マルチメディア通信と分散処理　34－6
（１９８７・７・２４）

階層型通信メモリバス方式（Ｈ－ＣＯＭ）を用いたプロセッサ間通信方式

加久間　勝　、　中川　秀敏　、　白石　一彦

（金沢工業大学）

プロセッサ数の多いマルチ・マイクロ・プロセッサ・システムを対象として、多数のプロセッサ間の情報交信を、効率的にかつ経済的に行える新しいプロセッサ間通信方式を提案している。第一は、対称個別通信メモリバス方式（ＳＩＣＯＭ）である。これは、各プロセッサ間をＩＣＭと呼ぶユニットで完全結合したものであり、処理のオーバーヘッドが小さく、高速な通信が期待できる。第２は、階層化した通信メモリバス方式（Ｈ－ＣＯＭ）である。これは、各プロセッサ間を疎結合にし、通信制御プロセッサ（ＣＰ）と主通信制御プロセッサ（ＭＣＰ）を経由させることにより、ＩＣＭの数を大幅に削減し、経済化を図っている。

"INTER-PROCESSOR COMMUNICATION SYSTEM BY HIERARCHICAL COMMUNICATION MEMORY PATH SYSTEM" (in Japanese)

by Masaru KAKUMA, Hidetoshi NAKAGAWA, Kazuhiko SHIRAISHI

(Kanazawa Institute of Techonology, 7-1,Ogigaoka,Nonoichi-machi Ishikawa-ken, 921, Japan)

THis paper discribes two new inter-processor communication systems for the multi-micro processor system. The one is called SICOM(Symmetrical Individual Communication Memory Path System). In this system,all proces-sors are connected by ICM(Individual Communication Memory) each other. SICOM system is expected less communication delay time and less process-ing over-head. The other is called H-COM(Hierarchical Communication Memory Path System).This system consists of three stages that are perip-heral processor(PE),communication processor(CP) and main communication prosseor(MCP). PE is connected to CP through ICM,and CP is connected to MCP through ICM. H-COM is less expencive system than SICOM.

1・まえがき

電子計算機が商用化されて以来、単体としての電子計算機の処理能力の向上に、不断の努力がはらわれてきた。その発展に寄与したのは、主として電子部品技術の進歩によるものであって、計算機アーキテクチャの変革によるものではなかった。しかし、最近のＬＳＩ技術の進歩によって、これまでのシングルプロセッサの構成から、マイクロプロセッサ程度の処理装置を多数使用し、これらの処理装置を並行実行させる事により、電子計算機の処理能力を向上させる複合計算機システムが着目されるようになってきた。これらのシステムには、ノイマン型計算機をベースとした負荷分散を図った負荷分散型マルチマイクロプロセッサシステムや、処理装置にそれぞれ異なる機能を分担させて並行処理する機能分散型マルチプロセッサシステムがあり、さらに並行処理が容易なデータ駆動原理を基礎にした非ノイマン型のデータフローマシンなどがある。

これらのシステムは広義の意味でのマルチプロセッサシステムであり、処理装置相互間で情報を更新しながら処理を実行するから、処理装置相互間の情報交信（以下プロセッサ間通信方式と称す）の性能の是非が、システムの性能を支配する重要な要因になるものと考えられる。

今後は処理装置（以下プロセッサと称す）の数がますます増大するものと考えられるので、これら多数のプロセッサ間の通信を効率的にかつ経済的に行える新しいプロセッサ間通信方式の開発が必要になってくる。この場合、システム開発を容易に進めるために各プロセッサの内部処理とプロセッサ間の通信制御との独立性を高めること、及び通信時のプロセッサでのオーバーヘッドを大幅に減少させることが重要である。

プロセッサでの独立性を高める方法としては、プロセッサ間通信に関する制御を、各プロセッサでの内部処理と全く分離し、専用のハードウェアで行う方法が考えられる。本論文で述べている対称個別通信メモリバスシステム（ＳＩＣＯＭ）と階層型通信メモリバスシステム（Ｈ－ＣＯＭ）はこの理念を基礎としたものである。

