Rapport sur le manuscrit de Thèse présenté par

Fabrice Dupros

pour l'obtention du Doctorat de l'Université de Bordeaux I

Rapporteur : Stéphane GENAUD,
Maître de conférences habilité à diriger les recherches,
Université de Strasbourg.

Titre du document évalué :

Contribution à la modélisation numérique de la propagation des ondes sismiques sur architectures multicœurs et hiérarchiques

Le document présenté par Fabrice Dupros décrit les recherches qu'il a menées dans le cadre de sa thèse sous la direction conjointe de Dimitri Komatitsch, Professeur à l'Université de Pau, membre de l'équipe projet INRIA Magique 3D, et Jean Roman, professeur à l'Université de Bordeaux I, responsable de l'équipe projet INRIA HIEPACS.

Contexte de l'étude

Le cadre général du travail est celui de la mise en oeuvre efficace d'applications de modélisation sismique sur des architectures de calcul modernes. Ce travail étudie les moyens d'améliorer l'efficacité de méthodes numériques classiques, la méthode des éléments finis, et celle des différences finies fréquement utilisées dans ce type d'application, pour des processeurs multi-cœurs disposant d'une hériarchie de mémoires.

Le travail touche un sujet crucial et d'actualité qui concerne toute l'industrie du logiciel et du matériel : comment programmer efficacement des ordinateurs dont l'architecture est composée de différents niveaux de parallélisme (par exemple plusieurs processeurs contenant chacun de plus en plus de cœurs), et de mémoires organisées hiérarchiquement. Si la multiplication des unités de calcul est la seule voie envisageable aujourd'hui pour croître en puissance de calcul, les exploiter au mieux est particulièrement difficile. C'est pourtant un défi que la simulation numérique, l'un des domaines qui peut bénéficier le plus de puissance de calcul accrue, doit relever.

Analyse du document présenté

Le document est organisé en deux parties comptant 6 chapitres qui décrivent le cœur du travail.

Un chapitre d'introduction générale, très concis, présente d'abord les enjeux de la simulation numérique en géophysique et les caractéristiques des méthodes numériques proposées dans la littérature. Il esquisse ensuite l'évolution des architectures de calcul, multi-cœurs, avec des temps d'accès à la mémoire non-uniformes (NUMA). Suit un premier chapitre (*Cadre de l'étude*) qui reprend en détails ces éléments. Il décrit plus précisément les méthodes des différences finies et des éléments finis et le

traitement particulier des bords du domaine (conditions aux limites), puis expose les évolutions marquantes des architectures de calcul, notamment celles concernant l'organisation de la mémoire par rapport aux unités de calcul. Logiquement, les évolutions des modèles de programmation pour exprimer le parallélisme sont introduites à la suite. On peut regretter que la discussion de cet état de l'art ne soit pas plus argumentée concernant les alternatives non choisies pour le reste de l'étude, comme les modèles de programmation de type PGAS. Davantage de références internationales auraient donné plus de poids aux choix des solutions étudiées en profondeur par la suite. Ces solutions sont néanmoins parfaitement judicieuses car reconnues au meilleur niveau international et parfois même présentes au sein de logiciels majeurs, comme hwloc, sous-projet d'OpenMPI. Ce chapitre pose au final une hypothèse forte du travail, qui est que le multi-threading est le meilleur modèle de programmation candidat. Il est invoqué, à juste titre, l'argument que le modèle à passage de messages (MPI) prépondérant jusqu'à récemment, induit des surcoûts prohibitifs en mémoire (zones d'échanges). Cependant, l'idée d'utiliser MPI dans un mode hybride n'est pas écartée, annonçant la solution retenue au chapitre 5.

La première partie traite ensuite de la modélisation élastodynamique. Elle est constituée de quatre chapitres.

Le chapitre 2 traite de l'équilibrage de charge en prenant pour motivation les calculs plus longs sur les bords des domaines (limites absorbantes) qu'ailleurs. M. Dupros justifie le choix d'un découpage quasi-statique permettant de ne pas perdre les avantages liés à la régularité de la grille, puis il propose un code simplifié (un Jacobi) présentant une structure de boucles imbriquée similaire au code visé, afin d'évaluer diverses implémentations multi-threadées. Les tests menés comparent le comportement des threads POSIX avec GNU OpenMP et des environnements de recherche de l'Université de Bordeaux (GOMP, ForestGOMP utilisant les threads MARCEL, puis MPC), sous deux angles : l'opportunité de faire du multi-threading imbriqué et la capacité à équilibrer dynamiquement la charge (automatiquement via les directives). Dans ce cas test faisant abstraction des accès mémoire, l'évaluation montre qu'une efficacité quasi idéale peut être atteinte dès qu'on a au moins un thread par cœur, mais que l'imbrication des threads est mal supportée. Dans ces expériences, la démarche de réduire au minimum le noyau de calcul pour isoler l'effet observé lors de l'évaluation est très pertinente.

