theory, by scaffolding the learning of increasingly complex affordances in the environment.

In computational models of strategy selection for the problem of integer addition, Shrager and Siegler proposed a mechanism that maintains the concurrent exploration of alternative strategies with use frequencies that are

proportional to their performance for solving a particular problem. This mechanism was also used by Chen and Siegler to interpret an experiment with 1.5- and 2.5-year-olds that had to retrieve an out-of-reach toy, and where they could use one of several available strategies that included leaning forward to grasp a toy with the hand or using a tool to retrieve the toy.

In a paper that we presented at the The Sixth Joint IEEE International Conference on Developmental Learning and Epigenetic Robotics, Cergy-Pontoise, France, September 19-22nd [82], we studied tool use discovery and considered other mechanisms of strategy selection and evaluation. In particular, we presented models of curiosity-driven exploration where strategies are explored according to the learning progress/information gain they provide (as opposed to their current efficiency to actually solve the problem). In these models, we defined a curiosity-driven agent learning a hierarchy of different sensorimotor models in a simple 2D setup with a robotic arm, a stick and a toy. In a first phase, the agent learns from scratch how to use its robotic arm to control the tool and to catch the toy, and in a second phase with the same learning mechanisms, the agent has to solve three problems where the toy can only be reached with the tool (See Fig. 9). We showed that agents choosing strategies based on learning progress also display overlapping waves of behavior compatible with the one observed in infants, and we suggested that curiosity-driven exploration could be at play in Chen and Siegler's experiment, and more generally in tool use discovery.
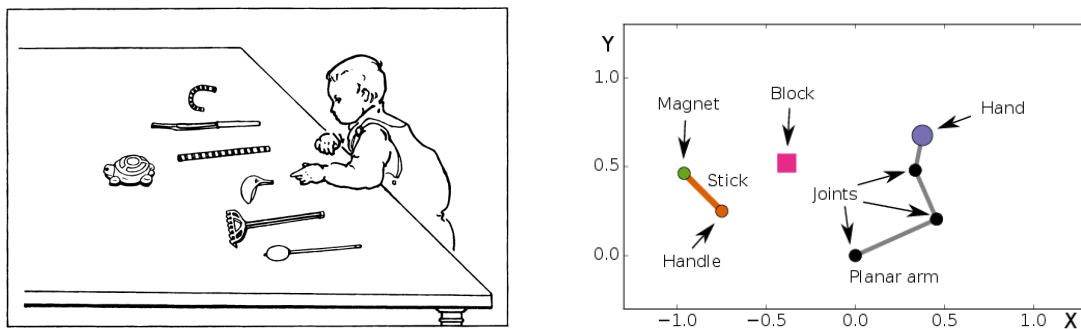

Figure 9. Left: Chen and Siegler's experimental setup with 1.5 and 2.5 years old babies who have to pick the good toy to retrieve an interesting toy. Right: Simulated robotic setup with a 3 DOF robotic arm that has 2 strategies to retrieve a toy: either grasp it with the hand, or use the stick to pull the toy.

7.1.2.2. Curiosity-driven developmental processes and their role in development and evolution of language

Infants' own activities create and actively select their learning experiences. In a collaboration with Linda Smith, we have analyzed recent models of embodied information seeking and curiosity-driven learning and have showed that these mechanisms have deep implications for development and evolution. In [69], we have discussed how these mechanisms yield self-organized epigenesis with emergent ordered behavioral and cognitive developmental stages. We described a robotic experiment that explored the hypothesis that progress in learning, in and for itself, generates intrinsic rewards: the robot learners probabilistically selected experiences according to their potential for reducing uncertainty. In these experiments, curiosity-driven learning led the robot learner to successively discover object affordances and vocal interaction with its peers. We explain how a learning curriculum adapted to the current constraints of the learning system automatically formed, constraining learning and shaping the developmental trajectory. The observed trajectories in the robot experiment share many properties with those in infant development, including a mixture of regularities and diversities in the developmental patterns. Finally, we argued that such emergent developmental structures can guide and constrain evolution, in particular with regards to the origins of language.

7.1.3. Computational Models Of Developmental Exploration Mechanisms in Vocal Babbling and Arm Reaching in Infants

Participants: Pierre-Yves Oudeyer [correspondant], Clement Moulin-Frier, Freek Stulp, Jules Borchard.

7.1.3.1. Proximodistal Freeing of DOFs in Motor Learning as an Emergent Property of Stochastic Optimization Principles

To harness the complexity of their high-dimensional bodies during sensorimotor development, infants are guided by patterns of freezing and freeing of degrees of freedom. We have formulated and studied computationally the hypothesis that such patterns, such as the proximodistal freeing of degrees of freedom when learning to reach, can emerge spontaneously as the result of a family of stochastic optimization processes, without an innate encoding of a mat- urational schedule. In particular, we present simulated experiments with a 6-DOF arm where a computational learner progressively acquires reaching skills through adaptive exploration, and we show that a proximodistal organization appears spontaneously, which we denote PDFF (ProximoDistal Freezing and Freeing of degrees of freedom). We also compare the emergent structuration as different arm structures are used – from human-like to quite unnatural ones – to study the effect of different kinematic structures on the emergence of PDFF.

7.1.3.2. Emergent Jaw Predominance in Vocal Development through Stochastic Optimization

Infant vocal babbling is strongly relying on jaw oscillations, especially at the stage of canonical babbling, which underlies the syllabic structure of world languages. We have proposed, modelled and analyzed an hypothesis to explain this predominance of the jaw in early babbling. This hypothesis states that general stochastic optimization principles, when applied to learning sensorimotor control, automatically generate ordered babbling stages with a predominant exploration of jaw movements in early stages, just like they generate proximo-distal organization of exploration in arm reaching as described in the paragraph above. In particular, such stochastic optimization principles predominantly explore jaw movement at the beginning of vocal learning, and when close to the rest position of the vocal tract, as it impacts the auditory effects more than other articulators.

7.1.4. Learning and Teaching in Adult-Child and Human-Robot Interaction

Participants: Anna-Lisa Vollmer [correspondant], Pierre-Yves Oudeyer.

7.1.4.1. Pragmatic Frames

One of the big challenges in robotics today is to learn from human users that are inexperienced in interacting with robots but yet are often used to teach skills flexibly to other humans and to children in particular. A potential route toward natural and efficient learning and teaching in Human-Robot Interaction (HRI) is to leverage the social competences of humans and the underlying interactional mechanisms. In this perspective, we propose 'pragmatic frames' as flexible interaction protocols that provide important contextual cues to enable learners to infer new action or language skills and teachers to convey these cues. Following the concept developed in the field of developmental linguistics [117], we define a pragmatic frame to be an interaction protocol negotiated over time between interaction partners. We further specify a Pragmatic Frame to especially involve an observable **coordinated sequence of behaviors** and also relevant **cognitive operations**.
Figure 10 depicts the book reading frame Bruner observed in his studies on word learning.

At home, a mother is sitting on the sofa with her child on her lap and she is holding a picture book in front of them. The mother points to the book and says "look!" to direct the child's attention. The child then gazes to the image. And the mother asks "What's that?", prompting the child's performance. The child answers with babble strings and smiles, maybe "auo". "Yes, a pineapple!" The mother gives positive feedback and the correct label. "Anappo", again babble strings and smiles. And the mother gives positive feedback. This stable sequence that the child is familiar with helps the child to participate and to pick up the only variable information he or she is supposed to learn. We argue that this frame also triggers the relevant cognitive functions to process the information.
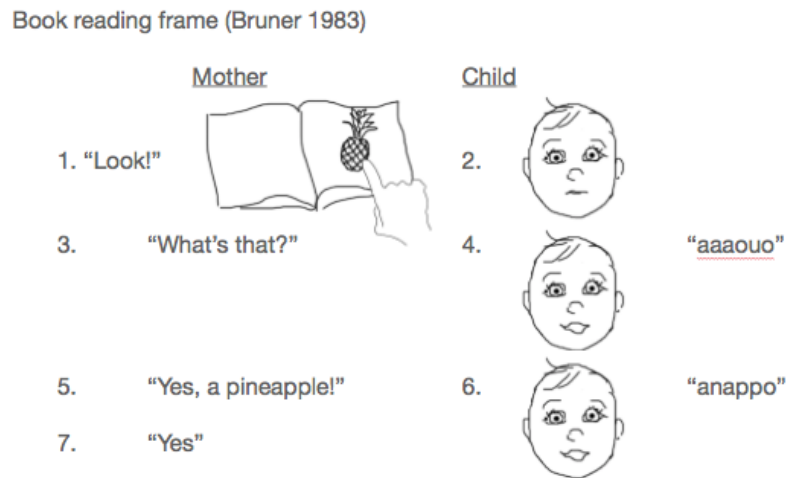
Book reading frame (Bruner 1983)

Mother   Child

1. "Look!"   2.

3.   "What's that?"   4.   "aaaouo"

5.   "Yes, a pineapple!" 6.   "anappo"

7.   "Yes"

Figure 10. Example of a learning/teaching pragmatic frame.

Our results in 2016 have been twofold. First, in a paper published in Frontiers in Psychology [70], we have given a theoretical account of pragmatic frames as an alternative to the mapping metaphor which posits that children learn a word by mapping it onto a concept of an object or event. However, we believe that a mapping metaphor cannot account for word learning, because even though children focus attention on objects, they do not necessarily remember the connection between the word and the referent unless it is framed pragmatically, that is, within a task. Word learning with pragmatic frames occurs as children accomplish a goal in cooperation with a partner. We elaborate on pragmatic frames, offer some initial parametrizations of the concept, and embed it in current language learning approaches.

Second, aiming at leveraging the concept of pragmatic frames for Human-Robot Interaction, we published an article in Frontiers in Neurorobotics [71] in which we study a selection of HRI work in the literature which has focused on learning–teaching interaction and analyze the interactional and learning mechanisms that were used in the light of pragmatic frames. This allows us to show that many of the works have already used in practice, but not always explicitly, basic elements of the pragmatic frames machinery. However, we also show that pragmatic frames have so far been used in a very restricted way as compared to how they are used in human–human interaction and argue that this has been an obstacle preventing robust natural multi-task learning and teaching in HRI. In particular, we explain that two central features of human pragmatic frames, mostly absent of existing HRI studies, are that (1) social peers use rich repertoires of frames, potentially combined together, to convey and infer multiple kinds of cues; (2) new frames can be learnt continually, building on existing ones, and guiding the interaction toward higher levels of complexity and expressivity. To conclude, we give an outlook on the future research direction describing the relevant key challenges that need to be solved for leveraging pragmatic frames for robot learning and teaching.

7.1.5. *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 underlie 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 [168]. 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 [30], [131], [140]. This approach has been mostly studied for single robotic agents learning sensorimotor affordances [153], [38]. However active learning might represent an evolutionary advantage for language formation at the population level as well [49], [170].

Naming Games are a computational framework, elaborated to simulate the self-organization of lexical conventions in the form of a multi-agent model [169]. 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 [53], [46], [122] or the hearer [90] actively controlling vocabulary growth.

We study here how the convergence time and the maximum level of complexity scale with population size, for three different strategies (one with random topic choice and two with active topic choice) detailed in table 11 .

Naive (random)	Success Threshold	Minimal counts
$m \leftarrow \mathrm{random}(\mathcal{M})$	**if** $\mathrm{mean}\left(\frac{succ(i)}{succ(i)+fail(i)}\right)_{i\in\mathcal{LM}} \geq \boldsymbol{\alpha}$: $\quad m \leftarrow \mathrm{random}(\mathcal{UM})$ **else:** $\quad m \leftarrow \mathrm{argmin}_{i\in\mathcal{LM}}\left(\frac{succ(i)}{succ(i)+fail(i)}\right)$	**if** $\forall i \in \mathcal{LM}\ succ(i) > \boldsymbol{n}$: $\quad m \leftarrow \mathrm{random}(\mathcal{UM})$ **else:** $\quad m \leftarrow \mathrm{argmin}_{i\in\mathcal{LM}}(succ(i))$
$\mathcal{M}$: all meanings, $\mathcal{LM}$: labeled meanings, $\mathcal{UM}$: unlabeled meanings, μ: vocabulary size (# word-meaning associations) $succ$: # successful interactions per meaning, $fail$: # failed interactions per meaning		

Figure 11. Strategies: Choice of meaning m. Both active strategies use a parameter (α and n), which is each time chosen optimal in our simulations.

As for the version of the Naming Game used in our work, the scenario of the interaction is described in [90]. Vocabulary is updated as described in the Minimal Naming Game, detailed in [177]. In our simulations, we choose to set $N = M = W$, where N is the population size, M the number of meanings, and W the number of possible words. The computed theoretical success ratio of communication is used to represent the degree of convergence toward a shared lexicon for the whole population. A value of 1 means that the population reached full convergence. Complexity level of an individual lexicon is measured as the total number of distinct associations between meanings and words in the lexicon, or in other words: memory usage.

We show here (see figures 12 ,13) that convergence time and maximum complexity are reduced with active topic choice, a behavior that is amplified as larger populations are considered. The minimal counts strategy yields a strictly minimum complexity (equal to the complexity of a completed lexicon), while converging as fast as the success threshold strategy. Further work will deal with other variants of the Naming Game (with different vocabulary update, population replacement, and different ratio for N, M and W). For the moment only the hearer's choice scenario is studied, because of its high robustness to changes in parameter values for the different strategies [90].

7.2. Lifelong Robot Learning And Development Of Motor And Social Skills

7.2.1. *Intrinsically Motivated Multitask Reinforcement Learning*

Participants: Sébastien Forestier [correspondant], Pierre-Yves Oudeyer, Fabien Benureau.
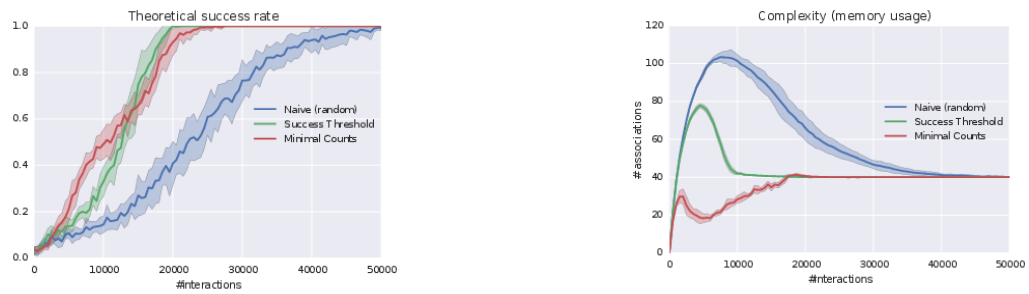
Figure 12. Strategy comparisons, in terms of convergence time (theoretical success ratio) and complexity level (memory usage). In this case, the hearer is the one choosing the topic. M=W=N=40, averaged over 8 trials
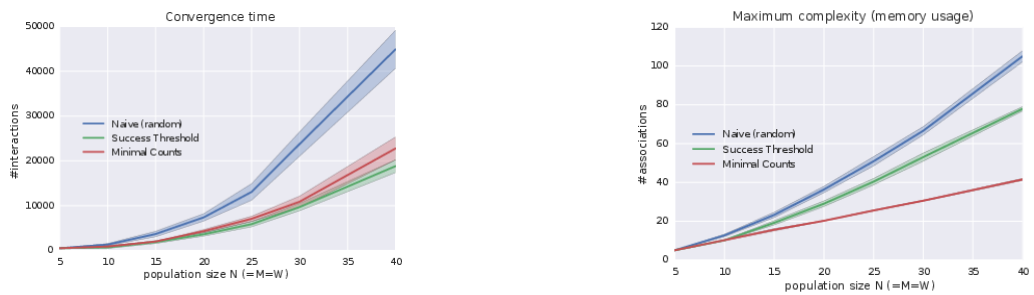

Figure 13. Scaling of maximum memory usage and convergence time for the different strategy, in function of population size. In this case, the hearer is the one choosing the topic. M=W=N, averaged over 8 trials.

7.2.1.1. Intrinsically Motivated Exploration of Spaces of Parameterized Skills/Tasks and Application to Robot Tool Learning

A major challenge in robotics is to learn 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 problems 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 [109].

In this work we study algorithms used by a learner to explore high-dimensional structured sensorimotor spaces such as in tool use discovery. We consider goal babbling architectures that were designed to explore and learn solutions to fields of sensorimotor problems, i.e. to acquire inverse models mapping a space of parameterized sensorimotor problems/effects to a corresponding space of parameterized motor primitives. However, so far these architectures have not been used in high-dimensional spaces of effects. Here, we show the limits of existing goal babbling architectures for efficient exploration in such spaces, and introduce a novel exploration architecture called Model Babbling (MB). MB exploits efficiently a modular representation of the space of parameterized problems/effects. We also study an active version of Model Babbling (the MACOB architecture). We compared those architectures in a simulated experimental setup with an arm that can discover and learn how to move objects using two tools with different properties, embedding structured high-dimensional continuous motor and sensory spaces (See Fig. 14).

7.2.1.2. Transfer Learning through Measures of Behavioral Diversity Generation in Autonomous Exploration

The production of behavioral diversity – producing a diversity of effects – is an essential strategy for robots exploring the world when facing situations where interaction possibilities are unknown or non-obvious. It allows to discover new aspects of the environment that cannot be inferred or deduced from available knowledge. However, creating behavioral diversity in situations where it is most crucial – new and unknown ones – is far from trivial. In particular in large and redundant sensorimotor spaces, only small areas are interesting to explore for any practical purpose. When the environment does not provide clues or gradient toward those areas, trying to discover those areas relies on chance. To address this problem, we introduce a method to create behavioral diversity in a new sensorimotor task by re-enacting actions that allowed to produce behavioral diversity in a previous task, along with a measure that quantifies this diversity. We have shownd that our method can learn how to interact with an object by reusing experience from another, that it adapts to instances of morphological changes and of dissimilarity between tasks, and how scaffolding behaviors can emerge by simply switching the attention of the robot to different parts of the environment. Finally, we show that the method can robustly use simulated experiences and crude cognitive models to generate behavioral diversity in real robots. This work was published in [62].

We presented the results at the IEEE/RSJ International Conference on Intelligent Robots and Systems, Daejeon, Korea, October 9-14th [81].

7.2.2. Social Learning of Interactive Skills

Participants: Manuel Lopes [correspondant], Thibaut Munzer, Marc Toussaint, Li Wang Wu, Yoan Mollard, Baptiste Busch, Jonathan Grizou, Marie Demangeat, Freek Stulp.

7.2.2.1. Relational Activity Processes for Modeling Concurrent Cooperation

In human-robot collaboration, multi-agent domains, or single-robot manipulation with multiple end-effectors, the activities of the involved parties are naturally concurrent. Such domains are also naturally relational as they involve objects, multiple agents, and models should generalize over objects and agents. We propose a novel formalization of relational concurrent activity processes that allows us to transfer methods from standard relational MDPs, such as MonteCarlo planning and learning from demonstration, to concurrent cooperation domains. We formally compare the formulation to previous propositional models of concurrent
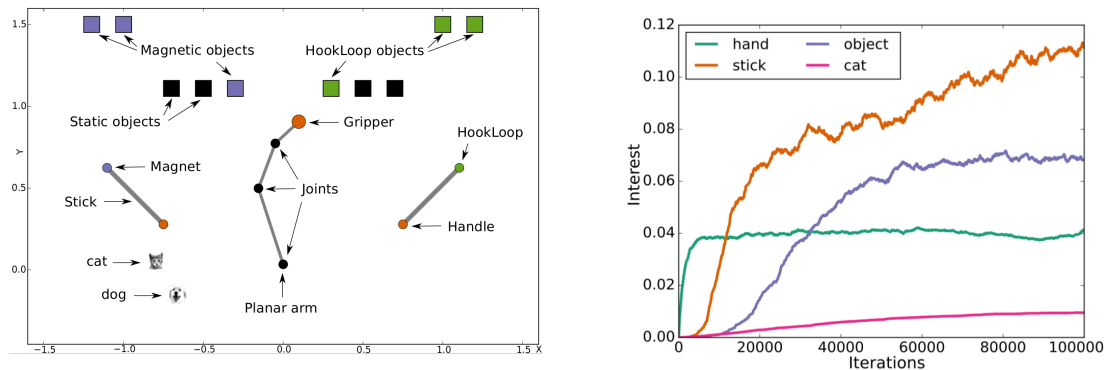
Figure 14. Left: Simulated robotic setup with a robotic arm that can grab tools to retrieve some interesting objects among a set of controllable and non-controllable (cat and dog) objects. Right: Evolution of the self-measured learning progress to move objects, with the MACOB active exploration architecture. The learning progress to explore objects is increasing for the tool and toy objects and stays low for the uncontrollable animals.

decision making and demonstrate planning and learning from demonstration methods on a real-world human-robot assembly task. A paper summarizing this research has been publish to the *International Conference on Robotics and Automation* (ICRA) 2016 [84].

7.2.2.2. Interactive Behavior Learning for Cooperative Tasks

This work goal is to propose a method to learn cooperative behavior to solve a task while performing the task with the user. The proposed approach reuses previous work on learning policy for RAP. The main differences are: i) formulate the problem as a cooperative process. In MDP and RAP, it is assumed that there is one central decision maker. However, in a cooperative both the robot and the operator are taking decisions. ii) estimating the confidence. A Query by Bagging approach has been used where many policies are learned from a subset of the data. Their potential disagreement allows quantifying the confidence. iii) Using the confidence for autonomous acting and for query making. Based on the confidence, the robot either act before acting or ask confirmation before acting.

Results show that using an interactive approach require less instruction from the user while producing a policy that makes fewer mistakes. We developed a robotic implementation 15 using a Baxter robot. A first article resulting from this work focusing on interactive preferences learning have been submitted to the *International Conference on Robotics and Automation* (ICRA) 2017 and a video demonstration can be view at : https://vimeo.com/182913540. A broader journal article is in preparation. We also conducted a user study to evaluate the impact of interactive learning on naïve users acceptation and performances.

7.2.2.3. Legible Motion

In a human-robot collaboration context, understanding and anticipating the robot intentions ease the completion of a joint-task. Whereas previous work has sought to explicitly optimize the legibility of behavior, we investigate legibility as a property that arises automatically from general requirements on the efficiency and robustness of joint human-robot task completion.

Following our previous work on legibility of robot motions [56], we have conducted several user experiments to analyze the effects of the policy representation on the universality of the legibility.

This work lead to a submission of a journal article to the International Journal of Social Robotics (IJSR) under the special issue: Towards a Framework for Joint Action. The article has been accepted with minor revisions and is currently in the final stage of the review process.
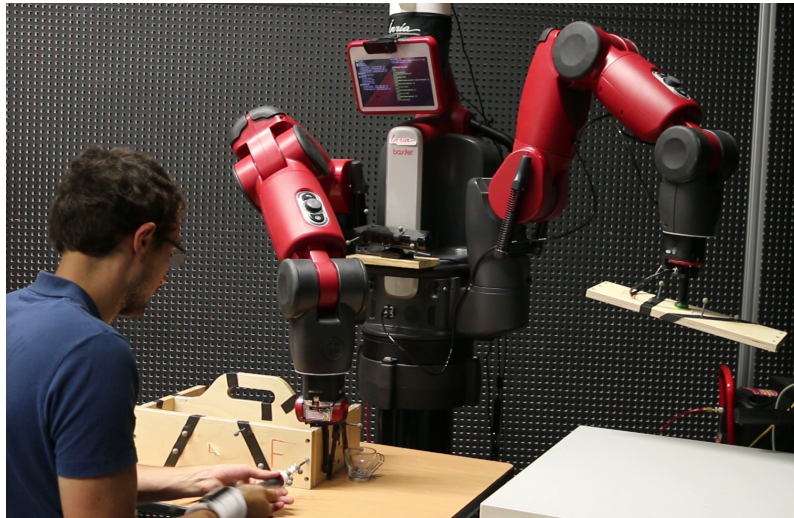
Figure 15. Interactive cooperative task learning.

7.2.2.4. Postural optimization for a safe and comfortable human-robot interaction

When we, humans, accomplish a task our body posture is (partially) constrained. For example, acting on an object constrains the pose of the hand relatively to the object, and the head faces the object we are acting upon. But due to the large number of degrees of freedom (DOF) of the human body, other body parts are unconstrained and several body postures are viable with respect to the task. However, not all of them are viable in terms of ergonomics. Using a personalized human model, observational postural assessment techniques can be automatized. Optimizing the model body posture is then the logical next step to find an ergonomically correct posture for the worker to accomplish a specific task.

To optimize the subject's model to achieve a specific task, we define an objective function that minimizes the efforts of the whole body posture, based on the Rapid Entire Body Assessment (REBA) technique [135]. The objective function also account for visibility of the target object and worker's laterality. We have also implemented an automatic assessment of the worker's body posture based on the REBA method.

Using a spherical object, carried by a Baxter humanoid robot as illustrated in Fig. 16 , we mimic an industrial scenario where the robot helps the worker by positioning and orienting an object in which the worker has to insert specific shapes. In a user-study with forty participants, we compare three different robot's behaviors, one of them being the result of the postural optimization of the subject's personalized model. By the mean of a survey session, and the online assessment of the subject's posture during the interaction, we prove that our method leads to a safer posture, and is perceived as more comfortable.

This work has been submitted to the IEEE Robotics and Automation Letters (RA-L) with the ICRA conference option and is currently under review.

7.3. Representation Learning

Participants: David Filliat [correspondant], Celine Craye, Yuxin Chen, Clement Masson, Adrien Matricon, Freek Stulp.

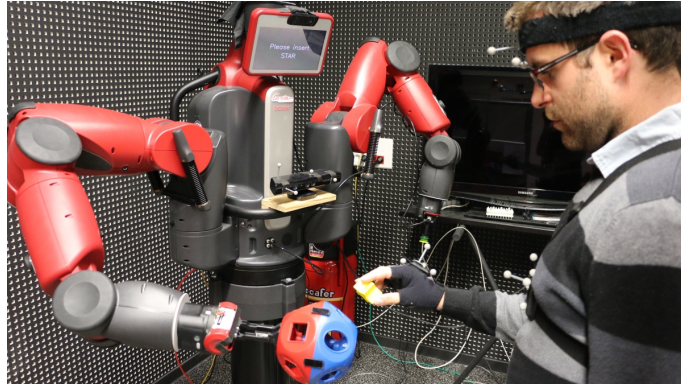### 7.3.1. *Incremental Learning of Object-Based Visual Saliency*

Figure 16. Representation of the setup considered in the user study. The robot presents to the user a spherical ball in which multiple shapes can be inserted. Final pose of the object is calculated from the user posture at his current location. Body motions during the insertion are recorded using a suit made from OptiTrack markers.

Searching for objects in an indoor environment can be drastically improved if a task-specific visual saliency is available. We describe a method to learn such an object-based visual saliency in an intrinsically motivated way using an environment exploration mechanism. We first define saliency in a geometrical manner and use this definition to discover salient elements given an attentive but costly observation of the environment. These elements are used to train a fast classifier that predicts salient objects given large-scale visual features. In order to get a better and faster learning, we use intrinsic motivation to drive our observation selection, based on uncertainty and novelty detection. Our approach has been tested on RGB-D images, is real-time, and outperforms several state-of-the-art methods in the case of indoor object detection. We published these results in two conferences [78],[77].

7.3.2. *Cross-situational noun and adjective learning in an interactive scenario*

Learning word meanings during natural interaction with a human faces noise and ambiguity that can be solved by analysing regularities across different situations. We propose a model of this cross-situational learning capacity and apply it to learning nouns and adjectives from noisy and ambiguous speeches and continuous visual input. We compared two different topic models for this task: Non Negative Matrix Factorization and Latent Dirichlet Association. We present experiments on learning object names and color names showing the performance of these model on realistic data and show how active learning can be used to speed-up learning by letting the learner choose the objects to be described. We published these results in a conference paper [75]

7.3.3. *Learning representation with gated auto-encoders*

We investigated algorithms that would be able to learn relevant visual or multi-modal features from data recorded while the robot performed some task. Representation learning is a currently very active research field, mainly focusing on deep-learning, which investigates how to compute more meaningful features from the raw high dimensional input data, providing a more abstract representation from which it should be easier to make decision or deduction (e.g classification, prediction, control, reinforcement learning). In the context of robotics, it is notably interesting to apply representation learning in a temporal and multi-modal approach exploiting vision and proprioception so as to be able to find feature that are relevant for building models of the robot itself and of its actions and their effect on the environment. Among the many existing approaches, we decided to explore the use of gated auto-encoders [104], a particular kind of neural networks including multiplicative connections, as they seem well adapted to this problem. Preliminary experimentations have been carried out with gated auto-encoders to learn transformations between two images. We observed that

Gated Auto-Encoders (GAE) can successfully find compact representations of simple transformations such as translations, rotation or scaling between two small images. This is however not directly scalable to realistic images such as ones acquired by a robot's camera because of the number of parameters, memory size and compational power it would require (unless drastically downsampling the image which induces sensible loss of information). In addition, the transformation taking an image to the next one can be the combination of transformations due to the movement of several object in the field of view, composed with the global movement of the camera. This induces the existence of an exponential number of possible transformations to model, for which the basic GAE architecture is not suited.

7.3.4. *Incremental Learning in high dimensions*

Participants: Alexander Gepperth [correspondant], Cem Karaoguz.

7.3.4.1. Incremental learning in data spaces of high dimensionality

Currently existing incremental learning algorithms in robotics have achieved a relatively high degree of usability due to the reduction of free model parameters in such approaches LWPR. Indeed, such algorithms are usually applied to low-dimensional tasks such as graspin with very good success, as the incremental learning paradigm is very appropriate to the robotics domain in general, especially in interactive scenarios. On the other hand, the partitioning of input space that is performed by LWPR and related approaches fails to be applicable if data dimension exceeds 50 elements since the used covariance matrices grow quadratically in size w.r.t. data dimensionality. Therefore, especially the incremental treatment of visual information is difficult, particularly for recognition and classification of objects or obstacles in general. To remedy this, we developed the incremental learning algorithm PROPRE [130] of fixed model complexity that can easily deal with data dimensionalities of 10.000 and beyond, where the only assumption is the same that is explicitly made for LWPR: that the data has structure, i.e., lies on a low-dimensional sub-manifold. We demonstrated the feasibility of the algorithm on several realistic datasets, on the one hand MNIST and on the other hand a much more challenging visual pedestrian pose recognition task from the intelligent vehicle domain[65].

7.3.4.2. Incremental learning with two memory systems

In order to increase PROPRE's ability to react quickly to changes in data statistics (e.g., a newly added visual class) while at the same time avoiding fast forgetting, a second, short-term memory system was proposed for PROPRE in [65]. This short-term memory is filled when task failures occur and is used to re-train the incremental long-term memory at a later time and on a slower time scale. In this way, abrupt changes in data statistics maybe immediately reacted upon, whereas the long-term memory can retain its stability that ensures that any forgetting happens gradually, on a determined time scale.

7.3.4.3. Steps towards incremental deep learnig

Since PROPRE is a neural architecture with just one hidden layer, its capacity is limited. This is why steps were taken to create deeper hierarchies with PROPRE in afashion totally analogous to current deep learning approaches. First of all, it was shown that a deep PROPRE architecture can achieve the same classification accuracy on MNIST as a shallow one but at a significantly lower computational cost [86]. Furthermore, it was shown that a deep PROPRE architecture is capable of change detection at multiple levels, a prerequisite for incremental learning [87]. Next steps will consist of creating regular deep PROPRE architectures and testing them on curently accepted machine learning benchmark tasks.

7.3.4.4. Real-world application of incremental learning

In [88], the incremental PROPRE algorithm was applied to object recognition and detection problems in the domain of intelligent vehicles. I was shown that, by re-casting pedestrian detection as an incremental learning problem where the background class is added only after learnig the pedestrian class, the number of required model resources for representing the background is reduced, and better accuracy can be obtained.

7.3.5. *Measuring Uncertainty in Deep Learning Networks*

Participants: Florian Golemo [correspondant], Manuel Lopes.

As precursor to the main objective of the IGLU project, we investigated methods that would enable deep neural networks to judge their knowledge about a domain.

Neural networks, especially deep ones, have been shown to be able to model arbitrarily complex problems, and thus offer powerful tools for machine learning. Yet they come with a significant flaw of not being inherently able to represent certainty of their predictions. By adding a measure of uncertainty to neural networks, this technology could be applied to autonomous exploration and open-ended learning tasks.

Thus the goal of this project was to find a method to measure how much knowledge a neural network has about about an unlabeled data item (measure of uncertainty), and to apply this new measure in an active learning context. The objective of the latter was to demonstrate the efficiency in handpicking interesting data, to optimally extend the system's own capabilities.

We were successful in finding a measure of uncertainty that would reliably distinguish data that the network has seen before, from data that was generally unfamiliar to the network. This measure was created by measuring the entropy of the network's last layer across a batch of stochastic samples generated by adding Poisson noise to the inputs.

The measure failed however to outperform random sampling in several active learning scenarios. Yarin Gal published related work as part of his dissertation [129] after this project was concluded. He elaborated that deep neural networks are very effective in canceling out input noise. The author suggested to use existing "Dropout" layers instead for stochastic sampling, but he reaches the same conclusion of using the last layer entropy as measure of uncertainty.

7.3.6. *Learning models by minimizing complexity*

We introduce COCOTTE (COnstrained Complexity Optimization Through iTerative merging of Experts), an iterative algorithm for discovering discrete, meaningful parameterized skills and learning explicit models of them from a set of behaviour examples. We show that forward-parameterized skills can be seen as smooth components of a locally smooth function and, framing the problem as the constrained minimization of a complexity measure, we propose an iterative algorithm to discover them. This algorithm fits well in the developmental robotics framework, as it does not require any external definition of a parameterized task, but discovers skills parameterized by the action from data. An application of our method to a simulated setup featuring a robotic arm interacting with an object is shown. This work was published in a conference paper [83]

7.4. Applications for Robotic myoelectric prostheses: co-adaptation algorithms and design of a 3D printed robotic arm prosthesis

Participants: Pierre-Yves Oudeyer [correspondant], Manuel Lopes, Mathilde Couraud, Sebastien Mick, Aymar de Rugy, Daniel Cattaert, Florent Paclet.

