

TM-1109

1.5M LIPS
AIプロセッサの開発

町田 浩久、安藤 秀樹、中西 知嘉子
前田 敦、中島 浩、中屋 雅夫 (三菱)

September, 1991

© 1991, ICOT

ICOT

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

(03)3456-3191~5
Telex ICOT J32964

Institute for New Generation Computer Technology

1. 5 M L I P S A I プロセッサの開発

A 1.5MLIPS 40-bit AI Processor

町田 浩久 安藤 秀樹 中西 知嘉子 前田 敦
Hirohisa Machida, Hideki Ando, Chikako Nakanishi, Atsushi Maeda,

中島 浩 中屋 雅夫
Hiroshi Nakashima, and Masao Nakaya

三菱電機株式会社
Mitsubishi Electric Corporation

1. はじめに

PrologやLisp等の記号処理用言語のための専用プロセッサはAIに関わる研究と開発を促進し、その実用化と普及に大きな役割を果たしている。専用プロセッサの最大の特徴は処理の高速性にあるが、そのためには言語に特化したハードウェアによる並列処理が重要となる。Prologに代表される論理型言語においては、言語に特有でかつ頻繁に行われる処理、たとえばデータのタイプ判定、及びデレファレンスと呼ばれる処理があり、これは汎用プロセッサにとって非常に負荷が大きい処理となっている。そのため特別なハードウェアを付加した専用プロセッサが必要となる。この付加するハードウェアと汎用プロセッサが処理する通常のハードウェアとは密接に関係しているため、付加するハードウェアを汎用のマイクロプロセッサの周辺回路として実現することは困難であった。しかし最近ではASIC技術の進歩もあり、多くのトランジスタを集積したVLSIの設計もある程度容易になり、Ivory[1]、Pegasus[2]、PSI-II[3]のようなAIプロセッサが開発された。

我々は、より高機能（並列性の増大とマシンサイクル及び実行ステップ数の縮小）への要求を満たすため、PUと呼ぶAIプロセッサを開発した。PUチップはPIM/mやAIワークステーションのキーデバイスとなるCISC型の専用マイクロプロセッサである。PIM/m

とは第5世代コンピュータプロジェクトで開発されている並列推論マシン[4]であり、2次元格子状に256個の要素プロセッサを結合した疎結合のマルチプロセッサシステムである。PUチップはPIM/mに搭載される場合にはKL1[5]を、AIワークステーションに搭載される場合にはESP[6]をというように複数の論理型言語を処理することができる。

PUチップではマシンサイクルを縮小するため、5ステージのパイプラインを導入した。さらに論理型言語の実行に必須であるデータのタイプ判定とデレファレンスもパイプライン化したため、実行ステージの負荷を大きく軽減でき、実行ステップ数を減少させることができた。また、最新の微細プロセスを使用することで、さらにマシンサイクルを小さくすることができた。これらアーキテクチャの改良と、デバイス技術の進歩による遅延時間短縮との相乗効果により、ベンチマークプログラム（append）での評価結果はPSI-IIと比較すると、性能が3～6倍に向上している。そしてPUチップはAIワークステーション搭載時に、1.5MLIPSという現在では最高レベルの推論速度を達成することができた。

以下この論文では、PUチップの性能を達成するための新しいパイプラインアーキテクチャと、ASICプロセッサに有効なクロック手法及び0.8 μ m CMOS技術の特徴について述べる。

2. アーキテクチャ

P Uチップは32Kワードの外部WCSに書えられたマイクロプログラムに制御される40ビットのパイプライン型のマイクロプロセッサ[7]である。この章ではそのハードウェアアーキテクチャについて述べる。

2.1 タグアーキテクチャ

P r o l o g等の論理型言語では変数に対する型宣言がないため、変数は任意の型のデータを保持することができ、かつ一般にはコンパイル時に変数の型を予測することができない。従って、変数のデータの型を実行時に定め、かつ判定する処理が必要となるが、これを実現するものがタグである。

