

GPUコンピューティング (CUDA) 講習会

GPUとGPUを用いた計算の概要

丸山 直也

スケジュール

- **13:20-13:50 「GPUを用いた計算の概要」**
 - 担当 丸山
- **13:50-14:30 「GPUコンピューティングによるHPCアプリケーションの高速化の事例紹介」**
 - 担当 青木
- **14:30-14:40 休憩**
- **14:40-17:00 「CUDAプログラミングの基礎」**
 - 担当 丸山

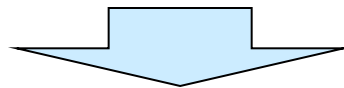
GPUコンピューティング

- GPUを一般アプリケーションの高速化に適用
 - GPUを計算アクセラレータとして利用
- GPGPU (General-Purpose Computing on GPU)とも言われる
- 2000年前半から研究としては存在。2007年にCUDAがリリースされてから大きな注目



計算加速器 (アクセラレータ)

- Cell, GPU, GRAPE, ClearSpeed, FPGA, ...
- 汎用CPUとは別に特定の計算のオフロードが可能なプロセッサ
- 汎用CPUと比較して高性能and/or低消費電力
- HPCではベクトル演算に特化したアクセラレータが注目



- ハイブリッドコンピューティング
 - 汎用CPUとアクセラレータの組み合わせ
 - HPCにおける最近の最もホットなトピックの1つ

例：Roadrunner at LANL

- Opteron + *PowerXCell 8i*
- 史上初ペタフロップ超えマシン
 - 1.105 PFLOPS (LINPACK)
- 現在世界最速スパコン
 - 2008年6月より2009年6月までTOP500ランキングにて1位



2009/11/25

例: TSUBAME @ 東工大GSIC

- Opteron (> 10K cores) + ClearSpeed (> 600) + NVIDIA Tesla (> 600)
- Peak: 170 TFLOPS (DP), Linpack: 87.01 TFLOPS (56th at Nov '09 TOP500)

Tesla S1070

- 4 Tesla cards in a 1U node
- Connected to host machines via PCIe extension cables



TSUBAME 1.2. The most Heterogeneous Supercomputer in the world

- Three node configurations with four different processors → **>30,000 cores, ~170TFlops system**



SunFire X4600+ 2 TESLAs + ClearSpeed

- Opteron 2.4GHz 16 cores
 - TESLA S1070 (30cores) 2boards
 - ClearSpeed X620 (2cores) 1board
- 78 cores, 330 Gflops peak
- x 318nodes



SunFire X4600+ClearSpeed

- Opteron 2.4GHz 16 cores
 - ClearSpeed X620 (2cores) 1board
- 18 cores, 157 Gflops peak
- x 330nodes

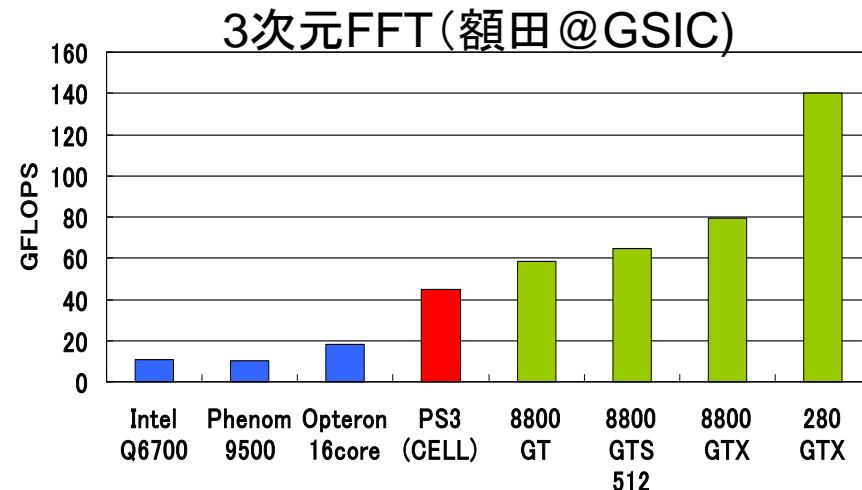


SunBlade X6250 (TSUBASA cluster)

- Xeon 2.83GHz 8 cores
- 8 cores, 90.7 Gflops peak
- x 90nodes

なぜGPU？

- CPUを大幅に上回る計算性能＋メモリバンド幅
 - Tesla
 - 1 TFOPS (SP) / 90 GFOPS (DP)
 - 100 GB/s
 - Core 2 Quad @ 3 GHz
 - 96 GFLOPS (SP) / 48 GFOPS (DP)
 - < 10 GB/s
- 物理シミュレーションなどの多くのデータ並列なアプリ



GPUコンピューティング

- GPUを一般アプリケーションの高速化に適用
 - GPU→アクセラレータと呼ばれるものの一種
- GPGPU (General-Purpose Computing on GPU)とも言われる
- 2000年前半から研究としては存在。2007年にCUDAがリリースされてから大きな注目

