

# エクサスケール・コンピューティングへの挑戦

小柳義夫 (神戸大学)

理化学研究所計算科学研究機構により神戸に建設された「京コンピュータ」は、2011年に10.51 PFlopsという世界一の性能をマークした後、2012年から供用開始され、さまざまな科学技術上の問題解決に活用されている。しかし精度を上げ複雑な対象に対応するには、すでに性能の不足が痛感されている。現在、2020年頃を目指して「京」の100倍程度の実効性能を持つエクサスケールのコンピュータを建設するプロジェクトが、アメリカ、ヨーロッパ、中国などで進められており、日本でもその動きが始まっている。社会的な制約から、「京」と同程度の予算、消費電力、床面積で建設することが求められているので、直面する技術的チャレンジは大きい。半導体技術の進歩は急速であるが、エクサスケールは市場の動向だけで実現する性能ではない。本稿では、どのような技術的なチャレンジがあり、それにどう立ち向かおうとしているか、また、そのような巨大な計算性能でどんな社会的・学術的な課題を解決しようとしているのか、世界の動きとあわせて解説する。

## 計算科学技術をめぐる状況

最初の電子式計算機が出現したのは1940年代後半であるが、それから長足の進歩を遂げ、以来60年余りの間に約13桁の飛躍的性能向上を実現し、科学技術全体に大きな影響を与えた。コンピュータの高性能化により、大規模な計算によるシミュレーションは現代の科学技術における重要な手法の1つとなっている。コンピュータによる大規模計算という新しい方法論によって得られる知の体系を計算科学技術と呼ぶ。

### ▶ 計算科学技術の意義

計算科学技術の進展は広い分野に及んでいる。気象・気候や地震・津波といった自然現象、実験できないような高温、高圧、高速、長時間などの極限環境での現象、生命現象のように複雑な現象、星や銀河の形成といったスケールの違う現象などを、基本法則によりコンピュータ上に実現することが可能になる。これにより、現象を理解し、未来を予測するばかりか、現実には起こらない状況さえ作り出すことができる。また、種々の物質材料における原子や分子の挙動を、かなり大きな系まで量子力学によりシミュレーションできるようになり、新しい材料の設計や薬の候補物質の探索に用いられている。

基礎研究のみならず、産業界のものづくりの現場において、設計、試作、試験、評価のプロセスの一部をシミュレーションで代替することにより、製品の設計サイクルを速めるとともに、創造性、生産性を高めることが可能になりつつある。また、交通、市場、伝染病のような社会現象の解析にも用いられる。

このように、大規模なコンピュータを利用した計算科学技術は、学術の振興のみならず、産業競争力の強化や、安心な国作りの観点からも重要な方法論となっており、その重要性は今後さらに高まるものと考えられる。以下、性能は特記のない限りLinpackベンチマークの性能で示すが、今後は、Linpackベンチマークの性能と実アプリケーションの性能とが「京」以上に乖離することが想定されるので、「エクサフロップス」ではなく「エクサスケール」というぼかした表現がよく用いられる。

