Activités P2IO autour du calcul parallèle

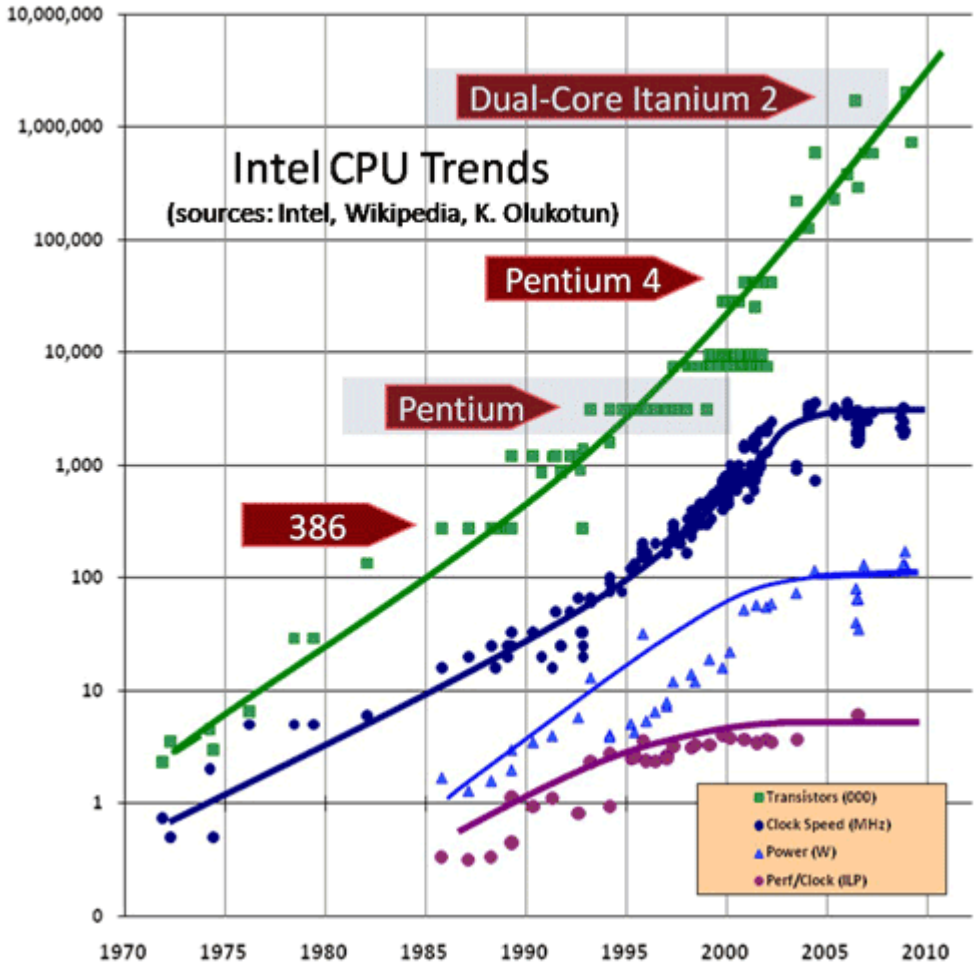
Instructions vectorielles, accélérateurs, mémoire partagée et/ou distribuée… vers un matériel hétérogène et une programmation hybride.

Remerciements : Ivana Hrivnacova, Benoit Blossier, Sacha Brun, Claude Mercier, Olivier brand-Foissac, Xavier Grave, Gilles Grasseau, Arnaud Beck, Christophe Diarra, Antoine Pérus, Sébastien Binet, Joël Surget, Charles Loomis, Eric Legay.

