世界のスーパーコンピュータと京速コンピュータ

森 智弥，伊藤 冬子
Tomoya MORI，Fuyuko ITO

1　はじめに

近年，科学技術の発展に伴い，工学的問題の複雑化が進んでいる．例えば，技術の発達により，飛行機が開発されたが，より速く飛行するためにはエンジンの開発や改良等の新たな問題を解く必要がある．これらの大規模で複雑な問題に対処するためには計算機の能力が必要不可欠である．しかし，一般的なパーソナルコンピュータ等を用いると莫大な時間がかかってしまうため，それらの問題を短時間で解くために開発されたのがスーパーコンピュータである．本稿では，世界のスーパーコンピュータとその現状について述べた後，日本が開発を進めている京速コンピュータとそれを実現するための技術について述べる．

2　スーパーコンピュータ

スーパーコンピュータとは，一般的なコンピュータに比べ，演算処理速度が非常に高速でその時代の最新技術が投入された最高性能の計算機のことである．日本の文部科学省の科学技術・学術審査会では 2005 年の時点において，1.5TFlops 以上の演算性能を持つ計算機をスーパーコンピュータと定義している．

3　世界のスーパーコンピュータ

3.1　Top500

Top500[5] とは世界中のスーパーコンピュータの性能を計測し，そのランキングを公表するプロジェクトのことである．Top500 では，連立一次方程式の解を求める Linpack ベンチマークをを用いて，全世界のスーパーコンピュータの性能を計測している．ランキングは毎年 6 月と 11 月に上位 500 台が発表される．

3.2　国別で見るスーパーコンピュータ

スーパーコンピュータの製造が最も盛んな国は米国である．2008 年 11 月の最新の Top500 リストにおいて，米国のスーパーコンピュータは 290 台もランクされており，その全体に占める割合は 58 %にもなる．Top500 リスト首位である IBM のスカラ型スーパーコンピュータ Roadrunner もその一つである．その演算処理能力は Linpack ベンチマークテストにおいて 1.105PFlops と報告されている．

一方，日本は 17 台しかランクされておらず，その占有率はわずか 3 %程度である．2008 年 11 月の時点において最も上位にランクされている日本のスーパーコンピュータは東京大学情報基盤センターの T2K Open Supercomputer (Todai Combined Cluster) で 27 位．その演算性能は 82.984TFlops である．また，東京工業大学の TSUBAME1.2 は 77.48TFlops で 29 位，筑波大学の T2K システムは 32 位である．

中国の最高位は 10 位の上海超級計算センターに設置された Dawning 5000A システムであり，演算性能は 180.6TFlops である．これは米国以外のシステムとして，さらに Windows Cluster としても最高位である．また科学院の DeepComp 7000 システムが 19 位にランクされており，勢力を伸ばしていることが分かる．

さらに，インドの TATA の EKA システムが 13 位に，ドイツの FZJ の JUGENE システムが 11 位に，フランスの Jade システムが 14 位にランクされている．

Top500 における国別のシェア率の推移を Fig. 1 に示す．縦軸は Top500 にランクインした各国のマシンの台数，横軸は時間を表している．米国は 1993 年から 2008 年にかけて変わらず Top500 リストの大半を占めている．一方で，日本は占有率を落としているが中国は 2001 年以降占有率を上げている．

また，スーパーコンピュータの利用目的も国別で異なる．米国は軍事，ゲノムの解析，天文学への利用が予想され，中国でも軍事利用が予想される．日本ではゲノムの解析や自然現象の解析等に利用される．

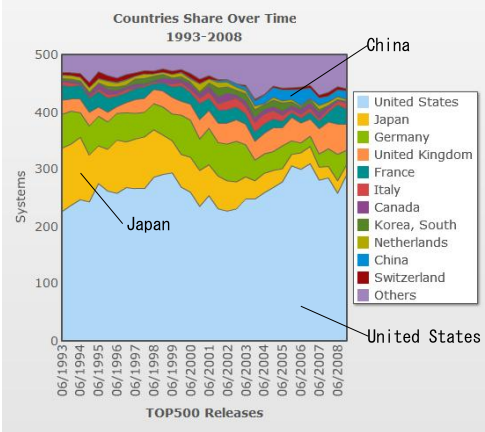

Fig.1　国別のシェア率 (参考文献 [5] より引用)

