資　料

ハイエンドコンピューティング技術に関する調査研究 I

－ペタフロップスを中心とする－

平成 12 年 3 月

財団法人 日本情報処理開発協会
先 端 情 報 技 術 研 究 所

KEIRIN

この事業は、競輪の補助金を受けて実施したものです。

ハイエンドコンピューティング技術調査ワーキンググループ

主査	山口　喜教	筑波大学 電子・情報工学系　教授
委員	秋山　　泰	技術研究組合　新情報処理開発機構 並列応用つくば研究室　室長
委員	天野　英晴	慶應義塾大学理工学部 情報工学科　助教授
委員	笠原　博徳	早稲田大学理工学部電気電子情報工学科　教授
委員	久門　耕一	(株)富士通研究所 コンピュータシステム研究部　主管研究員
委員	妹尾　義樹	NEC　C&C メディア研究所 高性能コンピューティング　研究マネージャー
委員	関口　智嗣	通産省工業技術院　電子技術総合研究所 情報アーキテクチャ部　主任研究官
委員	近山　　隆	東京大学　新領域創成科学研究科　教授
委員	中島　　浩	豊橋技術科学大学　情報工学系　教授
委員	花輪　　誠	（株）日立製作所　中央研究所 エンタープライズシステム研究部　主任研究員
委員	福井　義成	（株）東芝 IS センター エンジニアリングシステム部 主査
委員	古市　昌一	三菱電機(株)　情報技術総合研究所 電子システム部　主席研究員
委員	横川　三津夫	日本原子力研究所 地球シミュレータ開発特別チーム　副主任研究員
幹事	岡崎　文夫	（財）日本情報処理開発協会　先端情報技術研究所 技術調査部　主任研究員
幹事	岡田　恵太	（財）日本情報処理開発協会 先端情報技術研究所 技術調査部　主任研究員

ハイエンドコンピューティング技術に関する調査研究　I
－ペタフロップスマシンを中心にして－

目　　次

第１章　まえがき

本報告書は、先端情報技術研究所（AITEC）内に設置された「ハイエンド・コンピューティング技術調査ワーキンググループ」、略して HECC（High End Computing and Communication）WG での議論に基づき、各委員の意見をまとめたものである。このワーキンググループは、昨年度までの「ペタフロップスマシン技術調査ワーキンググループ」をベースにして、より広範囲な視点で先端的なコンピュータ技術の調査を行うことを目的として設置された。したがって、本ワーキンググループ（WG）のメンバーも、若干の委員の交代などはあるが、昨年までの「ペタフロップスマシン技術調査ワーキンググループ」の委員をベースに構成されている。具体的には、大学、メーカ等の若手研究者でアーキテクチャ、ソフトウェア、アプリケーションの各分野において実際に研究や開発に携わっている方々12 名から成っている。

昨年度までの「ペタフロップスマシン技術調査ワーキンググループ」においては、ペタフロップス級のスーパーコンピュータを実現するための技術や日米格差などについて、近未来の情報処理産業を見据えてどのような技術開発を行うべきかについてヒアリングや議論を行い、報告書としてまとめてきた。特に、昨年度は、情報処理技術に関する技術階層を示し、各階層において、技術的な重要課題はどのようなものであるかという点を念頭において議論を進めた。そして、近未来の情報処理におけるアプリケーション分野を考慮にいれて、情報処理技術の重要課題について議論を深めた。

新たにスタートした、HECC ワーキンググループでは、ペタフロップス計算機の開発やそれに関連するソフトウェア、応用技術という観点にとらわれず、ハイエンドアーキテクチャ、ハイエンドハードウェアコンポーネンツ、アルゴリズム研究を含む基礎研究、ソフトウェアおよびハイエンドアプリケーションなどを対象として、調査や議論を行うこととした。まず、議論の手がかりとして、この分野で先頭を走っていると考えられている米国のハイエンド・コンピューティング研究開発計画を参考にして、議論をすすめることとした。そして、そのための具体的な材料として、米国連邦政府のコンピューティング・情報・通信委員会（CIC）がまとめている通称 Blue Book と呼ばれているドキュメント（詳しくは付属資料 2 を参照）を参考資料とした。Blue Book で指摘されている技術項目について、特に各委員から見て技術的に興味があり、市場動向などを視野に入れて見たときに市場にインパクトを与えそうな分野やテーマについて、それらの技術的内容および成果の調査・評価と我が国の技術との格差などについて各委員を中心に議論することとした。本報告書は、これらの議論に基づいて、各委員の HECC に関

する見解をまとめたものである。

本報告書では、まず第２章に、米国の HECC に関する政府支援の状況や戦略に関して AITEC がまとめたものを掲載する。第３章では、これを踏まえて各委員の報告ならびに意見を中心にまとめてある。この章では、Blue Book で示された内容と、各委員の考える現在の日本の技術状況や各委員の研究内容などの比較を中心に記述していただいた。なお、本報告書における具体的な記述に関しては、各委員の信じる技術論・方法論に関する積極的記述を書いていただいている。そのために、本報告書では、各報告ごとにそれを報告した各委員の記名がなされている。また、この章には、本ワーキンググループでのいくつかの議論の種となる外部講師の方々の講演内容について、講演者にお願いして執筆していただいた。ここで、あらためて感謝したい。

本報告書が、わが国のハイエンド・コンピューティングの研究開発に向けた技術開発や政策の一助になれば幸いである。

（山口喜教主査）

第 2 章　米国の政府支援ハイエンドコンピューティング研究開発動向

2.1　概要

2.1.1　背景

米国の景気は 1991 年 3 月以降 2000 年の２月まで「インフレなき経済成長」を 106 ヶ月以上継続するという米国史上初めての記録をうち立てた。

このような未曾有の経済発展を実現したドライビングフォースは、インターネットを利用したいわゆるサイバー空間での経済活動、e-コマースをはじめとするデジタルエコノミーによるものである。インフレなき経済成長をクリントン大統領は 2001 年度大統領教書、予算教書で「ニューエコノミー」と表現し、1995 年以降の経済成長の３分の１が情報技術産業によるものであり、また情報技術によって創出された仕事の賃金は民間部門の平均賃金よりも 80%も高い実績をもたらしたと主張している。

90 年代米国繁栄の礎とも言えるインターネットは、元々1961 年に米国国防総省の防衛高等研究庁（DARPA：Defense Advanced Research Projects Agency）が有事の際にも全てがダウンしてしまうことのないようなネットワークを目指して開発を始めたものである。その後、主にビジネスとは縁のない大学や研究所で使われて来たが、1987 年に UUNET テクノロジー社が電子メール、電子ニュースを利用できる UUCP サービスを開始し、1993 年には Mosaic という最初の WEB ブラウザがイリノイ大学の学生によって無償公開され、1995 年にはネットスケープ社が Netscape Navigator 2.0 を発売するに及んで、インターネットの利用は一般の人々にも急速に広まるようになってきた。また、WWW 上の文書記述言語も HTML の拡張版である XML の出現により、ユーザ独自のタグも定義可能になって使いやすくなってきたことに加え、実際のビジネス活動のいろいろな場面でも十分使えるアプリケーションが多く提供されてきており、いまやインターネットを利用した活動は経済以外でも多くの分野で活発になりつつある。

このインターネットの例に見るように、30 年前に連邦政府の支援で行った基礎研究開発段階では誰も考えていなかった全く新しい用途が現在出現している。WEB ブラウザなどの使いやすいツールやアプリケーションも多く出てきて、それらが国民に広く使われたことによってアマゾン・ドット・コムのようなオンラインショッピングが普及した。このようにデジタルエコノミーが国民生活に急速に普及したことが、1990 年代の米国の繁栄につな

がっているという認識に基づき、1999 年２月に出された大統領直属情報技術諮問委員会（PITAC ： President's Information Technology Advisory Committee）の報告書（Information Technology Research: Investing in Our Future）では「連邦政府は民間企業ではなかなか手の付けることのできないタイプの基礎研究で成果が出るのに 20 年､30 年かかると見込まれるリスクの高い基礎研究への支援をより重視すべきである」という勧告が提出された。これを受けて米国政府は「21 世紀に向けた情報技術」(IT^2：Information Technology for the Twenty-first Century)イニシアチブを設け、2000 年度予算では前年度 HPCC 予算の 28%に当たる３億 6,600 万ドルを追加投資するための予算要求を行った。

また、PITAC の勧告に沿ったかたちで情報技術分野に対する連邦政府支援を継続することを目的として、センセンブレナー下院議員が委員長を務める科学委員会で「ネットワーキングおよび情報技術研究開発法（NTIRD 法: Networking and Information Technology Research and Development Act)」案が提案され、下院を通過した。

2.1.2　本章の構成

以下、本章では、1990 年代、特にクリントン政権下の米国における情報技術分野での柱とも言えるハイエンドコンピューティングの開発の流れをほぼ時系列的に概観する。先ず 2.2 節でどのような政府支援をしてきたかを関連法案でおさらいし、その次に 2.3 節で NTIRD 法案も含め情報技術分野に焦点を絞って 2001 年度予算での基礎研究開発の最新状況、取り上げられているテーマを紹介する。2.4 節では現時点までに判明している情報をもとに ASCI（Accelerated Strategic Computing Initiative）計画の現状、ある程度姿が見えてきているペタフロップスマシンの開発動向および量子コンピューティング、DNA コンピューティングなど最先端の研究を含む国防総省（DOD）の防衛高等研究庁（DARPA）が行っている基礎研究の一つである UltraScale Computing を中心に述べる。

なお、本章執筆に当たっては、連邦政府コンピューティング・情報・通信小委員会（CIC：Subcommittee on Computing, Information, and Communications R&D）がまとめている通称ブルーブックと呼ばれている大統領予算教書への科学技術に関する補足ドキュメント（Blue Book 2000）[1]の他、大統領予算教書、連邦政府発表資料などインターネットで公開されている情報を主なソースとして利用した。

2.1.3　まとめと提言

1990 年代の米国におけるハイエンドコンピューティングシステムへの取り組みの底流にあるものは 1960､1970 年代に行った政府支援の基礎研究があればこそ現在の米国の繁栄があるという認識である。そして今後とも米国が世界のリーダとして発展し続けるには基礎研究、特に情報技術分野の基礎研究への投資が国家として不可欠であるということを大統領自ら遊説などいろいろな機会を利用して国家のビジョンを国民に分かりやすく訴え続けている。

本報告の中では、イニシアチブを「計画」と訳したりしているが、筆者はこれでは言葉不足で不適切であると思っている。そもそもイニシアチブとは、ある明確な意志を持って取り組むということであるから、ASCI 計画は、「コンピュータの計算能力を市場の開発計画よりも何倍も早く引き上げ（Accelerated)、核兵器開発保全という国家戦略に沿って（Strategic)、核実験をシミュレート出来るような計算能力を実現する（Computing）ための国家プロジェクト（Initiative)」とでも表現すべきものだと考える。イニシアチブとはこのように、トップダウン的に国家目標を大統領が明確に示したものなのである。このイニシアチブのもとで政府主導による産官学の緊密な連携が米国では可能となっていると考えている。

また研究開発を実行する仕組み上で特徴的なのは、ペタフロップスマシンの開発に見るように、達成すべき目標に対して同時にかつ競争的に複数の実現技術が研究開発されている点にある。一本に絞り込んだ計画では失敗したときのリスクは途方もなく大きいが、複数の候補が競い合っている場合にはどれか１つがつぶれたとしても、手段は異なるが別の技術で目的は達成できるのである。

もう１つの特徴は、HTMT アーキテクチャのように、ひとつひとつがその分野でその時点の最先端の要素技術であり、それらを統合して１つの大きなシステムを構築するというようなチャレンジングな開発が存在するという点である。要素技術の世界だけに止まらずに、それらが相互作用し連携することによる波及効果は計り知れないインパクトをもたらすことが期待できるからである。この２つの特徴に共通なものは技術の過去から未来へ向けての時間的連続性、周辺技術への拡散・伝播という意味の空間的連続性であるように思われる。

我が国においても、国研、大学、企業研究所で半ば独自に行っている先端分野の基礎研究の発展的な相互連携を可能とするような場を提供するプロジェクトの創出が必要であると思われる。

また、社会生活についても21世紀の生活環境はデジタル化された社会になるのは間違いのないところである。あらゆるコンテンツ、情報、金が電子化され、国家の物理的な境の区別なくサイバースペースを行き来し、日常生活も今までの生活様式とは著しく異なるものになることは想像に難くない。社会生活が豊かに、便利になると同時に、国家を含む社会組織にはセキュリティ、個人にはセキュリティとプライバシーの確保が重要な課題となってくる。分かりやすい例として、その秘匿機能は国家社会の安全保障、認証機能はe-コマースの基盤になっている暗号を考えてみても、コンピュータの計算能力の向上による短時間での暗号解読という大きな危険に脅かされるようになってきている。

報告の中では触れなかったが、米国では、2000年1月7日にクリントン大統領が21世紀でのサイバーセキュリティーを促進するために「情報システム防衛のための国家計画（National Plan for Information Systems Protection）」というイニシアチブを打ち出し、2001年度予算として23億3,000万ドルを要求している。

我が国においても、サイバーテロといった不測の事態にも対処し21世紀においても先進国として国家の繁栄と安全を持続するには、ハイエンドコンピューティングを含めた情報技術研究開発に関する国家目標を明確に設定し、その必要性を国民に分かりやすく説明することや情報公開を徹底して国民の理解を求めることが重要である。これは政府の果たすべき国民への責務であると考える。

2.2　1990 年代の米国政府のハイエンドコンピューティング研究開発支援政策

米国政府が行っているハイエンドコンピューティング研究開発支援の詳細については、昨年度の報告書[2]に詳細に報告されているので、ここではこれからの説明に必要な事項についてのみ要点を簡略に記述する。

(1) 連邦政府組織 (http://www.ccic.gov/orgchart.html)

1976 年に設置された科学技術政策局 (OSTP：Office of Science and Technology Policy) は科学技術に関する政策および予算について大統領に助言する組織であり、国家科学技術委員会 (NSTC ： National Science and Technology Council) を統括している。この国家科学技術委員会は連邦政府として科学技術への投資についての明確な目標を立てるために、閣僚レベルで科学、宇宙、技術政策を連邦政府として横断的にコーディネートすることをそのタスクとして 1993 年にクリントン政権により設立された。

ハイエンドコンピューティングに対応するものとして、国家科学技術委員会配下の技術委員会 (CT ： Committee on Technology) の中に設置されているコンピューティング・情報・通信小委員会 (CIC：Subcommittee on Computing, Information, and Communications R&D) がある。

この CIC 小委員会でマルチエージェンシーの高性能コンピューティングおよび通信 (HPCC：High Performance Computing and Communications) 計画を立て、予算化、実施、レビューを行っている。ここでまとめられた通称ブルーブックが予算教書への科学技術に関する補足をしているドキュメントであり、一般に公開されている。このブルーブックの内容を調査することによりアメリカの高性能コンピューティング開発動向をある程度読みとることが出来る (http://www.ccic.gov/pubs/blue00/)。

(2) 政策の流れ

1990 年代の米国におけるアメリカのハイエンドコンピューティング開発を目的とした政策、主なイニシアチブを表 2.1 に示した。なお年号は特に断らない限り予算年度ではなく暦年である。

ここ 10 年間の米国におけるハイエンドコンピューティング開発の道筋は 1991 年当時上

院議員であったゴア現副大統領が提案した高性能コンピューティング法（High Performance Computing Act of 1991）で決定的になり、HPCC 計画、CIC 計画と発展し、1999 年からは「21 世紀基礎研究ファンド」として予算項目にあげられ、基礎研究への投資も手厚くするようになった。更に大統領直属情報技術諮問委員会（PITAC ： President's Information Technology Advisory Committee）の報告を受けて「21 世紀に向けた情報技術：IT^2」イニシアチブを作り、民間企業では実施が難しい、行われることの少ないハイリスクかつ長期間の基礎研究に重点を置いた追加予算措置をとった。PITAC 勧告値では 2000 年度予算での 1999 年度予算に対する HPCC 予算増額は、４億 7,200 万ドルであったが、予算要求額は３億 6,600 万ドル、認可額は PITAC 勧告値の半額２億 3,600 万ドルとなった（http://www.ccic.gov/it2/）。

過去に連邦政府が行った基礎研究支援の成果であるインターネットを利用した経済活動により建国以来最長期間の経済成長を遂げたという事実を背景にして、21 世紀での国家繁栄のためには科学技術の基礎研究がますます重要になるという認識のもと、下院の科学委員会では、情報技術分野への研究開発予算を単年度ではなく、向こう 5 年間にわたり予算を認めて、より長期の開発を実施しやすくすることを目的とした「ネットワーキングおよび情報技術研究開発法(NTIRD 法 ： Networking and Information Technology Research and Development Act)」案を提案し、2000 年 2 月には下院を通過し上院に送付された。

表 2.1　1990 年代米国の政策

西暦	政策(法案、主なイニシアチブなど)
1991 年	高性能コンピューティング法 (HPC 法 ：High Performance Computing Act of 1991) 成立（ただし 5 年間の時限立法）　同時に HPCC 計画立案
1991 年	ペタフロップスイニシアチブ発足
1995 年	加速的戦略的コンピューティングイニシアチブ（ASCI ： Accelerated Strategic Computing Initiative）　本報告書では ASCI 計画と称す
1996 年	CIC 計画（HPCC 計画の継承計画） 96 年度で HPC 法が失効したが、立法措置はとらず。
1998 年	次世代インターネット研究法（NGIR 法 ：Next Generation Internet Research Act of 1998）成立。1991 年の HPC 法の修正。
1999 年	「21 世紀に向けた情報技術：IT^2」イニシアチブ 2000 年度予算として HPCC 計画予算とは別枠で予算要求。
2000 年	NTIRD 法（Networking and Information Technology Research and Development Act）案 2 月下院通過。 1991 年の HPC 法の修正。2000～2004 年度の支出認可案。

また、クリントン政権発足当時から取り組んでいる包括的核実験禁止条約（CTBT ： Comprehensive Test Ban Treaty）を批准するという目的に密に対応した活動として、エネルギー省（DOE ： Department of Energy）の行っている核兵器保全管理計画（Stockpile Stewardship and Management Program）を高性能コンピューティングでバックアップする加速的戦略的コンピューティングイニシアチブ（ASCI：Accelerated Strategic Computing Initiative）がある。この ASCI 計画は 1995 年から 2004 年までの 10 年間で 10 億ドルを投資するという計画である。

2.3　2001 年度予算に見る研究開発

米国の 2001 年度予算案は 2000 年２月７日にクリントン大統領によって発表された（http://w3.access.gpo.gov/usbudget/fy2001/pdf/budget.pdf)。

各省庁からも議会説明用の予算資料が公表されている。それらとは別に、クリントン大統領は年明けから重点施策説明のための遊説を全国各地で行っていて、発言内容は文書化され国務省から公開されている（http://usinfo.state.gov/)。以上の公開情報をもとに情報技術分野でどのような研究開発を計画しているのかという点に的を絞って見ていく。

以下に､内容に重複があるがクリントン大統領が説明の中で強調している 21 世紀基礎研究ファンド（21st Century Research Fund）と情報技術研究開発（Information Technology Research & Development）を先に見て、その次に大統領予算教書に示された連邦政府予算案（Budget of the United States Government)、IT2 計画の強化継続策の「ネットワーキングおよび情報技術研究開発法（Networking and Information Technology Research and Development Act)」案、DOE の ASCI 計画関連予算の概要を順次紹介する。

連邦政府の研究開発予算内容を細かに紹介した文書は身近にあまりないと思われるので、冗長の感はあるが取り組みテーマがある程度把握できるようにした。

2.3.1　予算の概要

2001 年度の連邦政府予算は総額 1 兆 8,350 億 3,300 万ドル、そのうち研究開発費予算は 853 億 3,300 万ドルであり、4.6%に当たる。研究開発費予算の内訳は次の表 2.2 のようになっている。基礎研究は 13 億ドル増と大きくのびている。

表 2.2　研究開発費予算の推移

（単位:100 万ドル）

種別	1999 年度 実算	2000 年度 見込み	2001 年度 要求	2000 年度比 増分(%)
基礎研究	17,468	19,027	20,328	7
応用研究	15,915	17,193	18,026	5
開発	44,302	44,017	44,321	1
機器類	1,015	1,026	1,137	11
設備類	1,612	1,427	1,521	7
合計	80,342	82,744	85,333	3

（出典:予算教書 99 ページ Table 5-2 より抜粋）

以下の説明での予算金額は、切り口がイニシアチブによるもの、あるいは担当省庁によるものなどが適宜用いられているので注意していただきたい。

2.3.2　21 世紀基礎研究ファンド（21st Century Research Fund）

428 億 9,500 万ドルの予算要求で、非軍事研究としては過去最大の前年度比増額(29 億ドル)要求となっている。主要計画は下記の４項目である。

(1)　NIH でのバイオ医療研究に 10 億ドルの増額。糖尿病、脳機能障害、ガン、遺伝子薬品、疾病予防計画、AIDS ワクチン開発に関する研究支援。

(2)　新しく起こした国家ナノテクノロジー計画に４億 9,500 万ドル。トランジスタとインターネットが情報時代を切り開いたように、原子分子を操作するナノテクノロジーは 21 世紀を革命的に変化させる。議会図書館の蔵書を角砂糖一つの大きさに格納できる分子コンピュータといったブレークスルーが期待できる。

(3)　大学ベースの基礎研究を活性化するために NSF へ６億 7,500 万ドル増額。科学と工学のバランスの取れた支援を行う。

(4)　情報技術の基礎研究に５億 9,400 万ドル増額し、トータルで 23 億ドルの支援を行う。高速無線ネットワーク、ハリケーンや竜巻をより高精度に予測するためのスーパーコンピュータ、救命医薬品の開発への支援を行う。次項に内容の詳細を示す。

2.3.3　情報技術研究開発(IT R&D)のハイライト
(http://www.ccic.gov/highlights/itrd-4pager.pdf)

これは前項の（4）に対応する計画であり、情報技術分野の研究開発費は 2000 年度比５億 9,400 万ドル増の 23 億 1,500 万ドルの予算要求となっている（表 2.3)。増額の内訳を見ると NSF が２億 2,300 万ドル、DOE が１億 5,000 万ドル、DOD が１億 1,500 万ドル、NASA が 5,600 万ドル、DHHS が 4,200 万ドルとなっている。

表 2.3　情報技術研究開発予算要求額

（単位:100 万ドル）

省庁	2000 年度予算	2001 年度予算	増分(%)
DOC(NOAA、NIST)	36	44	22
DOD(DARPA、NSA、URI)	282	397	41
DOE	517	667	29
EPA	4	4	0
DHHS(NIH、AHRQ)	191	233	22
NASA	174	230	32
NSF	517	740	43
合計	1,721	2,315	35

2001 年度の重点分野として以下の 11 のテーマがあがっている。

(1)　最先端コンピューティング開発チーム ： 国家として最もさし迫った情報分野での課題を解決するためのツール作成を情報科学者、数学者および医学研究、気象モデリング、天文学の専門家が一緒に進められるような新しい協力関係をサポートする。これにより情報科学が進歩し、応用分野でのブレークスルーが期待できる。

(2)　最先端コンピュータモデリングとシミュレーション用インフラ ： 民間研究者用に NSF として 2 番目のテラスケールコンピューティング設備を用意する。

(3)　より信頼性の高いソフトウエア ： 公共インフラ障害を発生させるなどの社会経済に影響を及ぼすソフトウエアバグを無くすように、工業製品の設計テストツールのように生産的かつ予測可能なソフトウエア設計、テスト方法を開発する。

(4)　データの格納、管理および保存 ：NASA の地球観測衛星は 1 年間に議会図書館が持っている情報の 3 倍のデータを発生させている。このような多量のデータを PC のハードディスク程度の大きさの装置に格納する技術開発。

(5)　インテリジェントマシンとロボットネットワーク ：NASA では未知の環境でも知的に順応的に自己完結的に動作し、集団でも動作できるような宇宙探査機を必要としている。

(6)　ユビキタスコンピューティングと無線ネットワーク ： 周囲に埋め込まれたコンピュータシステムは人の声、身振り、接触を通して人と通信でき、人の意図を読みとり必要な情報を自動的に供給出来るようになる。更に自動車や飛行機の運航をより安全に管理することが出来、安全でより信頼性のある医療施設を構築できる。

(7)　情報のセキュリティとプライバシーの管理と保証 ： コンピューティング技術と通信技術が一体になると、データは意図せぬ使用にさらされることが多くなってくる。ｅ-コマース、ソフトウエアの配布、ネットワーク保護やデジタル透かしでのデータを保護するために、DOD、NIST は継続して暗号の研究を行う。

(8)　未来世代コンピュータ ： 量子コンピューティングや分子あるいはナノテクノロジーによる電子回路を使って、今まで計算不能であった気候モデリング、生態系のフルスケールシミュレーション、宇宙動力学や外科的シミュレーションを可能にする。バイオインフォマティック研究は膨大な量のゲノム情報の管理を可能にする。

(9)　広帯域光ネットワーク ：DARPA が支援している研究で、現状のスピードより 1,000 倍速い光ネットワークが可能となった。エンドユーザがこのメリットを得るためには更に光スイッチの改良が必要である。この技術はヘルスケア、航空管制など民間部門のインフラに応用できる。

(10)　社会、経済および労働力への情報技術の関わり合い ： NSF は、日常生活における情報技術の急速な浸透により生じている見込みと社会、経済および労働力のチャレンジを同定、理解し予測し立ち向かうための研究投資を行う。

(11)　新世代研究者の教育と訓練 ： 情報技術研究と教育の容量を大きくするには新しい研究者が必須である。NSF、DOE、NIH が関連する専門分野をまとめることにより多分野に通用する能力を持った新世代の研究者を訓練することが出来る。

2.3.4　連邦政府予算案 (Budget of the United States Government)
(http://w3.access.gpo.gov/usbudget/fy2001/pdf/budget.pdf)

研究開発予算は、予算教書第４章「21 世紀のアメリカを強化するには」の第５節「研究を促進する」に 12 ページにわたって取り上げられている。その前文には米国経済の発展に対する科学技術の貢献が書かれており、1999 年に、「アメリカのための 21 世紀基礎研究ファンド（the 21st Century Research Fund for America)」をおこして、NIH（National Institutes of Health:国立衛生研究所)、NSF（National Science Foundation:科学基金)、DOE（Department of Energy:エネルギー省）での基礎研究を中心にコンピュータ、通信、エネルギー、環境その他の分野への調和しかつバランスの取れた資源の投資戦略を採ってきていると主張している。

ここで言っている「調和しかつバランスの取れた」という意味は、ヘルスケアを例に取ると、その発展には医学医療分野だけではなく、その周辺技術である医療イメージング技

術やコンピュータを活用して初めて実現できる新薬の早期開発およびヒトゲノムのマッピング技術などのブレークスルーにも大いに依存しているのだという認識を指している。要するに基礎および応用の両面にまたがり、相互に関連する領域の研究開発を上手に組み合わせて、得られる成果を増幅しようという姿勢である。

2.3.4.1 科学技術イニシアチブ（The Science and Technology Initiative）
－戦略的成長への果敢な道のり－

ベースになっているのは前述の「21 世紀基礎研究ファンド」であり、2001 年度では 2000 年度比 29 億ドル増（７％増）の 429 億ドルを予算要求している。特に（1）基礎的かつ長期の研究への投資を強化する、(2）ヘルスケア研究と他の領域とのバランスを取る、(3）大学ベースの研究を強調する、(4）国家としてプライオリティーの高い戦略的な研究への支援増加をその狙いとしている。2001 年度に予算が大幅に増加したナノテクノロジー開発支援はその好例である。

そのほかバイオ関連や既に実施している情報技術分野の研究開発に対しても基礎研究の強化、先端的スーパーコンピュータ応用プログラムへの支援増加も行っている。対象となっているのは、NSF、DOE、NASA（National Aeronautics and Space Administration:航空宇宙局)、DOD（Department of Defense:国防総省)、DOC（Department of Commerce:商務省）などの省庁である。

全研究開発予算（Research and Development Investments）は 2000 年度比３％増の 853 億 3,300 万ドルとなっており、そのうち非軍事関連予算は 51％である。

科学技術イニシアチブで米国の重点テーマとして取り上げられている個別イニシアチブを抜き出すと次のようになる（予算教書 100 ページ Table 5-3)。

・国家ナノテクノロジーイニシアチブ　４億 9,500 万ドル
・情報技術イニシアチブ（IT^2）　８億 2,300 万ドル
・気候変動技術イニシアチブ　14 億 3,200 万ドル
・省庁連携教育研究イニシアチブ　5,000 万ドル
　（以上 NSTC 取りまとめ）
・Space Launch イニシアチブ（NASA）　２億 9,000 万ドル
・国家基礎研究イニシアチブ（USDA）　１億 500 万ドル
・インテリジェント交通システムイニシアチブ（DOT）　３億 3,800 万ドル

以下に、情報技術分野に関連するものを中心に予算主要項目の内容を紹介する。

(1) 基礎研究の強化および連邦政府研究ポートフォリオのバランス

2000 年度比 13 億ドル増(7%増)の 203 億ドルを基礎研究への支援にあてる。国家的にプライオリティーの高い基礎研究を行っている NIH、NSF、DOE、NASA が対象になっている。

(2) 大学ベースの基礎研究強化

大学での研究を支えている助成金は競争的なものであるため、科学の新しい概念を創造したり研究領域を広げることに役立っていると同時に次世代の科学者、技術者の育成の場を提供しているという特別な役割を持っている。この助成金は 2000 年度比 13 億ドル増（８％増）の 178 億ドルの予算要求となっていて、主として NSF、NIH、DOE を通して支給されている。連邦政府が支援している大学での研究に占める NSF の支援比率は健康関連以外の基礎研究分野で 50%以上になり、有能な学生を科学技術の道に進ませることへの動機付けにもなっている。

(3) NSTC が推進する主要なマルチエージェンシー研究イニシアチブ

情報通信分野への支援増加の他、2001 年度から新たに「国家ナノテクノロジーイニシアチブ(National Nanotechnology Initiative)」と「生物ベース製品およびバイオエネルギー(Biobased Products and Bioenergy)」という二つの計画をスタートさせた。

(a) ナノテクノロジー研究

原子または分子レベルで物質を操作することにより、今までにない新しい階層のデバイスを作り出すことを可能にするような新規の性質、プロセス、現象を研究し、その研究成果を情報技術をはじめ国防、宇宙開発などのあらゆる分野に必要となる電子/電気光技術のたゆまない改善に結びつけるといった計画となっている。医学の分野では、人間の細胞程度の大きさの検知器によってガン細胞を検出する手法も考えられている。４億 9,500 万ドルの予算を提案している。

(b) クリーンエネルギー（生物ベース製品およびバイオエネルギー）

化石エネルギー資源に変わる生物ベースの製品やバイオエネルギーの使用を現状の３倍にするという大統領命令 13134 号に基づき予算として２億 8,900 万ドル提案されている。穀物の生産性向上や収穫技術の改善に対しても支援を行う。

(c) 情報技術分野の研究開発

2001 年度予算での範囲は、過去 10 年間にわたって実施されてきた HPCC 計画(次世代インターネット（NGI）を含む）と 2000 年度予算で出来た IT^2 計画を合併して Information Technology R&D（IT R&D）と新しい計画名称になっている。この

計画では IT^2 計画に８億 2,300 万ドル、NGI に 8,900 万ドルを含む総額で 23 億ドルの予算を提案している。

今後更に強力になるハイエンドコンピューティングシステム、先端性能を持つ地球規模のネットワーク技術、ソフトウエア開発技術やアプリケーションソフトの進歩、広範囲に分散されている知識ディポジトリーへのアクセス方法とその管理およびヒューマンインタフェースの進歩に貢献できる、コンピューティング、情報および通信分野への長期の支援を行う。

(d) 大量破壊兵器配備および基幹的インフラストラクチャ防衛研究開発（Weapons of Mass Destruction（WMD）Preparedness and Critical Infrastructure Protection R&D)

後者の基幹的インフラストラクチャ防衛研究開発は、国家、国民生活の基幹をなす電力、通信、情報、交通などのシステムをいわゆるサイバーテロの攻撃から守るために官産学協同で取り組むものであり、６億 600 万ドルの予算提案をしている。

そのほか、情報技術分野以外では以下の計画がある。

(e) 気候変動技術イニシアチブ（CCTI: Climate Change Technology Initiative）

(f) 新世代乗用車パートナーシップ（PNGV: Partnership for a New Generation of Vehicles)

(g) エコシステム課題のための総合科学（ISEC: Integrated Science for Ecosystem Challenges)

(h) 健康に関する基礎研究（Fundamental Health Research)

(i) 教育技術と省庁連携教育研究イニシアチブ（Education Technology and the Interagency Education Research Initiative)
デジタルデバイドを解消するための施策の一つ。

(j) 地球規模変化の研究計画（USGCRP: U.S. Global Change Research Program)

2.3.4.2 連邦政府機関での研究開発投資

(1) NIH（National Institutes of Health） 国立衛生研究所

・バイオ医療研究(糖尿病、脳障害、ガン、遺伝子薬品、疾病予防、生物医療情報技術、ナノテクノロジー、AIDS ワクチン)に 2000 年度予算比 10 億ドル増額

・研究者主体の、ピアレビュー型研究助成方式の堅持

(2) NSF (National Science Foundation)　科学基金

・広範囲にわたる科学技術研究に対し 457 億ドル(前年度比 17%増)

・IT R&D 分野、特に長期間のコンピュータ科学研究と先進スーパーコンピュータを科学者が利用出来るようにするために７億４千万ドル

(3) DOE (Department of Energy)　エネルギー省

・物理、化学、生物、材料、環境およびコンピュータ科学研究計画に 31 億５千万ドル（前年度比 13%増）

(4) NASA (National Aeronautics and Space Administration)　航空宇宙局

・現在推進中の計画および新規計画に対して 140 億ドル（前年度比３%増）

(5) DOD (Department of Defense)　国防総省

・基礎研究に 12 億ドル、応用研究に 31 億ドル、先端技術開発に 32 億ドル

・研究は物理科学に重点を置き、国家の工学、数学およびコンピュータ科学分野での活動に不可欠である

(6) USDA (Department of Agriculture)　農務省

(7) DOC (Department of Commerce)　商務省

・NIST の ATP に対して 1 億 7,600 万ドル(前年度比 23%増)

・2001 年から始める新しい研究はナノテクノロジーと情報科学

(8) USGS (Department of the Interior's U.S. Geological Survey) 内務省地質調査部

(9) EPA (Environmental Protection Agency)　環境保護庁

(10) Department of Transportation　運輸省

・ITS (Intelligent Transportation System) イニシアチブに 3 億 3,800 万ドル

(11) Department of Veterans Affairs' Medical Research　復員軍人省医療研究

(12) Department of Education　教育省

2.3.5　ネットワーキングおよび情報技術研究開発法案

2001 年度予算教書とは別に、IT^2 計画の強化継続策として、2000 年度から 2004 年度の５年間にわたり計画的に情報技術分野への政府支援を行うことを目的とした「ネットワーキングおよび情報技術研究開発法 (NITRD 法 : Networking and Information Technology Research and Development Act)」案が第 106 議会下院本会議に上程され、2000 年２月 15 日下院を通過し上院に送付された。

この法案は、下院科学委員会が提案したもので1990年代米国での高性能コンピューティング開発の端緒となった 1991 年の「高性能コンピューティング法（HPC 法 : High-Performance Computing Act)」(1998年の「次世代インターネット研究法（NGIR法 : Next Generation Internet Research Act)」により既に修正されている）を今回更に修正してNSF、NASA、DOE、NIST、NOAA、EPA、NIHの研究開発支出を2000年度から2004年度までの５年間についてあらかじめ認可しようというものである。この背景には、過去の情報科学、コンピューティング科学分野での基礎研究の成果が現在の国家繁栄をもたらしたという事実、そしてそれが将来の米国の経済発展に最も重要な要因であり国家として最優先で支援すべきものであるという認識と、IT^2 計画もそうであったが、今までの予算措置が単年度であったために、研究のための資源が十分に投入し難いという状況改善の狙いがあるように思われる。

2000 年２月 15 日下院で承認された「ネットワーキングおよび情報技術研究開発法（Networking and Information Technology Research and Development Act)」案での予算認可案を表2.4に示す。

表 2.4 「ネットワーキングおよび情報技術研究開発法」案での予算認可案

（単位:百万ドル）

	2000年度認可案	2001年度認可案	2002年度認可案	2003年度認可案	2004年度認可案
NSF	580.00	699.30	728.15	801.55	838.50
研究設備(注1)	70.00	70.00	80.00	80.00	85.00
NASA	164.40	201.00	208.00	224.00	231.00
DOE	60.00	54.30	56.15	65.55	67.50
NIST	9.00	9.50	10.50	16.00	17.00
NOAA	13.50	13.90	14.30	14.80	15.20
EPA	4.20	4.30	4.50	4.60	4.70
NIH	223.00	233.00	242.00	250.00	250.00
HPC(注2)	1,124.10	1,285.30	1,343.60	1,456.50	1,508.90
DOE	25.00	15.00	15.00	----	----
NSF	25.00	25.00	25.00	----	----
NIH	7.50	0.00	0.00	----	----
NASA	7.50	10.00	10.00	----	----
NIST	7.50	5.50	5.50	----	----
NGIR(注3)	72.50	55.50	55.50	----	----
NITRD計	1,196.60	1,340.80	1,399.10	1,456.50	1,508.90

(注1) テラスケールコンピューティング実現のための研究設備開発助成金
(注2) 'High-Performance Computing Act of 1991'への修正値
2000年度～2004年度追加
(注3) 'Next Generation Internet Research Act of 1998'への修正値
2001年度、2002年度追加

NIH については、HPC 法の 205A セクションとして今回新たに追加されたものであるが、バイオ医療および行動科学研究での計算技術とソフトウエアツールの進歩と応用拡大がその目的となっている。

2.3.6　ASCI 計画

クリントン政権発足時からの国家目標であった包括的核実験禁止条約（CTBT：Comprehensive Test Ban Treaty）を批准するという目的に密に対応した活動である ASCI 計画については、連邦政府予算案には説明がないため DOE の予算要求案からその動きを見ることにする（http://www.cfo.doe.gov/budget/01budget/highlite/hilite01.pdf）。

（1）2001 年度予算における ASCI 計画の位置づけ

ASCI 計画の存在理由となっている核兵器保全管理計画（Stockpile Stewardship and Management Program）への見直しが行われ、ASCI 計画は 2000 年度予算での国家防衛権限法（National Defense Authorization Act for FY2000）に基づいて 2000 年３月１日付で DOE 内に設置された半自立的なエージェンシーである国家核安全保障管理部（NNSA: National Nuclear Security Administration）管理下に置かれることになった（http://www.nnsa.doe.gov/）。

この結果、ASCI 計画は、NNSA の防衛計画課がとりまとめ部署となり、兵器に関する活動（Weapon Activities）の４つの主要コンポーネントの１つである保全管理（Stewardship Operations and Maintenance）の更に３つのサブコンポーネントを構成している２つのサブコンポーネントの作戦行動（Campaigns）と技術基地と設備の配備（RTBF：Readiness in Technical Base and Facilities）で管理されることになった。DOE の予算要求説明書を見る限りでは新体制に移行しても既に定められている ASCI 計画の目標値は変更を受けずに 2004 年まで推進されることになっているようである。

（2）2001 年度予算

今までの ASCI 計画に含まれていた具体的な計画は 2001 年度では作戦行動サブコンポーネントの防衛アプリケーションとモデリング（Defense Applications and Modeling）および RTBF サブコンポーネントの先端的シミュレーションとコンピューティング（Advanced Simulation and Computing）で予算化されており、予算要求額は５億 9,520 万ドル（2000

年度比 8500 万ドル増）となっている。

主な実施内容としては 1999 年 9 月 29 日に RFP（Request for Proposal）を締め切った Los Alamos 研究所の ASCI T30 ハードウエア（30 テラフロップスプラットフォーム）調達、Lawrence Livermore 研究所の 10 テラフロップスプラットフォーム（ASCI White）改良の完成が計算結果を視覚化し理解を助ける VIEWS、改良されたソフトウエアを使ったプログラムコーディング開発が盛り込まれている。

2.4　ハイエンドコンピューティング開発状況

米国におけるハイエンドコンピューティング開発の主なトピックスは、既に報告書[2][3][4]にまとめられているので詳細はそちらを参照いただくとして､本報告書ではテラフロップス以上の処理能力を持つハイエンドマシンの開発状況全般について現時点での最新状況を報告することを狙いとする。また、Blue Book 2000[1]にあがっている主要開発テーマについては本報告書の付属資料２としてまとめて報告しているので、本節では主要なハイエンドコンピューティング開発に限定する。

2.4.1　ASCI 計画（http://www.llnl.gov/asci/）－テラフロップスマシンの開発－

本項では、1999 年 11 月に公表された TOP500 Supercomputer（http://www.top500.org/）で１位から３位を独占し､現在実用レベルで世界の最先端を行っている ASCI マシンについて開発状況をまとめた。

DOE の ASCI 計画を簡単に表現すると、核兵器の開発、保管を行う上で、従来は実際に核爆発を伴う実験を通して確認をしてきた核爆発の威力とその影響、材料の強度、物性などを高性能コンピューティングシステムでシミュレーションしてしまおうというものである。そして、このようなシミュレーションで要求される大規模なコンピューティング能力を民間企業､大学が持っているの最先端の技術力をベースにして実現しようというものである。

ASCI 計画の特徴は､DOE 管轄の Los Alamos､Sandia､Lawrence Livermore の３つの Defense Programs laboratories と University of Illinois、 University of Chicago、California Institute of Technology、Stanford University、University of Utah の諸大学との連携（Academic Strategic Alliance Program（ASAP））があげられる。もう一つの特徴としては、プラットフォームは IBM、SGI、Intel 等民間企業がその時点で持っている最先端技術を適用した市販品（COTS ： commercial（commodity）-off-the-shelf）を使用し、かつその性能を最大限引き出すことが出来るようにソフトウエアも同時に開発している点である。

ソフトウエア関連の活動分野は問題解決環境(Problem Solving Environment（PSE）)、数値計算環境（Numerical Environment for Weapons Simulation（NEWS））、遠距離・分散コンピューティング（Distance Computing and Distributed Computing（DISCOM））、可視化（Visualization)、応用（Application）である。計画内容の詳細は ASCI の WEB サイトを参照願いたい（http://www.llnl.gov/asci/index.html)。

DOE は､ASCI 計画で達成したコンピューティング環境を ASAP 加盟の５大学の研究者にも

一部開放し、ハイエンドコンピューティング分野のハードウエア、アプリケーションソフトウエア開発の便に供している。上記の各大学で開発されたアプリケーションは軍事機密には指定されていないため、非軍事の科学研究へも適用されている。

図 2.1 の開発ロードマップにあるように、2004 年に 100 テラフロップスの計算能力達成が ASCI 計画の最終ゴールである。この最終ゴールに向けて、ASCI 計画の中に PathForward プログラムが新たに設けられ、1998 年から 2001 年までの 4 年間で 30 テラフロップスマシンを実現するための大規模スケーラブルコンピュータシステム構築活動を始めている。このプロジェクトには DEC が 256 ノードの AlphaServer SMP のインターコネクト開発、IBM が 100 テラフロップスまでサポート可能な高速かつ低レイテンシのスケーラブルなスイッチング技術の開発、SGI/Cray が信号処理とインターコネクト技術の開発、SUN がハードウエア、ソフトウエア両面にわたる実行可能性の総合評価を行うといった分担で参加している。