Together with the Hybrid team at INCIA, CNRS, the Flowers team continued to work on establishing the foundations of a long-term 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. In particular, in 2016 two lines of work were achieved, concerning two important scientific challenges, and in the context of one PEPS CNRS projects:

First, a new version of the experimental setup was designed to allow fast prototyping of 3D printed robotic prostheses. This work was based on the use of the Poppy open-source modular platform, and resulted in a functional prototype. A video demonstrations is available at: https://github.com/s-mick

Second, we have designed various control models allowing to transform signals coming from the human arm (either measured through EMGs or direct force sensors) and we have studied the influence of control modes on usability in the operation of a robotic arm prosthesis. In this context, we designed an experimental framework centered on a target-reaching task, and carried out tests with healthy subjects. The usability assessment relies on performance metrics on one hand, and a post-experiment questionnaire on another hand, in order to explore the multiple dimensions of the system's usability rather than focus only on measurements evaluating skills and performances. The code associated to this experimental setup is open-source and available at https://github.com/s-mick.

7.5. Applications for Educational Technologies

7.5.1. *Multi-Armed Bandits for Adaptive Personalization in Intelligent Tutoring Systems*

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

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 present 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 introduce 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 is evaluated in a scenario where 7-8 year old schoolchildren learn how to decompose numbers while manipulating money. Systematic experiments are presented with simulated students, followed by results of a user study across a population of 400 school children. [14]

7.5.1.2. A Comparison of Automatic Teaching Strategies for Heterogeneous Student Populations

Online planning of good teaching sequences has the potential to provide a truly personalized teaching experience with a huge impact on the motivation and learning of students. In this work we compare two main approaches to achieve such a goal, POMDPs that can find an optimal long-term path, and Multi-armed bandits that optimize policies locally and greedily but that are computationally more efficient while requiring a simpler learner model. Even with the availability of data from several tutoring systems, it is never possible to have a highly accurate student model or one that is tuned for each particular student. We study what is the impact of the quality of the student model on the final results obtained with the two algorithms. Our hypothesis is that the higher flexibility of multi-armed bandits in terms of the complexity and precision of the student model will compensate for the lack of longer term planning featured in POMDPs. We present several simulated results showing the limits and robustness of each approach and a comparison of heterogeneous populations of students.

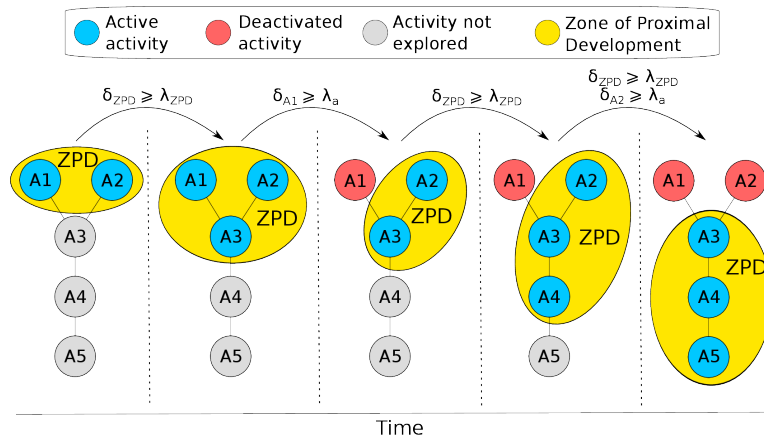This work has been publish and presented at Educational Data Mining 2016 conference in Raleigh, USA [76].

Figure 17. ZPDES exploration of an activity graph, with δ_{ZDP} the success rate over all active activities, λ_{ZPD} the threshold to expand the ZPD, δ_{Ax} the success rate for the activity Ax, and λ_a the threshold to reach to deactivate an activity.

Github link of the experiments paper code : https://github.com/flowersteam/kidlearn/tree/edm2016

7.5.1.3. The KidBreath project

To create learning contents linked to asthma to personalize it like mathematics activities in Kidlearn project [14] we used recommandation criterias in Therapeutic Education Program for asthma kids made by Health High Autority. Following an approach of participatory design [114], contents were validated by medical experts like health educators, pulmonoligists and pediatrics. Then, we conducted a workshop with forty kids aged 8 in order to iterate over the application interfaces and evaluate enjoy about it with observations. Finally, we realized a focus group with 5 asthma kids to validate the global comprehension of a part of the content. It revealed that children wanted more contents about the crisis treatment and how the asthma works in the human system (verbatims).

In a preliminary study, we experimented two conditions in 20 control children (with 3 asthma kids), one giving the possibility of choosing activities like the child wants, and one no giving this choice (activities displayed in random). No significative difference between the two groups, but results showed KidBreath was easy to use with scores > 75 using System Usability Scale [115]. Based on Cordova and Lepper works to evalueate motivation and knowledge with similar system and population [121], children had their disease knowledge increased with just one week use and were motivated using it. Finally, asthma kids showed they were more engaged than healthy kids and used KidBreath more seriously (stayed in breaks). These results was presented in the 5th edition of Serious Games in Medicine Conference in Nice.

We presented Thesis project in some events this year, with one publication surbmitted and validated:

- 2nd Meeting for Aquitaine and Euskadi companies in Biology and Health Between, February 11th 2016 in San sebastian (poster),
- Hackathon of innovation in pulmonary diseases, Respirhacktion, September 16th to 18th 2016 in Paris (project development in hackathon),
- 5th Conference in Health Ergonomics and Patient Safety, October 5th to 7th 2016 (poster) [102],
- Learning Lab day, November 16th in Inria Paris (oral presentation of project),
- 5th edition of Serious Games in Medicine Conference, December 2nd to 3rd 2016 in Nice Sofia-Antipolis University (oral presentation).

7.5.2. *Poppy Education: Designing and Evaluating Educational Robotics Kits*

Participants: Pierre-Yves Oudeyer [correspondant], Didier Roy, Théo Segonds, Stéphanie Noirpoudre, Marie Demangeat, Thibault Desprez, Matthieu Lapeyre, Pierre Rouanet, Nicolas Rabault.

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.

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, engineer schools, ...). For secondary education and higher education, scientific literacy centers, Fablabs.

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

- Poppy Humanoid, a robust and complete robotics platform designed for genuine experiments in the real world and 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. 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, which is a variant of Scratch. Its features allow a thorough introduction of information technology.

- 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. It is possible in the classroom to work on the simulated model and then allow students to run their program on the physical robot.

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

This project is user centred design. The pedagogical tools of the project (real and virtual robots, pedagogical activities, etc.) are being created directly with the users and evaluated in real life by experiments. For our experimentations in the classroom we are mainly using the robot Poppy Ergo Jr (real and virutal) and Snap! Our purpose is to improve this pedagogical tools and to create pedagogical activities and resources for teachers.

- A pedagogical working group:

  At the beginning of the project, we established a pedagogical working group of 12 volunteers, teachers from different level (mainly high school teachers of the Aquitaine region) to help to design educational activities in line with the needs of the school curriculum and to test them in the classroom.

  At the beginning of the second year of the project we added 7 other teachers from different background (middle-school and high school teachers) into the group to add more diversity.

  We organised some training to help them to discover and learn how to use the robotics platform, then we met monthly to exchange about the project and to get some feedbacks from them.

  You can see the videos of pedagogical robotics activities here:
  https://www.youtube.com/playlist?list=PLdX8RO6QsgB7hM_7SQNLvyp2QjDAkkzLn

- Experiment and Evaluate the pedagogical kits:

  Some engineer of the Poppy Education team went to visit the teachers in their school to see and to evaluate the pedagogical tools (robot and activities) in real contexts of use.

Figure 18. Experiment robots and pedagogical activities in classroom

In addition to the observations in classroom, two trainee students of Master 2 in cognitive sciences (M. Demangeat, D. Thibaut) have established an experimental protocol to evaluate the utility and the integration of the pedagogical kits in class. They created and filled out questionnaires by teachers and students. The analyzes of the results are presented in their paper thesis.

This experimentations are helping us to understand the educational needs, to create and improve the pedagogical tools.

7.5.2.2. 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.

Last year student of "bachelor degree" (ENSAM-Talence) have designed, manufactured, assembled and programmed 4 different solutions to replace the fixed hand of Poppy by a gripper device: https://www.youtube.com/watch?v=DZjGaJk2fQk
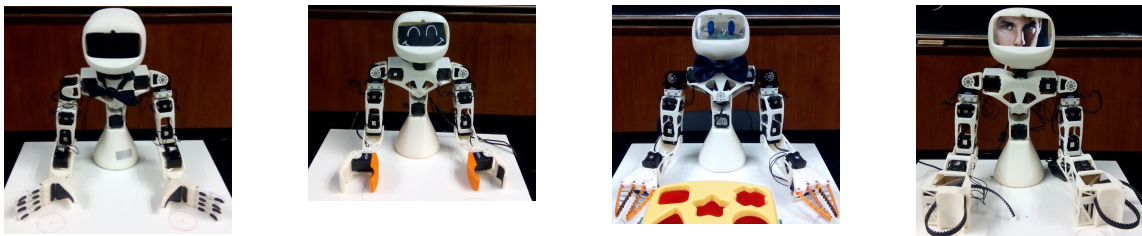

Figure 19. 4 grippers hands designed by students

- Audiovisual Students project

  Students from the BTS audiovisual of Saint-Genes La Salle have created a complete video report on the Poppy project to highlight the use in education and art:
  https://www.youtube.com/watch?v=NMwwH7AWO2Q

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

  The project "Poppy enters the dance" (Canope 33) uses the humanoid robot Poppy, 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 an overview of the project with kindergarten students :
  https://www.youtube.com/watch?v=XB9IXwcfJo0

7.5.2.3. Created pedagogical documents and resources

- Rebuilt the documentation of Poppy-project

  It was necessary for us to have an accessible and clear documentation to help teachers to use and create projects with the robots in the classroom so we rebuilt the existing documentation of the robotics platform Poppy. We added and improve the contents and we used the platform gitbook :
  https://docs.poppy-project.org/en/

- Pedagogical booklet

  The pedagogical booklet [96] brings together all the pedagogical activities and project testing in the classroom. It provides guided activities, small challenges and projects to become familiar with the Poppy Ergo Jr robot and the Programming language Snap!
  https://drive.google.com/file/d/0B2jV8VX-lQHwTUxXZjF3OGxHVGM/view

Figure 20. Pedagogical booklet: learn to program the robot Poppy Ergo Jr in Snap!

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

- Guide on the pedagogical use of the kit Poppy Ergo Jr in classroom

  We wrote an article [95] to explain how to use the Robot Ergo Jr in a classroom. It includes a summary of the characteristics of the robots, activities example and give all the necessary resources: https://pixees.fr/dans-la-famille-poppy-je-voudrais-le-robot-ergo-jr

- Demonstration guide to introduce the project

  This document is for people who already have a little experience with the Poppy Ergo Jr robot and snap! and wishing to present the project (i.e: to a colleague/acquaintance, on a exhibition stand, during a conference).

  The purpose of this document is to provide the necessary elements to enable the Poppy Education project to be presented through the use of Poppy Ergo Jr. robot. The key points of the Poppy Education project and the features of Poppy Ergo Jr kit are presented as well as examples of demonstrations of educational activities (videos and snap! projects) and educational projects (videos). An example of structuring a demo is provided at the end of the document. https://forum.poppy-project.org/t/guide-de-demo-du-kit-pedagogique-poppy-ergo-jr-version-beta/2698

- Model of pedagogical activities sheet

  We designed a model of pedagogical activity sheet. It helps us to get back the various activities and allows to have a homogeneous presentation. It is simpler to share and get back the creations of each. https://forum.poppy-project.org/t/modele-de-fiche-pedagogique-telechargeable-pour-les-activites-robotiques/2706

7.5.2.4. Scientific mediation

To promote educational uses of the platform, we participated in events (conference, seminar etc.).

We participated as well at some workshops to introduce students to robotics and programming.

7.5.2.5. Symposium robotics

We organized a symposium robotics (http://dm1r.fr/colloque-robotique-education/) that present research results and feedback on the use of Poppy and Thymio robots in education (other robots have been discussed, such as BeeBot and Metabot), from kindergarten to higher education, The Centers for Scientific and Technical Culture.

It was a 2 day event : 200 participants, 40 speakers (conferences and workshops).

Poppy Education team and the working group teachers helped with the organisation of the event and during the event (talk and workshops).

All conference videos are available on the web : https://www.youtube.com/watch?v=prFmC-BpdY8&index=1&list=PL9T8000j7sJBC_H3L_hS-i4Ltlh1Fz2FY

7.5.3. IniRobot: Educational Robotics in Primary Schools

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

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

IniRobot Project consists to produce and diffuse a pedagogical kit for teachers and animators, to help 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 enquiry-based educational approach, where kids are led to construct their understanding through practicing an active investigation methodology within teams. It is based on the use of the Thymio II robotic platform. More details about this projects were published in RIE 2015 [50], which presents the detailed pedagogical objectives and a first measure of results showing that children acquired several robotics-related concepts. See also http://www.inirobot.fr.

Deployment: After 24 months of activity, IniRobot is used by about 1400 adults and 16 000 children in 54 cities of France. Example of action in university: MEEF teacher training for the hope of Aquitaine. Example of action in school: training of all Gironde Pedagogical ICT Advisors, covering nearly 1000 schools. Example of action in the extracurricular time: training 82 facilitators TAP cities of Talence, Pessac, Lille, ..., CDC Gates of inter-seas. Example of national action: Training of the digital mediators of the 8 Inria centers.

7.5.3.1. Partnership

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

7.5.3.2. Created pedagogical documents and resources

- IniRobot pedagogical kit [94]: This pedagogical booklet provides activities scenarized as missions to do. A second pedagogical booklet has been also created by three pedagogical advisers for primary school, with pedagogical instructions and aims, under ou supervision. http://tice33.ac-bordeaux.fr/Ecolien/ASTEP/tabid/5953/language/fr-FR/Default.aspx A new pedagogical kit is in progress, Inirobot Scratch, which will propose activities with Scratch and Snap! and Thymio robot.

- Inirobot website and forum http://www.inirobot.fr With this website, teachers, animators and general public can download documents, exchange about their use of inirobot' kit.

- Publication about Inirobot and Poppy Education A poster and talk were produced in Didapro-Didastic 6 Conference in Namur (Belgium) on 2016 January. [99]

7.5.3.3. Scientific mediation

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

7.5.3.4. Symposium robotics

With Poppy Education, Inirobot is a main line in our colloquium "Robotics and Education" (http://dm1r.fr/colloque-robotique-education/)

<div align="center">

GEOSTAT Project-Team

</div>

6. New Results

6.1. Automatic segmentation of activation periods in an electrogram during atrial fibrillation

Participants: G. Attuel, H. Yahia.

Experiments show that the multiscale properties displayed in signals recording the electrical activity of the heart during (atrial) fibrillation are of the type *out of equilibrium dynamics*. These dynamics have common features, possibly shared by "subclasses": "$1/f$ power spectrum", large proability laws for the distributions of the amplitude increments, multifractal spectra. Theoretically these dynamics are at least the result of a competition between elastic energies and disorder, which leads to the emergence of collective behaviours. Mathematically, the universality classes involved are not those corresponding to the central limit theorem, but generalize it in its more elaborated forms (Levy & Gnedenko). A class has recently been described completely: directed polymers on a random medium. Scaling exponents are known, together with the fluctuations' statistics. Large deviation theory plays a central role. The fixed point of the associated dynamics is that of KPZ (Kardar-Parisi-Zhang). We have indications that heartbeat dynamics in episodes of atrial fibrillation belongs to that class. In such a context, the questions raised relate to finite size effects when asymptotic convergence is slow. From an experimental point of view, the problem of determining universality classes is very hard. However, it is possible to formulate hypothesis on the universality class so as to extract important information from acquisition signals. A good modus operandi consists in using key properties of a model stated *a priori*, to combine them with experimental signal analysis in order to produce *a posteriori* characteristics of interest. Last year, we developed the first model of cardiac dynamics compatible with observed data. The model allows us to test the efficiency of a combined methodology using singularity exponents and Bayesian analysis. This has led us to a first automatic method able to identify periods of cardiac activity and make the distinction with measure noise. From this, the fine automatic determination of activity periods become tractable. This will lead to an automatic quantitative determination of fragmentation, hence opening the way for a determination of universality classes. We illustrate some steps in the figures below.

Publications: [24], [23], [25]

6.2. Pathological voice classification

Participants: K. Daoudi, N. Brodu.

We propose a fully reproducible speech-based technique for objective differential diagnosis between progressive supranuclear palsy (PSP) and multiple system atrophy (MSA). Our technique yields a classification mean accuracy of 86.1% which is a significant improvement as compared to a recent pioneer study on this task. We also show that information extracted through a variety of speech tasks can be used to estimate the degree of Parkinson's disease severity.

Publications: [18], [20], [https://hal.inria.fr/hal-01360038].

6.3. Temporal evolution of coastal upwelling

Participants: A. El Aouni, K. Minaoui, A. Tamim, K. Daoudi, H. Yahia, A. Atillah, D. Aboutajdine.
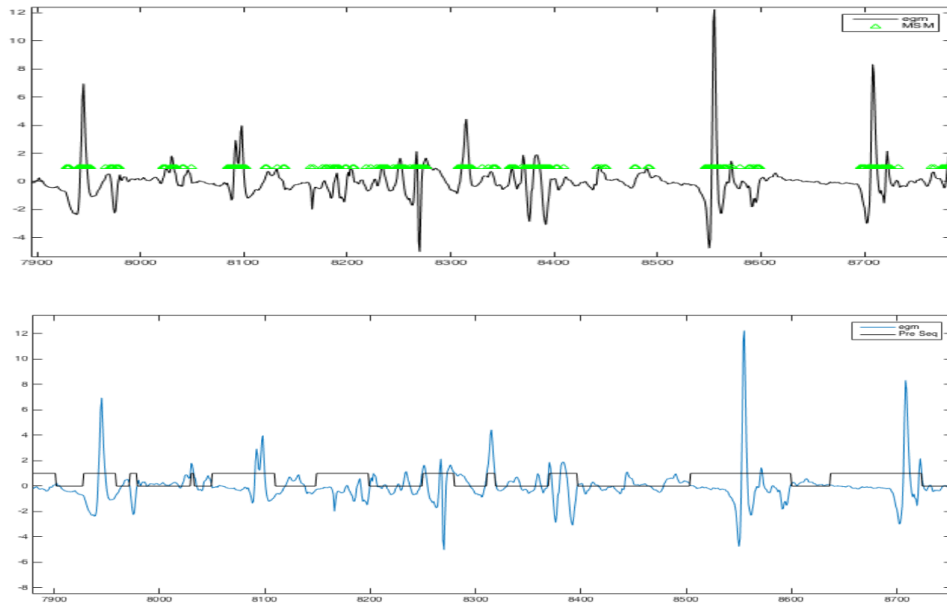
Figure 1. Egm during fibrilllation (in black). Density of the most singular manifold (in green). Result of a 2-state HMM (in blue).
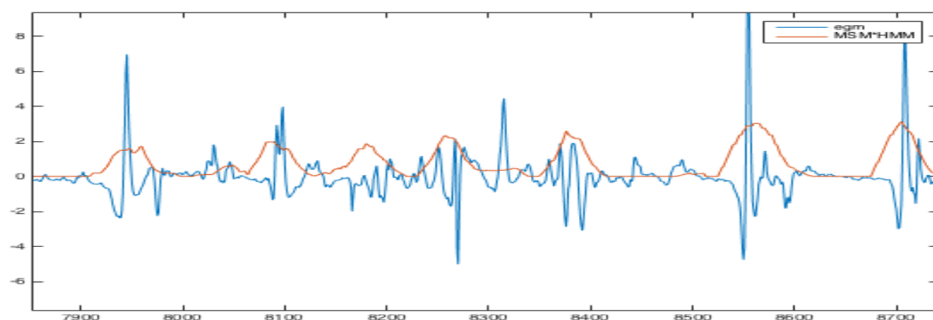

Figure 2. Egm during fibrillation (in blue). Signal of activation probability computed with the result of an HMM and the singularity density (in red).
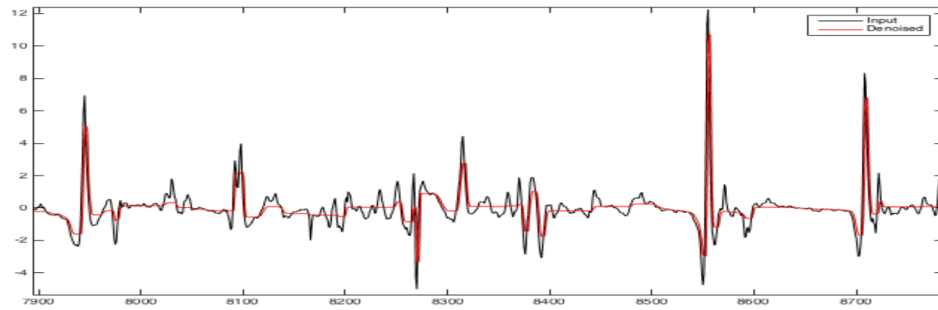
Figure 3. Egm during fibrillation (in black). Signal denosised (norm L^p) (in red).
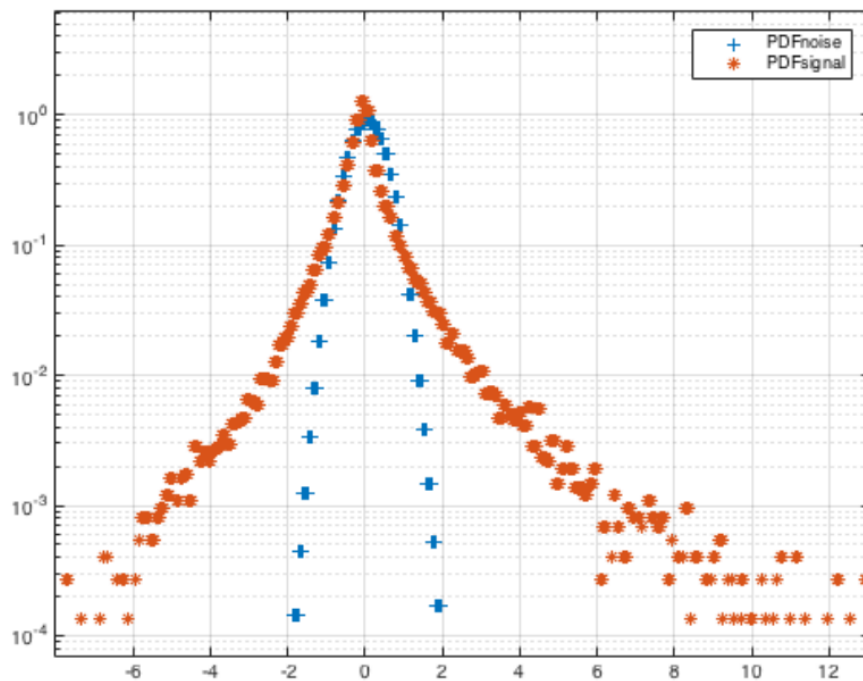

Figure 4. Probability distribution of estimated noise (in blue). Distribution of estimated active dynamics (in red).
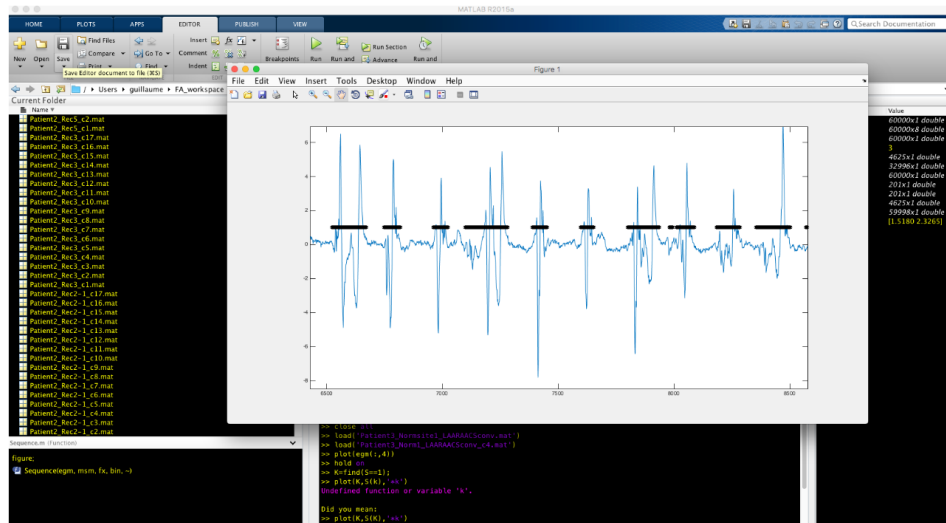
Figure 5. Snapshot of the software (written partially in Matlab and which makes use of FluidExponents). The image shows the result of sequencing, with average confidence level, of a real egm

We present a new methodology to derive rigorous SST-based coastal upwelling index for the purpose of conducting a saisonal variablity of upwelling area along the Moroccan Atlantic coast. The method is based on the scientific knowledge of upwelling area and its spatial dis- tribution provided by expert oceanographers.The latter consists in automatically identify and extract the region covered by the upwelling waters in the costal ocean of Morocco using the Fuzzy c-means algorithm and finding regions of homogeneous pixels. Then Region Growing process is used to filter out the remaining noisy structures in the offshort waters. The methodology is used to provide a satistical view of the spatial and temporal variability of the Moroccan upwelling activity. The relevance of the proposed Coastal Upwelling Index (CUI) is evaluated by an oceanographer using 86 8-days sea surface temperature images and it is shown to be superior to that of the standard upwelling index.

Publication: [https://hal.inria.fr/hal-01424036].

6.4. Non-local and low rank approach for integrability

Participants: H. Badri, H. Yahia.

A formulation is proposed which consists in a sparse gradient data-fitting term to handle outliers together with a gradient-domain non-local low-rank prior.

Publication: [15].

6.5. Low-Rankness transfer for realistic denoising

Participants: H. Badri, H. Yahia, D. Aboutajdine.

Current state-of-the-art denoising methods such as non-local low-rank approaches are mainly tuned to work with uniform Gaussian noise corruption and known variance, which is far from the real noise scenario. Noise level estimation is already a challenging problem and denoising methods are quite sensitive to this parameter. Moreover, these methods are based on shrinkage models that are too simple to reflect reality, which results in over-smoothing of important structures such as small-scale text and textures. We propose a new approach for more realistic image restoration based on the concept of low-rankness transfer (LRT).

Publication: [14].

6.6. Multiscale methods for Earth Observation data

Participants: H. Yahia, N. Brodu, V. Garçon, J. Sudre, S. Kumar Maji, D. Singh, K. Daoudi, D. Aboutajdine. Earth observation data of different kinds are tested for super-resolution or analysis using the multiscale approaches developped in the team. This paragraph is mainly concerned with the publications of last year results.

Publications: [https://hal.inria.fr/hal-01254482], [https://hal.archives-ouvertes.fr/hal-01425021], [https://hal.inria.fr/hal-01287182], [https://hal.inria.fr/hal-01287181], [https://hal.inria.fr/hal-01426666], [16], [27], [30], [28], [31], [21].

6.7. Signal analysis of ultrasonic dental response

Participants: H. Yahia, G. Rosi, S. Jaffard, S. Seuret.

The long-term success of a dental implant is related to the properties of the bone-implant interface. It is important to follow the evolution of bone remodeling phenomena around the implant. Methods based on ultrasound wave propagation were already successfully used by collaborators, in the qualitative and quantitative evaluation of primary and secondary stability of dental implants. Results, numerical and experimental, are analysed with signal processing tools based on multifractal methods. Analysis of the first results shows that these methods are potentially efficient in this case because they can explore and exploit the multi-scale structure of the signal.

Publication: [22].

6.8. Complexity in Electrophysiological Dynamics

Participants: O. Pont, H. Yahia, B. Xu.

Action potentials play an important role in the dynamics of cell-cell communication and they are thus of key relevance in neural tissues. We show that typical real-world electrophysiological signals, with smooth deviations from harmonicity, are typically well described with just the first few terms and result in a rather compact, sparse representation. In particular, we have done an analysis of FitzHugh-Nagumo impulse trains; we have found that 3 anharmonic terms reconstruct better than an equivalent 8-term Fourier representation, with less than half the PSNR and no artifacts from Gibbs phenomenon.

Publication: [17].

6.9. Nonlinear trend removal and heart rate variability analysis

Participants: B. Xu, R. Dubois, O. Pont, H. Yahia.

Publication: [29].

<p style="text-align:center;color:red;font-weight:bold;">HIEPACS Project-Team</p>

7. New Results

7.1. High-performance computing on next generation architectures

7.1.1. *Numerical recovery strategies for parallel resilient Krylov linear solvers*

As the computational power of high performance computing (HPC) systems continues to increase by using a huge number of cores or specialized processing units, HPC applications are increasingly prone to faults. In this paper, we present a new class of numerical fault tolerance algorithms to cope with node crashes in parallel distributed environments. This new resilient scheme is designed at application level and does not require extra resources, i.e., computational unit or computing time, when no fault occurs. In the framework of iterative methods for the solution of sparse linear systems, we present numerical algorithms to extract relevant information from available data after a fault, assuming a separate mechanism ensures the fault detection. After data extraction, a well chosen part of missing data is regenerated through interpolation strategies to constitute meaningful inputs to restart the iterative scheme. We have developed these methods, referred to as Interpolation-Restart techniques, for Krylov subspace linear solvers. After a fault, lost entries of the current iterate computed by the solver are interpolated to define a new initial guess to restart the Krylov method. A well suited initial guess is computed by using the entries of the faulty iterate available on surviving nodes. We present two interpolation policies that preserve key numerical properties of well-known linear solvers, namely the monotonic decrease of the A-norm of the error of the conjugate gradient or the residual norm decrease of GMRES. The qualitative numerical behavior of the resulting scheme have been validated with sequential simulations, when the number of faults and the amount of data losses are varied. Finally, the computational costs associated with the recovery mechanism have been evaluated through parallel experiments.

More details on this work can be found in [7].

7.1.2. *Interpolation-restart strategies for resilient eigensolvers*

The solution of large eigenproblems is involved in many scientific and engineering applications when for instance, stability analysis is a concern. For large simulation in material physics or thermo-acoustics, the calculation can last for many hours on large parallel platforms. On future large-scale systems, the mean time between failures (MTBF) of the system is expected to decrease so that many faults could occur during the solution of large eigenproblems. Consequently, it becomes critical to design parallel eigensolvers that can survive faults. In that framework, we investigate the relevance of approaches relying on numerical techniques, which might be combined with more classical techniques for real large-scale parallel implementations. Because we focus on numerical remedies we do not consider parallel implementations nor parallel experiments but only numerical experiments. We assume that a separate mechanism ensures the fault detection and that a system layer provides support for setting back the environment (processes, . . .) in a running state. Once the system is in a running state, after a fault, our main objective is to provide robust resilient schemes so that the eigensolver may keep converging in the presence of the fault without restarting the calculation from scratch. For this purpose, we extend the interpolation-restart (IR) strategies initially introduced for the solution of linear systems in a previous work to the solution of eigenproblems in this paper. For a given numerical scheme, the IR strategies consist of extracting relevant spectral information from available data after a fault. After data extraction, a well-selected part of the missing data is regenerated through interpolation strategies to constitute a meaningful input to restart the numerical algorithm. One of the main features of this numerical remedy is that it does not require extra resources, i.e. , computational unit or computing time, when no fault occurs. In this paper, we revisit a few state-of-the-art methods for solving large sparse eigenvalue problems namely the Arnoldi methods, subspace iteration methods and the Jacobi-Davidson method, in the light of our IR strategies. For each considered eigensolver, we adapt the IR strategies to regenerate as much spectral information as possible. Through extensive numerical experiments, we study the respective robustness of the resulting resilient schemes with respect to the MTBF and to the amount of data loss via qualitative and quantitative illustrations.

More details on this work can be found in [8].

7.2. High performance solvers for large linear algebra problems

7.2.1. *Exploiting Kepler architecture in sparse direct solver with runtime systems*

Many works have addressed heterogeneous architectures to exploit accelerators such as GPUs or Intel Xeon Phi with interesting speedup. Despite researches towards generic solutions to efficiently exploit those accelerators, their hardware evolution requires continual adaptation of the kernels running on those architectures. The recent Nvidia architectures, as Kepler, present a larger number of parallel units thus requiring more data to feed every computational units. A solution considered to supply enough computation has been to study problems with large number of small computations. The batched BLAS libraries proposed by Intel, Nvidia, or the University of Tennessee are examples of this solution. We have investigated the use of the variable size batched matrix-matrix multiply to improve the performance of a the `PaStiX` sparse direct solver. Indeed, this kernel suits the super-nodal method of the solver, and the multiple updates of variable sizes that occur during the numerical factorization.

These contributions have been presented at the PMAA'16 conference [28].

7.2.2. *Blocking strategy optimizations for sparse direct linear solver on heterogeneous architectures*

The preprocessing steps of sparse direct solvers, ordering and block-symbolic factorization, are two major steps that lead to a reduced amount of computation and memory and to a better task granularity to reach a good level of performance when using BLAS kernels. With the advent of GPUs, the granularity of the block computations became more important than ever. In this paper, we present a reordering strategy that increases this block granularity. This strategy relies on the block-symbolic factorization to refine the ordering produced by tools such as `METIS` or `Scotch`, but it does not impact the number of operations required to solve the problem. We integrate this algorithm in the `PaStiX` solver and show an important reduction of the number of off-diagonal blocks on a large spectrum of matrices. This improvement leads to an increase in efficiency of up to 10% on CPUs and up to 40% on GPUs.

These contributions have been presented at the SIAM PP'16 conference [35] and an extended paper has been submitted to SIAM Journal on Matrix Analysis and Applications [49].

7.2.3. *Sparse supernodal solver using hierarchical compression*

In the context of FASTLA associate team, during the last 3 years, we are collaborating with Eric Darve, professor in the Institute for Computational and Mathematical Engineering and the Mechanical Engineering Department at Stanford, on the design of a new efficient sparse direct solvers.

Sparse direct solvers such as `PaStiX` are currently limited by their memory requirements and computational cost. They are competitive for small matrices but are often less efficient than iterative methods for large matrices in terms of memory. We are currently accelerating the dense algebra components of direct solvers using hierarchical matrices algebra. In the first step, we are targeting an $O(N^{4/3})$ solver. Preliminary benchmarks indicate that a speed up of 2x to 10x is possible (on the largest test cases).

In the context of the FASTLA team, we have been working on applying fast direct solvers for dense matrices to the solution of sparse direct systems. We observed that the extend-add operation (during the sparse factorization) is the most time-consuming step. We have therefore developed a series of algorithms to reduce this computational cost. We presented a new implementation of the `PaStiX` solver using hierarchical compression to reduce the burden on large blocks appearing during the nested dissection process. To improve the efficiency of our sparse update kernel for both BLR (block low-rank) and HODLR (hierarchically off-diagonal low-rank), we are now investigating to BDLR (boundary distance low-rank) approximation scheme to preselect rows and columns in the low-rank approximation algorithm. We also have to improve our ordering strategies to enhance data locality and compressibility. The implementation is based on runtime systems to exploit parallelism.

