

《特 集》

2010年：シミュレーションの旅

—コモディティ技術による PetaFLOPS コンピューティングへの道しるべ—

松 岡 聡*

ABSTRACT Commodity High-Performance Computing which utilizes commodity computing building blocks for high-performance computing is expected to reduce the cost of computing by a factor of over ten thousand over the next ten years, implementing the so-called Petaflops computing as well as making Terascale computing prevalent. As a result, simulations of unprecedented scale or resolution will become possible, making the role of simulation ever more important in science and technology. We attempt to predict the advances of computing power by exploring the technical trends, and investigate how such advances will affect to revolutionize simulation.

1. はじめに

筆者は、ソフトウェア技術、特に並列処理ソフトウェア、および高性能オブジェクト指向ソフトウェアの研究に従事してきた。近年では、「コモディティ高性能計算」を旗印に、大規模超並列計算が普遍的に提供される計算基盤環境を目指し、幾つかのプロジェクトを推進している。これらに関しては後に詳しく述べるが、全て今後10年間の「超 Moore 状態」の計算パワーの指数的な進化に対応する超高性能ソフトウェア技術である。これらは汎用技術によるコモディティな高性能計算機技術、および超高速ネットワーク技術に支えられており、今後10年間に計算のコストを数万分の1にして、PetaFLOPS 級の計算を実現し、さらに TeraFLOPS 級の計算を一般的にする、と予測される。これにより、シミュレーションの世界でも従来には考えられなかった大規模あるいは高精密な計算が可能となり、これによりシミュレーションの役割がさらに重要になると考えられる。

本稿では、そのような立場から、今後数年間の計算パワーの進歩の予測と、それを可能にする新しい技術トレンドに関して述べ、かつ、それらがシミュレーション技術に与える革命的な影響の可能性を議論した

い。筆者自身は詳しいシミュレーション技術自体は専門ではないが、本稿が莫大な計算パワーを生かすべく新たなシミュレーション技術の研究の展開の火種ともなれば大変幸いである。

基本的に、計算性能の向上は PC レベルのコモディティマイクロプロセッサとネットワークの激的な速度向上、それによってもたらされる超巨大超並列 PetaFLOPS 級マシンおよびコモディティクラスターの普及、さらに超高速インターネット/WAN 技術によってもたらされる“Grid”計算（超広域高性能計算）である。どれも、従来の高性能ベクトルマシンに代表される、特定の計算（特にベクトル計算）を高価な技術を用いて高速化する、いわゆる「スパコン技術」ではなく、本来はより日常生活を支える大量の電子工業生産品と通信インフラ、いわゆるコモディティ技術の大胆かつ大規模な活用技術が中心となっている。それゆえ、進歩が急激になっており、シミュレーションもそのトレンドに乗じることで、莫大な計算パワーを安価に手に入れることができる。

2. 計算性能向上の展望：Moore の法則と「超 Moore 状態」によるコモディティ計算技術の躍進

ご存知の方も多いと思うが、計算機の世界では米国 Intel 社の創始者である *Gordon Moore* が1960年代に唱えた「Moore の法則」を満たしながら進化してきた。Moore の法則は「計算機の性能は3年間で4倍になる」という非常に単純なものである。Moore の

2010: A Simulation Roadmap –A Road To PetaFLOPS Using Commodity Technology. By *Satoshi Matsuoka* (Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology).

*東京工業大学大学院情報理工学研究所/JST

法則は指数則であり、かつ経験則である：つまり、当然無限の性能向上は（少なくとも現状の計算機テクノロジーの延長線上では）物理限界からどこかで破綻するのは間違いないのであるが、それが過去30年間以上、つまり約 2^{20} 倍以上の性能向上が人為的に技術の進歩によって維持されて来た。世界最初の計算機は1946年に米国で開発された ENIAC であると言われており、ちょっと前までは「ENIAC が今やマッチの頭程度に凝縮されている」と言われたものであるが、現在ではそれどころか、数 cm 四方に1億トランジスタ以上、つまり、ENIAC が数千台以上入り、さらにそれが Moore の法則に従って進歩しつづけている。

さて、高性能な並列計算を成立させるには、(1)実際に計算を行うための高速なプロセッサ技術、(2)データを供給するための高バンド幅のメモリ技術、および(3)並列処理時にプロセッサ間で高速にデータを交換するための低レイテンシ、高バンド幅のネットワーク技術の3点が重要である。従来はこれらのハイエンドな技術はスパコンなど、高価な特殊ハードウェアに限られてきたが、近年ではハイエンドとコモディティの技術格差が大幅に縮まるか、あるいはほとんどなくなる状態になってきた。これらを以下に順に説明する。

