

1) Etat civil

Né le 13 octobre 1971 à Montreuil (93), marié, 1 enfant.

Adresse personnelle : 1c rue du clos du Haut Carré, 33400 Talence.

Adresse professionnelle : Université Bordeaux, 351 cours de la Libération, 33405 Talence Cedex.

2) Parcours Professionnel

Situation professionnelle actuelle

Statut et fonction: Maître de conférences HC HDR

Etablissement: Université de Bordeaux - France

Date d'entrée en fonction: septembre 2000

Expériences professionnelles antérieures

Date début	Date fin	Etablissement	Fonction et statut
2015	-	CEA CESTA	Conseiller scientifique pour le HPC
2000	-	Université de Bordeaux	Maître de Conférences HC HDR
			Enseignement au département informatique de l'IUT
			Recherche LaBRI (UMR 5800) et Inria projet HiePACS
2011	2012	Inria projet Bacchus	Délégation
2003	2005	Inria projet ScAIApplix	Délégation

3) Encadrement d'étudiants et de jeunes chercheurs

J'encadre actuellement deux thèses :

- **la thèse de Clément Richefort financée sur bourse CEA [directeur, encadrement à 100%]**. Il travaille sur le développement de nouvelles méthodes de résolution de problèmes pour l'électromagnétisme avec une scalabilité accrue. La piste de recherche la plus prometteuse est la méthode multigrille. Le principe de cette méthode est l'utilisation d'une collection de problèmes grossiers qui permettent d'accélérer le calcul de la solution fine. C'est une méthode itérative, qui est éprouvée pour des problèmes elliptiques, et ayant une scalabilité optimale. Cependant, il est connu que les méthodes multigrilles ne sont pas performantes pour des problèmes à noyaux oscillants tels que l'électromagnétisme ou l'acoustique (ces équations conduisant à des problèmes indéfinis). Un travail important de recherche est donc nécessaire pour trouver des opérateurs multigrilles qui soient performants pour ce type de problèmes.
- **la thèse d'Eragul Korkmaz financée sur le projet ANR SASHIMI [directeur, encadrement à 50%]**. Elle s'est intéressée dans la première partie de sa thèse aux alternatives aux factorisations QR-SVD qui présentent l'avantage de garantir la qualité de la compression après troncature, mais les inconvénients d'une complexité importante et d'une portabilité difficile sur les accélérateurs de type GPU. Nous avons réalisé une étude comparative des solutions basées sur des techniques de randomisation et intégré les noyaux de compression PQRCP (factorisation QR avec pivotage), RQRCP (avec randomisation), TQRCP (avec troncature), et RQRRT (avec rotation). Dans la seconde partie de sa thèse, nous avons proposé une méthode pour identifier les blocs avec de faibles taux de compression. Nous montrons que la factorisation LU incomplète par blocs permet d'identifier, à faible coût, la plupart de ces blocs non compressibles et ainsi obtenir un bon compromis entre efficacité et réduction mémoire.

J'ai co-encadré 6 thèses soutenues :

- **la thèse de Grégoire Pichon [124] financée sur un contrat DGA, soutenue en novembre 2018 [directeur, encadrement à 50%]**. Il est actuellement maître de conférences à l'ENS-Lyon.
Le sujet de recherche concerne l'exploitation des techniques de compression Block Low-Rank (BLR) pour la résolution de grands systèmes linéaires creux. En pratique, les blocs les plus volumineux sont compressés avec une décomposition en valeurs singulières (SVD) ou un algorithme de type Rank-Revealing QR (RRQR). Deux versions du solveur BLR ont été développées. La stratégie « minimal memory » permet de réduire l'empreinte mémoire avec un surcoût en temps limité. L'intérêt principal est que des problèmes qui étaient trop volumineux pour être résolus

auparavant peuvent maintenant passer en mémoire. La seconde stratégie « just-in-time » permet de réduire le temps de résolution, avec un d'impact plus limité sur la consommation mémoire.

- **la thèse de Salli Moustafa [125] financée sur un contrat de collaboration passé avec EDF-R&D, soutenue en décembre 2015 [encadrement à 50%].** Il est actuellement consultant chez ANEO après avoir travaillé chez EDF. L'objectif de cette thèse était la mise au point d'une méthode de résolution massivement parallèle de l'équation de transport neutronique. La réduction du temps de retour à moins d'une heure, pour des calculs impliquant plus de 10 milliards d'inconnues, a constitué une avancée importante pour la simulation des coeurs des centrales nucléaires. Pour y parvenir, il était nécessaire d'utiliser les techniques de parallélisation les plus avancées impliquant plusieurs algorithmes parallèles imbriqués ainsi que plusieurs paradigmes de programmations parallèles (passage de message, multi-threading et instructions SIMD).
- **la thèse d'Astrid Casadei [126] financée sur une allocation de recherche ministérielle, soutenue en octobre 2015 [encadrement à 80%].** Elle travaille actuellement chez Naval Group. Dans le contexte des méthodes hybrides basées sur une approche par complément de Schur, il est critique de contrôler l'équilibrage de la taille des interfaces. Dans cette thèse, nous avons étudié une piste prometteuse qui consiste à prendre en compte la taille réelle des interfaces dans l'étape récursive d'une méthode de dissection emboîtée. Cette activité de recherche a été menée en collaboration avec Y. Saad de l'Université du Minnesota et E. Ng dans le cadre de l'équipe associée FastLA. Trouver de nouveaux algorithmes et les implémenter de manière générique dans les principaux outils de partitionnement de graphe étaient à la fois un défi et une avancée majeure pour la résolution de très grands systèmes linéaires creux hors de portée des méthodes directes.
- **la thèse de Xavier Lacoste [127] financée sur le projet ANR ANEMOS, soutenue en février 2015 [encadrement à 80%].** Il travaille actuellement chez TOTAL. Il s'agissait d'adapter des algorithmes d'ordonnancement pour tirer parti d'un environnement d'exécution exploitant les architectures mixte CPUs/GPGPUs. Nous avons étudié la possibilité de remplacer les ordonnanceurs dynamiques basés sur des stratégies internes par des supports d'exécution génériques. Sur des noeuds de calcul disposant d'accélérateurs, ces supports d'exécution offrent la possibilité de dérouler le graphe de tâches de la factorisation numérique en s'exonérant la prise en compte de la complexité du matériel qui est prise en charge par le moteur d'exécution qui analyse le graphe et ses dépendances. Les premiers résultats ont montré qu'une approche basée sur un DAG (Directed Acyclic Graph) offre un cadre de programmation uniforme et portable pour réaliser du calcul haute performance sur des problèmes irréguliers d'algèbre linéaire creuse sur des nœuds hétérogènes.
- **la thèse de Bruno Lathuilière [128] financée sur un contrat de collaboration passé avec EDF-R&D, soutenue en janvier 2010 [encadrement à 80%].** Il travaille actuellement chez EDF. Les calculs de réactivité constituent une brique fondamentale dans la simulation des coeurs des réacteurs nucléaires. Ceux-ci conduisent à la résolution de problèmes aux valeurs propres généralisées via l'algorithme de la puissance inverse. A chaque itération, on est amené à résoudre un système linéaire de manière approchée via un algorithme d'itérations imbriquées. Au cours de cette thèse, nous avons étudié une méthode de décomposition de domaine de type Schur dual. Plusieurs placements de l'algorithme de décomposition de domaine au sein du système d'itérations imbriquées ont été considérés. Les résultats obtenus permettent d'envisager l'industrialisation de la méthodologie associée.
- **la thèse de Mathieu Faverge [129] financée sur le projet ANR NUMASIS, soutenue en décembre 2009 [encadrement à 50%].** Il est actuellement maître de conférences à Bordeaux INP.

Les nouvelles architectures de calcul intensif intègrent de plus en plus de microprocesseurs qui eux-mêmes intègrent un nombre croissant de coeurs de calcul. Cette multiplication des unités de calcul dans les architectures a fait apparaître des topologies fortement hiérarchiques. Nous avons étudié un ordonnancement dynamique adapté aux architectures NUMA (Non-Uniform Memory Acces) pour un solveur direct creux supernodal. Les structures de données du solveur, ainsi que les schémas de communication ont dû être modifiés pour s'adapter aux caractéristiques de ces architectures et à un ordonnancement dynamique. Nous nous sommes également intéressés à l'adaptation dynamique du grain de calcul pour exploiter au mieux les architectures multi-coeurs et la mémoire partagée.

4) Encadrement de développements technologiques

Je suis l'architecte et j'assure le développement ainsi que le suivi de la bibliothèque PaStiX [130] qui implémente un solveur haute performance pour la résolution de grands systèmes linéaires basé sur une approche directe supernodale. Ces travaux seront menés dans un contexte de concurrence forte avec des équipes internationales renommées. Ce solveur est intégré dans plusieurs autres logiciels, qui contribuent à sa diffusion indirecte.

5) Responsabilités collectives

- Membre des comités de lecture pour les revues : TOMS (ACM), SISC (SIAM), JCP (Elsevier), Parallel Computing (Elsevier), Computer & Fluids (Elsevier), Concurrency and Computation: Practice and Experience (Wiley), Numerical Methods in Fluids (Wiley), Numerical Methods in Engineering (Wiley).
- Membre des comités de lecture pour les conférences : IPDPS (IEEE), EuroPar (Springer), PMAA, PARENG, Compas.
- Depuis septembre 2015, **je suis conseiller scientifique pour le domaine du HPC au CEA CESTA** (Centre d'Etudes Scientifiques et Techniques d'Aquitaine), un des établissements de la Direction des Applications Militaires (DAM). Mon activité de prospective scientifique auprès des ingénieurs me permet de transmettre les grandes orientations prises par la communauté internationale dans le domaine du HPC et plus particulièrement en algorithmique numérique et parallèle.
- Je suis candidat pour prendre la **présidence du comité thématique CT6 au GENCI** lors du prochain renouvellement courant 2022.
- Entre 2008 et 2016, **j'étais expert au GENCI pour le comité thématique "maths-informatique" CT6** avec pour mission d'étudier les demandes d'allocation de ressources de calcul lors des appels nationaux pour les supercalculateurs du CINES, CCRT et IDRIS. Pendant 12 ans, je me suis occupé des demandes de ressources informatiques de l'équipe sur les centres de calcul nationaux CINES, IDRIS et CCRT maintenant regroupés dans le cadre d'un appel commun sous la direction du GENCI (Grand Equipement National de Calcul Intensif).
- Membre de l'organisation de la 9^{ième} édition de la conférence "Parallel Matrix Algorithms and Applications" qui a lieu tous les deux ans et qui s'est déroulée à Bordeaux en juillet 2016.
- Organisateur du 3^{ième} workshop HOSCAR en septembre 2013 à Bordeaux, un projet collaboratif porté par Stéphane Lanteri entre le CNPq (Brésil) et Inria.
- Co-organisateur de la conférence internationale "Preconditioning Techniques For Scientific And Industrial Applications" qui a lieu tous les deux ans et qui s'est déroulée à Bordeaux en mai 2011.

6) Management

- Depuis janvier 2021, je participe aux groupes de travail "Energie" et "Sciences et Usages du Numérique" dans le cadre du **projet Exascale français**.
- Depuis janvier 2020, **je suis responsable du département SATANAS du LaBRI** qui regroupe les équipes projets Inria HiePACS, Storm et Tadaam, soit environ 25 permanents et 25 non-permanents. J'étais le responsable adjoint de SATANAS depuis 2011, et à ce titre, je participe au conseil scientifique de LaBRI. **J'ai représenté le département SATANAS lors de l'évaluation HCERES du LaBRI en janvier 2021.**
- Membre du groupe de travail sur les infrastructures Inria, piloté par F. Desprez et L. Saccavini, et dont les conclusions ont été remises en novembre 2019. A cette occasion, j'ai présenté PlaFRIM (Plateforme Fédérative pour la Recherche en Informatique et Mathématiques) lors du workshop TILECS (Towards an Infrastructure for Large-Scale Experimental Computer Science) à Grenoble en juillet 2019.
- Membre de la CDT (Commission de Développement Technologique qui examine en particulier les demandes ADT et ATT) pour le centre Inria de Bordeaux depuis 2015.
- Membre du conseil scientifique et du bureau du MCIA (Mésocentre de Calcul Intensif Aquitain) entre 2008 et 2016.

