3T-3

可変構造型並列計算機の通信システム

森眞一郎　甲斐康司　村上和彰　福田晃　末吉敏則　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：Multiplexed Crossbar Network）で相互接続し，PE間結合形態およびメモリ構造を可変と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．本システムではPE間通信をメッセージ交換により実現している．本稿では，このメッセージ交換を行なうための具体的な通信制御方式，および通信システムについて述べた後，その性能予測を行なう．

2. 通信制御方式

本システムでは，当初，MC-net上での通信プロトコルとして，SEND信号とACK信号のハンドシェイクによるクロック非同期転送方式を採用する予定であった．しかし，この方式では，クロスバーLSIおよびPEのMC-netインターフェースの動作周波数を20MHzとしても，MC-netの最大転送速度は10MByte／secにも満たず，クロスバーLSIの性能を十分引き出すことができない．これは，ハンドシェイク方式では，PE－クロスバーLSI間を接続するケーブルの遅延時間（往復で100nS以上の見積り）が，SEND信号のサイクル時間を決定することによる．このような性能低下を防ぎ，クロスバーLSIを最大限に動作させるために，ハンドシェイク方式をやめてパイプライン制御によるクロック同期転送方式を採用することにした．この方式は，PE間を結ぶMC-net上の通信路を遅延要素ごとに分割して，これを1つのパイプラインステージと見なし，ステージ間にあるラッチを同一クロックで制御する方式である．このときパイプライン・ピッチは各ステージの遅延のうち最大のもので決まる．

本MC-netでは図1に示すように，通信路を

①MCU（MS）－クロスバーLSI入力ポート間

②クロスバーLSIの入出力ポート間

③クロスバーLSI出力ポート－MCU（MR）間

の3つのステージに分割し，3段パイプライン構成としている．また，MC-netは半二重回線であるためパイプライン上でも双方向のデータ転送を可能にしなければならない．そこで，本システムでは，ハンドシェイクによる方向制御シーケンスをデータ転送とオーバーラップして実行し，これを可能にしている．なお，クロック周波数は20MHzとし，最大転送速度20MByte／secを予定している．

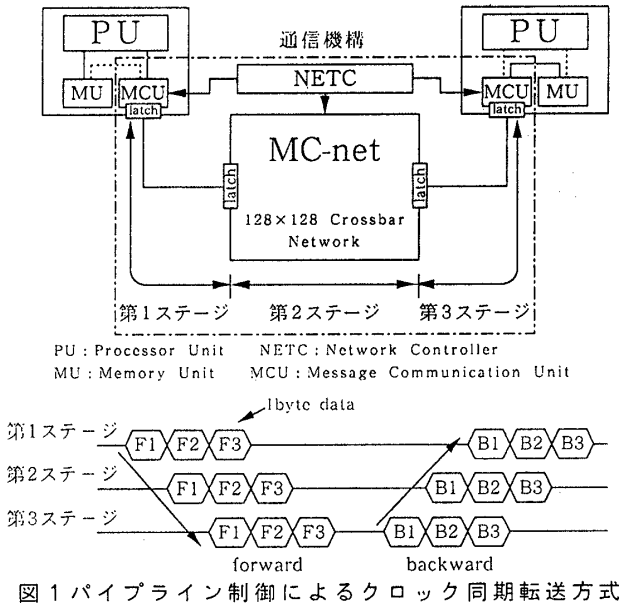

図1 パイプライン制御によるクロック同期転送方式

3. 通信システム

本システムの通信システムは図1に示すように，大きく3つに分割できる．以下，それらの機能について簡単に説明する．

3.1 ネットワーク・コントローラ（NETC）

通信システム全体を制御するユニットである．ホストからの指示に従ってネットワークを任意の構成に設定するよう，集中制御を行う．具体的には，通信機構を統括する2種類のシステムクロックの作成，MC-netのモード制御などの機能がある．

3.2 多重化クロスバー網（MC-net）

8×8のクロスバーLSIを256個用いて構成した128×128のクロスバー網である．MC-netは，各クロスバーLSIがもつ時分割多重化機能（プリセットモード時）により，各種のPE間結合形態をエミュレートすることができる．以下MC-netの3つの動作モードについて述べる．

①デマンドモード：PEからの回線接続要求に対して回線の接続を動的に行なう．

②プリセットモード：プログラムの実行前に，PE間の接続形態を時系列としてクロスバーLSI内のスイッチ制御メモリに記憶させ，実行時には，このメモリの内容を順次読み出し，その内容に従って，クロスバーLSI内およびMC-net全体の回線の接続を行う．

Communication System of the Reconfigurable Parallel-Processor

Shin-ichirou MORI, Kouji KAI, Kazuaki MURAKAMI, Akira FUKUDA, Toshinori SUEYOSHI, and Shinji TOMITA
Kyushu University

③パーティション・モード：MC-netをデマンドモードで動作する領域とプリセットモードで動作する領域とに分割し、両モードが同時に混在することを許すモードである.

3.3 メッセージ通信ユニット（MCU）

MCUはPUとMC-netとの間のインターフェースの役割を果たしており，以下の3つの機能を提供している.

