RESEARCH CENTRE

Bordeaux - Sud-Ouest

IN PARTNERSHIP WITH:

Université de Bordeaux, CNRS, Institut Polytechnique de Bordeaux

2021
ACTIVITY REPORT

Project-Team

HIEPACS

High-End Parallel Algorithms for Challenging Numerical Simulations

IN COLLABORATION WITH: Laboratoire Bordelais de Recherche en Informatique (LaBRI)

DOMAIN

Networks, Systems and Services, Distributed Computing

THEME

Distributed and High Performance Computing

Contents

Project-Team HIEPACS

Creation of the Project-Team: 2010 January 01

Keywords

Computer sciences and digital sciences

A1.1.4. – High performance computing

A1.1.5. – Exascale

A1.1.9. – Fault tolerant systems

A6.2.5. – Numerical Linear Algebra

A6.2.7. – High performance computing

A7.1. – Algorithms

A8.1. – Discrete mathematics, combinatorics

A8.2. – Optimization

A9.2. – Machine learning

A9.7. – AI algorithmics

Other research topics and application domains

B3.3.1. – Earth and subsoil

B3.6. – Ecology

B3.6.1. – Biodiversity

B4.2.2. – Fusion

B5.5. – Materials

B9.5.1. – Computer science

B9.5.2. – Mathematics

B9.5.4. – Chemistry

1 Team members, visitors, external collaborators

Research Scientists

- Luc Giraud [Team leader, Inria, Senior Researcher, HDR]
- Emmanuel Agullo [Inria, Researcher]
- Olivier Beaumont [Inria, Senior Researcher, HDR]
- Olivier Coulaud [Inria, Senior Researcher, HDR]
- Lionel Eyraud-Dubois [Inria, Researcher]

Faculty Members

- Aurélien Esnard [Univ de Bordeaux, Associate Professor]
- Mathieu Faverge [Institut National Polytechnique de Bordeaux, Associate Professor]
- Abdou Guermouche [Univ de Bordeaux, Associate Professor, HDR]
- Pierre Ramet [Univ de Bordeaux, Associate Professor, HDR]

Post-Doctoral Fellows

- Pierre Esterie [Univ de Bordeaux, until Oct 2021]
- Mohammad Issa [Inria, until Jun 2021]

PhD Students

- Tobias Castanet [Univ de Bordeaux, until Oct 2021]
- Jean Francois David [Inria, from Sep 2021]
- Marek Felsoci [Inria]
- Martina Iannacito [Inria]
- Esragul Korkmaz [Inria]
- Aboul Karim Mohamed El Maarouf [IFPEN]
- Romain Peressoni [Univ de Bordeaux]
- Clément Richefort [CEA, Nov. 2021 to Jan. 2022]
- Alena Shilova [Inria]
- Nicolas Venkovic [CERFACS, until Sep 2021]
- Mathieu Verite [Inria]
- Yanfei Xiang [Inria]
- Xunyi Zhao [Inria, from Sep 2021]

Technical Staff

- Tony Delarue [Inria, Engineer]
- Rémi Duclos [Inria, Engineer, from Mar 2021]
- Pierre Esterie [Inria, Engineer, from Nov 2021]
- Gilles Marait [Inria, Engineer]
- Matthieu Simonin [Inria, Engineer, until Oct 2021]

Interns and Apprentices

- Leo Bertheas [Inria, from Jun 2021 until Sep 2021]
- Nolan Bredel [Inria, from May 2021 until Sep 2021]
- Kea Horvath [Inria, from May 2021 until Jun 2021]
- Tom Moënne Loccoz [Inria, from May 2021 until Sep 2021]

Administrative Assistant

- Chrystel Plumejeau [Inria]

Visiting Scientists

- Cristina Boeres [Universidade Federal Fluminense, From Nov. 2021 to Jan. 2022]
- Jesus Camara [Université de Murcia, until May 2021]
- Rebello Vinod [Universidade Federal Fluminense, From Nov. 2021 to Jan. 2022]

External Collaborators

- Jean Rene Poirier [INP Toulouse, HDR]
- Guillaume Sylvand [Airbus]

2 Overall objectives

2.1 Introduction

Over the last few decades, there have been innumerable science, engineering and societal breakthroughs enabled by the development of High Performance Computing (HPC) applications, algorithms and architectures. These powerful tools have provided researchers with the ability to computationally find efficient solutions for some of the most challenging scientific questions and problems in medicine and biology, climatology, nanotechnology, energy and environment. It is admitted today that *numerical simulation is the third pillar for the development of scientific discovery at the same level as theory and experimentation.* Numerous reports and papers also confirm that very high performance simulation will open new opportunities not only for research but also for a large spectrum of industrial sectors.

An important force which has continued to drive HPC has been to focus on frontier milestones which consist in technical goals that symbolize the next stage of progress in the field. In the 1990s, the HPC community sought to achieve computing at a teraflop rate and and exascale machines are now expected in the next few months/years.

For application codes to sustain petaflops and more in the next few years, hundreds of thousands of processor cores or more are needed, regardless of processor technology. Currently, a few HPC simulation codes easily scale to this regime and major algorithms and codes development efforts are critical to achieve the potential of these new systems. Scaling to exaflop involves improving physical models, mathematical modeling, super scalable algorithms that will require paying particular attention to acquisition, management and visualization of huge amounts of scientific data.

In this context, the purpose of the HIEPACS project is to contribute performing efficiently frontier simulations arising from challenging academic and industrial research. The solution of these challenging problems require a multidisciplinary approach involving applied mathematics, computational and computer sciences. In applied mathematics, it essentially involves advanced numerical schemes. In computational science, it involves massively parallel computing and the design of highly scalable algorithms and codes to be executed on emerging hierarchical many-core, possibly heterogeneous, platforms. Through this approach, HIEPACS intends to contribute to all steps that go from the design of new high-performance more scalable, robust and more accurate numerical schemes to the optimized implementations of the associated algorithms and codes on very high performance supercomputers. This research will be conduced on close collaboration in particular with European and US initiatives and in the framework of EuroHPC collaborative projects.

The methodological part of HIEPACS covers several topics. First, we address generic studies concerning massively parallel computing, the design of high-end performance algorithms and software to be executed on future extreme scale platforms. Next, several research prospectives in scalable parallel linear algebra techniques are addressed, ranging from dense direct, sparse direct, iterative and hybrid approaches for large linear systems. We are also interested in the general problem of minimizing memory consumption and data movements, by changing algorithms and possibly performing extra computations, in particular in the context of Deep Neural Networks. Then we consider research on N-body interaction computations based on efficient parallel fast multipole methods and finally, we address research tracks related to the algorithmic challenges for complex code couplings in multiscale/multiphysic simulations.

Currently, we have one major multiscale application that is in *material physics*. We contribute to all steps of the design of the parallel simulation tool. More precisely, our applied mathematics skill will contribute to the modeling and our advanced numerical schemes will help in the design and efficient software implementation for very large parallel multiscale simulations. Moreover, the robustness and efficiency of our algorithmic research in linear algebra are validated through industrial and academic collaborations with different partners involved in various application fields. Finally, we are also involved in a few collaborative initiatives in various application domains in a co-design like framework. These research activities are conducted in a wider multi-disciplinary context with colleagues in other academic or industrial groups where our contribution is related to our expertises. Not only these collaborations enable our expertise to have a stronger impact in various application domains through the promotion of advanced algorithms, methodologies or tools, but in return they open new avenues for research in the continuity of our core research activities.

Thanks to the two Inria collaborative agreements such as with Airbus/Conseil Régional Grande Aquitaine and with CEA, we have joint research efforts in a co-design framework enabling efficient and effective technological transfer towards industrial R&D. Furthermore, thanks to the past associate team FASTLA we contribute with world leading groups at Berkeley National Lab and Stanford University to the design of fast numerical solvers and their parallel implementations.

Our high performance software packages are integrated in several academic or industrial complex codes and are validated on very large scale simulations. For all our software developments, we use first the experimental platform PLAFRIM, the various large parallel platforms available through GENCI in France (CCRT, CINES and IDRIS Computational Centers), and next the high-end parallel platforms that will be available via European and US initiatives or projects such as PRACE.

3 Research program

3.1 Introduction

The methodological component of HIEPACS concerns the expertise for the design as well as the efficient and scalable implementation of highly parallel numerical algorithms to perform frontier simulations. In order to address these computational challenges a hierarchical organization of the research is considered. In this bottom-up approach, we first consider in Section 3.2 generic topics concerning high performance computational science. The activities described in this section are transversal to the overall project and their outcome will support all the other research activities at various levels in order to ensure the parallel scalability of the algorithms. The aim of this activity is not to study general purpose solution but rather to address these problems in close relation with specialists of the field in order to adapt and tune advanced approaches in our algorithmic designs. The next activity, described in Section 3.3, is related to the study of parallel linear algebra techniques that currently appear as promising approaches to tackle huge problems on extreme scale platforms. We highlight the linear problems (linear systems or eigenproblems) because they are in many large scale applications the main computational intensive numerical kernels and often the main performance bottleneck. These parallel numerical techniques will be the basis of both academic and industrial collaborations, some are described in Section 4.1, but will also be closely related to some functionalities developed in the parallel fast multipole activity described in Section 3.4. Finally, as the accuracy of the physical models increases, there is a real need to go for parallel efficient algorithm implementation for multiphysics and multiscale modeling in particular in the context of code coupling. The challenges associated with this activity will be addressed in the framework of the activity described in Section 3.5.

Currently, we have one major application (see Section 4.1) that is in material physics. We will collaborate to all steps of the design of the parallel simulation tool. More precisely, our applied mathematics skill will contribute to the modelling, our advanced numerical schemes will help in the design and efficient software implementation for very large parallel simulations. We also participate to a few co-design actions in close collaboration with some applicative groups. The objective of this activity is to instantiate our expertise in fields where they are critical for designing scalable simulation tools. We refer to Section 4.2 for a detailed description of these activities.

3.2 High-performance computing on next generation architectures

Participants: Emmanuel Agullo, Olivier Beaumont, Olivier Coulaud, Pierre Esterie, Lionel Eyraud-Dubois, Mathieu Faverge, Luc Giraud, Abdou Guermouche, Gilles Marait, Pierre Ramet, Nick Schenkels, Alena Shilova, Mathieu Verite.

The research directions proposed in HiePACS are strongly influenced by both the applications we are studying and the architectures that we target (i.e., massively parallel heterogeneous many-core architectures, ...). Our main goal is to study the methodology needed to efficiently exploit the new generation of high-performance computers with all the constraints that it induces. To achieve this high-performance with complex applications we have to study both algorithmic problems and the impact of the architectures on the algorithm design.

From the application point of view, the project will be interested in multiresolution, multiscale and hierarchical approaches which lead to multi-level parallelism schemes. This hierarchical parallelism approach is necessary to achieve good performance and high-scalability on modern massively parallel platforms. In this context, more specific algorithmic problems are very important to obtain high performance. Indeed, the kind of applications we are interested in are often based on data redistribution for example (e.g., code coupling applications). This well-known issue becomes very challenging with the increase of both the number of computational nodes and the amount of data. Thus, we have both to study new algorithms and to adapt the existing ones. In addition, some issues like task scheduling have to be restudied in this new context. It is important to note that the work developed in this area will be applied for example in the context of code coupling (see Section 3.5).

Considering the complexity of modern architectures like massively parallel architectures or new generation heterogeneous multicore architectures, task scheduling becomes a challenging problem which is central to obtain a high efficiency. With the recent addition of colleagues from the scheduling community (O. Beaumont and L. Eyraud-Dubois), the team is better equipped than ever to design scheduling algorithms and models specifically tailored to our target problems. It is important to note that this topic is strongly linked to the underlying programming model. Indeed, considering multicore and heterogeneous architectures, it has appeared, in the last five years, that the best programming model is an approach mixing multi-threading within computational nodes and message passing between them.

In the last five years, a lot of work has been developed in the high-performance computing community to understand what is critic to efficiently exploit massively multicore platforms that will appear in the near future. It appeared that the key for the performance is firstly the granularity of the computations. Indeed, in such platforms the granularity of the parallelism must be small so that we can feed all the computing units with a sufficient amount of work. It is thus very crucial for us to design new high performance tools for scientific computing in this new context. This will be developed in the context of our solvers, for example, to adapt to this new parallel scheme. Secondly, the larger the number of cores inside a node, the more complex the memory hierarchy. This remark impacts the behavior of the algorithms within the node. Indeed, on this kind of platforms, NUMA effects will be more and more problematic. Thus, it is very important to study and design data-aware algorithms which take into account the affinity between computational threads and the data they access. This is particularly important in the context of our high-performance tools. Note that this work has to be based on an intelligent cooperative underlying run-time (like the tools developed by the Inria STORM Project-Team) which allows a fine management of data distribution within a node.

Another very important issue concerns high-performance computing using "heterogeneous" resources within a computational node. Indeed, with the deployment of the GPU and the use of more specific co-processors, it is important for our algorithms to efficiently exploit these new type of architectures. To adapt our algorithms and tools to these accelerators, we need to identify what can be done on the GPU for example and what cannot. Note that recent results in the field have shown the interest of using both regular cores and GPU to perform computations. Note also that in opposition to the case of the parallelism granularity needed by regular multicore architectures, GPU requires coarser grain parallelism. Thus, making both GPU and regular cores work all together will lead to two types of tasks in terms of granularity. This represents a challenging problem especially in terms of scheduling. From this perspective, we investigate new approaches for composing parallel applications within a runtime system for heterogeneous platforms.

In the context of scaling up, and particularly in the context of minimizing energy consumption, it is generally acknowledged that the solution lies in the use of heterogeneous architectures, where each resource is particularly suited to specific types of tasks, and in a fine control at the algorithmic level of data movements and the trade-offs to be made between computation and communication. In this context, we are particularly interested in the optimization of the training phase of deep convolutional neural networks which consumes a lot of memory and for which it is possible to exchange computations for data movements and memory occupation. We are also interested in the complexity introduced by resource heterogeneity itself, both from a theoretical point of view on the complexity of scheduling problems and from a more practical point of view on the implementation of specific kernels in dense or sparse linear algebra.

In order to achieve an advanced knowledge concerning the design of efficient computational kernels to be used on our high performance algorithms and codes, we will develop research activities first on regular frameworks before extending them to more irregular and complex situations. In particular, we will work first on optimized dense linear algebra kernels and we will use them in our more complicated direct and hybrid solvers for sparse linear algebra and in our fast multipole algorithms for interaction computations. In this context, we will participate to the development of those kernels in collaboration with groups specialized in dense linear algebra. In particular, we intend develop a strong collaboration with the group of Jack Dongarra at the University of Tennessee and collaborating research groups. The objectives will be to develop dense linear algebra algorithms and libraries for multicore architectures in the context the PLASMA project and for GPU and hybrid multicore/GPU architectures in the context of the MAGMA project. A new solver has emerged from the associate team, Chameleon. While PLASMA and MAGMA focus on multicore and GPU architectures, respectively, Chameleon makes the most out of heterogeneous architectures thanks to task-based dynamic runtime systems.

