

# Blue Gene/L

～世界最速スーパーコンピュータ～

平尾 洋樹, 梶原 広輝

Hiroki HIRAO, Hiroki KAJIWARA

## 1 はじめに

近年, 科学技術の発達に伴い, 分子レベルでのミクロなシミュレーションや地球規模での現象シミュレーションなど, 対象とする問題の大規模化・複雑化が進んでいる。そのような背景から, 高性能な計算資源, すなわちスーパーコンピュータの需要が高まっている。

本報告では, 世界最速の「Blue Gene/L」とともに, スーパーコンピュータ全体の動向について述べる。

## 2 TOP500

世界中のコンピュータの性能ランキング「TOP500 Computer Sites」<sup>1)</sup>の最新リスト(2005年6月)をTable 1に示す。上位の大半をIBMが占めている。

Table 1 TOP500 BEST5 (単位: TFlops)

| Rank | Computer / Processors            | Rmax  | Rpeak |
|------|----------------------------------|-------|-------|
| 1    | Blue Gene/L-eServer / 65,536     | 136.8 | 183.5 |
| 2    | BGW-eServer / 40,960             | 91.2  | 114.6 |
| 3    | Columbia-SGI Altix / 10,160      | 51.8  | 60.9  |
| 4    | Earth-Simulator / 5,120          | 35.8  | 40.9  |
| 5    | MareNostrum-JS20 Cluster / 4,800 | 27.9  | 42.1  |

1位の「Blue Gene/L」は, Linpack テストのスコアを2004年11月の70.72TFlopsから136.8TFlopsに伸ばし, 2回連続で世界最速にランキングされた。2位もBlue Gene Solutionの「BGW」で, IBMのThomas J. Watson Research Centerで使用されている。2004年6月まで5回連続首位を保ち同年11月に3位であったNECの「地球シミュレータ」が4位に後退した。

また, 全体的な性能の底上げも顕著で, 500位の「TSE1200」でも1.166TFlopsのスコアを記録しており, 今やTOP500入りするために1TFlopsを超える性能が必要になる。この性能は, 1993年6月に初めてTOP500が作られたときの全500システムの合計と同等である。

## 3 Blue Gene Solution

「Blue Gene Solution」とは, IBMが1999年から始めたプロジェクトであり, 最終的な目標は1PFlops(1秒間に $10^{15}$ 回の計算)を行えるシステムの構築である。地下核実験の代替として核兵器の実験シミュレーションなど科学技術計算を行うことを主たる目的としている。

「Blue Gene/L」はこのプロジェクトの最初のシステムであり, 2005年にローレンス・リバモア国立研究所に納品されたマシンである。その他にも, 「Blue Gene/P」は2006～2007年にピーク性能1PFlops, 実効性能300TFlopsを, 「Blue Gene/Q」は2007～2008年にピーク性能3PFlops, 実効性能1PFlopsを実現する。

## 4 Blue Gene/Lの構成

Blue Gene/Lは2.8GFlops, 0.7GHzで駆動するプロセッサが集積された密結合タイプのクラスタコンピュータである。1チップに2プロセッサとキャッシュメモリが搭載され, このチップ2個とメモリが1枚のカードに搭載される。このカード16個が1つのノードボードに搭載され, これが32枚収納されたキャビネットが1つの単位となり, システムはこのキャビネット64台から構成される。よって, 最終構成でのプロセッサの数は

$$2 \times 2 \times 16 \times 32 \times 64 = 131,072$$

(CPU/chip)(chips/card)(cards/node)(nodes/cabinet)(cabinets)(processors)である。この全体構成をFig. 1に示す。

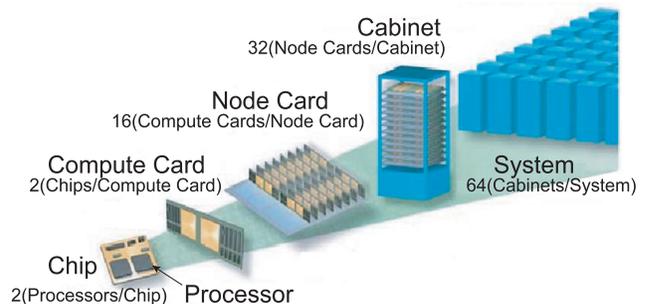


Fig. 1 Blue Gene/Lのハードウェア構成

これが完成すれば, 国内最速のスーパーコンピュータ, 地球シミュレータの9倍の速度でピーク性能360TFlopsを, 消費電力3分の1, 床面積17分の1で達成する。

ちなみに, 2005年6月に世界最速を達成したシステムは, 開発段階中で最終構成の2分の1の規模であった。1システム32キャビネット, 集積プロセッサ数65,536で, 実効性能136.8TFLOPSを達成した。

過去このように多数のノードからなるシステムは例がなく, さらにノードあたりのメモリが512MBと小さい。そのため高い実効性能を出すためには, これまで以上に高度なタスク割り当て制御が必要であると思われる。

