3J-08

ＰＣクラスタにおけるＩｎｆｉｎｉＢａｎｄを用いたハイブリッド並列処理法の提案

津田　伸生†　　諸角　義志‡
金沢工業大学†

1. はじめに

近年，マルチコア及びマルチプロセッサが一般的となり，PC クラスタの性能も向上している．しかし，プロセッサの多重化に伴い，処理時間全体におけるボトルネックである通信時間が増加している．

本研究ではこのボトルネックを解消するため，PC クラスタを構成するネットワークケーブルを高速通信が可能な InfiniBand[1]の使用及び，粗粒度は MPI，細粒度は OpenMP で分割するハイブリッド並列処理法を提案し，その有効性を評価した．

2. InfiniBand の概要

nfiniBand とは，イーサネットに代わる次世代インターフェイス技術である．複数の規格があり，本研究では QDR 規格の InfiniBand HCA（InfiniBand のネットワークインターフェイス），スイッチを使用している．表 1 に現在市場に出ている InfiniBand の規格のリストを示す．

表 1　InfiniBand の各規格の帯域幅

	SDR	DDR	QDR	FDR
1レーン	2.5Gbps	5Gbps	10Gbps	14Gbps
HCA(4x)	10Gbps	20Gbps	40Gbps	56Gbps

InfiniBand の利点を以下に示す．

・高バンド幅
　片方向最大 56Gbps のノード間転送速度を実現している（FDR）．ギガビットイーサネットの片方向最大 1Gbps と比べ，大幅に向上している．

・低レイテンシ
　データ転送時の遅延が小さく，小さなデータ通信が頻発した時に効果を発揮する．

・高信頼性
　InfiniBand スイッチ，HCA でエラーの検出と訂正を行う．

・低 CPU オーバーヘッド
　RDMA によりメモリ間で直接データ転送を行う事ができ，CPU の割り込みがほとんど発生しない．レイテンシの短縮にもなる．

3 MPI/OpenMP ハイブリッド並列処理の概要

マルチスレッド並列処理を実現する指示文又は基盤のことで，他のマルチスレッド(POSIX スレッド，TBB)と比較し，プログラミングがかなり容易に行えるという特徴がある．

3.1 ハイブリッド化のメリット・デメリット

メリット
・高並列処理のプログラミングが容易
・既存プログラムの拡張が容易
・通信待ち時間の減少

デメリット
・マルチスレッドによるオーバーヘッドの増加
・メモリの局所性低下による処理遅延

4. 研究環境

本研究では新たにＰＣクラスタを構築し，GPU と HCA 及びそれらのドライバをインストールした．ギガビットイーサネットと InfiniBand の選択はプログラム実行時のオプションで変更している．表 2 に PC クラスタの仕様を示す．

表 2　PC クラスタの仕様

サーバ	DELL PowerEdge T410
OS	CentOS 6.3 x86_64
CPU	Xeon® E5506 2.13GHz 4コア x 2ソケット 計8コア
Memory	16GB
GPU	Nvidia GT640 GDDR3 1GB
Network	GigabitEthernet 1Gb/s
	InfiniBand QDR 40Gb/s
HCAドライバ	MLNX_OFED_LINUX-1.5.3-4.0.42-rhel6.3-x86_64
MPI	OpenMPI 1.4.3
GPGPU	CUDA 5.5
台数	4台

5. 比較実験

PC クラスタで並列処理したときに通信時間の比率の多い問題を解くことが InfiniBand の有効性を評価しやすくなる．本研究では天体シュミレーション[2]について比較実験を行った．

5.1 天体シミュレーション

天体シミュレーションでは，速度，質量，座標を定義した質点を配置し，時間経過による質点の座標を求める．その際に使用する質点が受ける力を計算する方程式を示す．

$$F_i = \sum_{\substack{1\le j\le N \\ j\neq i}} f_{ij} = Gm_i \cdot \sum_{\substack{1\le j\le N \\ j\neq i}} \frac{m_j r_{ij}}{\|r_{ij}\|^3} \quad \cdots式1$$

式 1 により 2 点間の加速度を計算して速度，座標を補正する．また，分割単位ごとに質点数を分担し，計算ステップごとに通信を行う．

ギガビットイーサネットと InfiniBand を使用して計算し，処理時間を比較した．質点数 400，ループ回数 1000 で実行した結果を図 2 に示す．

ギガビットイーサネットではノード数が多くなると通信時間の割合が大きくなり，処理時間の短縮率は悪い．特に 3 ノードや 4 ノードでの実行時は通信時間が全体の半分以上を占め，1 ノードよりも処理時間が増加している．InfiniBand を使用した時は通信時間が大幅に短縮し，ノード数に応じた処理時間短縮ができている．

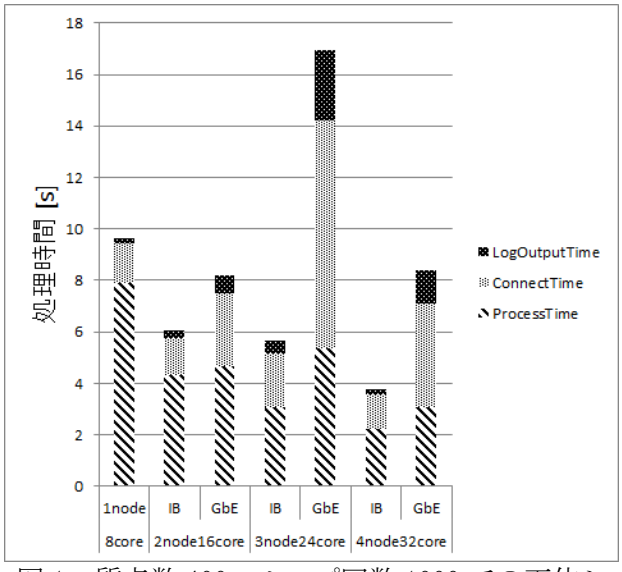

図 1　質点数 400，ループ回数 1000 での天体シミュレーションの処理時間の比較

６．まとめと今後の課題

PC クラスタを InfiniBand で接続することで，複数ノードでの並列処理時間が大幅に短縮できることがわかった．今後今回よりさらに多数のノードを InfiniBand で接続することで更なる高性能化が期待できる．

今後はハイブリッド並列処理の実装や，InfiniBand の持つ GPU のメモリに直接アクセス機能である GPU Direct を使い，PC クラスタのさらなる高速化を実現したい．

参考文献

[1] http://www.mellanox.co.jp/infiniband/
[2] 東雄也・津田伸生：”MPI/OpenMPI によるハイブリッド並列処理の試み” JHES2012，F-17.

A Hibrid Apoaroch for Parallel Proccessing using a PC Cluster Providing InfiniBand Interconnection Network

†Tsuda Nobuo，Morokado Yoshiyuki
‡Kanazawa Institute of Technology