A more prospective objective is to study the resiliency in the context of large-scale scientific applications for massively parallel architectures. Indeed, with the increase of the number of computational cores per node, the probability of a hardware crash on a core or of a memory corruption is dramatically increased. This represents a crucial problem that needs to be addressed. However, we will only study it at the algorithmic/application level even if it needed lower-level mechanisms (at OS level or even hardware level). Of course, this work can be performed at lower levels (at operating system) level for example but we do believe that handling faults at the application level provides more knowledge about what has to be done (at application level we know what is critical and what is not). The approach that we will follow will be based on the use of a combination of fault-tolerant implementations of the run-time environments we use (like for example ULFM) and an adaptation of our algorithms to try to manage this kind of faults. This topic represents a very long range objective which needs to be addressed to guaranty the robustness of our solvers and applications.

Finally, it is important to note that the main goal of HIEPACS is to design tools and algorithms that will be used within complex simulation frameworks on next-generation parallel machines. Thus, we intend with our partners to use the proposed approach in complex scientific codes and to validate them within very large scale simulations as well as designing parallel solution in co-design collaborations.

3.3 High performance solvers for large linear algebra problems

Participants: Emmanuel Agullo, Olivier Coulaud, Tony Delarue, Mathieu Faverge, Aurélien Falco, Marek Felsoci, Luc Giraud, Abdou Guermouche, Esragul Korkmaz, Gilles Marait, Van Gia Thinh Nguyen, Jean Rene Poirier, Pierre Ramet, Cristobal Samaniego Alvarado, Guillaume Sylvand, Nicolas Venkovic, Yanfei Xiang.

Starting with the developments of basic linear algebra kernels tuned for various classes of computers, a significant knowledge on the basic concepts for implementations on high-performance scientific computers has been accumulated. Further knowledge has been acquired through the design of more sophisticated linear algebra algorithms fully exploiting those basic intensive computational kernels. In that context, we still look at the development of new computing platforms and their associated programming tools. This enables us to identify the possible bottlenecks of new computer architectures (memory path, various level of caches, inter processor or node network) and to propose ways to overcome them in algorithmic design. With the goal of designing efficient scalable linear algebra solvers for large scale applications, various tracks will be followed in order to investigate different complementary approaches. Sparse direct solvers have been for years the methods of choice for solving linear systems of equations, it is nowadays admitted that classical approaches are not scalable neither from a computational complexity nor from a memory view point for large problems such as those arising from the discretization of large 3D PDE problems. We will continue to work on sparse direct solvers on the one hand to make sure they fully benefit from most advanced computing platforms and on the other hand to attempt to reduce their memory and computational costs for some classes of problems where data sparse ideas can be considered. Furthermore, sparse direct solvers are a key building boxes for the design of some of our parallel algorithms such as the hybrid solvers described in the sequel of this section. Our activities in that context will mainly address preconditioned Krylov subspace methods; both components, preconditioner and Krylov solvers, will be investigated. In this framework, and possibly in relation with the research activity on fast multipole, we intend to study how emerging $\mathcal{H}$-matrix arithmetic can benefit to our solver research efforts.

3.3.1 Parallel sparse direct solvers

For the solution of large sparse linear systems, we design numerical schemes and software packages for direct and hybrid parallel solvers. Sparse direct solvers are mandatory when the linear system is very ill-conditioned; such a situation is often encountered in structural mechanics codes, for example. Therefore, to obtain an industrial software tool that must be robust and versatile, high-performance sparse direct solvers are mandatory, and parallelism is then necessary for reasons of memory capability and acceptable solution time. Moreover, in order to solve efficiently 3D problems with more than 50 million unknowns, which is now a reachable challenge with new multicore supercomputers, we must achieve good scalability in time and control memory overhead. Solving a sparse linear system by a direct method is generally a highly irregular problem that induces some challenging algorithmic problems and requires a sophisticated implementation scheme in order to fully exploit the capabilities of modern supercomputers.

New supercomputers incorporate many microprocessors which are composed of one or many computational cores. These new architectures induce strongly hierarchical topologies. These are called NUMA architectures. In the context of distributed NUMA architectures, in collaboration with the Inria STORM team, we study optimization strategies to improve the scheduling of communications, threads and I/O. We have developed dynamic scheduling designed for NUMA architectures in the PaStiX solver. The data structures of the solver, as well as the patterns of communication have been modified to meet the needs of these architectures and dynamic scheduling. We are also interested in the dynamic adaptation of the computation grain to use efficiently multi-core architectures and shared memory. Experiments on several numerical test cases have been performed to prove the efficiency of the approach on different architectures. 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 block low-rank compression techniques.

In collaboration with the ICL team from the University of Tennessee, and the STORM team from Inria, we are evaluating the way to replace the embedded scheduling driver of the PaStiX solver by one of the generic frameworks, PaRSEC or StarPU, to execute the task graph corresponding to a sparse factorization. The aim is to design algorithms and parallel programming models for implementing direct methods for the solution of sparse linear systems on emerging computer equipped with GPU accelerators. More generally, this work will be performed in the context of the ANR SOLHARIS project which aims at designing high performance sparse direct solvers for modern heterogeneous systems. This ANR project involves several groups working either on the sparse linear solver aspects (HIEPACS and ROMA from Inria and APO from IRIT), on runtime systems (STORM from Inria) or scheduling algorithms (HIEPACS and ROMA from Inria). The results of these efforts will be validated in the applications provided by the industrial project members, namely CEA-CESTA and Airbus Central R & T.

3.3.2 Hybrid direct/iterative solvers based on algebraic domain decomposition techniques

One route to the parallel scalable solution of large sparse linear systems in parallel scientific computing is the use of hybrid methods that hierarchically combine direct and iterative methods. These techniques inherit the advantages of each approach, namely the limited amount of memory and natural parallelization for the iterative component and the numerical robustness of the direct part. The general underlying ideas are not new since they have been intensively used to design domain decomposition techniques; those approaches cover a fairly large range of computing techniques for the numerical solution of partial differential equations (PDEs) in time and space. Generally speaking, it refers to the splitting of the computational domain into sub-domains with or without overlap. The splitting strategy is generally governed by various constraints/objectives but the main one is to express parallelism. The numerical properties of the PDEs to be solved are usually intensively exploited at the continuous or discrete levels to design the numerical algorithms so that the resulting specialized technique will only work for the class of linear systems associated with the targeted PDE.

In that context, we continue our effort on the design of algebraic non-overlapping domain decomposition techniques that rely on the solution of a Schur complement system defined on the interface introduced by the partitioning of the adjacency graph of the sparse matrix associated with the linear system. Although it is better conditioned than the original system the Schur complement needs to be precondition to be amenable to a solution using a Krylov subspace method. Different hierarchical preconditioners will be considered, possibly multilevel, to improve the numerical behaviour of the current approaches implemented in our software library MaPHyS. This activity will be developed further developped in the H2020 EOCOE2 project. In addition to this numerical studies, advanced parallel implementation will be developed that will involve close collaborations between the hybrid and sparse direct activities.

3.3.3 Linear Krylov solvers

Preconditioning is the main focus of the two activities described above. They aim at speeding up the convergence of a Krylov subspace method that is the complementary component involved in the solvers of interest for us. In that framework, we believe that various aspects deserve to be investigated; we will consider the following ones:

- Preconditioned block Krylov solvers for multiple right-hand sides. In many large scientific and industrial applications, one has to solve a sequence of linear systems with several right-hand sides given simultaneously or in sequence (radar cross section calculation in electromagnetism, various source locations in seismic, parametric studies in general, ...). For "simultaneous" right-hand sides, the solvers of choice have been for years based on matrix factorizations as the factorization is performed once and simple and cheap block forward/backward substitutions are then performed. In order to effectively propose alternative to such solvers, we need to have efficient preconditioned Krylov subspace solvers. In that framework, block Krylov approaches, where the Krylov spaces associated with each right-hand side are shared to enlarge the search space will be considered. They are not only attractive because of this numerical feature (larger search space), but also from an implementation point of view. Their block-structures exhibit nice features with respect to data locality and re-usability that comply with the memory constraint of multicore architectures. We will continue the numerical study and design of the block GMRES variant that combines inexact break-down detection, deflation at restart and subspace recycling. Beyond new numerical investigations, a software implementation to be included in our linear solver libray Fabulous originately developed in the context of the DGA HIBOX project and further developped in the LynCs (Linear Algebra, Krylov-subspace methods, and multi-grid solvers for the discovery of New Physics) sub-project of PRACE-6IP.

- Extension or modification of Krylov subspace algorithms for multicore architectures: finally to match as much as possible to the computer architecture evolution and get as much as possible performance out of the computer, a particular attention will be paid to adapt, extend or develop numerical schemes that comply with the efficiency constraints associated with the available computers. Nowadays, multicore architectures seem to become widely used, where memory latency and bandwidth are the main bottlenecks; investigations on communication avoiding techniques will be undertaken in the framework of preconditioned Krylov subspace solvers as a general guideline for all the items mentioned above.

3.3.4 Eigensolvers

Many eigensolvers also rely on Krylov subspace techniques. Naturally some links exist between the Krylov subspace linear solvers and the Krylov subspace eigensolvers. We plan to study the computation of eigenvalue problems with respect to the following two different axes:

- Exploiting the link between Krylov subspace methods for linear system solution and eigensolvers, we intend to develop advanced iterative linear methods based on Krylov subspace methods that use some spectral information to build part of a subspace to be recycled, either though space augmentation or through preconditioner update. This spectral information may correspond to a certain part of the spectrum of the original large matrix or to some approximations of the eigenvalues obtained by solving a reduced eigenproblem. This technique will also be investigated in the framework of block Krylov subspace methods.

- In the context of the calculation of the ground state of an atomistic system, eigenvalue computation is a critical step; more accurate and more efficient parallel and scalable eigensolvers are required.

3.3.5 Fast Solvers for FEM/BEM Coupling

In this research project, we are interested in the design of new advanced techniques to solve large mixed dense/sparse linear systems, the extensive comparison of these new approaches to the existing ones, and the application of these innovative ideas on realistic industrial test cases in the domain of aeroacoustics (in collaboration with Airbus Central R & T).

- The use of $\mathscr{H}$-matrix solvers on these problems has been investigated in the context of the PhD of A. Falco. Airbus CR&T, in collaboration with Inria Bordeaux Sud-Ouest, has developed a task-based $\mathscr{H}$-matrix solver on top of the runtime engine StarPU. Ideas coming from the field of sparse direct solvers (such as nested dissection or symbolic factorization) have been tested within $\mathscr{H}$-matrices.

- The question of parallel scalability of task-based tools is an active subject of research, using new communication engine such as NewMadeleine, that will be investigated during this project, in conjunction with new algorithmic ideas on the task-based writing of $\mathscr{H}$-matrix algorithms.
- Naturally, comparison with existing tools will be performed on large realistic test cases. Coupling schemes between these tools and the hierarchical methods used in $\mathscr{H}$-matrix will be developed and benched as well.

3.4 High performance Fast Multipole Method for N-body problems

Participants: Emmanuel Agullo, Olivier Coulaud, Pierre Esterie, Guillaume Sylvand.

In most scientific computing applications considered nowadays as computational challenges (like biological and material systems, astrophysics or electromagnetism), the introduction of hierarchical methods based on an octree structure has dramatically reduced the amount of computation needed to simulate those systems for a given accuracy. For instance, in the N-body problem arising from these application fields, we must compute all pairwise interactions among N objects (particles, lines, ...) at every timestep. Among these methods, the Fast Multipole Method (FMM) developed for gravitational potentials in astrophysics and for electrostatic (coulombic) potentials in molecular simulations solves this N-body problem for any given precision with $O(N)$ runtime complexity against $O(N^2)$ for the direct computation.

The potential field is decomposed in a near field part, directly computed, and a far field part approximated thanks to multipole and local expansions. We introduced a matrix formulation of the FMM that exploits the cache hierarchy on a processor through the Basic Linear Algebra Subprograms (BLAS). Moreover, we developed a parallel adaptive version of the FMM algorithm for heterogeneous particle distributions, which is very efficient on parallel clusters of SMP nodes. Finally on such computers, we developed the first hybrid MPI-thread algorithm, which enables to reach better parallel efficiency and better memory scalability. We plan to work on the following points in HiePACS .

3.4.1 Improvement of calculation efficiency

Nowadays, the high performance computing community is examining alternative architectures that address the limitations of modern cache-based designs. GPU (Graphics Processing Units) and the Cell processor have thus already been used in astrophysics and in molecular dynamics. The Fast Mutipole Method has also been implemented on GPU. We intend to examine the potential of using these forthcoming processors as a building block for high-end parallel computing in N-body calculations. More precisely, we want to take advantage of our specific underlying BLAS routines to obtain an efficient and easily portable FMM for these new architectures. Algorithmic issues such as dynamic load balancing among heterogeneous cores will also have to be solved in order to gather all the available computation power. This research action will be conduced on close connection with the activity described in Section 3.2.

3.4.2 Non uniform distributions

In many applications arising from material physics or astrophysics, the distribution of the data is highly non uniform and the data can grow between two time steps. As mentioned previously, we have proposed a hybrid MPI-thread algorithm to exploit the data locality within each node. We plan to further improve the load balancing for highly non uniform particle distributions with small computation grain thanks to dynamic load balancing at the thread level and thanks to a load balancing correction over several simulation time steps at the process level.

3.4.3 Fast multipole method for dislocation operators

The engine that we develop will be extended to new potentials arising from material physics such as those used in dislocation simulations. The interaction between dislocations is long ranged ($O(1/r)$) and anisotropic, leading to severe computational challenges for large-scale simulations. Several approaches based on the FMM or based on spatial decomposition in boxes are proposed to speed-up the computation. In dislocation codes, the calculation of the interaction forces between dislocations is still the most CPU time consuming. This computation has to be improved to obtain faster and more accurate simulations. Moreover, in such simulations, the number of dislocations grows while the phenomenon occurs and these dislocations are not uniformly distributed in the domain. This means that strategies to dynamically balance the computational load are crucial to achieve high performance.

3.4.4 Fast multipole method for boundary element methods

The boundary element method (BEM) is a well known solution of boundary value problems appearing in various fields of physics. With this approach, we only have to solve an integral equation on the boundary. This implies an interaction that decreases in space, but results in the solution of a dense linear system with $O(N^3)$ complexity. The FMM calculation that performs the matrix-vector product enables the use of Krylov subspace methods. Based on the parallel data distribution of the underlying octree implemented to perform the FMM, parallel preconditioners can be designed that exploit the local interaction matrices computed at the finest level of the octree. This research action will be conduced on close connection with the activity described in Section 3.3. Following our earlier experience, we plan to first consider approximate inverse preconditionners that can efficiently exploit these data structures.

