【公報種別】特許法第１７条の２の規定による補正の掲載
【部門区分】第７部門第３区分
【発行日】平成17年1月6日(2005.1.6)

【公表番号】特表2004-507163(P2004-507163A)
【公表日】平成16年3月4日(2004.3.4)
【年通号数】公開・登録公報2004-009
【出願番号】特願2002-520493(P2002-520493)
【国際特許分類第７版】
　Ｈ０４Ｌ　12/56
【ＦＩ】
　Ｈ０４Ｌ　12/56　　Ｆ

【誤訳訂正書】
【提出日】平成14年4月18日(2002.4.18)
【誤訳訂正１】
【訂正対象書類名】明細書
【訂正対象項目名】全文
【訂正方法】変更
【訂正の内容】
【書類名】明細書
【発明の名称】スイッチ・ファブリック・チップセット・システムを介した送信用にパケットを準備するためのシステム、方法および製造品
【特許請求の範囲】
【請求項１】スイッチ・ファブリックを介した送信用にデータグラムを準備するための方法であって、
（ａ）ヘッダ部およびボディ部を有するデータグラムを受信すること、
（ｂ）前記ヘッダ部を前記ボディ部から分離すること、
（ｃ）前記ヘッダ部を処理のために転送すること、
（ｄ）前記ボディ部をメモリに格納すること、
（ｅ）前記ヘッダ部を処理後に受信すること、および
（ｆ）前記処理されたヘッダ部を前記メモリに格納することを含む方法。
【請求項２】前記メモリにおける空間を、前記ヘッダ部を格納するために、前記ボディ部を前記メモリに格納する前に確保し、前記確保された空間へのリンクを、前記確保された空間を位置指定するために生成することをさらに含む請求項１に記載の方法。
【請求項３】前記処理されたヘッダを格納することが、前記リンクを利用して前記確保された空間を位置指定すること、および前記処理されたヘッダを前記確保された空間に格納することをさらに含む請求項２に記載の方法。
【請求項４】前記ボディ部が複数のセルに分割され、前記ボディ部を前記メモリに格納することが、各セルを前記メモリに順次格納すること、および前記順次格納されたセルの間のリンクを生成することをさらに含む請求項１に記載の方法。
【請求項５】前記処理されたヘッダ部が、前記パケットがルーティングされるべき先の前記スイッチ・ファブリック・システムにおける宛先を指示する宛先ベクトルを含む請求項１に記載の方法。
【請求項６】前記宛先ベクトルが３２ビットのデータを含む請求項５に記載の方法。
【請求項７】前記データグラムを、前記スイッチ・ファブリックを介した後続の送信のために、出力優先キューに割り当てることをさらに含む請求項１に記載の方法。
【請求項８】スイッチ・ファブリックを介した送信用にデータグラムを準備するためのシステムであって、
ヘッダ・キューおよびメモリに結合されたイングレス・ダイレクト・メモリ・アクセス（

ＤＭＡ）を有するポート・コントローラを含み、前記ヘッダ・キューはパケット・プロセッサに結合され、前記ポート・コントローラは、データグラムを受信するための論理を有し、前記データグラムはヘッダ部およびボディ部を含み、さらに、
前記イングレスＤＭＡは、前記データグラムの前記ヘッダ部を前記ボディ部から分離すること、前記ヘッダ部を処理のために前記ヘッダ・キューへ前記パケット・プロセッサによって転送すること、前記ボディ部をメモリに格納すること、前記ヘッダ部を前記ヘッダ・キューから処理後に受信すること、および前記処理されたヘッダ部を前記メモリに格納することを行うように適合されているシステム。
【請求項９】前記ヘッダ部を格納するために、前記ボディ部を前記メモリに格納する前に前記メモリにおける空間を確保する論理、および、前記確保された空間を位置指定するために、前記確保された空間へのリンクを生成する論理をさらに含む請求項８に記載のシステム。
【請求項１０】前記処理されたヘッダを格納することが、前記リンクを利用して前記メモリにおいて前記確保された空間を位置指定すること、および前記処理されたヘッダを前記確保された空間に格納することをさらに含む請求項９に記載のシステム。
【請求項１１】前記ボディ部が複数のセルに分割され、前記ボディ部を前記メモリに格納することが、各セルを前記メモリに順次格納すること、および前記順次格納されたセルの間のリンクを生成することをさらに含む請求項８に記載のシステム。
【請求項１２】前記処理されたヘッダ部が、前記パケットがルーティングされるべき先の前記スイッチ・ファブリック・システムにおける宛先を指示する宛先ベクトルを含む請求項８に記載のシステム。
【請求項１３】前記宛先ベクトルが３２ビットのデータを含む請求項１２に記載のシステム。
【請求項１４】前記データグラムを、前記スイッチ・ファブリックを介した後続の送信のために、出力優先キューに割り当てるための論理をさらに含む請求項８に記載のシステム。
【請求項１５】スイッチ・ファブリックを介した送信用にデータグラムを準備するためのコンピュータ・プログラム製品であって、
（ａ）データグラムを受信するためのコンピュータ・コードを含み、前記データグラムはヘッダ部およびボディ部を有し、さらに、
（ｂ）前記ヘッダ部を前記ボディ部から分離するためのコンピュータ・コードと、
（ｃ）前記ヘッダ部を処理のために転送するためのコンピュータ・コードと、
（ｄ）前記ボディ部をメモリに格納するためのコンピュータ・コードと、
（ｅ）前記ヘッダ部を処理後に受信するためのコンピュータ・コードと、
（ｆ）前記処理されたヘッダ部を前記メモリに格納するためのコンピュータ・コードとを含む、コンピュータ・プログラム製品。
【請求項１６】前記ヘッダ部を格納するために、前記ボディ部を前記メモリに格納する前に前記メモリにおける空間を確保するためのコンピュータ・コード、および、前記確保された空間を位置指定するために、前記確保された空間へのリンクを生成するためのコンピュータ・コードをさらに含む請求項１５に記載のコンピュータ・プログラム製品。
【請求項１７】前記処理されたヘッダを格納するための前記コンピュータ・コードが、前記リンクを利用して前記確保された空間を位置指定するためのコンピュータ・コード、および、前記処理されたヘッダを前記確保された空間に格納するためのコンピュータ・コードをさらに含む請求項１６に記載のコンピュータ・プログラム製品。
【請求項１８】前記ボディ部が複数のセルに分割され、前記ボディ部を前記メモリに格納するための前記コンピュータ・コードが、各セルを前記メモリに順次格納するためのコンピュータ・コード、および、前記順次格納されたセルの間のリンクを生成するためのコンピュータ・コードをさらに含む請求項１５に記載のコンピュータ・プログラム製品。
【請求項１９】前記処理されたヘッダ部が、前記パケットのルーティング先である前記スイッチ・ファブリック・システムにおける宛先を指示する宛先ベクトルを含む請求項１に

