高速の光ファイバループを用いたプロセッサ間結合方式に関する検討

小柳津 育郎　　魚住 栄市　　星子 隆幸
（日本電信電話公社 横須賀電気通信研究所）

1．はじめに

近年、計算機処理量の増大、システムの高信頼化と段階的成長等への対処を多数台の疎結合マルチプロセッサ構成による負荷分散化で解決し、また、ソフトウェアの開発費削減・生産性向上をプロセッサの分散化（機能分散）により解決していく傾向が顕著である。

機能分散化の例としては、

・マルチメディア化に対応するための通信処理の専用プロセッサへの分離。

・アーキテクチャの異なるマシン上に構築されたソフトウェアやデータベースをジョブ転送又はファイル転送により有効利用するための異機種間接続。

などのシステム構成があげられる。この様なシステムの複合構成化に対応していくため、同一局舎内又は隣接局舎間に分散設置された数十台程度のプロセッサを完全群接続し、任意のプロセッサ間で高速な通信を可能にする方式の実現が要望される。

本稿は、上記のようなシステムの複合構成化を経済的に実現可能にするためのプロセッサ間の接続方式、即ち任意のプロセッサ間のN対N通信が可能な、高速で高効率なプロセッサ間結合装置（以下ＰＣＵと称す）の機能・構成条件について提案するものである。

なお、ＰＣＵ機能条件のうちソフトウェアインタフェースに関する機能の詳細、ＰＣＵ－ＰＣＵ間のデータリンク制御に関する機能の詳細については、それぞれ別稿で述べる。

2．システム構成条件

2．1　プロセッサ接続インタフェース

プロセッサ接続インタフェースとして、回線結合、チャネル結合、メモリ結合の各インタフェースが考えられる。これら各インタフェースのうち、回線結合インタフェースは転送速度、通信に要するソフトウェアオーバヘッド等に問題があり、高速のプロセッサ間通信には不向きである。メモリ結合は、転送速度、通信に要するソフトウェアオーバヘッドの観点からは最も高速性に優れた方式であるが、性能レンジやアーキテクチャが異なる多数のプロセッサ間接続に適用するには不向きである。

一方、チャネル結合は、数メガバイト／秒程度の高速転送が可能であり、標準的なチャネルインタフェース（Ｉ／Ｏインタフェース）が提供されている利点がある。更に、チャネル間アダプタ（ＣＴＣＡ等）に複数サブチャネルを設置し、各サブチャネルを介してタスク間等でデータの直接転送を行えば、単一サブチャネルでメッセージバッファを介した通信と比べて、適用する通信種別によっては通信処理に要するソフトウェアオーバヘッドを１桁以上低減可能との報告例もある（参考文献(1) 参照）。従って、高速性・汎用性ともに優れたプロセッサ接続を経済的に実現するインタフェースとしてチャネル結合インタフェースが適していると考えられる。

2．2　ＰＣＵ－ＰＣＵ間インタフェース

2．2．1　前提条件

(1) ＰＣＵ－ＰＣＵ間は高速性・経済性に優れたデータリンク制御手順（参考文献(3) 参照）により所定のフレーム単位に情報転送を行う。

(2) ＰＣＵを介したプロセッサ間通信において、最も高速性が要求される排他制御情報等のフレーム伝送が性能上のネック要因にならないためには１００Ｍビット／秒程度の伝送速度を必要とする。

(3) ＰＣＵ－ＰＣＵ間の接続距離≧１Ｋｍ

2．2．2　インタフェース諸元

表1にＰＣＵ－ＰＣＵ間のインタフェース諸元を示す

表1　インタフェース諸元

項　目	諸　元	理由／背景
伝送方式／伝送路	光ファイバケーブル（直列式）	上記前提条件(2)(3)の同時満足（図1 参照）
トポロジー	ループ	100メガビット/秒程度の高速伝送を数十台のPCU 間で光信号を用いて信頼性良く且つ経済的に実現するには、バス、スターは不向き
アクセス方式	トークンパッシング	①完全分散制御（集中局無し） ②ループ長が1 ～2Km のシステムが多く、ループ1周に500～1000bit の情報を載せれる程度なので、TDMA,スロッティッドリング方式では分割損があり伝送効率が悪い。