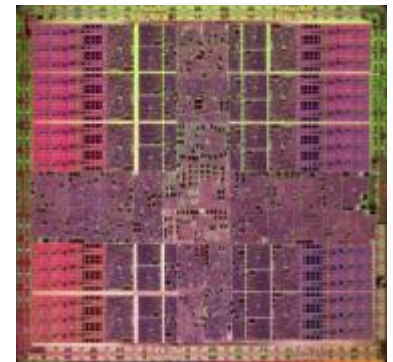


GPUコンピューティング：ハードウェア

- NVIDIA GPU
 - GeForceシリーズ： 一般のPCに搭載されているタイプで、比較的安価。GeForce 8800 GTXよりCUDAを実行可能
 - Teslaシリーズ： GPUコンピューティング専用ハードウェア（ディスプレイ出力無し）。高価だがより高信頼（といわれている）。TSUBAMEに搭載
- AMD/ATI GPU
 - Radeonシリーズ
 - FireStreamシリーズ

Tesla 10 (T10)

- NVIDIA G200系アーキテクチャによるHPC向けプロセッサ
 - コンシューマ向け → GeForce 280 GTX
- 240コア @ 1.29-1.44 GHz
- 4GB memory, 102 GB/s
- Peak: 1 TFLOPS (SP), 90 GFLOPS (DP)
- 製品
 - Tesla C1060: PCIe card
 - Tesla S1070: 1U system with 4 C1060 cards
- GeForceとの違い
 - ビデオ出力無し
 - 品質 (NVIDIAによる全品検査 vs ボードメーカーによるサンプル検査)
 - 価格 (C1060 @ \$1,700, GTX 280 @ \$400)



NVIDIA T10

Fermi

- NVIDIAの次世代GPUアーキテクチャ
- Tesla系は2010年春ごろから出荷？
GeForce系はもっと早い？
- 新機能 & 改善点
 - 倍精度性能の改善（単精度の半分）
 - 16SM, 512コア
 - ECCサポート
 - C++サポート



Tesla 20 シリーズ

- Tesla C2050 & C2070
 - グラフィックスボード型
 - 520 GFLOPS – 630 GFLOPS
 - 3GB – 6GB
 - \$2499 & \$3999
- Tesla S2050 & S2070
 - 4枚のGPUボードを搭載した1Uシステム
 - 2.1 TFLOPS – 2.5 TFLOPS
 - 12GB – 24GB
 - \$12995 & \$18995

GPUコンピューティング：ソフトウェア

- NVIDIA CUDA
 - 2007年2月にNVIDIAが自社のGPU向けにリリース
 - C/C++の言語拡張
 - NVIDIAのGPU専用
 - 最も普及
- OpenCL
 - Appleによる提案に始まり、標準化団体により制定
 - 言語自体はベンダー非依存
 - Snow Leopardに標準搭載
 - NVIDIA/ATI GPU、x86 CPU向けSDKが利用可能
 - 普及はまだこれから
- その他
 - Brook/Brook+, RapidMind, DirectX Compute, etc.

OpenCL

- Khronos 標準化団体によって策定されたGPUコンピューティングのための共通仕様
 - KhronosはOpenGLを策定した団体
- 同一のソースコードでNVIDIA GPU、AMD GPU、Intel/AMD CPUなどで動作
- ただし、同一のプログラムがすべての環境に適しているわけではない
 - オンチップメモリのサイズ、ベクター長など
 - 特にベンダーの異なるGPU間では結局異なるプログラムを書くことに
- 本講習会で取り上げる内容は基本的にOpenCLプログラミングでも有効
 - 細かな技術的な違いはあるものの、概念的なレベルでは同じ

OpenCL vs. CUDA



- OpenCLの利点
 - ベンダー非依存
 - NVIDIA/ATI GPU以外でも動作
 - Cell, Intel Larrabee?
 - 業界標準
- CUDA の利点
 - NVIDIA GPUの最先端の機能を利用可能
 - CUDA3.0ではデバッガー等の機能も大幅に拡充
 - cf. OpenGL vs. DirectX
 - プログラミングの簡便さ
 - OpenCLはCUDAのドライバーAPIと同様にカーネル呼び出し等が煩雑
 - これまでの知識・経験の蓄積 (OpenCLも普及すれば時間の問題)

DirectCompute

- Microsoft独自のDirectX 11に追加されたGPUコンピューティングのための仕様
 - OpenGL vs. DirectX
 - OpenCL vs. DirectCompute
- DirectX 11をサポートするGPU+Windowsで利用可能
- CUDA、OpenCLとの違い
 - DirectXとよりシンプルかつ密な連携が可能(例えば3Dゲーム中のAIの計算などをGPUで高速化など)
 - Windowsのみだが、WindowsのDirectXにおける開発に慣れた人であればハードルはCUDA等より低い

おまけ

- SC09でのNVIDIA関連の講演等のビデオ
 - <http://nvidia.fullviewmedia.com/sc09/agenda.html>