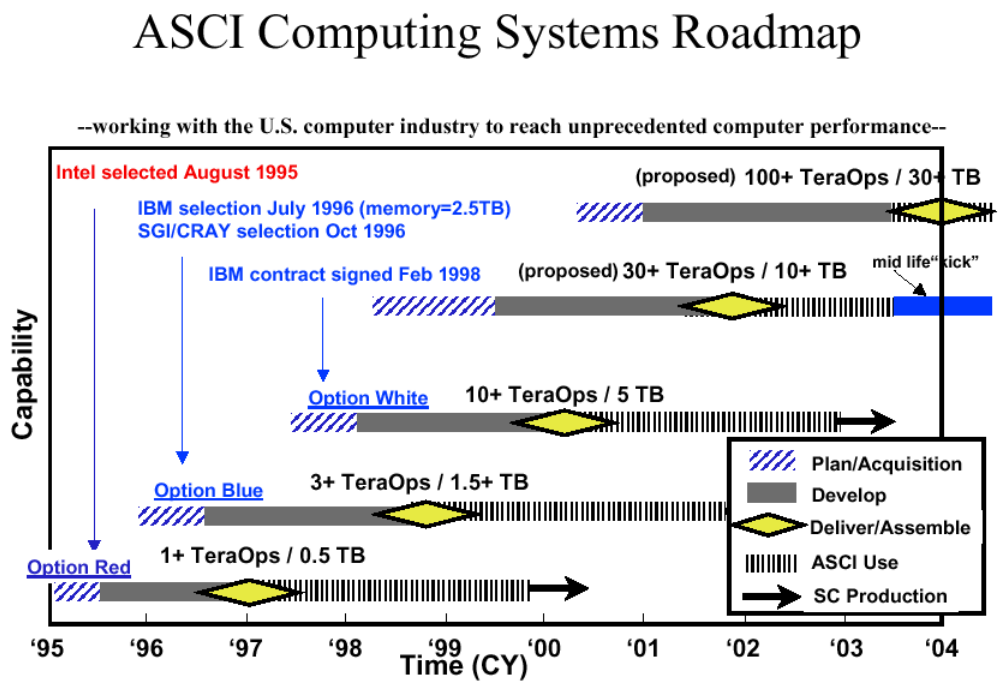


図 2.1　ASCI 計画開発ロードマップ

（出典:P.H.Smith, PetaFLOPS Initiative (1999)）

2000 年３月時点で開発中のものおよび公募を締め切ったものを含めた各プラットフォームの概要を表 2.5 にまとめた。

2001 年度予算では、2.3.6 (2) に書いたように、1999 年 9 月 29 日に RFP (Request for

Proposal）を締め切った Los Alamos 研究所の ASCI T30 ハードウエア（30 テラフロップスプラットフォーム）調達、Lawrence Livermore 研究所の 10 テラフロップスプラットフォーム（ASCI White）改良の完成が計算結果を視覚化し理解を助ける VIEWS、改良されたソフトウエアを使ったプログラムコーディング開発が計画されている。

1999 年 11 月に公表された TOP500 Supercomputer では、ASCI Red、ASCI Blue Pacific、ASCI Blue Mountain が１位から３位を独占した（http://www.top500.org/）。

表 2.5　ASCI プラットフォームの概要

名称	Red	Blue Pacific	Blue Mountain	White	T30
設置研究所	Sandia	Lawrence Livermore	Los Alamos	Lawrence Livermore	Los Alamos
メーカ	Intel	IBM	SGI	IBM	未定
使用 MPU	9,536 × Pentium Ⅱ Xeon	5,856×Power PC	Origin2000 MIPS R10000	8,192×Power3-Ⅱ	未定
目標性能	1.8Tflops メモリ 606GB Disk 容量 40TB	3.1Tflops メモリ 2.6TB Disk 容量 75TB	3.1Tflops メモリ 2.5TB Disk 容量 75TB	10.2Tflops メモリ 2.5TB Disk 容量 75TB	30+Tflops
実績	3.2Tflops (‘99/10 月)	3.9Tflops (‘98/10 月)	3.1 Tflops (‘98 年)	----	----

（注）性能はピーク性能値である。

2.4.2　ペタフロップスマシンの開発 [5]

上に見たように、ASCI 計画は 2004 年の 100 テラフロップス達成に向けて計画通り順調に進捗しており、1999 年度予算のブルーブックでは、2007 年迄には１ペタフロップスでの持続的処理が達成できるかも知れないと書かれている。

テラフロップスマシン、ペタフロップスマシン開発の位置づけを改めて図 2.2 に示す。ペタフロップスシステムの開発に関しては 2.2 節の表 2.1 に示したように既に 1991 年にイニシアチブが作られ、ASCI 計画などと連動しながら、更に高度の技術レベルを目指すものとして進められて来たようである。ペタフロップスイニシアチブではペタフロップスシステムを実現するものとして、COTS クラスタ、超並列（MPP:Massively Parallel Processor）システムアーキテクチャ、ハイブリッドテクノロジーカスタムアーキテクチャ（後に Hybrid Technology Multithreaded（HTMT）Architecture）の３方式を実現技術候補として検討が進められた。以下簡単にこれらの実現技術の内容を見ていく。表 2.6 に各方式の概要を示す。

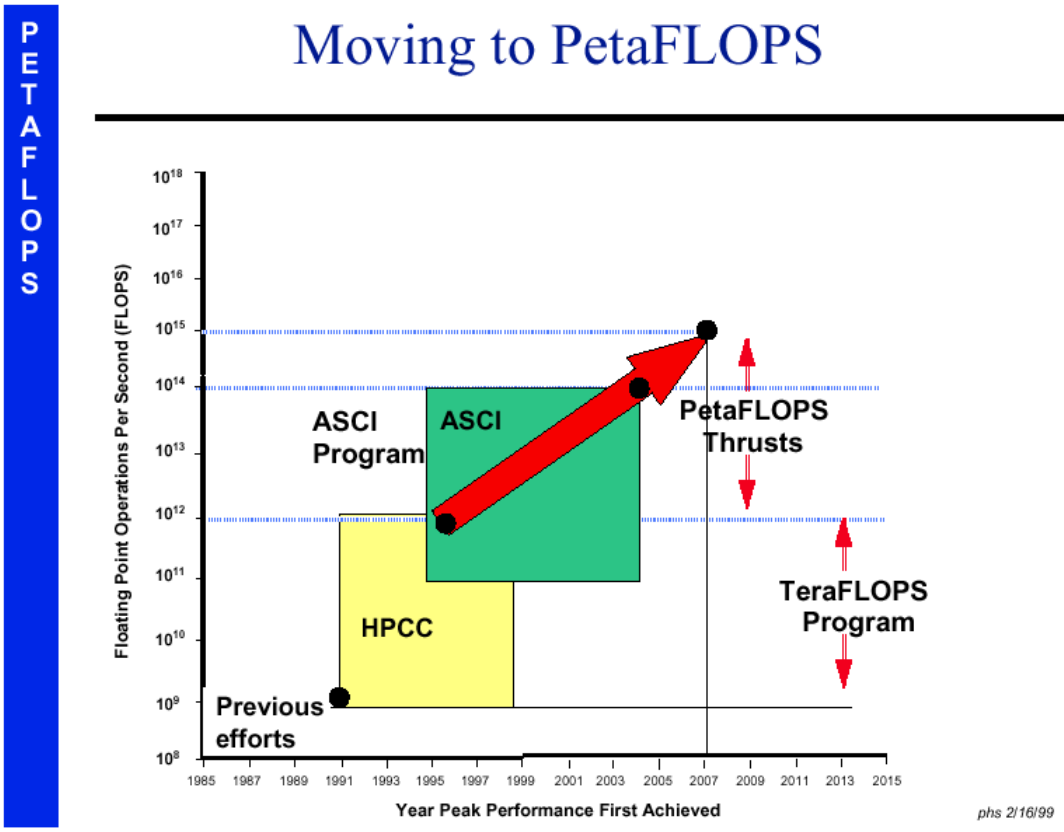


図 2.2　高性能コンピューティング開発の位置づけ

（出典:P.H.Smith, PetaFLOPS Initiative (1999)）

表 2.6　ペタプロップス実現技術の比較

主要項目	COTS クラスタ	超並列アーキテクチャ	HTMT アーキテクチャ
プロセッサ	3GHz　10Gflops	3GHz　10Gflops	150GHz　600Gflops
プロセッサ数	100,000	100,000	2,048
メモリ	32TB DRAM 40ns	32TB DRAM 40ns	16TB PIM-DRAM 80ns
相互接続	ハイパキューブ 20Gbps/ch	フレームスイッチ 128Gbps/ch	Data Vortex 500 Gbps/ch >10Pbps bijection bw
2 次ストレージ	1PB　1ms	1PB　1ms	1PB　1μs
マルチプロセッサ方式	分散型 3 レベルキャッシュ 1 レベル DRAM	分散共有型	共有型 4 階層
レイテンシ管理	ソフトウエア	Cache coherence protocol	Multithreaded /percolation

（出典:Thomas Sterling, Challenges to Petaflops (1999) を一部手直し）

(1) COTS クラスタ

代表的なものとしては ASCI Red の他に航空宇宙局 (NASA : National Aeronautics and Space Administration) が支援している「ベオウルフ (Beowulf)」が知られている。このもとになっているのは、1994 年に Thomas Sterling と Don Becker によって 16 個の Intel DX4 プロセッサと 10Mbit/s Ethernet で構成された PC クラスタである。「ベオウルフ(Beowulf)」は、並列マシンの分類からすると、超並列マシンとワークステーションネットワークの中間に位置づけられるものであり、天体物理学分野に特化した専用システムである。最近の情報では 10 テラプロップスレベルの性能を実現し、更にアルゴリズムを改善することにより 100 倍スピードアップし、ペタフロップスを目指している。

PC クラスタは、PC の性能が格段に向上してきたことと、必要な能力をもったスーパーコンピュータを安価にスケーラブルに構成できる点が魅力となって大学を中心とした研究コミュニティーで、米国ばかりでなく全世界的に普及している。2000 年３月には全米コンピュータ科学連合 (National Computer Science alliance) は基礎研究用の PC ネットワークを IBM の Netfinity PC サーバ 256 台で構成すると発表した。OS として Linux を使い、計画されている演算性能は 375 ギガフロップスである。

我が国でも同様の研究プロジェクトが大学や研究機関で進められている。主要なプロジェクトとして次のものが知られている。

1) 東京大学が中心となって推進している日本科学技術振興会未来開拓学術研究「計算科学」の中で「次世代並列計算機開発」があり、100 テラフロップスオーダの処理能力を持つサブペタフロップスマシンの実現をその目標としている。
(http://olab.is.s.u-tokyo.ac.jp/cse/index-j.html)

2) 理化学研究所では専用の LSI (WINE チップ) を特別に開発し、それを 2688 個並列接続した分子動力学シミュレーション専用の WINE－2 システムを作ったが、そのピーク性能は 50Tflops 超を実現した (1999 年)。今後は分子動力学シミュレーションだけでなく、電子状態のシミュレーションやゲノム系列の解析用の専用計算機を開発する提案をしており、目標性能を１ペタフロップスにおいている。
(http://www.riken.go.jp/r-world/info/release/press/1999/1116/index.html)

3) そのほか、技術研究組合　新情報処理開発機構 (RWCP)、科学技術振興事業団 ERATO 北野共生システムプロジェクトなどの研究機関や大学でも必要な性能を比較的安価にかつ容易に実現できるコンピューティング環境としての魅力から PC クラスタの研究開発、応用が盛んに行われている。

(2) 超並列 (MPP:Massively Parallel Processor) システムアーキテクチャ

代表例は、ASCI Blue、ASCI White である。この延長線上であると考えられる１ペタフロップスマシン Blue Gene を５年後の 2004 年に実現すると発表した (1999 年 12 月)。IBM は政府支援を受けずに独自の開発としている。Blue Gene は従来のシリコン技術を使い、「SMASH (Simple, Many and Self-Healing)」アーキテクチャと呼ぶ全く新しいアーキテクチャを採用する。１ペタフロップス級の性能を実現するために１ギガフロップの演算能力を持つ CPU100 万個を並列に接続し、800 万以上のスレッドを同時実行可能にすると共に、プロセッサやスレッドのエラーの影響を自動的に修復する自己修復機能を持っている。PIM (Processor in Memory) も新技術として採用されている。完成時の最初の仕事はタンパク質の折り畳み構造の解明になる予定 (http://www.ibm.com/news/1999/12/06.phtml)。

(3) HTMT アーキテクチャ

Blue Book 2000 の HECC 研究開発のトップで説明されている。ハイブリッドというその名の通り、超電導技術 (超電導 RSFQ プロセッサ)、光インターコネクション技術、高速半導体技術や磁気記憶技術等の異種技術を組み合わせたものであり、国家安全局 (NSA : National Security Agency)、航空宇宙局 (NASA : National Aeronautics and Space Administration)、防衛高等研究庁 (DARPA：Defense Advanced Research Projects Agency) などからのファンドを受け研究開発を行っている。開発のとりまとめ責任者は NASA ジェット推進研究所の Thomas Sterling である (http://htmt.caltech.edu/)。

開発計画は 2007 年のシステム完成させるまで全５フェーズの計画となっているが、2000 年２月の時点ではフェーズ３ (1999 年７月～2000 年９月末) に入っており、700 万ドルを使って実施した過去２年間の基礎検討結果を詳細にわたり更に検討の詰めを行っているようである。

PITAC 報告書の中で、ペタプロップスレベルの性能を実現するのは、それ自体が目的ではなく、技術を発展推進するための原動力として認識すべきと主張している。このことに意を強くしてか、HTMT プログラムマネージャ達はファンドを更に増額してくれるのならば 2004 年にはプロトタイプのペタフロップスシステムを実現できるということをほのめかしている。

今までは超電導 RSFQ プロセッサを使う HTMT アーキテクチャがペタフロップスマシン実現の主流として見られていた。しかし、(2) で書いたように従来のシリコン技術による MPP マシンの実現計画がでてきた以上は、システムに内在する極低温から常温にわたる 300 度の温度差を克服して、開発してきた個々の要素技術をいかにして組み上げるかといった相互接続、実装技術を確立することがペタフロップス級の性能を実現するために必要と思われる。

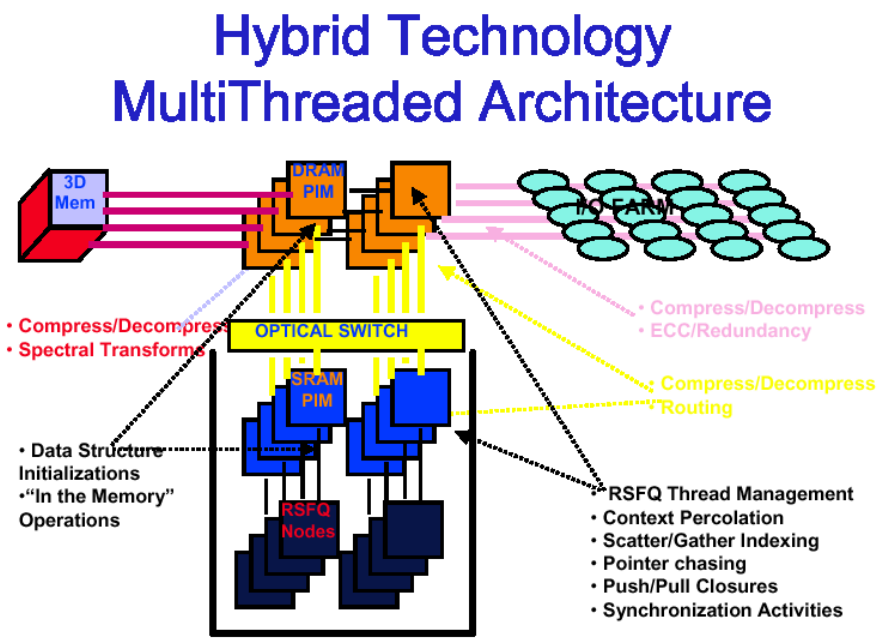


図 2.3　HTMT アーキテクチャのブロックダイアグラム

（出典: Thomas Sterling, Larry Bergman;A Hybrid Technology Approach to Petaflops Computing (1999)）

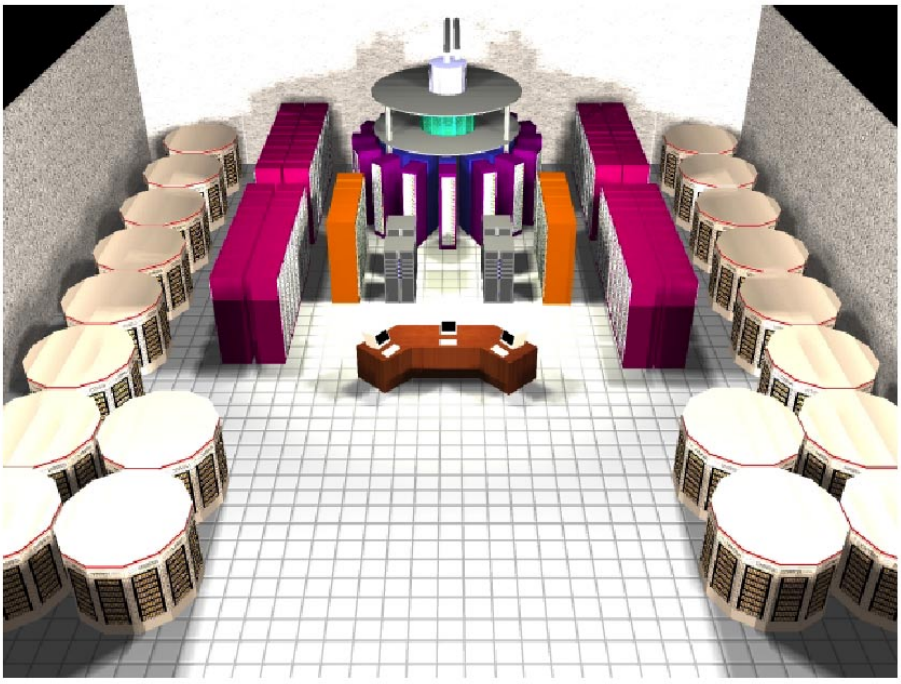


図 2.4　HTMT システムの完成予想図

（出典: Thomas Sterling, Larry Bergman;A Hybrid Technology Approach to Petaflops Computing (1999)）

2.4.3　未来世代コンピューティングーエクサフロップスへ

前項で紹介したのは、ASCI 計画のように実際に稼働しているか、あるいは姿がある程度見えてきている次世代超高性能システムであった。本項ではさらなる性能向上を目指して研究が進められているトピックスを紹介する。

VLSI など半導体デバイスの分野では、「半導体の性能と集積度は、18 ヶ月ごとに 2 倍になる」というムーアの法則に従ってチップサイズ、配線を微細化する方向で性能向上を図ってきた。この延長で行くと、現状のサブミクロンオーダ（１ミリの１万分の１）からデカナノオーダ（１ミリの 10 万分の１）になるのも時間の問題となってきている。更に微細化が進めば、配線のスケールも原子サイズに近づき、デバイスの性能を支配する物理法則もミクロ世界の法則、つまり量子力学に従うものになるはずである。

このような状況も一つの契機となり、しかしミクロ世界の分野へは民間の研究投資もあまりつぎ込まれない基礎研究的性格が強いことから、米国政府は「21 世紀に向けた情報技術：IT^2」イニシアチブの中で革命的コンピューティング（Revolutionary computing）として生物コンピューティング、シングルエレクトロントランジスタ、量子コンピュータなどをあげ、これらは政府が支援すべきものとしている。NSF の 2001 年度予算要求説明書を読めば 2.3 節で紹介した 2001 年度に新たに創設された「国家ナノテクノロジーイニシアチブ」における材料プロセス、デバイス面の研究開発は量子コンピュータ製造プロセスを目標とした未来世代コンピューティング対応であると考えられる。

（http://www.nsf.gov/bfa/bud/fy2001/pdf/budget01.pdf）

本項では未来世代コンピューティングを目指した研究例として DARPA で推進されている UltraScale Computing を中心に最近の研究動向をごく簡単に紹介する。

（1）UltraScale Computing 研究開発の目標

（http://www.sainc.com/arpa/ultrascale/index.htm）

この研究では従来の材料、プロセスにとらわれることなく、性能とコストパフォーマンスの良い先端的な計算技術を開発することが目的であるとされ、現在行っているのは既知の有望先端研究の評価および、それ以外の全く新規なものの探索と評価である。数値目標を表 2.7 に示す。（http://www.sainc.com/arpa/ultrascale/ slides/note001.htm）

表 2.7　UltraScale Computing の性能目標

項目	性能目標
性能	1 Exaflops (= 1,000 Petaflops)以上
コンピューティングユニット数	10^{11} processors 以上
素子の大きさ	原子間距離レベル
消費電力	10^{19} ops/joule 以下
複雑さ	手に負えない複雑さを扱うことができるようにする
賢さ	推量や創造性をもたせる

(2) 計画内容 (http://www.sainc.com/arpa/ultrascale/detail.htm)

計算の新規モデルの研究と計算実行のための新規物理メカニズムの研究からなっており、取り上げられているテーマは前者では並列コンピューティング、群 (swarm) コンピューティング、量子コンピューティング、後者では DNA コンピューティング、細胞工学、神経ネットワークである。これらの研究は 1996 年から 2000 年までの 5 カ年計画となっている。

a. 計算の新規モデル (表 2.8)

1) 数万から数十万という非常に多くのプロセッシングユニットを使用するローエンド側では、並列計算のオーバヘッドの増加を抑えることの出来る新規のアーキテクチャを開発すること。

2) 数百万プロセッサの場合では、新たに出現すると考えられる望ましい振る舞いを引き出すための言語、プロトコルおよびアルゴリズムの開発が必要になる。

3) 更に複雑度が大きくなったケースでは、量子コンピューティングの可能性を評価するために量子状態の重ね合わせやエンタングルメントを探求する。

表 2.8　計算の新規モデルの概要

研究テーマ	研究開発項目	2000 年度目標
並列コンピューティング	Fixed Array, Adaptable Software Continuum Computer Architecture (CCA) (研究契約先:カリフォルニア工科大学)	100teraflop 性能を検証
群 (swarm) コンピューティング	アモーファスコンピューティング プログラム可能材料開発技術、アルゴリズム、アーキテクチャ (研究契約先:マサチューセッツ工科大学)	100 万素子のアモーファスアレーの検証
量子コンピューティング	量子コンピューティング用エンジンの基礎開発 (研究契約先:カリフォルニア工科大学)	量子コンピューティングの実現度決定

b. 計算実行のための新規物理メカニズム（表 2.9）

1） 分子レベル：DNA 分子のデータストレージ、プロセッシング用途の研究

2） 単細胞機構：遺伝子構造操作によるパターン繰り返し、新規製造方法、直接計算の研究

3） 多細胞レベル：電子回路と直接に信号をやりとりできる神経網の人工環境（試験管）あるいはシリコン上での作成

表 2.9　計算実行のための新規物理メカニズムの概要

研究テーマ	研究開発項目	2000 年度目標
DNA コンピューティング	計算実行技術開発と軍事 NP-hard 問題解決戦略立案	軍事アプリケーションの DNA コンピューティングへのポーティング
細胞工学	計算実行および低コスト計算エレメント製造のための生物工学開発	低コスト計算エレメント製造の実証
神経ネットワーク	電子回路と直接相互作用する合成神経網の試験管培養	2 細胞神経網とシリコンロジック間の計算の実証

(3)　最近の研究動向[6]

1985 年にオックスフォード大学の David Deutsch が考案した量子チューリングマシンをベースにして計算理論を研究する分野を量子コンピューティングという。

(http://www.qubit.org/resource/deutsch85.pdf)

David Deutsch の考案からおよそ 10 年後、1994 年に AT&T の Peter Shor が量子コンピュータを用いれば整数の素因数分解が小さな誤り確率で高速に行えることを示した。

(http://xxx.lanl.gov/abs/quant-ph/9508027)

量子コンピュータは複数個の電子、イオンなどの量子力学的状態（qubit）の重ね合わせを利用して並列計算を実行する。56 qubit の場合 2^{56}（7.2057×10^{17}）通りの計算を一度に処理してしまう。このことは桁数の大きい整数の因数分解は現状のスーパーコンピュータを使っても解読に非常に長い時間がかかるので安全だとしている暗号の信頼性を揺るがすことになり、国家安全保障上からも量子コンピューティングの研究がにわかに活発になってきている。

現時点での量子コンピューティング研究の状況は、アルゴリズム、誤り訂正理論研究、ゲート設計および基本的なデバイスを作成して動作原理確認を行っている段階であるが、大学、国立研究所および民間企業においても盛んに研究が行われているようである。WEB で公開されている代表的な研究機関を以下にあげておく。

・Quantum Computation/Cryptography at Los Alamos　http://qso.lanl.gov/qc/

・Quantum Information and Computation

http://theory.caltech.edu/~quic/index.html

QUIC は（2）の DARPA Ultrascale Computing Program のファンドを受けているものでカリフォルニア工科大学、南カリフォルニア大学の共同プロジェクトである。

・The Stanford-Berkeley-MIT-IBM NMR Quantum Computing Project

http://squint.stanford.edu/

このプロジェクトも DARPA の支援を受けたもので、スタンフォード大学、カリフォルニア大学バークレー校、マサチューセッツ工科大学、IBM の共同で推進されている。

・IBM

http://www.research.ibm.com/quantuminfo/

Quantum Cryptography、Quantum Teleportation、特定計算の指数関数的スピードアップが主テーマとなっている。

2000 年度版ブルーブックに記載されている量子コンピューティング関係の研究動向を以下に記す。そのほかの新しい概念としてバイオコンピュータ技術、光コンピュータ技術というキーワードはでているが、その内容は記載されていない。[3]

（ア）量子コンピューティング

国家安全局（NSA: National Security Agency）のサポートで、9 つを超える数の大学と企業が研究を行っている。期待されている研究成果は以下の通り。

・古典的な情報通信理論の量子版を開発する。将来の量子コンピューティングの設計と評価に役立てる。

・「隠れたサブグループ（hidden subgroup)」問題を新しい量子アルゴリズムの設計に適用する。

・従来型コンピュータと量子コンピュータの間に位置するモデルで達成できる計算能力を研究する。

・光子間に量子論理ゲートを導入するために新しいパルス レーザー技術を検証する。

・３つの光子のもつれを作成する。

・コヒーレントな量子状態における少数イオンに対する操作を達成する。

・量子の重ね合わせ、もつれ、多重 qubit 操作の証拠を明らかにする。

・結果として生じる量子のコヒーレンスを特徴づける。

将来の研究では1-qubitの操作を実証すること、原子を捕らえる光格子法を使って2-qubitの操作を達成する実験を行うこと、すでに達成されているよりもっと詳しく１揃いのqubitのダイナミクス（動力学）をシミュレートすることを計画している。

NASA の量子コンピューティング プロジェクトの最初のフェーズ（５年間）は1999 年度に終了。実用的な量子コンピュータの開発には 20 年かそれ以上の年数の研究が必要かもしれないと見ている。

（イ）量子テレポーテーション

DOE のオークリッジ国立研究所（ORNL：Oak Ridge National Laboratory）は、量子コンピューティングと通信の分野の研究領域である量子テレポーテーションを研究する最先端の研究所を設立した。高出力の 1,000 兆分の１秒パルスレーザーを使ってシグナル発信機と発信されたシグナルが量子力学的にもつれ合った状態での実験を行っている。

（ウ）量子暗号化

NSA では、基本的な保護モードとして量子物理学を使用するキー交換技法を採用することによって、攻撃に影響されることのない新しいクラスの暗号化について研究を行っている。最近の研究では 47 キロメートルの光ファイバー・ケーブル上でのキー交換実現の可能性を示した。

我が国においても、量子コンピューティングは電総研、理化学研究所、大学、日本電気、三菱電機の研究所などで活発に研究が行われており、これからの成果が期待できるのではないかと思われる。

参考文献

[1] 2000 年度版ブルーブックの和訳は下記のところで公開されている
http://www.icot.or.jp/FTS/Ronbun/bluebook2000-J.htm
[2] ペタフロップスマシン技術に関する調査研究　Ⅲ,1999 年 3 月
[3] ペタフロップスマシン技術に関する調査研究　Ⅱ,1998 年 3 月
[4] ペタフロップスマシン技術に関する調査研究　Ⅰ,1997 年 3 月
[2]～[4]は下記のところで公開されている
http://www.icot.or.jp/FTS/REPORTS/Report-index-J.html#10nendo
[5] Thomas Sterling, Paul Mesina and Paul H. Smith, ペタフロップスコンピュータ

（筑波出版会、1997）

[6] 量子コンピューティングを含む計算理論の解説書

西野哲朗、中国人郵便配達問題=コンピュータサイエンス最大の難関（講談社、1999）

David Deutsch , The Fabric of Reality : The Science of Parallel Universes-And Its Implications (Scholastic Paperbacks,1998)

A. ヘイ、R. アレン、ファインマン計算機科学（1999, 岩波書店）

第 3 章　わが国における研究開発の状況　－課題と展望－

平成 11 年度のワーキンググループの活動は、米国の最新研究開発動向を参考に、また過去 3 年間のワーキンググループの活動もふまえ、次のような活動方針とした。

【方針】

米国の HPCC プログラムの中で特に技術的に興味あるものであり、標準化の動き、市場動向なども視野に入れて見たときに市場にインパクトを与えそうな分野やテーマについてワーキンググループメンバーによる技術的内容および成果の調査・評価と我が国の技術との格差などについて議論する。また、テーマによっては適宜外部講師による講演も実施する。

活動結果は我が国が行うべき情報技術開発重点テーマのマップ作成の基礎データとして活用する。

本章の報告は上記の活動方針に沿って、各委員の研究テーマに関連する 2000 年度版ブルーブックで取り上げられているものの技術的内容および成果の調査・評価と我が国の技術との格差などについて見解を執筆していただいた。

平成 11 年度ワーキンググループ委員が実際に携わっている研究テーマと、2000 年度版ブルーブックを参考にした分類項目との対応付けをしてみると表 3.1 のようになっている。

今年度は 2 名の方に外部講師をお願いし、最新状況を講演していただいた。講演内容は本章の報告としてまとめていただいた。

（１）工藤知宏　講師(新情報開発機構)：「ローカルエリアで高性能並列計算を実現するネットワーク　RHiNET」

（２）田中秀俊　講師(三菱電機)：「健康と気象のデータマイニング」

表 3.1　ワーキンググループ委員テーマ

Blue Book 2000	委員担当テーマ
(1)アーキテクチャ	**天野委員**:PC レベルに繋がる光チップがほぼできた。現在、突出した性能だが使用法が不明。このあたりも考慮し、アーキテクチャ・クラスタコンピューティング・光ネットワークをサーベイし、WG に貢献できないかと考える。分野的には(1)。 **久門委員**:メモリ・クラスタコンピューティング。 **中島委員**:最近の Processor In Memory 関連研究
(2)ハードウエアコンポーネンツ	**花輪委員**: (2)あたり。(1)の Commodity あたりも興味はある。
(3)基礎理論およびアルゴリズム	**笠原委員**: (3)、(4)の間。コンパイラ関係・クラスタ・グローバルコンピューティング・スケジューリング関係。 **近山委員**: Continuum Computing で、その上でのソフトウェアモデル。分野は(3)並列分散コンピューティング付近?
(4)並列分散ソフトウエア	**妹尾委員**: (4)あたり。コンパイラに軸を置き、グローバル・分散処理を横目に睨んだあたり。 **関口委員**: グローバルコンピューティングとそのアプリケーションあたりでネットワークを含めた 5～10 年先のもの。(4)をベースに(5)にもからむか。 **古市委員**:異機種分散シミュレーション接続アーキテクチャ HAL と並列分散シミュレーションへの応用
(5)可視化技術・シミュレーション技術・アプリケーション	**秋山委員**:蛋白質あたりをメインに(7)の DNA コンピューティングあたり。 **福井委員**:シミュレーション技術・アプリケーションあたりをベースに。場合により、(アプリケーション側からみた)モデリング。 **横川委員**:シミュレーションの現状と将来の目標をあわせてどのくらいの性能が必要か、どういうものに役に立つのかを考える。
(6)ASCI 計画	
(7)その他量子/DNA コンピューティングなど	**秋山委員:** DNA コンピューティング

3.1　概要

3.1.1　ハイエンドコンピューティング国家戦略の必要性

笠原博徳　委員

ここでは、Information Technology Frontiers for a New Millennium（大統領 2000 年度予算教書の補足）における米国におけるハイエンドコンピューティングに関する提案をまとめると共に、我が国におけるハイエンドコンピューティング戦略の必要性について論じる。

3.1.1.1　NSTC（National Science and Technology Council）国家科学技術委員会

まず Information Technology Frontiers for a New Millennium の発行母体は

NSTC（National Science and Technology Council）国家科学技術委員会

であり、これは

OSTP（Presidential（White House）Office of Science and Technology Policy）

の下部組織として、大統領が、科学、宇宙、技術を連邦政府横断的にコーディネートするため 1993 年に設立したものである。

この委員会の議長は大統領であり、委員は副大統領、科学技術担当大統領補佐官、閣僚秘書、特に深く科学技術に関連するエージェンシィ長、ホワイトハウス官僚等から成っている。委員会の主目的は、

・情報技術及び保険衛生から輸送システムの改善
・基礎研究の強化に至る分野に対する連邦政府科学技術投資のための明確な国家目標の確立

である。

NSTC は、下記の５つのコミッティを持つ

- Committee of Environment and Natural Resources
- Committee on International Science, Engineering, and Technology
- Committee on National Security
- Committee on Science
- *Committee on Technology （CT）

この５つのコミッティの内、今回の調査の対象となる CT（テクノロジ委員会）は、さらに下記の７つの R&D サブコミッティを持っている。

- Interagency Working Group for the U.S. Innovation Partnership
- Interagency Working Group on Critical Infrastructure Protection R&D
- Partnership for New Generation of Vehicles
- Subcommittee on Building and Construction
- Subcommittee on Materials
- Subcommittee on Transportation R&D
- *Subcommittee on Computing,Information, and Communication (CIC) R&D

3.1.1.2　CIC（コンピューティング、情報、通信 R&D サブコミッティ）と HPCC R&D

この内、ハイエンドコンピューティングを扱う、CIC（コンピューティング、情報、通信 R&D サブコミッティ)は､HPCC R&D に参加する 12 のエージェンシィの代表と OMB(White House Office of Management and Budget）と OSTP からの代表で構成されている。この **CIC** は、さらに下記のような５つのワーキンググループと複数のサブグループを持っている。

- High End Computing and Computation (HECC) WG
- Large Scale Networking (LSN) WG
    - High Performance Networking Applications Team (HPNAT)
    - Information Security Team (IST)
    - Joint Engineering Team (JET)
    - Networking Research Team (NRT)
- High Confidence Systems (HCS) WG
- Human Centered Systems (HuCS) WG
- Education, Training, and Human Resources (ETHR) WG

その他に FISAC (Federal Information Services and Council) がある。

現在の **HPCC　R&D プログラム**（以前は **CIC** と呼ばれた）は、前記 CIC の WG に対応した５つのプログラムコンポーネントエリア(PCA)により構成されている。

このHPCC R&Dマルチエージンシィ R&Dプログラムの2000年度予算は、$1462 Millionで、$1314 Million であった1999FY（予算年度）に比べ11%増加している。

また、NCO /CIC（The National Coordination Office for Computing, Information, and Communications：コンピューティング、情報、通信に関する国家調整局）という組織が、CIC R&D サブコミッティのサポートを行っている。また、この組織は同時に後述するPITAC（大統領の情報テクノロジ・アドバイザリ委員会)、IT^2 （21世紀イニシアティブ情報テクノロジ・プロジェクト）のとりまとめもサポートしている。

なお、PITACは、このHPCCにおけるNCOと各エージェンシィ代表を集めたワーキンググループの構成は、エージェンシィ間協力の良いモデルであり、今後 IT（情報テクノロジ）R&D全般に取り入れられるべきであると提案している。

3.1.1.3　PITAC (President's Information Technology Advisory Committee)

1997 年２月に Clinton 大統領が、将来の IT R&D 強化を目指し、**High Performance Computing, Communications, Information Technology, Next Generation Internet** に関するアドバイス並びに、連邦政府の役割に関する評価、を受ける目的で設置したコミッティであり、大学、産業界リーダ26人から構成されている。

主要な活動としては

1998年８月　　暫定レポート発行

11月４日　SC98（オーランド）のタウンホールミーティングで公開

1999年２月４日　“Information Technology Research: Investing in Our Future”を大統領に提出が挙げられる。

(1) PITACの評価結果、提案の概要

PITACによる主な評価結果と提案の概要は以下である。

- 情報技術の経済的・戦略的重要性から、情報技術 R&D に関する連邦予算を増大させるべき。
- 現在のIT研究に関する連邦のサポートは、産業界の競争的環境の影響から短期ミッ

ションオリエンテッドなものに集中し、長期・ハイリスクの研究サポートは削減されており、不適切である。

・連邦政府は戦略的なイニシアティブを確立し、2004 年度までに$1.37billion（約 1400 億円）程度増大させ、その増額した分は夢のような研究・ハイリスクな基礎研究を促進するために使うべき。そのために、研究期間の延長、研究サポートの方法の多様化を考えるべき。

・NSF が基礎研究でリーダーシップを取るべき。

・IT R&D のための Senior Policy Official を選定すべき。

・HPCC プログラムコーディネーションを全ての主要 IT R&D に拡張すべき。

・研究目的と研究費に関する年次報告を行うべき。

等である。

(2) High-End Computing における重点研究課題

また、PITAC は、高速計算とデータ転送能力を持つ超高速コンピューティングシステムが、気象・環境予測、高度生産デザイン、新薬の開発、国家クリティカルインタレストのサポートのために必須であると提言している。

特にこの提言の中では、以下の項目に速やかに研究費を重点配分すべきと主張している。

- 革新的コンピューティング技術・アーキテクチャ
  - 将来のアプリケーションが要求するメモリバンド幅確保のためのレーテンシを隠蔽するメモリ階層設計、光、量子、バイオ、ニューロコンピューティング、ハード・ソフト協調型アーキテクチャ

- ハイエンドコンピューティング性能向上のためのソフトウェア R&D

  High-End Computer の性能及び効率向上のためのソフトウェアに投資すべき
  - 特にシステムソフト：言語*、コンパイラ*、OS、ランタイムライブラリ、ファイルシステム、I/O ドライバ、デバッガ、チューニングツール*

    (*コンパイラに関しては、国内でも通産省アドバンスト並列化コンパイラプロジェクトでサポート予定)

  ソフトが petaops システムの開発においてより重要な役割を果たす。

- ハード／ソフト戦略のバランスをとり、2010 年までに実アプリケーションに対して petaops あるいは petaflops の持続性能を得るための High-End Computing 研究を立ち上げるべき

→この提言は Blue Book 中で記載されている HTMT (Hybrid Technology Multithreaded) architecture プロジェクトの推進にも連携している。

HTMT (Hybrid Technology Multithreaded) マシンプロジェクト

Caltech/NASA JPL の Dr.Thomas Sterling を中心に多くの研究機関が連携して、Petaops を目指し、超伝導技術、光インタコネクト（データボルテックス、3D ホログラフィックメモリ)、ハイスピード VLSI 技術、大容量磁気ストレージ技術等の革新的な技術を組み合わせハイパフォーマンス・コンピュータを開発するプロジェクトで、1999 年度の成果により 2000 年度よりプロトタイプの開発を開始することになっている。

また、Blue Book 中では 2006 年頃の完成を目標としているが、今年の 5 月 31 日から 6 月 1 日に早稲田大学国際会議場にて開催される 2000 年記念並列処理シンポジウム JSPP2000 での Dr.Thomas Sterling 招待講演アブストラクトでは 2005 年完成予定と 1 年計画が前倒しされている。

- 科学・工学分野の研究をサポートするために最もパワフルな High-End Computing の獲得
  - NSF PACI (Partnerships in Advanced Computational Infrastructure) Centers, NSF National Center for Atmospheric Research, DoE ASCI に予算を重点配分すべき。それらに加え DoD, DoE, NASA もミッションオリエンテッドな研究のためにマシンを獲得すべき。
  - HECC WG のコーディネーション機能を全主要政府 High-End Computing プロジェクト（DoE ASCI, DoD HPCMP: High Performance Computing Modernization Program, NASA 関連プロジェクト等を含めて）へ適用すべき。

3.1.1.4 IT^2 : Information Technology for Twenty First Century Initiative

PITAC の提案を受けて、大統領は 2000 年度　R&D 予算のハイライトとして IT^2（Information Technology for Twenty First Century Initiative）を計画している。これは、21 世紀に向けた情報革命を持続するために、基礎情報科学における知識ベースを進歩させ、次世代研究者の育成のために$366 Million の追加投資を行うというものである。

この研究費は、NGI (Next Generation Internet), DoE ASCI を含めた HPCC に加え、以下の３つのキーエリアに配分されるべきと述べられている。

・コンピューティングと通信分野における基礎的な進歩を導く長期的 IT 研究
・ 国民に利益をもたらす科学的、工学的発見を促進するツールとしての先端的コンピューティングインフラストラクチャ
・情報革命がもたらす社会・経済効果に関する研究及び大学における IT 従事者の育成

また、議会で承認されれば、HPCC　R&D プログラムと統合し進められるべきと提案されている。

3.1.1.5 提案

上記、米国　Information Technology Frontiers for a New Millennium、PITAC(President's Information Technology Advisory Committee) 提言で記載されている事項を踏まえ、我が国においても 21 世紀の IT を中心とした産業競争力の確保・強化、基礎研究の強化を行っていくためには、ハイエンドコンピューティングに関する中・長期戦略を考えていくことが必須と思われる。

具体的にハイエンドコンピューティングに関連して至急考慮すべき事項には、以下のようなものが挙げられる。

① HPCC R&D プログラムの様に省庁横断的な研究プロジェクトにより、21 世紀の情報産業並びに High End Computer を使用して製品開発を行う産業、21 世紀を支える基礎研究を促進するために日本でも High End Computing 研究をさらに進めるべき。

これに関連して国内でもミレニアムプロジェクトのような省庁横断的なプロジェクトが行われている。このような努力をさらに進めていくべきと考える。

② HECC のような関連省庁代表が委員となる委員会の設置を検討すべき。

③ PITAC に対応するような日本の将来戦略を提案し、現在の省庁にわたる研究プロジェクトの評価を行うような仕組みを検討すべき。

④ 米国では PITAC による 2010 年までに実効性能 PFLOPS を達成するハイパフォーマンス・コンピュータを開発するという提言を受け、 2005 年 PFLOPS（ピーク性能と考えられる）を目指した HTMT プロジェクトがスタートしようとしている。これに対し、日本では 2001 年度の地球シミュレータ以降の大規模計画は策定されていない。
至急検討を開始すべきである。

この内容にも関連し、本年６月１日に早稲田大学国際会議場で開催される JSPP2000 ミレニアム記念スーパーパネルディスカッション“ペタフロップスへの道”で、我が国の産学を代表する研究者・技術者を集め、