記載のコンピュータ・プログラム製品。
【請求項２０】前記データグラムを、前記スイッチ・ファブリックを介した後続の送信のために、出力優先キューに割り当てるためのコンピュータ・コードをさらに含む請求項１に記載のコンピュータ・プログラム製品。
【発明の詳細な説明】
【０００１】
（関連出願の相互参照）
本願は、２０００年８月１１日出願の「ＧＩＧＡＢＩＴ　ＥＴＨＥＲＮＥＴ（登録商標）　ＩＰ　ＳＷＩＴＣＨ　ＣＨＩＰＳＥＴ」という名称の米国仮特許出願第６０／２２５０３４号の利点を主張するものであり、これは全体としてあらゆる目的のために参照により本明細書に組み込まれる。
【０００２】
（発明の分野）
本発明は一般にデータ・ルーティング・システムに関し、より詳細には、ネットワーク・スイッチを介したパケットのルーティングに関する。
【０００３】
（発明の背景）
イーサネット（登録商標）は、幅広く導入されているローカル・エリア・ネットワーク技術である。米国電気電子学会（ＩＥＥＥ）８０２．３の規格において指定されたイーサネットは当初、Ｘｅｒｏｘによって開発され、次いでさらにＸｅｒｏｘ、ＤＥＣおよびＩｎｔｅｌによって開発された。イーサネットＬＡＮは通常、同軸ケーブルまたは特殊なグレードのツイスト・ペア線を使用する。一般に導入されたイーサネット・システムは１０ＢＡＳＥ－Ｔと呼ばれ、最大１０Ｍｂｐｓの伝送速度を提供する。デバイスがケーブルに接続され、キャリア検知多重アクセス衝突検出（ＣＳＭＡ／ＣＤ）プロトコルを使用して、アクセスのために競争する。高速イーサネットまたは１００ＢＡＳＥ－Ｔは最大１００メガビット毎秒の伝送速度を提供し、ＬＡＮバックボーン・システムのために使用することができ、１０ＢＡＳＥ－Ｔカードを有するワークステーションをサポートする。ギガビット・イーサネットは、さらにより高レベルのバックボーン・サポートを１０００メガビット毎秒（１ギガビットまたは１０億ビット毎秒）で提供する。
【０００４】
高速イーサネットは１００メガビット毎秒のデータレートを提供するローカル・エリア・ネットワーク伝送規格である（「１００ＢＡＳＥ－Ｔ」と呼ばれる）。既存の１０メガビット毎秒（１０ＢＡＳＥ－Ｔ）のイーサネット・カードを有するワークステーションを、高速イーサネット・ネットワークに接続することができる。（１００メガビット毎秒は共用データレートであり、各ワークステーションへの入力は１０Ｍｂｐｓカードによって制約される）。ギガビット・イーサネットはローカル・エリア・ネットワーク伝送規格であり、１０億ビット毎秒（１ギガビット）のデータレートを提供する。ギガビット・イーサネットはＩＥＥＥ８０２．３規格において定義されている。ギガビット・イーサネットはエンタープライズ・バックボーンとして使用することができる。ギガビット・イーサネットを光ファイバで搬送することができる（銅媒体上では非常に短い距離で可能）。１０および１００Ｍｂｐｓカードを有する既存のイーサネットＬＡＮは、ギガビット・イーサネット・バックボーンに組み込むことが可能である。
【０００５】
データは、パケット交換網を介して、デジタル信号を使用して送信される。パケット交換網では、ユーザが同じパスを同時に共有することができ、データ・ユニットが移動するルートは、状態が変化すると変わることがある。パケット交換では、メッセージがある数のバイトのユニットであるパケットに分割される。送信側および宛先のネットワーク・アドレスがパケットに追加される。各ネットワーク・ポイントがパケットを見て、これを次にどこへ送信するべきかを調べる。同じメッセージにおけるパケットは異なるルートを移動する可能性があり、それらが送信された順序と同じ順序で到着しない可能性がある。宛先