P I M/mでは、図1に示すように1ワードの幅は40ビットであり、その内の8ビットがデータの型を示すタグとして用いられている。すなわち図1の(a)は整数の1991であり、(b)は1991番地にあるリストセルへのポインタであることがタグによって示されている。(c)は参照ポインタと呼ばれる特殊なポインタであり、データの操作をポインタが指しているアドレスのデータに対して行うことを示している。このように変数のデータの型を実行時に定め、かつ判定する処理をデータタイプ判定と呼ぶ。

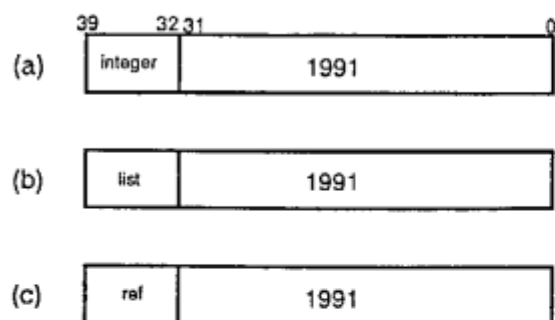


図1 データの内部表現

また一般には任意段数の参照ポインタの連鎖の末尾に操作すべきデータがある。図2は論理型言語に特有のユニフィケーションと呼ばれる処理である。2つの変数が等しいかどうかの判断をする処理、あるいは一方の変数を他方に一

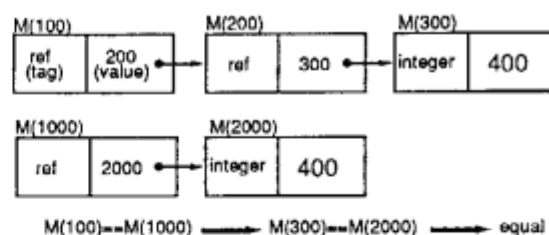


図2 ユニフィケーション

致させる処理の際には参照ポインタをたどりその末尾を見つけるためのデレファレンスという処理が必要となる。論理型言語の処理では、このタグの操作が重要であり、P Uチップの設計に際してもその高速化に重点をおいた。

2.2 ハードウェア構成

P Uチップは図3に示すようにD、A、R、S、Eの5つのパイプラインステージによって構成されている。

Dステージには命令デコードのためのRAMテーブルがあり、マイクロプログラムの実行開始アドレスや下流のステージを制御するためのナノコードが格納されている。デコードをRAMで構成したことは、複数言語のサポート(E S P、K L 1、L i s p等)、及びマイクロプログラム開発に多大な効果をもたらしている。

Aステージではオペランドのアドレス計算を行い、Rステージはその結果に基づき、主記憶からオペランドを読み出す。オペランドのアドレスはナノコードに従い以下の資源をソースとして計算される。

- (1) 命令のオペランドフィールド
- (2) プログラムカウンタ
- (3) レジスタファイル(A R F)
- (4) アドレスレジスタ

ここでA R Fというのは下位のEステージにあるレジスタファイルのコピーである。なお、Aステージでは分岐を含む命令フェッチの制御も行う。

Sステージはナノコードにしたがって、以下の資源の中から3つのオペランドを選択してEステージに引き渡す。

- (1) 命令フィールド
- (2) 主記憶上のオペランド及びそのアドレ

- (3) レジスタファイル (RF)
- (4) ワーキングレジスタ (WR)
- (5) 特殊レジスタ (SR)

このオペランドをセットアップする操作は、通常はRステージに相当する部分で行われるが、PUでは特にSステージを設けている。これは主記憶から読み出したオペランドのタグにより、Eステージで実行されるマイクロプログラムの開始番地を修飾する機能があり、頻繁に行われるデータタイプ判定をパイプライン化するためである。