まずＳＩＣＯＭ方式は、個別メモリユニット（ＩＣＭ）と称する２者間のアクセス制御機構を持つ、一種の共有メモリを、各プロセッサ間に配置して相互通信を行う方式である。この方式では、各プロセッサ間にそれぞれ個別の通信バスが設置されるので、

（１）　バス系の輻輳による待ち合わせが発生しない。
（２）　データ転送制御手順が不要になる。

などの利点が生じ、大幅なオーバーヘットの削減が可能になる。　しかし、ＳＩＣＯＭ方式は各プロセッサ間に個別にＩＣＭユニットを配置するため、ｎ台のプロセッサの場合、${}_nC_2$個すなわちｎ（ｎ－１）／２個のＩＣＭユニットがn^2に比例して増大し、経済的な問題が生じる。そこで、ＳＩＣＯＭ方式の構成を変え、ｎ台のプロセッサに対してｎ台程度のＩＣＭユニットで相互通信を可能にする方式として、Ｈ－ＣＯＭを考案した。

ＳＩＣＯＭ方式は完全対称（密結合）の構成を採っているのに対し、Ｈ－ＣＯＭ方式はＳＩＣＯＭ方式の利点を十分に生かし、２階層からなるモジュール結合（疎結合）の構成を採っている。　これは、１６台の周辺プロセ

セッサをＩＣＭユニット結合した、通信制御プロセッサＣＰを１つのグループとし、更にこれを１６個のＩＣＭユニットで結合した主通信制御プロセッサＭＣＰの２段構成となり、２７２個のＩＣＭユニットを用いて最大２５６台のプロセッサ間通信を可能にしている。

２．　ＳＩＣＯＭ方式

２．１　ＳＩＣＯＭ方式による電子計算機システムの構成

ＳＩＣＯＭ（Symmetrical Indived-ual Communication Memory Path System）によるシステム構成を図１に示す。各周辺プロセッサ（ＰＥ）はプロセッサ間に配置された個別通信メモリユニット（ＩＣＭユニット）によって完全結合の構成をとり、プロセッサ間の通信はこのＩＣＭユニットをアクセスすることにより行われる、実行プログラムはホストコンピュータ（ＨＣ）より、それぞれＨＣとＰＥに対応するＩＣＭユニットに分割されて供給される、各ＰＥはその供給されたプログラムに基づいて各ＰＥとプロセッサ間通信を行いながら処理を進めていく、その実行結果はＩＣＭユニットを通してＨＣへ送られるようにシステムが構成されている。　このように各プロセッサは内部処理を中断することなくプロセッサ間での通信を行うことができる。また、各プロセッサはＩＣＭユニットにより完全結合の構成を採っているためハードウェア構成を変えることなくソフトウェアの変更のみで、アレイプロセッサやパイプラインプロセッサは勿論、データフローマシンとしても使用することが可能である。

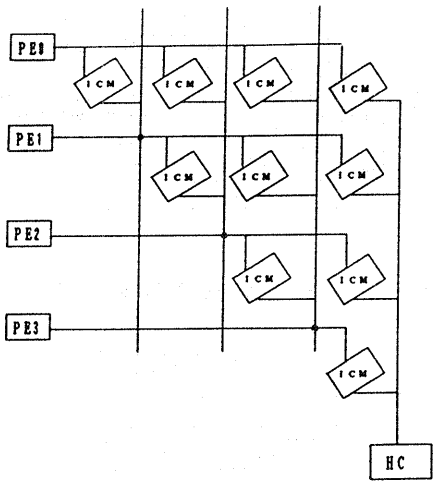


図１　ＳＩＣＯシステム構成

２．２　ＳＩＣＯＭの動作

ＩＣＭユニット内のデータの流れを図２に、制御動作フローを図３に示す。

各ＰＥから対応する通信メモリをリードするときは、相手の状態をかく乱することなく同時にかつ自由にアクセスできる。ＰＥから通信メモリにライトされたデータはメモリに格納されると同時にバッファに対してもそのデータとアドレスがセットされる。さらにその対応するＰＥの通信メモリへそのＰＥがメモリフェッチしていない周期を利用して、バッファ内のアドレスに基ずいて対応する通信メモリへデータを転送する。試作したシステムのサイクルスチールのタイミングチャートを図４に示す。

