GPU同士を直接接続したスーパーコンピュータを作るスイッチPEACH(1/2)

GPU搭載スーパーコンピュータ

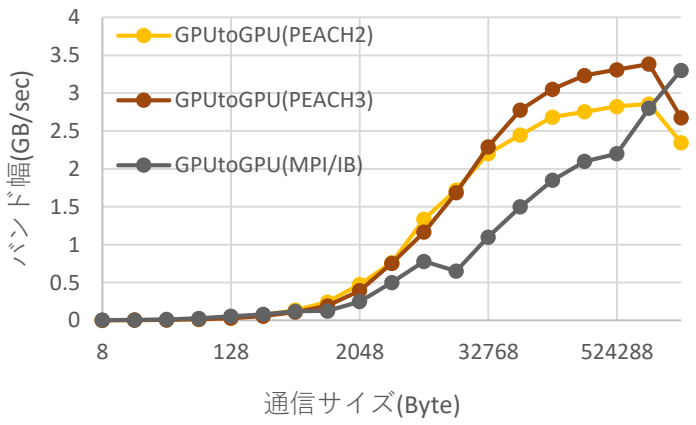
近年のスーパーコンピュータはアクセラレータ(GPU,FPGA,etc…)を利用し計算性能を向上させているものが多くあります.特にGPUはコストパフォーマンスに優れ多くのシステムで利用されています.しかし,GPUはCPUとの通信や他ノードのGPUとの通信を行う場合遅延が大きく,通信が性能低下の原因となってしまいます.

GPU間をつなぐスイッチングハブ,PEACH3

PEACH3はAltera社のStratix V FPGAを利用しています.ホストマシンと接続するためのPCIeカードエッジ(Nポート)と,他ノードと接続するPCIeコネクタ(E,Wポート)を持ち,E,Wポート間をPCIeケーブルで接続します.PEACH3には512GBという大規模なPCIeアドレス空間を割当て,接続された全ノードのCPU,GPU,PEACH3のアドレスを保持します.そのアドレスを利用し,DMAを行うのですが,GPUメモリに対するDMAには,NVIDIAが提供するGPUDirectRDMAを利用しています.ルーティングは宛先アドレス上位1ビットを確認するだけで良いので,とても高速に行うことが出来ます.

PEACH3のバンド幅

PEACH3による通信バンド幅を示しています.比較のため,前バージョンのPEACH2と,Infinibandの結果も示しています.グラフよりPEACH3がほぼ全域に渡り一番良い性能を示しています.



研究者名

金田隆大 志村英樹 天野英晴

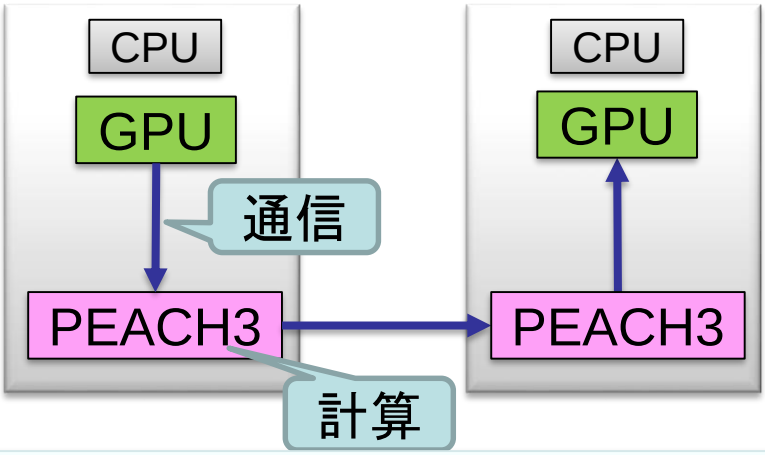
お問合せ先

hlab_ac-crest@am.ics.keio.ac.jp

GPU同士を直接接続したスーパーコンピュータを作るスイッチPEACH(2/2)

Accelerator in Switch

ハイエンドなFPGA(Field Programmable Gate Array)は、従来、ネットワークスイッチやルータとして主に利用されて来ましたが、最近はCPUやGPUが苦手な計算を加速するアクセラレータとしての利用が期待されています。しかし、行列演算など多数の浮動小数演算器を並列に使う場合は、GPUに比べて性能が上がりません。そこで、我々は、従来得意とするスイッチの中にアクセラレータを組み込み、交信中のデータに対して処理を施す方式を提案しています。これがアクセラレータインスイッチです。スーパーコンピュータ用のスイッチPEACH3の中に2種類のアクセラレータを組み込んだ例を紹介します。



作成モジュール例(Allreduce)

Allreduceは指定した各ノードのデータの総和を計算し,各ノードに再び書き戻すという計算が必要です.特にGPUメモリ上の場合,メモリアクセスが多く必要で低速になってしまいます.そこでGPUからPEACH3がデータを読み出し,PEACH3間の通信と計算のみで総和求めた後GPUに結果を書き戻すことで高速に総和を求めることが出来ます.

研究者名

金田隆大 志村英樹 天野英晴

お問合せ先

hlab_ac-crest@am.ics.keio.ac.jp