3.3　上位計算機に見るスーパーコンピュータの現状

3.3.1　プロセッサ

スーパーコンピュータのアーキテクチャはスカラ型とベクトル型に大きく分けることが出来る．スカラ型計算機とは 1 つの命令で 1 データを処理するプロセッサを持つ計算機のことであり，大きなデータを細分化して処理し，逐次的に処理する計算に適している．一方，ベクトル

型計算機は 1 命令で複数データを処理するプロセッサを持つ計算機のことであり, 似たような処理を複数同時に処理することが出来るので大規模行列計算などに適している．また, 近年では GPGPU(General Purpose Graphics Processing Unit) などの特定用途向けプロセッサも数多く利用されるようになってきている．

現在，世界のスーパーコンピュータのプロセッサアーキテクチャの主流はスカラ型である．2008 年 11 月の Top500 において，ベクトル型は日本の地球シミュレータの 1 台のみであり，残り 499 台がスカラ型である．その理由は，ベクトル型計算機は汎用で安価なプロセッサを使うスカラ型計算機に比べて高価になることに加え，現在ではスカラ型プロセッサの並列接続が容易となり，ベクトル型計算機と同等の速度での演算が可能になったためである．また，スカラ型計算機とされているが，実際には CPU の内部でスカラ型とベクトル型のハイブリッド構成になっている場合が増えてきている．

3.3.2 ハイブリッド型のプロセッサ構成

現在，Top500 の首位になっている Roadrunner は各ノードに AMD 社の Opteron プロセッサと IBM 社の PowerXCell 8i プロセッサを搭載したハイブリッド型の構成になっている．ハイブリッド型の長所はシステムが複雑な算術演算をセグメントに分割し，各セグメントを最も効率的に処理するように割り振ることができることである．

Opteron の特徴はマイクロプロセッサがメモリに直接アクセスできること，メモリのデータを読み出す際の遅延が小さいことである．一方，PowerXCell 8i の特徴は IBM 社の Cell プロセッサと同様に 1 個の PPE（Power PC Processor Elements）プロセッサコアと 8 個の SPE（Synergistic Processor Elements）プロセッサコアから成るが，演算性能が Cell に比べて約 5 倍の向上（107GFlops）以上に強化されていることである．以上の特徴から，通常の演算処理やファイルの入出力，通信処理は Opteron に，複雑な処理や繰り返しの処理は PowerXCell 8i に割り当てられる．しかし，複数の種類のプロセッサを使用するため，プログラミングが複雑となることが短所である．

3.3.3 インターコネクト

インターコネクトについては Gigabit Ethernet が主流で，そのシェア率は約 56 %，次いで Infiniband が約 28 %である．一般に InfiniBand より Gigabit Ethernet の方が遅延が大きいが，Gigabit Ethernet のシェア率の方が高い理由は，ラック間のケーブル配線，管理の拡張，ソフトウェアやハードウェアのアップグレードなどを考慮に入れた場合に優れているためである．また，コスト面についても Gigabit Ethernet の方が Infiniband に比べ，安価で入手できる点も理由の一つである．

3.3.4 ストレージシステム（pNFS）

膨大な量のデータを読み書きするためには大容量かつ高速にデータを転送できるストレージの構築が必要である．そこで Roadrunner では新しいストレージシステム pNFS（parallel Network File System）を採用している．従来のシステムの場合，データの送受信は，クライアントとストレージの中間に置かれていたサーバを仲介して行っていた．これは扱うデータ量が増大するとシステムのボトルネックとなる．

pNFS では最初にクライアントはサーバに欲しいデータの情報が書かれたメタデータを渡す．pNFS のアーキテクチャを Fig. 2 に示す．この後，サーバはストレージ上でのデータの格納先を確認し，そのデータに対する処理を行う．pNFS では，クライアントとストレージ間にサーバが介在しないため，直接データのやり取りが行われる．このため，より高速なデータ処理を実現し，ボトルネックを解消することができる．

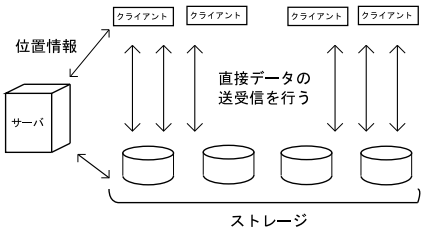

Fig.2 pNFS(参考文献 [8] を参考に自作)

