(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4166285号
(P4166285)
(45)発行日　平成20年10月15日(2008.10.15)　(24)登録日　平成20年8月8日(2008.8.8)

(51)Int.Cl.　F I
H04L 12/56 (2006.01)　H04L 12/56 F

請求項の数 8　(全 28 頁)

(21)出願番号	特願平10-535788
(86)(22)出願日	平成10年1月30日(1998.1.30)
(65)公表番号	特表2001-511979(P2001-511979A)
(43)公表日	平成13年8月14日(2001.8.14)
(86)国際出願番号	PCT/US1998/001988
(87)国際公開番号	WO1998/036536
(87)国際公開日	平成10年8月20日(1998.8.20)
審査請求日	平成16年12月20日(2004.12.20)
(31)優先権主張番号	60/038,025
(32)優先日	平成9年2月14日(1997.2.14)
(33)優先権主張国	米国(US)

(73)特許権者　591016172
アドバンスト・マイクロ・ディバイシズ・インコーポレイテッド
ＡＤＶＡＮＣＥＤ　ＭＩＣＲＯ　ＤＥＶＩＣＥＳ　ＩＮＣＯＲＰＯＲＡＴＥＤ
アメリカ合衆国、９４０８８－３４５３　カリフォルニア州、サニィベイル、ピィ・オゥ・ボックス・３４５３、ワン・エイ・エム・ディ・プレイス、メイル・ストップ・６８（番地なし）
(74)代理人　100064746
弁理士　深見　久郎
(74)代理人　100085132
弁理士　森田　俊雄

最終頁に続く

(54)【発明の名称】バッファを再要求するための方法および装置

(57)【特許請求の範囲】
【請求項１】
パケット交換ネットワーク内においてネットワークスイッチのためにフレームをストアするバッファを再要求するための構成であって、
前記ネットワークスイッチからもはや送信される必要がないフレームがストアされた第１のバッファのバッファメモリにおける場所をポイントするフレームポインタをキューとして維持するように構成されたリクレームキューと、
前記ネットワークスイッチから送信されるフレームのコピーの数のカウント値をストアするよう構成された探索可能な第１のメモリと、
バッファメモリにおけるバッファの場所をポイントするバッファポインタをキューとして維持するように構成されたフリーバッファプールと、
フレームのフレームポインタを受信してそのフレームのコピーを送信し、このコピーの送信が、送信されるそのフレームの最後のコピーであるかどうかを判断するよう前記第１のメモリを探索し、そうであれば、前記ネットワークスイッチから前記そのフレームの最後のコピーを送信し、そのバッファの内容を送信した後に、前記フレームをストアする各バッファの前記バッファポインタを前記フリーバッファプールに送るように構成されたバッファマネージャと、
複数のエントリをストアするように構成されたマルチコピーキューとを含み、各エントリはフレームポインタおよび関連のコピー数を含み、正のコピー数は、前記関連のフレームポインタによってポイントされた前記フレームの送信されるコピーの数を示し、負のコピ

ー数は、前記関連のフレームポインタによってポイントされた前記フレームのコピーの送信を表わしており、
前記バッファマネージャはさらに、
前記フレームのコピーの送信が前記ネットワークスイッチによって送信されるそのフレームの最後のコピーではないと前記第１のメモリの探索により判断したときに、負のコピー数を有する前記エントリのうちの１つを前記マルチコピーキューにロードし、
前記マルチコピーキューを出るエントリのコピー数を検査し、正のコピー数を有する前記エントリを前記第１のメモリにストアし、負のコピー数を有するエントリについてはそのフレームポインタに等しいフレームポインタを有するエントリを前記第１のメモリ内において探索し、
前記第１のメモリを探索することにより、前記第１のメモリのエントリが、負のコピー数を有する前記エントリの前記フレームポインタに等しいフレームポインタを有することが示されたときに、前記第１のメモリの前記エントリの前記コピー数が１よりも大きい場合には、前記第１のメモリのエントリの前記コピー数をデクリメントするように構成される、構成。
【請求項２】
前記バッファマネージャは、前記第１のメモリを探索することにより、前記第１のメモリの前記エントリが、負のコピー数を有する前記エントリの前記フレームポインタに等しいフレームポインタを有することが示されたときに、前記第１のメモリの前記エントリの前記コピー数が１に等しい場合には、前記第１のメモリの前記エントリを削除し、前記第１のメモリから削除された前記エントリの前記フレームポインタを前記リクレームキューにロードするようにさらに構成される、請求項１に記載の構成。
【請求項３】
前記バッファマネージャは、前記リクレームキューから出るフレームポインタを得て、前記フレームポインタによってポイントされたフレームをストアする各バッファの前記バッファポインタを前記フリーバッファプールに送るようにさらに構成される、請求項２に記載の構成。
【請求項４】
単一のフレームをストアするバッファがリンクトリストを形成するようにともに繋がれ、前記フレームの前記フレームポインタは前記リンクトリストを形成する前記バッファのうちの第１のものの前記バッファポインタに等しく、前記リンクトリストの各バッファは、そのフレームに関する前記リンクトリスト内の次のバッファをポイントするバッファポインタを含み、前記リンクトリスト内の最後のバッファは、前記次のバッファへの前記バッファポインタの代わりに、前記リンクトリストの最後を示す表示を含む、請求項３に記載の構成。
【請求項５】
前記バッファマネージャは、前記リンクトリスト内の前記バッファを検査し、前記フレームの前記リンクトリスト内の前記バッファの前記バッファポインタを取出し、前記フリーバッファプールにそれらを送るようにさらに構成される、請求項４に記載の構成。
【請求項６】
前記バッファマネージャによって受信された、送信のためのフレームの各フレームポインタは、前記フレームのコピーが１つ送信されるかまたは前記フレームのコピーが複数送信されるかを示すビットを有し、前記バッファマネージャは、前記ビットを検査し、前記フレームのコピーが複数送信されることを前記ビットが示す場合にのみ前記第１のメモリの探索を行ない、前記フレームの１つのコピーが送信されることを前記ビットが示す場合には、前記フレームの前記１つのコピーを送信し、そのバッファの内容を送信した後に、前記フレームをストアする各バッファの前記バッファポインタを前記フリーバッファプールに送るようにさらに構成される、請求項５に記載の構成。
【請求項７】
前記フリーバッファプールに結合され、データをストアするために使用できるようになっ

た前記バッファのバッファポインタを前記フリーバッファプールに戻すように構成された再要求回路と、
データのストアに使用できるフリーバッファを提供するよう、前記構成の起動時にバッファポインタを発生するカウンタと、
前記フリーバッファプールの出力に結合された第１の入力と、前記カウンタの出力に結合された第２の入力と、前記第１の入力と前記第２の入力との間で選択を行なう制御入力とを備えるマルチプレクサとをさらに含む、請求項１に記載の構成。
【請求項８】
前記構成の起動後に、データのストアにバッファが必要とされると、前記カウンタが最大カウント値に達するまで、前記マルチプレクサの前記第２の入力を選択し、その後、ストアのためにさらなるバッファが必要とされると、前記第１の入力を選択するための手段をさらに含む、請求項７に記載の構成。
【発明の詳細な説明】
発明の分野
この発明はデータストアの分野に関し、特に、データをストアするために用いられるバッファを再要求して、種々のデータをストアするためにバッファを再使用することができるようにするための方法および装置に関する。
背景技術
パケット交換ネットワーク（たとえばイーサネットネットワーク）などの受信および転送を行なう多くのシステムにおいては、データを転送する前に一時的にデータをストアするためにメモリバッファが用いられる。たとえば、パケット交換ネットワークでは、データのフレームはネットワークスイッチで受信され、このネットワークスイッチは適切なポートにフレームを転送し、フレームはデータがスイッチで受信されている間に個々のバッファにロードされる。フレームのサイズは通常、個々のバッファの容量よりも大きいため、単一のフレームのデータをストアするために複数のバッファが用いられる。
受信される都度データをバッファすることにより、種々の特徴の中でもとりわけ、頑健なエラー検査がフレームに対して行なわれるようになり、送信および受信ポート間でのレートを整合させることも可能になる。データのフレームがスイッチに到達すると、現在使用可能である（すなわち「フリーな」）バッファだけがデータのストアに用いられるため、データが送信される前にデータが上書きされることはない。データが送信済みか、またはさもなければ最早必要でなくなった場合、データをストアするバッファは新しいフレームのデータのストアのために再使用することができる。また、入来するフレームのデータをストアするためにバッファを確保しておくことは可能であるが、エラー検査により、または受信されたデータの量が不十分であることにより、フレームが破棄されてしまう。この状況においても、バッファは再使用のために戻されるべきである。
効率を高め、可能な限り入来フレームのストアに使用できるバッファを確保するために、再使用のためにバッファを戻すことは、それらが最早必要でなくなった後にできるだけ迅速に行なわれるべきである。フレーム全体が送信された後に複数のバッファが単一のフレームをストアできるようにするシステムでは、スイッチはフレームの送信を認識し、そのチェーンをストアしている個々のバッファを再要求するプロセスを開始する。同じフレームのコピーが複数送信される場合、スイッチは、フレームをストアしたバッファを再要求するプロセスを開始する前にそのフレームのすべてのコピーが送信されるまで待機する。したがって、フレームのコピーが１つ送信されようと、フレームのコピーが複数送信されようと、同じ再要求手順が用いられるであろう。
メモリにデータをストアするためにバッファを用いるシステムにおける別の考慮点は、これらのバッファをストアするためのポインタを発生することである。メモリ中のバッファがある場所をポイントするバッファポインタはシステムの起動時に発生され、バッファプールにロードされ、データをストアするためにバッファが用いられる際にそこから取出される。使用後にバッファを再要求すると、バッファポインタはバッファプールに戻される。システムが多数のバッファを使用する場合、システムの起動時に最初にバッファポイン

タを発生し、バッファプールにそれらをロードすることは比較的時間のかかるプロセスである。

ＥＰ－Ａ－０　６２２　９２２には、通信システムにおいてデータをマルチキャストする際にバッファを解放するための典型的な公知の機構が開示されている。その図７Ａを参照して、最後のメッセージの送信後に、直接制御ブロックはそのメッセージを含むバッファをポイントする。その後、間接制御ブロックが解放され、直接制御ブロックから複製の数が読出され、これは、送信されたものが最後のメッセージであったことを示す１に等しいであろう。複製の数は０までデクリメントされる。その後、複製の数は０に等しいと判断され、この判断に応答して初めて直接制御ブロックが解放される。

発明の概要

メモリのバッファからデータを送信する際に素早い態様でバッファを再要求し、他のデータをストアするためにバッファを再使用できるようにするための構成および方法が必要である。

この必要性および他の必要性は、パケット交換ネットワークにおけるネットワークスイッチに対し、フレームをストアするバッファを再要求するための構成を提供するこの発明によって満たされる。この構成は、フレームがストアされる初めのバッファのバッファメモリ内の場所をポイントするフレームポインタをキューとして維持するよう構成されたリクレームキューを含む。探索可能な第１のメモリは、ネットワークスイッチから送信されるべきフレームのコピー数のカウント値をストアする。フリーバッファプールは、バッファメモリにおけるバッファの場所をポイントするバッファポインタをキューとして維持するよう構成される。バッファマネージャは、フレームのコピーを送信するためにそのフレームのフレームポインタを受信し、このコピーの送信が、送信されるべきフレームの最後のコピーであるかどうかを判断するために第１のメモリを探索する。この送信が最後のコピーに関するものであれば、バッファマネージャはそのフレームの最後のコピーがネットワークスイッチから送信されるようにし、フレームをストアする各バッファのバッファポインタを、そのバッファの内容が送信された後にフリーバッファプールに送る。

この発明の利点の１つは、送信されることとなる、次に送信されるフレームのコピーが、送信されるべき最後のコピーであるかと判断することが可能である場合、フレームをストアするバッファが、各バッファの内容が送信された後に再要求されることである。これにより、フレーム全体が送信されるまで待機し、そのフレームのコピーがこれ以上送信されないと判定して、そのときにはじめてフレームをストアしたバッファの再要求を開始する場合よりも、より素早く再使用のためにバッファが戻されるようになる。

前述の必要性および他の必要性は、ネットワークスイッチからのフレームの送信の後にそのフレームをストアしているバッファを再要求するための方法を提供するこの発明の別の実施例によって満たされる。発明の方法では、送信されるフレームのコピーが、送信されるべきフレームの最後のコピーであるかどうかを判断しようとする。送信されるフレームのコピーが送信されるべきフレームの最後のコピーであると判断された場合、フレームをストアしているバッファは、ネットワークスイッチから各バッファの内容が送信された後に再要求される。しかしながら、送信されるフレームのコピーが送信されるべきフレームの最後のコピーでないと判断された場合、フレームのコピーの送信を示す表示が、ネットワークスイッチからフレームが送信された後にキューとして維持される。また、送信されるフレームのコピーが送信されるべきフレームの最後のコピーであるかどうかを判断することができない場合にも、フレームのコピーの送信を示す表示が、ネットワークスイッチからフレームが送信された後にキューとして維持される。

この発明のこの実施例の１つの利点は、フレームの送信状態、すなわちそれが最後のコピーであるかどうかを判断する試みがなされることである。判断可能である場合、各バッファの内容が送信された直後にバッファを再要求することにより、再要求プロセスの速度を高めることができる。これが可能なのは、バッファのデータがそのフレームの別のコピーを送るために必要ではないことがわかっているためである。しかしながら、それが最後のコピーでないとき、またはそれが最後のコピーかどうか判断できない場合、送信を示す表