では、メッセージのパケットが収集され、元のメッセージに再び組み立てられる。
【０００６】
階層化は、プログラミングを別々のステップに編成することであり、これらは順次実行され、ある量の情報の送信または受信などの全体の機能が完了するまで各ステップの結果を次のプログラムまたは層へ渡す特定のインターフェイスによって決められる。通信プログラムはしばしば階層化される。通信プログラムのための参照モデル、開放型システム間相互接続（ＯＳＩ）は、通信交換の両側にある２つのプログラムが等しいセットの層を使用する階層化されたプロトコルのセットである。ＯＳＩは７つの層を含み、それぞれが、プログラム対プログラムの通信をコンピュータの間で行うために実行されなければならない異なる機能を担っている。伝送制御プロトコルおよびインターネット・プロトコル（ＴＣＰ／ＩＰ）は、２層セットのプログラムの一例であり、インターネット通信のためのトランスポートおよびネットワーク・アドレス機能を提供するものである。
【０００７】
スイッチは、データのユニットをその次の宛先へ送信するためのパスまたは回路を選択するインターネットワーキング・デバイスである。スイッチはルータの機能も含み、ルート、および特に、どの隣接ネットワーク・ポイントにデータが送信されるべきであるかを決定する。階層化の開放型システム間相互接続（ＯＳＩ）通信モデルでは、スイッチはデータ・リンク層－第２層に関連付けられる。スイッチは、ネットワーク層－第３層のルーティング機能を実行することもできる。第３層スイッチは時としてＩＰスイッチとも呼ばれる。
【０００８】
ＯＳＩのデータ・リンク層－第２層は、データをネットワークにおける物理リンクを横断して移動させることに関係する。ネットワークでは、スイッチを、データ・メッセージを第２層レベルでリダイレクトするデバイスとすることができ、これは宛先の媒体アクセス制御（ＭＡＣ）アドレスを使用して、どこにメッセージをダイレクトするべきであるかを決定する。データリンク層は２つの副層を含み、これらはＩＥＥＥ－８０２のＬＡＮ規格において記載されている。すなわち媒体アクセス制御（ＭＡＣ）および論理リンク制御（ＬＬＣ）である。データ・リンク層は、最初の接続がセット・アップされていることを確かめ、出力データをデータ・フレームに分割し、データがうまく到着したという受信側からの応答を処理する。これはまた、フレームの特殊な位置のビット・パターンを解析することによって入力データがうまく受信されたことも保証する。
【０００９】
ＯＳＩのネットワーク層－第３層は、ネットワークにおける隣接ノードのアドレスを知ること、ルートおよびサービス品質を選択すること、およびローカル・ホスト・ドメインのための入力メッセージを認識し、かつトランスポート層へ転送することに関係する。スイッチは第３層デバイスであり、第３層の機能を実行することができる。ＩＰアドレスは第３層アドレスである。
【００１０】
スイッチングは、その本質により、過渡的なデータ・フロー（ｔｒａｎｓｉｅｎｔ ｄａｔａ ｆｌｏｗｓ）を含む。ギガビット速度で、これらの一時的なものは非常に高速に、かつ非常にランダムな方法において発生する。過去においては、スイッチ・データ・フローのどのコンポーネントがダイレクトされるべきであるかを十分に早く決定して、スイッチを通過するデータ・フローを最大限にし、かつスイッチにおいて妨害となるトラフィックを低減することは困難であった。
【００１１】
（発明の概要）
スイッチ・ファブリックを介した送信用にデータグラムを準備するためのシステム、方法および製造品を提供する。一般に、ヘッダ部およびボディ部を有するデータグラムが最初に受信される。ヘッダ部はボディ部から分離され、次いで処理のために転送される。次いで、分離されたボディ部がメモリに格納される。処理後、ヘッダ部が受信され、メモリに

格納される。
【００１２】
本発明の一実施態様では、ヘッダ部を格納するために、ボディ部をメモリに格納する前にメモリにおける空間を確保することができ、確保された空間へのリンクも、確保された空間を位置指定するために生成される。このような実施態様におけるオプションとして、処理されたヘッダを格納することが、リンクを利用して確保された空間を位置指定すること、次いで、処理されたヘッダを確保された空間に格納することを含んでもよい。
【００１３】
本発明の一態様では、ボディ部を複数のセルに分割することができる。このような態様では、ボディ部をメモリに格納することがさらに、各セルをメモリに順次格納すること、および順次格納されたセルの間のリンクを生成することを含むことができる。本発明の別の態様では、処理されたヘッダ部が、パケットがルーティングされるべき先のスイッチ・ファブリック・システムにおける宛先を指示する宛先ベクトルを含むことができる。このような態様では、宛先ベクトルが３２ビットのデータを含むことができる。
本発明の別の実施態様では、データグラムを、スイッチ・ファブリックを介した後続の送信のために、出力優先キューに割り当てることができる。
【００１４】
以上および他の特徴、態様、および利点は、以下の詳細な説明、添付の請求項、および添付の図面から、より良く理解される。
【００１５】
（詳細な説明）
イーサネット・スイッチ・ファブリック・チップセット・システムは、スイッチ・ファブリック・トポロジ、および、３２ギガビットのイーサネット・データ・ストリームのワイヤスピード、ノンブロッキング、第２層および第３層のスイッチングを支持することができるマルチプロセッサ・ポート・コントローラを使用して開示されている。これは、ネットワーク・プロセッサとして当技術分野で知られているものに拡張することができる。スイッチング技術の基本的な理解のために、読者は、Ｒｉｃｈ　Ｓｅｉｆｅｒｔ（Ｊｏｈｎ　Ｗｉｌｅｙ＆Ｓｏｎｓ，Ｉｎｃ．２０００）によるＴｈｅ　Ｓｗｉｔｃｈ　Ｂｏｏｋ：ｔｈｅ　ｃｏｍｐｌｅｔｅ　ｇｕｉｄｅ　ｔｏ　ＬＡＮ　ｓｗｉｔｃｈｉｎｇ　ｔｅｃｈｎｏｌｏｇｙを指示される。これは、あらゆる目的のために、参考文献によって完全に本明細書に組み込まれている。
【００１６】
本明細書に記述したシステムにより、どのスイッチ・ファブリック構成要素の過渡的なデータ・フローをダイレクトすることができるかに関する決定は、他のデータ・フロー・トラフィックによるブロッキングを防止しながら、スイッチ・ファブリック・システムを通るデータ・フローを最大にするのに十分な速さで実施することが可能である。これは、過渡現象が生じる前に可能な限り多くの決定を実施し、かつ、考慮の決定を、スイッチ・ファブリックを通る経路全体ではなく、単に経路の次のスイッチ・ファブリック構成要素とすることによって、達成される。この解決法の一部として、スイッチ・ファブリックを通して状況と制御の情報を送り返すことに備え、かつ、制御経路における初期接続手順の必要性を軽減または排除することを支援するために、データ経路の他に、分離制御経路を提供することが可能である。この解決法の他の一部として、制御経路を使用して更新することができる各スイッチ・ファブリック構成要素の内側に、スイッチング・テーブルを設けることが可能である。
【００１７】
（システム構成）
図１は、本発明の実施形態による、マルチポート・スイッチ・アーキテクチャ１００の概略的なブロック図である。一般に、このマルチポート・スイッチ・アーキテクチャ１００は、スイッチ・ファブリック１０４を介して互いに相互接続された複数のポート・コントローラ（例示した実施形態では３２のポート・コントローラ）１０２、ポート・コントロ