3.3.5 Windows Cluster

Windows Cluster は 2008 年 11 月の Top500 リストにおいて 5 台ランクされているが 10 位以内にランクされるのは初めてである．Windows Cluster がこのように上位にランクされるようになった背景には Windows HPC Server 2008 の登場がある．Windows HPC Server 2008 にはユーザのクラスタ環境を Linpack ベンチマーク向けに自動調整する HPL Wizard などの管理ツールやネットワーク関連の必要最小限の環境のみをインストールする Sever Core 等がある．Server Core を利用すると GUI のグラフィカルな表示やその他アプリケーションなどにリソースを割くことなく，サーバーとしての機能に処理を集中させることができる．

4 日本のスーパーコンピュータ

日本のスーパーコンピュータで最新の Top500 リストにおける最上位は 27 位であるが，かつて日本のスーパーコンピュータである地球シミュレータが 2002 年から 2 年間世界一の演算処理能力として Top500 の首位にランクされ，大規模気象シミュレーション等の様々な分野で応用された．近年, 日本は Top500 リストにおいて再び首位を獲得すべく，国家プロジェクトとして京速コンピュータの開発を行っている．

4.1 京速コンピュータ

京速コンピュータとは 1 秒間に 1 京（10 ペタ＝ 10^{16}）回の浮動小数点演算を行うスーパーコンピュータのことで，文部科学省のプロジェクトとして理化学研究所，富士通，NEC，日立製作所が共同開発しており，科学技術，学術研究，産業，医，薬など広汎な分野で世界をリードし

続けることを目指している．完成後は，生命科学分野においては遺伝子，細胞，臓器などの人体スケールでのシミュレーション，また地球科学分野においては断層モデルの可視化，長期的な気象予測シミュレーション等への利用を目的としている．立地は神戸市の人工島・ポートアイランドであり，2012 年頃完成を目処に開発が進められている．現在，システム開発としてはシステム演算部，制御フロントエンド部，共有ファイルはすべて詳細設計段階であり，今年度から試作，評価に移行される．施設については計算機棟を建設中であり，現在基礎工事段階である．

現在，京速コンピュータ開発にあたり，ハードウェアに関する以下の要素技術の研究開発が実施されている．

- 低電力高速デバイスの研究開発
- 超高速コンピュータ用光インターコネクトの開発
- ペタスケール・システムインターコネクト技術の開発
- 並列コンピュータ内相互結合網 IP 化による実行効率最適化方式の開発

以降，京速コンピュータのシステムの基本構成と複合シミュレーションについて述べた後，低電力高速デバイスと超高速コンピュータ用光インターコネクトの開発について述べる．

4.1.1 システムの基本構成

シミュレーションの多様性に応えるために，京速コンピュータはスカラ型とベクトル型の複合型となっている．これにより，従来困難であった複雑かつ大規模なシミュレーションが実行可能になるとされている．具体的には，ナノテクノロジー分野において現在の平均的なスーパーコンピュータでは 150 年かかるタンパク質の構造解析の計算を京速コンピュータは 6 カ月で実行することができる．また，ナノ電子デバイス解析においては現在 2000 原子程度の解析しか出来ないが京速コンピュータでは 10 万原子の解析まで可能になる．

スカラ型とベクトル型の複合型にはデメリットも存在する．例えば，プログラミングは機種毎に最適化条件が異なるのでアルゴリズムが複雑化する．また，複合型計算機では最も遅い部分が総合性能に大きな影響を及ぼすことを考えると，2 機種複合型の場合は 2 機種とも効率化しないとシステム全体が高効率にならない．つまり，複合型システムを効率よく管理，運用するためのソフトウェアの開発も必要となる．

4.1.2 複合シミュレーション

京速コンピュータはスカラ型とベクトル型の複合型であるため両演算部を同時に使用する複合シミュレーションを行うことが可能である．複合シミュレーションの流れを Fig. 3 に示す．

複合シミュレーションでは，ある時刻ごとに出力される途中結果の逐次的なデータ解析に最適なシミュレーションを実行することができ，各演算部を連携させることで一連のデータ処理の短縮化ができる．さらに，大規模かつ長時間シミュレーションの途中結果をモニタリングすることで計算の中断や，実験のパラメータの変更をすることが可能になり，実験の効率化や資源の有効活用に繋がる．