まず、(1)のプロセッサ技術では、近年ではコモディティプロセッサの速度向上に対する激烈な競争と技術革新により、いわゆる「超 Moore 状態」になっているのが現状である。過去には、シミュレーションが主に必要とする浮動小数点演算性能に関しては、高額な、いわゆるベクトル型計算機（スパコン）の独壇場であったが、ここ数年では様々なニーズと技術革新に支えられ、PCなどに用いられるコモディティプロセッサの浮動小数点演算性能（FP 性能）は Moore の法則を超える速度で向上してきた。通常の x86 系のプロセッサの計算速度は、わずか6年前では Intel 486DX2-66 が数十 MFLOPS 程度であったが、現在 AMD Athlon 1.2 GHz では軽く 2.4 GFLOPS を超える；これは数年前の単体の Cray C90 を超えるものであり、最新型のベクトル型計算機のピーク比較してもせいぜい数分の一である。しかもこの秋に登場する Intel Pentium 4 では、ピーク FP 性能は 3-4 GFLOPS に達する。また多少高額になるが、Itanium (Intel), Alpha (Compaq), Power 4 (IBM), UltraIII (SUN) など、ワークステーション (WS) 級の CPU も同様あるいはそれを超えた性能を発揮する。

しかも、半導体技術と計算機アーキテクチャの進歩により、このような劇的な性能向上が少なくとも今後

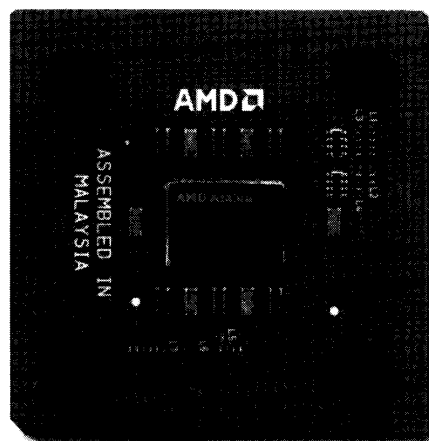


図1 AMD Athlon Processor 1.2 GHz

10年間継続的に起きうる、と予測されている。例えば、現状においてわが国で最速の計算機は高エネルギー物理学研究所にある日立の SR8000-F1 スパコンであり、ピーク性能は 1 TeraFLOPS 強である。しかしコモディティプロセッサ単体の倍精度演算のピーク性能は、2002年後半には最高で 8 GFLOPS 以上であると予測されている。したがって、2002年の時点で、PC を密結合した百数十 node 程度のいわゆる PC クラスタ計算機においてピーク性能は「軽く」TeraFLOPS を超える可能性が高く、そのコストは高々数千万円程度であると予測される。つまり、本稿が出版されて1-2年以内に、TeraFLOPS 級の計算機が1/100程度のコストパフォーマンスの低下によって大学の学科や研究室レベルでも手が届くようになるのである。これは、現状の安価な PC やワークステーションレベルの計算性能の約1000倍以上に達する。

プロセッサのピーク性能が上昇しても、(2)のメモリバンド幅が向上しなければ、限られたアルゴリズムしか性能向上しない。かのスパコンの代表格である Cray 計算機の開発者の故 Seymour Cray は、HPC アーキテクチャの肝は「It's the Bandwidth」であると断言したことでも有名である。現状の PC ではメモリバンド幅は 1 GB/秒以下、実効では 300 MB~500 MB/秒程度と低いのが欠点とされている。特に一部のシミュレーションはメモリバンド幅に性能が依存するので、そのような状況ではコモディティハードウェアは性能が発揮できなかった。

幸い、この1-2年でコモディティにおいてもメモリバンド幅の急激な向上が予測されている。Intel Pentium 4+850 チップセットの組み合わせは2チャンネル