3.5 Load balancing algorithms for complex simulations

Participants: Aurélien Esnard, Pierre Ramet.

Many important physical phenomena in material physics and climatology are inherently complex applications. They often use multi-physics or multi-scale approaches, which couple different models and codes. The key idea is to reuse available legacy codes through a coupling framework instead of merging them into a stand-alone application. There is typically one model per different scale or physics and each model is implemented by a parallel code.

For instance, to model a crack propagation, one uses a molecular dynamic code to represent the atomistic scale and an elasticity code using a finite element method to represent the continuum scale. Indeed, fully microscopic simulations of most domains of interest are not computationally feasible. Combining such different scales or physics is still a challenge to reach high performance and scalability.

Another prominent example is found in the field of aeronautic propulsion: the conjugate heat transfer simulation in complex geometries (as developed by the CFD team of CERFACS) requires to couple a fluid/convection solver (AVBP) with a solid/conduction solver (AVTP). As the AVBP code is much more CPU consuming than the AVTP code, there is an important computational imbalance between the two solvers.

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 simulation and not only for each stand-alone code, as it is most 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, which can drastically decrease the overall performance. Therefore, we argue that it is required to model the coupling itself in order to ensure a good scalability, especially when running on massively parallel architectures (tens of thousands of processors/cores). In other words, one must develop new algorithms and software implementation to perform a *coupling-aware* partitioning of the whole application. Another related problem is the problem of resource allocation. This is particularly important for the global coupling efficiency and scalability, because each code involved in the coupling can be more or less computationally intensive, and there is a good trade-off to find between resources assigned to each code to avoid that one of them waits for the other(s). What does furthermore happen if the load of one code dynamically changes relatively to the other one? In such a case, it could be convenient to dynamically adapt the number of resources used during the execution.

There are several open algorithmic problems that we investigate in the HIEPACS project-team. All these problems uses a similar methodology based upon the graph model and are expressed as variants of the classic graph partitioning problem, using additional constraints or different objectives.

3.5.1 Dynamic load-balancing with variable number of processors

As a preliminary step related to the dynamic load balancing of coupled codes, we focus on the problem of dynamic load balancing of a single parallel code, with variable number of processors. Indeed, if the workload varies drastically during the simulation, the load must be redistributed regularly among the processors. Dynamic load balancing is a well studied subject but most studies are limited to an initially fixed number of processors. Adjusting the number of processors at runtime allows one to preserve the parallel code efficiency or keep running the simulation when the current memory resources are exceeded. We call this problem, *MxN graph repartitioning*.

We propose some methods based on graph repartitioning in order to re-balance the load while changing the number of processors. These methods are split in two main steps. Firstly, we study the migration phase and we build a “good” migration matrix minimizing several metrics like the migration volume or the number of exchanged messages. Secondly, we use graph partitioning heuristics to compute a new distribution optimizing the migration according to the previous step results.

3.5.2 Load balancing of coupled codes

As stated above, the load balancing of coupled code is a major issue, that determines the performance of the complex simulation, and reaching high performance can be a great challenge. In this context, we develop new graph partitioning techniques, called *co-partitioning*. They address the problem of load balancing for two coupled codes: the key idea is to perform a "coupling-aware" partitioning, instead of partitioning these codes independently, as it is classically done. More precisely, we propose to enrich the classic graph model with *inter-edges*, which represent the coupled code interactions. We describe two new algorithms, and compare them to the naive approach. In the preliminary experiments we perform on synthetically-generated graphs, we notice that our algorithms succeed to balance the computational load in the coupling phase and in some cases they succeed to reduce the coupling communications costs. Surprisingly, we notice that our algorithms do not degrade significantly the global graph edge-cut, despite the additional constraints that they impose.

Besides this, our co-partitioning technique requires to use graph partitioning with *fixed vertices*, that raises serious issues with state-of-the-art software, that are classically based on the well-known recursive bisection paradigm (RB). Indeed, the RB method often fails to produce partitions of good quality. To overcome this issue, we propose a *new* direct k-way greedy graph growing algorithm, called KGGGP, that overcomes this issue and succeeds to produce partition with better quality than RB while respecting the constraint of fixed vertices. Experimental results compare KGGGP against state-of-the-art methods, such as Scotch, for real-life graphs available from the popular *DIMACS'10* collection.

3.5.3 Load balancing strategies for hybrid sparse linear solvers

Graph handling and partitioning play a central role in the activity described here but also in other numerical techniques detailed in sparse linear algebra Section. The Nested Dissection is now a well-known heuristic for sparse matrix ordering to both reduce the fill-in during numerical factorization and to maximize the number of independent computation tasks. By using the block data structure induced by the partition of separators of the original graph, very efficient parallel block solvers have been designed and implemented according to super-nodal or multi-frontal approaches. Considering hybrid methods mixing both direct and iterative solvers such as MaPHyS, obtaining a domain decomposition leading to a good balancing of both the size of domain interiors and the size of interfaces is a key point for load balancing and efficiency in a parallel context.

We intend to revisit some well-known graph partitioning techniques in the light of the hybrid solvers and design new algorithms to be tested in the Scotch package.

4 Application domains

4.1 Material physics

Participants: Olivier Coulaud, Pierre Esterie.

Due to the increase of available computer power, new applications in nano science and physics appear such as study of properties of new materials (photovoltaic materials, bio- and environmental sensors, ...), failure in materials, nano-indentation. Chemists, physicists now commonly perform simulations in these fields. These computations simulate systems up to billion of atoms in materials, for large time scales up to several nanoseconds. The larger the simulation, the smaller the computational cost of the potential driving the phenomena, resulting in low precision results. So, if we need to increase the precision, there are two ways to decrease the computational cost. In the first approach, we improve algorithms and their parallelization and in the second way, we consider a multiscale approach.

A domain of interest is the material aging for the nuclear industry. The materials are exposed to complex conditions due to the combination of thermo-mechanical loading, the effects of irradiation and the harsh operating environment. This operating regime makes experimentation extremely difficult and we must rely on multi-physics and multi-scale modeling for our understanding of how these materials behave in service. This fundamental understanding helps not only to ensure the longevity of existing nuclear reactors, but also to guide the development of new materials for 4th generation reactor programs and dedicated fusion reactors. For the study of crystalline materials, an important tool is dislocation dynamics (DD) modeling. This multiscale simulation method predicts the plastic response of a material from the underlying physics of dislocation motion. DD serves as a crucial link between the scale of molecular dynamics and macroscopic methods based on finite elements; it can be used to accurately describe the interactions of a small handful of dislocations, or equally well to investigate the global behavior of a massive collection of interacting defects.

To explore i.e. to simulate these new areas, we need to develop and/or to improve significantly models, schemes and solvers used in the classical codes. In the project, we want to accelerate algorithms arising in those fields.

We focus on the following topics (in particular in the currently under definition OPTIDIS project in collaboration with CEA Saclay, CEA Ile-de-france and SIMaP Laboratory in Grenoble) in connection with research described at Sections 3.4 and 3.5.

- The interaction between dislocations is long ranged ($O(1/r)$) and anisotropic, leading to severe computational challenges for large-scale simulations. In dislocation codes, the computation of interaction forces between dislocations is still the most CPU time consuming and has to be improved to obtain faster and more accurate simulations.

- In such simulations, the number of dislocations grows while the phenomenon occurs and these dislocations are not uniformly distributed in the domain. This means that strategies to dynamically construct a good load balancing are crucial to acheive high performance.

- From a physical and a simulation point of view, it will be interesting to couple a molecular dynamics model (atomistic model) with a dislocation one (mesoscale model). In such three-dimensional coupling, the main difficulties are firstly to find and characterize a dislocation in the atomistic region, secondly to understand how we can transmit with consistency the information between the two micro and meso scales.

4.2 Co-design of algorithms in scientific applications

Participants: Emmanuel Agullo, Olivier Beaumont, Lionel Eyraud-Dubois, Mathieu Faverge, Marek Felsoci, Luc Giraud, Gilles Marait, Pierre Ramet, Matthieu Simonin, Guillaume Sylvand, Alena Shilova.

4.2.1 High performance simulation for ITER tokamak

Scientific simulation for ITER tokamak modeling provides a natural bridge between theory and experimentation and is also an essential tool for understanding and predicting plasma behavior. Recent progresses in numerical simulation of fine-scale turbulence and in large-scale dynamics of magnetically confined plasma have been enabled by access to petascale supercomputers. These progresses would have been unreachable without new computational methods and adapted reduced models. In particular, the plasma science community has developed codes for which computer runtime scales quite well with the number of processors up to thousands cores. The research activities of HiePACS concerning the international ITER challenge have started in the Inria Project Lab C2S@Exa in collaboration with CEA-IRFM and were related to two complementary studies: a first one concerning the turbulence of plasma particles inside a tokamak (in the context of GYSELA code) and a second one concerning the MHD instability edge localized modes (in the context of JOREK code). The activity concerning GYSELA was completed at the end of 2018.

Other numerical simulation tools designed for the ITER challenge aim at making a significant progress in understanding active control methods of plasma edge MHD instability Edge Localized Modes (ELMs) which represent a particular danger with respect to heat and particle loads for Plasma Facing Components (PFC) in the tokamak. The goal is to improve the understanding of the related physics and to propose possible new strategies to improve effectiveness of ELM control techniques. The simulation tool used (JOREK code) is related to non linear MHD modeling and is based on a fully implicit time evolution scheme that leads to 3D large very badly conditioned sparse linear systems to be solved at every time step. In this context, the use of PaStiX library to solve efficiently these large sparse problems by a direct method is a challenging issue.

This activity continues within the context of the EoCoE2 project, in which the PaStiX solver is identified to allow the processing of very larger linear systems for the nuclear fusion code TOKAM3X from CEA-IRFM. Contrary to the JOREK code, the problem to be treated corresponds to the complete 3D volume of the plasma torus. The objective is to be competitive, for complex geometries, compared to an Algebraic MultiGrid approach designed by one partner of EoCoE2.

4.2.2 Numerical and parallel scalable hybrid solvers in large scale calculations

Parallel and numerically scalable hybrid solvers based on a fully algebraic coarse space correction have been theoretically studied and various advanced parallel implementations have been designed. Their parallel scalability has been initially investigated on large scale problems within the EoCoE project thanks to a close collaboration with the BSC and the integration of MaPHyS within the Alya software. This activity will further develop in the EoCoE2 project. The performance has also been assessed on PRACE Tier-0 machine within a PRACE Project Access through a collaboration with CERFACS and Laboratoire de Physique des Plasmas at Ecole Polytechnique for the calculation of plasma propulsion. A comparative parallel scalability study with the Algebraic MultiGrid from Petsc has been conducted in that framework.

4.3 Aeroacoustics Simulation

This domains is in the context of a long term collaboration with Airbus Research Centers. Wave propagation phenomena intervene in many different aspects of systems design at Airbus. They drive the level of acoustic vibrations that mechanical components have to sustain, a level that one may want to diminish for comfort reason (in the case of aircraft passengers, for instance) or for safety reason (to avoid damage in the case of a payload in a rocket fairing at take-off). Numerical simulations of these phenomena plays a central part in the upstream design phase of any such project. Airbus Central R & T has developed over the last decades an in-depth knowledge in the field of Boundary Element Method (BEM) for the simulation of wave propagation in homogeneous media and in frequency domain. To tackle heterogeneous media (such as the jet engine flows, in the case of acoustic simulation), these BEM approaches are coupled with volumic finite elements (FEM). We end up with the need to solve large (several millions unknowns) linear systems of equations composed of a dense part (coming for the BEM domain) and a sparse part (coming from the FEM domain). Various parallel solution techniques are available today, mixing tools created by the academic world (such as the Mumps and Pastix sparse solvers) as well as parallel software tools developed in-house at Airbus (dense solver SPIDO, multipole solver, $\mathcal{H}$-matrix solver with an open sequential version available online). In the current state of knowledge and technologies, these methods do not permit to tackle the simulation of aeroacoustics problems at the highest acoustic frequencies (between 5 and 20 kHz, upper limits of human audition) while considering the whole complexity of geometries and phenomena involved (higher acoustic frequency implies smaller mesh sizes that lead to larger unknowns number, a number that grows like f^2 for BEM and f^3 for FEM, where f is the studied frequency). The purpose of the study in this domain is to develop advanced solvers able to tackle this kind of mixed dense/sparse linear systems efficiently on parallel architectures.

4.4 Optimization for Deep Convolutional Neural Networks

The training phase of Deep Convolutional Neural Networks represents nowadays a significant share of the computations performed on HPC supercomputers. It introduces several new problems concerning resource allocation and scheduling issues, because of the specific pattern of task graphs induced by the stochastic gradient descent and because memory consumption is particularly critical when performing training. As of today, the most classical parallelization methods consists in partitioning mini-batches, images, filters,... but all these methods induce high synchronization and communication costs, and only very partially resolve memory issues. Within the framework of the Inria IPL on HPC Big Data and Learning convergence, we are working on re-materialization techniques and on the use of model parallelism, in particular to be able to build on the research that has been carried out in a more traditional HPC framework on the exploitation of resource heterogeneity and dynamic runtime scheduling.

5 Highlights of the year

5.1 Publication at NeurIPS

We have published our work on *Efficient Combination of Rematerialization and Offloading for Training DNNs* [22] in the NeurIPS conference. NeurIPS is the main and most prestigious conference of the machine learning community. This publication highlights our 3-year effort to use our scheduling expertise from the HPC environment to obtain meaningful contributions for the machine learning community. In addition, we expect this to boost the visibility of the associated Rotor software within the machine learning community.

6 New software and platforms

6.1 New software

6.1.1 AVCI

Name: Adaptive vibrational configuration interaction

Keywords: Vibrational spectra, Eigen value

Functional Description: A-VCI is a theoretical vibrational spectroscopy algorithm developed to effectively reduce the number of vibrational states used in the configuration-interaction (CI) process. It constructs a nested basis for the discretization of the Hamiltonian operator inside a large CI approximation space and uses an a-posteriori error estimator (residue) to select the most relevant directions to expand the discretization space.

The Hamiltonian operator consists of 3 operators: a harmonic oscillator sum, the potential energy surface operator and the Coriolis operators. In addition, the code can compute the intensity of eigenvectors.

The code can handle molecules up to 10 atoms, which corresponds to solving an eigenvalue problem in a 24-dimensional space.