On ne rase plus gratis

Herb Sutter, 2005

- *Fréquence d'horloge et puissance électrique stagnent.*
- *Le nombre de transistors continue sa croissance :*
  - **Multiplication des coeurs**
  - **Ajout de capacités de calcul vectoriel au sein des coeurs : SSE, AVX.**
  - Augmentation des caches.
- *Du supercalculateur au téléphone, il faut maintenant programmer “parallèle”.*



QCD sur réseau

- *La saga des APE (Array processor experiment)*
  - Puce dédiée au QCD crée dans les années 80 par des italiens.
  - vers 1995 : LPT et IRISA rejoignent la collaboration.
  - Vers 2005 : dernière génération, les APEnext.
  - Algorithme spécifique : Hybrid Monte-Carlo.
  - Langage de programmation dédié : TAO.
  - Une source d'inspiration pour les BlueGene d'IBM.
- *Localement*
  - Laboratoires impliqués : LPT, LAL, IRFU/SPhN.
  - Projets : QCDNext, PetaQCD, Nepal.
  - Ressources : 1024 nœuds APEnext.
- *Perspectives*
  - Aujourd'hui : utilisation des 65536 coeurs de BlueGene/Q de l'IDRIS.
  - Demain : machines CPU-GPU ?

Coast

- *~40 personnes impliquées (IRFU/SAP, IRFU/SEDI)*
  - Profils : ingénieurs algorithme, physiciens developpeurs et physiciens utilisateurs.
- *Simulation à toutes les échelles de* l'astrophysique
  - Grandes structures de l'univers, fusion de galaxies.
  - Etoiles sur la séquence principale ou dans les phases avancées.
  - Disques d'accrétion, formation des étoiles, milieu interstellaire.
  - Anneaux de saturne.
- *Ressources matérielles*
  - En interne : 1200 cœurs pour la simulation, plusieurs noeuds pour la visualisation/analyse (0.5 à 1 To de ram, cartes type quadro + fermi/kepler), couplés à quelques centaines de To de stockage. Réflexion en cours pour obtenir des financements récurrents.
  - GENCI : ~40-50% des demandes au CP4.
  - PRACE : déjà obtenu 36 Mh.
- *Logiciels maison*
  - En production : Ramses (MPI), Heracles (MPI), ASH (MPI), Sdvision (IDL).
  - En développement : variantes hybrides soit MPI-OpenMP, soit Cpu-Gpu.

Geant4

- ***Geant4 10.0** livré le 6 décembre 2013*
  - A la demande, possibilité de simuler les évènements en parallèle.
  - Implémenté à l'aide de multi-threading de type POSIX.
  - Les données en lecture seule sont partagées (géométrie, cross-sections…)
  - Reproductibilité des résultats
    (un générateur aléatoire par thread + un germe prédéfini par évènement).
  - Excellente efficacité et passage à l'échelle, en multi-cœurs et sur Xeon Phi.
  - Possibilité de croiser avec MPI.
  - Prototypes en cours utilsant TBB, les GPUs…
  - Tous les contributeurs doivent maintenant faire du code « Thread-Safe » : **IRFU/SPhN, IPNO, LAL, LLR**, LAPP, LPC-Clermont, CENBG,
- ***hGATE**, dérivé de GATE, dérivé de Geant4*
  - Une implémentation hybride CPU/GPU pour les applications d'imagerie et de thérapie.
  - Partenaires : **IMNC, DSV/I2BM/SHFJ**, CPPM, IPHC, …

GridCL

- *Objectif premier*
  - Evaluer l'utilisation de matériel many-core hétérogène, via OpenCL, au sein d'une grille.
- *Partenaires : **LLR**, **IAS**, LAL, IMNC, IRFU, IPNO, LPT…*
- *Ressources partagées (financement P2IO)*
  - **2** noeuds sandy-bridge, dotés chacun de **2 NVidia K20**, connectés en **Infiniband**.
  - **2** noeuds sandy-bridge, dotés chacun de **2 Intel 5110P**, connectés en **Infiniband**.
  - Bientôt : **1** noeud ivy-bridge, doté de **6 NVidia Titan**.
  - Logiciel : **OpenCL**, CUDA 5, Intel Cluster Studio XE, CAPS OpenACC.
- *Expérimentations*
  - Pour CMS : reconstruction des traces dans les collisions d'ions lourds.
  - Pour CTA : traitement de signaux de télescope.
  - Pour SDO : traitement d'images satellitaires sur matériel hétérogène.
  - Evaluation et enrichissement du banc d'essai SHOC.
- *Dissémination*
  - Atelier OpenMP/MPI/OpenCL aux JDEVs 2013
  - A venir : formation sur les outils de profilage Intel