ーラ１０２に接続されたネットワーク制御プロセッサ１０６、およびスイッチ・ファブリック１０４を備える。図１に示したように、要素間の接続（例えば１０８、１１０、１１２）は双方向である。

【００１８】

さらに図１を参照すると、このマルチポート・スイッチ・アーキテクチャ１００の特定の実施形態では、スイッチ・ファブリックは、クロス・ネットワーク構成／トポロジとして知られている双方向コンジット（例えばコンジット１１０）を介して共に接続されている複数のスイッチ・ファブリック構成要素（例示した実施形態では１２のスイッチ・ファブリック構成要素）１０４ａ、１０４ｂ、１０４ｃ、１０４ｄ、１０４ｅ、１０４ｆ、１０４ｇ、１０４ｈ、１０４ｉ、１０４ｊ、１０４ｋ、１０４ｌを備えている。図１に示した特定の構成は、スイッチ・ファブリック構成要素（１０４ａ、１０４ｂ、１０４ｃ、１０４ｄおよび１０４ｉ、１０４ｊ、１０４ｋ、１０４ｌ）の外側ステージの対を接続するスイッチ・ファブリック構成要素（１０４ｅ、１０４ｆ、１０４ｇ、１０４ｈ）の中間ステージを有する３つのステージまたは層のクロス・ネットワーク・トポロジとして知られている。

【００１９】

スイッチ・ファブリック１０４のクロス・ネットワーク・トポロジにより、各外側ステージのスイッチ・ファブリック構成要素１０４ａ、１０４ｂ、１０４ｃ、１０４ｄおよび１０４ｉ、１０４ｊ、１０４ｋ、１０４ｌは、独立したコンジットを介して、中間ステージのスイッチ・ファブリック構成要素１０４ｅ、１０４ｆ、１０４ｇ、１０４ｈの各々に直接結合される。例えば、図１では、スイッチ・ファブリック構成要素１０４ａは、それを４つの中間ステージのスイッチ・ファブリック構成要素１０４ｅ、１０４ｆ、１０４ｇ、１０４ｈに結合する４つのコンジットを有する。逆に、スイッチ・ファブリック１０４のクロス・ネットワーク・トポロジにより、中間ステージのスイッチ・ファブリック構成要素１０４ｅ、１０４ｆ、１０４ｇ、１０４ｈの各々は、独立したコンジットを介して、外側ステージのスイッチ・ファブリック構成要素１０４ａ、１０４ｂ、１０４ｃ、１０４ｄ、１０４ｉ、１０４ｊ、１０４ｋ、１０４ｌの各々に直接結合される。したがって、図１に示した実施形態では、中間ステージのスイッチ・ファブリック構成要素１０４ｅ、１０４ｆ、１０４ｇ、１０４ｈは、外側ステージのスイッチ・ファブリック構成要素１０４ａ、１０４ｂ、１０４ｃ、１０４ｄ、１０４ｉ、１０４ｊ、１０４ｋ、１０４ｌの各々への直接的で独立した接続を有する。

【００２０】

また、スイッチ・ファブリックを横切る単一のパケットのために、単一の外側ステージのスイッチ／ファブリック構成要素を備える入力ステージ、中間ステージのスイッチ・ファブック構成要素を備える中間ステージ、および、スイッチ・ファブリック構成要素間の双方向の接続性のために残りの外側ステージのスイッチ・ファブリック構成要素を備える外側ステージを有するような、図１に示した特定の実施形態を考慮することが可能であることに留意されたい。

【００２１】

各外層のスイッチ・ファブリック構成要素１０４ａ、１０４ｂ、１０４ｃ、１０４ｄ、１０４ｉ、１０４ｊ、１０４ｋ、１０４ｌは、直接それに結合／接続された複数のポート・コントローラを有する。図１に示した実施形態では、外層のスイッチ・ファブリック構成要素の各々は、１つのコンジット（例えばコンジット１０８）が各ポート・コントローラに対応する、４つのコンジットによって結合された４つのポート・コントローラを有する。例えば、スイッチ・ファブリック１０４ａは、４つの別々の双方向コンジットによってそれに結合されたポート・コントローラ１０２ａ、１０２ｂ、１０２ｃ、１０２ｄを有する。この実施形態の一態様では、ポート・コントローラ１０２は、すべて、ギガビットのポート・コントローラを備えている。

