解　説

3. アーキテクチャ

3.3 数値処理マシン†

金田悠紀夫††

1. はじめに

多数の演算プロセッサを結合した高並列プロセッサシステムを実現して高速数値計算を行うことはコンピュータ・アーキテクチャ研究の夢であり ILLIAC Ⅳ システムを始めとして先駆的ないくつかの実例を上げることができる．しかしハードウェア技術が十分に発達していなかった時代においてはシステム規模の増大にともなう厚い壁につき当り十分な成果が得られずに難行する場合が多かった．しかし現在の VLSI 技術の進歩が新しい可能性を生みだし，VLSI 化された多数の高性能演算プロセッサを用いた高並列マシンの実現が試みられるようになってきた．第2の並列計算機システムのブーム到来ともいえる．すでに商用化されているシステム，完成して稼動しているシステム，提案・設計段階のシステムなど多く存在するが，ここでは二つのタイプに分けて解説する．

（1） 多目的型高並列計算機システムとも呼ぶべきもので高性能のマイクロプロセッサを高並列に結合することにより実現し，数十〜数百 MFLOPS (Mega Floating Point Operations Per Second) の性能を目指している．幅広い応用とコストパーフォーマンス向上を目的としている．

（2） 特殊目的型計算機システムで用途を特別なものにしぼり特別に設計した超高速演算プロセッサ多数を用いて構成する．現在のスーパコンピュータよりも10倍以上高性能な 10 GFLOPS や，100 GOPS (Giga Operations Per Second) の性能を出せる．

前者の例として BBN (Bolt Beranek & Newman) 社のバタフライ (Butterfly) パラレルプロセッサ[1]，IBM 社の RP 3[2)~4)]，筑波大学の PAX[5),6)] を取り上げる．後者の例として IBM 社の GF 11[4),7)] と富士通が開発した FFT マシン FX[8)] を取り上げる．

2. 多目的型高並列計算機システム

計算指向で処理時間の長い科学技術計算において一つのタスクをいくつかのサブタスクに分けて多数の演算プロセッサで並行して実行させることにより計算時間の大幅な短縮を目指すいわゆる MIMD（多重命令列・多重データ列）型のマシンで，幅広い応用に対して効率よく対応できることを目指して設計されている．

演算プロセッサとして16ビット系もしくは32ビット系の高性能マイクロプロセッサと数値演算用の付加プロセッサを用いているものが多く，プロセッサ間結合方式，メモリの共有方式に高速性，汎用性，ハードウェア量増大の抑制の観点から工夫がほどこされているのが特徴である．

2.1 バタフライパラレルプロセッサ[1)]

BBN 社が開発した高並列プロセッサで商用化されており **図-1** に示すプロセッサノード（1ボードコンピュータ）をバタフライスイッチと呼ぶ高性能のスイッチネットワークによって相互結合したシステムである．16 台のプロセッサノードから構成されるシステ

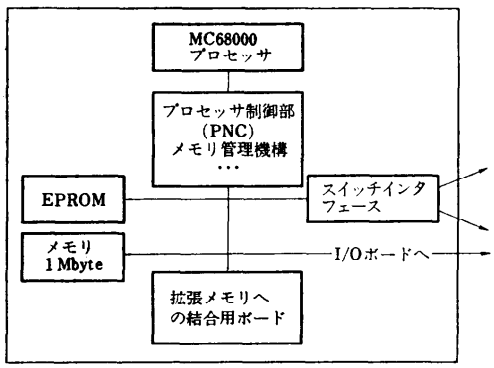
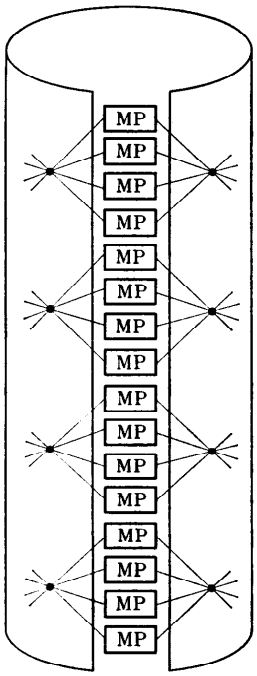


図-1 バタフライプロセッサ用プロセッサノード

† Parallel Machine for Numerical Calculation by Yukio KANEDA (Faculty of Systems Engineering, Kobe University).