Eステージは、フェーズ1とフェーズ2の並行に動作する2つのブロックから構成され、マイクロプログラムによって制御される。フェーズ1には命令レジスタ、RF、WR、及びSRがあり、この中から2つのものが選択されてフェーズ2に送られる。なお、マイクロプログラムルーチンの最後のマイクロ命令では、フェーズ1はSステージによって制御され、オペランドのセットアップのために使用される。

フェーズ2には、メモリアドレスレジスタ (MAR)、メモリデータレジスタ (MDR)、及び作業用レジスタ (REG) がそれぞれ2つある。フェーズ2ではこの中から2つのレジスタを選択し、ALUで演算した結果をフェーズ1及びフェーズ2のレジスタに書き込む。

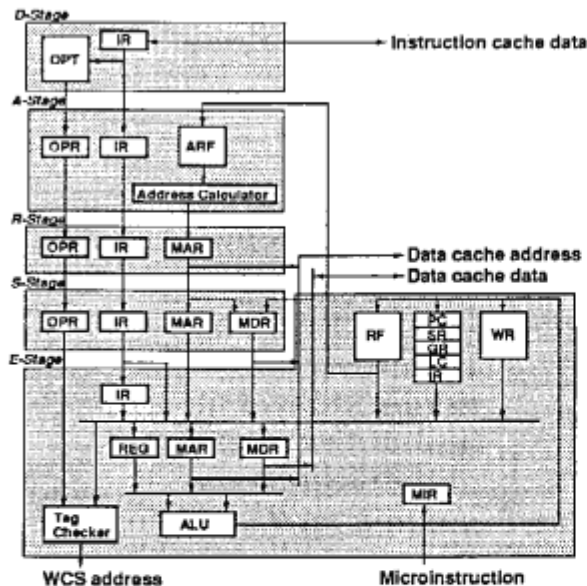


図3 PUチップの構成

2. 3 データタイプ判定とデレファレンス

データタイプ判定とデレファレンスはデータのタグをチェックして、その結果で制御フローをいろいろと変更することで実行される。Sステージには、データタイプ判定のための機能として以下のものがある。

- (1) 主記憶上のオペランドのタグと即値との比較結果に基づく、Eステージで実行されるマイクロプログラムの実行開始番地の修飾。
- (2) 主記憶上のオペランドのタグによる多方向分岐のセットアップ。
- (3) Eステージに引き渡すオペランドのタグと即値との比較結果に基づく2方向分岐のセットアップ。

これらの機能のうち、(1)と(2)を実現するためには、RステージとEステージの間に特別なステージが必要となる。

Sステージのもう1つの重要な機能として、デレファレンス機能がある。RFを元データとするデレファレンスの場合、元データのタグが参照ポインタである時のみRステージで主記憶の読み出しが行われる。さらに、その結果が参照ポインタである時には、Sステージにより参照ポインタでないデータが見つかるまでデレファレンスが行われる。

3. レイアウト設計手法

図4にPUチップのチップ写真を、表1にチップ諸元を示す。設計期間短縮のため、PUチップはセルベース設計方式によって自動でレイアウト生成を行った。PUチップは50種類のポリセル12、000個と、モジュールジェネレータにより生成されたRAMやPLAから構成される。レイアウト設計と検証に要した期間は約2週間である。しかし、このようにチップサイズが大きく高性能なVLSIを1度で自動設計を行うには、チップ内のクロックスキュー等が問題となる。この章では、その解決のためにPUチップに用いたクロック手法について述べる。

表1 チップ諸元

チップサイズ	16.3 mm x 13.6 mm
トランジスタ数	384k
random logics	110k
RAMs	270k
PLAs	4k
パッケージ	361-pin PGA
データ幅	40bit
ALU	32bit
パイプライン段数	5 stage
動作周波数	16.7 MHz
電源電圧	5 V
消費電力(16.7MHz動作時)	2.5 W
システムパフォーマンス	1.51 MLIPS