Publications: hal-01310708, hal-02518243, hal-01534134, hal-02524533

Author: Olivier Coulaud

Contact: Olivier Coulaud

Partner: IPREM

6.1.2 Chameleon

Keywords: Runtime system, Task-based algorithm, Dense linear algebra, HPC, Task scheduling

Scientific Description: Chameleon is part of the MORSE (Matrices Over Runtime Systems @ Exascale) project. The overall objective is to develop robust linear algebra libraries relying on innovative runtime systems that can fully benefit from the potential of those future large-scale complex machines.

We expect advances in three directions based first on strong and closed interactions between the runtime and numerical linear algebra communities. This initial activity will then naturally expand to more focused but still joint research in both fields.

1. Fine interaction between linear algebra and runtime systems. On parallel machines, HPC applications need to take care of data movement and consistency, which can be either explicitly managed at the level of the application itself or delegated to a runtime system. We adopt the latter approach in order to better keep up with hardware trends whose complexity is growing exponentially. One major task in this project is to define a proper interface between HPC applications and runtime systems in order to maximize productivity and expressivity. As mentioned in the next section, a widely used approach consists in abstracting the application as a DAG that the runtime system is in charge of scheduling. Scheduling such a DAG over a set of heterogeneous processing units introduces a lot of new challenges, such as predicting accurately the execution time of each type of task over each kind of unit, minimizing data transfers between memory banks, performing data prefetching, etc. Expected advances: In a nutshell, a new runtime system API will be designed to allow applications to provide scheduling hints to the runtime system and to get real-time feedback about the consequences of scheduling decisions.

2. Runtime systems. A runtime environment is an intermediate layer between the system and the application. It provides low-level functionality not provided by the system (such as scheduling or management of the heterogeneity) and high-level features (such as performance portability). In the framework of this proposal, we will work on the scalability of runtime environment. To achieve scalability it is required to avoid all centralization. Here, the main problem is the scheduling of the tasks. In many task-based runtime environments the scheduler is centralized and becomes a bottleneck as soon as too many cores are involved. It is therefore required to distribute the scheduling decision or to compute a data distribution that impose the mapping of task using, for instance the so-called “owner-compute” rule. Expected advances: We will design runtime systems that enable an efficient and scalable use of thousands of distributed multicore nodes enhanced with accelerators.

3. Linear algebra. Because of its central position in HPC and of the well understood structure of its algorithms, dense linear algebra has often pioneered new challenges that HPC had to face. Again, dense linear algebra has been in the vanguard of the new era of petascale computing with the design of new algorithms that can efficiently run on a multicore node with GPU accelerators. These algorithms are called “communication-avoiding” since they have been redesigned to limit the amount of communication between processing units (and between the different levels of memory hierarchy). They are expressed through Direct Acyclic Graphs (DAG) of fine-grained tasks that are dynamically scheduled. Expected advances: First, we plan to investigate the impact of these principles in the case of sparse applications (whose algorithms are slightly more complicated but often rely on dense kernels). Furthermore, both in the dense and sparse cases, the scalability on thousands of nodes is still limited, new numerical approaches need to be found. We will specifically design sparse hybrid direct/iterative methods that represent a promising approach.

Overall end point. The overall goal of the MORSE associate team is to enable advanced numerical algorithms to be executed on a scalable unified runtime system for exploiting the full potential of future exascale machines.

Functional Description: Chameleon is a dense linear algebra software relying on sequential task-based algorithms where sub-tasks of the overall algorithms are submitted to a Runtime system. A Runtime system such as StarPU is able to manage automatically data transfers between not shared memory area (CPUs-GPUs, distributed nodes). This kind of implementation paradigm allows to design high performing linear algebra algorithms on very different type of architecture: laptop, many-core nodes, CPUs-GPUs, multiple nodes. For example, Chameleon is able to perform a Cholesky factorization (double-precision) at 80 TFlop/s on a dense matrix of order 400 000 (i.e. 4 min 30 s).

Release Contributions: Chameleon includes the following features:

- BLAS 3, LAPACK one-sided and LAPACK norms tile algorithms - Support QUARK and StarPU runtime systems and PaRSEC since 2018 - Exploitation of homogeneous and heterogeneous platforms through the use of BLAS/LAPACK CPU kernels and cuBLAS/MAGMA CUDA kernels - Exploitation of clusters of interconnected nodes with distributed memory (using OpenMPI)

URL: https://gitlab.inria.fr/solverstack/chameleon

Contact: Emmanuel Agullo

Participants: Cédric Castagnede, Samuel Thibault, Emmanuel Agullo, Florent Pruvost, Mathieu Faverge

Partners: Innovative Computing Laboratory (ICL), King Abdullha University of Science and Technology, University of Colorado Denver

6.1.3 Diodon

Name: Diodon

Keywords: Dimensionality reduction, Data analysis

Functional Description: Most of dimension reduction methods inherited from Multivariate Data Analysis, and currently implemented as element in statistical learning for handling very large datasets (the dimension of spaces is the number of features) rely on a chain of pretreatments, a core with a SVD for low rank approximation of a given matrix, and a post-treatment for interpreting results. The costly part in computations is the SVD, which is in cubic complexity. Diodon is a list of functions and drivers which implement (i) pre-treatments, SVD and post-treatments on a large diversity of methods, (ii) random projection methods for running the SVD which permits to bypass the time limit in computing the SVD, and (iii) an implementation in C++ of the SVD with random projection at prescribed rank or precision, connected to MDS.

URL: https://gitlab.inria.fr/afranc/diodon

Authors: Alain Franc, Florent Pruvost, Romain Peressoni, Romain Peressoni

Contact: Alain Franc

6.1.4 DPLASMA

Name: Distributed Parallel Linear Algebra Software for Multicore Architectures

Keywords: Dense linear algebra, Linear algebra

Functional Description: DPLASMA is the leading implementation of a dense linear algebra package for distributed heterogeneous systems. It is designed to deliver sustained performance for distributed systems where each node featuring multiple sockets of multicore processors, and if available, accelerators like GPUs or Intel Xeon Phi. DPLASMA achieves this objective through the state of the art PaRSEC runtime, porting the PLASMA algorithms to the distributed memory realm.

URL: http://icl.cs.utk.edu/parsec/index.html

Author: Mathieu Faverge

Contact: Mathieu Faverge

6.1.5 Fabulous

Name: Fast Accurate Block Linear krylOv Solver

Keywords: Numerical algorithm, Block Krylov solver

Scientific Description: Versatile and flexible numerical library that implements Block Krylov iterative schemes for the solution of linear systems of equations with multiple right-hand sides

Functional Description: Versatile and flexible numerical library that implements Block Krylov iterative schemes for the solution of linear systems of equations with multiple right-hand sides. The library implements block variants of minimal norm residual variants with partial convergence management and spectral information recycling. The package already implements regular block-GMRES (BGMRES), Inexact Breakdown BGMRES (IB-BMGRES), Inexact Breakdown BGMRES with Deflated Restarting (IB-BGMRES-DR), Block Generalized Conjugate Residual with partial convergence management and Inexact Breakdown Block Generalized Conjugate Residual with inner Orthogonalisation and Deflated Restarting (IB-BGCRO-DR). The C++ library relies on callback mechanisms to implement the calculations (matrix-vector, dot-product, ...) that depend on the parallel data distribution selected by the user.

Release Contributions: Added the IB-BGCRO-DR algorithm

URL: https://gitlab.inria.fr/solverstack/fabulous/

Publication: hal-01067159

Contact: Luc Giraud

Participants: Emmanuel Agullo, Luc Giraud, Gilles Marait, Cyrille Piacibello, Matthieu Simonin

6.1.6 MAPHYS

Name: Massively Parallel Hybrid Solver

Keyword: Parallel hybrid direct/iterative solution of large linear systems

Functional Description: MaPHyS is a software package that implements a parallel linear solver coupling direct and iterative approaches. The underlying idea is to apply to general unstructured linear systems domain decomposition ideas developed for the solution of linear systems arising from PDEs. The interface problem, associated with the so called Schur complement system, is solved using a block preconditioner with overlap between the blocks that is referred to as Algebraic Additive Schwarz. A fully algebraic coarse space is available for symmetric positive definite problems, that insures the numerical scalability of the preconditioner.

The parallel implementation is based on MPI+thread. Maphys relies on state-of-the art sparse and dense direct solvers.

MaPHyS is essentially a preconditioner that can be used to speed-up the convergence of any Krylov subspace method and is coupled with the ones implemented in the Fabulous package.

URL: https://gitlab.inria.fr/solverstack/maphys

Publications: hal-01379227, hal-01399203

Contact: Emmanuel Agullo

Participants: Emmanuel Agullo, Luc Giraud, Matthieu Kuhn, Gilles Marait, Louis Poirel

6.1.7 MAPHYS++

Name: Massively Parallel Hybrid Solver in modern C++

Keywords: Linear Systems Solver, Parallel computing, C++, Parallel numerical solvers, Sparse Matrices, Hybrid direct iterative method

Scientific Description: Maphys++ is a parallel linear solver for the resolution of large hollow linear systems that attempts to best combine the advantages of direct methods in terms of robustness and iterative methods in terms of memory footprint and adapted precision. It implements a modern C++ interface (C++17/20), giving the user a wide range of solving methods : efficient direct solver (wrapping MUMPS, Pastix...), iterative solver (CG, GMRES...) and also allowing for a combination of those (hybrid solve using the Schur complement, with adapted preconditioners). Parallelism is based on domain decomposition methods and is implemented in MPI.

Functional Description: MAPHYS++ is a numerical library for solving large sparse linear systems arising from modeling complex phenomena. The library is designed to be used on parallel computers and allows to control the expected accuracy. In a more technical way, the library has a modern C++ interface (C++17/20), allowing a large composability in the solution methods: efficient direct solvers (wrappers to MUMPS, Pastix, ...), iterative solvers (CG, GMRES, ...), as well as the possibility of combining these (hybrid solving using the Schur complement or directly the initial matrix, associated with various preconditioners allowing to control the condition number and consequently the convergence rate in the symmetric positive definite case). The parallelism exploits the domain decomposition principle and implemented in MPI.

News of the Year: The redesign of the package has been completed New release : 1.0 End of 1st development cycle.

URL: https://gitlab.inria.fr/solverstack/maphys/maphyspp

Contact: Emmanuel Agullo

Participants: Emmanuel Agullo, Luc Giraud, Gilles Marait, Matthieu Simonin, Louis Poirel

6.1.8 StarPart

Keywords: High performance computing, HPC, Parallel computing, Graph algorithmics, Graph, Hypergraph

Functional Description: StarPart is a flexible and extensible framework that integrates state-of-the-art methods for graph partitioning and sparse matrix ordering. More precisely, StarPart is a framework that offers a uniform API to manipulate graph, hypergraph and mesh structures. It is designed to be easily extensible by adding new methods and to plug all these methods into a comprehensive framework. It is initially designed to provide graph partitioning and sparse matrix ordering methods, that come from sate-of-the-art software such as Metis, Scotch, Patoh, Zoltan, etc. Besides, it provides some facilities for IO, diagnostic, benchmark, visualization (VTK, SVG, ...). StarPart is the core of the MetaPart project. It is built upon the LibGraph library.

URL: https://gitlab.inria.fr/metapart/starpart

Contact: Aurélien Esnard

Participant: Aurélien Esnard

6.1.9 MPICPL

Name: MPI CouPLing

Keywords: MPI, Coupling software

Functional Description: MPICPL is a software library dedicated to the coupling of parallel legacy codes, that are based on the well-known MPI standard. It proposes a lightweight and comprehensive programing interface that simplifies the coupling of several MPI codes (2, 3 or more). MPICPL facilitates the deployment of these codes thanks to the mpicplrun tool and it interconnects them automatically through standard MPI inter-communicators. Moreover, it generates the universe communicator, that merges the world communicators of all coupled-codes. The coupling infrastructure is described by a simple XML file, that is just loaded by the mpicplrun tool.

URL: https://gitlab.inria.fr/esnard/mpicpl

Contact: Aurélien Esnard

Participant: Aurélien Esnard

6.1.10 OptiDis

Keywords: Dislocation dynamics simulation, Fast multipole method, Large scale, Collision

Functional Description: OptiDis is a new code for large scale dislocation dynamics simulations. Its purpose is to simulate real life dislocation densities (up to 5.1022 dislocations/m-2) in order to understand plastic deformation and study strain hardening. The main application is to observe and understand plastic deformation of irradiated zirconium. Zirconium alloys are the first containment barrier against the dissemination of radioactive elements. More precisely, with neutron irradiated zirconium alloys we are talking about channeling mechanism, which means to stick with the reality, more than tens of thousands of induced loops, i. e. 100 million degrees of freedom in the simulation. The code is based on Numodis code developed at CEA Saclay and the ScalFMM library developed in HiePACS project. The code is written in C++ language and using the last features of C++11/14. One of the main aspects is the hybrid parallelism MPI/OpenMP that gives the software the ability to scale on large cluster while the computation load rises. In order to achieve that, we use different levels of parallelism. First of all, the simulation box is distributed over MPI processes, then we use a thinner level for threads, dividing the domain by an Octree representation. All theses parts are controlled by the ScalFMM library. On the last level, our data are stored in an adaptive structure that absorbs the dynamics of this type of simulation and manages the parallelism of tasks..

URL: https://gitlab.inria.fr/optidis/optidis

Publication: hal-01086371

Contact: Olivier Coulaud

Participant: Olivier Coulaud

Partner: CEA

6.1.11 PaStiX

Name: Parallel Sparse matriX package

Keywords: Linear algebra, High-performance calculation, Sparse Matrices, Linear Systems Solver, Low-Rank compression

Scientific Description: PaStiX is based on an efficient static scheduling and memory manager, in order to solve 3D problems with more than 50 million of unknowns. The mapping and scheduling algorithm handles a combination of 1D and 2D block distributions. A dynamic scheduling can also be applied to take care of NUMA architectures while taking into account very precisely the computational costs of the BLAS 3 primitives, the communication costs and the cost of local aggregations.

Functional Description: PaStiX is a scientific library that provides a high performance parallel solver for very large sparse linear systems based on block direct and block ILU(k) methods. It can handle low-rank compression techniques to reduce the computation and the memory complexity. Numerical algorithms are implemented in single or double precision (real or complex) for LLt, LDLt and LU factorization with static pivoting (for non symmetric matrices having a symmetric pattern). The PaStiX library uses the graph partitioning and sparse matrix block ordering packages Scotch or Metis.

The PaStiX solver is suitable for any heterogeneous parallel/distributed architecture when its performance is predictable, such as clusters of multicore nodes with GPU accelerators or KNL processors. In particular, we provide a high-performance version with a low memory overhead for multicore node architectures, which fully exploits the advantage of shared memory by using a hybrid MPI-thread implementation.

The solver also provides some low-rank compression methods to reduce the memory footprint and/or the time-to-solution.

