

HPC 向け変動精度計算基盤および複数 GPU 間ストリーム処理機構の

開発

Development of Computing Platform based on Transprecision and Inter-GPU Streaming Mechanism for HPC

埴 敏博

東京大学

1. 研究目的

FPGA は、AI を始めとした特定用途のアクセラレータとして活用され始めている。OpenCL などの高位合成が可能な開発環境の普及に伴い、限られた電力制約の下 HPC アプリケーションに対してカスタマイズ可能なアクセラレータとして FPGA への期待が高まっている。そこで我々は、1. 変動精度による HPC アプリケーションの精度評価および性能・電力評価、2. 複数ノード GPU 間のデータ転送に伴うストリーム処理機構の実現、を目指して研究を行っている。

本研究では、宇宙物理のアプリケーション、数値計算などを対象に、FPGA へのオフローディングの実現と変動精度の適用を目指している。さらに、筑波大と理研 R-CCS で開発されている、OpenCL で記述可能な FPGA 間通信フレームワーク CIRCUS (Communication Integrated Reconfigurable CompUting System) の導入に加えて、ホスト上の GPU との連携についても検討する。

2. 研究成果の内容

高性能計算における計算時間の短縮や消費エネルギー削減に向けてアプリケーション中の各フェーズで必要とする精度に合わせて計算を行う、変動精度演算が注目を集めている。しかし、現在の CPU や GPU では、あらかじめ用意された精度の演算器を使う必要があり、任意精度を実現するためにはエミュレーションを行わなければならない。従って、アプリケーション全体で変動精度演算の効果を検証するのは困難であった。そこで本年度は、FPGA 上に高位合成 (HLS: High Level Synthesis) によって任意精度演算器を実装し、ホストからオフロード実行できることを示した。

具体的には、様々な精度の演算器を、Verilog HDL によるハードウェア記述の実装と C++ による実装をそれぞれライブラリとして用意し、OpenCL 演算カーネルに組み込むことによって実現し、評価を行った。さらに、このカーネルを、GEMM、3 次元ステンシル計算に組み込んで評価を行った。その結果、期待通りの任意精度の演算器を実装することができ、実行の結果、仮数ビット長とアプリケーションによって生じる

誤差を確認した。また、ビット数が少ないほど演算器の数を増やすことができるため、並列性を高め、アプリケーション全体のスループットを向上できることも確認した。ただし、用いた Intel Stratix10 GX FPGA に内蔵された演算器の構造と、アプリケーションの実装の問題から、浮動小数点フォーマットが IEEE754 準拠の FP32, FP16 のときの性能が高い傾向が見られた。

3. 学際共同利用が果たした役割と意義

本プロジェクトでは、GPU と FPGA との協調計算を目指しており、GPU と FPGA の複合クラスタとしては筑波大 CCS の Cygnus 以外に存在しない。また、筑波大、理研 R-CCS と協力して研究を進める上での共通プラットフォームとしても Cygnus が最適な環境である。そのため、本研究の遂行には学際共同利用が必要不可欠である。

4. 今後の展望

これまでに FPGA に OpenCL, C++を用いて変動精度のためのオフロード演算環境を実装し実現可能性を示した。今後は、実アプリケーションのカーネルへの適用を行っていく予定である。

5. 成果発表

(1) 学術論文

(2) 学会発表

- ① T. Hara, T. Hanawa, Transprecision Calculation Platform Offloaded on FPGA, xSIG 2021, 2021 年 7 月（発表決定, Best Master's Student Award)

- ② 原 忠辰, 埜 敏博, FPGA による変動精度演算に向けた実装方法の検討, 情報処理学会ハイパフォーマンスコンピューティング研究会, 2020 年 12 月

(3) その他

使用計算機	使用計算機 に○	配分リソース*	
		当初配分	追加配分
Cygnus	○	5000	
Oakforest-PACS			

※配分リソースについてはノード時間積をご記入ください。