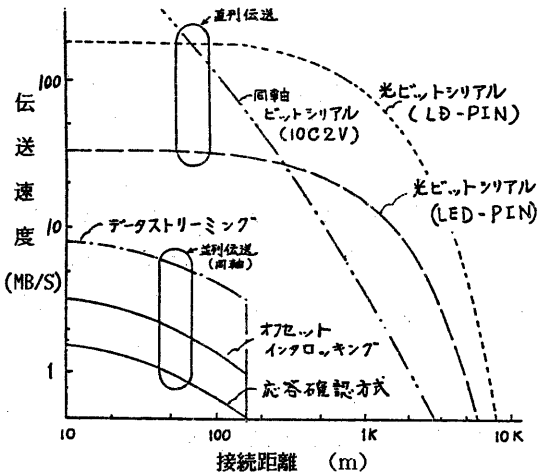


図1　伝送速度と接続距離

ＰＣＵを介したプロセッサ間結合システムの構成概要を図２に示す。

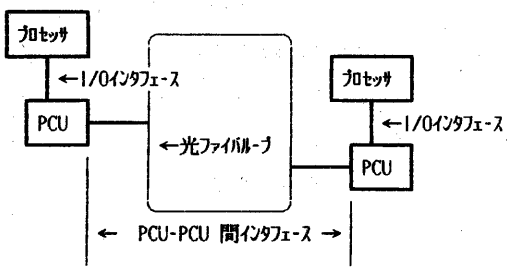


図2　システム構成概要

3．通信制御に関する対ソフトウェアインタフェース

3．1　基本方針

ループ等の共通伝送路を介したプロセッサ間通信におけるデータ転送は、通信制御プログラムが管理する送受信バッファ間で一旦データ転送を行い、割込み等を契機に同バッファより各ユーザ領域に移送する方式が一般的である。本方式は、高トラヒック、長データ転送の環境下におかれると送受信バッファ管理、割込み処理、データ移送等に要する通信処理ステップが性能上のネック要因として無視し得なくなる。

この為、１対１のプロセッサ間通信で適用されている手法（チャネル間アダプタに複数設置したサブチャネルを介し個々のユーザ領域間でデータの直接転送を行う手法）をＮ対Ｎ通信にも適用可能になるように拡張し、上記通信処理ステップの削減による通信の効率化を可能にする。

3．2　通信効率化のための機能条件

3．2．1　複数サブチャネルによる多重通信

複数サブチャネルを利用した通信効率化のためには、ＷＲＩＴＥ／ＲＥＡＤコマンドによるデータ転送において、先行発行コマンド側はチャネルを早期に解放すると共に相手プロセッサへの割込み無しに相手チャネルからのコマンド発行を待合せ、この間に別のサブチャネル上でのコマンド実行処理を可能にすることにより通信の多重度をあげる必要がある。この待合せ機能を用いたサブチャネル間の通信動作例を図３に示す（RET はコマンド再試行機能を用いたチャネル早期解放指示、RRは受信可能を示す応答、ACK は受信完了を示す応答）。

3．2．2　通信パスの設定管理

複数サブチャネルを用いたプロセッサ間通信では、ユーザソフトが意識する論理的通信パスを通信制御プログラムがサブチャネルにマッピングし、該サブチャネルを実通信パスとして多重通信を行う。１対１通信の場合はチャネル間アダプタのサブチャネルをＩＯＤアドレスで

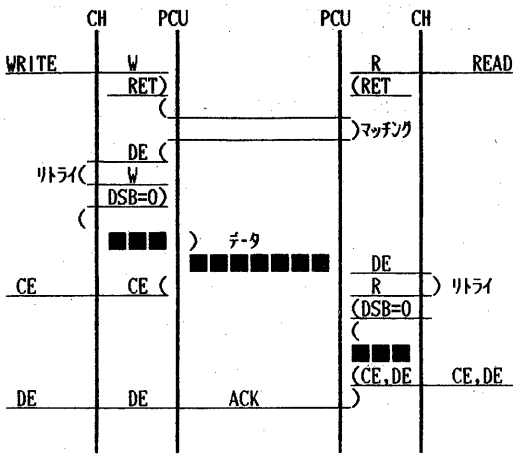


図3　サブチャネル間通信動作例

選択指定することにより一意に実通信パスが決定されるが、N対N通信の場合は2台のPCUのサブチャネル間で実通信パスを設定するためには、互いに通信相手のPCUアドレス並びにPCUサブチャネルアドレス情報を持つことが必要となる。表2にN対N通信でのアドレス指定に関する問題点の所在とその対処法を示す。

