科学研究費助成事業　　研究成果報告書

平成　２６　年　　６　月　　６　日現在

機関番号：　１２６０８
研究種目：　挑戦的萌芽研究
研究期間：　2011～2013
課題番号：　２３６５００１２

研究課題名（和文）高度なＧＰＵプログラミング手法の開拓

研究課題名（英文）Development of novel GPU programming techniques

研究代表者
額田　彰（NUKADA, AKIRA）

東京工業大学・学術国際情報センター・特任准教授

研究者番号：４０５４５６８８

交付決定額（研究期間全体）：（直接経費）　　2,800,000円、（間接経費）　　840,000円

研究成果の概要（和文）：2006年にNVIDIA社が汎用計算向けのGPU環境としてCUDAを公開して以来、長い計算時間を要する多くの処理がGPUに移植され高速化を実現してきた。スーパーコンピュータを用いた大規模な計算ではGPU間のデータ転送速度が特に重要になる。中でも高速フーリエ変換は特にこのGPU間の転送量が多く、通信パターンも全対全と効率が低下しやすい。スーパーコンピュータTSUBAME2.0においてホストとデバイス間のPCI-Express転送やノード間のInfiniBandネットワーク通信を適切にスケジューリングする手法を提案し、多数のユーザのジョブ間で共有されるネットワーク網の性能を引き出すことに成功した。

研究成果の概要（英文）：In 2006, NVIDIA presented CUDA as GPU computing environment for genericl computations, and after that many time-consuming applications are ported to GPU and achieved extreme speed-ups. For large-scale computation using supercomputers, efficient data transfer between GPUs is the most important. FFT is used in many scientific simulations and requires all-to-all communications between GPUs. We presented a novel scheduling techniques of PCI-Express data transfer between host and GPU, and InfiniBand data transfers between nodes. As a result, we could achieve scalable performance even if many other users' jobs are running on the shared network.

研究分野：　総合領域

科研費の分科・細目：　情報学・ソフトウェア

キーワード：　並列処理・分散処理　ＧＰＧＰＵ

１．研究開始当初の背景

GPGPU と呼ばれる GPU を用いた汎用計算はここ数年で急激に普及してきた．特に NVIDIA 社が CUDA という新しい GPU アーキテクチャとそのソフトウェア開発環境を提供してから多くのユーザが容易に GPU 用のプログラムを開発できるようになったためである．GPU の演算性能やメモリバンド幅は CPU のそれらを遥かに上回り，特に実行時間が長いシミュレーション等の計算を対象に GPU を用いた高速化が実現されている．GPU は比較的安価であるため手軽に導入することができ，また電力効率に優れるために今後主要な計算資源として用いられることは確実視されている．

２．研究の目的

ここ数年で GPU を用いた汎用計算は急激に普及してきたが，今年になって研究分野としては減速したように見える．多くのユーザはそれほど高いプログラミング技術を持っていないため，簡単に GPU に移植可能な計算のみが対象となっている面もある．しかしそれ以上に GPU の潜在的な機能を過小評価している感がある．本研究ではこのような GPU の一歩進んだ活用技術を開拓し，GPU でより広範囲なアプリケーションの高速化を可能にする．

３．研究の方法

GPU 計算の一番の問題点は GPU に搭載されるメモリの容量がホストのそれと比べて極めて少ないことである。多くのアプリケーションを GPU に移植しようとしても GPU のメモリ不足で実現に至らなかったり、またはホストと GPU 間で何度もデータのやり取りをすることになる場合も多い。
GPU に搭載されるメモリの容量を増やすことはその製造上難しい。そこで使用する GPU の数を増やすことが通常の選択である。その結果、GPU 間のデータ転送が必要になる。単一のノードに接続可能な GPU 数は限られるため、複数ノードを利用することになり、単純なノード間データ転送に加えて、ホストと GPU 間のデータ転送と多段階のデータ転送が必要になる。これらを効率よくスケジューリングする手法の研究を行う。

