2016 年 2 月 4 日

平成 27 年度　情報・通信工学科　コンピュータサイエンスコース　卒業研究発表会

DVFS 手法による GPGPU の低消費電力化

情報・通信工学科　学籍番号:1211023　成見研究室　上原衛

1　はじめに

近年、GPU の演算性能の飛躍的な向上や消費電力あたりの演算性能が高いことから、GPU を従来のグラフィック処理以外の汎用的な計算に応用する GPU コンピューティングが広く普及してきた。しかし,GPU コンピューティングの普及に伴い、より高い演算性能をもつ計算機が必要になっている.

単純に GPU のコア数を増やすという方法だと GPU が必要とする膨大な消費電力の問題によって、性能は向上できても現実的に実現が難しくなっているという問題がある. その他の方法として、GPU 上でのメモリアクセス手法などを実行プログラム側で工夫する方法では、プログラムの作成に高い技術力が必要という問題がある. 他に有効な手法として DVFS（Dynamic Voltage and Frequency System）という GPU コアやメモリの動作周波数や供給電圧を変更することで無駄な消費電力のみを削減するという方法がある.

本研究では,GPU が本来持っている高い演算性能を犠牲にすることなく、動作するプログラムの特徴によって発生する無駄な消費電力のみを削減する DVFS 手法を用いることで、プログラム実行時の電力効率の向上を目指す。

2　既存研究

Mei.X らは, 演算性能の低下が供給電圧と動作周波数を変更していない場合と比較し 4%以下になる範囲で、消費電力が最も削減できる最適な供給電圧と動作周波数の値を計測し、37 個のベンチマークプログラムにおいて、最適な供給電圧と動作周波数を手動で設定することで平均 19.28%の消費電力の削減に成功している [3]。本研究とは演算性能を維持しつつ低消費電力化をおこなうという主旨は本研究と一致しているが、最適な電圧と周波数の予測と GPU への設定を自動的に行う点で異なる。

L.Tang らは、プログラムの実行状況や GPU の供給電圧、動作周波数、消費電力などの値を計測機器を駆使することで測定し、プログラムの実行状況や GPU のコアとメモリの供給電圧と動作周波数の関係を分析したモデルを作成することで、そのモデルを利用した DVFS システムの開発を行っている [4]。本研究とは、開発する DVFS システムが GPU ドライバから得られる情報のみを用いている点で異なる。

3　システムの概要

本研究で開発した CUDA DVFS システムは、CUDA モニタープログラムとシステムマネージャ、GPU コントローラから成り立つ。そのシステム全体の概要を図 1 に示す。

CUDA モニタープログラムの実体は、演算性能の測定機能と CUDA DVFS システムとの通信機能が実装されたヘッダファイルである。このヘッダファイルを既存の CUDA プログラムへ手動で導入し、必要な機能を追加することで CUDA DVFS システムを簡単に利用することが可能となる。

システムマネージャと GPU コントローラは、CUDA モニタープログラムを経由して CUDA プログラムから送信されてくるプログラムの実行開始通知や演算性能の測定データを受信するアプリケーションである。システムマネージャは受信した実行開始通知によって GPU コントローラの制御・管理を行い、GPU コントローラは測定データと現在の GPU の調整状況 (フェーズ) に応じて、GPU の供給電圧と動作周波数を最適な組み合わせへ変更する。動作周波数の変更は NVIDIA Inspector[5] を用いて行う。

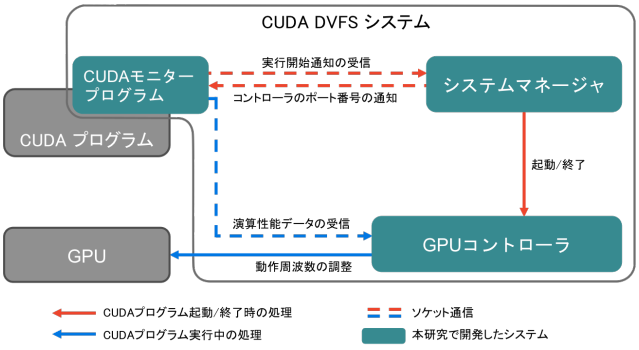

図 1: システムの概略図

表 1: nbody における消費電力削減手法の比較

	消費電力 [W]	演算性能 [回/s]	電力効率 [回/W]	電力効率比
動作周波数最大	121	2.47	0.020413	1.000000
GPU ドライバ	115	2.37	0.020608	1.009553
CUDA DVFS システム	105	2.26	0.021524	1.054426