4．伝送制御手順

SBSA	状態情報	DA	SBDA
0	----------	A	a
1	----------	B	b
2	----------	C	c
・ ・	・ ・	・	・

SBSA：送信元PCUサブチャネルアドレス
DA　：送信先PCUアドレス
SBDA：送信先PCUサブチャネルアドレス

図4　CTBL構成イメージ

4．1　データリンクレベルプロトコルの基本設計

標記プロトコル設定にあたって以下の項目に関する評価検討が必要となる。

①適用領域

②コストパフォーマンス

③汎用性・拡張性

上記観点から、IEEE　802委員会のTOKEN RING標準化仕様案を、高速のプロセッサ間通信への適用を考慮して提案する仕様（PCU仕様；概要後述、詳細は参考文献(3) 参照）と比較すると表3に示す結果が得られる。

表2　N対N通信でのアドレス指定に関する問題点とその対処法

項番	問題点	対処法	評価／理由	記事
1	既存のチャネル仕様では、チャネル配下のデバイスとして自側PCUのサブチャネルを指定できるのみで、通信相手のPCU並びにPCUサブチャネルのアドレス指定が出来ない。	案A：通信用コマンドは起動時データのメモリ読出しを行い、該データで左記アドレス指定を行う。 案B：①PCUに各サブチャネル対応に相手PCU並びにPCUサブチャネルアドレスを記憶し管理するテーブル（CTBL）を設置　*1 ②通信の前処理として、専用コマンドによりCTBL設定し実通信パス確立 ③1対1通信の場合と同様に自側PCUサブチャネルアドレスのみを指定。通信相手アドレスはCTBL検索によりPCUが自動設定	案B：better （理由） ・1対1通信と等価な通信効率実現可能 ・1対1通信との親和性が高く通信制御プログラムの一部流用も可能	*1:CTBL の構成ｲﾒｰｼﾞは図4参照
2	ループに接続される全PCU相互の通信が可能なため、実通信パス設定が複雑。特にシステム更改、障害時等のパス再設定処理が複雑化し影響範囲も大きい。	案A：ソフトで予め余裕を持った実通信パスの設定を行い、システム更改、障害等の影響を最小限にとどめる。 案B：通信相手PCUの指定を行えば送受信双方のPCUサブチャネルの対応関係はPCUハードが自動選定　*2	案B：better （理由） ・ソフト論理の簡素化 ・システムジェネレーションの簡素化 ・トラヒックに即応した最適パス設定の容易化	*2: 本機能の詳細は参考文献(2) 参照

高速の光ファイバループを対象としたＬＡＮ標準化仕様が具体化してない現状では、純粋なコスト性能比較でＩＥＥＥ　８０２委員会のＴＯＫＥＮ　ＲＩＮＧ標準化仕様案より本稿で提案するＰＣＵ仕様の方が優れている。

４．２　送信権の管理

送信権の移動契機に着目した場合、送信権制御手法として以下の３案が考えられる。

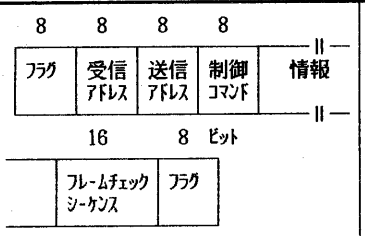
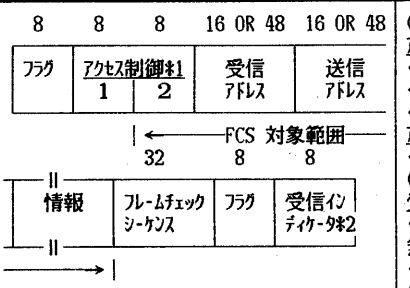
案Ａ：送信待ち状態にある通信フレームを全て送出し終わった時点で送信権を放棄する。

