

解 説

● 分子科学のための電子計算機システムの条件†



柏 木 浩††

1. はじめに

分子科学研究所（分子研）の電子計算機センターは化学と生物学のための大型計算センターとして国際的にも例のない存在である。分子科学は人間の生活に密接した基礎科学の分野であるために数限りない問題が存在している。計算機を十分に使用できないために潜在的であった化学計算が分子研センターの発足により一挙に解放された感がある。これはある程度予測された事態ではあるが、発足後3年に満たないうちに早くも近い将来のより大規模でより経済的な計算機システムの構想を立てることが必要になった。そのために1981年の7月から9月にかけてアメリカとヨーロッパをまわり、科学計算の将来を考えてみようということになったわけである。この特集のスーパーコンピュータについては日本における一般の計算機利用者の関心はまだ薄いようである。8月末に北イングランドの古都チェスターで開かれたベクタプロセッサとパラレルプロセッサに関する学会には、数百人の計算機科学者、基礎科学の計算機利用者、計算機会社の技術者が集まり、CRAY-1、DAP、AP-190 Lなどのアーキテクチャ、計算処理能力、応用のためのアルゴリズムなど実際経験に基づいた研究が多数報告されている。帰国して間もないため未整理のままであるが、分子科学計算のための将来の計算機システムについて必要なことがらを検討してみたい。

2. 分子研電子計算機センターの現状

分子研センターの利用状況を見ると分子科学研究者の計算意欲がどれほど大きかったか、また現在の最高速のシーケンシャル計算機（HITAC M-200 H など）ではその意欲を満たし切れないことがはっきりする。

分子研電子計算機センターは分子科学の大規模理論計算、実験データの収集と解析、プログラムライブラリとデータベースの開発維持、基礎生物学研究所と生理学研究所の計算処理を目的としている。1979年1月に HITAC M-180 2台のマルチプロセッサシステムで運転を開始し、同じ年の9月には完全無人運転システムを完成させ¹⁾、1980年4月には1台の M-180 を M-200 H に置き換え、現在は図-1の LCMP (Loosely coupled multiprocessor) システムで運転を行っている。

このセンターは分子科学者の指導の下に設立され、理論化学者が直接設計に携わっている。単純労働なしで長時間ジョブを効率よく実行するための無人運転システムは一週間単位の連続運転を実現している。このシステムはジョブが少なくなったとき、または障害が発生した場合には自動的に運転を停止する。分子研センターはオープンセルフサービス方式を採用し、計算機が動いている限り自由に利用することができる。この方式をサポートするためのいくつかのソフトウェアも開発され利用されている^{1), 2)}。連続運転による長時間ジョブの実行の容易さ、数 GB のディスクの開放、オープンセルフサービスシステムなどのために大学のセンターでは実行し難い規模の計算が容易にできるようになっている。

利用者は約400名で大半が全国各地の大学及び研究所の分子科学研究者である。表-1に見られるように使用された CPU 時間の増加はきわめて急速である。55年度の CPU 時間の中約5,000時間が M-200 H で使用された時間で、一年が8,000余時間であることを考えると最大限に近い利用率に達している。研究者の願望に応えるためには近い将来に現在の10~100倍の能力を持った電子計算機システムが必要である。

3. 分子科学における大規模計算の特徴

分子研の計算機は多種類の計算に利用されているがここでは一番大きな比重を占めている非経験的な分子

† Necessary Conditions of the Computer System for Molecular Science by Hiroshi KASHIWAGI (Computer Center, Institute for Molecular Science).