Some contributions have already been presented at the workshops on Fast Solvers [32], [31], [30]. This work is a joint effort between Professor Darve's group at Stanford and the Inria HiePACS team within FASTLA.

7.2.4. Hierarchical hybrid sparse linear solver for multicore platforms

The solution of large sparse linear systems is a critical operation for many numerical simulations. To cope with the hierarchical design of modern supercomputers, hybrid solvers based on Domain Decomposition Methods (DDM) have been been proposed. Among them, approaches consisting of solving the problem on the interior of the domains with a sparse direct method and the problem on their interface with a preconditioned iterative method applied to the related Schur Complement have shown an attractive potential as they can combine the robustness of direct methods and the low memory footprint of iterative methods. In this report, we consider an additive Schwarz preconditioner for the Schur Complement, which represents a scalable candidate but whose numerical robustness may decrease when the number of domains becomes too large. We thus propose a two-level MPI/thread parallel approach to control the number of domains and hence the numerical behaviour. We illustrate our discussion with large-scale matrices arising from real-life applications and processed on both a modern cluster and a supercomputer. We show that the resulting method can process matrices such as tdr455k for which we previously either ran out of memory on few nodes or failed to converge on a larger number of nodes. Matrices such as Nachos_4M that could not be correctly processed in the past can now be efficiently processed up to a very large number of CPU cores (24 576 cores). The corresponding code has been incorporated into the `MaPHyS` package.

More details on this work can be found in [44]

7.2.5. Task-based conjugate gradient: from multi-GPU towards heterogeneous architectures

Whereas most parallel High Performance Computing (HPC) numerical libaries have been written as highly tuned and mostly monolithic codes, the increased complexity of modern architectures led the computational science and engineering community to consider more mod- ular programming paradigms such as task-based paradigms to design new generation of parallel simulation code; this enables to delegate part of the work to a third party software such as a runtime system. That latter approach has been shown to be very productive and efficient with compute-intensive algorithms, such as dense linear algebra and sparse direct solvers. In this study, we consider a much more irregular, and synchronizing algorithm, namely the Conjugate Gradient (CG) algorithm. We propose a task-based formulation of the algorithm together with a very ne instrumentation of the runtime system. We show that almost optimum speed up may be reached on a multi-GPU platform (relatively to the mono-GPU case) and, as a very preliminary but promising result, that the approach can be e ectively used to handle heterogenous architectures composed of a multicore chip and multiple GPUs. We expect that these results will pave the way for investigating the design of new advanced, irregular numerical algorithms on top of runtime systems.

More details on this work can be found in [42]

7.2.6. Analysis of rounding error accumulation in conjugate gradients to improve the maximal attainable accuracy of pipelined CG

Pipelined Krylov solvers typically offer better scalability in the strong scaling limit compared to standard Krylov methods. The synchronization bottleneck is mitigated by overlap- ping time-consuming global communications with useful computations in the algorithm. However, to achieve this communication hiding strategy, pipelined methods feature multiple recurrence re- lations on additional auxiliary variables to update the guess for the solution. This paper aims at studying the influence of rounding errors on the convergence of the pipelined Conjugate Gradient method. It is analyzed why rounding effects have a significantly larger impact on the maximal attainable accuracy of the pipelined CG algorithm compared to the traditional CG method. Fur- thermore, an algebraic model for the accumulation of rounding errors throughout the (pipelined) CG algorithm is derived. Based on this rounding error model, we then propose an automated residual replace- ment strategy to reduce the effect of rounding errors on the final iterative solution. The resulting pipelined CG method with automated residual replacement improves the maximal attainable accuracy of pipelined CG

to a precision comparable to that of standard CG, while maintaining the efficient parallel performance of the pipelined method.

More details on this work can be found in [46].

7.2.7. *Nearly optimal fast preconditioning of symmetric positive definite matrices*

We consider the hierarchical off-diagonal low-rank preconditioning of symmetric positive definite matrices arising from second order elliptic boundary value problems. When the scale of such problems becomes large combined with possibly complex geometry or unstable of boundary conditions, the representing matrix is large and typically ill-conditioned. Multilevel methods such as the hierarchical matrix approximation are often a necessity to obtain an efficient solution. We propose a novel hierarchical preconditioner that attempts to minimize the condition number of the preconditioned system. The method is based on approximating the low-rank off-diagonal blocks in a norm adapted to the hierarchical structure. Our analysis shows that the new preconditioner effectively maps both small and large eigenvalues of the system approximately to 1. Finally through numerical experiments, we illustrate the effectiveness of the new designed scheme which outperforms more classical techniques based on regular SVD to approximate the off-diagonal blocks and SVD with filtering.

This work is a joint effort between Professor Darve's group at Stanford and the Inria HiePACS team within FASTLA. More details on this work can be found in [41].

7.2.8. *Robust coarse spaces for abstract Schwarz preconditioners via generalized eigenproblems*

The solution of large sparse linear systems is one of the most important kernels in many numerical simulations. The domain decomposition methods (DDM) community has de- veloped many efficient and robust solvers in the last decades. While many of these solvers fall in Abstract Schwarz (AS) framework, their robustness has often been demonstrated on a case-by-case basis. In this paper, we propose a bound for the condition number of all deflated AS methods pro- vided that the coarse grid consists of the assembly of local components that contain the kernel of some local operators. We show that classical results from the literature on particular instances of AS methods can be retrieved from this bound. We then show that such a coarse grid correction can be explicitly obtained algebraically via generalized eigenproblems, leading to a condition number independent of the number of domains. This result can be readily applied to retrieve the bounds previously obtained via generalized eigenproblems in the particular cases of Neumann-Neumann (NN), additive Schwarz (aS) and optimized Robin but also generalizes them when applied with approximate local solvers. Interestingly, the proposed methodology turns out to be a comparison of the considered particular AS method with generalized versions of both NN and aS for tackling the lower and upper part of the spectrum, respectively. We furthermore show that the application of the considered grid corrections in an additive fashion is robust in the aS case although it is not robust for AS methods in general. In particular, the proposed framework allows for ensuring the robustness of the aS method applied on the Schur complement (aS/S), either with deflation or additively, and with the freedom of relying on an approximate local Schur complement, leading to a new powerful and versatile substructuring method. Numerical experiments illustrate these statements.

More details on this work can be found in [45]

7.3. High performance fast multipole method for N-body problems

7.3.1. *Task-based fast multipole method*

With the advent of complex modern architectures, the low-level paradigms long considered sufficient to build High Performance Computing (HPC) numerical codes have met their limits. Achieving efficiency, ensuring portability, while preserving programming tractability on such hardware prompted the HPC community to design new, higher level paradigms. The successful ports of fully-featured numerical libraries on several recent runtime system proposals have shown, indeed, the benefit of task-based parallelism models in terms of performance portability on complex platforms. However, the common weakness of these projects is to

deeply tie applications to specific expert-only runtime system APIs. The OPENMP specification, which aims at providing a common parallel programming means for shared-memory platforms, appears as a good candidate to address this issue thanks to the latest task-based constructs introduced as part of its revision 4.0. The goal of this paper is to assess the effectiveness and limits of this support for designing a high-performance numerical library. We illustrate our discussion with the `ScalFMM` library, which implements state-of-the-art fast multipole method (FMM) algorithms, that we have deeply re-designed with respect to the most advanced features provided by OPENMP 4. We show that OPENMP 4 allows for significant performance improvements over previous OPENMP revisions on recent multicore processors. We furthermore propose extensions to the OPENMP 4 standard and show how they can enhance FMM performance. To assess our statement, we have implemented this support within the KLANG-OMP source-to-source compiler that translates OPENMP directives into calls to the `StarPU` task-based runtime system. This study, [38] shows that we can take advantage of the advanced capabilities of a fully-featured runtime system without resorting to a specific, native runtime port, hence bridging the gap between the OPENMP standard and the very high performance that was so far reserved to expert-only runtime system APIs.

7.3.2. *Task-based fast multipole method for clusters of multicore processors*

Most high-performance, scientific libraries have adopted hybrid parallelization schemes - such as the popular MPI+OpenMP hybridization - to benefit from the capacities of modern distributed-memory machines. While these approaches have shown to achieve high performance, they require a lot of effort to design and maintain sophisticated synchronization/communication strategies. On the other hand, task-based programming paradigms aim at delegating this burden to a runtime system for maximizing productivity. In this article, we assess the potential of task-based fast multipole methods (FMM) on clusters of multicore processors. We propose both a hybrid MPI+task FMM parallelization and a pure task-based parallelization where the MPI communications are implicitly handled by the runtime system. The latter approach yields a very compact code following a sequential task-based programming model. We show that task-based approaches can compete with a hybrid MPI+OpenMP highly optimized code and that furthermore the compact task-based scheme fully matches the performance of the sophisticated, hybrid MPI+task version, ensuring performance while maximizing productivity. In [40] we illustrate our discussion with the ScalFMM FMM library and the StarPU runtime system.

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

7.4.1. *Load Balancing for Coupled Simulations*

In the field of scientific computing, the load balancing is an important step conditioning the performance of parallel programs. The goal is to distribute the computational load across multiple processors in order to minimize the execution time. This is a well-known problem that is unfortunately NP-hard. The most common approach to solve it is based on graph or hypergraph partitioning method, using mature and efficient software tools such as Metis, Zoltan or Scotch. Nowadays, numerical simulation are becoming more and more complex, mixing several models and codes to represent different physics or scales. Here, the key idea is to reuse available legacy codes through a coupling framework instead of merging them into a standalone application. For instance, the simulation of the earth's climate system typically involves at least 4 codes for atmosphere, ocean, land surface and sea-ice . Combining such different codes are still a challenge to reach high performance and scalability. In this context, one crucial issue is undoubtedly the load balancing of the whole coupled simulation that remains an open question. The goal here is to find the best data distribution for the whole coupled codes and not only for each standalone code, as it is usually done. Indeed, the naive balancing of each code on its own can lead to an important imbalance and to a communication bottleneck during the coupling phase, that can dramatically decrease the overall performance. Therefore, one argues that it is required to model the coupling itself in order to ensure a good scalability, especially when running on tens of thousands of processors. In this work, we develop new algorithms to perform a coupling-aware partitioning of the whole application.

Surprisingly, we observe in our experiments that our proposed algorithms do not highly degrade the global edgecut for either component and thus the internal communication among processors of the same component is still minimized. This is not the case for the *Multiconst* method especially as the number of processors increases. Regarding the coupled simulation for the real application AVTP-AVBP (provided by Cerfacs), we noticed that one may carefully decide the parameters of the co-partitioning algorithms in order not to increase the global edgecut. More precisely, the number of processors assigned in the coupling interface is an important factor that needs to be determined based on the geometry of the problem and the ratio of the coupling interface compared to the entire domain. Again, we remark that our work on co-partitioning is still theoretical and further investigation should be conducted with different geometries and more coupled simulations that are more or less coupling-intensive.

This work corresponds to the PhD of Maria Predari, defended on December 9$^{\text{th}}$ 2016.

7.5. Application Domains

7.5.1. Material physics

7.5.1.1. Molecular vibrational spectroscopy

Quantum chemistry eigenvalue problem is a big challenge in recent research. Here we are interested in solving eigenvalue problems coming from the molecular vibrational analysis. These problems are challenging because the size of the vibrational Hamiltonian matrix to be diagonalized is exponentially increasing with the size of the molecule we are studying. So, for molecules bigger than 10 atoms the actual existent algorithms suffer from a curse of dimensionality or computational time.

A new variational algorithm called adaptive vibrational configuration interaction (A-VCI) intended for the resolution of the vibrational Schrödinger equation was developed. The main advantage of this approach is to efficiently reduce the dimension of the active space generated into the configuration interaction (CI) process. Here, we assume that the Hamiltonian writes as a sum of products of operators. This adaptive algorithm was developed with the use of three correlated conditions i.e. a suitable starting space ; a criterion for convergence, and a procedure to expand the approximate space. The velocity of the algorithm was increased with the use of a posteriori error estimator (residue) to select the most relevant direction to increase the space. Two examples have been selected for benchmark. In the case of H_2CO, we mainly study the performance of A-VCI algorithm: comparison with the variation-perturbation method, choice of the initial space, residual contributions. For CH_3CN, we compare the A-VCI results with a computed reference spectrum using the same potential energy surface and for an active space reduced by about 90 %. This work was published in [9].

7.5.1.2. Dislocations

We have focused on the improvements in collision detection in the Optidis Code. Junction formation mechanisms are essential to characterize material behavior such as strain hardening and irradiation effects. Dislocations junctions appear when dislocation segments collide with each other, therefore, reliable collision detection algorithms must be used to detect an handle junction formations. Collision detection is also a very costly operation in dislocation dynamics simulations, and performance must be carefully optimized to allow massive simulations.

During the first year of this PhD thesis, new collision algorithms have been implemented for the Dislocation Dynamics code OptiDis. The aim was to allow fast and accurate collision detection between dislocation segments using hierarchical methods. The complexity to solve the N-body collision problem can be reduced to O(N) using spatial partitioning; computation can be accelerated using fast-reject techniques, and OpenMP parallelism. Finally, new collision handling algorithms for dislocations have been implemented to increase the reliability of the simulation.

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

7.5.2.1. Interior penalty discontinuous Galerkin method for coupled elasto-acoustic media

We introduce a high order interior penalty discontinuous Galerkin scheme for the nu- merical solution of wave propagation in coupled elasto-acoustic media. A displacement formulation is used, which allows for the solution of the acoustic and elastic wave equations within the same framework. Weakly imposing the correct transmission condition is achieved by the derivation of adapted numerical fluxes. This generalization does not weaken the discontinuous Galerkin method, thus hp-non-conforming meshes are supported. Interior penalty discontinuous Galerkin methods were originally developed for scalar equations. Therefore, we propose an optimized formulation for vectorial equations more suited than the straightforward standard transposition. We prove consis- tency and stability of the proposed schemes. To study the numerical accuracy and convergence, we achieve a classic plane wave analysis. Finally, we show the relevance of our method on numerical experiments.

More details on this work can be found in [47].

7.5.2.2. High performance simulation for ITER tokamak

Concerning the `GYSELA` global non-linear electrostatic code, the efforts during the period have concentrated on 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 had been carried on in the framework of Fabien Rozar's PhD in collaboration with `CEA-IRFM` (defended in November 2015) and is continued in the PhD of Nicolas Bouzat funded by IPL `C2S@ExA`. The scientific objectives of this new work will be first to consolidate the parallel version of the gyro-average operator, in particular by designing a scalable MPI+OpenMP parallel version and 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.

7.5.2.3. 3D aerodynamics for unsteady problems with bodies in relative motion

The first part of our research work concerning the parallel aerodynamic code `FLUSEPA` has been to design an operational MPI+OpenMP version based on a domain decomposition. We achieved an efficient parallel version up to 400 cores and the temporal adaptive method used without bodies in relative motion has been tested successfully for complex 3D take-off blast wave computations. Moreover, an asynchronous strategy for computing bodies in relative motion and mesh intersections has been developed and has been used for 3D stage separation cases. This first version is the current industrial production version of `FLUSEPA` for Airbus Safran Launchers.

However, this intermediate version shows synchronization problems for the aerodynamic solver due to the time integration used. To tackle this issue, a task-based version over the runtime system `StarPU` has been developed and evaluated. Task generation functions have been designed in order to maximize asynchronism during execution while respecting the data pattern access of the code. This led to the re-factorization of the `FLUSEPA` computation kernels. It's clearly a successful proof of concept as a task-based version is now available for the aerodynamic solver and for both shared and distributed memory. It uses three parallelism levels : MPI processes between sub-domains, `StarPU` workers in shared memory (for each sub-domain) themselves running OpenMP parallel tasks. This version has been validated for large 3D take-off blast wave computations (80 millions of cells) and is much more efficient than the previous MPI+OpenMP version: we achieve a gain in computation time equal to 70 % for 320 cores and to 50 % for 560 cores. The next step will consist in extending the task-based version to the motion and intersection operations. This work has been carried on in the framework of Jean-Marie Couteyen's PhD (defended in September 2016) in collaboration with Airbus Safran Launchers ([2], [17]).

<p style="text-align:center; color:red;">**LFANT Project-Team**</p>

6. New Results

6.1. Class invariants in genus 2

Abelian surfaces, or equivalently, Jacobian varieties of genus 2 hyperelliptic curves, offer the same security as elliptic curves in a cryptographic setting and often better efficiency, and could thus be an attractive alternative. The theory of complex multiplication can be used to obtain cryptographically secure curves. Relying on Shimura reciprocity for Siegel modular forms, we have developed the necessary mathematical theory in [24]. It requires deeper algebraic reasoning than for elliptic curves: Ideals of the endomorphism rings of the abelian varieties are no more two-dimensional modules over the integers, but two-dimensional projective modules over quadratic number rings. We succeed in proving results adapted from the elliptic curve case by suitably normalising quadratic forms over number rings and using strong approximation. The result is an elegant theory that leads to clearly formulated and practical algorithms, which we illustrate by examples.

6.2. Elliptic curve and Abelian varieties cryptology

Participant: Damien Robert.

The paper [15] in which David Lubicz and Damien Robert explain how to improve the arithmetic of Abelian and Kummer varieties has been published in the journal Finite Fields and Their Applications. The speed of the arithmetic is a crucial factor in the performance of cryptosystems based on abelian varieties. Depending on the cryptographic application, the speed record holders are elliptic curves (in the Edwards model) or the Kummer surface of an hyperelliptic curves of genus 2 (in the level 2 theta model). One drawback of the Kummer surface is that only scalar multiplications are available, which may be a problem in certain cryptographic protocols. The previous known models to work on the Jacobian rather than the Kummer surface (Mumford coordinates or the theta model of level 4) are too slow and not competitive with elliptic curves. This paper explains how to use geometric properties (like projective normality) to speed up the arithmetic. In particular it introduces a novel addition algorithm on Kummer varieties (compatible addition), and uses it to speed up multi-exponentiations in Kummer varieties and to obtain new models of abelian surfaces in which the scalar multiplication is as fast as on the Kummer surface.

Theta functions, and in particular the Dedekind eta function, are at the heart of complex multiplication constructions of curves. They can be written as sparse power series with coefficients ± 1. In [23] we devise optimised addition sequences for the occurring exponents, with a proof relying on classical number theory, which help us gain a factor of 2 compared to the standard approach and which is validated in practice by our two independent implementations. Using an approach from computer algebra and a proof relying on analytic number theory, we obtain another factor of 2.

6.3. Symbolic computation

The article [27], of which F. Johansson is a coauthor, was published. The article describes SymPy, an open source computer algebra system written in pure Python. It is built with a focus on extensibility and ease of use, through both interactive and programmatic applications. These characteristics have led SymPy to become a popular symbolic library for the scientific Python ecosystem. This paper presents the architecture of SymPy, a description of its features, and a discussion of select submodules. The supplementary material provide additional examples and further outline details of the architecture and features of SymPy.

Hypergeometric functions are among the most important mathematical functions, with a wide range of applications in everything from physics to number theory. The practical computation of such functions is a challenging problem. The preprint [26]. presents an efficient implementation of hypergeometric functions in arbitrary-precision interval arithmetic. The functions $_0F_1$, $_1F_1$, $_2F_1$ and $_2F_0$ (or the Kummer U-function) are supported for unrestricted complex parameters and argument, and by extension, we cover exponential and trigonometric integrals, error functions, Fresnel integrals, incomplete gamma and beta functions, Bessel functions, Airy functions, Legendre functions, Jacobi polynomials, complete elliptic integrals, and other special functions. The output can be used directly for interval computations or to generate provably correct floating-point approximations in any format. Performance is competitive with earlier arbitrary-precision software, and sometimes orders of magnitude faster. We also partially cover the generalized hypergeometric function $_pF_q$ and computation of high-order parameter derivatives.

The preprint [25] is the corresponding paper for the software Arb developed by F. Johansson. Arb is a C library for arbitrary-precision interval arithmetic using the midpoint-radius representation, also known as ball arithmetic. It supports real and complex numbers, polynomials, power series, matrices, and evaluation of many special functions. The core number types are designed for versatility and speed in a range of scenarios, allowing performance that is competitive with non-interval arbitrary-precision types such as MPFR and MPC floating-point numbers. This paper discusses the low-level number representation, strategies for precision and error bounds, and the implementation of efficient polynomial arithmetic with interval coefficients.

6.4. Logarithmic Class Groups

Logarithmic class groups and units, introduced by Jaulent in 1994, are an intriguing ℓ-adic variation on the classical class and unit groups related to Iwasawa theory and the wild kernels of algebraic K-theory. These $\mathbb{Z}_\ell$-modules of finite type provide direct access to invariants studied in standard conjectures about $\mathbb{Z}_\ell$-extensions. In [12] we devised a new algorithm to explicitly compute them in subexponential time under standard conjectures (GRH and Gross-Kuz'min) and to validate unconditionaly the computed results (now in exponential time). The algorithm has been implemented in the PARI/GP system.

6.5. Class groups and other invariants of number fields

The article by H. Cohen and F. Thorne on Dirichlet series associated to quartic fields with given cubic resolvent has been published. This article gives an explicit formula for the Dirichlet series $\sum_K |\Delta(K)|^{-s}$, where the sum is over isomorphism classes of all quartic fields whose resolvent field is isomorphic to a fixed cubic field k.

The article [22] by H. Cohen and F. Thorne generalizes the work of A. Morra and the authors, on giving explicit formulas for the Dirichlet series generating function of D_ℓ extensions of odd prime degree ℓ with given quadratic resolvent. Over the course of the proof, the authors explain connections between their formulas and the Ankeny-Artin-Chowla conjecture, the Ohno-Nakagawa relation for binary cubic forms, and other topics.

In her thesis, Iuliana Ciocanea-Teodorescu describes algorithms that answer questions arising in ring and module theory. The first main result of this thesis concerns the module isomorphism problem, how to compute a set of generators of minimal cardinality, and how to construct projective covers and injective hulls. The thesis also describe tests for module simplicity, projectivity, and injectivity, and constructive tests for existence of surjective module homomorphisms between two finite modules, one of which is projective. As a negative result, the problem of testing for existence of injective module homomorphisms between two finite modules, one of which is projective, is NP-complete. The last part of the thesis is concerned with finding a good working approximation of the Jacobson radical of a finite ring, that is, a two-sided nilpotent ideal such that the corresponding quotient ring is almost semisimple. The notion used to approximate semisimplicity is that of separability.

In her thesis [11], Pinar Kiliçer determines all CM curves of genus 2 defined over the reflex field. This extends the previous CM class number one problem for elliptic curves which asked to find all elliptic curves defined over the rationals with non-trivial endomorphism ring.

6.6. Number and function fields

The article [13] written by J. Brau and J. Nathan on "Elliptic curves with 2-torsion contained in the 3-torsion field" has been published. This article study the modular curve $X'(6)$ of level 6 defined over $\mathbb{Q}$ whose $\mathbb{Q}$-rational points correspond to j-invariants of elliptic curves E over $\mathbb{Q}$ for which $\mathbb{Q}(E[2])$ is a subfield of $\mathbb{Q}(E[3])$. The authors characterize the j-invariants of elliptic curves with this property by exhibiting an explicit model of $X'(6)$. $X'(6)(\mathbb{Q})$ then gives an infinite family of examples of elliptic curves with non-abelian "entanglement fields," which is relevant to the systematic study of correction factors of various conjectural constants for elliptic curves over $\mathbb{Q}$.

<p style="text-align:center;color:red;">**MAGIQUE-3D Project-Team**</p>

6. New Results

6.1. Seismic Imaging and Inverse Problems

6.1.1. *Time-harmonic inverse problem*

Participants: Hélène Barucq, Florian Faucher.

We study the seismic inverse problem for acoustic and elastic medium associated with the time-harmonic wave equation, and the underlying recovery of geophysical parameters. We employ Full Waveform Inversion (FWI) where the multi parameters reconstruction is based on iterative minimization techniques. This inverse problem shows a Lipschitz stability where the stability constant is related to the (conditional) lower bound of the Fréchet derivative, when assuming a piecewise constant representation of the parameters. We successively estimate the stability constant for different model partition in order to control the convergence of the scheme. Hence we define a multi-level (multi-scale, multi-frequency) algorithm where the natural progression of frequency is paired with the model partition. The method is implemented and numerical experiments are performed for elastic medium reconstruction, in particular for realistic geophysical situations.

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

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

We investigate the inversion of a series of parameters in the context of a 2D elasto-acoustic scattering problem. The inverse problem is solved by using a Newton-like method, where the shape of the scatterer is assumed to be Lipschitz-continuous. Herein, we want to recover the shape and the material parameters in the case of isotropic and anisotropic materials. Based on the different influences of these parameters on the far field pattern, the final goal is to propose an iterative algorithm to retrieve the parameters separately, by devoting some iterations to the reconstruction of the shape and the others to the determination of the parameters. On the other hand, due to the difficulties to retrieve the material parameters, the penetrability of scatters have been studied. The conclusion has been that the recovery of material parameters can be feasible, provided that the scattered waves are not completely reflected. The results of this work have been presented to the conference Inverse Problems for PDE in Bremen, Germany [24].

6.2. Mathematical modeling of multi-physics involving wave equations

6.2.1. *A study of the numerical robustness of single-layer method with Fourier basis for multiple obstacle scattering in homogeneous media*

Participants: Hélène Barucq, Juliette Chabassier, Ha Pham, Sébastien Tordeux.

We investigate efficient methods to solve direct and inverse problems for the propagation of acoustic wave in strongly inhomogeneous media in low-frequency regime. We start our investigation with inhomogeneities created by compactly-supported and non-overlapping obstacles. With a large number of small obstacles, optimized softwares based on Finite Element Method (FEM) lose their robustness. As an alternative, we work with an integral equation method, which uses single-layer potentials and truncation of Fourier series to describe the scattered field. We limit our numerical experiments to disc-shaped obstacles. We first compare our method with Montjoie (a FEM-based software); secondly, we investigate the efficiency of different solver types (direct and iterative) in solving the dense linear system generated by the method. We observe that the optimal choice depends on the distance between obstacles, their size and number, and applications.

6.2.2. *Derivation and validation of impedance transmission conditions for the electric potential across a highly conductive casing*

Participants: Hélène Barucq, Aralar Erdozain, David Pardo, Victor Péron.

Borehole resistivity measurements are a common procedure when trying to obtain a better characterization of the Earth's subsurface. The possible risk of having borehole collapses makes the employment of a casing very suitable for this type of scenarios. Such casing protects the borehole but it also highly complicates the resistivity measurements due to the thinness of the casing and the large contrasts between the conductivities of the casing and the rock formations.

This work is motivated by realistic configurations where the resistivity of the casing is proportional to the cube of the thickness of the casing. In this framework, our aim is to derive Impedance Transmission Conditions (ITCs) for the electromagnetic field across such a casing. As a first approach we derive ITCs for the electric potential. We consider a transmission problem for the static case of the electric potential, set in an axi-symmetric borehole shaped domain. This domain is composed of three different subdomains, the interior part of the borehole, the rock formations and the metallic casing.

In this framework, we address the issue of ITCs using two different approaches. The first one consists in deriving ITCs across the casing itself, whereas the second approach tackles the problem by deriving ITCs on an artificial interface located in the middle of the casing. We derive different models for the two considered approaches and we numerically assess them with a finite element method implementation. Then we perform a comparison on these models by showing the advantages and drawbacks of each model. Finally, we show an application to a borehole through-casing resistivity measurement scenario. This work delivers stability results and error estimates, leading to convergence of each approximate model. All the details regarding this work can be found in [43].and [10]. In addition it has been presented to the WONAPDE Conference [29].

6.2.3. *Semi-analytical solutions for asymptotic models for the electric potential across a highly conductive casing*

Participants: Hélène Barucq, Aralar Erdozain, Ignacio Muga, Victor Péron.

This work is performed in the framework of borehole through-casing resistivity measurements. A transmission problem for the electric potential is considered, where one part of the domain is a high-conductive casing. Numerical instabilities are created during the numerical simulations when such a casing is present in the configuration. Therefore, three different asymptotic models derived in [43] are considered, which are composed of impedance conditions specially designed to avoid the casing. These models correspond to approximations of orders one, two and four.

In this work, we employ analytical methods for the aforementioned asymptotic models, which provide a consistent solution to test and verify the numerical solutions (Finite Element Method). In addition, these methods are computationally cheaper than the purely numerical methods. The standard method we follow consists in employing cylindrical coordinates and assuming material homogeneity in the vertical and angular variables. The source term is represented as a Dirac distribution. Under these conditions, we represent the solution to our problem as an inverse Fourier integral in the vertical variable, and a Fourier series in the angular variable.

Numerical tests are carried out to compare with Finite Element solutions. Several difficulties have to be taken into account during the implementation of the semi-analytical solutions, like the treatment of the Dirac distribution and the presence of singularities when the Fourier variable tends to zero. These difficulties are also addressed in this work which is detailed in [10].

6.2.4. *Numerical investigation of instabilities of Perfectly Matched Layers coupled with DG-schemes in elastodynamic*

Participants: Hélène Barucq, Lionel Boillot, Henri Calandra, Julien Diaz, Simon Ettouati.

We observed long-term numerical instabilities when DG-schemes are coupled with PML in elastodynamic, even with isotropic media. To investigate the causes of this instabilities, we have led a series of numerical experiments with elasticus 5.1 . The conclusion was that the instabilities only appear in truly elastodynamic media (i.e. when the velocities of S waves is positive) and that different factors impact the stability : the heterogeneities of the domain, the choice of the fluxes, the boundary conditions, the use of unstructured meshes... In the best scenario, using a cartesian grid with periodic boundary conditions for an homogeneous medium and centered fluxes, we did not observe instabilities. However, changing only one element of the configuration made the instabilities appear. Our conclusion is that we need a very particular flux in the PML that should be able to handle the heterogeneities of the domain and the structure of the mesh. This flux should also be adapted to discretize the boundary condition. We are now working on the design of this flux.

6.2.5. *Elasto-acoustic coupling*

Participants: Hélène Barucq, Lionel Boillot, Henri Calandra, Julien Diaz, Simon Ettouati.

Last year, we developed a Discontinuous Galerkin Method for the elastoacoustic coupling in time domain. The proposed solution methodology in general and can be applied to any kind of fluxes. The method had been implemented in Elasticus 5.1 and we have transfered it into the Total plateform TMBM-DG 5.4 .

In [23], we have considered elastoacoustic coupling with curved interfaces and we have proposed a solution methodology based on Finite Element techniques, which allows for a flexible coupling between the fluid and the solid domain by using non-conforming meshes and curved elements. Differently from other non-conforming approaches proposed so far, our technique is relatively simpler and requires only a geometrical adjustment at the coupling interface at a preprocessing stage, so that no extra computations are necessary during the time evolution of the simulation. This work, has been achieved in collaboration with Angel Rodriguez Rozas, former post-doc of the team.

6.2.6. *Atmospheric radiation boundary conditions for helioseismology*

Participants: Hélène Barucq, Juliette Chabassier, Marc Duruflé.

Modeling acoustic wave propagation inside a celestial body (as the Sun) prompts the question of imposing an adequate boundary condition. Classical atmosphere models suppose an exponential decay of the medium density and a constant wave celerity outside a given radius. This work proposes several radiation boundary conditions that mimic the presence of such an atmosphere and assesses their behavior numerically in radial and axisymmetric configurations.

6.2.7. *Hybrid discontinuous finite element approximation for the elasto-acoustics.*

Participants: Hélène Barucq, Henri Calandra, Julien Diaz, Elvira Shishenina.

Discontinuous finite element methods proved their accuracy and flexibility, but they are still criticized for the number of degrees of freedom which they use: it is much higher than the ones of the conventional methods based on continuous approximations.

Thus hybrid methods have been developed and their integration into the DIP is under way, both in the acoustic and elastic domains.

The global purpose of this work is to develop a new approach for solving wave equation in discontinuous function spaces. This will provide all propagators already developed in the CARBON platform. Possible directions in this research are for example the development of a Trefftz type approximation for elasto-acoustics, coupling with VEM, HDG.

Our current work is concentrated on using Trefftz method. The main idea of the method is that chosen basis functions of Trefftz approximation space are discrete local solutions of the initial equations to be solved.

The possible advantages of Trefftz type approximations compared to the standard ones are: 1) better orders of convergence; 2) flexibility in the choice of basis functions; 3) low dispersion; 4) incorporation of wave propagation directions in the discrete space; 5) adaptivity and local space-time mesh refinement.

The particularity of Trefftz methods is that in case of applying to time-dependent problems they require a space-time mesh.

We studied theory of application of the method to the coupled acoustic system, and implemented numerically Trefftz method to solve the first-order 1D acoustic wave propagation system. The obtained results were presented during annual workshop in Houston organized by Depth Imaging Partnership between Inria and Total.

6.3. Supercomputing for Helmholtz problems

6.3.1. *Extend task-based node parallelism to cluster level: applications to geophysics*

Participants: Emmanuel Agullo, Lionel Boillot, George Bosilca, Henri Calandra, Corentin Rossignon.

The context of this work is to replace static parallelism based on MPI + threads and/or CUDA by dynamic task-based parallelism on top of runtime systems. On a previous work, we demonstrated the speed-up of the new solution when applied to geophysics, at a node level. Moreover, this task paradigm proved its flexibility on several architectures such as ccNUMA big nodes or many-core Intel Xeon Phi co-processors.

We extended this principle to a set of nodes, eventually heterogeneous, in order to measure performance at a cluster level. Preliminary results on few homogeneous nodes were encouraging, ie still faster than pure MPI. Unfortunately, the geophysics algorithm being too repetitive, the load-balancing issue which can be removed within a node (i.e. between cores) comes back between nodes when they are numerous or few but heterogeneous. This is due to the work-stealing feature of the task paradigm which is by default enabled at the node level only.

To overcome this problem, we extended the work-stealing feature to cluster level. To do that we used the task identification by geometrical sub-meshes to detect candidates that can be exchanged between nodes. Then, we compared PAPI counters on these tasks to find the best choice. Finally, we use a separate task-based program to automatically do the main code task update. Preliminary results show clear improvement of load-balancing at cluster level.

This work has been presented to the conferences Rice Oil&Gas[34] and SIAM-PP (Parallel Processing) [35].

6.3.2. *Numerical libraries for hybrid meshes in a discontinuous Galerkin context*

Participants: Hélène Barucq, Lionel Boillot, Aurelien Citrain, Julien Diaz.