示がキューとして維持される。この表示は、送信されたコピー数のカウント値を維持するために用いられ得る。したがって、状況によっては、この発明は再要求プロセスの速度を高める。

発明のいくつかの実施例においては、フレームの単一のコピーのみが送られるか、または複数のコピーが送られるかが判断される。このような実施例では、単一のコピーの送信時にはバッファを再要求するプロセスの速度を高めることが可能である。なぜなら、バッファの内容はそのフレームの他のコピーの送信には必要でないことがわかっているからである。したがって、フレームの単一のコピーのバッファの再要求は、（複数のコピーフレームの最後のコピーであることがわかっている場合の）複数のコピーフレームの最後のコピーと同じ態様で行なうことができる。

また、起動時に時間のかからない、システムの起動時にバッファポインタを発生するための構成および方法が必要である。

この必要性および他の必要性は、メモリ内にデータをストアするためのバッファが使用できるようにするための構成を提供するこの発明によって満たされる。この構成は、複数のバッファポインタをストアするよう構成されたフリーバッファプールを含む。各バッファポインタは、関連のフリーバッファ、すなわちデータをストアするために現在使用できるバッファがある、メモリ中の別々の場所をポイントする。データをストアするために使用できるようになったバッファのバッファポインタをフリーバッファプールに戻す再要求回路が提供される。カウンタはこの構成の起動時にバッファポインタを発生して、データのストアに使用できるフリーバッファを提供する。マルチプレクサは、フリーバッファプールの出力に結合された第１の入力と、カウンタの出力に結合された第２の入力と、第１の入力と第２の入力との間で選択を行なう制御入力とを有する。

この発明のこの実施例の利点は、システムの起動時に、バッファポインタが使用できるようになる前に発生されてフリーバッファプールにロードされるのでは用できるようになる前に発生されてフリーバッファプールにロードされるのではなく、必要に応じて発生されることである。３２，０００個のバッファポインタなどといった、発生され得るバッファポインタの数が非常に多い場合、データをストアするためのバッファの使用を開始する前に、これらのポインタのすべての発生およびロードを待機することは時間の無駄である。この発明では、バッファポインタのすべてが発生されるのを待機することなく、カウンタによって発生されたバッファポインタがポイントするバッファをデータのストアのために直ちに用いることができる。最終的に、スイッチの動作時にバッファポインタのすべてが発生されると、カウンタは最早必要でなくなり、マルチプレクサはバッファポインタのすべてがフリーバッファプールから取出されるように切換わる。

この発明の前述および他の特徴、局面および利点は、添付の図面に関連して読むとこの発明の以下の詳細な説明からより明らかとなるであろう。

【図面の簡単な説明】

図１は、この発明の実施例に従って構成されたパケット交換システムのブロック図である。

図２は、この発明の実施例に従って構成され、図１のパケット交換システムに用いられるマルチポートスイッチのブロック図である。

図３は、この発明の実施例に従って構成された、図２のマルチポートスイッチのスイッチサブシステムの概略図である。

図４は、この発明の実施例に従って構成された、図３のスイッチサブシステムの単一の出力キューを示すブロック図である。

図５は、この発明の実施例による第１のタイプの出力キューを詳細に示す図である。

図６は、この発明の実施例による第２のタイプの出力キューを詳細に示す図である。

図７は、この発明の実施例に従って構成された、外部メモリのオーバフロー領域を詳細に示す図である。

図８は、この発明に採用されるリンクトリストデータ構造のブロック図である。

図９は、この発明の実施例によるフレームバッファヘッダフォーマットを概略的に示す図

である。
図１０は、この発明の実施例に従って構成された、図４のスイッチサブシステムのマルチコピー、リクレームおよびフリーバッファプール領域を詳細に示す図である。
図１１は、この発明の実施例に従って構成されたフリーバッファプール構造のブロック図である。
図１２は、この発明の実施例に従って構成されたマルチコピーキューのブロック図である。
図１３は、この発明の実施例に従って構成されたマルチコピーキャッシュの概略図である。
図１４は、この発明の実施例に従って構成された、スイッチサブシステムのバッファマネージャのキュー部およびポートベクタＦＩＦＯのブロック図である。
例示的な実施例の詳細な説明
イーサネット（ＩＥＥＥ　８０２．３）網などのパケット交換ネットワークにおけるスイッチを例に挙げてこの発明を説明する。しかしながら、以下に詳細に説明するように、この発明は他のパケット交換システムおよび一般的な他のタイプのシステムにも適用可能であることが明らかとなるであろう。
図１は、この発明が有利に採用され得る例示的なシステムのブロック図である。例示的なシステム１０はイーサネット網などのパケット交換ネットワークである。パケット交換ネットワークは、ネットワークステーション間でのデータパケットの通信を可能にする統合マルチポートスイッチ（ＩＭＳ）１２を含む。ネットワークはたとえば１０Ｍ＼ｂｐｓのネットワークデータレートでデータの授受を行なう２４個の毎秒１０メガビットの速度（Ｍ＼ｂｐｓ）のネットワークステーション１４と、１００Ｍ＼ｂｐｓのネットワーク速度でデータパケットの授受を行なう２つの１００Ｍ＼ｂｐｓネットワークステーション２２といった、種々の構成を有するネットワークステーションを含み得る。したがって、スイッチ１２はネットワークステーション１４または２２から受けたデータパケットをイーサネットプロトコルに基づく適切な宛先に選択的に転送する。
開示される実施例によると、１０Ｍ＼ｂｐｓネットワークステーション１４は媒体１７を介して、かつ半二重イーサネットプロトコルに従って、スイッチ１２に対してデータパケットの授受を行なう。イーサネットプロトコルＩＳＯ／ＩＥＣ　８８０２－３（ＡＮＳＩ／ＩＥＥＥ　Ｓｔｄ．８０２．３，１９９３Ｅｄ．）は、すべてのステーション１４が等しくネットワークチャネルにアクセスできるようにする半二重媒体アクセス機構を規定する。半二重環境のトラヒックは媒体１７と区別されたりまたはそれより優先されることはない。各ステーション１４はむしろ、媒体上のトラヒックを認識するために搬送波感知多重アクセス／衝突検出（ＣＳＭＡ／ＣＤ）を用いるイーサネットインタフェースカードを含む。媒体上の受信搬送波がデアサートされたことを感知することによりネットワークトラヒックの不在が検出される。送信するデータを有するステーション１４はすべて、パケット間ギャップ期間（ＩＰＧ）として公知である、媒体上の受信搬送波がデアサートされた後、予め定められた時間だけ待機することにより、チャネルにアクセスしようとする。複数のステーション１４がネットワーク上に送信するデータを有する場合、ステーションの各々が、媒体上の受信搬送波の、デアサートが感知されたことに応答してＩＰＧ期間の後に送信を行なおうとするため、衝突が生じる。したがって、送信ステーションは、別のステーションが同時にデータを送信することにより衝突が生じていないかを判断するために媒体を監視する。衝突が検出されれば、両方のステーションが停止し、ランダムな期間だけ待機し、再度送信を試みる。
１００Ｍ＼ｂｐｓネットワークステーション２２は好ましくは、提案されているフロー制御によるイーサネット規格ＩＥＥＥ　８０２．３ｘ全二重－草案（０．３）に従う全二重モードで動作する。全二重環境は各１００Ｍ＼ｂｐｓネットワークステーション２２とスイッチ１２との間に双方向ポイントツーポイント通信リンクを設け、スイッチ１２およびそれぞれのステーション２２は衝突することなくデータパケットの送受信を同時に行なうことができる。１００Ｍ＼ｂｐｓネットワークステーション２２の各々は、１００ベース

－ＴＸ、１００ベース－Ｔ４または１００ベース－ＦＸタイプの１００Ｍ／ｂｐｓ物理（ＰＨＹ）装置２０を介してネットワーク媒体１７に結合される。スイッチ１２は、物理装置２０への接続をもたらす媒体独立インタフェース（ＭＩＩ）２４を含む。１００Ｍ／ｂｐｓネットワーク２２は他のネットワークへの接続のためのサーバまたはルータとして実現され得る。

図１に示されるように、ネットワーク１０は、スイッチ１２と１０Ｍ／ｂｐｓステーション１４との間で送信されたデータパケットの時分割多重化および時分割非多重化を行なう一連のスイッチトランシーバ１６を含む。磁気変成器モジュール１９は媒体１７上の信号の波形を維持する。スイッチ１２は、時分割多重化プロトコルを用いて単一のシリアルノンリターンツーゼロ（ＮＲＺ）インタフェース２３を介して各スイッチトランシーバ１６に対するデータパケットの送受信を行なうトランシーバインタフェース１８を含む。スイッチトランシーバ１６はシリアルＮＲＺインタフェース２３からパケットを受信し、受信されたパケットを非多重化し、ネットワーク媒体１７を介して適切なエンドステーション１４にそのパケットを出力する。開示される実施例によると、各スイッチトランシーバ１６は独立した４つの１０Ｍ／ｂｐｓツイストペアポートを有し、スイッチ１２が必要とするＰＩＮの数が４分の１に減少するようにするシリアルＮＲＺインタフェースを介する４：１多重化を用いる。

スイッチ１２は、意思決定エンジン、切換エンジン、バッファメモリインタフェース、構成／制御／状態レジスタ、管理カウンタ、ならびにネットワークステーション１４および１２のためのイーサネットポート間でデータパケットの経路制御を行なうためのＭＡＣ（媒体アクセス制御）プロトコルインタフェースを含む。スイッチ１２はまた、インテリジェントな切換決定を行ない、後に説明するように、外部の管理エンティティに管理情報ベース（ＭＩＢ）オブジェクトの形式で統計的なネットワーク情報を与えるための優れた機能を有する。スイッチ１２はさらに、スイッチ１２のチップサイズを最小にするためにパケットデータの外部ストアおよびスイッチ論理を可能にするインタフェースを含む。たとえば、スイッチ１２は、受信したフレームデータ、メモリ構造およびＭＩＢカウンタ情報をストアするための外部メモリ３６へのアクセスをもたらす同期型ダイナミックＲＡＭ（ＳＤＲＡＭ）インタフェース３４を含む。メモリ３６は２Ｍｂまたは４Ｍｂのメモリサイズを有する８０、１００または１２０ＭＨｚ同期型ＤＲＡＭであってもよい。

スイッチ１２はさらに、外部管理エンティティが管理ＭＡＣインタフェース３２によってスイッチ１２の全体的な動作を制御できるようにする、管理ポート３０を含む。スイッチ１２は、ＰＣＩホストおよびブリッジ２８を介して管理エンティティがアクセスできるようにするＰＣＩインタフェース２６をさらに含む。これに代えて、ＰＣＩホストおよびブリッジ２８が複数のスイッチデバイス１２に対する拡張バスとしての役割を果たしてもよい。

スイッチ１２は、１つのソースから少なくとも１つの宛先ステーションに受信データパケットを選択的に送信する内部意思決定エンジン（図２）を含む。内部意思決定エンジンには外部ルールチェッカが代用されてもよい。スイッチ１２は外部ルールチェッカインタフェース（ＥＲＣＩ）４０を含み、これは内部意思決定エンジンの代わりにフレーム転送決定を行なうために外部ルールチェッカ４２が用いられるようにする。したがって、フレーム転送決定は、内部切換エンジンまたは外部ルールチェッカ４２のいずれかによって行なわれ得る。

スイッチ１２は、ポートごとのステータスをクロックに合せて出力しＬＥＤ外部論理４６を駆動する、ＬＥＤインタフェース４４をさらに含む。ＬＥＤ外部論理４６は人間が読取ることができるＬＥＤディスプレイエレメント４８を駆動する。発振器３８はスイッチ１２のシステム機能に４０ＭＨｚのクロック入力を与える。

図２は、図１の統合マルチポートスイッチ（ＩＭＳ）１２のブロック図である。スイッチ１２はそれぞれの１０Ｍ／ｂｐｓネットワークステーション１４間で半二重のデータパケットの送受信を行なうための２４個の１０Ｍ／ｂｐｓ媒体アクセス制御（ＭＡＣ）ポート５０（ポート１から２４）と、それぞれの１００Ｍ／ｂｐｓネットワークステーション間

で全二重のデータパケットの送受信を行なうための２つの１００Ｍ／ｂｐｓ　ＭＡＣポート５３（ポート２５および２６）とを含む。上述のとおり、管理インタフェース３０もまたＭＡＣ層プロトコル（ポート０）に従って動作する。ＭＡＣポート５０、５３および３０の各々は、受信先入れ先出し（ＦＩＦＯ）バッファ５２と送信ＦＩＦＯ５４とを有する。ネットワークステーションからのデータパケットは対応のＭＡＣポートで受信され、対応の受信ＦＩＦＯ５２にストアされる。受信されたデータパケットは対応の受信ＦＩＦＯ５２から外部メモリインタフェース３４に出力されて、外部メモリ３６にストアされる。受信されたパケットのヘッダもまた、内部ルールチェッカ５８または外部ルールチェッカインタフェース４０のいずれかである、意思決定エンジンに転送され、いずれのＭＡＣポートからデータパケットが出力されるかを決定する。具体的には、パケットヘッダは、スイッチ１２が内部ルールチェッカ５８または外部ルールチェッカ４２を用いて動作するよう構成されているか否かに依存して、内部ルールチェッカ５８または外部ルールチェッカインタフェース４０に送られる。内部ルールチェッカ５８および外部ルールチェッカ４２は、所与のデータパケットに関する宛先ＭＡＣポートを決定するための意思決定論理を提供する。したがって、意思決定エンジンは、単一ポート、マルチプルポートまたは全ポート（すなわちブロードキャスト）のいずれかに所与のデータパケットを出力し得る。たとえば、各データパケットにはソースおよび宛先アドレスを有するヘッダが含まれ、意思決定エンジンは宛先アドレスに基づいて適切な出力ＭＡＣポートを特定する。これに代えて、宛先アドレスは、適切な意思決定エンジンが複数のネットワークステーションに対応するものと特定するバーチャルアドレスに対応してもよい。これに代えて、受信されたデータパケットは、（１００Ｍ／ｂｐｓステーション２２のうちの１つのルータを介する）別のネットワークまたは所定のグループのステーションを特定するＩＥＥＥ　８０２．１ｄプロトコルに準拠するＶＬＡＮ（バーチャルＬＡＮ）タグ付フレームを含んでもよい。したがって、内部ルールチェッカ５８または外部ルールチェッカ４２のいずれかがインタフェース４０を介して、バッファメモリ３６に一時的にストアされたフレームが単一のＭＡＣポートまたは複数のＭＡＣポートに出力されるべきかを決定する。

