

RESEARCH CENTRE

**Inria Centre
at the University of Bordeaux**

IN PARTNERSHIP WITH:

Airbus Central Research & Technology,
Centre Européen de Recherche et de
Formation Avancée en Calcul Scientifique

2023

ACTIVITY REPORT

Project-Team
CONCACE

**Numerical and Parallel Composability for
High Performance Computing**

DOMAIN

**Networks, Systems and Services,
Distributed Computing**

THEME

**Distributed and High Performance
Computing**

Inria

Contents

Project-Team CONCACE	1
1 Team members, visitors, external collaborators	2
2 Overall objectives	3
3 Research program	5
3.1 Numerical algebra methodologies in model and data-driven scientific computing	5
3.1.1 Scientific computing in large size linear algebra	6
3.1.2 Scientific computing in large dimension multi-linear algebra	6
3.1.3 Scientific continuum between large size and large dimension	7
3.2 Composition of parallel numerical algorithms from a sequential expression	7
4 Application domains	8
4.1 Aeroacoustics Simulation	8
5 New software, platforms, open data	9
5.1 New software	9
5.1.1 compose	9
5.1.2 ScalFMM	9
5.1.3 CPPDiodon	10
5.1.4 FMR	10
6 New results	11
6.1 Towards a direct task-based solver for sparse/dense FEM/BEM linear systems	11
6.2 On the Arithmetic Intensity of Distributed-Memory Dense Matrix Multiplication Involving a Symmetric Input Matrix	11
6.3 Task-based parallel programming for scalable matrix product algorithms	12
6.4 Combining reduction with synchronization barrier on multi-core processors	12
6.5 A note on GMRES algorithm in Tensor Train format for the solution of parametric linear systems	12
6.6 On some orthogonalization schemes in Tensor Train format	12
6.7 Neural network preconditioned subspace methods for the solution of the Helmholtz equation	13
6.8 Machine learning techniques to predict the rank H-matrices	13
6.9 Multivariate extensions of the Multilevel Best Linear Unbiased Estimator for ensemble- variational data assimilation	13
6.10 A filtered multilevel Monte Carlo method for estimating the expectation of discretized random fields	14
6.11 Multilevel Surrogate-based Control Variates	14
6.12 Inferences in Hybrid Bayesian Networks using Quadrature Rules	15
7 Bilateral contracts and grants with industry	15
7.1 Bilateral Grants with Industry	15
8 Partnerships and cooperations	15
8.1 International initiatives	15
8.1.1 Participation in other International Programs	15
8.2 European initiatives	16
8.2.1 H2020 projects	16
8.2.2 Other european programs/initiatives	18
8.3 National initiatives	18
8.4 Regional initiatives	22

9 Dissemination	22
9.1 Promoting scientific activities	23
9.1.1 Scientific events: organisation	23
9.1.2 Scientific events: selection	23
9.1.3 Journal	23
9.1.4 Scientific expertise	23
9.1.5 Research administration	23
9.2 Teaching - Supervision - Juries	24
9.2.1 Teaching	24
9.2.2 Supervision	24
9.2.3 Juries	24
10 Scientific production	25
10.1 Major publications	25
10.2 Publications of the year	26
10.3 Cited publications	28

Project-Team CONCACE

Creation of the Project-Team: 2022 September 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
- A6.3. – Computation-data interaction
- A7.1. – Algorithms
- A8.2. – Optimization
- A8.10. – Computer arithmetic
- A9.2. – Machine learning
- A9.7. – AI algorithmics
- A9.10. – Hybrid approaches for AI

Other research topics and application domains

- B3.3.1. – Earth and subsoil
- B3.6. – Ecology
- B3.6.1. – Biodiversity
- B4.2.2. – Fusion
- B5.2.3. – Aviation
- B5.5. – Materials
- B9.5.1. – Computer science
- B9.5.2. – Mathematics
- B9.5.4. – Chemistry
- B9.5.6. – Data science

1 Team members, visitors, external collaborators

Research Scientists

- Luc Giraud [Team leader, INRIA, Senior Researcher, HDR]
- Carola Kruse [Team leader, CERFACS, Senior Researcher]
- Guillaume Sylvand [Team leader, AIRBUS, Senior Researcher]
- Emmanuel Agullo [INRIA, Researcher]
- Pierre Benjamin [AIRBUS, Senior Researcher]
- Olivier Coulaud [INRIA, Senior Researcher, HDR]
- Sofiane Haddad [AIRBUS, Senior Researcher]
- Paul Mycek [CERFACS, Senior Researcher]

Post-Doctoral Fellows

- Marvin Lasserre [INRIA, Post-Doctoral Fellow]
- Maksym Shpakovych [INRIA, until Sep 2023]
- Yanfei Xiang [INRIA, Post-Doctoral Fellow, from Dec 2023]

PhD Students

- Theo Briquet [INRIA, from Oct 2023]
- El Mehdi Ettaouchi [EDE, from Dec 2023]
- Marek Felsoci [INRIA, until Jan 2023]
- Antoine Gicquel [INRIA, from Nov 2023]
- Romain Peressoni [INRIA, until Apr 2023]
- Amine ZEKRI [LAPLACE, from Nov 2023, Toulouse]

Technical Staff

- Ludovic Coutès [SED, Inria, Engineer, part-time]
- Pierre Estérie [INRIA, Engineer, until Oct 2023]
- Gilles Marait [SED, Inria, Engineer, full-time]
- Florent Pruvost [SED, Inria, Engineer, part-time]

Interns and Apprentices

- Mahamat Younous Abdraman [INRIA, Intern, from May 2023 until Jul 2023]
- Aymane Abibi [INRIA, Intern, from Jun 2023 until Sep 2023]
- Abdessamad El Hajjami [INRIA, Intern, from Jun 2023 until Sep 2023]
- Antoine Gicquel [INRIA, Intern, from Apr 2023 until Sep 2023]
- Alexandre Malhene [INRIA, Intern, from Jul 2023 until Jul 2023]
- Ziad Zahi [INRIA, Intern, from Jun 2023 until Sep 2023]

Administrative Assistant

- Flavie Blondel [INRIA]

External Collaborators

- Hadrien Godé [Cerfacs, from Dec 2023]
- Jean-Rene Poirier [TOULOUSE INP, from May 2023, HDR]
- Ulrich Rüde [Friedrich-Alexander-Universität Erlangen & Cerfacs, HDR]

2 Overall objectives

Over the past 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 enabled researchers to find computationally efficient solutions to some of the most challenging scientific questions and problems in medicine and biology, climate science, nanotechnology, energy, and environment – to name a few – in the field of **model-driven computing**. Meanwhile the advent of network capabilities and IoT, next generation sequencing, ... tend to generate a huge amount of data that deserves to be processed to extract knowledge and possible forecasts. These calculations are often referred to as **data-driven calculations**. These two classes of challenges have a common ground in terms of numerical techniques that lies in the field of linear and multi-linear algebra. They do also share common bottlenecks related to the size of the mathematical objects that we have to represent and work on; those challenges retain a growing attention from the computational science community.

In this context, the purpose of the concace project, is **to contribute to the design of novel numerical tools for model-driven and data-driven calculations** arising from challenging academic and industrial applications. The solution of these challenging problems requires a multidisciplinary approach involving applied mathematics, computational and computer sciences. In applied mathematics, it essentially involves **advanced numerical schemes both in terms of numerical techniques and data representation of the mathematical objects** (e.g., compressed data, low-rank tensor [57, 64, 53] low-rank hierarchical matrices [55, 41]). In computational science, it involves large scale parallel heterogeneous computing and **the design of highly composable algorithms**. Through this approach, concace intends to contribute to all the steps that go from the design of new robust and accurate numerical schemes to the flexible implementations of the associated algorithms on large computers. To address these research challenges, researchers from Inria, Airbus Central R&T and Cerfacs have decided to combine their skills and research efforts to create the Inria concace project team, which will allow them to cover the entire spectrum, from fundamental methodological concerns to full validations on challenging industrial test cases. Such a joint project will enable a real synergy between basic and applied research with complementary benefits to all the partners. The main benefits for each partner are given below:

- Airbus Central R&T
 - Push our specific needs and use-cases towards the academic world to stimulate research in particular directions;
 - Remain at the level of the scientific state of the art, this collaboration allows us to facilitate the feedback by exposing directly our challenges and industrial applications to eventually facilitate the transfer of research in our design tools;
 - The Inria research model will naturally be extended to Airbus, allowing for the multiplication of ambitious, very upstream and long-term research, while at the same time directly applying to the needs expressed by Airbus;
 - Benefit from the very high-level international network of the Inria team (e.g., Univ. of Tennessee Knoxville, Barcelona supercomputing center, Julich supercomputing center, Lawrence Berkeley National Lab, Sandia National Lab, etc.).

- Cerfacs
 - Join forces, in terms of skills and expertise, with Inria and Airbus to make faster and more effective progress on the research areas addressed by the team;
 - Bring scientific challenges from industrial applications through our privileged relationship with our industrial partners;
 - Reciprocally, promote the developed methodologies and the obtained results towards our industrial partners;
 - Naturally interact with the national and european HPC ecosystems, as a member of the EuroHPC national competence center on HPC, to promote the research activities and tools of the team and to meet novel scientific challenges where our methodologies or tools apply.
- Inria
 - Reinforce the impact of our research through a direct contact and close interactions with real scientific and technical challenges;
 - Feed the virtuous feedback cycle between academic research and industrially-relevant applications enabling the emergence of new research avenues;
 - Create a privileged space for an open scientific dialogue enabling the fostering of existing synergies and to create new ones, in particular when one of the industrial partners is a large group whose spectrum of scientific problems is very broad.

In addition to the members of these entities, two other external collaborators will be strongly associated: Jean-René Poirier, from Laplace Laboratory at University of Toulouse) and Oguz Kaya, from LISN (Laboratoire Interdisciplinaire des Sciences du Numérique) at University of Saclay.

The scientific objectives described in Section 4 contain two main topics which cover numerical and computational methodologies. Each of the topic is composed of a methodological component and its validation counterpart to fully assess the relevance, robustness and effectiveness of the proposed solutions. First, we address numerical linear and multilinear algebra methodologies for model- and data-driven scientific computing. Second, because there is no universal single solution but rather a large panel of alternatives combining many of the various building boxes, we also consider research activities in the field of composition of parallel algorithms and data distributions to ease the investigation of this combinatorial problem toward the best algorithm for the targeted problem.