Elasticus team code 5.1 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). We are finalizing now the unstructured-goal.

The purpose is to use the h-adaptivity of discontinuous Galerkin method to easily encircle unstructured tetrehedra with hexahedra to form hybrid meshes (so in 3D). In addition, it would be interesting to couple numerical methods depending on the element types.

6.3.3. *Code transfer: TMBM-DG/THBM into Total R&D environment*

Participants: Lionel Boillot, Julien Diaz.

The goal of the DIP collaboration between Total and Inria is to transfer the validated research codes. At first, DIVA-DG has been created in conjunction with Total developers team. It concerns the time modeling of wave propagation. Then, we forked it into Elasticus code to focus on mathematical research at the Inria side. Finally, once validated, we managed its transfer into the recent Total R&D environment (so instead of DIVA template, we moved to TMBM template) to form the TMBM-DG 5.4 code. The entire code has been transferred now, including unit tests and full documentation.

In the meantime, another code emerged within the DIP collaboration, THBM, concerning the frequency modeling of wave propagation. The development is directly done since the beginning in the Total R&D environment. An important part is already validated while research still continues.

6.3.4. *Hybridizable Discontinuous Galerkin methods for solving the elastic Helmholtz equations*

Participants: Marie Bonnasse-Gahot, Henri Calandra, Julien Diaz, Stéphane Lanteri.

The advantage of performing seismic imaging in frequency domain is that it is not necessary to store the solution at each time step of the forward simulation. Unfortunately, the drawback of the Helmholtz equations, when considering 3D realistic elastic cases, lies in solving large linear systems. This represents today a challenging task even with the use of High Performance Computing (HPC). To reduce the size of the global linear system, we developed a Hybridizable Discontinuous Galerkin method (HDGm). It consists in expressing the unknowns of the initial problem in function of the trace of the numerical solution on each face of the mesh cells. In this way the size of the matrix to be inverted only depends on the number of degrees of freedom on each face and on the number of the faces of the mesh, instead of the number of degrees of freedom on each cell and on the number of the cells of the mesh as we have for the classical Discontinuous Galerkin methods (DGm). The solution to the initial problem is then recovered thanks to independent elementwise calculation. This results were presented in a submitted paper.

Moreover, as the HDG global matrix is very sparse, we focus on a suitable solver for this kind of matrix. We tested two linear solvers: a parallel sparse direct solver MUMPS (MUltifrontal Massively Parallel sparse direct Solver) and a hybrid solver MaPHyS (Massively Parallel Hybrid Solver) which combines direct and iterative methods. We compared the performances of the two solvers when solving 3D elastic waves propagation over HDGm. These comparisons were presented at the MATHIAS 2016 conference and at the DIP Workshop [36], [37]

6.3.5. *A Symmetric Trefftz-DG Formulation based on a Local Boundary Element Method for the Solution of the Helmholtz Equation.*

Participants: Hélène Barucq, Abderrahmane Bendali, M'Barek Fares, Vanessa Mattesi, Sébastien Tordeux.

A general symmetric Trefftz Discontinuous Galerkin method is built in [12] for solving the Helmholtz equation with piecewise constant coefficients. The construction of the cor- responding local solutions to the Helmholtz equation is based on a boundary element method. A series of numerical experiments displays an excellent stability of the method relatively to the penalty parameters, and more importantly its outstanding ability to reduce the instabilities known as the "pollution effect" in the literature on numerical simulations of long-range wave propagation.

6.4. Hybrid time discretizations of high-order

6.4.1. *High order time discretization for dissipative wave equations.*

Participants: Juliette Chabassier, Julien Diaz, Anh-Tuan Ha, Sébastien Imperiale.

Magique-3D team is interested in numerical methods for wave propagation in realistic media, which are naturally dissipative in many application cases. In this internship, we wish to investigate several dissipation models, that lead to Partial Differential Equations with different structures. The simplest model is the scalar wave equation with homogeneous and constant damping $\frac{\partial^2 u}{\partial t^2} + R\frac{\partial u}{\partial t} - \Delta u = f$. In order to approach the complexity of the propagating medium and its geometry, high order finite elements in space are used. Once the spatial discretization is fixed, we get a differential equation of the kind $\frac{d^2 u_h}{dt^2} + B_h\frac{du_h}{dt} + A_h u_h = f_h$, where the mass matrix is the identity thanks to the mass lumping technique followed by a renormalization, B_h is the dissipation matrix and A_h the stiffness matrix. Classically, this equation is discretized in time with a centered and second order finite difference scheme known as the θ-scheme ($\theta > 0$)

$$\frac{u_h^{n+1} - 2u_h^n - u_h^{n-1}}{\Delta t^2} + B_h\frac{u_h^{n+1} - u_h^{n-1}}{2\Delta t} + A_h\left(\theta u_h^{n+1} + (1 - 2\theta)u_h^n + \theta u_h^{n-1}\right) = f_h^n \qquad (1)$$

In order to preserve the precision obtained with high order finite elements in space, we wish to design higher order time discretizations, while preserving some interesting mathematical properties as the dissipation of a discrete energy, and an efficiency close to the one observed for the second order scheme. More precisely, if $\theta = 0$ and B_h is diagonal, scheme (1) only requires the inversion of a diagonal matrix at each time step.

We want to use the technique of the modified equation, which consists in compensating the first term of the consistency error of a low order discretization, by adding a well chosen new term. If $\theta = 0$, this approach leads to the following fourth order accurate in time scheme

$$\left(I_h + \frac{\Delta t^2}{12} B_h\right) \frac{u_h^{n+1} - 2u_h^n - u_h^{n-1}}{\Delta t^2} + \left[B_h + \frac{\Delta t^2}{12} \left(B_h A_h - A_h B_h\right)\right] \frac{u_h^{n+1} - u_h^{n-1}}{2\Delta t} + A_h u_h^n - \frac{\Delta t^2}{12} A_h^2 u_h^n = \widetilde{f}_h^n$$

Even if B_h is diagonal, A_h and B_h do not commute in general. We propose to replace the matrix $B_h A_h - A_h B_h$, potentially hard to invert, by an approximated matrix, easy to invert, without deteriorating the consistency of the scheme.

An article is being written and will be submitted soon.

6.4.2. *High order conservative explicit and implicit schemes for wave equations.*

Participants: Juliette Chabassier, Sébastien Imperiale.

In 2016 we have studied the space/time convergence of a family of high order conservative explicit and implicit schemes for wave equations. An original proof of convergence has been proposed and provides an understanding of the lack of convergence of some schemes when the time step approaches its greatest admissible value for stability (CFL condition). An article has been submitted.

6.4.3. *Efficient high order implicit time schemes for Maxwell's equations.*

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

The Padé approximant is well known to be one of the best approximation of an exponential function which is involved in the exact solution of the linear ODE (Ordinary Differential Equations):

$$y'(t) = Ay(t) + F(t)$$

where A is a given matrix (usually coming from finite element discretization) and F is a term source. The numerical solution can be constructed by approximating the exponential function using the diagonal Padé approximant:

$$R(z) = \frac{P_m(z)}{Q_m(z)}$$

The function $R(z)$ is a fraction involving two polynomial P_m and Q_m of same degree and approximating the exponential. The corresponding scheme is implicit and A-stable in the sense of Dahlquist. The associated stability function is the same as the stability function of the Gauss-Runge-Kutta schemes. However, Gauss-Runge-Kutta schemes can be used to handle non-linear ODEs, but they are too expensive to use in practice. The diagonal Padé schemes presented here can be seen as a simplification of Gauss-Runge-Kutta schemes in the case of linear ODE. We have proposed an efficient way to implement the diagonal Padé schemes with an accurate approximation of the source term to keep the correct order of accuracy.

The main drawback of Padé schemes is that the denominator $Q_m(z)$ has distinct roots. It implies that we have to solve distinct linear systems at each time step. As a result, we have also studied the case where the denominator has an unique real root γ :

$$R(z) = \frac{N(z)}{(1 - \gamma z)^m}$$

The numerator N is then found to obtain the "best" approximation of the exponential under the constraint of the A-stability property of the underlying schemes. The obtained schemes have been called Linear Singly Diagonal Implicit Runge-Kutta schemes (Linear SDIRK) since they share the same property as SDIRK (a unique linear system to solve several times) but they can be applied only to linear ODEs. We provide a performance assessment of different implicit schemes (Padé schemes, SDIRK and Linear SDIRK). The comparison criteria are based on the amplitude and phase errors which are reliable gauges of accuracy when approximating waves problems. The Linear SDIRK schemes and the diagonal Padé schemes have been implemented in the code Montjoie. We have performed numerical experiments in 1-D and 2-D for Maxwell's equations to validate these schemes and compare their efficiency.

This work has been presented at the conference ICOSAHOM [28], the colloquium Inter'Actions en Mathématiques Lyon 2016 and the Mathias annual Total seminar [27].

6.4.4. *Optimized high-order explicit Runge-Kutta-Nyström schemes.*
Participants: Marc Duruflé, Mamadou N'Diaye.

In this work we propose a high order time integration explicit scheme to solve a second order derivative non-linear ordinary differential equation (ODE)

$$y'' = f(t, y)$$

To solve this family of ODEs, explicit one-step Runge-Kutta-Nyström have been proposed by Hairer et al. The stability condition (CFL) associated with these schemes have been studied for order 3, 4 and 5 by Chawla and Sharma. In this work, we have extended the stability studies for high order. We proposed optimal coefficients for Runge-Kutta-Nyström schemes of order 6, 7, 8 and 10 which have been obtained by optimizing the CFL. With the obtained optimal CFL, these schemes are well suited for stiff problems where the stability condition is restrictive. These schemes have been implemented in the code Montjoie.

Numerical experiments have been conducted in 1-D for the non-linear Maxwell's equation and show that obtained Runge-Kutta-Nyström schemes of order 7 is quite efficient. This work has been presented at the conference ICOSAHOM [48].

<p style="text-align:center; color:red">**MANAO Project-Team**</p>

7. New Results

7.1. Analysis and Simulation

7.1.1. Principles of Light Field Imaging

Light field imaging offers powerful new capabilities through sophisticated digital processing techniques that are tightly merged with unconventional optical designs. This combination of imaging technology and computation necessitates a fundamentally different view of the optical properties of imaging systems and poses new challenges for the traditional signal and image processing domains. We aimed to provide a comprehensive review [14] of the considerations involved and the difficulties encountered in working with light field data during 25 years of research.

7.1.2. Physically-Based Reflectance Model Combining Reflection and Diffraction

Reflectance properties express how objects in a virtual scene interact with light; they control the appearance of the object: whether it looks shiny or not, whether it has a metallic or plastic appearance. Having a good reflectance model is essential for the production of photo-realistic pictures. Measured reflectance functions provide high realism at the expense of memory cost. Parametric models are compact, but finding the right parameters to approximate measured reflectance can be difficult. Most parametric models use a model of the surface micro-geometry to predict the reflectance at the macroscopic level. We have shown [26] that this micro-geometry causes two different physical phenomena: reflection and diffraction. Their relative importance is connected to the surface roughness. Taking both phenomena into account, we develop a new reflectance model that is compact, based on physical properties and provides a good approximation of measured reflectance.

7.1.3. Multi-Scale and Structured SV-BRDF Model for Scratched Materials

We developed a Spatially-Varying BRDF model tailored to the multi-scale rendering of scratched materials such as metals, plastics or finished woods. Our approach takes advantage of the regular structure of scratch distributions to achieve high performance without compromising visual quality (fig. 8). The BRDF for a single scratch is simulated using an optimized 2D ray-tracer and compactly stored in a three-component 2D texture. In contrast to existing models, our approach takes into account all inter-reflections inside a scratch, including Fresnel effects. At render time, the SV-BRDF for the scratch distribution under a pixel or ray footprint is obtained by linear combination of individual scratch BRDFs. Our model can be evaluated using both importance and light sampling, in direct and global illumination settings. Our approach provides users with controls over the profile, micro-BRDF, density and orientation of scratches. All these material properties are updated at interactive rates. This work has been published at ACM Siggraph 2016 [16] and one our result has been selected as the cover of the ACM Siggraph 2016 proceedings. It is part of the PhD Thesis "Control of anisotropic materials appearance" [11] defended this year.

7.1.4. Cues for Perception of Appearance

Thanks for the FP7 ITN PRISM, we have participated to several user studies to understand the perception of an object appearance. First, for fluids and other deformable materials, we find [25] that observers show a high degree of constancy in matching the viscosity across the different variations. However, volume differences between test and match stimulus, especially with static stimuli, caused large effects of over- and under-estimation of viscosity. We also find that a number of cues related to curvatures, periodic movements of the liquids, and the way they spread out predict aspects of the observer's performance, but that humans achieve better constancy than the cues predict.
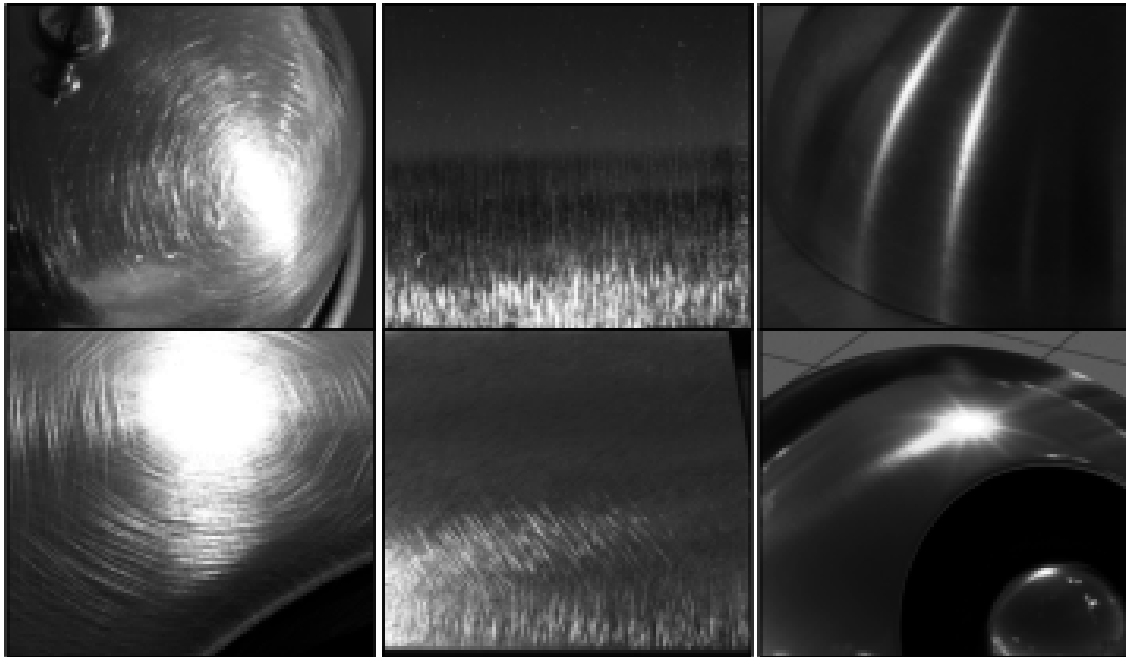
Figure 8. Our scratch BRDF (top) can reproduce several effects similar to real photographs (bottom).

We have also investigated gloss haze [22]. The results reveals that haziness is a distinct visual dimension orthogonal to the commonly studied glossiness and blurriness. Coatedness appears to be nearly synonymous with haziness, as this is one of the main physical causes of haze in real world materials. Polish seems to be a combination of glossiness and haziness, as materials go from dull to hazy to highly glossy during the physical polishing process. The inferred tactile quality of friction is apparently uncorrelated with haziness. Our results demonstrate that haze is indeed a distinct perceptual dimension of gloss, which is systematically related to the kurtosis of the specular lobe.

7.2. From Acquisition to Display

7.2.1. *Spatial Augmented Reality*

Spatial augmented reality allows to improve or modify the perception of the reality with virtual information displayed directly in the real world, using video-projection. Many fields such as tourism, entertainment, education, medicine, industry or cultural heritage may benefit from it. Recent computer science techniques allow to measure, analyse and visualise the geometry of the surface of real objects, as for instance archeological artefacts. We have proposed a SAR interaction and visualisation technique (part of the PhD thesis "Interaction techniques, personalized experience and surface reconstruction for spatial augmented reality" [12] defended this year) that combines the advantages of the study of both real and 3D archeological artefacts. Thus, we superimpose on the object an expressive rendering based on curvatures with SAR, allowing for example to show details of engravings. Next, we simulate the use of a flashlight with the help of a 6-degree-of-freedom controller. The user can then specify the area on the object to be augmented and adjust the various necessary parameters of the expressive rendering. One of the main caracteristics of SAR is to enable multiple users to simultaneously participate to the same experience. However, depending on the target application, this can be seen as a drawback.

We have also proposed a new display device [27] that allows to create experiences in SAR that are both multiuser and personalised by taking into account the user point of view. In order to do so, the projection display, set in front of the object to augment, is made from a material that is both retro-reflective and semi-transparent. We suggest two different uses of this new device, as well as two scenarios of application.

7.2.2. Isotropic BRDF Measurements

Image-based BRDF measurements on spherical material samples present a great opportunity to shorten significantly the acquisition time with respect to more traditional, non-multiplexed measurement methods for isotropic BRDFs. However, it has never been analyzed deeply, what measurement accuracy can be achieved in such a setup; what are the main contributing uncertainty factors and how do they relate to calibration procedures. We have developed [20] a new set of isotropic BRDF measurements with their radiometric and geometric uncertainties acquired within such an imaging setup. We have analyzed the most prominent optical phenomena that affect measurement accuracy and pave the way for more thorough uncertainty analysis in forthcoming image-based BRDF measurements. Our newly acquired data with their quantified uncertainties will be helpful for comparing the quality and accuracy of the different experimental setups and for designing other such image-based BRDF measurement devices.

7.3. Rendering, Visualization and Illustration

7.3.1. Cache-friendly Sampling

Monte-Carlo integration techniques for global illumination are popular on GPUs thanks to their massive parallel architecture, but efficient implementation remains challenging. The use of randomly de-correlated low-discrepancy sequences in the path-tracing algorithm allows faster visual convergence. However, the parallel tracing of incoherent rays often results in poor memory cache utilization, reducing the ray bandwidth efficiency. Interleaved sampling [65] partially solves this problem, by using a small set of distributions split in coherent ray-tracing passes, but the solution is prone to structured noise. On the other hand, ray-reordering methods [83] group stochastic rays into coherent ray packets but their implementation add an additional sorting cost on the GPU [74], [50]. We have introduced [19] a micro-jittering technique for faster multi-dimensional Monte-Carlo integration in ray-based rendering engines. Our method, improves ray coherency between GPU threads using a slightly altered low-discrepancy sequence rather than using ray-reordering methods. Compatible with any low-discrepancy sequence and independent of the importance sampling strategy, our method achieves comparable visual quality with classic de-correlation methods, like Cranley-Patterson rotation [66], while reducing rendering times in all scenarios.

7.3.2. Multi-Resolution Meshes for Feature-Aware Hardware Tessellation

Hardware tessellation is de facto the preferred mechanism to adaptively control mesh resolution with maximal performances. However, owing to its fixed and uniform pattern, leveraging tessellation for feature-aware LOD rendering remains a challenging problem. In [15], we relax this fundamental constraint by introducing a new spatial and temporal blending mechanism of tessellation levels, which is built on top of a novel hierarchical representation of multi-resolution meshes. This mechanism allows to finely control topological changes so that vertices can be removed or added at the most appropriate location to preserve geometric features in a continuous and artifact-free manner (cf. Figure 9). We then show how to extend edge-collapse based decimation methods to build feature-aware multi-resolution meshes that match the tessellation patterns. Our approach is fully compatible with current hardware tessellators and only adds a small overhead on memory consumption and tessellation cost. This work as been published at Eurographics 2016 [15].

7.3.3. Shape Depiction for Transparent Objects

Shading techniques are useful to deliver a better understanding of object shapes. When transparent objects are involved, depicting the shape characteristics of each surface is even more relevant. We have developed [21] a method for rendering transparent scenes or objects using classical tools for shape depiction in real time. Our method provides an efficient way to compute screen space curvature on transparent objects by using a novel
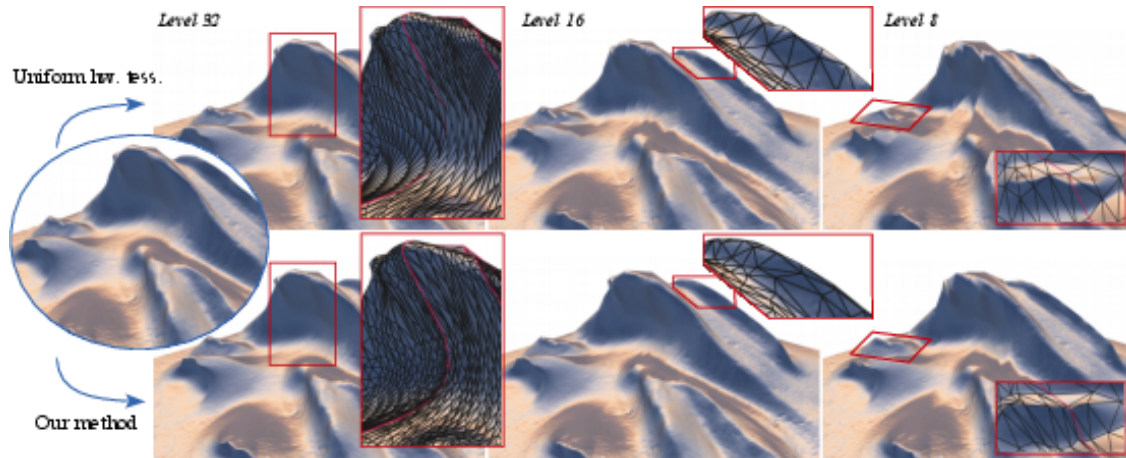
*Figure 9. Uniform hardware tessellation (**top**) fails at representing accurately sharp features and areas of high curvature, such as the top and deep part of the drifts, which produces tessellation artifacts. Our method (**bottom**) better preserves those regions by adapting the triangle size and aligning their edges with those features.*

screen space representation of a scene derived from Order Independent Transparency techniques. Moreover, we propose a customizable stylization that modulates the transparency per fragment, according to its curvature and its depth, which can be adapted for various kinds of applications.

7.4. Editing and Modeling

7.4.1. Flow-guided Warping for Image-based Shape Manipulation

Manipulating object shape in images usually require a-priori on their 3D geometry, and either user interactions or huge databases of 3D objects. In collaboration with the Maverick team (Inria Rhone Alpes), we have developed a method that manipulates perceived object shape from a single input color image without the need of addional 3D information, user input or 3D data. The key idea is to give the illusion of shape sharpening or rounding by exaggerating orientation patterns in the image that are strongly correlated to surface curvature (fig. 10). We build on a growing literature in both human and computer vision showing the importance of orientation patterns in the communication of shape, which we complement with mathematical relationships and a statistical image analysis revealing that structure tensors are indeed strongly correlated to surface shape features. We then rely on these correlations to introduce a flow-guided image warping algorithm, which in effect exaggerates orientation patterns involved in shape perception. We evaluate our technique by 1) comparing it to ground truth shape deformations, and 2) performing two perceptual experiments to assess its effects. Our algorithm produces convincing shape manipulation results on synthetic images and photographs, for various materials and lighting environments. This work has been published at ACM Siggraph 2016 [17].

7.4.2. Local Shape Editing at the Compositing Stage

Modern compositing software permit to linearly recombine different 3D rendered outputs (e.g., diffuse and reflection shading) in post-process, providing for simple but interactive appearance manipulations. Renderers also routinely provide auxiliary buffers (e.g., normals, positions) that may be used to add local light sources or depth-of-field effects at the compositing stage. These methods are attractive both in product design and movie production, as they allow designers and technical directors to test different ideas without having to re-render an entire 3D scene. In this work, we extended this approach to the editing of local shape: users modify the rendered normal buffer, and our system automatically modifies diffuse and reflection buffers to

(a) Input image - ©Expertissim   (b) Shape sharpening   (c) Shape rounding

Figure 10. Our warping technique takes as input (a) a single image (Jules Bennes, after Barye: "walking lion") and modifies its perceived surface shape, either making it sharper in (b) or rounder in (c).

provide a plausible result. Our method is based on the reconstruction of a pair of diffuse and reflection prefiltered environment maps for each distinct object/material appearing in the image. We seamlessly combine the reconstructed buffers in a recompositing pipeline that works in real-time on the GPU using arbitrarily modified normals. This work as been published at the Eurographics Symposium on Rendering [24].

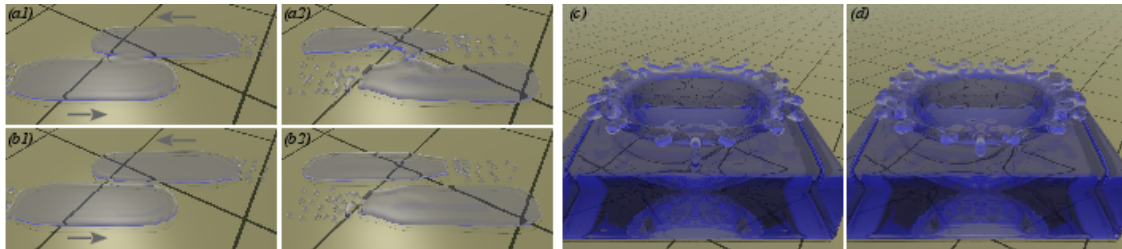### 7.4.3. Topology-Aware Neighborhoods for Point-Based Simulation and Reconstruction

Figure 11. Two SPH fluid simulations using a standard Euclidean particle neighborhood (a,c), and our new topological neighborhood (b,d). On the left, two fluid components are crossing while moving in opposite directions. Our new neighborhood performs accurate merging computations and avoids both unwanted fusion in the reconstruction and incorrect fluid interaction in the simulation. On the right, our accurate neighborhoods lead to different shape of the splash, and enable the reconstruction of the fluid with an adequate topology while avoiding bulging at distance.

Particle based simulations are widely used in computer graphics. In this field, several recent results have improved the simulation itself or improved the tension of the final fluid surface. In current particle based implementations, the particle neighborhood is computed by considering the Euclidean distance between fluid particles only. Thus particles from different fluid components interact, which generates both local incorrect behavior in the simulation and blending artifacts in the reconstructed fluid surface. In collaboration with IRIT, we developed a better neighborhood computation for both the physical simulation and surface reconstruction steps (fig. 11). Our approach tracks and stores the local fluid topology around each particle

using a graph structure. In this graph, only particles within the same local fluid component are neighbors and other disconnected fluid particles are inserted only if they come into contact. The graph connectivity also takes into account the asymmetric behavior of particles when they merge and split, and the fluid surface is reconstructed accordingly, thus avoiding their blending at distance before a merge. In the simulation, this neighborhood information is exploited for better controlling the fluid density and the force interactions at the vicinity of its boundaries. For instance, it prevents the introduction of collision events when two distinct fluid components are crossing without contact, and it avoids fluid interactions through thin waterproof walls. This leads to an overall more consistent fluid simulation and reconstruction. This work as been published at the Eurographics/ ACM SIGGRAPH Symposium on Computer Animation [18].

<p style="text-align: center; color: red">MEMPHIS Project-Team</p>

7. New Results

7.1. Hybrid POD/DNS: application to aeroelastic windturbine blade

Some new techniques related to Reduced Order Modelling have been developed in the framework of the EU project AEROGUST. The first proposed approach is based on a domain decomposition method in which a POD [56], [43], [44], [37], [36] model is dynamically coupled with a CFD solver [39], [38]. This tool can be used to perform predictive simulations thanks to the fact that the non-linear effects related to new working conditions are directly captured by the CFD solver while the far field region can be efficiently described by the POD model. The hybrid technique has been extended to gust simulations by the introduction of forcing terms which can describe perturbations coming from the far field.

The domain decomposition approach has been proposed also inside an iterative procedure named "numerical zoom" which is based on the use of several mesh levels. This procedure is repeated several times in order to focus the degrees of freedom of the discretisation in the region close to the body.

Finally, the POD method has been proposed also for the acceleration of CFD solver for incompressible flows. The solution of the Poisson problem on the pressure variable which appears in incompressible solvers can be quite time consuming. The proposed approach consists in searching the solution of the Poisson problem in the space spanned by the POD basis. This is done by substituting the POD expansion in the Poisson equation and minimizing the residuals. The robustness of the method has been improved by introducing a check on the quality of the Poisson solution (based on the divergence of the velocity field at the end of the correction step) and a dynamic update of the POD basis.

The domain decomposition approach with the forcing terms has been used to simulate the effects of a gust on a wind turbine blade in a simplified configuration at low Reynolds number. The numerical zoom procedure is applied by coupling a DNS simulation with a POD description of the far field. The solution obtained on three levels of mesh is reported in Figure 9 in which the vortex structures are shown according to the q-criterion. The bending of the blade is described by a non-linear beam model. Figure 10 shows the shapes of the blade without loads, in the chosen working condition and during the gust.
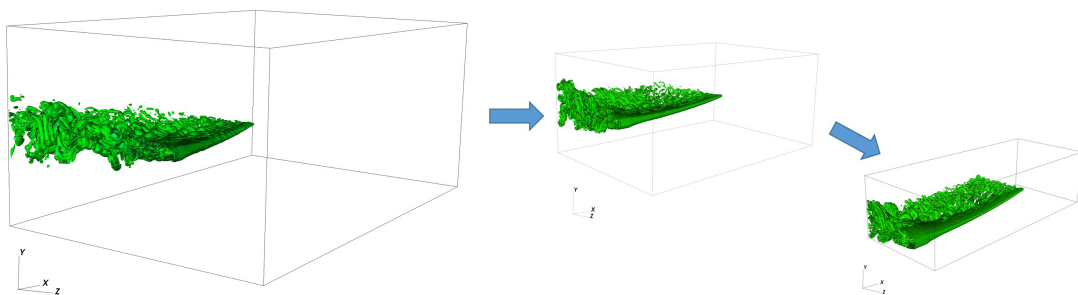

Figure 9. Numerical zoom on the wind turbine blade simulation (vortex structure visualised by q-criterion)

Figure 10. Blade shape without loads (top), in the chosen working condition (middle) and during the gust (bottom)

7.2. Discretization of the Laplacian operator using a multitude of overlapping cartesian grids

A new finite-difference approach to solve a Laplacian operator has been developed, using patches of overlapping grids where a fined level is needed, leaving coarser meshes in the rest of the computational domain. These overlapping grids will have generic quadrilateral shapes (as shown in figure 11).
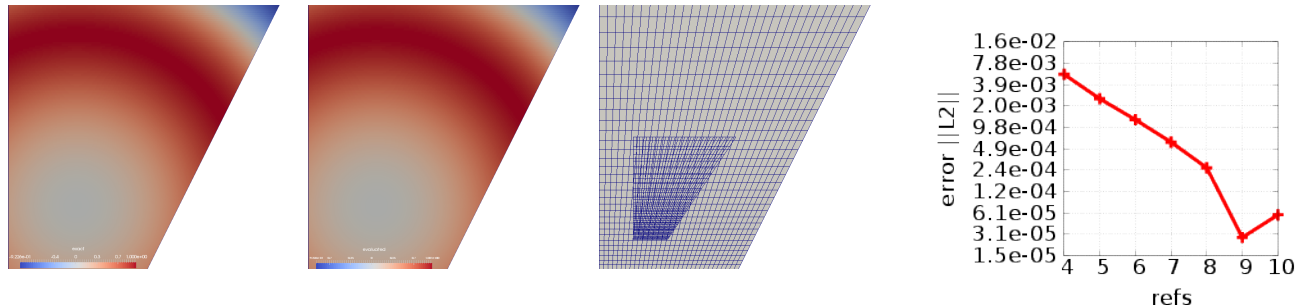
(a) Exact: (b) Evaluated   (c) Numerical grid   (d) Numerical error
$sin((x - 0.5)^2 + (y - 0.5)^2)$

Figure 11. results

A monolithic approach is used to solve the algebraic equations, applying restriction and prolongation operators to fill the non-diagonal blocks of the resulting matrix. These operators works on data structures communicated between the different grids using ad hoc parallel inter-communicators, as shown in figure 12 . 12 .

Works are going on to change the solver from the finite-difference approach to a finite-volume one, and to implement the 3D case. The use of a finite-volume solver can benefit from the usage of octree patches instead of cartesian grids, obtaining a more accurate refining and a greater precision.

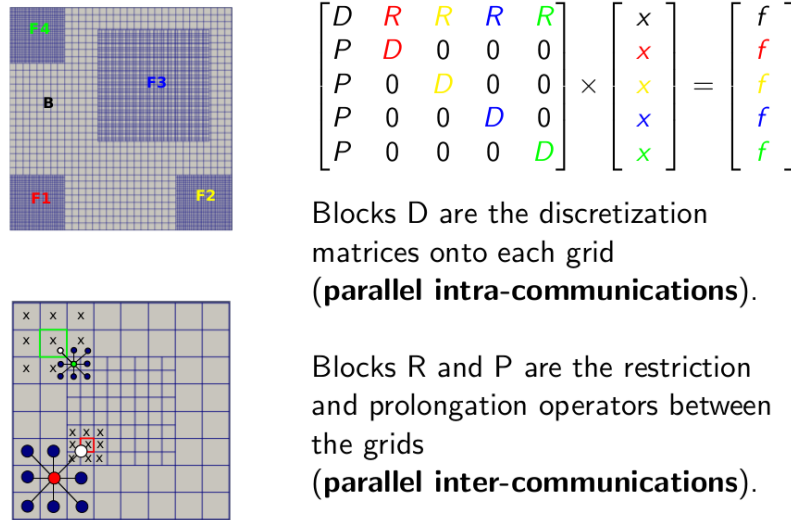## 7.3. Numerical simulation of a biomimetic LVAD developed by CorWave

$$
\begin{bmatrix}
D & R & R & R & R \\
P & D & 0 & 0 & 0 \\
P & 0 & D & 0 & 0 \\
P & 0 & 0 & D & 0 \\
P & 0 & 0 & 0 & D
\end{bmatrix}
\times
\begin{bmatrix}
x \\ x \\ x \\ x \\ x
\end{bmatrix}
=
\begin{bmatrix}
f \\ f \\ f \\ f \\ f
\end{bmatrix}
$$

Blocks D are the discretization matrices onto each grid (**parallel intra-communications**).

Blocks R and P are the restriction and prolongation operators between the grids (**parallel inter-communications**).

