

分散型共有メモリをもつ

6T-2 データパラレル並列計算機のアーキテクチャ

村松 晃¹⁾、吉原郁夫¹⁾、濱中直樹²⁾、田中輝雄²⁾(株)日立製作所 システム開発研究所¹⁾、中央研究所²⁾

1. はじめに

並列計算機の構成一要素プロセッサ(P E)台数、結合ネットワーク等一を意識せず、対象とする問題だけを意識してプログラムを書くためには、共有メモリ方式の並列計算機が望ましい。一方、絶対性能向上のためにP E台数を増加させていくと、共有メモリ方式ではアクセス競合によるオーバーヘッドが大きくなりすぎて、非実用的となる。キャッシュを設けても、各P Eがアクセスするデータに強い局所性がないと、有効ではない。以下では、このようなジレンマを解決するために、分散メモリ上でデータを共有し、P Eが処理するデータに著しい局所性がなくても分散メモリの利点が活かせるようなメモリ方式(分散型共有メモリ)と、このメモリ方式を用いたデータパラレル計算モデル¹⁾に基づく並列計算機(データパラレル計算機)のアーキテクチャおよびマシンモデルについて述べる。

2. 分散型共有メモリのアーキテクチャ

各P Eが、局所メモリをアクセスする程度の速度でアクセスできる物理的に分散した共有メモリについて述べる。このメモリにおいては、各P Eは同一データを並列に読むことができるので、アクセス競合は発生しない。

2. 1 アドレス空間とアドレス変換

本メモリはアドレスコードの先頭にプロセッサ番号を付した多重局所空間として構成する。プロセッサ番号をALLと指定してWRITEすると、全局空間の同一アドレスに書き込みが行われる。各局所空間をセグメンテーションし、アドレスコードのセグメントフィールドの動的アドレス変換により物理アドレスを得る。

2. 2 コピー制御

各セグメントは他の局所空間にそのコピーを置くことができる。コピーセグメントと親セグメントのアドレスの差、およびプロセッサ番号の値の差を

コピーアドレス情報と呼ぶ。コピーアドレス情報は、セグメントテーブルにリンクしたファンアウトテーブルと呼ぶ固定領域に格納し、書き込みアドレスの変換時にその内容を読みだす。書き込みアドレスと自プロセッサ識別子を読みだされたコピーアドレス情報に加えれば、多重局所空間中のコピーアドレスが得られる。これを処理コード、データと併せて相互結合ネットワークに送出しコピーの更新を行う(図1)。以上の処理を、命令パイプラインのアドレス変換ステージでハードウェアにより実行することにより、書き込み命令実行時に自動的にコピーの更新動作が起動される。コピーが複数個ある場合には、順に更新する。データに付けたタグ等を経由してトークンを授受することにより、コピーと親のコヒーレンシーをとる。これをデータフロー同期と呼ぶ。本メモリでは、このように各セグメント単位にコピーを持たせることができるので、個々の共有配列を異なるセグメントにアサインすれば、複数のP Eが同時に同一配列データを参照することができ、アクセス競合が低減される。

2. 3 リモートオペレーション

コピー制御が有効であるのは、コピー関係をコンパイラが認識できる場合、すなわち、配列が解析可能な添字式で表現されている場合である。間接参照の場合には、データとそのデータを定義するプログラムを常に同じP Eに割り付けることができないため、他のP EのデータAを定義する必要性が生ずる。このとき、 $A = A + B$ のようなインプレース型演算は、相手のメモリを長時間ロックするために著しいオーバーヘッドを生ずる。本方式では、AのアドレスとBの値、演算命令 '+' を相手に送って処理を代理実行させるリモートオペレーション命令を用意し、この問題を解消している。

3. データパラレル計算機のマシンモデル

配列データとその定義手続きを一体化したデータオブジェクトへのメッセージ送信文を含む逐次処理プログラムをホスト計算機で実行し、並列性の抽出はデータオブジェクトを格納するP E側で行うデータパラレル計算モデル準拠のマシンモデルについて述べる。

An Architecture of Data Parallel Computer with
Sharable Distributed Memory System

Akira MIRAMATSU, Ikuo YOSHIHARA, Naoki HAMANAKA,
Teruo TANAKA, Hitachi, LTD.

