

CALCUL HAUTE PERFORMANCE ET ACCÉLÉRATEURS GPU

H. JOURDREN
CEA - DAM - Île-de-France

Dans le domaine du "calcul haute performance", l'arrivée de cartes graphiques dédiées au calcul numérique, programmables en langage de haut niveau, a constitué un fait marquant de l'année 2007. Objet de travaux académiques depuis plusieurs années, l'approche *GPGPU* – *General Purpose Graphic Processing Unit* – peut dorénavant être mise en œuvre, et évaluée sur des problèmes de petite taille, dans un langage de programmation proche de langages généralistes tels que C ou C++.

Les premières évaluations réalisées en 2007 ont porté sur la résolution des équations de la dynamique des gaz. Plusieurs équations d'état en formulation analytique ou tabulée ont été testées sur carte graphique, ainsi que des schémas numériques précis en hydrodynamique compressible monofluide. Les cartes utilisées étaient de type nVidia 8800 GTX (*grand public*) ou Quadro (*professionnel*). Les tests étaient menés en mode "simple précision GPU" : calculs en simple précision sans compatibilité IEEE, et sans code de correction d'erreur (ECC) pour la mémoire de la carte graphique.

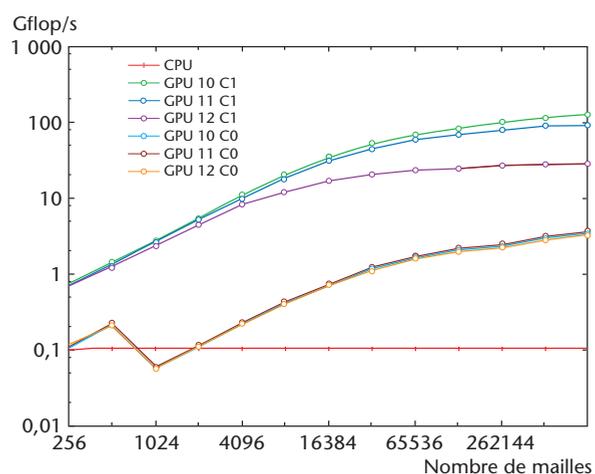
Les résultats reportées sur la figure concernent l'équation d'état *JWL* de produits de détonation, coûteuse en temps calcul. Dans le cas favorable d'un grand nombre de mailles résidant déjà en mémoire sur la carte graphique (*environ 1 million de mailles*), nous relevons des performances comprises entre 60 et 100 Gflop/s, selon le mode d'adressage utilisé. De tels chiffres ont été également atteints – voire dépassés – avec plusieurs schémas hydrodynamiques sur des maillages structurés. Ces chiffres élevés sont à comparer aux 6 Gflop/s crête obtenus actuellement par cœur de calcul sur un processeur généraliste. Ces résultats sont donc tout à fait encourageants.

Avec l'arrivée de prochaines générations de cartes offrant un mode double précision *GPU* matériel – calculs double précision sur accélérateurs *GPGPU*

non conformes à la norme *IEEE* 64 bits – et des capacités mémoire plus élevées, il conviendra de surveiller l'envolée annoncée des performances crêtes, ainsi que l'évolution des coûts (*cartes professionnelles*), et de la consommation électrique.

Associée à des processeurs multi-cœurs généralistes, l'accélération *GPGPU* pourrait devenir intéressante en calcul scientifique, pour des applications lourdes, comme la dynamique des gaz, l'élasticité, ou l'hydrodynamique. Sur des problèmes de grande taille en schémas explicites, des modifications profondes des architectures des codes parallèles sont à prévoir pour une utilisation optimale. Au-delà de la réécriture des noyaux de calcul sous forme de longues boucles vectorielles, des techniques complexes de chargement et de déchargement de sous-domaines sont à envisager, car l'ensemble des sous-domaines ne peut pas résider en mémoire de façon permanente sur les cartes accélératrices.

Comme la vectorisation ou la gestion "out of core" de sous-domaines dans les années 80 et 90, le développement de nouveaux modèles de programmation hybride, incluant processeurs multi-cœurs et accélérateurs, s'impose comme thème de recherche pour le calcul haute performance.



Figure

Calcul de l'équation d'état *JWL* sur accélérateur graphique nVidia G80. Les courbes représentent les performances en Gflop/s, 32 bits, en simple précision non-IEEE, en fonction du nombre de mailles.

CPU : référence processeur généraliste ; I0 : adressage direct ; I1 : faux adressage indirect ($index(i) = i$) ; I2 : adressage indirect ; C0 : données résidant côté CPU, avec déport PCI Express 4 Go/s vers l'accélérateur ; C1 : données résidant en mémoire sur l'accélérateur.