複合シミュレーションの具体例としては，太陽電池の設計が挙げられる．スカラ部では透明電極材料の電子構造計算，電解質内のヨウ素イオンドリフトの古典分子動力学シミュレーションなどが行われ，ベクトル部では動的量子力学による励起エネルギー，高効率色素分子設計の分子軌道計算などが行われる．

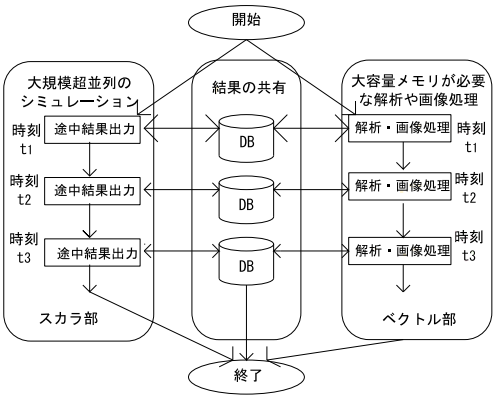

Fig.3 複合シミュレーションの流れ (参考文献 [13] を参考に自作)

4.1.3 低電力高速デバイスの開発

スーパーコンピュータの開発において，高速化の最大の障壁となるのは LSI の消費電力である．プロセッサ中のコア数の増加が計算機の演算処理能力の高速化に繋がるため，スーパーコンピュータを開発する際には半導体の微細化が行われる．しかし，微細化されるに従い，リーク電流が大幅に増大し，消費電力や発熱量の増加や発熱にともなう素子の劣化等を引き起こし，高速化の障害となる．そのため，半導体の消費電力を低減させることがリーク電流対策と高速化の最大の焦点となっている．

京速コンピュータの開発においても，実効性能あたりの LSI 消費電力を低減する技術の研究開発が進められている．従来のデバイスと新構造デバイスを Fig. 4 に示す．その手法は，まず，プレーナ構造（各層が平板状に積み重なる半導体構造）の SOI(Silicon on Insulator) を基本に，埋込み酸化膜を薄膜化して基板電圧によってしきい値電圧（トランジスタがオン状態になる電圧）を制御可能な構造とする．なお，SOI とは基板のチャネルの下に絶縁体を形成して，リーク電流による電子回路の誤作動を防ぐ技術のことである．そして，動作パターンに合わせてゲート電極とシリコン基板との間に電位差を生じさせることで細かい電圧制御を実現する．これは絶縁体が薄い場合のみ実現可能となる．しきい値電圧を制御することで製造後にプロセスや動作条件のばらつきが原因で生じるリーク電流を抑止することができる．この新構造トランジスタは SOTB（Silicon on Thin Buried Oxide) と呼ばれている．SOTB を採用すると LSI の消費電力を 2 分の 1 から 3 分の 1 に低減される．

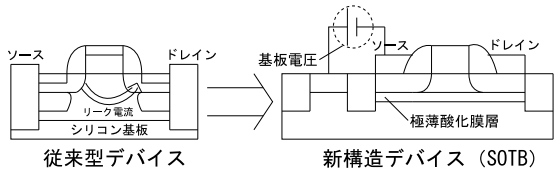

Fig.4　従来デバイスと SOTB デバイス (参考文献 [9]) を参考に自作)

4.1.4　光インターコネクト技術の開発

実効性能で PFlops 級のマシンを実現するに当たり，数千〜数万台規模の計算ノードを相互接続するシステムインターコネクトの高性能化は最重要課題のうちの一つである．そこで，研究開発されているのが光インターコネクト技術である．光インターコネクトとは電気信号に代わって，光で信号を伝送する技術のことで，次世代スーパーコンピュータでは，1 信号あたり 20Gbps の速度と，高密度実装化で超高速の伝送性能を目指している．