### ▶ エクサスケール・コンピュータはなぜ必要か

2013年11月現在、ペタフロップス (PFlops =

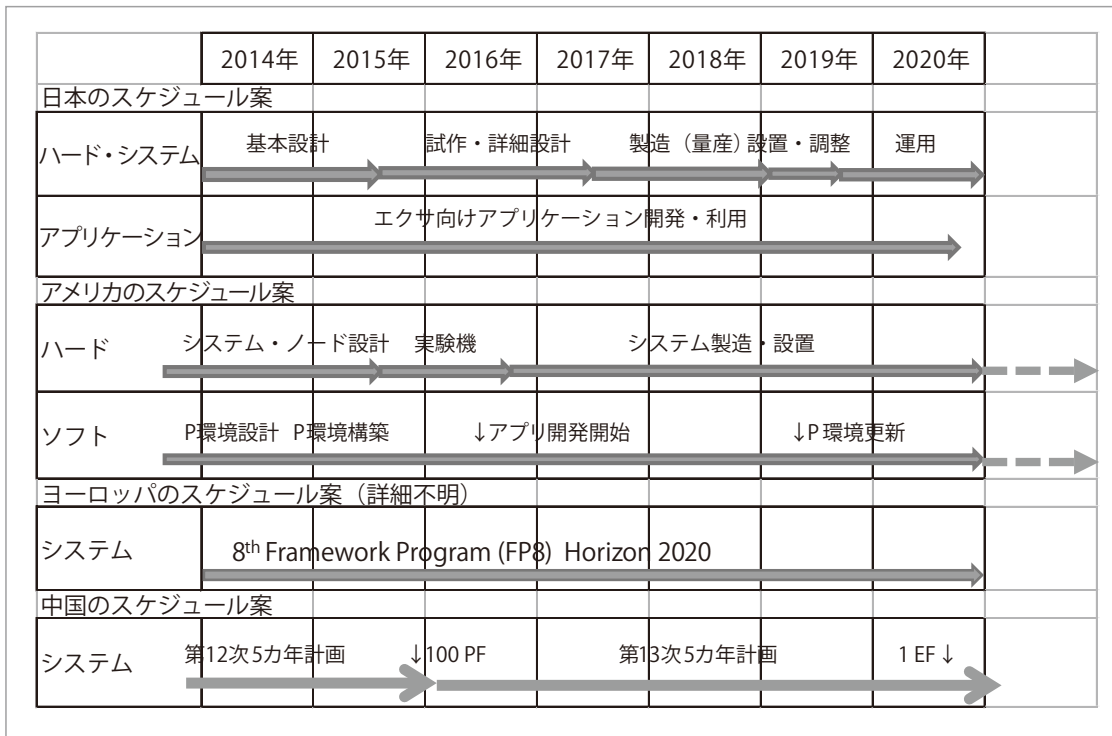


図-1 各国のエクサスケールに向けての開発計画。いずれも提案段階で、確定したものではない。

10<sup>15</sup> Flops) を超えるスーパーコンピュータは世界に30台以上設置され、さまざまな研究開発に利用されているが、まだまだ十分ではない。たとえば、870mメッシュの全球気象シミュレーションは、雲の構造まで入れて計算できるので正確な予報ができるが、「京」全体を使っても1日分の計算に20時間ほどかかり気象予報には使えない。少なくとも10倍以上実効性能の高いコンピュータが必要である。津波の浸水予測は「京」で1.5日かかるが、1時間以内で実行できることが要求される。また「京」で40日を要する自動車の衝突解析は10時間以内で可能となれば、衝突試験のコストや時間が大幅に削減できる。創薬においては、新薬のスクリーニングのための分子動力学計算のさらなる高速化が必要とされている。

このように、「京」によるシミュレーションは、基礎科学から産業界まで多くの成果を創出しているが、計算能力への要求は止まるところを知らない。

## 世界の状況

計算科学技術の重要性が増加している中で、スーパーコンピュータの開発・利用は世界的に進められている<sup>1)</sup>。図-1に各国のエクサスケールを目指す計

画案を示す。いずれも提案段階で、承認された最終案ではない。年号も暦年と会計年度が混ざっている。

### ▶ アメリカの状況

1980年代の日本のベクトルコンピュータの技術開発に驚いたアメリカは、1991年にAl Goreの提唱によりThe High Performance Computing Act of 1991を成立させ、国家的投資によってスーパーコンピュータを開発設置するHPCC (High Performance Computing and Communication) 計画が発足した。以来アメリカは、世界のスーパーコンピュータの総計算性能の約半分を定常的に占めている。

またアメリカは世界トップのシステムを組織的に開発してきている。ペタフロップスを目指す最初の会議がPasadenaで開かれたのは1994年2月であるが、この頃の世界最高性能が航空宇宙技術研究所(日本)のNWT(「数値風洞」)の124 GFlops (GFlops=10<sup>9</sup> Flops)であったことを考えると、約1万倍速いコンピュータを構想していたことになる。その先見の明には驚かされる。1 PFlopsを現実に超えたのは14年後の2008年6月のRoadrunner (Los Alamos 国立研究所, 1.026 PFlops)であった。ちなみに日本で最初に1 PFlopsを超えたのは、東工大