URL: https://gitlab.inria.fr/solverstack/pastix

Contact: Pierre Ramet

Participants: Tony Delarue, Grégoire Pichon, Mathieu Faverge, Esragul Korkmaz, Pierre Ramet

Partners: INP Bordeaux, Université de Bordeaux

6.1.12 pmtool

Keywords: Scheduling, Task scheduling, StarPU, Heterogeneity, GPGPU, Performance analysis

Functional Description: Analyse post-mortem the behavior of StarPU applications. Provide lower bounds on makespan. Study the performance of different schedulers in a simple context. Provide implementations of many scheduling algorithms from the literature

News of the Year: Included many new algorithms, in particular online algorithms Better integration with StarPU by accepting .rec files as input

URL: https://gitlab.inria.fr/eyrauddu/pmtool

Publications: hal-01386174, hal-01878606

Contact: Lionel Eyraud Dubois

Participant: Lionel Eyraud Dubois

6.1.13 rotor

Name: Re-materializing Optimally with pyTORch

Keywords: Deep learning, Optimization, Python, GPU, Automatic differentiation

Scientific Description: This software implements in PyTorch a new activation checkpointing method which allows to significantly decrease memory usage when training Deep Neural Networks with the back-propagation algorithm. Similarly to checkpointing techniques coming from the literature on Automatic Differentiation, it consists in dynamically selecting the forward activations that are saved during the training phase, and then automatically recomputing missing activations from those previously recorded. We propose an original computation model that combines two types of activation savings: either only storing the layer inputs, or recording the complete history of operations that produced the outputs (this uses more memory, but requires fewer recomputations in the backward phase), and we provide in https://hal.inria.fr/hal-02352969 an algorithm to compute the optimal computation sequence for this model.

Our PyTorch implementation processes the entire chain, dealing with any sequential DNN whose internal layers may be arbitrarily complex and automatically executing it according to the optimal checkpointing strategy computed given a memory limit. In https://hal.inria.fr/hal-02352969, through extensive experiments, we show that our implementation consistently outperforms existing checkpoint-ing approaches for a large class of networks, image sizes and batch sizes.

Functional Description: Allows to train very large convolutional networks on limited memory by optimally selecting which activations should be kept and which should be recomputed. This code is meant to replace the checkpoint.py utility available in pytorch, by providing more efficient rematerialization strategies. The algorithm is easier to tune: the only required parameter is the available memory, instead of the number of segments.

URL: `https://gitlab.inria.fr/hiepacs/rotor`

Publication: hal-02352969

Contact: Lionel Eyraud Dubois

Participants: Olivier Beaumont, Alena Shilova, Alexis Joly, Lionel Eyraud Dubois, Julien Herrmann

6.1.14 ScalFMM

Name: Scalable Fast Multipole Method

Keywords: N-body, Fast multipole method, Parallelism, MPI, OpenMP

Scientific Description: ScalFMM is a software library to simulate N-body interactions using the Fast Multipole Method. The library offers two methods to compute interactions between bodies when the potential decays like 1/r. The first method is the classical FMM based on spherical harmonic expansions and the second is the Black-Box method which is an independent kernel formulation (introduced by E. Darve @ Stanford). With this method, we can now easily add new non oscillatory kernels in our library. For the classical method, two approaches are used to decrease the complexity of the operators. We consider either matrix formulation that allows us to use BLAS routines or rotation matrix to speed up the M2L operator.

ScalFMM intends to offer all the functionalities needed to perform large parallel simulations while enabling an easy customization of the simulation components: kernels, particles and cells. It works in parallel in a shared/distributed memory model using OpenMP and MPI. The software architecture has been designed with two major objectives: being easy to maintain and easy to understand. There is two main parts:

the management of the octree and the parallelization of the method the kernels. This new architecture allow us to easily add new FMM algorithm or kernels and new paradigm of parallelization.

Functional Description: Compute N-body interactions using the Fast Multipole Method for large number of objects

URL: `https://gitlab.inria.fr/solverstack/ScalFMM`

Contact: Olivier Coulaud

Participants: Bramas Bérenger, Olivier Coulaud, Pierre Estérie

6.1.15 VITE

Name: Visual Trace Explorer

Keywords: Visualization, Execution trace

Functional Description: ViTE is a trace explorer. It is a tool made to visualize execution traces of large parallel programs. It supports Pajé, a trace format created by Inria Grenoble, and OTF and OTF2 formats, developed by the University of Dresden and allows the programmer a simpler way to analyse, debug and/or profile large parallel applications.

URL: https://solverstack.gitlabpages.inria.fr/vite/

Contact: Mathieu Faverge

Participant: Mathieu Faverge

6.1.16 PlaFRIM

Name: Plateforme Fédérative pour la Recherche en Informatique et Mathématiques

Keywords: High-Performance Computing, Hardware platform

Functional Description: PlaFRIM is an experimental platform for research in modeling, simulations and high performance computing. This platform has been set up from 2009 under the leadership of Inria Bordeaux Sud-Ouest in collaboration with computer science and mathematics laboratories, respectively Labri and IMB with a strong support in the region Aquitaine.

It aggregates different kinds of computational resources for research and development purposes. The latest technologies in terms of processors, memories and architecture are added when they are available on the market. It is now more than 1,000 cores (excluding GPU and Xeon Phi) that are available for all research teams of Inria Bordeaux, Labri and IMB. This computer is in particular used by all the engineers who work in HiePACS and are advised by F. Rue from the SED.

URL: https://www.plafrim.fr/en/home/

Contact: Olivier Coulaud

7 New results

7.1 High-performance computing on next generation architectures

7.1.1 Resiliency in numerical algorithm design for extreme scale simulations

This work is based on the seminar titled “Resiliency in Numerical Algorithm Design for Extreme Scale Simulations” held March 1-6, 2020 at Schloss Dagstuhl, that was attended by all the authors. Advanced supercomputing is characterized by very high computation speeds at the cost of involving an enormous amount of resources and costs. A typical large-scale computation running for 48 hours on a system consuming 20 MW, as predicted for exascale systems, would consume a million kWh, corresponding to about 100k Euro in energy cost for executing 1023 floating-point operations. It is clearly unacceptable to lose the whole computation if any of the several million parallel processes fails during the execution. Moreover, if a single operation suffers from a bit-flip error, should the whole computation be declared invalid? What about the notion of reproducibility itself: should this core paradigm of science be revised and refined for results that are obtained by large scale simulation? Naive versions of conventional resilience techniques will not scale to the exascale regime: with a main memory footprint of tens of Petabytes, synchronously writing checkpoint data all the way to background storage at frequent intervals will create intolerable overheads in runtime and energy consumption. Forecasts show that the mean time between failures could be lower than the time to recover from such a checkpoint, so that large calculations at scale might not make any progress if robust alternatives are not investigated. More advanced resilience techniques must be devised. The key may lie in exploiting both advanced system features as well as specific application knowledge. Research will face two essential questions: (1) what are the reliability requirements for a particular computation and (2) how do we best design the algorithms and software to meet these requirements? While the analysis of use cases can help understand the particular reliability requirements, the construction of remedies is currently wide open. One avenue would be to refine and improve on system- or application-level checkpointing and rollback strategies in the case an error is detected. Developers might use fault notification interfaces and flexible runtime systems to respond to node failures in an application-dependent fashion. Novel numerical algorithms or more stochastic computational approaches may be required to meet accuracy requirements in the face of undetectable soft errors. These ideas constituted an essential topic of the seminar. The goal of this Dagstuhl Seminar was to bring together a diverse group of scientists with expertise in exascale computing to discuss novel ways to make applications resilient against detected and undetected faults. In particular, participants explored the role that algorithms and applications play in the holistic approach needed to tackle this challenge. This article gathers a broad range of perspectives on the role of algorithms, applications, and systems in achieving resilience for extreme scale simulations. The ultimate goal is to spark novel ideas and encourage the development of concrete solutions for achieving such resilience holistically.

For more information on this work we refer to [15].

7.1.2 Resilience and fault tolerance in high-performance computing for numerical weather and climate prediction

Progress in numerical weather and climate prediction accuracy greatly depends on the growth of the available computing power. As the number of cores in top computing facilities pushes into the millions, increased average frequency of hardware and software failures forces users to review their algorithms and systems in order to protect simulations from breakdown. This report surveys hardware, application-level and algorithm-level resilience approaches of particular relevance to timecritical numerical weather and climate prediction systems. A selection of applicable existing strategies is analysed, featuring interpolation-restart and compressed checkpointing for the numerical schemes, in-memory checkpointing, user-level failure mitigation and backup-based methods for the systems. Numerical examples showcase the performance of the techniques in addressing faults, with particular emphasis on iterative solvers for linear systems, a staple of atmospheric fluid flow solvers. The potential impact of these strategies is discussed in relation to current development of numerical weather prediction algorithms and systems towards the exascale. Trade-offs between performance, efficiency and effectiveness of resiliency strategies are analysed and some recommendations outlined for future developments.

For more information on this work we refer to [16].

7.1.3 TEXTAROSSA: Towards EXtreme scale Technologies and Accelerators for euROhpc hw/Sw Supercomputing Applications for exascale

To achieve high performance and high energy efficiency on near-future exascale computing systems, three key technology gaps needs to be bridged. These gaps include: energy efficiency and thermal control; extreme computation efficiency via HW acceleration and new arithmetics; methods and tools for seamless integration of reconfigurable accelerators in heterogeneous HPC multi-node platforms. TEXTAROSSA aims at tackling this gap through a co-design approach to heterogeneous HPC solutions, supported by the integration and extension of HW and SW IPs, programming models and tools derived from European research.

More information on this project can be found in [21].

7.2 High performance solvers for large linear algebra problems

7.2.1 Comparison of coupled solvers for FEM/BEM linear systems arising from discretization of aeroacoustic problems

When discretization of an aeroacoustic physical model is based on the application of both the Finite Elements Method (FEM) and the Boundary Elements Method (BEM), this leads to coupled FEM/BEM linear systems combining sparse and dense parts. In this work, we propose and compare a set of implementation schemes relying on the coupling of the open-source sparse direct solver MUMPS with the proprietary direct solvers from Airbus Central R&T, i.e. the scalapack-like dense solver SPIDO and the hierarchical $\mathscr{H}$-matrix compressed solver HMAT. For this preliminary study, we limit ourselves to a single 24-core computational node.

For more information on this work we refer to [26] and [28].

7.2.2 Recycling Krylov subspace strategies for sequences of sampled stochastic elliptic equations

We are interested in the quantification of uncertainties in discretized elliptic partial differential equations with a random coefficient field. In sampling-based approaches, this relies on solving large numbers of symmetric positive definite (SPD) linear systems with different matrices. In particular, we consider the case in which these operators are sampled by Markov chain Monte Carlo, which leads to sequences of correlated matrices. We investigate recycling Krylov subspace strategies for the iterative solution of sequences of linear systems formed with such matrices. The linear systems are solved using initialized conjugate gradient (Init-CG) methods, where approximate eigenvectors of the previously sampled operator are used to set an initial guess, and deflated conjugate gradient (Def-CG) methods, where the Krylov subspace is augmented with these vectors. The following aspects of eigenvector approximation, and their effect on deflation and initialization, are investigated in this context: (i) projection technique, and (ii) refreshing strategy of the eigen-search space. Our numerical experiments show that, when not using a preconditioner, these aspects only impact convergence behaviors of Def-CG at the early stages of the sampling sequence. Second, unlike sequences with multiple right-hand sides and a constant operator, our experiments with multiple matrices show that, even for highly correlated matrices, Init-CG does not reproduce the convergence behavior of Def-CG. Finally, the limits of deflation used as a means to compensate for the inefficiency of block-Jacobi (bJ) preconditioners are investigated. For small systems, using a bJ preconditioner while deflating with at least as many approximate eigenvectors as the number of bJ blocks achieves similar convergence behaviors to PCG with a constant algebraic multigrid (AMG) preconditioner. For larger systems, although the effect of deflation is improved when using the right refreshing strategy of the eigen-search space, the combination of deflation with bJ preconditioners does not scale as well as using PCG with a constant AMG preconditioner based on the median realization of the coefficient field.

For more information on this work we refer to [33].

7.2.3 A block minimum residual norm subspace solver with partial convergence management for sequences of linear systems

We are concerned with the iterative solution of linear systems with multiple right-hand sides available one group after another, including the case where there are massive number (like tens of thousands) of right-hand sides associated with a single matrix so that all of them cannot be solved at once but rather need to be split into chunks of possible variable sizes. For such sequences of linear systems with multiple left and right-hand sides, we develop a new recycling block generalized conjugate residual method with inner orthogonalization and inexact breakdown (IB-BGCRO-DR), which glues subspace recycling technique in GCRO-DR [SIAM J. Sci. Comput., 28(5) (2006), pp. 1651–1674] and inexact breakdown mechanism in IB-BGMRES [Linear Algebra Appl., 419 (2006), pp. 265–285] to guarantee this new algorithm could reuse spectral information for subsequent cycles as well as for the remaining linear systems to be solved. Related variant IB-BFGCRO-DR that suits for flexible preconditioning is devised to cope with constraints on some applications while also enabling mixed-precision calculation, which provides advantages in speed and memory usage over double precision as well as in perspective of emerging computing units such as the GPUs.

For more information on this work we refer to [31] that is a preliminary version of a paper recently accepted in SIMAX.

7.2.4 Deciding Non-Compressible Blocks in Sparse Direct Solvers using Incomplete Factorization

Low-rank compression techniques are very promising for reducing memory footprint and execution time on a large spectrum of linear solvers. Sparse direct supernodal approaches are one of these techniques. However, despite providing a very good scalability and reducing the memory footprint, they suffer from an important flops overhead in their unstructured low-rank updates. As a consequence, the execution time is not improved as expected. In this paper, we study a solution to improve low-rank compression techniques in sparse supernodal solvers. The proposed method tackles the overprice of the low-rank updates by identifying the blocks that have poor compression rates. We show that the fill-in levels of the graph based block incomplete LU factorization can be used in a new context to identify most of these non-compressible blocks at low cost. This identification enables to postpone the low-rank compression step to trade small extra memory consumption for a better time to solution. The solution is validated within the PaStiX library with a large set of application matrices. It demonstrates sequential and multithreaded speedup up to 8.5×, for small memory overhead of less than 1.49× with respect to the original version.

For more information on this work we refer to [25].

7.3 Linear algebra in tensor spaces

7.3.1 Solving eigenvalue problems using contour integration in Tensor Train format

In high-dimension, solving an eigenvalue problem encounters several issues related to the curse of dimensionality especially when only eigenvalues in a given interval are desired. To overcome these difficulties, we consider the FEAST algorithm that is developed for solving a Hermitian eigenproblem inside a given interval and based on a contour integration that projects the matrix pencil onto the subspace associated with the eigenpairs is the desired interval. We also consider the tensor-train decomposition (TT) for the operators and the eigenvectors to overcome the curse of dimensionality and reduce the required storage. In this paper, we extend the FEAST algorithm to match with the TT representation of operators and vectors to form an algorithm that computes the eigenvalue problems in an interval based on operations in TT-format. The proposed algorithm is applied on some high-dimensional problems, including the Laplacian operator and the vibrational Hamiltonian, and it shows a high efficiency and accuracy. We validate the results by comparing with an analytical solution for Laplacian operator, and with an existing method in TT-format for computing certain minimal eigenpairs.