To illustrate on a single but representative example of **model-driven problems** that the joint team will address we can mention one encountered at Airbus that is related to large aero-acoustic calculations. The reduction of noise produced by aircraft during take-off and landing has a **direct societal and environmental impact on the populations (including citizen health)** located around airports. To comply with new noise regulation rules, novel developments must be undertaken to preserve the competitiveness of the European aerospace industry. In order to design and optimize new absorbing materials for acoustics and reduce the perceived sound, one must be able to simulate the propagation of an acoustic wave in an aerodynamic flow: The physical phenomenon at stake is **aero-acoustics**. The complex and chaotic nature of fluid mechanics requires simplifications in the models used. Today, we consider the flow as non-uniform only in a small part of the space (in the jet flow of the reactors mainly) which will be meshed in volume finite elements, and everywhere else the flow will be considered as uniform, and the acoustic propagation will be treated with surface finite elements. This brings us back to the solution of a linear system with dense and sparse parts, an atypical form for which there is no "classical" solver available. We therefore have to work on the coupling of methods (direct or iterative, dense or sparse, compressed or not, etc.), and to **compose different algorithms** in order to be able to handle very large industrial cases. While there are effective techniques to solve each part independently from one another, there is no canonical, efficient solution for the coupled problem, which has been much less studied by the community. Among the possible improvements to tackle such a problem, **hybridizing simulation and learning** represents an alternative which allows one to reduce the complexity by avoiding as much as possible local refinements and therefore reduce the size of the problem.

Regarding **data-driven calculation**, climate data analysis is one of the application domains that generate huge amounts of data, either in the form of measurements or computation results. The ongoing effort

between the climate modeling and weather forecasting community to mutualize digital environment, including codes and models, leads the climate community to use finer models and discretization generating an ever growing amount of data. The analysis of these data, mainly based on classical numerical tools with a strong involvement of linear algebra ingredients, is facing new scalability challenges due to this growing amount of data. Computed and measured data have **intrinsic structures** that could be naturally exploited by **low rank tensor representations** to best reveal the hidden structure of the data while addressing the scalability problem. The close link with the **CECI team at Cerfacs** will provide us with the opportunity to study novel numerical methodologies based on tensor calculation. Contributing to a better understanding of the mechanisms governing the climate change would obviously have significant **societal and economical impacts on the population**. This is just an illustration of a possible usage of our work, we could also have possibly mentioned an on-going collaboration where our tools will be used in the context of a steel company to reduce the data volume generated by IoT to be transferred on the cloud for the analysis. The methodological part described in Section 4 covers mostly two complementary topics: the first in the field of numerical scientific computing and the second in the core of computational sciences.

To sum-up, for each of the methodological contributions, we aim to find at least one dimensioning application, preferably from a societal challenge, which will allow us to validate these methods and their implementations at full-scale. The search for these applications will initially be carried out among those available at Airbus or Cerfacs, but the option of seeking them through collaborations outside the project will remain open. The ambition remains to develop generic tools whose implementations will be made accessible via their deposit in the public domain.

3 Research program

The methodological component of our proposal concerns the expertise for the design as well as the efficient and scalable implementation of highly parallel numerical algorithms. We intend to go from numerical methodology studies to design novel numerical schemes up to the full assessment at scale in real case academic and industrial applications thanks to advanced HPC implementations.

Our view of the research activity to be developed in **CONCACE** is to systematically assess the methodological and theoretical developments in real scale calculations mostly through applications under investigations by the industrial partners (namely Airbus Central R&T and Cerfacs).

We first consider in Section 4.1 topics concerning parallel linear and multi-linear algebra techniques that currently appear as promising approaches to tackle huge problems both in size and in dimension on large numbers of cores. We highlight the linear problems (linear systems or eigenproblems) because they are in many large scale applications the main bottleneck and the most computational intensive numerical kernels. The second research axis, presented in Section 4.2, is related to the challenge faced when advanced parallel numerical toolboxes need to be composed to easily find the best suited solution both from a numerical but also parallel performance point of view.

In short the research activity will rely on two scientific pillars, the first dedicated to the development of new mathematical methods for linear and multilinear algebra (both for model-driven and data-driven calculations). The second pillar will be on parallel computational methods enabling to easily compose in a parallel framework the packages associated with the methods developed as outcome of the first pillar. The mathematical methods from the first pillar can mathematically be composed, the challenge will be to do on large parallel computers thank to the outcome of the second pillar. We will still validate on real applications and at scale (problem and platform) in close collaborations with application experts.

3.1 Numerical algebra methodologies in model and data-driven scientific computing

At the core of many simulations, one has to solve a linear algebra problem that is defined in a vector space and that involves linear operators, vectors and scalars, the unknowns being usually vectors or scalars, *e.g.* for the solution of a linear system or an eigenvalue problem. For many years, in particular in model-driven simulations, the problems have been reformulated in classical matrix formalism possibly unfolding the spaces where the vectors naturally live (typically 3D PDEs) to end up with classical vectors

in R^n or C^n . For some problems, defined in higher dimension (*e.g.*, time dependent 3D PDE), the other dimensions are dealt in a problem specific fashion as unfolding those dimensions would lead to too large matrices/vectors. The concace research program on numerical methodology intends to address the study of novel numerical algorithms to continue addressing the mainstream approaches relying on **classical matrix formalism** but also to investigate alternatives where the structure of the underlying problem is kept preserved and all dimensions are dealt with equally. This latter research activity mostly concerns **linear algebra in tensor spaces**. In terms of algorithmic principles, we will lay an emphasis on hierarchy as a unifying principle for the numerical algorithms, the data representation and processing (including the current hierarchy of arithmetic) and the parallel implementation towards scalability.

3.1.1 Scientific computing in large size linear algebra

As an extension of our past and on-going research activities, we will continue our works on numerical linear algebra for model-driven applications that rely on classical vectorial spaces defined on R^n and C^n , where vectors and matrices are classical sparse or dense objects encountered in regular numerical linear algebra computations.

The main numerical algorithms we are interested in are:

- **Matrix decompositions** including classical ones such as the QR factorization that plays a central role in block Krylov solvers [37, 52], randomized range finder algorithms [39, 40], to name a few, as building orthonormal basis of subspaces guarantees numerical robustness. But also other factorizations, not used in classical linear algebra for model-driven calculation, such as non-negative factorization encountered in data-science for multi-variable analysis [51, 45].
- **Iterative solvers** both for linear system solutions and for eigenproblems. Regarding linear systems, we will pay a particular attention to advanced numerical techniques such as multi-level preconditioning, hybrid direct-iterative (both algebraic and PDE driven interface boundary conditions) and the solution of augmented systems (*e.g.*, Karush-Kuhn-Tucker or KKT) [58, 59]. We will investigate variants of nested subspace methods, possibly with subspace augmentation or deflation. In the multiple right-hand sides or left-hand sides cases, we will further study the possible orthogonalization variants and the trade-off between the associated parallel scalability and robustness. A particular attention will be paid to the communication hiding approaches and the investigation of their block extensions. For eigenproblem solutions, we will consider novel nested subspace techniques to further extend the numerical capabilities of the recently proposed AVCI [63, 60] technique as well as countour based integral equations (that intensively use linear systems techniques mentioned above).

In that context, we will consider the benefit of using hybridization between simulation and learning in order to reduce the complexity of classical approaches by diminishing the problem size or improving preconditioning techniques. In a longer term perspective, we will also conduct an active technological watch activity with respect to quantum computing to better understand how such a advanced computing technology can be synergized with classical scientific computing.

3.1.2 Scientific computing in large dimension multi-linear algebra

This work will mostly address linear algebra problems defined in large dimensional spaces as they might appear either in model-driven simulations or data-driven calculations. In particular we will be interested in tensor vectorial spaces where the intrinsic mathematical structures of the objects have to be exploited to design efficient and effective numerical techniques.

The main numerical algorithms we are interested in are:

- Low-rank tensor decompositions for model- and data-driven, some of them rely on some numerical techniques considered in the previous section [47, 50];
- Extension of iterative numerical linear solvers (linear systems and eigensolvers) to tensor vectorial spaces to handle problems that were previously vectorized to be amenable to solution by classical linear algebra techniques;

- Study preconditioning and domain decomposition techniques suited for the solution of stochastic PDEs (encountered in some Uncertainty Quantification context) [68] leading to large dimension or preconditioning based on a low-rank approximation of the tensorization of the dense matrix in Boundary Element Method solver [35, 38, 65].

3.1.3 Scientific continuum between large size and large dimension

Novel techniques for large size and large dimension problems tend to reduce the memory footprint and CPU consumption through data compression such as low-rank approximations (hierarchical matrices for dense and sparse calculation, tensor decomposition [49, 66, 61]) or speed up the algorithm (fast multipole method, randomized algorithm [56, 62] [67, 40] to reduce the time and energy to solution. Because of the compression, the genuine data are represented with lower accuracy possibly in a hierarchical manner. Understanding the impact of this lower precision data representation through the entire algorithm is an important issue for developing robust, “accurate” and efficient numerical schemes for current and emerging computing platforms from laptop commodity to supercomputers. Mastering the **trade-off between performance and accuracy** will be part of our research agenda [43, 46].

Because the low precision data representation can have diverse origins, this research activity will naturally cover the multi-precision arithmetic calculation in which the data perturbation comes entirely from the data encoding, representation and calculation in IEEE (or more exotic Nvidia GPU or Google TPU) floating point numbers. This will result in variable accuracy calculations. This general framework will also enable us to address soft error detection [34] and study possible mitigation schemes to design resilient algorithms.

3.2 Composition of parallel numerical algorithms from a sequential expression

A major breakthrough for exploiting multicore machine [42] is based on a data format and computational technique originally used in an out-of-core context [54]. This is itself a refinement of a broader class of numerical algorithms – namely, “updating techniques” – that were not originally developed with specific hardware considerations in mind. This historical anecdote perfectly illustrates the need to separate data representation, algorithmic and architectural concerns when developing numerical methodologies. In the recent past, we have contributed to the study of the sequential task flow (STF) programming paradigm, that enabled us to abstract the complexity of the underlying computer architecture [33, 32, 31]. In the concace project, we intend to go further by abstracting the numerical algorithms and their dedicated data structures. We strongly believe that combining these two abstractions will allow us to **easily compose toolbox algorithms and data representations in order to study combinatorial alternatives towards numerical and parallel computational efficiency**. We have demonstrated this potential on domain decomposition methods for solving sparse linear systems arising from the discretisation of PEDs, that has been implemented in the `maphys++` parallel package.