の TSUBAME 2.0 (2010) である。

さらにアメリカは、エクサフロップス (EFlops = 1,000 PFlops) を目指す会議を 2004 年 10 月に Santa Fe で開催しているが、この頃は日本の「地球シミュレータ」の 35.86 TFlops (TFlops=10<sup>12</sup> Flops) が世界の最高性能であり、約 3 万倍速いコンピュータを構想していた。ちなみに日本では、このころやっと文部科学省情報科学技術委員会が次世代スーパーコンピューティング技術の検討 (これが「京コンピュータ」に結実) に入ったところであった。日米の長期戦略の落差は歴然である。アメリカの研究者は図-1 のように 2020 年にエクサスケールの実現を計画しているが、連邦議会は 2013 年末、エネルギー省が 10 年計画で高性能なコンピュータを開発する計画を承認した。アメリカの計画ではソフトにおいてプログラミング (P) 環境の構築に重点が置かれているのが特徴である。

### ▶ ヨーロッパの状況

ヨーロッパでは、国家間連携によりスーパーコンピュータの整備・利用が進められており、世界全体の約 20% を占めている。特に 2008 年にはヨーロッパ各国のスーパーコンピュータを連携利用する PRACE (Partnership for Advanced Computing in Europe) という枠組みを構築し、その中でペタフロップス級のシステムの整備を行っている。また 2010 年～2011 年には EESI (European Exascale Software Initiative) 1 を実施し、引き続いて EESI2 においてエクサスケール実現の課題抽出とロードマップ作成を継続している。EU は 2014 年から 8th Framework Program (FP8), 別名 Horizon 2020 を進めており、その中でエクサスケールの実現を目指すようであるが、具体的な取り組みは検討中である。以前から Mont Blanc, DEEP, CRESTA などエクサスケール技術のためのハードやソフトのプロジェクトを進めている。

### ▶ 中国等の状況

中国は、1986 年 3 月に鄧小平国家主席の主導で決定された「高技術研究発展計画要領 (通称 863 計

画)」に基づき、近年急速にスーパーコンピュータの開発・利用を進めており、2013 年 6 月には天河 2 号が 33.9 PFlops で世界一となった。性能合計も世界全体に対して 20% を占めている。今後、第 12 次 5 年計画が完了する 2015 年までに 100 PFlops、第 13 次 5 年計画の終わる 2020 年までにエクサスケールの実現を目指している。

このほか、ロシアやインドなどでもスーパーコンピュータの整備や利用が積極的に進められている。

## 日本の取り組み

### ▶ 日本のこれまでの状況

日本では、2006 年度から「京」の開発・整備を行い、「京」を中核として国内の大学等のコンピュータやストレージを高速ネットワークで結び、多様なニーズに応える HPCI (High Performance Computing Infrastructure) の構築とその利用促進を図ってきた。「京」は 2011 年 11 月に 10.51 PFlops を達成するとともに、優れた実効効率や信頼性を実現した。「京」を含む HPCI は 2012 年 9 月に運用を開始し、成果を出し始めている。2011 年に策定された第 4 期科学技術基本計画においても、世界最高水準の HPC 技術を国家安全保障・基幹技術として位置づけ、国として強力に推進していくことが謳われている。

たしかに日本の計算資源の使用量は年率 1.8 倍の伸びで増加している。しかし、世界の総計算資源の中での日本の割合は、「地球シミュレータ」や「京」などの出現により一時的に増加するものの、長期的には減少傾向になっている。2013 年 11 月の Top500 によると、日本の性能総和は 22 PFlops で世界の 9% しかない。今後の計算科学技術のインフラを全体としてどのように維持・発展させていくか、エクサスケールを目指してどのように取り組んでいくかが重要な課題である。このような状況のなかで、日本でもエクサスケールを目指す活動が遅まきながら始まった。

### ▶ エクサスケールへの研究者の動き

2008 年 11 月に Austin で開催された SC08 国際



会議において、アメリカを中心に、日本やヨーロッパが加わって IESP (International Exascale Software Project) が発足した。これはエクサスケールを目指して、OS、コンパイラ、ミドルウェアなどのシステム・ソフトウェアを共同で開発するプロジェクトである。スーパーコンピュータの開発において、ハードウェアはベンダ間、国家間の競争が激しく、それぞれの手法で世界最高性能を目指す、システム・ソフトウェアについては共同で開発することにより、人手やコストを削減することを期待している。日本の研究者がエクサスケールの活動を始めたのはこの頃であろう。IESP はその後、アメリカ、ヨーロッパ、日本で数回の会議を開催した。

