サイエンティフィック・システム研究会
2010 年度科学技術計算分科会会合資料

2010 年 10 月 20 日

原子力研究開発を推進するスーパーコンピュータシステムと原子力アプリケーション

平山 俊雄 ・ 清水 大志 ・ 久米 悦雄
日本原子力研究開発機構・システム計算科学センター

[アブストラクト]

原子力機構では、計算科学を活用した原子力の研究開発を加速するために、本年、3 月には、Altix3700 Bx2（13TFlops）に替えて、国内最高性能となるピーク性能 200 TFlops の Linux クラスターシステム（BX900）とピーク性能 12 TFlops の次世代計算機プロトタイプ機（FX1）からなる 2 種類のスーパーコンピュータを導入し、それらの運用を開始した。BX900 は、旺盛な計算需要に応えるために、また FX1 は次世代計算機（京コンピュータ）の利用に向けた原子力アプリケーションのチューニングを目的としている。

BX900 については、その利用開始後 3 日目で 90%を超える利用率を達成し、その後も高い利用状況が続いている。これは、原子力計算科学研究者が数千コアを用いた大規模並列計算コードの開発能力と利用技術力を既に兼ね備えていることを表している。次世代計算機向け原子力コードのチューニングについては、FX1 を用いたプロファイリングデータに基づく性能予測を行い、1PFlops までの良好な並列性能を推定した。また、新旧 3 機種のスーパーコンピュータの基本性能を比較した結果についても言及する。

[キーワード]

原子力計算科学、次世代計算機、ペタフロップス計算、BX900、FX1、ハイブリッド並列

1. はじめに

計算科学技術は「理論」及び「実験」と並ぶ第 3 の研究手法として、21 世紀の先端的研究のフロンティアを切り開くための重要な基盤技術となっている。特に、原子力のような巨大技術においては、安全面や時間・空間の制約により実験が困難な場合が多く、計算科学技術は従来から重要な研究手段となっている。すなわち、計算科学技術は、原子力研究開発の効率化、原子力施設の安全評価、国際競争力強化のカギを握る技術となっている。原子力委員会においても、その原子力政策大綱の中で「シミュレーション技術の高度化による大規模な技術システム開発の効率化」を求めている。

原子力関連二法人の統合と独立行政法人化を経て、旧二法人が所有していたスーパーコンピュータの資源について、整理統合と合理化を果たし、膨大な計算需要に応えるための新たな設備を導入すると共に、京コンピュータの利用に向けた準備を整えた。

2. 新スーパーコンピュータの性能比較

BX900 の Linpack 性能は 191.40 TFlops、実行効率 95.66 %であり、2010 年 6 月時点の TOP500 によれば世界 22 位、実行効率世界 2 位であった。FX1 については、11.66 TFlops、実行効率 90.37 %である。

第 1 図に、新旧 3 機種の HPC Challenge Benchmark の性能を示す。ここで、Altix3700Bx2(Itanium2)は更新前の機種で、ピーク性能 13 TFlops である。Altix3700Bx2 に比較して、BX900 は演算性能、通信性能ともにコアあたりのピーク性能差（～2 倍）だけの改善がみられる。ただし、プロセス間通信のバンド幅および遅延については、ソケットあたりのコア数の増加やインターコネクトの構造の違いのため、単純に比較することができない。

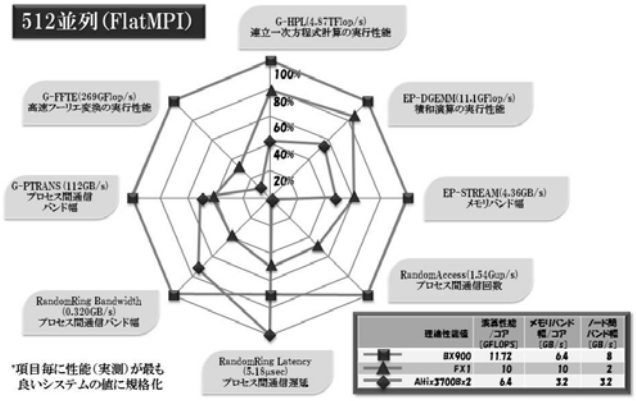


図1：新スーパーコンピュータの基本性能
HPC Challenge Benchmarkの結果

FX1 については、実効メモリバンド幅及びプ

ロセス間通信性能が BX900 と比較して低位にある。特に、実効メモリバンド幅については、BX900 がカタログ性能の 70%程度出ているのに対し、FX1 では 30%以下となっている。この問題は、メモリコントローラチップが CPU の外にあることに起因しているようだが、京コンピュータでは CPU 内蔵となっており、本問題は解決しているとのことである。