†† 神戸大学工学部システム工学科



P：プロセッサ　M：メモリ

図-2 バタフライパラレルプロセッサの構成
(16プロセッサ)

ムの例を図-2 に示す．すべてのメモリはいずれかのプロセッサノードにローカルであるが，各プロセッサはマシン中の任意のメモリに対してバタフライスイッチを通じてリモート参照できる．1～256 セットのプロセッサノードがバタフライスイッチと呼ばれる多段スイッチにより完全結合されているのでプログラマから見たアクセス時間が若干長くなることを除けばリモート参照（約 5～6 μsec）もローカル参照（約 2 μsec）も差はない．

プロセッサノードはモトローラ社製 16 ビットマイクロプロセッサ MC 68000，1 M バイトの主メモリ（4 M バイトまで拡張可能），プロセッサ制御部（PNC），メモリ管理部，I/O バス，バタフライスイッチへのインタフェース部から構成されている．PNC はスイッチを介して転送されるメッセージの送受を取扱いプロセッサ間の同期と通信に必要なテスト & セット命令やスケジューラをサポートしている．プロセッサノードの処理能力は 0.5 MIPS 程度であるが，現在では 32 ビット系マイクロプロセッサ MC 68020 と MC 68881 浮動小数点演算付加プロセッサから成るボードが開発されている．

バタフライスイッチとスイッチ上でのパケットの伝搬の様子を 図-3 に示す．バタフライスイッチはパケットスイッチの技法を取入れて高速・高信頼・経済的なプロセッサ間通信を実現している．スイッチはノードの集りで“シリアルデシジョン”形のネットワークとなっており，その構成が高速フーリエ変換のバタフライ演算機構に似ているのでその名前が付けられている．図-3 の 16 プロセッサシステムでは各スイッチは 4×4 が 1 グループとなってカスタム LSI 化されてい

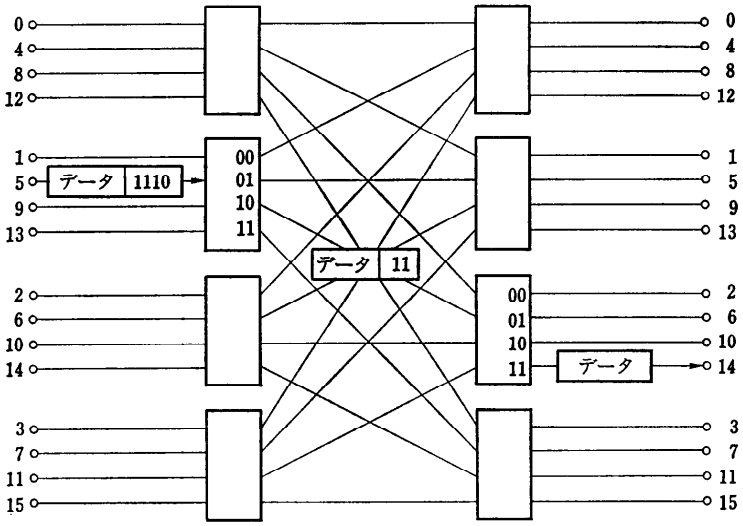


図-3 バタフライスイッチとパケットの伝搬

る．8つのチップが1枚のプリント基板にパッケージされていて16入力16出力のスイッチになっている．パケットに付加されたプロセッサアドレスに従って目的のプロセッサへルーティングされる．(図-3 に示すように付加プロセッサアドレスの値によって 4×4 のスイッチの切換えを行いルーティングされる．) 各プロセッサ間のデータ転送レイトは 32 Mbits/sec であり 16 台プロセッサシステムでは 512 Mbits/sec，64 台プロセッサシステムでは 2,048 M bits/sec となる．スイッチの素子数は N 台のプロセッサで $N \log_4 N$ のサイズになっておりクロスバーの場合の $\sim N^2$ に比して大幅に少ないといえる．

同一ノードへのパケットが重なり衝突が発生した場合には一方を一時的に遅らせて再転送するという機能を持っている．またプロセッサの台数が増加してきたときにはスイッチの信頼性向上のため全プロセッサペア間に複数の経路ができるように冗長なスイッチングノードを付加して対応している．

ソフトウェアはC言語で記述されており，そのほかに Lisp や Fortran も用いることができる．応用プログラムは Butterfly chrysalis と呼ぶオペレーティングシステムの制御下で動くが，ソフトウェアの開発はフロント・エンドマシンの DEC VAX または Sun ワークステーション上の 4.2 BSD UNIX (Tm) 上で開発される．したがって UNIX 上のツールを利用することができる．バタフライシステムはイーサネットまたはシリアルターミナルインタフェースでフロントエンドプロセッサに付加される．したがってソフトウェア開発はフロントエンドプロセッサとバタフライプロセッサとで相互通信しながら行うこととなる．性能であるが，400×400 の行列の積を実行した場合，128 台で並列計算を行うと1台の場合に比し120倍になるというデータが公表されている[1]．