案Ｂ：優先制御等の付加機能の実現を容易にするため、送出した１フレーム受信時点で送信権を放棄する（例；ＩＥＥＥ　８０２仕様）。

案Ｃ：１フレーム送出時点で即送信権を放棄する。

このうち、案Ａは特定のＰＣＵに長時間送信権が占有されることがあり、送信権監視等の制御が複雑化する。

表３　　データリンクレベルプロトコルの比較

項　目		ＰＣＵ仕様	ＩＥＥＥ　８０２仕様	記　事
フレーム構成				（※1） アクセス制御1 ・優先制御 ・トークン/フレーム識別 ・トークン 監視 アクセス制御2 ・フレーム 種別識別 （※2） 受信インディケータ ・フレーム 異常の有無識別 ・受信応答 （アドレス,フレームコピー）
性能	伝送効率	高トラヒック，長ループ，短データになる程ＩＥＥＥ８０２仕様より優れた特性を示す。※3　○	低トラヒック，短ループ，長データの領域ではＰＣＵ独自仕様と同等の特性を示す。　△	（※3） 詳細は参考文献（3）参照
	伝送帯域	ビットレートと同程度の帯域　※4　○	フラグの識別にマンチェスター符号の使用を前提とした方式を採っているため、ビットレートの２倍の帯域を必要とする。従って、高ビットレートのデータ転送には不向き。　△	（※4） 符号形式については 6章参照
ハード量（データリンク 制御基本部）※5		１（正規化済み）　○	～　２　△	（※5） データリンク制御基本部 ・送信件管理 ・アドレス 付加・比較 ・シリアル・パラレル変換 ・FCS付加・検査 ・フラグ発生・検出
コスト		－　○	①時点では低速のシールド撚対線のみを対象としており、１００メガビット／秒程度の高速の光ファイバを対象としたものはない。　→ハード量大により高コスト化 ②ＬＥＤ適用困難、受信応答制御の高速化により高コスト化　△	
信頼性		①フレーム全体がＦＣＳの対象 ②ＬＥＤ適用可　○	①フレームの一部がＦＣＳの対象外　△	
適用領域の拡張性		・高速性の確保、長時間の送信権保留無しによりプロセッサ間通信以外への適用性も高い　○	・優先制御機能適用により、多種類の通信効率化に優れている。　◎	

（注）◎：適用性に優れている　　○：適用性あり　　△：適用性にやや問題あり

また、伝送速度（１００Mビット／秒）とチャネルーＰＣＵ間の転送速度（～２４Mビット／秒）との性能差を考えると、特定のＰＣＵが長時間送信権を占有すればチャネルーＰＣＵ間の転送速度が性能上のネック要因となり、伝送速度を生かしたループ全体のスループット向上が実現出来ない。

案Ｂは優先制御等の付加機能を実現する上で有効な手法であるが、音声情報等のリアルタイム性の高い情報転送を除けば優先制御等の必要性は少ないと考えられ、少なくともプロセッサ間通信に限って言えば機能的に冗長である。また、案Ｂは送信権獲得までの待ち時間が案Ｃと比べて長くなる傾向がある。案Ｂと案Ｃのメッセージ通過時間をトランザクション、ループ長、メッセージ長をパラメータとして比較すれば図５に示す結果が得られる（ＶＬ：ループ伝送速度、ＶＩ：チャネルーＰＣＵ間転送速度、Ｌ：ループ長、Ｔpcu ：ＰＣＵ内処理時間、Ｄ：ＰＣＵ内伝送遅延、Ｂ：メッセージ転送単位；詳細は参考文献(3) 参照）。全般的に案Ｃの方がメッセージ通過時間が短く、特に高トラヒック、長ループ、短メーツセージになる程その傾向が顕著となる。