・日本においてペタフロップスを目指すようなハイパフォーマンス・コンピュータの開発は必要か？
・必要な場合には、小さな市場、莫大な開発費を考えてどのように開発すべきか？
国家プロジェクト？、企業間連携？、産官学の協力？　など、どのようなアプローチを取るべきか？
また、開発時期はいつが適切か？
・日本において開発する場合にはどのような目的（アプリケーション）で使用されるべきか？　またそのようなマシンが世界にどのように貢献できるか？
・アプリケーションを設定する場合、アプリケーション研究・開発者とアーキテクチャ、システムソフトウェア研究・研究開発者の協力のあり方
・ペタフロップスあるいはさらに先のハイパフォーマンス・コンピュータを開発する際には、どのようなアーキテクチャ、システムソフトウェア、アプリケーションにどのような技術的課題があるか？
・もし国内でのハイパフォーマンス・コンピュータの開発が必要ない場合には、情報産業は何を牽引力として技術開発を行うべきか？

などが議論される予定である。

このような学会活動も通して、ハイパフォーマンス・コンピュータに限らず、21 世紀の IT を支える中・長期技術は何か、またどのように振興すべきかを産官学で継続的に検討していくことが、21 世紀も日本が持続的産業競争力をもつために必須であると考えられる。

⑤ PITAC、HECC が推進する High End Machine 用ソフトウェアの重要性を認識し、ソフトウェアに関する研究をさらに促進すべきであると考えられる。

3.2　アーキテクチャおよびハードウェア

3.2.1　PIM アーキテクチャの動向と展望

中島　浩　委員

PIM（Processor-In-Memory）アーキテクチャとは、近年のプロセッサ／メモリ混載技術を活用し、メモリ・チップ上に（比較的小規模の）プロセッサを搭載して、メモリ機能の高度化を図ろうというものである。したがって PPRAM などのプロセッサ／メモリ混載アーキテクチャがどちらかと言えばプロセッサを主体とする（Large）Memory on Processor Chip であるのに対し、メモリが主体という点で MBP や FLASH などの所謂プロトコル・プロセッサと近親性がある。ただし、これらのプロセッサでは（実装上の制約もあって）メモリ・バンド幅がさほど大きくないのに対し、PIM ではチップ内の大きなバンド幅を生かすようなアプローチとなっている点が異なる。

さて、PIM が目的としているメモリ機能の高度化について、HECC-WG の前身である Peta-FLOPS WG の平成９年度報告書において、筆者は「メモリ・アーキテクチャの展望と課題」と題して、以下のような議論を行った[1]。すなわち、

・超大容量メモリの必要性
・演算／メモリ・スループットのバランス確保
・von Neumann bottleneck の存在

を前提とするならば、von Neumann bottleneck の解消／緩和が Peta-FLOPS アーキテクチャの鍵であり、そのためには bottleneck 通過物（現状ではアドレスやデータ）の変質が必要であると主張した。またその候補となるアーキテクチャとして、以下の三つのものを取り上げて議論した。

(1) Visible Memory System

キャッシュに代表される不可視のメモリ・システム要素を可視化し、何らかの制御を可能とするアーキテクチャとして JUMP-1 [2] などを取り上げ、可視要素を含む非均一メモリ・システムの統一的ビューの設定がソフトウェアによる安定的利用に不可欠であると主張した。その後 SCIMA [3] をはじめとして、メモリ性能の向上を目的とした種々の可視メモリ要素が提案されているが、統一的ビューの設定に関する提案は見られない。

(2) Memory with Program

プログラム実行のための制御機能の一部をメモリに移管する先鋭的アプローチとして UPCHMS [4] を取り上げ、アクセスのためにメモリに渡される情報をアドレス列よりも

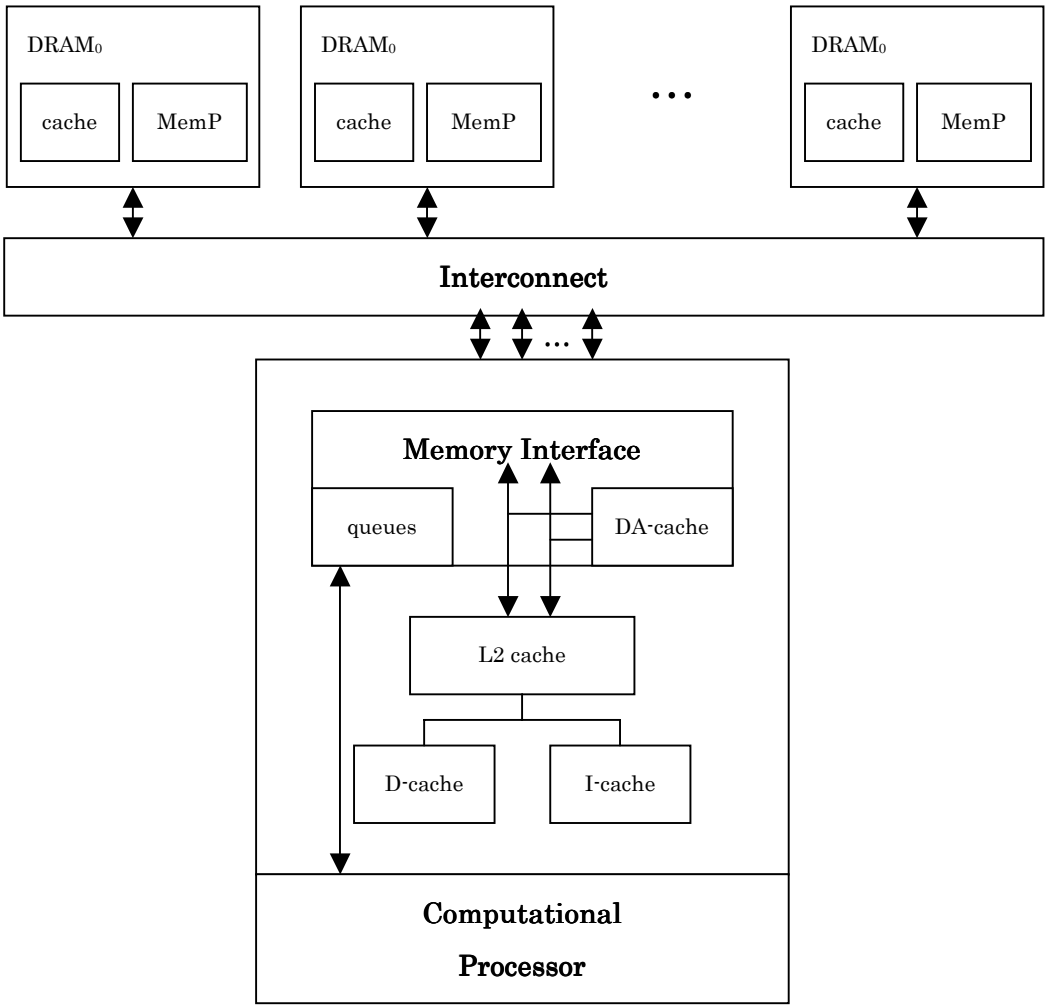


図１　DA-DRAM system [6]

十分小さくできる可能性を議論した。

(3) Code (PC) Flow Model

(2) をさらに発展させたアプローチとして、メモリ上の個々のオブジェクト（あるいはオブジェクトの集合）が固有のプロセッサを持ち、プログラム・コード（あるいはプログラム・カウンタ）と実行コンテクストの組がオブジェクト間を移動するようなモデルを提言した。

今回の報告で Processor-In-Memory (PIM) アーキテクチャについて議論するために、最近の研究動向を調査したところ、上記の (2) や (3) のアプローチに基づくものが見出され、２年前の提言が必ずしも的外れではなかったことが明らかになった。以下、(2) については Veidenbaum らによる DA-DRAM [5,6] を、また (3) については HTMT アーキテクチャにおける smart RAM [7,8,9] を例にとって議論する。

3.2.1.1 DA-DRAM

DA-DRAM (Decoupled Access DRAM) は、現在の高性能 uniprocessor system をさらに高性能化することを主な目的としたもので (マルチプロセッサについても可能性は否定していない)、図 1 に示すように計算の主体である 1 個の computational processor (ComP) と、メモリ・アクセスのためのプロセッサである memory processor (MemP) を持つ複数の DRAM から構成される。ComP は superscaler などのような一般的なプロセッサであり、キャッシュあるいはキューを介して MemP との間の通信を行う。MemP は単純な整数プロセッサであり、アドレス計算を含むメモリ・アクセス操作全般 (特に読出) を担当し、ComP が必要とするデータをプログラムの先行実行によってプリフェッチして ComP に供給する。

ComP と MemP の間の通信をキャッシュ／キューのどちらを経由して行うかは、アドレス計算の複雑さによって決定される。スカラあるいは配列要素のように比較的単純にアドレスが求まる場合には、ComP と MemP の双方でアドレス計算を行い、データはキャッシュを介して受け渡される。たとえば

```
for (i=0;i<N;i++)  y[i]=a*x[i]+y[i];
```

は、以下のように ComP/MemP で分担実行される。

MemP

```
for (i=0;i<N;i++) {
    send_to_cache(x[i]);
    send_to_cache(y[i]);

}
```

ComP

```
for (i=0;i<N;i++) {
            X=recv_from_cache(x[i]);
            Y=recv_from_cache(y[i]);
        y[i]=a*X+Y;
}
```

この例でのキャッシュの役割はデータの再利用というよりもむしろ addressable なバッファである。また再利用を考える場合も、データの存在が確定的ではないキャッシュよりも、UPCHMS や SCIMA のように高速バッファ・メモリを用いるほうが勝るのではないかと思われる (UPCHMS のようにバッファにもプロセッサを置く必要はなかろうが)。

なお、この例では明らかに MemP におけるメモリ・アクセス・スループットが ComP での演算スループットよりも大きく劣るが、このスループット差は複数の MemP が並列実行する一種のインタリーブによって補われる。M 個の MemP により並列実行するには、個々の MemP が

```
for (i=m;i<N;i+=M) {
    send_to_cache(x[i]);
    send_to_cache(y[i]);
}
```

のように、サイクリック分割に基づく並列化コードを実行すればよい。しかし［5］ではより単純に

```
for (i=0;i<N;i++) {
    if (is_local(&x[i])) send_to_cache(x[i]);
    if (is_local(&y[i])) send_to_cache(y[i]);
}
```

という一種の SIMD 的な実行で十分であると主張している。

一方上記のプログラムは、以下のようにキューを使った通信によっても実現できる。

MemP

```
for (i=0;i<N;i++) {
    send_to_queue(x[i]);
    send_to_queue(y[i]);

}
```

ComP

```
for (i=0;i<N;i++) {
          X=recv_from_queue();
          Y=recv_from_queue();
    y[i]=a*X+Y;
}
```

このキューを使った通信は、アドレス計算が複雑である場合により有効であると［6］では主張している。すなわち複雑なアドレス計算は（若干奇異に思えるかもしれないが）MemP で行い、ComP には計算すべきデータのみが送られる。たとえば以下のような疎行列 A と密ベクトル V の乗算を考える。

```
for (i=0;i<N;i++) {
    ip=0.0;
    for (j=head[i];j<head[i+1];j++)
        ip=ip+As[j]*V[idx[j]];
    AV[i]=ip;
}
```

この例ではAのi 行目の非零要素がAs[head[i]]から連続して格納されており、As[j]の列番号は idx[j]に保持されている。このプログラムでは、V の要素のアドレス計算にインデックス配列参照が必要であるため、以下に示すようにアドレス計算は MemP のみが行い、ComPはキューを介してデータを受け取る。

MemP

```
send_to_queue(head[0]);
for (i=0;i<N;i++) {

  send_to_queue(head[i+1]);
  for (j=head[i];j<head[i+1];j++) {
    send_to_queue(As[j]);
    send_to_queue(V[idx[j]]);

  }


}
```

ComP

```
h1=recv_from_queue();
for (i=0;i<N;i++) {
  ip=0.0;
  h2=recv_from_queue();
  for (j=h1;j<h2;j++) {
    a=recv_from_queue();
    v=recv_from_queue();
    ip=ip+a*v;
  }
  AV[i]=ip;
  h1=h2;
}
```

この例についての MemP の並列実行も、前述の SIMD 的なやり方で行うことができる。ただしV[idx[j]]のアクセスはidx[j]を保持するMemPが行う必要があり、そのため一般には遠隔アクセスとなる。

リスト構造のようなポインタを用いたアクセスについては [5,6] では触れられていないが、たとえば

```
for (p=head;p!=NULL;p=p->next)
    compute(p->elem);
```

のようなプログラムを

MemP

```
send_to_queue(head);
for (p=head;p!=NULL;p=p->next) {
    send_to_queue(p->elem);

    bloadcast(p->next);
}
```

ComP

```
p=recv_from_queue();
for (;p!=NULL;) {
    e=recv_from_queue();
    compute(e);
    p=recv_from_queue();
}
```

のようにMemP間でも通信を行いながら実行することは可能であろう。

さて、上記のプログラムは全て、MemP が制御フローを完全に把握できるものである。しかしComPによる演算結果に制御フローが依存する場合、たとえば

```
while (e>eth) { ...; e=...; }
```

のような収束計算を行う場合、MemPのComPに対する制御依存による遅延が問題となる。そこで［6］では、MemPが多重分岐予測を行うことによって、制御依存遅延の問題を解決できると主張している。すなわち上記のループを

ComP

```
while (pred(TC)) {
    ...

}
```

MemP

```
while (e>eth) {
    ...
    TC=!(e>eth);
}
```

として実行し、分岐予測（pred(TC)）の精度が高ければ、十分な性能が得られるとしている。精度について［6］では、過去 18 回の分岐履歴に基づいて８つ先の分岐を予測した場合、70～80%程度の正解率であることを示している。

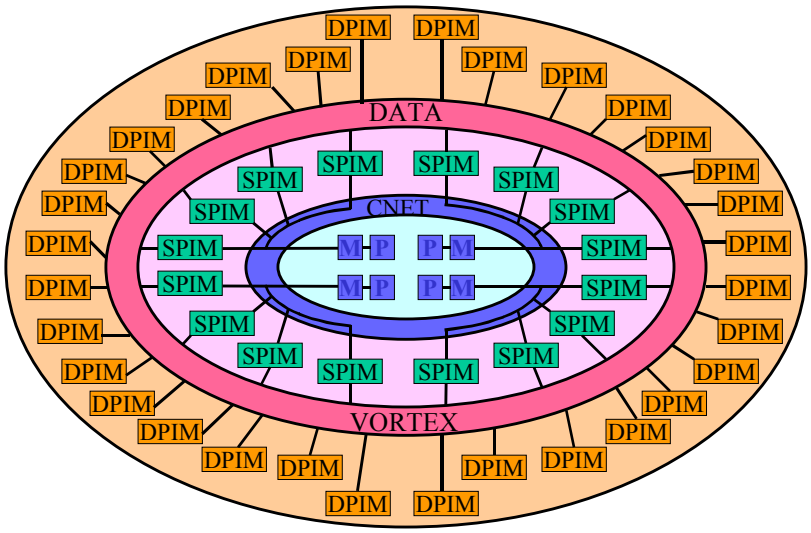


図２　HTMT アーキテクチャの概要［7］

3.2.1.2　Smart RAM

Smart RAM は HTMT（Hybrid Technology Multi-Threaded Architecture）[7,8,9]の構成要素である。HTMT は図２に示すように

・超伝導素子 RSFQ（Rapid Single Flux Quantum）で構成された 66GHz のプロセッサ SPELL

・超伝導素子で構成された１MB の高速 RAM（CRAM）

・PIM アーキテクチャによる４×64MB の smart SRAM（SPIM）のクラスタ

・PIM アーキテクチャによる８×512MB の smart DRAM（DPIM）のクラスタ

から構成されている。これらは全て 4,096 個ずつ存在し、以下のネットワークにより相互接続される。

・SPELL-CRAM 間：2TB/sec の RSFQ

・SPELL/CRAM 相互：512GB/sec の CNET

・SPELL/CRAM-SPIM 間：160GB/sec の CNET

・SPIM-DPIM 間および相互：80GB/sec の光ファイバ（Data Vortex）

SPIM/DPIM には、SPELL と SRAM/DRAM の間のレイテンシ隠蔽を目的として、それぞれ固有のプロセッサが搭載されている。プロセッサは ASAP（At the Sense Amps Processor）と呼ばれ、その名のとおり RAM モジュールのセンスアンプ入出力に直結され、大きなアクセスバンド幅（[7]によれば 1K bit）を活用できるようにしている。プロセッサの詳細な構成

は「データフローのアレイとマルチスレッド CPU の組合せ」[9]という程度しかわからないが、SIMD や VLIW 的な操作に対応できるものであるという。

PIM 相互および PIM と SPELL の間の通信は、parcel (Parallel Communication Elements) と呼ばれる高機能メッセージによって行われる。parcel はメモリ上のオブジェクトに宛てたメソッドとみなすことができ、PIM や SPELL によってスレッドとして実行される。この parcel を用いた DRAM ⇔ SRAM ⇔ CRAM のデータ転送は percolation と呼ばれる。

たとえば [7] に示された行列積 $C = A \times B$ の例では $M \times M$ の行列 A の percolation は以下のように行われる (図３)。

（１）A は t^2 個の小行列 $A_{i,j}$ に分割され、$A_{i,j}$ はさらに s^2 個の小行列 $a_{\alpha,\beta}$ に分割される。

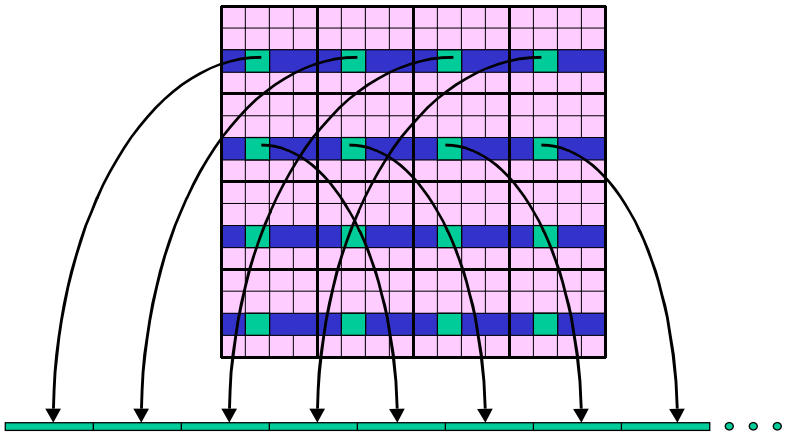


図３ HTMT による行列積での percolation [7]

(1) 一つの DRAM には $\alpha = m \pmod s$ なる $a_{\alpha,\beta}$ が全て格納される。

(2) 一つの SRAM には $\alpha = m \pmod s$ かつ $\beta = n \pmod s$ なる $a_{\alpha,\beta}$ が DRAM から「内向きに」percolation される。

(3) 一つの SPELL は $\alpha = n \pmod s$ かつ $\beta = n \pmod s$ なる C の小行列 $c_{\alpha,\beta}$ の生成を担当し、$a_{\alpha,\beta}$ と $b_{\alpha,\beta}$ が SRAM から「内向きに」percolation される。

(4) 小行列、$a_{\alpha,\beta}$ と $b_{\alpha,\beta}$ は隣接する SPELL 間で循環的にシフトされ、$C_{i,j,k} = A_{i,k} \times B_{k,j}$ の乗算が行われる。これを全ての k について繰り返すことにより $C_{i,j}$ が求められる。

[7]では前述の RAM 容量に基づき、$M = t \times s \times 125 = 26 \times 64 \times 125 = 208{,}000$ とし、か

つ RAM のアクセス時間と SPELL の演算スループットについて以下の仮定を置いて、HTMT の行列積性能を見積もっている。

・DRAM ： 7cycle@450MHz / 1K-bit = 0.97ns / W

・SRAM ： 3cycle@450MHz / 1K-bit = 0.42ns / W

・SPELL ： 5FLOP / cycle@66GHz = 330GFLOPS

見積もりの結果は 1.1 PFLOPS となり、ピーク性能である 1.35 PFLOPS の約 80 % が得られると主張している。

3.2.1.3　まとめ

今回の報告では、PIM を用いた高性能計算の例として、DA-DRAM と HTMT について述べた。どちらのアーキテクチャも von Neumann ボトルネックを通過する情報を変質させることによりボトルネックの緩和を図っており、[1]で示した展望に添った形で研究が進められていることは興味深い。

これらのアーキテクチャの有効性についての検討は現時点では十分でなく、一般的に高性能計算に有用であると立証されているわけではない。特に一般性について、今後どのように研究が展開するかを見守っていく必要があろう。

参考文献

[1] 中島浩：メモリ・アーキテクチャの展望と課題、ペタフロップスマシン技術に関する調査研究、3.1.4 節, pp. 54-60, Mar. 1998.

[2] 五島正裕, 森眞一郎, 中島浩, 富田眞治：Virtual Queue: 超並列計算機向きメッセージ通信機構, 情報処理学会論文誌, Vol. 37, No. 7, pp. 1399-1408, Jul. 1996.

[3] 大河原英喜, 中村宏, 吉江友照, 金谷和至：ハイパフォーマンスコンピューティングに適したメモリ階層の初期評価, 情報処理学会 ARC 研究報告, ARC-133-10, Mar. 1999.

[4] 牧晋広, 岡本秀輔, 曽和将容：ユーザプログラム制御階層メモリシステム, 情報処理学会論文誌, Vol. 37, No. 10, pp.1512-1526, Oct. 1996.

[5] A. V. Veidenbaum and K. A. Gallivan：Decoupled Access DRAM Architecture, IWIA'97, Oct. 1997.

[6] A. V. Veridenbaum ： Increasing the Lookahead of Multilevel Branch Prediction, IWIA'98, Oct. 1998.

[7] J. Nelson, G. R. Gao, P. Merkey, T. Sterling, Z. Ruis and S. Ryan ： Performance

Prediction for the HTMT: A Programming Example, TFP'99, Feb. 1999.

[8] K. B. Theobald, G. R. Gao and T. L. Sterling : Superconducting Processors for HTMT: Issues and Challenges, FRONTIERS'99, Feb. 1999.

[9] P. Kogge, PIM Architecture to Support Petaflops Level Computation in the HTMT Machine, IWIA'99, Nov. 1999.

3.2.2　スーパースイッチをお手元に

天野英晴　委員

3.2.2.1　はじめに

近年のコンピュータは、ハイエンドのスーパコンピュータの性能の向上もさることながら、一般のユーザが広く使うことのできるPCの性能の向上が著しいことが特徴である。このため、一時代前のスーパコンピュータ並の性能を持つPCを個人が利用することが可能となり、マルチメディア処理の発達とコンピュータの利用形態の変化をもたらしている。

これが可能になったのは、CMOS LSI の高速化と高集積化が、ECL を中心としたスーパコンピュータの実装技術に迫る域に達しつつあることによる。ECLを用いたスーパコンピュータの実装技術は、性能だけを考えると、現在の高速PCのCPUが達成した性能を、ずっと前に実現したのだが、サイズ、消費電力、コストを考えると一般ユーザが広く用いることができるような代物ではなかった。すなわち、近年急激に進んだのは、スーパコンピューティングの大衆化であるといえる。

一方で、高速スイッチとリンクについても似たような傾向が現れている。従来から転送速度が1Gbpsに及ぶリンクとスイッチは、GaAsやECL等の技術を用いれば実現可能であった。しかし、これらのスイッチは特殊な基板と実装技術を必要とし、コスト的にも数千万円のレベルであり、主にATM交換器として特定の基地局に装備されるものであった。

ところが、最近は、これと同等の交換能力を持つスイッチが純粋なCMOSの技術で実現可能となり、PC と同様の個人レベルで用いることができるようになりつつある。しかも、この周辺の技術に関しては日本は、かなりの域に達しており、世界を主導することができる位置にいる。本稿では、高速スイッチとリンクについて、最近の状況をまとめると共に、筆者が共同開発しているRHiNETについて紹介する。

3.2.2.2　最近の高速スイッチの状況

3.2.2.2.1　高速スイッチのそれぞれの分野

高速スイッチ、高速リンクは、主として以下の分野で開発されている。

(1) 高速計算機のComponentとして: 大型計算機、大規模並列計算機などで、構成要素を接続するため用いられる。一つの Component で多数用いる場合は、コストはある程度制限される。以下のような例が挙げられる。短距離の転送なので、bit幅が大きめで、周波数帯

は200MHz程度。両方向同時転送が可能である。デッドロックを防ぐための仮想チャネルが複数装備され、場合によっては適応型ルーティングも備えている。高機能なスイッチが多いが、接続形態はメッシュ等簡単な形で済む。

SGI SPIDER: 20bit幅で200MHz、両方向同時、両エッヂ転送。8Gbps

Intel Cavallino: 16bit幅で200MHz、両方向同時転送。3.2Gbps

SR2200: ハイパクロスバ用スイッチ 10Gbps

(2) ATM スイッチ: 電話交換網、広域ネットワークに用いられる Asynchronous Transfer Mode のパケット（セル）を交換するスイッチ。スイッチ自体の交換能力よりも、バッファの位置（入力か出力か、両方に置くか)、バッファ容量、制御方法が問題になる。パケットの種類により廃棄可能かどうか等の転送の質を保証するのが最近のスイッチの特色である。

NTT ATM スイッチ: 単体のスイッチは 2.5Gbps４x４だが、MCM (Multi-Chip Module) の利用により10Gbpsの転送速度を実現。

Si/SiGe CPS スイッチ: 特殊トランジスタの利用により10Gbpsのリンクを４本交換可能。

(3) Cluster Computing, Local Computer Network用: PC/WSを接続するためのスイッチ、リンク。かつては、低速のものが多かったが、最近は高速化が進んでいる。柔軟に接続できる可能性がある。

Myrinet SW: クラスタ構成用のSAN(System Area Network)として広く用いられている。16x16のスイッチで、各リンクは、リンクにはリボン状のケーブルで9bit幅で80MHzで動作し、1.2Gbpsを実現する。

RHiNET-1/SW: 机上に普通に用いられている PC を接続して大規模計算を行なうシステム。光ケーブルあるいは Gigabit Ethernet 用ケーブルを用いる。９bit 幅で 133MHz, 1.4Gbpsのリンクを８入出力交換する。

RHiNET-2/SW: RHiNET-1/SWの高速版。詳細は後述。

これらのスイッチをまとめて図 1 に示す。右方向ほどそれぞれのリンクの転送容量が大きく、上方に行くほど交換するリンク数が多い。リンク当たりの転送容量は10Gbpsに達し、交換リンク数は４-16である。

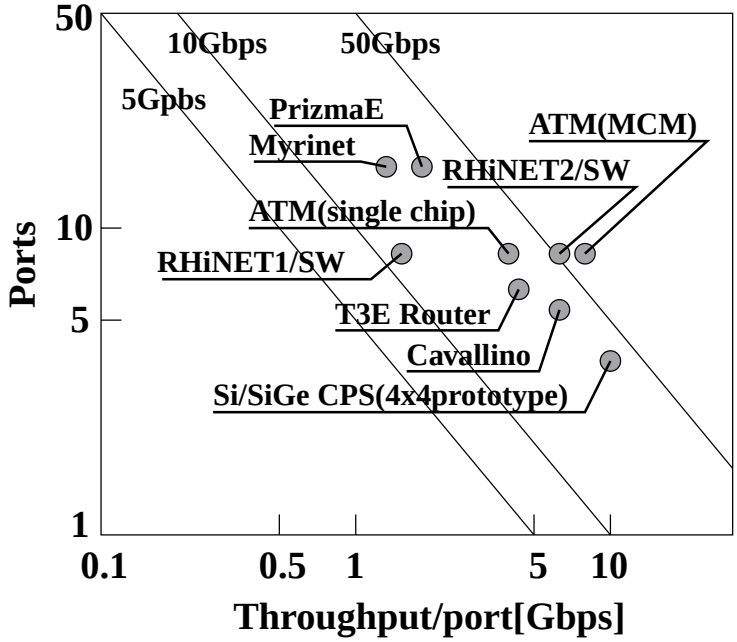


図１　スイッチチップの交換能力の比較

3.2.2.2.2　注目の高速化技術

3.2.2.2.1 で紹介したように、現状ではリンクの転送周波数はだいたい 800MHz-１GHz 程度まで可能になっているが、これは（1）リンクの転送速度の限界（2）スイッチの交換性能の限界による。

(1) リンク転送速度

リンク転送速度は、転送周波数が上がるにつれ、Component 内といえどもある程度の距離が必要な場合は、光転送技術が有望となる。米国の光転送技術開発プログラムである OMNET Program[1]は、以下の技術に重点を置いている。

(a) **Parallel Optics:** リボンケーブルを用いて並列転送を行なう（もちろんクロックも同じケーブルで送る)。高速転送時には bit 間の Skew が問題となる。短距離転送向き。OMNET Program では 1.25Gbps の転送速度で 12bit の転送を実現している。

(b) **Wide WDM:** Wavelength Division Multiplexing、すなわち一本のケーブル上に異なる複数の波長の光を載せる技術。2.5Gbps で 1280/1300/1320/1340nm の４波長を載せることが目標である。

(2) スイッチの交換性能

いかに高速スイッチを用いても、800MHz-１GHz の転送速度でやってくるパケットをその

まま交換できるはずはなく、Elastic Buffer で周波数揺らぎを抑え、Demultiplexer で bit 幅を大きくする代わりに周波数を落す。多くの場合、200MHz まで周波数を落してスイッチを行なう。しかし、この方法では１GMz 以上は難しい。

(a) **光スイッチの利用:** 光転送を行なうのならば電気的に変換しなければ、速度を落す必要はない。しかし、光スイッチの利用は、光の減衰の問題等で周波数が１Gbps に達するような場合はまだ困難である。

(b) **特殊半導体の利用:** Si/SiGe HBT (Heterogeneous Binary Transistor) Technology[2] は、トランジスタの片方の Si 領域に Ge を Doping する新しいトランジスタの方式である。高速だけではなく、低電力に特徴があり、８GHz で 4 入出力のスイッチで２W の消費電力を実現可能である。

3.2.2.2.3 Information Technology Frontiers for a New Millennium を読んで

米国の将来技術の発展の方向を示すドキュメントの中で、高速スイッチ、リンクに関する記述は、少なく、かつむしろ Conservative であるように思える。

(1) **High-end Hardware Component:** 超伝導クロスバスイッチ
2.5Gbps の転送速度のリンク 128 本を交換するクロスバを新デバイスの利用により実現する構想。

(2) **大規模ネットワーキング:** Parallel Optics の利用と、波長多重 (Wavelength Division Multiplexing) による超高速ネットワークの実現。

前者は、特に新デバイスを用いる必要はないように思える。後者は、技術的な方向性をきちんととらえているが、挑戦的なテーマではない。全体としてこのドキュメントの中で、物理層に近い高速転送技術と高速スイッチ技術は、さほど重視されていない、ということが読みとれると思う。

3.2.2.3 高速スイッチの大衆化

今までに紹介した高速転送、交換技術は、一部はスーパコンピュータ的なものであるが、一面、通常の CMOS 技術の発達により、超高速スイッチを、PC と同様の個人レベルで用いることができるようになりつつある。筆者が共同開発している RHiNET は、このような超高速スイッチを身近に用いることが鍵となるプロジェクトである。

3.2.2.3.1　LASN（Local Area System Network）と RHiNET

超高速スイッチが身近に使えるようになれば、クラスタ構成用に専用 PC を用いるのではなく、図２に示すように、フロアやビルに分散して配置されている PC を接続し、PC クラスタマシンと同等のパフォーマンスを獲得することが夢ではない。現在、事業部によっては、百台に及ぶ PC がフロア内に配置されているが、大体の場合は、これらの PC はアイドル状態にあって、その計算能力をフルに発揮していない。超高速スイッチによってこれらが強力なスーパコンピュータに生まれ変われば、様々な用途に用いることができるだろう。

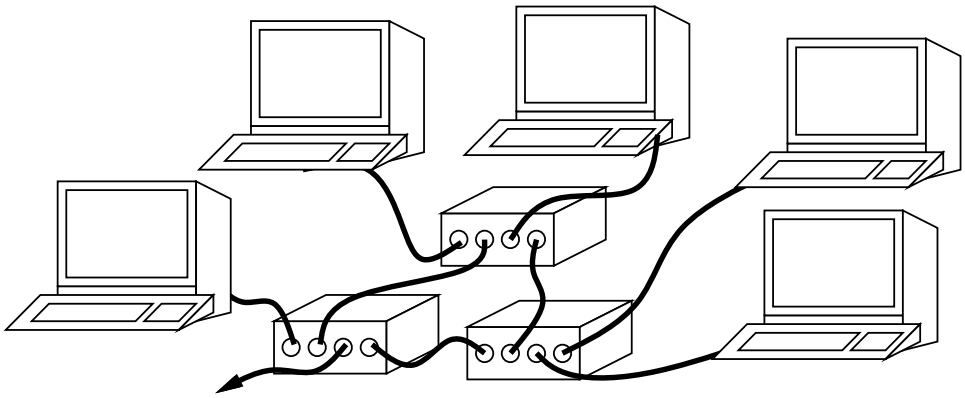


図２　LASN の概念図

しかし、このことを実現するためには、ネットワーク自体に問題がある。現在、WS/PC クラスタは Myrinet のような専用の SAN(System Area Network)を用いて実装されているが、これは伝搬特性上長い距離を引き回すようにはできていない。一方で、LAN（Local Area Network）は、最近、高速、大容量のものが出現しているが、パケットの転送法の問題や、一定のエラーレートを想定してソフトウェアレイヤを用いるため、どうしても転送レイテンシィが大きくなってしまう。

そこで、この両者の利点を兼ね備え、SAN と同様の信頼性のある低遅延通信を、ビル内やフロア内に分散して配置された PC 間で行なうネットワークが、LASN（Local Area System Network）である。我々は RWC のプロジェクトの一環として、LASN を実現するための RHiNET（RWCP High Performance Network）と呼ばれるシステムを開発している。LASN の要求事項を解決するために、RHiNET では現在次のような戦略をとっている。

1. レイテンシが大きい store and forward routing やマルチキャスト時のデッドロック回避が困難でチャネルの占有数が多い wormhole routing を利用せず、asynchronous

wormhole routing を用いる。

2. パケットの廃棄や望まない順序の入れかえを行わず、ハードウエアでのエラーレートを低くすることで、上位プロトコル層における通信品質補償に必要な通信コストを極力小さくする。また、パケットの破棄を許さず通信コストを抑えるため、デッドロックフリーを保障する。これには，構造化チャネル法の利用が効果的である。
3. ビル内やフロア内に分散して配置された計算機を接続するために、ループを含むある程度自由なトポロジを許し、100m 程度の延長距離と並列処理で要求される十分な bi-section bandwidth を確保する。

現在、LASN のプロトタイプとしてこれらの特徴を全て備えた RHiNET-1 を実装し、稼働試験を行っている。このためのスイッチである RHiNET-1/SW[3] は 0.35μm CMOS エンベッデッドアレイを用いて構成した 1 チップスイッチで，８×８のクロスバを内蔵し、リンク当り 130MHz で９bit データを交換することができる。

3.2.2.3.2 RHiNET-2/SW

RHiNET-1/SW は現在の PCI バスの性能を考えると十分な性能を持つ。しかし、これだけではすぐにも普及が予想される 64bit 化された高速 PCI バスには対応できないし、超高速スイッチとしてインパクトを与えるものではない。そこで、さらに性能を向上させた RHiNET-2 の実装を行っている。このスイッチ RHiNET-2/SW は、高速 CMOS 実装技術、基板実装技術により、最大８Gbps の転送能力を持つリンクを８入出力持つ世界最高速のスイッチの一つである[4]（ただし、技術的な困難は光インターコネクト日立研究所が解決したものであり、我々の手柄ではない)。

とはいえ、スイッチの設計自体にも LASN を実現するため、様々な工夫が施されている。まず、縮約構造化チャネル法により、トポロジフリーなネットワークを構築でき、その下でどの様なルーティングを行ってもデッドロックフリーが保障される。また、外部に簡単な制御用プロセッサを備えることで、複雑な形状の接続に対しても、ルーティングテーブルを書き換えて、自動的に対処することができる。100m に及ぶ線路長に対応するため、拡張 slack buffer 方式と称するハンドシェークを行っており、このハンドシェークパケットを利用して、活線挿抜、故障の自動検出等も可能である。

さらに、RHiNET-2/SW では新たに、以下の工夫を施している。

(a) 高速リンクへの対応のため、RHiNET-1/SW では、チップ外に配置した SRAM をチップ内に組み込み、性能向上を図っている。

(b)それぞれの flit に ECC を付けて自動訂正を行なうことで信頼性の向上を図っている。

(c) 高速リンクだけではなく、低速リンクに対応するため、３種類のビットレートに対応することができる。

RHiNET-2/SW の構成を図３に示す。RHiNET-2/SW では、0.25μm CMOS プロセスの利用により、800Mbps ９bit でパケットを入力して内部では 200MHz の動作速度で交換する。

パケットはもっとも早い場合、19clock で次のスイッチに送ることができる。ランダムロジック部のゲート数は合計で 421,000 ゲート、その他に９個の 144bit×４Kwords メモリと 16 個の 36bit × 16words メモリ、１個の 12bit × 256words メモリが配置されている。ダイの様子を図４に、スイッチのチップの写真を図５に示す。

RHiNET-2/SW は、光インターコネクトの部分に日立の Parallel Optical Fiber Interface MDR/MDS4212A を用いており、この部分が十数万円程度のコストを要するが、チップ自体は通常の CMOS 技術を利用しており、最近の高速プロセッサに比べて特にコストを要するわけではない。

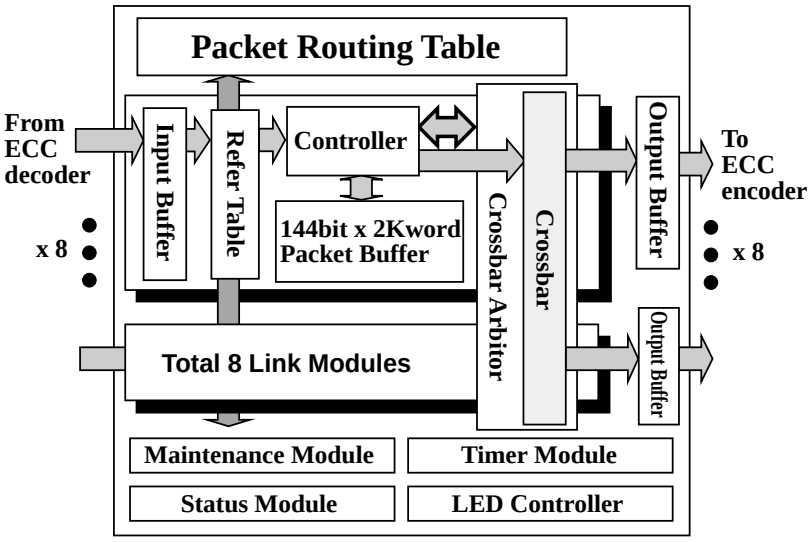


図３　RHiNET-2 / SW の構成

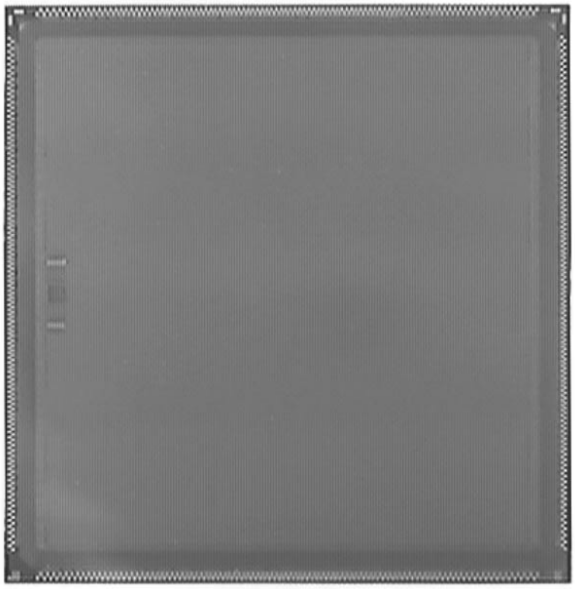


図 4　RHiNET-2 / SW のダイ

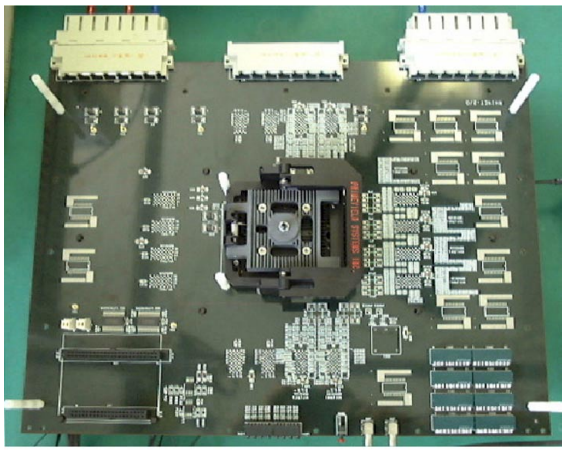


図 5　RHiNET-2 / SW の基板

3.2.2.4　おわりに

超高速スイッチの最近の状況をまとめ、さらに数十万程度のコストで身近に超高速スイッチを用いることを可能とする RHiNET-2 に関して紹介した。現状では、光転送技術およびスイッチに関しては、日本の技術は米国に比べても先行していると考えられる。また、米国は他ほどこの分野に力を入れていないこともあり、付け入る隙は大いにあると思う。

ただし、問題は、SANや高速LANの分野での標準化がほとんど米国主導で決まってしまっていることである。さらに、大衆化した超高速スイッチにより計算機の利用状況にインパクトを与えるためには、当然のことながらシステム、アプリケーションソフトウェアを考える必要があり、この点を強化することが今後の課題である。頼もしいことに、研究レベルでは日本はこの点でも決して遅れをとっているわけではないので、この分野は全体としてきわめて有望である。

参考文献

[1] L. Buckman, D. Dolfi, K. Giboney, B.Lemoff, J.Straznicky, K. Wu:
Multi-channel High Speed Optical Interconnects
Proceedings of Hot Interconnects VII,
pp. 145--162(1999).

[2] http://www.eng.auburn.edu/department/ee/amstc/extras/SiGe.html

[3] H.Nishi, K.Tasho, T.Kudoh, H.Amano:
"RHiNET-1/SW: One-chip switch ASIC for a local area system network,"
Proc. COOL Chips III, Apr. 2000 to appear

[4] S.Nishimura, T.Kudoh, H.Nishi, K.Harasawa, N.Matsudaira, S.Akutsu, K.Tasyo, H.Amano:
"Network switch using parallel optical interconnection for high performance parallel processing using PCs,"
Proc. 6th International Conference on Parallel Interconnect, pp.5-12,
Oct. 2000.