LPaSo (LHC Parallel Software)

- *Objectif*
  - Gagner en performance en exploitant les cœurs multiples, les instructions vectorielles, les accélérateurs, …, afin de pouvoir absorber la montée en luminosité du LHC.
- *Partenaires*
  - **LAL (ATLAS, LHCb), LLR (CMS), IRFU/SPP (ATLAS)**
  - LPNHE (LHCb), LPC-Clermont (ATLAS)
  - LPTHE
  - LRI, LIMOS
- *Thématiques*
  - Parallélisme de tâches avec GaudiHive (ATLAS, LHCb).
  - Déclenchement haut-niveau avec GPUs (LHCb).
  - Traitement de données avec accélérateurs (CMS).
  - Parallélisation/vectorisation des outils d'analyse statistique.
  - Parallélisation/vectorisation de FastJet.
- *Candidature ANR « défi de tous les savoirs »*

HEP Software Collaboration

- *Motivation*
  - Pour exploiter efficacement les nouveaux matériels, et garder le contact avec les autres communautés scientifiques, notre patrimoine logiciel vieillissant a besoin d'une refonte profonde (C++11, parallélisme sous toute ses formes).
- *Proposition*
  - Transformer le « Concurrency Forum » en collaboration plus formelle, afin d'apporter plus de reconnaissance aux contributeurs, de solliciter des fonds auprès de H2020 et NSF/DOE, d'être plus attractif auprès de l'industrie.
- *Work Packages*
  - Etudes R&D courtes sur les alternatives matérielles et logicielles.
  - Remaniement des bibliothèques et boites à outils existantes.
  - Développement de nouveaux composants logiciels d'intérêt général.
  - Constitution d'une infrastructure d'essai matérielle (Xeon/Phi, AMD, NVidia, ARM, …) et logicielle (compilateurs, déboggeurs, profileurs,…).
  - Déploiement d'outils et processus communs (dépôts, système d'intégration continue, ...).
  - Expertise, consultance et accompagnement auprès des expériences.
- *Réunion de lancement au CERN 3-4 Avril*

Bienvenue dans la jungle

Herb Sutter, 2012

- *Tendances technologiques*
  - Multiplication des cœurs, allongement des vecteurs, mélange d'unité de calcul différentes, générales et spécialisées => **matériel hétérogène**
  - Besoin de gérer ce matériel hétérogène en mélangeant plusieurs modes de programmation => **programmation hybride**
  - Les gains de performance rendent certaines tâches habituellement offline envisageables en online. Apparition de chips multi-core embarqués.
- *Les laboratoires de P2IO au cœur de la bataille*
- *Impact pour les chercheurs développeurs*
  - Moins d'objets, de pointeurs… retour des tableaux de float !
  - Sans doute la fin de la mémoire unique, globale, uniforme.
  - L'option facile, pour les tâches génériques : utiliser des bibliothèques déjà parallélisées.
  - Pour les codes "maison" : besoin de s'intéresser au minimum à la programmation "thread-safe", puis éventuellement à la programmation par directives (OpenMP, OpenACC), puis à la programmation plus bas niveau d'accélérateurs (OpenCL, CUDA) et de processus communiquants (MPI).

Des questions ?

Activités P2IO autour du parallélisme

Diapositives annexes : activités par laboratoire.

