

GPU クラスタにおける並列一次元 FFT の実現と評価

Title: Implementation and evaluation of parallel 1-D FFT on GPU clusters

高橋大介

Daisuke Takahashi

筑波大学システム情報系

Faculty of Engineering, Information and Systems, University of Tsukuba,
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

本発表では、GPU クラスタにおいて並列一次元 FFT を実現し評価した結果について報告する。近年、GPU (Graphics Processing Unit) の高い演算性能とメモリバンド幅に着目し、これを様々な HPC アプリケーションに適用する試みが行われている。また、GPU を搭載した計算ノードを多数接続した GPU クラスタも普及が進んでおり、2013 年 6 月の TOP500 リストでは NVIDIA Tesla K20X GPU を搭載した GPU クラスタである Titan が第 2 位にランクされている。

これまでに GPU クラスタにおける並列三次元 FFT の実現は行われているが[1, 2]、並列一次元 FFT については著者の知る限りまだ実現されていないのが現状である。

本研究では、GPU クラスタにおいて並列一次元 FFT の実現を行った。GPU クラスタにおいて並列一次元 FFT を行う際には、全対全通信が 3 回行われることから、計算時間の大部分が全対全通信によって占められることになる。さらに CPU と GPU 間を接続するインターフェースである PCI Express バスの理論ピークバンド幅は PCI Express Gen 2 x 16 レーンの場合には一方向あたり 8GB/sec となっていることから、CPU と GPU 間のデータ転送量を削減することも重要になる。

GPU 上のメモリを MPI により転送する場合、基本的には(1) GPU 上のデバイスメモリから CPU 上のホストメモリへデータをコピーする、(2) MPI の通信関数を用いて転送する、(3) CPU 上のホストメモリから GPU 上のデバイスメモリにコピーする、という手順で行う必要がある。この場合、CPU と GPU のデータ転送を行っている間は MPI の通信が行われれないという問題がある。

そこで、CPU と GPU 間のデータ転送とノード間の MPI 通信をパイプライン化してオーバーラップすることができる MPI ライブラリである MVAPICH2 を用いることで、この問題を解決した。さらに、FFT の処理において出現する行列の転置の処理を GPU 上で行うなどの工夫も行った。

実現した並列一次元 FFT を GPU クラスタである HA-PACS (268 ノード, 4288 コア, 1072GPU) の 128 ノードを用いて性能評価を行った。その結果、 2^{34} 点倍精度複素数 FFT において約 741 GFlops の性能を得ることができた。

[1] Y. Chen, X. Cui and H. Mei: Large-Scale FFT on GPU Clusters, Proc. 24th ACM International Conference on Supercomputing (ICS'10) (2010).

[2] A. Nukada, K. Sato and S. Matsuoka: Scalable Multi-GPU 3-D FFT for TSUBAME 2.0 Supercomputer, Proc. 2012 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC'12) (2012).