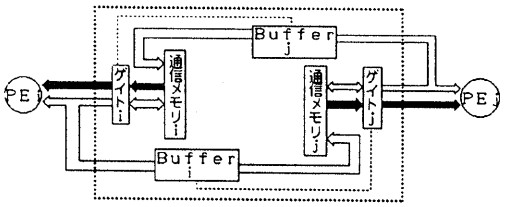


図２　データ転送過程概念図

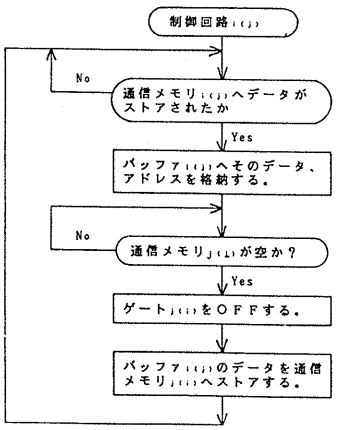


図３　制御動作フロー

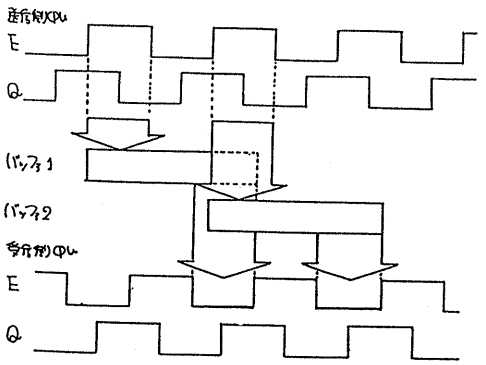


図４　サイクルスチールの
タイミングチャート

２．３　ＩＣＭユニットの構成

図５にＩＣＭユニットの全体ブロック図を示す。ＩＣＭユニットは図１で示した様に２つのＣＰＵの間で双方向通信を可能にするユニットでありどちらのＣＰＵも送信側、または受信側になることがある、従ってＡブロック、Ｂブロックが完全対称の構成となり、各ＰＥ間に配置され相互通信を可能にする。

メモリーコントロール回路は、ＣＰＵアクセスの時、及びサイクルスチールの時、両方の信号を得てメモリのチップセレクト信号及びライト信号を生成するためのものです。

リクエスト回路は、ＣＰＵからの書き込みが発生した場合、それを相手側のスチールコントロール回路に知らせる信号と相手側のバッファへの書き込みタイミングを発生する。

スチールコントロール回路では、相手側のリクエスト回路で発生した信号を受けて、スチールコントロール信号を発生すると共に２つのバッファへの切り替えを行う。

バスセレクト回路は、メモリに接続されているアドレスバス及びデータバスを、ＣＰＵからの書き込み時にはＣＰＵに、メモリースチール時にはバッファからの出力へ切り替えを行う。

