並列推論マシンPIM

3B-6 —中期PIMのハードウエア構成について—

松本 明、六沢 一昭、後藤 厚宏

(財)新世代コンピュータ技術開発機構

1．はじめに

ＩＣＯＴでは，要素プロセッサ（ＰＥと略す)を100台程度の規模で接続した中期ＰＩＭの研究開発を進めている。中期ＰＩＭでは，並列処理方式の研究と，それを反映したハードウェア・アーキテクチャの研究との協調関係を重要視している[１]。

ハードウェア・アーキテクチャの研究テーマには，ＰＥ単体の高速化に関するものと，それらの結合方式に関するものとがある。ＰＥ単体に関しては，タグ付データの操作機構，パイプライン処理，さらにコンパイラの最適化能力を十分引き出す機械語の導入等による高速化手法が明らかになってきた。一方，並列推論マシンの立場からみると，ＰＥ単体以上に重要な研究テーマとして，ＰＥとＰＥ，或いはＰＥとメモリ等の結合方式が挙げられる。但し，ＰＥの結合方式には，未解決な問題点が多いのが現状である。

そこで，本稿では，主に，①中期ＰＩＭにおける共有メモリを介したＰＥの結合方式，②共有メモリ・アクセスを高速化する並列キャッシュ・メモリ，③共有メモリのロック機構等について報告する。

2．中期ＰＩＭのハードウェア構成（図１）

2.1．クラスタと階層構造

並列ソフトウェアの立場からは，複数のＰＥ間に距離が全く無いモデル，または，距離の差の少ないモデルが良い。しかし,100台規模の中期ＰＩＭのハードウェアを考えると，ある程度の距離の差が生じることはやむをえない。そこで，核言語第一版(ＫＬ１)の処理系では，予め，次に述べるハードウェア上のＰＥ間の距離を考慮したものとなっている。

現在検討中のＫＬ１処理系で，意識しているハードウェア構成は，図１に示すように，

① 要素プロセッサ（ＰＥ）

② 共有メモリを介して，10台程度のＰＥを同一距離で結合した，密結合マルチプロセッサからなるクラスタ

③ クロスバー等の上位ネットワークで，10個程度のクラスタを結合した，疎結合マルチプロセッサからなるシステム

の３レベルからなる，階層構造をしたハードウェアである。

2.2．共有メモリとローカル・メモリ

ＫＬ１のクラスタ内処理方式は，共有メモリ・ベースであるが，ゴール・レコード等のように，各ＰＥだけがアクセスするデータは，ＰＥ毎のローカル・メモリに切り出す検討を進めている[２]。これは，ＰＥ毎のローカル・メモリは，一般的に共有メモリよりも，アクセス速度が速いので，メモリ・アクセスの高速化に有効な処置である。また，後述（第３節）するが，共有バスの使用率を下げるためにも，ＰＥ毎に局所的なデータは，できるだけローカル・メモリに切り出して格納した方が良い。なお，試算では，全メモリ・アクセスの約１／３は，ローカル・メモリに対するアクセスである。

2.3．アドレス変換機構

ソフトウェアに負担をかないで，有限の物理メモリを有効利用するために，ハードウェアによる論理アドレスと物理アドレスの変換は必要であると考える。共有メモリ用のアドレス変換機構を置く位置は，ＰＥ側ではなく，共有メモリ側に置く方式とする。これは，ＰＥ側に置く方式では，論理／物理アドレスの変換表が複数存在するため，それら

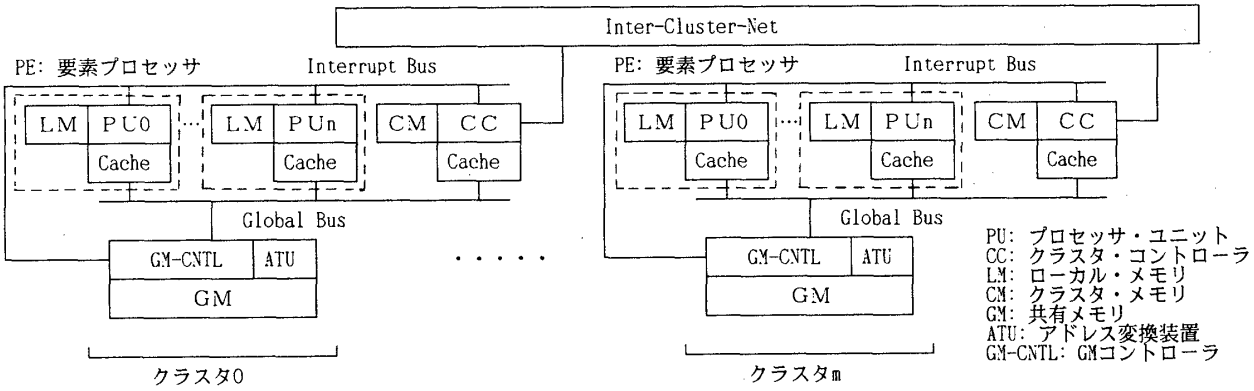

図１．中期ＰＩＭのハードウェア構成

Parallel Inference Machine : PIM
-On Hardware Configuration of the Intermediate Stage PIM-
Akira MATSUMOTO, Kazuaki ROKUSAWA, Atsuhiro GOTO
Institute for New Generation Computer Technology (ICOT)