7) Collaborations

Collaborations internationales ayant données lieu à des séjours courts et à des séminaires :

- **Collaboration avec Stanford** : Depuis 2013, dans le cadre de l'équipe associée FAST-LA, **j'ai mis en place une collaboration avec Eric Darve** sur l'application des techniques de compression utilisant des matrices hiérarchiques pour l'accélération et la réduction de la complexité des solveurs directes. Une *bourse Chateaubrillard* a été accordée en 2014, dans le cadre de cette collaboration, à Clif Robert Dudley, étudiant de master à Stanford. Cet échange a permis de définir un premier prototype utilisé comme base pour le solveur développé par Grégoire Pichon dans sa thèse [50].

- **Collaboration avec l'ICL** : Depuis 2011, dans le cadre de l'équipe associée MORSE, **j'ai mis en place une collaboration avec Jack Dongarra¹ qui a obtenu cette année le prix Turing**. Mon premier doctorant, Mathieu Faverge, a effectué un post-doctorat de 2 ans à l'ICL (Innovative Computing Laboratory). Nous avons ainsi utilisé l'ordonnanceur générique PaRSEC développé par l'équipe de G. Bosilca, afin de proposer la première implémentation d'un solveur creux direct sur des architectures multiCPUs et multiGPUs [21].
- **Collaboration avec le BSC** : Depuis 2015, dans le cadre du projet JLESC, un laboratoire commun (Urbana-Champaign University, Argonne National Laboratory, Barcelona Supercomputing Center (BSC), Jülich Supercomputing Centre et RIKEN Center for Computational Science) autour du calcul intensif, j'ai développé une collaboration avec Guillaume Houzeau au BSC sur l'intégration du solveur PaSTiX dans la plateforme HPC multi-physiques Alya [37].
- **Collaboration avec le JAEA** : En 2007, j'ai participé au lancement d'un projet de collaboration avec le JAEA (Japan Atomic Energy Agency, qui participe au projet ITER), initié par M. Dayd (directeur de l'IRIT et actuellement chargé de mission "grilles de calcul et supercalculateurs" par le CNRS). Le JAEA a accès à l'infrastructure du projet NAREGI (NAtional REsearch Grid Initiative), un équivalent japonais du projet Grid'5000 (maintenant SILECS) en France, et développe également ses propres solveurs en algèbre linéaire pour ses applications de simulation numérique. Cette collaboration a donné lieu à plusieurs présentations [24, 25, 93].
- **Collaboration NSF/Inria** : Nos travaux ont permis d'établir des échanges internationaux dans le cadre d'un contrat NSF/Inria et lors de deux séjours à l'université de Minneapolis, sur invitation de Y. Saad². J'ai participé à l'équipe associée Phyleas, portée par J. Roman, impliquant Inria et l'université du Minnesota. Mes travaux sur les factorisations incomplètes [9] sont issues de ces échanges.

Collaborations nationales ayant données lieu à des déplacements sur les sites des partenaires et à des présentations :

- **ANR JCJC18 SASHIMI** : Ce projet continuera les travaux préliminaires sur les formats à un niveau (BLR) en passant à des formats hiérarchiques, qui offrent de meilleurs taux de compression et une plus grande réduction des coûts asymptotiques. Les travaux seront réalisés dans la bibliothèque PaSTiX qui est implémentée au-dessus de supports d'exécution. Cela permettra une meilleure adaptation à l'irrégularité des noyaux de calcul hiérarchiques. Le solveur sera étudié sur des architectures à mémoire partagée et distribuée et de nouvelles solutions dédiées à la distribution des données et à l'ordonnancement des calculs seront étudiées. Le résultat final sera comparé aux principaux compétiteurs en termes de précision, temps de calcul et consommation mémoire.
- **ANR MN13 SOLHAR** : Ce projet avait pour but d'étudier et de mettre au point des algorithmes et des modèles de programmation parallèles pour implanter des méthodes directes de résolution de systèmes linéaires creux sur les plates-formes émergentes équipées d'accélérateurs. Le but à long terme de ce projet est de développer une solution logicielle fournissant un solveur basé sur les méthodes directes pour résoudre des systèmes creux d'équations linéaires. A ce jour, les approches proposées pour atteindre cet objectif sont essentiellement basées sur un simple modèle de délégation de certaines tâches de calcul aux accélérateurs et requièrent une optimisation manuelle du code. Au contraire, l'approche proposée s'articule autour de trois axes de recherche: algèbre linéaire, moteurs d'exécution, ordonnancement. La suite à ce projet (**ANR SOLHARIS**) a été acceptée en 2019 avec les mêmes partenaires.
- **ANR MN11 ANEMOS** : Il s'agissait de la suite du projet **ANR ASTER** dans laquelle nous avons fait des progrès significatifs dans la compréhension de la physique inhérente aux stratégies de contrôle actif des instabilités MHD du bord du plasma dans les Tokamaks (et ITER en particulier). **J'étais le responsable scientifique pour le partenaire Inria/LaBRI**. Nous avons pu proposer l'adaptation un solveur direct creux aux architectures multicœurs/multigpus et les résultats ont été présentés au workshop HCW de la conférence IPDPS [21].
- **ANR COSINUS08 PETAL** : Mes travaux sur les factorisations incomplètes de type ILU s'inscrivaient aussi dans le projet PETAL de l'appel à projets 2008 de l'ANR "Cosinus". Il s'agissait en particulier de comparer nos approches avec celles développées par d'autres membres du projet. Ce projet a été renouvelé en 2010 (**ANR PETALH**). En particulier, une méthode a été développée pour réduire les surcoûts mémoires générés lors de la construction du complément de Schur dans un solveur direct.
- **ANR CIS06 SOLSTICE** : Mes travaux de recherche s'inscrivaient également dans le projet SOLSTICE de l'appel à projets 2006 de l'ANR "Calcul intensif". L'objectif de ce projet était la conception et la mise en œuvre haute performance de solveurs linéaires parallèles efficaces pour résoudre des problèmes scientifiques complexes multi-physiques et multi-échelles de très grande taille, et leur intégration effective dans des codes applicatifs académiques et industriels dans le but de faire des simulations "grands challenges".

¹https://en.wikipedia.org/wiki/Jack_Dongarra

²https://en.wikipedia.org/wiki/Yousef_Saad

- **ANR CIS06 ASTER** : Pour l'appel 2006 de l'ANR "Calcul intensif", avec Guido Huysmans, nous avons monté le projet ASTER dont j'étais le **responsable scientifique pour le partenaire Inria/LaBRI**. Ce projet visait à développer et implémenter des méthodes pour améliorer les codes en Magnéto-Hydro-Dynamique afin de permettre la simulation des instabilités dans les plasmas. La compréhension de ces instabilités est très limitée et les pertes d'énergie induites sont un réel problème pour le réacteur ITER. Nous avons en particulier travaillé sur la prise en compte de techniques de type raffinement de maillages adaptatifs et sur l'adaptation de nos solveurs. Les premiers résultats ont été présentés dans la revue *Plasma Physics and Controlled Fusion* [7].
- **ANR CIS05 NUMASIS** : Dans le cadre de l'appel 2005 de l'ANR "Calcul intensif", j'ai co-encadré la thèse de Mathieu Faverge financée sur le projet NUMASIS. Nous nous sommes intéressés aux problèmes d'algorithmique parallèle et aux modèles de programmation sur architectures NUMA. Cela concernait aussi bien les algorithmes spécifiques au domaine applicatif de la sismique que les noyaux et bibliothèques de calcul classiques qui interviennent dans la conception des codes scientifiques de simulation.

8) Diffusion de l'information scientifique

- Depuis janvier 2021, j'anime le **groupe de travail hebdomadaire de l'équipe HiePACS/TOPAL**. Un résumé des activités passées est disponible sur la page : <https://topal.gitlabpages.inria.fr/topal-working-group/>
- Participation à une table ronde lors des rencontres Inria Industrie en 2013 pour témoigner sur l'expérience du dispositif doctorant-conseil et sur le transfert industriel avec une PME.
- Publication d'un article [82] dans la revue CHOCS en 2012. Il s'agit de la revue scientifique et technique de la DAM, d'un bon haut niveau scientifique tout en étant accessible à un public non expert http://www-physique-chimie.ccea.fr/science-en-ligne/docs/chocs/Chocs_41.pdf.

9) Visibilité

- J'ai obtenu une **bourse de l'Université de Stanford** pour présenter mes travaux au séminaire du département "Mechanical Engineering" (ME395) en avril 2018.
- J'étais rapporteur et membre du jury de la thèse de Gilles Moreau (ENS-Lyon, décembre 2018).
- En dehors des jurys des thèses que j'ai encadrées, j'ai participé en tant qu'examineur aux jurys des thèses de :
 - Ashish Bhole (Université Côte d'Azur, novembre 2021),
 - Théo Mary (Université de Toulouse, novembre 2017), lauréat du prix SIF Gilles Kahn en 2018,
 - Stanislas Pamela (Université de Provence Aix-Marseille, septembre 2010).

10) Éléments divers

- Lors de mon service militaire, j'étais scientifique du contingent détaché au CEA CESTA dans le groupe "Analyse Numérique". J'ai travaillé à la mise en place d'un solveur parallèle en algèbre linéaire dense sur les supercalculateurs de la DAM Cray T3D puis T3E. Le code ARLENE (électromagnétisme) utilise encore une partie de ce travail et est l'un des codes de simulation du service les plus utilisés en production.
- Dans le cadre de l'initiative HPC-PME (BPI, Maison de la simulation, GENCI), **j'ai mis en place un contrat d'expertise avec la société Algo'Tech**.
- Avancement de grade au titre du CNU (MCF hors classe) en décembre 2017.
- Prime d'encadrement doctoral et de recherche (PEDR) depuis 2017.

DESCRIPTION SYNTHÉTIQUE DE L'ACTIVITÉ D'ENSEIGNEMENT

[Page web résumant mes activités d'enseignement : <https://ramet.gitlab.io/>]

J'effectue mon service statutaire au département informatique de l'IUT de l'université Bordeaux. Je dispense, en particulier, des cours en programmation, bases de données, systèmes et réseaux. De par mes activités d'enseignement dans des filières informatiques variées, j'ai été amené à créer des supports et à dispenser un grand nombre de contenus de cours, de l'algorithmique numérique à la programmation système, en passant par les bases de données et la cryptographie. Bien sûr, je dispose du recul et des connaissances nécessaires pour assurer des enseignements autour des thèmes de l'architecture, du système et du calcul parallèle.

Je dispense également un cours sur l'ordonnancement et l'algèbre linéaire parallèle en 3e année à l'ENSEIRB dans la filière informatique (option calcul parallèle), ainsi qu'un module d'algorithmique numérique en 1re année et un module de cryptographie en 2e année du cycle ingénieur.

Dernièrement, **je m'ai monté un module d'introduction à l'IA et à l'apprentissage automatique**. L'objectif est de présenter, à travers des exemples et des projets d'application, les grandes catégories de méthode d'apprentissage supervisé et non supervisé. Parmi les approches classiques, les algorithmes de "k plus proches voisins", les arbres de décision et le "deep learning" sont naturellement mis en oeuvre sur les frameworks d'apprentissage disponibles (TensorFlow, PyTorch, Keras ...) avec le langage Python et ses bibliothèques associées. Le contenu est centré sur 3 modules : l'analyse de données et les méthodes de réduction de dimension, les réseaux de neurones appliqués à la classification et à la régression, les méthodes d'optimisation et les techniques d'apprentissage par renforcement.

Depuis 2017, **je suis le responsable des poursuites d'études** pour les étudiants du département informatique.

