

TM-0696

マルチPSIの基本構成と今後の展開

益田嘉直(主査), 内田俊一

March, 1989

©1989, ICOT

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03) 456-3191 ~ 5
Telex ICOT J32964

Institute for New Generation Computer Technology

「マルチ P S I の基本構成と今後の展開」

益田 嘉直* 内田 俊一**

*三菱電機㈱コンピュータ製作所

**勸新世代コンピュータ技術開発機構

〔内容梗概〕

第五世代コンピュータ・プロジェクトでは、1990年代の高度情報化社会で必要となる並列推論技術の確立を目指して並列推論マシンの研究開発を進めている。ここでは、本プロジェクトを推進している嶺新世代コンピュータ技術開発機構（ICOT）が中心となって開発したマルチP S Iと呼ぶ並列推論実験機の基本構成と今後の展開について述べる。

＜開発の背景＞

第五世代コンピュータ・プロジェクトは、通商産業省の情報化関連政策の一環として、1990年代に必要な知識情報処理システムの実現を目指し、1982年より開始された国家プロジェクトである。

本プロジェクトは、前期3年、中期4年、後期3年の10年計画で進めており、最終的な目標として論理型プログラミングをベースとする大規模な知識情報処理を効率よく高速に実行する並列推論マシンの研究開発がある。前期計画における並列推論マシンの研究では、主にマシン・アーキテクチャの研究に重点が置かれてきたが、その過程で論理プログラミングの並列実行に関するソフトウェア面での研究の重要性が認識されるようになってきた。そのため、本プロジェクトを推進している嶺新世代コンピュータ技術開発機構（ICOT: Institute for New Generation Computer Technology）では、本格的な並列推論マシンの開発に先駆けて、並列ソフトウェアの研究開発を主目的とした並列推論マシンのパイロットシステムとして、マルチP S Iシステムの開発を行った。

＜マルチP S Iの特長＞

マルチP S Iは、前期計画で開発した逐次型推論マシンP S IのCPUをLSI化により改良小型化したものを要素プロセッサとし、それらを専用の接続ネットワーク制御機構により最大64台まで接続した並列計算機（図1）であり、以下のような特長を持っている。

- (1) マルチP S Iは、並列論理型言語KL1（Kernel Language Version 1）を高速に実行するマルチプロセッサであり、並列推論OSであるPIMOSを搭載する。PIMOS（KL1言語で記述）は、マルチプロセッサ用OSではあるが、いわゆる分散OSではなく集中型単一OSである。すなわち、OSとその上で動作する応用プログラムに関する限り、対象となるマルチプロセッサ・システムは1台の計算機とみなされる。
- (2) 並列ソフトウェアの重要研究課題である負荷分散方式を考慮し、Processing Power Plane と呼ばれる仮想平面を用いる負荷分散方式をサポートするためのハードウェア機

構を備えている。この負荷分散方式は、実行負荷の各要素プロセッサに対する割付けを局所的な情報のみで動的に変更することで負荷の均衡を得る方式である。

- (3) 全要素プロセッサは二次元格子状に接続され、プロセッサ間通信を高速化するために全系同期のクロックで動作する。また、マンマシン・インタフェースとして最大4台のフロントエンドプロセッサ（P S I - II）が接続可能であり、入出力制御およびマルチ P S I 本体の監視と保守を行う機能がある。

< システムの概要 >

マルチ P S I は並列論理型言語 K L 1 を効率よく高速に実行するマルチプロセッサであり、並列 O S である P I M O S を搭載する並列推論マシンのパイロットシステムである。

(1) 並列論理型言語 K L 1 の高速実行

K L 1 言語は、G H C（Guarded Horn Clauses）と呼ばれる並列論理型言語をベースとしており、その言語体系は図2のようにになっている。各要素プロセッサは、隣接する4つの要素プロセッサとの間に通信チャンネルを持っており、K L 1 を効率的に処理するために開発された K L 1 - B という機械語命令を高速に実行する。K L 1 - B は、W A M（Warren Abstract Machine）と同様な体系を持つ抽象機械語命令であるが、ユニフィケーションを利用したプロセス間の同期メカニズム、並列実行されるゴールのスケジューリング、実時間ガベージコレクションなど、K L 1 特有の機能を持っている。さらに、プロセッサ間のユニフィケーションや、プロセッサに対するゴールの割当処理など並列プロセッサに必要な機能も K L 1 - B によって実現される。