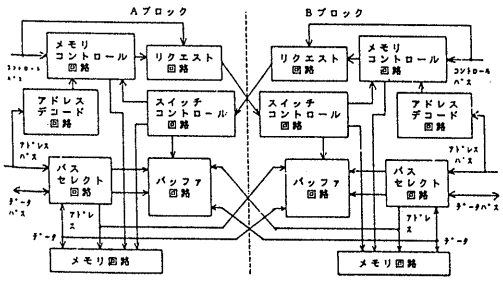


図５　ＩＣＭユニットのブロック図

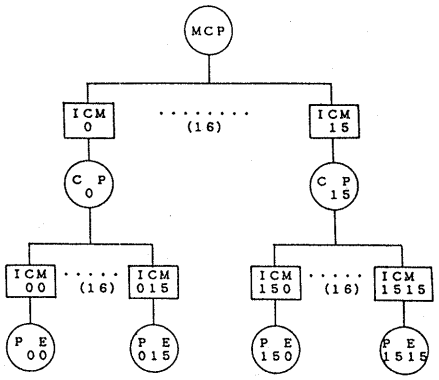
3. Ｈ－ＣＯＭ方式

3．1 Ｈ－ＣＯＭによる電子計算機システムの構成

Ｈ－ＣＯＭ（Hierachical Communication Memory Path System）によるシステム構成を図６に示す。Ｈ－ＣＯＭは１６台の周辺プロセッサをＩＣＭユニットにより結合した通信制御プロセッサ（ＣＰ）を１つのグループとし、さらにこれを１６個のＩＣＭユニットで結合した、主通信制御プロセッサ（ＭＣＰ）の２段構成となる。従って２７２個のＩＣＭユニットを使い２５６台のプロセッサ間通信が可能となる。

１つのグループ内の通信はＰＥ－ＩＣＭ－ＣＰ－ＩＣＭ－ＰＥの順で通信が行われ、他グループとの通信にはＰＥ－ＩＣＭ－ＣＰ－ＩＣＭ－ＭＣＰ－ＩＣＭ－ＣＰ－ＩＣＭ－ＰＥの手順で通信が行われる。

ＣＰとＭＣＰは通信制御のみを専用に行うプロセッサであり、今回の試作ではＰＥと同じマイクロプロセッサを用い、プログラム制御方式によって通信制御を行っている。データ転送遅延時間に関する性能の面では、ＣＰ及びＭＣＰの性能が支配的になる。当初から予想したとうりであるが、通信制御をソフトウェアで行う今回の方式は、転送遅延時間が大きくなり、スループットも減少し、実用性の面では問題がある。しかし、この通信制御のアルゴリズムは簡単なものであるから、ＣＰ及びＭＣＰを電子回路による布線論理制御方式で実現し、性能の大幅な改善を図ることは、比較的容易であると想定している。



ＣＰ：通信制御プロセッサ
ＭＣＰ：主通信制御プロセッサ

図６ Ｈ－ＣＯＭのシステム構成図

3．2 通信制御方式

プロセッサ間通信の制御を行う方式として、リングバッファ方式とフラグセンス方式を検討した。

3．2．1 リングバッファ方式

（１）構成の概要

リングバッファ方式は、ＩＣＭの通信メモリ上にリングバッファを構成しメッセージパケット単位でデータの転送を行うものである。図７にメッセージパケットの形式を示す。パケット形式はヘッダが３バイト、データ（ＤＡＴ）が最大６１バイトまでの可変長とする。ヘッダ情報は、受信ＰＥ番号（ＲＮ）、送信ＰＥ番号（ＳＮ）、メッセージデータ長（ＬＥＮ）から成り、ＰＥ番号は上位４ビットがグループ番号、下位４ビットがグループ内の相対ＰＥ番号（０～１５）を表現する。

ＩＣＭ上でのリングバッファの構成を図８に示す。バッファは固定長のバッファを複数個組合せ、読み出し用ポインタ（ｒ）と書き込み用ポインタ

（ｗ）を付けリングバッファを構成する。　リードポインタ（ｒ）とライトポインタ（ｗ）は、図の矢印の方向へ進むが互いに相手を追い越すことはな区、待ち行列制御をＩＣＭ内で行うことができ、プロセッサの通信からの独立性が高くなり、非同期通信が可能と成る。　なお、リードの時はポインタを１つ更新した場所より読み込み、ライトの時はそのポインタの示す場所に書き込む。実際にＩＣＭ上では通信メモリを送信側と受信側に分割してそれぞれにリングバッファを構成する、又リングバッファを制御するリードポインタ、ライトポインタも同じく通信メモリ上に配置する。