外部ルールチェッカ４２を使用することにより、容量の増加、およびフレームが外部メモリに完全にバッファされる前にフレーム転送決定を可能にし、かつスイッチ１２がフレームを受信する順からは独立した順で決定が行なわれるようにする、決定キューのうちランダムな順序付け、といった利点がもたらされる。

意思決定エンジン（すなわち内部ルールチェッカ５８または外部ルールチェッカ４２）は、データパケットを受信すべき各ＭＡＣポートを特定するポートベクタの形式で転送決定をスイッチサブシステム５６に出力する。ルールチェッカからのポートベクタは、外部メモリ３６にデータパケットをストアするアドレス場所と、データパケットを受信して送信するためのＭＡＣポート（たとえばＭＡＣポート０から２６）の識別子とを含む。スイッチサブシステム５６はポートベクタに特定されたデータパケットを外部メモリインタフェース３４を介して外部メモリ３６から取出し、取出されたデータパケットを特定されたポートの適切な送信ＦＩＦＯ５４に与える。

付加的なインタフェースにより管理および制御情報が与えられる。たとえば、管理データインタフェース５９は、ＭＩＩ管理仕様（ＩＥＥＥ　８０２．３ｕ）に従うスイッチトランシーバ１６および１００Ｍ／ｂｐｓ物理装置２０と制御およびステータス情報をスイッチ１２が交換できるようにする。たとえば、管理データインタフェース５９は、双方向管理データＩＯ（ＭＤＩＯ）信号経路に時間基準を与える管理データクロック（ＭＤＣ）を出力する。

ＰＣＩインタフェース２６は、ＰＣＩホストプロセッサ２８によって内部ＩＭＳステータスおよび構成レジスタ６０にアクセスし、かつ外部メモリ３６にアクセスするための、３２ビットＰＣＩ改訂２．１に適合したスレーブインタフェースである。ＰＣＩインタフェース２６は複数のスイッチデバイスのための拡張バスとしての役割も果たし得る。管理ポート３０は標準７ワイヤ反転シリアルＧＰＳＩインタフェースを介して外部ＭＡＣエンジンにインタフェースされ、標準ＭＡＣ層プロトコルによりホストコントローラがスイッチ

１２にアクセスできるようにする。
図３は、この発明の例示的な実施例に従う、図２のスイッチサブシステム５６をより詳細に説明する図である。図２に示されるマルチポートスイッチ１２の他のエレメントは、スイッチサブシステム５６とこれらの他のエレメントとの接続を示すために図３に再度示される。スイッチサブシステム５６はフレームの受信および転送を行なうためのコアスイッチングエンジンを含む。スイッチングエンジンを実現するために用いられる主な機能ブロックは、ポートベクタＦＩＦＯ７０と、バッファマネージャ７２と、複数のポート出力キュー７４と、管理ポート出力キュー７５と、拡張バスポート出力キュー７７と、フリーバッファプール１０４と、マルチコピーキュー９０と、マルチコピーキャッシュ９６と、リクレームキュー９８とを含む。これらの機能ブロックの動作および構成は後により詳細に説明するが、まず、個々のエレメントに関する後の説明に関連性を持たせるために、図３のスイッチサブシステム５６の全体像を簡単に説明する。
ポートからマルチポートスイッチ１２に入るフレームには基本的に２つのタイプがある。すなわち、単一コピーフレームとマルチコピーフレームとである。単一コピーフレームは、マルチポートスイッチ１２によって他の１つのポートにのみ送られることとなる、ポートで受信されたフレームである。これとは対照的に、マルチコピーフレームは、１つのポートで受信され、１つより多い数のポートに送信されるフレームである。図３では、各ポートは別個のＭＡＣ５０によって表わされ、それ自体の受信ＦＩＦＯ５２および送信ＦＩＦＯ５４を有する。
単一コピーまたはマルチコピーであるフレームは内部ＭＡＣエンジン５０によって受信される。フレームパケットがポートで受信されると、それは受信ＦＩＦＯ５２に置かれる。各フレームはヘッダを有し、これは、内部ルールチェッカ５８または外部ルールチェッカ４２のいずれかのルールチェッカに与えられる。ルールチェッカ４２または５８は、ヘッダの情報に基づいて、フレームパケットがどこから送り出されるかを決定し、すなわちいずれのポートを介してフレームパケットが送信されるかを決定する。
ルールチェッカ４２または５８が転送決定を行なうのと同時に、バッファマネージャ７２はフリーバッファプール１０４からフリーバッファポインタを得る。このフリーバッファポインタは、受信ＦＩＦＯ５２によってフレームがストアされることとなる外部メモリ３６の場所である。バッファマネージャ７２によってフリーバッファポインタがフリーバッファプール１０４から得られると、フリーバッファポインタによってポイントされるバッファはもはやフリーであるとは考えられない。フレームデータは、直接メモリアクセス（ＤＭＡ）トランザクションでデータバス８０を介して受信ＦＩＦＯ５２から外部メモリ３６に転送される。フレームはフリーバッファプール１０４から得られたフリーバッファポインタがポイントする場所にストアされるが、後に説明するように、フレームをストアするために多くの他のバッファが用いられてもよい。
ヘッダデータの他に、ルールチェッカ４２または５８はバッファマネージャ７２からのフリーバッファポインタも受信する。このフリーバッファポインタはここではフレームポインタと呼ばれる。なぜなら、フレームがストアされる外部メモリ３６でのメモリ場所をポイントするからである。ルールチェッカ４２または５８は、転送決定を行ないかつ「ポートベクタ」の形式で転送命令を発生するためにヘッダ情報を用いる。図示される例示的な実施例では、ポートベクタは、フレームが転送されるべき各出力ポートに対してセットされたビットを備えた２８ビットベクタである。この全体像での例では、受信されたフレームは単一フレームであると想定する。したがって、ルールチェッカ４２または５８によって生成されたポートベクタには１つのビットしかセットされない。ポートベクタにセットされたビットはポートのうち特定的なものに対応する。
ルールチェッカ４２または５８はポートベクタＦＩＦＯ７０にポートベクタおよびフレームポインタ（ならびに制御操作コードおよびＶＬＡＮインデックス）を置く。ポートベクタはポートベクタＦＩＦＯ７０によって検査され、ポートベクタに関連したフレームポインタがどの特定の出力キュー７４に入力されるべきかを決定する。ポートベクタＦＩＦＯ７０は適切な出力キュー７４の一番上にフレームポインタを置く。これによりフレームの

送信がキューとして維持される。
ある時点で、フレームポインタは出力キュー７４を通過した後に出力キュー７４の一番下まで到達する。バッファマネージャ７２はそれが出力キュー７４の一番下まで到達したときにフレームポインタを取り、フレームポインタ書込バス８６を介して正しいポートの適切な送信ＦＩＦＯ５４にそのフレームポインタを送る。これによりフレームの送信がスケジュールされる。フレームポインタによってポイントされた外部メモリ３６での場所からＤＭＡトランザクションにおいて読出されたフレームデータは、適切な送信ＦＩＦＯ５４に置かれ後に送信される。
マルチコピー送信は、ポートベクタが、フレームがそれらから送信されることとなるマルチプルポートを示す、セットされた複数のビットを有する点を除いて、単一コピー送信と同様である。フレームポインタは適切な出力キュー７４の各々に置かれ、対応の送信ＦＩＦＯ５４から送信される。
バッファマネージャ７２は特殊な制御キューを用い、すなわち、フリーバッファプール１０４と、マルチコピーキュー９０と、リクレームキュー９８と、マルチコピーキャッシュ９６とを用いて、受信フレームをストアするためにバッファを割当て、フレームがその指定された出力ポートに送信されると再度使用できるようバッファを取出すプロセスを管理する。後により詳細に説明するが、バッファマネージャ７２はまた、出力キュー７４ならびに制御キュー１０４、９０および９８のために外部メモリ３６に「オーバフロー」領域を維持する。
この動作上の全体像を背景として、以下にスイッチサブシステム５６の個々のセクションおよびさまざまな局面をより詳細に説明する。これらの局面のうち最初に説明するものは、この発明のさまざまな出力キュー７４の構造である。１０Ｍｂ／ｓポートおよび１００Ｍｂ／ｓ出力ポートに指定される出力キュー７４の他に、管理ポート３０のために出力キュー７５が設けられ、拡張ポート２６のために出力キュー７７が設けられる。これらの出力キュー７５および７７は出力キュー７４と同じ外部構成を有するが、後に説明するように、異なった内部構成を有する。
図４は、この発明の実施例に従う出力キュー７４の外部構成を示すブロック図である。図４から明らかなように、この発明の出力キュー７４は３部構成である。性能を最も高くするためには、チップ上のキュー構造のすべてを保持することが好ましいが（マルチポートスイッチ１２を参照）、チップの占有面積に関する費用は非常に高い。これにより、チップが多数のエントリの切換を行ない、それらをキューとして維持する必要があるときにはジレンマが生じる。この発明は、チップ上に高性能な小容量セクションを含み、チップ外にオーバフロー領域を含む、単一の出力キューを与えることによりこのジレンマを解消する。オーバフロー領域は、チップ上の領域よりも比較的性能が低いにも関わらず、所要の大容量のキューとしてキューが役割を果たすようにする。
図４の実施例に従うこの発明の単一論理出力キュー７４は３つの物理セクションを有する。これらには、出力キュー書込側７６と、出力キュー読出側７８と、外部メモリ３６にある出力キューオーバフロー領域（全体が１１０として示される）とが含まれる。出力キュー７４のすべてに関する外部メモリ３６へのアクセスは、前述のとおり外部メモリインタフェース３４を介するものである。この発明は、現在の外部メモリのバースト的な性質を使用し、（フレームポインタなどの）データが、チップ１２を外部メモリ３６に接続するバス８４を介してバースト状にチップの内外からオーバフローキュー領域１１０に送られるようにする。
出力キュー書込側７６および出力キュー読出側７８はチップ１２上にある。書込側７６および読出側７８は小さくて値段の高い資源であると考えられる。これとは対照的に、出力キュー７４の第３の部分を形成するオーバフロー領域１１０は大きくて比較的安価である。書込側７６および読出側７８により高い性能がもたらされ、オーバフロー領域を通る経路によっては低性能で大容量の経路がもたらされる。
動作時に、出力キュー書込側７６はエントリを受信する。この発明に従うマルチポートスイッチ１２の例示的な実施例では、エントリは、フレームの最初の２５６バイトがストア

される外部メモリの第１のバッファをポイントするフレームポインタである。しかしながら当業者には、出力キューの構成７４はエントリとしてのフレームポインタに制限されず、マルチポートスイッチおよび他の技術の両方において、他のタイプのエントリをキューとして維持することに広く適用可能であることが明らかであろう。
エントリが出力キュー書込側７６内を完全に移動し、その一番下まで到達すると、出力キュー７４に関連した制御論理はエントリをどう処理するか決定する。出力キュー読出側７８にスペースがあれば、出力キュー７４のオーバフロー領域１１０は空いており、１つまたはそれ以上のエントリが出力キュー書込側７６から出力キュー読出側７８に直接送られる。書込側７６から読出側７８に直接エントリを送ることはすべてチップ１２上で行なわれるため、エントリは低レイテンシーで素早く完全に送られる。
出力キュー読出側７８がいっぱいであり、出力キュー書込側７６に少なくとも１バーストサイズの量のデータ（たとえばエントリの１６バイト分）があれば、データはその出力キュー７４のオーバフロー領域１１０にバースト状に書込まれる。出力キュー読出側７８がいっぱいであり、かつ出力キュー書込側７６にはまだ１バーストサイズの量のデータがないときは、エントリは出力キュー書込側に留まりさらに処理は行なわれない。最終的には、出力キュー読出側７８は空になり、出力キュー読出側７８に１バーストサイズの量のデータを収容する十分なスペースが生まれ、かつオーバフロー領域１１０にデータがあるときがくると、オーバフロー領域１１０から出力キュー読出側７８に１バーストのデータが与えられる。
出力キュー構成において、読出側７８は伝統的なキューとほぼ同様に作用する。なぜなら、エントリが１つずつ取出されるのはこの部分からであるからである。出力キュー書込側７６は主に、データをバーストに組立てて外部メモリ３６に書込むための回収機能を果たす。したがって、この発明は単一の事象（エントリを出力キュー７４に置くこと）をバースト事象に変える。書込側７６は、蓄積されたデータが必要に応じて外部メモリ３６のオーバフロー領域１１０にバーストされるようにする。比較的稀な場合にのみ必要となる機能に高価なチップ資源を提供するのではなく、輻輳時にオーバフロー領域１１０が安価なストレージを提供する。この発明はチップ外のオーバフロー領域１１０を使用するが、この領域１１０のアクセスは、１度に多くのバイトの情報をバーストすることにより効率よく行なわれる。これは、単一のエントリがキューに対して書込まれたり読出されたりする従来のキュー構造とは対照的である。
動作時に、出力キュー７４に到達するエントリが多ければ、これらのエントリはオーバフロー領域１１０に置かれ、チップ上のキュー７８のオーバフローを回避するようにする。したがって、この発明のキュー構造を用いるとフレームの廃棄が大幅に防止される。また、オーバフロー領域１１０のためのメモリの合計量は、外部メモリ３６のサイズを変更することにより容易に変更可能である。さらに、個々の特定のオーバフロー領域１１０のサイズは、出力キュー７４の性能に影響を及ぼすことなくキューのサイズをカスタマイズするためにプログラム可能である。
典型的に、キューは、先入れ先出し構成を有する順序づけられた構成である。しかしながら、リクレームキュー９８およびフリーバッファプール１０４などのいくつかのタイプのキューでは、エントリの順序は問題ではない。書込側１００から読出側１０２にデータを直接送信することが可能であれば、この発明はそのキューに関するオーバフロー領域を迂回して情報がこの経路に直接送信されるようにする。これは、情報が順番によって影響を受けない限り、関連のオーバフロー領域に情報がある場合でも可能である。たとえば、バッファの再要求は順番によって影響を受けない。なぜなら、バッファがフレームにストアされる必要がなくなった後に、最終的にバッファがフリーバッファプール１０４のフリーリストに戻される順番は、いかなるものでも許容されるからである。したがって、データが順番によって影響を受けない場合に外部メモリ３６のリクレームキュー９８のオーバフロー領域１１０への書込についての帯域幅が生じるのを回避するために、読出側１０２にさらなるエントリのためのスペースがあるものと想定して、書込側１００から読出側１０２に情報が直接送られる。リクレームキュー９８は順番によって影響を受けないデータを