2.2 RP 3 システム[2)-4)]

RP 3 (Research Parallel Processor Prototype) プロジェクトは IBM 社のワトソン研究センタで研究が進められており，高並列処理のハードウェアおよびソフトウェア研究の実験用機として開発されている．

このマシン開発の目的として

(1) 汎用のマイクロプロセッサ群を用いて多目的な並列処理を行うことの可能性の追求

(2) 並列処理の応用，言語，アルゴリズム，アーキテクチャの研究の道具とする

の二つが上げられている．

RP 3 の全体構成を 図-4 に示す．最大 512 台の汎用 32 ビットマイクロプロセッサまで拡張できるシステムで 64 台のマシンを当初開発することを計画している．スイッチを介してメモリを共有する形をとって

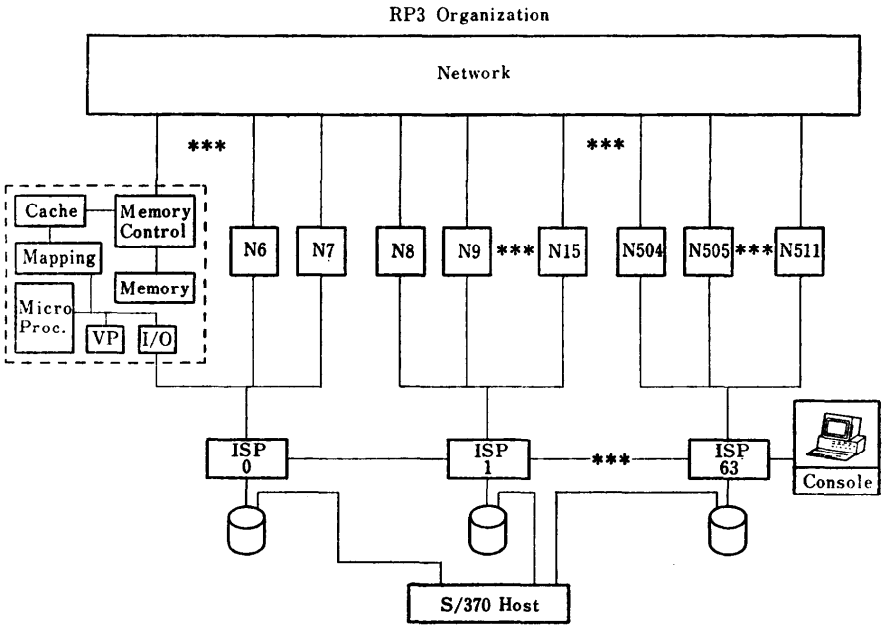


図-4 RP 3 システムの構成 (文献 4) より)

いる．512 台のシステムが完成したときに期待される性能を示すと以下のようになる．

・ピーク性能 1,300 MIPS，代表的な科学技術計算でキャッシュのヒット率を考慮に入れると約 1,000 MIPS となる．

・主メモリは 1〜2 G バイト，32 K バイトのキャッシュがプロセッサごとに実装．

・13 G バイト／秒のプロセッサ間通信ネットワーク．

・簡単なベクトルプロセッサを各処理ノードに付加することにより 800 MFLOPS の演算速度を実現する．

・メモリアクセス時間比はキャッシュ，ローカルメモリ，共有メモリで 1：10：16 を想定している．

・192 M バイト／秒のピーク I/O を 64 の独立な I/O チャネル群により実現する．

・ニューヨーク大学で開発中のウルトラコンピュータで採用されている[9] 高機能な“統合形”プロセッサ間スイッチを採用する．

・メモリ管理に柔軟性があり，ソフトウェアの制御で実行時にローカルメモリ，グローバルメモリの指定が可能である．

本システムでは全主メモリは各 PME (Processor-Memory Element) に含まれておりローカルメモリの部分とグローバルメモリの部分に分かれている．プロセッサからのメモリ参照はメモリマッピング処理（通常のセグメント，ページマッピング）を通ってキャッシュに入る．ページマップにはページやセグメントがキャッシュ化可能か否かの情報を持っている．すなわちプライベートなデータはキャッシュ化できるが共有のデータはキャッシュ化できない．これらの指定はプログラマがデータ宣言で行うことになるが動的に一時共有データをキャッシュ化することもプログラマにとって可能である．またローカルメモリとグローバルメモリの境界をプログラマが動的にプログラム実行時に変えることができるようになっておりきわめて柔軟性が高くなっている．

メモリ参照が他 PME ノードのメモリの場合にはそのアクセスは“パケット”としてメモリ制御部からネットワークへ送出される．参照結果はネットワークを通って元の PME に返送されてくる．