3.2.3　計算機間を接続して高性能並列処理を実現するネットワーク RHiNET

新情報処理開発機構　工藤知宏　講師

3.2.3.1　はじめに

技術研究組合新情報処理開発機構並列分散システムアーキテクチャつくば研究室では光インターコネクション日立研究室および慶應義塾大学理工学部天野研究室と共同で、PC や WS を接続して高性能並列処理を可能にするためのネットワーク RHiNET（RWCP High Performance Network）の開発を行なっている。RHiNET は、我々が提案する LASN と呼ぶ新しいネットワーククラスに属し、PC や WS の I/O バスに装着されるネットワークインタフェースと、高速なネットワークスイッチ、そしてそれらの間を接続する光インタコネクションから成り、図１に示すようにビル内やフロア内に配置された PC や WS を接続して高性能並列処理を行うことを目指している。

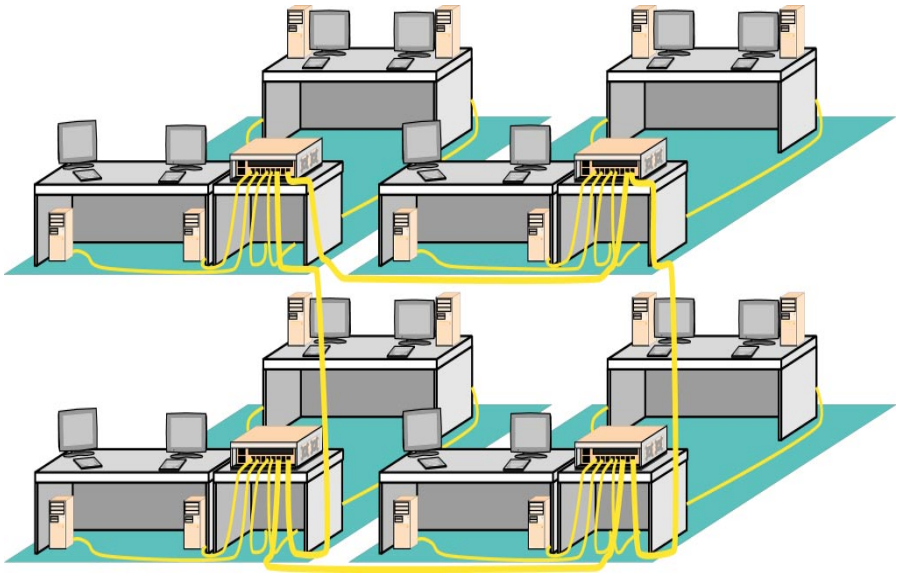


図１　RHiNET/LASN 概念図

3.2.3.2　LASN

近年、PC を多数使ったクラスタコンピューティングと呼ばれる並列処理が注目を集めている。従来のクラスタシステムは、大きく 100Base-T や Gigabit Ethernet などの LAN を用いて PC 間を結合したものと、Myrinet[1]などの SAN(System Area Network)を用いて PC 間

を結合したものにわかれる。

SAN は一般に wormhole や virtual-cut-through 等のルーティングを採用しており、低遅延と高い bi-section バンド幅を提供する一方で、配線の延長距離やトポロジにおける制約が厳しいため、集約型の構成をとらざるを得ない。

一方 LAN を用いたシステムでは、ビル内やフロア内に分散して配置されて利用されている計算機を接続して並列処理を行なうことも可能である。Gigabit Ethernet や 1.06Gbps の Fibre Channel などが実用・普及段階に入り、LAN 環境でも数ギガビットクラスの高速な通信が可能となっている。しかし、従来の LAN では物理層でのパケットの廃棄や順序の入れ換えを許しているため、信頼性のある通信を行う場合には、上位レイヤにおいて再送や順序の入れ換えを行う TCP/IP などのプロトコルを用いる必要がある。しかしこの様なプロトコルはホストプロセッサが行う処理のオーバヘッドが大きく、数ギガビットのバンド幅を持つネットワークを有効利用するだけの実効バンド幅を得ることが困難である。このように、LAN は高性能並列計算のためのネットワークとしては問題がある。

そこで、我々は SAN と LAN の利点を併せ持つ新しいネットワーククラス、LASN（Local Area System Network）を提唱している。LASN はフロアやビルに分散して配置されている PC を接続し、PC クラスタと同様の並列処理を行うシステムである。

- SAN と同様の信頼性のある低遅延通信を行う
- SAN と同様の bi-section バンド幅を持つ
- LAN と同様にビル内やフロア内に分散して配置された PC を相互に結合する

このような特徴を持つ LASN は次のような利点を持つ。

- 日常の業務で利用されている PC を、クラスタの構成要素として用いることができる。常時全ての PC がフル稼働しているとは考え難く、余剰計算能力を集めることで性能の高い並列マシンを安価で構築できる可能性がある。
- 異なる設置スペースにあるサーバやクラスタマシンの計算能力を集約することができる。
- ベクトルマシン、MPP、DSP など、異なる性能や得意分野を持つ計算機を集約することができる。
- 特別に設置スペースを設ける必要がなくなる。

RHiNET（RWCP High Performance Network）は LASN に属するネットワークであり、並列

処理をサポートする。RHiNET はネットワークとネットワークインタフェース（NI）から構成され、上記 LASN の性質に加えて NI が並列処理をサポートする様々な機能を持つ。

RHiNET の開発は、１Gbps クラスの光インタコネクションを用いる RHiNET-1 と、８Gbps クラスの光インタコネクションを用いる RHiNET-2 の２段階で行なっている。これまでに RHiNET-1 のハードウェアの開発を完了し、現在 RHiNET-2 の開発を行なっている。

RHiNET は LASN のインプリメンテーションであり、LASN の要求事項を解決するために、RHiNET では、次のような戦略をとっている。

(1) パケットの廃棄や望まない順序の入れ換えを行わず、ハードウェアでのエラーレートを低くすることで、上位プロトコル層における通信品質補償に必要な通信コストを極力小さくする。また、パケットの破棄を許さず、通信コストを抑えるため、デッドロックフリーを保障する。Myrinet などの SAN では、Wormhole routing や Virtual cut through routing を用いることにより、遅延（レイテンシ）が小さな通信を実現している。これらのルーティング法では、ノード間のパケット伝達路同士が環状の関係になるとパケットがネットワーク中に滞留したまま動かなくなってしまうデッドロックが発生する可能性がある。これを解決するもっとも簡単な方法は動かなくなったパケットを廃棄することであるが、この場合パケットの到着を確認し、必要に応じて再送する機構が必要となり、並列処理には適さない。そこで、これまでの System Area Network では、ネットワークのトポロジを限定して、伝達路間に環状の依存関係が生じないように伝達路を定めることによりデッドロックを回避してきた。しかし、フロア内やビル内の計算機群を接続するには、より自由なネットワークトポロジが求められる。そこで、RHiNET では、Asynchronous wormhole routing と構造化チャネル法を用い、多数の Virtual Channel (VC) を各入力ポートに用意することにより自由なトポロジにおいてデッドロックフリーを実現する。

(2) ビル内やフロア内に分散して配置された計算機を接続するために、ループを含むある程度自由なトポロジを許し、100m 程度の延長距離と、並列処理で要求される十分な bi-section bandwidth を確保する。

3.2.3.3 RHiNET-1[2][3]

RHiNET-1 は、１～1.3Gbps の光インタコネクションまたは Gigabit Ethernet で用いられる光リンクを用いる。このためのネットワークインタフェース RHiNET-1/NI と、スイッチ ASIC RHiNET-1/SW を開発した。

3.2.3.3.1　RHiNET-1/NI

RHiNET-1/NI（図２）は、PCI バスに装着されるネットワークインタフェースで、コントローラに PLD（Programmable Logic Device）を用いており、PLD の設定を変更することにより様々な通信プロトコルを実装、評価することができる。

アドレス変換テーブル（TLB）を内部に持ち、任意の大きさの領域のゼロコピー通信をサポートする。４つのメモリバンクに同時アクセスでき、アドレス変換などの操作とデータ転送を同時に行なうことが出来る。アプリケーション開発は message passing（MPI）ベース、共有メモリ（Open/MP）ベースの両方のスタイルで記述でき、専用の LINUX のライブラリおよびデバイスドライバを用いる。

このプリミティブ処理は NI で行われ、これらを組み合わせることにより、マルチタスク環境での zero-copy 通信を実現する。

図３に RHiNET-1/NI のブロック図を示す。RHiNET-1/NI は、パケット送受信部（Interconnection Interface）、PCI バス制御部（PCI Bus Interface）、プリミティブ処理部（Function Unit）およびプリミティブ処理用に用いるメモリから構成される。高速動作が要求されるパケット送受信部および PCI バス制御部は、高速な QuickLogic 社のアンチヒューズ型 FPGA を用い、プリミティブ処理部は、in-system programming が可能な FPGA（Altera 社 CPLD）を用いている。

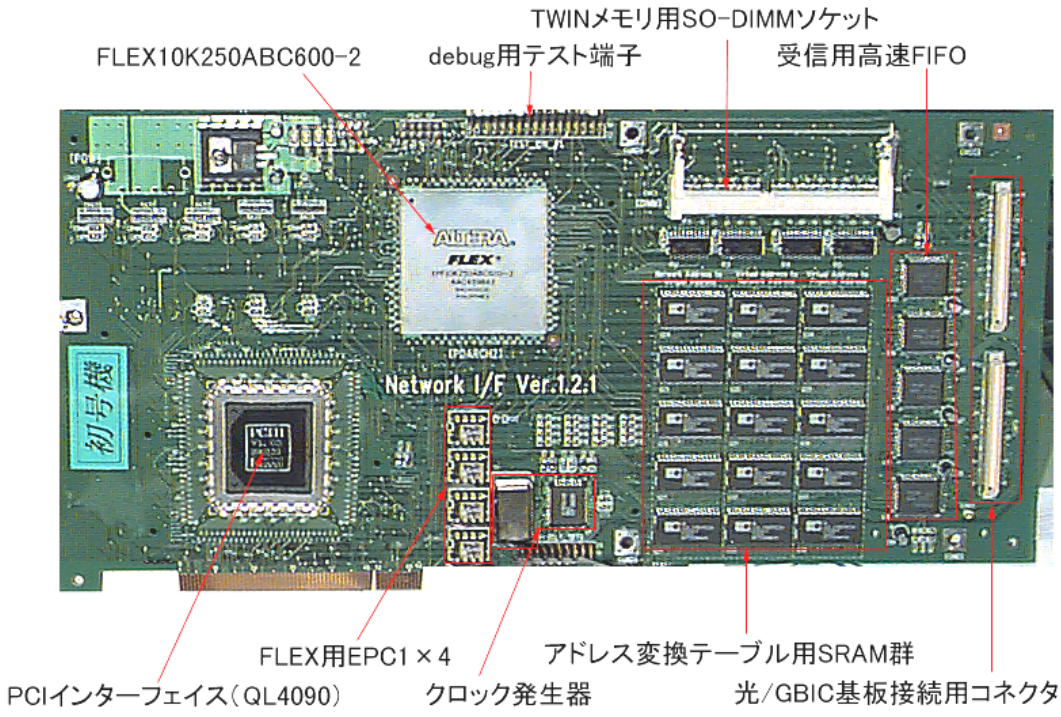


図２　RHiNET-2/NI

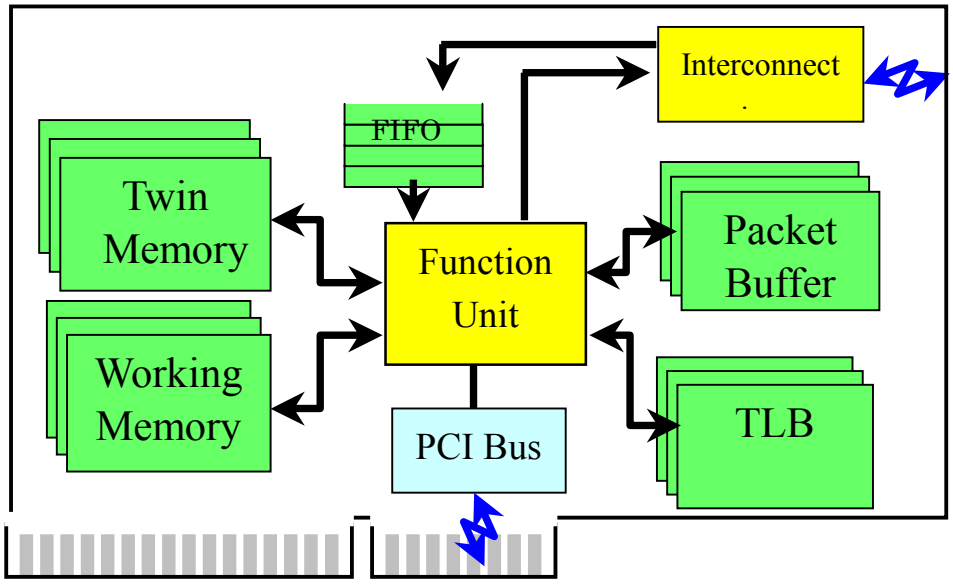


図３　RHiNET-1/NI　ブロック図

3.2.3.3.1.1　RHiNET-1/NI 上でのプリミティブ処理

ローカルノード（Initiator）からリモートノード（Remote）にパケットを転送する push プリミティブの実行（図４）を例に RHiNET-1/NI の動作を解説する。

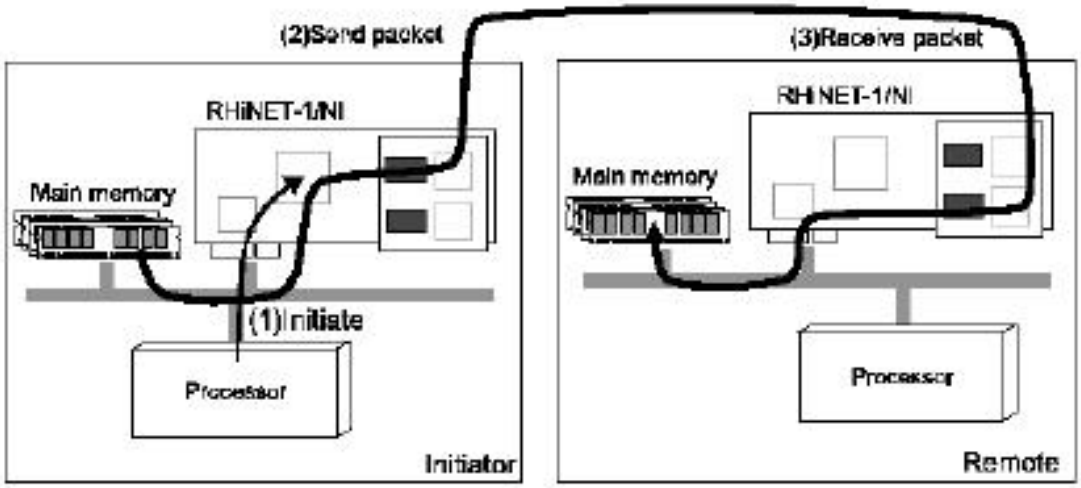


図４　push プリミティブの動作

(1) 起動

プリミティブ処理部は、プリミティブ起動に必要な情報を格納するプリミティブ起動用情報レジスタを持つ。ユーザプロセスが、ライブラリ経由でプリミティブ実行に必要な情報（送信データアドレス、宛先プロセス ID、サイズ等）をプリミティブ処理部のレジスタに書き込むことで、プリミティブが起動される。

プリミティブ処理部は、ボード上の TLB を参照して仮想アドレスを物理アドレスに変換した後、PCI バス制御部に転送すべきアドレスおよびサイズを与えることで、DMA 転送の要求を出す。PCI バス制御部とプリミティブ処理部は、データ転送用のローカルバスと、DMA 転送起動用の DMA バスで二重に接続されており、データ転送中も次回の DMA の起動をかけることができる。

(2) パケット送信

PCI バス制御部は、要求に従い主記憶からプリミティブ処理部にデータ転送を開始する。この間、プリミティブ処理部は、パケットヘッダを準備し、パケット送受信部の送信用 FIFO にヘッダの送信を開始する。DMA 転送された 32bit データは、即座に送信用 FIFO に入れられる。パケット送信部は、送信用 FIFO にデータが入り次第、32bit データを光インターコネクトで転送する 8bit 形式に変換し、転送を開始する。DMA による転送が終了した場合、プリミティブ処理部は、主記憶の所定の番地に対して、終了通知フラグを書き込むために、PCI バス制御部に対して再び DMA 要求を出す。この処理は単に転送終了を示すフラグを転送するだけのために DMA を起動するため、一見無駄が多いように見える。しかし、ホストプロセッサが NI 上のフラグをポーリングすると、その度に PCI バスを利用するため、データ転送の効率が落ちる。一方、メインメモリ上のフラグのポーリングではキャッシュにヒットするため、このような損失を避けることができる。

(3) パケット受信

リモートノードでは、まずパケット受信部が光インターコネクトから受け取ったデータを 32bit の形に変換し、受信 FIFO に入れる。受信 FIFO に少しでもデータが入れば、リモートノードのプリミティブ処理部が起動される。プリミティブ処理部は、ヘッダ中のリモートアドレスを、送信側同様に TLB を参照することにより、物理アドレスに変換して、PCI バス制御部に対して DMA 要求を発生する。PCI バス制御部は、DMA 要求を受け付けると、受信 FIFO 内のデータを主記憶に転送する。

一連の操作は、PCI バス制御部中で PCI バスの転送を効率化のためにデータをバッファする以外は、ほとんどの部分でデータをバッファリングすることなしに、パケット転送を行

うことができる。また、プリミティブ処理部の構造は、in-system programming 可能であり、様々な機能のプリミティブを状況に応じて実行することができる。さらに、RHiNET-1/NI は、SDRAM から成る TWIN メモリと呼ばれるメモリを装備している。このメモリを利用することにより、複数のプロセッサが同じリージョンに書き込むことが可能な multiple writer protocol をサポートすることができる。

3.2.3.3.2 RHiNET-1/SW

RHiNET-1/SW は 0.35μm CMOS エンベッデッドアレイを用いて構成した 1 チップスイッチで、8 x 8 のクロスバを内蔵している。RHiNET-1/SW および RHiNET-1/NI は、自由なトポロジと、長距離のパケット転送に対応するために、以下の技法を用いた。

(1) 縮約構造化チャネル法と VCC

asynchronous wormhole routing を用いると、転送パス間にループ依存関係があった場合、デッドロックが生じる可能性がある。パケットを廃棄することなくこのデッドロックを回避する手法として、構造化チャネル法が提案されている。これはネットワークの直径に等しい数の仮想チャネルを用意し、スイッチを経由するごとに異なる番号の仮想チャネルを用いる方法である。RHiNET でも、デッドロック回避のために構造化チャネル法を用い、パケットがスイッチを通過するごとに 1 大きい番号の仮想チャネルを使用する。

しかし、構造化チャネル法には、ネットワークの規模（最大直径）が仮想チャネルの数で制限されてしまうという問題がある。そこで、RHiNET-1/SW では外部に大容量の SRAM を持たせ、必要なチャネルのみをチップ内のバッファに割り当てる VCC (Virtual Channel Cache、仮想チャネルキャッシュ）を採用した。この方法により、少量のチップ内バッファで多数のチャネルを実現している。

また、分岐のないスイッチ（他のスイッチと接続されているポートが 2 以下のスイッチ）を経由しても、異なる番号の仮想チャネルを用いる必要がないことに着目して、縮約構造化チャネル法を提案して用いた。この方法では、全てのパケットは、他のスイッチへのリンクを 3 以上持つスイッチを通過した時にのみ、使用する仮想チャネルの番号を増やすことにより、必要な仮想チャネル数を減らすことができる。

(2) 拡張 slack buffer

RHiNET では、ビル内やフロア内に設置された計算機間を接続するため、最大 100m のリンク長（ホストやスイッチ間の線路長）を想定している。100m の光ファイバでは入力端から

出力端までの伝送には約 500ns を要する。従って、ハンドシェイクには、リンクの往復分のレイテンシと回路の動作時間を合わせた 1.5μ秒程度を要する。受信側はパケット送信停止の要求を送信側に送った後でも、この間に受信するデータを受け取らなくてはならない。このため、受信側はリンク中にあるパケットを受信するのに十分な容量がパケットバッファに残っているうちに、送信側にパケット送信をこれ以上行わないように要求する必要がある。これは window によるフロー制御の一種であり、Myrinet で用いられている手法で slack buffer と呼ばれる。RHiNET では構造化チャネル法を用いるので、パケットバッファは仮想チャネルごとに必要であり、送受信のハンドシェイク操作を仮想チャネルごとに行うように拡張する。この方式を拡張 slack buffer と呼ぶ。

これらの方法の採用により、RHiNET-1 は、リンク当たり 1.33 Gbps のバンド幅を持つ LASN を実現した。

3.2.3.4 RHiNET-2[4]

将来の PC や WS の性能向上とこれに伴うネットワークへの要求性能を考えると、さらに高速なネットワークが必要になることが予測される。

並列処理の応用分野には、行列計算など大量のメモリコピーを高速に行うことが必要になる問題が多い。システム全体の性能を上げるには、メモリが提供するバンド幅に近い速度でメモリ間コピーを行うことができるネットワークが求められる。PC のメモリのバンド幅は既存のものでも最大 8.5Gbps（133MHz、64bit）に達している。さらに、複数のメモリバンクや RAMBUS 方式のメモリを採用するなどした、より高いバンド幅を持つものが登場しつつある。

一方、ネットワークインタフェースが装着される PC の I/O 部にも、PCI や InfiniBand[5] などの 10Gbps クラスのバンド幅の規格が策定されつつある。近い将来、オフィスや研究室などに設置された通常の PC や WS が数ギガビットクラスのネットワークインタフェースを装備できるようになると考えられる。

このような状況では、近い将来にも、RHiNET-1 をはるかに上回る性能を持つネットワークが必要である。そこで、より高速な光インタコネクションを用いてさらに性能の高いネットワークを持つ RHiNET-2 の開発を行っている。

3.2.3.4.1 RHiNET-2/NI

ネットワークインタフェース RHiNET-2/NI は、コントローラに ASIC を用いる。ASIC 内部に大容量のメモリとプロトコル処理プロセッサを持ち、プロトコル処理の内容はプログラムすることが出来る。８Gbps クラスの光インタコネクションに直接接続可能である。64bit/66MHz の PCI インタフェースおよび外付け SDRAM のインタフェース、専用プロセッサを内蔵している。また、大容量の SRAM を内蔵することにより基本的な処理をチップ内で行なうことができる。現在設計を進めており、2000 年度中の稼働を目指している。

3.2.3.4.2 RHiNET-2/SW

RHiNET-2/SW では、RHiNET-1 から以下の点について改良を行なった。

(1) 光インタコネクトモジュール

RHiNET-2 では、日立製作所製のインタコネクトモジュール MDS4212A/MDR4212A を用いた。このモジュールは、12 組の発振波長 1310nm 端面発光レーザアレイとフォトダイオードを並列駆動することで、小型パッケージ（3.9cc）で 8.8Gbps の大容量光データ接続を実現している。１本のパラレルファイバケーブルで、クロックと 11bit の 800Mbps のデータを転送することができ、最大 100m の伝送が可能である。

(2) 大容量オンチップメモリの採用

RHiNET-2/SW では、転送レートの向上にともない slack buffer のために必要なメモリ量（VC あたりのメモリの大きさ）も大きくなる。一方、0.18μmルールの CMOS を用いることにより、オンチップメモリ上にポートあたり 512Kbits の仮想チャネルのためのバッファを持つ。これを用いて、４Kbytes の仮想チャネルをポート毎に 16 個装備した。これにより８Gbps、100m のリンクで waveform pipeline と拡張 slack buffer を実現した。

(3) 柔軟なルーティングテーブルの採用

RHiNET-2/SW では、パケットのルーティングはヘッダに格納されたルーティング ID によってルーティングテーブルを参照して決定される。ルーティングテーブルは、どの出力ポートにパケットを送り出すかを示す bitmap を返す。これは、8 つの出力ポートとおよびスイッチに付随するメンテナンスプロセッサそれぞれにパケットを出力するかどうかを示す９bit の bitmap である。従って、bitmap で複数の出力先が指定されていれば（すなわち 1 であるビットが複数あれば）マルチキャストとなる。

RHiNET-2/SW のルーティングテーブルにはパケットバッファと同じ大きさのメモリを用いており、65536 のエントリを持つ。ルーティングテーブルの設定は、メンテナンスプロセッサもしくはパケットにより行なうことができる。宛先ノード ID を用いてテーブルを参照すれば、最大 65536 のノードにまでパケットをルーティングできる。

システム全体のノード数が少ない時には、単一宛先とマルチキャスト、ブロードキャストのエントリを混在させたり、同一宛先への複数の経路を設定するなどの柔軟なルーティングが可能である。256 ノード以下のシステムでは、宛先と送り先のノード番号の組によりルーティングを行うことも可能になる。

(4) 仮想ネットワーク

一般には仮想チャネルは単一ネットワーク上に互いに影響されない複数の仮想ネットワークを実現することにより、パケットのルーティングに優先順位を設けたり QoS を保証するなどの用途に用いられることが多い。RHiNET では、仮想チャネルをデッドロック回避に用いており、RHiNET-2/SW は各入力ポートに 16 個の仮想チャネルを備えて直径の大きなネットワークにも対応している。しかし、ネットワークの直径が大きくない場合には、これらの仮想チャネルを仮想ネットワークの実現に用いることが考えられる。

仮想ネットワークを用いることにより RHiNET の通信プリミティブを効率的に処理することができる。RHiNET のプリミティブは複数の request と acknowledge パケットにより構成されるが、これらの request や acknowledge パケットそれぞれに異なる仮想ネットワークを割り当てれば、request パケットに全く阻害されることなく acknowledge パケットを受け取ることができる。従来ではプリミティブ間のデッドロック回避の為に、ネットワークインタフェースにイベントキューを設け、さらにプリミティブの発行数を制限する必要があったが、仮想ネットワークの利用により、この制限を無くすることができる。

(5) 伝送路のエラーレートとエラー訂正

RHiNET では、ネットワークの信頼性が高いことを利用して、再送などの動作を省くことにより高速転送を行う。RHiNET-2/SW で使用する光インタコネクションモジュールの単体での Bit Error Rate (BER) は 10^{-20} 以下である。これは、1000 台の PC で構成される程度の規模で RHiNET システムを構築し、全てのスイッチが最大転送バンド幅で 1000 時間動作し続けた場合を想定しても、システム全体で１つ以下のエラーしか発生しないことを意味する。また、RHiNET-2/SW では ASIC パッケージの I/O や基板上の配線を伝達される電気信号も非常に伝送レートが高いことを考慮して、エラー訂正符号 (ECC) を用いることによりさらに信頼性を向上させている。

(6) マルチビットレート

RHiNET-2/SW は、様々な価格や性能を持つインタコネクトの混在を許すため、異なる転送周波数での交換を可能とするマルチビットレート転送機能を持つ。これにより、RHiNET-1/NI と同様のアーキテクチャを持つ比較的低速なネットワークインタフェースにも接続可能である。

RHiNET-2/SW は、電気的特性を調べるための最初のプロトタイプによる実証試験を終え、論理動作を含めて動作する最終プロトタイプを用いた調整を行なっている。図５に RHiNET-2/SW チップのダイの写真を示す。このチップは日立製作所製の 16.5mm 角の 0.18μ ルールエンベッデッドアレイである。

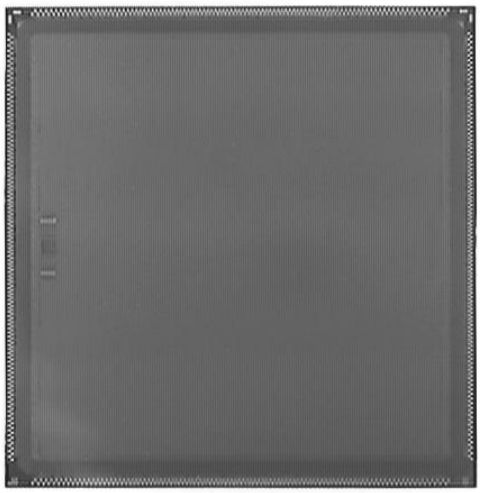


図５　RHiNET-2/SW チップ　ダイ

3.2.3.5　おわりに

現在、RHiNET-1 システムを用いてソフトウェアの開発とシステムの検証を行なうとともに、RHiNET-1 での経験に基づき RHiNET-2 の開発を行なっている。RHiNET-2 システムは当初 RHiNET-2/SW に RHiNET-1/NI を改良したネットワークインタフェースをつないで運用し、2000 年度中には ASIC コントローラを用いた RHiNET-2/NI を用いて稼働する予定である。

参考文献

[1] http://www.myri.com/

[2] 山本淳二他「高性能並列計算機用ネットワーク RHiNET-1 の実装と評価」情報処理学会

研究報告　2000-ARC 137-11, 2000 年 3 月

[3] 西宏章他「仮想チャネルキャッシュを持つネットワークルータの構成と性能」、並列処理シンポジウム JSPP'99 Vol 99 No.6. pp.71-78

[4] 西宏章他「「LASN 用 8Gbps/port 8x8 one-chip スイッチ RHiNET-2/SW」、並列処理シンポジウム JSPP 2000(採録)

[5] http://www.sysio.org/

3.2.4　プロセッサ技術およびストレージ技術に関する研究開発の状況　－課題と展望－

花輪　誠　委員

3.2.4.1　ペタフロップスマシンの実現は目前

ハイエンドコンピューティングマシンのハードウェアについて考える際、最近のマイクロプロセッサの動作周波数の飛躍的な向上について触れざるを得ない。当面のハイエンドマシンの目標スペックをペタフロップスとすると、100GFLOPS のプロセッサを 10,000 個、並列動作させれば達成可能である。ここで、100GFLOPS のプロセッサの実現可能性を考えてみると、20 個の演算器を 5GHz、または、32 個の演算器を３GHz で動作できるか？を考えれば良い。

先ず、「１プロセッサ内で演算器を 20～30 個程度、並列に動作できるか」であるが、最近の家庭用ゲーム機向けＬＳＩですら、16 個程度の演算器を並列動作させている訳であるから、ハイエンドコンピューティングの分野では、問題無く実現可能であろう。

次に、動作周波数であるが、今年の 2 月に開催された ISSCC2000（半導体回路の国際学会）において、GHz 級のマイクロプロセッサの発表が相次いで行われている。これを見ると、３～５GHz 動作も、間もなく実現可能であると考えられる。

ISSCC2000 は、正に、GHz プロセッサの学会と言っても過言ではなかった。テクニカルセッションでは、１GHz 級プロセッサで３件、760～780MHz 級で２件の発表があり、テクノロジーディレクションでは、4.5GHz で動作する演算器の試作結果の報告があった。また、ワークショップでは、マルチ GHz プロセッサの設計技術の議論が行われた。

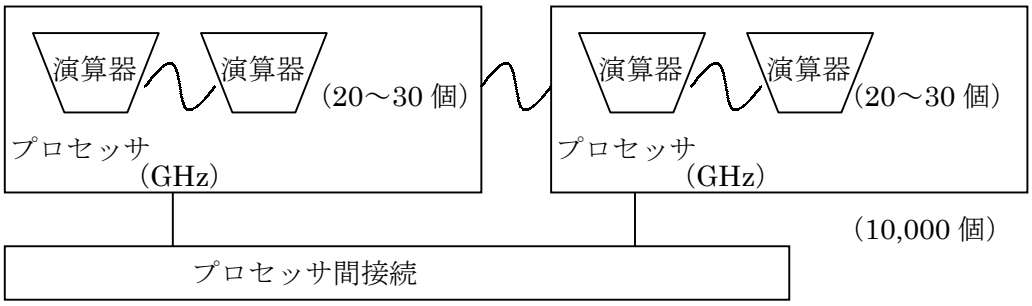


図１　ペタフロップスマシン

ISSCC2000 での GHz プロセッサは、全て、CMOS-LSI で実現されている。HTMT (Hybrid Technology Multithreaded architecture) で検討されているような超伝導技術を持ち出すまでもなく、一般のワークステーションやパソコンに使われているマイクロプロセッサの延長線上で実現できる訳である。

ペタフロップスマシンの実現は目前である。一方、ペタフロップスマシンを如何に産業に活用していくかと言う質問に対しては、明確な解答を持ち合わせていない。ペタフロップスマシンの利用技術を確立するには、先ず、実在のマシンを手に入れる必要がある。

国の予算で国内にペタフロップスマシンを 10 台程度、設置し、それを国民が活用して、新製品の開発、新事業モデルの構築などを推進し、GDP を１％拡大できないものだろうか。これが可能なら、1 台数百億円のペタフロップスマシン 10 台でもトータル数千億円であり、国の経済対策として見れば実現可能な計画と思える。

ハイエンドコンピューティングについて、何か提案させてもらえるとしたら、２～３年後の運用開始を目指して、国内に数台のペタフロップスマシンを配置し、経済活動に有効な利用技術を立ち上げることを提案したい。

ペタフロップスマシンの活用案を広く公募し、有効と判断できる案件に、CPU 利用時間を予算として割り当てることにしたい。使用料金は格安にして、結果として生まれた新製品や新事業モデルの利益から、一部を回収する方式、言わば、「出世払い」型が良い。

新規の事業モデルを構築する際、市場動向シミュレーションや、実績データからのデータマイニングなど、ペタフロップスマシンを活用した大規模計算を実施することにより、事業化リスクを大幅に低減できるかもしれない。

ハイエンドコンピューティングの利用技術とマシンの構築技術が車の両輪となって、互いに加速し合うポジティブフィードバックを形成しないと、最近の国内計算機業界の閉塞感を打破できない。

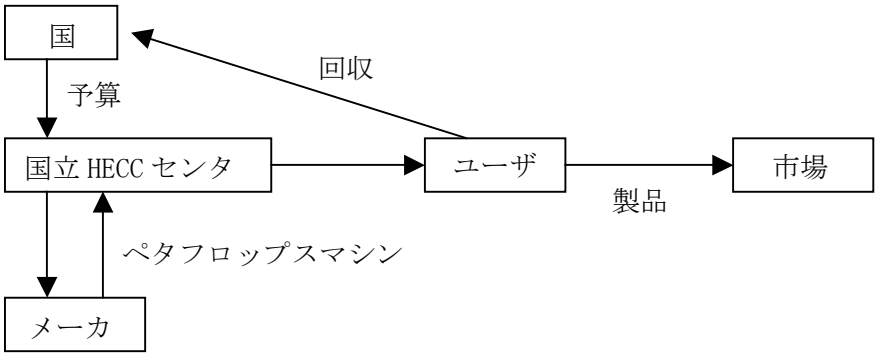


図２　ペタフロップマシン計画

3.2.4.2　プロセッサ技術に関する研究開発の状況 －課題と展望－

前項で、GHz 動作のプロセッサが ISSCC2000 で多数、発表されたと述べた。しかし、これらは、全て米国のプロセッサメーカの発表である。日本からの発表では、450MHz 程度である。

マイクロプロセッサの動作周波数を向上させるには、高性能 CMOS デバイス技術と高周波数対応設計技術の両方が必要である。高性能 CMOS デバイス技術は、最先端の微細加工半導体製造プロセスラインの構築にかかっている。特に、DRAM 等のメモリ LSI 製造ラインでなく、プロセッサ等のロジック LSI の製造ラインの構築が重要である。米国メーカは、早くからロジック LSI の生産に特化していた。

また、この製造ラインの構築に必要な投資ができるか否かも重要なファクタである。特に、最近の微細加工プロセスでは、1 ラインの構築に 1,000 億円規模の投資が必要である。世界市場を相手にして、大量の生産規模を確保することによって、次世代技術に対する投資が可能になる。

国内の半導体産業は、以前は、DRAM 主体の事業であったが、近年、台湾や韓国のメーカの追い上げが厳しく、システム LSI への展開が急務である。ロジック LSI 向け高性能 CMOS デバイス技術の開発は、先延ばしできない重要なコアコンピタンスである。

この分野に関しては、半導体のリソグラフィ技術など、国家レベルの将来への布石はある程度、打たれていると認識している。

設計技術に関しては、大きな格差を認めざるを得ない。先ず、人材面では、大学における LSI 設計技術教育の歴史の違い、更に、世界中からカリフォルニア州やテキサス州などに人材が集中してくる状況などが、格差の源であろう。

CAD / DA などツール類に関しても、新技術を次から次へと生み出すベンチャ企業、更に、ベンチャ企業の技術が有効と判断したら迅速に吸収合併する大手ベンダによって、格差が生じているものと考えられる。

一方、国内のマイクロプロセッサメーカの強みは、低消費電力化技術である。情報家電を始めとした民生品向けのマイクロプロセッサで、第一に必要とされるからである。言わば、低消費電力化技術の開発投資は、回収可能な投資であるが、高周波数化技術の開発は、国内メーカにとって、回収可能な開発投資ではなかったと言うことに他ならない。

前項で提案したペタフロップスマシン全国配備計画が実行されれば、国内メーカにとっても、高周波数化技術の開発投資が回収可能な投資になる。実際、米国では ASCI 計画によって、マイクロプロセッサの開発に弾みがついたと見ることもできる。

ISSCC2000 で発表された GHz プロセッサの設計は、DA ツールによる自動設計ではなく、人海戦術による力業的な設計である。これらの設計を自動化する研究に対して、米国では NSF がサポートしている。国内においても、抜本的な加速策が必要である。

以下、ISSCC2000 で発表された GHz プロセッサに関して、概要を紹介しておく。

表１　ISSCC2000 における GHz 級マイクロプロセッサの例

会社名	プロセッサ	動作周波数など	主な発表内容
Compaq	Alpha	1GHz@1.65V 65W 0.18μm 13.1x14.7mm^2 15.2M Trs. Al 7 層	SPECint/fp2000＝700/900(見積り) SPECint/fp95＝60/110(見積り)
Sun	UltraSPARC III	1GHz@1.6V <80W 0.15μm 244mm^2 23M Trs. Al 7 層	現世代に比べクロック周波数は 1.5 倍 ILP は 1.15 倍 パイプラインステージを９から 14 に増加
Intel	IA-32	1GHz 1.1～1.7V 0.1818μm 106mm^2 28M Trs. Al 配線	SPECint/fp2000： 733MHz 品に対して 1.17/1.22 倍
IBM	S390 G6	760MHz@1.9V 32W@637MHz 0.22μm 14.6 x 14.7mm^2 25M Trs. Cu ６層	1 プロセッサで 200 MIPS(S390) 12way SMP で 1600 MIPS(S390)
Motorola	PowerPC	780MHz@1.5V 0.18μm Cu ６層	クロック分配系を工夫
日立	64b RISC	450MHz@1.8V 0.20μm 28.3M Trs. Al ７層	MOS トランジスタの閾値をチューニング
IBM	64b 浮動小数点乗算器	3.3～4.5GHz 1.5V 0.18μm	非同期回路技術

3.2.4.3　ストレージ技術に関する研究開発の状況 －課題と展望－

ペタフロップスマシンでは、メインメモリとしてペタバイト級、磁気ディスクとしてエクサバイト級のストレージが必要であろう。また、バンド幅としては、100TB/s 級のアクセススループットが必要である。ペタフロップスマシンは、10,000 個のプロセッサを想定しているので、磁気ディスクも 1,000 から 10,000 個の並列構成が想定される。

ストレージに対する大容量並列アクセス技術の確立に向け、米国 DOE は、HPSS（High Performance Storage System）の開発を推進している。実際の活動母体は、サンディエゴスーパーコンピュータセンタであり、IBM 社の他、Sun 社、Storage Technology Corporation 社なども参画している。

最近、データウェアハウス、データマイニングなど非科学技術計算分野以外でも、大量のデータを扱うことが多くなっている。データそのものに価値がある時代である。今後は、生のデータを大量に蓄積していくことが要求されるであろう。

ストレージ装置の方も、ディスクドライブの大容量化とディスクアレイ装置の普及、更に、ストレージ エリア ネットワーク（SAN）の導入などによるアクセススループットの高度化が進展している。

従来、オペレーティングシステムの Ｉ／０ 処理は、シーケンシャルな実行が基本であったので、これを並列化し、データの整合性を維持していくためには、新たな仕組みが必要になってくるものと思われる。日本が得意とする光ファイバ接続との連携により、これらの課題を解決できるのではないかと考えられる。

HPSS の開発においても、実際にハードウェア システムを構築して、計算センタの運用を実際に行いながら、新規機能の開発がフェーズ分けされながら、推進されている。第 1 項で提案したように、先ず、ペタフロップスマシンによるハイエンドコンピューティングサイトを構築し、その上で、オペレーティングシステムを始めとしたソフトウェアの開発を推進していく必要がある。

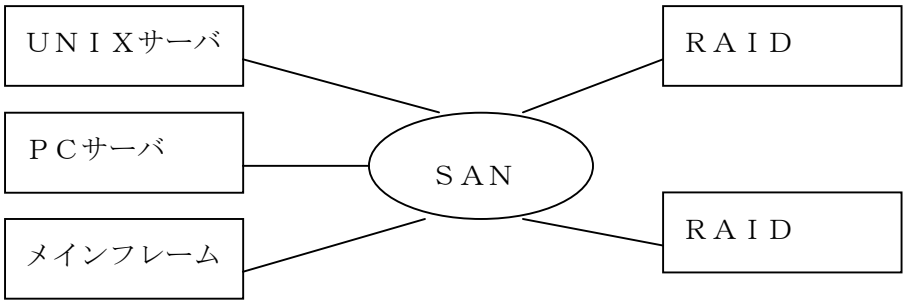


図３　ストレージ エリア ネットワーク（ＳＡＮ）

3.2.4.4　まとめ

ハイエンドコンピューティングマシンの当面の目標であるペタフロップスマシンの実現性について検討した。最近のマイクロプロセッサ高性能化の勢いを見ると、5GHz 程度で動作する CMOS プロセッサも近い将来に実現可能であると予想される。その結果、１プロセッサ当り 100GFLOPS 程度は、実現可能であり、これを 10,000 個、並列動作させれば、ペタフロップスマシンは構成可能であろう。

ハイエンドコンピューティング分野における研究開発を加速させるために、マシンを実際に構築し、利用技術を早急に立ち上げる計画を提案した。最近のインターネット データセンタの普及と共に、ハイエンドコンピューティングの産業利用が更に加速されるものと予想できる。利用技術と構築技術を車の両輪に、ポジティブフィードバックにより研究開発を加速させたい。

プロセッサ技術への開発投資は、最近の国内メーカにとっては、苦しい状況が続いている。開発成果の市場投入、開発投資の回収と言ったループが構築できていない。上記のペタフロップスマシン計画を利用して、弾みを付けられないものかと考えている。

一方、我が国のマイクロプロセッサ技術力の強みに、低消費電力化技術がある。これを伸ばして行き、民生機器分野での競争力を向上すると共に、今後ハイエンドコンピューティング分野においても、低消費電力化技術は重要な技術の一つになるであろうから、上手に育成していく必要がある。

ストレージ技術に関しては、磁気ディスクの驚異的な容量拡大と言うシーズと共に、データウェアハウス、データマイニング技術の一般化により大容量化のニーズも拡大している。課題は、大量のデータを如何に高速にオペレーションするかであり、今後、この分野の技術開発が重要になってくるであろう。

今回は、ネットワークについて余り配慮できなかったが、本来は、グローバルコンピューティング等の検討が重要である。今後の検討課題としたい。

参項文献

[1] ISSCC2000 Digest of Technical Papers, IEEE SSCS, Feb. 2000

[2] http://www.sdsc.edu/hpss

[3] http://www.llnl.gov/asci

3.3　並列分散システム

3.3.1　高性能計算連続体　　― NPACI プログラムの概要 ―

近山　隆　委員

3.3.1.1　はじめに

National Partnership for Advanced Computational Infrastructure (NPACI) は米国 National Science Foundation (NSF) のプロジェクトである。このプロジェクトは、地理的に離れた計算資源を有機的に結合し、いつ・どこからでも同じやり方で高性能計算ができる新たな研究インフラストラクチャを築くことにある。