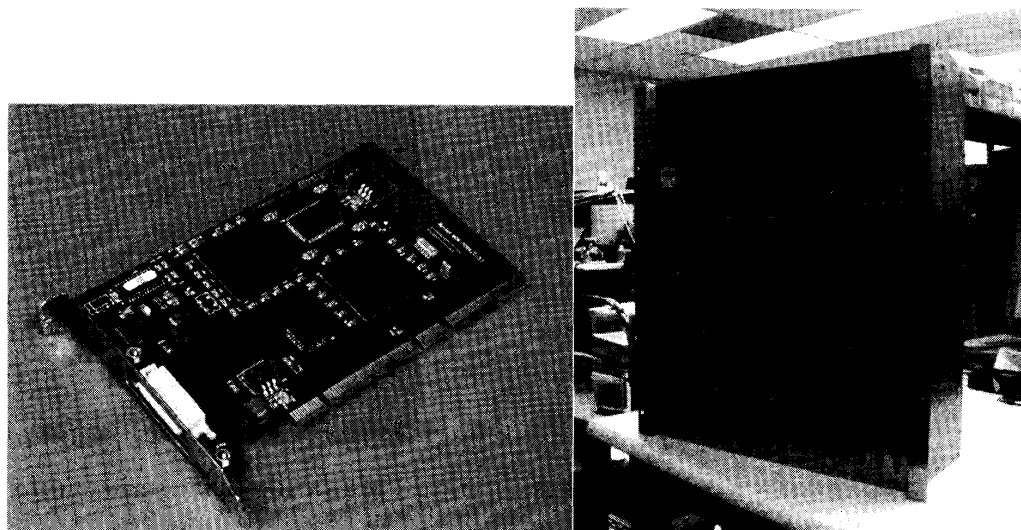


図2 Myricom 社 Myrinet 2000 のネットワークカード, およびネットワークスイッチ

の DRDRAM を用い, 3.2 Gbyte/秒のメインメモリのバンド幅を持つ. 一方, AMD Athlon は DDR-SDRAM を用い, 2.1 Gbyte/秒を確保している. これらにより, 従来はメモリバンド幅の性能限界により WS と比較して低かった SpecFP2000 などの浮動小数点系のベンチマークが WS 並に改善されると期待されている. さらに DRDRAM DDR-SDRAM 両陣営ともコスト削減とバンド幅向上に熾烈な戦いを繰り広げており, さらにマルチチャンネルの RDRAM や Quad-Data Rate の SDRAM も技術的に検討されている. 無論, これらは従来のマルチバンクメモリと比較して, ランダムアクセス時のレーテンシが大きいなどの問題点があるが, プリフェッチを活用する, 超高速なプロセッサ内の大容量 L2 キャッシュを使う, チップセット内に数 MB から数十 MB の L3 キャッシュ相当の高速メモリを内蔵する, などの方策により, その莫大なバンド幅を生かすことが可能となる.

並列計算機を実現するための(3)の相互結合ネットワークもコモディティ化が進んでいる. すでに 100Base-T のネットワークカードは数千円程度で購入可能であり, 百ポート規模の大規模スイッチでもポートあたりのコストは数千円となっているので, 100ノード程度の PC クラスタ型並列計算機はネットワークのコストはノードあたり 1 万円程度である. ただし, 100 MBps の実効バンド幅では, モンテカルロ法などの通信をあまり必要のない計算は大丈夫であるものの, より密な計算ではネットワークがボトルネックとなり, プロセッサ速度の向上があまり意味をなさなく

なる.

幸い, 100Base-T よりコスト高にはなるものの, PC レベルのコモディティハードウェアネットワークにより遥かに強大なバンド幅を実現できる. ハイエンドの Ethernet の規格である Gigabit Ethernet は, 1 GBps のバンド幅を実現し, 近年ではカードも比較的安価となっているので, 小～中規模の PC クラスタの高性能ネットワークとして有望視されている. しかしながら, 1000 Base 規格の問題点は, (1)並列計算機の専用ネットワークと比較して, 通信レーテンシが大きい (10倍以上), (2)ネットワークスイッチ, 特に大規模スイッチが現状では大変高価, (3)実効スループットが必ずしも高くない (60 MBps 程度) と, 数百～数千プロセッサの大規模 PC クラスタを実現するには幾つかの問題がある.

さらに大規模なクラスタでは, Myrinet (www.myri.com), Giganet (www.giganet.com) など, 専用の高速ネットワークが開発されている. このうち Myrinet の最新型の Myrinet 2000 は, 超並列型スパコンに匹敵する 2 GBps のバンド幅, 数 μ 秒の低レーテンシ, および 1 万ノード以上にスケラブルなネットワーク構成が可能なおと, ノードごとのコストがスイッチを含んで 20 万～30 万円程度であることとあいまって, 大規模 PC クラスタ構築用に多く用いられている. 今後は Infiniband などのネットワーク技術も登場が予定されており, さらに急速な性能向上と低コスト化, それによるコモディティの超並列計算が現実のものとなることが期待されている.