RP 3 には 2 組のネットワークがあり PME 間を結合している．一つはバイポーラ技術を用いた高速のスイッチでシステムのボトルネックにならないようにするため設けたものである．変形オメガネットワークとなっており 4-way のスイッチノード群から構成された 4 段のスイッチとなっており，各入口端一出口端間に二つずつの経路がある（図-5）．ネットワークの各ポートは 4 重に多重化されており 4 つのプロセッサが付加される．ネットワークは実際には 128 経路となっているので 512 プロセッサ構成となる．1 ポートが 4 プロセッサに多重化されているのはスイッチが高速であるからで 4 プロセッサを接続しても効率の低下が生じないためである．実際には二つのスイッチ網が存在しデッドロック発生の可能性をなくすため，それぞれを要求 (request) と応答 (replies) に用いている．

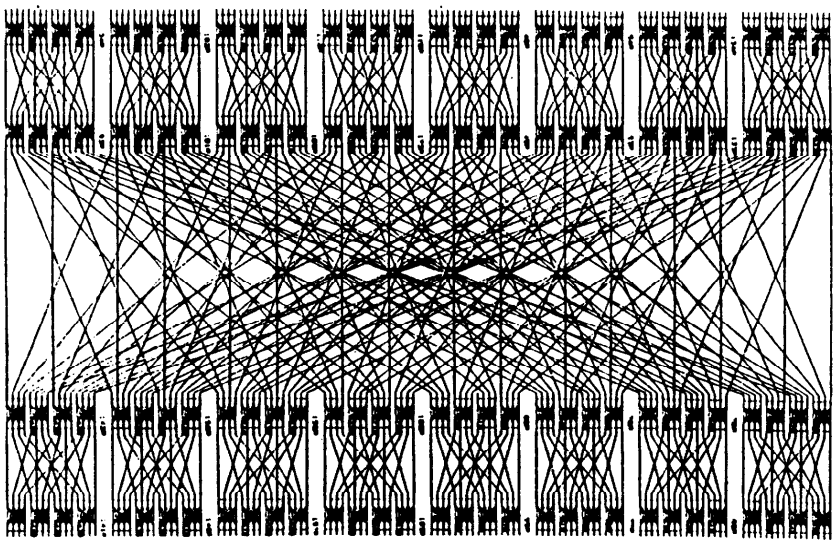


図-5 RP 3 の高速スイッチネットワーク（文献 2) より）

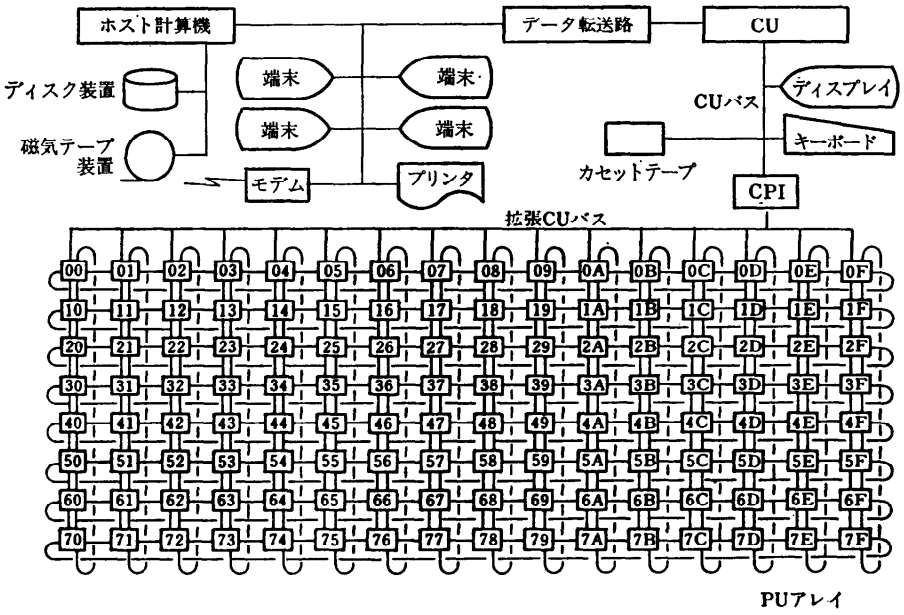


図-6 PAX-128 の全体構成（文献 5）より）

一方同期用ネットワークはより低速であるがより高集積な FET 技術を用いている．特徴として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で提案されているメモリアクセス要求統合化手法を採用している．この手法は，多数台のプロセッサが同期をとり合うため test and set 命令の拡張形命令である fetch-and-add 命令や fetch-and-op 命令[9]を同一メモリロケーション* に対してほぼ同時に実行したとき，これらの同一メモリロケーションへのアクセス要求をスイッチ上で統合し，統合が行われたという事実を各スイッチ・ノード上の待ちバッファに格納しておく．一つに統合されたアクセス要求はメモリに転送され，アクセスが実行され，その応答を返す際に各スイッチノード上で要求元のプロセッサに分配して返送するという手法である．これはメモリアクセスの衝突がわずかなパーセンテージでも存在すると大規模な並列計算機システムにおいてはシステム全体の性能が大幅にダウンするので，これを防止するためである[10]，

