6T-7

並列計算機H2Pの要素プロセッサ間非同期データ転送方式

中越順二　田中 輝雄　濱中 直樹　面田耕一郎

(株)日立製作所 中央研究所

1. はじめに

並列処理では出来るだけ多くの要素プロセッサ(PE)を結合し,それらを効率よく並列に動作させることにより逐次処理に比べて飛躍的に高い性能が得られる可能性があると言われている.

しかし,実際には,PE間データ転送オーバヘッド,PE間の負荷の不均一,プロセス切り換えオーバヘッドなどの要因により,その効率が低下している.

本報告ではPE間データ転送オーバヘッドの削減に注目し,並列計算機H2P[1]で検討したPE間非同期データ転送方式について述べる.

2. 基本構成

並列計算機H2Pシステムの主要部は多数台のPEとネットワークから成り,図1のように構成される.

PEはn×nの2次元状に配置され,X方向とY方向のそれぞれに設けた小規模なクロスバスイッチに接続する.

クロスバスイッチの乗り換えはPEで行うが,PEの命令処理とは非同期に実行する.従って,他PEの命令処理を中断させることなく任意PE間でのデータ転送が可能となる(このネットワークを2次元ハイパクロスバスイッチ[2]とよぶ).このネットワークにおいて,たとえばPE-aからPE-dのデータ転送では2回の転送と1回のクロスバスイッチの乗り換えで可能である.

図2にPEの構成を示す.PEはノイマン型計算機で,PE内にローカル記憶を持ち,他PEとは独立に動作可能である.また,PEにはベクトル処理機能を内蔵し,科学技術計算を高速に行う.ネットワーク制御ではPE内の命令処理とは独立に,他PEとのデータ転送の送受信,および,X/Y方向のクロスバスイッチの乗り換えを行う.

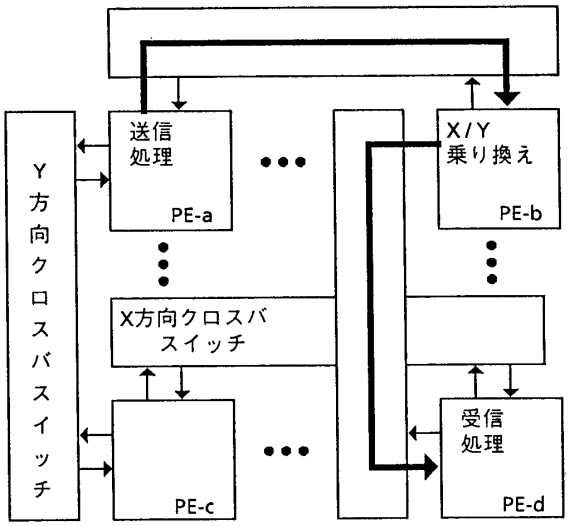


図1　主要部の構成とネットワークの動作

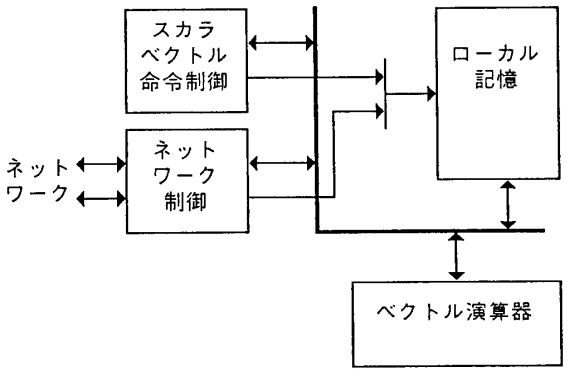


図2　PEの構成

3. PE間非同期データ転送方式

このデータ転送方式では,転送データにデータ識別子を付け,送信側PEと受信側PE間で非同期にデータ転送を行う[3].また,受信側PEには受信用のバッファを設け,受信側PEはデータ識別子によりそのバッファを検索する.

受信バッファはローカル記憶内に構成し,処理内容に応じて領域を確保する.受信バッファの各エント

Asynchronous Data Transfer Method of The Parallel Processor H2P,
Junji NAKAGOSHI, Teruo TANAKA, Naoki HAMANAKA, Koichiro OMODA
Central Research Laboratory, Hitachi Ltd.