†† 分子科学研究所電子計算機センター

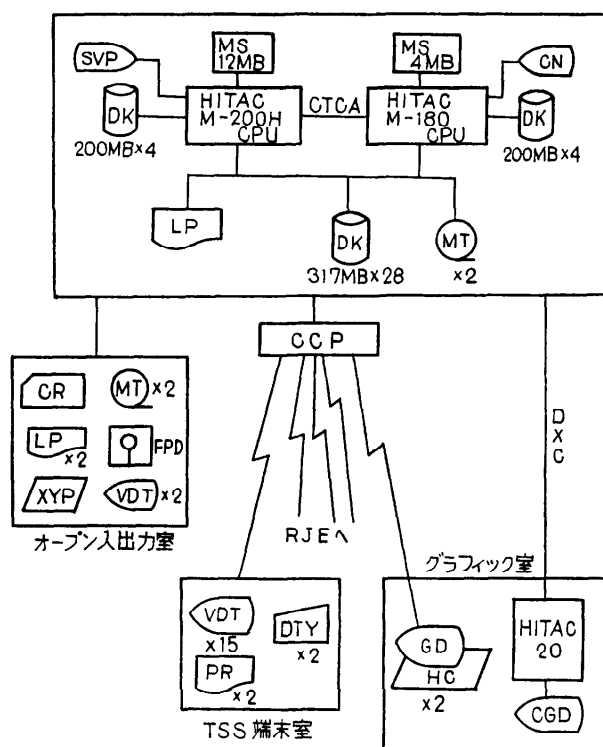


図-1 分子研センターのシステム概略図

表-1 分子研センターにおけるジョブ処理状況の推移

	ジョブ処理件数	CPU 時間
53 年度 (3 カ月)	41,521件	509H
54 年度	155,980	2,405
55 年度	183,840	5,405

(注) CPU 時間は M-200H の処理速度を基準としている。
(M200H : M180 = 2.7 : 1 として換算)

軌道計算 (ab initio MO 計算) について計算上の特徴を検討してみよう。この方法は分子の多電子波動関数を求め、分子の諸物性や化学反応の解析、予測などに有力な方法である。波動関数を展開するための原子軌道の数を N とすると $N(N+1)/2$ 次元の行列が必要である。スーパーマトリックスと呼ばれるもので分子軌道を決定するものである。電子相関を取り込むための標準的な方法としては配置間相互作用 (CI: Configuration Interaction) の方法がある。この方法では考慮する電子配置の数を次元とする行列の固有値問題を解く必要がある。ほとんどの計算は線形演算に帰着するので簡単そうに見えるが行列の次元の大きさが問題である。

たとえば、緑色植物の光合成の反応中心の計算を考えてみよう。光合成の初期過程は太陽光によるクロロフィルの励起と反応中心における電子の分離により開始される。このメカニズムを明らかにするために反応中心の波動関数を求めるとしよう。反応中心はクロロフィルダイマと考えられるので原子軌道の数 N はおよそ 10^3 になる。したがって、スーパーマトリックスの次元は $\sim 10^6$ 、要素の数は $\sim 10^{12}$ になる。この中大部分の要素はゼロに近いので、実際に処理しなければならない要素の数は 10^{10} ぐらいであろう。CI 計算の行列も同様のものを想定しておこう。スーパーマトリックスについては筆者自身が 10^8 程度の非ゼロ要素を実際に取り扱っている。CI 計算の行列についても欧米において次元が 10^5 を越える計算が実行されている。したがって、前述の規模の行列は近い将来の目標として過大なものではない。したがって、将来の計算機の必要条件は、

- (1) 10^8 次元のスパースな行列の演算を実際の時間でい。
- (2) 演算速度と釣り合った I/O 速度の大記憶を持ち。
- (3) プログラミングが容易である。

ことである。

4. CRAY 型のベクトルプロセッサ

1970 年代の中頃から世界最高速の電子計算機として CRAY-1 はスーパーコンピュータの代名詞になっている。他の計算機会社もパイプライン方式のベクトルプロセッサを開発し、スーパーコンピュータの時代が開花しようとしている。CRAY-1 は現在世界で 30 余台が稼動しているにすぎない。CRAY-1 を利用している分子科学者はきわめて限られている。アメリカにおいても CRAY-1 は原子力関係の研究などに置かれていて、わずかの分子科学者がそれを利用しているにすぎない。ヨーロッパではイギリスの Daresbury 研究所と西ドイツの Max Prank 研究所で CRAY-1 による分子科学計算が行われている。Daresbury 研究所はマンチェスターの近くにおいて Science and Engineering Research Council (SERC) に属する研究機関であり、日本の分子研や高エネルギー研究所と類似の役割をはたしている。Daresbury の CRAY-1 はイ