キューとして維持するタイプのキューの一例である。しかしながら、順番によって影響を受けない他の多くのタイプのデータが種々の適用例で可能であるため、この発明のこの特徴は、他のタイプのデータをキューとして維持するキューにおいて有用性を見出す。
図１および図２に示されるこの発明の例示的な実施例のマルチポートスイッチには２８個の出力キュー（各々が出力ポートと関連する）があり、すなわち、１０Ｍｂ／ｓユーザポートに関するものが２４個、１００Ｍｂ／ｓサバポートに関するものが２つ、管理ポートに関するものが１つ、そして拡張バスポートに関するものが１つある。出力キュー７４、７５および７７は、フレームポインタが送信のためにキューとして維持されるときにそれらに一時的なストレージを提供する。キュー作業は、転送ポートベクタに示されるさまざまな出力キュー７４、７５および７７に対してポートベクタＦＩＦＯ７０がフレームポインタを書込むという形態をとる。
この発明のある好ましい実施例では、さまざまな出力キュー７４、７５および７７は以下のフィールドのうちいくつかまたはすべてを含む。すなわち、単一コピービットと、フレームポインタと、制御操作コードまたは制御信号と、ＶＬＡＮ（バーチャルローカルエリアネットワーク）インデックスとである。単一コピービットは１つの出力ポートにのみ転送されることとなるフレームを示す。フレームポインタは外部メモリ３６のフレームをポイントする。制御操作コードはフレームに関する特定的な情報（すなわち新たに得たフレームなど）を識別する。制御信号は制御操作コードからの情報を用いて、送信前にポートによってフレームがいかに処理されるかを示す。ＶＬＡＮインデックスは、外部へのフレームに（必要であれば）挿入されるべきＶＬＡＮタグに対する基準を与える。しかしながら、この発明は種々のタイプのフィールドを有する他の出力キューにも適用可能であるため、これらのフィールドは例としてのみのものである。
第１のタイプの出力キュー７４、すなわち１０Ｍｂ／ｓポート出力キューの例示的な実施例の内部構成が図５に示される。１０Ｍｂ／ｓ出力キュー７４は１０Ｍｂ／ｓポートに転送されることとなるフレームのエントリを保持する。これらのキューの出力キュー書込側７６は３２個のエントリを保持し、出力キュー読出側７８は図示される例示的な実施例において１６個のエントリを保持するが、考えられる他のサイズのものもこの発明の範囲内である。１０Ｍｂ／ｓ出力キュー７４は単一コピービットとフレームポインタ（１４ビット）とを含む。この発明のマルチポートスイッチの例示的な実施例では、１０Ｍｂ／ｓポートにはＶＬＡＮタグがないためＶＬＡＮインデックスは必要ない。
第２のタイプの出力キュー７４、すなわち１００Ｍｂ／ｓポート出力キューの例示的な実施例の内部構成が図６に示される。１００Ｍｂ／ｓポート出力キューは１００Ｍｂ／ｓポートに転送されることとなるフレームのエントリを保持する。出力キュー書込側７６はこのタイプの出力キューに６４個のエントリを保持し、出力キュー読出側は１６個のエントリを保持する。各エントリはＶＬＡＮインデックスと、部分的な制御操作コード（ビット４－０）と、単一コピービットと、フレームポインタとを含む。
外部メモリ３６の例示的なマップが図７に示される。外部メモリ３６の全体の容量はたとえば４Ｍｂであるが、種々の実施例において他の容量のメモリが採用されてもよい。この発明に従ってオーバフロー領域に外部メモリ３６を使用することにより、外部メモリを変更するだけで出力キューのサイズを増減することができる。これは、キューとして維持する容量全体がチップの製造時に設定される、キュー構成がすべてチップ上にあるシステムよりも有利である。
スイッチ１２のストア要件を満たすために、外部メモリ３６の例示的な実施例は下記の領域にスペースを割当てる。すなわち、フリーバッファプールオーバフロー１２０と、リクレームキューオーバフロー１２２と、マルチコピーキューオーバフロー１２４と、管理ポート出力キューオーバフロー１２６と、１０Ｍｂ／ｓおよび１００Ｍｂ／ｓ宛先ポートの各々のための個々の出力キューオーバフロー１２８と、拡張バスポート出力キューオーバフロー１３０と、ＭＩＢカウンタ１３２と、グローバルフレームバッファプール１３４とである。
メモリ領域全体のＢＡＳＥアドレスはチップ上のレジスタ６０の中のメモリベースアドレ

スレジスタ内でプログラム可能である。外部メモリマップ内の各領域のＢＡＳＥアドレスはレジスタセット内でプログラム可能である。領域長レジスタは不要である。所与の領域の長さは、マッピング内のその領域のＢＡＳＥアドレスから次の領域のＢＡＳＥアドレスまでの領域に等しい。
個々のオーバフロー領域の長さ（したがって容量）がプログラム可能であるため、各キューの容量全体がプログラム可能である。この発明のこの特徴により、必要に応じて容量の増大した特定の出力キューを提供するようにスイッチをカスタマイズすることが可能になる。
したがって、チップ１２上の制御キューに適合しない後続のオーバフロー領域ストアエントリは外部メモリ３６に置かれる。フリーバッファプールオーバフロー領域１２０はアドレスポインタをグローバルフレームバッファプール１３４中の未使用のバッファにストアする。リクレームキューオーバフロー領域１２２は、必要でなくなったリンクトリストチェーンにフレームポインタをストアする。マルチコピーキューオーバフロー領域１２４は（キューとして維持されたフレームポインタについては）コピーナンバー「≧１」を、かつ（うまく送信されたフレームについては）コピーナンバー「－１」を付してフレームポインタをストアする。
後続のオーバフロー領域は、チップ上に入らない出力キューのエントリをストアする。管理ポート出力キューオーバフロー領域１２６は管理ポートへの送信を待機するフレームポインタをストアする。出力キューオーバフロー領域１２８は適切な１０Ｍｂ／ｓまたは１００Ｍｂ／ｓポートへの送信を待機するフレームポインタをストアする。拡張バスポート出力キューオーバフロー領域１３０は拡張バスポートへの送信を待機するフレームポインタをストアする。
ＭＩＢカウンタ領域１３２は、スイッチ１２によって周期的に更新されるポートごとの統計をすべて含む。スイッチ１２はＭＩＢ統計をストアするための８ビットおよび１６ビットカウンタをチップ上に維持する。スイッチ１２はＭＩＢデータの損失を防止するために要求される周波数で、外部メモリ３６の３２ビットまたは６４ビットのＭＩＢカウンタを更新する。
グローバルフレームバッファプール１３４は、受信されたフレームデータをストアするリンクトリストのバッファを含む。任意の時点で、これらリンクトリストは有効フレームデータと無効になったバッファとを含み、無効になったこれらのバッファは、バッファマネージャ７２によってフリーバッファプール１０４に戻されるか、またはＰＣＩホストプロセッサ２８の所有となる。
次に図８を参照して、いずれかのＭＡＣポートまたはＰＣＩバスから受信されたフレームデータは、この発明の例示的な実施例におけるリンクトリストデータ構成のフォーマットで外部メモリ３６にストアされる。リンクトリストを生成するために用いられるバッファ１４０の長さは２５６バイトであるが、発明の種々の実施例では他の長さのバッファ長さが採用されてもよい。これらのバッファ１４０の各々へのアドレスポインタはスイッチ１２内のフリーバッファプール１０４によってストアされる。
スイッチ１２のポートのうち１つにフレームが受信されると、バッファマネージャ７２はフリーバッファプール１０４からアドレスポインタを要求し、バッファ１４０をリンクしてフレームをストアするようにする。フレームをストアする外部メモリ３６の第１のバッファに対するアドレスポインタが、そのフレームに対するフレームポインタになる。フレームポインタは、送信されることとなるフレームをキューとして維持するためのスイッチサブシステム５６において用いられる。
バッファ１４０は、メモリの次のバッファの場所を示す各バッファヘッダ１４２のアドレスポインタによって互いに繋がれる。バッファヘッダ１４２はまた、バッファ１４０に含まれるフレームデータに関する他の情報を含む。図９ａの例示的なバッファヘッダフォーマットに示されるように、先頭のバッファのヘッダは１２バイトである。図９ｂに示されるように、後の各バッファのヘッダは４バイトである。外部メモリバーストは、２バンク×１６バイトの長さであるため、各バッファの実際のフレームストア容量は２５６Ｂ－１

６Ｂ＝２４０Ｂである。
図９ａおよび図９ｂに示されるように、先頭および後のバッファヘッダフォーマットは下記のフィールドを含む。
バッファフォーマットビット：どのバッファフォーマットが使用中であるかを示す。１は１２バイトの長さの先頭バッファフォーマットを示す。０は４バイトである後のバッファのフォーマットを示す。バッファを繋ぐ際に残りのバッファの各々に関して用いられる。
Ｅビット（フレームマーカの最後）：フレームに関する最後のバッファであることを示す。Ｅビットがセットされていれば、チェーンにはこれ以上バッファはない。
Ｃビット（ＣＲＣエラー検出）：ＣＲＣエラーが受信機によって検出されたことを示す。Ｃビットが検出されると、送信機能は反転されたＣＲＣを意図的に送信する。
Ｌビット（整列エラー）：フレーム整列エラーが（ＣＲＣエラーとともに）受信フレームに検出されたことを示す。
Ｏビット（受信ＦＩＦＯオーバフロー）：受信ＦＩＦＯがオーバフローし、バッファのデータが有効でないかもしれないことを示す。
バッファ長さ：バッファヘッダの後の最初のバイトから始まる、バッファのデータフィールドにおいて有効なバイトの合計数。この長さにはオフセットバイト値は含まれるべきではない。
次のバッファポインタ：次のバッファに対するポインタ。次のバッファポインタはＥビットがセットされているときには有効でない。
オフセットバイト数：バッファのフレームデータセクションにおいてフレームの最初のバイトが始まる場所を示す。０のオフセットは、データがバッファヘッダ１４２の後の最初のビットで始まることを意味する。０のオフセットは、データがバッファの１６番目のバイトに後続するバイトで始まることを示す。オフセットが０でない値の場合、フレームデータは１６Ｂ＋バッファの始まりからのオフセットの後に始まる。送信機能はオフセットバイトフィールドに示されるバイト数だけ飛び越す。
Ｐビット（ポートタイプ）：入来する受信フレームのポートタイプを示す。０は１０Ｍｂ／ｓポートを示し、１は１００Ｍｂ／ｓポートを示す。このビットは、フレームが完全に受信されて外部メモリ３６にバッファされる前に、フレームを拡張バスに転送するようスイッチ１２をプログラミングする際に、タイムスタンプフィールドに関連してホスト２８によって用いられる。
Ｔビット：受信されたフレームのタイプを示す。タグ付またはタグ付でない場合がある。１はタグ付のフレームであることを示し、ＶＬＡＮ識別子フィールドは受信ＶＬＡＮ　ＩＤを含む。０はタグ付でないフレームを示し、ＶＬＡＮ　ＩＤは有効でない。
受信ポート番号：フレームが受信されたポート番号を示す。
ＶＬＡＮ識別子：「タグ付」ポートから受信されたＶＬＡＮ　ＩＤ。フレームがタグ付でないポートから受信される場合、このフィールドは無効である。
Ｒビット（ＣＲＣ再計算）：ＣＲＣを除去し送信機能において再計算する必要があることを示す。スイッチ１２はタグ付フレームが受信されるとこのビットをセットする。さらに、ホスト２８がフレームの内容を修正した場合、ホスト２８はこのビットをセットしなければならない。スイッチ１２がフレームを送信すると、スイッチ１２はこのビットを検査して、既存のＣＲＣを送信するか、ＣＲＣを除去してＣＲＣを再計算するかを判断する。
Ａビット（ＣＲＣ追加）：フレームデータの最後にＣＲＣがないことを示す。ホストはメモリに（ＣＲＣなしの）フレームを生成し、このビットをセットすることができる。スイッチ１２はフレームの送信時にＣＲＣを発生して追加する。Ａビットがセットされている場合、フレームの長さにはＣＲＣは含まれるべきでない。
Ｆビット（フォーマットビット）：フレーム長／タイムスタンプフィールドを特定する。０はフィールドが入来フレームのタイムスタンプであることを示す。１はフィールドが受信フレームのフレーム長であることを示す。
フレーム長／タイムスタンプ：Ｆビットに依存する。Ｆビットがクリアされていると、このフィールドは受信フレームの最初からのタイムスタンプを表わす。タイムスタンプは１