以上の理由により、上記３案の中では案Ｃがプロセッサ間通信への適用を主目的とした送信権制御手法としては最も優れている。

４．３　応答方式

データリンクレベルで受信フレームに対する応答の返し方として、以下の３案が考えられる。

案Ａ：ＩＥＥＥ　８０２仕様と同様に、フレーム末尾の受信インディケータビットを受信側がＯＮにすることにより応答を返す。

案Ｂ：受信側が送信権獲得時、応答フレームにて応答を返す。

案Ｃ：送信側がフレーム送出時、該フレームに後続した領域を応答領域として確保し、受信側は該応答領域に応答フレームを挿入することにより応答を返す。

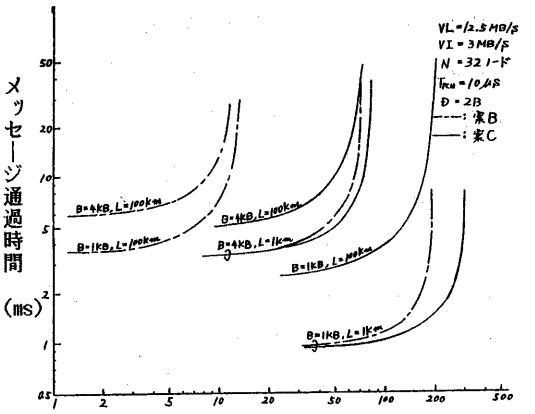


図５　メッセージ通過時間

上記３案のうち、案Ａは受信インディケータビットのＯＮ制御を１００Mビット／秒の伝送速度に同期して高速に行うことが必要となり、ハード実現上の負担が大きい。　また、受信インディケータビットをＦＣＳの対象とすることが困難で、応答の信頼性に問題がある。

案Ｂは、ＦＣＳエラー等のフレーム受信異常があった場合、送信側での該フレーム再送が望まれるが、送信側で再送の要否を判断出来るまでの時間が長く、送信バッファの利用効率が悪くなる。また、トラヒックが集中した場合も正常に応答フレームが返せるように、沢山の応答フレームを記憶する手段が必要となる。

案Ｃは、若干の余裕を持った応答領域の設置が必要となり、実質的なフレーム長増大による伝送効率低下を伴うが他案に見られる大きな問題点は無い。

以上の理由により、１００Mビット／秒程度の高速伝送への適用を対象とした場合、上記３案の中では案Ｃが最も優れている。

４．４　長データ転送方式

ＰＣＵ－ＰＣＵ間のデータ転送方式として、大別してソフトが指定するデータ長に制限を付けずハードで一定長に分割して送る方式と、ソフトが指定するデータ長に制限を付け、ハードは制限長内のデータを一括して送る方式とが考えられる。

後者の方式は、バッファ管理、再送制御が前者と比べて簡単で、関連ファームウェア量も１Ｋステップ程度の低減が見込まれる。しかし、メッセージ通過時間、ソフトウェアインタフェースで以下の問題が指摘出来る。

(1) 一括転送方式は、４Ｋバイトのメッセージ長（チャネルーＰＣＵ間の転送速度：３Ｍバイト／秒、分割転送単位：１Ｋバイト）で６０～７０％程度分割転送方式よりメッセージ通過時間が長くなる（詳細は参考文献(3)参照）。これは、ＬＣＭＰによる排他制御情報を回覧する場合など応答時間の伸びがＣＰＵ処理上のネック要因となることが予想されることから、極力低減化することが望ましい。

(2) プロセッサ間通信に関する従来のソフトウェア制御で、データ転送長に関する制限は必ずしも統一化されていない。これに制限を付けた場合、分割転送単位にプロトコルヘッダが付き、該分割のためのプロトコル（トランスポートレベル）が必要となる場合がある。

(3) チャネル制御方式ではデータチェイン指定が可能であり、転送データ長については制限がない方式が一般的である。

ハードの作りの容易性・経済性を重視するか、性能・ソフトインタフェースを重視するかで優劣の判断が別れるが、汎用性の高いプロセッサ間結合への適用を志向するのであれば、ソフトインタフェースがフリーな分割転送方式を採用する必要がある。

