

CUDAを用いたGPUによるフラグメント分子軌道法の高速化

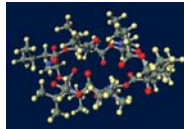
利用課題責任者 株式会社クロスアビリティ 古賀良太

背景 Background

1. 近年、主にハードウェアに関する計算機環境の目覚ましい変化により、タンパク質をはじめとする巨大分子に対して計算化学的手法が適用できるようになってきた。
2. High Performance Computingを指向したグラフィックカードを用いるGeneral Purpose GPU(GPGPU)計算が注目されており、NVIDIA社によるCUDA環境を用いて各分野で相当の高速化が達成されている。
3. 第一原理の量子化学計算においても、CUDA環境による高速化の研究例が示されている。

本研究の目的 Purpose of this Research

Yasuda(2008)の研究成果をもとに、FMO高速化に必須であるCUDAを用いた各種量子化学計算に結合可能なGPGPU対応の汎用モジュールを開発し、パフォーマンス測定する。



実装方針 Plan of Coding

1. 電子状態計算のエネルギーおよびその勾配を計算する。
2. GPUの加速対象は、計算コストの高い近接クーロン項および交換行列、格子点上の電子密度、密度汎関数の基底空間への変換とエネルギー勾配計算における二電子積分の部分に限り、その他の部分はホストのCPUで計算する。
3. GPU上では高速な単精度計算のみ行う。その上で十分な計算精度を維持できるようにする。
4. ノード内マルチコアCPU、ノード内マルチGPU、ノード間並列計算に対応する。
5. 非同期でCPUとGPUの計算を同時実行し、計算資源を効率よく使用する。
6. 汎用モジュール"XA-CUDA-QM"として、他の量子化学計算ソフトウェアに組み込み可能な形式で開発する。

計算アルゴリズム Algorithms

二電子積分:

クーロン行列の計算はDirect J Engine法と高速多重極子法(FMM)を用いることとし、計算コストの高いJ Engineでの近接クーロン項の計算をGPGPUで高速化する。

primitive Hermite Gaussianを使うことによって、計算速度を落とさず、メモリ使用量を大幅に削減できる。

$$J_{ab} = \sum_{cd} (ab|cd) D_{cd} = \sum_p E_{ab}^p \sum_q (p|q) D_q$$

ホストメモリとの通信は、 p, q, D_q, J_p であり、計算量に比べて小さい。

積分の計算は、メモリ使用量の少ないGauss-Rys求積法を用いた。

$p \leftrightarrow q$ の積分対称性は使わない。

計算精度のために積分値をシュワルツ不等式により評価し、大きな積分値を持つ項のみCPUで倍精度計算する。CPUで計算する必要のある項は10%以下であり、GPUと並列に計算されるため、計算時間のロスはない。

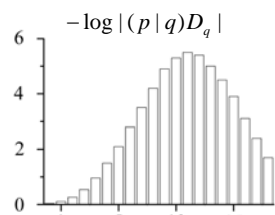


図1. Valinomycinでの項数 ($\times 10^8$, 収束D使用)

なお、手元のマシンで行った2010年6月11日現在の実装では、ハートリー・フォック交換行列(K行列)を積分対称性を活用して行っており、Fermi Core GPU(GTX470)のL1/L2 cacheを活用すべくデバイスメモリアクセスを頻用して高速化を図っている。

交換相関項:

- DFT法の交換相関項の計算は、
1. Grid上の電子密度とその勾配の計算 $\rho(r_i) = \sum_k D_{kl} \chi_k(r_i) \chi_l(r_i)$
 2. Grid上の汎関数値 f_i, g_i の計算
 3. 交換相関行列の計算 $v_{kl} = \sum_i \{f_i \chi_k(r_i) + g_i \cdot \nabla \chi_k(r_i)\} \chi_l(r_i)$

からなる。このうち計算コストの大きい1と3をGPUで加速する。

ホストメモリとの通信は、 χ, D, ρ, f, g, v であり、計算量に比べて小さい。

さらに、Grid上で10以上の値を持つ χ のみ使うことによって、精度を落とさず、約2倍の加速をした。

エネルギー勾配:

エネルギー勾配計算においては二電子積分の計算が70~80%を占める。計算方法はSCFの二電子積分の計算方法と同じであるが、角運動量子数が1つ大きいGPUカーネルが必要となる。

結果 Results

【ベンチマーク環境】

Valinomycinのエネルギー勾配全体計算をTSUBAMEで評価した。

CPU : AMD Opteron 2.4 GHz 4 core

GPU : Nvidia Tesla C1060 x 2

PGI fortran / CUDA 2.2 / ATLASでGaussianをコンパイル

表1. TSUBAMEによる評価 (修正test397)

	計算時間[秒]	収束エネルギー[a.u.]
Gaussian 03 rev.B.01	1159.09	-3772.609959
Gaussian + XA-CUDA-QM (2GPU)	407.52	-3772.609077

密度汎関数はPBE/PBE、3-21G基底を使用

GAMESS+XA-CUDA-QMを結合したものをGPGPUクラスターで評価し、1フラグメント1残基の問題サイズが小さい場合においてもCPUマルチコアに比して、もしくは併せて高速化させる必要がある。なお、TSUBAME利用後ではあるが、手元のマシン(Core2Quad, + GTX470)で2010年6月11日時点のK行列高速化の結果を表2に示す。Fermi CoreのL1/L2 cacheを活用した。

表2 : K行列の計算時間(1CPU vs 1GPU, 単位: 秒)

Software Processor	GAMESS Core 2	GAMESS Xeon	GAMESS + XA-CUDA-QM GTX470(Fermi)
Glycine	0.161	0.171	0.013
Glutamine	1.425	1.44	0.098
Tryptphan	4.415	4.393	0.275

Summary

GPGPUを用いた汎用量子化学計算モジュールを開発し、CPUでは最速なアルゴリズムを採用している著名な量子化学計算ソフトウェアに結合した。これによりエネルギー勾配計算全体の2~3倍の高速化成功し、FMO on GPUの加速に目処を立てた。

今後の課題 Future Theme

1. d軌道以上の高角運動量子数の基底を含む二電子積分の計算
2. 精度向上のためのCPU, GPUの使い分け
3. 複数フラグメントのCPU + GPUによる同時処理
4. 二次摂動項のGPU高速化
5. FMO環境静電ポテンシャル計算のGPU高速化
6. TSUBAME 2.0によるFermi Core GPU (CUDA 3.0, Tesla C2050/C2070) での大規模評価