μｓの分解能を有する。Ｆビットがセットされている場合には、ＣＲＣおよび受信されたＶＬＡＮタグの全てを含む受信フレームの長さの合計が示される。フレームが受信されると、スイッチ１２は（タイマレジスタからの）タイムスタンプでこのフィールドをマークする。フレームが完全に受信される前に拡張バスフレームを転送するようホスト２８によってスイッチ１２がプログラミングされている場合、フレームデータを過度に読出すことなく外部メモリ３６から取出すことができるデータを測定するために（受信ポートの速度とともに）タイムスタンプを用いることができる。フレーム全体が受信されると、スイッチ１２はフレーム長をこのフィールドに書込みＦビットをセットする。

コピー数：ポートベクタＦＩＦＯ７０によって送信されるようにうまくキューとして維持されたコピーの数を示すために用いられる。このフィールドは、バッファマネージャ７２が、新しいエントリのためのマルチコピーキャッシュ９６にスペースを設ける必要がある場合に、フレームポインタのコピー数をストアするために用いられる。

図１０は図３のスイッチサブシステム５６のいくつかの要素を示す詳細図である。これらの要素は、フレーム記憶のためのバッファを与えるため、かつ、バッファがフレーム記憶のためにもはや必要とされなくなるとこれらのバッファを再要求し、再び使用可能にするために用いられる。上述のように、各出力キュー７４、７５（出力キュー７７を除く）はフレームポインタをバッファマネージャ７２に渡し、バッファマネージャ７２はフレームポインタが指すフレームの送信をスケジュールする。バッファマネージャ７２は、１）スイッチ１２の内部バスを管理し、２）出力キュー７４への／からのフレームポインタのキュー入れ／出しを容易にし、３）バッファの位置を決め、フリーバッファプール１０４に戻すために制御キュー９０、９８を管理し、４）外部メモリ３６を出入りするデータの流れを制御し、５）ＭＩＢおよびオーバーフロー領域を含むメモリ構造を維持するという機能を制御する。バッファマネージャ７２は全アクセスを外部メモリ３６に割当てるためのスケジューラ機能を含む。これらのアクセスには、１）受信されたフレームデータを記憶バッファ１４０に書込み、２）送信のために記憶バッファ１４０からフレームデータを読出し、３）出力キュー７４および制御キュー９０、９８のためのオーバーフロー領域の各々にフレームポインタを維持し（すなわち、書込み、読出す）、４）ＭＩＢカウンタを更新することが含まれる。

バッファマネージャ７２が所与のフレームポインタを適切な全出力キュー７４、７５にコピーした後、ポートベクタＦＩＦＯ７０がコピーの数（「コピー数」）を計算し、フレームポインタおよびコピー数をマルチコピーキュー９０の書込側に入れる。コピー数は、フレームが転送されるべきでないことを示す「０」、単一コピー送信を示す「１」、またはマルチコピー送信を示す「＞１」であり得る。これらの３つの場合を以下に説明する。

コピー数が「０」であり、フレームポインタがセットされたビットのないヌル転送ポートベクタを有することが意味されているとき、ポートベクタＦＩＦＯ７０はフレームポインタをリクレームキュー９８の書込側１００に直接渡す。バッファマネージャ７２がリクレームキュー９８を処理するときは、以下に述べるようにバッファマネージャ７２がバッファのリンクトリストチェーンを解体し、各「フリー」バッファごとのアドレスポインタをフリーバッファプール１０４の書込側１０６に戻す。

コピー数が「１」の単一コピー送信のとき、ポートベクタＦＩＦＯ７０はフレームポインタ、制御信号／制御操作コードおよびＶＬＡＮインデックスを適切なポートの出力キュー７４にコピーする。ポートベクタＦＩＦＯ７０は出力キュー７４内の単一コピービットをセットして（図５および図６参照）、これが単一の送信であることを示す。バッファマネージャ７２はそのポートの出力キュー７４からフレームポインタおよび単一コピービットを読出すと、上述のように送信をスケジュールする。バッファマネージャ７２は、フレームがストアされている外部メモリ３６において最初のバッファの位置を決めるためにフレームポインタを用いる。バッファマネージャ７２はこの最初のバッファからバッファヘッダを読出し、最初のバッファからデータを捕捉し、このデータを適切なＭＡＣ送信ＦＩＦＯ５４に入れる。フレームが複数バッファにおよぶ場合を想定すると、そのフレームのためのチェーン内の全バッファを見つけ、送信するためのアドレスを、後続バッファへのリ

ングがバッファマネージャ７２に与える。データが送信のためにＦＩＦＯ５４に一旦置かれると、バッファは不使用となり、フリーバッファプール１０４に戻され、結果として別のフレームデータをストアするために再割当される。
コピー数が１よりも大きいとき、ポートベクタＦＩＦＯ７０はフレームポインタ、ＶＬＡＮインデックスおよび制御信号／制御操作コードを適切な出力キュー７４の各々にコピーする（キュー７４に言及する際には、キュー７５、７７もまた言及されている）。ポートベクタＦＩＦＯ７０は出力キュー７４内の適切なフレームポインタのための単一コピービットをクリアし、コピー数が「＞１」であるフレームポインタをマルチコピーキュー９０の書込側９２に入れる。
バッファマネージャ７２がフレームポインタおよびクリアされた単一コピービットを出力キュー７４の１つから読出すたびに、バッファマネージャ７２はフレームの送信をスケジュールするが、コピー数「１」のフレームポインタを有するエントリがあるかどうかマルチコピーキャッシュ９６をも調べる。コピー数「１」のフレームポインタがマルチコピーキャッシュ９６に見つかれば、バッファマネージャ７２は、フレームの単一コピー送信の場合と同様に、送信のためにフレームをスケジュールし、送信の間にバッファを再要求する。しかしながら、フレームポインタがマルチコピーキャッシュ９６にないか、マルチコピーキャッシュ９６におけるフレームポインタのコピー数が１よりも大きければ、バッファマネージャ７２はフレームを送信するがバッファを再要求しない。送信を成功させた後、バッファマネージャ７２はフレームポインタのコピーをコピー数「－１」とともにマルチコピーキュー９０の書込側９２に入れる。
マルチコピーフレームが送信されるたびに、バッファマネージャ７２はマルチコピーキャッシュ９６内にコピー数「１」のフレームポインタを見つけられなかったならば、フレームポインタのコピーをマルチコピーキュー９０に入れる。したがって、いかなる所与の時間でも、マルチコピーキュー９０はコピー数が「１」よりも大きいフレームポインタ、および／または、各々コピー数が「－１」である、同じフレームポインタのいくつかのコピーを含むことができる。
バッファマネージャ７２は不使用となったバッファを再要求するためにマルチコピーキュー９０およびマルチコピーキャッシュ９６を絶えず処理する。バッファマネージャ７２はマルチコピーキュー９０を処理し、コピー数「＞１」のフレームポインタを読出すと、この新しいエントリ（フレームポインタおよびコピー数）をマルチコピーキャッシュ９６に入れようと試みる。マルチコピーキャッシュ９６がフルであれば、バッファマネージャ７２はその新しいフレームポインタのためにスペースを設ける。バッファマネージャ７２は「より古い」マルチコピーキャッシュエントリを読出し、外部メモリ３６内のそのバッファヘッダ内のこのエントリに対するコピー数を更新し、このエントリをマルチコピーキャッシュ９６からクリアする。マルチコピーキャッシュ９６内に使用可能な空きができると、バッファマネージャ７２はマルチコピーキュー９０からの新しいエントリをマルチコピーキャッシュ９６に入れることができる。
バッファマネージャ７２がマルチコピーキュー９０を処理し、コピー数「－１」のフレームポインタを読出すと、それはマルチコピーキャッシュ９６を探索して、デクリメントまたはデリートするためにコピー数「≧１」の対応するフレームポインタアドレスを探す。
バッファマネージャ７２がフレームポインタの一致を見つければ、それは１）コピー数が「＞１」であればマルチキャッシュのフレームポインタをデクリメントするし、または２）コピー数が「１」であればマルチコピーキャッシュのフレームポインタ／コピー数エントリをデリートし、フレームポインタをリクレームキュー９８に入れる。
一致するフレームポインタが見つからなければ、バッファマネージャ７２はコピー数を求めて外部メモリ３６（図９参照）におけるフレームポインタのバッファヘッダを探索する。メモリ内のコピー数が「１」であれば、バッファマネージャ７２はフレームポインタをリクレームキュー９８に入れる。メモリ内のコピー数が「＞１」であれば、バッファマネージャ７２はこのコピー数のフレームポインタをマルチコピーキャッシュ９６に入れ、そのコピー数をデクリメントする。

バッファマネージャ７２は、フレームポインタを読出してから、リンクトリストチェーンをたどり、バッファをフリーバッファプール１０４に戻すことによって、リクレームキュー９８を絶えず処理する。この作用は、ヌルポートベクタを有し、ポートベクタＦＩＦＯ７０によってリクレームキューに入れられていたフレームか、マルチコピー転送ベクタを有し、全コピーの送信を完了したフレームかのためのバッファを戻すのみである。単一コピーフレームにリンクされたバッファは、上述のようにそのフレームが送信されるときにフリーバッファプール１０４へと直接戻される。

出力キュー７４と外部メモリ３６内のそのオーバーフロー領域１１０とがフルであるために、ポートベクタＦＩＦＯ７０が単一コピー転送ベクタのためのフレームポインタを出力キュー７４に入れることができなければ、そのフレームは廃棄される。フレームポインタはリクレームキュー９８に戻され、フレームの廃棄がスイッチの管理資源によって記録される。

１つ以上の出力キュー７４と外部メモリ３６内のそれらのオーバーフロー領域１１０とがフルであるために、ポートベクタＦＩＦＯ７０がマルチコピー転送ベクタのための１つ以上のフレームポインタを入れることができなければ、そのフレームは使用可能なスペースのある出力キューへと転送されるのみであり、マルチコピーキュー９０に入れられたコピー数はうまく入れられたフレームポインタを反映するのみである。フレームポインタが入れられなかったことは、フレームポインタがキューに入れられなかった各ポートごとにスイッチ管理資源によって記録される。全出力キュー７４と外部メモリ３６内のそれらのオーバーフロー領域１１０とがフルであるためにポートベクタＦＩＦＯ７０がマルチコピー転送ベクタのためのどのフレームポインタも入れることができなければ、そのフレームポインタはリクレームキュー９８に渡され、スイッチ管理資源にはそれに従い通知される。

マルチコピーキュー９０は、フレームをストアするために用いられる全バッファ（すなわち、アドレスポインタ）がフリーバッファプール１０４に戻され得る前に、特定のマルチコピーフレームの何回の送信が完了されなければならないかを追跡するためにバッファマネージャ７２が用いる高優先順位キューである。この出力キューの書込側９２および読出側９４はそれぞれ６４エントリおよび１６エントリを保持する。マルチコピーキュー９０はマルチコピーキャッシュ９６に入力を与え、マルチコピーキャッシュ９６はいつバッファを再要求するかを決定するためにバッファマネージャ７２によって用いられる。マルチコピーキューの内部構造を図１２に示す。

出力キュー７４にうまく入れることができたフレームポインタの数に基づいて、ポートベクタＦＩＦＯ７０はフレームのフレームポインタコピーと「＞１」であるコピー数とをマルチコピーキュー９０に入れる。特定のポートの出力キュー７４がフルであれば、ポートベクタＦＩＦＯ７０はフレームポインタのコピーをその出力キュー７４に入れることができず、したがって、これをコピー数を決定する際の成功した事象として含めることはできない。

バッファマネージャ７２が出力キューフレームポインタを読出し、単一コピービットが「０」である（すなわち、マルチコピー）ことを見つけるたびに、それは、これが最後の送信であることを示すコピー数「１」のフレームポインタを求めてマルチコピーキャッシュを調べる。この一致が見つからなければ、各バッファの内容が送信された後に不使用になったバッファをフリーバッファプール１０４に与えることによって、バッファマネージャ７２は単一コピー送信の場合と同様にフレームを送信し、バッファを再要求する。一致が見つかれば、バッファマネージャ７２はマルチコピーフレームを送信し、コピー数「－１」のフレームポインタのコピーをマルチコピーキュー９０に入れる。拡張バス出力キュー７５または管理ポート出力キュー７７へとキューに入れられたフレームのためのマルチコピーフレームポインタの（ＰＣＩインターフェイス２６を介しての）使用をホストが終了すると、ホストはコピー数「－１」のフレームポインタのコピーをフレームポインタレジスタを介してマルチコピーキューへと書込む。このレジスタは図２におけるレジスタ６０のブロックに示されるレジスタの１つである。