Entre 2007 et 2014, **j'ai assuré la direction des licences professionnelles** (Systèmes Informatiques et Logiciels) pour les spécialités ACPI (Assistant Chef de Projet Informatique) et DAWIN (Développeur en Applications Web et Images Numériques).

En collaboration avec E. Fleury, nous avons construit et présenté une nouvelle offre de formation (SSRL pour Spécialiste en Sécurité des Réseaux et Logiciels) au niveau licence professionnelle autour de la sécurité dans les réseaux et logiciels en nous appuyant sur une partie des contenus et de l'expérience du master CSI (Cryptologie et Sécurité Informatique).

J'ai une expérience d'enseignement à l'étranger ayant dispensé des cours au Gabon (niveau licence) en 2006 et des cours au Vietnam (niveau licence et master en anglais) entre 2008 et 2016.

J'ai une grande expérience d'enseignement de la programmation par projets que j'exploite dans le contexte des codes développés pour le CEA ou pour ITER. La confrontation des méthodes agiles et classiques pour les projets de programmation doit s'inscrire dans une approche réunissant les composantes de développement (Dev) et l'exploitation des systèmes (Ops). Je partage ma propre expérience en développement logiciel en abordant les problématiques du déploiement et de la reproductibilité des résultats. La confiance dans la qualité d'un programme passe par l'utilisation des méthodes d'intégration continue ainsi que par la maîtrise des processus de gestion de configuration, de revue de code, et de livraison automatique.

Dans le cadre de la mise en place du **BUT (Bachelor Universitaire de Technologie)**, l'équipe pédagogique du département informatique de l'IUT de Bordeaux a décidé de proposer 2 des 4 parcours définis nationalement. Il s'agit du parcours A (Réalisation d'applications : conception, développement, validation) et du parcours D (Intégration d'applications et management du système d'information) qui couvre la plus grande partie des compétences des membres du département. **J'ai ainsi piloté, cette année, la mise en place de 2 nouvelles ressources :**

- Exploitation d'une base de données : l'objectif étant à la fois de permettre aux étudiants de monter en compétence dans l'écriture de requêtes avancées, mais aussi de les initier aux techniques de visualisation de données simples, structurées ou géographiques.
- Qualité de développement : l'objectif est de permettre aux étudiants de monter en compétence dans l'utilisation des outils de gestion de versions, de les sensibiliser à la production de tests unitaires et à la problématique de non-régression.

[Page web résumant mes activités recherche : <https://ramet.github.io/>]

Mots-clés : **simulation numérique, parallélisme, calcul haute performance, recouvrement calcul/communication, irrégularité, hétérogénéité, ordonnancement, distribution de données, algèbre linéaire creuse.**

Ma problématique de recherche concerne les problèmes du calcul haute performance et plus spécialement le calcul parallèle scientifique pour les problèmes réguliers et irréguliers et je m'intéresse en particulier aux problèmes de communication, de distribution et de prise en compte de l'hétérogénéité des plates-formes de calcul dont l'architecture est en permanente évolution.

Lorsque **j'ai rejoint l'équipe Inria ScAIApplix**, j'ai contribué à la résolution de simulations numériques complexes avec une approche couplant les mathématiques appliquées et l'informatique. **J'ai ensuite intégré l'équipe Inria Bacchus** pour me rapprocher des applications, en particulier la simulation MHD pour le projet ITER. **J'étais dans l'équipe Inria HiePACS** afin de développer de nouveaux algorithmes et méthodes numériques qui répondent aux besoins des simulations numériques modernes. **Je viens d'intégrer l'équipe Inria TOPAL** pour compléter mes travaux en algèbre linéaire aux applications manipulants des tenseurs (en particulier pour l'apprentissage automatique). De nouvelles contraintes, liées aux supports d'exécution, à la compression de données et à la sobriété énergétique, devront être prises en compte.

► Dans mes travaux de thèse, j'ai présenté des contributions concernant le recouvrement calcul/communication. J'ai proposé un modèle ainsi qu'un schéma de résolution permettant de déterminer analytiquement ou numériquement la taille optimale de paquets qui maximise le recouvrement dans les algorithmes macro-pipelines réguliers.

Ce travail s'inscrit dans une démarche plus générale du recouvrement sur des architectures totalement hétérogènes et pour des applications irrégulières. J'ai débuté une étude de modélisation et la recherche des tailles de paquets qui minimisent le temps d'exécution pour un pipeline sur une plate-forme hétérogène. Enfin, dans le cas d'une architecture hétérogène ou hiérarchique (typiquement un cluster de nœuds multicœurs), j'ai été amené à considérer le problème du recollement de pipelines présentant des grains de calcul différents.

► Ma principale contribution consiste à concevoir, développer et faire évoluer le solveur de la bibliothèque PaStiX qui fut le premier solveur direct à résoudre des systèmes avec plusieurs dizaines de millions d'inconnues, issus d'applications industrielles en trois dimensions [10].

Avec l'émergence des architectures à base de nœuds multicœurs (allant jusqu'à une centaine de cœurs par nœud), je travaille avec Mathieu Faverge sur une version dédiée de PaStiX pour de telles plateformes de calcul. Ainsi nous nous sommes orientés vers une version de programmation hybride Thread/MPI du solveur qui a pour principe de partager les données d'un même nœud multicœurs via l'utilisation de threads POSIX.

Le solveur PaStiX est utilisé en production dans des codes du CEA CESTA [12] et fait l'objet de collaborations avec le CEA CADARACHE autour de la fusion contrôlée et du réacteur ITER [7]. Il est également utilisé dans des applications du CERFACS pour la simulation de l'évolution du climat avec Météo-France.

► Mes travaux de recherche concernent également l'étude de préconditionneurs basés sur une factorisation parallèle incomplète par blocs. Le but de ces travaux est d'utiliser au mieux la technologie algorithmique parallèle par bloc qui est utilisée pour les solveurs directs pour développer des préconditionneurs parallèles de type Cholesky incomplet de manière adaptative et paramétrée, et qui soient robuste numériquement, suffisamment économes en mémoire, très performants en temps CPU et présentant une bonne scalabilité.

Ces préconditionneurs sont alors intégrés dans des implémentations par blocs de méthodes de Krylov du type gradient conjugué ou GMRES. On recherche donc un compromis entre une diminution importante de la taille mémoire pour stocker les facteurs de la méthode directe, et une conservation d'une certaine dose de remplissage pour exploiter suffisamment les effets superscalaires dans les calculs BLAS3 du préconditionneur et atteindre globalement de bonnes performances en temps. Nous conservons les techniques de renumérotation et de distribution/ordonnancement du solveur direct pour avoir une implémentation parallèle efficace du calcul du préconditionneur et des itérés. Un point crucial est donc le choix de la partition initiale pour la factorisation symbolique incomplète. Nous avons proposé une méthode pour trouver des supernœuds approximatés pour la factorisation $ILU(k)$. Les détails de cette démarche et l'analyse des résultats de convergence pour des systèmes de grandes tailles ont été publiés dans la revue *Parallel Computing* [9].

► J'ai également proposé des optimisations algorithmiques pour les principaux solveurs linéaires creux hybrides directs itératifs (tels que HIPS, MaPHYs, PDSLIN ou ShyLU) qui sont basés sur une décomposition de domaine et une approche par « complément de Schur ». Bien que ces solveurs soient moins coûteux en temps et en mémoire que leurs homologues directs, ils ne sont néanmoins pas exempts de surcoûts. Nous nous sommes intéressés à la question de l'équilibrage de la charge que pose la décomposition de domaine pour le calcul parallèle.

Ce problème revient à partitionner le graphe d'adjacence de la matrice en autant de parties que de domaines désirés. Nous

avons mis en évidence le fait que pour avoir un équilibrage correct des temps de calcul lors des phases les plus coûteuses d'un solveur hybride, il faut à la fois équilibrer les domaines en termes de nombre de nœuds et de taille d'interface locale. Cependant, les partitionneurs de graphes tels que `Scotch` et `MeTiS` ne s'intéressent toutefois qu'au premier critère (la taille des domaines) dans le contexte de la renumérotation des matrices creuses. Nous avons proposé plusieurs variantes des algorithmes existants afin de prendre également en compte l'équilibrage des interfaces locales et ces modifications ont été implémentées dans le partitionneur `Scotch` [20].

► Pour des architectures hétérogènes multiCPUs et multiGPUs, j'ai étudié les bénéfices et les limites que peut apporter le remplacement de l'ordonnanceur natif, très spécialisé, du solveur `PASIX` par deux systèmes d'exécution génériques: `PaRSEC` et `StarPU`. Pour cela, l'algorithme doit être décrit sous la forme d'un graphe de tâches qui est fourni aux systèmes d'exécution qui peuvent alors calculer une exécution optimisée de celui-ci pour maximiser l'efficacité de l'algorithme sur la machine de calcul visée. Une étude comparative des performances a été menée sur différentes architectures et l'analyse met en évidence les performances comparables des versions utilisant les systèmes d'exécution par rapport à l'ordonnanceur natif optimisé pour `PASIX`. De plus ces implémentations permettent d'obtenir une accélération notable sur les machines hétérogènes en utilisant les accélérateurs tout en masquant la complexité de leur utilisation au développeur [21].

► Enfin, ces dernières années, je me suis intéressé aux techniques de compression pour matrices denses globalement regroupées sous le vocable de H-matrices. Ces techniques reposent sur une représentation "data sparse" des opérateurs linéaires et exploitent l'information géométrique du maillage sous-jacent pour définir le "clustering", i.e., la structure hiérarchique de la matrice dense. J'ai proposé de prolonger ces idées dans le cadre purement algébrique des méthodes de factorisation de matrices creuses en substituant le calcul matriciel dense "classique" par du calcul H-matriciel afin de réduire l'empreinte mémoire des solveurs directs creux qui constitue le principal frein au passage à l'échelle de ces approches pour des simulations 3D. L'utilisation des familles de matrices hiérarchiques repose sur une approximation de rang faible des blocs denses dont la structure est définie récursivement.

Un solveur utilisant le format de compression *Block Low-Rank* (BLR) a été développé dans le cadre de la thèse de Grégoire Pichon. En pratique, les supernœuds les plus volumineux sont compressés en format *low-rank* avec une décomposition en valeurs singulières (SVD) ou un algorithme de type Rank-Revealing QR (RRQR). Deux versions du solveur *low-rank* ont été développées. La stratégie *Minimal Memory* permet de réduire l'empreinte mémoire avec un surcoût en temps limité. L'intérêt principal est que des problèmes qui étaient trop volumineux pour être résolus auparavant peuvent maintenant passer en mémoire. La seconde stratégie *Just-In-Time* permet de réduire le temps de résolution, sans avoir d'impact important sur la consommation mémoire [3].

Une des problématiques de la compression *low-rank* est le *clustering* des inconnues. En effet, les séparateurs issus de la *dissection emboîtée* doivent être redécoupés en un ensemble de blocs pour exhiber la nouvelle partition *low-rank*. Une approche classique consiste à effectuer un *kway* partitionnement sur les séparateurs, pour former des clusters compacts avec peu de voisins. Cependant, une telle approche ne prend pas en compte les contributions qui vont arriver sur ce séparateur. Nous étudions actuellement de nouvelles heuristiques pour isoler de bons clusters dans le séparateur ainsi qu'un ensemble d'éléments pas (ou peu) compressibles. L'étude actuelle montre des gains en temps et en mémoire.

L'étude de l'ordering des inconnues, actuellement réalisée avec la *dissection emboîtée* au travers du partitionneur `Scotch`, pourrait permettre d'exhiber des blocs plus compressibles. Les techniques de partitionnement actuelles ont pour objectif de réduire le remplissage tout en garantissant un bon niveau de parallélisme. En les couplant avec des critères sur la compression (i.e., des distances dans le graphe), il est imaginable de fournir une meilleure numérotation des inconnues dans un contexte de compression *low-rank* [4].