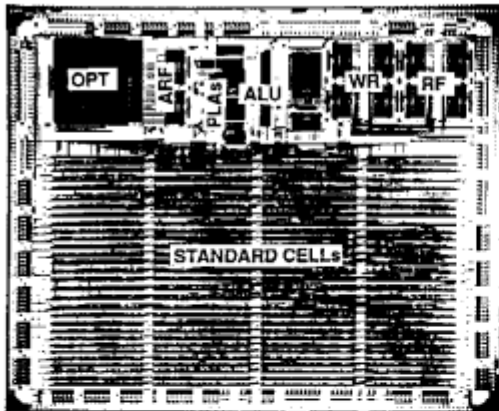


図4 チップ写真

3. 1 クロックの問題点

自動レイアウト手法を利用したVLSIの設計においては、レイアウト設計を階層的にブロックに分割し、クロック配線をツリー状に分配する方法がよく使われ、良好な結果を得ている[8,9]。しかし、この方法ではそれぞれのバッファに接続される負荷のバランスや配線の長さを制御したり、レイアウト実行後にバッファのサイズを調整したりする必要がある。この調整が商用の自動配置配線プログラムにとってはやっかいな処理となる。また集中バッファ方式では、チップ内のクロックスキューが大きくなる。

PUチップのようなASICプロセッサは開発期間の短縮が最重要課題である。我々はレイアウト設計にかかる時間をできるだけ短縮し、短い開発時間を機能・論理設計に費やすことを目標にした。また効果的な階層分割を考慮しなければ、バス線が階層間で2重に配線されることなどもあり、階層設計ではチップ面積が大きくなる傾向がある。そのためレイアウト設計手法には階層的設計手法を用いず、1度の自動配置配線設計を選択した。そこで、問題になるチップ内のクロックスキューを小さくするため次のような方法を用いた。

3. 2 PUチップのクロック手法

クロックスキューを最小にするため、図5に示すように2段のクロックドライバを用いて、2段目のクロックドライバ回路をチップの両サイドに配置した。2段目クロックドライバ回路

からフリップフロップまでのクロック線の抵抗を下げるため、垂直チャンネルの線幅を広く(10 μ m、第2層メタル使用)し、水平チャンネル内で両バッファの出力を結合した。水平チャンネルは容量を小さくするため1.4 μ m幅で配線した。クロックドライバ回路のサイズはpチャンネルトランジスタが3200 μ m、nチャンネルトランジスタが1600 μ mである。駆動するゲート数(最大600/クロック)、及び配線経路から求められる容量値と抵抗値でSPICEシミュレーションを実行することで、このトランジスタサイズを決定した。

実際のチップをEBテストで実測した結果を図6に示す。図に示すように2nsec以内のクロックディレイと1nsec以内のクロックスキューを達成することができた。この簡単なクロック方法は、これからの大規模ASICプロセッサに十分適用が可能と考えている。