NPACI の実現を目指す新たな目標のひとつに、伝統的な数値計算のためのインフラストラクチャと、比較的最近発展してきた大量のデータを中心とした処理のインフラストラクチャの統合がある。今後の高性能計算では、数十から数百テラバイト級のデータを扱うことは必須になる。NPACI の中心的なサイトである San Diego Supercomputer Center では 1998 年時点で 50TB のデータを高性能記憶システム上で運用しており、2000 年にはペタバイト級のデータを保持できる見込みである。このようなローカルな記憶容量の拡大のみならず、遠隔地にあるデータへの円滑なアクセスも重要であり、NPACI の目標のひとつとなっている。

計算技術の発展により、理論と実験に加えて、計算機によるシミュレーションが科学の手段として重要になってきている。NPACI ではこの理論・実験・シミュレーションを円滑に結合する情報技術体系（図 1）の確立を目指している。

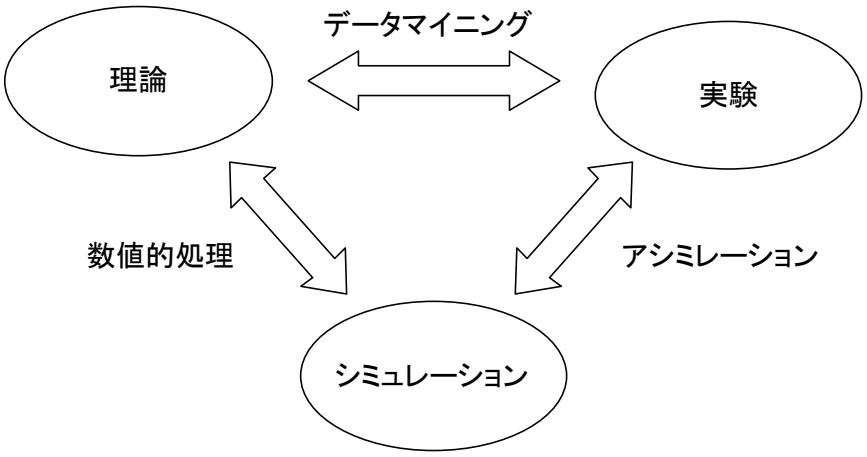


図 1　理論・実験・シミュレーションを統合する情報技術

NPACI には数多くのサブプロジェクトがあるが、本稿ではアーキテクチャとメタシステムについて述べる。

3.3.1.2　アーキテクチャ

過去十年程度の間に高性能計算機システムの性能は約千倍向上した。この性能向上がどのようにしてもたらされてきたかを概観し、この性能向上速度を今後十年継続するためには計算機アーキテクチャ上どのような変更が必要になるか、それがシステム全体にどのような影響をもたらすかについて予測する。

・大規模並列計算機

高性能計算機の実現に並列処理を用いることはすでに常識となっているが、将来ともこの傾向は強まっていくであろうと予測できる。しかし、今後の半導体技術の更なる進展によってプロセッサの性能が向上し、また結合できるプロセッサの数が増大していくにつれて、プロセッサの性能を十分発揮させるだけのデータを供給する技術が性能向上の鍵となることは間違いない。

並列計算機の黎明期には、複数のプロセッサを専用のハードウェアで、特殊な形態で結合するのが一般的であった。近年、並列計算が一般化するに伴い、複数のプロセッサ間結合にも汎用品（Myrinet など）が用いられることが多くなってきている。また、独立した計算機システムを接続するためのネットワーク機器も高性能化しており、汎用の LAN 機器での結合も費用対性能比の面で有利になってきている。

・階層メモリ技術

プロセッサの高速化に比して、メモリはさほど高速化していない。プロセッサの速度向上が年率 60%にも上るのに対し、DRAM の速度向上は 7%に過ぎない。大規模な並列計算機システムでは、一次キャッシュ中のデータへのアクセスが１～2 サイクルで済むのに対し、遠く離れたメモリにあるデータのアクセスには数十万サイクルもかかることになる。このギャップを埋める技術が、高性能アーキテクチャ実現の鍵となると予想できる。

そのためには、現在用いられているよりもさらに深い階層構造を持つメモリシステムが必要になる。こうした階層メモリもアクセスの局所性が低くては十分な性能を発揮できない。そこで、キャッシュなどの階層をうまく利用するアルゴリズムや、メモリ階層を意識したコンパイル技術の開発が重要になる。

また、遠隔メモリのアクセス遅延が性能低下をもたらさないためには、アクセス遅延の間に他の有用な計算をする必要があり、このためには多数の実行スレッドを低オーバヘッドで切り替えられるアーキテクチャが求められる。

別の方向として､プロセッサーメモリ間のスループットを改善するためにはProcessor In Memory と呼ばれる、プロセッサと DRAM を同じチップ上に実装する方式も注目されている。DRAM のアクセス速度のボトルネックはプロセッサとのインタフェースを整えることにあり、同じチップ上に実装することによってこの点を大きく改善できる可能性がある。

・ベクトル計算機

大規模並列計算機だけではなく、ベクトル計算機も進歩を続けている。しかし、近年ベクトル機のクロック向上は頭打ちの状況で、もはや最大級の問題を扱うシステムではないといえる。一方、ベクトル機においても並列度を上げて性能を向上させる手法が一般化してきている。また、コスト的に有利な CMOS 技術への移行の動きも顕著である。

前述のとおりベクトル機はもはやハイエンドとはいえなくなってきているが、ソフトウェア資産は潤沢で、豊富な応用ソフトウェアの供給や安定したシステムソフトウェアを有する強みから、今後しばらくは重要な地位を占めつづけるであろう。

・性能向上の見通し

過去十年で高性能計算機システムはピーク性能において約千倍の向上を果たした。さらに千倍の向上を果たして、ペタフロップス級の性能を実現するには、何が必要であろうか。CMOS 技術の進展は 10～15 年程度で限界に達する。SIA は、今後十年間のクロック速度向上は 3 倍以内であろうと予測している。一方、ジョセフソン素子など、原理から異なる新素子技術に基づくシステムが商用になるには 10～15 年が必要であろう。

今後の性能向上のための要素技術としては、CMOS に代わる素子技術の利用によるクロック速度向上、プロセッサアーキテクチャの改良によるクロックあたりの性能向上、そしてシステムアーキテクチャの改良による向上が考えられるが、今後十年のこれらの要素による性能向上は、十倍程度にとどまるであろう。となると、ペタフロップス級の性能を実現するには、現在の百倍程度高い並列性を持つシステムが必要になる。

・入出力

仮にペタフロップス級の計算機システムが実現できたとすると、従来と同様の応用ソフトウェアを動かすには 34 テラバイト程度のメモリが必要となる。これと見合う量のオンラインディスクとしては 1 ペタバイト程度が適切である。2007 年のディスク技術では、これ

は 45GB ドライブ 22,000 台と予測され、その合計バンド幅 352GB/sec と予測される。

34TB のチェックポイントダンプを 5 分で取れるためには 113GB/sec のバンド幅が必要である。これは 22,000 台のディスクをピークの 1/3 の性能で並列動作させることになり、これは容易なことではない。

一方、1PB のディスクのバックアップには exabyte 級のアーカイブ装置が必要になる。これには、今日のテープとロボット技術では 500,000 立方フィート（14,000m³）の容積が必要になる。こうした点の解決も重要な課題である。

・まとめ

高性能計算機システムの性能を従来のペースでは向上させていくことは非常に難しく、性能向上には複雑度の増加が伴うことを避けられない。このため、今後の高性能システム実現にはソフトウェア面の非常な努力が必要になる。オペレーティングシステムは、非常に多数のスレッドと複雑な階層メモリを管理しなければならない。コンパイラや実行時システムは、プロセッサ内の高い並列性や、多層にわたる階層メモリを有効に扱う必要がある。たとえこうした基本ソフトウェアについての改善を施せたとしても、高性能を実現するために応用ソフトウェア開発にかかる負担は現在よりもさらに大きなものとならざるを得ない。

現在の技術の延長では 10 年程度で性能向上の限界が来るであろう。さらにその先の高性能化を求めるには、従来と異なる新たな技術体系の開発が急務である。

3.3.1.3　メタシステム

メタシステムは、ネットワーク上に広域分散した数多くのファイル、データベース、計算機、入出力機器などを統合し、あたかもひとつの巨大な計算機環境を利用しているような使い勝手を利用者に提供するシステムである。メタシステムを用いれば、利用者はデータがどこにあるのか、それをどの計算機で処理をするのか、ついてはどのデータとどのプログラムをどこにコピーする必要があるのか、といった物理的な層を意識せずに処理を行なうことができる。

ここでは NPACI のメタシステムを構成する要素である、共有永続オブジェクト空間、透過的遠隔実行、広域キュー、広域並列処理、そしてメタ応用について概説する。

・共有永続オブジェクト空間

共有永続オブジェクト空間は、メタシステムのもっとも基本的な機能である位置透過性を実現するものである。すなわち、さまざまな資源がどこにあるのかを意識せずに、利用者が使えるようにする機能である。ファイルアクセスについては従来からNFSやAndrewのような分散ファイルシステムが使われてきたが、共有オブジェクト空間においては、ファイルはもちろんのこと、実行中のタスクなどに至るまで、適宜名前を与え、権限をもつ利用者間で共有するようになる。

こうしたオブジェクトは、たとえばファイルひとつをとっても、単一のインタフェースを持つとは限らない。同じ物理的構成を持つファイルを、異なるインタフェースを持つクラスでオブジェクト指向に抽象化できる。たとえば、特定の応用に特化したファイルなら、通常のread, write, seekなどの他に、2次元の配列データを表すものとして扱う、行単位、列単位、あるいは部分配列に対する入出力などのインタフェースを持つことも考えられる。キャッシュやプリフェッチの仕方を利用者が指定すること、入出力を応用プログラムの実行と非同期的に行なうこと、複数の入出力要求をキューイングすることなども考えられる。逆に、異なるフォーマットで格納したデータに対して、同じインタフェースを提供することも考えられる。さらに、受動的にデータを提供するだけでなく、自律的に動作するようなオブジェクトも考えられる。

・透過的遠隔実行

透過的な遠隔実行はやや高度なサービスになる。現在普及しているシステムでは、利用者が自分のプログラムをどこで実行するか（自分の目の前にあるワークステーションか、近くの高性能計算機か、それとも遠隔地の計算センターか）を決めなくてはならない。この際には、どこで実行すれば転送時間まで含めてもっとも早く結果が得られるか、遠隔システムを利用するのに技術的問題はないか、アカウントはあるのかなど、さまざまな点を考慮に入れた判断が必要になる。

メタシステムにおいては、利用者は単にどのプログラムを実行するかを指定しさえすればよい。システムがその利用者が利用可能なホストの中から適切なものを自動的に選び、必要なプログラムの転送も自動的に行なって、プログラムを実行してくれる。データは上述の共有永続オブジェクト空間から入力し、結果も同様に共有永続オブジェクト空間に格納する。

・広域キュー

これは現在でもよく用いられているバッチシステムを、分散した複数のシステムに広げ

て運用するものである。メタシステムにおけるバッチシステムは、単一の機種に対するものではなく、多様な機種の計算機が交じり合った環境でも利用できるようにする。

・広域並列処理

メタシステムは与えられた問題をうまく分割可能な場合、複数のホストに仕事を割り振って実行することにより、非常に大規模な並列処理を実現することができる。もちろんこのような分割がうまくできない問題も少なくないが、独立性の高い部分計算からなるデータ並列問題については、比較的容易にこのような並列化が可能である。

・メタ応用

メタ応用とは、従来個別の要素問題として扱われてきたような複数の計算処理からなる、全体としてひとつの応用である。それぞれの要素問題の特性は異なり、たとえば共有メモリ並列処理に向いた問題、ベクトル計算機に向いた問題、分散した超並列処理に向いた問題が含まれていることが考えられる。これらを単一のホストで処理するのは得策ではない。システムはこうした問題の性格を考慮に入れて実行方針を決めねばならない。

これは容易に解決できる問題ではないが、NPACI が目標としている機能のひとつである。

・まとめ

メタシステムの技術は実験的な状況から、実用に耐える技術へと成熟しつつある。メタシステムの持つべき様々な機能が実用に供されるようになれば、科学技術における計算処理の生産性は飛躍的に向上するであろう。

参考文献

Sidney Karin and Susan Graham, Eds., “The High-Performance Computing Continuum”, Communications of the ACM, Vol. 41, No. 11, November 1998.

3.3.2　ハイエンド計算におけるグローバルコンピューティング

関口智嗣　委員

3.3.2.1　はじめに

ハイエンド計算技術はハイパフォーマンスコンピューティング（HPC）を実現する上で今後重要な課題である。応用としての計算科学技術は純粋な真理追究の科学のみならず、生産・加工・設計・製造等のいずれにおいても実際の産業活動に直結している（図１）。ハイエンド計算技術はこの計算科学を支持する情報基盤システムである。すなわち、ハイエンド計算を通じて高性能計算機資源の利用が社会生活においても非常に重要な意味を持っていることを示している。このため、ハイエンド計算技術は大企業だけではなくベンチャーや中小企業も技術革新のためにはこの技術導入が必至である。ところが、スーパーコンピュータのような高性能計算機はパソコンと較べて価格性能比が悪いので、企業による個別導入が困難になっている。スーパーコンピュータはスーパーコンピュータらしく利用される局面、例えば数ギガバイト級の大規模メモリを必要とする計算とか絶対性能が必須で時間的制約がある計算など他では代替が不可能で、かつミッションクリティカルな場合には不可欠な計算能力を提供する。一方で、ハイエンド計算技術はハイエンド計算のみならず、ダウンスケールした波及効果として、より価格性能比の優れた計算技術が望まれている。

本稿では、ハイエンド計算技術に求められる今後の技術としてグローバルコンピューティング技術を紹介し、その必要性と将来像について述べる。

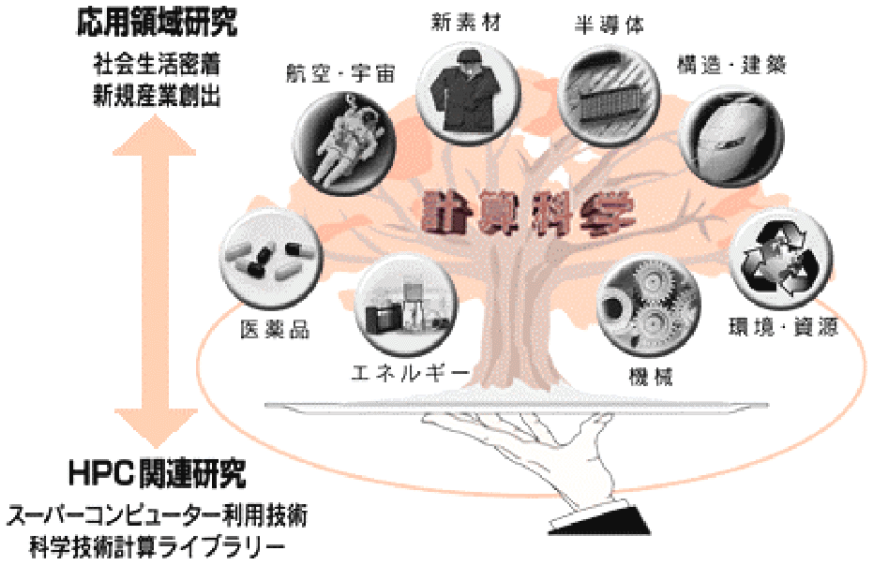


図１計算科学と HPC

3.3.2.2　グローバルコンピューティング技術概観

グローバルコンピューティング（Global Computing）関連技術は様々な呼称を持つ。例えば、コンピュータの物理的配置や接続を隠蔽するメタコンピューティング（Meta Computing）と呼ばれたり、グリッド（GRID, iGrid）と呼ぶ場合には仮想的な計算資源の供給インフラを構築するという意味合いが強い。さらに、ネットワーク可用サーバ(Network enabled server）という計算サービスやダトゥール（Datorr: Desktop access to remote resources）やポータル（Computing Portal）と呼ばれることもある。いずれも、定義が明確なものではないが、広域に敷設された高速ネットワークを利用したコンピュータの集合体をインフラとして構築または利用する技術であることは疑いない。

グローバル コンピューティングは様々な側面を有している。高速計算はひとつの明白な応用であるが、もちろん、それだけではない。いくつかの具体的応用システムイメージについて示す。

3.3.2.2.1　スーパーコンピュータ・アンサンブル

当初の広域分散コンピューティングのイメージである。世界中に分散されているスーパーコンピュータをネットワークで接続し、仮想的に超大規模スーパーコンピュータを作り出し、これまでに絶対解くことができなかったような問題への挑戦である。広域分散コンピューティングにより、これまでのコンピューティング環境をはるかに凌ぐ計算パワー（ピーク性能）を提供することが目標である。

Supercomputing'99 においてこうした大規模計算の実験が行われた。プロジェクトチームはドイツの HLRS（High Performance Computing Center Stuttgart）を中心に、イギリスの CSAR/MCC（Computer Services for Academic Research/Manchester）、アメリカの PSC（Pittsburgh Supercomputer Center）と日本の工業技術院先端情報計算センター（TACC）によるメンバー構成であった。デモンストレーションでは HLRS、CSAR/MCC、PSC の T3E と TACC の「SR8000」（64 ノード）とを高速ネットワークで接続し、接続されたコンピュータの総合最高性能が 2.2 TFLOPS コンピュータとして、分子動力学、計算流体力学のシミュレーションを実行した（図２）。具体的な計算例としてはまだまだ小規模であったが、こうしたスーパーコンピュータをネットワーク接続して利用する可能性を示したものとして、また国際的な協力下において実現したことは興味深い。

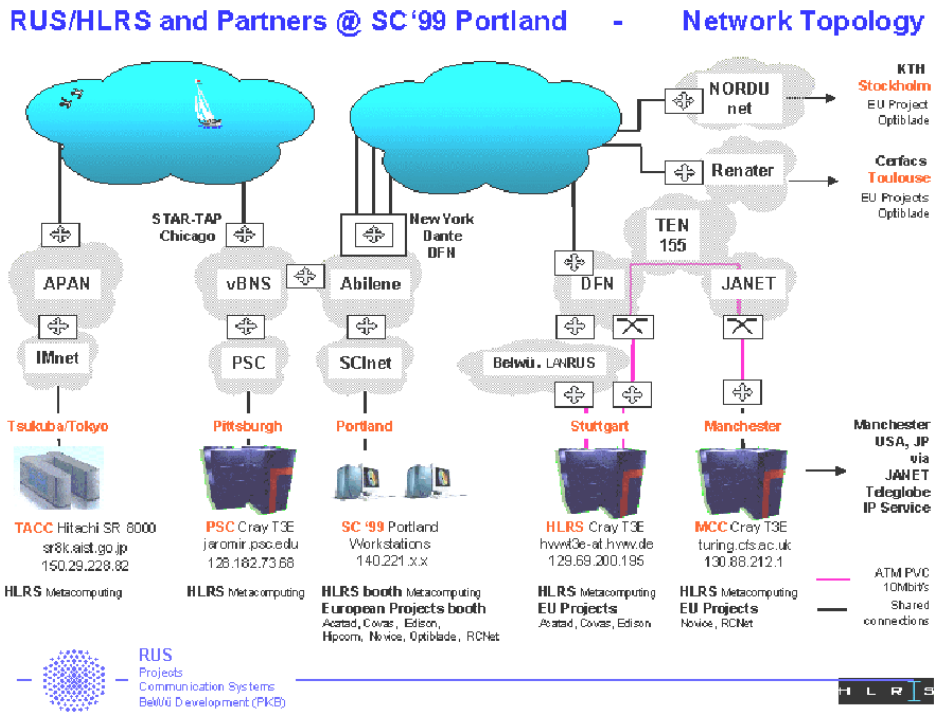


図２　ネットワークの構造

3.3.2.2.2　広域分散コンピュータオーケストラ

一方、グローバルコンピューティングは世界中の遊休計算機を有効に活用して、無尽蔵の要求に対して高いスループットを提供することを目指している。同じくSupercomputing'99で行われたデモでは世界10カ国に設置されたワークステーションやPC UNIX機の合計150台上に計算処理を分割して分散配置を行った。今回のこの挑戦の面白いところは世界中の計算機を集めてくることにほとんどお金がかかっていないことである。並列性の高いモンテカルロ計算であったが、世界中の遊休計算機を活用すれば大規模計算ができる可能性を示した。システムの提供方法などは個人による依頼ベースであり、システム的な対応ではなく、技術的な問題は山積みである。例えばアカウントを作成する必要があること、事情により供出が困難になってもそれをマスタ側に伝える方法がないこと、別のプログラムを実行させようと思ったときには全部の計算機でそのプログラムを用意する必要があること、などまだまだ実用的にはほど遠いものがある。しかし、インターネットの世界でハウジングビジネスが成立しているのであるから、計算そのものをサービスとして受けるという科学技術計算におけるASP（Application Service Provider）が成立することが期待される。

3.3.2.3　グローバルコンピューティング技術動向

グローバルコンピューティング（広域分散計算）の近年の発展はコンピューティング技術の高性能化とネットワーク技術の高速化という技術的発展ばかりではなく、これらが社会的にも普及し、従来のコンピュータが占有的資産であり排他的に用いることから共有的資産で Give & Take により活用される資産であるという認識の中で現在注目を集めている技術である。もちろん、広い意味でインターネット技術の一応用に過ぎないのではないか、なんら新しいものを産み出していないのではないかという批判があるのも確かである。しかし、インターネットばかりではなくや地域ネットワーク、職場や大学のキャンパスエリアネットワーク、ローカルエリアネットワークなどがこうしたグローバル化を実現するための基盤を提供することで初めて実現が間近になった技術である。これらの基盤を用いることで、電子商取引、電子政府他のインターネットビジネスチャンスは大きく広がってきた。ここではあえてグローバルコンピューティングが「コンピューティング」とあえて付けているところに着目したい。すなわち、コンピュータ生来でかつ人間の苦手な計算処理技術を中心として広い意味のインターネット技術への応用を目指している。安易な言い方をすれば、これまでの Web がデータアクセスを中心に考えていたのに対して、計算処理における Web を目指していると考えてもよい。このために必要な要素技術と実際の応用技術、これらの統合化技術が求められている。

代表的なグローバルコンピューティングシステムを下記に示す。また、下記を含めて代表的なプロジェクトの名称等を図３に示す。

・Akenti　　http://www-itg.lbl.gov/Akenti/　　アメリカ
- 分散ネットワーク環境でのスケーラブルなセキュリティサービスを提供するセキュリティモデルおよびアーキテクチャ。

・Albatross　　http://www.cs.vu.nl/albatross/　　オランダ
- グローバルコンピューティング環境におけるクラスタシステム（広域クラスタシステム）上でのプログラミング方法や、高性能を得るための方法などに関する研究プロジェクト。

・AppLeS　　http://apples.ucsd.edu/　　アメリカ
- アプリケーションレベルでのスケジューリングに着目した研究プロジェクト。

・Condor　　http://www.cs.wisc.edu/condor/　　アメリカ
- 分散配置された多数のワークステーション等を使って高スループットな計算を行

なうための開発環境等の枠組みを提供するシステム。

・EuroTools　http://www.irisa.fr/EuroTools/　ヨーロッパ各国

- ヨーロッパ全体で高性能計算・ネットワークを利用するためのプロジェクト。

・Globus　http://www.globus.org/　アメリカ

- アメリカを中心とした、現在最も大きなプロジェクト。グローバルコンピューティング環境のソフトウェアインフラストラクチャを構築するためのツール群を提供。

・Grid Forum　http://www.gridforum.org/　世界中

- グローバルコンピューティングシステムの標準化を目指し、様々な技術要素に関して標準化策定など交換を行なう会合。

・IceT　http://www.mathcs.emory.edu/icet/　アメリカ

- グローバルコンピューティング環境における仮想的な計算機の併合/分割や、コードやデータの可搬性などの機能を備えたメタコンピューティングシステム。

・IPG　http://www.nas.nasa.gov/Groups/Tools/IPG/　アメリカ

- コンセントにプラグを差し込むだけで誰もが電力の供給を受けることができるのと同様に、国中/世界中の情報を簡単に引き出すことが出きるような枠組み、仕組みを提供するシステムに関する研究プロジェクト。

・Legion　http://legion.virginia.edu/　アメリカ

- デスク上のパソコン/ワークステーションから世界中の資源にアクセスすることのできる、オブジェクトベースのメタシステムソフトウェアプロジェクト。

・NetSolve　http://www.cs.utk.edu/netsolve/　アメリカ

- ネットワークを介して遠隔地にある科学技術計算用ライブラリを利用するためのクライアント・サーバモデルに基づくシステム。

・Ninf　http://ninf.etl.go.jp/　日本

- 電総研を中心として開発中の日本発のシステム。遠隔手続き呼び出しの手法を利用し、プログラマが容易に遠隔地にあるハードウェア、ソフトウェア、データベース等を利用することのできるシステム。

・PACX-MPI　http://www.hlrs.de/structure/organisation/par/projects/pacx-mpi　ドイツ

- グローバルコンピューティング環境用に拡張したMPI通信ライブラリ。

・UNICORE　http://www.kfa-juelich.de/unicore/　ドイツ

- ユーザが遠隔地にある計算資源にリモートジョブエントリを行うシステム。

研究開発中のシステム

- Akenti
- AppLeS
- Arcade
- CIF
- Condor
- CUMULVUS
- EveryWare
- Globus
- Habanero
- Harness
- IceT
- IPG NAS-NASA
- JINI
- METODIS
- EuroTools
- Llava
- Legion
- NCSA Workbench Project
- NEOS
- NetSolve
- NINF
- Ninja
- PAWS
- PARDIS
- POEMS
- Sweb
- Teraweb
- UNICORE
- WebFlow

図３　代表的なグローバルコンピューティングに関する研究名一覧

グローバルコンピューティングはこれらのプロジェクトは単独で広域分散コンピューティングに関する全ての機能を有するわけではない。それぞれのプロジェクトがそれぞれの役割を担っている。グローバルコンピューティングシステムにおいても階層化モデルにより表現することが可能であり、それを図４に示す。

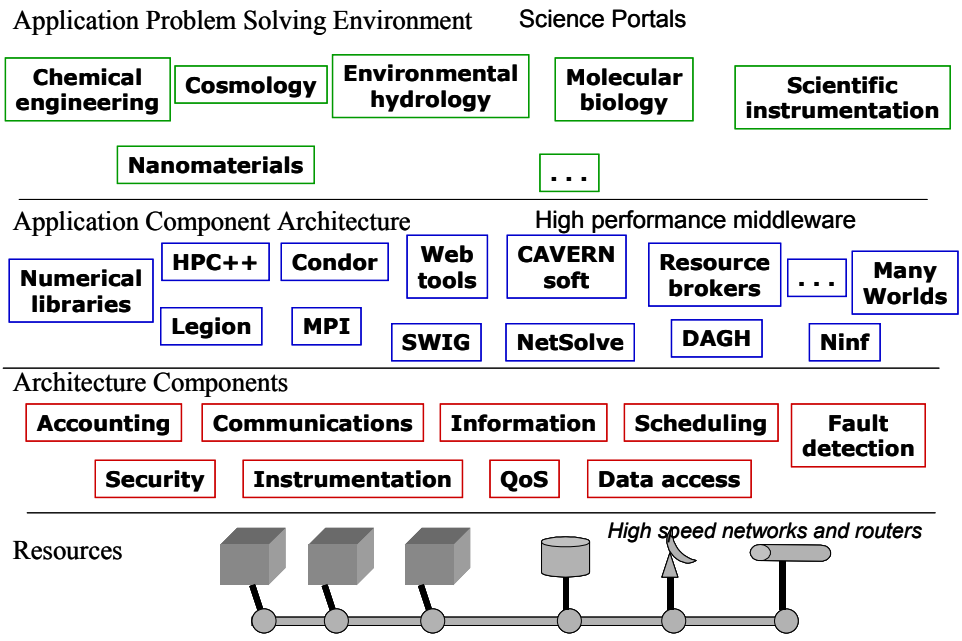


図４　グローバルコンピューティングシステムの階層化モデル

当初、広域分散コンピューティングに関する研究が盛んになり始めた頃は、主にミドルウェア層に位置するシステムが盛んに研究・開発されてきた。現在でも数多くのシステムが実績を残している。図３に示したシステムの多くも、このミドルウェア層に位置することになる。これらのシステムは広域分散コンピューティング環境で動くアプリケーションの作成や、実際に動かす際の手助けをしてくれるシステムであり、プログラマが直接触れることになる部分である。このミドルウェア層に位置するシステムの中から、遠隔地にある計算資源をアクセスするための仕組みを提供しているNinfと、並列プログラミングに良く利用されるメッセージ通信ライブラリである MPI をグローバルコンピューティング環境向けに実装したPACX-MPIに関して紹介する。

また、米国でのスーパーハイウエー構想に関して I-WAY 実験ネットワーク上で様々なツールが開発されてきた。Globus はこれらのツール群をとりまとめたプロジェクトであり、1997年に発足した。Globusプロジェクトはグローバルコンピューティングのソフトウェアインフラストラクチャを構築するためのツール群として近年非常に注目されている Globus Metacomputing Toolkitの構築をはじめとし、短期間で数多くの実績を残している。現在ミドルウェア層に位置する多くのシステムが、何らかの形でGlobus Toolkitを利用するようになってきている。そこで、最後にGlobus Toolkitに関して紹介する。

【Ninf: ネットワーク数値情報ライブラリ】

いわゆるスーパーコンピュータ（スパコン）や大規模クラスタシステムなどの高性能計算システムは、そう簡単に手に入れられるものではない。日本でもそれらのシステムを導入している企業、研究機関、大学等は限られている。それら高性能計算システムの魅力は、何と言っても高い計算処理能力にある。大規模計算においてはスパコンを使えば数分、数時間で答えを求めることが可能となる。従来の分散処理は単に処理時間の短縮を求めていたが、こればかりではなく、高品質ソフトウェアを簡便に共有することが可能となる。例えばあるスパコンに搭載されている計算ライブラリを使えば、非常に精度の高い計算結果を得られるなどである。また、計算のみならず、同様な API によりどこかのサイトが保持している大規模データベースに手元の端末から簡単にアクセスする事も可能である。

このような要求を満たすシステム、つまり手元にある計算資源（PC やワークステーション）上でユーザインタフェースを提供し、遠隔地にある高性能計算機（スパコンなど）を利用するシステムのことを Datorr (Desktop Access to Remote Resources) という。Ninfはこの Datorr システムの１つであり、電子技術総合研究所で当初設計、開発された。現在はお茶の水女子大学、東京工業大学等とも共同で開発を行っている。

Ninf は広域ネットワーク上の数値計算ライブラリや科学技術計算に必要な数値情報データベースを通じて、主に科学技術計算分野の情報ならびに計算資源を提供・共有する仕組みを備えている。Ninf の基本システムはクライアント/サーバモデルに基づいて設計されており、図５に示すように Ninf クライアント、Ninf サーバ、要素ルーチン、メタサーバの 4 つの要素から構成される。

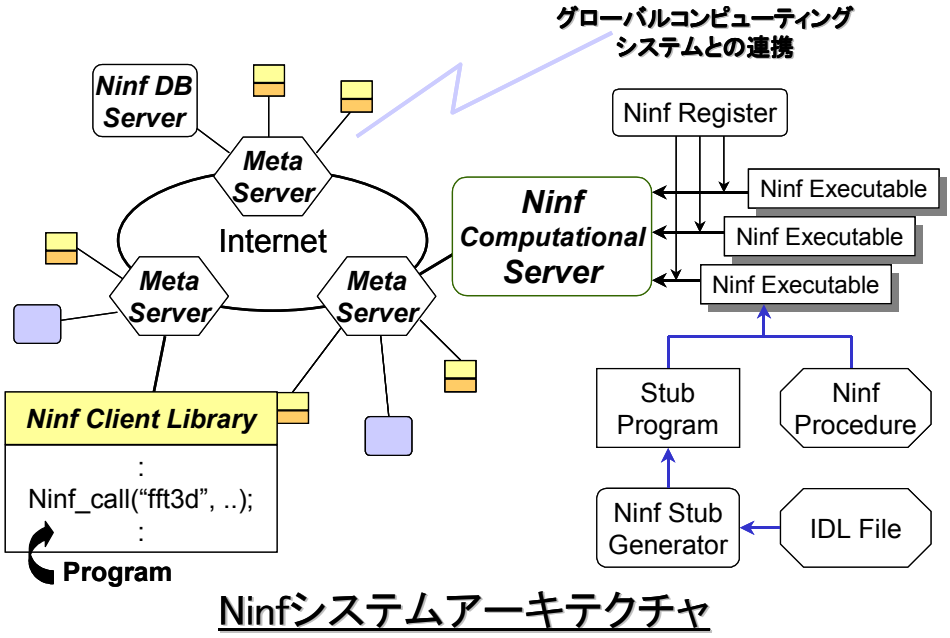


図５　Ninf のシステムアーキテクチャ

Ninf システムは、「計算サービスを提供する側（サーバ）」に対し、手元の端末（クライアント）から計算要求を発信し、遠隔地にある計算資源を利用することを可能とするシステムである。しかも、サービスが増えた場合でも従来の SunRPC 等とは異なり、サーバ側だけで更新が行える。

【PACX-MPI：グローバルコンピューティング環境向けに実装された MPI】

メッセージ通信を用いた並列プログラミングを広域分散システム上に展開するために標準的な MPI を拡張したものが PACX-MPI である。MPI は最も標準的でもっとも多く利用されているメッセージ通信ライブラリであり、C 言語や Fortran などで利用することができる。MPI は利用する計算機のアーキテクチャに応じた方法でメッセージ通信を行うように実装部分が留保されている。例えば共有メモリ計算機においては、共有メモリを介したり、ネットワークで接続されたクラスタシステムにおいては TCP/IP による通信を利用したりする

のと同様に広域分散システムにおけるMPIのデザインがいくつかある。ただし、MPIは元来グローバルコンピューティング環境を意識して実装されていたものがなく、既存の MPI をグローバルコンピューティング環境で利用するのはアーキテクチャの異なる計算機間での利用や遠隔システム上でジョブプロセスを生成する方法に関して障害があり、また性能の面においても問題がある。

PACX-MPI はドイツの HLRS (High Performance Computing Center Stuttgart) で開発された、グローバルコンピューティング環境向けに実装された MPI である。PACX-MPI が提供するライブラリ関数の API やセマンティクスは MPI に準拠しており、MPI を使って書かれたプログラムならばソースコードに修正を加えることなく、PACX-MPI のライブラリをリンクすることにより広域分散コンピューティング環境でプログラムが動作。この場合、広域分散で問題になるのがネットワークの遅延に起因する性能である。例えばそれらの並列計算機がインターネットで接続されている場合、異なる計算機上で動くプロセス間では TCP/IP を用いて通信することになる。一方、構成要素の並列計算機ノードには、その計算機固有の方法で高速なメッセージ通信を行なうことのできる MPI が実装されている。内部の通信とノード間の通信を区別し、性能を維持するため PACX-MPI は同じ計算機上で動作するプロセス間ではその計算機固有な MPI を使って通信し、異なる計算機上で動作するプロセス間では TCP/IP を使って通信を行う。図６に計算機内部とノード間 MPI の関係を示す。

PACX-MPI は現在並列プログラミングの際に非常に多く用いられている MPI をソースコードレベルでの互換性を保ちつつ広域分散コンピューティング環境で効率良い動作を目指したシステムであり、次に述べる MPICH-G (MPICH on Globus Device) と並んで広域分散コンピューティング環境上でのプログラミングの標準である。

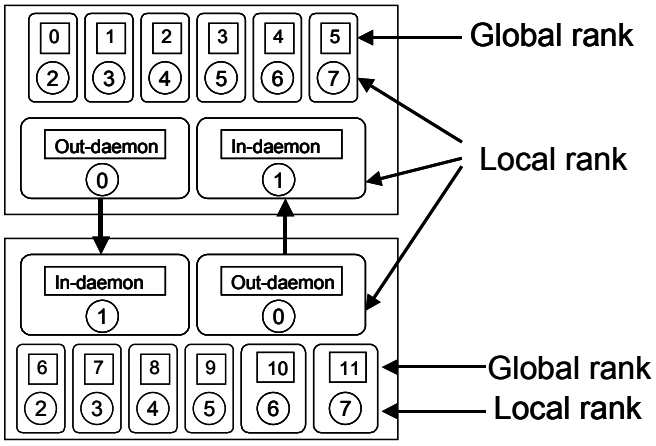


図６ PACX-MPI の構造

【Globus Metacomputing Toolkit】

Ninf や PACX-MPI などミドルウェア層に位置するシステムを構築するためには、ユーザ認証、通信、遠隔計算機上でのプロセス生成、計算資源に関する情報サービス等の要素技術が必要となる。個別ではなく共通のツール群の提供を目指し 1996 年頃に開始された Globus プロジェクトは並列・分散計算、ネットワーク、セキュリティといった様々な分野の研究者達が参加するプロジェクトである。現在も広域分散コンピューティングの分野におけるトップレベルの研究者達がこのプロジェクトに参加している。Globus プロジェクトの大きな産物の 1 つに、Globus Metacomputing Toolkit（以下 Globus Toolkit）がある。Globus Toolkit はユーザ認証システム、通信ライブラリ、計算資源管理機構といった広域分散コンピューティングシステムの構築に必要とされる基本的なサービスの集まり（toolkit）である。Globus Toolkit が提供するツールを用いて上位レベル（ミドルウェア層など）にグローバルコンピューティングシステムを構築することができる。例えば、通信やユーザ認証に Globus Toolkit を用いて MPICH を実装した MPICH-G (MPICH Globus Device) などがある。

前述の Ninf も Globus Toolkit が提供する通信ライブラリを使って実装したバージョンなどが存在する。このように Globus Toolkit は広域分散コンピューティングシステムに必要とされる様々な基本的サービスを提供しており、ソフトウェアインフラストラクチャの事実上の標準となってきている。

1998 年 10 月に Globus Toolkit の Version 1.0.0 がリリースされ、昨年 12 月に Version 1.1.1 がリリースされた。図 7 に、Globus Toolkit が提供するサービスの一覧を示す。

サービス	名前	概　要
資源管理	GRAM	リソースの割り当ておよびプロセス生成
通信	Nexus	Unicast/Multicast 通信サービス
情報	MDS	システムの構造および状態に関する情報へのアクセス
セキュリティ	GSI	authentication などのセキュリティサービス
状態管理	HBM	システムの状況サービス
遠隔データアクセス	GASS	データへのリモートアクセスサービス
実行ファイル管理	GEM	実行ファイルの構築、キャッシングおよび配置

図 7　Globus Toolkit のコアサービス

Globus が提供するこれらのサービスは必要に応じて個別に利用することができるようになっていて、既存のアプリケーションへのインクリメンタルな導入が可能となっている点が特徴的である。

3.3.2.4　ハイエンド計算におけるグローバルコンピューティング

ハイエンド計算において高性能計算機は一朝一夕には導入ができず、また利用技術を蓄積することが困難である。しかし、時代の流れとして、HPC 技術のアウトソーシングが必須である。すなわち、通常はインハウスにあるパソコンや中規模サーバで普段の仕事をするが、ミッションクリティカルの局面ではどこか、まさに地理的、組織的にもどこかの計算能力資源への要求が高まる。この要求と余剰の計算能力を保有しているサーバ群との間でなんらかの受給を満たす市場原理が働くことが期待される。サーバ側はその余剰分に関する情報を交換用マーケットに流しておき、エージェントはクライアント側の計算資源要求をその質に応じて満足するよう振る舞う。なんらかの交渉メカニズムにより合意に達すればクライアントへサーバを紹介する。

サーバはクライアントを集めようと思えば、より高品位なサービスを提供することに努力をするであろう。例えば、裸の CPU 利用権ではなく、ある特定のアプリケーションをパッケージにして利用のノウハウまで併せて提供するようになる。いわゆる ASP の HPC 版に近いイメージである。また、マーケットを成立させるためには、標準的な手順の確立が急務だと考えられる。これらを実現するにはグローバルコンピューティング技術が不可欠となる（図８）。

公的機関としてグローバルコンピューティングの市場機能を構築するニーズは緊急事態策として、緊急事態に陥った場合の危機管理体制下の準備をする必要がある。すなわち、例えば、ある事故により環境汚染が予想される場合には、これまでに計算された様々なシナリオと、現時点で得られる測定値を比較検討しながら、一刻も速い計算とその結果に対応させる対策が必要となる。こうした場合には、国内だけではなく、世界中へＣＰＵを提供することによる貢献が期待される。また、それが可能となるようなシステムを構築するべきである。

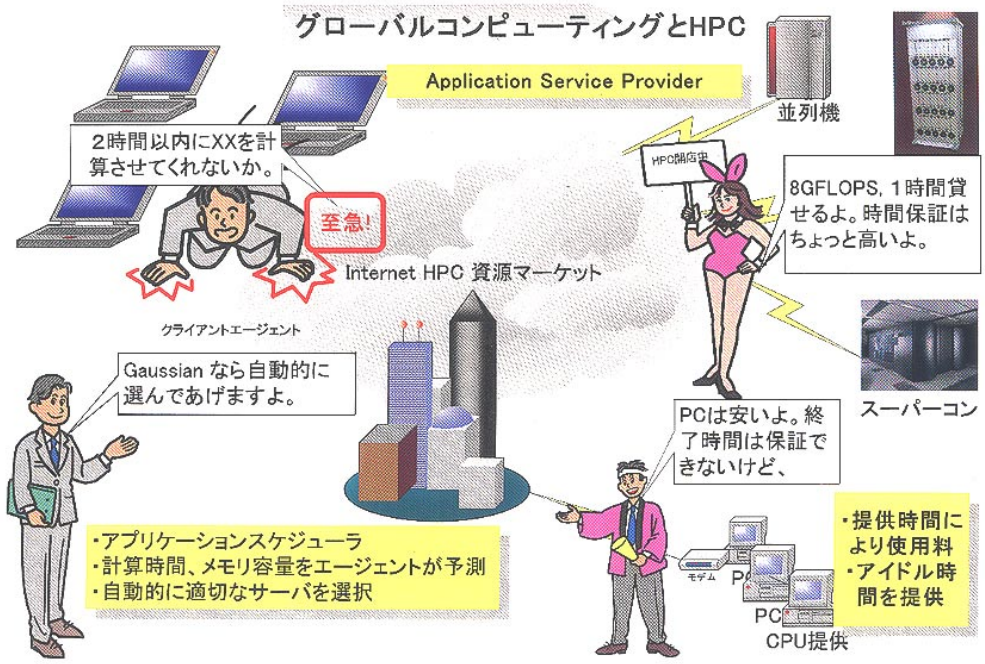


図８　ハイエンドコンピューティング市場

3.3.2.5　グローバルコンピューティングテストベッド

現状の問題点は研究者の絶対的な不足である。広域分散コンピューティングの関連技術を試験的に実行する環境が必要である。単にどこかの WWW サイトを訪れて、必要なソフトを持ってきて、インストールすれば済むほど単純ではない。少なくとも、計算サーバ、高速ネットワーク、クライアントの３者を維持することが必要である。すなわち、自分でサーバとクライアントを兼ねてしまっては本質的な広域分散コンピューティングではない。広域分散コンピューティング技術が発展するためには、魅力的なアプリケーションソフトや開発環境を維持提供し、多くのクライアントを抱えて、実際に有効性と利便性をユーザに展開しないといけない。