The scientific report describing this work should be published in early 2022.

7.3.2 Iterative linear solver in tensorial spaces

In this work, we studied the solution of linear systems defined in tensor spaces of different dimensions. Specifically, we considered nested subspace techniques in tensor format for the solution of 3d (PDEs in a 3-dimensional space), 4d (parametric PDEs, or 3D PDEs with multiple right-hand sides), 5d (time-dependent parametric PDEs) linear problems. In particular, we have derived bounds to evaluate the quality of the solutions computed in low-rank tensor format compared to their more classical linear algebra counterparts.

The scientific report describing this work should be published in early 2022.

7.4 Tensor calculations for data analysis

7.4.1 High Order Singular Value Decomposition for Plant Biodiversity Estimation

We propose a new method to estimate plant biodiversity with Rényi and Rao indexes through the so called High Order Singular Value Decomposition (HOSVD) of tensors. Starting from NASA multispectral images we evaluate biodiversity and we compare original biodiversity estimates with those realised via the HOSVD compression methods for big data. Our strategy turns out to be extremely powerful in terms of storage memory and precision of the outcome. The obtained results are so promising that we can support the efficiency of our method in the ecological framework.

For more information on this work we refer to [17].

7.4.2 Correspondence Analysis to multiway data-sets through High Order SVD

In this work we propose an extension of Correspondence Analysis (CA) to tensors through High Order Singular Value Decomposition (HOSVD) from a geometric viewpoint. Correspondence analysis is a well-known tool, developed from principal component analysis, for studying contingency tables. Different algebraic extensions of CA to multi-way tables have been proposed over the years, nevertheless neglecting its geometric meaning. Relying on the Tucker model and the HOSVD, we propose a direct way to associate with each tensor mode a point cloud. We prove that the point clouds are related to each other. Specifically using the CA metrics we show that the barycentric relation is still true in the tensor framework. Finally two data sets are used to underline the advantages and the drawbacks of our strategy with respect to the classical matrix approaches.

For more information on this work we refer to [30].

7.5 Applications domains

7.5.1 A note on the strong parallel scalability of numerically scalable Poisson linear solvers

In the context of a parallel plasma physics simulation code, we perform a qualitativeperformance study between two natural candidates for the parallel solution of 3D Poisson problemsthat are multigrid and domain decomposition. We selected one representative of each of thesenumerical techniques implemented in state of the art parallel packages and show that dependingon the regime used in terms of number of unknowns per computing cores the best alternative interms of time to solution varies. Those results show the interest of having both types of numericalsolvers integrated in a simulation code that can be used in very different configurations in termsof selected modelisations, problem sizes and parallel computing platforms.

For more information on this work we refer to [29].

7.5.2 The JOREK non-linear extended MHD code and applications to large-scale instabilities and their control in magnetically confined fusion plasmas

JOREK is a massively parallel fully implicit non-linear extended magneto-hydrodynamic (MHD) code for realistic tokamak X-point plasmas. It has become a widely used versatile simulation code for studying large-scale plasma instabilities and their control and is continuously developed in an international community with strong involvements in the European fusion research programme and ITER organization. This article gives a comprehensive overview of the physics models implemented, numerical methods applied for solving the equations and physics studies performed with the code.

For more information on this work we refer to [19].

7.6 Optimization for the training phase of deep convolutional networks

7.6.1 Pipelined Model Parallelism: Complexity Results and Memory Considerations

The training phase in Deep Neural Networks has become an important source of computing resource usage and the resulting volume of computation makes it crucial to perform efficiently on parallel architectures. Data parallelism is the most widely used method, but it requires to replicate the network weights on all processors, and to perform collective communications of the network weights. In this context, model parallelism is an attractive alternative, in which the different layers of the network are distributed over the computing processors. Indeed, it is expected to better distribute weights (to cope with memory problems) and it eliminates the need for large collective communications since only forward activations are communicated. However, to be efficient, it must be combined with a pipelined approach, which in turn induces new memory costs. We have thus worked to formalize pipelined model parallelism as a scheduling problem, to establish its complexity, and to analyze the consequences of the assumptions that are typically performed in practical solutions such as Pipedream.

More information on this work can be found in [23].

7.6.2 Efficient Combination of Rematerialization and Offloading for Training DNNs

Rematerialization and offloading are two well known strategies to save memory during the training phase of deep neural networks, allowing data scientists to consider larger models, batch sizes or higher resolution data. Rematerialization trades memory for computation time, whereas Offloading trades memory for data movements. As these two resources are independent, it is appealing to consider the simultaneous combination of both strategies to save even more memory. We precisely model the costs and constraints corresponding to Deep Learning frameworks such as PyTorch or Tensorflow, we propose optimal algorithms to find a valid sequence of memory-constrained operations and finally, we evaluate the performance of proposed algorithms on realistic networks and computation platforms. Our experiments show that the possibility to offload can remove one third of the overhead of rematerialization, and that together they can reduce the memory used for activations by a factor 4 to 6, with an overhead below 20

More information on this work can be found in [22].

7.7 Scheduling

7.7.1 READYS: A Reinforcement Learning Based Strategy for Heterogeneous Dynamic Scheduling

We have proposed READYS, a reinforcement learning algorithm for the dynamic scheduling of computations modeled as a Directed Acyclic Graph (DAGs). Our goal is to develop a scheduling algorithm in which allocation and scheduling decisions are made at runtime, based on the state of the system, as performed in runtime systems such as StarPU or ParSEC. Reinforcement Learning is a natural candidate to achieve this task, since its general principle is to build step by step a strategy that, given the state of the system (the state of the resources and a view of the ready tasks and their successors in our case), makes a decision to optimize a global criterion. Moreover, the use of Reinforcement Learning is natural in a context where the duration of tasks (and communications) is stochastic. We propose READYS that combines Graph Convolutional Networks (GCN) with an Actor-Critic Algorithm (A2C): it builds an adaptive representation of the scheduling problem on the fly and learns a scheduling strategy, aiming at minimizing the makespan. A crucial point is that READYS builds a general scheduling strategy which is neither limited to only one specific application or task graph nor one particular problem size, and that can be used to schedule any DAG. We focus on different types of task graphs originating from linear algebra factorization kernels (CHOLESKY, LU, QR) and we consider heterogeneous platforms made of a few CPUs and GPUs. We first propose to analyze the performance of READYS when learning is performed on a given (platform, kernel, problem size) combination. Using simulations, we show that the scheduling agent obtains performance very similar or even superior to algorithms from the literature, and that it is especially powerful when the scheduling environment contains a lot of uncertainty. We additionally demonstrate that our agent exhibits very promising generalization capabilities. To the best of our knowledge, this is the first paper which shows that reinforcement learning can really be used for dynamic DAG scheduling on heterogeneous resources.

More information on this work can be found in [24].

8 Bilateral contracts and grants with industry

8.1 Bilateral contracts with industry

Some on the ongoing PhD thesis are developed within bilareal contract with industry for PhD advisory:

Airbus for the PhD thesis of Marek Felsoci,

IFPEN for the PhD of Aboul-Karim Mohamed El Maarouf,

CEA-Cesta for the PhD of Clément Richefort.

9 Partnerships and cooperations

9.1 International initiatives

9.1.1 Inria associate team not involved in an IIL or an international program

MOLIERE

Title: Memory Optimization for paraLlel traIning of dEep neuRal nEtworks

Duration: 2020 ->

Coordinator: Ivan Oseledets (I.Oseledets@skoltech.ru)

Partners:

- Skoltech – Skolkovo Institute of Science and Technology

Inria contact: Olivier Beaumont

Summary: Training for Deep Learning Networks (DNNs) has become a major compute intensive application. Nevertheless, even if training is still in general performed on small clusters of GPU machines, the use of large HPC infrastructures is becoming popular, in particular because they offer high bandwidth and low latency networks. An important limitation of hyper-parameter tuning and data parallelism approaches is that they do not help in considering larger models by solving memory issues during the training phase. Indeed, in both settings, the whole set of weights has to be stored on all participating resources. In general, the memory consumed by the training phase consists in two main parts. The first part is related to the storage of the parameters of the network and is directly related to the size of the model. For both hyper-parameters and data parallelism approaches, these weights have to be replicated on every node, and, in the case of data parallelism, they even have to be aggregated and broadcast on the network after each parallel mini-batch training phase. The second source of memory consumption is due to the storage on each node of all the forward activations (i.e. all the outputs of the different stages of the network) until the backward phase. This part is directly proportional to the size of the mini-batch itself. In this context, our goal in the framework of this associated team is to provide new techniques to both increase parallelism while limiting memory requirements, by exploring techniques based on compression and tensor train decomposition and checkpointing and offloading techniques to control the memory consumed during the training phase.

9.2 European initiatives

9.2.1 FP7 & H2020 projects

EoCoE-2

Title: Energy oriented Centre of Excellence for computer applications

Duration: 2018-2022

Coordinator: CEA

Inria coordinator: Bruno Raffin

HIEPACS **contact:** Luc Giraud

Partners:

- AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE, L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE (Italy)
- BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION (Spain)
- CENTRE EUROPEEN DE RECHERCHE ET DE FORMATION AVANCEE EN CALCUL SCIENTIFIQUE (France)
- CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS (France)
- COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES (France)
- CONSIGLIO NAZIONALE DELLE RICERCHE (Italy)
- FORSCHUNGSZENTRUM JULICH GMBH (Germany)
- FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V. (Germany)
- MAX-PLANCK-GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV (Germany)
- RHEINISCH-WESTFAELISCHE TECHNISCHE HOCHSCHULE AACHEN (Germany)
- THE CYPRUS INSTITUTE (Cyprus)
- UNIVERSITA DEGLI STUDI DI ROMA TORVERGATA (Italy)
- UNIVERSITA DEGLI STUDI DI TRENTO (Italy)

- UNIVERSITE LIBRE DE BRUXELLES (Belgium)
- UNIVERSITE PARIS-SUD (France)
- UNIVERSITY OF BATH (UK)

Inria contact: Bruno Raffin (Datamove)

Summary: The aim of the present proposal is to establish an Energy Oriented Centre of Excellence for computing applications, (EoCoE). EoCoE (pronounce "Echo") will use the prodigious potential offered by the ever-growing computing infrastructure to foster and accelerate the European transition to a reliable and low carbon energy supply. To achieve this goal, we believe that the present revolution in hardware technology calls for a similar paradigm change in the way application codes are designed. EoCoE will assist the energy transition via targeted support to four renewable energy pillars: Meteo, Materials, Water and Fusion, each with a heavy reliance on numerical modelling. These four pillars will be anchored within a strong transversal multidisciplinary basis providing high-end expertise in applied mathematics and HPC. EoCoE is structured around a central Franco-German hub coordinating a pan-European network, gathering a total of 8 countries and 23 teams. Its partners are strongly engaged in both the HPC and energy fields; a prerequisite for the long-term sustainability of EoCoE and also ensuring that it is deeply integrated in the overall European strategy for HPC. The primary goal of EoCoE is to create a new, long lasting and sustainable community around computational energy science. At the same time, EoCoE is committed to deliver high-impact results within the first three years. It will resolve current bottlenecks in application codes, leading to new modelling capabilities and scientific advances among the four user communities; it will develop cutting-edge mathematical and numerical methods, and tools to foster the usage of Exascale computing. Dedicated services for laboratories and industries will be established to leverage this expertise and to foster an ecosystem around HPC for energy. EoCoE will give birth to new collaborations and working methods and will encourage widely spread best practices.

PRACE 6IP

Title: PRACE Sixth Implementation Phase

Duration: 2019-2022

Partners: see list

Inria contact: Luc Giraud

Summary: The mission of PRACE (Partnership for Advanced Computing in Europe) is to enable high-impact scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society. PRACE seeks to realise this mission by offering world class computing and data management resources and services through a peer review process. PRACE also seeks to strengthen the European users of HPC in industry through various initiatives. PRACE has a strong interest in improving energy efficiency of computing systems and reducing their environmental impact. The objectives of PRACE-6IP are to build on and seamlessly continue the successes of PRACE and start new innovative and collaborative activities proposed by the consortium. These include: assisting the development of PRACE 2; strengthening the internationally recognised PRACE brand; continuing and extend advanced training which so far provided more than 36 400 person·training days; preparing strategies and best practices towards Exascale computing, work on forward-looking SW solutions; coordinating and enhancing the operation of the multi-tier HPC systems and services; and supporting users to exploit massively parallel systems and novel architectures. The activities are designed to increase Europe's research and innovation potential especially through: seamless and efficient Tier-0 services and a pan-European HPC ecosystem including national capabilities; promoting take-up by industry and new communities and special offers to SMEs; assistance to PRACE 2 development; proposing strategies for deployment of leadership systems; collaborating with the ETP4HPC, CoEs and other European and international organisations on future architectures,

RISC2

Title: A network for supporting the coordination of High-Performance Computing research between Europe and Latin America

Type: H2020 (Coordinated Support Action)

URL: See also: list

Duration: 2021 - 2023

Coordinator: Barcelona Supercomputing Center (Spain)

Inria coordinator: Stéphane Lanteri

HIEPACS **contact:** Luc Giraud

Partners:

- Forschungzentrum Julich GMBH (Germany)
- Inria (France)
- Bull SAS (France)
- INESC TEC (Portugal)
- Universidade de Coimbra (Portugal)
- CIEMAT (Spain)
- CINECA (Italy)
- Universidad de Buenos Aires (Argentina)
- Universidad Industrial de Santander (Columbia)
- Universidad de le Republica (Uruguay)
- Laboratorio Nacional de Computacao Cientifica (Brazil)
- Centro de Investigacion y de Estudios Avanzados del Instituto Politecnico Nacional (Mexico)
- Universidad de Chile (Chile)
- Fundacao Coordenacao de Projetos Pesquisas e Estudos Tecnologicos COPPETEC (Brazil)
- Fundacion Centro de Alta Tecnologia (Costa Rica)

Summary: Recent advances in AI and the Internet of things allow high performance computing (HPC) to surpass its limited use in science and defence and extend its benefits to industry, healthcare and the economy. Since all regions intensely invest in HPC, coordination and capacity sharing are needed. The EU-funded RISC2 project connects eight important European HPC actors with the main HPC actors from Argentina, Brazil, Chile, Colombia, Costa Rica, Mexico and Uruguay to enhance cooperation between their research and industrial communities on HPC application and infrastructure development. The project will deliver a cooperation roadmap addressing policy-makers and the scientific and industrial communities to identify central application areas, HPC infrastructure and policy needs.

