

マルチPSIのネットワーク・ハードウェア構成

7B-2

益田 嘉直 瀧 和男 木村 康則
 (三菱電機) (財)新世代コンピュータ技術開発機構

1. はじめに

第五世代コンピュータ・プロジェクトの並列推論マシンの研究開発において、論理型プログラミングの並列実行に関するソフトウェア面での研究開発の重要性が認識されるようになって来ている。そのため、ICOTでは並列ソフトウェアの開発環境を構築するために、逐次型推論マシンPSIを複数台接続したマルチPSIシステムの開発を進めている。

マルチPSIシステムは第1版と第2版の2段階に分けて開発を進めているが[1],[2]、本稿で紹介する第1版は現在のPSIマシン6~8台程度を接続ネットワーク・ハードウェアにより格子状に接続した構成をとる。本稿では、ネットワーク・ハードウェアの設計及び試作を行ったので、その構成、ネットワーク制御方式及び制御ファームウェアについて述べる。

2. 接続ネットワーク・ハードウェアの構成

接続ネットワーク・ハードウェアはPSIマシン6~8台程度を格子状に接続するためのハードウェアで、CPUの内部バスのオプション・スロットに実装され、隣接の4台のプロセッサと接続するための4本のケーブルが引出される。これらの引出し口をそれぞれチャンネルと呼び、データはパリティビットを含め10ビット単位で並列転送されチャンネルの一方あたりあたりの転送能力は約500KB/secである(図1)。

各チャンネルは、それぞれ10ビットの送信データ線と受信データ線を独立に持っており、チャンネルの出入り口には受信インタフェース、送信インタフェースがあり受信、送信の転送制御を行っている。また、チャンネル信号線上のデータはハンドシェイク制御で10ビット毎に非同期並列転送さ

れる。接続ネットワーク・ハードウェアとCPUのインタフェース部分には4KBの容量を持つ送信用、受信用FIFO型のバッファ(CPU Readバッファ及びCPU Writeバッファ)を設置しており、また各チャンネルの送出部分にも256BのFIFO型のバッファ・メモリを設けた。これらのFIFO型バッファのサイズは性能評価その他のために変更可能としている。

エラー検出機能としては、パリティエラー、パケットエラーのチェック機能を備えており、動作状態でエラーが検出されると、接続ネットワーク・ハードウェアは停止状態になる。また、モード制御レジスタやシーケンス制御レジスタ等各種レジスタが搭載されているが、大部分のレジスタはファームウェアによりアクセスが可能である。

3. ネットワーク制御方式

マルチPSI第1版のネットワークの形状は格子型とし、プロセッサ間通信は共有メモリは置かず、メッセージ・パケットの交換による方式を採った。データ・パケットは可変長であり、各パケットの先頭には行先CPU番号の情報を持つ。10ビットのデータのうち最上位ビット(ビット9)はパリティで、ビット8はパケットの先頭/終了か、途中データかを示し、先頭/終了データの場合にはビット7で先頭/終了の区別を行う(図2)。

表1. 接続ネットワーク・ハードウェアの主な仕様

項目	内容
接続方式	格子型(メッシュ構造)
交換方式	メッセージ・パケット交換
データ転送方式	ハンドシェイク制御 (10ビット毎の非同期並列転送)
転送能力	約500KB/sec(各チャンネル)
最大ノード交換転送能力	2.5MBytes/sec
データ転送幅	10ビット(パリティビットを含む)
パケット形式	可変長(先頭に行先CPU番号の情報)
エラー検出機能	パリティエラー、パケットエラー (強制エラー機能有)

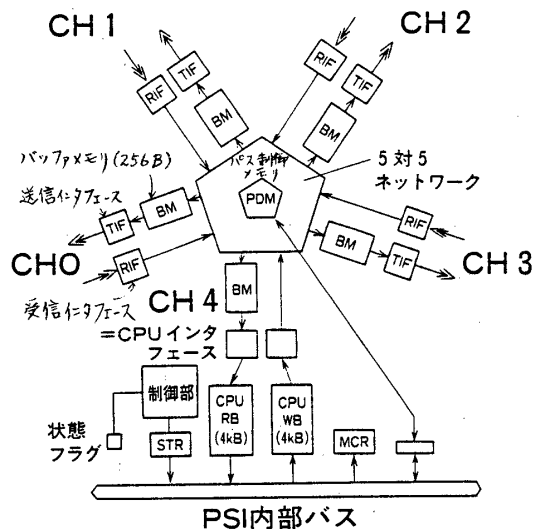


図1. 接続ネットワーク・ハードウェアの構成

ネットワークの構成は図1に示すように4本のチャンネル(CH0～CH3)とCPUインタフェースの5対5の接続形態となっており、行先CPU番号から送出先チャンネルを選択する操作、いわゆるルーチングの制御はバス制御メモリを用いて行われる。すなわち、各々のチャンネルから到着したパケットの行先CPU番号を、CPU番号と送出先チャンネルの対応表を記憶しているバス制御メモリを用いて認識し、それぞれ対応した送出先へ再送出する。この時、パケットがそれぞれ異なる送出先へ再送出され、行先チャンネルが競合しない場合は転送は同時並列的に実行され、各接続ネットワーク・ハードウェア(ノードに対応)当りの最大ノード交換転送能力は2.5MB/secとなる。この方式では横方向優先などの簡単なルーチング戦略を固定的に用いるだけでデッドロックが回避される。