Regarding the abstraction of the target architecture in the design of numerical algorithms, the STF paradigm has been shown to significantly reduce the difficulty of programming these complex machines while ensuring high computational efficiency. However, some challenges remain. The first major difficulty is related to the scalability of the model at large scale where handling the full task graph associated with the STF model becomes a severe bottleneck. Another major difficulty is the inability (at a reasonable runtime cost) to efficiently handle fine-grained dynamic parallelism, such as numerical pivoting in the Gaussian elimination where the decision to be made depends on the outcome of the current calculation and cannot be known in advance or described in a task graph. These two challenges are the ones we intend to study first.

With respect to the second ingredient, namely the abstraction of the algorithms and data representation, we will also explore whether we can provide additional separation of concerns beyond that offered by a task-based design. As a seemingly simple example, we will investigate the possibility of abstracting the matrix-vector product, basic kernel at the core of many numerical linear algebra methods, to cover the case of the fast multipole method (FMM, at the core of the ScalFMM library). FMM is mathematically a block matrix-vector product where some of the operations involving the extra-diagonal blocks with hierarchical structure would be compressed analytically. Such a methodological step forward will consequently allow the factorisation of a significant part of codes (so far completely independent

because no bridge has been made upstream) including in particular the ones dealing with \mathcal{H} -matrices. The easy composition of these different algorithms will make it possible to explore the combinatorial nature of the possible options in order to best adapt them to the size of the problem to be treated and the characteristics of the target computer. *Offering such a continuum of numerical methods rather than a discrete set of tools is part of the team's objectives* It is a very demanding effort in terms of HPC software engineering expertise to coordinate the overall technical effort.

We intend to strengthen our engagement in reproducible and open science. Consequently, we will continue our joint effort to ensure consistent deployment of our parallel software; this will contribute to improve its impact on academic and industrial users. The software engineering challenge is related to the increasing number of software dependencies induced by the desired capability of combining the functionality of different numerical building boxes, e.g., a domain decomposition solver (such as `maphys++`) that requires advanced iterative schemes (such as those provided by `fabulous`) as well as state-of-the-art direct methods (such as `pastix`, `mumps`, or `qr_mumps`), deploying the resulting software stack can become tedious [36].

In that context, we will consider the benefit of using hybridization between simulation and learning in order to reduce the complexity of classical approaches by diminishing the problem size or improving preconditioning techniques. In a longer term perspective, we will also conduct an active technological watch activity with respect to quantum computing to better understand how such a advanced computing technology can be synergized with classical scientific computing.

4 Application domains

We have a major application domain in acoustic simulations that is provided by Airbus CR & T and a few more through collaborations in the context of ongoing projects, that include: plasma simulation (ESA contract and ANR Maturation), Electric device design (ANR TensorVim) and nanoscale simulation platform (ANR Diwina).

4.1 Aeroacoustics Simulation

Participants: Emmanuel Agullo, Carola Kruse, Paul Mycek, Pierre Benjamin, Marek Felsoci, Luc Giraud, Gilles Marait, Guillaume Sylvand.

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 [44]. 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.

5 New software, platforms, open data

Most of the software packages we develop are deployed using Guix-HPC [22].

5.1 New software

5.1.1 compose

Name: Numerical and parallel composability for high performance computing

Keywords: Numerical algorithm, Parallel computing, Linear algebra, Task-based algorithm, Dense matrix, Sparse matrix, Hierarchical matrix, FMM, C++

Functional Description: Composable numerical and parallel linear algebra library

URL: <https://gitlab.inria.fr/compose/compose>

Contact: Emmanuel Agullo

5.1.2 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.

The version 3.0 of the library is a partial rewriting of the version 2.0 in modern C++ (C++17) to increase the genericity of the approach. This version is also the basic framework for studying numerical and parallel composability within Concace.

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

Release Contributions: ScalFmm is a high performance library for solving n-body problems in astrophysics and electrostatics. It is based on the fast multipole method (FMM) and is highly parallel

News of the Year: Performance improvements in version 3.0. For the moment, this version only considers the interpolation approach. New features - the target particles can be different from the source particles - possibility to consider a non-mutual approach in the direct field - the low rank approximation of the transfer operator is taken into account.

URL: <https://gitlab.inria.fr/solverstack/ScalFMM>

Publications: [hal-00911856](#), [hal-01359458](#), [hal-00746089](#), [hal-01517153](#), [hal-01474556](#), [hal-01517153](#)

Contact: Olivier Coulaud

Participants: Olivier Coulaud, Pierre Est erie

5.1.3 CPPDiodon

Name: Parallel C++ library for Multivariate Data Analysis of large datasets.

Keywords: SVD, PCA

Scientific Description: Diodon provides executables and functions to compute multivariate data Analysis such as: Singular Value Decomposition (SVD), Principal Component Analysis (PCA) and variants (with different pre-treatments), Multidimensional Scaling (MDS), Correspondence Analysis (CoA), Canonical Correlation Analysis (CCA, future work), Multiple Correspondence Analysis (MCoA, future work). All these methods rely on a Singular Value Decomposition (SVD) of a 2D matrix. For small size matrices the SVD can be directly computed using a sequential or multi-threaded LAPACK solver such as OpenBlas or Intel MKL. For large matrices the SVD becomes time consuming and we use a Randomized Singular Value Decomposition method (rSVD) instead of the exact SVD which implementation is given by the FMR library. FMR can perform computations of the rSVD on parallel shared and distributed memory machines using adequate parallel dense linear algebra routines internally such as OpenBlas or Intel MKL on a shared memory node and Chameleon for distributed memory nodes (MPI).

Functional Description: Dimension reduction by multivariate data analysis. Diodon is a list of functions and drivers that implement in C++ and Python (i) pre-processing, SVD and post-processing with a wide variety of methods, (ii) random projection methods for SVD execution which allows to circumvent the time limitation in the calculation of the SVD, and (iii) a C++ implementation of the SVD with random projection to an imposed range or precision, connected to the MDS, PCA, CoA.

Release Contributions: Initial release of cppdiodon : a parallel C++ library for Multivariate Data Analysis of large datasets. Contains methods to compute Singular Value Decomposition (SVD), Randomized SVD, Principal Component Analysis (PCA), Multidimensional Scaling (MDS) and Correspondence Analysis (CoA). Handles text and hdf5 files. Parallel (mpi, threads, cuda) randomized SVD and EVD (for symmetric matrices) provided by FMR. Use multithreaded Lapack or Chameleon (distributed systems + GPUs).

URL: <https://gitlab.inria.fr/diodon/cppdiodon>

Publication: hal-03773985

Authors: Olivier Coulaud, Florent Pruvost

Contact: Olivier Coulaud

Partner: INRAE

5.1.4 FMR

Name: Fast Methods for Randomized numerical linear algebra

Keyword: SVD

Scientific Description: Fast Dense Standard and Randomized Numerical Linear Algebra is a library that allows to compute singular values or eigenvalues of large dense matrices by random linear algebra techniques. It is based on the random projection method (Gaussian or fast Hadamard/Fourier) or row/column selection (Nystrom method and variants). The library is developed in C++ and proposes a shared memory parallelization and a distributed approach with Chameleon (<https://gitlab.inria.fr/solverstack/chameleon>).

Functional Description: Fast Dense Standard and Randomized Numerical Linear Algebra is a library that allows to compute singular values or eigenvalues of large dense matrices by random linear algebra techniques. It is based on the random projection method (Gaussian or fast Hadamard/Fourier) or row/column selection (Nystrom method and variants). The library is developed in C++ and proposes a shared memory parallelization and a distributed approach with Chameleon (<https://gitlab.inria.fr/solverstack/chameleon>).

URL: <https://gitlab.inria.fr/compose/legacystack/fmr>

Publications: [hal-01228519](#), [hal-01334747](#), [hal-03773985](#)

Contact: Olivier Coulaud

Participants: Olivier Coulaud, Florent Pruvost, Romain Peressoni

6 New results

Participants: All team members.

6.1 Towards a direct task-based solver for sparse/dense FEM/BEM linear systems

We are interested in the direct solution of very large linear systems composed of both sparse and dense parts. Coupled hollow/dense systems appear in various physical problems, such as the simulation of acoustic wave propagation around aircraft. To produce a physically realistic result, the number of unknowns in the system can be extremely high, making its handling a real challenge. Thanks to the building blocks provided by state-of-the-art hollow and dense solvers, we can compose a coupled hollow/dense solver. To reduce the computation time and memory consumption of direct methods, some solvers implement advanced features such as numerical compression, out-of-core computation and distributed memory parallelism. These functionalities can be easily applied within the individual building blocks, but this is not trivial at the articulation between the hollow solver bricks and the dense solver bricks. Their programming interface (API) has not been designed for this purpose. We have previously proposed solver coupling schemes that still allow the use of these well-optimized solvers with advanced functionalities. The idea is to apply the existing API to carefully selected sub-arrays of coupled systems so as to take full advantage of digital compression and out-of-core computation in both shared and distributed memory. Although capable of handling considerably larger coupled systems compared to the state of the art, these schemes remain sub-optimal due to intrinsic design limitations. We therefore explore an alternative coupling scheme based on direct task-based solvers that use the same execution engine. The aim is to improve composability and facilitate data passing between hollow and dense solvers for more efficient computation. Before considering the integration of this approach into the complex code of a full community solver, we implemented a proof-of-concept without some advanced features. A preliminary experimental study enabled us to validate our prototype and demonstrate its competitiveness against other approaches.

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

6.2 On the Arithmetic Intensity of Distributed-Memory Dense Matrix Multiplication Involving a Symmetric Input Matrix

Dense matrix multiplication involving a symmetric input matrix (SYMM) is implemented in reference distributed-memory codes with the same data distribution as its general analogue (GEMM). We show that, when the symmetric matrix is dominant, such a 2D block-cyclic (2D BC) scheme leads to a lower arithmetic intensity (AI) of SYMM than that of GEMM by a factor of 2. We propose alternative data distributions preserving the memory benefit of SYMM of storing only half of the matrix while achieving up to the same AI as GEMM. We also show that, in the case we can afford the same memory footprint as GEMM, SYMM can achieve a higher AI. We propose a task-based design of SYMM independent of the data distribution. This design allows for scalable A-stationary SYMM with which all discussed data distributions, may they be very irregular, can be easily assessed. We have integrated the resulting code in a reduction dimension algorithm involving a randomized singular value decomposition dominated by SYMM. An experimental study shows a compelling impact on performance.