Figure 12. Monolithic approach

We just started a collaboration with the PME CorWave. The CorWave LVAD utilizes an undulating disc wave pumping mechanism, replacing the high speed, high shear impeller of current continuous flow rotary pumps. Louis de Lillers, the CorWave project manager, has contacted MEMPHIS to perform numerical simulations and optimizations of their LVAD. This collaboration has started with an industrial PhD (Cifre, Antoine Fondaneche). Figure 13 shows preliminary results (proof of concept) for the CorWave LVAD obtained with the code NaSCar described in [38].
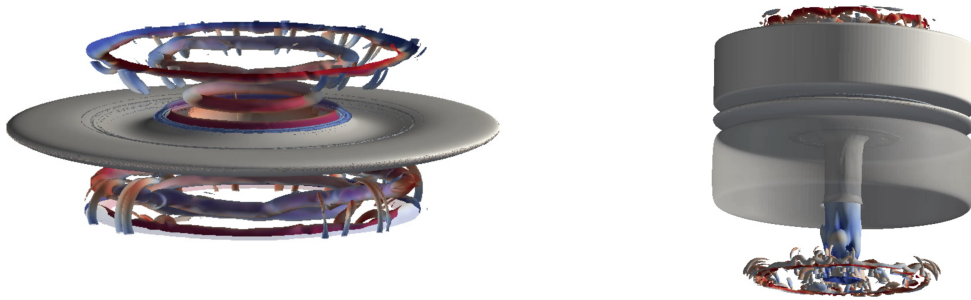

Figure 13. Preliminary results (proof of concept) for the CorWave LVAD. Left: oscillating membrane, right: the whole pump system.

7.4. A sharp Cartesian method for incompressible flows with large density ratios

We have developed and validated a new Cartesian method for bifluid incompressible flows with high density ratios. The specificity of the method relies on a sharp second order numerical scheme for the spatial resolution of the discontinuous elliptic problem for the pressure, that was developed in [42]. The Navier-Stokes equations are integrated in time thanks to a fractional step method based on the Chorin scheme and discretized in space on a Cartesian mesh. The bifluid interface is implicitly represented using a level set function. The numerical tests show the improvements due to this sharp method compared to classical first order methods. As an illustration, we present here numerical results for the dam break test case.

This test case is studied in [53] and [41], and based on experiments conducted in [49]. The initial configuration is a water column at rest in air. The initial height and width of the column are both 5.715 cm. The domain size is 40 cm×10 cm. The value of the physical parameters are

$$
\begin{cases}
\rho_{water} = 1000 \ kg/m^3, \\
\mu_{water} = 1.137 \times 10^{-3} \ kg/ms, \\
\rho_{air} = 1.226 \ kg/m^3, \\
\mu_{air} = 1.78 \times 10^{-5} \ kg/ms, \\
\sigma = 0.0728 \ kg/s^2 \\
g = -9.8m/s^2
\end{cases}
\tag{2}
$$

We present in Figure 14 the interface evolution at non-dimensional times $T = t\sqrt{g/h} = 0, 1, 2, 3, 4$, with h the initial height of the water column. The computations are performed with 256×64 points.
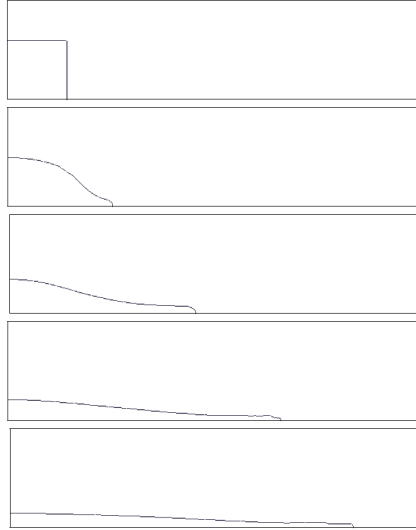

Figure 14. Evolution of the interface for the dam break problem at non-dimensional times
$T = t\sqrt{g/h} = 0, 1, 2, 3, 4.$

In Figure 15 , we plot the evolution in time of the water front, compared to the experimental results of [49], and to the results obtained for the Ghost-Fluid method and the conservative method of Raessi and Pitsch [53]. We observe that the front propagation is in agreement with the experimental results and the results of the conservative method presented in [53]. It means that, though the method is not strictly conservative, the numerical errors due to momentum transfer across the interface are not large enough to slow down the

propagation of the front. It is not the case for instance for the Ghost-Fluid method, as it can be noticed in Figure 15 and has been reported in [53].
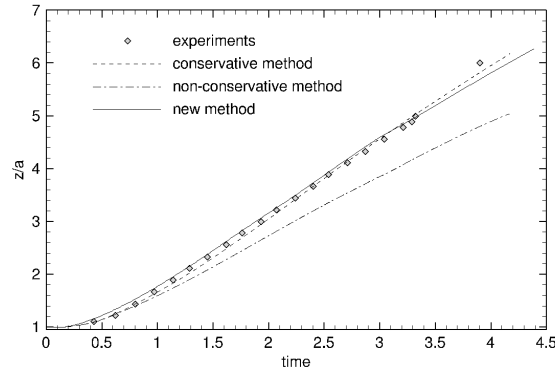

Figure 15. Evolution of the front of propagation: comparison between experimental date and several numerical methods: the Ghost Fluid method (non-conservative method), the conservative method of Raessi and Pitsch and our new method, The dimensionless location of the front $\frac{z}{a}$ is plotted as a function of the dimensionless time $t\sqrt{q}h$.

7.5. Platooning of trucks on highways

In the context of energy saving, the platooning of ground vehicles on top of a road, in particular highways has been studied. The numerical simulations are performed in 2D and 3D for up to 10 billions unknowns on 384 cores. The goal is to have trucks autonomously following their leader to form a road train in order to improve traffic flow efficiency and to reduce oil consumption. Thus the distance between trucks is short. For instance a gain of about 40% can be obtained on the drag coefficient of the followers when the distance between trucks is equal to 1.8125 their height (see the figure 16), that is approximately eight meters. Even the leader has a lower drag coefficient (-10%) as the first follower compresses the flow in its wake. In the figure it is clearly shown that the pressure gradients inside the gap between the vehicles are much lower than in front of the leader.
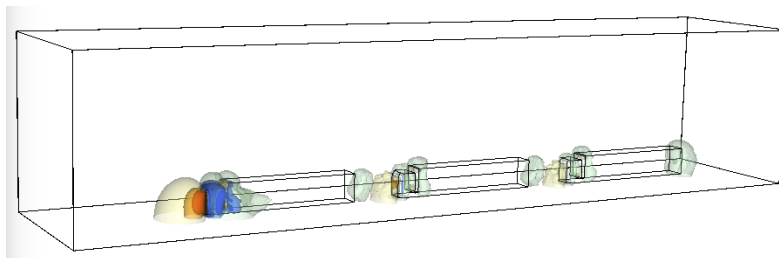

Figure 16. Mean pressure contours of the mean flow around three simplified European tractor-trailer geometries with a distance equal to 1.8125 their height on top of a road in three dimensions.

7.6. Non-linear elasto-plastic dynamics of compressible materials

We describe a numerical model to simulate the non-linear elasto-plastic dynamics of compressible materials. The model is fully Eulerian and it is discretized on a fixed Cartesian mesh. The hyperelastic constitutive law considered is neohookean and the plasticity model is based on a multiplicative decomposition of the inverse deformation tensor. The model is thermodynamically consistent and it is shown to be stable in the sense that the norm of the deviatoric stress tensor beyond yield is non increasing. The multimaterial integration scheme is based on a simple numerical flux function that keeps the interfaces sharp. Numerical illustrations in one to three space dimensions of high-speed multimaterial impacts in air are presented.

In TC4 an iron sphere is impacting an aluminium plate immersed in air. The computational domain is $[-0.3, 0.7] \times [-0.4, 0.4] \times [-0.4, 0.4]$m. The initial velocity of the projectile is $1000 m.s^{-1}$. The computation is performed on a $500 \times 400 \times 400$ mesh with 216 processors. Homogeneous Neumann conditions are imposed on the left and right borders and cantilever on the others.

The results are given in Fig 17 where we present the Schlieren results on the vertical symmetry plane and the material interfaces. As in the 2D case the projectile perforates the aluminium plate which is strongly stretched. The breaking of the plate at final time is due to the level set function resolution.
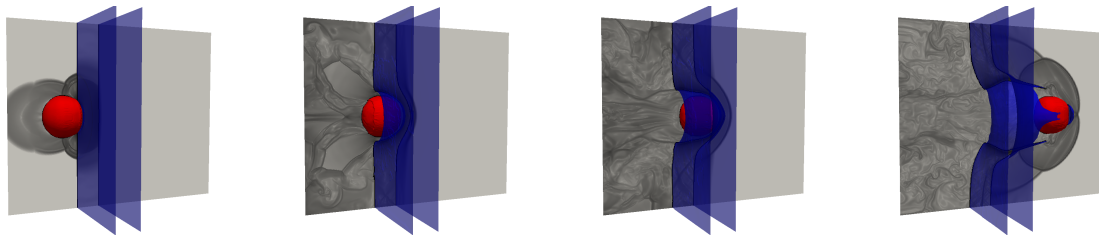

Figure 17. Schlieren representation on the vertical symmetry plane and the material interfaces.

7.7. Hierarchical grids: applications with quadtrees/octrees

A first application with a specific method is the resolution of the incompressible Navier-Stokes equations. A Navier-Stokes solver dealing with quadtrees has been implemented in parallel this year. The overall aim will be to model in 3D the flow over a wind turbine using an Octree grid. On the figure 18 can be seen an example of QuadTree mesh. A Finite Volume Semi-Lagrangian scheme is used.

First, the order of convergence of the Laplacian Solver discretization on QuadTrees has been computed and compared with those obtained with other schemes as explained previously. The method for solving the Laplacian Solver is named Diamond method and consists in using a dual mesh and considering that the Gradients are constants inside. The order of convergence of 2 has been obtained. The order of convergence of the overall Navier-Stokes resolution on QuadTree meshes has been computed and the order of 2 is get as can be seen on figures 18 . For the L_∞ norm, the order of 2 can't be reach caused by the loss of accuracy when a gap in refinement level occurs.

For now the grid is fixed, so the next step will be to refine and coarsen the grid following the position of the obstacle and the "interesting" areas. The work will then go on with the implementation of this Navier-Stokes solver with adaptive QuadTree meshes in 3D.

A second application is phase changing material. We consider problems governed by linear elliptic equations with discontinuity interfaces across the domain. The equation coefficients, the solution and its normal
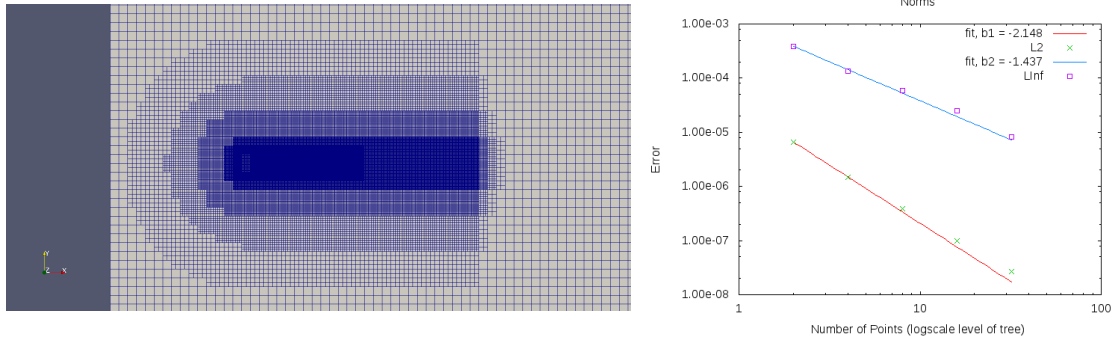
(a) Example of QuadTree mesh for circular cylinder   (b) Norms as a function of number of points on grid

Figure 18.

derivative can undergo a jump across these internal boundaries. We present a compact second-order finite-difference scheme on a tree-based adaptive grid that can be efficiently solved in parallel. The main idea is to optimize the truncation error of the discretization as a function of the local grid configuration.

The variable coefficient heat diffusion problem we consider is modeled by:

$$-\text{div}(\kappa(\overrightarrow{x})\nabla u(\overrightarrow{x})) = g(\overrightarrow{x}), \quad \text{in } D, \tag{3}$$

$$R(\kappa\partial_{\overrightarrow{n}}u(\overrightarrow{x}))_S = [u], \text{ on } \gamma \tag{4}$$

$$[\kappa(\overrightarrow{x})\partial_{\overrightarrow{n}}u(\overrightarrow{x})] = 0, \quad \text{on } \gamma \tag{5}$$

Where $\overrightarrow{x} = (x, y, z)$ are the spatial coordinates and $\kappa(\overrightarrow{x})$ is piecewise continuous on each subdomain but it may be discontinuous across γ (the boundary of the D subdomain that contains the discontinuities through). A cell-centered investigation often leads to a symmetric linear system, since the relation between two neighbors is reflective. Considering the configuration in Fig. 19 it is natural to define the discretization at c_4 in terms of the others.
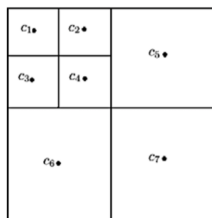

Figure 19. A test configuration centerd in c_4.

Let h be the side length of the cell c_4. To obtain the existence of a linear consistent scheme we must be able to find the coefficients a_i such that:

$$u_{xx} + u_{yy} = a_1 u_1 + a_2 u_2 + a_3 u_3 + a_4 u_4 + a_5 u_5 + a_6 u_6 + a_7 u_7 + O(h)$$

A complete Taylor's analysis on all the involved neighbors, applying them relative linear combinations of the expansions, implies that the coefficients a_i must satisfy the following linear system:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & -h & 0 & -h & \frac{3h}{2} & -\frac{h}{2} & \frac{3h}{2} \\ 0 & h & h & 0 & \frac{h}{2} & -\frac{3h}{2} & -\frac{3h}{2} \\ 0 & \frac{h^2}{2} & 0 & \frac{h^2}{2} & \frac{9h^2}{8} & \frac{h^2}{8} & \frac{9h^2}{8} \\ 0 & -h^2 & 0 & 0 & \frac{3h^2}{4} & \frac{3h^2}{4} & -\frac{9h^2}{4} \\ 0 & \frac{h^2}{2} & \frac{h^2}{2} & 0 & \frac{h^2}{8} & \frac{9h^2}{8} & \frac{9h^2}{8} \end{pmatrix} \begin{pmatrix} a_4 \\ a_1 \\ a_2 \\ a_3 \\ a_5 \\ a_6 \\ a_7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

In the example above there are seven concerned points, so, we can determine infinite solutions of the complete system but we search a unique one. Let M be the constraints matrix, $\overrightarrow{a}$ the weights vector, $\overrightarrow{f}$ the right hand side vector for consistency and $F(\overrightarrow{a})$ a weights function. The problem to minimize has the Lagrangian form:

$$\mathcal{L} = F(\overrightarrow{a}) - \overrightarrow{\lambda}(M\overrightarrow{a} - \overrightarrow{f}) \tag{6}$$

We write the minimization problem (5) in matrix form like:

$$Ax = b \Leftrightarrow \begin{cases} \frac{\partial F(\overrightarrow{a})}{\partial \overrightarrow{a}} - M^T \overrightarrow{\lambda} & = 0 \\ M\overrightarrow{a} & = \overrightarrow{f} \end{cases}$$

This choice allows us to implement a scheme always consistent case by case.

Following a consistence proof several tests have been produced to strenghten our method like penalization and different kinds of model until a modeling of the problem (7.7) step by step. We built a new cell centered finite difference method able to:

- be consistent and locally convergent to second order on balanced grids;

- simplify and promote an AMR approach along discontinuity;

- solve the coupled problems (2)-(3) and (2)-(4);

- a first consistent result on the complete model (2)-(3)-(4).

The finite different method presented here has been compared with two other methods finite volume scheme (Fig. 20). The first one has been implemented by Marco Cisternino and the diamonds one by Claire Taymans (see previous sections). All the three methods stick on the same grid and they use PABLO's data structure with its parallel balance.
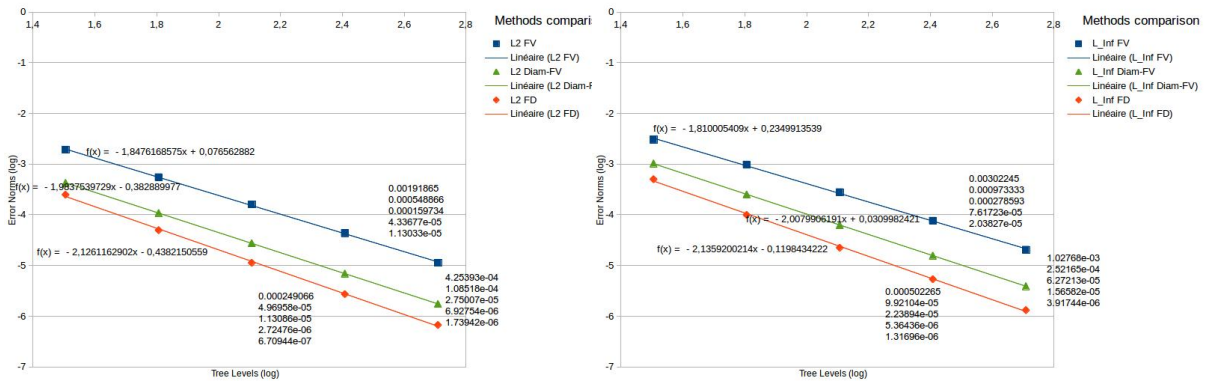
(a) L_2 norm comparison   (b) L_∞ norm comparison

Figure 20. Comparison between several numerical approaches.

<p align="center">**MNEMOSYNE Project-Team**</p>

7. New Results

7.1. Overview

This year we first explored two main loops of cerebral architecture, the limbic and motor loops, 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 motor loop (*cf.* § 7.3) considers the evolution of sensorimotor learning, from goal-directed behaviors to habitual behaviors.

We also began this year to study some characteristics of the systemic integration of our models (*cf.* § 7.4), raising the question of the conditions of autonomous learning and certain global characteristics such as neuromodulation.

Finally, we study the links between our bio-inspired modeling work and Machine Learning (*cf.* § 7.5), revisiting this latter domain in the light of the principles highlighted by our models.

7.2. The limbic loop

We explored the limbic loop by describing a series of neural mechanisms that propose how responding conditioning results from interactions between the amygdala, the nucleus accumbens and the limbic pole of the frontal cortex. In our models [1], this learning is also fed by exchanges with the hippocampus (episodic memory) and the sensory cortex (semantic memory) and we studied the major role of acetylcholine in these exchanges. This also allowed us to address the difficult question of the articulation between the respondant and operant conditioning in particular in the nucleus accumbens. We proposed an original mechanism whereby noradrenaline could modulate the balance between exploration and exploitation [12] based on an assessment of the level of uncertainty and its impact on performance [13].

Also in connection with this loop, we studied the dynamics of dopamine release in the midbrain, considered to play an essential role in the coding of the prediction error. This model developed in the framework of our collaboration with India (*cf.* § 9.3) proposes to introduce into the classical circuit, new actors (such as the pedunculopontine tegmental nucleus in the brainstem) and new functions (dissociation of amplitude and timing of the reward), that we will seek to corroborate in the future.

Lastly, we carried out a thorough study about the behavior of our model of associative memory in the hippocampus [23], and particularly about its resistance to interference.

7.3. The motor loop

The nervous system structures involved in decision making constitute a circuit formed by the basal ganglia, the cortex, the thalamus and their numerous interconnections. This circuit can be described as a set of loops operating in parallel and interacting at different points. The decisions and therefore the actions of an individual emerge from the interactions between these loops and the plasticity of their connections. These emerging behaviors and arising learning processes are addressed through a closed-loop approach in which the theoretical model is in constant interaction with the environment of the task. To this end, neural modeling and dedicated analysis software tools were developed in the laboratory, at the level of the neuronal circuit. We have explored this year the dynamics of information flows within this circuit through a computational model described at the neuron and synapse level. Taking into account previous experimental observations from primates and earlier computational models, we incrementally developed a network capable of learning to perform behavioral tasks under several protocols and conditions [5]. The development of this computational model was conducted in parallel with the development of an experimental model of decision making in the salamander (Pleurodeles waltlii) [2]. The result here is a computational model of learning and decision making in the basal ganglia that allows for the testing of experimental hypotheses and also to conduct in silico pathophysiological or pharmacological investigations at the cellular level.

7.4. Systemic integration

We have worked this year on the integration of goal-oriented and habitual behaviors, two modes of learning associated to the motor loop. There is an apparent contradiction between experimental data showing that the basal ganglia are involved in goal-oriented and routine behaviors and clinical observations. Lesion or disruption by deep brain stimulation of the globus pallidus interna has been used for various therapeutic purposes ranging from the improvement of dystonia to the treatment of Tourette's syndrome. None of these approaches has reported any severe impairment in goal-oriented or automatic movement. To solve this conundrum, we trained two monkeys to perform a variant of a two-armed bandit-task (with different reward contingencies). Bilateral inactivation of the globus pallidus interna, by injection of muscimol, prevents animals from learning new contingencies while performance remains intact, although slower for the familiar stimuli. We replicate in silico these data by adding lateral competition and Hebbian learning in the cortical layer of the theoretical model of the cortex–basal ganglia loop that provided the framework of our experimental approach [7]. These results suggest that a behavioral decision results from both the cooperation (acquisition) and competition (expression) of two distinct but entangled memory systems, the goal-directed system and the habitual system that may represent the two ends of the same graded phenomenon.

We began our first works of systemic integration associating our models developed in the limbic and motor loops, for the study of the taking into account of the uncertainty in the selection of the action [1]. This preliminary work using the VirtualEnaction platform (*cf.* § 6.4) will be continued this year with a PhD that begins.

We have more generally proposed a study [11], analyzing the role of neuromodulation in adaptation to uncertainty, whose potential systemic impact is evident, particularly because it provides precious characteristics for autonomous learning [10].

7.5. Machine Learning

In Machine Learning, we were interested this year in two phenomena for which we consider classical paradigms of modeling and for which we wonder how they could be adapted by bio-inspiration.

The first paradigm concerns the manipulation of temporal sequences. In a perspective of better understanding how brain learn structured sequences we extended a model on syntax acquisition using the Reservoir Computing framework (using random recurrent networks) [16], [9], [19], [20]. The extended model is also used in a Human-Robot Interaction architecture to enable users to use more natural language with robots [14], [15], [18]. This work will be extended with our collaborators at the University of Hamburg (*cf.* § 9.3).

In an industrial application for the representation of electrical diagrams (*cf.* § 8.1), we also study how recurrent layered models can be trained to run through these schemes for prediction and sequence representation tasks.

The second paradigm concerns the extraction of characteristics and the use of hierarchical networks, as in the case of deep networks. An industrial application whose study has just begun (*cf.* § 9.2) will lead us to revisit these models to make them more easily usable in constrained frameworks, for example with limited size corpuses.

<p align="center" style="color:red">MONC Project-Team</p>

7. New Results

7.1. Free boundary problem for cell protrusion formations: theoretical and numerical aspects

Authors: Olivier Gallinato, Masahito Ohta, Clair Poignard, Takashi Suzuki

In this paper, a free boundary problem for cell protrusion formation is studied theoretically and numerically. The cell membrane is precisely described thanks to a level set function, whose motion is due to specific signalling pathways. The aim is to model the chemical interactions between the cell and its environment, in the process of invadopodia or pseudopodia formation. The model consists of Laplace equation with Dirichlet condition inside the cell coupled to Laplace equation with Neumann condition in the outer domain. The actin polymerization is accounted for as the gradient of the inner signal, which drives the motion of the interface. We prove the well-posedness of our free boundary problem under a sign condition on the datum. This criterion ensures the consistency of the model, and provides conditions to focus on for any enrichment of the model. We then propose a new first order Cartesian finite-difference method to solve the problem. We eventually exhibit the main biological features that can be accounted for by the model: the formation of thin and elongated protrusions as for invadopodia, or larger protrusion as for pseudopodia, depending on the source term in the equation. The model provides the theoretical and numerical grounds for single cell migration modeling, whose formulation is valid in 2D and 3D. In particular, specific chemical reactions that occured at the cell membrane could be precisely described in forthcoming works. Journal: Journal of Mathematical Biology, Springer Verlag (Germany), 2016, <10.1007/s00285-016-1080-7> lien hal: https://hal.inria.fr/hal-01412264v1

7.2. Mathematical model for transport of DNA plasmids from the external medium up to the nucleus by electroporation

Authors: Michael Leguèbe, M Notarangelo, Monika Twarogowska, Roberto Natalini, Clair Poignard

This work is devoted to modelling gastrointestinal stromal tumour metastases to the liver, their growth and resistance to therapies. More precisely, resistance to two standard treatments based on tyrosine kinase inhibitors (imatinib and sunitinib) is observed clinically. Using observations from medical images (CT scans), we build a spatial model consisting in a set of non-linear partial differential equations. After calibration of its parameters with clinical data, this model reproduces qualitatively and quantitatively the spatial tumour evolution of one specific patient. Important features of the growth such as the appearance of spatial heterogeneities and the therapeutical failures may be explained by our model. We then investigate numerically the possibility of optimizing the treatment in terms of progression-free survival time and minimum tumour size reachable by varying the dose of the first treatment. We find that according to our model, the progression-free survival time reaches a plateau with respect to this dose. We also demonstrate numerically that the spatial structure of the tumour may provide much more insights on the cancer cell activities than the standard RECIST criteria, which only consists in the measurement of the tumour diameter. Finally, we discuss on the non-predictivity of the model using only CT scans, in the sense that the early behaviour of the lesion is not sufficient to predict the response to the treatment. Journal: Mathematical Medicine and Biology, Oxford University Press (OUP), 2016, <10.1093/imammb/dqw002> lien hal: https://hal.inria.fr/hal-01380292

7.3. Free boundary problem for cell protrusion formations: theoretical and numerical aspects

Authors: Olivier Gallinato, Masahito Ohta, Clair Poignard, Takashi Suzuki

In this paper, a free boundary problem for cell protrusion formation is studied theoretically and numerically. The cell membrane is precisely described thanks to a level set function, whose motion is due to specific signalling pathways. The aim is to model the chemical interactions between the cell and its environment, in the process of invadopodia or pseudopodia formation. The model consists of Laplace equation with Dirichlet condition inside the cell coupled to Laplace equation with Neumann condition in the outer domain. The actin polymerization is accounted for as the gradient of the inner signal, which drives the motion of the interface. We prove the well-posedness of our free boundary problem under a sign condition on the datum. This criterion ensures the consistency of the model, and provides conditions to focus on for any enrichment of the model. We then propose a new first order Cartesian finite-difference method to solve the problem. We eventually exhibit the main biological features that can be accounted for by the model: the formation of thin and elongated protrusions as for invadopodia, or larger protrusion as for pseudopodia, depending on the source term in the equation. The model provides the theoretical and numerical grounds for single cell migration modeling, whose formulation is valid in 2D and 3D. In particular, specific chemical reactions that occured at the cell membrane could be precisely described in forthcoming works. Journal: Journal of Mathematical Biology, Springer Verlag (Germany), 2016, <10.1007/s00285-016-1080-7> lien hal: https://hal.inria.fr/hal-01412264v1

7.4. Spatial modelling of tumour drug resistance: the case of GIST liver metastases Mathematical Medicine and Biology Advance

Authors: Guillaume Lefebvre, François Cornelis, Patricio Cumsille, Thierry Colin, Clair Poignard, Olivier Saut

This work is devoted to modelling gastrointestinal stromal tumour metastases to the liver, their growth and resistance to therapies. More precisely, resistance to two standard treatments based on tyrosine kinase inhibitors (imatinib and sunitinib) is observed clinically. Using observations from medical images (CT scans), we build a spatial model consisting in a set of non-linear partial differential equations. After calibration of its parameters with clinical data, this model reproduces qualitatively and quantitatively the spatial tumour evolution of one specific patient. Important features of the growth such as the appearance of spatial heterogeneities and the therapeutical failures may be explained by our model. We then investigate numerically the possibility of optimizing the treatment in terms of progression-free survival time and minimum tumour size reachable by varying the dose of the first treatment. We find that according to our model, the progression-free survival time reaches a plateau with respect to this dose. We also demonstrate numerically that the spatial structure of the tumour may provide much more insights on the cancer cell activities than the standard RECIST criteria, which only consists in the measurement of the tumour diameter. Finally, we discuss on the non-predictivity of the model using only CT scans, in the sense that the early behaviour of the lesion is not sufficient to predict the response to the treatment. Journal: Mathematical Medicine and Biology, Oxford University Press (OUP), 2016, <10.1093/imammb/dqw002> lien hal: https://hal.inria.fr/hal-01380292

7.5. Mathematical modeling of cancer immunotherapy and synergy with radiotherapy

Team participant: S. Benzekry Other participants: R. Serre, N. André, J. Ciccolini, D. Barbolosi (SMARTc, Inserm, Marseille, FR), L. Padovani, X. Muracciole (Radiotherapy Unit, La Timone Hospital, Marseille, FR), F. Barlési (Multidisciplinary Oncology and Therapeutic Innovations Unit, AP-HM, Marseille, FR) and C. Meille (Roche Pharmaceutics, Basel, Switzerland)

Combining radiotherapy with immune checkpoint blockade may offer considerable therapeutic impact if the immunosuppressive nature of the tumor microenvironment (TME) can be relieved. In this study, we used mathematical models, which can illustrate the potential synergism between immune checkpoint inhibitors and radiotherapy. A discrete-time pharmacodynamic model of the combination of radiotherapy with inhibitors of the PD1–PDL1 axis and/or the CTLA4 pathway is described. This mathematical framework describes how a growing tumor first elicits and then inhibits an antitumor immune response. This antitumor immune response is described by a primary and a secondary (or memory) response. The primary immune response appears first

and is inhibited by the PD1–PDL1 axis, whereas the secondary immune response happens next and is inhibited by the CTLA4 pathway. The effects of irradiation are described by a modified version of the linear-quadratic model. This modeling offers an explanation for the reported biphasic relationship between the size of a tumor and its immunogenicity, as measured by the abscopal effect (an off-target immune response). Furthermore, it explains why discontinuing immunotherapy may result in either tumor recurrence or a durably sustained response. Finally, it describes how synchronizing immunotherapy and radiotherapy can produce synergies. The ability of the model to forecast pharmacodynamic endpoints was validated retrospectively by checking that it could describe data from experimental studies, which investigated the combination of radiotherapy with immune checkpoint inhibitors. In summary, a model such as this could be further used as a simulation tool to facilitate decision making about optimal scheduling of immunotherapy with radiotherapy and perhaps other types of anticancer therapies.

7.6. Non-Standard Radiotherapy Fractionations Delay the Time to Malignant Transformation of Low-Grade Gliomas

Team participant: S. Benzekry. Other participants: A. Henares-Molina, V.M. Perez-Garcia and A. Martinez-Gonzalez (Môlab, Universidad de Castilla-La Mancha, Ciudad Real, Spain) P.C. Lara (Radiation Oncology, Las Palmas University Hospital, Las Palmas, Spain), M. Garcia-Rojo (Pathology department, Jerez de la Frontera Hospital, Jerez de la Frontera, Spain)

Grade II gliomas are slowly growing primary brain tumors that affect mostly young patients. Cytotoxic therapies (radiotherapy and/or chemotherapy) are used initially only for patients having a bad prognosis. These therapies are planned following the "maximum dose in minimum time" principle, i. e. the same schedule used for high-grade brain tumors in spite of their very different behavior. These tumors transform after a variable time into high-grade tumors, what decreases significantly the patient's life expectancy. In this paper we study mathematical models describing the growth of grade II gliomas in response to radiotherapy. We find that protracted metronomic fractionations, *i.e.* therapeutical schedules enlarging the time interval between low-dose radiotherapy fractions, may lead to a better tumor control without an increase in toxicity. Other non-standard fractionations such as protracted or hypoprotracted schemes may also be beneficial. The potential survival improvement depends on the tumor proliferation rate and can be even of the order of years. A conservative metronomic scheme, still being a suboptimal treatment, delays the time to malignant progression of at least one year when compared to the standard scheme.

7.7. Model-driven optimization of antiangiogenics + cytotoxics combination in breast cancer mice treated with bevacizumab and paclitaxel

Team participant: S. Benzekry. Other participants: S. Mollard (CRUK, Cambridge, UK), J. Ciccolini, D-C Imbs, R. El Cheikh, D. Barbolosi (SMARTc, Inserm, Marseille, FR)

Bevacizumab is the first-in-class antiangiogenic drug administrated concomitantly with cytotoxics. Several reports have shown that antiangiogenics could induce a transient phase of vascular normalization, thus ensuring a better drug delivery provided that cytotoxics administration is delayed. However, determining this best sequence is challenging. We have developed a simple mathematical model describing the impact of antiangiogenics on tumor vasculature. A 3.4 days delay between bevacizumab and paclitaxel was first proposed by the model as an optimal sequence. To test its relevance, 84 mice were orthotopically xenografted with human MDA-231Luc+ breast cancer cells. Two different sets of experiments were performed, based upon different bevacizumab dosing (10 or 20 mg/kg) and inter-cycle intervals (7 or 10 days), comprising several combinations with paclitaxel. Results showed that scheduling bevacizumab administration 3 days before paclitaxel improved antitumor efficacy (48% reduction in tumor growth as compared with concomitant dosing, p<0.05) while reducing metastatic spreading. Additionally, bevacizumab alone could lead to more aggressive metastatic disease with shorter survival in animals. Our phenomenological model was able to fit e perietal data a d provided insight o the underlying d a i s of as ulature's a ilit to deliver the cytotoxic agent. Final simulations suggested a new, data-informed optimal sequence of 2.4 days. Our data suggest that concomitant

dosing between bevacizumab and paclitaxel could be a sub-optimal strategy at bedside. In addition, this proof of concept study suggests that mathematical modelling could help to identify the best sequence among a variety of possible alternate treatment modalities, thus refining the way experimental or clinical studies are conducted.

7.8. Dynamics of concomitant resistance: data, theories and mathematical modeling

Team participant: S. Benzekry Other participants: C. Lamont, L. Hlatky, P. Hahnfeldt (Center of Cancer and Systems Biology, Boston, USA)

In mice bearing two tumors implanted simultaneously, tumor growth was suppressed in one of the two tumors. Three theories of this phenomenon were advanced and assessed against the data. As formalized, the two models of competition for nutrients and indirect angiogenesis-regulated inhibition were not able to explain the growth behavior as well as a third model based on direct systemic inhibition. The superior model offers a depiction of concomitant resistance that provides an improved theoretical basis for tumor growth control that may also find utility in therapeutic planning to avoid post-surgery metastatic acceleration.