2010年には、石川裕(東大)らを中心に SDHPC(戦略的高性能計算システム開発) ワークショップが始まり、これまで10回の会合を重ねている。5年後の HPC 像についての自由な討論から始まったが、実質的にはエクサスケールを目指すものであった。

### ▶ エクサスケールへの政府の動き

HPCI 計画や HPCI 戦略プログラムなどを推進してきた文部科学省の HPCI 計画推進委員会では、2011年4月に「今後の HPC 技術の研究開発の検討ワーキンググループ」を設置し、内外の研究開発状況の調査と今後の課題の検討を行った。7回の会議の検討結果を踏まえ、7月から「アプリケーション作業部会」と「コンピュータアーキテクチャ・コンパイラ・システムソフトウェア作業部会」を、若手世代を中心に組織することとした。前者では、今後5～10年の科学的、社会的に重要な課題と、それを解決するために必要な計算機資源について詳しく分析し、後者は、将来の HPC アーキテクチャの可能性を技術的に分析した。短い時間であったが、集中的に検討を進め、翌年3月には「今後の HPC 技術開発に関する報告書」<sup>2)</sup> および2つのロードマップ白書<sup>3)</sup>を公開した。これは日本でエクサスケールを目指したはじめての公式プロジェクトであった。

この成果の上に文部科学省は、10年後の日本の社会的・科学的課題を予測し、その解決を目指す

HPC システムの高度化に必要な技術的知見を獲得するための調査研究 (Feasibility Study, 通称 FS) を2012年度と2013年度に実施することとなった。公募に基づき、アプリケーション1チームと、システム設計研究3チームが選定され、活動を行っている。筆者はそれまで、エクサスケールにおいては応用分野ごとにその特徴に合った別々のコンピュータを設計しなければならないのではないかと漠然と考えていたが、FSなどの研究の結果、全部と言わないまでもある程度広い分野をカバーする「汎用的」なコンピュータ・アーキテクチャという方向性が見えてきたように思われる。しかしエクサスケールでは実効性能がアプリケーションによって大きく異なるので、一律に「京」の100倍とはならず、それぞれのアプリケーションがどのような性能で実行できるかはアーキテクチャに依存する。

### ▶ ワーキンググループでの検討

同時に、2012年2月の HPCI 計画推進委員会において、下部組織として、「今後の HPCI 計画推進のあり方に関するワーキンググループ」(WG)を設置することを決定し、筆者が座長に指名された。WGは2013年度末に報告書を提出する予定であるが、1年余りの議論を経た段階で、2013年5月に中間報告案をまとめ、パブリックコメントを経て、7月に中間報告を公表した<sup>4)</sup>。中間報告では、フラッグシップシステムを中心とするリーディングシステム群を開発する方向を勧告している。

### ▶ リーディングシステムとフラッグシップシステム

「京」やその他のスーパーコンピュータは多くの成果を出してきたが、計算資源に対する利用者のニーズは高く、今後とも増大し、しかも多様化していくことが予想される。そのため、日本の計算科学技術を支える計算資源としては、世界トップレベルのスーパーコンピュータとともに、次のレベルのスーパーコンピュータを複層的に配置し、全体として多様なニーズに対応するグランドデザインを建てなければならない。振り返ると、「地球シミュレータ」

(2002) は当時世界で断然トップのスーパーコンピュータであったが、それに続く計算資源の開発設置には遅れをとったと言わなければならない。

図-2に、日本の計算科学技術システムのグランドデザインの概念図を示す。まず日本の計算科学全体を牽引し、科学技術の新たな展開を切り拓いていくシステムとして、世界トップレベルの高い性能を持ち、最先端の技術を利用して開発されるいくつかのシステムを、「リーディングシステム」として整備していくことが必要である。これは、日本の最先端システムとして、さまざまな分野における先端的な大規模計算を効率的に実行するために、広い分野での高い実効性能とメモリ容量を有していることが求められる。