入出力は 8 PME ごとに一つの入出力サポートプロセッサ (ISP) が装備されており，ISP は S/370 へのチャネルインタフェースを持っている．

512 PME システムは物理的には 64 PME のサブシステム 8 台を直径約 35 フィートの円型にフロア上に配置することにより構成している．

オペレーティングシステムとしては UNIX (Tm) BSD 4.2 を予定しており，高並列オペレーションや共有メモリのサポートなどの付加機能を付けたものとなっている．プログラミング言語としては標準の言語である C, Fortran, PASCAL に共有変数宣言を加えたものを予定している．

2.3 筑波大学の PAX[5),6)]

筑波大学で開発された PAX は我が国における並列計算機システム研究のパイオニアとしての役割をはたしておりその成果は高く評価できる．

現在までに PAX-9, PAX-32, PAX-128 の 3 台が製作されており，並列処理を行う 2 次元直角格子状に配置された複数の PU（プロセッシングユニット），この PU 群を制御する CU（コントロールユニット），およびそれらのソースプログラムのコンパイルあるいはアセンブル，オブジェクトプログラムのロード，データの入力，計算結果の出力を行うホスト計算機からなる．ここでは最も新しい PAX-128 を取上げる．PAX-128 は名前が示すように PU 台数が 128 台で図-6 のように 2 次元格子状に結合されている．

PU は並列処理を実行するマイクロコンピュータで MPU として MC 68 B 00 を用いており，浮動小数点演算を行うために付加プロセッサ Am 9511 A-4 を用いている（図-7）．LM（ローカルメモリ）は各 PU

* 同期のためやロックのためにシステムが用意した共有データで，多数のプロセッサが頻繁にアクセスする可能性が多い．

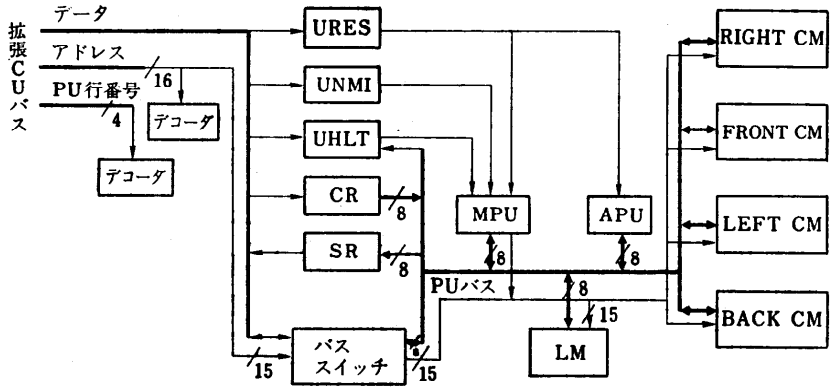


図-7 PU の構造 (文献 5) より)

表-1 PAX のハードウェア (文献 5) より)

	PAX-9	PAX-32	PAX-128
ホスト計算機	汎用計算機 M 150 F		TI 990/20
CU	マイクロコンピュータ コスモターミナル D		
PU プロセッサ	MC 6800	MC 6800	MC 68 B 00
算術演算ユニット	な し	Am 9511 A	Am 9511 A-4
PU アレイ規模	3×3	8×4	8×16
性能 (MFLOPS)	0.01	0.5	4
プログラム開発言語	アセンブラ	SPLM	SPLM
消費電力 (kW)		0.5	1.2
IC 数	57 個/PU	63/PU 個	97 個/PU
PU の大きさ (cm×cm)	20×20	31×31	18×31
PU アレイの大きさ	40×50×30	70×70×130	150×150×160
実装メモリ容量	5.25 KB/PU	18 KB/PU	32 KB/PU

固有のメモリで，各 PU のデータおよびプログラムを格納する．したがって各 PU は独自にプログラムを実行できる．CM (通信用メモリ) は前後左右の隣接 PU との通信用メモリで，互いにデータ交換ができる．

