

Mini-symposium EFFCALC
Calcul Scientifique : efficacité calculatoire des méthodes numériques

Mini-symposium porté par le GdR Calcul

Résumé

L'évolution actuelle et prévisible des architectures de machines est en train d'imposer une révision de nombreuses méthodes numériques. En particulier les notions de parallélisation, de localité des données et de vectorisation deviennent essentielles pour tirer le maximum de performance des calculateurs. Les laboratoires d'informatique et de mathématiques développent des techniques pour le calcul hautes performances qui interviennent à des niveaux différents, sans forcément beaucoup d'interaction mutuelle. L'objectif de ce mini-symposium est d'aborder des thématiques à l'interface des deux disciplines et de présenter des points de vue complémentaires sur l'efficacité calculatoire des méthodes numériques et les optimisations automatiques proposées par la compilation de stencils et le tiling.

Organisateur(s)

1. **Violaine Louvet**, CNRS / GRICAD.
2. **Alain Darte**, ENS Lyon / LIP.
3. **Thierry Dumont**, Université Lyon 1 / ICJ.

Liste des orateurs

1. **Alain Darte, Thierry Dumont**, LIP / ICJ
Titre : Repenser les schémas numériques à l'aune de la localité et de la vectorisation.
2. **Florian De Vuyst**, CMLA / ENS Paris-Saclay
Titre : Schémas numériques et modèles de performances associés. Concepts et exemples.
3. **Emmanuel Agullo**, HiePACS / INRIA Bordeaux - Sud-Ouest
Titre : Programmation et passage à l'échelle de bibliothèques mathématiques numériques à base de tâches sur moteur d'exécution.
4. **Philippe Helluy**, IRMA / Université de Strasbourg
Titre : Palindromic discontinuous Galerkin method.

Violaine Louvet, Grenoble Alpes Recherche - Infrastructure de Calcul Intensif et de Données, violaine.louvet@univ-grenoble-alpes.fr

Alain Darte, Laboratoire de l'Informatique du Parallélisme, alain.darte@ens-lyon.fr

Thierry Dumont, Institut Camille Jordan, tdumont@math.univ-lyon1.fr

1 Repenser les schémas numériques à l'aune de la localité et de la vectorisation

Alain Darte, Thierry Dumont

L'évolution de l'architecture des machines (incluant tous les types de matériels et d'environnements, des CPU et GPU à l'Exascale), de leurs coeurs de calcul et de leur hiérarchie mémoire, nécessite de repenser les schémas de discrétisation et les algorithmes en termes de localité (spatiale et temporelle) et de vectorisation, vers des méthodes si possible plus vectorielles et plus denses.

Parmi les différentes possibilités, deux voies sont complémentaires :

- l'utilisation de méthodes numériques adaptées (les méthodes de type Galerkin discontinu et les méthodes explicites stabilisées en sont des exemples),
- l'utilisation de techniques d'optimisation de code automatiques ou semi-automatiques (par exemple l'approche polyédrique et le "tiling").

Nous donnerons une courte introduction à ces techniques.

2 Schémas numériques et modèles de performances associés. Concepts et exemples

Florian De Vuyst

Il est acquis aujourd'hui que les prochaines architectures de processeurs seront dotées de multiples unités de traitement éventuellement hétérogènes avec différents niveaux de mémoires et de caches. Au delà des différentes options d'optimisations de compilation, il n'est plus du tout clair de savoir comment va se comporter un schéma numérique sur ce type d'architectures, et quelle performance (en millions de mises à jour par cellules par secondes pour des problèmes d'évolution par exemple) va être atteinte. Les premiers modèles de performance analytiques [Roofline, Williams et al. 2009] sont apparus pour permettre de mieux comprendre les facteurs clés de performance pour les non-experts. Aujourd'hui, des raffinements de ces modèles ([Execution Cache Memory ECM model, Stengel et al. 2015], [Automatic Cache-aware Roofline, Denoyelle et al. 2016] permettent de tenir compte des tailles et vitesses des différents niveaux de cache. Dans notre équipe, nous avons tenté de dériver un modèle "roofline" ainsi qu'un modèle ECM pour un schéma numérique complexe de type Lagrange+projection à variables décalées d'ordre deux, utilisés dans des les codes de production d'hydrodynamique multimatériaux. Il a été possible de construire des modèles quantitatifs pour ce type de schéma numérique. Le modèle roofline est suffisamment précis pour les kernels compute-bound et présente des taux d'erreur importants pour les kernels memory-bound. Ce défaut est corrigé avec le modèle ECM, ce qui valide le besoin de bien comprendre la dynamique des données dans les différents niveaux de cache. L'analyse des modèles de performance obtenus nous a orientés vers la dérivation d'une nouvelle famille de schémas numériques – les schémas Lagrange-Flux – qui sont une variante des schémas Lagrange+projection mais où les performances et la scalabilité multi-coeurs sont bien meilleurs. Il s'avère que les schémas Lagrange-Flux ont enfin un potentiel de généralisation à d'autres problèmes de nature hyperbolique. Cette expérience aux frontières des mathématiques et de l'informatique nous a convaincu que la méthodologie schéma/performance permet d'explorer des pistes nouvelles/innovantes en méthodes computationnelles. Travail en collaboration avec Thibault Gasc, Renaud Motte, Mathieu Peybernes et Raphael Poncet.

3 Programmation et passage à l'échelle de bibliothèques mathématiques numériques à base de tâches sur moteur d'exécution

Emmanuel Agullo

La complexité architecturale des super-calculateurs modernes a amené la communauté des développeurs de bibliothèques scientifiques à envisager des paradigmes de programmation de relativement haut niveau et à s'appuyer sur des entités tierces pour déployer leurs calculs sur les machines sous-jacentes. Dans cet exposé, nous allons présenter un tel paradigme, la programmation à base de tâches, et son utilisation pour

développer des solveurs d'algèbre linéaire dense et creux. Nous allons montrer que nous pouvons allier productivité (programmation séquentielle à base de tâches) et haute performance (en nous appuyant sur un moteur d'exécution). Nous illustrerons cet exposé avec le solveur d'algèbre linéaire dense Chameleon, le solveur creux `qr_mumps` et la bibliothèque `ScalFMM` implantant une méthode multipôle rapide. Nous montrerons comment ces bibliothèques écrites à base de tâches sont parallélisées automatiquement lors de l'exécution par le moteur d'exécution `StarPU` sur des super-calculateurs homogènes et hétérogènes.

4 Palindromic discontinuous Galerkin method

Philippe Helluy

We present a high order scheme for approximating kinetic equations with stiff relaxation. The objective is to provide efficient methods for solving the underlying system of conservation laws. The construction is based on several ingredients : (i) a high order implicit upwind Discontinuous Galerkin approximation of the kinetic equations with easy-to-solve triangular linear systems ; (ii) a second order asymptotic-preserving time integration based on symmetry arguments ; (iii) a palindromic composition of the second order method for achieving higher orders in time. The method is then tested at orders 2, 4 and 6. It is asymptotic-preserving with respect to the stiff relaxation, accepts high CFL numbers and is well adapted to parallel optimizations.