並列推論マシンPIM/pの概要

5N-1

後藤 厚宏 (ICOT)　篠木 剛　久門 耕一　服部 彰 (富士通株式会社)

1. はじめに

ICOTでは第5世代コンピュータの中核となる並列推論マシンの研究開発を進めている。本研究においては, 核言語(KL1), オペレーティングシステム(PIMOS)とともに, 論理型の枠組の基で並列推論マシンを開発することにより, 並列ソフトウェアからハードウェアまでの一貫性を高めることを重視してきた。

PIM/pは, 並列推論マシンのパイロットマシンの一つであり, 100台規模の要素プロセッサからなる。本マシンの開発においては, 今後の並列処理研究のベースとして利用できるように, 実質性能の高いパイロットマシン(10–20M rps)を目指した。単体でも十分に高性能な要素プロセッサと核言語KL1の並列処理のためのハードウェア機構を導入した。本報告では, 並列推論マシンアーキテクチャの基本要素とPIM/pにおける実現方式の概要について述べる。

2. 並列推論マシンアーキテクチャの基本要素

2.1 KL1の抽象命令とコンパイラによる最適化

KL1の効率的実行には, コンパイル時における最適化と, 実行時の動的データ型判定のハードウェア支援が必要である。KL1-B[2]は, コンパイラによってリダクションの制御や単一化を最適化することを目指したKL1の抽象命令である。KL1の単一化は, 受動部と能動部および引数の型に従って分類し, 操作を効率化できる。KL1-Bは, それぞれの単一化の意味に対応した単一化命令を持ち, コンパイル時の解析によって使い分ける。

ほとんどのKL1-B命令は実行時の動的データ型判定を含む。また, 判定によって選択する操作が多岐にわたる高機能命令も多い。このため, 要素プロセッサにタグアーキテクチャを採用するとともに, 高機能命令を効率良く実現できることが必要である。

2.2 MRBによる実行時のガーベジコレクション

KL1はメモリの破壊的書き換えを許さないため, 単純に実装すると急速にメモリを消費し, ガーベジコレクション(GC)が頻繁に起こってしまう。通常のGCは, メモリ参照の局所性が悪く, キャッシュのように参照の局所性を利用する機構が有効に働かなくなる。また, 疎結合システムでは, 一つのノードがGCを起こすと他のノードとの間の通信ができなくなり, 全体性能の低下を招く。このため, 並列推論マシンにおいては, 実行時に不要になったメモリ領域を要素プロセッサ毎に回収し, 再利用することが必須である。メモリ再利用方式としては, データへの参照数を効率良く管理できるMRB方式[1]を用いる。MRBによる再利用操作はKL1-Bの単一化命令の中に組み込まれて実行される。

2.3 局所性のあるプロセッサ間接続方式

並列推論マシンでは, ゴール単位の負荷分散とユニフィケーションのためのプロセッサ間通信によってKL1の並列処理を進める。並列処理の効果を引き出すために, 通信コストに局所性のある接続方式を用いて要素プロセッサを接続する。負荷分散においては, 要素プロセッサ内での局所的な処理, および比較的通信コストの小さい近傍のプロセッサ群による局所的処理を利用する。

3. PIM/pの全体構造と特徴

3.1 クラスタによるプロセッサの階層的接続

PIM/pにおいては, 図1に示すように, 要素プロセッサ(PE)をクラスタによって階層的に接続し, 通信コストに局所性のある接続方式を実現する。PIM/pの本体は, ネットワークによって結合された16台のクラスタで構成される。各クラスタは, 共有バス / 共有メモリによって密に結合された8台のPEからなる。クラスタ内では一つのアドレス空間を共有し, レスポンスが高速で通信コストが小さいプロセッサ間の並列処理が可能である。

一方, クラスタ間に渡る分散ユニフィケーションはネットワークを介した非同期メッセージ通信によって行う。ここで, クラスタ毎のメモリ再利用を十分に行うために, クラスタ間に渡る変数への参照ポインタを表管理する。クラスタ間に渡る並列処理はクラスタ内に比べて通信コストが高いが, クラスタ毎のメモリ管理の独立性, クラスタ数の拡張性が得られる。

各PEはFEPのためのポートを持っており, PIM/pの用途に応じて複数のFEP(または入出力機器)を接続することができる。

3.2 RISCと高機能命令を融合した要素プロセッサの命令セット

PIM/pのPEは, RISCの利点とマイクロプログラムの利点を融合した命令セットを持つ[4]。RISC指向の命令(単命令)はマシンサイクルの短縮とハードウェアコストの節約が可能である。PIM/pの単命令には, MRBによる実行時のメモリ再利用やデレファレンスを支援する命令があり, 4段のパイプラインを用いて, 最高1マシンサイクルに1命令の割合で処理することが可能である。マシンサイクルとしては50ナノ秒を目標としている。

The Overview of Parallel Inference Machine PIM/p
Atsuhiro Goto (ICOT)
Tsuyoshi Shinogi, Kouichi Kumon, Akira Hattori
(Fujitsu Limited)