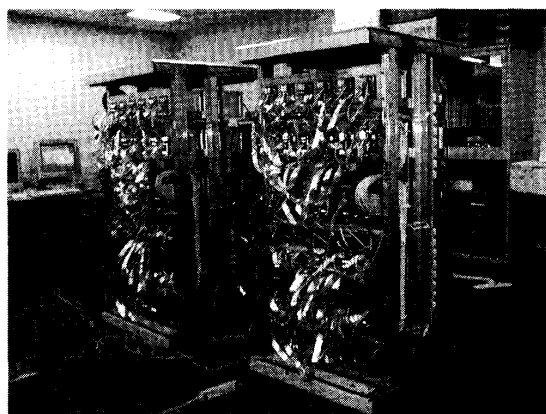


図3 RWCP PC-Cluster-2 計算機 (128プロセッサ)

これらのテクノロジーをベースに、大規模クラスタ計算機を構築するプロジェクトが多く進行している。多くの場合、これらは従来のスパコンに匹敵する計算パワーを数十分の1のコストで実現する。

クラスタ関係の研究は数多く、本稿ではとても紹介しきれないが、有名なものとしては Beowul Project がある¹⁾。まだ、世界最大のクラスタは Sandia 国立研究所の Cplant²⁾で、現状では2000プロセッサを越えている；ただし、CplantはプロセッサとしてPCではなくワークステーション用のAlphaを用いているので、高速であるがそれほど安価ではない。わが国で代表的なクラスタ研究としては、新情報処理開発機構(RWCP)のPCクラスタのプロジェクト³⁾がある。ここでは、単にクラスタ計算機のハードウェアを開発するだけでなく、単一の計算機に近い形で使うための様々なOS・ミドルウェア・プログラム言語などのソフトウェアの研究開発を行っている。RWCは種類のクラスタを作成しており、現状では100プロセッサ規模のクラスタであるが、来年早々に1000プロセッサ規模のクラスタを目指している。これは、密結合されたクラスタ計算機では世界最大の一つになるであろうと予測される。

本研究室においても、後の超広域高性能計算(The Grid)の計算リソースとして、数十から数百ノード規模のクラスタを構築し、種々の研究を行っている⁴⁾。写真では二台のクラスタが写っており、左のProsperoクラスタは現状では66ノード/132プロセッサであるが、本稿の出版と前後して192ノード/256プロセッサに拡張され、その際のピーク性能は200 GigaFLOPSを超える。また、64ノード規模のクラスタ計算機をさらに複数構築して高速ネットワークで接

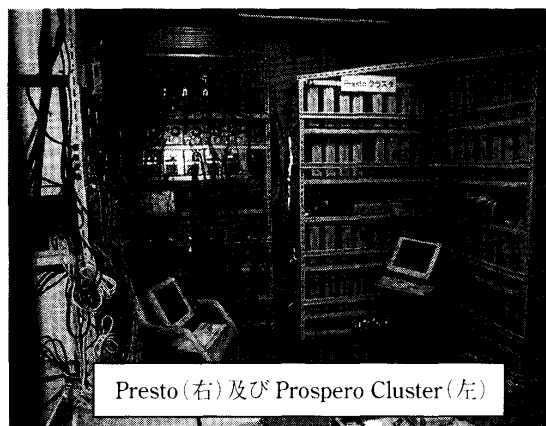


図4 東工大松岡研究室のクラスタ計算機

続し、広域環境で仮想的な一つのクラスタとみなして計算を行う研究も遂行している。詳細は研究室のホームページに譲るが、平成12年度末までに合算では400プロセッサ/ピーク400 GigaFLOPS程度になる予定である。

3. PetaFLOPS へ：超巨大超並列マシンプロジェクトの数々

コモディティテクノロジーで速度向上が図られても、それを上回る速度で計算科学を主体としたシミュレーションの計算パワーに対する要求が増加しているのが現状である。例えば、気象シミュレーションの場合、3次元のメッシュ精度を10倍にするだけで、演算量はざっと1000倍近くにもなってしまう。しかも、地球温暖化などのシミュレーションでは、海流の時定数が数百年単位と長く、大気と地上、および海流の相互作用を加味しながら(Computational Couplingと呼ばれる)の並列シミュレーションが必要とされる。また、単純な流体ではなく、様々な複雑な化学反応が起きることを加味する必要がある場合もある。他の分野も同様で、より精密な自然現象に近づけようという要求に呼応して、必要な計算パワーは指数的に上昇しつつあるのが現状であろう。

単純なムーアの法則の速度向上だけでは、とてもこれらの要求はまかないきれない：1000倍の速度向上には12年間必要になる。必要な計算パワーがGigaFlopsからTeraFlops、さらにPetaFlops(毎秒 10^{15} 演算)に至るにつれ、高性能プロセッサを結合した超巨大な超並列並列マシンが計画されている。これらのマシンの主な用途は科学シミュレーションであるが、従来は非現実的だった種類の問題の計算シミュレーションが

可能になると期待されている。

筆者の個人的な意見としては、このような今後のハイエンドな計算もコモディティなテクノロジーが主流となるという予測であるが、現状ではまだそうではない。しかし、PetaFLOPSを目指すハイエンドの計算技術を鑑みると、コモディティ技術と本質的な差異はなく、1-2年の遅延で性能的にマッチングすることが可能である。ここで紹介する超 TeraFLOPS から PetaFLOPS を目指す幾つかのプロジェクトは、数年後のコモディティコンピューティングであることを忘れてはならない。