3. 京コンピュータの利用に向けたチューニングと性能予測

将来において膨大な計算需要を擁している核融合コードについて、 PFlops 級の計算を想定した並列性能および実行性能を BX900 および FX1 を用いて評価した。

BX900 の 16,384 コアを用いて並列性能を評価した結果、全てのコアに対して MPI プロセスを割り当てる FlatMPI では MPI プロセス間通信量の増大により、全体処理性能は大幅に低下する。したがって、大規模並列計算ではスレッド並列を併用する Hybrid 並列化を用いることで、MPI プロセス数を削減することが必須となる。なお、富士通 Fortran の自動並列化は核融合コードについてはかなりの部分で有効であった。また、メモリ使用量を軽減する上でも、Hybrid 並列化は有効である。

第 2 図は、京コンピュータ 4096 並列（～500TFlops）上での性能予測結果である。FX1 の 256CPU（1024 コア）を用いて性能予測をするため、4096 並列の 1 プロセスあたりの問題規模を保ったまま、weak scaling を仮定して 4096 プロセス並列から 256 プロセスにスケールダウンをする。次に、FX1 の 256CPU を用いて取得したプロファイリング情報に対して、FX1 から京コンピュータへの性能変換表を用いて演算時間を推定した。通信時間については、当該コードの通信の種類、データ長、通信回数を調査し、インターコネクトのトポロジー（仮定）をもとに机上計算により推定している。

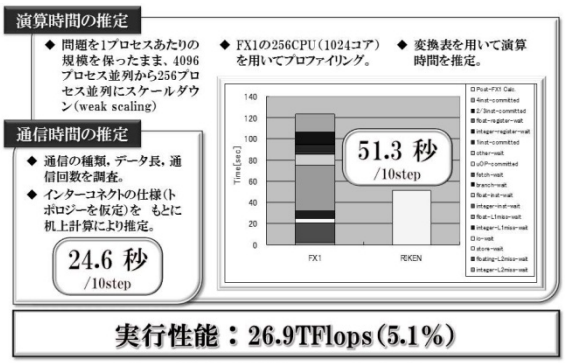
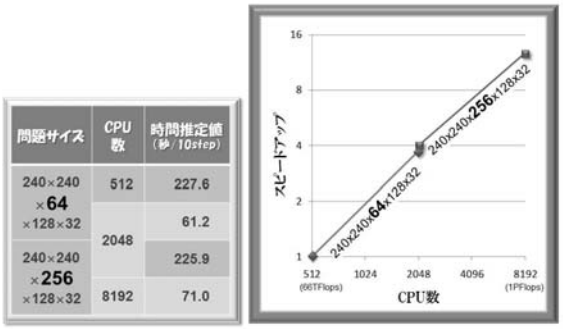


図2：性能予測；FX1から京へ
（4096並列、～500TFlops）

演算性能については、京コンピュータにおいてプロセスあたり 2.4 倍程度の高速化が期待できる。京コンピュータのコア数が、4 コア（FX1）から 8 コア（京）と 2 倍になっていることから、演算性能は十分にスケールしているといえる。しかし、1 コアあたりの性能はすでに飽和しつつあることに留意する必要がある。したがって、京コンピュータの利用にあたっては、コア数が増えた時のコードの並列化性能が特に重要になる。

第 3 図は、京コンピュータの 1PFlops について、上記推定手法に基づいた 2 段階の Strong Scaling である。1PFlops においても良好な並列化性能が期待できることがわかる。



問題サイズ	CPU数	時間推定値（秒/10step）
240×240×64×128×32	512	227.6
	2048	61.2
		225.9
240×240×256×128×32	8192	71.0

図3：1PFlopsまでのスケーラビィリティ

4. まとめ

日本原子力研究開発機構では、「シミュレーション技術の高度化による大規模な技術システム開発の効率化」を求めた原子力政策大綱に基づき、計算科学的手法を用いた原子力研究開発を推進してきた。その結果、年間の論文数の 20%以上が計算科学的手法を用いた研究に移行している。しかし、数 PFlops に達する膨大な潜在需要に応えるためには、計算機資源の不足は深刻な状況であり、京コンピュータの積極的利用が不可欠となっている。

京コンピュータのプロトタイプ・アーキテクチャの FX1 を用いたコード分析とチューニングは、京コンピュータの利用に向けた準備作業として極めて有効であることを確認した。また、原子力コードについて FX1 を用いたチューニングを行い、1PFlops までの良好な並列性能を確認した。