9.3 National initiatives

9.3.1 ANR

OPERA (Adpative planar optics - ANR ASTRID Maturation)

Duration: 2019 – 2022

Coordinator: Stéphane Lanteri (Atlantis - SAM)

HIEPACS **contact:** Luc Giraud

Summary: In the OPERA project, we are investigating and optimizing the properties of planar photonic devices based on metasurfaces using numerical modelling. The scientific and technical activities that constitute the project work programme are organized around 4 main workpackages. The numerical characterization of the optical properties of planar devices based on metasurfaces, as well as their optimization are at the heart of the activities and objectives of two horizontal (transversal) workpackages. These numerical methodologies will be integrated into the DIOGENeS software framework that will eventually integrates (1) discontinuous Galerkin-type methods that have been tested over the past 10 years for the discretization of Maxwell equations in time and frequency regimes, mainly for applications in the microwave band, (2) parallel resolution algorithms for sparse linear systems based on the latest developments in numerical linear algebra, (3) modern optimization techniques based on learning and metamodeling methods and (4) software components adapted to modern high performance computing architectures. Two vertical workpackages complete this program. One of them aims to demonstrate the contributions of methodological developments and numerical tools resulting from transversal workpackages through their application to diffusion/radiation control by passive planar devices. The other, more prospective, concerns the study of basic building blocks for the realization of adaptive planar devices.

SASHIMI**: Sparse Direct Solver using Hierarchical Matrices**

Duration: 2018 – 2022

Coordinator: Mathieu Faverge

Summary: Nowadays, the number of computational cores in supercomputers has grown largely to a few millions. However, the amount of memory available has not followed this trend, and the memory per core ratio is decreasing quickly with the advent of accelerators. To face this problem, the SaSHiMi project wants to tackle the memory consumption of linear solver libraries used by many major simulation applications by using low-rank compression techniques. In particular, the direct solvers which offer the most robust solution to strategy but suffer from their memory cost. The project will especially investigate the super-nodal approaches for which low-rank compression techniques have been less studied despite the attraction of their large parallelism and their lower memory cost than for the multi-frontal approaches. The results will be integrated in the PaStiX solver that supports distributed and heterogeneous architectures.

SOLHARIS**: SOLvers for Heterogeneous Architectures over Runtime systems, Investigating Scalability**

Duration: 2018 – 2022

Coordinator: Alfredo Buttari (IRIT)

HiePACS **contact:** Abdou Guermouche

Partners:

- IRIT Institut de Recherche en Informatique de Toulouse
- Inria Bordeaux - Sud-Ouest and Lyon
- Airbus Central R&T
- CEA Commissariat à l'énergie atomique et aux énergies alternatives

Summary: The SOLHARIS project aims at addressing the issues related to the development of fast and scalable linear solvers for large-scale, heterogeneous supercomputers. Because of the complexity and heterogeneity of the targeted algorithms and platforms, this project intends to rely on modern runtime systems to achieve high performance, programmability and portability. By gathering experts in computational linear algebra, scheduling algorithms and runtimes, SOLHARIS intends to tackle these issues through a considerable research effort for the development of numerical algorithms and scheduling methods that are better suited to the characteristics of large scale, heterogeneous systems and for the improvement and extension of runtime systems with novel features that more accurately fulfill the requirements of these methods. This is expected to lead to fundamental research results and software of great interest for researchers of the scientific computing community.

9.3.2 FUI

ICARUS: Intensive Calculation for AeRo and automotive engines Unsteady Simulations

Duration: 2018 – 2022

Coordinator: SAFRAN

Inria contact: Aurélien Esnard

Partners:

- CENAERO
- CERFACS
- CORIA
- DISTENE
- GDTECH
- IFPEN
- ONERA
- SAFRAN
- SIEMENS

Summary: Large Eddy Simulation (LES) is an increasingly attractive unsteady modelling approach for modelling reactive turbulent flows due to the constant development of massively parallel supercomputers. It can provide open and robust design tools that allow access to new concepts (technological breakthroughs) or a global consideration of a structure (currently processed locally). The mastery of this method is therefore a major competitive lever for industry. However, it is currently constrained by its access and implementation costs in an industrial context. The ICARUS project aims to significantly reduce them (costs and deadlines) by bringing together major industrial and research players to work on the entire high-fidelity LES computing process by:

- increasing the performance of existing reference tools (for 3D codes: AVBP, Yales2, ARGO) both in the field of code coupling and code/machine matching;
- developing methodologies and networking tools for the LES;
- adapting the ergonomics of these tools to the industrial world: interfaces, data management, code interoperability and integrated chains;
- validating this work on existing demonstrators, representative of the aeronautics and automotive industries.

9.3.3 Inria Project Labs

Challenge HPC BigData

Duration: 2018 – 2022

Coordinator: Bruno Raffin

HIEPACS **contact:** Olivier Beaumont & Olivier Coulaud

Inria teams:

- KerData
- SequeL
- Sierra
- Tau
- Zenith
- Parietal
- TADaaM
- HiePACS
- Storm

Summary: The goal of the Challenge on HPC-BigData is to gather teams from the HPC, Big Data and Machine Learning (ML) areas to work at the intersection between these domains. HPC and Big Data evolved with their own infrastructures (supercomputers versus clouds), applications (scientific simulations versus data analytics) and software tools (MPI and OpenMP versus Map/Reduce or Deep Learning frameworks). But Big Data analytics is becoming more compute-intensive (thanks to deep learning), while data handling is becoming a major concern for scientific computing. Within the IPL, we are in particular involved in a tight collaboration with Zenith Team (Montpellier) on how to parallelize and how to deal with memory issues in the context of the training phase of Pl@ntnet (`https://www.plantnet.org`). Alexis Joly (Zenith) co supervises with Olivier Beaumont the PhD Thesis of Alena Shilova. We are also collaborating with Sequel (Nathan Grinsztajn and Philippe Preux) and Tadaam (Emmanuel Jeannot) teams on the design of dynamic runtime schedulers based on reinforcement learning.

9.4 Regional initiatives

Title: HPC-Ecosystem

Duration: 2018-2020

Coordinator: Emmanuel Agullo

Partners:

- STORM, TADAAM from Inria Bordeaux Sud-Ouest
- Airbus
- CEA-CESTA

Description: Numerical simulation is today integrated in all cycles of scientific design and studies, whether academic or industrial, to predict or understand the behavior of complex phenomena often coupled or multi-physical. The quality of the prediction requires having precise and adapted models, but also to have computation algorithms efficiently implemented on computers with architectures in permanent evolution. Given the ever increasing size and sophistication of simulations implemented, the use of parallel computing on computers with up to several hundred thousand computing cores and consuming / generating massive volumes of data becomes unavoidable; this domain corresponds to what is now called High Performance Computing (HPC). On the other hand, the digitization of many processes and the proliferation of connected objects of all kinds generate ever-increasing volumes of data that contain multiple valuable information; these can only be highlighted through sophisticated treatments; we are talking about Big Data. The intrinsic complexity of these digital treatments requires a holistic approach with collaborations of multidisciplinary teams capable of mastering all the scientific skills required for each component of this chain of expertise. To have a real impact on scientific progress and advances, these skills must include the efficient management of the massive number of compute nodes using programming paradigms with a high level of expressiveness, exploiting high-performance communications layers, effective management for intensive I / O, efficient scheduling mechanisms on platforms with a large number of computing units and massive I / O volumes, innovative and powerful numerical methods for analyzing volumes of data produced and efficient algorithms that can be integrated into applications representing recognized scientific challenges with high societal and economic impacts. The project we propose aims to consider each of these links in a consistent, coherent and consolidated way. For this purpose, we propose to develop a unified Execution Support (SE) for large-scale numerical simulation and the processing of large volumes of data. We identified four Application Challenges (DA) identified by the Nouvelle-Aquitaine region that we propose to carry over this unified support. We will finally develop four Methodological Challenges (CM) to evaluate the impact of the project. This project will make a significant contribution to the emerging synergy on the convergence between two yet relatively distinct domains, namely High Performance Computing (HPC) and the processing, management of large masses of data (Big Data); this project is therefore clearly part of the emerging field of High Performance Data Analytics (HPDA).

10 Dissemination

10.1 Promoting scientific activities

10.1.1 Scientific events: organisation

Member of the organizing committees

Luc Giraud is member of the Gene Golub SIAM Summer School. The eleventh Gene Golub SIAM Summer School was entilted “Theory and Practice of Deep Learning”.

10.1.2 Scientific events: selection

Chair of conference program committees

SBAC-PAD'21 (O. Beaumont, Program Co-Chair)

Member of the conference program committees

IPDPS'21 (O. Beaumont, O. Coulaud, L. Eyraud-Dubois, M. Faverge, A. Guermouche), PDSEC'21 (O. Coulaud, L. Giraud), SC'21 (O. Beaumont)

10.1.3 Journal

Member of the editorial boards

- O. Beaumont is Associate Editor in Chief for the Journal of Parallel and Distributed Computing (Elsevier).
- L. Giraud is member of the editorial board of the SIAM Journal on Scientific Computing (SISC).

Reviewer - reviewing activities

The members of the HIEPACS project have performed reviewing for the following list of journals: Computer and fluids, SIAM J. Matrix Analysis and Applications, SIAM J. Scientific Computing, Cluster Computing, Concurrency and Computation: Practice and Experience, ACM Transactions on Parallel Computing, Applied Numerical Mathematics, Journal of Computational Physics, Journal of Scheduling,

10.1.4 Scientific expertise

- Olivier Beaumont acted as expert for 2020 FET Open Evaluation Committee.

- Luc Giraud is member of the board on Modelization, Simulation and data analysis of the Competitiveness Cluster for Aeronautics, Space and Embedded Systems. He is also member of the scientific council of the ONERA Lab LMA2S (Laboratoire de Mathématiques Appliquées à l'Aéronautique et au Spatial).

  He also acted as expert for the Czech Science Foundation.

- Pierre Ramet is Scientific Advisor at the CEA-DAM CESTA.

- Guillaume Sylvand is Expert in Numerical Simulation and HPC at Airbus.

10.1.5 Research administration

- P. Ramet is the head of the SATANAS (Supports and Algorithms for High Performance Numerical Applications) team from LaBRI UMR CNRS 5800.

10.2 Teaching - Supervision - Juries

10.2.1 Teaching

- Undergraduate level/Licence:
  - A. Esnard: System programming 36h, Computer architecture 40h, Network 23h, C programming 35h at Bordeaux University. He is also responsible for the second year of the computer science degree (L2 Informatique), which involves managing about 200 students each year.
  - M. Faverge: Programming environment 26h, Numerical algorithmic 40h, C projects 25h at Bordeaux INP (ENSEIRB-MatMeca).
  - A. Guermouche: System programming 36h at Bordeaux University.
  - P. Ramet: System programming 24h, Databases 32h, Object programming 48h, Distributed programming 16h, Cryptography 16h, Introduction to AI Deep Learning and Data Analytics 16h at Bordeaux University.

- Post graduate level/Master:
  - E. Agullo: Operating systems 24h at Bordeaux University ; Dense linear algebra kernels 8h, Numerical algorithms 30h at Bordeaux INP (ENSEIRB-MatMeca).
  - O. Coulaud: Paradigms for parallel computing 8h, Introduction to Tensor methods 4h at Bordeaux INP (ENSEIRB-MatMeca).
  - A. Esnard: Network management 27h, Network security 27h at Bordeaux University; Programming distributed applications 35h at Bordeaux INP (ENSEIRB-MatMeca).
  - L. Eyraud-Dubois and Olivier Beaumont: Approximation and BigData 24h at Bordeaux University.
  - M. Faverge: System programming 72h, Linear Algebra for high Performance Computing 13h at Bordeaux INP (ENSEIRB-MatMeca). He is also in charge of the master 2 internship for the Computer Science department at Bordeaux INP (ENSEIRB-MatMeca). Starting in September, he is in charge with Raymond Namyst of the High Performance Computing - High Performance Data Analytics option at ENSEIRB-MatMeca. This is a common training curriculum between the Computer Science and MatMeca departments at Bordeaux INP and with Bordeaux University in the context of the Computer Science Research Master.
  - Alena Shilova and Olivier Beaumont: Deep Learning Frameworks, at Bordeaux INP (ENSEIRB-MatMeca), 20h.
  - Olivier Beaumont, Sketching and Streaming Algorithms, ENS Lyon, 8h.
  - L. Giraud: Introduction to intensive computing and related programming tools 30h, INSA Toulouse; On mathematical tools for numerical simulations 10h, ENSEEIHT Toulouse.

  - A. Guermouche: Network management 92h, Network security 64h, Operating system 24h at Bordeaux University.
  - P. Ramet: Cryptography 20h and Numerical algorithms 40h at Bordeaux INP (ENSEIRB-Matmeca).

- High School teachers:

  - A. Esnard, M. Faverge, and A. Guermouche participated to the training of the High School teachers (DIU Enseigner l'Informatique au Lycée) in computer science for the new computer science program starting in September 2019.

10.2.2 Supervision

- PhD in progress: Tobias Castanet; Replication algorithms for multi-player virtual worlds; started Sep. 2019; O. Beaumont, N. Hanusse (LaBRI), C. Travers (Bordeaux INP - LaBRI).
- PhD in progress: Marek Felsoci; Fast solvers for high-frequency aeroacoustics; started Oct. 2019; G. Sylvand, E. Agullo.
- PhD in progress: Martina Iannacito; Linear solvers in tensorial format for high dimensional problems; started Oct 2019; O. Coulaud, L. Giraud.
- PhD in progress: Esragul Korkmaz; Sparse linear solver and hierachical matrices; started Nov. 2018; M. Faverge, P. Ramet.
- PhD in progress: Romain Peressoni; Fast multidimensional scaling method for the study of biodiversity; started Oct 2019; E. Agullo, O. Coulaud, A. Franc (PLEIADE).
- PhD in progress: Aboul-Karim Mohamed El Maarouf; Parallel fine grain imcomplete LU factorization for the solution of sparse linear systems; started: Dec. 2019; L. Giraud, A. Guermouche.
- PhD in progress: Clément Richefort; Multigrid methods applied to electromagnetism problems; started Nov. 2021; P. Ramet, M. Lecouvez (CEA Cesta).
- PhD in progress: Alena Shilova; Scheduling for deep learning applications; started Oct. 2018; L. Eyraud-Dubois, O. Beaumont.
- PhD in progress: Nicolas Venkovic; Domain decomposition techniques for the solution of stochastics elliptic PDEs; started Nov. 2018; L. Giraud, P. Mycek (Cerfacs).
- PhD in progress: Mathieu Vérité; Static allocation algorithms for scheduling High-Performance applications; started Sept. 2019; L. Eyraud-Dubois, O. Beaumont.
- PhD in progress: Yanfei Xiang; Solution of large linear systems with massive numbers of right-hand sides. Started Nov. 2019; L. Giraud, P. Mycek (Cerfacs).