（２）　動作の概要

ＣＰは各ＰＥの通信バッファのポインタをルックイン方式で走査し通信要求が発生していたならば受信プロセッサへ転送する。その時、受信プロセッサがグループ内であるかグループ外であるか判断し、グループ内であればその宛先のプロセッサへパケットを転送する。グループ外であるならＭＣＰにパケットを転送する、以後この動作を繰り返す。ＭＣＰは各ＣＰからの送信バッファをルックイン方式で走査し通信要求が発生したならば宛先となるプロセッサのあるグループのＣＰへパケットを転送する、以後この動作を繰り返す。

ＰＥｉからＰＥｊについての転送手順について述べる。この場合ＰＥｉとＰＥｊが同一グループ間での転送なのか、他グループに対しての転送かで、その通信処理は約３倍に増加する。

（Ⅰ）ＰＥｉはＩＣＭｉにメッセージパケットを生成し、送信側ライトポインタを更新する。

（Ⅱ）ＣＰはＩＣＭｉから通信要求を確認すると、宛先のＩＣＭｊが転送可能か判断し、可能であれば転送を行い、ＩＣＭｉの送信側リードポインタとＩＣＭｊの送信側ライトポインタを更新する。

（Ⅲ）ＰＥｊは受信要求を確認するとＩＣＭｊにメッセージパケットを取り込み、ＩＣＭｊの受信側リードポインタを更新する。

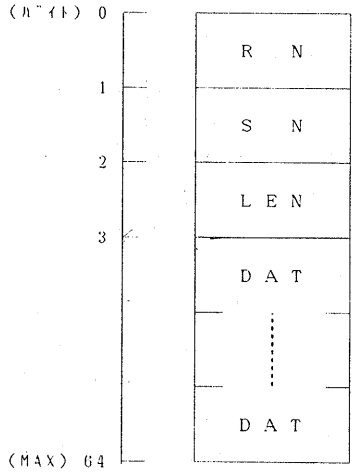


図７　パケットの形式

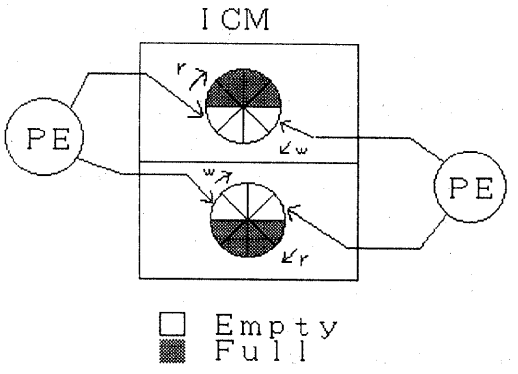


図８　ＩＣＭ内のリングバッファの構成

３．２．２　フラグセンス方式

（１）構成の概要

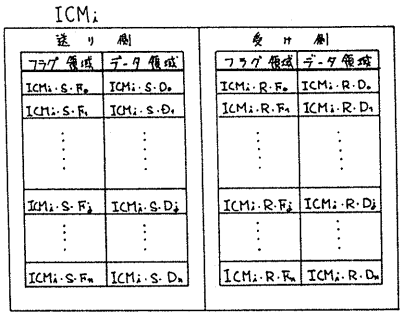
フラグセンス方式は、ＩＣＭ内の通信メモリを仮想的に対応する周辺プロセッサ（ＰＥ）に対するバッファメモリとして固定的に分割し、転送データの有無をフラグによって表示する方式である。ＣＰとＰＥ間のＩＣＭ内の通信メモリを図９のように分割する。

