

NVIDIA NVLink ブリッジの効果について

2022年6月10日

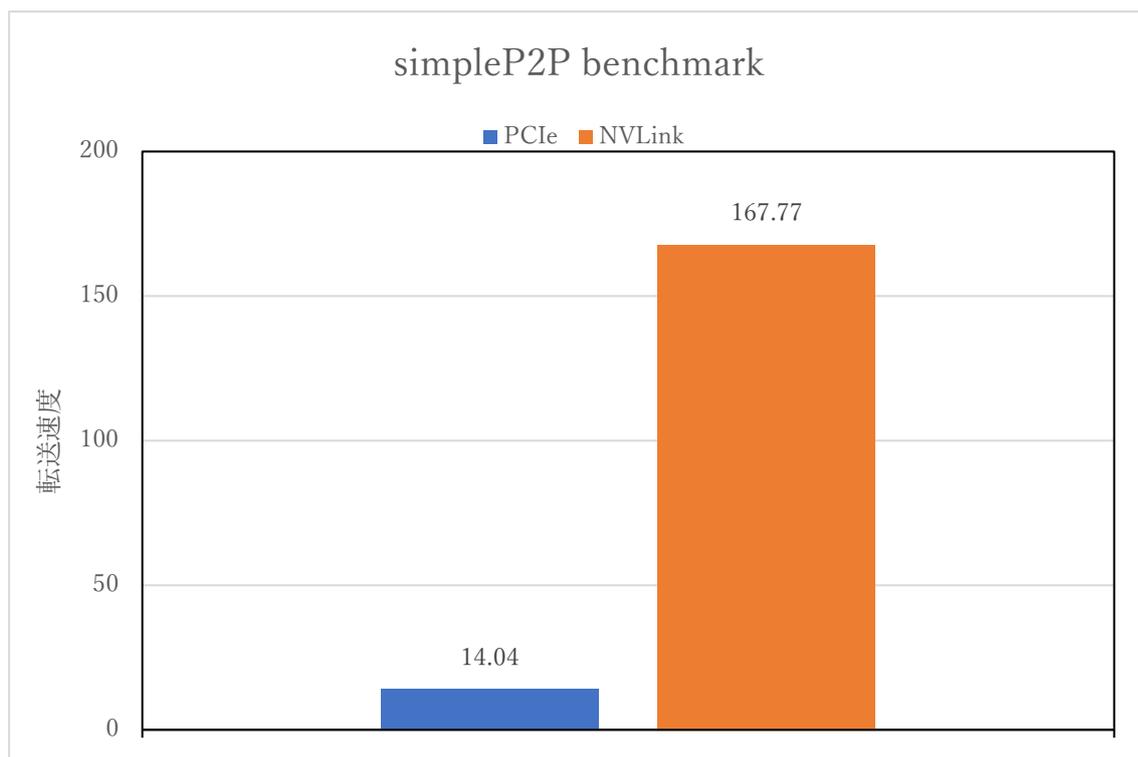
ビジュアルテクノロジー株式会社

1 はじめに

GPU (NVIDIA A100 40GB PCIE) の NVLink ブリッジ利用に関する CUDA アプリケーションの実行性能への影響について、下記の通り、検証を行いました。

2 GPU 間のデータ転送速度の比較

2個の GPU 間のデータ転送速度を測定する CUDA Toolkit のサンプルプログラム”simpleP2P”を実行し、NVLink ブリッジの有無での転送速度を比較した棒グラフを以下に示します。

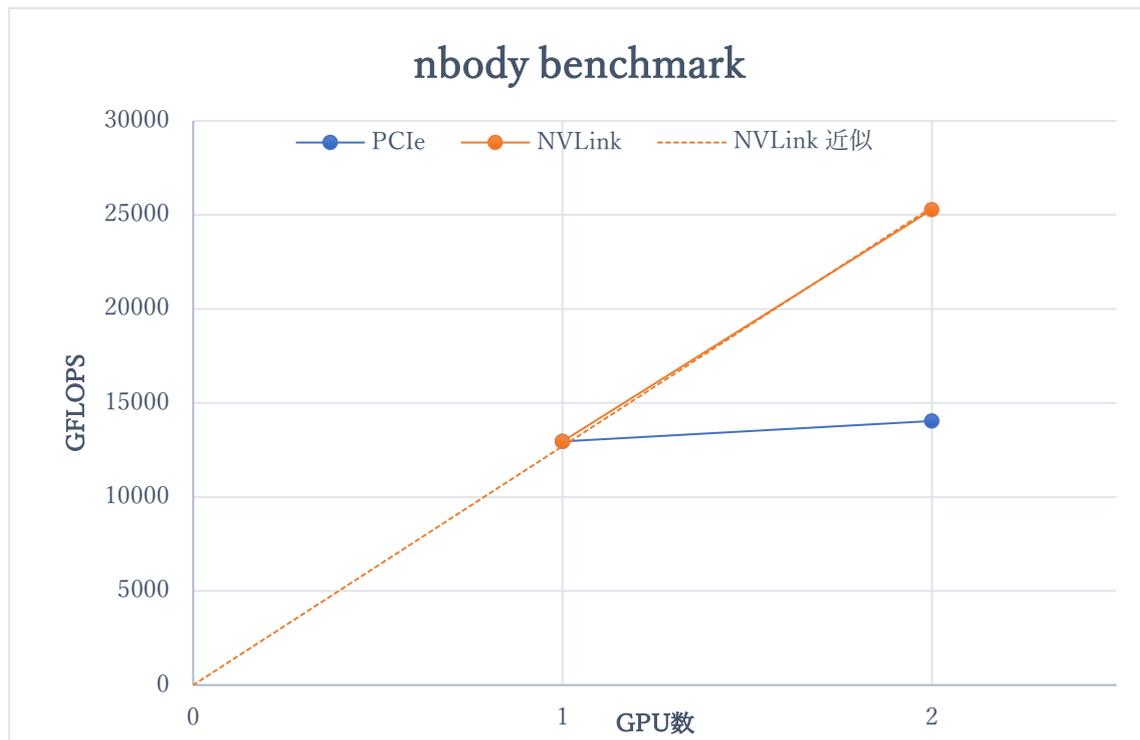


このグラフからは、ブリッジの追加によりブリッジなし (PCIe バス経由でデータ転送) の 10 倍以上の速度でデータを転送できることがわかります。

3 複数 GPU を使用するアプリケーションの実行性能の比較

GPU を使用して多体問題を近似的に解く CUDA Toolkit のサンプルプログラム”nbody”は複数の GPU を使用して計算を行い、実行性能 (FLOPS 値) を GFLOPS 単位で表示することができます。

GPU を 1 基使用して nbody を実行した場合の FLOPS 値、ブリッジなしで GPU を 2 個使用した場合の FLOPS 値、ブリッジありで GPU を 2 基使用して実行した場合の FLOPS 値のグラフを以下に示します。



ブリッジなしで GPU を 2 個使用してもほとんど FLOPS 値が増加しないこと (青グラフ)、ブリッジありで使用すると FLOPS 値が約 1.7 倍に増加 (橙グラフ) することがわかります。

複数 GPU を使用して nbody を実行すると GPU 間のデータの移動が大量に発生するため、ブリッジなしの場合は PCIe バスの速度に律速されると考えられます。

また、GPU 0 個での FLOPS 値を 0 と仮定して NVLink 使用時の FLOPS 値の近似直線を求めると、GPU 数 1 と 2 での近似値が実測値と非常に近い値となります (橙点線グラフ)。つまり、NVLink の使用することで GPU 数に比例して nbody の実行性能が向上すると考えられます。

以上