【００２２】

スイッチ・ファブリックのクロス・ネットワーク・トポロジの特徴は、ノンブロッキング

・トポロジであるということである。すなわち、各ポート・コントローラから同じスイッチ・ファブリック構成要素に直接結合されていない他のすべてのポート・コントローラへ、中間ステージのスイッチ・ファブリック構成要素１０４ｅ、１０４ｆ、１０４ｇ、１０４ｈのそれぞれを通る経路が存在する。例えば、スイッチ・ファブリック構成要素１０４ａに接続されたポート・コントローラ１０２ａ、１０２ｂ、１０２ｃ、１０２ｄは、中間ステージのスイッチ・ファブック構成要素１０４ｅ、１０４ｆ、１０４ｇ、１０４ｈの各々を介して、他の外側ステージのスイッチ・ファブリック構成要素１０４ｂ、１０４ｃ、１０４ｄ、１０４ｉ、１０４ｊ、１０４ｋ、１０４ｌのすべてに結合されたポート・コントローラに接続されている。

【００２３】

図１に示した構成１００の好ましい実施形態では、全体で３２ギガビットのイーサネット・ポート・コントローラのためのスイッチ・ファブリック構成要素の各外側ステージに提供された１６ギガビットのイーサネット・ポート・コントローラが存在することが可能である。この好ましい実施形態では、１６ギガビットのイーサネット・ポート・コントローラの各セットからの最大スループットの要求は、１６×２（２倍）ギガビット・イーサネット・データ・ストリーム＝３２Ｇｂｐｓ＋［（３２Ｇｂｐｓ＋オーバーヘッド）＋（３２Ｇｂｐｓ＋オーバーヘッド）］の両方向における最大スループットの要求に対するオーバーヘッド＝６４Ｇｂｐｓ＋オーバーヘッドとなることが可能である。そのような好ましい実施形態では、３２ギガビットのイーサネット・ポート構成におけるスイッチ・ファブリック構成要素の各々は、１６Ｇｂｐｓにオーバーヘッドを加えたスループット能力を有する変更された８ポートのクロスバー・スイッチ・ファブリックとすることが可能である。スイッチ・ファブリックの中間ステージは、これらのデバイスの４つを備えるので、中間ステージのスループットは、全体で６４Ｇｂｐｓにオーバーヘッドを加えたものとすることが可能である。

【００２４】

図２は、本発明の実施形態による、他のマルチポート・スイッチ・アーキテクチャ２００の概略的なブロック図である。この実施形態では、マルチポート・スイッチ・アーキテクチャ２００は、単一のスイッチ・ファブリック構成要素２０４を備えるスイッチ・ファブリックを介して共に相互接続された４つのポート・コントローラ２０２ａ、２０２ｂ、２０２ｃ、２０２ｄを備える。そのような実施形態の一態様では、ポート・コントローラ２０２ａ、２０２ｂ、２０２ｃ、２０２ｄは、図１に示したポート・コントローラのアーキテクチャと同様のアーキテクチャを有する高速イーサネット・ポート・コントローラを備えることが可能である。

【００２５】

図３は、本発明の実施形態による、ギガビット・アップリンキング３００の概略的なブロック図である。アップリンキング３００は、双方向に一緒に結合された第１ポート・コントローラ３０２ａと第２ポート・コントローラ３０２ｂを備える。そのような実施形態の一態様では、ポート・コントローラの一方は、図２に示したマルチポート・スイッチ・アーキテクチャ２００で使用したタイプのような高速イーサネット・ポート・コントローラ３０２ａを備えることが可能であり、他のポート・コントローラ３０２ｂは、図１に示したマルチ・ポート・スイッチ・アーキテクチャ１００で使用したタイプのようなギガビット・ポート・コントローラを備えることが可能である。

【００２６】

（ポート・コントローラ）

各ポート・コントローラは、パイプライン処理と呼ばれる技術を実施する。パイプライン処理は、複数の命令が実行の際に部分的に重なる実施技術である。パイプラインは、一連のステージとして定義することができ、各ステージで何らかの作業が行われる。一般にポート・コントローラは、物理インターフェイス（ＰＨＹ）からイーサネット・パケットを受け取り、それらを処理し、スイッチ・ファブリックに送るための第１のパイプラインを有する。ポート・コントローラはまた、類似するが複雑さのより低い第２のパイプライン

も有し、これはスイッチ・ファブリックからイーサネット・パケットを受け取り、それらを物理インターフェイス（ＰＨＹ）に送るためのものである。

【００２７】

一般に、ポート・コントローラの好ましいバージョンは少なくとも２つある。第１の好ましいバージョンは、単一（二重）ギガビット・イーサネット・ラインに対してワイヤ速度のパケット処理およびその他のポート・コントローラ機能を提供することができる。第２の好ましいバージョンは、８本の高速イーサネット（またはイーサネット）・ラインに対して同じ機能を提供することができる。これらのデバイスは、スイッチ・ファブリックによって構成して、３２本のギガビット・イーサネット・ラインまで、または２５６本の高速イーサネット・ラインまでのワイヤ速度の第２層／第３層スイッチングを実現することができる。

