

# [NVIDIA intensifie sa stratégie HPC avec des avancées dans les GPU et les puces ARM](#)

Dans le cadre du SC11 de Seattle, NVIDIA a fait plusieurs annonces, détaillées ici sur deux pages.

## **Un nouveau standard pour la programmation parallèle**

La firme propose tout d'abord une solution logicielle permettant de faciliter la mise au point d'applications massivement parallèles, aptes à tourner sur ses GPU.

Cette volonté se traduit par l'introduction de l'OpenACC, un standard soutenu par des acteurs comme Cray, PGI et CAPS Entreprise. Cette technique permet d'insérer des directives dans du code classique, lesquelles seront converties par un compilateur compatible OpenACC en optimisations dédiées aux GPU et aux CPU multicœurs.

Le code demeure pour le reste inchangé et sera toujours compilable avec des outils courants. Cette technique permet donc de coder le logiciel de façon standard, puis de passer par une phase de mesure des performances et d'optimisation des parties critiques du logiciel. Nvidia promet que cette approche permettra aux développeurs de multiplier par deux la vitesse d'une application effectuant des calculs en seulement quatre semaines.

Une version d'essai du PGI Accelerator (un premier produit C et Fortran compatible avec l'OpenACC) est accessible [depuis cette page web](#) du site de NVIDIA.

## **Un compilateur rénové pour les cinq ans de CUDA**

La technologie CUDA de la firme vient de souffler sa cinquième bougie. À cette occasion, le Dr. **Ian Buck**, l'inventeur de CUDA et le directeur général du GPU Computing chez Nvidia [revient sur cette avancée technologique](#) qui a changé le monde des composants graphiques, comme celui des supercalculateurs.

Pour fêter dignement cet événement, la *release candidate* de CUDA 4.1 fait son entrée. Elle introduit un compilateur *open source* basé sur LLVM. Ce dernier est non seulement ouvert, mais il permet aussi d'améliorer la vitesse des applications d'un maximum de 10 %.

Un nouvel outil de profilage permettra pour sa part de mieux mesurer les performances du code écrit. Enfin, plus de 1000 nouvelles fonctions font leur entrée dans la librairie de traitement d'images fournie avec CUDA.

Dernier point, les utilisateurs pourront dorénavant s'équiper avec une nouvelle génération de stations de travail, combinant un GPU orienté graphismes (une carte Quadro) et un GPU orienté calculs (une carte Tesla C2075 créditée d'une puissance de 515 gigaflops en double précision). Dell, Fujitsu, HP et Lenovo proposent d'ores et déjà des machines 'Maximus' dans leurs catalogues respectifs.

## CUDA à l'attaque des ARM

La firme vient de publier les résultats de son dernier trimestre : 1,07 milliard de dollars de chiffre d'affaires. Une valeur en hausse de 26,3 % depuis 2010. Le Tegra 2 a sans conteste eu un rôle important à jouer dans ces bons résultats. [La sortie récente du Tegra 3](#) devrait confirmer cette tendance.

Nvidia ne compte pas s'arrêter là. La société a en effet annoncé [dès janvier 2011](#) son intention d'investir rapidement les marchés des PC de bureau, des serveurs et des supercalculateurs avec des composants ARM conçus à cet effet. Les puces de cette nouvelle famille intégreront des cœurs ARM et un puissant GPU.

Dans l'attente, des solutions de test commencent à poindre leur nez. L'Italien Seco propose ainsi [une carte de développement](#) qui sera prochainement adaptée pour proposer un processeur Tegra 3 et un GPU NVIDIA. Le tout sera livré avec une mouture ARM de CUDA. Cette carte permettra de mettre en valeur l'efficacité énergétique combinée des deux technologies. Arrivée prévue au premier semestre 2012.

Notez que la firme ne souhaite pas encore fournir de *roadmap* concernant les futurs composants du « *Project Denver* ».

## Un premier supercalculateur à base de puces ARM

Sous l'impulsion du BSC (*Barcelona Supercomputing Center*), un premier supercalculateur à base de technologies ARM est en train de voir le jour. La machine de test [du projet Mont-Blanc](#) comprend actuellement des GPU pilotés par 256 processeurs ARM.

Une première étape pour cette initiative de grande envergure, qui vise à proposer dès 2014 des *clusters* capables d'afficher une efficacité énergétique quatre à dix fois supérieure à celle des modèles actuels. Sachant que la puissance électrique de ces machines se compte aujourd'hui en mégawatts, on comprend pourquoi le BSC souhaite accélérer la transition vers le couple ARM+GPU.

À plus long terme, une réduction de la consommation électrique d'un facteur allant de 15 fois à 30 fois a été évoquée. Objectif, fournir un supercalculateur d'un exaflops avec une consommation de 20 MW soit 50 gigaflops/watt, contre 1,7 gigaflops par watt pour les meilleurs *clusters* actuels.