この他に、図2における K L 1 - C はフラット G H C をベースとした中核言語のことであり、K L 1 - P は並列マシン上での負荷分散などを効率よく実行するためにプロセスの割り当てや優先度設定を行うプラグマ（pragma）と呼ばれるものである。また、K L 1 - U は K L 1 - C にモジュール化機能等を追加したユーザ言語のことである。

(2) 並列推論 O S である P I M O S 実装

P I M O S の役割は、計算資源（C P U 資源、メモリ資源、入出力機器）を管理し、ユーザプログラムの実行を効率的に行うための環境を提供することである。この役割の中でも、ユーザの過ちからシステム全体を、ひいてはユーザ自身を守ることが最も基本的な機能である。

P I M O S はマルチ P S I 上に実装され、その構成は図3に示すようにマルチ P S I 本体の上で動作する P I M O S 本体、フロントエンドプロセッサの上で動作する C S P 部分と F E P 部分から構成される。C S P は P I M O S の立上げ時点に、イニシャルブ

プログラムのローディングや、接続ネットワークのルーティングのためのパステーブルを設定のほか、要素プロセッサの診断や監視を行う。FEPはPIMOSの立上げが終わると、本体上のPIMOSと一本のストリームによる通信路により結合され、以後PIMOS本体に対する入出力管理機能を持つ要素プロセッサとして、ファイル入出力やウィンドウシステムを用いた対話インタフェースとして働く。FEPの機能の実装はSIMPOSを用い、ESPで記述されている。

(3) 負荷分散方式の実験システム

マルチPSIの開発目的の1つとして負荷分散方式の研究があげられる。この負荷分散方式は、処理能力が均一に分布した平面PPP (Processing Power Plane) を仮定し、動的に負荷の再分配を行う方式であり、近傍のPE間の局所的な負荷の情報のみで負荷の分散をおこなうことを特徴としている。

並列推論マシンのように大規模なマルチプロセッサ・システムにおいてシステム全体の処理能力を効果的に使用するためには、各PEの実行負荷を均衡させることが重要な課題である。一般に画像処理等の処理・データに高い規則性のある応用では、各PEの実行負荷は比較的容易に予測することができ、負荷の分配は予めプログラム中で記述することができる。しかし、知識情報処理等の処理・データ共に規則性の低い応用では、各PEに対する実行時の負荷を予測することは極めて難しい。

PPPを用いる方式では、初期状態で各PEに対して適当に負荷を分配し、実行時に隣接したPE同士の局所的な負荷の不均衡に応じて、動的な負荷の再分配をすることによって各PEに対する負荷を均一にする。マルチPSIでは、結合方式としては比較的単純な二次元格子型を採用し、上記の負荷分散方式をサポートするためのハードウェア機構を接続ネットワーク制御部に試験的に実装している。

<システム構成>

マルチPSIの要素プロセッサ1台分は、PSIのCPUをLSI化により改良小型化したCPU部、80Mバイトメモリ及び接続ネットワーク制御部から成り、これら8台分が1筐体に納められてマルチPSI本体を構成する。図1に示すようにマルチPSI本体は、2列並べて最大8筐体まで結合可能であり、本体とフロントエンド・プロセッサを接続することによりマルチPSIシステムが実現される。図4に64台の要素プロセッサを二次元格子状に接続した最大システムの構成を示すが、本体部は入出力機器を持たないため、システムの入出力処理はフロントエンド・プロセッサにより行われる。

要素プロセッサPEは、図5に示すLSI化された要素プロセッサ用プリント基板4枚とメモリのプリント基板4枚の計8枚のプリント基板から構成される。各要素プロセッサは、図6に示すようにLSI化により改良小型化したPSIのCPUに、接続ネットワー