PAX は偏微分方程式系の近接作用を特徴とする分布系の問題とアルゴリズムを念頭において設計されており，Odd-Even SOR 法による2次元ポアソン方程式の求解などが最も適した応用といえる．このほか文献 5) には多くの科学技術計算への適用例が上げられており大変興味深い．

現在の PAX は浮動小数点演算ユニットとして 0.03 MFLOPS 程度の Am 9511 A-4 を用いているため高速な並列処理システムとはいえない．AMD 社や Weitek 社は 10 MFLOPS 近い超高速の素子を提供し始めている．PAX 様の構成に上記超高速の素子を採用した形態のマシンの実現が望まれる．

3. 超高速の特殊目的型並列プロセッサ

ここでは応用を単一にしぼり徹底的に高速化を図ることにより 10 GFLOPS 以上や 100 GOPS (Giga Operations Per Second) 以上の計算性能を出している超高速の特殊目的用並列プロセッサの例を上げる．前者は IBM 社が開発を進めている GF 11 システムであり，後者は富士通が開発した専用ディジタル信号処理装置 FX である．

3.1 GF 11 システム[4),7)]

物理学上の量子色力学 (QCD : Quantum Chromo Dynamics) と呼ばれる問題を数値計算により解くために設計されたスーパコンピュータで，ピーク性能で 11.5 GFLOPS，平均性能で 10 GFLOPS を出すことを目指している．本マシンは SIMD 型でそれぞれが 20 MFLOPS の演算能力を持つ要素プロセッサ 576 (通常 64 は予備プロセッサとなっている) を多段スイッチネットワークで結合した構成になっている (図-8)．

QCD における計算量はたとえば陽子・中性子などの質量を求めるのに 3×10^{17} 回の演算が必要で，この計算は 100 MFLOPS のスーパコンピュータでも 100 年かかることになる．GF 11 はこれを 1 年程度に短縮することを目指しており，1 年間は故障なしで動き続けるよう予備プロセッサを付加するなどして信頼性の向上を図っている．

本マシンは SIMD 型で中央のコントローラが制御 RAM から 50 ns ごとに並列型マイクロコード (180 ビット幅) を読み出し各要素プロセッサに放送して制

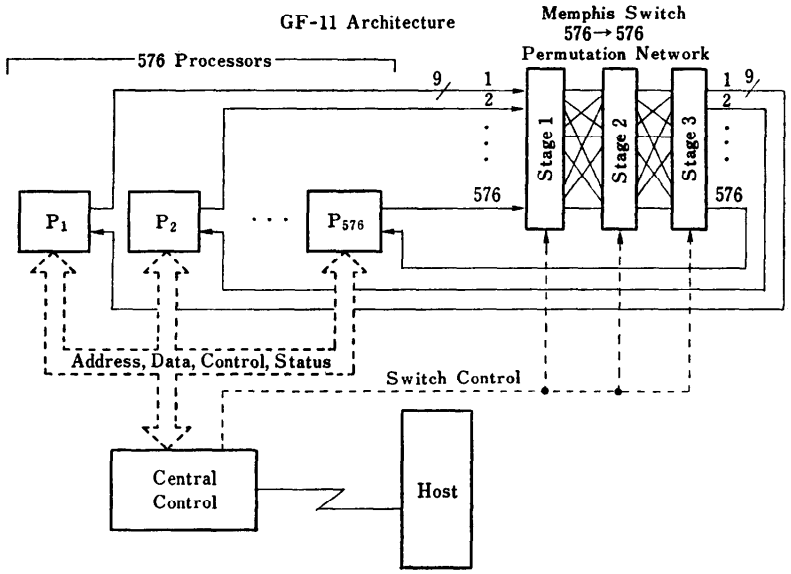


図-8 GF 11 のアーキテクチャ（文献 4）より）

御する．

QCD 計算を 8×8×8 台のプロセッサで計算しており，3次元の最短近傍の相互結合がされている．本システムではこれを3段の交替型ネットワークで実現しており，全体で 11.5 G バイト／秒のバンド幅を持っている．スイッチ要素は 24×24 のクロスバー・スイッチで 24 個並んで1段を構成し，これが3段結線されている．要素プロセッサにはウェイテック社の 32 ビット浮動小数点演算チップ群を用いており乗算器2個，ALU 2個で構成されており，それぞれが 5 MFLOPS の演算速度を持つ．さらにサイクル時間 12.5 nsec 256 語のスクラッチパッドメモリ，64 k バイトの高速スタティック RAM，256 k バイト～2 M バイトの DRAM から構成されている．

3.2 超高速並列ディジタル信号処理装置 FX[8]