7.9. Modeling spontaneous metastasis following surgery: an in vivo-in silico approach

Team participant: S. Benzekry. Other participants: A. Tracz, M. Mastri, R. Corbelli and J. Ebos (Roswell Park Cancer Institute, Buffalo, USA) D. Barbolosi (SMARTc, Inserm, Marseille, FR)

Rapid improvements in the detection and tracking of early-stage tumor progression aim to guide decisions regarding cancer treatments as well as predict metastatic recurrence in patients following surgery. Mathematical models may have the potential to further assist in estimating metastatic risk, particularly when paired with in vivo tumor data that faithfully represent all stages of disease progression. Herein we describe mathematical analysis that uses data from mouse models of spontaneous metastasis developing after surgical removal of orthotopically implanted primary tumors. Both presurgical (primary tumor) and postsurgical (metastatic) growth was quantified using bioluminescence and was then used to generate a mathematical formalism based on general laws of the disease (*i.e.* dissemination and growth). The model was able to fit and predict pre-/post-surgical data at the level of the individual as well as the population. Our approach also enabled retrospective analysis of clinical data describing the probability of metastatic relapse as a function of primary tumor size. In these data-based models, inter-individual variability was quantified by a key parameter of intrinsic metastatic potential. Critically, our analysis identified a highly nonlinear relationship between primary tumor size and postsurgical survival, suggesting possible threshold limits for the utility of tumor size as a predictor of metastatic recurrence. These findings represent a novel use of clinically relevant models to assess the impact of surgery on metastatic potential and may guide optimal timing of treatments in neoadjuvant (presurgical) and adjuvant (postsurgical) settings to maximize patient benefit.

7.10. Computational Trials: Unraveling Motility Phenotypes, Progression Patterns, and Treatment Options for Glioblastoma Multiforme

Team participants: Thierry Colin, Olivier Saut. Other participants: Fabio Raman, Elizabeth Scribner, Olivier Saut, Cornelia Wenger, Hassan Fathallah-Shaykh.

Glioblastoma multiforme is a malignant brain tumor with poor prognosis and high morbidity due to its invasiveness. Hypoxia-driven motility and concentration-driven motility are two mechanisms of glioblastoma multiforme invasion in the brain. The use of anti-angiogenic drugs has uncovered new progression patterns of glioblastoma multiforme associated with significant differences in overall survival. Here, we apply a mathematical model of glioblas- toma multiforme growth and invasion in humans and design computational trials using agents that target angiogenesis, tumor replication rates, or motility. The findings link highly- dispersive, moderately-dispersive, and hypoxia-driven tumors to the patterns observed in glioblastoma multiforme

treated by anti-angiogenesis, consisting of progression by Expand- ing FLAIR, Expanding FLAIR + Necrosis, and Expanding Necrosis, respectively. Further- more, replication rate-reducing strategies (e.g. Tumor Treating Fields) appear to be effective in highly-dispersive and moderately-dispersive tumors but not in hypoxia-driven tumors. The latter may respond to motility-reducing agents. In a population computational trial, with all three phenotypes, a correlation was observed between the efficacy of the rate- reducing agent and the prolonga- tion of overall survival times. This research highlights the potential applications of computational trials and supports new hypotheses on glioblastoma multiforme phenotypes and treatment options.

N

PHOENIX Project-Team

7. New Results

7.1. Tablet-Based Activity Schedule in Mainstream Environment for Children with Autism and Children with ID

Including children with autism spectrum disorders (ASD) in mainstream environments creates a need for new interventions whose efficacy must be assessed in situ. This article presents a tablet-based application for activity schedules that has been designed following a participatory design approach involving mainstream teachers, special education teachers, and school aides. This application addresses two domains of activities: classroom routines and verbal communications. We assessed the efficiency of our application with two overlapping user studies in mainstream inclusion, sharing a group of children with ASD. The first experiment involved 10 children with ASD, where five children were equipped with our tabled-based application and five were not equipped. We show that (1) the use of the application is rapidly self-initiated (after 2 months for almost all the participants) and (2) the tablet-supported routines are better performed after 3 months of intervention. The second experiment involved 10 children equipped with our application; it shared the data collected for the five children with ASD and compared them with data collected for five children with intellectual disability (ID). We show that (1) children with ID are not autonomous in the use of the application at the end of the intervention, (2) both groups exhibited the same benefits on classroom routines, and (3) children with ID improve significantly less their performance on verbal communication routines. These results are discussed in relation with our design principles. Importantly, the inclusion of a group with another neurodevelopmental condition provided insights about the applicability of these principles beyond the target population of children with ASD.

7.2. Self Determination-Based Design To Achieve Acceptance of Assisted Living Technologies For Older Adults

Providing technological support to assist older adults in their daily activities is a promising approach to aging in place. However, acceptance is critical when technologies are embedded in the user's life. Recently, Lee et al. established a connection between acceptance and motivation. They approached motivation via the Self-Determination Theory (SDT): the capacity to make choices and to take decisions. This paper leverages SDT to promote a new design style for gerontechnologies that consists of principles and requirements. We applied our approach to develop an assisted living platform, which was used to conduct a six-month field study with 34 older adults. We show that self-determination is a determining factor of technology acceptance. Furthermore, our platform improved the self-determination of equipped participants, compared to the control group, suggesting that our approach is effective. As such, SDT opens up new opportunities for improving the design process of gerontechnologies.

7.3. Frameworks compiled from declarations: a language-independent approach

Programming frameworks are an accepted fixture in the object-oriented world, motivated by the need for code reuse, developer guidance, and restriction. A new trend is emerging where frameworks require domain experts to provide declarations using a domain-specific language (DSL), influencing the structure and behaviour of the resulting application. These mechanisms address concerns such as user privacy. Although many popular open platforms such as Android are based on declaration-driven frameworks, current implementations provide ad hoc and narrow solutions to concerns raised by their openness to non-certified developers. Most widely used frameworks fail to address serious privacy leaks, and provide the user with little insight into application behaviour. To address these shortcomings, we show that declaration-driven frameworks can limit privacy

leaks, as well as guide developers, independently from the underlying programming paradigm. To do so, we identify concepts that underlie declaration-driven frameworks, and apply them systematically to both an object-oriented language, Java, and a dynamic functional language, Racket. The resulting programming framework generators are used to develop a prototype mobile application, illustrating how we mitigate a common class of privacy leaks. Finally, we explore the possible design choices and propose development principles for developing domain-specific language compilers to produce frameworks, applicable across a spectrum of programming paradigms.

7.4. Analysis of How People with Intellectual Disabilities Organize Information Using Computerized Guidance

Access to residential settings for people with intellectual disabilities (ID) contributes to their social participation, but presents particular challenges. Assistive technologies can help people perform activities of daily living. However, the majority of the computerized solutions offered use guidance modes with a fixed, unchanging sequencing that leaves little room for self-determination to emerge. The objective of the project was to develop a flexible guidance mode and to test it with participants, to describe their information organization methods. This research used a descriptive exploratory design and conducted a comparison between five participants with ID and five participants with no ID. The results showed a difference in the information organization methods for both categories of participants. The people with ID used more diversified organization methods (categorical, schematic, action-directed) than the neurotypical participants (visual, action-directed). These organization methods varied depending on the people, but also on the characteristics of the requested task. Furthermore, several people with ID presented difficulties when switching from virtual to real mode. These results demonstrate the importance of developing flexible guidance modes adapted to the users' cognitive strategies, to maximize their benefits. Studies using experimental designs will have to be conducted to determine the impacts of more-flexible guidance modes.

7.5. Leveraging Declarations over the Lifecycle of Large-Scale Sensor Applications

Masses of sensors and actuators are being deployed in our daily environments to provide innovative services for such spaces as parking lots, buildings, and railway networks. Yet, to realize the full potentials of these sensor network infrastructures, services need to be developed. Service development raises a number of challenges due to existing approaches that are often low level and network/hardware-centric. This paper proposes a high-level approach to the development of large-scale orchestrating applications. It revolves around a declaration language that allows to express the sensor-network dimensions of an application (sensor discovery, delivery models, actuation process). These declarations define the behavior of an application with respect to the sensor network infrastructure. We demonstrate the key relevance of these declarations at every stage of an application lifecycle, from design to runtime. In doing so, declarations allow to match the sensor-network behavior of an application to the target infrastructure. Our approach summarizes and puts in perspective our development of industrial case studies and our experience in using a commercially-operated sensor infrastructure.

7.6. Improving the Reliability of Pervasive Computing Applications By Continuous Checking of Sensor Readings

This paper shows that context-aware applications commonly make implicit assumptions about a sensor infrastructure. Because context-awareness critically relies on these assumptions, the developer typically need to ensure their validity by encoding them in the application code, polluting it with non-functional concerns. This defensive programming approach can be avoided by formulating these assumptions aside from the application, thus factorizing them as an explicit model of the sensor infrastructure. This model can be expressed as a set of rules and can be checked automatically and continuously to ensure the reliability of a sensor infrastructure, both at installation time and during normal functioning. The usefulness of our approach

is demonstrated in the domain of assisted living for seniors. We applied it to sensor data collected in the context of a 9-month field study of an assisted living platform, deployed at the home of 24 seniors. We show that several kinds of sensor malfunctions could have been identified upon their occurrence, thanks for our continuous checking, and resolved.

7.7. Designing Parallel Data Processing for Large-Scale Sensor Orchestration

Masses of sensors are being deployed at the scale of cities to manage parking spaces, transportation infrastructures to monitor traffic, and campuses of buildings to reduce energy consumption. These large-scale infrastructures become a reality for citizens via applications that orchestrate sensors to deliver high-value, innovative services. These applications critically rely on the processing of large amounts of data to analyze situations, inform users, and control devices. This paper proposes a design-driven approach to developing orchestrating applications for masses of sensors that integrates parallel processing of large amounts of data. Specifically, an application design exposes declarations that are used to generate a programming framework based on the MapReduce programming model. We have developed a prototype of our approach, using Apache Hadoop. We applied it to a case study and obtained significant speedups by parallelizing computations over twelve nodes. In doing so, we demonstrate that our design-driven approach allows to abstract over implementation details, while exposing architectural properties used to generate high-performance code for processing large datasets.

7. New Results

7.1. Clavispora lusitaniae genome

Clavispora lusitaniae is an ubiquist environmental ascomycetous yeast, with no known specific ecological niche. It can be isolated from different substrates, such as soils, waters, plants, and gastrointestinal tracts of many animals including birds, mammals and humans. In immunocompromised hosts, *C. lusitaniae* can be pathogenic and is responsible for about 1% of invasive candidiasis, particularly in pediatric and onco-haematology patients [24].

So far, two strains have had their genomes sequenced: ATCC 42720, isolated from the blood of a patient with myeloid leukemia [29], and MTCC 1001, a self-fertile strain isolated from citrus [27]. We performed the genome assembly of the *C. lusitaniae* type strain CBS 6936 [37], isolated from citrus peel juice.

Illumina sequences were obtained by our collaborator (T. Noel, UMR 5234 CNRS Université Bordeaux) and we ran the assemblies using several assemblers, e.g. MINIA [21], MIRA [20] and SPAdes [19]. Each assembly gave sequence scaffolds colinear with the already sequence genome of ATCC 42720. However the number of scaffolds varied dramatically in assemblies. SPAdes gave the best results by an order of magnitude (MINIA: 2913, MIRA: 930, SPAdes: 53). This last assembly will serve as a basis for further experiments and SNP detections in mutants strains derived from CBS 6936.

This work is a collaboration between Pleiade team, UMR 5234 CNRS/Université de Bordeaux, and MIAT INRA.

7.2. Elaeis guineensis transcriptome

The mesocarp of the oil palm tree (*Elaeis guineensis* Jacq.) contains oil up to 90% of its dry weight and is the oil richest known vegetal tissue [23]. Our goal is to understand how this tissue achieves this result, in order to increase, by synthetic biology approaches, the yield in oil for crops grown in Europe. In a first milestone, oil palm tree genes relevant for oil synthesis from fatty acids will be transiently expressed in tobacco leaves.

In order to select relevant gene candidates, we performed transcriptome assemblies on high quality Illumina sequences. As an annotated genome is available [34], we performed genome-guided assemblies with TRINITY assembler [26]. The sequence reads (total 300 millions) came from 5 RNA independent isolates from 3 tissues: leaf (2 isolates), kernel (1 isolate) and mesocarp (2 isolates). Transcriptomes coming from duplicate isolates show a good level of overlap as regards the predicted transcripts.

Using expression specificity, abundance and transcript annotation from the genome, we selected genes or transcrpt isoforms candidates, as well as transcription factors. A set of 19 sequences has been retained for expression in tobacco leaves and is under genetic engineering processing.

The 5 transcriptomes were fused into a pan-transcriptome which will be used in another angle of the project. The yield and the composition of oil measured from wild-type palm tree specimens varies dramatically, indicating a high level of bio-diversity. We are currently doing a sampling campaign in Africa, RNA extracts will be sequenced and compared to our pan-genome to serve in variant detection (SNPs) and association genetics studies.

This work is a collaboration between Pleiade team and UMR 5200 CNRS/Université de Bordeaux.

7.3. A Geometric View of Diversity

Diversity may be understood as a set of dissimilarities between objects. The underlying mathematical construction is the notion of distance. Knowing a set of objects, on which pairwise distances can be measured, it is possible to build a Euclidean image of it as a point cloud in a space of relevant dimension. The objects under study are microbial communities, given as a set of reads produced by NGS technologies. Distances between specimen are computed as genetic distances between associated reads (so called amplicon approach). Then, the diversity of a community can be associated with the shape of the point cloud built from such distances. Such an embedding is classically implemented by MDS (Multidimensional Scaling). Such an approach triggers two methodological questions, addressed in 2016:

- the numerical solution is through finding the eigenvectors and eigenvalues of a large, full, symmetric matrix. Current algorithm, parallelized or not, are in complexity $\mathcal{O}(n^3)$ if n is the number of specimen on which to study patterns of diversity, i.e. the size of the matrix. This is not feasible for dataset produced by NGS technologies, which can assemble 10^5 to 10^6 sequences. We have set up a collaboration with a PhD student in HIEPACS Team (Inria Bordeaux SO), Pierre Blanchard, to develop a connection between random projection methods and MDS. Random Projection Methods are methods relying on Johnson-Lindenstrauss Lemma, which states that the likelihood that the distances are very well conserved is very high when projecting a point cloud in a space of very large dimension (say n) to a random space of large dimension (say, proportional to $\log n$) . This permits to compute eigenvectors and eigenvelaues in space of reasonable dimension. The method for MDS has been studied by Pierre Blanchard, under supervision by Olivier Coulaud and Alain Franc, and proved to be surprisingly efficient and precise. This work builds one chapter of the PhD thesis of P. B. to be defended by early 2017. This collaboration has lead to a joint poster at Platform for Advanced Scientific Computing (PASC), June 2016, Lausanne, Switzerland.

- The eigenvalues of the matrix under study can be positive or negative. Positive eigenvalues lead to Euclidean structure behind MDS. Classically, negative eigenvalues are ignored. We have begun a study on the role of negative eigenvalues in the discrepancy between Euclidean distances computed between points in MDS, and genetic distances between reads produced by NGS, which adds to the well understood discrepancy in MDS due to dimension reduction. This has lead to three seminars or presentations:

    – a seminar at MIAT research Unit, Toulouse, on February 19

    – a seminar at LABRI on April, 28

    – a presentation at the days of mathematics and computing sciences organized by MIA INRA division (INRA global meeting), on October, 5

    These three events have permitted to "polish" the analysis of the problem through several exchanges, and to orientate its study towards quadratic embedding, or isometry into pseudo-euclidean spaces.

7.4. Topological Data Analysis

Leyla Mirvakhabova has defended and obtained her BSc memoir on "Distance geometry and biodiversity patterns" at the Department of Mathematics of the National Research University - Higher School of Economics, Moscow. Here is the abstract: In this work, we study the biodiversity of the tree species in French Guiana. We consider the matrix of the pairwise genetic distances between the 1501 species. The distances are measured by using the Smith-Waterman algorithm applied to the trnH regions - the chloroplasts of trees. The aim of the project is to analyze the shape of a point cloud in a Euclidean space built from the pairwise distances. To study the structure of the point cloud built from the given distances, we have considered the following methods: Hierarchical Clustering, Multidimensional Scaling (MDS), Nonlinear Mapping (NLM), t-Distributed Stochastic Neighbor Embedding (t-SNE), and Topological Data Analysis (TDA). For the first four methods, we used the Python package scikit-learn 0.17.1 implementations and have written the program for the TDA algorithm. All of these methods were tested on the given dataset. This work has been performed as part of a collaborative research project of the PLEIADE team in the Inria Bordeaux – Sud-Ouest (supervisor Alain Franc) with

the Laboratory of System Biology and Computational Genetics in the Vavilov Institute of General Genetics (supervisor Ivan Kulakovskiy).

7.5. Bespoke tools for comparative genomics

Large-scale comparison of strains of cell factory species is an indispensible tool for understanding the genetic origin of phenotypic variability, and can considerably optimize the selection and construction of high-performing industrial strains. For example, in oenological applications new strains may be selected based on their influence on aroma, their adaptation to grape musts, or their robustness during fermentation. In oil production applications, new strains may be selected based on their yield, or on the saturation degree of the lipids, or on their growth characteristics. Comparative genomics has proven quite effective in understanding cell factory diversity [1], [6], [5], [36], [31]. A typical project will involve 500 segregants and 50 genomes. Accurate and rapid analysis of the concomitant data volumes requires efficient tool sets that must be adapted to the real use cases of the industrial application.

PLEIADE addresses this problem through the definition of bespoke software systems that associate integrated sets of tools, including its Magus software platform (section 6.1). A key challenge in defining this kind of integrated system is the need to connect the components. We develop configuration formalisms whose solutions are orchestrations of weakly-coupled microservices running in independant containers. These services may be data banks, genome browser and visualization software, workflow tools like Galaxy, machine learning algorithms for classification, or shared workbooks like Jupyter or Zeppelin. By formalizing the connections between services, we can simplify deployment, and also create an opportunity for *continuous deployment*.

7. New Results

7.1. Alpha release of the T-calculus

One of the main achievement of the PoSET project in 2016 is the alpha release of the T-calculus [15] that not only implements the tiled front-end programing interface that was proposed earlier [10], [8], but also an original mide-end programing interface for implementing interactive behavior and the related categorical combinators that allows for effectively running these high level constructs.

7.2. A new collaboration with Bernard Serpette

A new collaboration with Bernard Serpette also aim at developing a formal models for the T-calculus semantics [27], [25]. Though at its birth, such an approach eventually reveals rather deep connection with Matsikoudis and Lee works on causal functions semantics [33], openning new perspectives towards higher-order timed programing.

<p style="text-align:center">**POTIOC Project-Team**</p>

7. New Results

7.1. HOBIT

Participants: David Furio, Benoit Coulais, Martin Hachet

Practical work in optics learning allows supporting the construction of knowledge, in particular when the concept to be learned remains diffuse. To overcome the limitations of the current experimental setups, we have designed a hybrid system that combines physical interaction and numerical simulation. This system relies on 3D-printed replicas of optical elements, which are augmented with pedagogical information (see Figure 3). In a first step, we have focused on the well-known Michelson interferometer experiment, widely studied in under graduate programs of Science. A 3-months user study with 101 students and 6 teachers showed that, beyond the practical aspects offered by this system, such an approach enhances the technical and scientific learning compared to a standard Michelson interferometer experiment. A second version of HOBIT is currently being developed. This new version will let us simulate and augment multiple experiments related with optics, like polarization or Young's interferometer.

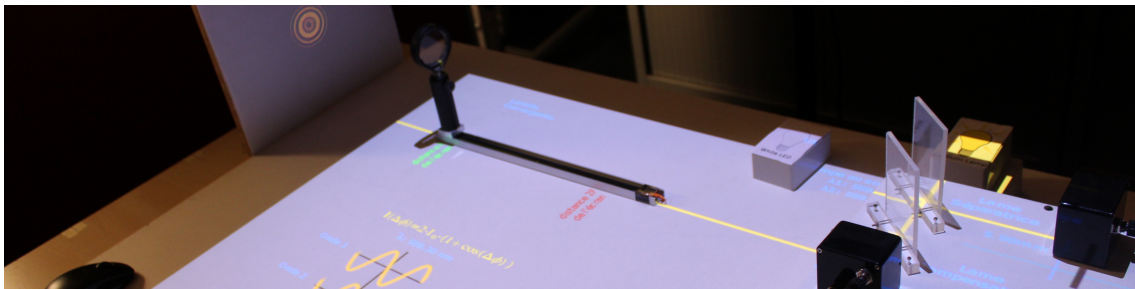A paper presenting HOBIT has been (conditionaly) accepted at ACM CHI 2017.

Figure 3. HOBIT: Hybrid Optical Bench for Innovative Teaching

7.2. Inner Garden

Participants: Joan Sol Roo, Renaud Gervais, Jeremy Frey, Martin Hachet

Digital technology has completely integrated our daily lives; we use it for entertainment, productivity and our social lives. However, the potential of leveraging technology to improve its users' overall happiness and life satisfaction is still largely untapped. Mindfulness, the act of paying a deliberate and non-judgmental attention to the present moment, has been shown to have a positive impact on a person's subjective well-being. With this in mind we created Inner Garden, an ambient mixed reality installation, inspired by a zen garden, comprised of an augmented sandbox along with a virtual reality modality to support mindful experiences (Figure 4 . By shaping the sand, the user creates a living miniature world that is projected back onto the sand. Moreover, using a VR headset, she can take a moment to herself by actually going inside her own garden to meditate. The natural elements of the garden are connected to real-time physiological measurements, such as breathing, helping staying focused on the body. We evaluated the system through a first user study and consulted meditation teachers, who envisioned the use of the garden in their teaching, especially for novice practitioners. The reception of the system seems to indicate that technology can, when designed carefully, both engage the users and foster well-being.

Figure 4. Inner Garden, an ambient mixed reality installation to support mindful experiences

A paper presenting Inner Garden has been (conditionaly) accepted at ACM CHI 2017.

7.3. Hybridation of Spatial Augmented Reality and Virtual Reality

Participants: Joan Sol Roo and Martin Hachet

Spatial Augmented Reality (SAR) allows a user, or a group of users, to benefit from digital augmentations embedded directly into the physical world. This enables co-located information and unobstructed interaction. On the other hand, SAR suffers from limitations that are inherently linked to its physical dependency, which is not the case for see-through or immersive displays. In this work, we explore how to facilitate the transition from SAR to VR, and vice versa, integrating both into a unified experience (Figure 5). We developed a set of interaction techniques and obtained first feedback from informal interviews.

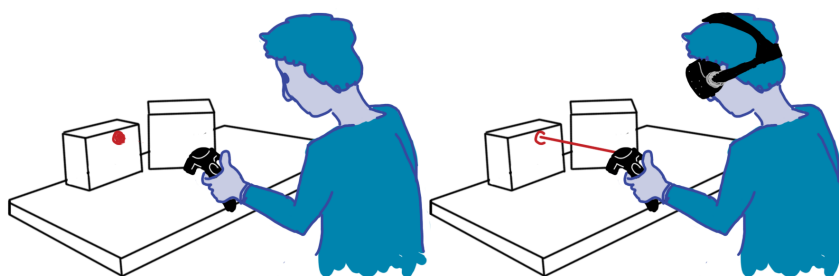A technote presenting this work has been (conditionaly) accepted at IEEE 3DUI 2017.

Figure 5. Transition from Spatial Augmented Reality to Virtual Reality

7.4. Augmented Reality and Tangible User Interfaces for Understanding Astronomy

Participants: Robin Gourdel, Benoit Coulais, Jeremy Laviole, Martin Hachet

We have worked with Stephanie Fleck (Université de Lorraine) to improve the leraning environment AIBLE she had imagined (see http://stefleck4.wixsite.com/aible/-propos2-cw4c). We have designed Helios, an augmented reality platform that aims at enhancing the understanding of abstract concepts in astronomy, specifically for primary schools' curriculum with children aged from 8 to 11. In order to provide physical evidence for the influence of sunlight on the Earth and the Moon, and of the consequences of their relative positions, the learning tasks are designed on inquiry-based learning principles. Children have to test their own hypotheses by using tangible props and a set of cards that trigger dedicated pedagogical activities (e.g. seasons and the Earth revolution around the Sun, lunation origin, Earth rotation and time measurement).

Helios basically consists of a standard laptop computer, a webcam, printable AR markers placed on tangible props and on dedicated pedagogical cards (See Figure 6). The (virtual) celestial bodies are displayed on the screen, and many visual feedback help understanding various phenomena (e.g. shadow cones, time zones, and so on). In [13], we discuss how such an approach allows learners to better understand abstract phenomena.
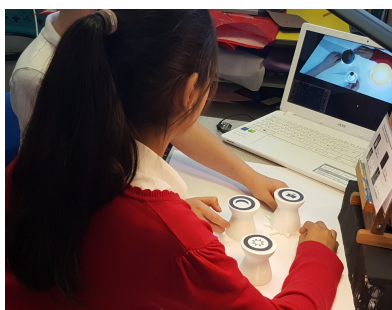

Figure 6. Helios: Manipulation of tangible objects and visualization of an augmented scene to learn astronomy.

7.5. Collaboration in VR

Participants: Damien Clergeaud and Pascal Guitton

The Airbus company regularly uses virtual reality for design, manufacturing and maintenance. We work with them on collaborative interaction in order to enable an efficient collaboration between operators immersed in the same virtual environment from remote locations and with heterogeneous equipment (large displays, CAVE, HMD). More precisely, we have developped tools to point and manipulate 3D objects, to remotely visualize the virtual environment, to be aware of remote manipulations and to describe tools and components trajectories (Figure 7). These tools have been validated by Airbus experts and the next step is to integrate them in their global process.

We are also working on Through-The-Lens Interaction techniques to ease the collaboration in some asymmetric tasks that requires a guide and an operator. Through-The-Lens techniques enable the guide to interact with the surroundings of the operator in order to help him in the task he has to perfrom. A paper presenting such a technique has been (conditionaly) accepted at IEEE 3DUI 2017.

7.6. Interactive 3D Environments for Immersive Musical Performances

Participants: Martin Hachet

Together with Florent Berthaut (Univeristé Lille 3), we presented a set of works that attempts to extent the frontiers of music creation as well as the experience of audiences attending digital performances. Indeed, the power of interactive 3D graphics, immersive displays, and spatial interfaces is still under-explored in domains where the main target is to enhance creativity and emotional experiences. The goal of our work is to connect sounds to interactive 3D graphics that musicians can interact with and the audience can observe [11].
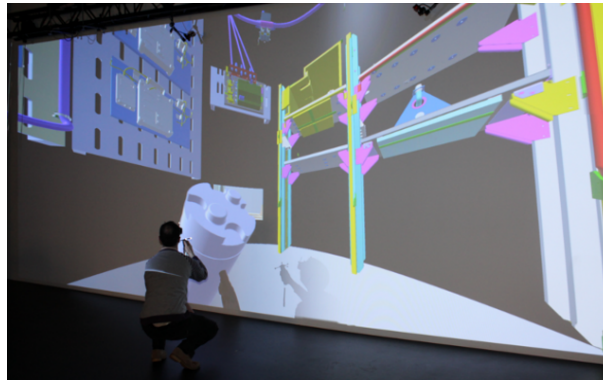
Figure 7. An immersed user has to perform a virtual task in a complex environment. In order to help the user to be fully aware of the VE, another immersed operator may guide him using a Through-The-Lens metaphor.

7.7. Multisensory Maps and 3D Printed Interactive Maps for Visually Impaired People

Participant: Anke Brock

Visually impaired people face important challenges related to orientation and mobility. Accessible geographic maps are helpful for acquiring knowledge of an urban environment. Historically, raised-line paper maps with braille text have been used, but these maps possess significant limitations. For instance, only a small percentage of the visually impaired population reads braille. Recent technological advances have enabled the design of accessible interactive maps with the aim to overcome these limitations. We designed Mappie, an accessible interactive map prototype based on the use of a multi-touch screen with a raised-line map overlay and speech output (Figure 8). Then, we deployed Mappie in a class of seven children and one caretaker during three months. Our formative study showed promising results and allowed insights in the design of accessible interactive maps, such as using olfactory and gustatory modalities to foster reflective learning, and using tangible objects to support storytelling. Following this first study, we designed MapSense as an extension of Mappie. MapSense uses the same hardware and interaction techniques as Mappie, but additionally provides fourteen "Do-It-Yourself" conductive tangibles. Some tangibles could be filled with scented material, such as olive oil, smashed raisins or honey, thus creating a real multi-sensory experience. The map was explored during two classes of three hours separated by a week, taught conjointly by a locomotion trainer and a specialized teacher. We observed that the map and tangible objects triggered strong positive emotions and stimulated spatial learning as well as creativity of the visually impaired students. This work has been conducted as part of the PhD thesis of Emeline Brulé under the supervision of Gilles Bailly and Annie Gentes, and in collaboration with the IRIT research lab in Toulouse. It has been published at CHI'16 [20].

As part of the postdoc of Stephanie Giraud at IRIT Toulouse under the supervision of Christophe Jouffrais, we have investigated how to print entire interactive maps in 3D, allowing users to construct a city like using a puzzle. We have conducted a user study comparing an interactive map that has been entirely 3D printed to a raised-line map with braille text (Figure 8 left). Our results suggest that the interactive map is significantly more effective for providing spatial knowledge than a tactile paper map with braille.

7.8. Navidrone

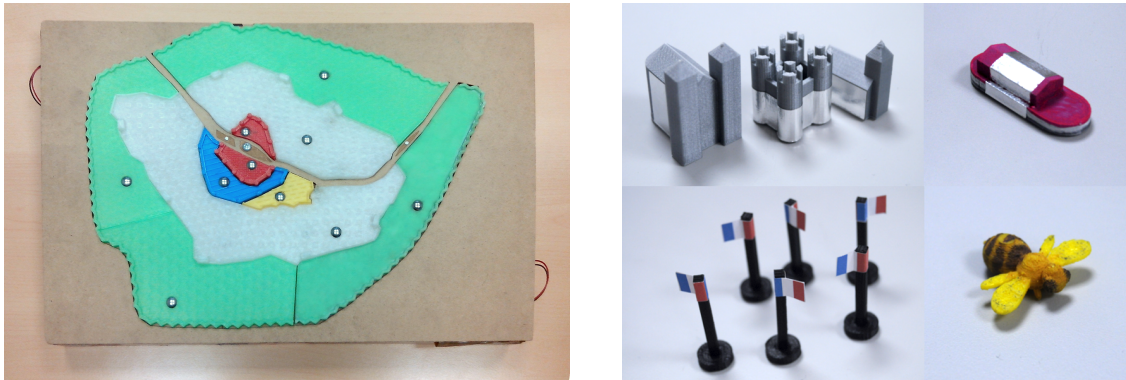**Participants**: Julia Chatain, Anke Brock, Martin Hachet

Figure 8. (Left) 3D printed interactive map, (Right) 3D printed tangibles for multisensory maps

With recent technological advances, the shapes of mobile devices are evolving. For example, we now see the emergence of new types of devices in form of autonomous aerial vehicles (drones) that become available in our everyday environment. As drones are becoming increasingly autonomous, it is crucial to understand how interaction with such devices will happen. These new devices, allow us to imagine new contexts of map usage, as for instance drone-based autonomous tour guides ((Figure 9). In order for those scenarios to happen, many problems need to be investigated. From a perspective in Human-Computer Interaction (HCI), the first questions to study are related to suitable input and output techniques. We iteratively designed interaction techniques for Navidrone, a drone-based autonomous tour guide. This work has been done in collaboration with the Prof. James Landay and Dr. Jessica Cauchard from the Stanford HCI Group.

Figure 9. Sketch showing the Navidrone concept: users interact with maps projected from drones by using their phones.

7.9. Accessibility of e-learning systems

Participants: Pierre-Antoine Cinquin and Pascal Guitton

E-learning systems, such as MOOC or serious games, are increasingly taking part in training process. Unfortunately, like most digital systems, they suffer from a lack of accessibility, in particular for people with cognitive disabilities (e.g. who have limited attention and memory). In this project, we develop a framework based on various disciplinary fields (education, cognitive sciences, human factors) but also participatory design research with students with disabilities to design interfaces promoting e-learning accessibility. From this 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. We are currently working on the analysis of this first play in order to enhance Aïana by designing new interaction modalities.
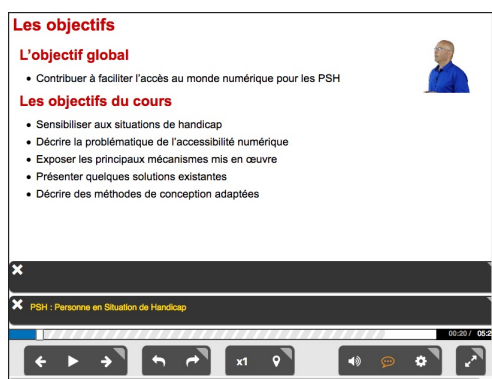

Figure 10. The Aïana MOOC player.

7.10. Teegi, a tangible EEG interface for scientific outreach

Participants : Thibault Lainé, Renaud Gervais, Jérémy Frey, Hugo Germain, Fabien Lotte, Martin Hachet

Teegi is an interactive systems that combines electroencephalographic (EEG) recordings and tangible interaction in order to let novices learn about how their brain works. By displaying EEG activity in real time on a support that is easy to manipulate and to comprehend, Teegi is a good tool for scientific outreach, that raises public interest.

While last year we developed a semi-spherical display based on LEDs, we push the project further during 2016 and built a complete autonomous puppet (Figure 11). The robot can move its two hands independently or its feet, and it can close its eyes. Beside the display of EEG activity, Teegi can react accordingly to the brain patterns recorded in real time from the user.

We demonstrated this new prototype in various occasions over the year, during public events such as "Fête de la Science" in La Cité des Sciences in Paris, a manifestation that gathered thousands of visitors (See Section 10.3 "Popularization").