Finalement, l'ensemble de ces travaux vise à répondre au défi majeur de concevoir et réaliser des solveurs robustes numériquement sur des supports d'exécution capables de passer à l'échelle et de repousser les limites des codes industriels existants en utilisant pleinement l'ensemble des ressources calculatoires hétérogènes et hiérarchiques.

J'ai eu l'occasion d'encadrer le travail de très nombreux étudiants dans le cadre de mes travaux de recherche et de développement logiciel :

- Ingénieurs :
 - T. Delarue de 2019 à 2021, dans le cadre du projet EoCoE II, sur une ré-implémentation du parallélisme distribué (MPI) dans la nouvelle version du solveur PaStiX (release 6.1 <https://gitlab.inria.fr/solverstack/pastix/-/releases>).
 - M. Hastaran, en 2016, dans le cadre du projet EoCoE, sur la mise en place d'un ensemble de tests de non-régression et d'outils de post-analyse pour les campagnes de tests sur des calculateurs parallèles.
 - T. Terraz, en 2014, dans le cadre du projet Fortissimo, sur la redéfinition de l'API utilisateur du solveur PaStiX afin d'être compatible avec un fonctionnement en mode SAAS.
 - X. Lacoste, en 2008, sur l'implémentation d'une fonctionnalité de type Out-of-Core.
 - D. Lecas, en 2005, sur l'implémentation des factorisations avec pivotage statique.
- Post-doctorant :
 - H. Sellama, en 2009, sur l'implémentation des étapes de raffinement et de dé-raffinement de maillage des éléments finis de type Bézier dans le code JOREK.
- Etudiants en dernière année de master :
 - O. Zenati, en 2012 dans le cadre d'une collaboration avec ICL, sur l'expression du parallélisme via des supports d'exécution génériques, pour l'algorithme de la factorisation LU dense avec pivotage numérique.
 - N. Schied, en 2006, sur la parallélisation d'une méthode multigrille.
- Etudiants en dernière année d'ingénieur :
 - N. Bredel, en 2021, sur l'intégration du support low-rank dans le mode mémoire distribué de PaStiX.
 - T. Moenne-Loccoz, en 2021, sur l'intégration de l'ordonnanceur HeteroPrio pour améliorer les performances de PaStiX avec StarPU.
 - C. Soyez-Martin, en 2018, sur une analyse des noyaux de compression utilisant les techniques de randomisation.
 - V. Bridonneau, en 2018, sur l'implémentation parallèle des étapes du raffinement itératif, puis dans le cadre d'une collaboration avec l'ICL en 2019.
 - H. Vikstrom en 2016, sur l'implémentation d'un nouvel algorithme pour la recherche des valeurs propres.
 - A. Bellot, en 2016, sur la gestion des degrés de liberté variables dans l'API utilisateur de PaStiX.
 - G. Pichon, en 2014, sur une implémentation portable des méthodes de raffinement itératif.
 - A. Khabou, en 2009, sur des prétraitements numériques et symboliques pour résoudre des systèmes non symétriques.
 - D. Genet, en 2008, sur la parallélisation de l'étape de factorisation symbolique.
 - G. Caramel, en 2005, sur la parallélisation d'un code de mécanique des fluides avec le CEA CESTA, puis dans le cadre d'un contrat avec EDF.
 - F. Huard, en 2005, sur l'intégration du solveur PaStiX dans le code ODYSSEE développé au CEA CESTA.
 - N. Frezier, en 2005, sur l'intégration du solveur PaStiX dans la plateforme du projet GRID TLSE.

Contribution 1 : PaStiX: un solveur de référence en algèbre linéaire creuse

1. Description de la contribution

Le solveur PaStiX que je développe peut être vu comme le démonstrateur de mon expertise dans l'exploitation optimale des différentes architectures parallèles introduites lors des 20 dernières années. Il est utilisé en production pour des codes de simulations développés par la DAM (Direction des Applications Militaires).

Les thèses de Mathieu Faverge, Xavier Lacoste et Grégoire Pichon se sont appuyées directement sur ce solveur pour démontrer l'impact des contributions proposées. L'ensemble des projets ANR (NUMASIS, SOLSTICE, ANEMOS/ASTER, PETAL/PETALH, SOLHAR/SOLHARIS et SASHIMI) auxquels j'ai participé étaient étroitement associés aux travaux développés autour de ce solveur.

2. Contribution personnelle de la candidate / du candidat

Je suis l'architecte et j'assure le développement ainsi que le suivi de la bibliothèque PaStiX [130] qui implémente un solveur haute performance pour la résolution de grands systèmes linéaires basé sur une approche directe supernodale.

Un site web donnant l'accès au téléchargement et à la documentation de l'interface est disponible sur : <https://gitlab.inria.fr/solverstack/pastix>. Ce logiciel est maintenant placé sous licence libre CeCILL-C et la première distribution publique date de septembre 2006. Cette distribution, qui représente actuellement plus de 120000 lignes de code, est constamment mise à jour pour tenir compte des améliorations algorithmiques apportées par mes travaux, ce processus étant matérialisé par la mise à disposition de la communauté de versions successives avec 2 révisions majeures par an. La distribution logicielle PaStiX est très largement diffusée et utilisée, et a fait l'objet de plus de 60000 téléchargements (dans le TOP20 des logiciels développés sur la forge Inria).

Ce solveur est reconnu par la communauté pour être un des plus performants sur les architectures récentes constituées de nœuds multicœurs. Très récemment, nous avons pu proposer la première implémentation d'un solveur creux direct sur des architectures hétérogènes disposant d'accélérateurs de type GPU.

3. Originalité et difficulté

Tous ces travaux seront menés dans un contexte de concurrence forte avec des équipes internationales renommées avec qui je développe des échanges réguliers, entre autres :

- MUMPS: Patrick Amestoy et Jean-Yves l'Excellent, startup MUMPS, FR
- SuperLU: Sherry Li, Lawrence Berkeley National Laboratory, USA
- STRUMPACK: Pieter Ghysels, Lawrence Berkeley National Laboratory, USA
- SuiteSparse (Matlab): Tim Davis, Texas university, USA
- TAUCS (Mathematica): Sivan Toledo, Tel-Aviv university, IL
- WSMP (IBM): Anshul Gupta, IBM Watson, USA
- PARDISO (INTEL): Olaf Schenk, Lugano university, CH
- HSL (NAG): Jennifer Scott, Rutherford Appleton, UK

4. Validation et impact

L'ensemble de ces travaux vise à répondre au défi majeur de concevoir et réaliser des solveurs robustes numériquement sur des supports d'exécution capables de passer à l'échelle et de repousser les limites des codes industriels existants en utilisant pleinement l'ensemble des ressources calculatoires telles que les CPUs, les GPUs et autres accélérateurs. Cette activité de recherche s'inscrit naturellement dans la volonté de faire converger le calcul intensif et le traitement des données massives. L'algèbre linéaire étant au cœur des techniques d'apprentissage automatique, c'est un élément clé pour les applications mises en avant dans de nombreux domaines cibles de l'IA et du BigData.

L'objectif de produire un logiciel visant une audience assez large nous a conduit à respecter les bonnes pratiques et standards modernes de développement : une gestion de version sous Gitlab, des "dashboards" (SonarQube, CDash, Coverity), une intégration continue (Gitlab-runner avec plus de 6000 tests unitaires/fonctionnalités), une gestion des contributions et demandes des utilisateurs (à travers les workflow "Issues" et "Merge Request"), un processus de livraison semi-automatique et une génération automatique de la documentation.

5. Diffusion

Ce solveur est intégré dans plusieurs autres logiciels, qui contribuent à sa diffusion indirecte, par exemple :

- PETSc (<http://www.mcs.anl.gov/petsc>) : une bibliothèque de fonctions en C permettant de gérer des vecteurs et des matrices creuses et de résoudre les systèmes linéaires correspondants avec des solveurs directs ou itératifs ;
- Eigen (<http://eigen.tuxfamily.org>) : une bibliothèque template en C++ d'algèbre linéaire.

Enfin, parmi les principaux utilisateurs, nous pouvons citer ceux avec qui nous entretenons des collaborations scientifiques régulières :

- CEA CESTA : pour des codes en production de mécanique des structures et d'électromagnétisme ;
- CEA CADARACHE : pour un code de Magnéto-Hydro-Dynamique dont le but est de contribuer au dimensionnement et au contrôle du réacteur ITER ;
- CERFACS : pour un code de simulation de l'évolution du climat avec Météo-France ;
- Algo'Tech : dans le cadre de l'initiative HPC PME, la société Algo'Tech utilise notre solveur dans ses codes de simulation de systèmes électriques et électromagnétiques pour ses clients issus aussi bien de l'aéronautique que de l'automobile.

Notre solveur est également intégré dans plusieurs frameworks comme Trilinos (<https://trilinos.github.io/>) ou Kratos (<https://www.cimne.com/kratos/>). Il est également disponible pour la communauté des utilisateurs de GNU Octave (<https://www.gnu.org/software/octave/>), avec qui nous avons des contacts réguliers pour maintenir une version pour l'OS Windows.

Une trace d'exécution de PASTIX est par exemple utilisée pour illustrer le HPC par Inria : <https://www.inria.fr/fr/calcul-haute-performance>.

Contribution 2 : Collaboration dans le domaine du HPC pour le CEA CESTA et CADARACHE

1. Description de la contribution

Il s'agit d'une action de valorisation de mes travaux de recherche et de transfert de compétence dans des codes de simulation pour des applications en électromagnétisme et en physique des plasmas.

2. Contribution personnelle de la candidate / du candidat

Contrat de valorisation avec le CEA CESTA : Dans le cadre de mes travaux de recherche, j'ai participé à des contrats [123, 121, 120, 119] avec obligation de résultat passés avec le CEA CESTA. Il s'agissait d'études d'applicabilité de nos méthodes aux problèmes étudiés par l'équipe "Analyse Numérique" du CESTA et de transférer nos développements logiciels. Le solveur PaStiX est depuis utilisé en production pour des codes de simulations d'électromagnétisme développés au CESTA. J'ai également travaillé sur la parallélisation du code MIRO, un code qui simule la propagation d'un faisceau laser dans une chaîne de composants optiques pour le laser Mégajoule [122].

Un autre type de couplage mixant des méthodes directes et multigrilles a été regardé dans les travaux de thèse de M. Chanaud. Dans un souci de robustesse, une idée originale est de conserver un solveur direct parallèle pour effectuer les résolutions sur le niveau grossier de la méthode multigrille. L'application d'une méthode directe sur le maillage grossier constitue le point d'entrée de l'algorithme multigrille et permet également de distribuer le système. En particulier, la distribution des données issues du solveur direct parallèle PaStiX est utilisée pour piloter la définition et le traitement des niveaux plus fins. Cette étude réalisée pour la résolution des équations de Maxwell tridimensionnelles mérite d'être poursuivie et étendue à d'autres applications du CESTA.

Transfert industriel avec le CEA CADARACHE : Depuis 2006, j'ai assuré le transfert et le suivi du solveur PaStiX dans le code de simulation JOEREK qui a pour objectif d'améliorer la compréhension des instabilités au bord des plasmas, en particulier pour le contrôle du réacteur ITER utilisant le principe de la fusion thermonucléaire. Cette collaboration nous a conduits à améliorer l'efficacité de l'interface proposée entre le solveur et le code de simulation, tant du point de vue des performances que de la consommation mémoire. Suite à des séminaires au CEMRACS, j'ai également initié une collaboration avec Patrick Tamain, qui développe le code TOKAM3X et utilise le solveur PaStiX pour la résolution de systèmes linéaires. Ce travail s'est traduit par une action commune dans le projet EoCoE II.

J'ai encadré le postdoctorat d'Hocince Sellama financé sur le projet ANR ASTER. Le travail consistait à intégrer une procédure de raffinement dans le code JOEREK, tout d'abord au début de la simulation (pour l'équilibre) puis en apportant des modifications pour obtenir un raffinement adaptatif (au cours de la simulation). Les étapes de raffinement et de dé-raffinement des éléments finis de type Bézier ont été implémentées pour le calcul à l'équilibre. Le raffinement adaptatif (au cours de la simulation) est également disponible et les premières simulations illustrant l'intérêt de l'approche ont été menées comme par exemple l'injection de "glaçons" pour le contrôle de la densité du plasma. Enfin, le traitement particulier de la région du point-X dans géométrie des Tokamaks était une difficulté qui a été prise en compte lors du raffinement.

