

スーパーコンピュータ「京」の開発状況について

横川三津夫[†]

[†] 理化学研究所次世代スーパーコンピュータ開発実施本部 〒650-0047 神戸市中央区港島南町 7-1-26

E-mail: [†] yokokawa@riken.jp

あらまし スーパーコンピュータ「京」は、8万台以上の計算ノード、2階層のファイルシステム、制御及び管理サーバ群、フロントエンドサーバ群からなる大規模分散メモリ型スーパーコンピュータであり、理化学研究所と富士通が共同で開発している。昨年11月に目標性能である LINPACK 性能 10 ペタフロップス以上を達成し、TOP500 において2期連続で第一位となった。今年6月にはシステムが完成する予定である。本講では、スーパーコンピュータ「京」の開発状況について述べる。

キーワード スーパーコンピュータ, 京, 10 ペタフロップス, 国家基幹技術, HPCI

1. はじめに

計算機シミュレーションによって様々な現象の解明を試みる「計算科学」は、第3の科学手法として確立されてきた。これは、現象等をモデル化した偏微分方程式を計算機で解く数値計算法の発展だけでなく、計算機それ自体の急速な性能向上に依るところが大きい。特に最先端の技術を用いた浮動小数点演算性能の高いコンピュータ、いわゆるスーパーコンピュータは、計算機シミュレーションの高精度化、高精細化を格段に進展させ、より現実的な現象再現を行うための強力な基盤ツールとして、今や必要不可欠なものとして認められている。

我が国は、平成18年3月の第3期科学技術基本計画の中で、スーパーコンピューティング技術を国家基幹技術として位置付け、その基盤ツールとしてのスーパーコンピュータの開発を開始した。理化学研究所(理研)は、この開発主体として、平成18年に次世代スーパーコンピュータ開発実施本部を設置し、LINPACK 性能 10 ペタフロップスの計算能力を持ち、かつ様々なアプリケーションにおいて高い性能が得られるような汎用性のあるシステムの開発を進めてきた。

この結果、理研と富士通が共同で開発しているスーパーコンピュータ「京(けい)」は、2011年11月に目標性能の LINPACK 性能 10 ペタフロップスを達成するとともに、第38回 TOP500 リストにおいて、昨年6月に引き続いて2回目の第1位を獲得した。TOP500 リストとは、LINPACK と呼ばれる連立一次方程式を解くベンチマークプログラムの実行速度を指標として、世界のスーパーコンピュータの上位500位までを定期的にランク付けするものであり、1年に2回発表される。今回の LINPACK 登録値は、京の計算性能を確認する

ために、864 筐体 (IO ノード用の CPU を含め CPU 総数 88,128 個) の構成による計測結果である。LINPACK 性能値 10.51 ペタフロップス (毎秒 1.051 京回の浮動小数点演算数)、実行効率 93.2% と高水準の記録を観測した。また、計測時間は約 29.5 時間に達し、88,128 個の CPU を搭載したシステムが故障せずに動作した点は高く評価できる。

また、スーパーコンピュータの総合的な性能を評価するベンチマーク「HPC チャレンジ」の4部門においても、第1位を獲得した。HPC チャレンジベンチマークは、科学技術計算で多用される計算パターンから抽出した28項目により、CPU 性能、メモリアクセス性能及び計算ノード間のネットワーク性能などを評価するものであり、その中の4部門①Global HPL (大規模な連立1次方程式の求解における演算速度)、②Global RandomAccess (並列プロセス間でのランダムメモリアクセス性能)、③EP STREAM(Triad) per system (多重負荷時のメモリアクセス速度)、④Global FFT (高速フーリエ変換の総合性能) において最も性能の高いものに HPC チャレンジ賞 (クラス1) が与えられるものである。これにより、「京」は種々のアプリケーションにおいても高い性能が得られる可能性を示すことができた。

本稿では「京」のシステム概要について述べる。

2. 京のシステム概要

京は、82,944 個の計算ノード (他に 5,184 個の IO ノードがある) から成る大規模分散メモリ型並列計算機システムである。2010年9月に第一号筐体が兵庫県神戸市の計算科学研究機構にある計算機室に設置された後、約10カ月をかけてすべての計算機筐体を設置され