7.11. Neuroergonomy

Participants : Jérémy Frey, Martin Hachet, Fabien Lotte

3D user interfaces are increasingly used in a number of applications, spanning from entertainment to industrial design. However, 3D interaction tasks are generally more complex for users since interacting with a 3D environment is more cognitively demanding than perceiving and interacting with a 2D one. As such, it is essential that we could finely evaluate user experience, in order to propose seamless interfaces. To do
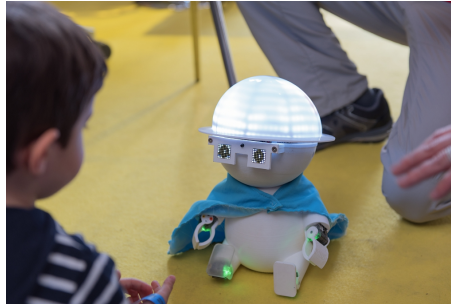
Figure 11. Teegi is a "Tangible EEG interface" that displays cerebral activity in real time by the mean of electroencephalography. A new robotized version can move and react accordingly to the brain patterns of the user, helping to explain to novices how the brain works.

so, a promising research direction is to measure users' inner states based on brain signals acquired during interaction, by following a neuroergonomics approach. Combined with existing methods, such tool can be used to strengthen the understanding of user experience.

In [15][26], we reviewed the related work in this area. We summurized what has been achieved and the new challenges that arise. We described how a mobile brain imaging technique such as electroencephalography (EEG) brings continuous and non-disruptive measures. EEG-based evaluation of users can give insights about multiple dimensions of the user experience, with realistic interaction tasks or novel interfaces. We investigate four constructs: workload, attention, error recognition and visual comfort. Ultimately, these metrics could help to alleviate users when they interact with computers.

Such advance in the understanding of the users will eventually come forward thanks to the increasing dissemination of non-invasive brain imaging devices that record electrical activity onto the scalp. In [24][23] we compared side by side two EEG amplifiers, the consumer grade OpenBCI and the medical grade g.tec g.USBamp. We suggested how an affordable and open-hardware device could facilitate, beside neuroergomomics, the appearance of various brain-computer interfaces applications.

7.12. Physiological computing

7.12.1. *Physiological computing*

Participants : Jérémy Frey

While physiological sensors enter the mass market and reach the general public, they are still mainly employed to monitor health. Over the course of a thesis that explored the new possibilities offered by physiological computing in terms of communication and social presence, we described several use-cases involving the externalization of inner states through novel user interfaces.

For example, we created an application that uses heart rate feedback as an incentive for social interactions. A traditional board game was "augmented" through remote physiological sensing (Figure 12), using webcams to account for the subtle changes in blood flow that occur with each heartbeat. Projection helped to conceal the technological aspects from users and merged the biofeedback with the physical environment. We detailed how players reacted – stressful situations could emerge when users are deprived from their own signals – and we gave directions for game designers to integrate physiological sensors.

We envisioned a second application, that merges virtual reality, interactive fiction and physiological computing in order to craft *truly* immersive stories; narratives that evolve depending both on the actions and on the inner states of the user/reader, stretching a medium that shaped for ages humanity (Figure 13) [32].

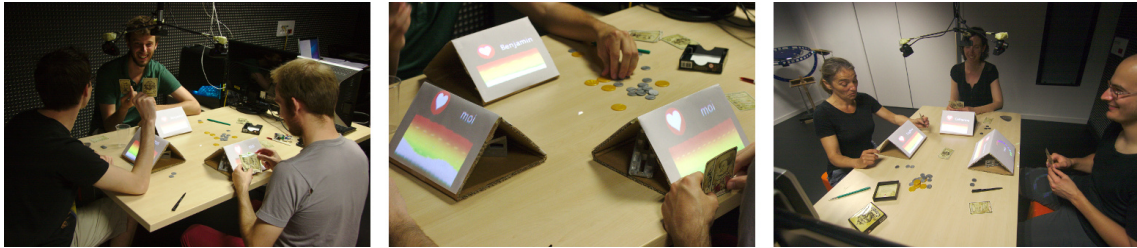*Figure 12. We augmented a traditional board game with remote physiological monitoring and projection to demonstrate how physiological computing could be used to foster new interactions between people and increase social presence.*
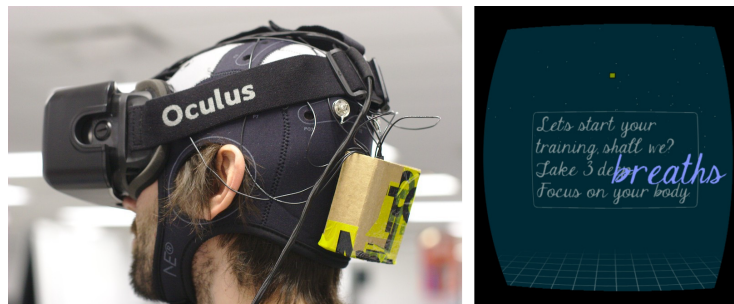

Figure 13. A combination of physiological sensors and head-mounted display (left) is used to immerse the reader in a narrative that reacts to gaze and to bodily activity (right).

7.13. EEG signal classification for BCI based on Riemannian geometry

Participants : Fabien Lotte

Although promising from numerous applications, current Brain-Computer Interfaces (BCIs) still suffer from a number of limitations. In particular, they are sensitive to noise, outliers and the non-stationarity of ElectroEncephaloGraphic (EEG) signals, they require long calibration times and are not reliable. Thus, new approaches and tools, notably at the EEG signal processing and classification level, are necessary to address these limitations. Riemannian approaches, spearheaded by the use of covariance matrices, are such a very promising tool slowly adopted by a growing number of researchers. We proposed a review of how these approaches have been used for EEG-based BCI, in particular for feature representation and learning, classifier design and calibration time reduction. Finally, we also identified relevant challenges and promising research directions for EEG signal classification in BCIs, such as feature tracking on manifold or multi-task learning [18].

7.14. Understanding Mental Imagery-based Brain-Computer Interface user-training

Participants : Camille Jeunet, Fabien Lotte

Mental Imagery-based Brain-Computer Interface (MI-BCI) enable their users to send commands to computer by imagining mental tasks (i.e., by performing MI) that are recognized in their brain signals. This type of BCI requires user training, and this training is currently poorly understood, and we basically do not know, who can learn MI-BCI control, what is to learn and how to learn it efficiently. Moreover, we have shown that current MI-BCI training protocols were both theoretically and practically inappropriate, and that there is a lack of fundamental knowledge on BCI user training, which prevents us from designing better user training approach [12].

In order to address these points, we first proposed a review and classification of cognitive and psychological predictors of MI-BCI performance. Three categories were defined: the user-technology relationship, attention and spatial abilities. The user-technology relationship refers to personality traits and states that influence users' perception of the technology and consequently impact the way they will interact with the technology (i.e., their feeling of being in control, their self-efficacy, etc.). The attention category gathers, among others, users' attentional abilities, motivation and engagement towards the task. These elements are essential to learn, whatever the skill targeted. They are also closely related to the user-technology relationship (for instance, feeling in control will increase users' engagement towards the task, thus they will allocate more attentional resources to the task). Finally, spatial abilities are the ability to produce, manipulate and transform mental images, which is closely related to the ability to control an MI-BCI. The description of these categories and of their neurophysiological correlates enabled us to submit ideas to improve MI-BCI user-training. For instance, we explained how to reduce computer-anxiety and increase the sense of agency, notably through the use of a positively biased feedback for novice users. Also, we proposed solutions to raise and improve attention, e.g., using neurofeedback or meditation. Finally, we argued that spatial abilities could be trained to improve users' capacity to perform mental imagery and consequently, potentially improve their MI-BCI performance [17].

We also did a review of the literature of current training protocols (published as a book chapter in [41]) which suggests that these protocols are, at least theoretically, inappropriate to acquire a skill and thus that they could be one of the factors responsible for inefficient MI-BCI user-training. In particular, participants are most of the time provided with uni-modal and evaluative feedback while literature recommends multi-modal, informative and supporting feedback. Although instructive, these insights only provide theoretical considerations about the flaws associated with the feedback approaches used in MI-BCI. It was therefore necessary to *concretely* assess whether standard MI-BCI feedback is appropriate to train a skill, and to what extent the feedback impacts BCI performance and skill acquisition. In order to experimentally evaluate the extent to which such a feedback has an impact on their ability to acquire a skill, we used it to teach users to perform simple motor tasks. Results (N=53 participants) revealed that with this feedback, 17% did not manage to learn the skill. This suggests that current BCI feedback is most probably suboptimal to teach a skill. A sub-group of our participants (N=20) then took part in a motor-imagery based BCI experiment. Results showed that those who struggled during the

first experiment improved in performance during the second, while the others did not. We hypothesised that these results are linked to the considerable cognitive resources required to process this feedback [16].

It should be noted that there are many connections between BCI user training, and neurofeedback training for clinical applications, both field aiming at training their users to perform self regulation of their brain activity. We have therefore shown how these two field share fundamental research questions on BCI user training, and how they can both benefit from each other [10].

7.15. Improving Mental Imagery BCI user-training & feedback

7.15.1. Spatial Ability Training Protocol

Participants : Suzy Teillet, Camille Jeunet, Fabien Lotte

The results of one of our previous studies suggested that users' MI-BCI performance correlates with their spatial abilities [34], which is consistent with the literature. This result was replicated in a second study in a pure motor-imagery based BCI [16]. We thus decided to explore the potential causal relationship between both: would an improvement of spatial abilities lead to better MI-BCI performances? To try to answer this question, we designed and implemented a spatial ability (SA) training procedure (see Figure 14). Then, we performed two user studies to validate the SA training procedure: results suggest that it efficiently improves participants' SA [29]. Consequently, we included this SA training procedure in an MI-BCI protocol. Results (N=24) showed no difference in classification accuracy between participants performing 6 MI-BCI sessions and those performing 3 SA and 3 MI-BCI sessions. Nonetheless, SA training intensity impacted users' progression, and neurophysiological analyses provided us with valuable insights into brain pattern evolution throughout the training process.

Figure 14. One item per exercise included in the Spatial Ability training: the shape on top is the target, and the participant must identify the two shapes that are identical to the target among the four below.From the left to the right are displayed the shapes, matrices, cubes, arms exercises.

7.15.2. PEANUT: Personalised Emotional Agent for Neurotechnology User-Training

Participants : Léa Pillette, Camille Jeunet, Boris Mansencal, Fabien Lotte

Mental-Imagery based Brain-Computer Interfaces (MI-BCI) are neurotechnologies enabling users to control applications using their brain activity alone. These neurotechnologies are very promising. However, existing training protocols do not enable every user to acquire the skills needed to use them. Indeed, those protocols are not consonant with psychological field recommendations. In particular, the current protocols do not provide social presence and emotional support to the user. Therefore, we designed and tested PEANUT, the first Learning Companion dedicated to the improvement of MI-BCI user-training. PEANUT has been designed throughout a combination of recommendations from the literature, the analysis of data from previous experiments and user-studies. He provides emotional support using spoken sentences, such as "'C'est avec la pratique que l'on progresse"', and facial expressions. Experiments were conducted in order to assess his influence on user's performance and experience. The first results indicate that PEANUT improves the user experience. Indeed, people who trained with PEANUT found it was easier to learn and memorize how to use the MI-BCI system and rated themselves more efficient and effective than participants who had no learning
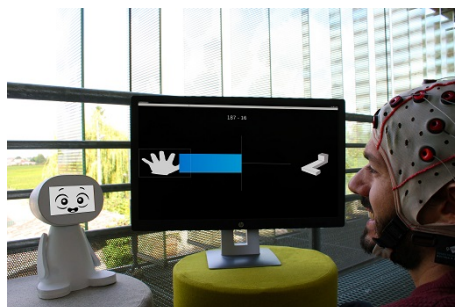
Figure 15. A participant taking part in a Brain-Computer Interface training process during which he learns to perform different mental imagery tasks (here, imagining a left-hand movement) to control the system. Along the training, PEANUT (on the left) provides the user with social presence and emotional support adapted to his performance and progression.

companion. These results indicate that using PEANUT does benefit MI-BCI user training. Future research will keep focusing on how to provide adapted cognitive and emotional feedback to MI-BCI users thanks to the use of learning companions.

7.16. Adaptive BCI training and operation

Participants : Jelena Mladenović, Jérémy Frey, Fabien Lotte

7.16.1. A generic framework for adaptive EEG-based BCI training and operation

There are two main approaches engaged in improving BCI systems: (i) improving the machine learning techniques, and the newly introduced (ii) improving human learning, by using the knowledge from instructional design and positive psychology. Both agree that the system needs to be adapted to the user but rely on different sources of adaptation: the machine for the former and the brain for the latter. In particular, machine learning algorithms should adapt to non-stationary brain signals, while human learning approaches should adapt the system to the various users' skills and profiles. Including both aspects of adaptation would give rise to a system ready to be used in real life conditions. However, a major obstacle lies in the large spectrum of sources of variability during BCI use, ranging from (i) imperfect recording conditions (e.g., environmental noise, humidity, static electricity etc. to (ii) the fluctuations in the user's psychophysiological states, due to e.g., fatigue, motivation or attention. For these reasons a BCI has not yet proved to be reliable enough to be used outside the laboratory. Particularly, it is still almost impossible to create one BCI design effective for every user, due to large inter subject variability. Therefore, the main concerns are to create a more robust system with the same high level of success for everyone, at all times, and to improve the current usability of the system. This calls for adaptive BCI training and operation.

We propose a conceptual framework which encompasses most important approaches to fit them in such a way that a reader can clearly visualize which elements can be adapted and for what reason. In the interest of having a clear review of the existing adaptive BCIs, this framework considers adaptation approaches for both the user and the machine, i.e., referring to instructional design observations as well as the usual machine learning techniques. It provides not only a coherent review of the extensive literature but also enables the reader to perceive gaps and flaws in current BCI systems, which would, hopefully, bring novel solutions for an overall improvement.

The framework (see Figure 16) has a hierarchical structure, from the lowest level elements which endure rapid changes, to the highest level elements which change at a much slower rate. It is comprised of: (i) one or several BCI systems/pipelines; (ii) a user model, whose elements are arranged according to different time scales ; (iii) a task model, enabling the system adaptation with respect to the user model; (iv) the conductor, an intelligent agent which implements the adaptive control of the whole system. A book chapter on this framework was submitted to a handbook on BCI.
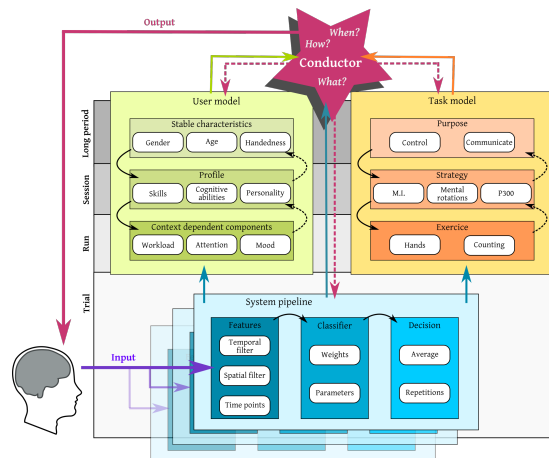

Figure 16. Multiple signals (input) maybe observed and processed in parallel in order to infer complementary states or intents, at the trial wise time scale. All the information extracted from these parallel pipelines may trigger the up-dating of the user or task model, which in turn might yield a decision from the conductor to take action, such as adapting one of the systems or the output, or modifying the task or the user model.

7.16.2. Adapting BCI Feedback based on Flow Theory

Using BCI systems can be very frustrating for people because it is not trivial and so it takes time to master. Differently from other learning procedures, BCIs do not have enough, if any explanatory feedback in assisting the learning of users. Also, as the feedback is not engaging the user's mind might easily wander off, which highly affects the system's accuracy as well as the person's learning pace. For this reason it takes more time to train a user to understand the procedure and have control over the system. Hence, we want to create an immersive and playful environment to attract the user's attention and help them learn with less effort and frustration.

We rely on the theory of Flow, introduced by Csikszentmihalyi in the 1970s. Flow is a state of enjoyment while effortlessly focused on a task so immersive that one looses the perception of time. In order to fulfil these requirements, we choose the users to be involved in an open-source video game called Tux Racer. Also, to ensure the maximal attention of the users, the game difficulty adapts according to users performance in real-time.

7.17. Brain-Computer Interfaces 1 and 2: foundations, methods, practice and applications

Participants : Jérémy Frey, Camille Jeunet, Fabien Lotte.

Figure 17. A subject using motor imagery, i.e., imagining left or right hand movements to manipulate Tux to catch fish.

Together with Maureen Clerc (Inria Sophia) and Laurent Bougrain (Inria Nancy), we co-edited the first book on Brain-Computer Interfaces in French [50], [51], this book being also translated into English [48], [49]. It is published in two volumes, and co-written with researchers from all over France from many disciplines related to BCI. It covers both theoretical and practical aspects, as well as the neuroscience, mathematics, psychology, computer science, engineering, and ethical aspects of BCI. It is aimed at being a key resource for anyone who wants to start BCI research or want to deepen their knowledge in the many aspects of this exciting discipline.

Figure 18. The two volumes of the French version of the BCI book we edited.

<p style="text-align:center; color:red;">**REALOPT Project-Team**</p>

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 [29] 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 [18], [30] 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 neighborhood 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. Aggregation Techniques

We have developed [6] a general solution framework based on aggregation techniques to solve NP-Hard problems that can be formulated as a circulation model with specific side constraints. The size of the extended Mixed Integer Linear Programming formulation is generally pseudo-polynomial. To efficiently solve exactly these large scale models, we propose a new iterative aggregation and disaggregation algorithm. At each iteration, it projects the original model onto an aggregated one, producing an approximate model. The process iterates to refine the current aggregated model until the optimality is proved.

The computational experiments on two hard optimization problems (a variant of the vehicle routing problem and the cutting-stock problem) show that a generic implementation of the proposed framework allows us to outperform previous known methods.

We have applied this aggregation method to reduce the size of column generation (CG) models for covering problems in which the feasible subsets depend on a resource constraint [10]. The aggregation relies on a correlation between the resource consumption of the elements and the corresponding optimal dual values. The resulting aggregated dual model is a restriction of the original one, and it can be rapidly optimized to obtain a feasible dual solution. A primal bound can also be obtained by restricting the set of columns to those saturated by the dual feasible solution obtained by aggregation. The convergence is realized by iterative disaggregation until the gap is closed by the bounds. Computational results show the usefulness of our method for different cutting-stock problems. An important advantage is the fact that it can produce high-quality dual bounds much faster than the traditional lagrangian bound used in stabilized column generation.

7.3. Review of Algorithmic Enhancements for Benders Decomposition

In Benders decomposition approach to mixed integer programs, the optimization is carried in two stages: key first-stage decision variables are optimized using a polyhedral approximation of the full-blown problem projection, then a separation problem expressed in the second-stage variables is solved to check if the current first-stage solution is truly feasible, and otherwise, it produces a violated inequality. Such cutting-plane algorithms suffer from several drawbacks and may have very bad convergence rates. We have reviewed [23] the battery of approaches that have been proposed in the literature to address these drawbacks and to speed-up the algorithm. Our contribution consists in explaining these techniques in simple terms and unified notations, showing that in several cases, different proposals of the literature boil down to the same key ideas. We classify methods into specific initialization mode, stabilization techniques, strategies to select the separation point, and cut generation strategies. Where available, we highlight numerical benchmarks that have resulted from such enhancements.

7.4. Vehicle Routing Problems

Given a directed graph $G = (V, A)$, a cost function c associated with the arcs of A, and a set of precedence constraints $B \subset V \times V$, the Precedence Constrained Asymmetric Traveling Salesman Problem (PCATSP) seeks for a minimum cost Hamiltonian circuit, starting at node 1, and such that for each $(i, j) \in B$, the node i is visited before node j. There are many ways of modelling the ATSP and several for the PCATSP. In [20] we present new formulations for the two problems that can be viewed as resulting from combining precedence variable based formulations with network flow based formulations. Indeed, the former class of formulations permits to integrate linear ordering constraints. The motivating formulation for this work is a complicated and "ugly" formulation that results from the separation of generalized subtour elimination constraints presented. This so called "ugly" formulation exhibits, however, one interesting feature, namely the "disjoint subpaths" property that is further explored to create more complicated formulations that combine two (or three) "disjoint path" network flow based formulations and have a stronger linear programming bound. Some of these stronger formulations are related to the ones presented for the PCATSP and can be viewed as generalizations in the space of the precedence based variables. Several sets of projected inequalities in the space of the arc and precedence variables are obtained by projection from these network flow based formulations. Computational results for the ATSP and PCATSP evaluate the quality of the new models and inequalities.

The Dial-a-Ride Problem is a variant of the pickup and delivery problem with time windows, where the user inconvenience must be taken into account. In [17], ride time and customer waiting time are modeled through both constraints and an associated penalty in the objective function. We develop a column generation approach, dynamically generating feasible vehicle routes. Handling ride time constraints explicitly in the pricing problem solver requires specific developments. Our dynamic programming approach for pricing problem makes use of a heuristic dominance rule and a heuristic enumeration procedure, which in turns implies that our overall branch-and-price procedure is a heuristic. However, in practice our heuristic solutions are experimentally very close to exact solutions and our approach is numerically competitive in terms of computation times.

In [22], [21], we consider the problem of covering an urban area with sectors under additional constraints. We adapt the aggregation method to our column generation algorithm and focus on the problem of disaggregating the dual solution returned by the aggregated master problem.

7.5. Production Scheduling Problems

We have considered [7] the flowshop problem on two machines with sequence-independent setup times to minimize total completion time. Large scale network flow formulations of the problem are suggested together with strong Lagrangian bounds based on these formulations. To cope with their size, filtering procedures are developed. To solve the problem to optimality, we embed the Lagrangian bounds into two branch-and-bound algorithms. The best algorithm is able to solve all 100-job instances of our testbed with setup times and all 140-job instances without setup times, thus significantly outperforming the best algorithms in the literature.

In [9], we address a multi-activity tour scheduling problem with time varying demand. The objective is to compute a team schedule for a fixed roster of employees in order to minimize the over-coverage and the under-coverage of different parallel activity demands along a planning horizon of one week. Numerous complicating constraints are present in our problem: all employees are different and can perform several different activities during the same day-shift, lunch breaks and pauses are flexible, demand is given for 15 minutes periods. Employees have feasibility and legality rules to be satisfied, but the objective function does not account for any quality measure associated with each individual's schedule. More precisely, the problem mixes simultaneously days-off scheduling, shift scheduling, shift assignment, activity assignment, pause and lunch break assignment.

To solve this problem, we developed four methods: a compact Mixed Integer Linear Programming model, a branch-and-price like approach with a nested dynamic program to solve heuristically the subproblems, a diving heuristic and a greedy heuristic based on our subproblem solver. The computational results, based on both real cases and instances derived from real cases, demonstrate that our methods are able to provide good quality solutions in a short computing time. Our algorithms are now embedded in a commercial software, which is already in use in a mini-mart company.

7.6. Scheduling and Placement for HPC

With the complexification of the architecture of HPC nodes (multicores, non uniform memory access, GPU and accelerators), a recent trend in application development is to explicitly express the computations as a task graph, and rely on a specialized middleware stack to make scheduling decisions and implement them. Traditional algorithms used in this community are dynamic heuristics, to cope with the unpredictability of execution times. In [12], we analyze the performance of static and hybrid strategies, obtained by adding more static (resp. dynamic) features into dynamic (resp. static) strategies. Our conclusions are somehow unexpected in the sense that we prove that static-based strategies are very efficient, even in a context where performance estimations are not very good. We also present and generalize HeteroPrio, a semi-static resource-centric strategy based on the acceleration factors of tasks. In [19], we generalize this strategy to platforms with more than two types of resources. This allows to use intra-task parallelism by grouping several CPU cores together. In [27], we prove tight approximation ratios for HeteroPrio in the context of independent tasks, providing a theoretical insight to its good practical performance.

Another study [26] focuses on the memory-constrained case, where tasks may produce large data. A task can only be executed if all input and output data fit into memory, and a data can only be removed from memory after the completion of the task that uses it as an input data. There is a known, polynomial time algorithm [55] to minimize the peak memory used on one machine for the cases where the input graph is a rooted tree. We generalize in [26] to the variant where the input graph is a directed series-parallel graph, and propose a polynomial time algorithm. This allows to solve this practical problem in two important classes of applications.

In [13], we consider the static problem of data placement for matrix multiplication in heterogeneous machines, so as to optimize both load balancing and communication volume. This is modeled as a partitioning of a square into a set of zones of prescribed areas, while minimizing the overall size of their projections onto horizontal and vertical axes. We combine two ideas from the literature (recursive partitioning, and optimal solution structure for low number of processors) to obtain a non-rectangular recursive partitioning (NRRP), whose approximation ratio is $\frac{2}{\sqrt{3}} \simeq 1.15$, improving over the previous 1.25 ratio. Moreover, we observe on a large set of realistic platforms built from CPUs and GPUs that this proposed NRRP algorithm allows to achieve very efficient partitionings on all considered cases. In [14], we consider the generalization of this problem to the three dimensional case. We prove the NP-completeness of the problem, and propose a generalisation of NRRP with a $\left(\frac{5}{6}\right)^{\frac{2}{3}}$ approximation ratio.

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 [15], we design a Two-Echelon Distribution Network that helps synchronize 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 [8], 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 $k = 3, 4, 5$.

7.8. Strategic Planning of Phytosanitary treatments in Wineries

In [16], we consider planning phytosanitary treatments in a vineyard. We are given a set of diseases (or requests) that must be treated for each site. Product mixtures are defined by their composition of active components, and their duration of protective power for each request. Machines are available to spread the mixtures on the sites. The time horizon is divided in time periods. Sites are partitioned in sectors. The objective of the problem is to minimize the machine leasing costs, their travel cost to sectors and the costs related to the product use. To solve this problem, we use a column generation approach where the machine policy and the product order policy are pure master decisions, while treatment planning decisions are made in individual pricing subproblems associated with each site. We developed a dedicated dynamic program to solve the pricing subproblems.

7.9. Two-dimensional Guillotine-cut Bounded Knapsack Problem

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 [28], 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.

7.10. Matching-Based Allocation Strategies in Cloud Platforms

MapReduce is a well-know framework for distributing data-processing computations on parallel clusters. In MapReduce, a large computation is broken into small tasks that run in parallel on multiple machines, and scales easily to very large clusters of inexpensive commodity computers. Before the Map phase, the original dataset is first split into chunks, that are replicated (a constant number of times, usually 3) and distributed onto the computing nodes. During the Map phase, nodes request tasks and are allocated first tasks associated to local chunks (if any). Communications take place when requesting nodes do not hold any local chunk anymore. In [25], we provide the first complete theoretical data locality analysis of the Map phase of MapReduce, and more generally, for bag-of-tasks applications that behaves like MapReduce. We show that if tasks are homogeneous (in term of processing time), once the chunks have been replicated randomly on resources with a replication factor larger than 2, it is possible to find a priority mechanism for tasks that achieves a quasi-perfect number of communications using a sophisticated matching algorithm. In the more realistic case of heterogeneous processing times, we prove using an actual trace of a MapReduce server that this priority mechanism enables to complete the Map phase with significantly fewer communications, even on realistic distributions of task durations.

7.11. On Sets Avoiding Distance 1

In a joint work with C. Bachoc, T. Bellitto and P. Moustrou [11], we consider the maximum density of sets avoiding distance 1 in $\mathbb{R}^n$. Let $||.||$ be a norm of R^n and $G_{||.||}$ be the so-called unit distance graph with the points of R^n as vertex set and for edge set, the set of pairs $\{x, y\}$ such that $||x - y|| = 1$. An independent set of $G_{||.||}$ is said to avoid distance 1.

Let $||.||_E$ denote the Euclidean norm. For $n = 2$, the chromatic number of $G_{||.||_E}$ is still wide open: it is only known that $4 \leq \chi\left(G_{||.||_E}\right) \leq 7$ (Nelson, Isbell 1950). The *measurable* chromatic number χ_m of the graph $G_{||.||}$ is the minimal number of *measurable* stable sets of $G_{||.||}$ needed to cover all its vertices. Obviously, we have $\chi\left(G_{||.||_E}\right) \leq \chi_m\left(G_{||.||_E}\right)$. For $n = 2$, $5 \leq \chi_m\left(G_{||.||_E}\right)$ (Falconer 1981).

Let $m_1\left(G_{||.||}\right)$ denote the maximum density of a measurable set avoiding distance 1. We have $\frac{1}{m_1\left(G_{||.||}\right)} \leq \chi_m\left(G_{||.||}\right)$. We study the maximum density m_1 for norms defined by polytopes: if P is a centrally symmetric polytope and x is a point of R^n, $||x||_P$ is the smallest positive real t such that $x \in tP$. Polytope norms include some usual norms such as the L^1 and L^∞ norms.

If P tiles the space by translation, then it is easy to see that $m_1\left(G_{||.||_P}\right) \geq \frac{1}{2^n}$. C. Bachoc and S. Robins conjectured that equality always holds. We show that this conjecture is true for $n = 2$ and for some polytopes in higher dimensions.

<p style="text-align: center; color: red;">**SISTM Project-Team**</p>

7. New Results

7.1. High dimensional data

Approaches Applied in Genomics Context [13]

Motivation: The association between two blocks of ?omics? data brings challenging issues in computational biology due to their size and complexity. Here, we focus on a class of multivariate statistical methods called partial least square (PLS). Sparse version of PLS (sPLS) operates integration of two datasets while simultaneously selecting the contributing variables. However, these methods do not take into account the important structural or group effects due to the relationship between markers among biological pathways. Hence, considering the predefined groups of markers (e.g. genesets), this could improve the relevance and the efficacy of the PLS approach. Results: We propose two PLS extensions called group PLS (gPLS) and sparse gPLS (sgPLS). Our algorithm enables to study the relationship between two different types of omics data (e.g. SNP and gene expression) or between an omics dataset and multivariate phenotypes (e.g. cytokine secretion). We demonstrate the good performance of gPLS and sgPLS compared with the sPLS in the context of grouped data. Then, these methods are compared through an HIV therapeutic vaccine trial. Our approaches provide parsimonious models to reveal the relationship between gene abundance and the immunological response to the vaccine.

Combining clustering of variables and feature selection using random forests: the CoV/VSURF procedure [26]

High-dimensional data classification is a challenging problem. A standard approach to tackle this problem is to perform variables selection, e.g. using step-wise or LASSO procedures. Another standard way is to perform dimension reduction, e.g. by Principal Component Analysis or Partial Least Square procedures. The approach proposed in this paper combines both dimension reduction and variables selection. First, a procedure of clustering of variables is used to built groups of correlated variables in order to reduce the redundancy of information. This dimension reduction step relies on the R package ClustOfVar which can deal with both numerical and categorical variables. Secondly, the most relevant synthetic variables (which are numerical variables summarizing the groups obtained in the first step) are selected with a procedure of variable selection using random forests, implemented in the R package VSURF. Numerical performances of the proposed methodology called CoV/VSURF are compared with direct applications of VSURF or random forests on the original p variables. Improvements obtained with the CoV/VSURF procedure are illustrated on two simulated mixed datasets (cases n>p and n<p)

Arbres CART et Forêts aléatoires, Importance et sélection de variables [27]

Two algorithms proposed by Leo Breiman : CART trees (Classification And Regression Trees for) introduced in the first half of the 80s and random forests emerged, meanwhile, in the early 2000s, are the subject of this article. The goal is to provide each of the topics, a presentation, a theoretical guarantee, an example and some variants and extensions. After a preamble, introduction recalls objectives of classification and regression problems before retracing some predecessors of the Random Forests. Then, a section is devoted to CART trees then random forests are presented. Then, a variable selection procedure based on permutation variable importance is proposed. Finally the adaptation of random forests to the Big Data context is sketched.

Comments on: " A Random Forest Guided Tour " [8]
This paper is a comment on the survey paper by Biau and Scornet (2016) about random forests. We focus on the problem of quantifying the impact of each ingredient of random forests on their performance. We show that such a quantification is possible for a simple pure forest , leading to conclusions that could apply more generally. Then, we consider " hold-out " random forests, which are a good middle point between " toy " pure forests and Breiman's original random forests.

Targeting HIV-1 Env gp140 to LOX-1 Elicits Immune Responses in Rhesus Macaques. [18]
Improved antigenicity against HIV-1 envelope (Env) protein is needed to elicit vaccine-induced protective immunity in humans. Here we describe the first tests in non-human primates (NHPs) of Env gp140 protein fused to a humanized anti-LOX-1 recombinant antibody for delivering Env directly to LOX-1-bearing antigen presenting cells, especially dendritic cells (DC). These data, as well as the safety of this protein vaccine, justify further exploration of this DC-targeting vaccine approach for protective immunity against HIV-1.

Significant changes in HIV-1 Capsid stability induced by common CTL-driven viral sequence mutations. [46]
HIV-1-infected individuals with protective HLA class I alleles exhibit better control of viremia and slower disease progression. Virus control in these individuals has been associated with strong and potent HIV-1-specific cytotoxic-T-lymphocyte (CTL) responses restricted by protective HLA alleles, but control of viremia also occurs in the presence of selected CTL escape mutations. Taken together, these data demonstrate that CTL-driven escape mutations within p24 Gag restricted by protective HLA class I alleles have a significant impact on capsid stability that might contribute to the persistent control of viral replication observed despite viral escape from CTL responses.

Optimization and evaluation of luminex performance with supernatants of Peripheral Blood Mononuclear Cell culture. [48]
The Luminex bead-based multiplex assay is useful for quantifying immune mediators such as cytokines and chemokines. Cross-comparisons of reagents for this technique from different suppliers have already been performed using serum or plasma but rarely with supernatants collected from antigen-stimulated peripheral blood mononuclear cells (PBMC). Here, we first describe an optimization protocol for cell culture including quantity of cells and culture duration to obtain reproducible cytokine and chemokine quantifications. Then, we compared three different Luminex kit suppliers.

7.2. Modeling biomarkers and Mecanistic modeling

- **Dynamic models for estimating the effect of HAART on CD4 in observational studies: Application to the Aquitaine Cohort and the Swiss HIV Cohort Study. [15]**
  Highly active antiretroviral therapy (HAART) has proved efficient in increasing CD4 counts in many randomized clinical trials. Because randomized trials have some limitations (e.g., short duration, highly selected subjects), it is interesting to assess the effect of treatments using observational studies. This is challenging because treatment is started preferentially in subjects with severe conditions. This general problem had been treated using Marginal Structural Models (MSM) relying on the counterfactual formulation. Another approach to causality is based on dynamical models. We present three discrete-time dynamic models based on linear increments models (LIM): the first one based on one difference equation for CD4 counts, the second with an equilibrium point, and the third based on a system of two difference equations, which allows jointly modeling CD4 counts and viral load. We also consider continuous-time models based on ordinary differential equations with non-linear mixed effects (ODE-NLME). These mechanistic models allow incorporating biological knowledge when available, which leads to increased statistical evidence for detecting treatment effect. Because inference in ODE-NLME is numerically challenging and requires specific methods and softwares, LIM are a valuable intermediary option in terms of consistency, precision, and complexity. We compare the different approaches in simulation and in illustration on the ANRS CO3 Aquitaine Cohort and the Swiss HIV Cohort Study.