ク制御及びメンテナンスバスとのインタフェースを備えたものである。P Eに含まれる論理回路のほとんどは、図5に示すように15石のLSIにより実現されている。そのうちの9石は8,000ゲート(CPU部)、2石は20,000ゲート(接続ネットワーク制御部)の集積度を持つCMOSゲートアレーを用いて作られている。

CPU部のデータ幅は40ビットであり、そのうちの8ビットはデータの属性を示すタグとして用いられる。CPU部は、タグの抽出・判定などのタグ操作用ハードウェアを持ち、KL1の高速実行に大きな役割を果たしている。表1にハードウェアの基本仕様、また図8にケーブル接続した様子(1筐体)を示す。

(1) 接続ネットワーク制御部

接続ネットワーク制御部は、図7に示すように隣接する4つのP Eとの間の通信チャネル(CH₀～CH₃)とCPUとのインタフェースからなる5つの通信路を、相互に接続するためのスイッチ回路を中心に構成されている。各々の通信チャネルは、10ビット幅のデータ(パリティビットを含む)の送受信を独立して行うことができ、その転送速度は1チャネル当たり5Mバイト/秒である。各通信チャネルの出力ポートには、ネットワークの混雑を緩和するための48バイトのFIFOバッファ(OB₀～OB₃)が備えられている。CPUとのインタフェースには、WB及びRBという4KバイトのFIFOバッファが備えられており、CPUは送出すべきパケットを順次WBに書き込み、1つのパケットが完全に書き込まれると自動的に送信が開始される。また、到着したパケットはRBに書き込まれ、CPUにパケット到着を示す割り込みが報告される。

スイッチ回路は、各通信チャネルから受信したパケット及びCPUがWBに書き込んだパケットのあて先に従って、パケットの送信またはRBへの書き込みを行う。また、送信先が競合しないパケットは独立に送信することができる。あて先に従った送信先の決定は、PT(Path Table)と呼ばれるテーブルの索引により行われる。

(2) メンテナンスバス機構

マルチPSIのようなマルチプロセッサ・システムにおいては、システムの立上げ、障害管理などを行うために、各P E個別にメンテナンスを行える手段が必要である。このシステムでは、接続ネットワーク制御用の基板内にメンテナンスバス用のインタフェース回路を搭載し、フロントエンド・プロセッサから各P E個別にメンテナンスできるようになっている。メンテナンスバスは、フロントエンド・プロセッサと各P E間を1本の非同期バスでいもづる式に接続したものであり、フロントエンド・プロセッサがマスタでP Eがスレーブであるような単一マスタ複数スレーブ方式のバスである。

このメンテナンスバスを通して、接続ネットワーク制御とは別にフロントエンド・プロセッサから各P Eの内部資源(レジスタ、バッファなど)の読出し/書き込み、各P Eへのブートストラップ、イニシャライズ、異常検出などが行える。

<今後の展開>

第五世代コンピュータ・プロジェクトにおける最終的な目標である並列推論マシンについてのハードウェア面では、本格的な大規模並列推論マシンのハードウェア試作が進行中である。また、ソフトウェア面では、マルチPSIによる64台規模の並列ハードウェア・システムの完成ならびにPIMOSの試作により、本格的な並列ソフトウェアの試作実験を行う上台が整って来ている。

並列応用ソフトウェアの本格的な研究開発については、今後PIMOSを実装するマルチPSIシステム上で行われるが、既に「詰め碁」、「詰め込みパズル」、「最適経路問題」ならびに「自然言語構文解析プログラム」などの試作により並列推論ソフトウェアに関する問題点の明確化や評価が開始されている。

負荷分散方式については、

(1) 負荷をどのように定義するか。

(2) 負荷調整によるバスターブルの書換え中、パケットの転送をどうするか。

等、解決しなければならない各種の課題があるが、今後も前述の負荷分散方式の検討評価が引続き行われ、並列推論マシンにおける負荷分散方式の確立が行われる予定である。

[参考文献]