（Ⅰ）送り側（ＰＥ→ＣＰ）と受け側（ＣＰ→ＰＥ）に２分割する。

（Ⅱ）それぞれをさらにフラグ領域とデータ領域に分割する。

（Ⅲ）送り側のフラグ領域のアドレスは受信ＰＥの装置番号と１対１に対応させる。

（Ⅳ）受け側のフラグ領域のアドレスは、送信ＰＥの装置番号と１対１に対応させる。



ICMi·S·Fj ： ICMi·Send·Flag j
＝ PEi から PEj への送信データに対するフラグ領域

ICMi·S·Dj ： ICMi·Send·Data j
＝ PEi から PEj への送信データのバッファ領域

ICMi·R·Fj ： ICMi·Receive·Flag j
＝ PEi の PEj からの受信データに対するフラグ領域

ICMi·R·Dj ： ICMi·Receive·Data j
＝ PEi の PEj からの受信データのバッファ領域

図９　ＩＣＭ内の分割

（２）　動作の概要

ＰＥｉからＰＥｊへデータを転送する場合の通信の動作を図１０に示す。

（Ⅰ）ＰＥｉはＰＥｊとＣＰ間に配置されたＩＣＭｉの送り側のフラグ領域のｊ番地をフェッチし、空いていればフラグを立てると共に、データ領域のｊ番地に転送するデータを格納する。

（Ⅱ）ＣＰは常時ＩＣＭの送り側のフラグ領域を走査しており、ＩＣＭｉのｊ番地にフラグを検出すると、

（a）ＩＣＭｊの受け側のフラグ領域のｉ番地のフラグを調べ空いているかどうか判断する。

（b）空いていると、ＩＣＭｉの送り側のデータ領域のｊ番地のデータをＩＣＭｊの受け側のデータ領域のｉ番地に転送する。

（c）ＩＣＭｊの受け側のフラグ領域のｉ番地にフラグを立てると共にＩＣＭｉの送り側のフラグ領域のｊ番地のフラグをクリアする。

（Ⅲ）ＰＥｊは常時ＩＣＭｊの受け側のフラグ領域を走査しており、ｉ番地にフラグを検出すると、

（a）ＰＥｉからの転送データと判断する。

（b）ＩＣＭｊのデータ領域のｉ番地のデータをＰＥのメモリへ転送する。

（c）ＩＣＭｊのフラグ領域のｊ番地のフラグをクリアする。

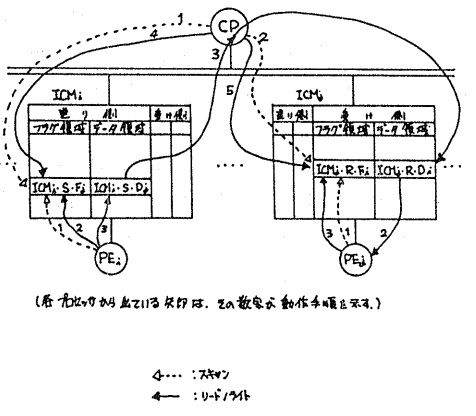


図１０　ＰＥｉからＰＥｊに対する通信動作

4. むすび

高速で通信オーバーヘッドの小さいプロセッサ間通信方式であるＳＩＣＯＭ方式では、プロセッサ間をＩＣＭにより完全結合するために、ＩＣＭの数がプロセッサ数の２乗に比例する経済的な欠点がある。これを改善するために、ＩＣＭを階層構造にした新しいプロセッサ間通信方式であるＨ－ＣＯＭ方式を提案し、その構成概要を説明した。この方式ではＩＣＭの数がほぼプロセッサ数に類似するので経済化の目的は達成しているが、ＣＰ及びＭＣＰでの制御をプログラム制御で行っているため、当初から予想したことではあるが転送遅延時間が増大する欠点がある。この問題を解決するためには、ＣＰ及びＭＣＰでの制御を電子素子を用いた布線論理化することが必要である。

ＳＩＣＯＭ方式およびＨ－ＣＯＭ方式について、ＩＣＭを試作し、周辺プロセッサにＭＣ６８０９を用いた小規模のシステムを構築し、機能の正常性を確認した。

参考文献

（１）中川、加久間：対称個別通信メモリバスシステム（ＳＩＣＯＭ）によるプロセッサ間通信方式、　信学会研究資ＳＥ８３－３８（１９８３）

（２）加久間、中川：対称個別通信メモリバスシステム（ＳＩＣＯＭ）による並行処理計算機の構成、　信学会研究資ＥＣ８４－２０（１９８４）