こうした開発を促進するには公共的サービスとして計算サーバ構築し、ネットワーク経由でアクセスしても常に安定した運営の継続が必要だと考える（図９）。計算サーバなどもある程度の計算能力がないと、クライアントユーザにとってはネットワーク経由でアクセスする魅力がない。これまでにもコンピューティングテストベッドの維持・構築を小規模で行ってきたが、広く使えるように整備が必要だと考えている。特に計算サーバを強化した大規模なテストベッドを構築し、多くのアプリケーションを創出することが、世界に対する貢献である。

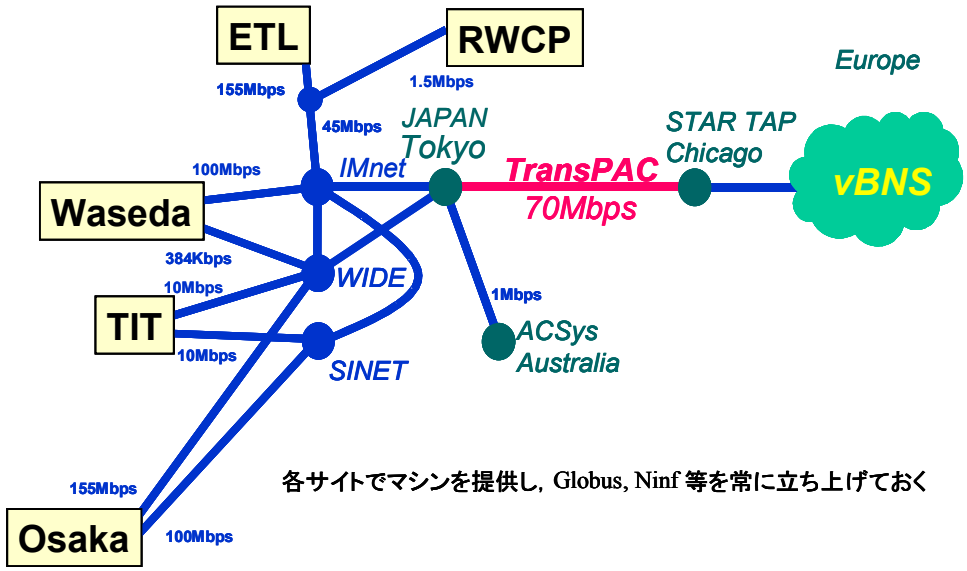


図 9　グローバルコンピューティングテストベッド

グローバルコンピューティングはハイエンドコンピューティングの将来に必須な技術のひとつであり、かつ現在各地で精力的な研究開発が進められている。また国際的な協力関係が大きく期待される分野である。このため、国内におけるカウンタパートの設定はもとより、グローバルコンピューティングテストベッドを含めた国内の環境の整備、また aGrid ともいうべきアジア地区におけるテストベッドと研究開発を緊急に整備する必要がある。また、非常に動きの早い分野であるため、毎年定例になっている会合がいくつかある。今後定期的に開かれるであろうものは Grid Forum である。すでに昨年の 6 月と 10 月に開催され、参加者は 150 名くらいであるが、いくつかのワーキンググループに分散して最新情報の交換、標準化に関する議論を行っている。現在のワーキンググループは下記の 9 つである。

- Scheduling Working Group (Sched-WG)
- Grid Information Service Working Group (GIS-WG)
- Security Working Group (Security-WG)
- Remote Data Access Working Group (Data-WG)
- Application and Tools Requirements Working Group (Apps-WG)
- End-to-end Performance Working Group (Perf-WG)
- Advanced Programming Models Working Group (Models-WG)
- Account Management Working Group (Accounts-WG)
- User Services Working Group (Users-WG)

2000 年度のグローバルコンピューティング関連会議の予定では下記のものがあり、いずれもフォローを続ける必要があると考える。

・ 3/15-17 グローバルとクラスタコンピューティングに関するワークショップ（WGCC2000)、つくば市
・ 3/22-24 グリッドフォーラム（Grid Forum)、米国サンディエゴ市
・ 7/10-12 グリッドフォーラム、米国レッドモンド市
・ 7/18-21 インターネットカンファレンス （INET2000)、横浜パシフィコ
・11/4-10 SC2000、米国ダラス市

3.3.3 大規模並列シミュレーションのための High Level Architecture

古市昌一　委員

3.3.3.1 はじめに

現代社会において、計算機とネットワークは水道や電気と同様社会基盤として重要な役割を果たしている。その中でも最も重要な役割は、人と人、人とシステム、システムとシステムの間のコミュニケーションのための道具としての利用であろう。その次に重要なのは、多様で複雑化するシステムを効率良く設計・製造し、運用開始後における教育・訓練において必須となる、シミュレーションを行うための道具としての利用であると考える。以下に計算機シミュレーションの目的を分類した例を示す。

・自然および物理現象の解明や予測：風洞実験、気象予測、宇宙の進化など
・システムや社会現象の解明や予測：道路交通、株取引、戦略など
・教育および訓練、エンタテイメント：フライトシミュレータ、ゲームなど

シミュレーションを利用する理由は、1）実際に実現するのが不可能、2）安全性や予算面で現実的に実現するのが不可能、3）一度しか操作を実現できず、繰り返し経験を積むことが重要、などが考えられる。

ここで、例えばある自治体が新しい空港の建設を計画した場合のシナリオを考えてみる。すると、以下に列挙するように新空港の開港までには様々な目的でシミュレーションが行われ、各種のシミュレータを用いて解析や訓練が繰り返されるであろう。更に、運用開始後も引き続きシミュレーション技術が活用されることになると予測される。

・空港ビルの構造解析シミュレータによる強度評価
・道路交通シミュレータによる周辺道路の渋滞予測と、新道路建設計画立案
・航空交通流シミュレータによる滑走路長と本数の立案
・避難シミュレータによる空港ターミナルの避難誘導路の立案
・航空管制官の訓練用シミュレータによる管制官の訓練
・旅客機の操縦シミュレータによるパイロットの新空港離着陸訓練
・気象予報や乱気流予測シミュレータによる航空管制業務支援

このように、新空港建設一つを例としても様々な種類の計算機シミュレーションが不可欠であり、今後もシミュレーション技術の発展と実行用計算機の性能向上に対する要求はとどまることがない。

計算機の性能向上は、プロセッサの性能向上と並列処理技術の発展が期待できる。しかし、今後益々多様化するシミュレーションシステムをその都度新規に開発していては、開発コストが増すばかりで効率良いシステム開発を行うことはできない。

例えば、米国国防総省においては長年に渡って多数の解析評価用や訓練用のシミュレーションシステムを開発し、1998 年 3 月の時点では約 800 種類が実際に運用中だと報告されている[1]。

東西冷戦終結後の米国では、防衛産業の民需転換と国際競争力の強化を目的とし、国防予算の多くが情報産業の育成に注入されている。特に、1995 年に出された"Modeling and Simulation Master Plan"[2] 以降、モデリング&シミュレーション技術の育成に対する投資額が大幅に増大し、結果として軍隊の近代化を促すと同時に、今日の米国情報技術(IT)産業の隆盛となって効果が現れているのは間違いないであろう。

そこで、本報告では米国におけるモデリング& シミュレーション技術の育成の中核となって現在研究開発が進行中の、HLA(High Level Architecture) の紹介を通して、HECC のアプリケーション分野の一つとしての並列分散シミュレーション技術に関して考察する。

3.3.3.2 High Level Architecture

HLA は 1995 年に米国国防総省が提案し[3,4]、SISO (Simulation Interoperability Standardization Organization) [5] が中心となって IEEE 1516 標準化[6] が推進されている、異機種シミュレーションシステム間の接続仕様である。特に、多種多様なシミュレーションシステムが利用されている防衛分野において、システムの再利用性と相互接続性を高め、今後の開発・保守コストを低減できるという効果があると共に、他の多くの分野においても同様の効果を期待することができる。

図１に HLA が対象とするアプリケーション分野を示す。図中 DIS (Distributed Interactive Simulation) [7, 8]と ALSP (Aggregate Level Simulation Protocol) [9]はそれぞれ現在防衛用の分散シミュレーションシステムで使われている接続仕様で、米国国防総省下で現在利用されている全てのシステムは、2000 年度中に HLA 化が完了するよう作業が進められている。

元来 HLA は既存の異機種シミュレータを多数共通の接続仕様で連接し、全体として一つの分散協調型シミュレータとして機能させるためのアーキテクチャである。一方、大規模なシミュレーションプログラムを分散計算機環境や共有メモリ型の並列計算機上で実行するための、論理プロセス (LP) 間の同期とデータ交換プロトコルとして HLA を利用しよう

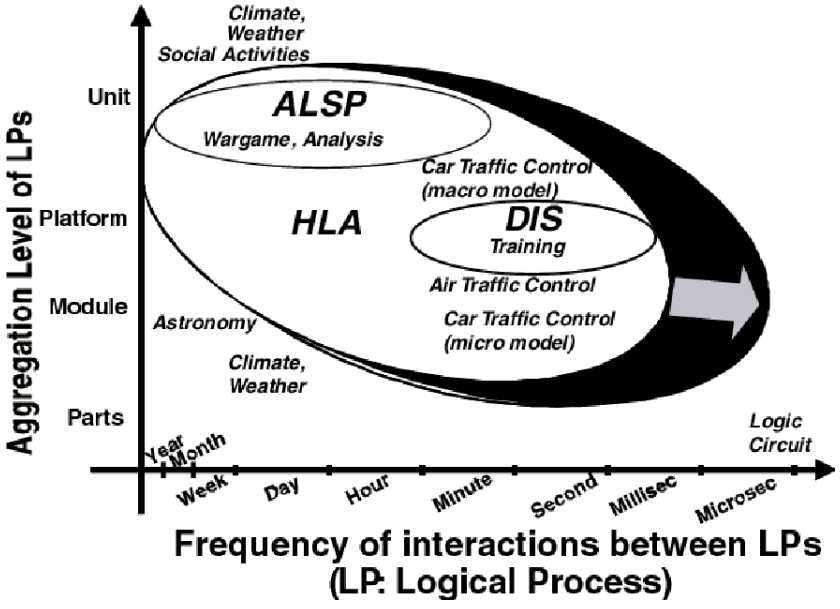


図１ HLA が対象とするアプリケーションドメイン

という研究もなされている[10, 11, 12, 13]。これらの研究により、将来的には図 1 中黒く塗った部分に対象アプリケーションが拡大されるものと予想される。

図 2 に、HLA に基づく分散シミュレーションシステム (HLA ではフェデレーションと呼ぶ) の構成を示す。図中、HLA に準拠した各シミュレータ (HLA ではフェデレートと呼ぶ) は、ネットワークを介して接続された他のフェデレートと、それぞれ共通の API (Application Program Interface) を用いて通信及び同期を行う。API の下位では、HLA インタフェース仕様書[6]により規定された各サービス (通信) を実行するための RTI と呼ぶソフトウェアが動作する。

HLA では RTI の実装方式は規定されておらず、DMSO (Defense Modeling and Simulation Office) が開発してフリーで利用できる DMSO-RTI1.3 および 1.3NG[14,15] を代表に、eRTI[16,17]、pRTI[18] や Javelin[19]などが知られている。

フェデレートと RTI 間のサービスとしては、シミュレーション管理機能やオブジェクト管理機能をはじめとして、図 2 に記載した 6 種類が規定されている。

HLA の特長は、各フェデレートは他に対して公開するクラスと属性を RTI に対して宣言し、また自分が必要とするクラスと属性を購読宣言することにより、フェデレート間では互いに相手の所在や論理時刻の違いを意識することなく、RTI により通信と同期が行われる点である。例えば、図 2 でシミュレータ A がクラス B1 の購読宣言をしておくと、シミュレータ B 上でクラス B1 のインスタンスオブジェクトが生成されると RTI から B1 の生成が通知され、B1 を参照することが可能となる。

HLA と同様にオブジェクト指向で分散システムを構築するための標準仕様としては、CORBA (Common Object Request Broker) があるが、分散シミュレーションでは必須の時刻管理機能と、オブジェクトの所有権管理機能が CORBA には備わっていないため、これらをアプリケーションプログラム側で実現する必要がある[20]。

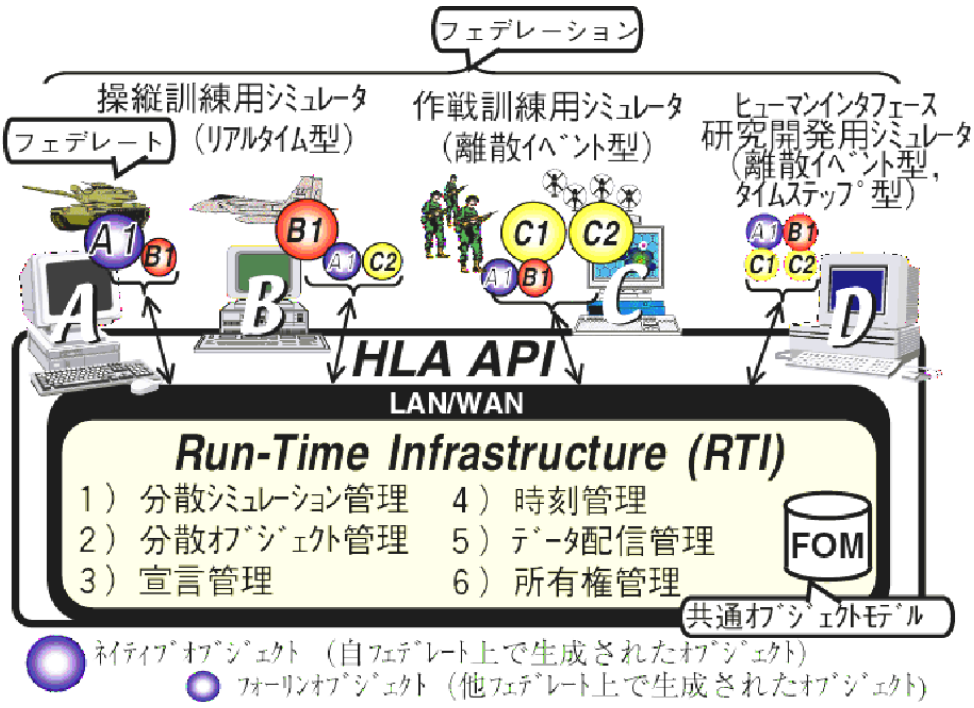


図 2　分散シミュレーション統合アーキテクチャ HLA

3.3.3.3　アプリケーション例

(1) JSIMS: Joint Simulation System

従来より米国国防総省において指揮幕僚訓練で用いられているシミュレータは、陸・海・空軍それぞれで全く異なるシステムが用いられてきた。しかし、近年各軍は共同で作戦を実施する機会が増えており、また、今後システムの改良や拡張をする上でも、できる限り仮想訓練環境は統一するのが望ましいと考えられている。

そこで、HLA が提案されるとほぼ同時期に研究開発が始まったのが JSIMS (図３) である。JSIMS の基本的な概念は、訓練用のシミュレータ、人工知能を応用した Synthetic Forces と、実際の指揮統制システムが同一の仮想環境を共有し、遠隔地に所在する多数の訓練生が同時に訓練を実施することができるための環境を構築することである。

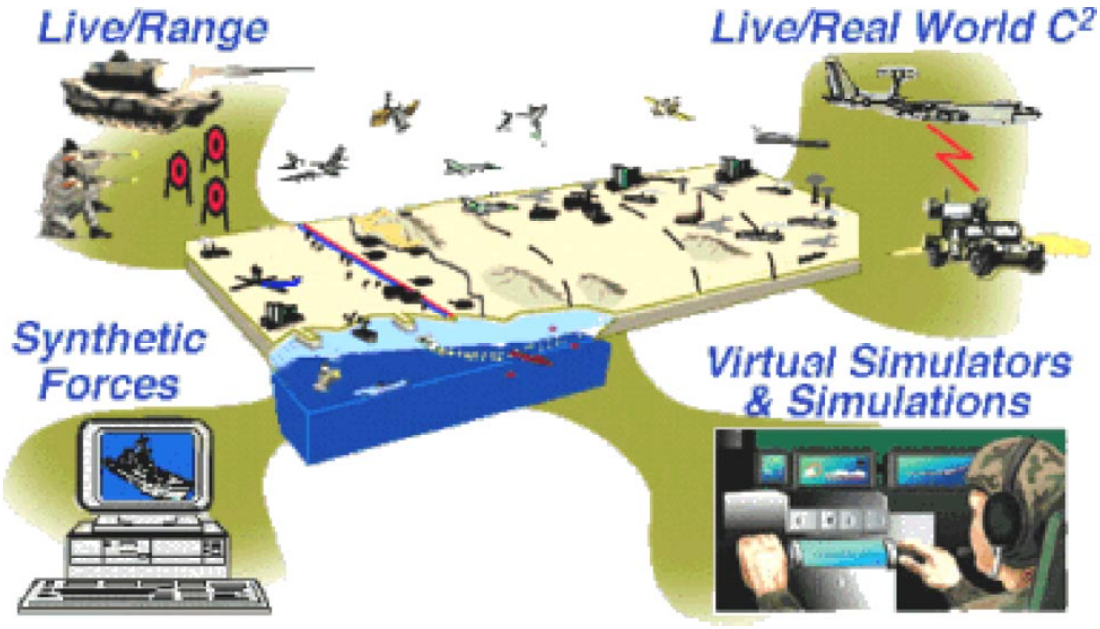


図３　JSIMS（http://www.jsims.mil/user!advo cacy/Program%20Review/j0008.htm より引用）

(2) SBA: Simulation Based Acquisition

ゴア米国副大統領の NPR（National Partnership for Reinventing Government）[21] では、調達コストの 25%削減を目標として掲げている。これに基づき、米国国防総省では調達のサイクルタイムの半減と発注者側の総コスト削減を目指している。そのためのキーとなるのが官民連携による調達におけるシミュレーションの活用である。

具体的には、デジタルモックアップ（DMU）の技術により、試作品を作ることなく、コンピュータ内部に構築した CAD モデルとその可視化（ビジュアライゼーション）ツールを駆使して、部品の組立や製品の運用、補修等の模擬を繰り返すことが可能となる。これにより、新製品の開発期間が劇的に短縮され、コストを大幅に削減することができる。

図４には DSMC (Defense Systems Management College) が作成した"Simulation Based Acquisition: A New Approach" (4-4 ページ) [22] の図を引用する。図からわかるように、例えば航空機を開発する過程において、設計、製造、操作系の評価、訓練、フライト試験に至るまで全てのフェーズは同一の分散協調型シミュレーション環境を共有し、各フェーズ間でデータ交換を行う。更に、フェーズ間を逆戻りすることも可能とすることにより、スパイラルな開発による柔軟なシステム開発を可能としている。

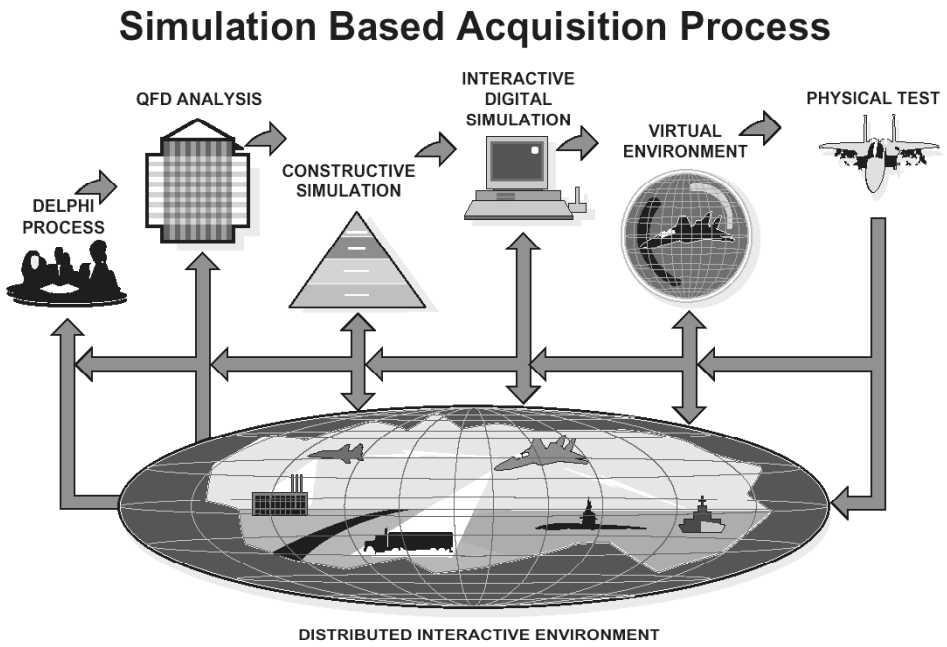


図４ SBA (http://www.dsmc.dsm.mil/pubs/mfrpts/mrfr10a.htm の pp.4-4 より引用)

米国国防総省では、全ての調達品に SBA 手法を適用すべく整備を進めており、シミュレータ間の連接とデータ共有の部分で HLA を利用することから、HLA の応用として最も巨大なアプリケーションの一つとして育とうとしている。

3.3.3.4 おわりに

HECC のアプリケーション分野の一つとして HLA をとらえ、HLA の概要を紹介するとともに、現在米国国防総省下で研究開発が進行中のアプリケーションの一部を紹介した。しかしながら、ここで紹介したアプリケーションは多数の異機種シミュレータを連接するため

の技術として HLA を利用した例であり、大規模並列分散シミュレーションの具体的事例紹介とはなっていない。これは、HLA を大規模並列分散シミュレーションの基盤として使うために必須の、実用的な性能を得られる HLA-RTI がまだ開発途上にあることと、プログラミング環境がまだ整備されてないからである。

既に、米国においては主要なシミュレーションシステム開発支援ツールの HLA 化が実現され､欧州の NATO 加盟各国や極東アジア諸国で HLA に関する研究が大変盛んになっている。一方、国内においては HLA に関する研究開発例は大変少ないのが現状である。今後シミュレーション技術が益々重要になる中、HECC のアプリケーションとして HLA は重要な位置を占めると考える。

参考文献

[1] Defense Modeling and Simulation Office: HLA Architecture Management Group Home page, http://hla.dmso.mil/amg/meetings/ (2000).

[2] Under Secretary of Defense for Acquisition and Technology: DoD 5000.59-P Modeling and Simulation Master Plan, Department of Defense, http://www.dmso.mil/documents/ (1995).

[3] Dahmann, J. S., Calvin, J. O. and Weatherly, R. M.：A Reusable Architecture for SIMULATIONS, Communications of the ACM, Vol. 42, pp. 79-84(1999).

[4] Fujimoto, R.:Parallel and Distributed Simulation Systems, John Wiley & Sons, Inc., pp. 209-221 (2000).

[5] SISO: Simulation Interoperability Standardization Organization, http://www.sisostds.org (2000).

[6] IEEE Std P1516: Draft Standard for Modeling and Simulation HLA -Federate Interface Specification-, http://www.sisostds.org/doclib/ (1998).

[7] IEEE Std 1278-1995: Standard for Information Technology, Protocols for Distributed Interactive Simulation, IEEE (1995).

[8] Hofer, R. C. and Loper, M. L.：DIS Today, Proc. of the IEEE, Vol. 83, pp. 1124-1137 (1995).

[9] Banks, J.: Handbook of Simulation -Principles, Methodology, Advances, Applications and Practice-, John Wiley & Sons, Inc., pp. 645-658 (1998).

[10] Mellon, L. and Itkin, D.: Cluster Computing in Large Scale Simulation, Proc. of the 1998, Fall Simulation Interoperability Workshop, pp. 1119-1127 (1998).

[11] Furuichi, M., Mizuno, M., Izumi, H., Ozaki, A., Takahashi, K., Satou, H., Matsukawa,H. and Aoyama, K.: The Applicability of High Level Architecture to Large Scale Parallel and Distributed Simulation, Proc. of the '99 Spring Simulation Interoperability Workshop, Vol. 2, pp. 466-473 (1999).

[12] Ferenci, S. L. and Fujimoto, R.:RTI Performance on Shared Memory and Message Passing Architecture, Proc. of the '99 Spring SIW Workshop (1999).

[13] Christensen, P. J., Hook, D. J. V. and Wolfson, H. M.: HLA RTI Shared Memory Communication, Proc. of the '99 Spring SIW Workshop , Vol. 2, pp. 571-574 (1999).

[14] Bachinsky, S.T., Mellon, L., Tarbox, G. H. and Fujimoto,R.: RTI2.0 Architecture, Proc. of the '98 Spring SIW Workshop, Vol. 2, pp. 871-880 (1998).

[15] Defense Modeling and Simulation Office: High Level Architecture Run-Time Infrastructure RTI 1.3-Next Generation Programmer's Guide, http://www.dmso.mil/hla/ (1999).

[16] Furuichi, M., Mizuno, M., Miyata, H., Miyazawa, M., Matsumoto, S. and Aoyama,K.: Design and Implementation of Experimental HLA-RTI Without Employing CORBA, Proc. of the 15th DIS Workshop, Vol. I, pp. 195-201 (1996).

[17] Furuichi, M., Mizuno, M. and Miyata,H.: Performance Evaluation Model of HLA-RTI and Evaluation Result of eRTI, Proc. of the '97 Fall Simulation Interoperability Workshop, Vol. II, pp. 1099-1109 (1997).

[18] Pitch Inc.: pRTI (portable RTI) Pro duct Information, http://www.pitch.se/prti (1999).

[19] Myjak, M. D., Sharp, S. and Briggs, K.: Javelin,Proc. of the '99 Spring SIW Workshop, Vol. 3, pp. 999-1007 (1999).

[20] 服部吉洋, 林富彦, 郡司勝, 村上栄一郎, 二宮勉, 宮園道明: CORBA を利用した分散協調型シミュレーション基盤, オブジェクト指向'97 シンポジウム, pp. 3-10 (1997).

[21] NPR: National Partnership for Reinventing Government, http://www.npr.gov (2000).

[22] Defense Systems Management College: Simulation Based Acquisition: A New Approach,http://www.dsmc.dsm.mil/pubs/mfrpts/mrfr10a.htm, pp. 4-4 (1998).

3.4　アプリケーション

3.4.1　HECC の必要分野とアプリケーションからの要望

福井義成　委員

3.4.1.1　望ましい HECC

3.4.1.1.1　HECC のニーズ

HECC が必要な分野には

（１）ミクロな現象とマクロな現象が組み合わされている場合

・ミクロな原理からマクロな現象を解明する場合

・乱流計算

（２）リアルタイム性が要求される分野

・気象予測等

（３）多くのケースの計算が必要な場合

・１ケースの計算はそれほど大きくないが、
計算のケース数が数千から数万ケースの場合

のようなものがある。これを前提にどのようなシステムが望ましいかについて述べる。

3.4.1.1.2　人間にとって使い易い HECC

計算機のユーザが問題解決のために計算機を使う場合、モデル（プログラム）記述が容易なことが大切である。ユーザにとって、演算が速いことが必要なわけではなく、解きたい問題が速く解けることが必要ある。HECC でのモデル（プログラム）記述はどのようにするのが、望ましいのか？

ここでは、HECC は並列計算機と仮定する。現在、普通に使われている手続き型言語（Fortran,C 等）はノイマン型計算機を前提にしている。プログラム（モデル）はどのデータをどのように加工するかの手順を計算機に指示することにより構成されている。現在のアプリケーションも多くのものはアルゴリズムもノイマン型計算機を前提にしている。手続き型言語での並列プログラムの作成はユーザにとって負担が大きい。また、並列計算機では、単独プロセッサでは問題とならない複数プロセッサ間のタイミングの問題も生じる

（図１）。タイミングにからむ問題は再現性が容易でないことが多く、これが手続き型言語での並列プログラミングの困難さの一因である。さらに、現在の並列計算機は各計算機ご

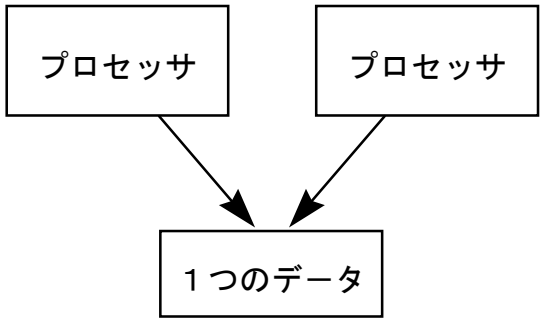


図１　複数プロセッサでのデータの競合

とに性能を発揮させるための手法が異なっていることが多く、ユーザにとって、負担となっている。

並列計算を使い易くするためには、並列計算機向きのモデル記述・アルゴリズムがあるべきである。 ユーザにとって、手続き型言語で並列プログラムを書くことは負荷が大きく、それが可能な人口は限られてしまう。並列処理のユーザを拡大するための障害になる。ユーザにとって、モデル記述とモデルの変更が容易で、並列性の抽出が計算機システムにとって容易であることが望ましい。

非手続き型でモデル（プログラム）を記述する方法も解決のための１つの方策である。非手続き型言語とは、プログラムを手続きではなく、関数型ないしはオブジェクト指向的に因果関係だけを任意の順序で記述することである（図２）。関数型ないしはオブジェクト

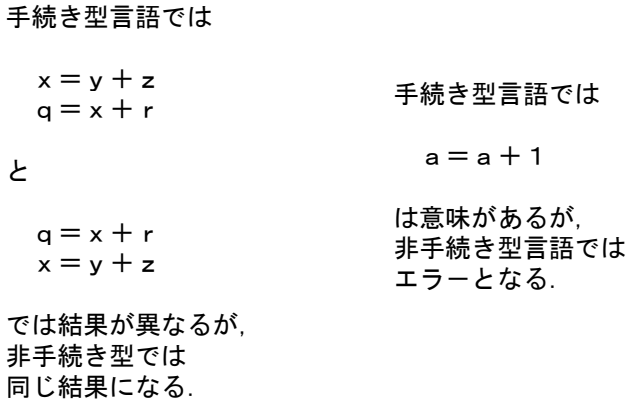


図２　手続き型言語と非手続き型言語

指向的な非手続き型言語で、並列プログラム（モデル）を記述することは人間の思考にあっており、非常に容易である。制御系で良く使用されるブロックダイアグラムが良い例である（図３）。

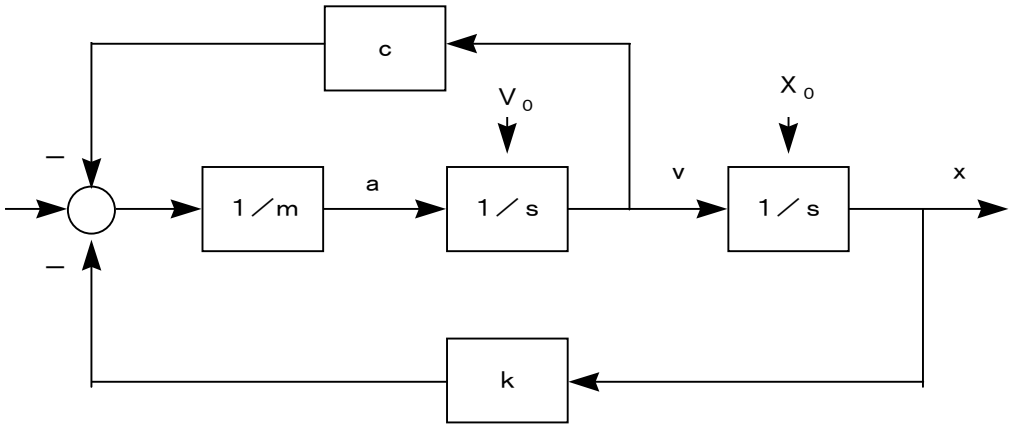


図３　振動子のブロックダイアグラム

手続き型言語ではモデルの計算順序を使用者が考慮しなければならず、小さなモデルでは容易であるが、大きなモデルでは難易度が増す。アプリケーションを開発するときのネックとなる。

別角度から計算の手順を考える。計算機でのシミュレーション手順（現象から計算結果の認識まで）を考えると図４のようになる[1)]。

図４で HECC が有効なのは、計算の部分であり、計算を効果的に進めるためには計算の前後の作業が効率的に行われなければ、全体として良い結果は得られない。特に、 HECC で計算の部分が高速化されることにより、計算の前後の使い易さがますます重要になる。 HECC の性能を十分発揮させるためには、計算だけの高速化だけではなく、計算の前後の処理の高速化が重要である。

計算機が高速化し、計算手法も高度化することにより、問題解決のフローの中で計算の占める比重が段々低下している。自動車の車体強度解析（衝突解析）では、計算よりもメッシュ作成の比重のほうが大きくなっている。計算は数日あれば終了するが、メッシュ作成には数ヶ月が必要になっている。このような部分にも計算機パワーを利用して短縮することも重要になっている。計算の高速化により、計算そのものよりも、モデル作成（メッシュ）に比重が移る。今後は、ますます、問題解決の全体の時間が重要になり、問題解決の手順の中で最も時間の掛かる部分を速くすることが重要となる（モグラたたき）。そのためには、数値的計算だけではなく、数式処理、人工知能等の非数値的処理も組み合わせて、

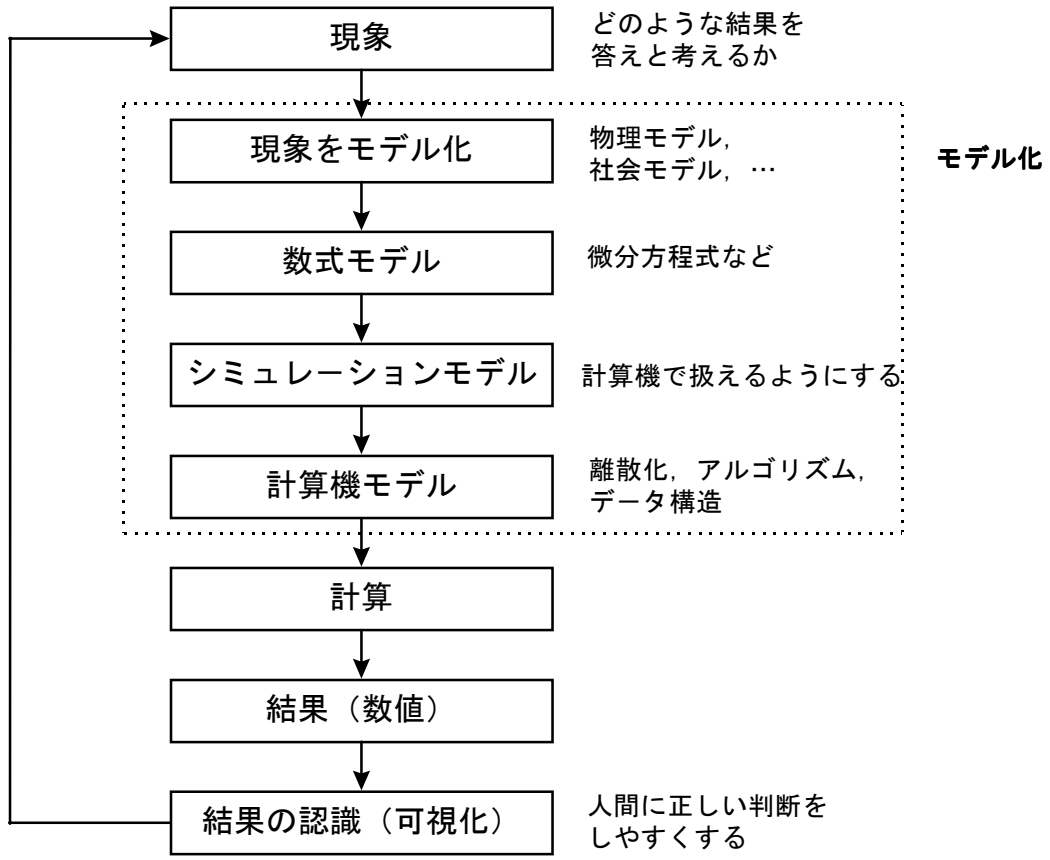


図 4　計算機で問題を解く手順

お互いの良いところを利用し、問題解決の全体の時間を短縮することが重要となる

3.4.1.1.3　並列化の方向

今後の高性能の計算機を構成するには並列化は不可欠である。そのため、並列化を推進するには、現在、高価な並列計算機の価格を安くする必要がある。 並列計算機を普及させて、価格を低減させることが重要である。そのためには並列化が行い易い分野から、並列化を行うべきである。パラメトリック・スタディなどは並列化しやすい分野である。現実的には、疎粒度の問題から手をつけることも必要である。(これは細粒度の並列化を否定しているのではない。並列処理を普及させると細粒度の並列化がより重要になる。まずは並列処理の普及という意見である。)

3.4.1.1.4　まとめ

HECC を使用するであろうユーザが本質的（直接的）に欲しいものは下記のようなものである。

（１）速く正しい結果の得られる道具

（２）問題の処理過程全体が速くなること

これを実現するためには、 HECC のような速い計算機とモデル（プログラム）記述手法（非手続き型、マクロレベルなど）の変更と大規模計算の前後の処理の高速化・高度化である。前後の処理の高速化・高度化には数値的処理だけではなく、非数値処理の高速化・高度化、認識（せまくは可視化）の高速化・高度化も重要である。

また、HECC が実現すると、現在、想定していた用途以外にも新たな用途が必ず出現する。それは過去の例が実証している。そのためには、 HECC が多くの人に利用可能となり、アプリケーションから見て有用なものであることが重要である。

3.4.1.1.5　付録（具体的に欲しい機能）

進展著しいデバイス技術で今後実現して欲しいこと（あるべき姿）とＨＰＣ分野で欲しい技術について述べる。また、記号処理についてもふれる[2)]。

（１）データの移動の構造を記述する手段

現在の計算機で大規模な計算（特に、ランダムなデータを扱う場合）を行っていて、最大の問題は演算ではなく、データの移動である。ノイマン型計算機ではメモリー内の番地を指定し、そのデータをメモリーからプロセッサ内のレジスターに移動する。レジスター内のデータ同士を演算し、結果を別のレジスターに格納し、その後、レジスターからメモリーにデータを転送するのが一般的である。演算は最近のプロセッサの高速化により、ほとんど問題とならない。メモリーとレジスター間の転送時間が最大の問題である。これはプロセッサに比べて、メモリーの遅さに原因がある。演算や分岐制御は完全にメモリーとレジスター間の転送に隠れてしまう場合もある。プロセッサとメモリーの速度差があることを是認すれば、この速度差を避ける手段を考える必要がある。普通、よく行われる計算機アーキテクチャ的解決方法はメモリー遅延の隠蔽である。

ソフトの面から考えると現在のプログラミング環境にはデータの移動の構造を明示的に記述する手段がない。そのため、必要以上にデータ移動の回数（データを示す番地デ

ータも含めて）が増えてしまうことがある。データの移動構造を明示的に記述できれば、計算機アーキテクチャ的にメモリー遅延の隠蔽も容易になる。

これを実現するにはソフト的な開発とそれを支援するハード的な開発が必要となる。

（２）可変データパス計算機

前項で述べたハード的な開発に利用できそうな技術が FPGA であろう。FPGA を利用したシステムでは命令体系等を可変にする研究は目にしているが、アプリケーションの性能を上げるためには、命令体系だけではなく、データパスを可変にした計算機が望まれる。可変データパスとは表現しているが、本当にデータパスを変更しないまでも、キャッシュの制御アルゴリズムやプリフェッチのアルゴリズムを指定できるだけでも非常に有効であろう。

（３）記号処理について

記号処理の場合、浮動小数点演算は非常に少ないので、文字処理が高速に行える計算機であることが必要である。具体的には LISP のような言語が高速に処理できる計算機である。記号処理の 1 分野である数式処理においては、ベクトル計算のような連続的なメモリーアクセスは少なく、多くの場合、ランダムなメモリーアクセスを行わなければならない。数式処理ではランダムなメモリーアクセスが高速な計算機が望ましい。

数式処理にはアルゴリズムを書いて計算するものと、「公式」を用意しておき、パターン認識で行うものがある。後者の場合、パターン処理が高速な計算機が必要になる。この点は翻訳や音声認識と共通な部分があるのではなかろうか。

大規模な科学技術計算に戻って考えると、数値的処理だけでは収束等に非常に時間のかかる計算も数式処理を上手く組み合わすことにより、収束性などを大幅に改善できる場合もある[3)、4)]。

3.4.1.2　仕組みへの要望

上で述べたように、計算機を使って問題を解くという原点に戻ると、計算そのものだけではなく、モデル化の問題、アルゴリズム、計算結果を正しく認識させる技術も重要である。基本的なモデリングについては、企業では難しいため、大学や政府の研究所での研究が必要である。

日米比較では多く（ほとんど）のアプリケーションが欧米製であり、日本はお寒い状況

である。上記のモデリングについても、米国、欧州での研究に遅れをとっている。米国の大学の場合、モデルを開発し、それをアプリケーションにのせるまでの仕事を行っている。日本でもそこまでやれるような体制（仕事の評価まで含めて）が必要である。また、すべての分野で欧米と対抗することは不可能であり、日本が得意とする分野で自主性を発揮すべきである。

今後は下記のような基礎的な分野での研究投資が重要である。

・計算機を構成する素子の基礎
・基本的なアーキテクチャ
・大規模計算のためのアルゴリズム（数値計算を含む）
・詳細な現象を表現できるようなモデル化技術（物理的面が大きい）

また、シミュレーション分野ではソフトの重要性が大きく、それも理論的なものだけではだめで、実際に利用可なレベルまでのソフトの開発が重要である。しかし、日本では大学等でのソフトに対する評価がされていないのが問題である。米国との比較においても最大の問題である。

技術開発にあたり、米国優位となる日米の差はソフト開発に対する評価と新しいソフトに対する先取性の有無であろう。これは、これまで日本が欧米に追いつくことだけに重点をおいてきたことの弊害であろう。

参考文献

[1] 福井義成, HPC を目指して, 情報処理学会 HPC 研究会資料 93-HPC-46, pp.1-6, 1993.
[2] 福井, 野寺, 久保田, 戸川, 新数値計算, 共立出版, 1999.
[3] Lecture Notes in Computer Science Vol.378 (Ed.)J.H.Davenport, Springer-Verlga 1989. pp.163-171. M.Suzuki, T.Sasaki, M.Sato and Y.Fukui : A Hybrid Algebraic-Numeric System ANS and Its Preliminary Implementation
[4] Journal of INFORMATION PROCESSING, Vol.13, No.4, pp.529-533, 1990. T.Sasaki, Y.Fukui, M.Suzuki and M.Sato: Proposal of a Scheme for LInking Different Computer Languages -from the Viewpoint of Algebraic- Numeric Computation

3.4.2　非数値計算の現状から見たハイエンドコンピューティング

久門耕一　委員

科学技術計算、特に基礎科学に属する分野に必要な高性能計算を行なうハイエンドコンピューティング向きの計算機は、商業的利益を見込むことが難しいため、長期的計画の元に公的機関が中心となって技術開発を進めて行く事が重要である。

現在まで、ベクトル計算機に代表されるモノリシックな構造を持つ計算手法が、科学技術計算における高性能計算の手法として、もっとも有効であった。ベクトル計算機は高性能技術計算に頻繁に出現する行列計算を効率良く実現するアーキテクチャであったためである。