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

6.3 Task-based parallel programming for scalable matrix product algorithms

Task-based programming models have succeeded in gaining the interest of the high-performance mathematical software community because they relieve part of the burden of developing and implementing distributed-memory parallel algorithms in an efficient and portable way. In increasingly larger, more heterogeneous clusters of computers, these models appear as a way to maintain and enhance more complex algorithms. However, task-based programming models lack the flexibility and the features that are necessary to express in an elegant and compact way scalable algorithms that rely on advanced communication patterns. We show that the Sequential Task Flow paradigm can be extended to write compact yet efficient and scalable routines for linear algebra computations. Although, this work focuses on dense General Matrix Multiplication, the proposed features enable the implementation of more complex algorithms. We describe the implementation of these features and of the resulting GEMM operation. Finally, we present an experimental analysis on two homogeneous supercomputers showing that our approach is competitive up to 32,768 CPU cores with state-of-the-art libraries and may outperform them for some problem dimensions. Although our code can use GPUs straightforwardly, we do not deal with this case because it implies other issues which are out of the scope of this work.

For more details on this work we refer to [13].

6.4 Combining reduction with synchronization barrier on multi-core processors

With the rise of multi-core processors with a large number of cores the need of shared memory reduction that perform efficiently on a large number of core is more pressing. Efficient shared memory reduction on these multi-core processors will help share memory programs being more efficient on these one. In this paper, we propose a reduction combined with barrier method that uses SIMD instructions to combine barriers signaling and reduction value read/write to minimize memory/cache traffic between cores thus, reducing barrier latency. We compare different barriers and reduction methods on three multi-core processors and show that proposed combining barrier/reduction method are 4 and 3.5 times faster than respectively GCC 11.1 and Intel 21.2 OpenMP 4.5 reduction.

For more details on this work we refer to [14].

6.5 A note on GMRES algorithm in Tensor Train format for the solution of parametric linear systems

We consider the solution of linear systems with tensor product structure using a GMRES algorithm. To cope with the computational complexity in large dimension both in terms of floating point operations and memory requirement, our algorithm is based on low-rank tensor representation, namely the Tensor Train format. In a backward error analysis framework, we show how the tensor approximation affects the accuracy of the computed solution. With the backward perspective, we investigate the situations where the $(d + 1)$ -dimensional problem to be solved results from the concatenation of a sequence of d -dimensional problems (like parametric linear operator or parametric right-hand side problems), we provide backward error bounds to relate the accuracy of the $(d + 1)$ -dimensional computed solution with the numerical quality of the sequence of d -dimensional solutions that can be extracted from it. This enables to prescribe convergence threshold when solving the $(d + 1)$ -dimensional problem that ensures the numerical quality of the d -dimensional solutions that will be extracted from the $(d + 1)$ -dimensional computed solution once the solver has converged. The above mentioned features are illustrated on a set of academic examples of varying dimensions and sizes.

6.6 On some orthogonalization schemes in Tensor Train format

In the framework of tensor spaces, we consider orthogonalization kernels to generate an orthogonal basis of a tensor subspace from a set of linearly independent tensors. In particular, we investigate numerically the loss of orthogonality of six orthogonalization methods, namely Classical and Modified Gram-Schmidt with (CGS2, MGS2) and without (CGS, MGS) re-orthogonalization, the Gram approach, and the Householder transformation. To tackle the curse of dimensionality, we represent tensor with low rank approximation using the Tensor Train (TT) formalism, and we introduce recompression steps in the

standard algorithm outline through the TT-rounding method at a prescribed accuracy. After describing the algorithm structure and properties, we illustrate numerically that the theoretical bounds for the loss of orthogonality in the classical matrix computation round-off analysis results are maintained, with the unit round-off replaced by the TT-rounding accuracy. The computational analysis for each orthogonalization kernel in terms of the memory requirement and the computational complexity measured as a function of the number of TT-rounding, which happens to be the computational most expensive operation, completes the study.

This work was presented in two international conferences [28, 29] from different scientific communities.

For more details on this work we refer to the revised version of the scientific report [48] to be published.

6.7 Neural network preconditioned subspace methods for the solution of the Helmholtz equation

In recent years, scientific machine learning, utilizing deep learning methodologies, has found widespread application in the fields of scientific computing and computational engineering. Nevertheless, while these data-driven deep learning solvers can be highly effective once appropriately trained, they often yield solutions of limited accuracy. Additionally, the computational expenses incurred during the training phase can be prohibitively high. In this talk, we first presents the details of training various learning solvers, incorporating with different neural network architectures, for solving the heterogeneous Helmholtz equation. Some mathematical ingredients from classical iterative solver are considered into the training phase to enhance robustness and speed. Moreover, once the neural network solvers are adequately trained, their inferences can be applied as a nonlinear preconditioner in the classical subspace methods, like the flexible GMRES or flexible FOM method. This presentation demonstrates the efficiency of employing neural networks as preconditioner and showcases the evident advantages of these neural network preconditioned approaches. They outperform both the newly emerging deep neural network methods and the classical subspace methods in both computational efficiency and solution accuracy.

For more details on this work we refer to the revised version of the scientific report [27, 30]

6.8 Machine learning techniques to predict the rank H-matrices

The discretization of spatial operators using boundary element techniques leads to dense linear systems. The representation of full matrices in \mathcal{H} -matrix (Hierarchical Matrices) format is based on a bisection of space, leading to a binary tree defined on the operator definition space. A block of the matrix representing the interaction between two subsets of unknowns can be interpreted as the interaction between two nodes of the binary tree. An admissibility condition is used to determine whether this matrix block admits a low-rank representation or whether this approximation should be considered at the level of the sons of these nodes. These admissibility conditions have been studied theoretically in certain configurations. The goal of this work is to propose an admissibility condition learned by a neural network from computationally inexpensive information. The training data will be extracted from a set of simulations performed by Airbus CR&T.

6.9 Multivariate extensions of the Multilevel Best Linear Unbiased Estimator for ensemble-variational data assimilation

Multilevel estimators aim at reducing the variance of Monte Carlo statistical estimators, by combining samples generated with simulators of different costs and accuracies. In particular, the recent work of Schaden and Ullmann (2020) on the multilevel best linear unbiased estimator (MLBLUE) introduces a framework unifying several multilevel and multifidelity techniques. The MLBLUE is reintroduced here using a variance minimization approach rather than the regression approach of Schaden and Ullmann. We then discuss possible extensions of the scalar MLBLUE to a multidimensional setting, i.e. from the expectation of scalar random variables to the expectation of random vectors. Several estimators of increasing complexity are proposed: a) multilevel estimators with scalar weights, b) with element-wise weights, c) with spectral weights and d) with general matrix weights. The computational cost of each method is discussed. We finally extend the MLBLUE to the estimation of second-order moments in

the multidimensional case, i.e. to the estimation of covariance matrices. The multilevel estimators proposed are d) a multilevel estimator with scalar weights and e) with element-wise weights. In large-dimension applications such as data assimilation for geosciences, the latter estimator is computationally unaffordable. As a remedy, we also propose f) a multilevel covariance matrix estimator with optimal multilevel localization, inspired by the optimal localization theory of Ménétrier and Auligné (2015).

For more details on this work we refer to the revised version of the scientific report [24]

6.10 A filtered multilevel Monte Carlo method for estimating the expectation of discretized random fields

We investigate the use of multilevel Monte Carlo (MLMC) methods for estimating the expectation of discretized random fields. Specifically, we consider a setting in which the input and output vectors of the numerical simulators have inconsistent dimensions across the multilevel hierarchy. This requires the introduction of grid transfer operators borrowed from multigrid methods. Starting from a simple 1D illustration, we demonstrate numerically that the resulting MLMC estimator deteriorates the estimation of high-frequency components of the discretized expectation field compared to a Monte Carlo (MC) estimator. By adapting mathematical tools initially developed for multigrid methods, we perform a theoretical spectral analysis of the MLMC estimator of the expectation of discretized random fields, in the specific case of linear, symmetric and circulant simulators. This analysis provides a spectral decomposition of the variance into contributions associated with each scale component of the discretized field. We then propose improved MLMC estimators using a filtering mechanism similar to the smoothing process of multigrid methods. The filtering operators improve the estimation of both the small- and large-scale components of the variance, resulting in a reduction of the total variance of the estimator. These improvements are quantified for the specific class of simulators considered in our spectral analysis. The resulting filtered MLMC (F-MLMC) estimator is applied to the problem of estimating the discretized variance field of a diffusion-based covariance operator, which amounts to estimating the expectation of a discretized random field. The numerical experiments support the conclusions of the theoretical analysis even with non-linear simulators, and demonstrate the improvements brought by the proposed F-MLMC estimator compared to both a crude MC and an unfiltered MLMC estimator.

For more details on this work we refer to the revised version of the scientific report [23]

6.11 Multilevel Surrogate-based Control Variates

Monte Carlo (MC) sampling is a popular method for estimating the statistics (e.g. expectation and variance) of a random variable. Its slow convergence has led to the emergence of advanced techniques to reduce the variance of the MC estimator for the outputs of computationally expensive solvers. The control variates (CV) method corrects the MC estimator with a term derived from auxiliary random variables that are highly correlated with the original random variable. These auxiliary variables may come from surrogate models. Such a surrogate-based CV strategy is extended here to the multilevel Monte Carlo (MLMC) framework, which relies on a sequence of levels corresponding to numerical simulators with increasing accuracy and computational cost. MLMC combines output samples obtained across levels, into a telescopic sum of differences between MC estimators for successive fidelities. In this paper, we introduce three multilevel variance reduction strategies that rely on surrogate-based CV and MLMC. MLCV is presented as an extension of CV where the correction terms devised from surrogate models for simulators of different levels add up. MLMC-CV improves the MLMC estimator by using a CV based on a surrogate of the correction term at each level. Further variance reduction is achieved by using the surrogate-based CVs of all the levels in the MLMC-MLCV strategy. Alternative solutions that reduce the subset of surrogates used for the multilevel estimation are also introduced. The proposed methods are tested on a test case from the literature consisting of a spectral discretization of an uncertain 1D heat equation, where the statistic of interest is the expected value of the integrated temperature along the domain at a given time. The results are assessed in terms of the accuracy and computational cost of the multilevel estimators, depending on whether the construction of the surrogates, and the associated computational cost, precede the evaluation of the estimator. It was shown that when the lower fidelity outputs are strongly correlated with the high-fidelity outputs, a significant variance reduction is obtained

when using surrogate models for the coarser levels only. It was also shown that taking advantage of pre-existing surrogate models proves to be an even more efficient strategy.