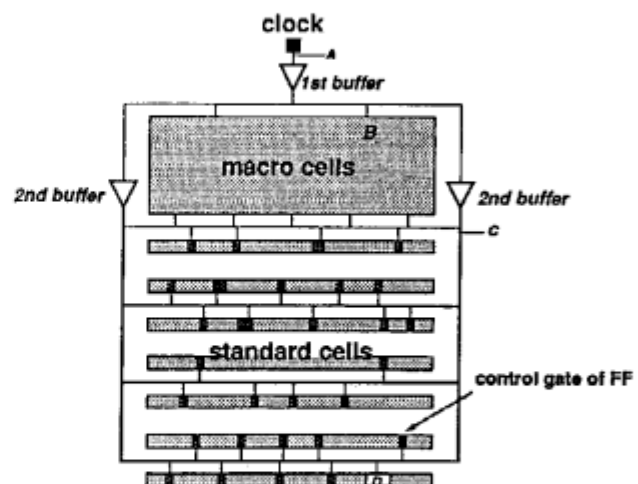


図5 PUチップのクロック分配

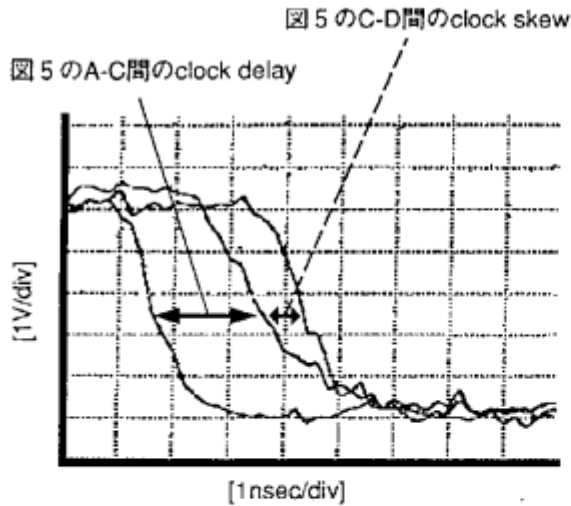


図6 EBテストで測定したチップ内部のクロック波形

4. プロセス技術

P Uチップはポリシリコン配線1層、メタル配線2層の0.8 μ m CMOS技術で製造した。表2に0.8 μ mプロセス技術をまとめる。ゲート長はnチャネル、pチャネルトランジスタとも0.8 μ mである。電源電圧5Vに対応して、ホットエレクトロンによる特性劣化を防止するため、nチャネルトランジスタにはn⁺をオーバーラップさせたLDD構造[10]を使用している。この構造はプロセス技術として斜め回転イオン注入法で形成したもので、これによってMOSトランジスタの信頼性を確保している。また、pチャネルトランジスタについてもLDD構造を採用する一方、チャネル部の不純物分布を最適化することで十分な耐圧を確保しながら高い電流駆動能力を得ている。

表2 プロセスの特徴

0.8 μ m Twin-well CMOS	
gate length (N/P)	0.8 μ m
gate oxide thickness	18nm
well depth (P-well)	4.0 μ m
(N-well)	2.5 μ m
diffusion space (P ⁺)	1.4 μ m
(N ⁺)	1.0 μ m
interlevel dielectric thickness	
(under 1st-metal)	0.7 μ m
(1st-metal / 2nd-metal)	0.8 μ m
metal pitch (1st-metal)	2.4 μ m
(2nd-metal)	3.2 μ m
contact hole size	0.8 μ m $\times$ 0.8 μ m
via hole size	0.8 μ m $\times$ 1.0 μ m

5. 性能評価と並列マシンへの応用

5.1 性能評価

図7に室温でのサイクル時間と電源電圧のシュムプロットを示す。P Uチップは外部WCS用のSRAMのアクセスタイムを考慮して、16.7MHzのサイクルタイムで使用している。そのときの室温における消費電力は電源電圧5Vで2.5Wである。

現在P UチップはA Iワークステーションに搭載されており、appendと呼ばれる処理において、1.5MLIPS (Mega Logical Inference Per Second)の処理性能を達成することができた。これは現時点のProlog処理マシンとしては最高性能のものである。

P Uチップの性能を評価するためP Uチップの前世代のA IコンピュータであるP S I - IIと比較した結果を表3に示す。逐次論理型言語ESPを用いてappend処理を実行したところ、総合的な性能では3.78倍に性能が向上した。性能向上の内訳は、0.8 μ mプロセスによる向上が1.67倍、パイプラインのアーキテクチャを導入したことによる向上が2.27倍であった。論理型言語を考慮したパイプラインアーキテクチャを導入して、サイクルタイムと実行ステップ数を減少させることができた。サイクルタイムは3/5に、実行ステップ数は11/15に向上した。