一方、計算機の商業利用の分野では、高性能計算とは、データベースシステムによるオンライントランザクション処理が重要であり、近年ではインターネットなどの通信基盤整備のために、計算能力要求の増大が飛躍的に増大してきた。オンライントランザクション処理は、科学技術計算とは動作が全く異なるものと考えられているが、実際には、以下に述べるように、離散系のシミュレータの動作と類似する点も多い。

本節では、代表的なオンライントランザクション処理ベンチマークである TPC-C ベンチマークにみる処理内容を特徴づけ、さらに、高性能科学技術計算との関係について検討を行なう。

3.4.2.1　ビジネスアプリケーションと従来型科学計算の相違

現在生産されている高性能計算機（サーバ）のほとんどは、技術系計算ではなくデータベースシステム（DBMS）によるオンライントランザクション処理に用いられている。一般に、オンライントランザクション処理は、I/O ボトルネックにより性能が律速され、CPU アーキテクチャ的には改善の余地がないと考えられやすいが、十分に性能チューニングされたシステム、その中でもトランザクション処理のベンチマーク測定を行なうような環境では、CPU の命令実行速度が性能を決定している。科学技術計算では、メモリが CPU にデータを供給するバンド幅で性能が律速されるが、オンライントランザクション処理でも同様に、キャッシュミスヒットによる CPU のパイプラインストールで処理性能が律速されるようなメモリバウンドな処理である。

他方、科学技術計算とは異なり、浮動小数演算の比重は小さく、命令のほぼすべてが整数系の演算に限られる。

3.4.2.2　オンライントランザクション処理と離散系シミュレーション

オンライントランザクション処理では、処理要求はシステム外部の端末から発生し、定められた時間（数秒）内に処理結果の応答を返す作業を繰り返す。個々のトランザクションは独立して処理ができるが、トランザクション単位で、扱うデータに対しての排他制御を頻繁に行なう。また、各々のトランザクション間の並列度を向上させるため、全体の排他制御は極力行なわず、扱う個々のデータ単位に排他制御を行なう。データベースシステム固有の基本特性としては、システムダウン時にもトランザクションは完全に実行されたか、まったく実行されなかったかのどちらかで無ければならない（原子性）ことや、全トランザクションに対するログを記録することが要求される。

オンライントランザクションでは、どのような要求がいつ端末から出されるのかは、確率的にしか与えられておらず、処理のスケジューリングは、動的に行なわざるを得ない。後に述べる代表的なベンチマークである TPC-C ベンチマークでは、外部からの 1 つのトランザクションを処理するために、ディスクアクセスが必須となる。ディスクへのアクセスを減らすため、システム内には膨大なキャッシュを持ってはいるが、１トランザクション当たり数十回のアクセスは必要になる。このディスクアクセスレイテンシは、ソフトウェアによるマルチスレッディング処理で隠蔽される。その結果、システム内のトランザクション処理の多重度は、プロセッサあたり 100 から数 100 程度は必要となる。

技術計算においても、このような構造を持つ計算として、分散非同期の離散的シミュレーションを行なうシミュレータが知られている。

3.4.2.3　オンライントランザクションベンチマーク（TPCC）における性能

オンライントランザクションベンチマークのひとつである、TPC-C ベンチマーク処理における並列効果を実際のベンチマーク結果から引用すると、以下の表にまとめられる。ここでは、比較を容易にするため、現状でもっとも高い性能値を持ち、かつ、同一 CPU で同一ソフトウェアによるベンチマーク結果が数多く提出されている Intel の Pentium-III 550MHz CPU を使用したシステムの実行結果を引用した。

表 1 と表 2 で CPU は同じだが、OS と DBMS が異なっているので、相互比較はできない。前半の表では、CPU 数を多くしていった時の単一 CPU 当たりの性能劣化、後半の表では、クラスタ構造を用いた場合の性能劣化を知ることができる。

TPMC=システムが 1 分間に処理したトランザクション数。

表 1　Pentium - III 550MHz 使用の SMP システムの TPCC 性能

システム名		TPMC	CPUs	TPMC/CPU	OS/DBMS
Compaq	ProLiant 8500-550-4P	26,686	4	6,671	MS SQL-EE7.0/NT4.0
HP	NetServer LH 6000	33,136	6	5,522	MS SQL-EE7.0/NT4.0
Compaq	ProLiant 8500-550-8P	40,697	8	5,087	MS SQL-EE7.0/NT4.0

表 2　Pentium - III 550MHz 使用の SMP/クラスタシステムの TPCC 性能

システム名		TPMC	CPUs	TPMC/CPU	OS/DBMS	クラスタ？
Unisys	e-@ction ES5085R Ent	48,768	8	6,096	MS SQL2000/Win2000	SMP
Compaq	ProLiant 8500-64P	152,208	64	2,378	MS SQL2000/Win2000	クラスタ
Compaq	ProLiant 8500-96P	227,079	96	2,365	MS SQL2000/Win2000	クラスタ

この表から、CPU 数が増加すると、システム性能は向上するが、SMP 構造であっても単一 CPU 当たりの性能は低下していることが分かる。この性能低下は、複数の CPU 間での排他制御の多発によるバストラフィックの増加により説明され、平均メモリアクセスレイテンシの増大による平均命令実行時間（CPI: Clock Per Instruction）の増加が主な要因である。

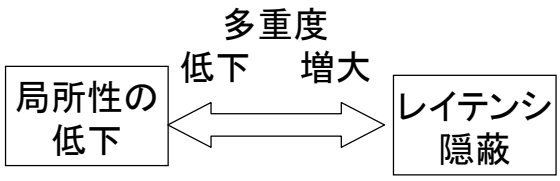
メモリアクセスレイテンシが増大した結果、キャッシュミスヒットによるパイプラインストールにより、スーパースカラ構造の CPU でも、CPI（平均命令実行クロック数:Clock Per Instruction）は 1 を切ることは出来ない。上記のベンチマーク実行時には、推定で CPI は 2.5 - 3.5 に達しており、CPU の稼働時間の 6 割以上は、メモリ待ちによるインターロック状態にある。

TPCC ベンチマークプログラムでは、各端末から発行されるトランザクションの 90%は、あらかじめ決められたデータベースにアクセスするよう局所化されているため、トランザクションの局所性は比較的高い。その結果、クラスタ構造による性能向上を見込むことができるが、SMP の場合の性能に比べ 1CPU あたり、約 1/3 程度の性能しか得られず、疎結合システムにするためのオーバヘッドが大きいことが分かる。しかし、この局所性を生かすためには、データの配置、スレッドの配置が難しく、この配置如何により性能が大きく左右されるため、SMP 構造でのシステム性能に比べ現実の性能を向上させるための手間が極めて大きい。

3.4.2.4　オンライントランザクションと大規模科学技術計算の相違

TPC-C ベンチマークの他の技術計算系ベンチマークとは異なる特徴として、計算能力の増大が実行速度の短縮ではなく、処理量の増大に向けられている事があげられる。現状の数千倍の演算能力を目指す高性能技術計算においても、計算時間の短縮だけではなく、正確な結果を得るために計算規模の増大に向かうと考えるのは自然である。

DBMS ではディスクアクセスレイテンシを隠蔽するためにソフト的なマルチスレッディングを実現しているが、結果的にデータの局所性が失われてしまう。今以上に、CPU 速度とディスク速度の乖離が大きくなると、多重度を上げなければならず、メモリ上のディスクキャッシュのヒット率が低下し、結果的に性能が低下する。ハードウェアマルチスレッディングを行なう場合もまったく同様の事が起きる。



局所性の向上とレイテンシの隠蔽は背反の関係であり、キャッシュの性能を期待する限り、最適な多重度はアプリケーション毎に異なるため、明確なターゲットアプリケーションを仮定しないと、各ハードウェアパラメータの設定は不可能である。

オンライントランザクションでは、端末からの要求により処理が開始されるが、トランザクション要求は統計的分布が与えられており、先行して実行を開始することが出来ない。

また、一つのトランザクションを実行する際のメモリアクセスも、非反復的で、統計的にしか振る舞いを知る方法が無いのでプリフェッチは困難である。他方、行列要素をすべてアクセスするような行列演算とは異なり、メモリアクセスの局所性はある程度確保できる。従って、キャッシュメモリによる主記憶アクセスの削減は可能で、現在の数 MB のキャッシュメモリでのミスヒット率は 0.5%前後である。キャッシュサイズの増大とともにミスヒット率は漸減するが、その減少率はあまり大きくない。

3.4.2.5　超高性能計算機構築に向けての課題

オンライントランザクション処理を非数値計算の代表例として考えると、現状の科学技術計算手法とは似た、あるいは異なった限界が分かる。

・全体の動作がモノリシック構造ではなく、疎粒度ながらかなりの分散処理が行われている結果、処理は全体としては同期的に行われておらず、競合するデータへのアクセスを調停するためのオーバヘッドが大きい。
・プリフェッチによるレイテンシ隠蔽はできず、マルチスレディングでの隠蔽を取らざるを得ない。元のプログラムに多くの並列度が含まれていたとしても、複数 CPU に複数の比較的独立なスレッドを割り当てるだけの並列度を抽出することが難しい。
・外部とのやり取りを基本としているため、実時間上の制約が加わる。
・クラスタ構造を取る事で、絶対的な性能向上を図る事が出来るが、性能を得るための手間が格段に大きくなる。

これらの事から、高性能技術計算においても、大規模な分散システムを作成するためのソフトウェア基盤の整備が必須であることが分かる。

現在まで、高性能並列計算機システムを構築する、大学間あるいは、国家プロジェクトが日本でも実施されてきた。それらは、CPU 数が数百台の規模ではあったが、ハードウェア主導型で作成された事が多く、ソフトウェア作成の段階で問題を生じた。ソフトウェアがわの立場からは、存在しない大規模ハードウェア向けのソフトウェア開発は不可能であるという見方もある。しかし、目標とすることは、高性能計算を必要とするアプリケーションに対し、要求に見合う高性能システムを提供することである。汎用計算機を目標とすることは、何のとりえもない計算機を作ることに等しい。従って、高性能並列計算機システム構築の最初の課題は、高性能計算機を必要とするアプリケーションを定め、そのアプリケーションに対し必要とされるシステムの設計を始めると言う手順でなければならない。

米国の ASCI 計画では、軍事目的の核兵器シミュレーションがこのようなアプリケーションとして筆頭に上げられている。ひとつでも成功例を作成することにより、そのシステム性能に対し魅力を感じる研究者を引き付け、周辺に応用を広げてゆくと言う戦略と考えられる。

日本において、アプリケーション主導型のシステム開発を進めるため、ソフトウェアとハードウェアの協調開発に適したアプリケーションの選定が高性能計算機システム作成への最初の課題だと考える。

3.4.3　気象，気候シミュレーションとモデル開発

横川三津夫　委員

3.4.3.1　はじめに

1992 年リオデジャネイロで開催された地球サミット（UNCED）において、温室効果ガス排出の影響をより正確に把握するための気候変動に関する国際連合枠組条約（UNFCCC：United Nations Framework Convention on Climate Change）が署名された、この動きは、1990 年に気候変動に関する政府間パネル（IPCC：International Panel on Climate Change）が発表した第一次評価報告書が契機となっている。1995 年の第二次評価報告書では、「中位のシナリオを用いれば、100 年後には地球全体平均気温が２℃上昇し、平均海面水位は約 50cm 上昇する」という衝撃的な内容が述べられており、人類の社会経済活動において発生する温室効果ガスが地球環境に多大な影響を与えることを警告している。UNFCCC の目的を具体化するために開催された温暖化防止京都会議（COP3：The Third Conference on the Parties）では、2008 年から 2012 年までの間に６種類の温室効果ガスの全体排出量を、1990 年比で少なくとも５％削減するという目標に同意しており、温室効果ガスの排出抑制による実質的な経済活動への影響が懸念される。地球温暖化の影響をより詳しく把握するためにも、地球規模の数値モデルによる数値シミュレーションの重要性が認識されており、この分野の研究開発の進展が望まれている。地球温暖化予測の最近の話題は、文献［1]、[2］に詳しい。

一方、世界各地で発生する異常気象についても解明されることが期待されており、特にエルニーニョ現象との関連性について関心が高い［3]、[4]。異常気象は、社会経済活動に影響を与えるばかりでなく、洪水や熱波による人的被害も大きい。このような現象に対しても数値モデルによる予測は重要さを増してきている。1992 年のブラジル北東部に出されたエルニーニョ現象による干ばつ予報によって、干ばつに強い作物の選択がなされ収穫を維持したという例が報告されており、これは数値モデル（数値シミュレーション）による異常気象予測が有効であることを示している。

ここ数年、地球環境変動に関する研究が注目されている背景には、地球温暖化予測や異常気象解明が、ある程度数値シミュレーションによって可能であると考えられているからであろう。しかし、地球規模の現象は、時間的、空間的にもスケールの極端に異なる現象が複雑に絡み合って生じており、これらの複雑な現象のメカニズム解明は非常に困難である。また、気象現象は本質的にカオスの性質をもっており、初期値依存性が大きいため、長期予報には本質的に限界がある。すなわち、「バタフライ効果」と呼ばれる例え（ブラジ

ルで一羽の蝶が羽ばたけばテキサスでトルネードが生じることがある）どおり、気象現象はある場所のちょっとした空気の流れの違いが遠く離れた気象現象に影響を与えうる性質を持っており、数値シミュレーションにおける初期値（多くは観測地からの推定値）の与え方によっては長期予報の結果が大きく異なってしまう。

しかし、気象や気候に関する数値シミュレーションがこのような性質を内包していたとしても、実験では現象の再現が極めて困難であるし、また理論的解析が不可能であるため、数値シミュレーションが唯一の解析方法と云える。数値シミュレーションの精度を向上させるためには、離散化空間の解像度を高くする必要があるが、現在のスーパーコンピュータにおいてさえ、その能力が十分でない。21 世紀に向けて、ハードウェア性能の強化ばかりか、効率よいソフトウェア開発環境など広範囲のハイエンドコンピューティング技術の重要性は言うまでもない。

3.4.3.2 気象、気候シミュレーションに関する米国の開発状況

米国においても、大気、海洋モデルに関する研究が進められている。IT２イニシアチブに関する Bluebook2000 には、大気、海洋モデルに関する研究開発として、

- NOAA による分散主記憶型並列計算機向け海洋大循環シミュレーション用プログラム MOM3（Modular Ocean Model, version 3）の開発
- 地球流体力学モデルを並列化するために NOAA / FSL が開発している気象モデリング用スケーラブルツール SMS（Scalable Modeling System）の開発
- 気象、海洋研究のコミュニティで広く利用されているデータ標準フォーマット netCDF について、NOAA / GFDL と LBNL が行っている並列 I/O 機能の研究
- NOAA / NCEP による並列計算機向けの気象予測モデルの開発
- NOAA / FSL が行っている飛行機に影響を与える気象の予測モデル RCU-2 の開発
- NOAA / NCEP によるハリケーン予測
- 実時間気象予測

についての記述があるが、詳しい内容が書かれていない。ここでは、IT２イニシアチブの一部である DOE の SSI（Scientific Simulation Initiative）のパンフレットに沿って、DOE が実施している気象、気候シミュレーションについて概観してみる。

当然の事ながら、地球環境変動の予測を行うことは緊急を要するとの立場であり、質的にも量的にも信頼のある予測が要求されている。このため、以下に示すような大気、海洋、海氷域および地表面を考慮した解像度の高い地球気候モデルの開発が目標となっている。

①信頼度の高い予測が可能なモデル

モデルに含まれる物理的プロセス、化学的プロセス及び生物学的プロセスは、十分理解が進んでいることが条件で、現在及び過去の気候データとの比較によって、モデルの妥当性が十分検証されている必要がある。

②定量的に評価可能なモデル

気候変動の予測値は、カオス的変動に基づく不確実さを集合（アンサンブル）平均によって評価でき、将来の経済成長予測とそれに起因する二酸化炭素や他の化学物質の放出予測が評価可能なモデルが必要である。

③解像度の高い局所域モデル

現状モデルでは解像度が低く、その予測値は地域的、地形、土壌、植生、工業地域などの要因で大きく変わってしまう。現在の大気モデルでは水平方向 300km 格子及び鉛直方向 20 層の格子、海洋モデルでは水平方向 100-200km 格子及び鉛直方向 30-40 層が用いられているが、細かい計算格子間隔が取れるかどうかは、局所的気候変動予測を精度良く行う上で重要である。現在のモデルでは、100 年の気候数値シミュレーションを行うのに最新のスーパーコンピュータでも１ヶ月は必要である。この格子間隔を現在の 10 分の１となる 30km にするのが長期的目標である。

スーパーコンピュータに対する要求性能については、「現在のモデルで高い解像度を達成しようとすると、計算機は現在の少なくとも 100 倍が必要である。また、現在のモデルに大気の化学変化及び海洋の生化学変化等のプロセスの追加を行うためには、さらに５倍以上の能力が必要である。さらに水平方向の格子数を２倍にし、それに応じて鉛直方向の格子を増加させるためには、数 10 倍の能力が必要である。結果として、１ペタフロップス以上の計算機が必要であると予測されるが、あと５年間では達成不可能と思われる。」と述べられている

ハードウェアだけではなくソフトウェアの進展も必要としており、その場合には、応用科学者と計算機科学者の協力が重要であるとしている。以下の内容が研究すべき課題としてリストアップされている。

・マイクロプロセッサの効率を３０％から５０％にあげるプログラミング手法
・モデル方程式に対する計算が簡単となる定式化
・並列計算機上の気候モデルに特化した数値解法
・高度なコンパイラ
・高度な数学ライブラリと通信方法

各論として、次のモデルに関する記述がある。

(1) 高解像度大気モデル

NCAR 開発の CCM3.6.6 を用いた高解像度シミュレーション結果

(2) 高解像度海洋モデル

POP (Parallel Ocean Program) を用いた 0.1 度格子のシミュレーション結果。衛星による観測結果と比較している。

(3) 海氷域モデル

気候シミュレーションにおける海氷域モデルの重要性を示す記述とロスアラモス国立研究所で開発された CICE (Sea Ice Model) によるシミュレーション結果

(4) 並列気候モデル (PCM：Parallel Climate Model)

大気海洋結合モデル PCM (CCM3＋POP＋改良モデル) の開発

(5) 局所モデル

3.4.3.3 数値モデルの開発について

日本においても、地球変動の解明を目標とした研究組織が発足している [5]。そこでは、

(1) アジア・太平洋域における気候変動の予測

(2) アジア地域における水循環の予測

(3) 地球温暖化の予測

(4) アジア・太平洋域における大気組成変動の予測

(5) アジア地域における生態系の変動の予測

(6) 地球内部変動メカニズムの解明

が具体的目標として設定されており、次世代の数値モデル開発も含まれている。数値モデルは今後の研究に欠くことの出来ない研究手段であるが、高性能計算を行う上では並列計算機の利用技術、もっと言うならば並列計算機そのものの知識が必要であり、計算機科学者との連携が重要となっている。

しかしながら、米国と比較して日本はソフトウェア開発の重要性の認識、大規模高性能計算への要求が低いように感じられる。とりわけ、適切な応用ソフトウェア開発体制を取ることができず、応用分野の研究者と計算機科学分野の研究者の協力関係が築けないように思う。

米国では、既に述べたように DOE だけでも気象、気候分野の数値シミュレーションが組織間の連携の下、または DOE 以外の組織との連携を取りながら、精力的に研究開発が実施

されている。それらの開発の多くは米国の研究コミュニティにおける数値シミュレーションのための共通基盤モデル（コミュニティモデル）と言えるものである。SC99（米国ポートランド）のパネル討論において、２つのコミュニティモデル構築の紹介があったが、米国ではコミュニティモデルを共同で開発し、それを育成する体制が取れるようである。

２つのコミュニティモデルの内､一つは宇宙気象学分野の NSF プロジェクト GEM(Geospace Environment Modeling) 計画で開発されている GGCM であり、太陽風の地球周辺部での挙動解明と宇宙気象予測を行うためのシミュレーションモデルである。もう一つは、地震機構の解明を目的に開発されている地震の大規模数値シミュレーションモデル GEM である。GEM では、オブジェクト指向プログラミングの考え方を取り入れ、計算手法として Fast multipole method、３次元有限要素法、適応的格子生成法を用いている。

他方、日本ではコミュニティモデルの開発体制が取れない。応用分野の研究者と高性能計算分野の研究者の連携はその重要性が認識されておらず、並列計算のためのプログラミングは計算機メーカの SE 頼みになっているのが現状である。これは、省庁間や研究コミュニティの壁、開発ソフトウェアの所有権帰属の問題等、複雑な問題に一部依るところが大きい。また、強力なリーダが欠如している。より高性能計算を追求していく上では、応用ソフトウェア開発体制を整備していくことが望まれる。

3.4.3.4　まとめ

今回の報告書をまとめるにあたって、米国の気候モデルに関する資料を読み返した。概して、パンフレット等はわずかな成果を大げさに記述するのが普通であり、実際どの程度有効なソフトウェア開発が行われているかは明らかでないが、応用ソフトウェア開発プロジェクトの旗の下、研究者が参集し、協力しながら成果を出していく体制を取っていることには感心している。

応用ソフトウェア開発の重要性を認識しつつ、望ましい開発体制、開発をリードできる人材の育成、開発資金の投入方法について見直す必要があると思う。

参考文献

[1] 三村，地球温暖化の影響，気象 42.9（1998）

[2] 真鍋，地球温暖化予測の現状，気象 42.8（1998）

[3] 木本，世界各地で発生する異常気象，気象 43.9（1999）

[4] 藤吉，異常気象の社会経済への影響，気象 43.10（1999）

[5] 松野，地球フロンティア研究システムの発足，天気，Vol.45，No.4（1998）

3.4.4　健康と気象のデータマイニング

三菱電機　田中秀俊　講師

ネットワーキングの普及と記憶装置の低価格化から、情報流通はその速度も量も指数関数的な増加を見せている。例えばコンビニエンスストアの中枢データウェアハウスには日に何百万ものトランザクションがひたすら堆積し、その解析をするにあたってさまざまな課題が生じている。理論的には、従来の統計手法を使えば、データがあればあるだけ強力な推定ができることになっている。しかし実際はそう簡単な話ではないらしい。問題は、統計手法でも学習手法でも何を使うにしても、とにかく予測をする、すなわち正解と過去事例が一緒に格納されているデータを元に、パターンを見つけ、新たな事例で正解を出す、それにはどうするか、という点に尽きる。それには何テラバイトもの高速処理という量の壁と、真に有用なパターンの抽出という質の壁を乗り越えて、死蔵データを財産に変える技術が求められている。前者はまさに高性能コンピュータの仕事であり、後者は人間と高性能コンピュータの協同作業になるだろう。

本項では、データマイニングと呼ばれて注目を集めている、上記のような大量データに基づく予測問題について、最近の研究動向を簡単に紹介する。そして筆者らの、テラバイトなどとはとても言えない小規模な人間とパソコンの協同作業の例として、健康診断データと気象データの解析例を紹介し、データマイニング最大の問題とも言われる、需要について触れる。

3.4.4.1　データマイニングの現況

データマイニングには予測と知識発見とがあると言われる[1]。予測問題は事例と解答がペアで蓄積されたデータベースを前提としていて、目標はそのデータベースにおける解答に相当するものを新事例から当てることにある。そしてデータベースあるいは分野の知識が未整理で予測には不十分な場合、知識発見問題と呼ばれるようである。アプリオリ法[2]はデータマイニング流行のきっかけを作り、並列処理技術も広く研究され、量の壁への挑戦がもっとも進んでいると考えてよいが、これは知識発見問題の解法であって、予測の前段階であるにすぎない。

一方、予測法として注目されているのは、分類問題ではサポートベクターマシン[3]、バギングやブースティングに代表されるコミッティー学習[4,5]、時系列予測問題においてはGARCH 法[6]ではないかと思われる。筆者らは時系列予測問題に隠れマルコフモデル[7]を適

用する方向で研究を行っている。

3.4.4.2　健康診断データマイニング

本例はまさにデータベースも分野の知識も未整理で予測には不十分な、知識発見問題の例である。対象に考えたのは、半年に一度の定期健康診断における各種検査結果と、その際に回収される生活習慣アンケートである。本格的に収集すれば数千人の結果を数年分というオーダーは可能だったが、とりあえず 5000 人分あまりを借りて、アプリオリ法を若干改良しその評価を行った。

一般にデータマイニングは、データの獲得、削減、データからのルール抽出、ルールの削減や解釈、そしてルール適用の５段階からなると言われている。本章ではその順に簡単に手法と結果の紹介を行う。

(1) データ獲得

対象データ数は 5770 レコード、属性数 561 で、その大半はアンケート項目回答（「いつも」「ときどき」「いいえ」からの三者択一）である。ここから、データのクリーニングとして 5770 レコードすべてにおいて値の記述されていない属性、値が 1 種類の属性など、属性出現頻度の観察には不要と思われるものを取り除いた。このときは、検査・アンケート・追加検診・体力測定などの複数の属性集合ごとに用意されたデータから、手間をかけて被験者毎の全属性に対応する長いレコードを作成した。しかし後の別データからの同様な作業との共通性を検討した結果、属性集合識別子、被験者識別子、属性集合に対応した値集合の三者からなるレコードを作成するのが、アプリオリ法のような属性値の同時出現頻度を観察するようなルール抽出には適しているとの結論を得ている[8]。

(2) データ削減

量の壁への挑戦自体も重要だが、併せてデータ量の削減の重要性も指摘される。これは質にも影響する本質的なものである。データ削減はデータ数の削減、データの属性の削減、データ属性のもつ値の種類の削減の３種類からなる。健康診断データにおいては、前述のような最低限の属性削減をした上で、連続値の離散化をユーザと対話的に実施することを試みた。データの平均と標準偏差から得られる適当なデフォルトの離散化用閾値をユーザに提示し、ユーザはそれを上書きできる。検査結果の中で異常値範囲が既知のものを、こ

の機能を用いて修正した。さらに、不健康な被験者に対象をしぼるべく、検査結果の頻度を分析して９以下のものをすべて削除して実験を行った。

(3) ルール抽出

ルール抽出については、アプリオリ法を改良して用いた[9]。データ削減を徹底的に行ったため Pentium Pro 200MHz から PentiumIII 400MHz 程度の計算機を用いて数分という程度の問題になった。ちなみにデータ削減を行わない場合は、数日～1 週間かかる。

(4) ルール削減・解釈

ルールは、アプリオリ法のパラメータ設定にもよるが、数万のオーダーで得られた。この中から有用なルールだけを残す方法として、ユーザによるルール削減とルール検索の手段を用意した。ユーザが一度見て不要と判断したルールを削除するフィルターや、ルールで言及する項目が互いに包含関係にある場合に、基準を設けて一方を削除する方式を試作した。

(5) ルール適用

得られたルールは産業医に確認を求め、有用そうに見えるルールがいくつかあるという指摘をしてもらった。白血球数と喫煙習慣の関係など、検査と生活習慣の関係を示したルールには有用なものが多そうだという見通しは得られた。しかしながら、ルールの形式的な削除では除けない大量の「あたりまえ」のルールに阻まれ、真に有用なルールを指摘しつくすことはできなかった。

3.4.4.3　気象予測データマイニング

高度 20km 付近の比較的風速の弱い成層圏に飛行船を浮かべ、そこを無人基地として局所的な観測や情報通信中継に役立てようという計画がある[10]。このような飛行船の運航、停留、回収には、風速の予測が不可欠である。この予測に用いることのできるデータには、地表付近に関してはアメダス、気象レーダ、GPV[11]などがあり、高層については高層気象観測ゾンデ[12]と、信楽にある MU レーダ[13]とがある。この中から入手の容易な高層気象観測ゾンデに基づき、風速予測問題を設定して実験を行った[14]。

気象庁が気象業務支援センターを通じて公開している観測資料 CDROM には、気象ロケット観測で得られた高度 20ｋｍ～約 60ｋｍにおける気温・風向・風速等の資料や、日本国内

の高層気象観測官署におけるレーウィンゾンデ観測とレーウィン観測で得られた観測資料などが収録されている。この中からレーウィンゾンデ観測による上空約 30ｋｍまでの指定気圧面における風速観測結果を解析に用いた。

統計手法には均質な母集団を前提としたものが多いが、母集団の均質性に基づかないデータモデルを用いた解析手法もよく用いられる。例えば複数の性質のデータが混合した母集団を想定するものがあり、音声認識の分野などでよく使われている。このような混合母集団モデルの中で、均質とみなせるデータ集団ひとつひとつを、本稿では要素母集団と呼ぶことにする。混合母集団モデルにおいて推定すべきパラメータは、均質とみなせる各要素母集団の平均や標準偏差と、それら要素母集団の混合比になる。このような要素母集団の性質とこの混合比とを同時に求める方法として、EM 法（Expectation Maximization Algorithm, 期待値最大化法）が知られている[15]。

実験では混合母集団モデルに対角共分散混合ガウシアンモデルを選択した。これは各要素母集団が共分散成分のない多次元正規分布をとるという、混合母集団モデルの中では非常に単純なモデルである。結果の例として、高層気象観測データを混合数 20 で分類した結果を図 1 に示す。これは指定気圧面 22 ヶ所の風速、計 3652 回分のデータを 20 種類に分類したものである。図 1 において、横軸は指定気圧面、縦軸には風速をとっている。凡例の部分の数字は何回分がその分類に帰属するかを表している。

また、各データがどの要素母集団に属する確率が最も高いかによって分類帰属を決め、データを分類の変遷の形に直して図 2 に示す。図 2 では横軸に 5 年分の時系列をとり、縦軸に 20 種類の分類を平均風速の順に並べている。

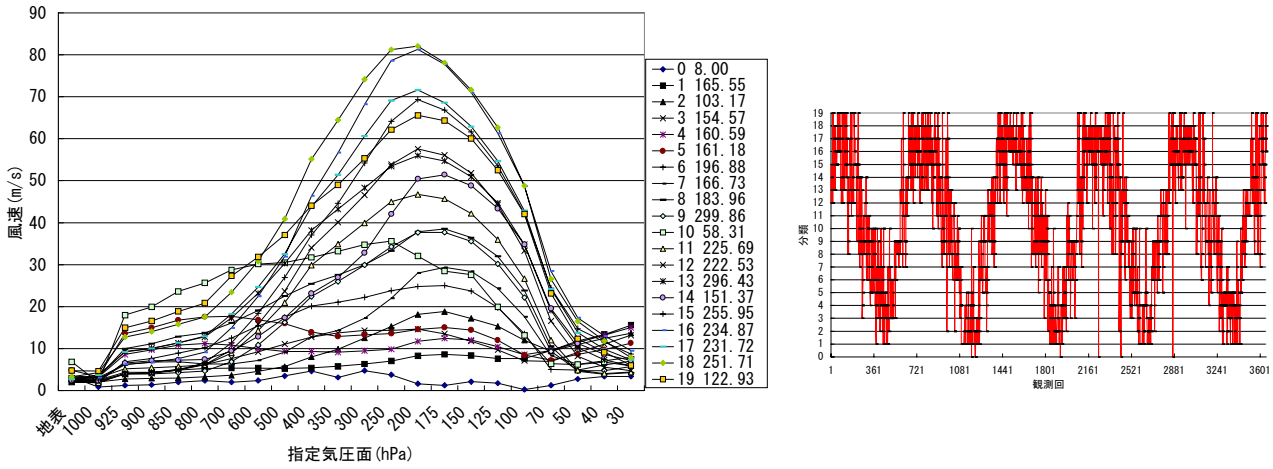


図 1　鹿児島上空 1993～1997 年の風速分布の 20 分類結果（左）

図 2　鹿児島上空 1993～1997 年の風速分布の 20 分類の変遷（右）

図 2 で示した分類変遷は各分類が確率的な出力分布を持つので、この上でマルコフ性を仮定した解析、例えば「ある分類の x 回後の分類」の分布を調べることは、隠れマルコフモデル[7]を想定して解析していることになる。この分布から、ある分類に属するデータの x 回後のデータについて区間推定ができる。図 3 はその例として、図 1 や図 2 の 20 分類に基づき、各分類から 1 回後～28 回後にどの分類になるかをそれぞれ集計して、その比率を用いて 1 回後～28 回後に風速が地表から高度 20km までにわたって 20m/s 以下である確率を示した。図 3 によれば、 1 回後には分類 1, 3, 4 が 80～95%、一方 28 回後には分類 1, 2, 4 が 60～65%の高い確率を示している。最も平均風速の低い分類 0 に属する 8 回は、実際に風の弱い時と、観測不能で 0 が記録されている時が混在していることが図 2 から窺える。

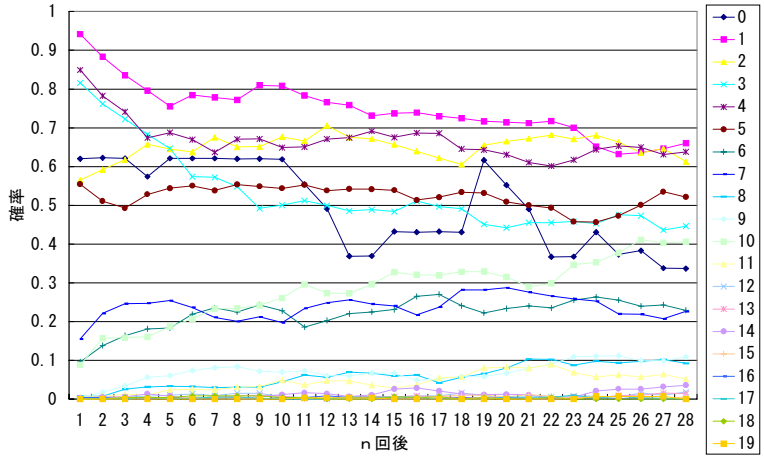


図 3　20 の分類のそれぞれについて n 回後の観測で全観測値が 20m/s 以下である確率

3.4.4.4　需要展望

本節ではデータマイニングの一実施例として、健康診断データと気象データの解析例を紹介した。健康診断データにアプリオリ法を適用した例では、データ削減を徹底的に行ったため Pentium Pro 200MHz から PentiumIII 400MHz 程度の計算機を用いて数分という程度の問題になったが、削減をしなかった場合にはすぐ数日を要する規模となった。これを例えば企業従業員全員を対象にするだけで、計算機への要求性能は数桁はねあがる。ただ、アプリオリ法に代表される知識発見段階の手法は、具体的応用までにもう一段階必要なことがほとんどで、現段階ではユーザを見つけるのが困難という問題がある。実際、データマイニングで成功している企業は、別の製品等の付属物として扱って成功しているだけとまで言われている。別の製品とは例えば、高性能並列計算機であったり、高性能データ可

視化ソフトウェアやグラフィクスマシンであったり、統計パッケージであったり、経営コンサルテーション業務だったりする。このことから、データマイニングはケーキのチェリーに過ぎないという言われ方もする。一方、気象データを例にとった予測問題は、そのまま応用が見えるため、需要が高くユーザがつきやすいと考えられる。気象データに限った話でも、今後局地予報や航空予報の分野に需要が高まると予想されており、その際には今回とは桁違いに大量の観測データについて同様の解析を行う必要が発生すると期待されている。

参考文献

[1] Weiss, S.M. and Indurkhya, N. *Predictive Data Mining*. Morgan Kaufmann Publishers Inc. (1998).

[2] Agrawal, R. Apirori. (1993).

[3] Schoelkopf,Burges & Smola(Ed.) *Advances in Kernel Methods*. The MIT Press. (1999).

[4] Breiman, L., Bagging Predictors. *Machine Learning*. 24 (2)123-140, (1996).

[5] Freund, Y. and Schapire, R.E., Experiments with a New Boosting Algorithm. *Proc. of the Thirteenth Intl. Conf. on Machine Learning*, Morgan Kaufmann, San Francisco, California. pp.148-156, (1996).

[6] 三浦. デリバティブの数理. 臨時別冊数理科学 SGC ライブラリ 6. サイエンス社. (1999).

[7] 中川. 確率モデルによる音声認識. 電子情報通信学会. (1988).

[8] 田中他. データマイニングのできるデータ形式について. データ工学研究会(1999).

[8] 三石他. Knodias におけるデータマイニング方式. 第 56 回情処全国大会 (1998).

[9] 第 1 回成層圏プラットフォームワークショップ予稿集(1999)

[10] http://www.jmbsc.or.jp/ (気象業務支援センター)

[11] 高層気象観測データ (CDROM 付属マニュアル)

[12] 廣田. 気象解析学. 東京大学出版会. (1999).

[13] 田中. 高層大気の風速解析と予測. 情処 数理モデル化と問題解決研究会 99-MPS-26 (1999).

[14] McLachlan, G.J. and Krishnan, T. The EM Algorithm and Extensions. John Wiley & Sons, Inc. (1997).

第４章　あとがき

本年度における HECC ワーキンググループの調査としては、過去３年間の「ペタフロップスワーキンググループ」における調査活動の結果およびその報告書を踏まえ、新たな視点から、高性能コンピューティングに関連する情報処理技術のあるべき姿を探るように努めた。具体的には、情報技術に関する大統領諮問委員会（PITAC: President's Information Technology Advisory Committee）の提言やこれに応えて、米国政府がまとめたいわゆる Blue Books を踏まえて、わが国の立場として、これから取り組むべき分野や技術的な課題、そして各種の問題点などについて議論を行ってきた。この報告書は、それらの議論を踏まえて各委員の見解をまとめたものであり、必ずしも各委員の意見が同じ方向を向いているわけではないが、全体を包含する考え方としてはある程度のコンセンサスがあると考えている。

昨年度までの「ペタフロップス WG」でも何度か指摘されていたが、今年度の「HECC WG」での調査などを総合的に勘案して、ハイエンドの計算機およびその上で効率的な処理を実行するために次の２つの点を解決することが重要であることが、より明らかになってきていると考えられる。その１つは、プロセッサの処理速度とメモリの速度とのギャップが年々高まってきており、このギャップを埋めるためのハードウェア、アーキテクチャそれにソフトウェアなどの各レベルでの新たな技術開発が必要とされている点である。もう 1 つは、高性能計算を行うために必要となる並列処理および分散処理のソフトウェア技術をさらに高める必要があるという点である。前者に関しては、プロセッサアーキテクチャの観点からは Processor In Memory と呼ばれるプロセッサと DRAM を同じチップ上に実装することによってプロセッサーメモリ間のスループットを改善する方式や、多数の実行スレッドを低オーバヘッドで切り替えられるアーキテクチャなどが有望視されている。同様に、ソフトウェアの観点からも、メモリ遅延を隠蔽するコンパイラ技術やマルチスレッドに対応したソフトウェア技術が求められている。すなわち、メモリ階層や並列処理におけるアクセス遅延などをうまく利用するアルゴリズムや、それらを意識したコンパイル技術の開発が重要であるということが指摘できる。これらの技術を総合的に統合するような開発計画として米国では、HTMT（Hybrid Technology Multithreaded Architecture）プロジェクトがすでに計画から実行段階に入っている。これは、超伝導デバイス、光インターコネクト、大容量ストレージ技術などに関して最新の技術を開発してペタフロップス級の計算機を開発しようとする野心的なプロジェクトであり、我が国としてこのようなプロジェクトの動向に

注目する必要がある。一方、後者に関しては、まず並列計算機の実効効率を高める技術が必要である。そのためのコンパイラ技術や最適化技術、あるいは並列化のためのプログラミングツールなどの果たす重要性が高い。これらの技術開発、とくにコンパイラ技術に関しては、通産省のアドバンスト並列化コンパイラプロジェクトが計画されており、その成果が期待される。また、ネットワークの高速化およびグローバル化に伴い、広域に敷設された高速ネットワークを利用したコンピュータの集合体をインフラとして構築または利用し、あたかもひとつの巨大な計算機環境を利用しているような使い勝手を利用者に提供するグローバルコンピューティングやメタコンピューティングと呼ばれる考え方やその関連技術が提案されており、これらの技術開発を進めることによって、HECC をより身近なものにすることが可能となる。

HECC 分野における技術的および産業的なポテンシャルを高めるためには、第３章の笠原委員や福井委員の指摘に見られるように、わが国における HECC 国家戦略の必要性や応用分野の研究者と高性能計算機分野の研究者の連携が必要である。この点は、昨年度の「ペタフロップス WG」の報告書でのあとがきでも指摘したことであるが、我が国としては、独自の立場で HECC 分野さらには情報処理技術全般に対する戦略を早急に立てるべき時に来ているのではないだろうか。わが国においても、省庁横断的な形で「情報通信産業技術戦略検討会」（座長、長尾真　京都大学総長）による報告書が平成 12 年の３月に出され、その中で情報処理に関する提言がまとめられている。米国政府に対する PITAC などの提言と比較すると、総花的で、概念的な面はあるが、このような方向性が現れたことは、大きな一歩であり、少なくてもこのような提言に沿った形でわが国の HECC 技術開発を含む情報通信に関する技術開発やそれに対する予算措置などを早急に講じることが必要であると考えられる。

（山口喜教主査）

付属資料１　平成 11 年度ワーキンググループ活動記録

※ワーキンググループ全５回は先端情報技術研究所（AITEC）内会議室にて開催

開催日	出席者(人数)	発表内容
第 1 回 99.10.12	主査（1） 委員（10） AITEC（5） 幹事（2） 計 18 名	・岡崎幹事「米国支援 ハイエンドコンピューティング研究開発動向」
第 2 回 99.11.29	主査（1） 委員（8） AITEC（7） 幹事（2） 計 18 名	・秋山委員「大規模並列計算機によるタンパク質情報解析」 ・近山委員「Computing Continuum -CACM Nov.1998 を中心に」 ・花輪委員「ストレージ システムの動向」
第 3 回 99.12.13	外部講師（1） 主査（1） 委員（8） AITEC（4） 幹事（2） 計 16 名	・工藤講師「ローカルエリアで高性能並列計算を実現するネットワーク　RHiNET」 ・古市委員「異機種分散シミュレーション接続アーキテクチャ HLA と並列分散シミュレーションへの応用」
第 4 回 00.01.24	外部講師（1） 主査（1） 委員（8） AITEC（7） 幹事（2） 計 19 名	・田中講師「健康と気象のデータマイニング」 ・中島委員「PIM(Processor In Memory)の研究動向」 ・妹尾委員「並列分散処理とそのプログラミング」 ・天野委員「高速スイッチ異機種格闘技戦」 ・笠原委員「ハイエンドコンピューティング国家戦略の必要性」
第 5 回 00.02.21	主査（1） 委員（8） AITEC（7） 幹事（2） 計 18 名	・関口委員「ハイエンド計算におけるグローバルコンピューティング」 ・久門委員「高速計算機とビジネスアプリケーション」 ・福井委員「HECC-WG 課題と要望」

付属資料２　2000年度Blue Bookにみる
米国ハイエンドコンピューティング研究開発技術動向

米国2000会計年度Blue Book(参考http://www.hpcc.gov/)の内容から、現在、米国が推進しているハイエンドコンピューティング分野に関する研究開発を紹介する。以下では、アーキテクチャ・ハードウェアコンポーネント・アルゴリズム・ソフトウェア・アプリケーションに分類し、それぞれ、スポンサーの省庁別にまとめてみた。