For more details on this work we refer to the revised version of the scientific report [25]

6.12 Inferences in Hybrid Bayesian Networks using Quadrature Rules

Probabilistic inference in high-dimensional continuous (or hybrid) domains is a challenging problem typically addressed through discretization, sampling, or reliance on often naive parametric assumptions. The drawbacks of these methods are well-known: slow computational speeds and/or highly inaccurate results.

This paper introduces a novel deterministic and general inference algorithm designed for hybrid Bayesian networks featuring both discrete and continuous variables. The algorithm avoids the discretization of continuous densities into histograms by employing quadrature rules to compute continuous integrals, thus transforming the process of marginalizing continuous random variables into summations. These summations are subsequently computed using classical sum-product algorithms within an auxiliary discrete Bayesian network, appropriately constructed for this purpose.

Numerous experiments are conducted using either the conditional linear Gaussian model for reference, or non-Gaussian models for the sake of generality. The algorithm shows remarkable performances both in speed and accuracy when compared with discretization, kernel smoothing or Gaussian assumption. This establishes the algorithm's efficacy across a spectrum of scenarios, and proving its potential as a robust tool for hybrid Bayesian network inferences.

7 Bilateral contracts and grants with industry

Participants: All permanent members.

7.1 Bilateral Grants with Industry

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

- Airbus CR&T for the PhD thesis of Marek Felsoci.

In addition two post-docs, namely Maksym Shpakovych and Marvin Lasserre, are funded by the "plan de relance".

8 Partnerships and cooperations

8.1 International initiatives

8.1.1 Participation in other International Programs

Participants: Emmanuel Agullo, Olivier Coulaud, Luc Giraud, Gilles Marait.

PHC Bosphore There is a continuous deployment of sensor devices in industrial manufactures to monitor the production processes. Beyond the real time monitoring of the infrastructures, the huge amount of data collected can be further exploited to forecast for instance the aging or the failures of some production tools using machine learning techniques. Classically, this data analysis is performed off-line using a cloud-based service and transferring the data from the production place to the processing place on the cloud is a major bottleneck.

This project aims to design a highly efficient and robust parallel non-linear dimensionality reduction approach for a generic cloud- based Industrial Internet of Things (IIoT) data processing system to reduce the volume of data transferred without compromising the accuracy of the target machine learning tasks. An interdisciplinary collaboration that addresses all the relevant issues from numerical linear algebra, parallel processing, machine learning, and IIoT systems is required to achieve this goal.

The project's main objective is to develop a novel eigendecomposition approach based on contour integrals and recycling Krylov subspaces to improve the robustness and efficiency of non-linear dimensionality reduction techniques. The resulting non-linear dimensionality technique will then be used to reduce the size of time-dependent IIoT data to decrease the data transfer costs without compromising the overall machine learning accuracy on the cloud side. To achieve the main objective of the project, partners will accomplish the following specific objectives: (i) Design of an efficient and scalable eigensolver, (ii) Integration of the efficient eigensolver to the non-linear dimensionality reduction methods, and (iii) Application of the proposed method to realistic IIoT system data.

Inria International Lab: JLESC The work between ANL and Inria was initiated in the context of the JLESC initiative. Compression is ubiquitous in scientific computing in general and numerical linear algebra in particular. One of the most well-known methods for compression in this latter field is truncated singular value decomposition (TSVD). TSVD allows for compressing matrices in some optimum sense. However, there are fewer techniques for compressing vectors. An old but currently intensively studied method is to design numerical algorithms able to use mixed-precision arithmetic. Still, data are stored in a way that sticks to the hardware processing capacity, typically under the form of 64-, 32-, and 16-bit words. The idea of variable-accuracy storage is instead to rely on a compressor such as SZ develop at ALN to compress vectors independently from hardware constraints and apply it to the solution of large sparse linear systems.

8.2 European initiatives

8.2.1 H2020 projects

Participants: Emmanuel Agullo, Olivier Coulaud, Luc Giraud, Carola Kruse, Gilles Marait.

RISC-2

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

Type: Coordinated Support Action

Duration: 2021 - 2023

Coordinator: Barcelona Supercomputing Center (Spain)

Inria coordinator: Stéphane Lanteri

Concace contact: Luc Giraud

Partners: Forschungszentrum 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.

EoCoE-3

Title: Energy oriented Centre of Excellence for computer applications

Duration: 2024-2026

Coordinator: CEA

Inria coordinator: Bruno Raffin

CONCACE contact: Emmanuel Agullo

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)
- UNIVERSITA DEGLI STUDI DI ROMA TORVERGATA (Italy)
- UNIVERSITA DEGLI STUDI DI TRENTO (Italy)
- UNIVERSITE LIBRE DE BRUXELLES (Belgium)
- UNIVERSITE PARIS-SUD (France)

Inria contact: Bruno Raffin (Datamove)

Summary: The Concace team (Inria, Cerfacs) participates in the Energy-oriented Centre of Excellence (EoCoE-III), starting in January 2024. The project applies cutting-edge exascale computational methods in its mission to accelerate the transition to the production, storage and management of clean, decarbonized energy. EoCoE-III is anchored in the High Performance Computing (HPC) community and targets research institutes and key commercial players who develop and enable energy-relevant numerical models to be run on exascale supercomputers, demonstrating their benefits for the net-zero energy transition. The project will draw on the experience of two successful previous projects EoCoE-I and -II, where a large set of diverse computer applications from four such energy domains achieved significant efficiency gains thanks to a multidisciplinary expertise in applied mathematics and supercomputing. EoCoE-III channels its efforts into 5 exascale lighthouse applications in the low-carbon sectors of Energy Materials, Water, Wind and Fusion. This multidisciplinary effort will harness innovations in computer science and mathematical algorithms within a tightly integrated co-design approach to overcome performance bottlenecks and to anticipate HPC hardware developments. A world-class consortium of 16 complementary partners forms a unique network of expertise in energy science, scientific computing and HPC, including 3 leading European supercomputing centres.

8.2.2 Other european programs/initiatives

Title: High Performance Spacecraft Plasma Interaction Software

Duration: 2022 - 2024

Funding: ESA

Coordinator: Sébastien Hess (ONERA)

Concace contact: Olivier Couland and Luc Giraud

Partners: • Airbus DS

- Artenum
- ONERA

Summary: Controlling the plasma environment of satellites is a key issue for the nation in terms of satellite design and propulsion. Three-dimensional numerical modelling is thus a key element, particularly in the preparation of future space missions. The SPIS code is today the reference in Europe for the simulation of these phenomena. The methods used to describe the physics of these plasmas are based on the representation of the plasma by a system of particles moving in a mesh (here unstructured) under the effect of the electric field which satisfies the Poisson equation. ESA has recently shown an interest in applications requiring complex 3D calculations, which may involve several tens of millions of cells and several tens of billions of particles, and therefore in a highly parallel and scalable version of the SPIS code.

8.3 National initiatives

MAMBO

Duration: 2018 – 2022

CONCACE contact: Guillaume Sylvand

Funding: DGAC

Partners:

- CEA

- Inria
- CNRS

Summary:**PEPR Numpex****Duration:** 2018 – 2022**CONCACE contact:** Emmanuel Agullo, Luc Giraud**Funding:** ANR**Partners:**

- CEA
- Inria
- CNRS

Summary: NumPEX is a French program dedicated to Exascale: High-performance computing (HPC), high-performance data analytics (HPDA), and Artificial Intelligence (AI) pose significant challenges across scientific, societal, economic, and ethical realms. These technologies, including modeling and data analysis, are crucial decision support tools addressing societal issues and competitiveness in French research and development. Digital resources, essential across science and industry, demand high-performance hardware. HPC enables advanced modeling, while HPDA handles heterogeneous and massive data. The solution to exploding demand is the upcoming “exascale” computers, a new generation with extraordinary capabilities.

In this context, the French Exascale program NumPEX aims at designing and developing the software components that will equip future exascale machines. NumPEX will deliver Exascale-grade numerical methods, softwares, and training, allowing France to remain one of the leaders in the field. It will contribute to take bridging the gap between cutting-edge software development and application domains to prepare the major scientific and industrial application codes to fully exploit the capabilities of these machines. Application domains of the NumPEX program include, but are not limited to, weather forecasting and climate, aeronautics, automotive, astrophysics, high energy physics, material science, energy production and management, biology and health.

Numpex is organized in 7 scientific pillar projects, we are directly involved in two of them namely:

- Exa-MA : Methods and Algorithms for Exascale;
- Exa-SofT : HPC softwares and tools.

TensorVIM**Duration:** 2023 – 2026**Coordinator:** LAPLACE**CONCACE contact:** Olivier Coulaud**Funding:** ANR**Partners:**

- Inria
- LAPLACE
- G2ELaB

Summary: The aim of this project is to develop high-performance computational tools for the rapid implementation of low-frequency electromagnetic simulations for electrical applications. We consider an approach based on volume integral methods using low-rank approximations. Instead of using classical compression techniques such as the fast multipole method or the hierarchical matrix approach, we propose to investigate the use of low-rank tensors to accelerate the computation of the solution of the linear system. The tools developed will be used for the modeling of various devices (PCB modeling, Electrical Machines) with the main goal of improving their energy performance.

Maturation

Title: MAssively parallel sparse grid PIC algorithms for low TemperatURe plAsmas SimulaTIONs

Duration: 2023 – 2026

Coordinator: Laurent Garrigues (Laplace)

CONCACE contact: Luc Giraud

Funding: ANR

Partners:

- Laplace Lab
- IMT

Summary: The simulation under real conditions of partially magnetized low temperature plasmas by Lagrangian approaches, though using powerful Particle-In-Cell (PIC) techniques supplemented with efficient high-performance computing methods, requires considerable computing resources for large plasma densities. This is explained by two main limitations. First, stability conditions that constrain the numerical parameters to resolve the small space and time scales. These numerical parameters are the mesh size of the grid used to compute the electric field and the time step between two consecutive computations. Second, PIC methods rely on a sampling of the distribution function by numerical particles whose motion is time integrated in the self-consistent electric field. The PIC algorithm remains close to physics and offers an incomparable efficiency with regard to Eulerian methods, discretizing the distribution function onto a mesh. It is widely and successfully operated for the discretization of kinetic plasma models for more than 40 years. Nonetheless, to spare the computational resources, the number of numerical particles is limited compared to that of the physical particles. Inherent to this “coarse” sampling, PIC algorithms produce numerical approximations prone to statistical fluctuations that vanish slowly with the mean number of particles per cell. The mesh accessible on typical high performance computing machines may 10^9 cells, which brings the mesh size close to the scale of the physics, but the mean number of numerical particles in each cell shall be limited, to mitigate the memory footprint as well as the computational time. A breakthrough is therefore necessary to reduce the computational resources by orders of magnitude and make possible the use of explicit PIC method for large scale and/or densities for 3D computations.