ギリスの大学間ネットワークの一つの核になっていて多数の科学者に開放されている。

今回の旅行目的の一つは CRAY-1 の使用経験を持つ欧米の分子科学者から CRAY-1 の分子科学計算への適応性を聞くことであった。Ab initio SCF MO CI 計算について CRAY-1 のベクタプロセッサの効果は次のように要約できる。

(1) シークエンシャル計算機用の従来のプログラムをそのまま走らせたのでは効果はほとんどない。

(2) 従来のプログラムパッケージの一部のサブルーチンを選択して DO LOOP のネスティングの変更を行うと、平均して 2 倍程度速くなる。

(3) プログラム全体のアルゴリズムをベクタプロセッサに適応させると 10 倍前後高速化する。

CRAY-1 の最高速度は約 150 MFLOPS で、それ自身のシークエンシャルプロセッサとしての速度の十数倍である。分子科学の活動的な研究グループはすでに FORTRAN で 10 万ステップぐらいのプログラムを持っているので、この財産を保存しながら CRAY-1 を利用すると(2)の場合の 2 倍程度の高速化しか得られない。より高速化するためには(3)の方針をとる必要があるが、相当に機種依存性の高いプログラムを作成しなければならない。ベクタプロセッサのユーザとその他の機種のユーザとのプログラムの交換は不利になる。ベクタプロセッサの時代がいつまで続くかという問題もある。

CPU の高速化にともなう問題はディスクなど周辺記憶装置の I/O 速度である。分子軌道計算では 3 MFLOPS の M-200H 計算機ですらかなりのジョブが I/O バウンドになっている。データ量の増大にともなってこの傾向は顕著になる。CPU 速度が数百 MFLOPS になっても I/O 速度が同じならばスループットは何ら改善されない。CRAY-1S という改良型ではオプションとして I/O プロセッサを付けることができる。主記憶とディスクの間に 8MB のバッファメモリを置いてディスク I/O を強化している。しかし、8MB を超えるデータセットを処理するジョブでは結局ディスクとバッファメモリの間の I/O がネックになる。

分子科学者の CRAY-1 への評価は大きく分かれる。分岐点は CRAY-1 がユーザに要求することを受け入れるか、そのような要求は過去及び未来への制約ととらえるかにある。もう一つ分岐点はそのユーザが多量データの I/O を必要としているかどうかにか

かっている。CRAY-1 のようなベクタプロセッサが科学技術計算の主力機になるかどうかはこれらの欠点をどこまで解決できるかにかかっている。

5. メガミニコンとアタッチド・アレイプロセッサ

アメリカで化学計算に使われている電子計算機は多様である。量子化学の分野ではメガミニコンを使用している研究グループが指導的な立場に立っている。この点が日本では見られないきわだった特徴である。アメリカでは大学などの大型計算機センターが必ずしも十分に大型でなく、かつ計算機使用料が高価であるため、コストパフォーマンスの良いメガミニコンを研究グループが所有して、自由に使いこなすのが有利な環境になっている。アメリカの計算機センターの第一印象はオペレータの数の多さである。CDC 6600, 7600 などが 1~2 台あるセンターには数十人から百人を超えるオペレータがいて、一日に数百回の磁気テープのマウントを行い、ラインプリンタの出力を手でちぎってユーザのボックスに振り分けている。このようなシステムでのサービスは硬直化しがちであり、使用料金は高価でユーザをメガミニコンの所有へと追いやる。