接続ネットワーク・ハードウェアの各チャンネルの送出部分にはネットワーク上を流れるデータのよどみの解消や、転送効率の向上のためにFIFO型のバッファ・メモリが設けられているが、実際には1つのメモリを5つに分け、各々のチャンネル用として時分割で用いている。各々のチャンネルから入力されたデータがどのような手順で転送されるかを以下に示す。(シーケンス制御)

- (1) 接続ネットワーク・ハードウェアは動作状態に入るとCH0からCH3及びCH4(自CPUとのインタフェース)の順にデータ到着の有無をチェックし、次にCH0からCH4の順にバッファ・メモリ(BH)のデータが出力可能かどうかチェックする。
- (2) データが到着し、かつそれが最初のデータの時はバス制御メモリを引いて転送先チャンネルを求める。そして、そのチャンネルがビジーでなければ、チャンネルの使用権を得てデータ到着チャンネルからの転送経路を開設し、Busyフラグをセットする。すでにビジーであればビジーが解けるまで待たされる。
- (3) バッファ・メモリへのデータの書込み、及びバッファメモリからのデータ送出は10ビット単位で時分割で行われ、送出先ネットワークがビジーでないか、または行先が自CPUでCPU Readバッファがfullでない場合、データ送出は積極的に行われる。

この他、CPUへパケットを讀込む際には完全なパケットのみをCPU Readバッファから高速に讀込む必要があるので未完パケットカウンタやパケット到着フラグを設けている。

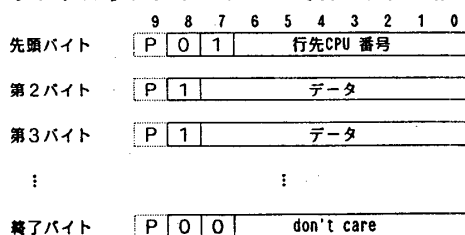


図2. パケット形式

4. ネットワーク制御ファームウェア

接続ネットワーク・ハードウェアには通常使用モードとハードウェアの初期化、エラー発生時の処理、保守時に使われる非通常使用モードがある。通常使用モードにおけるパケットの送出、受取りの処理はネットワーク制御ファームウェアによって行い、非通常使用モードにおけるレジスタ・アクセスの処理は標準の粗込述語で行っている。

通常使用モードで制御ファームウェアによって行われるパケットの送出、受取りの処理について述べる。

(1) パケットの送出処理

CPUからパケットをネットワークへ送出するためにはCPU内でパケットの形に整形されたデータを制御ファームウェアによりCPU Writeバッファに書込むことにより行う。この時、制御ファームウェアによりCPU Writeバッファの空き領域を調べてから実際の書込み処理を行う。また、パケットの先頭/終了マークバイトの付加なども制御ファームウェアで行う。

(2) パケットの受取り処理

ネットワークからCPUにパケットを受取るにはCPU Readバッファに転送されて来たパケットを制御ファームウェアでCPU内に読込むことにより行う。この時、CPU Readバッファ内には複数のパケットが入っていることが一般的であり、ネットワークから転送途中のパケット(未完パケットと呼ぶ)が入る可能性もある。従って、CPUにパケットを讀込む際には、まずCPU Readバッファに1ヶ以上の完全なパケットが到着したことを知らせるパケット到着フラグを制御ファームウェアで調べる。そして、このフラグがセットされている時のみ実際の読込み処理を行う。ここで未完パケットカウンタの内容を讀込むことにより、CPU Readバッファ内の完全なパケットのバイト数を計算する処理も制御ファームウェアで行う。

上記の処理を実現するために、各PSIマシン上にはネットワーク制御ファームウェアが粗込述語追加の形で実装される。

5. おわりに

マルチPSI第1版の接続ネットワーク・ハードウェアの構成、ネットワーク制御方式およびそのファームウェアについて述べた。現在はICOTにおいてPSIマシンを6台接続し、KL1処理系の開発およびテストを行っている。

最後に、ご指導ご鞭撻をいただいたICOT第4研究室内田俊一室長ならびに関係各位に深謝する。

〈参考文献〉

- [1] 瀧 和男、"並列ソフトウェア開発用マシン：マルチPSIシステム" 第32回情報全国大会、昭和61年3月
- [2] 木村 康則、他、"マルチPSIシステムとその接続方式"、本大会予稿集7B-1、昭和61年10月