５．障害処理方式

５．１　基本設計

ＰＣＵ障害は、大別して以下の３レベルに分類可能である。

レベル１：伝送路断等放置するとループシステム運用に致命的な支障がある障害

レベル２：ＰＣＵ－ＰＣＵ間の個々の通信に支障があり、障害原因又は障害の現象を個々の通信対応にソフト報告可能な障害

レベル３：ＦＣＳ異常等レベル２障害の原因となりうる障害であるが、個々の通信主体を限定出来ない障害

上記各レベルの障害は、障害レベルに応じて以下の対処が望まれる。

(1) レベル１障害

大別して、ソフトウェア又はオペレータ指示を介在してループ接続状態の切替えを行う方式と全てハードで該切替えを行う方式とが考えられる。前者は、障害発生時点からループ接続状態切替えまでの時間が長くなり、その間の通信全てに擾乱を与える可能性が高い。　また、プログラムが意識してＰＣＵのループ組込み、切離しを行う以外はループ接続状態をソフトが制御する必要性は低く、単に状態監視が出来れば運用上又は保守上問題は無い。従って、経済的なインプリメントが可能であれば後者の方式採用が望ましい。

(2) レベル２障害

通信主体が限定でき、障害が個々の通信に閉じていて他ＰＣＵの通信に擾乱を与えることが無いので、本障害検出時ソフトに障害発生とその原因を報告する手段を確立すればよい。

(3) レベル３障害

通信主体が限定出来ないので、障害検出時のソフト報告は無用な擾乱を与えない意味から必要性は低い。一方、本障害はレベル２障害の原因となる場合が多く、予防保全の観点からも障害発生を無視してしまうのは好ましくない。従って、障害発生をログ情報等により記憶しておき、必要に応じてソフトからのアクセス可能なようにしておくのが望ましい。

５．２　障害自動回復の機能条件

２重ループ（逆方向）の伝送路構成を前提にレベル１障害に対しＰＣＵハードで行う自動回復の機能条件を明らかにする。

レベル１障害としては、伝送路又はＰＣＵのリピータ部の障害によりクロック異常として検出されるものと、クロックは正常だがトークン消滅によりループ上から送

信権がなくなり、ループがハングアップしてしまうものとが考えられる。これら両障害の検出レベル並びに処理レベルを表4に示す。

表4に示す障害処理レベルに関し、ハードウェア量及び信頼性の観点から特殊パターンを用いてデータリンクレベルで実施するのが最も望ましい。クロック異常の検出レベルに関しては、ハードウェア量的にデータリンクレベルが望ましいが、障害箇所切分けの容易性、汎用的なプロトコル階層との親和性を考慮すると物理レベルの方が優れている。　従って、物理レベルでのクロック異常検出又はデータリンクレベルでの送信権消滅検出時、データリンクレベルでフレームを用いずループ接続状態の切替え制御を行う行う方式が優れている（具体的な障害自動回復手順については参考文献(2) 参照）。

6．伝送路符号

光伝送に適用性の高い符号を5種類選択し、各符号間の優劣を比較した結果を表5に示す。

最近、１００Mビット／秒程度の高速伝送速度域で経済性及び信頼性に優れたＬＥＤ適用の見通しが得られて

表4　　障害の検出／処理レベル

項番	障害種別	検出レベル	処理レベル	概　　要	利害得失		記　事
					利　点	欠　点	
1	クロック異常	物理レベル	物理レベル	リピータ部でクロック異常を検出し、同部でループ接続状態の切替え処理を実施	左記切替え処理に用いる情報のビットエラー耐力に優れている	リピータ部に障害回復用の制御回路を持つ必要があり、ハードウェア量増大	
2		物理レベル	データリンクレベル	リピータ部からのクロック異常検出通知を受けてデータリンクレベルでフレームを用いてループ接続状態の切替え処理を実施	データリンク制御用ファームウェアに左記切替え処理用のファームウェアを追加すればよく、ハード負担が小さい。	左記切替え処理に用いるフレームのビットエラー耐力に弱い。	
3		データリンクレベル	データリンクレベル	ＮＲＺ等の非マンチェスター系符号を用い、論理的に意味のあるデータはクロック異常時に発生するパターンを避けてコード化しておく。受信データのビットシーケンスが論理的に存在しえないものである場合、専用の特殊パターンを用いてループ接続状態の切替え処理実施	リピータ部のクロック異常検出回路を削減でき、ハードウェア量的に最も簡素化可能	符号依存性がある。	
4	送信権消滅	データリンクレベル	データリンクレベル	送信権監視を行い、一定時間以上のトークン無受信時トークン生成処理実施。本処理によっても送信権の回復がない場合、フレームを用いてループ接続状態の切替え処理実施	－	左記切替え処理に用いるフレームのビットエラー耐力に弱い。	
5		データリンクレベル	データリンクレベル	送信権監視を行い、一定時間以上のトークン無受信時トークン生成処理実施。本処理によっても送信権の回復がない場合、専用の特殊パターンを用いてループ接続状態の切替え処理実施	左記切替え処理に用いる特殊パターンのビットエラー耐力に優れている。	符号依存性又は特殊パターン検出の複雑化	