- **Use of dynamical models for treatment optimization in HIV infected patients : a sequential Bayesian analysis approach. [15]**
  The use of dynamic mechanistic models based on ordinary differential equations (ODE) has greatly improved the knowledge of the dynamics of HIV and of the immune system. Their flexibility for fitting data and prediction abilities make them a good tool for optimization of the design delivery and efficacy of new intervention in the HIV field. We present the problem of inference in ODE models with mixed effects on parameters. We introduce a Bayesian estimation procedure based on the maximization of the penalized likelihood and a normal approximation of posteriors, which is implemented in the NIMROD software. We investigate the impact of pooling different data by using a sequential Bayesian analysis (SBA), which uses posteriors of a previous study as new priors. We show that the normal approximation of the posteriors, which constrains the shape of new priors, leads to gains in accuracy of estimation while reducing computation times. The illustration is from two clinical trials of combination of antiretroviral therapies (cART): ALBI ANRS 070 and PUZZLE ANRS 104. This paper reproduces some unpublished work from my PhD thesis. It is an extension of my oral presentation on the same topic at the 47th Journées de Statistique organized by the French Statistical Society (SFdS) in Lille, France, May 2015, when being awarded the Marie-Jeanne Laurent-Duhamel prize.

- **Surveillance of $\gamma\delta$T Cells Predicts Cytomegalovirus Infection Resolution in Kidney Transplants. [11]**
  Cytomegalovirus (CMV) infection in solid-organ transplantation is associated with increased morbidity and mortality, particularly if a CMV mutant strain with antiviral resistance emerges. Monitoring CMV specific T cell response could provide relevant information for patient care. We assessed if V delta 2 neg gamma delta T cell kinetics in peripheral blood predict CMV infection resolution and emergence of a mutant strain in high risk recipients of kidney transplants, including 168 seronegative recipients receiving organs from seropositive donors and 104 seropositive recipients receiving antithymocyte globulins (R+/ATG). In conclusion, longitudinal surveillance of V delta 2 neg gamma delta T cells in recipients of kidney transplants may predict CMV infection resolution and antiviral drug resistance.

- **Early CD4+ T Cell Responses Are Associated with Subsequent CD8+ T Cell Responses to an rAd5-Based Prophylactic Prime-Boost HIV Vaccine Strategy. [12]**
  Initial evaluation of a candidate vaccine against HIV includes an assessment of the vaccine's ability to generate immune responses. However, the dynamics of vaccine-induced immune responses are unclear. We hypothesized that the IFN-gamma producing cytotoxic CD8+ T cell responses could be predicted by early IL-2 producing CD4+ helper T cell responses, and we evaluated this hypothesis using data from a phase I/II prophylactic HIV vaccine trial. The objective was to assess the dynamics after vaccination with a recombinant adenoviral serotype 5 (rAd5) HIV vaccine. Regression models confirmed this relationship with a significant association between the two markers. These results suggest an early and leading role of CD4+ T cells in the cellular response to the rAd5-rAd5 vaccine and in particular the stimulation of cytotoxic CD8+ T cell responses. These results could inform better timing of CD4+ T cell measurements in future clinical trials.

- **Reference curves for CD4 T-cell count response to combination antiretroviral therapy in HIV-1-infected treatment-naïve patients. [29]**
  The aim of this work was to provide a reference for the CD4 T-cell count response in the early months after the initiation of combination antiretroviral therapy (cART) in HIV-1-infected patients. All patients in the Collaboration of Observational HIV Epidemiological Research Europe (COHERE) cohort who were aged > 18 years and started cART for the first time between 1 January 2005 and 1 January 2010 and who had at least one available measurement of CD4 count and a viral load < 50 HIV-1 RNA copies/mL at 6 months (+- 3 months) after cART initiation were included in the study. Unadjusted and adjusted references curves and predictions were obtained using quantile regressions. Reference curves aid the evaluation of the immune response early after antiretroviral therapy initiation that leads to viral control.

- **Repeated Cycles of Recombinant Human Interleukin 7 in HIV-Infected Patients With Low CD4 T-Cell Reconstitution on Antiretroviral Therapy: Results of 2 Phase II Multicenter Studies. [17]**.
  Phase I/II studies in human immunodeficiency virus (HIV) infected patients receiving antiretroviral therapy have shown that a single cycle of 3 weekly subcutaneous injections of recombinant human interleukin 7 (r-hIL-7) is safe and improves immune CD4 T-cell restoration. Herein, we report data from 2 phase II trials evaluating the effect of repeated cycles of r-hIL-7 (20 microg/kg) with the objective of restoring a sustained CD4 T-cell count >500 cells/microL. INSPIRE 2 was a single-arm trial conducted in the United States and Canada. INSPIRE 3 was a 2 arm trial with 3:1 randomization to r-hIL-7 versus control conducted in Europe and South Africa. Participants with plasma HIV RNA levels <50 copies/mL during antiretroviral therapy and with CD4 T-cell counts between 101 and 400 cells/microL were eligible. A repeat cycle was administered when CD4 T-cell counts fell to <550 cells/microL. A total of 107 patients were treated and received 1 (n = 107), 2 (n = 74), 3 (n = 14), or 4 (n = 1) r-hIL-7 cycles during a median follow-up of 23 months. r-hIL-7 was well tolerated. Four grade 4 events were observed, including 1 case of asymptomatic alanine aminotransferase elevation. After the second cycle, anti-r-hIL-7 binding antibodies developed in 82% and 77% of patients in INSPIRE 2 and 3, respectively (neutralizing antibodies in 38% and 37%), without impact on the CD4 T-cell response. Half of the patients spent >63% of their follow-up time with a CD4 T-cell count >500 cells/microL. CONCLUSIONS: Repeated cycles of r-hIL-7 were well tolerated and achieved sustained CD4 T-cell restoration to >500 cells/microL in the majority of study participants.

7.3. Implication in analysis of results from Clinical trials and cohorts

- **Superior efficacy of an HIV vaccine combined with ARV prevention in SHIV challenged non-human primates. [38]**
  Although vaccines and antiretroviral (ARV) prevention have demonstrated partial success against human immunodeficiency virus (HIV) infection in clinical trials, their combined introduction could provide more potent protection. Furthermore, combination approaches could ameliorate the potential increased risk of infection following vaccination in the absence of protective immunity. We used a nonhuman primate model to determine potential interactions of combining a partially effective ARV microbicide with an envelope-based vaccine. These important findings suggest that combined implementation of new biomedical prevention strategies may provide significant gains in HIV prevention.

- **A Method to Estimate the Size and Characteristics of HIV-positive Populations Using an Individual-based Stochastic Simulation Model. [14]**
  It is important not only to collect epidemiologic data on HIV but to also fully utilize such information to understand the epidemic over time and to help inform and monitor the impact of policies and interventions. We describe and apply a novel method to estimate the size and characteristics of HIV-positive populations. In the pseudo-epidemic example, HIV estimates have narrower plausibility ranges and are closer to the true number, the greater the data availability to calibrate the model. We demonstrate that our method can be applied to settings with less data, however plausibility ranges for estimates will be wider to reflect greater uncertainty of the data used to fit the model.

- **Immunologic response in treatment-naïve HIV-2-infected patients: the IeDEA West Africa cohort. [9]**
  Response to antiretroviral therapy (ART) among individuals infected with HIV-2 is poorly described. We compared the immunological response among patients treated with three nucleoside reverse-transcriptase inhibitors (NRTIs) to boosted protease inhibitor (PI) and unboosted PI-based regimens in West Africa. In this observational study using African data, boosted PI-containing regimens had better immunological response compared to triple NRTI combinations and unboosted PI-based regimens at 12 months. A randomized clinical trial is still required to determine the best initial regimen for treating HIV-2 infected patients.

- **Intrinsic defect in keratinocyte function leads to inflammation in Hidradenitis suppurativa. [10]**
  Hidradenitis suppurativa (HS) is a chronic, inflammatory, debilitating, follicular disease of the skin. Despite a high prevalence in the general population, the physiopathology of HS remains poorly understood. The use of antibiotics and immunosuppressive agents for therapy suggests a deregulated immune response to microflora. These findings point out a functional defect of keratinocytes in HS leading to a balance prone to inflammatory responses. This is likely to favor a permissive environment for bacterial infections and chronic inflammation characterizing clinical outcomes in patients with HS.

- **Uptake of Home-Based HIV Testing, Linkage to Care, and Community Attitudes about ART in Rural KwaZulu-Natal, South Africa: Descriptive Results from the First Phase of the ANRS 12249 TasP Cluster-Randomised Trial. [34]**
  The 2015 WHO recommendation of antiretroviral therapy (ART) for all immediately following HIV diagnosis is partially based on the anticipated impact on HIV incidence in the surrounding population. We investigated this approach in a cluster-randomised trial in a high HIV prevalence setting in rural KwaZulu-Natal. We present findings from the first phase of the trial and report on uptake of home-based HIV testing, linkage to care, uptake of ART, and community attitudes about ART. Home-based HIV testing was well received in this rural population, although men were less easily contactable at home; immediate ART was acceptable, with good viral suppression and retention. However, only about half of HIV-positive people accessed care within 6 mo of being identified, with nearly two-thirds accessing care by 12 mo. The observed delay in linkage to care would limit the individual and public health ART benefits of universal testing and treatment in this population.

7.4. Conferences

Members of the team were involved in more than 20 talks during conferences and colloquium. In particular, [20], [21], [23], [19], [22], [24] and [25] have proceedings.

6. New Results

6.1. Automatic OpenCL Task Adaptation for Heterogeneous Architectures

OpenCL defines a common parallel programming language for all devices, although writing tasks adapted to the devices, managing communication and load-balancing issues are left to the programmer. In this work [11], we propose a novel automatic compiler and runtime technique to execute single OpenCL kernels on heterogeneous multi-device architectures. Our technique splits computation and data automatically across the computing devices. The technique proposed is completely transparent to the user, does not require off-line training or a performance model. It handles communications and load-balancing issues, resulting from hardware heterogeneity, load imbalance within the kernel itself and load variations between repeated executions of the kernel, in an iterative computation. We present our results on benchmarks and on an N-body application over two platforms, a 12-core CPU with two different GPUs and a 16-core CPU with three homogeneous GPUs.

6.2. Fast Forward Error Correction Codes

Erroc Correction Codes are essential for preserving data integrity in communications. These algorithms find errors due to noise in transmissions and correct these errors with a high probability. Several algorithms are used, with different capacities in term of correction and most of them are implemented in cell phones or satellites as ASICS. The need to handle many different usages, different contexts of use pushes towards software solutions. A larger spectrum of algorithms can be explored, in order to meet the expectations in terms of performance, power consumption and error correcting power. These new algorithms, for the 5G for instance, can then be either implemented in software (for large antenna for instance) or in hardware. In both case, software simulation is necessary in order to evaluate the properties of the new algorithms. We developed in collaboration with IMS new versions of algorithms and a new software, AFF3CT http://aff3ct. github.io/index.html, that allows the exploration of many different algorithmic variants and their evaluation. Two conference papers have been published on these new results [7][6].

6.3. Resource aggregation for task-based Cholesky Factorization

Hybrid computing platforms are now commonplace, featuring a large number of CPU cores and accelerators. This trend makes balancing computations between these heterogeneous resources performance critical. In a recent paper [8] we propose aggregating several CPU cores in order to execute larger parallel tasks and thus improve the load balance between CPUs and accelerators. Additionally, we present our approach to exploit internal parallelism within tasks. This is done by combining two runtime systems: one runtime system to handle the task graph and another one to manage the internal parallelism. We demonstrate the relevance of our approach in the context of the dense Cholesky factorization kernel implemented on top of the StarPU task-based runtime system.We present experimental results showing that our solution outperforms state of the art implementations. In addition, we realized an extended version of this paper submitted for review to the Parallel Computing journal special issue for HCW and HeteroPar 2016 workshops. In this new paper [19] we provide additional details on our contribution and propose a brand new study on the recent Intel Xeon Phi Knights Landing (KNL) where we show that we are able to outperform existing state of the art implementations on this platform thanks to our proposed technique.

6.4. Scheduling of Linear Algebra Kernels on Multiple Heterogeneous Resources

In this paper [5], we consider task-based dense linear algebra applications on a single heterogeneous node which contains regular CPU cores and a set of GPU devices. Efficient scheduling strategies are crucial in this context in order to achieve good and portable performance. HeteroPrio, a resource-centric dynamic scheduling strategy has been introduced in a previous work and evaluated for the special case of nodes with exactly two types of resources. However, this restriction can be limiting, for example on nodes with several types of accelerators, but not only this. Indeed, an interesting approach to increase resource usage is to group several CPU cores together, which allows to use intra-task parallelism. We propose a generalization of HeteroPrio to the case with several classes of heterogeneous workers. We provide extensive evaluation of this algorithm with Cholesky factorization, both through simulation and actual execution, compared with HEFT-based scheduling strategy, the state of the art dynamic scheduling strategy for heterogeneous systems. Experimental evaluation shows that our approach is efficient even for highly heterogeneous configurations and significantly outperforms HEFT-based strategy.

6.5. Analyzing Dynamic Task-Based Applications on Hybrid Platforms: An Agile Scripting Approach

In this paper [10], we present visual analysis techniques to evaluate the performance of HPC task-based applications on hybrid architectures. Our approach is based on composing modern data analysis tools (pjdump, R, ggplot2, plotly), enabling an agile and flexible scripting framework with minor development cost. We validate our proposal by analyzing traces from the full-fledged implementation of the Cholesky decomposition available in the MORSE library running on a hybrid (CPU/GPU) platform. The analysis compares two different workloads and three different task schedulers from the StarPU runtime system. Our analysis based on composite views allows to identify allocation mistakes, priority problems in scheduling decisions, GPU tasks anomalies causing bad performance, and critical path issues.

6.6. Distributed StarPU Scalability on Heterogeneous Platforms

The emergence of accelerators as standard computing resources on supercomput- ers and the subsequent architectural complexity increase revived the need for high-level parallel programming paradigms. Sequential task-based programming model has been shown to efficiently meet this challenge on a single multicore node possibly enhanced with accelerators, which moti- vated its support in the OpenMP 4.0 standard. In this paper, we show that this paradigm can also be employed to achieve high performance on modern supercomputers composed of multiple such nodes, with extremely limited changes in the user code. To prove this claim, we have extended the StarPU runtime system with an advanced inter-node data management layer that supports this model by posting communications automatically [16]. We illustrate our discussion with the task- based tile Cholesky algorithm that we implemented on top of this new runtime system layer. We show that it allows for very high productivity while achieving a performance competitive with both the pure Message Passing Interface (MPI)-based ScaLAPACK Cholesky reference implementation and the DPLASMA Cholesky code, which implements another (non sequential) task-based programming paradigm.

6.7. Controlling the Memory Subscription of Distributed Applications with a Task-Based Runtime System

The ever-increasing supercomputer architectural complexity emphasizes the need for high-level parallel pro- gramming paradigms. Among such paradigms, task-based programming manages to abstract away much of the archi- tecture complexity while efficiently meeting the performance challenge, even at large scale. Dynamic run-time systems are typically used to execute task-based applications, to schedule computation resource usage and memory allocations. While computation scheduling has been well studied, the dynamic management of memory resource subscription inside such run- times has however been little explored. This paper [12] studies

the cooperation between a task-based distributed application code and a run-time system engine to control the memory subscription levels throughout the execution. We show that the task paradigm allows to control the memory footprint of the application by throttling the task submission flow rate, striking a compromise between the performance benefits of anticipative task submission and the resulting memory consumption. We illustrate the benefits of our contribution on a compressed dense linear algebra distributed application.

6.8. StarPU Interfacing with GASPI/GPI2

A version of the distributed dependence support of StarPU has been ported by Corentin Salingue, under the supervision of Olivier Aumage on the high performance GASPI/GPI2 networking layer developed by the Fraunhofer institute in Germany. The GPI2 framework offers a lightweight communication interface specifically designed for thread enabled HPC applications. This work has been conducted as part of the H2020 INTERTWinE european project.

6.9. A Stencil DSEL for Single Code Accelerated Computing with SYCL

Stencil kernels arise in many scientific codes as the result from dis- cretizing natural, continuous phenomenons. Many research works have designed stencil frameworks to help programmer optimize stencil kernels for performance, and to target CPUs or accelerators. However, existing stencil kernels, either library-based or language-based necessitate to write distinct source codes for accelerated ker- nels and for the core application, or to resort to specific keywords, pragmas or language extensions. SYCL is a C++ based approach designed by the Khronos Group to program the core application as well as the application kernels with a single unified, C++ compliant source code. A SYCL application can then be linked with a CPU-only runtime library or processed by a SYCL-enabled compiler to automatically build an OpenCL accelerated application. Our contribution [13] is a stencil domain specific embedded language (DSEL) which leverage SYCL together with expression template techniques to implement statically optimized stencil applications able to run on platforms equipped with OpenCL devices, while preserving the single source benefits from SYCL.

6.10. Bridging the gap between OpenMP 4.0 and native runtime systems for the fast multipole method

With the advent of complex modern architectures, the low-level paradigms long considered sufficient to build High Performance Computing (HPC) numerical codes have met their limits. Achieving efficiency, ensuring portability, while preserving programming tractability on such hardware prompted the HPC community to design new, higher level paradigms. The successful ports of fully-featured numerical libraries on several recent runtime system proposals have shown, indeed, the benefit of task-based parallelism models in terms of performance portability on complex platforms. However, the common weakness of these projects is to deeply tie applications to specific expert-only runtime system APIs. The OpenMP specification, which aims at providing a common parallel programming means for shared-memory platforms, appears as a good candidate to address this issue thanks to the latest task-based constructs introduced as part of its revision 4.0. The goal of this paper [15] is to assess the effectiveness and limits of this support for designing a high-performance numerical library. We illustrate our discussion with the ScalFMM library, which implements state-of-the-art fast multipole method (FMM) algorithms, that we have deeply re-designed with respect to the most advanced features provided by OpenMP 4. We show that OpenMP 4 allows for significant performance improvements over previous OpenMP revisions on recent multicore processors. We furthermore propose extensions to the OpenMP 4 standard and show how they can enhance FMM performance. To assess our statement, we have implemented this support within the Klang-OMP source-to-source compiler that translates OpenMP directives into calls to the StarPU task-based runtime system. This study shows that we can take advantage of the advanced capabilities of a fully-featured runtime system without resorting to a specific, native runtime port, hence bridging the gap between the OpenMP standard and the very high performance that was so far reserved to expert-only runtime system APIs.

6.11. Hierarchical Tasks

Modern computing platforms are heterogeneous and the load balancing is more complex to reach high performance. We decided to deal with the granularity problem in the context of task paralleliism and in a dynamic way through the implementation of hierarchical tasks in StarPU runtime. The idea is to give the runtime the ability to control tasks submission in order to choose the good granularity at the right moment. The application describes a control graph and the runtime generates the computation tasks graph on-the-fly according to the state of the machine (available computing resources, memory consumption, ...). As a consequence the runtime is able to limit the size of the computation tasks graph without loosing parallelism. Some experiments have been done on a Cholesky application and in the qr-mumps software and show that the work of an application programmer can be alleviated and the granularity choice could be easily delegated to the task based runtime.

6.12. Software-Hardware Exploration for Read-Only Data

We have proposed a new way of managing the cache by exploiting the difference of behavior in the memory system between read-only data and read-write data. A division of the existing cache-based memory hierarchy is proposed in order to create a dedicated data path for read-only data. This proposition is similar to the existing separation at the L1-level between instruction and data caches. In order to justify this approach, an analysis performed on a set of benchmarks shows that read-only data count for significant part of the working set and are less reused than read-write data. A transparent solution is proposed based on specific compilation support to separate automatically the memory accesses of read-only data at L1-level. This organization exploits the properties of the different sub- workloads in order to increase the overall data locality and data reuse. Simulated in a multicore environment, the evaluation of the new memory organization shows reduction of L1 misses up to 28.5%. Moreover, the messages issued on the interconnection network can be reduced up to 14.7% without any penalty on the performance.

Besides the reduced miss-rate allows maintaining performance with smaller cache size on the read-write path while the properties of the read- only part can benefit of a simplified cache implementation despite a shared multicore access [1].

<p align="center" style="color:red">TADAAM Team</p>

7. New Results

7.1. Network Modeling

NETLOC (see Section 6.1) is a tool in HWLOC to discover the network topology. Our first work with NETLOC was to redesign it to be more efficient and more adapted to the needs. The code was cleaned and some dependencies were removed. We have added a display tool, that is able to show a network topology in a web browser where a user can interact with. It ran on one of the largest European supercomputer (the TGCC/Genci CURIE machine) and successfully modeled its 5200 nodes and its interconnection network (more than 800 switches).

Moreover, it is now possible to interact with Scotch from netloc. The first feature is to export a network topology, or even the current available topology given by the resource manager, into a SCOTCH architecture. Conversely, we can use SCOTCH tools in NETLOC for building a process mapping based on resources found by NETLOC and a process graph describing communications between processes. Tests conducted on a stencil mini-app have shown that the benefits are real and still needs more work.

7.2. Communication and computation overlap

To amortize the cost of communication in HPC application, programmers want to overlap communications with computation. To do so, they assume non-blocking MPI communications will progress in background. NewMadeleine, our communication library, is actually able to make communication progress in background so as to actually have overlap happen. However, not all MPI implementations are able to overlap communication and computation.

We have proposed [8] a benchmark to measure what really happens when trying to overlap non-blocking point-to-point communications with computation. The benchmark measures how much overlap happen in various cases: sender-side, receiver-side, datatypes likely to be offloaded onto NIC or not, multi-threaded computation, multi-threaded communication or not. We have benchmarked a wide panel of MPI libraries and hardware platforms, and thanks to low-level traces, explained the results.

7.3. Topology Aware Performance Monitoring

A tool has been developed to abstract performance metrics and map them onto the HWLOC (see Section 6.6) topology model of the system. During the year 2016, the tool has been entirely rewritten to release a more meaningful and stable programming abstraction, with off the shelf performance abstraction plugins and raw performance acquisition plugin [16]. A special effort has been carried out on output presentation by extending lstopo tool from hwloc into a library embedded in the monitoring tool to display performance metrics on the system topology. Another backend using R has also been developed for the purpose of post-mortem analysis and model extraction from abstract metrics of the topology.

7.4. Locality Aware Roofline Model

The years 2016 marked the achievement of our extension of the famous Cache Aware Roofline Model(CARM) and the associate tool. The latter model targets deep plateform and application analysis on multicore processors. Its model consist into a two-dimensions plane bound by several machine ceils and representative of scientific application workloads. Our extension validate the use of the CARM on emerging processors with heterogeneous memory subsystem, and extend the CARM methodology to encompass interconnection network, thus, enabling full modeling of shared memory systems [17]. This work is a collaboration with the INESC-ID research center under the NESUS project.

7.5. Performance Analysis of Electromagnetic Field Application on Large SMP Node

In the scope of the COLOC project we worked on understanding scalability issues of the efield application on a large shared memory system. Our analysis with above mentionned tools highlighted a potential bandwidth bottleneck. This problem can usually be tackled by the mean of threads and data mapping on respectively the machine cores and the memories. Unfortunately, those techniques can't be applied with this (closed source) application since the system does not allow to monitor memory accesses and traffic on the system.

7.6. Structural Modeling of Heterogeneous Memory Architectures

HWLOC (see Section 6.6) is the de facto standard tool for gathering information of parallel platform topologies. The advent of new memory architecture, with high-bandwidth and/or non-volatile memories cause the memory management subsytem complexity to increase. Indeed, besides taking care of allocating data buffers locally, developers also have to choose between different local memories with different performance and persistence characteristics. Moreover, the operating systems still cannot expose the full details about these technologies to applications. We modified the HWLOC tool to cope with these new needs in collaboration with Intel. This work led to the design a new structural model for platforms with heteregeneous memories [10].

7.7. Scalable Management of Platform Topologies

HWLOC (see Section 6.6) is used for gathering the topology of thousands of nodes in large clusters. Those nodes are now growing to hundreds of cores, making the overall amount of topology information non-negligible. We designed new ways to compress topologies, either lossless or lossy, for easier transfer between compute nodes and front nodes and more compact storage and manipulation [20]. We also studied the overhead of topology discovery on the overall execution time and showed that the Linux kernel is bottleneck on large nodes. It raised the need to use exported and/or abstracted topologies to factorize this overhead [11].

7.8. MPI One-side operations

MPI one-sided operations, aka Remote Memory Access (RMA), are direct read/write memory access to a remote node. Only one node (the origin) explicitly calls MPI operations, while communication progression is implicit for the other node (the target). These operations assume that the communication library is able to make communication progress in background.

Since MadMPI, the MPI implementation of NewMadeleine (see Section 6.2), extensively uses event-driven mechanism to reach asynchronous progression, we have [24] taken advantage of this property to implement MPI RMA operations in the library. This implementation keeps the overlap properties by asynchronously handle the messages exchanged by the applications. The addition also supports MPI_THREAD_MULTIPLE, for both shared and distributed memory contexts.

7.9. Topology and affinity aware hierarchical and distributed load-balancing

The evolution of massively parallel supercomputers make palpable two issues in particular: the load imbalance and the poor management of data locality in applications. Thus, with the increase of the number of cores and the drastic decrease of amount of memory per core, the large performance needs imply to particularly take care of the load-balancing and as much as possible of the locality of data. One mean to take into account this locality issue relies on the placement of the processing entities and load balancing techniques are relevant in order to improve application performance. With large-scale platforms in mind, we developed a hierarchical and distributed algorithm which aim is to perform a topology-aware load balancing tailored for Charm++ applications. This algorithm is based on both LibTopoMap for the network awareness aspects and on Treematch to determine a relevant placement of the processing entities. We show that the proposed algorithm improves the overall execution time in both the cases of real applications and a synthetic benchmark as well. For this last experiment, we show a scalability up to one millions processing entities [12].

7.10. Topology-Aware Data Aggregation for Intensive I/O on Large-Scale Supercomputers

Reading and writing data efficiently from storage systems is critical for high performance data-centric applications. These I/O systems are being increasingly characterized by complex topologies and deeper memory hierarchies. Effective parallel I/O solutions are needed to scale applications on current and future supercomputers. Data aggregation is an efficient approach consisting of electing some processes in charge of aggregating data from a set of neighbors and writing the aggregated data into storage. Thus, the bandwidth use can be optimized while the contention is reduced. In [13], we have taken into account the network topology for mapping aggregators and we propose an optimized buffering system in order to reduce the aggregation cost. We have validated our approach using micro-benchmarks and the I/O kernel of a large-scale cosmology simulation. We have showed improvements up to 15× faster for I/O operations compared to a standard implementation of MPI I/O.

7.11. Communication monitoring in OpenMPI

Monitoring data exchanges is critical when it comes to optimize process placement in a large scale environment. We participated in adding in Open-MPI, which is one of the major MPI implementation, a fine grain, point-to-point monitoring component that keeps track of message exchanges. Unlike implementations using PMPI operations, the layer in which this monitoring acts allow us to record at a lower level the effective data communications, for example, after the covering tree has been calculated. This component has been enriched with a complete coverage of collectives, point-to-point and one-sided communications. This component also reports informations about message sizes distribution. Monitored informations can be accessed by using MPI_Tools interface, or by dumping data in files.

7.12. Process Placement with TreeMatch

We released TreeMatch ver 0.4 in August. The new feature are: a new API, the handling oversubscribing (being able to map more processes that computing resources), fast exhaustive search (for small cases), K-partitioning in case of large arity of the tree, and a set of extensive tests.

7.13. Topology Aware Resource Management

SLURM is a Resource and Job Management System, a middleware in charge of delivering computing power to applications in HPC systems. Our goal is to take in account in SLURM placement process hardware topology but application communication pattern too. We have a new [9], [19] selection option for the cons_res plugin in SLURM. In this case the usually best_fit algorithm used to choose nodes is replaced by TreeMatch, an algorithm to find the best placement among the free nodes list in light of a given application communication matrix. We plan to release this work in the next release SLURM 17.02.

Fragmentation in cluster is one of the criteria important for administrator. Indeed, the way jobs are allocated impacts the global resource usage. Usually it is observed throught utilization of a cluster for a fixed load rate, but no metrics dedicated to fragmentation exist in litterature. Hence we construct several metrics to measure it. Our goal is to study the impact of our selection algorithm on fragmentation in comparison with other.

7.14. Impact of progress threads placement for MPI Non-Blocking Collectives

MPI Non-Blocking Collectives (NBC) allow communication overlap with computation. A good overlapping ratio is obtained when computation and communication are running in parallel. To achieve this, some implementations use progress threads to manage communication tasks. These threads should be bound on different cores to maximize the overlap. Thus, we elaborate several threads placement algorithms. These algorithms have been implemented within the MPC framework, using the HWLOC software to get a global view of the machine topology. We propose [18] a thread placement algorithm taking into account the NUMA topology of the machine in order to improve the overlapping ratio of non-blocking collective communications.

7.15. Hierarchical Communication Management in MPI

MPI, in its current state provides only a very limited set of functionnalities so as to allow the programmer to effectively leverage the physical characteristics of the underlying hardaware, such as the potentially complex memory hierarchy. The MPI philosophy being to be a hardware-agnostic interface, the challenge is therefore to propose an interface extension that offers the programmer significant control over the hardawre without dwelling too much into hardware details. We seek the right level of abstraction for this interface and the goal is push this proposal to the MPI Forum. This new interface is based on the concept of communicators, expands an already existing function available in the standard and also introduces a couple of helper functions. We have prototyped and drafted our proposal for the 2017 meetings of the forum.

7.16. Fully-abstracted approach for efficient thread binding in task-based model of programming

Task-based models and runtimes are quite popular in the HPC community. They help to implement applications with a high level of abstraction while still applying different types of optimizations. An important optimization target is hardware affinity, which concerns to match application behavior (thread, communication, data) to the architecture topology (cores, caches, memory). In fact, realizing a well adapted placement of threads is a key to achieve performance and scalability, especially on NUMA-SMP machines. However, this type of optimization is difficult: architectures become increasingly complex and application behavior changes with implementations and input parameters, *e.g* problem size and number of thread. Thus, by themselves task based runtimes often deal badly with this optimization and leave a lot of fine-tuning to the user. In this work [21], [25], [14], we propose a fully automatic, abstracted and portable affinity module. It produces and implements an optimized affinity strategy that combines knowledge about application characteristics and the architecture's topology. Implemented in the backend of our task-based runtime ORWL, our approach was used to enhance the performance and the scalability of several unmodified ORWL-coded applications: matrix multiplication, a 2D stencil (Livermore Kernel 23), and a video tracking real world application. On two SGI SMP machines with quite different hardware characteristics, our tests show spectacular performance improvements for this unmodified application code due to a dramatic decrease of cache misses. A comparison to reference implementations using OpenMP confirms this performance gain of almost one order of magnitude.

7.17. Multi-criteria graph partitioning for multi-physics simulations load balancing

A new set of algorithms has been designed to compute multi-criteria static mappings for the load balancing of multi-phisics simulations. The multi-criteria graph partitioning is known to be NP-hard, and there exist very few multi-criteria graph partitioners. Moreover, they focus on the edge-cut minimization instead of enforcing load balance. In practice, this strategy often leads to very unbalanced partitions, which are not useful for multi-physics simulations.

We have designed algorithms that focus on balancing several criteria at the same time to ensure that our results always match all balance criteria. We have implemented a prototype in Python to test these different heuristics. One of them, called PIERE, obtained good results [15], in term of balance as well as communication costs. PIERE uses the classic multilevel framework, but implements a new initial partitioning algorithm, which allows to find a balanced partition of the graph. The partition is then refined by local optimization heuristics that ensure the balance is kept for all criteria. This allow us to return a partition respecting the balance constraints. In [15], we compare against well-known partitioners that are SCOTCH and METIS, and highlight that, for a small mesh, the results exhibit a high discrepancy: each tool lacks of robustness.

PIERE outperformed the existing software METIS in our test cases, but there is room for improvement. We also verified the superiority of the hypergraph model over the graph model used by most partitioners. Meanwhile, we studied the source code of well known partitioners, namely METIS and SCOTCH, and we have identified a lot of algorithmic choices and internal parameters that are not described in their documentations. Carefully analyzing them helps us to clearly understand the differences of the different algorithms.

7.18. Scotch

In order to prepare for the inclusion of multi-criteria graph partitioning algorithms in SCOTCH, in the context of the PhD thesis of Rémi Barat, a new branch has been created in the SCOTCH repository. This new branch, labeled as 6.1, is the basis for the next main release of SCOTCH. The sequential graph structure has been adapted to handle graphs with multiple loads per vertex, and all the related algorithms have been adapted to take into account multiple vertex loads. This resulted in minimal updates in the interface of Scotch, with ful ascending compatibility. All of these modifications have been performed so as not to slow down significantly the algorithms in the most common case of graphs with single vertex loads.

7.19. PaMPA

Parallel remeshing has been improved. PaMPA coupled with Mmg (v5) remeshed a tetrahedral mesh from 43Melements to more than 1Belements on 280 Broadwell processors in 20 minutes. The resulted mesh, used by CERFACS, permitted one of the most finest simulation computed with LES (Large Eddy Simulation) on combustion.

The scalability of PT-SCOTCH scalability has been tested on the Curie cluster and compared to that of PARMETIS. These tests used DARI resources.

7.20. Originality of software works

Most judges have very little, if not none, knowledge on software developement. This results in misconceptions and mistakes regarding the application of copyright/author right (*droit d'auteur*) in court cases related to software. More generally, the concept of originality is misunderstood. While this criterion is meant in theory to separate works of the mind that are personal to an author (e.g., literary works), from creations of form that cannot, by nature, reflect the personality of their creator (e.g. mathematical tables), it is often used to qualify the degree of similarity between two different works, in the context of plagiarism. Also, the distinction between the realm of programs, that is, works of the mind, and that of algorithms, is not mastered. Algorithms belong to the *fonds commun*, a French term that has no equivalent in English and might be translated as "common pool". In order to help judges and lawmakers in understanding these notions, and articulate them, we have proposed a methodology for ruling software disputes. This methodology is solely based on the study of similarities in software code, since author right exclusively pertains to the level of the form [22].