3.1 ASCI プロジェクト

ASCI (Accelerated Strategic Computing Initiative)⁵⁾ は、米国、特に DoD (国防総省) および DOE (エネルギー省) が国策的に進める大規模計算のプロジェクトである。目的は核反応および核のストックパイルのシミュレーションがそもそもであるが、その他のエネルギー省の多くの科学シミュレーションに計算パワーを提供する。

ASCI の目標は、従来にはない規模の並列計算機を構築し、性能の指数的な向上を加速させることによって、ベタフロップス級の計算パワーを目指すことにある。その実行としては、三箇所の国立研究所 (Lawrence Livermore, Sandia, Los Alamos) および NSF の San Diego Supercomputing Center に段階的に数千プロセッサ以上の規模の計算機を設置し、プロセッサテクノロジーの進歩とともにアップグレードをしていく、というものである。

現状では、以下の ASCI マシンが設置されている：

- ASCI Red-Sandia 国立研究所, 9280 プロセッサ, ピーク 3.15 TeraFLOPS, Intel 製
- ASCI Blue Pacific SST—Lawrence Livermore 国立研究所, 5856 プロセッサ, ピーク 3.9 TeraFLOPS, IBM 製
- ASCI Blue Mountain—Los Alamos 国立研究所, 約 6000 プロセッサ, ピーク 1.5 TeraFLOPS, SGI 製
- ASCI Blue Horizon—San Diego Supercomputing Center, 1152 プロセッサ, ピーク 1.5 TeraFLOPS, IBM 製

また、以下のマシンが構築あるいは準備中である

- ASCI White—Lawrence Livermore 国立研究所, 8192 プロセッサ, ピーク 12 TeraFLOPS, IBM 製
- ASCI Q—Los Alamos 国立研究所, 12000 プロセッサ, ピーク 30 TeraFLOPS, Compaq 製
- ASCI Final—各研究所, プロセッサ数未定, ピーク 100 TeraFLOPS 超, 各社

ASCI は米国のスーパーコンピューティングのリーダーシップを維持するのも大きな目的であり、巨大な超並列計算機を各計算機メーカーが開発、設置している。Lawrence Livermore 研究所に設置された最新の ASCI White は IBM 製で、8192 個の最新の Power 3 プロセッサが高性能スイッチで結合されている超並列計算機である。そのピーク性能は 12 TeraFLOPS にもおよび、来春には世界最速の計算機となる予定である。ASCI White の設置面積はバスケットボールコート 2 面分と言われている。さらに、次の ASCI Q は Com-