3. 1 全体構成

本計算機システムはホスト計算機とPEアレイの2階層で構成し、両者を放送および全体同期用の信号線で結合する。PEアレイは均質型で、各PEは処理装置、局所メモリ、動的アドレス変換回路を含むコピー制御装置とから構成する。各PEは任意のPE間で通信が可能な相互結合ネットワーク(例えばハイバークロスバスイッチ⁶⁾と呼ぶネットワーク)で結合する(図2)。

3. 2 プログラム実行制御方式

並列処理対象の配列データは添字により分割し、各PEに分散配置する。また、これを定義する繰返しループの本体をデータオブジェクトの手続きとして全PEに放送しておく。ホスト計算機では入出力とスカラデータの処理を行ない、アレイに手続き名を放送して繰返しループの処理を起動させた後、その繰返しループの処理と並列実行可能なスカラ処理を実行し、バリアで停止する。各PEは自分の担当する添字範囲をもつ配列要素を定義し、終了信号を全体同期信号線に出力する。全PEおよびネットワークが停止した段階でホスト計算機はバリアを通過し、次の命令実行に進む⁷⁾。

4. おわりに

分散型共有メモリを持つデータパラレル計算機のアーキテクチャおよびマシンモデルの概要を報告した。このマシンモデルのハードウェア性能評価はシミュレータにより行なっている⁸⁾。

5. 参考文献

- 1) 村松、他：データパラレル計算モデルの提案、第36回全大、1988

- 2) 齊藤、他：分散型共有メモリのためのFORTRANインタフェース、第38回全大、1989
- 3) 前田、他：データパラレル並列計算機によるICCG法の並列処理、第38回全大、1989
- 4) 吉原、他：多次元FFTの並列処理、第38回全大、1989
- 5) 濱中、他：並列計算機H2Pのシステム構成、第38回全大、1989
- 6) 中越、他：並列計算機H2Pのプロセッサ間非同期データ転送方式、第38回全大、1989

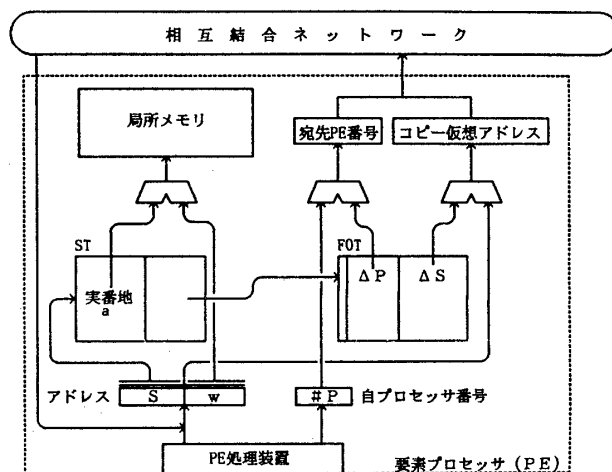


図1 アドレス変換とコピー制御

(ST:セグメントテーブル、FOT:ファンアウトテーブル)

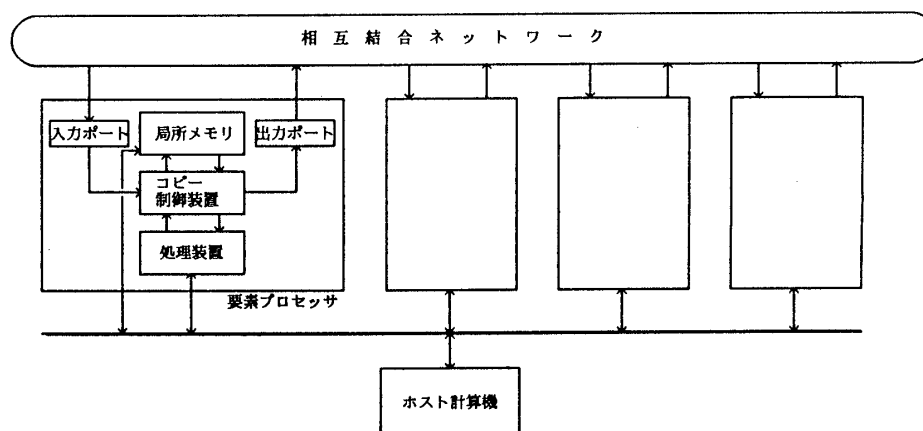


図2 データパラレル計算機