【００２８】

図４は、本発明の一実施形態によるポート・コントローラ４００のアーキテクチャの概略ブロック図である。ポート・コントローラ４００は、メディア・アクセス制御（ＭＡＣ）ユニット４０２、イングレス（ｉｎｇｒｅｓｓ）・ダイレクト・メモリ・アクセス（イングレスＤＭＡまたはＩＤＭＡ）ユニット４０４、ヘッダ・キュー４０６、パケット・プロセッサ・ユニット４０８、ダイナミック・ランダム・アクセス・メモリ（ＤＲＡＭ）コントローラ４１０、出力向けスケジューラ４１２、インターフェース送信器４１４、フリー・セル・マネージャ４１６、同期スタティック・ランダム・アクセス・メモリ（ＳＳＲＡＭ）コントローラ４１８、インタフェース受信器４２０、イグレス（ｅｇｒｅｓｓ）・コントローラ４２２、イグレス・バッファ４２４、およびネットワーク・インターフェイス４２６を備える。ＤＲＡＭコントローラ４１０は、外部パケット・バッファ／メモリと呼ばれる第１メモリに結合されるが、このメモリは、同期ＤＲＡＭ（ＳＤＲＡＭ）など何らかの種類のダイナミック・ランダム・アクセス・メモリでよい。ＤＲＡＭコントローラ４１０は、第１メモリ／外部パケット・バッファを制御し、外部パケット・バッファに対してデータの読取りおよび書込みを行う。ＳＳＲＡＭコントローラ４１８は、ポート・コントローラの第２メモリに結合されるが、このメモリは、何らかの種類の同期スタティック・ランダム・アクセス・メモリ（ＳＳＲＡＭ）であることが好ましい。

【００２９】

ＭＡＣインターフェイス４０２は、データ・パケットを送受信するためにイーサネット物理層インターフェイス（ＰＨＹ）に接続可能である。一実施形態では、ＭＡＣインターフェイス４０２は、ＧＭＩＩインターフェイスを備えたギガビットＭＡＣを含むことができる。ＭＡＣインターフェイス４０２は、イングレスＤＭＡ４０４に結合されて、ＭＡＣインターフェイスによって受信された入力データ・パケットをイングレスＤＭＡ４０４に送る。

【００３０】

イングレスＤＭＡ４０４は、ヘッダ・キュー４０６、ＤＲＡＭコントローラ４１０、およびフリー・セル・マネージャ４１６に結合されている。入力データ・パケットをＭＡＣインターフェイス４０２から受け取ると、イングレスＤＭＡ４０４は、パケットのボディ部分からパケットのヘッダ部分を分離する。次いで、パケットのヘッダ部分は、イングレスＤＭＡ４０４からヘッダ・キュー４０６に転送され、パケットのボディ部分は、外部パケット・バッファにリンク・リストで記憶されるようにＤＲＡＭコントローラ４１０に転送される。イングレスＤＭＡ４０４はまた、パケット・プロセッサ４０８によって処理されているヘッダ部分に対して、メモリ中の予約された空間へのリンクを定義することも担い、このリンクは、ヘッダ部分が処理された後でイングレスＤＭＡが処理済みヘッダ部分をメモリに記憶するのに使用する。イングレスＤＭＡはまた、パケットの記憶が完了したときに出力向けスケジューラ４１２に通知することも担う。

【００３１】

ヘッダ・キュー４０６は、パケット・プロセッサ・ユニット４０８に結合されている。好ましい一実施形態では、ヘッダ・キュー４０６は、パケット・プロセッサ・ユニット４０

８によって先入れ先出し（ＦＩＦＯ）処理されるように、イングレスＤＭＡ４０４から受け取ったヘッダを記憶する。

【００３２】

パケット・プロセッサ・ユニット４０８は、パケットを分析および処理する働きをする。一実施形態では、パケット・プロセッサは複数のパケット・プロセッサを含む。図４に示す実施形態では、パケット・プロセッサ・ユニットは２つのパケット・プロセッサP_0およびP_1、４０８ａおよび４０８ｂを含み、これらは、ヘッダ・キュー４０６中のヘッダを交互に処理する。

【００３３】

パケット・プロセッサ４０８は、入力パケットのヘッダ部分をどのように処理するか決定する。パケット・プロセッサ４０８によるヘッダ部分の処理は、ルックアップ・アドレス・テーブルを利用することによってパケットの宛先を決定し、パケットのヘッダ部分に内部タグ（宛先ベクトルを含む）を付け、次いで処理の完了をヘッダ・キュー４０６に通知することを含む。アドレス・ルックアップは、外部ＳＳＲＡＭに記憶された外部アドレス・テーブルにアクセスする（ＳＳＲＡＭコントローラ４１８を介して）ことによって行う（このＳＳＲＡＭは、他の必要情報を記憶するのにも使用される）。

【００３４】

出力向けスケジューラ４１２は、パケットが処理されたときにイングレスＤＭＡから通知を受け取る。また、出力向けスケジューラ４１２は、ＤＲＡＭコントローラ４１０を介して外部パケット・バッファから処理済みパケットをリトリーブできるように、ＤＲＡＭコントローラ４１０に結合されている。処理済みパケットをパケット・バッファからリトリーブすると、出力向けスケジューラ４１２は、ポート・コントローラ４００の外に送るために、リトリーブしたパケットをインターフェース送信器４１４（または定義された優先順位に応じてネットワーク・インターフェイス４２６）に転送する。

【００３５】

ＤＲＡＭコントローラ４１０は、イングレスＤＭＡ４０４から外部パケット・バッファにデータを書き込むための書込みコマンドを実行する。また、ＤＲＡＭコントローラ４１０は、出力向けスケジューラ４１２から読取りコマンドを受け取ると、パケット・バッファから出力向けスケジューラ４１２へのデータの読取りおよび転送も実行する。