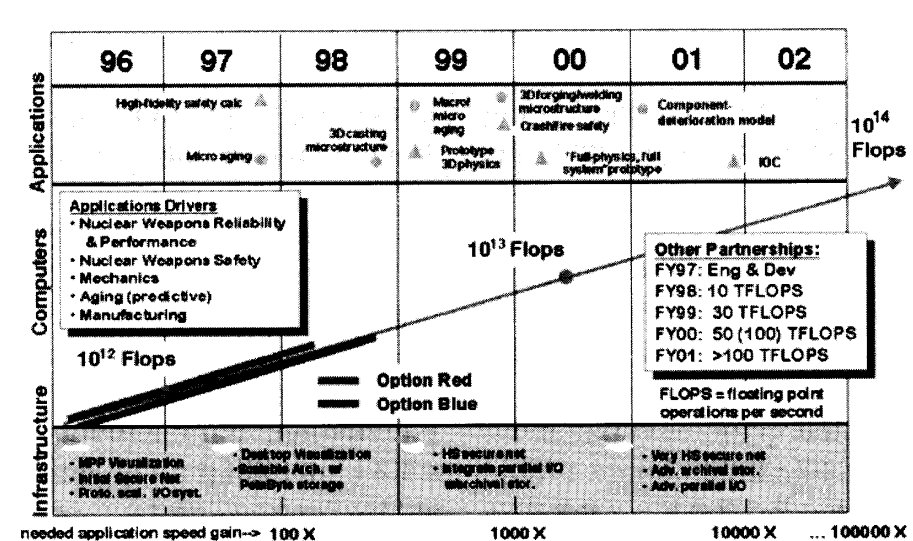


図5 ASCI プログラムのロードマップ⁵⁾

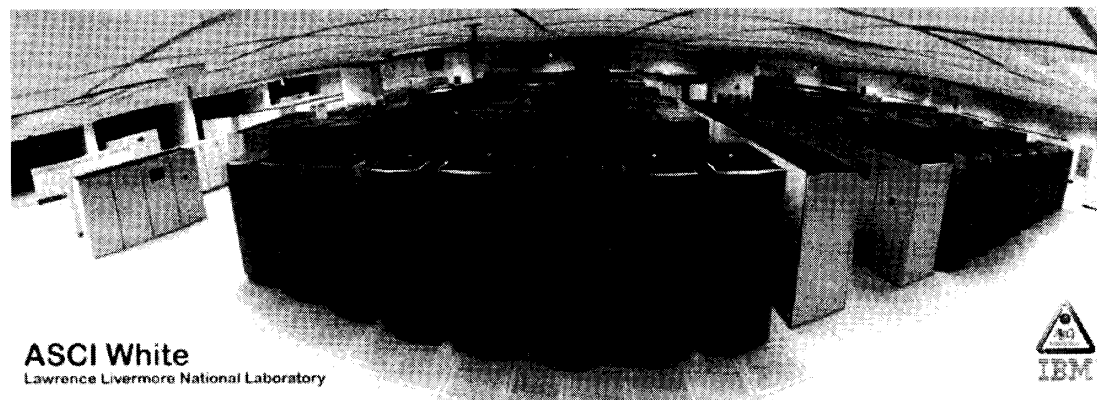


図6 世界最速の計算機 ASCI White (IBM 製) (2000 年12月現在)

paq 製で、12000 ノードプロセッサで30 TeraFlops を2002年に目指す。最終的には、ASCI Final では2005年までにピークで100 TeraFlops の超巨大計算機が誕生する予定である。

筆者は11月に Lawrence Livermore 国立研究所の Carol Hoover 氏を訪問し、現有の ASCI Blue Horizon SST, および構築中の ASCI White に関してたずねた。Hoover 氏は計算科学の部門のマネージャーで、主に並列の有限要素法ライブラリの研究開発も過去十年にわたり行っている。特に彼女らが開発した DYNA3D およびその並列版の ParaDyn ライブラリはクラッシュシミュレーションなどで、世界中で使用されているそうである。彼女によると、現行の5000プロセッサ級の Blue Horizon SST ですら、すでに満杯状態で、かつ1000プロセッサ規模を用いるプログラムが数多く走っているそうである。無論、問題が十分な並列性を持ち、並列化オーバーヘッドを防ぐための十分大きい問題サイズ、並びに高効率な並列化が必要であるが、多くのアプリケーションがその規模のプロセッサを使いこなせる Livermore 研究所のレベルの高さには感心した。

3.2 その他の米国のペタフロップスプロジェクト

ASCI では100 GFLOPS をコモディティテクノロジーで達成し、さらに PetaFLOPS をうかがうが、米国では HTMT という別の計画で1 PetaFlops を目指している。従来型のプロセッサを結合した「保守的な」ASCI マシンと比較して、超伝導 Josephson Junction 素子、プロセッサとメモリを1チップ内に混載する PIM (Processor-in-memory), 超高速光スイッチなど、先進的だがまだ超並列計算機では実証されていない技術の集合であり、2005年以降に一号機の完成を目指している。



図7 Lawrence Livermore 国立研究所の Carol Hoover 氏 (右) と筆者 (左)

米国ではさらに、IBM が Blue Gene という PetaFlops 計算機を計画している。これはゲノム計算、特にたんぱく質の構造解析等に特化した計算機で、PIM 技術により現段階では1チップあたり25個のプロセッサと100 Mbit のメモリを混載させる。1プロセッサのピーク性能は現段階では1 GFLOPS なので、これにより、1チップあたり25 GFLOPS の計算パワーを得る。マシン全体では100万プロセッサで1 PetaFLOPS を得る。専用マシンというのは、汎用の科学技術計算を行うには、あまりにもメモリが小さすぎるが(通常のスパコンの数千分の1)、ゲノムのシミュレーションには十分であるとされている。

3.3 地球シミュレータ

わが国においても、大規模並列計算プロジェクトとして、科学技術庁が地球シミュレータ⁶⁾を遂行している。アプリケーションとしては地球規模の気象シミュレーションがメインであるが、他の用途も考えられて

いる。ASCI やクラスタ計算機と異なり、地球シミュレータは数千台の従来のベクトル型スパコンを並列結合し、そのピーク性能は40 TeraFLOPS に上る。基本的に個々のノードはベクトル型のスパコンなので、多くのアプリケーションでピーク性能に近い性能が期待される。その反面、専用ハードウェアなので、コストは大幅に高く、同時期に近いピーク性能を達成するASCI Q の数倍のコストがかかると予測されている。無論ピーク性能のみではマシンの性能は語れず、むしろ実アプリケーションにおける性能が重要であるが、さらにこの数年間で登場する予定のコモディティハードウェアをベースとしたマシンに対してどの程度優位性を保てるかは不明である。