従来の LAN 等の電気スイッチと現在開発されている光パケットスイッチのアーキテクチャを Fig. 5 に示す．スーパーコンピュータには多くの電気スイッチが使用されているが，京速を実現させようとすると電気スイッチやケーブルの数が膨大となるため，消費電力が増加するといった問題が引き起こされる．しかし，光スイッチでは波長合波器において複数の波長の光信号を合成し，一つの光信号として送信するので，ケーブル本数を削減すること，また，一括スイッチによりスイッチ数，光電気変換モジュールを削減することが可能となる．ここで，スイッチングは，送信側で光信号に付与した符号ラベルと同じラベルを用いて復号処理を行った場合には高い信号ピークレベルの信号が出力されるが，違う符号ラベルを用いて復号を行うと信号レベルは低いままとなることから，信号レベルが高いときのみ当該信号を出力ポートに流すという手法によって実現される．

以上により，消費電力を電気スイッチのみを用いた場合の 3 分の 2 にまで低減できる．さらに，現在の電気伝送技術においては 5〜10Gbps が限界（多重信号，数十 cm 程度の伝送において）であるが，光伝送では 20Gbps 以上の高速化を実現することが可能となる．

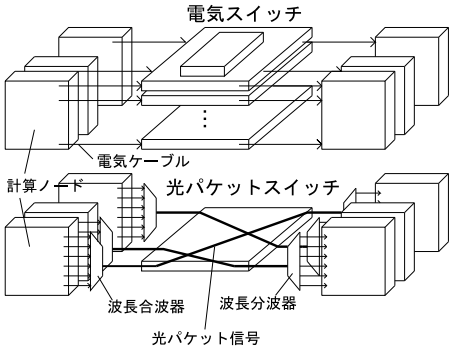

Fig.5　光インターコネクト (参考文献 [11]) を参考に自作)

5　今後の展望

日本以外でも京速コンピュータの開発が進められており，2009 年 2 月 3 日に米国エネルギー省の NNSA と IBM がスーパーコンピュータ Sequoia の開発を契約した．Sequoia のピーク浮動小数点演算性能は 20.13PFlops，計算ノード数は 98304 ノード，プロセサコア数は 160 万，メインメモリ量は 1.6PB と発表されている．発表に次世代 BlueGene という表現があることから Sequoia は BlueGene/Q システムと考えられ，日本の京速コンピュータと Sequoia との Top500 リストの首位争いが予想される．しかし，2009 年 5 月 13 日に京速コンピュータの開発から NEC と日立製作所が離脱することが発表がされた．NEC は本体製造に関する相当額の費用が今年度の業績に多大な影響を与えるためと説明している．NEC の撤退により，京速コンピュータは富士通によるスカラ型のみのシステムへ変更される可能性が高く，これからの動向が注目される．

スーパーコンピュータの性能については半導体以外の光インターコネクト，3 次元チップ，アクセラレータベース処理の技術向上が考えられる．将来これらの技術によりペタスケールを超え，エクサスケールの演算処理能力を持つスーパーコンピュータの開発が予想される．

参考文献

1) Cisco Systems,Inc
http://www.cisco.com/web/JP/solution/large_enterprise/enterprisearchitectures/datacenter/whitepaper/hcpnw_wp.html
2) HPCwire
http://www.hpcwire.com/features/Lawrence-Livermore-Prepares-for-20-
3) IT 用語辞典
http://e-words.jp/
4) Microsoft PressPass
http://www.microsoft.com/japan/presspass/detail.aspx?newsid=3588
5) Top500
http://www.top500.org/overtime/list/32/countries
6) YOMIURI ONLINE
http://www.yomiuri.co.jp/net/news/cnet/20090410-0YT8T00689.htm
7) スーパーコンピューティングの国家戦略
http://www.nsc.riken.jp/sympo2007/slide/seisaku-kouen.pdf
8) スケーラブルシステムズ株式会社
http://www.hp2c.biz/doc/HP2C.Biz/hp2c_panasas_pNFS.html
9) 低電力高速デバイス・回路技術・論理方式の研究開発
http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu2/006/shiryo/08032418/003.pdf
10) ペタ・スケール・コンピューティングに向けて
http://www.ksrp.or.jp/fais/sec/cell/fais/news/pdf/kenkyutheme/sanuki_workshop2008_v2.pdf
11) ペタスケール・システムインターコネクト技術の開発
http://www.psi-project.jp/images/event/ryouikimeeting_20060530.pdf
12) マイコミジャーナル
http://journal.mycom.co.jp/photo/column/architecture/150/images/002l.jpg
13) 理化学研究所提出資料 2
http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu2/toushin/07061321/002.pdf