This is the issue addressed within the MATURATION project aiming at introducing a new class of PIC algorithms with an unprecedented computational efficiency, by analyzing and improving, parallelizing and optimizing as well as benchmarking, in the demanding context of partially magnetized low temperature plasmas through 2D large scale and 3D computations, a method recently proposed in the literature, based on a combination of sparse grid techniques and PIC algorithm.

Diwina

Title: Magnetic Digital Twins for Spintronics : nanoscale simulation platform

Duration: 2023 – 2026

Coordinator: Institut Neel

CONCACE contact: Olivier Coulaud

Funding: ANR

Partners:

- CMAP, Institut Neel, Inria, SPINTEC

Summary: The DiWiNa project aims at developing a unified open-access platform for spintronic numerical twins, ie, codes for micromagnetic/spintronic simulations with sufficiently-high reliability and speed so that they can be trusted and used as reality. The simulations will be bridged to the advanced microscopy techniques used by the community, through plugins to convert the statics or time-resolved 3D vector- fields into contrast maps for the various techniques, including their experimental transfer functions. To achieve this, we bring together experts from different disciplines to address the various challenges: spintronics for the core simulations, mathematics for trust, algorithmics for speed, experimentalists for the bridge with microscopy. Practical work consists of checking the time-integration stability of spintronic torque involved in the dynamics when implemented in the versatile finite-element framework, improve the calculation speed through advanced libraries, build the bridge with microscopies through rendering tools, and encapsulate these three key ingredients into a user-friendly Python ecosystem. Through open-access and versatile user-friendly encapsulation, we expect that this platform is suited to serve the needs of the entire physics and engineering community of spintronics. The platform will be unique in its features, ranging from simulation to the direct and practical comparison with experiments. It will contribute to reduce considerably the number of experimental screening for the faster development of new spintronic devices, which are expected to play a key role in energy saving.

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

Duration: 2018 – 2023

Coordinator: Alfredo Buttari (IRIT)

CONCACE contact: Emmanuel Agullo

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.

8.4 Regional initiatives

Title: HPC-Ecosystem

Duration: 2018 – 2023

Coordinator: Emmanuel Agullo

CONCACE contact: Emmanuel Agullo

Partners:

- STORM, TADAAM, TOPAL from Inria Bordeaux Sud-Ouest
- Airbus Central R&T
- CEA Commissariat à l'énergie atomique et aux énergies alternatives

Summary: 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).

9 Dissemination

Participants: All permanent team members.

9.1 Promoting scientific activities

9.1.1 Scientific events: organisation

Member of the organizing committees

- Luc Giraud is member of the Gene Golub SIAM Summer School. The twelfth Gene Golub SIAM Summer School was entitled "Quantum Computing and Optimization", Lehigh University July 31 through August 11, 2023.
- Carola Kruse and Paul Mycek are members of the organising committee of the "Sparse Days 2023"

9.1.2 Scientific events: selection

Chair of conference program committees

- Compas: Emmanuel Agullo (parallelism chair of the steering committee)

Member of the conference program committees

- PDSEC: Olivier Coulaud, Luc Giraud,

Co-chair of conference proceedings

- ISC-HPC 2023: Carola Kruse

Reviewer

- ISC-HPC 2023: Carola Kruse for Birds of a feather submissions

9.1.3 Journal

Member of the editorial boards L. Giraud is member of the editorial board of the SIAM Journal on Scientific Computing (SISC).

Reviewer - reviewing activities Applied Mathematical Modelling, SIAM J. Scientific Computing, Mathematical Modelling and Numerical Analysis, ...

9.1.4 Scientific expertise

- Luc Giraud is
member of the board on Modelization, Simulation and data analysis of the Competitiveness Cluster for Aeronautics, Space and Embedded Systems.
member of the scientific council of the ONERA Lab LMA2S (Laboratoire de Mathématiques Appliquées à l'Aéronautique et au Spatial).
member of member of the scientific council of GDR Calcul.
- Guillaume Sylvand is expert in Numerical Simulation and HPC at Airbus. member of the scientific council of the ORAP.

9.1.5 Research administration

- Emmanuel Agullo is member of the CDT (Technological Development Commission) at inria Centre at the Bordeaux University.
- Luc Giraud is techniques pilot for the expert group for the evaluation of French research entities (UMRs and EAs) relatively to the protection of scientific and technological properties (PPST) on information and communication sciences and technologies (STIC).

9.2 Teaching - Supervision - Juries

9.2.1 Teaching

- 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 6 h at Bordeaux INP (ENSEIRB-MatMeca).
 - L. Giraud: Introduction to intensive computing and related programming tools 20h, INSA Toulouse; Advanced numerical linear algebra 10h, ENSEEIHT Toulouse.
 - C. Kruse: Advanced topics in numerical linear algebra, 10h, FAU Erlangen; Méthodes Itératives en Algèbre Linéaire, 14h, ENSEEIHT Toulouse.
 - P. Mycek: Multifidelity methods 25h, INSA Toulouse.

9.2.2 Supervision

- PhD completed: Nicolas Venkovic ; Preconditioning strategies for stochastic elliptic partial differential equations ; started Oct 2018 ; L. Giraud, P. Mycek, O. Le Maître (PLATON) ; defended Sep 11, 2023.
- PhD completed: Mohamed Anwar Abouabdallah ; Tensor-Train approach for inference in stochastic block models, application to biodiversity characterization ; started Oct 2019; O. Coulaud, A. Franc (PLEIADE), N. Peyrard (Inrae), defended on Feb. 2, 2023.
- PhD in progress: Théo Briquet; machine learning techniques for rank prediction of \mathcal{H} -matrices; started October 2023, L. Giraud, P. Mycek, G. Sylvand.
- PhD in progress: Mehdi El Ettaouchi; nonlinear domain decomposition techniques in geosciences; started March 2023, L. Giraud, C. Kruse, N. Tardieu (EDF).
- PhD completed: Marek Felsoci; Fast solvers for high-frequency aeroacoustics; started Oct. 2019; G. Sylvand, E. Agullo, defended on Feb. 22, 2023.
- PhD in progress: Antoine Gicquel; started Nov. 2023, O. Coulaud, B. Bramas; Acceleration of the matrix-vector product by the fast multipole method for heterogeneous machine clusters
- PhD completed: Romain Peressoni; Fast multidimensional scaling method for the study of biodiversity; started Oct 2019; E. Agullo, O. Coulaud, A. Franc (PLEIADE), defended on June 13, 2023.
- PhD completed: Aboul-Karim Mohamed El Maarouf; Parallel fine grain incomplete LU factorization for the solution of sparse linear systems; started: Dec. 2019; L. Giraud, A. Guermouche (Topal), defended on March 17, 2023.
- PhD in progress: Amine Zekri ; Low-rank Tensor Solver for magnetostatic problems for electric power applications, started October 2023; O. Coulaud, J.R. Poirier

9.2.3 Juries

PhD defense

- Nicolas Venkovic, "Preconditioning strategies for stochastic elliptic partial differential equations"; referees: Julien Langou, Anthony Nouy; members: Luc Giraud, Paul Mycek, Olivier Coulaud, Pietro Marco Congedo, Olivier Le Maître, Nicole Spillane; Université de Bordeaux, Spécialité mathématiques appliquées et calcul scientifique, Sep. 11, 2023.

- Mohamed Anwar Abouabdallah, "Tensor-Train approach for inference in stochastic block models, application to biodiversity characterisation"; referees: Sophie Donnet, Jean-René Poirier; members: Agnès Bouchez, Olivier Coulaud, Alain Franc, Nathalie Peyrard, Pierre-Henri Wuillemain; Université de Bordeaux, Spécialité informatique, Feb. 2, 2023.
- Karim Mohamed El Maarouf, "Incomplete factorization and solution of triangular systems for fine-grained parallelism computers"; referees: M. Damien Tromeur-Dervout, Pierre Jolivet; members: Jocelyne Erhel, Brice Goglin, Luc Giraud, David Goudin, Abdou Guermouche, Thomas Guignon; Université de Bordeaux, Spécialité informatique, March 17, 2023.
- Romain Peressoni, "Large Scale Multidimensional Scaling for the Study of Biodiversity"; referees: Emmanuel Paradis, Bruno Raffin; members: Emmanuel Agullo, Olivier Coulaud, Alain Franc, Sandrine Mouysset, Raymond Namyst, Gaël Varoquaux; Université de Bordeaux, Spécialité informatique, June 13, 2023.
- Marek Felsoci, "Fast solvers for high-frequency aeroacoustics"; referees: Jean-Yves L'Excellent (MUMPS Technologies), Ulrich Rüde (FAU); members: Emmanuel Agullo, Stéphanie Chaillat, Konrad Hinsin, David Goudin, Christian Pérez, Guillaume Sylvand; Université de Bordeaux, Spécialité informatique, Feb. 22, 2023.
- Mehdi Jadoui, "Robust Krylov solvers for solving partitioned and monolithic aero-structure coupled adjoint system"; referees: Luc Giraud, Michel Visonneau; members: Christophe Blondeau, Gilbert Rogé, Pierre Jolivet, François-Xavier Roux; Sorbonne Université, Spécialité Mathématiques appliquées, Nov. 16, 2023.
- Matthias Baray, "Tensorial approach for solving boundary integral equations in acoustics and electromagnetism"; referees: Christophe Geuzaine, Sébastien Tordeux; members: Nathalie Raveu, Jean-René Poirier, David Levadoux, Luc Giraud, Anthony Nouy, Gildas Kubické; INP Toulouse, Mathématiques appliquées, Jan. 17, 2023.
- Matthieu Gerest, "Using Block Low-Rank compression in mixed precision for sparse direct linear solvers"; referees: Iain Duff, Luc Giraud; members: Hélène Barucq, Olivier Boiteau, Fabienne Jézéquel; Théo Mary, Frédéric Nataf, Sorbonne Université, Spécialité informatique, Nov. 8, 2023.
- Clément Guillet, "Sparse approach to accelerate Particle-In-Cell method in 3D"; referees: Nicolas Crouseilles, Raphaël Loubère; members: Laurent Garrigues, Fabrice Deluzet, Charles Frédérique, Pierre Jolivet, Anne Bourdon, invited: Luc Giraud, Toulouse Université, Spécialité Mathématiques appliquées, June 28, 2023.
- Mohamed Amine Hamadi, "Krylov-based subspaces methods for large-scale dynamical systems and data-driven model reduction"; referees: Michela Redivo Zaglia, Luc Giraud, Giuseppe Rodriguez; members: Hassane Sadok, Khalid Jbilou, Ahmed Ratnani, Abdeslem Hafid Bentbib, Lahcen Laayouni, Université Mohammed VI Polytechnique et Université du Littoral, Spécialité Mathématiques appliquées, July 17, 2023.

