受付 ID	16a-58
分野	HPCS

密結合演算加速機構アーキテクチャに向けた アプリケーションの開発と性能評価

塙　敏博
東京大学情報基盤センター

１．研究目的

GPU に代表される演算加速装置は、その高い演算性能とメモリバンド幅、電力当たり性能のため HPC 用途のクラスタに搭載され広く用いられている。しかし、クラスタ上の演算加速装置間の通信では，これまでホストメモリを介した転送が必要であり，特に小データの転送ではレイテンシがボトルネックとなる。そこで，レイテンシとバンド幅の改善を目指した独自開発の演算加速装置向け専用相互結合機構 TCA(Tightly Coupled Accelerators)の開発を行っている。

本研究では，マルチノード・マルチ GPU を用いたアプリケーションとして，QCD や宇宙物理のアプリケーション，数値計算などを対象に， TCA に向けた変更を継続して行う。TCA はノードを超えた GPU 間通信のレイテンシを大きく改善することが可能であり，アプリケーションの強スケーリングにおける性能改善に期待されている。さらに，TCA と InfiniBand からなる複合ネットワークにおける通信の最適化について検討を継続する。TCA によるミニクラスタはノード数が制限されるため，それを超える規模のアプリケーションでは，GPU 間の通信には，従来と同様 InfiniBand を経由した MPI 通信を使って記述する。そこで TCA と MPI の組み合わせ手法についても検討している。

２．研究成果の内容

今年度は、TCA において集団通信関数を実装し、CG 法に適用して性能評価を行った。その結果、小さいデータサイズにおいては低レイテンシの効果が高く、MPI に比べて高い性能を得ることができた。

また Graph500 アプリケーションの実装を行い、PEACH3 との比較のため PEACH2 における性能評価を実施した。

一方、TCA/PEACH2 を利用するためには独自 の API を用いる必要があり、プログラミングコストが高く、既存のアプリケーションの移植も容易でないという問題があった。そこで PGAS 言語を対象とした通信ライブラリである GASNet に注目し，

TCA 用に実装した。これによって，GASNet を介して各種のソフトウェアとの互換性が生まれ，TCA/PEACH2 が広く利用できるようになる。プロトタイプ実装において、ノードを跨ぐ GPU 間の通信性能は TCA/PEACH2 を直接使用した場合の性能と 比較して，最小レイテンシの増大は 15%程度に抑えられ，ソフトウェア支援によって最大バンド幅は最大 2.1 倍の性能向上を達成した。また、PEACH2 には 4 チャネルの DMAC が実装されており，複数チャネルの DMAC を活用した転送による性能改善についての検討および実装を行った。その結果，有効な転送サイズの範囲は限られるが、1 チャネルしか利用しない場合と比べて最大で 1.4 倍のバンド幅が得られた。

さらに GPU カーネル内から MPI 通信の起動を可能とする GPU セルフ MPI 機構“GMPI”を引き続き開発した。姫野ベンチマークの性能評価を行った。

３．学際共同利用として実施した意義

本プロジェクトでは TCA における通信機能の実現および性能評価を目的としていたため、主として HA-PACS/TCA を用いた。TCA を搭載した GPU クラスタとしては唯一の環境であり、研究の遂行には学際共同利用プログラムが必要不可欠であった。本研究の成果は，TCA アーキテクチャ，ならびに HA-PACS/TCA における効率的な通信の実現につながっており、他の HA-PACS/TCA 利用者に対して、TCA のみならず、MPI における最適なパラメータなどフィードバックされている。

４．今後の展望

本プロジェクトにおいて基本性能に加えて実アプリケーションにおいても TCA の有用性が確認されている。また PEACH3 によりアプリケーション性能の向上も確認できた。今後の PACS-X に向けた、演算加速機構を支える通信機構の要素技術開発に成果を反映していく。

５．成果発表

(1) 学術論文

A) K. Matsumoto, N. Fujita, T. Hanawa, and T. Boku: Implementation and Evaluation of NAS Parallel CG Benchmark on GPU Cluster with Proprietary Interconnect TCA, Post-proceedings of International Meeting onHigh Performance Computing for Computational Science (VECPAR) 2016, the Springer Series Lecture Notes in Computer Science (LNCS),2016. (in Press)

(2) 学会発表

A) 佐藤 賢太，藤田 典久，塙 敏博，松本 和也，朴 泰祐, Khaled Ibrahim: 密結合並列演算加速機構 TCA による GPU 対応 GASNet の実装と評価,ハイパフォー

マンスコンピューティングと計算科学シンポジウム (HPCS)2016, pp. 68--76, 2016 年 6 月

B) Kenta Sato, Norihisa Fujita, Toshihiro Hanawa, Taisuke Boku, Khaled Z. Ibrahim, ``GPU-Ready GASNet Implementation on the TCA Proprietary Interconnect Architecture,'' International Conference on Computational Science and Computational Intelligence (CSCI2016), Dec. 2016..

(3) その他

A) Takahiro Kaneda, Chiharu Tsuruta, Toshihiro Hanawa and Hideharu Amano: Performance Evaluation of PEACH3: Field Programmable Gate Array Switch for Tightly Coupled Accelerators, The 7th International Symporium on Highly Efficient Accelerators and Reconfigurable Technologies (HEART 2016), short paper, Jul. 2016.

B)　Toshihiro Hanawa, Takahiro Kaneda, Chiharu, Tsuruta, Hideharu Amano, and Taisuke Boku: Performance Evaluation of Low-latency Inter-node Communication between GPUs using PEACH3, HPC in Asia Poster, in conjunction with International Supercomputing Conference (ISC'16), Frankfurt, Jun. 2016.

C) 金田 隆大, 鶴田 千晴, 塙 敏博, 天野 英晴: PEACH3 の性能測定, 情報処理学会研究報告, 2016-ARC-221(32), pp. 1--6, 松本市キッセイ文化ホール, 2016 年 8 月

D) 桑原 悠太, 塙 敏博 , 朴 泰祐: GPU クラスタにおける GPU セルフ MPI システム GMPI の予備性能評価, 2016-HPC-155(15), pp. 1--8, 松本市キッセイ文化ホール, 2016 年 8 月.

E) 佐藤 賢太, 藤田 典久, 塙 敏博, 朴 泰祐, Khaled Ibrahim:密結合並列演算加速機構 TCA における複数 DMAC の活用による GPU 対応 GASNet の性能改善,2016-HPC-156(5), pp. 1--8, 小樽経済センター, 2016 年 9 月

F) 金田 隆大 , 塙 敏博 , 天野 英晴:アプリケーション中の通信に PEACH3 を利用した場合の評価, 2017-SLDM-178(16)/2017-ARC-224(16), pp. 1--6, 慶応大, 2017 年 1 月.

使用計算機	使用計算機に○	配分リソース※
HA-PACS	○	140
HA-PACS/TCA	○	445
COMA	○	299
※配分リソースについては 32node 換算時間をご記入ください。		