リに対応して受信データが届いたか否かを示すタグをローカル記憶とは別に設ける.また,受信バッファは各PEで同一の領域とする.

以下,このデータ転送方式の手順について記す.

(1) 送信処理

送信側PEでは送信命令を解読するとその命令で指定される受信PE番号,データ識別子および転送データをネットワークに送出する.

また,この送信命令処理では受信バッファが各PEで同一の領域に確保されているためあらかじめ送信側PEでデータ識別子が正しく受信バッファの領域を示しているかをチェックすることができる.

これによりデータ識別子の誤りをいち早く検出でき,他のPEに悪影響を与えないため,並列プログラムのデバッグが容易となる.

(2) 受信処理1 (受信バッファへの書き込み)

転送データが受信側PEに届くと,図3に示すように,まず届いたデータ識別子をアドレスとしタグに1を書き込む.次にデータ識別子と受信バッファの開始アドレスを加算し,その結果をローカル記憶(受信バッファ)のアドレスとし受信データを書き込む.

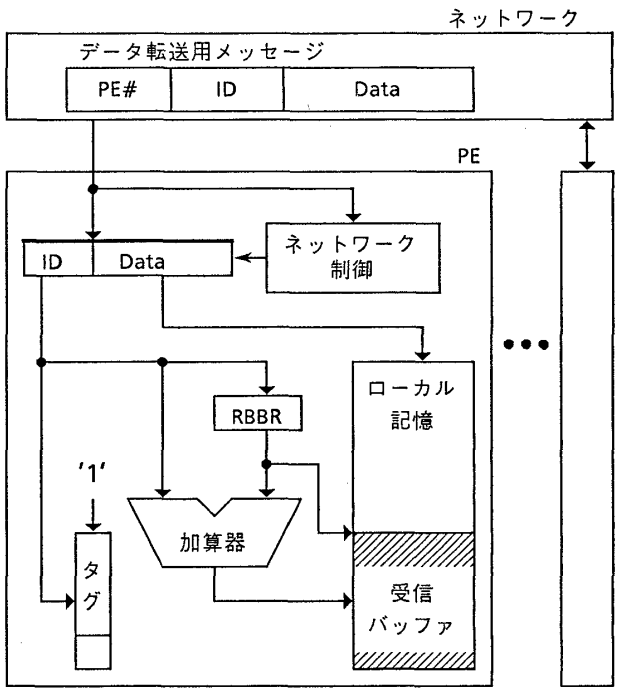
なお,この受信処理とは別に,非同期に行っている命令処理がローカル記憶をアクセスしている場合には,転送データの受信処理とでローカル記憶へのアクセスが競合する.このときは受信処理が遅れるが,それ以外は命令処理とオーバラップ可能である.

(3) 受信処理2 (受信バッファからの読み出し)

受信側PEが受信した転送データを必要とする場合は,受信命令で指定されたデータ識別子を用いてタグを読み出す.

読み出したタグが1(転送データが届いていることを示す)ならば,データ識別子と受信バッファの開始アドレスを加算し,その結果をローカル記憶(受信バッファ)のアドレスとし,転送データを読み出す.そして,後続の命令を実行する.タグが0(届いてないことを示す)ならば,転送データが到着するまで待つ.

上記ではスカラ処理の場合についてデータ転送方式を報告したが,ベクトル処理についても適用が可能である.



RBBR : 受信バッファの開始アドレスを示す
PE# : 受信PE番号
ID : データ識別子

図3 転送データの受信処理

このベクトル処理でのベクトルデータ転送では,転送データの到着待ちを必要最小限とするため以下のように制御する.

すなわち,ベクトルデータ全体の受信が終了していなくても一部のデータを受信すれば,後続の転送データの受信とオーバラップして,受信済みデータを用いた命令処理ができるようにする.

4. おわりに

本報告ではデータ転送オーバヘッドの削減に注目し,タグを用いたPE間非同期データ転送方式を提案した.

参考文献

[1] 濱中ほか,"並列計算機H2Pのシステム構成", 本大会予稿集掲載予定.

[2] 村松ほか, "キューブ系ネットワークの特性", 情報処理学会第37回全国大会予稿集 pp.192-193 (1988).

[3] 田中ほか, "データ転送オーバヘッドの削減を主眼とした並列アーキテクチャの提案", 情報処理学会第37回全国大会予稿集 pp.95-96 (1988).