10 Scientific production

10.1 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](https://doi.org/10.1137/130915662). URL: <https://hal.inria.fr/hal-00911856>.
- [2] E. Agullo, A. Buttari, A. Guermouche, J. Herrmann and A. Jego. 'Task-based parallel programming for scalable matrix product algorithms'. In: *ACM Transactions on Mathematical Software* (2023). URL: <https://hal.science/hal-03936659>.

- [3] 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](https://hal.inria.fr/hal-02074474). URL: <https://hal.inria.fr/hal-02074474>.
- [4] E. Bourne, P. Leleux, K. Kormann, C. Kruse, V. Grandgirard, Y. Güçlü, M. J. Kühn, U. Rüde, E. Sonnendrücker and E. Zoni. *Solver comparison for Poisson-like equations on tokamak geometries*. 23rd Sept. 2022. URL: <https://cea.hal.science/cea-03786723>.
- [5] B. Bramas, O. Coulaud and G. Sylvand. ‘Time-domain BEM for the wave equation on distributed-heterogeneous architectures: A blocking approach’. In: *Parallel Computing* 49 (July 2015), pp. 66–82. DOI: [10.1016/j.parco.2015.07.005](https://hal.inria.fr/hal-01200727). URL: <https://hal.inria.fr/hal-01200727>.
- [6] F. Casenave, A. Ern and G. Sylvand. ‘Coupled BEM-FEM for the convected Helmholtz equation with non-uniform flow in a bounded domain’. In: *Journal of Computational Physics* 257.A (15th Jan. 2014), pp. 627–644. DOI: [10.1016/j.jcp.2013.10.016](https://hal.science/hal-00805477). URL: <https://hal.science/hal-00805477>.
- [7] S. Cools, E. F. Yetkin, E. Agullo, L. Giraud and W. Vanroose. ‘Analyzing the Effect of Local Rounding Error Propagation on the Maximal Attainable Accuracy of the Pipelined Conjugate Gradient Method’. In: *SIAM Journal on Matrix Analysis and Applications* 39.1 (13th Mar. 2018), pp. 426–450. DOI: [10.1137/17M1117872](https://hal.inria.fr/hal-01753411). URL: <https://hal.inria.fr/hal-01753411>.
- [8] D. A. Di Pietro, P. Matalon, P. Mycek and U. Rüde. ‘High-order multigrid strategies for HHO discretizations of elliptic equations’. In: *Numerical Linear Algebra with Applications* (22nd June 2022). DOI: [10.1002/nla.2456](https://hal.science/hal-03531293). URL: <https://hal.science/hal-03531293>.
- [9] L. Giraud, Y.-F. Jing and Y. 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* 43.2 (2022), pp. 710–739. DOI: [10.1137/21m1401127](https://hal.inria.fr/hal-03546496). URL: <https://hal.inria.fr/hal-03546496>.
- [10] Y. Jang, J. Bonelle, C. Kruse, F. Hülsemann and U. Ruede. *Fast Linear Solvers for Incompressible CFD Simulations with Compatible Discrete Operator Schemes*. 27th Apr. 2023. URL: <https://hal.science/hal-04087358>.
- [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 (7th June 2017). DOI: [10.1063/1.4984266](https://hal.inria.fr/hal-01534134). URL: <https://hal.inria.fr/hal-01534134>.

10.2 Publications of the year

International journals

- [12] V. Acary, P. Armand, H. Minh Nguyen and M. Shpakovych. ‘Second order cone programming for frictional contact mechanics using interior point algorithm’. In: *Optimization Methods and Software* (16th Jan. 2024), pp. 1–31. DOI: [10.1080/10556788.2023.2296438](https://hal.science/hal-03913568). URL: <https://hal.science/hal-03913568>.
- [13] E. Agullo, A. Buttari, A. Guermouche, J. Herrmann and A. Jego. ‘Task-based parallel programming for scalable matrix product algorithms’. In: *ACM Transactions on Mathematical Software* (2023). DOI: [10.1145/3583560](https://hal.science/hal-03936659). URL: <https://hal.science/hal-03936659>.
- [14] A.-k. Mohamed El Maarouf, L. Giraud, A. Guermouche and T. Guignon. ‘Combining reduction with synchronization barrier on multi-core processors’. In: *Concurrency and Computation: Practice and Experience* 35.1 (10th Jan. 2023), e7402. DOI: [10.1002/cpe.7402](https://inria.hal.science/hal-03948901). URL: <https://inria.hal.science/hal-03948901>.

International peer-reviewed conferences

- [15] E. Agullo, A. Buttari, O. Coulaud, L. Eyraud-Dubois, M. Faverge, A. Franc, A. Guermouche, A. Jago, R. Peressoni and F. Pruvost. 'On the Arithmetic Intensity of Distributed-Memory Dense Matrix Multiplication Involving a Symmetric Input Matrix (SYMM)'. In: *International Parallel and Distributed Processing Symposium*. IPDPS 2023 - 37th International Parallel and Distributed Processing Symposium. St. Petersburg, FL, United States, June 2023, pp. 357–367. URL: <https://inria.hal.science/hal-04093162>.
- [16] B. Gobé, J. Saucourt, M. Shpakovych, G. Maulion, D. Helbert, D. Pagnoux, V. Kermene and A. Desfarges-Berthelemot. 'Machine learning method to measure the transmission matrix of a multimode optical fiber without reference beam for 3D beam tailoring'. In: *Photonic West 2024 : Laser Resonators, Microresonators, and Beam Control XXVI*. Vol. Laser Resonators, Microresonators, and Beam Control XXVI. San Francisco, USA, United States, 27th Jan. 2024, Paper Number: 12871–31. URL: <https://hal.science/hal-04245837>.

National peer-reviewed Conferences

- [17] E. Agullo, A. Buttari, M. Felšöci and G. Sylvand. 'Vers un solveur direct à base de tâches pour des systèmes linéaires FEM/BEM creux/denses'. In: *ComPAS 2023 - Conférence francophone d'informatique en Parallélisme, Architecture et Système*. Annecy, France, 4th July 2023. URL: <https://inria.hal.science/hal-04178064>.

Doctoral dissertations and habilitation theses

- [18] M. Felšöci. 'Solveurs rapides pour l'aéroacoustique haute fréquence'. Université de Bordeaux, 22nd Feb. 2023. URL: <https://theses.hal.science/tel-04077474>.
- [19] A.-K. Mohamed El Maarouf. 'Incomplete factorization and solution of triangular systems for fine-grained parallelism computers'. Université de Bordeaux, 17th Mar. 2023. URL: <https://theses.hal.science/tel-04429547>.
- [20] R. Peressoni. 'Large Scale Multidimensional Scaling for the Study of Biodiversity'. Université de Bordeaux, 13th June 2023. URL: <https://theses.hal.science/tel-04212482>.

Reports & preprints

- [21] M. A. Abouabdallah, O. Coulaud, N. Peyrard and A. Franc. *Computing WSBM marginals with Tensor-Train decomposition*. 15th Jan. 2024. URL: <https://hal.inrae.fr/hal-04394024>.
- [22] C. Acary-Robert, L. Courtès, Y. Dupont, M. Felšöci, K. Hinsén, O. Lünsdorf, P. Prins, P. Swartvagher, S. Tournier and R. Wurmus. *Guix-HPC Activity Report 2021–2022*. Inria; Max Delbrück Center for Molecular Medicine; Utrecht Bioinformatics Center, 10th Feb. 2023. URL: <https://inria.hal.science/hal-04013734>.
- [23] J. Briant, P. Mycek, M. Destouches, O. Goux, S. Gratton, S. Gürol, E. Simon and A. Weaver. *A filtered multilevel Monte Carlo method for estimating the expectation of discretized random fields*. 2023. DOI: [10.48550/arXiv.2311.06069](https://doi.org/10.48550/arXiv.2311.06069). URL: <https://hal.science/hal-04337114>.
- [24] M. Destouches, P. Mycek and S. Gürol. *Multivariate extensions of the Multilevel Best Linear Unbiased Estimator for ensemble-variational data assimilation*. TR-PA-23-67. Cerfacs, 12th June 2023. DOI: [10.48550/arXiv.2306.07017](https://doi.org/10.48550/arXiv.2306.07017). URL: <https://hal.science/hal-04337141>.
- [25] M. R. El Amri, P. Mycek, S. Ricci and M. De Lozzo. *Multilevel Surrogate-based Control Variates*. 2023. URL: <https://hal.science/hal-04129561>.
- [26] Y. Jang, J. Bonelle, C. Kruse, F. Hülsemann and U. Ruede. *Fast Linear Solvers for Incompressible CFD Simulations with Compatible Discrete Operator Schemes*. 27th Apr. 2023. URL: <https://hal.science/hal-04087358>.
- [27] M. Shpakovych. *Neural network preconditioning of large linear systems*. RT-0518. Inria Centre at the University of Bordeaux, 23rd Oct. 2023, p. 36. URL: <https://inria.hal.science/hal-04254315>.

Other scientific publications

- [28] M. Iannacito, O. Coulaud and L. Giraud. *Orthogonalization schemes in tensor train format*. Glasgow, United Kingdom, 27th June 2023. URL: <https://inria.hal.science/hal-04394115>.
- [29] M. Iannacito, O. Coulaud and L. Giraud. *Orthogonalization schemes in tensor train format*. Eindhoven / Hybrid, Netherlands, 10th July 2023. URL: <https://inria.hal.science/hal-04394130>.
- [30] Y. Xiang, L. Giraud, P. Mycek and C. Kruse. *Hybridization of Machine Learning and Numerical Linear Algebra Techniques for Scientific Computing: Learned Minimum Residual Solvers for the Helmholtz Equations*. Amsterdam, Netherlands, 28th Feb. 2023. URL: <https://inria.hal.science/hal-04398087>.