3. Originalité et difficulté

4. Validation et impact

J'ai pu valoriser mes compétences dans le domaine des techniques de parallélisation dans le cadre de plusieurs contrats d'expertise, et depuis septembre 2015, je suis conseiller scientifique au CEA CESTA avec une expertise pour le domaine du HPC. *Les missions qui me sont confiées dans le cadre de cette expertise sont directement liées à mon projet de recherche développé dans la section suivante.*

À noter qu'une étude réalisée par des collègues du CEA CADARACHE a été menée en 2019 pour analyser l'impact des techniques de compression low-rank sur le code JOEREK. Cette étude intègre en particulier une comparaison avec le solveur concurrent MUMPS (arXiv:1907.13442).

5. Diffusion

L'article [7] qui présente la solution retenue pour préconditionner le système linéaire dans JOEREK est l'un de mes articles les plus cités (>100 d'après Scopus). Nous avons publié un article de vulgarisation dans la revue CHOCS [82] en 2012. Avec l'équipe en charge du développement du code ARLENE, j'ai participé à la rédaction d'un article [14] présentant la mise en œuvre d'un support d'exécution permettant la gestion des dépendances implicites dans le cadre de l'utilisation de formats hiérarchiques et pour des modèles de programmation en mémoire partagée et distribuée.

Contribution 3 : Collaboration académique avec l'Université de Stanford

1. Description de la contribution

Depuis 2013, dans le cadre de l'équipe associée FAST-LA, j'ai mis en place une collaboration avec l'équipe d'Eric Darve, de l'Université de Stanford, sur l'application des techniques de compression utilisant des matrices hiérarchiques pour l'accélération et la réduction de la complexité des solveurs directs.

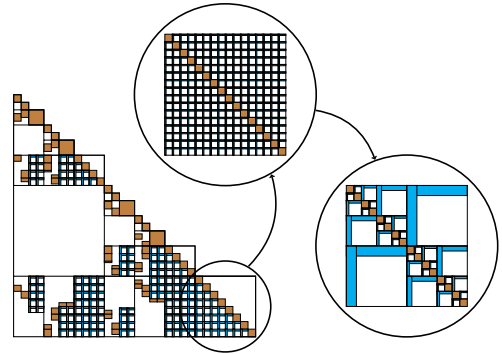
2. Contribution personnelle de la candidate / du candidat

De la communauté des méthodes intégrales ou méthodes des éléments de frontière a émergé, ces dernières années, des techniques de compression pour matrices denses globalement regroupées sous le vocable de \mathcal{H} -matrices. Ces techniques reposent sur une représentation "data sparse" des opérateurs linéaires et exploitent l'information géométrique du maillage sous-jacent pour définir le "clustering", i.e., la structure hiérarchique de la matrice dense.

Je propose donc de prolonger ces idées dans le cadre purement algébrique des méthodes de factorisation de matrices creuses en substituant le calcul matriciel dense "classique" par du calcul \mathcal{H} -matriciel afin de réduire l'empreinte mémoire des solveurs directs creux qui constitue le principal frein au passage à l'échelle de ces approches pour des simulations 3D. J'étudie l'impact numérique des techniques de compression exploitant l'arithmétique des matrices hiérarchiques (\mathcal{H} -matrices) dans les blocs denses qui apparaissent lors de la factorisation numérique. L'utilisation des familles de matrices hiérarchiques repose sur une approximation de rang faible des blocs denses dont la structure est définie récursivement.

Il est important de conserver un aspect algébrique pour cibler un spectre d'applications le plus large possible. Les formats hiérarchiques représentent des alternatives prometteuses en comparaison des formats à plat BLR (Bloc Low-Rank) en termes de complexité et de consommation mémoire.

Dans la thèse d'Eragul Korkmaz, on considère maintenant un format de compression de type \mathcal{H} , et plus particulièrement le format HODLR [?] introduit par E. Darve. Ce format utilisant des bases non imbriquées, il est plus facile de développer des noyaux efficaces basés sur des opérations BLAS 3.



3. Originalité et difficulté

Dans la thèse de Grégoire Pichon, pour une technique de compression BLR, nous avons montré que pour un ensemble de matrices issues de problèmes réels, nous pouvons réduire l'empreinte mémoire jusqu'à un facteur 4 en utilisant une stratégie qui utilise la compression de données dès la phase d'assemblage, et une accélération du temps de calcul pouvant atteindre un facteur 3.5 avec la stratégie qui minimise le nombre d'opérations.

4. Validation et impact

Une bourse Chateaubriand a été accordée en 2014 à Clif Robert Dudley, étudiant de master à Stanford.

J'ai été invité par l'Université de Stanford pour participer au séminaire du département "Mechanical Engineering" (ME395) en avril 2018. Durant cette collaboration, j'ai effectué 3 visites de 10 jours dans l'équipe d'Eric Darve pour rencontrer ses étudiants en thèse, en particulier Leopold Cambier avec qui nous avons échangé sur de nouvelles pistes pour réduire la complexité des solveurs directs.

5. Diffusion

Les travaux de thèse de Grégoire ont donné lieu à un très grand nombre de présentations en conférence et à 2 articles en revue SIAM SIMAX [4] en 2017 et Elsevier JOCS [3] en 2018. Nous avons également préparé un article [13] portant sur l'optimisation des algorithmes de « clustering ».

A l'issue de sa thèse DGA, Grégoire a obtenu un poste de maître de conférences à l'Université de Lyon 1 avec un rattachement au LIP (ENS-Lyon) dans l'équipe Inria ROMA.

Contribution 4 : Collaboration industrielle avec EDF

1. Description de la contribution

Les calculs de réactivité constituent une brique fondamentale dans la simulation des cœurs des réacteurs nucléaires. Ceux-ci conduisent à la résolution de problèmes aux valeurs propres généralisées via l'algorithme de la puissance inverse. A chaque itération, on est amené à résoudre un système linéaire de manière approchée via un algorithme d'itérations imbriquées. 2 thèses CIFRE ont permis d'améliorer le parallélisme des codes de simulation de l'état de l'art en neutronique. La première contribution correspond à une méthode adhoc de décomposition de domaine, et la seconde contribution porte sur l'utilisation des supports d'exécution pour optimiser un stencil 3D.

2. Contribution personnelle de la candidate / du candidat

En 2006, j'ai encadré le travail de Guilhem Caramel sur "l'optimisation des performances des outils de calcul de neutronique des cœurs" dans le cadre d'un contrat [118] passé avec EDF. Dans l'optique de trouver un algorithme parallélisable permettant de remplacer la méthode Gauss-Seidel pour la résolution des systèmes linéaires dans la boucle de la puissance inverse, nous avons étudié le comportement numérique de la méthode itérative GMRES dans le code de neutronique, suivant le nombre de groupes d'énergie.

Cette étude s'est prolongée dans le cadre de la thèse de Bruno Lathuillère [128] et du contrat de collaboration passé avec EDF-R&D. J'ai co-encadré ce travail portant sur la parallélisation d'une méthode de décomposition de domaine des équations la neutronique pour les études d'accident. Au cours de cette thèse, nous avons étudié une méthode de décomposition de domaine de type Schur dual. Plusieurs placements de l'algorithme de décomposition de domaine au sein du système d'itérations imbriquées ont été envisagés. Deux d'entre eux ont été implémentés et les résultats analysés. Le deuxième placement, utilisant les spécificités des éléments finis de Raviart-Thomas et de l'algorithme des directions alternées, a conduit à des résultats très encourageants permettant l'industrialisation de la méthodologie associée.

Dans le cadre d'un contrat de collaboration passé avec EDF-R&D, l'objectif de la thèse de Salli Moustafa [125] était la mise au point d'une méthode de résolution massivement parallèle de l'équation de transport neutronique. La réduction du temps de retour à moins d'une heure, pour des calculs impliquant plus de 10 milliards d'inconnues, a représenté une avancée importante pour la simulation des cœurs des centrales nucléaires. Pour y parvenir, il était nécessaire d'utiliser les techniques de parallélisation les plus avancées impliquant plusieurs algorithmes parallèles imbriqués ainsi que plusieurs paradigmes de programmations parallèles (passage de message, multi-threading et instructions SIMD). Une stratégie algorithmique adaptée au cas particulier des systèmes diffusifs allié à un très haut niveau de parallélisme a conduit à la réalisation d'un outil unique pour la simulation des centrales nucléaires.

3. Originalité et difficulté

Les calculs de transport par la méthode SN en neutronique forment un domaine de recherche où il existe une compétition internationale importante. Ces codes sont très consommateurs en temps de calcul et, depuis l'origine, ils sont utilisés comme benchmark pour évaluer les architectures HPC dont ils approchent les performances crêtes.

4. Validation et impact

La combinaison des développées dans le cadre de ces 2 thèses CIFRE nous a permis de concevoir une version massivement parallèle du solveur neutronique DOMINO. Les performances du solveur atteignent 33.9% de la performance crête théorique d'un supercalculateur composé de 768 cœurs. De plus, un calcul critique d'un réacteur de type REP 900MW à 26 groupes d'énergie mettant en jeu 10^{12} inconnues a été résolu en 46 minutes sur 1536 cœurs.

5. Diffusion

Ces travaux ont conduit à plusieurs publications dans des revues et conférences très sélectives.

Les résultats de Bruno ont été présentés lors des conférences CSE'08 [23], IMACS'08 [71], PMAA'08 [70] et *Reactor Physics* [67]. Un article a été publié dans la revue *Journal of Computational Physics* [6].

Les résultats de Salli ont été présentés lors de la conférence IPDPS'15 [19]. Deux articles ont été publiés dans les revues *Annals of Nuclear Energy* [5] et *Journal of Computational Physics* [2].

A l'issue de leur thèse CIFRE, Bruno Lathuillère a obtenu un poste chez EDF et Salli Moustafa est consultant chez ANEO.

Contribution 5 : Transfert industriel avec l'entreprise AlgoTech

1. Description de la contribution

Basée à Bidart (Pyrénées-Atlantiques), l'entreprise AlgoTech, qui emploie une quinzaine de personnes, souhaitait proposer à ses clients issus aussi bien de l'aéronautique que de l'automobile une solution pouvant simuler dès les phases de conception l'impact des phénomènes électromagnétiques sur les systèmes de câblage électrique, notamment de très grande taille et offrir ces services de simulation en mode Cloud.

Suite à une prise de contact dans le cadre de l'initiative HPC-PME (BPI, Maison de la simulation, GENCI), j'ai mis en place un contrat d'expertise avec le gérant de la société AlgoTech, Jacques Péré-Laperne.

2. Contribution personnelle de la candidate / du candidat

La PME a bénéficié d'un transfert de compétences dans le cadre de travaux sur le modèle du thésard-conseil réalisé en 2013 par mon doctorant Xavier Lacoste [116].

Suite à cette expertise, nous avons obtenu des supports européens avec les appels FORTISSIMO en 2014 et PRACE-4IP en 2015. J'ai notamment encadré le travail de Théophile Terraz recruté sur un contrat de 12 mois pour redéfinir l'API utilisateur du solveur afin d'être compatible avec un fonctionnement en mode SAAS (Software as a Service).

L'utilisation du solveur PaStiX a ainsi permis de réaliser le saut technologique nécessaire au développement d'une version du logiciel Simul'Elec adaptée au HPC.

3. Originalité et difficulté

Nous avons dû nous écarter des plateformes habituelles pour le calcul intensif et revoir le parallélisme pour une utilisation efficace du solveur sur un cluster de calcul administré par Bull/Atos en mode Cloud. La solution retenue consiste à déporter sur le cluster la partie la plus consommatrice en ressources CPU, ici la boucle en fréquences pour la résolution des second-membres multiples.

