可変構造型並列計算機のメッセージ・コプロセッサ

4L－8

森 眞一郎　上野智生　村上和彰　福田 晃　富田眞治
(九州大学)

1 はじめに

現在我々は，128台のPE（Processing Element）を，128×128の多重化クロスバー網（MC-net）で相互接続したマルチプロセッサシステム「可変構造型並列計算機」を開発中である.[1] 本システムは「適応型並列計算機」の構築を目的としたシステムであり，これを実現するために，相互結合網およびメモリにダイナミック・アーキテクチャを適用している．したがって任意の相互結合網を構成し，その上で密結合型，疎結合型あるいは両者混合型のマルチプロセッサを実現することが可能である．

本稿では，まず本システムのすべてのPE間の通信をつかさどるメッセージ通信ユニット（MCU）について述べ，次にメッセージ交換時にプロセッサのオーバヘッドを軽減するためのメッセージ・コプロセッサについて述べる．

2 メッセージ通信ユニット（MCU）

MCUは，プロセッサ・ユニット（PU）と多重化クロスバー網（MC-net）とのインターフェースとして，PUと並列に通信処理を行う.[2] MCUは図1に示すように，メッセージセンダ（MS），メッセージレシーバ（MR），および，メッセージ・コントローラで構成され，おのおの独立に動作可能である．

本システムでは，MC-net上で，高スループットを要求するプロセス間メッセージ交換，および，低遅延を要求するリモートメモリ・アクセス，といった性質の異なるPE間の通信を実現しなければならない．MCUは，これら全てのPE間の通信を3階層から成る通信プロトコルにより実現する．

MCUとMC-netとの物理的なインタフェースを行う第1層，通信を行うMCU間の論理的なインターフェースを行う第2層，およびPUとMCUとの物理的なインタフェースを行う第3層である．MCUは，これらのすべてのプロトコル処理をハードウェアにより実現している．さらに，PU上のプログラムに対してどの階層を見せるかで，MCUは図2に示す3種類のインタフェースを提供する．

①I／Oデバイス・インタフェース：PUが直接操作することが可能な最も低レベルなインタフェースである．このレベルの通信は後述するプリミティブ・メッセージの交換によってのみ可能である．

②リモートメモリ・インタフェース：他PEのローカルメモリにMC-netを介してアクセスする場合のインタフェースである．PUに対して通常のメモリ・アクセスと全く同様のインタフェースを提供する．

③メッセージ・コプロセッサ・インタフェース：MCU内のメッセージ・コントローラが提供する．このインタフェースに関しては以下で詳しく述べる．

3 メッセージ・コプロセッサ

一般の疎結合型マルチプロセッサでは，プロセス間メッセージ交換機能は通常OSの通信ハンドラが提供している．しかし，アプリケーションプログラムと通信ハンドラが同一プロセッサ上で動作することから，プロセス間メッセージ交換に伴う通信ハンドラの処理がオーバヘッドとなる可能性がある.[3]

そこで本システムでは，このオーバヘッドを軽減するため，プロセッサと並行動作可能で，通信ハンドラの機能を一部オフロードしたメッセージ・コプロセッサを導入した．

メッセージ・コプロセッサの機能としては，低遅延，高スループットなメッセージ処理機能に加え，本システムに特有な時分割多重化クロスバー網（MC-net）[4] の動作を鑑みた機能を設けている．以下に主な機能を示す．

①メッセージの解釈，処理
②メッセージの伝送
③バッファの管理およびメッセージのキューイング
④メッセージのマルチキャスト
⑤誤り検出および訂正
⑥時分割透過性の保証
⑦メッセージのルーティング
⑧プロセス間の同期

このような機能を実現するにあたり，高速かつ柔軟な処理を可能とするため，メッセージ・コプロセッサはマイクロプログラム制御による構成とした．また，メッセージ・コントローラとして1チップ・マイクロコントローラ（WSI社製PAC1000 [5]）を採用することで，低コストおよび低実装面積のコプロセッサを実現した．PAC1000のマイクロアーキテクチャは，64ビット巾水平型マイクロ命令，16ビットシーケンサ，16ビットALU，制御記憶容量1k語となっており，動作周波数はPUと同じく16.7MHzである．