Le chapitre 3 (*Prise en compte de l'affinité mémoire*) a pour objectif de clarifier les effets de la localité des données par rapport aux calculs. L'évaluation couvre un nombre conséquent (6) d'architectures NUMA différentes. L'expérimentation procède à nouveau par tests unitaires pour mettre en évidence les effets NUMA. L'étude de l'affinité compare l'environnement BubbleSched, permettant à l'utilisateur de délimiter des ensembles de threads, et à travers cela, la localité des données à certains ensembles, et la politique de placement des données par défaut de Linux. Les résultats montrent que des gains importants, de l'ordre de 10% (16 cœurs) à 40% (512 cœurs) sont envisageables.

Le chapitre 4 (*Décomposition espace-temps et réduction du trafic mémoire*) s'attache au problème des accès mémoires. L'exposé montre clairement l'aspect irrégulier des accès mémoire dans la méthode des différences finies. Il est ensuite proposé de mettre en œuvre la technique bien connue du *time skewing* (ré-organisation des accès) sur ce problème. L'exposé débouche sur l'énoncé d'un algorithme général qui découle des contraintes du cas d'espèce. Bien qu'introduit pédagogiquement, on aurait aimé que celui-ci soit formalisé. L'évaluation est approfondie, montrant des résultats convergents, sur le noyau Jacobi ainsi que sur le noyau du code sismique, d'un gain d'au moins 25% en cycles et défauts de cache. Ces résultats sont comparés à la littérature et sont convaincants quant au fait que le time skewing permet de réduire significativement l'impact de la hiérarchie mémoire. Le deuxième volet de l'expérience ajoute à la stratégie du *time skewing* la politique de placement mémoire. Ce choix permet de vérifier que les deux stratégies n'ont pas d'effets antagonistes. Les résultats finaux montrent une amélioration convergente du comportement sur la plupart des architectures. L'expérience, répétée

sur différentes architectures, met en évidence les différences très importantes de comportement, avec un gain quasi-nul quand le bus mémoire représente un goulot d'étranglement. L'expérience met en évidence l'impact du trafic mémoire en fonction des architectures.

Le chapitre 5 propose de combiner les modèles de programmation multi-threadés et ceux à passage de messages dans l'optique d'utiliser des clusters de nœuds multicœurs. Si cette approche dite *hybride* est bien connue, il n'est pas donné d'indication sur l'avantage qu'elle aurait en comparaison d'une version uniquement multi-threadée. La comparaison est faite avec une version purement MPI. Poursuivant le cheminement, une expérience est réalisée avec le code de sismique sur un exemple synthétique en utilisant un cluster complet. Les résultats montrent un déséquilibre de charge inférieur pour la version hybride, permettant un gain en performance de l'ordre de 25%.

La deuxième partie, beaucoup plus courte, traite de la méthode des éléments finis. On y trouve deux chapitres. Le chapitre 6 discute d'abord des problématiques liées aux noyaux de calculs matriciels, et en particulier le choix d'un solveur répondant aux exigences du cas applicatif. Pour cette partie applicative, il est souligné l'importance du partitionnement des éléments du maillage. La contribution de ce chapitre réside dans un algorithme de coloration de graphe original. L'algorithme permet de déterminer l'affectation des éléments de maillage aux cœurs avec l'objectif de minimiser le déséquilibre en calculs. Le chapitre 7 présente finalement une simulation numérique en vraie grandeur en utilisant 1024 cœurs avec des résultats convaincants.

Conclusion

Le document est très bien rédigé et agréable à lire. Fabrice Dupros y démontre des connaissances poussées à la fois dans le domaine des méthodes numériques utilisées pour la géophysique, et dans celui de la programmation parallèle des architectures de calcul moderne. Il faut souligner ici la grande variété des connaissances qu'il a fallu acquérir dans le domaine du parallélisme pour mettre en œuvre cette étude.

La contribution présentée est précieuse et assez rare : il est généralement difficile de comparer de manière rigoureuse, dans un vrai environnement applicatif, plusieurs solutions concurrentes. Ce travail est particulièrement intéressant aussi car la démarche adoptée pour le conduire est rigoureuse. Les différents aspects qui ont un impact sur la performance ont été isolés, étudiés séparément puis assemblés pour en comprendre les interactions. C'est donc une étude expérimentale scientifiquement bien fondée. On peut regretter cependant qu'une annexe ne liste pas les détails (numéros de versions, implémentations MPI) des logiciels utilisés, ceci à des fins de reproductibilité des expériences.

Les travaux ont fait l'objet de plusieurs publications (en particulier, une revue internationale du meilleur niveau, et deux conférences internationales). Considérant tous ces éléments, j'émets un avis favorable à la soutenance de la thèse de Fabrice Dupros en vue de l'obtention du titre de Docteur en informatique de l'université de Bordeaux I. La thèse peut être soutenue en l'état.

Stéphane GENAUD,
fait à Strasbourg, le 22 novembre 2010.