【００３６】

インターフェース送信器４１４は、出力向けスケジューラ４１２に結合されている。インターフェース送信器は、別のポート・コントローラまたはスイッチ・ファブリック・コンポーネントへの接続を可能にする外部入出力を有する。出力向けスケジューラ４１２は、処理済みパケットをインターフェース送信器４１４から外部入出力を介してスイッチ・ファブリック・コンポーネントまたは別のポート・コントローラに送ることができるようにするために、処理済みパケットをインターフェース送信器４１４に送る。

【００３７】

フリー・セル・マネージャ４１６は、イングレスＤＭＡ４０４および出力向けスケジューラ４１２に結合されており、それによりＤＲＡＭコントローラ４１０と通信することができる。フリー・セル・マネージャ４１６は、外部パケット・バッファ空間を管理し、データ記憶に利用可能なメモリの空間についてイングレスＤＭＡ４０４に通知し、データ（すなわち処理済みパケット）が出力向けスケジューラ４１２によってリトリーブされた後でメモリ空間を開放する。

【００３８】

インタフェース受信器４２０は、スイッチ・ファブリック・コンポーネントまたは別のポート・コントローラに接続可能な外部入出力を有し、スイッチ・ファブリック・コンポーネントまたは他のポート・コントローラからパケットを受け取ることができる。イグレス・コントローラ４２２は、インタフェース受信器４２０に結合されている。イグレス・コントローラ４２２には、イグレス・バッファ４２４およびＭＡＣインターフェイス４０２が結合されている。イグレス・コントローラ４２２は、インタフェース受信器から受け取

った送信パケットをイグレス・バッファ４２４に一時的に記憶し、次いでその後、ＰＨＹに送られるように、記憶したパケットをイグレス・バッファ４２４からＭＡＣインターフェイス４０２に転送する。

【００３９】

ネットワーク・インターフェイス・ユニット４２６は、外部入出力または内部チャネルを介してホスト・プロセッサとデータを交換する。ネットワーク・インターフェイス・ユニット４２６は、ホスト・プロセッサによって作成された制御パケットを分析し、内部レジスタを読み取り、ルックアップ・テーブルの内容を変更する。好ましい一実施形態では、ポート・コントローラのあらゆる内部ユニットをホスト・プロセッサによってネットワーク・インターフェイス・ユニット４２６を介して制御できるように、あらゆる内部ユニットはネットワーク・インターフェイス（ＮＰＩ）バスに接続される。

【００４０】

ＳＳＲＡＭコントローラ４１８は、第２メモリ（ＳＳＲＡＭ）に結合され、ＳＳＲＡＭへのアクセスを制御する。ネットワーク・インターフェイス・ユニット４２６とパケット処理ユニット４０８がＳＳＲＡＭコントローラに結合され、したがってＳＳＲＡＭにアクセスできる。ＳＳＲＡＭコントローラ４１８は、ネットワーク・インターフェイス・ユニットおよびパケット処理ユニットからの要求を調停する。

【００４１】

（パケット処理）

ポート・コントローラ４００において、パケットは、受信された順序に従ってパケット・メモリすなわちパケット・バッファに保存される。しかし、本発明の好ましい一実施形態では、入力パケットのヘッダ・フィールドとボディ・フィールドとの保存順序を変更することのできる、よりフレキシブルなパケット・ソーティング構成を利用することができる。

【００４２】

通常、イーサネット・ヘッダおよびＴＣＰ／ＩＰヘッダを有するパケットを処理するとき、従来型のスイッチではイーサネット・フィールドまたはＩＰフィールドだけを使用して、スイッチング方向を決定している。対照的に、本明細書で述べるスイッチ・ファブリック・システムの実施形態では、イーサネットとＩＰフィールドの両方、ならびにＴＣＰフィールドを使用してスイッチング方向を決定することができる。しかし、このような処理能力を用いると、このような順次パケット保存順序ルーチンは、ヘッダの処理に十分な時間および／または効率を提供することが不可能（または非実際的）であることにより、パケット処理中に獲得できる情報の量を制限する恐れがある。ヘッダ・フィールドとボディ・フィールドの保存順序を変更することにより、処理中により多くの情報を獲得することができ、それにより、ボディ・フィールドの保存中にヘッダ・フィールドを処理するための追加のタイミング余裕が与えられ、効率を最大限にしてメモリ・オーバヘッドを縮小することができる。

【００４３】

図５は、本発明の一実施形態による、スイッチ・ファブリックを介して送信するためのデータグラムを作成するプロセス５００のフローチャートである。まずオペレーション５０２で、ヘッダ部分およびボディ部分を有するデータグラムを受信する。オペレーション５０４および５０６で、ヘッダ部分をボディ部分から分離し、次いで処理のために転送する。オペレーション５０８で、分離したボディ部分をメモリに記憶する。オペレーション５１０および５１２で、ヘッダ部分を処理後に受け取り、メモリに記憶する。

【００４４】

本発明の一実施形態では、ボディ部分をメモリに記憶する前に、ヘッダ部分を記憶するためにメモリ中の空間を予約することができ、予約済み空間を突き止めるための、予約済み空間へのリンクも生成される。このような実施形態におけるオプションとして、処理済みヘッダを記憶することはさらに、リンクを利用して予約済み空間を突き止め、次いで処理済みヘッダを予約済み空間に記憶することを含むこともできる。