Les solutions logicielles d'AlgoTech étant développées dans le langage Delphi, nous avons été amenés à effectuer le premier portage du solveur PaStiX sous l'environnement Windows.

4. Validation et impact

Dans le cadre d'un accompagnement de l'Initiative HPC-PME (à présent SiMSEO), la société AlgoTech, spécialisée dans l'édition de logiciels de CAO électrique, a réalisé avec succès le passage au calcul intensif (HPC) de son logiciel de simulation.

Dernière étape de son projet, AlgoTech a été sélectionnée par le projet européen Fortissimo pour déployer son offre logicielle sur une plateforme Cloud commerciale. Elle est ainsi complètement intégrée et visible dans l'écosystème du calcul intensif européen.

5. Diffusion

Ce travail s'est traduit par une action communication relativement importante dans le cadre de l'initiative HPC-PME. Je peux citer en particulier l'article sur le site du GENCI "AlgoTech se câble au calcul intensif" <http://www.genci.fr/sites/default/files/Communiqu%C3%A9AlgoTech-def.pdf>.

Enfin, j'ai participé à une table ronde lors des rencontres Inria Industrie en 2013 pour témoigner sur l'expérience du dispositif doctorant-conseil et sur le transfert industriel avec une PME.

Xavier Lacoste est maintenant ingénieur chez Total et Théophile Terras est ingénieur dans l'équipe de Bruno Raffin, IPL HPC-BigData.

1. Publications caractéristiques

- [1] S. Moustafa, F. Févotte, M. Faverge, L. Plagne, P. Ramet. *Efficient Parallel Solution of the 3D Stationary Boltzmann Transport Equation for Diffusive Problems*. Journal of Computational Physics, 388(1):335-349, 2019. <https://hal.inria.fr/hal-02080624>
- [2] G. Pichon, E. Darve, M. Faverge, P. Ramet, and J. Roman. *Sparse Supernodal Solver Using Block Low-Rank Compression: design, performance and analysis*. International Journal of Computational Science and Engineering, 27:255-270, 2018. <https://hal.inria.fr/hal-01824275>
- [3] G. Pichon, M. Faverge, P. Ramet, and J. Roman. *Reordering strategy for blocking optimization in sparse linear solvers*. SIAM Journal on Matrix Analysis and Application, 38(1):226-248, 2017. <https://hal.inria.fr/hal-01485507v2>
- [4] G. Huysmans, Pamela S., E. van der Plas et P. Ramet. *Non-Linear MHD simulations of Edge Localised Modes (ELMs)*. Journal on Plasma Physics and Controlled Fusion, 51(12):124012, 2009. <https://hal.inria.fr/hal-02080624>
- [5] P. Hénon, P. Ramet, et J. Roman. *PaStiX: A High-Performance Parallel Direct Solver for Sparse Symmetric Definite Systems*. Parallel Computing, 28(2):301-321, 2002. <https://hal.inria.fr/inria-00346017>

2. Liste des publications

Avant 2012, l'ordre alphabétique a été utilisé pour l'apparition des auteurs dans les publications ci-dessous et correspondait aux pratiques de notre communauté.

Après 2012, j'ai pris l'habitude de conserver l'ordre alphabétique sauf pour le premier auteur, généralement le doctorant, lorsque son travail correspond à la contribution principale.

Voir <https://cv.archives-ouvertes.fr/pierre-ramet> pour une liste complète de mes publications et rapports de recherches.

2.1 Revues internationales

- [1] M. Hoelzl, G. Huysmans, S. Pamela, M. Bécoulet, et al.. *The JOREK non-linear extended MHD code and applications to large-scale instabilities and their control in magnetically confined fusion plasmas*. Journal of Nuclear Fusion, 61(6) 065001, 2021.
- [2] S. Moustafa, F. Févotte, M. Faverge, L. Plagne, P. Ramet. *Efficient Parallel Solution of the 3D Stationary Boltzmann Transport Equation for Diffusive Problems*. Journal of Computational Physics, 388(1):335-349, 2019.
- [3] G. Pichon, E. Darve, M. Faverge, P. Ramet, and J. Roman. *Sparse Supernodal Solver Using Block Low-Rank Compression: design, performance and analysis*. International Journal of Computational Science and Engineering, 27:255-270, 2018.
- [4] G. Pichon, M. Faverge, P. Ramet, and J. Roman. *Reordering strategy for blocking optimization in sparse linear solvers*. SIAM Journal on Matrix Analysis and Application, 38(1):226-248, 2017.
- [5] S. Moustafa, I. Dutka-Malen, L. Plagne, A. Poncot, and P. Ramet. *Shared Memory Parallelism for 3D Cartesian Discrete Ordinates Solver*. Annals of Nuclear Energy, 82:179-187, 2015.
- [6] M. Barrault, B. Lathuillère, P. Ramet et J. Roman. *Efficient Parallel Resolution of The Simplified Transport Equations in Mixed-Dual Formulation*. Journal of Computational Physics, 230(5):2004-2020, 2011.
- [7] G. Huysmans, Pamela S., E. van der Plas et P. Ramet. *Non-Linear MHD simulations of Edge Localised Modes (ELMs)*. Journal on Plasma Physics and Controlled Fusion, 51(12):124012, 2009.
- [8] R. Abgrall, R. Huart et P. Ramet. *Numerical simulation of unsteady MHD flows and applications*. MagnetoHydroDynamics Journal, 45(2):225-232, 2009.
- [9] P. Hénon, P. Ramet et J. Roman. *On finding approximate supernodes for an efficient ILU(k) factorization*. Parallel Computing, 34:345-362, 2008.
- [10] P. Hénon, P. Ramet, et J. Roman. *PaStiX: A High-Performance Parallel Direct Solver for Sparse Symmetric Definite Systems*. Parallel Computing, 28(2):301-321, 2002.

- [11] E. Caron, S. Chaumette, S. Contassot-Vivier, F. Desprez, E. Fleury, C. Gomez, M. Goursat, E. Jeannot, D. Lazure, F. Lombard, J.M. Nicod, L. Philippe, M. Quinson, P. Ramet, J. Roman, F. Rubi, S. Steer, F. Suter et G. Utard. *Scilab to Scilab//, the OURAGAN Project*. Parallel Computing, 27(11):1497-1519, 2001.
- [12] D. Goudin, P. Hénon, F. Pellegrini, P. Ramet, J. Roman et J.-J. Pesqué. *Parallel Sparse Linear Algebra and Application to Structural Mechanics*. Numerical Algorithms, 24:371-391, 2000.

2.2 Autres rapports de recherche et articles soumis

- [13] Pichon, G. and Darve, E. and Faverge, M. and Ramet, P. and Roman, J. *Supernodes ordering to enhance Block Low-Rank compression in sparse direct solvers*. Research Report 9238, <https://hal.inria.fr/hal-01961675>.
- [14] Augonnet, Cédric and Goudin, David and Kuhn, Matthieu and Lacoste, Xavier and Namyst, Raymond and Ramet, Pierre. *A hierarchical fast direct solver for distributed memory machines with manycore nodes*. Research Report, <https://hal-cea.archives-ouvertes.fr/cea-02304706>.
- [15] Sellama, Hocine and Huijsmans, Guido and Ramet, Pierre. *Adaptive mesh refinement for numerical simulation of MHD instabilities in tokamaks: JOEREK code*. Research Report 8635, <https://hal.inria.fr/hal-01088094>.

2.3 Conférence internationales avec comité de lecture

- [16] Korkmaz, E. and Faverge, M. and Pichon, G. and Ramet, P. *Deciding Non-Compressible Blocks in Sparse Direct Solvers using Incomplete Factorization*. HiPC 2021, 28th IEEE International Conference on High Performance Computing, Data, and Analytics, pages 1-10, Bangalore, India, décembre 2021.
- [17] C. Gou, A. Al Zoobi., A. Benoit, M. Faverge, L. Marchal, G. Pichon., P. Ramet. *Improving mapping for sparse direct solvers: A trade-off between data locality and load balancing*. EuroPar 2020 - 26th International European Conference on Parallel and Distributed Computing, Warsaw, Poland, pages 1-16, aout 2020.
- [18] G. Pichon, E. Darve, M. Faverge, P. Ramet, and J. Roman. *Sparse Supernodal Solver Using Block Low-Rank Compression*. 18th IEEE International Workshop on Parallel and Distributed Scientific and Engineering Computing (PDSEC 2017), Orlando, USA, juin 2017.
- [19] S. Moustafa, M. Faverge, L. Plagne, P. Ramet. *3D Cartesian Transport Sweep for Massively Parallel Architectures with PARSEC*. 29th IEEE International Parallel & Distributed Processing Symposium (IPDPS'15), pages 581-590, Hyderabad, India, mai 2015.
- [20] A. Casadei, P. Ramet, and J. Roman. *An improved recursive graph bipartitioning algorithm for well balanced domain decomposition*. 21st IEEE International Conference on High Performance Computing, pages 1-10, Goa, India, décembre 2014.
- [21] X. Lacoste, M. Faverge, P. Ramet, S. Thibault, and G. Bosilca. *Taking advantage of hybrid systems for sparse direct solvers via task-based runtimes*. HCW'2014 workshop of IPDPS, pages 29-38, Phoenix, USA, mai 2014.
- [22] G. Huysmans, Pamela S., E. van der Plas et P. Ramet. *Non-Linear MHD simulations of Edge Localised Modes*. 36th EPS Plasma Physics Conference, Sofia, Bulgarie, juin 2009.
- [23] M. Barrault, B. Lathuillère, P. Ramet et J. Roman. *A domain decomposition method applied to the simplified transport equations*. IEEE 11th International Conference on Computational Science and Engineering, Sao Paulo, Brazil, pages 91-97, juillet 2008.
- [24] Y. Caniou, J.-S. Gay et P. Ramet. *Tunable parallel experiments in a GridRPC framework: application to linear solvers*. VECPAR'08, LNCS 5336, pages 430-436, Toulouse, France, juin 2008.
- [25] N. Kushida, Y. Suzuki, N. Teshima, N. Nakajima, Y. Caniou, M. Dayde et P. Ramet. *Toward an International Sparse Linear Algebra Expert System by Interconnecting the ITBL Computational Grid with the Grid-TLSE Platform*. VECPAR'08, LNCS 5336, pages 424-429, Toulouse, France, juin 2008.
- [26] M. Faverge et P. Ramet. *Dynamic Scheduling for sparse direct Solver on NUMA architectures*. Proceedings of PARA'08, Trondheim, Norway, à paraître dans LNCS, mai 2008.
- [27] P. Hénon, P. Ramet et J. Roman. *Partitioning and Blocking Issues for a Parallel Incomplete Factorization*. PARA'06, Workshop on state-of-the-art in scientific computing, Umea, Suède, LNCS 4699, pages 929-937, juin 2006.
- [28] P. Hénon, P. Ramet et J. Roman. *On using an hybrid MPI-Thread programming for the implementation of a parallel sparse direct solver on a network of SMP nodes*. Sixth International Conference on Parallel Processing and Applied Mathematics, Workshop HPC Linear Algebra, Poznan, Pologne, LNCS 3911, pages 1050-1057, september 2005.

