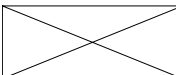


GPGPU技術の線形計算 への展開

A01 高橋班

電気通信大学情報理工学研究科

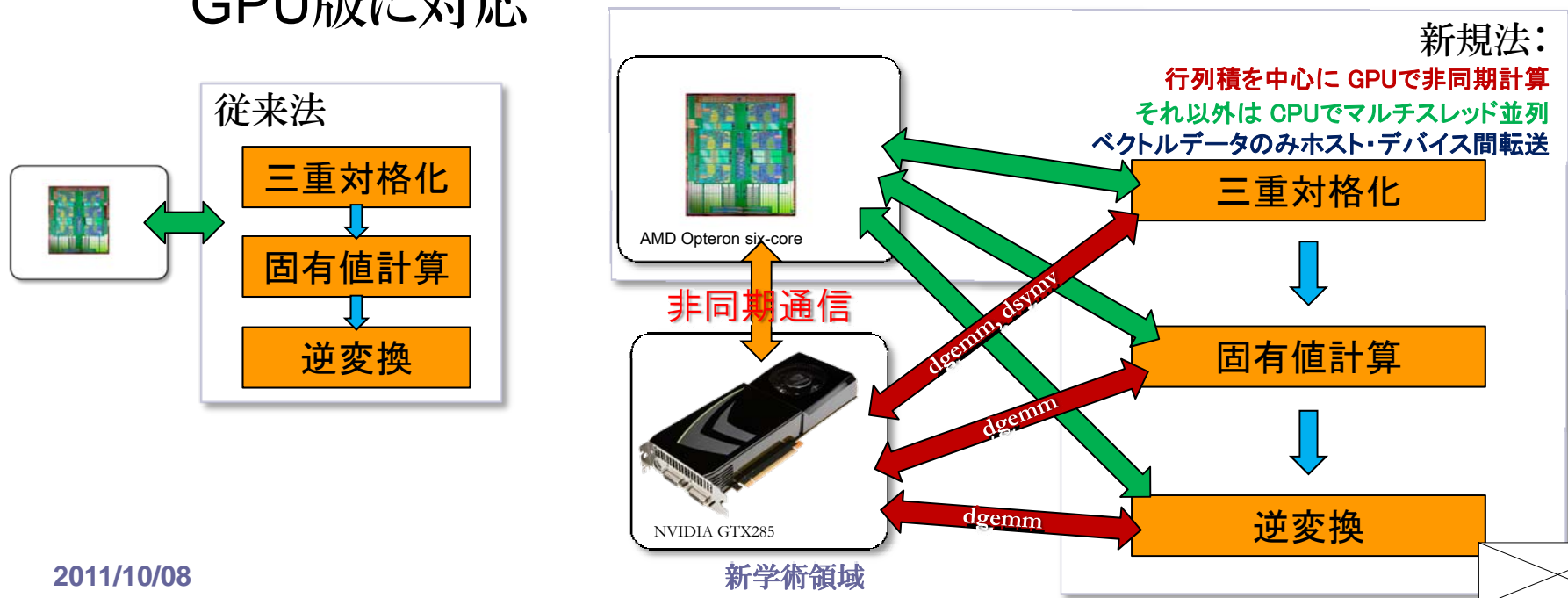
今村俊幸



H22年度-H23年度の研究から

GPU環境での線形計算ソフトの研究

- GPUで倍精度計算が可能になり, NVIDIA社Fermiコアで倍精度による応用計算が現実的に
- NVIDIA社のCUDA, CUBLASを利用することで、行列演算が容易にGPU化可能
- 本研究領域で重要性が高い固有値ソルバのマルチコア+GPU版に対応



関連研究

□ CUDA-BLAS

- NVIDIA社が提供するCUDA環境-BLAS。
- 基本中の基本ソフト BLAS関数を容易に置き換え可能

□ CULA

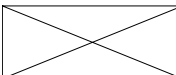
- 商用ソフト
- LAPACKの一部関数をサポート
- CUDA-BLASの独自最適化→自身のWhite Paperでは高性能をうたっている

□ MAGMA

- LAPACK, ScaLAPACKを開発したテネシー大学が開発中
- 新しいスレッド管理方式により高並列高性能をうたっている。
- サポート関数は一部であるが、それらは非常に高速