CSNSM

- *Projets*
  - Narval : massivement multi-tâches, grâce au support natif d'ADA et à son annexe pour l'informatique distribuée.
  - DCOD : fusion-refonte de NARVAL et ENX.
- *Compétences et savoir-faire présents*
  - ADA

IAS

- OpenCL, Cuda, PyOpenCL, PyCuda
  - 3 personnes
  - 2 codes solaire
- MPI : 2 équipes et 2 codes
  - 1 développement local run US (gros calculateur)
  - 1 développement et run IAS
- OpenMP : 1 equipe
  - code basé sur CAMP
- *Matériel acquis récemment*
  - 2 nœuds R720, 16 cœurs et 256 Go.
  - 2 nœuds R720, 16 cœurs et 256 Go **+ K20**
  - 2 nœuds R720, 16 cœurs et 256 Go **+ PHI**
  - 4 nœuds R820, 32 cœurs et 512 Go.
  - Scheduler en cours d'installation.

IMNC

- *hGATE*

IPNO

- *4 personnes impliquées*
- *Projets*
  - Geant4
  - O² : fusion des systèmes Online-Offline de la DAQ ALICE au CERN.
  - Optimisation/parallélisation de codes internes existants.
  - Exploitation d'un cluster MPI/HPC (utilisé avec les codes: VASP, CPMD, CP2K, MNCP, ...).
  - Evaluation à venir des GPU et de CUDA.
  - NARVAL.
- *Compétences et savoir-faire présents*
  - C/C++ & Threads Posix utilisés par Geant4
  - CoArray Fortran (MPI masqué)
  - Tasking Ada & Distributed Ada (Annexe E)
  - OpenMP
- *Technologies à surveiller/développer*
  - OpenCL, OpenACC, MPI.

IRFU

- *SPP*
- *SAP*
  - COAST
    - 1200 cœurs pour la simulation en astrophysique.
    - cluster OpenGL pour la visualisation en astrophysique.
- *SPhN*
  - calcul abinitio de la structure et des reactions nucléaires à basse énergie.
  - calculs sur réseaux QCD (structure > noyau).
  - études sur la physique des petits noyaux.
  - …
- *SACM*
  - 256 cœurs pour simulation d'accélérateurs.

LAL

- *Projets*
  - Dans le cadre de LPaSo
    - Parallélisme de tâches avec GaudiHive (ATLAS, LHCb).
    - Déclenchement haut-niveau avec GPUs (LHCb).
  - R&D optimisation/parallelisation tracking ATLAS.
  - Dans le cadre de GridCL
    - Portage many-core de RooFit
  - Geant4
- *Compétences et savoir-faire présents*
  - Multi-Threading
  - Hadoop
  - Go, Clojure
- *Technologies à surveiller/développer*
  - Go

LPT

- *~6 personnes impliquées*
- *Projets*
  - QIRAL (générateur de code parallèle à partir du langage mathématique)
  - ETMC (collaboration européenne Twisted Mass)
  - Simulation QCD sur réseau (Multi-threading, CUDA, MPI…)
  - Planck (Python parallele)
- *Partenariats*
  - QIRAL (**LPT, LAL**, INRIA Bordeau...)
  - ETMC (14 partenaires dont DESY Zeuthen)
  - Alpha (9 partenaires dont DESY Zeuthen)
  - FlaNPP (saveurs Neutrinos et Quarks) : proposition à l'ANR.
- *Compétences et savoir-faire présents*
  - Machines APE (Array Processor Experiment, langage Tao)
  - QIRAL (générateur de code parallèle à partir du langage mathématique)
- *Technologies à surveiller/développer*
  - The Maude System

LLR

- *~10 personnes impliquées*
- *Projets*
  - Pour Galop : simulation de l'accélération d'électrons par laser (avec code PIC MPI).
  - Dans le contexte de GridCL
    - Pour CMS ions lourds : reconstruction de traces (avec accélérateurs).
    - Pour CMS haute-luminosité : la méthode des éléments de matrice (avec accélérateurs).
    - Pour CTA : traitement des signaux de télescope (avec accélérateurs).
  - Pour ILC : (re)construction d'évènements en ligne (à venir, avec matériel many-core).
  - Geant4
- *Compétences et savoir-faire présents*
  - OpenCL, OpenMP, MPI, Hadoop.
- *Technologies à surveiller/développer*
  - OpenACC, C++11, LLVM, CUDA 6.