いる。 このＬＥＤ適用を前提に考えるとクロック速度２ｆを要する符号は適用困難である。 ｍＢｎＢ符号とｍＢ１Ｃ符号を比較すると、後者は直流成分変動を抑えるためのスクランブラが必要となり符号変換回路の規模が増大する。前者は直流成分を減らすことが可能で、符号変換回路規模も抑えることが可能である。従って、現時点での１００Ｍビット／秒域の高速伝送にはｍＢｎＢ符号が優れている。

表５　伝送路符号の比較

項目		mBnB (NRZ)	DPSK D-マンチェスタ	MPDC (MFM)	CMI	mB1C
クロック速度		nf/m ○	2f △	2f △	2f △	(m+1)f/m ○
伝送帯域		中 ○	大 △	小 ○	大 △	中 ○
DC成分変動		○	○	△	○	△
クロック抽出の容易性		○	○	△	○	○
変換回路	規模	○	○	○	○	△
	論理素子	ECL10K クラス ○	ECL100K クラス △	ECL100K クラス △	ECL100K クラス △	ECL10K クラス ○
使用可能発行素子		LED LD	LD	LD	LD	LED LD

（注）○：適用性あり　△：適用性にやや問題あり

７．おわりに

高速のプロセッサ間通信をＮ対Ｎ通信システムの中で経済的に実現するための方式条件を明らかにした。対ソフトウェアインタフェースは従来のチャネル間アダプタ（ＣＴＣＡ）等を介した１対１通信をＮ対Ｎ通信システム向きに拡張し、下位のデータリンクレベル以下の機能・構成は高速の光ファイバ通信向きに簡素化することの優位性を提言している。本提言に沿ったプロセッサ間結合方式は、任意のプロセッサ間で高速通信を志向した汎用性の高い方式として広く実用化が期待できる。

謝辞

本検討にあたり有意義なご指導、ご助言を頂いた関係各位に深謝します。

参考文献

(1) 中野、森：“疎結合計算機システムにおける高速計算機間通信方式”、情処第３２回計算機アーキテクチャ研究会、１９８１．

(2) 星子他：“Ｎ対Ｎのループシステムにおけるプロセッサ間通信効率化に関する検討”、情処分散処理システム研究会２３－６、１９８４．

(3) 魚住他：“高速光ファイバループ用データリンクプロトコルの検討”、情処分散処理システム研究会２３－７、１９８４．

(4) 星子他：“共通バスを介したプロセッサ間通信効率化の検討”、情処第２４回全国大会６Ｈ－６、１９８２

(5) 木村他：“リング状ネットワークシステム高信頼化に関する一検討”、情処第２６回全国大会３Ｇ－９、１９８３．

(6) 魚住：“トークンリングの送信権制御法に関する一検討”、情処第２７回全国大会３Ｊ－１、１９８３．

(7) Draft IEEE standard 802.5 Working Draft, Aug.5, 1983.

(8) 柏村：“ローカルエリアネットワークの標準化動向”、情処ローカルエリアネットワークシンポジウム論文集、pp19-26 、１９８３．

(9) H.Hatta,K.Yasue: “TECHNICAL CONSIDERATIONS AND IMPLEMENTATIONS ON AN OPTICAL TOKEN LOOP”, IFIP WG 6.4,Sept.,1983.