富士通が東京天文台野辺山宇宙電波観測所における5基のアンテナからなる開口合成型電波望遠鏡システムのために開発した専用ディジタル信号処理装置で，高並列パイプラインアーキテクチャを採用しており，総合演算能力は 120 GOPS にも相当し，最大 320 MHz の受信電波を実時間で 1,024 点数に分光し相関計算を行う．システムはフーリエ変換部（下部），相関部（X部），制御部から構成されている（図-9）．

1,024 点複素フーリエ変換を連続的に 3.2 μsec ご

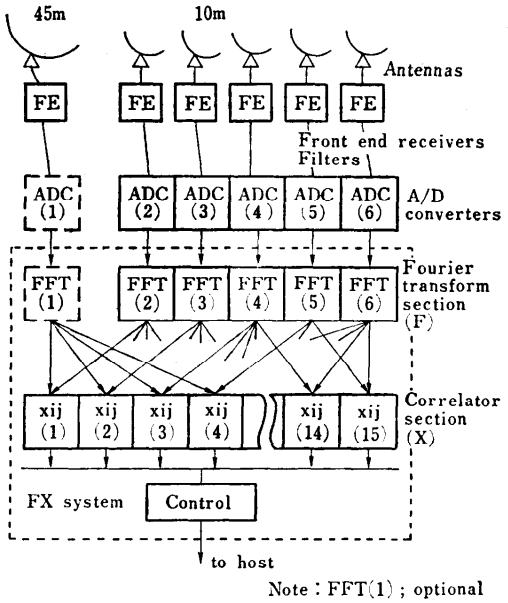


図-9 FX システムの全体構成（文献 8）より）

とに行うため FFT アルゴリズムを直接にハードウェアにマッピングしている．システム全体が長いパイプラインでデータは一方向へのみ流れる．

一般に N 点の離散フーリエ変換は $N=P\times Q$（P,Q は自然数）と分解できるとき，

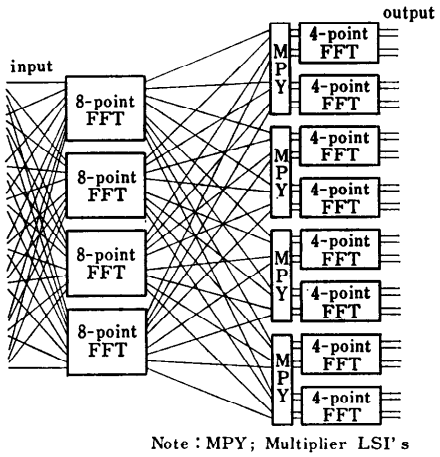


図-10 32点 FFT 回路 (文献 8) より)

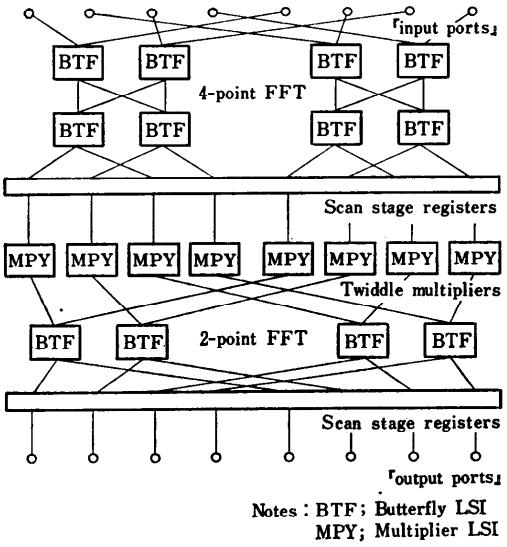


図-11 8点 FFT 回路 (文献 8) より)

ステップ1：Q 回の P 点 FFT

ステップ2：ステップ1の N 個の出力に対する位相回転 (ひねり係数乗算)

ステップ3：P 回の Q 点 FFT

に分解できる．この分解を再帰的に適用して，もとの N 点 FFT を多数回の小さな点数の FFT とひねり係数乗算に帰着させることができる．

1,024 点の分解を

$$1,024=32\times32=(8\times4)\times(8\times4)$$
$$=((4\times2)\times4)\times((4\times2)\times4)$$

としている．

第1段の分解は $1,024=32\times32$ であるから前述のステップ1とステップ3はいずれも32点 FFT となる．

32点 FFT は図-10 に示すように8点 FFT，位相回転，4点 FFT となり，8点 FFT は図-11 のように分解され，結局2点離散フーリエ変換回路バタフライ LSI (BTF) と乗算 LSI (MPY) を中心として構成される．また相関計算の複素共役乗算もパイプライン的に実行される．