同時に計算科学技術の諸分野における、演算性能、メモリ容量、メモリバンド幅に対する要求には幅があり、後述するような制約を考えると、1つのシステムだけですべての分野のさまざまな計算ニーズに的確に対応することが難しくなり、複数の特徴あるマシン群を求める意見もある。このため、高い計算性能を持ち、できるだけ幅広い分野をカバーする1つのシステムを「フラッグシップ」として整備・運用するとともに、フラッグシップでは実行性能の低い応用の処理性能を向上させる特徴的あるシステムを、リーディングシステムとして開発・整備することをも視野に入れている。たとえば、フラッグシップシステムが演算性能を重視したシステムとすると、メモリバンド幅を重視したシステムや、ビッグデータ解析に対応するシステムなどが、これを補完するリーディングシステムとして考えられる。

これに基づき、文部科学省は2020年頃までの実現を目指して、フラッグシップシステムの開発計画を、2014年度予算の概算要求として提出し、政府原案では12億円が認められた。図-1の日本の欄は概算要求書にあるスケジュールを示した。

## 解決すべき課題

日本を含め各国ともエクサスケール・コンピュータを、現在用いられているCMOS半導体技術の

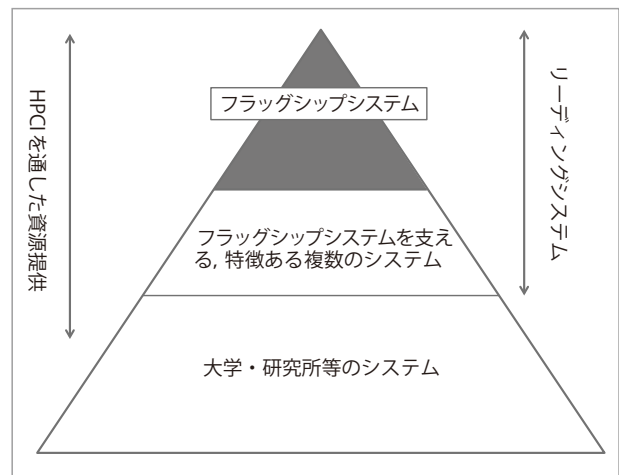


図-2 日本の計算科学技術システム開発・利用のグランドデザイン

延長により実現することを目指している。実現しなかったものの、1990年代のベタフロップス構想においては超伝導素子などのいわゆる破壊的技術(disruptive technology)を選択肢の1つとしていたのと対照的である。2020年頃には市場の半導体技術は10nmのレベルに達していると期待されるが、それでもエクサスケールの実現には多くの課題を解決しなくてはならない<sup>5)</sup>。代表的なものを挙げる。

### ▶ 消費電力の壁

「京」の消費電力は12.7MWである<sup>☆1</sup>が、そのままエクサスケールに外挿すれば1GWを超えてしまう。経済的にも環境保護の観点からも現在の電力を大幅に超えることは許されない。これは消費電力当たりの演算性能を100倍近く増強することを意味する。半導体技術の進歩を考えてもこの実現は容易ではない。現在の最先端のコンピュータにおいては、演算で消費される電力よりも、チップ間やノード間のデータ移動の電力の方が多く、エクサスケールではチップ内のデータ移動も演算より高コストになると予想される。データ移動を抑えるアーキテクチャやソフトウェア技術が重要になる。

現在の汎用プロセッサは、高いクロック周波数(Intel Xeonでは3~4GHz)、深いパイプライン構成、

☆1 大型火力発電所や原子力発電所の出力は1基1GW前後なのでその1%強にあたる。なお「京」ではバックアップ電源を兼ねて一部ガスタービンで発電し、廃熱を冷却に用いている(コジェネ)。

アウトオブオーダー実行などを活用し、レイテンシを最小にする構成をとっているが、大きな電力を消費する。これに対し GPU などの演算加速機構は、周波数を低くし (NVIDIA Tesla では 600 ~ 700 MHz)、パイプラインを浅くし、データを近くに置いて、インオーダーでマルチスレッド実行を行うことにより、レイテンシよりもスループットを重視しているの、演算当たりの電力は小さい。エクサスケールでは後者のアーキテクチャを活用せざるを得ない。