- (1) 内田：並列推論マシン，人工知能学会誌 Vol.2, No.4, 1987.
- (2) 瀧ほか：Multi-PSI システムの概要，情報処理第32回全国大会，1986.
- (3) 武田ほか：マルチPSI における実験的負荷分散メカニズム，情報処理学会研究報告 Vol.88, No.79 (88-ARC-73), 1988.
- (4) 益田ほか：マルチPSI における接続ネットワークの試作と評価，情報処理学会論文誌，Vol.29, No.10, 1988.

図表一覧

- 図1. マルチP S Iシステムの外観
- 図2. 並列論理型言語K L 1の言語体系
- 図3. マルチP S Iに実装されるP I M O Sの構成
- 図4. システム構成
- 図5. L S I搭載の要素プロセッサ用プリント基板
- 図6. 要素プロセッサの構成
- 図7. 接続ネットワーク制御部の構成
- 図8. 本体部のケーブル接続（後面パネル開）

- 表1. マルチP S Iのハードウェア基本仕様



図1. マルチPSIシステムの外観

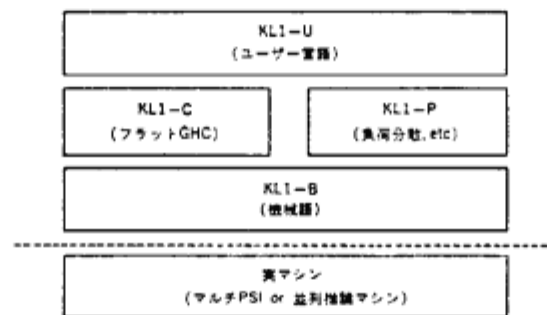


図2. 並列論理型言語KL1の言語体系

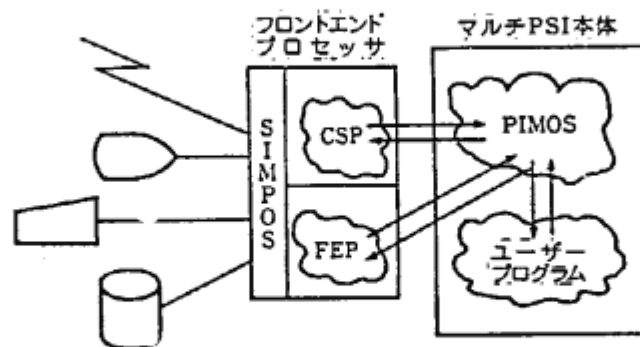


図3. マルチPSIに実装されるPIMOSの構成

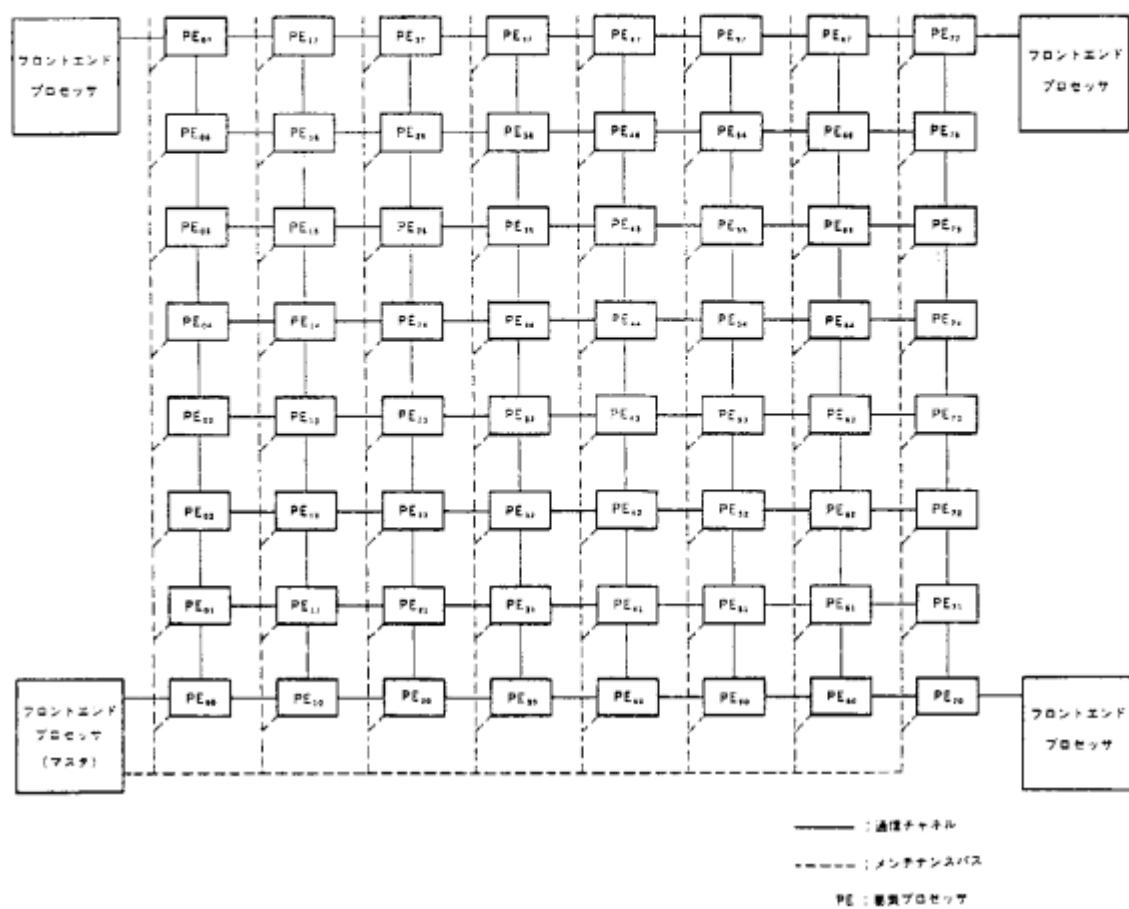


図 4. システム構成

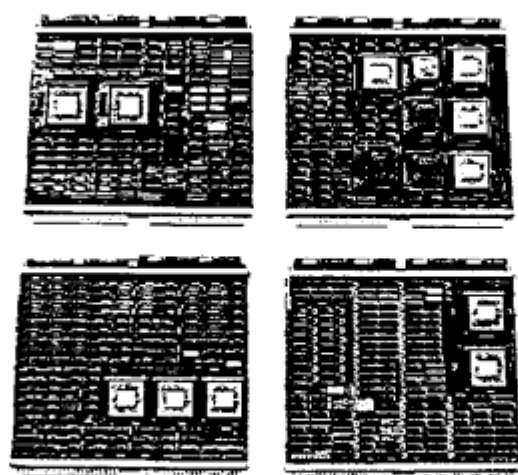


図 5. LSI搭載の要素プロセッサ用プリント基板

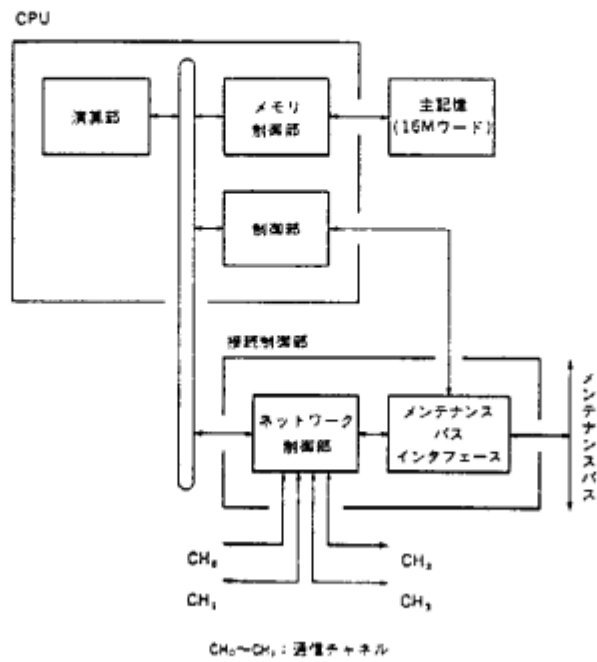


図 6. 要素プロセッサの構成

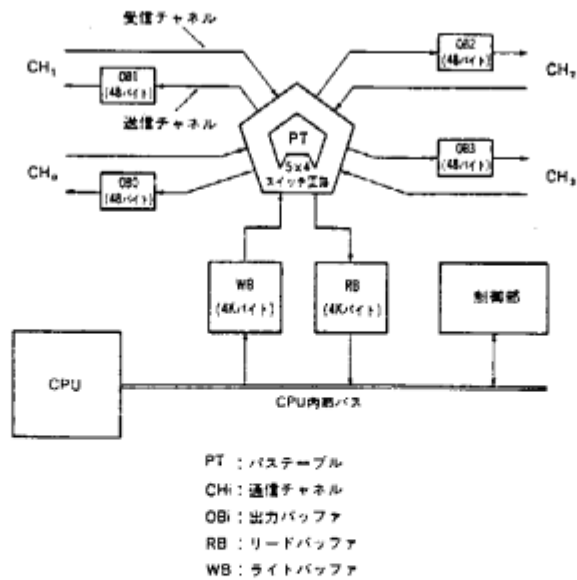


図 7. 接続ネットワーク制御部の構成

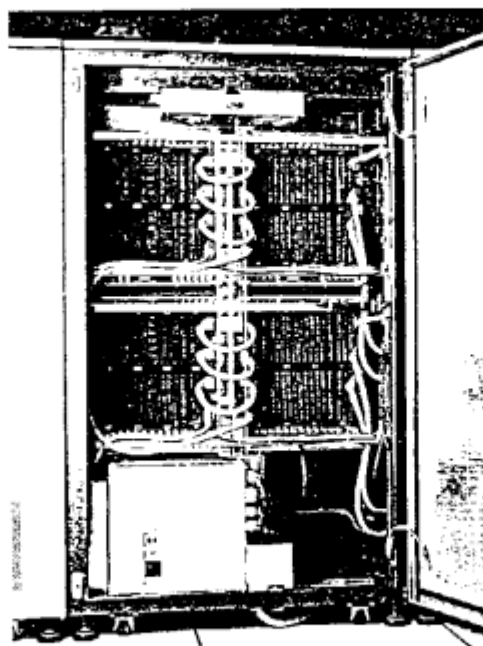


図 1. 本体部のケーブル接続 (後面パネル開)

表 1. マルチPSIシステムのハードウェア基本仕様

項 目		基 本 仕 様
システム構成	本体部筐体数	・最大8筐体(1筐体当たり8要素プロセッサ)
	要素プロセッサ数	・最大64台
	要素プロセッサ接続方式	・二次元格子型(8×8井盤目状)
	データ転送方式	・バイトシリアル転送 ・バスターブルによる自動経路開設
	転 送 能 力	・1チャンネル当たり 5Mバイト/秒 ・送受独立 ・複数経路同時開設可
	クロック制御	・基本動作周波数 5MHz ・PE間 同期/非同期モード 切替え可
要素プロセッサ	フロントエンドプロセッサ(FEP)	・改良小型化PSI(入出力装置含む)にインタフェースボード(2枚)付加 ・最大4台まで接続可 ・実行制御, 監視機能, 入出力制御
	CPU	・プリント基板3枚(改良小型化PSIのCPU部) ・8Kゲート CMOS G. A. 8種9石
	接続ネットワーク制御部	・プリント基板 1枚 ・20Kゲート CMOS G. A. 1種2石
	主 記 憶	・プリント基板 4枚(最大80Mバイト)
基本仕様	電 源 容 量	・3kVA/筐体
	寸 法・重 量 (1筐体あたり)	・860W×780D×1,400H(mm) ・280(kg)