出力キュー７４と同様に、マルチコピーキュー９０も入力経路および出力経路を備えて構

成される。入力経路または書込側により、ポートベクタＦＩＦＯ７０およびバッファマネージャはフレームポインタ／コピー数をマルチコピーキュー９０に入れることができる。出力経路または読出側により、マルチコピーキュー９０はフレームポインタ／コピー数をマルチコピーキャッシュ９６に入れることができる。マルチコピーキューオーバーフロー１２４と呼ばれる、フレームポインタ／コピー数のためのさらなるストレージが外部メモリ３６に設けられる。

フレームポインタ／コピー数が空のマルチコピーキュー９０に書込まれると、それらは読出側９４がフルになるまで書込側９２から読出側９４へと移動する。マルチコピーキュー９０の書込側９２に書込まれるさらなるフレームポインタ／コピー数は外部メモリ３６内のマルチコピーキューオーバーフロー領域１２４に入れられる。一旦マルチコピーキュー９０の読出側９４とそのオーバーフロー領域１２４とがフルになれば、マルチコピーキューに入れられるさらなるフレームポインタ／コピー数が書込側９２を満たし始める。

マルチコピーキュー９０を通過するフレームポインタの順序は、マルチコピーキューの読出側９４のスペースがクリアされると、フレームポインタ／コピー数がマルチコピーキューオーバーフロー領域１２４からマルチコピーキューの読出側９４へと移動し、マルチコピーキューの書込側９２からマルチコピーキューオーバーフロー領域１２４へと移動するようにして維持される。

マルチコピーキャッシュ９６はマルチコピーキュー９０と同様であるが、フレームポインタ／コピー数をスキャンするための探索可能な領域を設ける。マルチコピーキャッシュ９６は２５６までのエントリを保持する。バッファマネージャ７２はマルチコピーキュー９０からフレームポインタを読出し、コピー数が「＞１」または「－１」のいずれであるかによって、フレームポインタをマルチコピーキャッシュ９６に入れるかそれを処理するかする。

さらに、バッファマネージャ７２が出力キュー７４の読出側からフレームポインタを読出すごとに、バッファマネージャ７２は送信をスケジュールする。単一コピービットが「０」である（マルチコピーフレームを意味する）ならば、バッファマネージャ７２は、このフレームの最後の送信であることを示すコピー数「１」のフレームポインタを求めてマルチコピーキャッシュ９６をスキャンする。一致があれば、バッファマネージャ７２はフレーム送信の間にエントリを除去し、バッファをフリーバッファプールに戻す。一致がなければ、バッファマネージャは送信の終了時にコピー数「－１」のフレームポインタをマルチコピーキュー９０に入れる。

バッファマネージャ７２は周期的に、フレームポインタ／コピー数を読出し、それをマルチコピーキャッシュ９６に入れるか処理することによってマルチコピーキュー９０を処理する。これはフレーム送信から独立して行なわれる。バッファマネージャがコピー数「＞１」のフレームポインタを読出すか、コピー数「－１」のフレームポインタを読出すかによって２つの場合が引き続いて生じる。

１）バッファマネージャ７２がマルチコピーキュー９０からコピー数「＞１」のフレームポインタを読出す。マルチコピーキャッシュ９６に空きがあれば、それは新しいエントリを書込む。マルチコピーキャッシュ９６がフルであれば、バッファマネージャ７２はキャッシュ９６内のスペースをクリアしなければならない。これが行われるのは、マルチコピーキャッシュ９６からより古いフレームポインタ／コピー数の１つを読出し、外部メモリ３６内のフレームポインタのバッファヘッダをマルチコピーキャッシュ９６内のコピー数で更新し、このキャッシュエントリをデリートすることによってである。一旦スペースが生じると、新しいフレームポインタ／コピー数がマルチコピーキャッシュ９６に書込まれる。

２）バッファマネージャ７２がマルチコピーキャッシュ９０からコピー数「－１」のフレームポインタを読出す。バッファマネージャ７２はコピー数「≧１」の一致するフレームポインタを求めてマルチコピーキャッシュ９６を探索する。バッファマネージャ７２がマルチコピーキャッシュ９６内でフレームポインタの一致を見つけられるかどうかによって２つの場合が続く。

ａ）バッファマネージャ７２がフレームポインタの一致を見つける。マルチコピーキャッシュ９６のエントリのコピー数が「１」であれば、バッファマネージャ７２はマルチコピーキャッシュエントリをデリートし、フレームポインタをリクレームキュー９８に入れる。キャッシュエントリのコピー数が「＞１」であれば、バッファマネージャ７２はコピー数を「１」だけデクリメントする。
ｂ）バッファマネージャ７２がマルチコピーキャッシュ９６内でフレームポインタの一致を見つけられない。これは、一致するフレームポインタが外部メモリ３６内のフレームのリンクトリストチェーンのバッファヘッダに既に移動されていることを意味する。バッファマネージャ７２はバッファヘッダに行って、コピー数を読出さなければならない。（メモリ内の）この値が「１」であれば、フレームはもはや必要ではなく、バッファマネージャ７２はフレームポインタをリクレームキュー９８に入れる。（メモリ内の）この値が「＞１」であれば、バッファマネージャ７２は（外部メモリ３６内にあった）フレームポインタ／コピー数のコピーをマルチコピーキャッシュ９６に入れ、コピー数を「１」だけデクリメントする。マルチコピーキャッシュ９６がフルであれば、バッファマネージャはより古いフレームポインタ／コピー数の１つを外部メモリ３６に移動させることによってスペースをクリアする。
リクレームキュー９８はもはや必要とされないリンクトリストチェーンを指すフレームポインタを保持する。バッファマネージャ７２は、マルチコピーキャッシュを処理してフレームポインタのコピー数が「１」である（すなわち、フレームの最後の送信がうまく終った）ことを見出すと、フレームポインタのリクレームキューに書込む。さらに、ポートベクタＦＩＦＯ７０は、１）フレームポインタのポートベクタがヌルであるか、２）転送ベクタの全出力キューがフルであったのでフレームポインタがキューに入れられることができなかったという条件下で、フレームポインタをリクレームキュー９８に書込む。最後に、ホストは、拡張バス出力キュー７７または管理ポート出力キュー７５に対してキューに入れられた単一コピーフレームの使用を終えると、（フレームポインタレジスタを用いて）フレームポインタをリクレームキュー９８に書込む。
バッファマネージャ７２はリクレームキューのエントリを処理するとき、フレームポインタのリンクトリストチェーンをたどり、各バッファをフリーバッファプール１０４に戻す。リクレームキュー構造の内部構造は図示されないが、本発明の例示的実施例においてはフレームポインタ（１４ビット）のみを含む。リクレームキューの書込側１００は６４エントリを保持し、リクレームキューの書込側１０２は１６エントリを保持する。
出力キュー７４と同様に、リクレームキュー９８は入力経路および出力経路を備えて構成される。入力経路または書込側１００によってバッファマネージャ７２はフレームポインタをリクレームキュー９８に入れることができる。出力経路または読出側１０２によってバッファマネージャ７２はフレームポインタを読出し、関連の全バッファをフリーバッファプール１０４に戻すことができる。フレームポインタのためのさらなるストレージは外部メモリ３６内に設けられるリクレームキューオーバーフロー領域１２２内に設けられる。
フレームポインタが空のリクレームキュー９８に書込まれると、これらは読出側１０２がフルになるまで書込側１００から読出側１０２へと移動する。リクレームキュー９８の書込側１００に書込まれるさらなるフレームポインタは外部メモリ３６内のリクレームキューオーバーフロー領域１２２に入れられる。一旦リクレームキュー９８の読出側１０２およびオーバーフロー領域１２２がフルになると、リクレームキュー９８に入れられるさらなるフレームポインタが書込側１００を満たし始める。
図１１はフリーバッファプール１０４の内部構造の例示的実施例を示す。フリーバッファプール１０４は、外部メモリ３６内の全フリーバッファ１４０を指すアドレスポインタを含んだＦＩＦＯである。フレームが受信されると、バッファマネージャ７２は入来するデータをストアするためにフリーバッファプール１０４から使用可能なアドレスポインタを捕捉する。バッファマネージャ７２はまたフリーバッファプール１０４からのアドレスポインタを（要求される場合）ホストプロセッサ２８に割当てる。ホストは、直接入力／出

力スペースにおけるレジスタ６０の中のフリーバッファプールレジスタを読出すか書込むことによってアドレスポインタを要求するかそれらをフリーバッファプール１０４に戻すことができる。フリーバッファプール１０４の書込側１０６および読出側１０８は本発明の例示的実施例においては各々６４エントリを保持する。

フリーバッファプール１０４は（出力キュー７４と同様に）入力経路および出力経路を備えて構成される。入力経路または書込側１０６により、バッファマネージャ７２またはホスト２８はアドレスポインタをフリーバッファプール１０４へと入れることができる。フリーバッファプール１０４の出力経路または読出側１０８により、バッファマネージャ７２はアドレスポインタをホスト２８に与え、またはプール１０４からアドレスポインタを引出して受信フレームデータをストアすることができる。使用可能なアドレスポインタのさらなるストレージ、フリーバッファプールのオーバーフロー領域１２０は上述のように外部メモリ３６内に設けられる。

スイッチ１２が起動すると、フリーバッファプールは読出側１０８からアドレスポインタを発生する。フレームが入来するときにフリーバッファプール１０４内のフリーリストが読出される。書込側１０６にトラフィック要求を扱うのに十分なバッファポインタがなければ、オーバーフロー領域１２０がより多くのバッファポインタを得るためにアクセスされる。

本発明のある実施例は、スイッチ１２が開始されるとバッファポインタを与える有利な配置および方法を提供する。スイッチ１２が初めに電源投入されるとき、外部メモリ３６内のオーバーフロー領域１２０がバッファポインタを含むことは必要とされない。代わりに、バッファポインタはオンザフライで発生される。スイッチ１２は電源投入されるとバッファポインタを発生し、それをオーバーフロー領域１２０に入れることができるが、このようなポインタは１６，０００個または３２，０００個存在することがあり、これによってスイッチ１２の電源投入手順が遅くなるであろう。本発明は、電源投入時に全バッファがフリーであり、これらのバッファのアイデンティティが既知であるという事実を使用する。したがって、バッファポインタは電源投入後に必要とされるときに図１０に示されるようにカウンタ１８０を用いて発生される。

フリーリストカウント発生器１８０がマルチプレクサ１８２の入力に接続される。フリーバッファプール１０４のフリーリストが開始時に空であるので、フリーリストカウンタ１８０はバッファポインタを発生する。一旦フリーリストが最高カウントに達すると、それはこれ以上バッファポインタを発生しない。

フレームパケットがスイッチ１２において受信されると、フレームパケットは固定長バッファへと分解する。典型的にフレームはさまざまなサイズである。バッファは２５６バイトのサイズであり、バッファのデータ部分は２４０バイトである。バッファ内容の送信後、バッファポインタがリクレームキュー９８に入れられるか、または、バッファチェーンをたどることができるならばフリーバッファプール１０４のフリーリストに直接入れられる。スイッチ１２の動作の間、フリーバッファプール１０４に戻されるどのアドレスポインタも書込側１０６から読出側１０８へと移動する。読出側１０８がフルとなれば、さらなるアドレスポインタはオーバーフロー領域１２０に渡される。一旦読出側１０８およびオーバーフロー領域１２０がフルとなると、フリーバッファプール１０４に入れられるさらなるアドレスポインタがプール１０４の書込側１０６を再び満たし始める。

図１３は本発明の実施例に従うマルチコピーキャッシュ９６の内部配列の概略図である。上で簡単に述べたように、マルチコピーキャッシュ９６へのエントリの時間順が維持される。本発明では、このように時間順が維持されるのは先行技術におけるようなタイムスタンプによってではなく、メモリ内の物理的順序によってである。本発明のマルチコピーキャッシュ９６はまた有効性ビットの使用を避け、代わりに後述するように有効性を符号化する。

図１３を参照すると、マルチコピーキャッシュ９６は４ウェイセットアソシアティブメモリとして構成される。マルチコピーキャッシュ９６へのエントリは上述のようにフレームポインタとそのコピー数とを含む。フレームポインタの最下位６ビットが、エントリがス