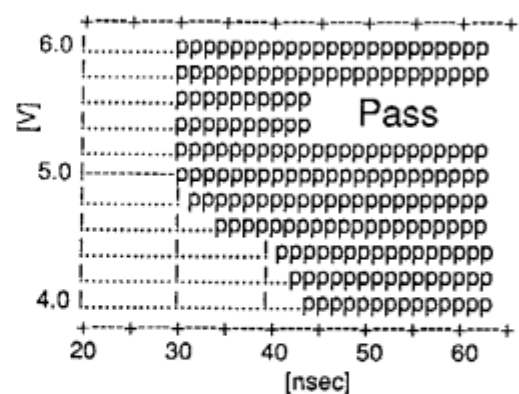


図7 Schmoor プロット
(cycle time vs. supply voltage)

表3 PUチップの性能比較

	サイクルタイム	ESPにおける append処理の 実行ステップ数
PSI-II (旧プロセッサ)	167nsec	15steps
0.8- μ m PSI-II	100nsec	15steps
PUチップ	60nsec	11steps

5. 2 並列推論マシン PIM/m

PUチップは並列推論マシン PIM/m の主要な構成要素でもある。PIM/m は、図8に示すように最大 16×16 の2次元格子状に256個の要素プロセッサを結合した疎結合のマルチプロセッサである。SCSIポート、ディスクなどの入出力装置やフロントエンドプロセッサ(FEP)を接続することができる。

PIM/mの要素プロセッサは、図9に示すように3つのVLSIチップPU、CU、NUと、浮動小数点プロセッサを中心に構成されている。CU(Cache Unit)チップは1Kwの命令キャッシュ、4Kwのデータキャッシュ(データアレイはチップ外)のほか、アドレス変換バッファや主記憶の制御回路からなる。NU(Network Control Unit)チップは、隣接する4プロセッサを結ぶネットワークチャネルのスイッチングを行うとともに、PUチップが行うメッセージの送受信を制御する。

PIM/mは、PSI-II用のプロセッサで構成されていた並列推論マシンmulti-PSIの4倍の要素プロセッサを持ち、1つの要

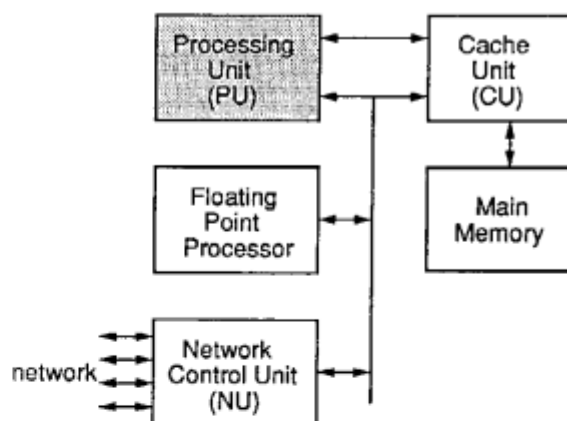


図9 PIM/mの要素プロセッサ

素プロセッサ自身では並列論理型言語KL1の処理において5倍の性能向上が達成できている。すでに要素プロセッサを32個接続した基本構成のPIM/mの動作は確認できており、256個構成のPIM/mは1992年3月に完成する予定である。

6. むすび

高性能な40ビットのパイプライン型のAIプロセッサを開発した。タグ処理を高速化したパイプラインアーキテクチャと、0.8 μ mのCMOS技術と、セルベース設計方式に有効なクロック手法を用いることで、append処理の性能において1.5MLIPSの高性能を実現できた。チップサイズは16.3mm $\times$ 13.6mmで、384,000個のトランジスタを集積している。

PUチップは最初のマスクですべての機能が満足であることが確認でき、SIMPOSと呼ばれるオペレーティングシステムを立ち上げることができた。これはセルベース設計方式を利用したことと、セルベース設計方式に適したクロック手法を用いてレイアウト設計時間を短時間で実現でき、機能設計や論理設計に多くの時間を費やすことができたことによる。

PUチップを搭載することでAIワークステーションPSI/UXをデスクトップサイズで実現することができた。またこのPUチップを使用することで、高性能な並列推論マシンを小さく収めることができ、大規模な並列処理が可能となる。