代表的なメガミニコンは VAX 11 とか HARRIS 4 などで名前の通り 1 MB 程度の主記憶を実装し、100 MB 前後のディスク記憶を持っているものが多い。スピードは 0.1~0.2 MFLOPS である。研究グループで所有している場合には利用の仕方が自由で長時間の使用ができ、研究の自由と独自性のためには有利である。しかし、一般にメモリが小さいこととスピードが遅いことのために大規模計算は不可能である。このためメガミニコンを使用しているグループの若い研究者には、分子研センターのユーザにとってはきわめて当り前の規模の計算も念頭にない。若い研究者の構想が小さな現実拘束されているのは将来の発展に望ましくない弱点である。

一方では計算機のグループ所有による研究という行き方に大きな力を与える計算機が急速に市場を拡大している。それは Floating Point Systems (FPS) 社の AP-120 B, AP-190 L などで代表されるアタッチド・アレイプロセッサである。計算速度はおおよそ 2 MFL-OPS で HITAC M-200 H や FACOM M-200 に匹敵するスピードである。これらはホストコンピュータを必要とするがコンパクトで(重さは人間一人の体重程度)価格は数千万円である。AP-120 B は VAX

11 や PDP 11 に接続可能。AP-190 L は IBM 370/168 などに接続できる。コストパフォーマンスは汎用大型機や CRAY-1 に比較して 10~100 倍である。ミニコン並の費用で VAX 11 などに大型機並のスピードを持たせてしまうのである。将来のメモリの増加の可能性も考慮すれば、研究室に一台という普及型計算機として科学研究の主力になる可能性を持っている。

6. パラレルプロセッサ

既存のパラレル計算機としては ILLIAC とか DAP の名前を挙げられるが、ILLIAC はテスト機としてその役割を終えようとしているし、DAP も CRAY を超えるほどの性能を持っていない。新しいパラレルプロセッサの開発としてはアメリカの Denelcor 社の HEP とニューヨーク大学のウルトラコンピュータ・プロジェクトがある。HEP は最大 16 個のプロセッサと 16 個の主記憶からなるパラレル計算機で 1,000 MB の主記憶を持つことができ、どのプロセッサもすべての主記憶との間でデータのアクセスができるように設計されている。この開発プロジェクトはコーネル大学の物理学者と計算機センターと Denelcor 社の三者で推進されようとしている。これらのプロジェクトは将来の目標として 1,000 個の規模のプロセッサからなる計算機のアーキテクチャを開発することを意図しているようである。大規模でかつ経済的な将来の計算機はパラレルプロセッサの形を取らざるを得ないのではなからうか。

7. 知性的な電子計算機を求む

電子計算機を大いに利用している人の実感は「計算

機にこき使われている」である。現在の計算機はその計算処理速度とか記憶能力の割には感覚、知性の面でまったく未発達である。新しい種類の計算をしようとする時（研究のために計算しようとするれば普通はそうなる）、まずプログラムを書かなければならない。たとえば、FORTRAN、素朴にしてエラーをなくすることが不可能に近い言語。TSS コマンド、言語学とは無縁な技術者が場当たりで決めた暗号の集合。ジョブ制御言語に至っては……。計算結果の取り出しはいまだにラインプリンタによる数値の打ち出しが標準である。原子軌道 10^3 個をベースとした分子軌道計算の場合、 10^3 個の固有値と、 $10^3 \times 10^3$ 個の固有ベクトルの成分の数値。ざっと一箱分のラインプリンタ用紙の山。いわゆるスーパーコンピュータについての暗いイメージである。

研究者の意志と思考に即応できる計算機は人間に近い感覚器官と知性を持つ必要がある。この問題は周辺機器の拡充では解決できない問題で、電子計算機中枢のアーキテクチャと不可分であろう。計算機を用いる研究の律速段階は、現在の計算機でもプログラムの製作と出力結果からのデータの選択にある。現在の大型機より 10~100 倍の計算処理能力のある計算機ではこれらの手続きにかかる労力は致命的な欠陥になるであろう。人間に近い感覚と知性が次世代の計算機の必要条件の一つである。

参 考 文 献

- 1) 分子研電子計算機センター：センター・レポート，No. 1 (1980).
- 2) 同上，No. 2 (1981).