### ▶ メモリの壁

回路の細密化によってプロセッサチップ当たりの演算性能は増大するが、メモリとプロセッサとの間のバンド幅はチップのピン数で制限される。バンド幅を総演算能力(ピーク値)で割った数値を B/F (Byte per Flop) と呼び、1 演算当たり何バイトの入出力が可能かを示す。「地球シミュレータ」では 4 であったが、「京」では 0.5 に減少している。エクサスケールではさらに減る可能性があり、利用の際の大きな制限になる。光入出力や 3 次元積層でバンド幅そのものを増やす可能性、プロセッサチップ内部に大きなキャッシュやバッファメモリを置いて、チップへの入出力を減らす可能性などが検討されている。

### ▶ 信頼性の壁

一般に故障の時間当たり発生率は回路の素子数に比例するので、エクサスケールでは大きな問題になる。「京」では高度な品質管理、実装技術、水冷による低温作動などの技術により、致命的な故障はかなり抑えられている (影響は 1% 以下)。エクサスケールでは、ハードウェアレベルの故障を減少させ、回避、回復させることはもちろん、OS やミドルウェアや応用ソフトウェアなどにおいても、耐故障性を高める技術を開発する必要がある。1 億次元の Linpack ベンチマークを 1 EFlops のコンピュータで実行すると 1 週間強かかる<sup>☆2</sup>が、エクサスケ

☆2 未知数  $n$  の連立 1 次方程式求解の演算量は  $2n^3/3$  なので、 $n=10^8$  なら演算量は  $6.67 \times 10^{23}$  である。これを 1 EFlops =  $10^{18}$  Flops で割ると、 $6.67 \times 10^5$  秒、すなわち 7.7 日となる。

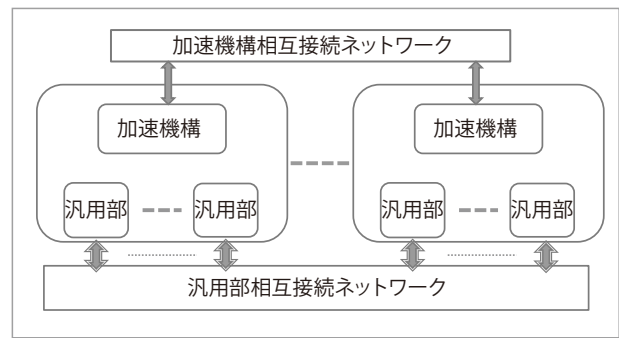


図-3 フラッグシップシステムのアーキテクチャ概念図

ールでは 1 週間無故障で稼働させることは困難ではないかと考えられている。

### ▶ プログラミングの壁

「京」の全系を利用するとスレッド数は 60 万以上になるが、エクサスケールではその 100 倍に及ぶことが想定されている。さらにメモリ階層の段数も増加する傾向にある。このようなシステムで効率よく動くプログラムをどう書くか、どこまでコンパイラなどシステム側が面倒を見て、どこまでユーザーが書けばいいのか、大きな問題である。MPI プログラミングがこの領域まで適用できるかについても疑問が出されている。

### ▶ フラッグシップシステムの方向性

設計の詳細はこれからであるが、以上の考察から汎用プロセッサと加速機構を併用したアーキテクチャが想定されている。加速機構のアーキテクチャは現在検討中であるが、GPU や Intel Xeon Phi のようなメニーコア、ClearSpeed のような演算器アレーなど種々の可能性がある。相互接続ネットワークとしては、汎用プロセッサを接続するネットワークとともに、加速機構を直接結合するネットワークも必要となるであろう。公表されている概算要求資料から推定される概念図を図-3 に示す。問題は汎用プロセッサと加速機構の性能比である。FS における検討によれば、加速機構で性能が出ない応用プログラムもあり、過度に加速機構に頼るアーキテクチャでは広い応用で活躍できない。たとえば、中国の天河 2 号では、ピーク性能が汎用 1 に対し加速機構 (Xeon