トアされるセットアソシアティブキャッシュ９６内の行を決定する。本発明の図示される実施例では、キャッシュ９６には６４行が存在するが、キャッシュサイズが大きくされれば他の行数も制限されない。
セットアソシアティブキャッシュ９６は４列に分割され、その各々が並行して探索される。バッファマネージャ７２がエントリをキャッシュ９６へとストアするとき、エントリは常に、第１の列の、フレームポインタの最下位６ビットによって示される行の最上位（５１：３９）ビットに入る。この行は読出され、全エントリが１３ビット分右にシフトされ、行は再び書込まれる。実際にキャッシュ９６に書込まれるエントリはフレームポインタの上位８ビットを含み、それはアドレスタグとフレームポインタに関連した５ビットコピー数を形成する。エントリがキャッシュ９６から読出されると、フレームポインタはキャッシュ９６の行数を指すビットおよびアドレスタグで再形成される。
行がフルであり、その行への新たなエントリが書込まれれば、キャッシュ９６内の最も古いエントリがキャッシュ９６から除去される。バッファヘッダ１４２に関して上述したように、除去されるフレームポインタに関連したコピー数は除去されるフレームポインタが指す外部メモリ内のフレームのバッファヘッダ１４２に書込まれる。したがって、外部メモリ３６にストアされるフレーム（すなわち、バッファ１４０）はコピー数をストアするためのマルチコピーキャッシュ９６のためのオーバーフロー領域となる。
本発明の有利な特徴の１つはセットアソシアティブキャッシュ９６に別個の有効ビットが存在しないことである。コピー数が０００００であるとき、エントリがもはや有効でないことをバッファマネージャ７２はわかっており、エントリをキャッシュ９６から除去する。これによってキャッシュ構成が簡素化される。本発明のキャッシュ９６の別の利点は非常に高速な探索が行なわれ得ることである。これは、バッファマネージャ７２がマルチコピーキュー９０を出たフレームポンタによって既に定められている単一の行を検査しさえすればよいためである。その行内の４つのエントリが並行して検査され、探索速度をさらに高める。４ウェイセットアソシアティブメモリとして説明しているが、これは例にすぎず、メモリは本発明の範疇から逸脱せずにｎウェイセットアソシアティブ方式となり得る。
上の説明から、本発明がキャッシュにおけるエントリの行ごとの物理的位置決めによってキャッシュエントリの時間順（エージ）を維持すると理解されるべきである。すなわち、キャッシュ内のエントリの物理的位置がエントリの相対的エージを示す。エントリはメモリにおけるエントリの物理的再順序付けによってエージングされる。
本発明のある実施例はポートごとにスイッチ１２によって切換えられるフレームのレイテンシをカスタマイズする。図１４を参照すると、ポートベクタＦＩＦＯ７０が受信ポートのプログラムされたスイッチモードを検査して、いつフレームポインタおよび関連の情報を送信ポートの適切な出力キュー７４へと入れるかを決定する。第１のモード（低レイテンシモード）では、ポートベクタＦＩＦＯ７０はいつフレームポインタを出力キュー７４に入れるかに対して制限を与えない。第２のモード（中間レイテンシモード）では、ポートベクタＦＩＦＯ７０はフレームの６４バイトが受信されて初めてフレームポインタを出力キュー７４に入れる。第３のモード（高レイテンシモード）では、ポートベクタＦＩＦＯ７０はフレームが完全に受信されて初めてフレームポインタを出力キュー７０に入れる。
いつポートベクタＦＩＦＯ７０がフレームポインタを出力キュー７４へと移動するかのタイミングを変えるいくつかの特殊な場合があり、それらは、１）第１または第２のモードの１０Ｍｂ／ｓポートから１００Ｍｂ／ｓポートへのフレーム転送と、２）管理ポート３０へのフレーム転送と、３）拡張バスポートへのフレーム転送とを含む。場合１）では、１０Ｍｂ／ｓポートから１００Ｍｂ／ｓポートへの速度不一致によって転送モードが強制的に第３の高レイテンシモードとされる。場合２）では、管理ポートへと移動する全フレームが第３のモードのフレームである。場合３）では、拡張バスポートへのどのフレーム転送も拡張バスポート２６のスイッチモードを用いる。マルチコピーポートベクタが特殊な場合のポートの１つを含む場合、ポートベクタ全体に対するフレームポインタのキュー

入れはポートベクタ内で表わされる最長レイテンシスイッチモードのそれになる。たとえば、フレームが第１または第２のモードのポートによって受信され、そのマルチコピー転送ポートベクタが管理ポート３０を含めば、スイッチモードは第３のモードである。この場合、フレームが完全に受信されて初めてフレームポインタのコピーが全出力キュー７４に入れられる。

スイッチモードをここでより詳細に説明する。入力（すなわち、受信）ポートに当てはまるスイッチモードが転送レイテンシ（一旦スイッチ１２がフレームを受信し始めるとどの程度後にスイッチ１２がフレームを転送するか）と出力ポートへのフラグメント／エラー伝搬を低減する能力とを決定する。第２の中間レイテンシモードは各ポートに対するデフォルトであるが、スイッチモードはレジスタ６０ではポートごとにプログラム可能である。

これら３つのモデルのすべてにおいて、内部ＭＡＣポートの受信ＦＩＦＯ５２で受信されるフレームデータはできるだけ早く外部メモリ５２内のバッファ１４０に転送される。ほぼ同時に、ルールチェッカ４２または５８が宛先アドレスおよびソースアドレス、受信ポート数、フレームポインタ、ならびにいくつかの付加的情報を受信し、適切なルックアップを行なう。一旦ルックアップが完了すると、ルールチェッカ４２または５８はフレームポインタおよび転送ポートベクタをポートベクタＦＩＦＯ７０に戻す。

ポートベクタＦＩＦＯはポートベクタ内で識別される出力ポートのための出力キュー７４の書込側７６にフレームポインタを入れる。受信ポートのスイッチモードは、ポートベクタＦＩＦＯ７０がポートベクタ（およびフレームポインタ）を受取るときから、それがフレームポインタを出力キュー７４に入れるときまでの間のレイテンシを規定する。これは以下の３つのモードに対して説明される。一旦フレームポインタが出力キュー７４の読出側７８に移動すると、バッファマネージャ７２はフレームポインタを読出し、送信をスケジュールする。バッファマネージャはフレームポインタによって特定されるアドレスからフレームデータを移動させ始める。一旦ＭＡＣポートの送信ＦＩＦＯ５４がその開始点に設定されると（そして、データ送信のために媒体が使用可能であると想定すると）、フレーム送信が始まる。

第１のモードは最低のレイテンシを与えるように設計される。フレームはライン－レート速度で受信され、転送される。この第１のモードにおいてはネットワークエラーに対する保護がなく、これは、フレームがフラグメント（すなわち、＜６４バイトの長さ）であるかＣＲＣエラーを含むかが判断され得る前にフレームが送信のためにキューに入れられるためである。第１のモードにおいて、フレーム受信は出力ポートでのフレーム送信が始まるまでに完了していないかもしれない。受信フレームが短すぎる場合または無効なＣＲＣで終る場合、受信ＭＡＣは外部メモリ３６内のバッファヘッダ１４２に印を付けてこれらの条件を示す。送信ＭＡＣは、後に短すぎるものか無効なＣＲＣで終るフレームの送信が始まればＭＡＣが無効なＣＲＣを発生することを保証する。送信ＭＡＣがフレーム送信を始めておらず、バッファヘッダ１４２が短すぎるものか無効なＣＲＣで終るフレームを示している場合、バッファマネージャ７２はフレームを出力ポートへと転送しない。

第２のモードはフレームを転送するための低レイテンシとあるネットワークエラーに対する保護とを与える。フレームは６４バイト以上が受信された後に受信され、転送される。これによってスイッチ１２がフレームのフラグメントをフィルタ処理する（すなわち、転送しない）ことが可能となるが、これは６４バイトよりも大きいＣＲＣエラーフレームを完全にはフィルタ処理しない。

第２のモードにおいては、受信ＭＡＣで６４バイトのしきい値を達成したフレームのフレームポインタは適切な出力キュー７４に入れられる。最小の６４バイトのしきい値を達成できないフレームはデリートされ、それらのフレームポインタは出力キュー７４に入れられない。６４バイト以上の受信フレームが無効なＣＲＣで終れば、受信ＭＡＣは外部メモリ３６内のバッファヘッダ１４２に印を付けてこの条件を示す。後に無効なＣＲＣで終る６４バイト以上のフレームの送信が開始されるときには、送信ＭＡＣは不良なＣＲＣで送信を終了する。送信ＭＡＣがフレーム送信を開始しておらず、バッファヘッダ１４２が無

効なＣＲＣで終るフレーム（６４ビット以上）であることを示している場合、バッファマネージャはフレームポインタを（単一コピー転送のための）リクレームキュー９８または（マルチコピー転送のための）マルチコピーキュー９６へと出力ポート７４への転送なしに戻す。

第３のモードは３つのモードの中で最高レベルのネットワークエラー保護を与えるがより高い転送レイテンシを有するストアアンドフォワードモードである。フレームは、スイッチ１２がそれらを出力ポートに転送する前に完全に受信される。このモードでは、スイッチ１２は転送の前に全てのフラグメントおよびＣＲＣエラーフレームをふるい分ける。第３のモードにおいて、一旦有効フレームが受信側でうまく完了すると（すなわち、有効なＣＲＣを持ち、６４バイト以上であると）、フレームポインタが適切な出力キュー７４に入れられる。受信エラー（無効ＣＲＣ、短すぎるもの（＞６４バイト）等）で終るフレームはデリートされ、それらのフレームポインタは出力キュー７４に入れられない。

ポートベクタＦＩＦＯ７０は、受信ポートの選択されたモードと受信されたデータ量とに依存してポートベクタを出力キュー７４に入れる決定を行なう。上述の実施例では、３つのしきい値があるが他の実施例では異なる数のしきい値が存在する。例示的実施例では、これらのしきい値は１）ｎ＜６４バイトであるようなｎバイト（たとえば６バイト）の受信、２）６４バイトの受信、および３）全フレームの受信である。

本発明はしきい値に基づいてフレームを出力キュー７４へと転送する。ポートベクタＦＩＦＯ７０は、受信されるデータタイプの量とポートがプログラムされたモードとに基づいて送信シーケンスを再び順序付ける。例示的実施例は受信されたデータの量に基づいて転送の決定を行なうが、本発明の他の実施例では、受信されるデータタイプのような他の要因に基づいて転送の決定が行われる。

本発明の転送方式を実施するにあたって、バッファマネージャ７２はフレームポインタを受信ポートと関連付ける、キャッシュメモリ（ＣＡＭ）１６１内のテーブル１６０を維持する。ポートベクタＦＩＦＯ７０が新しいポートベクタおよびフレームポインタをルールチェッカ４２または５８から受信するたびに、それは関連付けを行なって受信ポートがフレーム受信を終えたかどうかを判断し、終えていなければどれほどのフレームが既に受信されているかを判断する。ポートベクタＦＩＦＯ７０が受信ポートのアイデンティティに関する情報をルールチェッカ４２または５８から受信することはない。ポートベクタが受取る唯一のポートの何らかの識別を与える情報はフレームポインタである。

ポートベクタＦＩＦＯ７０はフレームポインタでアドレステーブル１６０に問合せをする。フレームがなお受信されていればアドレステーブルは受信ポートを戻し、またはアドレステーブル１６０はフレームポインタを見つけることができないときはフレームが既に受信されたことを意味する。一旦フレームが完全に受信されると、フレームポインタがアドレステーブル１６０から移動される。これは、第３のしきい値（フレーム完了）が満たされたことを意味する。したがって、フレームポインタは直ちに出力キュー７４に入れられ得る。

アドレステーブル１６０が受信ポートを戻せば、ポートベクタＦＩＦＯ７０がフレームポインタおよび関連の情報を保持領域１６２に入れ、その受信ポートからの２信号を監視し始める。これらの２信号は３つの事象のうちの１つを示す。第１の事象はポートがｎバイトを受信するときに示される。その時点で、そのポートが第１のモードにあれば、ポートベクタＦＩＦＯ７０がフレームポインタを適切な出力キュー７４に送ることによってその処理を開始する。受信ポートが第１のモードになければ、ポートベクタＦＩＦＯ７０は第２の事象の発生を示す信号が受信されるまで待機する。このポートが第２のモードにあれば、ポートベクタＦＩＦＯ７０はフレームポインタを保持領域１６２から解放し、適切な出力キュー７４に入れる。最後に、受信ポートが第３のモードにあれば、ポートベクタＦＩＦＯ７０はフレームが完全であることを示すフラグの受信を待つ。各受信ポート（図１４の参照番号１６４）がこのフラグを維持し、この情報をポートベクタＦＩＦＯ７０に提供する。フレームポインタに関連付けられたポートの決定はポートベクタＦＩＦＯ７０次第である。ポートベクタＦＩＦＯ７０は各ポートのモードを識別する情報を維持する。要

約すると、フレームポインタが受信されると、ポートベクタＦＩＦＯ７０は最初にバッファマネージャ７２のアドレステーブル１６０に問合せをして受信ポートを決定し、その受信ポートのためのモードを決定し、受信ポートからのフラグを監視し、モードおよびフラグに従ってフレームポインタを解放する。

バッファマネージャ７２はこの発明のスイッチ１２においてさまざまな機能を果たす。バッファマネージャ７２はこれらの機能を実現するための制御論理および／またはソフトウェアを含み、バッファマネジャ７２の機能に関する上記の説明があれば当業者なら容易に実現できる。

本発明が詳細に説明され、図示されたが、これは図示および例示のためのものにすぎず、限定するものとは理解されるべきでなく、本発明の精神および範疇が請求の範囲によってのみ規定されることが明らかに理解される。