本マシンは 3,700 個の CMOS LSI を使用した約 1,200 万ゲートのディジタルシステムである．計算精度が 6〜8 ビットであることも高速化に寄与しているが，商用のスーパコンピュータの FACOM VP-200 の 60 倍の速さで 1,024 点複素 FFT を実行しており，コンピュータと呼ぶより超高速専用電子回路とも呼ぶべきかもしれないが，超高速専用マシンの実例として興味深い．

4. ま と め

主として数値計算を対象として設計された高並列計算機システムについて論じてきた．VLSI 技術の持つ大きな可能性を生かす道として高並列計算機システムが上げられることは衆目の一致しているところで，多くの実験的な高並列システムの研究が強力に推進されている．

多目的型高並列計算機システムはコストパーフォーマンスの向上と処理性能の飛躍的向上とを目的として上げることができる．コストパーフォーマンスに関してはバタフライプロセッサを始めとして本稿では紹介しなかったが高性能な 16 ビットもしくは 32 ビットマイクロプロセッサ 10〜20 セットを高速共通バスに結合してマルチプロセッサを構成し，OS として UNIX をサポートしているものなどが市販されており，今後一段と普及してきて，中型・大型のコンピュータの強敵となる可能性が出てきている (Balance™ 8000 など)．

一方 RP 3 や PAX など数百台以上のプロセッサの結合を想定したものは現在のところまだ先進的な実験機の色あいが濃く，いかに多くの応用に効率よく適応させうるかが今後の研究課題であろう．その成否が決まるのにはもう少し年月が必要と考えられる．この場合プロセッサ群とメモリ群間，プロセッサ相互間の結合方式をいかにするかがキーポイントになる．現在のところポート当り数 M ビット/秒程度の高速転送能力を持つビットシリアル型の多段ネットワークが一つの有力な解と考えられる．

一方特殊目的専用マシンとしては 10 GFLOPS 以上，100 GOPS 以上という現在のスーパコンピュータに比しても 10 倍以上の性能を持つ専用マシンが大変に興味深い．本稿では紹介していないがプリンストン大学では最高速度 61.4 GFLOPS のスーパコンピュータ[11]を 32 ビット浮動小数点演算プロセッサ Am 29325 を最大 24 個使った 480 MFLOPS の処理能力を持つノードを 128 台コズミック・キューブ・ネットワークで結合したシステムとして開発を始めている．また同じく米国の Floating Point Systems, Inc. は 16 MFLOPS の単体プロセッサを最大 16,384 台をハイパーキューブ形で接続して処理速度 262 GFLOPS の性能を出せる T/40000 スーパコンピュータを開発している[12]．

現在のスーパコンピュータの能力をもってしても実用的な時間ではとうてい計算できない重要な問題はいくらでもあるので将来は TFLOPS (兆 FLOPS) 級マシンも実現してくるものと考えられる．今後の発展が大いに楽しみである．

参 考 文 献

1) Butterfly (Tm) Parallel Processor Overview, BBN Laboratories Incorporated (1985).
2) Pfister, G. F.: The Architecture of the IBM Research Parallel Processor Prototype (RP3), Proc. of 19 th HICSS, pp. 214-221 (1986).
3) Pfister, G. F. et al.: The IBM Research Parallel Processor Prototype (RP 3): Introduction and Architecture, Proc. of the 1985 International Conf. on Parallel Processing, pp. 764-789 (1985).
4) IBM Science Forum: 先進的計算機のアーキテクチャー (1985).
5) 星野：PAX コンピューター高並列処理と科学計算一，オーム社 (1985).
6) Hoshino, T. et al.: Highly Parallel Processor Array "PAX" for Wide Scientific Applications, 1983 Int. Conf. on Parallel Processing, pp. 95-105 (1983).
7) Beetem, J. et al.: The GF 11 Supercomputer, Proc. The 12th Annual International Symposium on Computer Architecture, pp. 108-115 (1985).
8) 中水，近田：超高速並列ディジタル信号処理装置，情報処理学会計算機アーキテクチャ研究会資料，59-1, pp. 1-8 (1985).
9) Gottlieb, A. et al.: The NYU Ultracomputer —Designing an MIMD Shared Memory Parallel Computer, IEEE Trans. Computer., Vol. C-32. No. 2, pp. 175-189 (1983).
10) Pfister, G. F. et al.: Hot Spot Contention and Combining in Multistage Interconnection Networks, Proc. of the 1985 International Conf. in Parallel Processing, pp. 790-797 (1985).
11) Electronics 18 (1985).
12) 日経エレクトロニクス，1986. 5. 5 (No. 394).

(昭和 61 年 5 月 9 日受付)