10.2.3 Juries

- Ashish Bhole, "Stabilized C1-bicubic finite element method for nonlinear MHD modelling of tokamak plasma"; referees: Eric Serre, Eric Nardon; members: Dinshaw Balsara, Hervé Guillard, Boniface Nkonga, Stanislas Pamela, Pierre Ramet; Université Cote d'Azur, 17 Nov. 2021.
- J. F. Reis, "Preconditioning of domain decomposition methods for stochastic elliptic equations"; referees: Anthony Nouy, Luc Giraud; members: frédéric Hecht, Nicole Spillane, Paul Mycek, Pietro Congedo, Olivier Le Maitre; Institut Polytechnique de Paris - Ecole polytechnique, Spécialité: mathématiques appliquées; 4 Oct. 2021.
- K.E. Prikopa, "Fault tolerant linear least squares solvers and matrix multiplication in parallel and distributed environments"; referees: Jesper Larson Traff, Luc Giraud; members: Helmut Hlavacs, Wielfrid Gansterer; University of Vienna; 14 Oct. 2021.

10.3 Popularization

10.3.1 Internal or external Inria responsibilities

Pierre Ramet is member of the CDT (Technological Development Commission) at inria Bordeaux since 2015.

11 Scientific production

11.1 Major publications

[1] E. Agullo, B. Bramas, O. Coulaud, E. Darve, M. Messner and T. Takahashi. 'Task-Based FMM for Multicore Architectures'. In: *SIAM Journal on Scientific Computing* 36.1 (2014), pp. 66–93. DOI: 10.1137/130915662. URL: https://hal.inria.fr/hal-00911856.

[2] E. Agullo, A. Buttari, A. Guermouche and F. Lopez. 'Implementing multifrontal sparse solvers for multicore architectures with Sequential Task Flow runtime systems'. In: *ACM Transactions on Mathematical Software* (July 2016). DOI: 10.1145/2898348. URL: https://hal.inria.fr/hal-01333645.

[3] E. Agullo, S. Cools, E. Fatih-Yetkin, L. Giraud, N. Schenkels and W. Vanroose. 'On soft errors in the conjugate gradient method: sensitivity and robust numerical detection'. In: *SIAM Journal on Scientific Computing* 42.6 (Nov. 2020). DOI: 10.1137/18M122858X. URL: https://hal.inria.fr/hal-03022845.

[4] E. Agullo, L. Giraud and Y.-F. Jing. 'Block GMRES method with inexact breakdowns and deflated restarting'. In: *SIAM Journal on Matrix Analysis and Applications* 35.4 (Nov. 2014), pp. 1625–1651. DOI: 10.1137/140961912. URL: https://hal.inria.fr/hal-01067159.

[5] E. Agullo, L. Giraud and L. Poirel. 'Robust preconditioners via generalized eigenproblems for hybrid sparse linear solvers'. In: *SIAM Journal on Matrix Analysis and Applications* 40.2 (2019), pp. 417–439. DOI: 10.1137/17M1153765. URL: https://hal.inria.fr/hal-02074474.

[6] O. Beaumont, L.-C. Canon, L. Eyraud-Dubois, G. Lucarelli, L. Marchal, C. Mommessin, B. Simon and D. Trystram. 'Scheduling on Two Types of Resources: a Survey'. In: *ACM Computing Surveys* 53.3 (May 2020). DOI: 10.1145/3387110. URL: https://hal.inria.fr/hal-02432381.

[7] O. Beaumont, L. Eyraud-Dubois and S. Kumar. 'Fast Approximation Algorithms for Task-Based Runtime Systems'. In: *Concurrency and Computation: Practice and Experience* 30.17 (Sept. 2018). DOI: 10.1002/cpe.4502. URL: https://hal.inria.fr/hal-01878606.

[8] O. Beaumont, L. Eyraud-Dubois and A. Shilova. 'Optimal GPU-CPU Offloading Strategies for Deep Neural Network Training'. In: *Euro-Par 2020 - 26th International Conference on Parallel and Distributed Computing*. Ed. by M. Malawski and K. Rzadca. Euro-Par 2020: Parallel Processing 12247. Warsaw / Virtual, Poland: Springer, Aug. 2020, pp. 151–166. DOI: 10.1007/978-3-030-57675-2_10. URL: https://hal.inria.fr/hal-02316266.

[9] O. Beaumont, T. Lambert, L. Marchal and B. Thomas. 'Performance Analysis and Optimality Results for Data-Locality Aware Tasks Scheduling with Replicated Inputs'. In: *Future Generation Computer Systems* 111 (Oct. 2020), pp. 582–598. DOI: 10.1016/j.future.2019.08.024. URL: https://hal.inria.fr/hal-02275473.

[10] A. Hugo, A. Guermouche, P.-A. Wacrenier and R. Namyst. 'Composing multiple StarPU applications over heterogeneous machines: A supervised approach'. In: *International Journal of High Performance Computing Applications* 28 (Feb. 2014), pp. 285–300. DOI: 10.1177/1094342014527575. URL: https://hal.inria.fr/hal-01101045.

[11] M. Odunlami, V. Le Bris, D. Bégué, I. Baraille and O. Coulaud. 'A-VCI: A flexible method to efficiently compute vibrational spectra'. In: *Journal of Chemical Physics* 146.21 (June 2017). DOI: 10.1063/1.4984266. URL: https://hal.inria.fr/hal-01534134.

[12] G. Pichon, E. Darve, M. Faverge, P. Ramet and J. Roman. 'Sparse supernodal solver using block low-rank compression: Design, performance and analysis'. In: *International Journal of Computational Science and Engineering* 27 (July 2018), pp. 255–270. DOI: 10.1016/J.JOCS.2018.06.007. URL: https://hal.inria.fr/hal-01824275.

[13] G. Pichon, M. Faverge, P. Ramet and J. Roman. 'Reordering Strategy for Blocking Optimization in Sparse Linear Solvers'. In: *SIAM Journal on Matrix Analysis and Applications.* SIAM Journal on Matrix Analysis and Applications 38.1 (2017), pp. 226–248. DOI: 10.1137/16M1062454. URL: https://hal.inria.fr/hal-01485507.

[14] M. Predari, A. Esnard and J. Roman. 'Comparison of initial partitioning methods for multilevel direct k-way graph partitioning with fixed vertices'. In: *Parallel Computing* (2017). DOI: 10.1016/j.parco.2017.05.002. URL: https://hal.inria.fr/hal-01538600.

11.2 Publications of the year

International journals

[15] E. Agullo, M. Altenbernd, H. Anzt, L. Bautista-Gomez, T. Benacchio, L. Bonaventura, H.-J. Bungartz, S. Chatterjee, F. M. Ciorba, N. Debardeleben, D. Drzisga, S. Eibl, C. Engelmann, W. N. Gansterer, L. Giraud, D. Göddeke, M. Heisig, F. Jézéquel, N. Kohl, S. Xiaoye, R. Lion, M. Mehl, P. Mycek, M. Obersteiner, E. S. Quintana-Ortí, F. Rizzi, U. Rüde, M. Schulz, F. Fung, R. Speck, L. Stals, K. Teranishi, S. Thibault, D. Thönnes, A. Wagner and B. Wohlmuth. 'Resiliency in numerical algorithm design for extreme scale simulations'. In: *International Journal of High Performance Computing Applications* (20th Sept. 2021). URL: https://hal.inria.fr/hal-03348787.

[16] T. Benacchio, L. Bonaventura, M. Altenbernd, C. D. Cantwell, P. D. Düben, M. Gillard, L. Giraud, D. Göddeke, E. Raffin, K. Teranishi and N. Wedi. 'Resilience and fault tolerance in high-performance computing for numerical weather and climate prediction'. In: *International Journal of High Performance Computing Applications* 35.4 (8th Feb. 2021), pp. 285–311. DOI: 10.1177/1094342021990433. URL: https://hal.inria.fr/hal-03138061.

[17] A. Bernardi, M. Iannacito and D. Rocchini. 'High Order Singular Value Decomposition for Plant Biodiversity Estimation'. In: *Bollettino dell'Unione Matematica Italiana* (10th June 2021). DOI: 10.1007/s40574-021-00300-w. URL: https://hal.inria.fr/hal-02385304.

[18] L. Giraud, Y.-F. Jing and Y.-F. Xiang. 'A block minimum residual norm subspace solver with partial convergence management for sequences of linear systems'. In: *SIAM Journal on Matrix Analysis and Applications* (2022). URL: https://hal.inria.fr/hal-03546496.

[19] M. Hoelzl, G. Huijsmans, S. Pamela, M. Bécoulet, E. Nardon, F. J. Artola, B. Nkonga, C. V. Atanasiu, V. Bandaru, A. Bhole et al. 'The JOREK non-linear extended MHD code and applications to large-scale instabilities and their control in magnetically confined fusion plasmas'. In: *Nuclear Fusion* 61.6 (20th May 2021), p. 065001. DOI: 10.1088/1741-4326/abf99f. URL: https://hal.inria.fr/hal-03352509.

[20] D. Rocchini, E. Thouverai, M. Marcantonio, M. Iannacito, D. Da Re, M. Torresani, G. Bacaro, M. Bazzichetto, A. Bernardi, G. M. Foody, R. Furrer, D. Kleijn, S. Larsen, J. R. M. H. Lenoir, M. Malavasi, E. Marchetto, F. Messori, A. Montaghi, V. Moudry, B. Naimi, C. Ricotta, M. Rossini, F. Santi, M. J. Santos, M. E. Schaepman, F. D. Schneider, L. Schuh, S. Silvestri, P. Simova, A. K. Skidmore, C. Tattoni, E. Tordoni, S. Vicario, P. Zannini and M. Wegmann. 'rasterdiv-An Information Theory tailored R package for measuring ecosystem heterogeneity from space: To the origin and back'. In: *Methods in Ecology and Evolution* 12.6 (2021), pp. 1093–1102. DOI: 10.1111/2041-210X.13583. URL: https://hal.archives-ouvertes.fr/hal-03244722.

International peer-reviewed conferences

[21] G. Agosta, D. Cattaneo, W. Fornaciari, A. Galimberti, G. Massari, F. Reghenzani, F. Terraneo, D. Zoni, C. Brandolese, M. Celino et al. 'TEXTAROSSA: Towards EXtreme scale Technologies and Accelerators for euROhpc hw/Sw Supercomputing Applications for exascale'. In: DSD 2021 - 24th Euromicro Conference on Digital System Design. Palermo / Virtual, Italy, 1st Sept. 2021. URL: https://hal.inria.fr/hal-03329640.

[22] O. Beaumont, L. Eyraud-Dubois and A. Shilova. 'Efficient Combination of Rematerialization and Offloading for Training DNNs'. In: NeurIPS 2021 - Thirty-fifth Conference on Neural Information Processing Systems. Virtual-only Conference, France, 6th Dec. 2021. URL: https://hal.inria.fr/hal-03359793.

[23] O. Beaumont, L. Eyraud-Dubois and A. Shilova. 'Pipelined Model Parallelism: Complexity Results and Memory Considerations'. In: Europar 2021. Proceedings of Europar 2021. Lisbon, Portugal: Springer, 2021. URL: https://hal.inria.fr/hal-02968802.

[24] N. Grinsztajn, O. Beaumont, E. Jeannot and P. Preux. 'READYS: A Reinforcement Learning Based Strategy for Heterogeneous Dynamic Scheduling'. In: IEEE Cluster 2021. Portland / Virtual, United States, 7th Sept. 2021. URL: https://hal.inria.fr/hal-03313229.

[25] E. Korkmaz, M. Faverge, G. Pichon and P. Ramet. 'Deciding Non-Compressible Blocks in Sparse Direct Solvers using Incomplete Factorization'. In: HiPC 2021 - 28th IEEE International Conference on High Performance Computing, Data, and Analytics. Bangalore, India: IEEE, 17th Dec. 2021, pp. 1–10. URL: https://hal.inria.fr/hal-03361299.

Conferences without proceedings

[26] E. Agullo, M. Felšöci and G. Sylvand. 'Comparison of coupled solvers for FEM/BEM linear systems arising from discretization of aeroacoustic problems'. In: COMPAS 2021 - Conférence francophone d'informatique en Parallélisme, Architecture et Système. Lyon / Virtuel, France, June 2021. URL: https://hal.inria.fr/hal-03264472.

Reports & preprints

[27] E. Agullo, M. Felšöci and G. Sylvand. *A comparison of selected solvers for coupled FEM/BEM linear systems arising from discretization of aeroacoustic problems*. RR-9412. Inria Bordeaux Sud-Ouest, June 2021, p. 52. URL: https://hal.inria.fr/hal-03263603.

[28] E. Agullo, M. Felšöci and G. Sylvand. *A comparison of selected solvers for coupled FEM/BEM linear systems arising from discretization of aeroacoustic problems: literate and reproducible environment.* RT-0513. Inria Bordeaux Sud-Ouest, June 2021, p. 100. URL: https://hal.inria.fr/hal-03263620.

[29] E. Agullo, L. Giraud, V. Joncquieres, G. Marait, L. Poirel, O. Vermorel and W. Villafana. *A note on the strong parallel scalability of numerically scalable Poisson linear solvers*. RR-9423. Inria Bordeaux - Sud Ouest, 22nd Sept. 2021, p. 31. URL: https://hal.inria.fr/hal-03352049.

[30] O. Coulaud, A. A. Franc and M. Iannacito. *Extension of Correspondence Analysis to multiway datasets through High Order SVD: a geometric framework*. RR-9429. Inria Bordeaux - Sud-Ouest; Inrae, 7th Nov. 2021. URL: https://hal.inria.fr/hal-03418404.

[31] L. Giraud, Y.-F. Jing and Y. Xiang. *A block minimum residual norm subspace solver for sequences of multiple left and right-hand side linear systems*. RR-9393. Inria Bordeaux Sud-Ouest, Feb. 2021, p. 60. URL: https://hal.inria.fr/hal-03146213.

[32] E. Korkmaz, M. Faverge, G. Pichon and P. Ramet. *Deciding Non-Compressible Blocks in Sparse Direct Solvers using Incomplete Factorization*. RR-9396. Inria Bordeaux - Sud Ouest, 2021, p. 16. URL: https://hal.inria.fr/hal-03152932.

[33] N. Venkovic, P. Mycek, L. Giraud and O. Le Maitre. *Recycling Krylov subspace strategies for sequences of sampled stochastic elliptic equations*. RR-9425. Inria Bordeaux - Sud Ouest, 5th Oct. 2021. URL: https://hal.archives-ouvertes.fr/hal-03366966.