4. Grid Computing による超広域高性能計算

今までは単一の計算機の高速化を述べてきたが、近年ではGigabit から Terabit への広域ネットワークのスケールアップに伴い、複数の広域の計算機を接合してTeraFLOPS あるいはPetaFLOPS を目指す動きもある。それらの総称を一般的にGrid コンピューティングあるいは単に“The Grid”と呼ぶ。より正確には、Grid コンピューティングとは、インターネットの普及と急激な高速化に伴い、TFLOPS 級の計算機、および数百T〜ペタバイト級のストレージを数Gigabit から数百Gigabit のWAN で結合することによって高性能計算およびそのデータセットを広範囲に提供するための計算インフラおよびパラダイムである⁷⁾。大規模計算だけではなく、超広帯域マルチメディアの連携による遠隔協調計算、多数のセンサのリアルタイム処理、およびDigital Library のテラバイト級のデータベース処理、なども含み、単に高性能計算機をremote login で遠隔で使うという話ではない。

米国ではこの2-3年間で急速にGridの研究開発が進み、The Next Generation Internet (NGI) およびInternet2などの超広帯域次世代インターネットと呼応して、急激にハード、ソフト、DB、および人的なインフラの整備が進行している。代表的なソフトウェアとしてはGlobus⁸⁾、Legion⁹⁾、Condor¹⁰⁾、Netsolve¹¹⁾などがあり、プログラムとしても米国の国立研間のAlliance、DOEのDOE2000、NSFのNPACI、NASAのIPG、DOE/DODのASCI DISCOMなどがあり、莫大な研究資金がぎぎとつぎ込まれつつある。最近ではIETF (インターネットの標準化委員会) にならない、Grid技術の研究および標準化を目指すThe Grid Forum

(<http://www.gridforum.org/>) も設立された。一方わが国のメーカ、大学、および研究所はこの流れからは取り残されていたが、今年から急速に関心が高まりだした。

当研究室では大規模コモディティクラスタを複数構築し、システムソフトウェアおよびテストアプリケーションを構築するとともに、通産省電子総合研究所らと共同研究でGrid コンピューティングのプロジェクトであるNinf¹²⁾を遂行している。また、科学技術庁の「さがけ21」の研究プログラムにおいて、先に述べた当研究室で構築されたクラスタや外部のクラスタを高速ネットワークとGridのソフトウェアを通じ連携しての使用を計画している。

Gridの研究はまだPetaFLOPSを記録した、という段階ではないが、様々な大規模科学技術計算およびシミュレーションの試みが報告されている。例えば、Globus Project⁸⁾においては、SF-Expressという離散事象シミュレーションにおいて、全米に分散したOrigin2000などのスパコンを複数結合し、記録的なサイズの問題のシミュレーションを行っている。また、CactusというGrid上の物理シミュレーションのフレームワークでは、大西洋をまたぐ複数のスパコンを結び、中性子星の衝突シミュレーションを行っている。

当研究室においても、当専攻の小島政和研究室らと共同し、クラスタ & Grid 環境で大規模な最適化問題を解く研究を行なっている。既に京都大学の藤澤氏が開発したSDPAを用い、報告されているものより大規模な非線形問題をNinfシステムを用いてProsperoクラスタで解いている。その際には問題によって128プロセッサで100倍程度の並列化効率を得ており、情報処理振興協会 (IPA) の次世代アプリケーション開発事業「広域分散コンピューティングアプリケーションの開発」にてさらにGrid環境でのテストが行われた。同様の手法で、同じく東工大の原研究室では、BMI最適化問題をNinfを用いて高速化している。さらに、我々の共同研究者のUCSDのBerman教授らのグループでは、MCellと呼ばれるNeurobiologyのシミュレーションを太平洋を越え、UCSDのクラスタ、Tennessee大学のクラスタ、ならびに我々のクラスタを結合して行い、良好な結果を得ている¹⁰⁾。

5. コモディティによる高性能計算の今後の研究課題と今後のシミュレーション

では、今後コモディティによる高性能計算の技術的

課題は何であろうか？ クラスタ計算機を鑑みれば、コストパフォーマンスからすれば従来のスパコンを圧倒するものの、その運用上の使いやすさや安定性はまだまだである。つまり、Single System Image を実現するソフトウェアのロバスト性、コンパイラやランタイムやその他ツール群を含む言語システムの性能・機能・安定性、I/O の Single View 化、MTBF を含む管理の容易性の向上などで、まだまだ長い年月をかけて開発されてきたベクトルスパコンには劣るといえよう。さらに、アルゴリズムやアプリケーションが、コモディティプロセッサの超高速大容量キャッシュ＋ショートベクトル処理＋粗粒度並列に対応せねばならず、かつその記述を容易にするソフトウェア技術が必要である。Grid ではさらに動的に変化するリソースの確保・スケジューリング、耐故障性の向上、セキュリティおよび性能との両立など、さらに技術的課題は多い。