４．研究成果

高速フーリエ変換（ＦＦＴ）は現在に至るまで数々のアプリケーションで用いられている計算の一つである。中でも３次元ＦＦＴは大規模シミュレーションに用いられることが多く、その高速化はとても大きな意義を持つ。昨年まで、NVIDIA や AMD などの GPU を対象に高速なアルゴリズムを提案しており、CPU と比べて何倍から何十倍の高速化を実現している。
実際にアプリケーションで用いる場合には複数の GPU を用いる場合がある。その理由には主に二つある。一つ目の理由はもちろん複数の GPU を用いることによるさらなる高速化である。二つ目の理由はデバイスメモリ容量の確保である。単体の GPU のメモリ容量はホスト CPU と比べると少なく、大規模アプリケーションを実行するには十分でないことが多い。
GPU のデバイスメモリはそれぞれ独立しており、複数 GPU を用いた計算では通常デバイスメモリ間でのデータ転送が必要になる。複数 GPU によるＦＦＴ計算の場合には、このデータ転送パターンが全ＧＰＵ間の全対全通信という非常に性能が出しにくいものになる。

TSUBAME 2.0 ではノード間は Fat-Tree 型トポロジーの InfiniBand ネットワーク２系統によって接続されているため、理論的にはノード数に比例した全対全通信性能を実現することも不可能ではないはずである。ところが実際に計測すると、Strong Scaling での性能評価では 64 ノードを越えたあたりから効率の低下が見られる。これには主に二つの要因がある。まずノード数を増やすことにより各相手ノードへ転送するデータ量が減るため、データ転送のオーバヘッドが相対的に大きくなり転送効率が低下する。もう１点として、使用するノード数が増えるにしたがってネットワーク内混雑に巻き込まれる可能性が向上することがあげられる。通常運用中の TSUBAME2.0 では自分の他にも様々なジョブが実行されており、それぞれの通信パターンで共有ネットワークを使用しているため、様々な影響を受ける。
このような状況下でも安定した性能を出すために通信アルゴリズムの改良を行う。(1) 小さいメッセージを効率よく転送するために MPI ライブラリではなく low level の IBverbs API を用い、(2) 混雑に巻き込まれた場合の影響を低減させるために複数の RDMA 転送を同時実行し、(3) 2 系統ある InfiniBand ネットワークを活用して、衝突が少なくなるように各相手ノードとの通信をそれぞれのネットワークに動的に振り分ける、という手法を用いる。その結果スケーラビリティは大きく向上し、256 ノード使用時に最大 4.8TFLOPS の性能を達成することができた。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　１　件）

[1] 遠藤 敏夫, 額田 彰, 松岡 聡. スーパーコンピュータ TSUBAME 2.0 における

Linpack 性能 1 ペタフロップス超の達成. 情報処理学会論文誌コンピューティングシステム, Vol. 4, No.4 (ACS 35), pp.169–179, 2011 年 10 月.

〔学会発表〕(計 16 件)

[1] Kento Sato, Akira Nukada, Naoya Maruyama, Satoshi Matsuoka. “I/O acceleration with GPU for I/O-bound Applications”, In GPU Technology Conference 2014, poster, San Jose, Mar. 24th, 2014.

[2] 遠藤 敏夫,額田 彰,松岡 聡. 「ウルトラグリーンスパコン TSUBAME2.5/TSUBAME-KFC」, 大学 ICT 推進協議会 2013 年度年次大会講演論文集, T5G-4, 千葉, 2013 年 12 月 18 日.

[3] 遠藤 敏夫, 額田 彰, 松岡 聡. 「TSUBAME-KFC: 液浸冷却を用いたウルトラグリーンスパコン研究設備」, ハイパフォーマンスコンピューティングとアーキテクチャの評価に関する北海道ワークショップ (HOKKE), 情報処理学会研究報告, Vol. 2013-ARC-199/HPC-142, 札幌, 2013 年 12 月 16 日.

[4] 住吉 優希, 長岡 駿希, 藤井 昭宏, 額田 彰, 田中 輝雄. 「APU 上の混合精度 AMG 法」, 情報処理学会研究報告, Vol. 2013-HPC-141, No. 13, pp. 1-7, 沖縄, 2013 年 9 月 30 日.