①OSのメッセージ交換機能のサポート：パケットの作成・分解，転送などの基本的な機能に加え，放送機能やマルチキャスト機能，さらにキューイング機能などをサポートする.

②MC-netとのインターフェース：NETCからの指示に従い，MC-netの動作モードに合ったデータ転送プロトコルでメッセージの転送を行なう.

③共有メモリウインドウ（SMW）へのアクセス：本システムは，分散メモリ構成を採るが，他PEのローカル・メモリ（リモート・メモリと呼ぶ）を直接アクセスすることを許している. MCUは，MC-netを用いてリモート・メモリ・アクセスを実現する.

4. 交換メッセージの種類

通信システムが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：通常のデータ送出メッセージの他に，他PEへの割込み，リセット，またはホールト，といった要求を伝えるためのメッセージが含まれる.

②SMWアクセス・メッセージ：メモリ共有型密結合マルチプロセッサにおいて必要なメッセージであり，SMWへのREAD／WRITEアクセス・メッセージ，Atomic LOAD−STORE（不可分）メッセージ，これらのアクセスに対するREPLYメッセージ，および2種類のキャッシュパージ・メッセージの5つのメッセージがある.

5. 通信システムの性能予測

以上述べた通信システムについて，プリセットモード時の性能予測を次の2つの項目について行なう.

5.1 プロセス間メッセージの転送速度

プロセス間メッセージ転送時のMC-netの実効転送速度を図2に示す. ここでは，受信側MCUの受信バッファには十分な空き領域が存在すること，を仮定している.

MC-netの物理的な転送速度20MByte／secに対して256Byteデータ転送時の実効転送速度（最大実効転送速度）は19.0MByte／secとなる。よって，MC-netの利用効率は95%となり，最大実効転送速度はハンドシェイクによる非同期転送時の2.1倍に向上する.

5.2 SMWアクセス時のアクセス時間

図3はアクセスサイズとアクセス時間の関係を表わしている. ただし，メモリ側PEでの内部バス競合は起こらないことを仮定している.

図3よりSMWアクセス時のキャッシュ，ローカルメモリ，グローバルメモリへの4Byte READアクセス時間比は約1：3：24となることがわかる. 本システムのキャッシュシステムはストアスルー方式を取っているため，実際には8Byteを超えるWRITEアクセスは生じない，従ってMC-netを経由したSMWアクセス時のアクセス時間はREADアクセス時1.31 μS（1Byte時）～2.86 μS（32Byte時），WRITEアクセス時1.31 μS（8Byte時）となる.

6. まとめ

以上，可変構造型並列計算機の通信システムについて述べた. パイプライン制御によるクロック同期転送方式を用いることにより2倍の通信速度が期待される.

参考文献

1）村上ほか："可変構造型並列計算機のシステム・アーキテクチャ"，情報処理学会「コンピュータアーキテクチャ」シンポジウム論文集，Vol.88，No.3，pp.165−174（1988年5月）

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能"，情報処理学会「並列処理」シンポジウム論文集（1989年2月）

3）濱口ほか："可変構造型並列計算機のプロセッサ・ユニット"，本大会論文集（1989年3月）

4）蒲池ほか："可変構造型並列計算機のメモリ・アーキテクチャ"，本大会論文集（1989年3月）

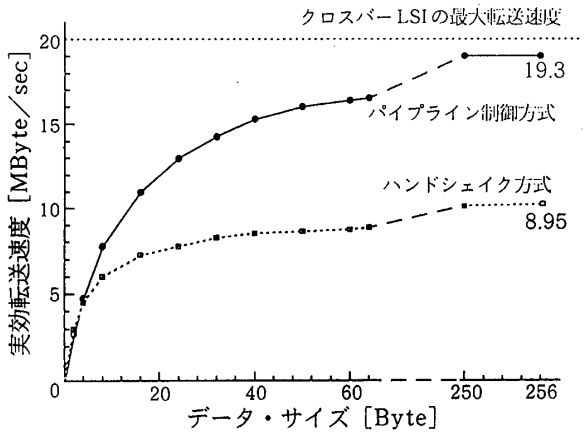

図2 プロセス間メッセージの実効転送速度

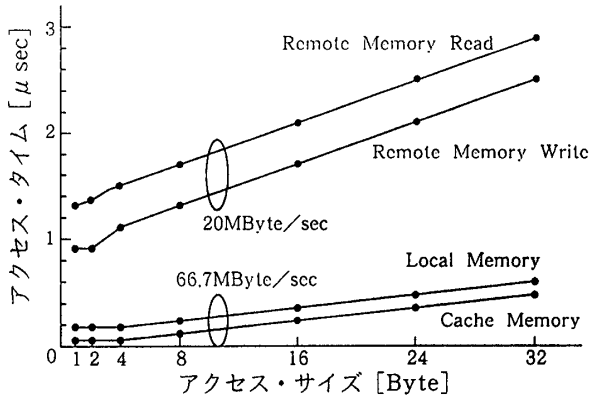

図3 共有メモリへのアクセス・タイム