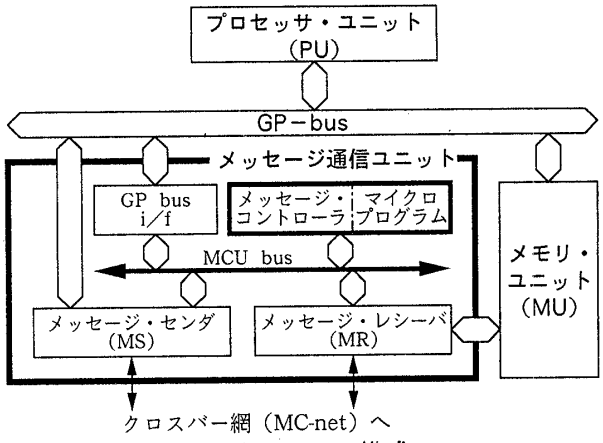

図1．PEの構成

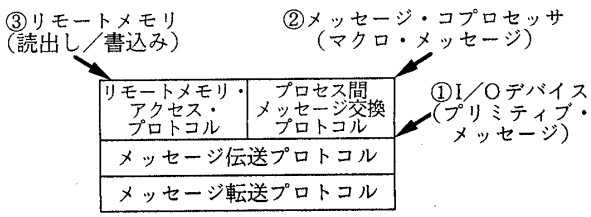

図2．MCUが提供するインタフェース

Message Coprocessor of the Kyushu University Reconfigurable Parallel Processor
Shin-ichiro MORI, Tomo-o UENO, Kazuaki MURAKAMI, Akira FUKUDA, Shinji TOMITA
Kyushu University

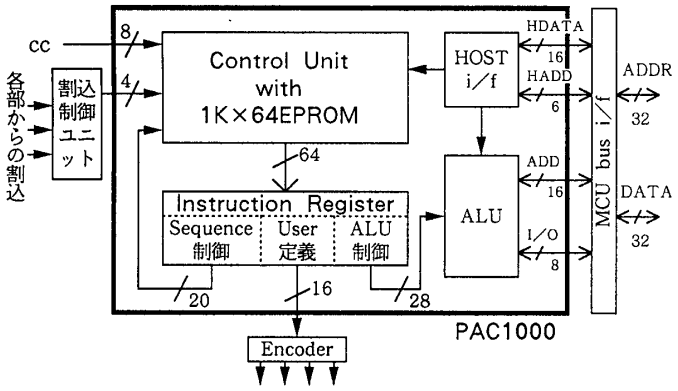

図3. メッセージ・コプロセッサの概略図

メッセージ・コプロセッサは，MSあるいはMRからの割込みが生じた場合，あるいはPUからのコマンド（マクロ・メッセージと呼ぶ）を受理した場合，に起動されメッセージ処理を行う．この処理に際し，PUとのメモリ競合を避けるためMCU内に専用のメモリ（受信バッファと併用，デュアルポート構成，容量64kB）をもたせ処理効率の向上を図っている．

4 メッセージ通信プリミティブ

メッセージ・コプロセッサはPUからマクロ・メッセージを受け取ると，そのメッセージを解釈し，処理を開始する．このとき，他PEのMCUとの通信が必要であれば，MCUのI／Oインターフェースを介してプリミティブ・メッセージを交換する．

以下，このプリミティブ・メッセージ，およびマクロ・メッセージについて述べる．

4. 1　プリミティブ・メッセージ

MCUがI／Oインタフェース・レベルでサポートするメッセージは大きく2種類に分けられる．

①プロセス間メッセージ，制御メッセージ

通常のデータ送出メッセージのほかに，他PEへの割込み，リセット，またはホルト，といった要求を伝えるためのメッセージが含まれる．いずれもOSまたはメッセージ・コプロセッサが明示的に伝送を依頼するメッセージである．

②SMWアクセス・メッセージ [2]