[5] Akira Nukada, Kento Sato and Satoshi Matsuoka. “Scalable Multi-GPU 3-D FFT for TSUBAME 2.0 Supercomputer”, In Proc. of 2012 ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC ’12), Salt Lake City, IEEE Press, Nov. 10th, 2012.

[6] Takashi Shimokawabe, Takayuki Aoki, Tomohiro Takaki, Akinori Yamanaka, Akira Nukada, Peta-scale GPU Computing of Phase-Field Simulation for Dendritic Solidification on the TSUBAME 2.0 supercomputer -- The 10th WORLD CONGRESS ON COMPUTATIONAL MECHANICS (WCCM 2012), July 8th, 2012, Sao Paulo, Brazil.

[7] 額田 彰, 「CUDA 版自動チューニング手法」, GPU Technology Conference Japan 2012, 六本木, 2012 年 7 月 26 日.

[8] 下川辺隆史, 青木尊之, 高木知弘, 山中晃徳, 額田彰. GPU スパコン TSUBAME 2.0 によるフェーズフィールド法を用いた 2 petaflops 樹枝状凝固成長計算, 第 17 回計算工学講演会論文集, Vol. 17, 京都, 2012 年 5 月 29 日.

[9] Akira Nukada, “Performance of 3-D FFT using Multiple GPUs with CUDA 4”, NVIDIA GPU Technology Conference 2012, San Jose, May 14th, 2012.

[10] Akira Nukada, Yutaka Maruyama, Satoshi Matsuoka. “High Performance 3-D FFT using multiple CUDA GPUs”, In Proceedings of the Fifth Workshop on General Purpose Processing using Graphics Processing Units (GPGPU-5) in conjunction with ACM ASPLOS XVII, London, UK, pp. 57-63, ACM Press, Mar. 3rd, 2012.

[11] 遠藤 敏夫,松岡 聡,額田 彰,長坂 真路,四津 匡康, “グリーンスパコン TSUBAME2.0 における電力危機対応運用, 情報処理学会研究報告, Vol. 2011-ARC-197/HPC-132, pp. 1–9, 札幌, 2011 年 11 月 28 日.

[12] Takashi Shimokawabe, Takayuki Aoki, Tomohiro Takaki, Akinori Yamanaka, Akira Nukada, Toshio Endo, Naoya Maruyama, and Satoshi Matsuoka, “Peta-scale Phase-Field Simulation for Dendritic Solidification on the TSUBAME 2.0 Supercomputer”, In Proc. of 2011 ACM/IEEE International Conference for High Performance, Networking, Storage, and Analysis (SC ’11), Seattle, ACM Press, Nov. 12th, 2011. (Technical Paper and Gordon Bell Award finalist.)

[13] Shuntaro Yamazaki, Akira Nukada, Masaaki Mochimaru, “Hamming Color Code for Dense and Robust One-shot 3D Scanning”, In Proc. of the 2011 British Machine Vision Conference, Dundee, Scotland, Springer, Aug. 29th, 2011.

[14] Akira Nukada, “Fast Fourier Transform for AMD GPUs”, AMD Fusion Developer Summit 2011, Bellevue, WA. June 16th, 2011.

[15] 遠藤 敏夫, 額田 彰, 松岡 聡. スーパーコンピュータ TSUBAME 2.0 における Linpack 性能 1 ペタフロップス超の達成. 先進的計算基盤システムシンポジウム (SACSIS2011)論文集, 東京, 2011 年 5 月 25 日.

[16] Akira Nukada, Hiroyuki Takizawa, and Satoshi Matsuoka. “NVCR: A Transparent Checkpoint-Restart Library for NVIDIA

CUDA”, In Proc. of 20th Heterogeneity in Computing Workshop (HCW 2011), in conjunction with IPDPS 2011, Anchorage, AK, USA, May 16th, 2011.

〔図書〕（計　0　件）

〔産業財産権〕
○出願状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者
額田　彰（NUKADA, AKIRA）
東京工業大学・学術国際情報センター・特任准教授
研究者番号：40545688

(2)研究分担者
（　　）

研究者番号：

(3)連携研究者
（　　）

研究者番号：