

発表概要

Unified Memory を利用した CUDA プログラミングと性能評価

土井 淳^{1,a)} 今井 晴基¹

2014年11月10日発表

CUDA を使用した GPU プログラミングにおいて、ホスト CPU と GPU デバイスそれぞれにおけるメモリの管理やそれぞれの間のデータ転送がプログラミングを複雑なものにしている。CUDA6 から実装された Unified Memory を用いることで、ホストとデバイスで同一のアドレス空間にアクセスができ、これによって GPU プログラミングが簡単になりより生産性が向上することが期待される。しかしながら、物理的には異なるメモリ空間に確保されたデータを参照するには、依然データ転送が発生し、手作業でデータ転送などを最適化したコードよりも性能が低下することが考えられる。本発表では、実際のアプリケーションプログラムを Unified Memory を用いて GPU 上で実装し、その性能を評価する。

Evaluation of CUDA Coding Using Unified Memory

JUN DOI^{1,a)} HARUKI IMAI¹

Presented: November 10, 2014

Complexity of GPU programming with CUDA is sometimes caused by the memory management on host CPU and GPU device and data transfer between them. By using unified memory implemented in CUDA6, we can access same address space both from host and device and it will make GPU programming easier and more productive. However the data transfer is still needed to access data allocated in different place implicitly and this may cause performance degradation from hand tuned codes. In this presentation, we implemented applications on CUDA with unified memory and evaluated the performance.

¹ 日本アイ・ビー・エム株式会社東京基礎研究所
IBM Research – Tokyo, Koto, Tokyo 135–8511, Japan

^{a)} doichan@jp.ibm.com