本システムを密結合システムとして動作させるときに必要なメッセージであり，共有メモリへのREAD／WRITEアクセス

表1 マクロ・メッセージ

①SEND, RECEIVE
SEND ({PEID | GRPID}, SIZE, ID, MSGPTR)
SHORT-SEND ({PEID | GRPID}, SIZE, ID, MSG)
AUTOROUTE-SEND ({PEID | GRPID}, SIZE, ID, MSGPTR)
MSG-RELAY ({PEID | GRPID}, SIZE, ID, MSGPTR)
RECEIVE (PEID)
②QUEUE操作
QALLOC ({PEID | GRPID})
QSTAT ({PEID | GRPID})
QDEL ({PEID | GRPID | *}, LVL)
GET-QLIST ()
③グループ定義
DEF-GRP (GRPLIST)
ADD-GRP (GRPID | GRPLIST)
④BLOCK転送
BLOCK-TRANSFER (TYPE, DEST, SRC, SIZE [, KEY])
⑤同期，制御
INT ({PEID | GRPID}, LVL)
MISC ({PEID | GRPID}, TYPE)

メッセージ等，7種のメッセージがある．この種のメッセージはPUのリモートメモリ・アクセスに伴いMCUで暗黙的に作成されるメッセージである．

4. 2　マクロ・メッセージ

メッセージ・コプロセッサが提供するコマンドである．表1に現在提供しているマクロ・メッセージを示す．以下これらについて簡単に説明する．

①SEND, RECEIVE：PUがメッセージの送受を行うためのコマンドである．メッセージは基本的にFCFSで処理されるが，メッセージサイズが32B以下のメッセージに対しては優先処理の指定ができる．また，MC-netが時分割多重化動作 [4] する場合，メッセージをあて先PEに直接送ることができない場合があるが，その場合には，ルーティングおよび中継を行うマクロ・メッセージを使用する．

②QUEUE：キューの割当て，削除等を行う．MC-netが時分割多重化動作する場合，PUは通信を行う必要があるPE対応に予めキューを割り当てる．キューが割り当てられると，PUは実際に通信が必要となった時点で，MC-netの動作を意識しなくてすむ．また，割り当てられたキューの管理はすべてメッセージ・コプロセッサが行う．ただし，MC-netを時分割多重化動作させない場合は，コプロセッサが静的に割り当てるキューのみを使用する．

③グループ定義：MC-netが提供するマルチキャスト機能を利用するためのグループ定義を行う．一度グループ定義を行うと，それ以後マルチキャストは1対1通信の場合と同一のコマンドのみで実行できる．

④BLOCK転送：メッセージ・コプロセッサをDMACとして利用する場合に使用する．ローカル・メモリが持つBLOCK－READ／WRITE機能を利用して高速なDMA転送を行う．

⑤同期，リモート制御：PE間の割込み，他PEの制御を行う．

これらのマクロ・プリミティブの機能は，マイクロプログラムにより柔軟に設定することが可能であり，必要に応じ順次，拡充および最適化を行う予定である．

5 まとめ

以上，可変構造型並列計算機のメッセージ・コプロセッサについて述べた．本システムを疎結合型マルチプロセッサとして使用した場合，プロセッサの処理と並列に，メッセージ・コプロセッサで通信処理を行わせることで，処理の高速化が図れる．現在，コプロセッサのマイクロプログラムを開発中である．

参考文献

[1] Murakami, K. et al.: The Kyushu University Reconfigurable Parallel Processor－Design Philosophy and Architecture－, Proc.IFIP 11th World Computer Congress, pp.995－1000 (1989).

[2] 森ほか："可変構造型並列計算機のPE間メッセージ通信機構"，情処論文誌,Vol.30, No.12 (1989)

[3] Peterson,J. et al.: A High－Speed Message－Driven Communication Architecture, Proc.1988 Int.Conf. Supercomputing, pp.355－366 (1988).

[4] 蒲池ほか：可変構造型並列計算機のネットワーク制御方式，信学技法，CPSY89－16 (1989).

[5] Waferscale Integration, Inc.: High－Performance Programmable Standalone Microcontroller (PAC), Waferscale Integration,Inc.(1988).