#### 4.1 プロセッサ

プロセッサには、PowerPC440 という 1999 年に発表された古いアーキテクチャが使われている。このコアに Double FPU(浮動小数点演算装置) と 2 次キャッシュを組合わせたプロセッサブロックを 2 つ集積している。Blue Gene/L では、PPC440A4FPUV1 と呼ばれる FPU コアを 2 つ搭載しているが、片方は主にノード間メッセージ転送用に使用されるため、通常演算に使用されるのは全体の半分となる。これが Primary と Secondary として動作することで、次の 3 つの動作が可能になる。

- 片方の FPU だけで演算する
- 両方の FPU を 1 つの命令で演算させる (SIMD 動作)
- 両方の FPU を別々の命令で演算させる (MIMD 動作)

これにより、低消費電力で高密度の組み込みが可能となり、プロセッサとメモリのバランスが向上している。

#### 4.2 チップ

PowerPC の CPU コアをベースに周辺回路に SoC を搭載することで、システム全体の回路を単純にしている。コンピュータの主要機能を 1 つのチップに詰め込んだ方が、当然小型化され MTBF(平均故障時間) も伸びるので安定度が高まる。ただし、SoC とはいえ周辺回路を全て統合するとダイサイズは肥大化しがちである。それを抑えるために回路規模の大きい Power シリーズではなく、さらに規模の小さい PowerPC が採用されている。

#### 4.3 ネットワーク

ノード同士の接続には、同時に 5 種類のネットワークが用いられている。ノード間の通信には 3 次元のトラス構造で接続されたリンクを使い、ネットワークや割り込みなどにはツリー構造のリンクを、システム制御にはスター型のリンクをそれぞれ利用するという仕組みである。目的に応じて異なる接続方法を用意することで、回路は単純になっている。ネットワークを Fig. 2 に示す。

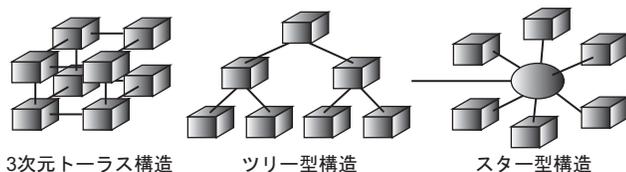


Fig. 2 Blue Gene/L のネットワーク

#### 4.4 OS

I/O ノードと計算ノードが役割を分担し、各ノードにはメッセージングに特化したプロセッサが割り当てられる。I/O ノードでは組み込み型 Linux が走り、計算ノードではハイパフォーマンス・カーネル (HPK) と呼ばれる、マルチスレッド処理に優れた IBM 独自仕様の軽量の OS を使用する。ユーザは外から見える I/O ノードの Linux だけを意識すればよく、計算ノードは見えない。

## 5 Blue Gene/L の利用例

高性能ながら小型で低価格、低消費電力という長所を併せ持つスーパーコンピューティング・テクノロジーは、様々なアプリケーションでより広範囲に利用されている。Blue Gene/L が国内で利用されている例を次に挙げる。

### 5.1 ニイウス株式会社 (NIWS)

ニイウスでは、Blue Gene/L を「グリッド/オートノミック・コンピューティングセンター (G/A センター)」と称される大型コンピュータのネットワーク内に設置し、顧客企業のアプリケーションを動かすプラットフォームとして提供する。事業戦略を大幅に転換中で、金融や医療分野に専門特化した業務アプリケーション提供や、オートノミックコンピューティング対応の高度な SE サービスなど、付加価値の高いビジネスへと移行している。

### 5.2 産業技術総合研究所 (産総研)

産総研では、「タンパク質立体構造予測」や、得られた立体構造と薬物候補化合物との相互作用の有無を調査する「タンパク質 - 化合物ドッキング」などバイオインフォマティクス分野の研究に活用していく。

## 6 今後の展望

現在、Blue Gene/L は IBM リサーチ部門のプロジェクトであるが、ゲノム研究、自動車の設計、金融、天気予報、流体力学など、極めて異質な複数の分野における最先端の課題に取り組むため、政府機関や大学の他、産業界も含めた世界中の研究者に利用されると予想される。

Blue Gene/L が完成すれば、先述したように、既存の性能の倍であるピーク性能 360TFlops を実現することになり、次回の TOP500 でも 1 位の座を死守するであろう。しかし、TOP500 全体の性能の底上げを考慮すると、10 年後の 2015 年頃には、最終構成の Blue Gene/L でさえ TOP500 から脱落してしまうことも考えられる。

日本の文部科学省は、ピーク性能が 10PFlops(1 秒間に  $10^{16}$  回) の次世代スーパーコンピュータ「京速計算機システム」の設計・開発を行っている。総事業費 1000 億円をかけ 10 年完成を目指し、米国がもつ世界最速のタイトル奪還に挑む。このように、各国が世界最速の称号を手にするべく、より速いスーパーコンピュータを作り続けるという傾向は今後も留まらないであろう。

## 参考文献

- 1) TOP500 Computer Sites  
<http://www.top500.org/>
- 2) Blue Gene ソリューション  
<http://www-6.ibm.com/jp/solutions/deepcomputing/solutions/bluegene.shtml>
- 3) Fall Processor Forum 2004  
<http://pcweb.mycom.co.jp/articles/2004/10/14/fpf/>