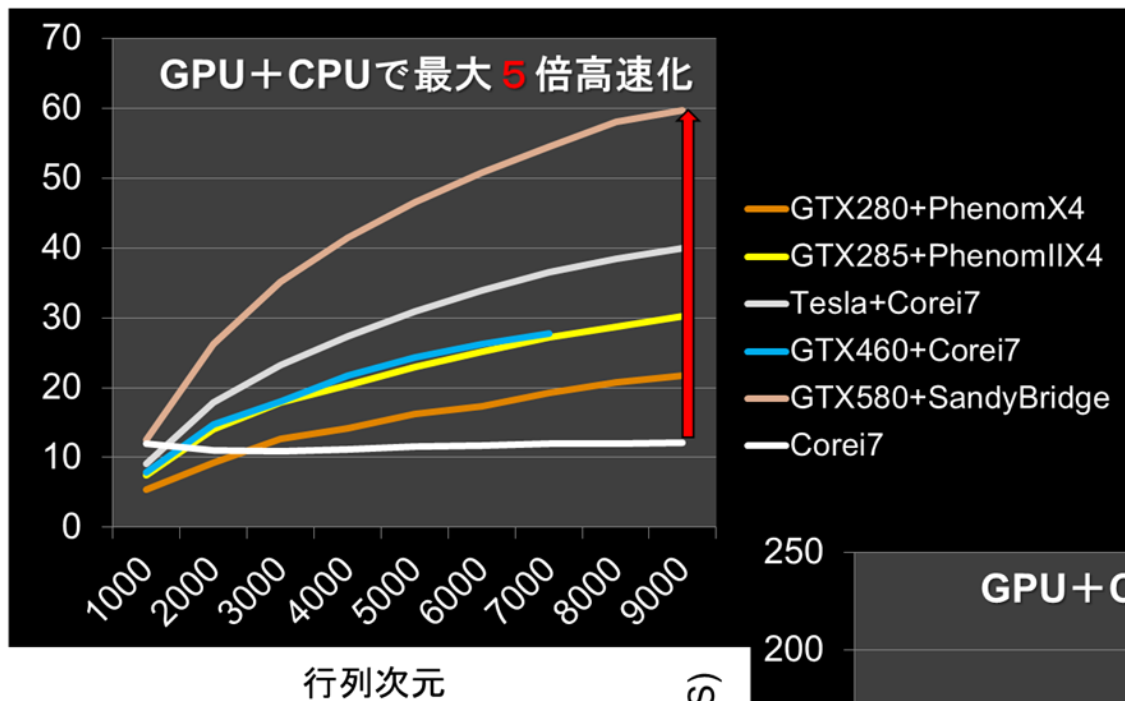
□ 今回の開発は:

- 固有値ソルバ(eigen-G)、スタンドアロン環境に限定して

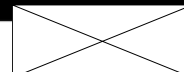
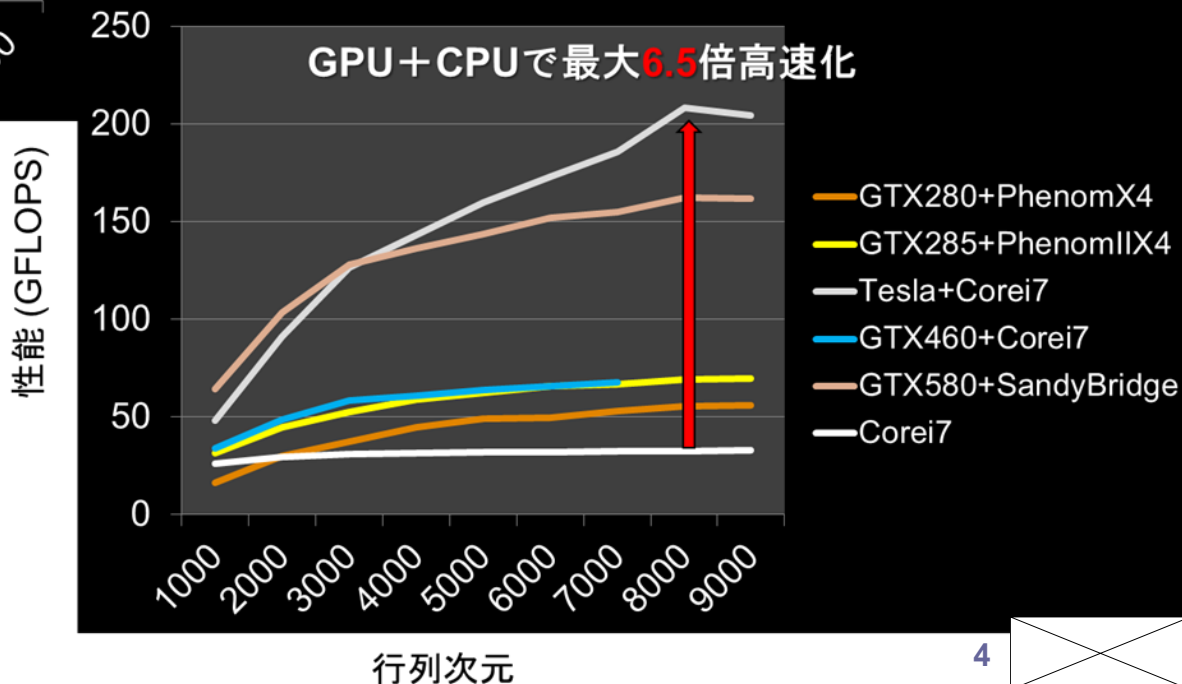