表 2: particles における消費電力削減手法の比較

	消費電力 [W]	演算性能 [回/s]	電力効率 [回/W]	電力効率比
動作周波数最大	108	25.04	0.231852	1.000000
GPU ドライバ	94	25.08	0.266809	1.150773
CUDA DVFS システム	83	23.92	0.288193	1.243004

4 DVFS 手法の詳細

CUDA DVFS システムの GPU コントローラでは、CUDA モニタープログラムから受信した演算性能の測定データとフェーズの状況により、GPU の供給電圧と動作周波数を最適な値へ変更する。フェーズはモニタリング開始、最大演算性能の測定、コアとメモリの動作周波数の調整、演算性能の監視、モニタリング終了の 5 段階のフェーズによって管理されている (図 2)。

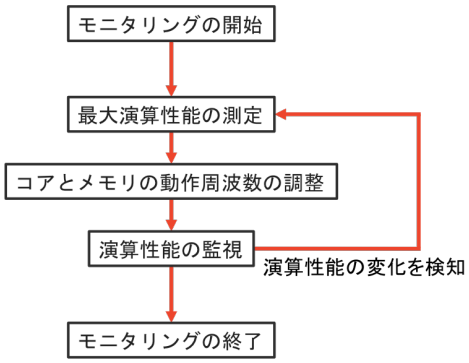

図 2: フェーズ管理の仕組み

5 システムの評価

本研究で開発した CUDA DVFS システムの評価実験として、GPU ドライバ標準の低消費電力化手法と DVFS システムの低消費電力化手法でそれぞれ低消費電力化を行い、CUDA プログラムをコアとメモリの最大動作周波数で実行した時に比べて、どちらの手法がより電力効率を向上させることができたかを検証した。この実験には、N 体シミュレーション nbody と粒子法シミュレーション Particles の 2 つの CUDA サンプルプログラムを使用し、その結果を表 1、表 2 に示す。

この結果から、CUDA DVFS システムでは GPU ドライバに比べて nbody では約 5.73 倍、Particles では約 1.61 倍の電力効率の向上が可能であることが確認でき、本研究で開発した CUDA DVFS システムは電力効率の向上性能において GPU ドライバよりも優れたシステムであることがわかった。

6 まとめと今後の課題

本研究では実行中の CUDA プログラムの演算性能をリアルタイムに監視し、最も電力効率が高い状態でプログラムの実行が可能な GPU のコアとメモリの動作周波数の組み合わせを発見するシステム、CUDA DVFS システムを開発した。しかし CUDA DVFS システムでは、既存の CUDA プログラムにいくつかの機能を追加する必要があることから、5%から 10%の性能低下があることや、低消費電力化を行うために要する時間が、Particles では 30.90s、nbody では 103.42s と非常に長い点が本研究の大きな課題となった。

参考文献

[1] “TSUBAME とは — [GSIC] 東京工業大学学術国際情報センター”,http://www.gsic.titech.ac.jp/tsubame, (最終アクセス日 2016 年 1 月 30 日)

[2] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton,“ImageNet Classification with Deep Convolutional Neural Networks”,Neural Information Processing Systems 2012, Lake Tahoe, 2012

[3] Mei,X., Yung,L.S., Zhao,K., Chu,X.,“A measurement study of GPU DVFS on energy conservation”,Workshop on Power-Aware Computing and Systems, HotPower'13, Pennsylvania, 2013

[4] L. Tang and Y.Zhang, “Low-Power Task Scheduling for GPU Energy Reduction“,http://netscale.cse.nd.edu/cms/pub/Edu/GradOSF11ProjProposal/Low-power Task Scheduling for GPU Energy Reduction.pdf,(最終アクセス日 2016 年 1 月 30 日)

[5] “Guide/NVIDIA Inspector - S.T.E.P. Project Wiki “, http://wiki.step-project.com/Guide:NVIDIA_Inspector, (最終アクセス日 2016 年 1 月 30 日)