図 1 整備中のスーパーコンピュータ「京」

た(図 1)。

一つの計算ノードは、一つの CPU (富士通製 SPARC64 VIIIfx), 16 ギガバイトのメモリ, 計算ノード間のデータ転送を行うインターコネク用 LSI (ICC: Inter-Connect Controller) で構成されている。CPU は、8 つのプロセッサコア, コア共有の 2 次キャッシュメモリ (6 メガバイト), メモリ制御ユニットを持っている。CPU の LSI の大きさは縦 22.7mm×横 22.6mm である。各コアは 4 つの積和演算器, 256 本の倍精度浮動小数点レジスタを持っており、一つの SIMD 命令 (ベクトル処理の一種) により、2 つの積和演算器を同時に動作させることができ、2 つの SIMD 命令を同時に実行することにより一つのコアはクロックサイクル毎に 8 個の浮動小数点演算ができる。したがって、コアの理論性能は 16GFLOPS, CPU (8 コア) の理論性能は 128GFLOPS となる。また、コア間の並列処理の同期を取るためのハードウェアバリア機構, 計算に必要なデータを事前にキャッシュに取り込むプリフェッチ機構, プログラマブルなキャッシュ制御を可能とするセクタキャッシュ機構など科学技術計算のための様々な機構を備えている。

計算ノード同士を繋ぐ接続 (インターコネク) は、新規に開発された Tofu (Torus Fusion) と呼ばれる 6 次元メッシュ/トーラス型の直接網であり、京の Tofu のサイズは 6 次元 (X, Y, Z, a, b, c) の各軸について $24 \times 18 \times (16+1) \times 2 \times 3 \times 2$ である。X, Z, b 軸がトーラス接続, Y, a, c 軸がメッシュ接続であり, Z 方向に一つの IO ノードが接続されている。各ノードの 10 本のリンクは、それぞれが別の計算ノードと繋がれており、各方向のリンクは片方向 5 ギガバイト/秒の理論帯域を持つ。

システムの利用イメージを図 2 に示す。計算ノード及びログインノードのオペレーティングシステムは、ユーザの利便性を考慮し、リナックス (Linux) とした。また、「京」が持つ高い計算性能を最大限に発揮させるために、CPU 内の自動並列化をサポートした最適化コンパイラ, デバッガ, 性能チューニングツール, 数値計算ライブラリを提供する。アプリケーション資産や

既存システムとの互換性を考慮し、科学技術計算の分野で広く使われているプログラミング言語 Fortran, C, C++ で書かれたプログラムが実行可能である。並列プログラミングには、標準ライブラリである MPI (メッセージパッシングインターフェース) とデータ並列言語の XPFortran をユーザに提供している。

ファイルシステムには、ローカルファイルシステムとグローバルファイルシステムの 2 階層ファイルシステムを採用し、グローバルファイルシステムとローカルファイルシステムとの間でファイルを転送する「ステージング方式」を採用した。ユーザのジョブ実行の前にユーザのファイルをグローバルファイルシステムからローカルファイルシステムへ転送し (ステージイン), ジョブ終了後の結果が格納されたファイルをローカルファイルシステムからグローバルファイルシステムに転送する (ステージアウト)。ステージングの指示は、ジョブ手順を記述するジョブスクリプトの中で指定し、ジョブマネージャがそれらを解釈して、ステージング動作を行う。

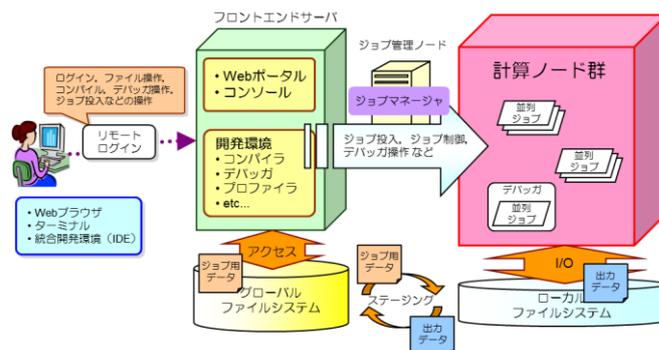


図 2 システム利用イメージ

3. 今後の予定

京は今年 6 月に完成し、秋には一般に公開する予定である。これに先立ち、平成 23 年 4 月より、ナノ分野、ライフサイエンス分野のグランドチャレンジ拠点、及び国が定めた戦略プログラムを担当する組織のユーザに対し、試験環境を提供し、それぞれのアプリケーションのチューニングが行われている。すでにいくつかの成果も報告されており、今後は、アプリケーションの成果が徐々に創出されることが期待される。

参考文献

- [1] Y. Hasegawa, et al., “First-principles calculations of electron states of a silicon nanowire with 100,000 atoms on the K computer” Proc. of the SC11, 2011.
- [2] 横川三津夫, 庄司文由, “京速コンピュータ「京(けい)」とは何か?”, 原子力学会誌 (アトモス), Vol. 52, No.12, pp.782-786, 2010.
- [3] 宇野篤也, 星屋隆之, “究極の Linux 機! スーパーコンピュータ世界一「京」の全貌”, 日経 Linux, 10 月号, pp.90-98, 2011.