サイエンティフィック・システム研究会　　2010 年 10 月 20 日
2010 年度科学技術計算分科会

次世代スパコン「京（けい）」注の言語処理系と性能評価

注：理化学研究所様が 2010 年 7 月に決定、発表した「次世代スーパーコンピュータ」の愛称）
※タイトルがプログラムから変更になっております。

林　正和
富士通株式会社 次世代テクニカルコンピューティング開発本部 ソフトウェア開発統括部

［アブストラクト］
次世代スパコン「京」への提供を目指して開発中の言語処理系（コンパイラ、ライブラリ、ツール）について、新 CPU/次期スパコンに盛り込んだ機能とその最適化及び効果について紹介する。また高並列プログラミングのチューニング方法について紹介する。

［キーワード］
スパコン、コンパイラ、性能評価、SPARC64、HPC-ACE、SIMD

１．はじめに

これまで、CPU は動作周波数を向上させることで性能向上を達成してきたが、近年動作周波数の向上は望めなくなっている。そのため、富士通は、コア数の増加と HPC 独自の命令を追加し、これらのコアを数万～数十万並列に処理することでの性能向上を狙った。

このようなアーキテクチャにおいて、プロセス並列（MPI）のみのプログラミングモデルでは、メモリ使用量やネットワーク通信量が増加し、数万超の並列では、性能向上の限界があると考えられ、次の 2 点を強化した。

1） プロセス並列とスレッド並列のハイブリッド並列

- VISIMPACT
  マルチコアを高速な 1CPU 化し、ハイブリッド並列化を容易にするアーキテクチャ及び自動並列コンパイラを開発。アーキテクチャの基本機能は、弊社 FX1 で実装し、効果を実証。次世代スーパーコンピュータ「京」でのベースアーキテクチャとして発展させている。
- HPC に向けてのコア強化
  - 汎用 CPU をベースに、レジスタ数拡張、柔軟な SIMD 演算器、ソフト制御可能なキャッシュ等の実行性能を高めるための機能(HPC-ACE)を追加した。（[1]）
  - 1CPU／1 ノード構成により高メモリバンド幅を確保。
    STREAM Triad 性能（メモリ）で、Peak　64GB/Sec　/　実効　46.6GB/sec を達成。

2） 数万超のネットワークが必要

- 6 次元メッシュ/トーラス（ユーザービューは 3 次元トーラス）
  通常の 3 次元トーラスではできない高い運用性や対故障性を 6 次元メッシュ/トーラスで実現。また、集団通信のアクセラレータ（Allreduce, バリア）をサポート。

本発表では、1)、2)を支える言語処理系に焦点をあてて発表する。（言語処理系は開発中であり、将来の機能変更があり得る。）

２．次世代スーパーコンピュータ「京」世代の言語処理系

超並列処理の実用化に向けて、プロセス/スレッドのハイブリッド実行モデルを容易に記述できるように、言語処理系は次の機能を有する。

1） コンパイラ

- Fortran/C/C++をサポート。標準規格だけでなく、デファクトな言語仕様もサポートすることで、広く使われているオープンソースを翻訳可能とする。
- 最適化機能としては、HPC-ACE の機能を活かすため、次に示すような最適化等を実施し、コア内の性能向上を図っている。
  - ◆ 拡張レジスタを有効に利用し、ループ最適化の範囲や対象を拡大することでコア内の演算待ちを減らす。
  - ◆ SIMD 演算を活用することで、実行命令数を減らす。
  - ◆ セクタキャッシュを有効に使うディレクティブを提供し、キャッシュの効率化を図る（将来は、コンパイラによるセクタキャッシュの自動利用も検討）
- ベクトル化を凌駕する自動並列機能に加え、OpenMP3.0 をサポート。

2) ライブラリ

- 新インターコネクトの特徴を活かし、数万プロセス並列を実用化する MPI を提供。
  - ◆1 対 1 通信
    ソフトウェアの階層構造をバイパスする特別な低遅延経路設定を実施。更に、送受信データの長さや配置に加え、ホップ数も考慮に加え、転送方式の切替えを最適化する。
  - ◆ 集団通信
    使用頻度高い関数（Bcast, Allgather, Allreduce, Alltoall 等）について、1 対 1 通信を利用せず、新インターコネクトの特徴を活かし、輻輳を抑える専用アルゴリズムを採用。また、新インターコネクトの高機能バリア通信(ハード実装)を利用
- システムにあわせてチューニングした数学ライブラリを提供。

3) 開発支援ソフトウェア

高並列プログラムのチューニングは、逐次性能と高並列性能を並行してチューニングし、測定するというサイクルを回す必要がある。そのための適切な情報を採取できるツールを提供する。また、各ユーザが使い慣れたISVツールとの連携を容易にするため、デファクトなインタフェースを採用した。

更に、高並列の場合はネットワークの輻輳が課題となる。ネットワークの輻輳は、①状態把握と ②対処の 2 ステップが必要である。それぞれについてのツールを検討中である。

３． 性能状況

（注：言語処理系は開発中のため、最終的な性能ではありません）

性能状況のサマリについて簡単に報告する。コア内の性能は、当社 FX1 機と比較することで評価している。コアの論理ピーク比が 1.6 倍（動作周波数:0.8 倍 × SIMD可:2 倍）に対し、平均で 1.5 倍を一つの性能ターゲットとしている。弊社のプログラムセット(140 本)で、2010/10 月時点では約 1.4 倍（平均）を達成。コンパイラの SIMD 生成時での改良や機能拡張により、更なる性能向上が見込まれる。

ノード内スケーラビリティについては、NAS PARALLEL BENCHMARKS([2])で 8 スレッドまでスケールすることを確認できた。MPI 通信性能については、現在評価中である。

以上

参考資料

[1] SPARC64™ VIIIfx 関連文書については、以下からダウンロードできます。
http://jp.fujitsu.com/solutions/hpc/brochures

[2] NAS PARALLEL BENCHMARKS の詳細は、次の URL を参照下さい。
http://www.nas.nasa.gov/Resources/Software/npb.html