固有値ソルバ[eigen_G]

性能 (GFLOPS)



- 三重対格化(左)
- 逆変換(右)



チューニングの結果 (CPU+GPU)

□ QuadCore CPU+GPU環境での固有値ソルバの性能

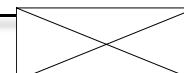
■ 上段:完全対格化に要する時間(秒), (LAPACK dsyevd相当)

■ 下段:計算性能(FLOPS)

CPUよりも2~3倍高性能

固有値計算に限定すれば **TeslaC2050 < GTX580**

	1000	2000	3000	4000	5000	6000	7000	8000
GTX285	0.379	1.596	4.080	8.267	14.60	23.02	34.37	48.35
/Phenom II X4	8.79G	16.7G	22.0G	25.8G	28.5G	31.2G	33.1G	35.2G
GTX460	0.349	1.492	3.971	7.989	14.19	22.85	31.85	N.A.
/Corei7	9.52G	17.8G	22.6G	26.7G	29.3G	31.4G	35.8G	---
TeslaC2050	0.294	1.160	2.865	5.559	9.871	15.19	22.64	30.79
/Coire7	11.3G	22.9G	31.4G	38.3G	42.2G	47.3G	50.4G	55.4G
GTX580	0.226	0.864	2.077	4.126	7.333	11.45	17.18	23.75
/SandyBridge	14.7G	30.8G	43.3G	51.6G	56.8G	62.8G	66.5G	71.8G
Only	0.144	1.22	---	9.28	---	---	---	67.4
SandyBridge	23.0G	21.7G	---	22.9G	---	---	---	25.2G



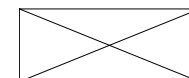
参考比較 (フリーソフトMAGMA1.0)

□ 倍精度完全対角化(dsyevd)

	1000	2000	3000	4000	5000	6000	7000	8000
TeslaC2050 /Coire7	.5277	5.607	3.307	7.330	12.78	20.87	35.50	45.33
GTX580 /SandyBridge	.4010	2.428	2.514	5.207	9.482	15.54	23.93	35.04
SandyBridge MKL11.1	.1454	1.025	3.677	8.789	17.33	29.73	47.28	69.71

※倍精度ソルバーとして良い性能。Eigen_GはMAGMAより30%程度高速な場合がある。

	1000	2000	3000	4000	5000	6000	7000	8000
TeslaC2050 /Coire7	0.294	1.160	2.865	5.559	9.871	15.19	22.64	30.79
	11.3G	22.9G	31.4G	38.3G	42.2G	47.3G	50.4G	55.4G
GTX580 /SandyBridge	0.226	0.864	2.077	4.126	7.333	11.45	17.18	23.75
	14.7G	30.8G	43.3G	51.6G	56.8G	62.8G	66.5G	71.8G



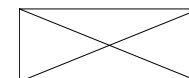
参考比較 (商用ソフトCULA)

- 倍精度完全対角化(dsyev): 【注意】 eigen-Gと異なるアルゴリズム(QR法)

	1000	2000	3000	4000	5000	6000	7000	8000
TeslaC2050 /Coire7	.9338	3.521	8.591	16.12	26.56	40.19	57.70	112.3
GTX580 /SandyBridge	.6705	2.679	6.264	11.75	19.53	29.86	43.35	60.22
SandyBridge MKL11.1	.3803	2.920	9.788	21.82	40.88	68.78	106.7	158.7

※倍精度ソルバーとして期待したほどの性能でないが、CPU版LAPACKよりも速い

	1000	2000	3000	4000	5000	6000	7000	8000
TeslaC2050 /Coire7	0.294	1.160	2.865	5.559	9.871	15.19	22.64	30.79
	11.3G	22.9G	31.4G	38.3G	42.2G	47.3G	50.4G	55.4G
GTX580 /SandyBridge	0.226	0.864	2.077	4.126	7.333	11.45	17.18	23.75
	14.7G	30.8G	43.3G	51.6G	56.8G	62.8G	66.5G	71.8G



まとめ

- Multi-{core,socket,GPU}向け線形数値計算ソフトウェアの最適実装・チューニング技術の研究
 - マルチコアCPU+GPUで固有値ソルバの高速化
 - 2000次元の対格化が1秒以下,4000次元が4秒程度
 - 過去の中規模問題は小規模問題となり、反復解法の内部に利用してもpayするレベル。
- 今後の課題：高性能数値計算ソフトウェアの応用
 - 他班のシミュレーションコードへの応用を視野に
 - アプリケーションサイドからの要求に応える観点から
 - スタンドアロン環境向けに限定したチューニング
 - 例えば100~1000次元の行列対角化を徹底的に最適化
 - 上記を達成するための汎用技術の開発
 - BLAS単位以外のCUDA並列化実装