10.3 Cited publications

- [31] E. Agullo, O. Aumage, B. Bramas, O. Coulaud and S. Pitoiset. ‘Bridging the gap between openMP and task-based runtime systems for the fast multipole method’. In: *IEEE Transactions on Parallel and Distributed Systems* 28.10 (2017). DOI: [10.1109/TPDS.2017.2697857](https://doi.org/10.1109/TPDS.2017.2697857).
- [32] E. Agullo, B. Bramas, O. Coulaud, E. Darve, M. Messner and T. Takahashi. ‘Task-based FMM for heterogeneous architectures’. In: *Concurrency and Computation: Practice and Experience* 28.9 (June 2016), pp. 2608–2629. DOI: [10.1002/cpe.3723](https://doi.org/10.1002/cpe.3723). URL: <http://doi.wiley.com/10.1002/cpe.3723>.
- [33] 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](https://doi.org/10.1137/130915662). URL: <https://hal.inria.fr/hal-00911856>.
- [34] 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](https://doi.org/10.1137/18M122858X). URL: <https://hal.inria.fr/hal-03022845>.
- [35] E. Agullo, E. Darve, L. Giraud and Y. Harness. ‘Low-Rank Factorizations in Data Sparse Hierarchical Algorithms for Preconditioning Symmetric Positive Definite Matrices’. In: *SIAM Journal on Matrix Analysis and Applications* 39.4 (Oct. 2018), pp. 1701–1725. URL: <https://hal.inria.fr/hal-01940053>.
- [36] 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*. Technical Report RT-0513. Inria Bordeaux Sud-Ouest, June 2021, p. 100. URL: <https://hal.inria.fr/hal-03263620>.
- [37] 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 (2014), pp. 1625–1651.
- [38] 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](https://doi.org/10.1137/17M1153765). URL: <https://hal.inria.fr/hal-02074474>.
- [39] P. Blanchard. ‘Fast hierarchical algorithms for the low-rank approximation of matrices, with applications to materials physics, geostatistics and data analysis’. PhD thesis. Bordeaux, 2017. URL: <https://tel.archives-ouvertes.fr/tel-01534930>.
- [40] P. Blanchard, O. Coulaud and E. Darve. *Fast hierarchical algorithms for generating Gaussian random fields*. Research Report 8811. Inria Bordeaux Sud-Ouest, Dec. 2015. URL: <https://hal.inria.fr/hal-01228519>.
- [41] S. Börm, L. Grasedyck and W. Hackbusch. *Hierarchical Matrices*. Tech. rep. 2003, pp. 1–173.
- [42] A. Buttari, J. Langou, J. Kurzak and J. Dongarra. ‘Parallel tiled QR factorization for multicore architectures’. In: *Concurrency and Computation: Practice and Experience* 20.13 (2008), pp. 1573–1590.

- [43] E. Carson, N. J. Higham and S. Pranesh. ‘Three-Precision GMRES-Based Iterative Refinement for Least Squares Problems’. In: *SIAM Journal on Scientific Computing* 42.6 (Jan. 2020), A4063–A4083. DOI: [10.1137/20m1316822](https://doi.org/10.1137/20m1316822).
- [44] F. Casenave, A. Ern and G. Sylvand. ‘Coupled BEM-FEM for the convected Helmholtz equation with non-uniform flow in a bounded domain’. In: *Journal of Computational Physics* 257.A (Jan. 2014). 23 pages, 9 figures, pp. 627–644. DOI: [10.1016/j.jcp.2013.10.016](https://doi.org/10.1016/j.jcp.2013.10.016). URL: <https://hal.archives-ouvertes.fr/hal-00805477>.
- [45] A. Cichocki, R. Zdunek, A. H. Phan and S.-i. Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley, 2009.
- [46] S. Cools, E. F. Yetkin, E. Agullo, L. Giraud and W. Vanroose. ‘Analyzing the Effect of Local Rounding Error Propagation on the Maximal Attainable Accuracy of the Pipelined Conjugate Gradient Method’. In: *SIAM Journal on Matrix Analysis and Applications* 39.1 (Mar. 2018), pp. 426–450. DOI: [10.1137/17M1117872](https://doi.org/10.1137/17M1117872). URL: <https://hal.inria.fr/hal-01753411>.
- [47] O. Coulaud, A. A. Franc and M. Iannacito. *Extension of Correspondence Analysis to multiway datasets through High Order SVD: a geometric framework*. Research Report RR-9429. Inria Bordeaux - Sud-Ouest ; Inrae, Nov. 2021. URL: <https://hal.inria.fr/hal-03418404>.
- [48] O. Coulaud, L. Giraud and M. Iannacito. *On some orthogonalization schemes in Tensor Train format*. Tech. rep. RR-9491. Inria Bordeaux - Sud-Ouest, Nov. 2022. URL: <https://hal.science/hal-03850387>.
- [49] A. Falco. ‘Comblent l’écart entre \mathcal{H} -Matrices et méthodes directes creuses pour la résolution de systèmes linéaires de grandes tailles’. Theses. Université de Bordeaux, June 2019. URL: <https://tel.archives-ouvertes.fr/tel-02183902>.
- [50] A. A. Franc, P. Blanchard and O. Coulaud. ‘Nonlinear mapping and distance geometry’. In: *Optimization Letters* 14.2 (2020), pp. 453–467. DOI: [10.1007/s11590-019-01431-y](https://doi.org/10.1007/s11590-019-01431-y). URL: <https://hal.inria.fr/hal-02124882>.
- [51] N. Gillis. *Nonnegative Matrix Factorization*. Society for Industrial and Applied Mathematics, Jan. 2020. DOI: [10.1137/1.9781611976410](https://doi.org/10.1137/1.9781611976410).
- [52] 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*. Research Report RR-9393. Inria Bordeaux Sud-Ouest, Feb. 2021, p. 60. URL: <https://hal.inria.fr/hal-03146213>.
- [53] L. Grasedyck, W. Hackbusch and B. Nr. ‘An Introduction to Hierarchical (H -) Rank and TT - Rank of Tensors with Examples’. In: *Computational Methods in Applied Mathematics* 11.329 (2011), pp. 291–304.
- [54] B. Gunter and R. Van De Geijn. ‘Parallel out-of-core computation and updating of the QR factorization’. In: *ACM Transactions on Mathematical Software (TOMS)* 31.1 (2005), pp. 60–78.
- [55] W. Hackbusch. *Hierarchical Matrices: Algorithms and Analysis*. 1st. Springer Publishing Company, Incorporated, 2015.
- [56] N. Halko, P.-G. G. Martinsson and J. A. Tropp. ‘Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions’. In: *SIAM Review* 53.2 (2011), pp. 217–288. DOI: [10.1137/090771806](https://doi.org/10.1137/090771806). URL: <http://arxiv.org/abs/0909.4061>.
- [57] T. G. Kolda and B. W. Bader. ‘Tensor Decompositions and Applications’. In: *SIAM Review* 51.3 (Aug. 2009), pp. 455–500. DOI: [10.1137/07070111X](https://doi.org/10.1137/07070111X). URL: <http://epubs.siam.org/doi/abs/10.1137/07070111X>.
- [58] C. Kruse, V. Darrigrand, N. Tardieu, M. Arioli and U. Rude. ‘Application of an iterative Golub-Kahan algorithm to structural mechanics problems with multi-point constraints’. In: *Adv. Model. Simul. Eng. Sci.* 7.1 (2020), p. 45. DOI: [10.1186/s40323-020-00181-2](https://doi.org/10.1186/s40323-020-00181-2). URL: <https://doi.org/10.1186/s40323-020-00181-2>.
- [59] C. Kruse, M. Sosonkina, M. Arioli, N. Tardieu and U. Rude. ‘Parallel solution of saddle point systems with nested iterative solvers based on the Golub-Kahan Bidiagonalization’. In: *Concurr. Comput. Pract. Exp.* 33.11 (2021). DOI: [10.1002/cpe.5914](https://doi.org/10.1002/cpe.5914). URL: <https://doi.org/10.1002/cpe.5914>.

- [60] V. Le Bris, M. Odunlami, D. Bégué, I. Baraille and O. Coulaud. ‘Using computed infrared intensities for the reduction of vibrational configuration interaction bases’. In: *Phys. Chem. Chem. Phys.* 22 (13 2020), pp. 7021–7030. DOI: [10.1039/DOCP00593B](https://doi.org/10.1039/DOCP00593B). URL: <http://dx.doi.org/10.1039/DOCP00593B>.
- [61] B. Lizé. ‘Résolution directe rapide pour les éléments finis de frontière en électromagnétisme et acoustique : \mathcal{H} -Matrices. Parallélisme et applications industrielles’. Theses. Université Paris-Nord - Paris XIII, June 2014. URL: <https://tel.archives-ouvertes.fr/tel-01244260>.
- [62] P.-G. Martinsson and J. Tropp. ‘Randomized Numerical Linear Algebra: Foundations & Algorithms’. In: (2020). eprint: [2002.01387](https://arxiv.org/abs/2002.01387). URL: <http://arxiv.org/abs/2002.01387>.
- [63] M. Odunlami, V. Le Bris, D. Bégué, I. Baraille and O. Coulaud. ‘A-VCI: A flexible method to efficiently compute vibrational spectra’. In: *The Journal of Chemical Physics* 146.21 (June 2017), p. 214108. DOI: [10.1063/1.4984266](https://doi.org/10.1063/1.4984266). URL: <http://aip.scitation.org/doi/10.1063/1.4984266>.
- [64] I. V. Oseledets. ‘Tensor-Train Decomposition’. In: *SIAM Journal on Scientific Computing* 33.5 (Jan. 2011), pp. 2295–2317. DOI: [10.1137/090752286](https://doi.org/10.1137/090752286). URL: <https://doi.org/10.1137/090752286>.
- [65] L. Poirel. ‘Algebraic domain decomposition methods for hybrid (iterative/direct) solvers’. Theses. Université de Bordeaux, Nov. 2018. URL: <https://tel.archives-ouvertes.fr/tel-02070618>.
- [66] J.-R. Poirier, O. Coulaud and O. Kaya. ‘Fast BEM Solution for 2-D Scattering Problems Using Quantized Tensor-Train Format’. In: *IEEE Transactions on Magnetics* 56.3 (Mar. 2020), pp. 1–4. DOI: [10.1109/TMAG.2019.2954584](https://doi.org/10.1109/TMAG.2019.2954584). URL: <https://hal.inria.fr/hal-03150956>.
- [67] G. Sylvand. ‘La méthode multipôle rapide en électromagnétisme. Performances, parallélisation, applications’. Theses. Ecole des Ponts ParisTech, June 2002. URL: <https://pastel.archives-ouvertes.fr/tel-00005683>.
- [68] N. Venkovic, P. Mycek, L. Giraud and O. Le Maitre. *Recycling Krylov subspace strategies for sequences of sampled stochastic elliptic equations*. Research Report RR-9425. Inria Bordeaux - Sud Ouest, Oct. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03366966>.