- [29] P. Hénon, F. Pellegrini, P. Ramet, J. Roman, et Y. Saad. *Applying parallel direct solver skills to build robust and highly performant preconditioners*. PARA'04, Workshop on state-of-the-art in scientific computing, Copenhagen, Danemark, LNCS 3732, pages 601-619, juin 2004.
- [30] O. Beaumont, P. Ramet et J. Roman. *Asymptotically optimal algorithm for Laplace task graphs on heterogeneous platforms*. Fifth International Conference on Parallel Processing and Applied Mathematics (PPAM), Czesochowa, Pologne, LNCS 3019, pages 880-887, septembre 2003.
- [31] P. Hénon, P. Ramet et J. Roman. *Efficient algorithms for direct resolution of large sparse system on clusters of SMP nodes*. SIAM Conference LA'2003, Williamsburg, USA, juillet 2003.
- [32] P. Hénon, P. Ramet et J. Roman. *PaStiX: A Parallel Direct Solver for Sparse SPD Matrices based on Efficient Static Scheduling and Memory Management*. SIAM Conference PPSC'2001, Portsmouth, Virginie, USA, mars 2001.
- [33] P. Hénon, P. Ramet et J. Roman. *PaStiX: A Parallel Sparse Direct Solver Based on a Static Scheduling for Mixed 1D/2D Block Distributions*. IPDPS'2000, Cancun, Mexique, LNCS 1800, pages 519-525, mai 2000.
- [34] P. Hénon, P. Ramet et J. Roman. *A Mapping and Scheduling Algorithm for Parallel Sparse Fan-In Numerical Factorization*. EuroPar'99, Toulouse, France, LNCS 1685, pages 1059-1067, septembre 1999.
- [35] F. Desprez, P. Ramet et J. Roman. *Optimal Grain Size Computation for Pipelined Algorithms*. EuroPar'96, Lyon, France, LNCS 1123, pages 165-172, septembre 1996.

2.4 Livres et chapitres de livre

- [36] O. Coulaud, L. Giraud, P. Ramet, and X. Vasseur. *Developments in Parallel, Distributed, Grid and Cloud Computing for Engineering*. Chapter Augmentation and Deflation in Krylov subspace methods, pages 249-275. Saxe-Coburg Publications, Kippen, Stirlingshire, United Kingdom, 2013.

2.5 Autres publications internationales (posters, articles courts)

- [37] P. Ramet et al. *Study of the recent developments around the PaStiX solver for the EoCoE project: distributed memory, runtime systems, and low-rank* 32nd International Conference on Parallel Computational Fluid Dynamics, Nice, France, mai 2021.
- [38] M. Faverge, E. Korkmaz, G. Pichon, and P. Ramet. *Recent Developments Around the Block Low-Rank PaStiX Solver* SIAM Conference on Parallel Processing for Scientific Computing, Seattle, USA, février 2020.
- [39] M. Faverge, G. Pichon, P. Ramet, and J. Roman. *Exploiting Parameterized Task-graph in Sparse Direct Solvers* SIAM Conference on Computation Science and Engineering, Spokane, USA, février 2019.
- [40] G. Pichon, E. Darve, M. Faverge, P. Ramet, and J. Roman. *Block Low-rank Algebraic Clustering for Sparse Direct Solvers* SIAM Conference on Computation Science and Engineering, Spokane, USA, février 2019.
- [41] G. Pichon, E. Darve, M. Faverge, P. Ramet, and J. Roman. *Block Low-rank Algebraic Clustering for Sparse Direct Solvers* PMAA'2018, Zurich, Suisse, juin 2018.
- [42] G. Pichon, E. Darve, M. Faverge, P. Ramet, and J. Roman. SIAM Conference on Computation Science and Engineering, Atlanta, USA, février 2017.
- [43] G. Pichon, M. Faverge, and P. Ramet. *Exploiting Modern Manycore Architecture in Sparse Direct Solver with Runtime Systems*. SIAM Conference on Computation Science and Engineering, Atlanta, USA, février 2017.
- [44] G. Pichon, M. Faverge, P. Ramet, and J. Roman. *Impact of Blocking Strategies for Sparse Direct Solvers on Top of Generic Runtime*. SIAM Conference on Computation Science and Engineering, Atlanta, USA, février 2017.
- [45] E. Darve, M. Faverge, G. Pichon, P. Ramet, and J. Roman. *Sparse Supernodal Solver Using Hierarchical Compression*. Workshop on Fast Direct Solvers, Purdue, USA, novembre 2016.
- [46] P. Ramet *On the use of low rank approximations for sparse direct solvers*. SIAM Annual Meeting, Boston, USA, juillet 2016.
- [47] M. Faverge, G. Pichon, P. Ramet, and J. Roman. *Impact of Blocking Strategies for Sparse Direct Solvers on Top of Generic Runtimes*. SIAM Conference on Parallel Processing for Scientific Computing, Paris, France, avril 2016.
- [48] E. Darve, M. Faverge, G. Pichon, P. Ramet, and J. Roman. *Exploiting H-Matrices in Sparse Direct Solvers*. SIAM Conference on Parallel Processing for Scientific Computing, Paris, France, avril 2016.

- [49] M. Faverge, G. Pichon, P. Ramet, and J. Roman. *Blocking strategy optimizations for sparse direct linear solver on heterogeneous architectures*. Sparse Days, Saint Giron, France, juin 2015.
- [50] M. Faverge, G. Pichon, P. Ramet, and J. Roman. *On the use of H-Matrix Arithmetic in PaStiX: a Preliminary Study*. Workshop on Fast Solvers, Toulouse, France, juin 2015.
- [51] X. Lacoste, M. Faverge, and P. Ramet. *A task-based sparse direct solver suited for large scale hierarchical/heterogeneous architectures*. SIAM Conference on Computation Science and Engineering, Salt Lake City, USA, février 2015.
- [52] A. Casadei, P. Ramet, and J. Roman. *Towards a recursive graph bipartitioning algorithm for well balanced domain decomposition*. SIAM Conference on Computation Science and Engineering, Salt Lake City, USA, février 2015.
- [53] P. Ramet. *On the design of parallel linear solvers for large scale problems*. International Congress on Industrial and Applied Mathematics, Pekin, China, aout 2015.
- [54] A. Casadei and P. Ramet. *Towards a recursive graph bipartitioning algorithm for well balanced domain decomposition*. International Congress on Industrial and Applied Mathematics, Pekin, China, aout 2015.
- [55] S. Moustafa, M. Faverge, L. Plagne, and P. Ramet. *Parallel 3D Sweep Kernel with PARSEC*. 16th IEEE International Conference on High Performance and Communications, workshop on HPC-CFD in Energy/Transport Domains, Paris, France, aout 2014.
- [56] A. Casadei, P. Ramet, and J. Roman. *Nested Dissection with Balanced Halo*. SIAM Workshop on Combinatorial Scientific Computing, Lyon, France, juillet 2014.
- [57] E. Agullo, M. Faverge, L. Giraud, A. Guermouche, P. Ramet, and R. Roman. *Toward parallel scalable linear solvers suited for large scale hierarchical parallel platforms*. WCCM-ECCM-ECFD, Barcelona, Spain, juillet 2014.
- [58] S. Moustafa, I. Dutka-Malen, L. Plagne, A. Poncot, and P. Ramet. *Shared Memory Parallelism for 3D Cartesian Discrete Ordinates Solver..* Joint International Conference on Supercomputing in Nuclear Applications + Monte Carlo, Paris, France, octobre 2013.
- [59] X. Lacoste, M. Faverge, and P. Ramet. *Sparse Linear Algebra over DAG Runtimes*. SIAM Conference on Computation Science and Engineering, Boston, USA, février 2013.
- [60] A. Casadei, L. Giraud, P. Ramet, and J. Roman. *Towards Domain Decomposition with Balanced Halo*. Workshop Celebrating 40 Years of Nested Dissection, Waterloo, Canada, juillet 2013.
- [61] P. Ramet. *From hybrid architectures to hybrid solvers*. Workshop Celebrating 40 Years of Nested Dissection, Waterloo, Canada, juillet 2013.
- [62] X. Lacoste, P. Ramet, M. Faverge, I. Yamazaki, G. Bosilca. *Toward a supernodal sparse direct solver over DAG runtimes*. PMAA'2012, London, England, juin 2012.
- [63] A. Casadei et P. Ramet. *Memory Optimization to Build a Schur Complement*. SIAM Conference LA'2012, Valencia, Spain, juin 2012.
- [64] X. Lacoste et P. Ramet. *Sparse direct solver on top of large-scale multicore systems with GPU accelerators*. SIAM Conference LA'2012, Valencia, Spain, juin 2012.
- [65] M. Faverge et P. Ramet. *Fine Grain Scheduling for Sparse Solver on Manycore Architectures*. SIAM Conference PPSC'2012, Savannah, USA, février 2012.
- [66] Y. Suzuki, N. Kushida, T. Tatekawa, N. Teshima, Y. Caniou, R. Guivarch, M. Dayde et P. Ramet. *Development of an International Matrix-Solver Prediction System on a French-Japanese International Grid Computing Environment*. Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2010, Tokyo, Japan, octobre 2010.
- [67] M. Barrault, B. Lathuilière, P. Ramet et J. Roman. *A Non Overlapping Parallel Domain Decomposition Method Applied to The Simplified Transport Equations*. International Conference on Mathematics, Computational Methods and Reactor Physics, New-York, USA, mai 2009.
- [68] R. Abgrall, O. Coulaud, P. Hénon, R. Huart, G. Huysmans, G. Latu, B. Nkonga, S. Pamela et P. Ramet. *Numerical simulation of tokamak plasmas*. 7th PAMIR International Conference on Fundamental and Applied MHD, Presqu'île de Giens, France, septembre 2008.

- [69] M. Faverge, X. Lacoste et P. Ramet. *A NUMA Aware Scheduler for a Parallel Sparse Direct Solver*. PMAA'2008, Neuchatel, Suisse, juin 2008.
- [70] M. Barrault, B. Lathuilière, P. Ramet et J. Roman. *A Domain Decomposition Method Applied to Large Eigenvalue Problems in Neutron Physics*. PMAA'2008, Neuchatel, Suisse, juin 2008.
- [71] M. Barrault, B. Lathuilière, P. Ramet et J. Roman. *A domain decomposition method for the resolution of an eigenvalue problem in neutron physics*. International Symposium on Iterative Methods in Scientific Computing (IMACS), Lille, France, mars 2008.
- [72] P. Hénon, P. Ramet et J. Roman. *On finding approximate supernodes for an efficient ILU(k) factorization*. PMAA'2006, Rennes, France, septembre 2006.
- [73] B. Braconnier, B. Nkonga, M. Papin, P. Ramet, M. Riccuto, J. Roman et R. Abgrall. *Efficient solution technique for low Mach number compressible multiphase problems*. PMAA'2006, Rennes, France, septembre 2006.
- [74] P. Hénon, F. Pellegrini, P. Ramet et J. Roman. *Blocking Issues for an Efficient Parallel Block ILU Preconditioner*. SIAM Conference On Preconditioning Techniques For Large Sparse Matrix Problems In Scientific And Industrial Applications, Atlanta, USA, mai 2005.
- [75] P. Hénon, P. Ramet et J. Roman. *A Blockwise Algorithm for Parallel Incomplete Cholesky Factorization*. PMAA'2004, Marseille, France, octobre 2004.
- [76] P. Hénon, B. Nkonga, P. Ramet et J. Roman. *Using of the High Performance Sparse Solver PaStiX for the Complex Multiscale 3D Simulations performed by the FluidBox Fluid Mechanics Software*. PMAA'2004, Marseille, France, octobre 2004.
- [77] P. Hénon, F. Pellegrini, P. Ramet, J. Roman et Y. Saad. *High Performance Complete and Incomplete Factorizations for Very Large Sparse Systems by using Scotch and PaStiX softwares*. SIAM Conference PPSC'2004, San Francisco, USA, février 2004.
- [78] P. Hénon, F. Pellegrini, P. Ramet et J. Roman. *Towards High Performance Hybrid Direct-Iterative Solvers for Large Sparse Systems*. SIAM Conference On Preconditioning Techniques For Large Sparse Matrix Problems In Scientific And Industrial Applications, Napa, USA, octobre 2003.
- [79] P. Hénon, P. Ramet et J. Roman. *Parallel factorization of very large sparse SPD systems on a network of SMP nodes*. PMAA'2002, Neuchâtel, Suisse, novembre 2002.
- [80] P. Hénon, P. Ramet et J. Roman. *PaStiX: A High-Performance Parallel Direct Solver for Sparse Symmetric Definite Systems*. PMAA'2000, Neuchâtel, Suisse, août 2000.
- [81] D. Goudin, P. Hénon, F. Pellegrini, P. Ramet, J. Roman et J.-J. Pesqué. *Description of the EMILIO Software Processing Chain and Application to Structural Mechanics*. PMAA'2000, Neuchâtel, Suisse, août 2000.