【図１】

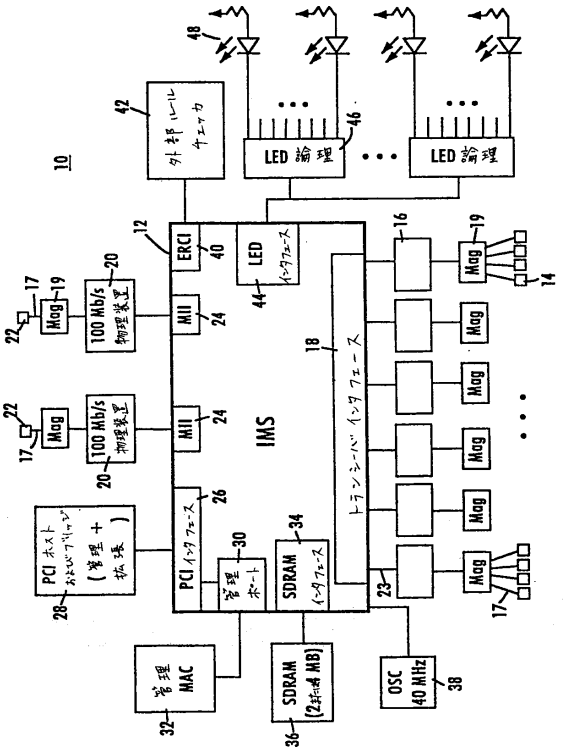
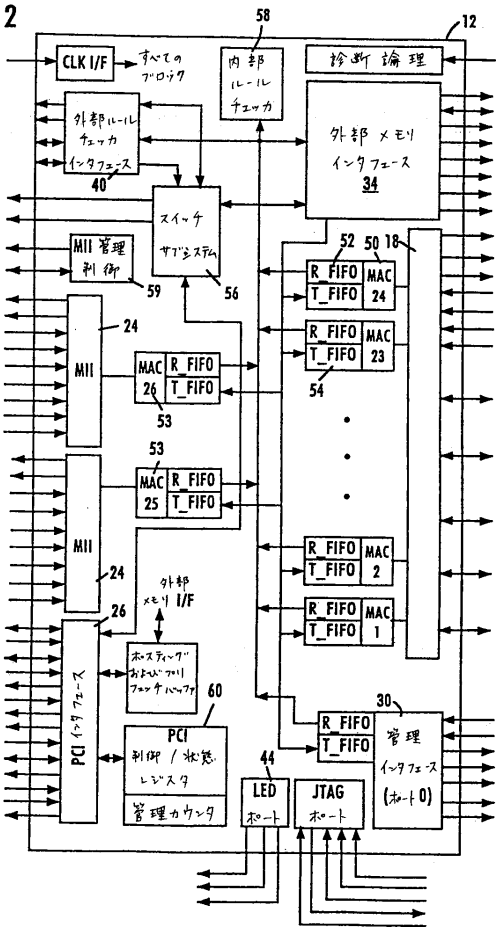


Fig. 1

【図２】

Fig. 2



【図３】

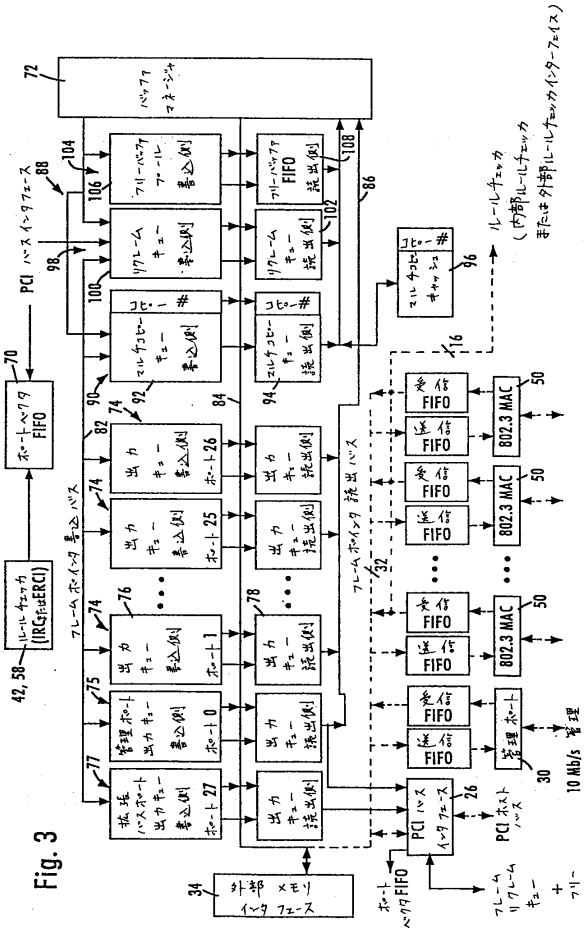


Fig. 3

【図４】

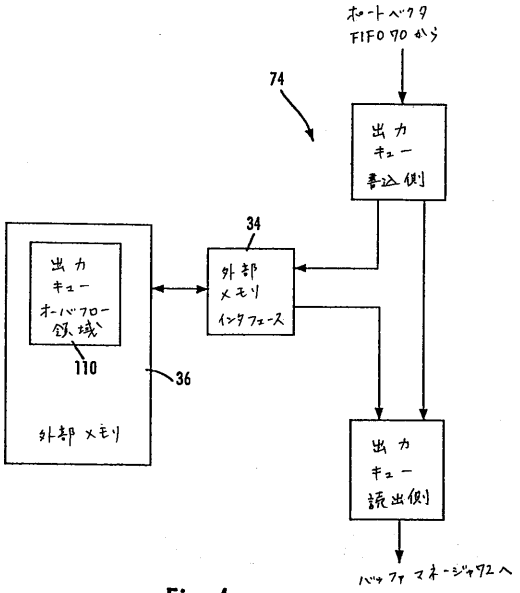


Fig. 4

【図５】

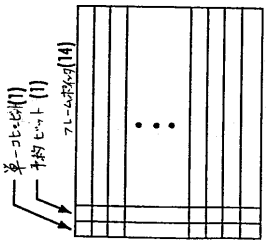


Fig. 5

【図６】

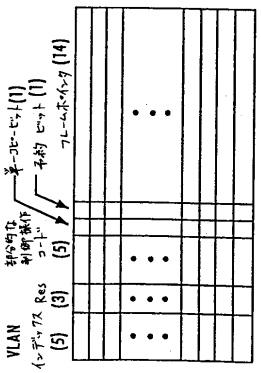


Fig. 6

【図７】

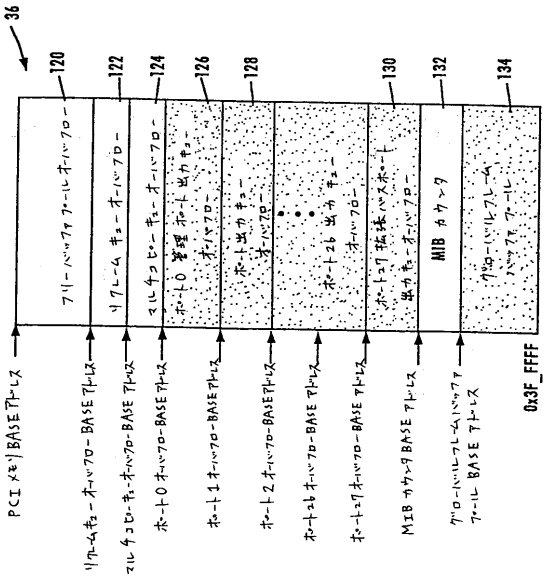


Fig. 7

【図８】

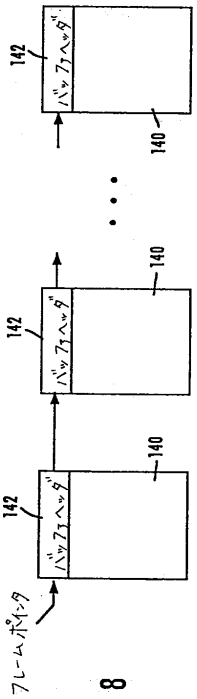


Fig. 8

【図９ａ】

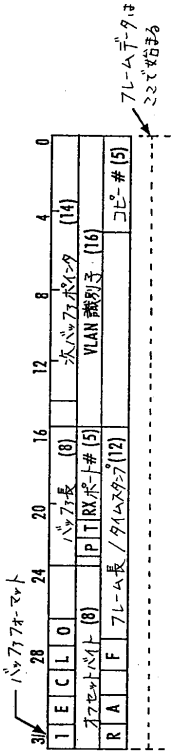


Fig. 9a

【図９ｂ】

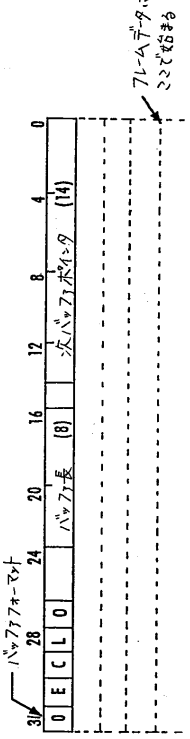


Fig. 9b

【図１０】

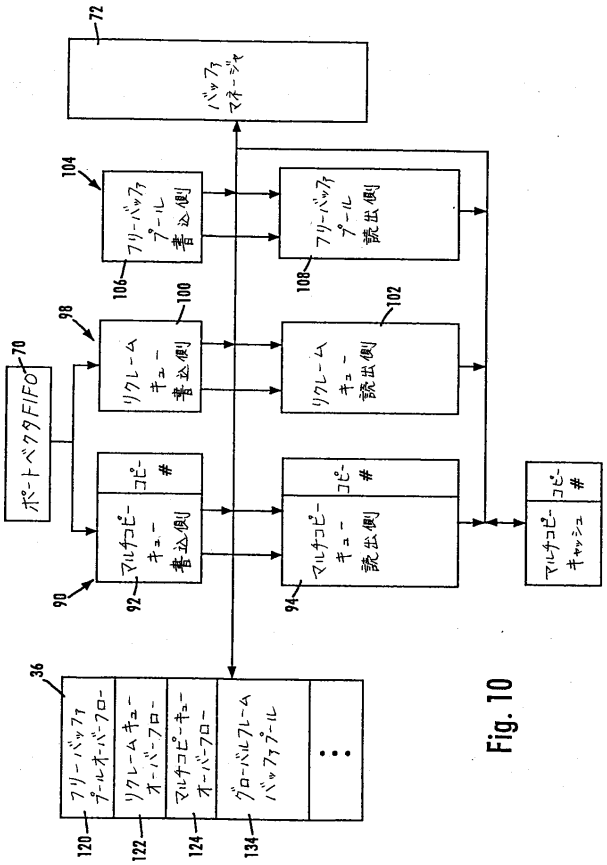


Fig. 10

【図１１】

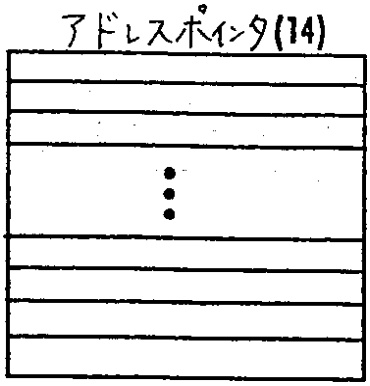


Fig. 11

【図１２】

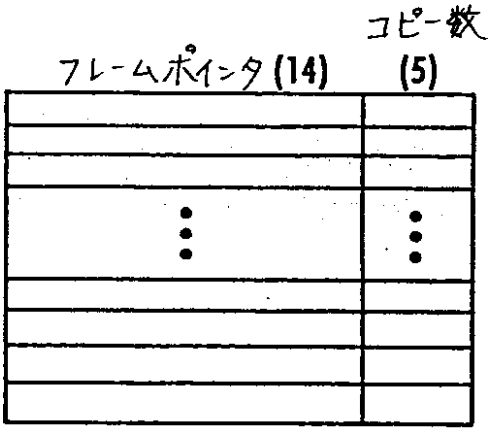


Fig. 12

【図１３】

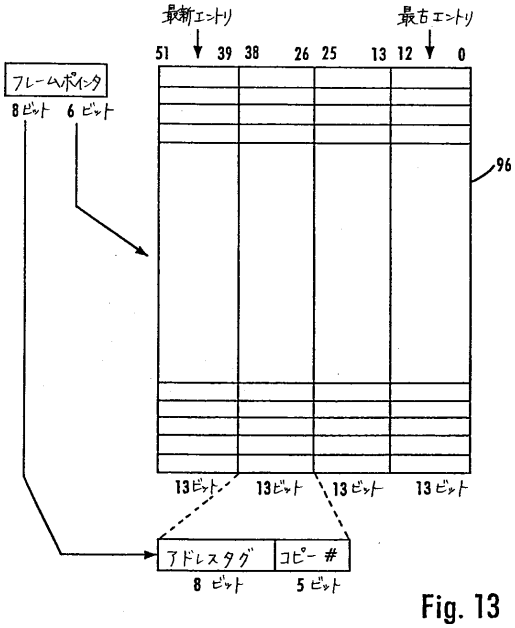


Fig. 13

【図１４】

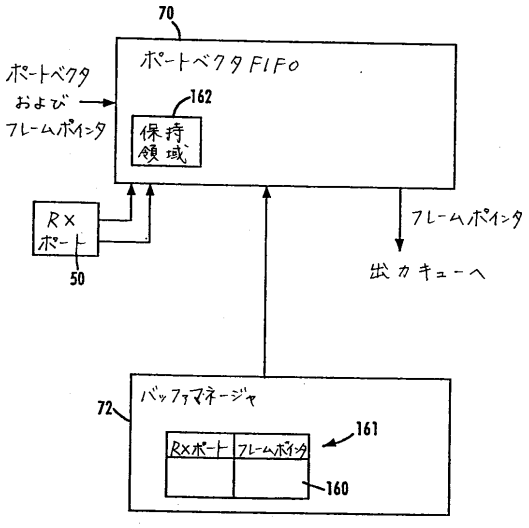


Fig. 14

フロントページの続き

(74)代理人 100096781
弁理士 堀井 豊
(72)発明者 エリムリ，バハディール
アメリカ合衆国、９４０４０ カリフォルニア州、マウンテン・ビュー、カリフォルニア・ストリート、２１０１、ナンバー・１０９
(72)発明者 ルナルデュー，トーマス・ジェファーソン
アメリカ合衆国、９５１１７ カリフォルニア州、サン・ノゼ、ブラックフォード・アベニュ、３７０１
(72)発明者 エグバート，チャンダン
アメリカ合衆国、９５１３２ カリフォルニア州、サン・ノゼ、ブルームズバリィ・ウェイ、３６３２

審査官 吉田 隆之

(56)参考文献 特開平４－１７５０３４（ＪＰ，Ａ）
特開平６－３３８８９９（ＪＰ，Ａ）
特開平８－８９０６（ＪＰ，Ａ）
特開平６－３３４６５２（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
H04L 12