1　アーキテクチャ

1.1　DARPA、NASA、NSA

(1) Hybrid Technology Multi-Threaded (HTMT) アーキテクチャ

DARPA、NASA、NSAによって資金提供を受けている研究者たちは、毎秒10^{15}の浮動小数点演算(1ペタフロップ)が可能なコンピュータシステムを構築する可能性を評価している。HTMTアーキテクチャは、改良された半導体技術と、要件を満たす最先端ハイブリッド技術（超伝導技術、光インターコネクト、高速VLSI、磁気記憶技術）を組み合わせる。
(http://htmt.jpl.nasa.gov/)

図1　HTMTアーキテクチャを採用したコンピュータシステムのプロトタイプ3-D図

1.2 NSA

(1) 量子コンピューティング (Quantum computing)

小型化しているコンピュータコンポーネントの現在のトレンドは、2020 年までに量子スケールに到達するだろうと示唆している。NSA のサポートにより、９以上の大学と企業が量子コンピュータ (QC) の研究を行なっている。

1.3 DOE

(1) 量子テレポーテーション (Quantum teleportation)

オークリッジ国立研究所 (ORNL) は、量子コンピューティングと通信分野の 1 研究領域である量子テレポーテーションを研究する最先端の研究所を設立した。ORNL は、シグナル発信機と発信されたシグナルが量子力学的にもつれ合った状態での実験をしている。実験には高出力の 1,000 兆分の 1 秒レーザーが使用される。

(2) ハイエンドアーキテクチャの評価

ORNL は、新しいハイエンドアーキテクチャ (例えば SRC-6 など) が DOE や政府の他のアプリケーションに対して適用性があるかを評価している。SRC-6 は商用のマイクロプロセッサ技術を新しい高性能なメモリアルゴリズムプロセッサ (MAP) サブシステム (特定のアプリケーション用に柔軟にハードウェアをカスタマイズできるようなプログラム可能なハードウェアコンポーネント) と結合する。ORNL は MAP プロトタイプサブシステムを評価しており、また、SRC-6 の受取りを計画している。

また、サンディエゴスーパーコンピュータセンター (SDSC) にある国立先端コンピュータ基盤パートナーシップ (NPACI) は、DOE の国立エネルギー研究所スーパーコンピュータセンター (NERSC) 、カリフォルニア工科大、ボーイングコンピュータサービスと一緒に、テラマルチスレッドアーキテクチャ (MTA) を評価している。このシステムは、プログラムを数百の実行スレッドに分割し、高度な並行処理を達成しようとしている。

1.4 NFS

(1) コモディティクラスタ (Commodity clusters)

NSF のサポートする高性能バーチャルマシン（HPVM）のソフトウェアプロジェクトは、商用デスクトップコンピュータを高性能クラスタとして使用できるようにする。NSF が基金を供給している国立コンピュータ科学同盟（NCSA）と NPACI で研究している研究者たちは、ワークステーション 100 台規模の Windows NT のスーパークラスタ（Intel Pentium Ⅱ）を構築し、複雑な天体物理学の流体力学のコード ZEUS-M を数ギガフロップのスピードで実行させた。HPVM はユーザーが簡単にメッセージパッシングインターフェイス(MPI)コードを NT クラスタ環境に移植できるようにする。

(http://access.ncsa.uiuc.edu/Features/HPVM/HPVM.html)

２　ハードウェアコンポーネント

2.1 NSA

(1) ダイヤモンドをベースにしたマルチチップモジュール（MCMs: Multi-Chip Modules）

NSA の MARQUISE/SOLITAIRE プログラムはダイヤモンドをベースにした MCM と薄いフィルムスプレー冷却を使っている高性能コンピュータを再パッケージする。SOLITAIRE プロトタイプは、ダイラスト(die-last)MCMs 、ダイヤモンドの基板（substrates）、スプレー冷却を合体させる完成度の高い高密度パッケージ技術を実証する。

(2) マイクロスプレー冷却（Micro spray cooling）

NSA のスプレー冷却パワーコンバータ研究プログラムは高密度、高効率パワーコンバータを開発している。これにはスプレー冷却と平面的にラミネートされたトランスを使用している。現在の研究は、ダイヤモンドの基板上でのシリコンまたはガリウム砒素（GaAs）のデバイスのハイブリッド集積化で起こる基本的な熱膨張係数の不一致を克服するために、等温スプレー冷却を適用することに焦点が当てられている。

(3) 超伝導クロスバースイッチ（Superconductive crossbar switch）

NSA はポートごとに毎秒 2.5 ギガビット（Gbps）で動作する 128×128 超伝導クロスバースイッチを作成している。それは自動ルーティング（self-routing）で、32×32 から 1024×1024 を超えるサイズまでスケーラブルで、レーテンシーが４ナノ秒より短い。

クロスバーエレクトロニクスは絶対４度で動作し、その入出力ポートは室温にあり、低温装置が冷却器で冷やされているが、標準的な室温の環境で使用することができる。もっ

と高速でサイズが拡張されたならば、このスイッチはHTMTアーキテクチャで使用するためのエレメントになるだろう。

(4) 光電子（Optoelectronic）の研究

NSA は、光学技術に関する研究をサポートしている。その技術はデータのルーティング処理を促進する論理関数を提供することによって理論的に将来の通信とコンピューティングシステムに対し数百Gbpsのデータレートをサポートすることができる。NSAはまたチップや半導体の光学増幅器に関する高性能な分光計の研究を指揮している。

2.3 DOE

(1) 高性能ストレージシステム（HPSS: High Performance Storage System）

DOE 国立研究所コンソーシアムは、IBM と協力して HPSS を開発してきた。おおよそ 20 サイトで２年を超える活動と展開によって、HPSSは高性能コンピューティングの世界でストレージシステムの標準となった。HPSS4.1 は分散ファイルシステム、file families、MPI-IO、スケーラブル計算、パフォーマンス改善などのサポートのような新しい特徴を持っている。サンマイクロシステムズ社とストレージテクノロジー社が、1999 会計年度にHPSSコミュニティに参加した。

３　基礎/アルゴリズム研究

3.1 NSF

(1) コンピューティングの理論

NSFは以下に示す領域のコンピューティング理論の基礎的な研究をサポートしている。

(a) アプリケーション固有の理論は、分子生物学、通信ネットワーク、コンピュータ言語学のような科学や工学の問題解決のためのモデル開発やテクニック開発をサポートする。

(b) 核となる理論は、計算の複雑さ、暗号法、インタラクティブな計算、計算の学習、並列/分散処理、ランダムデータの計算、オンライン処理、知識に関する推論をカバーする。

(c) 基礎的なアルゴリズムは、組合せ、近似、並行処理、オンライン、数値計算、幾何

学、グラフ理論のアルゴリズムを含み、複数のアプリケーションの領域にまたがっている。

(2) 数値、シンボリック、幾何学の計算

NSF は計算とグラフィックスに関する基礎研究をサポートしている。その中にはコンピュータ代数、物理的なプロセスの数値計算とモデリング、数学的な最適化、計算幾何学、画像処理、計算論理学における推論の演繹的手法、自動推論を含んでいる。NSF はまた、計算科学と計算工学をサポートするために研究成果を問題解決環境に統合するサポートをしている。たとえば製造設計、検査支援システム、プロトタイプ作成、設計検証のような科学的工学的システムにおける最新の計算やグラフィック技術の革新的なアプリケーションがある。

(3) 並列処理と分散処理コンピューティングにおける NFS の研究

NSF の 1999 会計年度の HPCC に関する研究は、並列処理モデル、科学コンピューティング用並列処理アルゴリズムとソフトウェア、動的コンパイルと並列コンパイルの最適化、分散処理用オペレーティングシステム、スーパースカラーアーキテクチャ、高性能メモリシステム、信号処理と通信システムの研究があり、特に無線情報処理技術とネットワークについて研究されている。2000 会計年度の研究計画には無線センター、ハードウェアとソフトウェアの共同設計、高性能科学用や商用アプリケーションが入っている。

(4) データマイニング (Data mining)

NCSA の最高のデータマイニングツールの中には、マシン学習アルゴリズムがある。このアルゴリズムは、プログラムが例から学習する反復処理に依存している。マシン学習アルゴリズムは戦術的戦略的意思決定に利用することができる。

また、天文学のデジタル天体観測や高エネルギー物理学などでは莫大な数の属性によって表現されているデータの中から関係するサブセットを特定する機能の研究が必要とされており、データをクラスタリングするためのスケーラブルな統計学的アルゴリズムがカリフォルニア大学デービス校の NPACI の中で開発されている。

3.2 DOE

(1) Whisker Weaving (WW)

DOE のサンディア国立研究所 (SNL) の数学的コンピュータ計算科学における研究は現在 WW という構造化されていない3-D 六面体のメッシュを生成するためのアルゴリズムとメッシュの最適化に焦点を当てている。メッシュに要する時間が、DOE の ASCI (Accelerated Strategic Computing Initiative) 計画の自動設計機能やその他の大きなプログラムを開発する上において現時点での最も大きな障害であるから、WW とメッシュ最適化の２つの潜在的インパクトは重要である。

(2) 大規模科学的工学的設計の最適化

SNL は非線形な最適化のためのオブジェクト指向ライブラリ OPT++を開発してきた。このライブラリには非束縛型と境界束縛型の両方の最適化に有効な様々な幅広い手法、計算化学問題に関する大規模な束縛型最適化のアルゴリズム、迅速強固な並列処理最適化に関する新しい手法が入っている。

４　ソフトウェア

4.1 DARPA, DOE, NSF

(1) Globus プロジェクト

DARPA、DOE、NSF によってサポートされている Globus プロジェクトは、アプリケーションが数十～数百の地理的に分散した計算に使用するリソースや情報、装置、ディスプレイをまとめるための計算グリッドと実行環境を構築するための基礎技術を開発している。南カリフォルニア大学の情報科学研究所とアルゴンヌ国立研究所 (ANL) から参加している Globus チームは、広範囲なアプリケーションのデモで SC98 の高性能コンピューティング (HPC) 努力賞を勝ち取った。このチームは Globus ユビキタススーパーコンピューティングテストベッド機構 (GUSTO) のテストベッドを示した。 (http://www-fp.globus.org/)

4.2 NSF, NIH

(1) コンピュータシミュレーションによる酵素の呼吸動作の検証

NSF や NIH がサポートしている研究者たちは、人体中の最も速くて最も効率的な酵素の一つにまつわる神秘について、強力なスーパーコンピュータを使って解き明かそうとしている。アセチルコリンエステラーゼ（AChE）は、神経伝達物資であるアセチルコリン（ACh）を分解するような化学反応を促進させることによって 1 つの神経細胞から次の神経細胞または筋肉細胞への伝達を直ちに停止するように働く。NSF がサポートしている NPACI で走っている大規模コンピュータシミュレーションを使って､研究者たちは AChE が｢呼吸する｣ことによってそのことを起こしているということを示した。

4.3 NSF

(1) 次世代ソフトウェア(NGS)プログラム

NSF の NGS プログラムは、設計、管理、コンピューティングシステムの制御のための性能工学技術を開発するために、計算グリッドや将来のペタフロッププラットフォームのような複合的なコンピューティングプラットフォーム上で実行する複合的なアプリケーションに対し実行時サポートを提供するための新しいシステムソフトウェアアーキテクチャを作成するために、研究を助成している。

(2) オペレーティングシステムとコンパイラ

オペレーティングシステムと分散システムに関する NSF の研究には次のものがある。ローカルエリアネットワーク（LAN）とワイドエリアネットワーク（WAN）のリソースの一様なアクセスと管理に対するメカニズムやアプリケーションプログラミングインターフェイス（API）。拡張可能なサービスを構築するためのミドルウェア基盤。新しいアプリケーションに対するリソース管理とサービスの質の要件。セキュリティと電子商取引。コンパイラやランタイムシステムに関する研究には次のものがある。動的コンパイル、記憶装置の一貫性と記憶階層効率のモデル、ワールドワイドウェブ上でのプログラミング用のコンパイラのサポート。

(3) ソフトウェア工学と言語

NSF は質の高いソフトウェアをベースにしたシステムの開発を行なうための基礎研究を

サポートする。これには、仕様と設計用の領域を特定した言語、ソフトウェアの設計と進化への構成的アプローチ、ソフトウェアのモジュール化と構成の問題、信頼と品質の強化、ソフトウェア開発の自動化のステージ、分散開発とソフトウェアのセキュリティを含む分散とネットワーク環境の問題、ソフトウェア工学とプログラミング言語のあらゆる面に関する公式的な基礎がある。

(4) 科学計算用ポータブル拡張可能ツールキット (PETSc)

PETSc は偏微分方程式による物理学現象の大規模シミュレーションのための一連の相互運用可能なソフトウェアコンポーネントである。PETSc には、データ構造と C、C++、Fortran で書かれたアプリケーションコードで使用される数値計算のカーネルの拡張セットがある。

PETSc 2.0 はすべてのメッセージパッシング通信用に MPI を使用し、ワークステーションのネットワークから大規模並列処理プロセッサまで持ち運べるようにする。PETSc プログラミングモデルはまた非均等メモリアクセス (NUMA) の共有メモリマシンに最も適している。

PETSc ソフトウェアの周りに構築された ANL-led マルチサイト計算科学プロジェクトは、空気力学や音響学 (NSF の多くの専門にまたがるチャレンジプログラム) 分野でのマルチモデル、マルチドメイン計算手法に関する研究を含んでいる。それはオールドドミニヨン大学、クーラント研究所、コロラド大学ボールダー、ノートルダム大学、ボーイングコンピュータサービスにおいて共同で行われている。

(5) 情報をベースにしたコンピューティング

発見や検索を自動化しているスーパーコンピュータのデータ管理技術は NPACI のデータ集約型コンピューティング環境に配置されつつある。インフラはアーカイブストレージシステム (HPSS)、データ移動システム (SDSC Storage Resource Broker)、メタデータカタログ (SRSC MCAT)、データ発見サービス用のデジタルライブラリ、データ表示のための情報仲介者 (mediators) で構成されており、結果として、科学的データセットの出版、従来のテキストと科学的データ収集の統合を可能にし、画像のデジタルライブラリをサポートする。

4.4 NSA

(1) プログラミングの方法と言語

NSA は、大量に並行処理や分散処理を行なう異種のコンピューティングプラットフォームや特定用途のプロセッサ向けの計算手法や言語に関する研究をサポートする。研究にはコンパイラ AC（そのターゲットは SGI/Cray の T3D や T3E アーキテクチャ、さらにその他のアーキテクチャ）の開発も含まれている。また、研究にはメッセージパッシングプログラミングの「将来の」モデルのインプリメントも含まれる。

4.5 NIST

(1) Java を利用した数値計算

ネットワークをベースにしたコンピューティングのための Java 言語と環境の急速で広範囲での採用は、科学用のソフトウェアの開発をサポートするための信頼性のある再利用可能な数値を扱うソフトウェアコンポーネントの需要をおこした。NIST は Java Grande Forum と一緒に作業し、高性能な数値計算において Java を使用できるようにする。

4.6 DOE

(1) Aztec 反復ソルバー

SNL は、線形システムを解くための最新式の反復法のライブラリである Aztec の研究開発を行なっている。Aztec は、最先端を行く反復法を提供することで、また並列処理の反復法に関連した扱いにくいプログラミングタスクから開放することで、アプリケーションエンジニアの作業を楽にしてきた。DOE の ASCI のアプリケーションを含む内部使用に加えて、全世界に 100 を超える Aztec の外部ユーザーがいる。

(2) スケーラブル非構造メッシュアルゴリズムとアプリケーション（SUMAA3d）

SUMAA3d は、超並列プロセッサ（MPP）からワークステーションのネットワークにわたる広範囲のアーキテクチャ上でのスケーラビリティを強調した分散メモリ型コンピュータによる非構造メッシュ計算用の一連のソフトウェアツールである。アプリケーションには、圧電結晶（modeling piezoelectric crystals）、高温超伝導体、ひび割れた圧力容器のモ

デリングがある。SUMAA3d プロジェクトは DOE の ANL、ペンシルベニア州立大学、ブリティッシュコロンビア大学、バージニア工科大の研究者たちの共同プロジェクトである。

最近は、大規模な偏微分方程式(PDE)アプリケーション用の最新大規模統合計算環境（ALICE）との相互運用性に精力を注いできた。現在は、SUMAA3d とソルバーや可視化コンポーネントとのインターフェイスをとること、今までに得られた知識を一般のメッシュコンポーネントの設計仕様を作成するために利用することに精力を注いでいる。SUMAA3d と科学計算用ポータブル拡張ツールキット（PETSc）と BlocSolve95 の間の効率的なインターフェイスがすでに完成している。

（3）適応可能な数値計算手法に対する動的負荷分散とデータマイグレーション

SNL はアプリケーションコードの MPSalsa、ALEGRA、SIERRA 用の非構造グリッドの階層構造の適応可能なメッシュの詳細化に投資している。これらのプロジェクトはそれぞれ四分木（quadtree）または八分木（octree）のデータ構造を用いて、詳細化したメッシュを表現する。その関連した問題を扱うために、Zoltan という動的負荷分散ツールが幅広いアプリケーションに利用されるように開発されている。

（4）ACTS（Advanced Computational Testing and Simulation）ツールキットと Globus

科学用の装置をリモートのスーパーコンピュータや可視化装置と結合すると、擬似リアルタイム(時間ではなくむしろ分刻み)やリモートの解析、画像処理、制御により、装置の性能を変えることができる。DOE の ACTS ツールキットは Globus のツールキットを結合して、このような組合せを実現する。Brilliant Xrays は現在 ANL の APS のような設備で利用できるが、高解像度の 3-D 断層写真撮影装置のデータを記録することを画像当たり１秒より短い時間で実行可能なものにする。この技術は生物学や考古学のデータを画像処理したりシミュレートしたりするのに使われている。Globus ソフトウェアは、探知器、２次記憶装置、スーパーコンピュータ、リモートユーザーの間でデータを効率よく移動させる。（http://www.nersc.gov/ACTS/）

（5）最新大規模統合計算環境（ALICE）

ANL の ALICE プロジェクトのゴールは、高性能数値計算アプリケーションを構築するために独自に開発されたソフトウェアを使用する時の障壁を取り除くことである。そしてテラフロップ規模の計算用のリソースや結果としての新しい科学的な洞察を一般に広く活用することのできる基礎を築くことである。ALICE の開発は大規模科学的アプリケーションの設計の妥当性と実用性を保証するだろう。（http://www-unix.mcs.anl.gov/alice/）

(6) ALICE メモリ “Snooper” (AMS)

AMS は１つの API で、計算処理の探索、モニタリング、デバッグツールを書く時に支援してくれる。AMS は MPI を使用した並列処理アプリケーションをサポートするクライアント/サーバー用のマルチスレッド API の１つである。 (http://www-fp.mcs.anl.gov/ams/)

(7) 分散問題解決システム

ORNL の NetSolve は種類の違うコンピュータとソフトウェアライブラリを統一された簡単に使用可能な計算サービスに変えようとしている。ネットワークで緩く接続された数知れないコンピュータのハードウェアやソフトウェアのリソースを集め、なじみのあるクライアントインターフェイスを使ってそれらを結合した能力を利用することができる。そしてスーパーコンピューティングを通常のネットワークのプラットホーム上で幅広いユーザーにとって透過的に利用可能なものにする。 (http://www.cs.utk.edu/netsolve/)

(8) Collaborative User Migration, User Library for Visualization and Steering (CUMULVS)

CUMULVS 共同コンピュータによる探索ツールは、大規模な分散シミュレーションに対して、リモートの可視化、探索、異質のチェックポイント機能または再開機能を提供する。CUMULVS は国を跨いで NSF の NPACI、DOE や DoD のサイトで使用されている。

(http://www.epm.ornl.gov/cs/cumulvs.html)

(9) ALICE Differencing Engine (ADE) 可視化ツールキット

ANL の研究者たちは ADE 可視化ツールキットを開発している。科学者やエンジニアが、没入型、対話型バーチャルリアリティ環境の中でベクトルデータやスカラーデータを含む最大 10 個のデータセットの中の対のものに関する差異を可視化することができる。

このツールキットは、DOE と Air Products and Chemicals, Inc.の共同プロジェクトで新しいアルミニウムの炉を設計するのに使われてきた。炉の効率性に対するそれぞれの燃料の種類の影響を調べるために、空気燃料、空気または酸素の混合、純酸素燃料のデータセットが表示され解析された。

4.7 NOAA

(1) 海洋モデル用分散メモリソフトウェア

NOAA の新しい Modular Ocean Model バージョン３ (MOM3) はモデル物理学を改良し、初めて分散メモリ型スーパーコンピュータで走らせることができる。メッセージパッシングを使ったテストは NOAA の Geophysical Fluid Dynamics Laboratory (GFDL) の SGI/Cray T3E-900 と T90 スーパーコンピュータで行われた。継続している研究の領域は順圧方程式のよりよいスケーリング、およびモデルの実行中にデータを効率よく出力するための並列処理技術を開発することを含んでいる。後者は、NERSC と共同で扱われているが、これはまた南洋の高解像モデルに発展していく。 (http://www.gfdl.gov/MOM/MOM.html)

(2) 気象モデリングのスケーラブルなツール

NOAA の予報システム研究所(FSL)は、並列コンピューティングアーキテクチャで地球物理学の流体力学モデルを開発し移植するためのツールの拡張可能なモデリングシステム (SMS) の最新コンポーネントを公表してきた。SMS は実験用のパブリックドメインソフトウェアとして使用することができる。 (http://www-ad.fsl.noaa.gov/ac/sms.html)

(3) 海洋および気象データの並列入出力 (I/O)

NOAA の GFDL の科学者たちは、ローレンスバークレー国立研究所 (LBNL) の DOE の科学者たちと共同して、海洋の画像処理や気象学の研究者コミュニティで広く使用される共通データフォーマットである netCDF の並列 I/O のインプリメントを開発している。

4.8 DARPA

(1) 新しい fill-reducing オーダリングアルゴリズム

DARPA が基金を提供している ORNL の研究者たちは、ゼロックスパロアルト研究センターとテネシー大学と共同で direct sparse ソルバーを固有の反復並列処理および拡張性とに組み合わせることによって前提条件付け反復法に基づいた modular parallel sparse matrix ソルバーのファミリーを開発している。この事が防衛および産業のエンジニアに並列マシンを使って大規模な問題を効率よく解くことができるようにする。

4.9 EPA

(1) APOALA

EPA がサポートしている APOALA プロジェクトは、複雑で大規模な環境でのプロセスを解析するために、時間的地理学的な情報/可視化環境を統合する機能を開発している。プロトタイプは、時空 4 次元のデータ、仮想の時空 4 次元の問い合わせ構築言語、データの効率的な検索と操作の並行化、従来型探索型の 3-D データの時系列的解析を楽にする可視化、地球のデータに対する時空 4 次元の可視化手法と緊密に組みあわされたデータの探索と知識の発見機能、これらに対する新しいデータベースモデルをまとめることになるだろう。(http://www.geovista.psu.edu/apoala/index.htm)

(2) ParAgent

EPA のサポートを受けているアイオワ州立大学で開発された ParAgent は、知識ベースのアプローチに従った特定種類のプログラムを自動並列化処理する対話型ツールである。ParAgent はウェブベースの性能モニタリング、性能解析、可視化のツールを含んでいる。現在のバージョンは有限差分法(FDM)をベースとするプログラム用に設計されており、NASA のアメス研究所にある 200Mhz Pentium II を搭載したワークステーションの 64 個のノードを持つクラスタの上の並列処理コンピューティング用にインプリメントされてきた。

(3) VisAnalysis Systems Technology/Space-Time Toolkit (VAST/STT)

EPA がサポートするアラバマ大学ハンツビル (UAH) の VAST チームは、マルチソースデータのビジュアルインテグレーションのための Java ベースソフトウェアを開発している。この VAST/STT は本来の時空 4 次元領域のデータを受け入れる。特に、STT は、マイクロセカンドから世紀の単位の時間分解能で、ダイナミックな空間のデータ間の関係を調べる時に使用される。

STT は、ナッシュビルの Southern Oxidant Study (SOS) のために、1995 年の２ヶ月のデータの相互関係付けによってテストされている。そのデータは航空機や人工衛星のセンサー、気象学的環境的モデルの出力、縦形ウィンドプロファイラ、点源測量、ドップラレーダーボリュームから取得された。 (http://vast.uah.edu/)

4.10 NASA

(1) 科学的可視化のための並列ボリュームレンダリングシステム (ParVox)

NASA の分散ボリュームレンダリングシステム ParVox は、遅いネットワークや下位機種のワークステーションを使っている時でも科学者のデスクトップ上で使用できる、時間とともに変化する大きなデータセットを扱う分散型可視化に関するソリューションを提供する。

NASA は現在 ParVox の機能的なパイプライン処理に関して作業している。1999 会計年度の大きなマイルストーンは非構造化グリッド 4-D データセットに対するレンダリング機能を追加することである。 (http://alphabits.jpl.nasa.gov/ParVox/)

５ アプリケーション

5.1 NSF, NIH

(1) エネルギー精製を支援するモデル構築(AMBER)

NSF と NIH がサポートする AMBER のコンピュータによる化学プロジェクトは、蛋白質の折り畳みと２つの与えられたホストに対するリガンド（配位子）（または１つのリガンドに対する２つのホスト）、の相対的なバインディング自由エネルギーを研究し、蛋白質と核酸の順序に依存した安定性を調査し、様々な液体の中で異なる分子の相対的溶媒和自由エネルギーを発見するためのソフトウェアを開発した。

最近、研究者たちは、villin headpiece subdomain と呼ばれる 36 個の残留蛋白質で分子の最も長い分子力学シミュレーションを行なうために AMBER を使用した。このシミュレーションはピッツバーグスーパーコンピューティングセンターで SGI/Cray の T3D と SGI/Cray Research の T3E で 100 日を超えて行なわれた。このシミュレーションでは、完全なマイクロ秒の間分子を追跡した。 (http://www.amber.ucsf.edu/amber/)

図２　コンピュータで処理した折り畳み状態のスナップショット

5.2　NSF

（1）血流を維持するコード(NekTar)

血液とその他の複雑な流体の流れをモデル化するための新しいNekTarソフトウェアは、動脈硬化症のような動脈病の外科処置に変化をもたらすかもしれない。このコードは、かつては磁気共鳴画像診断装置(MRI)のような「静止画像」フォーマットでしか正確に捉えられなかった血流に対し、アニメーションによるシミュレーションを作成することができる。ブラウン大学の研究とNCSAのシミュレーションが、科学者がもっと正確に多くの種類の流体の流れをモデル化することができ、コンピュータによる探索ができるようにする。

(http://www.cfm.brown.edu/people/tcew/nektar.html)

（2）ニューロンモデルの作成とテスト

コンピュータ計算モデルはニューロンの動きとネットワークにおけるニューロン間の通信を記述する。モデルは、生物の組織や生きている実験材料を使って行なうことが大変難しいかまたは不可能な場合に仮想的な実験を行なうために使用することができる。コンピュータによる集中シミュレーションはより大きな数のニューロンを電気的特性と化学的特性の両方のますます複雑で現実感のある特性のモデル化へ組み入れることができる。NSFが基金を提供しているNPACIのニューロサイエンスの研究者たちは高性能コンピューティングとインフラ資源を使いこの作業を行なっている。

（3）人間の脳を理解するためのロードマップ

NSFが基金を提供しているNPACIのニューロサイエンティストたちは脳における構造と接続のロードマップを生成するツールを作成している。研究者たちは、MRI スキャンや低

温断面切断脳 －脳を凍結し薄くスライスしたもの－ や高倍率の顕微鏡で収集したデータを解析するソフトウェアを作成している。このソフトウェアは脳の 3-D モデルを再生成することができ、特定の脳とその他のものとを比較するためのロードマップを生成することができる。

(4) 触媒作用の抗体

ウイルス、バクテリア、その他の有害な侵入者が血流の中に入ってきた時、免疫システムは、抗体を作成することによってそれらを撃退する。つまりは、侵入者をしっかりとつかんではなさないような形状(ちょうど鍵にぴったりあった錠)をした蛋白質が有害となる機能を固定して動けなくする。もしこの微少な侵入者の 3-D 的特徴にマッチした蛋白質を製造するためのこの驚くべき機能を利用することができたらどうであろう? この「触媒作用の抗体」が、合理的な薬の設計、言い換えれば他の分子と相互作用するように彫刻された 3-D 的特徴をもつ分子の作成、という明るい見通しを製薬産業に提供する。

(5) 赤色巨星のパルスを診る

NSF が資金提供しているミネソタ大学と NCSA の天体物理学者たちは赤色巨星の 3-D シミュレーションを作成した。それは非常に詳細でそれが脈打つところを観察することができる。これは、これらの大きくて脈打っている星の中の照度の違い（その「標準的なあかり」は宇宙の距離を測るために天文学者が拠り所にしている）を説明する助けになるかもしれない。またそれが、科学者たちが、ついには赤色巨星になるはずの我々の太陽から何を予測するべきかを知る助けになるであろう。

(6) Biology WorkBench

NSF がサポートしている NCSA の Biology WorkBench は生物学者にとって革新的なウェブベースのツールである。この WorkBench によって生物学者は多くの一般に普及している蛋白質や核酸の一連のデータベースを検索することができる。この成果は最初に 1996 年にリリースされ、多くの機能強化がなされた新しいバージョンの 3.0 が最近リリースされた。

(http://bioweb.ncsa.uiuc.edu/educwb/)

5.3 NASA

(1) NASA グランドチャレンジ

現在、NASA のゴダード宇宙飛行センター (GSFC) のテストベッドには、NASA が行なっているプログラムの他にグランドチャレンジチームをサポートしている 1,024 のプロセッサがある。NASA の地球と宇宙科学(ESS)Round-2 チームは SGI のアナリストと協力して、Cray T3E の上で地球と宇宙科学モデルをインプリメントしている。

(http://sdcd.gsfc.nasa.gov/ESS/grand-challenges.html)

(2) NASA インテリジェント総合環境（ISE）デモンストレーション

ISE プログラムは、多様な専門知識や技術や関心を持った科学者や技術者、エンジニアが１つのチームとして機能することのできるようなツールやプロセスを研究し、開発する予定である。地理的に分散した、職業的に多様な人材の間で NASA のミッションを定義、設計、実行、運用することをネットワーク環境による共同作業として機能することによって、開発に使われる時間が縮小され、ライフサイクルのコストが減少することになるだろう。

5.4 NOAA

(1) 国立環境予想センター（NCEP）

NOAA の NCEP は、並列処理スーパーコンピュータに気象予想モデルをインプリメントするための手法を開発するために海軍研究所(NRL)と NASA の GSFC と共同作業してきた。NCEP はそのターゲットマシンのアーキテクチャには依存しない最新の調達物に対するベンチマークを提供した。

(2) ハリケーンの予測

1998 年のシーズンはいくつかの大きなハリケーンが襲ってきた。それらは合衆国の南東部の財産に大きなダメージを与え、中央アメリカには多くの死傷者がでた。GFDL のハリケーン予測システムは NOAA の NCEP によってこれらの嵐の通り道を予測するためにモデル化の一部として広範囲にわたって使用されていた。

(3) リアルタイム気象予想

緊急を要する気象予想における進歩は 1 月にテキサス州ダラスで行われた American

Meteorological Society (AMS) のミーティングで実証例示された。入力データはNOAAの新しいドップラーレーダーシステムより得られた詳細なデータストリームであった。それぞれの観測日のリアルタイムで高解像度の数値化された気象予想は、NSF が資金提供している NCSA のほとんどノンストップで稼動している 128 個のプロセッサを搭載したSGI/Cray のスーパーコンピュータ Origin 2000 の全能力を使ってリモートに作成された。

5.5 FAA

(1) 航空機に影響する気象予想

FAA のサポートを受けて、NOAA の FSL は Rapid Update Cycle (RUC-2) のモデルを改良した。これにより気象学者が、氷結条件、風、雲、晴天乱流などの航空機に影響を与える気象をより高精度に予想することができるように支援する。RUC-2 は１時間毎（最初のバージョンは３時間）に走り、高精度で、より詳細なデータを取得し、より精確な情報を作成する。

5.6 DOE

(1) 近似 Green 関数を使用した効果的電子移動のシミュレーション

ORNL の研究者たちは、「近似 Green 関数」の手法によって（境界積分法のアプリケーションの中に、Green 関数が計算回数を減らすことができるものがあるという手法の）メリットが拡張可能であることを実証してきた。彼らはこの新しいアルゴリズムを、マイクロエレクトロニクスのワイヤーがボイド（空隙）によって変形された材料科学の電子移動問題のモデリングで使用してきた。

5.7 NIH

(1) Neuroimaging Analysis Center (NAC)

NAC は NIH の国立研究資源センター (NCRR) によってサポートされており、基礎神経科学と臨床神経科学に関する画像処理と解析手法を開発する。NAC は概念や手法の普及を重要視している。NAC の技術はアルツハイマー病や脳の老化の研究、精神分裂症や精神分裂

型の病気の形態計測、多くの硬化症の定量解析、神経外科学における対話的画像ベースの計画とガイダンスに使用されている。

(2) Laboratory of Neuro Imaging Resource (LONIR)

NIH がサポートしている LONIR は、十分に多次元の複雑さで脳の構造や機能を調査するための新戦略を開発する。LONIR は、脳のモデルを作成、解析、可視化、相互作用するための種々のツールとともに国際的な共同開発者のネットワークを提供する。この共同作業の主な焦点は４－Ｄの脳のモデルを開発することである。脳の構造と機能の関係の調査を 4 次元に拡大して、脳のモデルが発達や病気の中でダイナミックに変化する脳の構造の複雑なパターンを追跡し解析する。

(3) 仮想細胞 (Virtual Cell)

NIH がサポートしている National Resource for Cell Analysis and Modeling(NRCAM)は、細胞生物学的なプロセスをモデル化するための一般的なコンピュータによるフレームワークである仮想細胞を開発している。この新しい技術は、それらの個別の反応を記述している生化学的や電気生理学的なデータと、それらの細胞内分布を記述している実験用顕微鏡の画像データを結び付ける。仮想細胞の現在のアプリケーションには、神経芽細胞と心筋細胞におけるカルシウムの力学の研究、オリゴ樹木細胞を取り扱う細胞内 RNA の研究がある。 (http://www.nrcam.uchc.edu/)

5.8 NIST

(1) 光（学）情報処理 (Optical information processing)

NIST は、640×480 ピクセル以上のイメージを持つ（たとえば 1,000 人の顔の）大きな参照セットに備えて、入力のイメージ（ビデオレートで）をテストする光学パターン認識システムを設計している。NIST は、高速光学パターン認識のアプリケーションにおいて必要とされる 2 サブシステムのリアルタイムビデオシステムのテストベッドバージョンを開発してきた。NIST はそのサブシステムとコンポーネントを将来の設計作業のために評価する。

(2) マイクロマグネティック・モデリング (Micromagnetic modeling)

NIST は精確で効率的なマイクロマグネティック計算用の計算ツールを開発している。このような計算は、磁気ディスクドライブでより高密度で高速の読書き時間を達成するには

必須である。初期のNISTがマイクロマグネティック計算用のソフトウェアが如何に信頼性がないかということを示す研究をしたとすれば、NISTは現在、徹底的にテストされ、実験で計測した結果と比較され、広く公で使用可能になるはずの参照コードを開発している。

(3) 高速マシニングプロセスのモデリング

現在NISTでは、高速マシニングの力学を研究するために結合された実験用および解析用プログラムが進行している。その目標は精確な計測を提供することと、高速マシニングオペレーションの有限要素法ソフトウェアの予想能力を評価しながら基本的な金属切削プロセスのより良好な科学的理解を得ることである。NISTは直交マシニングの基本的なプロセスの数学モデルを導き出した。そしてそのシミュレーション用の有限要素法ソフトウェアは、NISTのスーパーコンピュータ上で開発されインプリメントされた。

5.9 EPA

(1) 有害物質制御戦略の開発

ノースカロライナ州立大学の研究者たちの、意思決定支援システム(DSSs)のプロトタイプでは、最新の空気品質モデルにおいて「コマンドとコントロール」や排出権取引プログラムのような二者択一の管理戦略が設計されテストされた。プロトタイプのDSSsは、費用対効果の高いベンチマーク戦略を見極めたり、設計目標の中でのトレードオフを特徴づけたり、設計プロセスの中で不確実性を考慮にいれるために使用できる最適化コンポーネントを持っている。それらは異質で分散したコンピュータネットワーク上の高性能コンピューティングを通して実用になる。

付属資料３　SC99 参加報告

― SC99: High Performance Networking and Computing Conference ―

ハイエンドコンピューティング技術調査ワーキンググループ(HECC-WG)の活動の一環として、1999 年 11 月 13 日～19 日に、米国オレゴン州ポートランドにて開催された Supercomputing Conference の“SC99”に参加したので報告する。

この Supercomputing の国際会議は、1988 年の第 1 回(Supercomputing'88)から毎年(大体 11 月に)開催されており、1999 年は 12 回目にあたる。この会議の履歴については、Web にて公開されているので、参考にして頂きたい。(http://www.supercomp.org/)

ちなみに 2000 年の会議は、“SC2000”と呼ばれ、11 月 4 日～10 日に米テキサス州ダラスで開催される。(http://www.sc2000.org/)

会議名: SC99 ― High Performance Networking and Computing Conference ―
(http://www.sc99.org/)
日　程: 1999 年 11 月 13 日～19 日
場　所: Oregon Convention Center ― 米オレゴン州ポートランド ―
(http://www.oregoncc.org/)
参加者: 岡崎文夫、岡田恵太 (HECC-WG 幹事)

会場の Oregon Convention Center

１．概要

今回参加した両幹事は、このSupercomputing Conferenceに初めて参加するため、16日からのTechnical ProgramやExhibitionだけでなく、14,15日のTutorialから聴講した(13日は受付のみ)。16日からのTechnical Sessionでは、専門的なTechnical Paperよりは、招待講演やパネル討議の聴講に時間を割いた。また、Exhibitionはこれらの聴講の合間に見学することとした。

会場のOregon Convention Centerは大きく立派な施設で、参加者も3,000～4,000人はいたように感じる。我々が聴講した14,15日のTutorialは、70～100人規模の部屋に、40～60名ほどの参加者であったが､16日の基調講演や招待講演は数百人はいたのではないかと思う。米国のIT研究開発にかかわっている人口の多さやパワーを実感した。

聴講の合間に見学したExhibitionは、150くらいの展示がされており、企業と研究機関とでエリアが分かれていた。企業では、IBM・Compaq・Hitachi・NEC・Fujitsuなどが大きなブースで展示し、研究機関の中では日本の新情報処理開発機構(RWCP)が広い面積をもっており、大勢のスタッフで頑張っておられた。その他、日本からはRISTのGeoFEMプロジェクトや電総研、埼玉大学の展示もあった。SC99に参加しておられたHECC-WG委員の方々に話を伺ったところ、例年に比べると展示見学の人数が少なくパワーも足りないとのことであった。

例年、11月中旬のポートランドは雨季にあたるようだが、到着した13日は暑いくらいの陽気であった。その後の開催期間は、曇りがちで小雨がぱらつくことが多かったが、思ったほど寒くはなく(5～10℃くらいか)､持っていった厚手のコートを着なくてもすむくらいであった。

２．聴講リスト

○Tutorial(14,15日)

- Real World Scalable Parallel Computing
  (Lawrence Livermore National Laboratory, Sandia National Laboratories)
- High Performance Computing Facilities for the Next Millennium
  (Berkeley National Laboratory)
- How to Run a Beowulf
  (California Institute of Technology)

・ Computational Biology and High Performance Computing
(Lawrence Berkeley National Laboratory (NERSC))

○Keynote Address and Invited Speaker (16,17,18 日)

・ Managing High Performance Creativity
(Donna Shirley, U.Oklahoma and Managing Creativity)
・ 100 Gbps or Bust: Building Wide Area High Performance Networks
(Dona Crawford, Sandia National Laboratories)
・ Supercomputing on Windows NT Clusters: Experience and Future Directions
(Andrew A. Chien, University of California, San Diego)
・ Merging Advanced Microscopy with Advanced Computing
(Mark H. Ellisman, University of California, San Diego)
・ The ABC's of Large-scale Computing and Visualization: ASCI, Brains, Cardiology, and Combustion
(Chris Johnson, University of Utah)
・ Human-competitive Machine Intelligence by Means of Parallel Genetic Programming
(John R. Koza, Stanford Medical School and Third Millennium Venture Capital Limited)
・ Achieving Petaflops-scale Performance through a Synthesis of Advanced Device Technologies and Adaptive Latency Tolerant Architecture
(Thomas Sterling, Caltech and NASA Jet Propulsion Laboratory)

○Panel(16,17,18 日)

・ Experiences with Combining OpenMP and MPI
・ digital.revolution.com: Transforming Science and Engineering
・ Internet2 Status and Plans
・ Data Mining: The New Frontier for Supercomputing?
・ Community Model Building
・ It's the Software, Stupid: What We Really Need for Super Computing

３．基調/招待講演

16 日の朝にあった Donna Shirley 女史の基調講演(Managing High Performance Creativity)では、NASA の JPL での火星探索プログラムにおける彼女のマネージャ経験に基づいたものであった。火星に 6 輪探索機を送るという前例のない独創的なプロジェクトを成し遂げるには、実に様々な考慮が必要なことを、豊富な写真を使い説明された。

続いての招待講演(100 Gbps or Bust: Building Wide Area High Performance Networks)では、Dona Crawford 女史が DOE の ASCI 計画の説明を行ない、今後の高速ネットワークで 100Gbps を達成させるという内容であった。

University of California, San Diego の Andrew A. Chien 教授の講演(Supercomputing on Windows NT Clusters)では、2000 年 Blue book(http://www.hpcc.gov/pubs/blue00/)にも記述のある Commodity clusters、High Performance Virtual Machine (HPVM)についての説明があり、彼らの行なってきた NT クラスタ構築が紹介された。

４．パネル討議

16 日に行われた“digital.revolution.com: Transforming Science and Engineering”は、1999 年のイニシアチブ「Information Technology for the Twenty-First Century(IT^2)」への勧告となった、米国大統領に提出されたレポート「Information Technology Research : Investing in Our Future」を作成した大統領情報技術諮問委員会(PITAC)のメンバによるパネル討議であった。パネラは、共同委員長の 1 人である Ken Kennedy(Rice U)のほか、David Cooper(LLNL)、Susan Graham(UC Berkeley)、Larry Smarr(NCSA)、Joe Thompson (Mississippi State U) で、前記レポート作成に関する話題が提供された。米国の、IT 分野におけるフロントランナーを維持し続けるために選ばれたメンバによる、IT 研究開発への意気込みが感じられた。

- IT^2 イニシアチブ: http://www.hpcc.gov/it2/
- PITAC レポート: http://www.hpcc.gov/ac/report/

  ※上記の日本語訳は、当研究所の Web に掲載されているので参考にして頂きたい。(http://www.icot.or.jp/の『解説論文/論説/OHP 資料集』のページ)

５．その他

18 日夜の Reception は、オレゴン科学産業博物館(http://www.omsi.edu/)を会場として実施された。各国からの参加者の交流の場とポートランド観光の一部として企画されたものと思うが、博物館の中で展示品を見ながらの立食や飲食は珍しく、印象に残るものであった。

ハイエンドコンピューティング技術に関する調査研究Ⅰ
－ペタフロップスを中心にして－

© 平成12年3月発行

発行所　財団法人　日本情報処理開発協会
先端情報技術研究所
東京都港区芝２丁目３番３号
芝東京海上ビルディング４階
TEL（03）3456-2511