の一貫性の保証が問題となるからである。

仮想記憶は，ガベージ・コレクションを考えると，オーバヘッドが大きくなるので採用しない。しかし，実共有メモリの容量は，可能な限り大きくする。

3. 並列キャッシュ・メモリとロック機構

3.1. 並列キャッシュ・メモリ

①共有メモリ・アクセスの高速化

中期ＰＩＭでは，共有メモリのアクセスを高速化するために，キャッシュ・メモリの導入を検討している。

②推論マシンとキャッシュ・メモリ

ＰＳＩ（逐次推論マシン）では，ハードウェア的手法によってメモリ・アクセスを高速化するキャッシュ・メモリを採用している。その評価によると，推論マシンが論理型言語を実行する場合における，キャッシュ・メモリの有効性が示されている。従って，並列推論マシンにおいても，逐次推論マシンと同等の効果が有ることが期待される。

③並列キャッシュ・メモリの課題

並列キャッシュ・メモリでは，複数のキャッシュ・メモリが同じデータのコピーを持つことがあり得る。このコピーの一つに書き込みが起きた場合には，他のキャッシュ・メモリのデータを無効化する等して，常に全ての並列キャッシュ・メモリ間で，データの一貫性を保証する必要がある。このデータの一貫性を保証するためには，自ＰＥが書き込んだアドレスを，他のＰＥに知らせる手段が必要である。

④中期ＰＩＭの並列キャッシュ・メモリ

クラスタ内では，共有バスを介して，各ＰＥのキャッシュ・メモリを接続する構成を取るので，あるＰＥの書き込みアドレスを，他の全てのＰＥが知ることが可能である。各ＰＥのキャッシュ・メモリ・コントローラは，データの一貫性を保証するために，自ＰＥのキャッシュ・メモリにコピーがあるアドレスに対して，他のＰＥによる書き込みを検出した場合には，自キャッシュのデータを無効化する。

共有バスを介して共有メモリに接続可能な最大ＰＥ数は，各ＰＥの共有バス使用率で決まる。このため，各ＰＥの共有バスの使用率を下げることが設計上のポイントとなる。そこで，キャッシュ・ブロックの状態にExclusive/Shared等の状態を追加して，バスの使用率を低くできる並列キャッシュ・メモリのバス・プロトコルを検討している。このバス・プロトコルは，多少複雑ではあるが，バスの使用率を低く抑えることができるため，クラスタ内の10台程度のＰＥを，一つの共有バスを介して，共有メモリに接続可能であると考えている。

3.2. ロック機構

①ロックの必要な処理

複数のプロセッサが，共有メモリをアクセスする処理方式で，２個以上のプロセス（プロセッサ）が共有メモリの同じ領域に対して書き込みを行う可能性がある場合には，これらのプロセス間の同期を取る必要がある。ロックは，この同期を取るための一つの手段である。

②ＫＬ１におけるロック処理と特徴

現在検討中のＫＬ１処理系で，ロックの必要な処理には，主に，未定義変数の処理と，制御構造（ゴールをメタ・レベルで管理するメタコール・レコード等）の処理の２種類がある。未定義変数の処理は，ロック頻度は高いが，変数ワードそのもの（１ワード）に対するロックで良く，ロック期間が短い。一方，制御構造に対する処理は，ロックの頻度は比較的低いが，制御構造全体のロックが必要で，ロック期間が比較的長くなる可能性がある。しかし，いずれの場合も，ロックするアドレスが実際に競合する確率は低いと考えられる。

未定義変数に対するロックでは，ＫＬ１の変数が単一代入であるという性質を利用したロックが考えられる。例えば，変数が未定義ならばロックをかけるが，既に変数の値が定まっているならば，ロックをかけずに単にデータを読み出すだけで良い。この高機能ロック処理を“Read&Lock if Undef”と呼ぶ。

ＫＬ１のロック処理には，上述した特徴があるため，次の方針でロック機構を設計している。

- 未定義変数に対する１ワードのロックは，頻度が高いため，高速実行できるハードウェア・ロックとする。さらに，“Read&Lock if Undef”のような高機能ロックを，ハードウェアでサポートすることも考えられる。
- 複数箇所の同時ロックは，頻度が比較的少ないため，実現可能であれば良い。ソフトウェア・ロックで実現する方式も考えられる。

③ロック機構と並列キャッシュ・メモリ

出現頻度の高い１ワードのロックを，高速実行するためには，ロック・メモリのように，アクセスが一箇所に集中する方式は良くないと考えられる。そこで，ＰＥ毎にロック機構を設ける分散制御方式を検討している。

ロック機構を分散制御する方式で，ロックをかける処理は，並列キャッシュ・メモリへの書き込みと同等な処理が必要である。このため，複数のＰＥにコピーのある領域に対してロックをかける場合には，並列キャッシュ・メモリへの書き込み処理と同様に，データの一貫性に関する問題が起こる。

このロックに係わる一貫性処理に要する共有バスの使用率を低く抑える方式として，自ＰＥだけがアクセスしている領域をロックする場合には，ロック／アンロックに関する情報を共有バスに出さない方式を検討している。

4. おわりに

中期ＰＩＭのハードウェア構成のうち，クラスタ内の要素プロセッサ結合方式に関する検討結果を述べた。その結果，共有バスを介して，10台程度の要素プロセッサを共有メモリに結合したクラスタは，並列キャッシュ・メモリを上手に設計することにより，十分実現可能であるとの見通しを得た。今後は，ソフトウェア・シミュレーションにより定量的評価を進めると共に，クラスタ間のハードウェア構造についても検討を進めて行く予定である。

〈参考文献〉

[1]後藤 他，“並列推論マシンPIM －中期構想－”，本大会予稿集 3B-5

[2]佐藤 他，“並列推論マシンPIM －中期PIMの処理方式について－”，本大会予稿集 3B-7