2.6 Revues nationales

- [82] M. Boulet, G. Meurant, D. Goudin, J.-J. Pesqué, M. Chanaud, L. Giraud, P. Hénon, P. Ramet et J. Roman. *Résolution des systèmes linéaires sur calculateurs pétaflopiques*. Revue CHOCS vol 41, revue scientifique et technique de la Direction des Applications Militaires, janvier 2012.

2.7 Conférence nationales avec comité de lecture

- [83] P. Hénon, P. Ramet. *Optimisation de l'occupation mémoire pour un solveur parallèle creux direct hautes performances de type supernodal*. RenPar'2002, Hamamet, Tunisie, avril 2002.
- [84] P. Hénon, P. Ramet. *PaStiX: Un solveur parallèle direct pour des matrices creuses symétriques définies positives basé sur un ordonnancement statique performant et sur une gestion mémoire efficace*. RenPar'2001, Paris, France, avril 2001.
- [85] P. Ramet. *Calcul de la Suite Optimale de Taille de Paquets pour la Factorisation de Cholesky*. RenPar'9, Lausanne, Suisse, pages 111–114, mai 1997.
- [86] P. Ramet. *Calcul de la Taille Optimale des Paquets pour les Algorithmes Macro-Pipelines*. RenPar'8, Bordeaux, France, pages 21–24, juin 1996.

2.8 Ateliers internationaux sur invitation

- [87] P. Ramet. *Heterogeneous architectures, Hybrid methods, Hierarchical matrices for Sparse Linear Solvers*. Seminar at Stanford, avril 2018.
- [88] P. Ramet. *From hybrid architectures to hybrid solvers*. Seminar at Stanford, juillet 2013.
- [89] P. Ramet. *Hybrid methods, Hybrid architectures, Hybrid compressions for sparse direct solvers*. Seminar at Stanford, novembre 2013.
- [90] P. Ramet. *Dynamic Scheduling for Sparse Direct Solver on NUMA and Multicore Architectures*. ComplexHPC meeting, Lisbon, Portugal, octobre 2009.
- [91] P. Hénon, P. Ramet et J. Roman. *A supernode amalgamation algorithm for an efficient block incomplete factorization*. Workshop on parallel iterative solvers and domain decomposition techniques, Minneapolis, USA, juillet 2008.
- [92] P. Hénon, P. Ramet et J. Roman. *A supernode amalgamation algorithm for an efficient block incomplete factorization*. Workshop PPAM'07, Gdansk, Pologne, septembre 2007.
- [93] P. Ramet. *High performances methods for solving large sparse linear systems - Direct and Incomplete Factorization*. Workshops ReDIMsOPS, Japan Atomic Energy Agency, Tokyo, Japon, mai 2007.
- [94] O. Czarny, G. Huysmans, P. Hénon et P. Ramet. *Improvement of existing solvers for the simulation of MHD instabilities*. Numerical flow models for controlled fusion, Porquerolles, France, avril 2007.
- [95] P. Hénon, F. Pellegrini, P. Ramet et J. Roman. *An efficient hybrid MPI/Thread implementation on a network of SMP nodes for the parallel sparse direct solver PaStiX: ordering / scheduling / memory management / out-of-core issues, and application to preconditioning*. Sparse Days and Grid Computing, Saint Giron, France, juin 2003.

2.9 Ateliers nationaux sur invitation

- [96] P. Ramet. *High performance computing in sparse linear algebra*. Séminaire du Laboratoire J.A. Dieudonné, Université Côte d'Azur, France, octobre 2021.
- [97] P. Ramet. *Sparse supernodal solver using block low-rank compression: Design, performance and analysis*. JOEREK development meeting, Cadarache, France, novembre 2019.
- [98] P. Ramet. *Utilisation de techniques de compression low-rank pour un solveur parallèle direct creux*. Journées Ondes du Sud-Ouest, Le Barp, France, mars 2019.
- [99] M. Faverge, G. Pichon, P. Ramet, and J. Roman. *Utilisation de la compression low-rank pour réduire la complexité du solveur PaStiX*. JCAD'2018 : Journées Calcul et Données, Lyon, France, octobre 2018.
- [100] G. Pichon, M. Faverge, P. Ramet, and J. Roman. *Utilisation de la compression Block Low-Rank pour accélérer un solveur direct creux supernodal*. COMPAS'2018 - SOLHAR final meeting, Toulouse, France, juillet 2018.
- [101] P. Ramet. *Solveurs Directs*. Maison de la Simulation : Formation en Algèbre Linéaire Creuse Parallèle, Montpellier, France, mars 2015.
- [102] P. Ramet. *On the design of parallel linear solvers for large scale problems*. Formation CNRS, Journée problème de Poisson, Paris, France, janvier 2015.
- [103] P. Ramet. *PaStiX: Parallel Sparse Matrix Package*. JDEV2015 : Journées Développement Logiciel, Bordeaux, France, juillet 2015.
- [104] P. Ramet. *Task-based linear solvers for modern architectures*. 7th ITER International School, High Performance Computing in Fusion Science, Aix-en-Provence, France, août 2014.
- [105] P. Ramet. *Solveurs Directs*. Maison de la Simulation : Formation en Algèbre Linéaire Creuse Parallèle, Paris, France, mars 2014.
- [106] P. Ramet. *Hybrid methods, Hybrid architectures, Hybrid compressions for sparse direct solvers*. Journée Scientifique du MCIA, février 2014.
- [107] P. Ramet. *Solveurs Directs*. Maison de la Simulation : Formation en Algèbre Linéaire Creuse Parallèle, Paris, France, mars 2013.

- [108] P. Ramet. *Méthodes directes et hybrides pour des solveurs creux adaptés aux machines multiCPUs/multiGPUs*. 3ième Ecole Thématique de Simulation Numérique, Frejus, France, juillet 2013.
- [109] P. Ramet. *Sparse direct solver on top of large-scale multicore systems with GPU accelerators*. CEMRACS'2012, Méthodes numériques et algorithmes pour architectures pétaflopiques, Marseille, France, août 2012.
- [110] P. Ramet. *Solveurs Directs*. Maison de la Simulation : Formation en Algèbre Linéaire Creuse Parallèle, Bordeaux, France, novembre 2011.
- [111] P. Ramet. *Linear algebra and sparse direct methods*. Séminaires de l'école MFN 2011 sur les méthodes et algorithmes pour le calcul haute performance, Roscoff, France, juin 2011.
- [112] P. Ramet. *Formation Parallélisme*. CEMRACS'2010, Modèles Numériques pour la Fusion, Marseille, France, août 2010.
- [113] P. Ramet. *Ordonnancement dynamique dans le solveur PaStiX pour des machines NUMA et multicœurs*. Formation CNRS, Solveurs de systèmes linéaires de grande taille : les avancées récentes, Lyon, France, novembre 2010.
- [114] P. Ramet. *Résolution de Systèmes Linéaires, Algorithmes et Parallélisme*. Formation CNRS, Informatique Scientifique pour le Calcul, Sète, France, octobre 2008.
- [115] P. Ramet et J. Roman. *Méthodes directes hautes performances de résolution en algèbre linéaire creuse*. Ecole CEA-EDF-Inria sur le calcul scientifique intensif, Rocquencourt, France, novembre 2006.

2.10 Rapports de fin de contrat

- [116] M. Alaya, M. Faverge, X. Lacoste, A. Péré-Laperne, J. Péré-Laperne, P. Ramet, and T. Terraz. *Simul'Elec and PASTIX interface specifications*. Algo'Tech, 2015.
- [117] M. Faverge, X. Lacoste, P. Ramet, and T. Terraz. *Etude de la factorisation directe hétérogène et de la factorisation incomplète sur solveur PaStiX appliquées à des systèmes issus de problèmes du CEA/CESTA*. CEA/DAM/CESTA, 2015.
- [118] G. Caramel et P. Ramet. *Optimisation des performances des outils de calcul de neutronique des coeurs*. E.D.F. / SINETICS, 2007.
- [119] P. Hénon, P. Ramet et J. Roman. *Evaluation des performances de la version SMP du solveur PaStiX de la chaîne logicielle EMILIO dans l'environnement du code ODYSSEE du CESTA*. CEA/DAM/CESTA, 2005.
- [120] P. Hénon, F. Pellegrini, P. Ramet et J. Roman. *Etude sur l'applicabilité de méthodes itératives nouvelles aux problèmes du CESTA*. CEA/DAM/CESTA, 2004.
- [121] P. Hénon, P. Ramet et J. Roman. *Amélioration et Extension du Solveur Direct Parallèle pour Grandes Matrices Creuses du CESTA*. CEA/DAM/CESTA, 2003.
- [122] D. Lecas et P. Ramet. *Parallélisation du code MIRO*. CEA/DAM/CESTA, 2001.
- [123] D. Goudin, P. Hénon, F. Pellegrini, P. Ramet et J. Roman. *Mise en oeuvre d'une Bibliothèque d'Outils pour la Résolution par Méthode Directe de Grands Systèmes Linéaires Creux Symétriques Définis Positifs sur Machine Parallèle*. CEA/DAM/CESTA, 2000.

2.11 Encadrement de thèses

- [124] G. Pichon. *On the use of low-rank arithmetic to reduce the complexity of parallel sparse linear solvers based on direct factorization techniques*. PhD thesis, LaBRI, Université Bordeaux, Talence, France, novembre 2018.
- [125] S. Moustafa. *Massively Parallel Cartesian Discrete Ordinates Method for Neutron Transport Simulation*. PhD thesis, LaBRI, Université Bordeaux, Talence, France, décembre 2015.
- [126] A. Casadei. *Optimizations of hybrid sparse linear solvers relying on Schur complement and domain decomposition approaches*. PhD thesis, LaBRI, Université Bordeaux, Talence, France, octobre 2015.
- [127] X. Lacoste. *Scheduling and memory optimizations for sparse direct solver on multi-core/multi-gpu cluster systems*. PhD thesis, LaBRI, Université Bordeaux, Talence, France, février 2015.
- [128] B. Lathuillère. *Méthode de décomposition de domaine pour les équations du transport simplifié en neutronique*. PhD thesis, LaBRI, Université Bordeaux, Talence, France, janvier 2010.
- [129] M. Faverge. *Ordonnancement hybride statique-dynamique en algèbre linéaire creuse pour de grands clusters de machines NUMA et multi-cœurs*. PhD thesis, LaBRI, Université Bordeaux, Talence, France, décembre 2009.

3. Développements technologiques

[130] PaStiX. Un solveur haute performance pour la résolution de grands systèmes linéaires creux basé sur une approche directe supernodale (voir la contribution 1).

Disponible à partir de l'adresse suivante : <https://gitlab.inria.fr/solverstack/pastix>.

Dépôt APP effectué par Inria (IDDN.FR.001.230016.000.S.C.2008.000.31235).

Self-Assessment³ : A-5, SO-4, SM-4, EM-4, SDL-5

Own-Contribution : DA-4, CD-3, MS-3, TPM-4

4. Impact socio-économique et transfert

Comme mentionné précédemment dans le dossier (voir les contributions 2, 4 et 5 pour mes collaborations avec le CEA, EDF et AlgoTech), j'ai été amené à collaborer avec de nombreux partenaires industriels pour des transferts de compétences acquises dans le domaine du calcul haute performances et des transferts technologiques pour le solveur PaStiX. Certains transferts ont conduit à des contrats d'expertise rémunérés avec livrables [116, 117, 118, 119, 120, 121, 122, 123]. Les deux thèses CIFRE avec EDF étaient associées à un contrat d'accompagnement respectant l'accord cadre passé entre EDF et Inria.

³Auto-appréciation selon le canevas disponible dans le document « Criteria for Software Self-Assessment » disponible à l'URL https://www.inria.fr/sites/default/files/2019-10/Criteria_software_self_assessment.pdf.