とどのつまり、今後は高性能計算の計算は（アプリケーションを含み）ソフトウェアの研究開発が最重要課題となる。それゆえ上記の問題は本質的にソフトウェア技術に対する大きなチャレンジであり、それを解決することによって、本質的に圧倒的にコストパフォーマンスが高く、かつテクノロジーカーブに載って性能が向上するクラスタ & Grid 技術がスパコンを一気に凌駕するであろう。それと共に、超高性能計算が圧倒的に低コスト化・普及し、市居の技術者でも TeraFLOPS 級のシミュレーションがデスクトップから容易に行えるようになる。これは、計算機によるシミュレーションが第一次革命、WS や PC によってパーソナルな環境でのシミュレーションが普通になったのを第二次革命とすれば、これは第三次の革命ともいえよう。

ただし、これはシミュレーションソフトウェアを開発する側にとっても、新しい計算環境に向けてプログラミング手法を対応させていかななくてはならないことを意味する。例えば、今後は数値アルゴリズムでは、プロセッサのキャッシュマネージメントをうまく行うようなアルゴリズム以外は今後実用的には無価値になる可能性があるし、並列プログラミングはメッセージパッシングなどのコードの書き換えを伴う。しかしながら、2010 年にはコモディティ技術により PetaFLOPS が達成されようとしており、その莫大な計算資源を活用するには安い「出費」といえよう。今

後計算パワーの指数的上により、シミュレーション技術自身にさらなる革命的な研究が芽生えることを期待したい。

（なお、クラスタ計算機対スパコンの技術的詳細については、筆者らによる¹⁴⁾をご覧くださいければ幸いです。）

参考文献

- 1) The Beowulf Project, <http://www.beowulf.org> (2000年11月現在)
- 2) The CPlant Project, <http://www.cs.sandia.gov/cplant/> (2000年11月現在)
- 3) 技術研究組合新情報処理開発機構 (RWCP), 並列分散システムソフトウェアつくば研究室, <http://pdswww.rwcp.or.jp/home.html> (2000年11月現在)
- 4) <http://matsu-www.is.titech.ac.jp> (2000年11月現在)
- 5) Accelerated Strategic Computing Initiative (ASCI), <http://www.llnl.gov/asci/> (2000年11月現在)
- 6) 科学技術庁 地球シミュレータ計画, <http://www.sta.go.jp/sonota/sonota/es00519/>
- 7) I. Foster and C. Kesselman (eds.): The Grid: Blueprint for a New Computational Infrastructure. Morgan Kaufmann, San Francisco (1999)
- 8) I. Foster and C. Kesselman: Globus: A Metacomputing Infrastructure Toolkit, Intl J. Supercomputer Applications, 11(2): 115/128 (1997), <http://www.globus.org/>
- 9) M. J. Lewis and A. Grimshaw: The Core Legion Object Model, Proc. IEEE Symp. on High Performance Distributed Computing (HPDC-5), Los Alamitos, California (1996. 8), <http://legion.virginia.edu>
- 10) M. Litzkow, M. Livny, and M. W. Mutka: Condor-A Hunter of Idle Workstations, Proc. 8th Int. Conf. of Distributed Computing Systems, 104/111 (1988. 6) <http://www.cs.wisc.edu/condor/>
- 11) H. Casanova and J. Dongarra: NetSolve: A Network Server for Solving Computational Science Problems, Int. Journal of Supercomputer Applications and High Performance Computing, 11(3), 212/223 (1997). Also in Proc. IEEE Supercomputing '96, Pittsburgh. (<http://www.cs.utk.edu/netsolve/>).
- 12) A. Takefusa, S. Matsuoka, H. Ogawa, H. Nakada, H. Takagi, M. Sato, S. Sekiguchi, and U. Nagashima: Multi-client LAN/WAN performance analysis of Ninf: a High-Performance Global Computing System, In Proc. IEEE Supercomputing '97 (1997), <http://ninf.etl.go.jp/>
- 13) Henri Casanova et al.: The AppLeS Parameter Sweep Template: User-Level Middleware for the Grid, Proceedings of IEEE Supercomputing 2000, Dallas, Texas (2000. 11), CD-ROM proceedings.
- 14) 朴, 渡辺, 松岡, 関口: インタラクティブエッセイ: これでいいのか? 日本のスパコン, 情報処理, 41-07. (<http://www.ipsj.or.jp> に完全版掲載) .