Phi) が7となっており、加速機構の割合が大きい。多くの応用での加速機構の活用が課題となる。消費電力も考慮の上、適切な設計が必要である。

消費電力は30～40 MWと予想されている。このようなハードウェアを使いこなすためにはプログラミング環境の設計が重要であるが、この1台だけではなく、将来にわたって有効な統一のプログラミングモデルを提供する必要がある。同時に、エクサスケール・システムを活用して世界最高水準の研究成果を創出するために、アプリケーションを開発するとともに、革新的なアルゴリズムに取り組む必要がある。

## 展望

エクサスケール・コンピュータは各国で構想されているが、日本では、第四期科学技術基本計画において、世界最高水準のスーパーコンピュータを国家基幹技術として位置づけ、国として戦略的に開発整備することが計画されている。これにより、科学技術イノベーションに適した環境を創出し、さまざま

な科学的・社会的課題の解決を先導し、科学技術の振興、産業競争力の強化、安全・安心を保障する国作り等に貢献することを目指している。

### 参考文献

- 1) 宇川 彰 他：岩波講座計算科学，第1巻，計算の科学，付録A「スーパーコンピュータの歩み」岩波書店（2013）。
- 2) 今後のHPCI技術開発に関する報告書，<http://open-supercomputer.org/wp-content/uploads/2012/03/FutureHPCI-Report.pdf>
- 3) 計算科学研究ロードマップ白書，<http://open-supercomputer.org/wp-content/uploads/2012/03/science-roadmap.pdf>  
HPCI技術ロードマップ白書，<http://open-supercomputer.org/wp-content/uploads/2012/03/hpci-roadmap.pdf>
- 4) 今後のHPCI計画推進の在り方について（中間報告），[http://www.mext.go.jp/b\\_menu/shingi/chousa/shinkou/028/gaiyou/1337595.htm](http://www.mext.go.jp/b_menu/shingi/chousa/shinkou/028/gaiyou/1337595.htm)
- 5) 小柳義夫 他：岩波講座計算科学，別巻，スーパーコンピュータ，第4章「スパコンの進化とエクサフロップスに向けた今後の課題」岩波書店（2012）。

（2013年10月30日受付）

小柳義夫（正会員） oyanagi@people.kobe-u.ac.jp

昭和46年東京大学大学院博士課程修了，理学博士。同大助手，高エネルギー物理学研究所助手，筑波大学講師，助教授，教授を経て，平成3年より東京大学理学部情報科学科教授。平成18年より工科院大学情報学部教授・学部長，東京大学名誉教授。平成23年より現職。

### 【用語集】

- [Linpack ベンチマーク]：密行列係数の連立1次方程式（サイズは任意）を Gauss の消去法で解く問題でコンピュータの浮動小数演算性能を測定するベンチマーク。倍精度演算（64ビット）で行う。
- [Flops (floating operations per second)]：演算速度の単位。倍精度浮動小数の加減算・乗算を1秒間に何回実行できるかを示す。演算器が全部作動すると仮定した場合をピーク性能または理論性能，Linpack ベンチマークで測定した場合を Linpack 性能という。
- [CMOS (Complementary MOS)]：半導体デジタル回路において MOSFET を相補型に配置したゲート構造。初期のメインフレームやベクトル計算機で用いられたバイポーラより消費電力が少ないので広く用いられているが，現在では高速化，高密度化により消費電力や発熱量が再び限界に近づいている。
- [アウトオブオーダー実行]：プログラムの命令の順序ではなく，結果が変わらない範囲で順序を変更して実行を高速化すること。プログラムの順序のまま実行することをインオーダー実行という。
- [レイテンシ (演算の)]：ある命令を発行してから実行結果が得られるまでの待ち時間。
- [スループット (演算の)]：多数の命令群を実行したとき，時間当たり平均的に実行される命令数。
- [MPI (Message Passing Interface)]：分散メモリ並列処理においてメッセージ通信を行うための標準規格。1対1通信を基本としているのでノード数が巨大になると実装が困難になる。
- [GPU (graphic processing unit)]：元来，PC や携帯機器の画像生成のための専用チップを意味するが，本稿ではそれを一般的な科学技術計算用にさらに高速化したものを指す。その意味で GPGPU (general purpose GPU) とも呼ばれる。