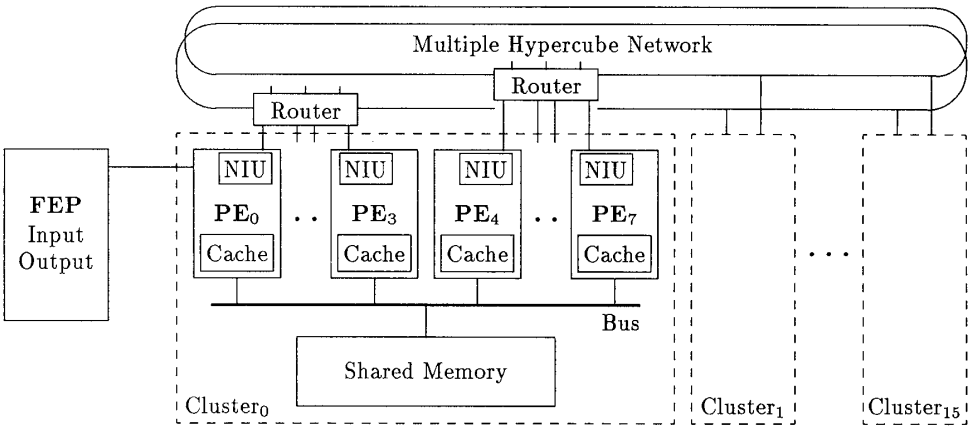

図 1: PIM/p の全体構造

KL1-B の高機能命令はPIM/pのマクロ命令を利用して実装する。マクロ命令は引数として指定されたレジスタのデータ型判定に基づき, 小さいコストでマクロの本体の条件呼び出しができる。マクロの本体は通常の単命令とほぼ同様の内部命令によって記述し, 各 PE の内部命令メモリから供給されるため, マクロ本体の実行中は共有メモリからの命令フェッチが節約できる。

このような KL1 向き命令セットと短いマシンサイクルにより, 実行時の GC コストを含めた要素プロセッサの単体性能として, append において 600K rps 程度, 実質 200–500K rps を予定している。

3.3 KL1 向き一貫性キャッシュと排他制御機構

PE 毎の処理効率を向上させるために, クラスタ内で一貫性が保持できるライトバック型のキャッシュを各 PE に設けた。ここで, クラスタ内での要素プロセッサ台数に比例した並列処理性能を得るためには, 共有バスネックの回避が必要である。そこで, バスの転送能力を高くするとともに, KL1 の並列処理の特性を活かしてバストラヒックを節約するキャッシュコマンドを用意した[4]。

共有メモリ上の変数を介した通信では, メモリアクセスの排他制御機構が重要である。そこで, 各 PE のキャッシュのブロック状態を利用したロック機構を設け, 低コストの排他制御を可能とした。

3.4 要素プロセッサ毎のクラスタ間メッセージ処理

PIM/pのクラスタ間に渡るメッセージ通信においては, パケットの組み立て / 読み取りをハードウェアで支援することが重要である。さらに, クラスタ間で相互に参照されるデータへのポインタを表管理する操作が必要である。そこで, クラスタ間に渡る並列処理性能をクラスタ内の高い並列処理性能に見合うだけ向上させるために, クラスタ内の各 PE にネットワークへの接続口 (NIU: Network Interface Unit) を用意した[3]。 NIU は CPU のコプロセッサとしてバイト / ワード変換機能とルータとのパケット送受を代行する。これにより, 各 PE がクラスタ間通信を必要とした時点で自らクラスタ間通信を行うことが可能になった。

3.5 構造バッファプールによるハイパキューブ網

クラスタ間ネットワークにおいては, 実効的なスループットを高めるとともに, ハードウェア実装上の問題を解決する必要がある。そこで, ネットワークのトポロジとしてノード間距離が小さいハイパキューブ網を採用し, ルーティングの自由度を高めた。さらに, デッドロックとネットワーク上のホットスポットを回避するために, そのルータには可変長パケットに対応できる構造バッファプール機構を設けた。

ハイパキューブの次元数はクラスタ数に対応した 4 次元 (+予備 1) であり, 電気リンクと光リンクの両方を検討している。各クラスタ性能に対応するために, ネットワーク全体を 2 重化し, 各クラスタでは 4 個の PE(NIU) 毎に一つのネットワークルータに接続する。ネットワークの通信路は 1 バイト幅であり, クラスタ当たりの通信容量は最大 40M バイト / 秒である。

4. おわりに

並列推論マシンの基本構造とパイロットマシンPIM/pの主な特徴について述べた。これまでに命令仕様の設計とハードウェアの詳細設計を終了した。現在, 要素プロセッサおよびネットワークの LSI 開発と実用性を重視した実装設計を進めている。

謝辞: 日頃, ご指導ご助言を頂く ICOT 内田第 4 研究室室長, 富士通研究所 林 人工知能研究部長, ならびに ICOT と富士通の PIM 研究開発メンバに感謝致します。

参考文献

[1] T. Chikayama, et al. Multiple Reference Management in Flat GHC. In *Proc. of the 4th ICLP*, 1987.

[2] Y. Kimura, et al. An Abstract KL1 Machine and its Instruction Set. In *Proc. of the 1987 SLP*, 1987.

[3] 久門 他 並列推論マシン PIM/p のネットワーク. 本大会予稿集.

[4] 篠木 他 並列推論マシン PIM/p の要素プロセッサのアーキテクチャ. 本大会予稿集.