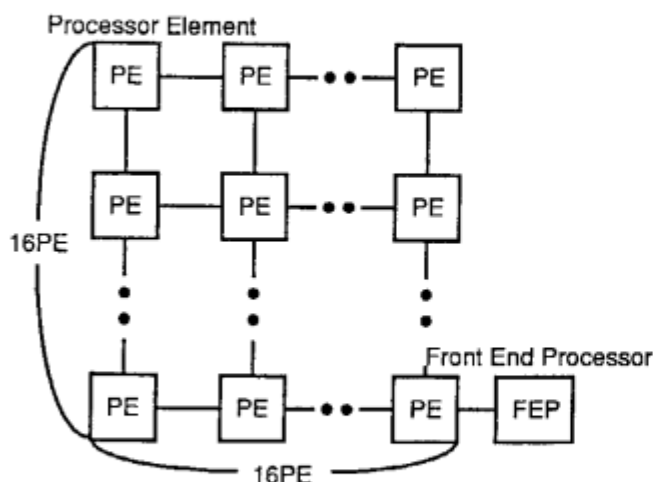


図8 並列推論マシン PIM/m

謝 辞

最後に本プロセッサの開発に当たり、御支援いただいた当社LSI研究所の小宮所長、中野副所長、及び堀場部長に感謝致します。

また、本プロセッサの製造に御協力頂いた製造グループの関係各位、及び評価に尽力頂いた、常友力氏に感謝致します。

参考文献

- [1] Baker, C., Chan, D., Cherry, J., Corry, A., Efland, G., Edwards, B. Matson, M. et al., "The Symbolics Ivory Processor: A 40 Bit Tagged Architecture Lisp Microprocessor," in Proc. of Intl. Conf. on Computer Design, pp. 512-515, 1987.
- [2] Seo, K. and Yokota, T., "Design and Fabrication of Pegasus Prolog Processor," in Proc. of Intl. Conf. on VLSI, 1989.
- [3] Nakashima, H. and Nakajima, K., "Hardware Architecture of the Sequential Inference Machine :PSI-Ⅱ," in Proc. of 4th Symp. on Logic Programming, pp. 104~113, 1987
- [4] Uchida, S., Taki, K., Nakajima, K., Goto, A., and Chikayama, T., "Research and Development of the Parallel Inference System in the Intermediate Stage of the FGCS Project," in Proc. of the Intl. Conf. on Fifth Generation Computer Systems 1988, pp.16~36, 1988.
- [5] Chikayama, T., Sato, H., and Miyazaki, T., "Overview of the Parallel Inference Machine Operating System(PIMOS)," in Proc. of Intl. Conf. on Fifth Generation Computer Systems 1988, pp. 208-229, 1988.
- [6] Chikayama, T., "Unique Features of ESP," in Proc. of Intl. Conf. on Fifth Generation Computer Systems 1984, pp. 292-298, 1984.
- [7] Nakashima, H., Takeda, Y., Nakajima, K., Andou, H., and Furutani, K., "A Pipelined Microprocessor for Logic Programming Languages," in Proc. of Intl. Conf. on Computer Design, pp.355-359, 1990.
- [8] Tokumaru, T., Masuda, E., Hori, C., Miyata, M., and Iwamura, J., "Design of A 32bit Microprocessor, TX1," in Symp. on VLSI Circuits Dig. Tech. Papers, pp. 33-34, 1988.
- [9] Boon, S., Bulter, S., Byrne, R., and Setering, B., "Hige Performance Clock Distribution for CMOS ASICs," in Proc. of the 12th annual Custom Integrated Circuits conf., pp. 15.4.1-15.4.4, 1989.
- [10] Inuishi, M., Mitsui, K., Komori, S., Shimizu, M., Oda, H., Mitsuhashi, J., and Tsukamoto, K., "Optimum Design of Gate/n- Overlapped LDD Transistor," in Symp. on VLSI Technology Dig. Tech. Papers, pp. 33-34, 1989.