【００４５】
本発明の一態様では、ボディ部分を複数のセルに分割することができる。このような態様では、ボディ部分をメモリに記憶することはさらに、各セルを順次メモリに記憶し、その順次記憶したセルの間のリンクを生成することを含む。本発明の別の態様では、処理済みヘッダ部分は、パケットのルーティング先となるスイッチ・ファブリック・システム中の宛先を示す宛先ベクトルを含むことができる。このような態様の好ましい一実施形態では、宛先ベクトルは３２ビットのデータを含むことができる。
【００４６】
本発明の別の実施形態では、後でスイッチ・ファブリックを介して送信するために、データグラムを出力優先順位キューに割り当てることができる。
【００４７】
図６は、本発明の一実施形態による例示的なパケット６００の概略図である。パケットは、パケット交換ネットワーク上の発信源と宛先との間でルーティングされるデータの単位として定義することができる。パケット６００は、ヘッダ部分またはヘッダ・フィールド（「ヘッダ」）６０２と、ボディ部分またはボディ・フィールド（「ボディ」）６０４を含む。ヘッダ部分６０２は、スイッチングに使用される制御情報を含み、ボディ６０４は一般に、パケットの残りを含む。
【００４８】
図６に示す例示的なパケット中では、ヘッダ部分６０２は処理済みであり、パケットの宛先を示す宛先ベクトル６０６と、パケットのソースを示すソース・ベクトル６０８と、パケットの長さを示す長さベクトル６１０（これらは総称して「内部タグ」または「内部タグ・フィールド」とも呼ばれる）を含む。ヘッダ部分６０２は、パケットのＭＡＣ／ＴＣＰ／ＩＰおよび何らかのアプリケーション・データも含むことができる。
【００４９】
図７は、本発明の一実施形態による、パケット６００を処理するプロセス７００のフローチャートである。オペレーション７０２で、入力パケットを受信する。入力パケットは、ヘッダ部分とボディ部分を含む。オペレーション７０４で、パケットのヘッダ部分をボディ部分から分離する。次いでオペレーション７０６で、分離したボディ部分をメモリに記憶する。オペレーション２３０８で、好ましくはオペレーション２３０６と同時に、ヘッダ部分を処理する。オペレーション２３１０で、後続の（または第２の）パケットを受信し、この第２のパケットに対してプロセス７００を繰り返す。第１の入力パケットのヘッダ部分を処理すると、次いでオペレーション７１２で、処理済みヘッダ部分をメモリに記憶する。
【００５０】
図８は、本発明の一実施形態による、図７で述べたパケット処理プロセス７００をポート・コントローラ４００中で示した第１の概略流れ図である。入力パケット６００が、ＭＡＣインターフェイス４０２を介してポート・コントローラの中に受信され、ＭＡＣインターフェイス４０２は、パケットをイングレスＤＭＡ４０４にルーティングする（矢印８０２参照）。イングレスＤＭＡ４０４中で、パケットのヘッダ部分６０２がパケットのボディ６０４から分離される（オペレーション７０４参照）。次いでイングレスＤＭＡ４０４は、ヘッダ部分６０２を処理のためにヘッダ・キュー４０６に転送し（矢印８０６参照）、ボディ部分６０４を、ＤＭＡコントローラ４１０を介してメモリ（「パケット・メモリ」）８０４に記憶する（矢印８０８およびオペレーション７０６参照）。
【００５１】
ヘッダ部分６０２を処理するために（オペレーション７０８参照）、ヘッダ・キュー４０６は、パケット・プロセッサ・ユニット４０８中の利用可能なパケット・プロセッサにヘッダ部分６０２を転送する（矢印８１０参照）。ヘッダ部分がパケット・プロセッサ・ユニット４０８によって処理された後、処理済みヘッダ部分はヘッダ・キュー４０６に戻される（矢印８１２参照）。
【００５２】

図９は、本発明の一実施形態による、図７で述べたパケット処理プロセス７００をポート・コントローラ４００中で示した第２の概略流れ図である。処理済みヘッダ部分は、ヘッダ・キュー４０６に戻った後、続いてイングレスＤＭＡ４０４およびＤＲＡＭコントローラ４１０を介してパケット・メモリ８０４に記憶される（矢印９０２参照）。

【００５３】

イングレスＤＭＡ４０４は、処理済みヘッダ部分を記憶すると、パケットの送信準備ができたことを出力向けスケジューラ４１２に通知し、パケットをパケット・メモリ８０４からリトリーブするための情報を出力向けスケジューラに提供する（矢印９０４参照）。出力向けスケジューラ４１２は、パケットを送信する準備ができたとき、パケットをパケット・メモリ８０４からリトリーブして、インターフェース送信器４１４に転送する。インターフェース送信器４１４は、ポート・コントローラ４００からパケットを出力する（矢印９０６参照）。

【００５４】

図１０は、本発明の一実施形態による、ポート・コントローラに入力される複数の入力パケットの処理および記憶の一例を示す概略図である。この図では、第１の入力パケット１００２はポート・コントローラのイングレスＤＭＡ４０４によって受信され、その後に第２のパケット１００４、第３のパケット１００６、第４のパケット１００８が続く。上記の通り、各パケットはヘッダ部分（１００２ａ、１００４ａ、１００６ａ、１００８ａ）とボディ部分（１００２ｂ、１００４ｂ、１００６ｂ、１００８ｂ）から構成される。各パケットが受信される際、イングレスＤＭＡ４０４によってボディ部分からヘッダ部分が分離される。次いで分離されたボディ部分はパケット・メモリ８０４に記憶され、一方、ヘッダ部分はパケット・プロセッサ・ユニット４０８による処理のために送られる（注意：簡約化のため、図１０ではヘッダ・キュー４０６は省略した）。ヘッダ部分のそれぞれの処理が完了すると、そのヘッダ部分のそれぞれはイングレスＤＭＡ４０４を介してパケット・メモリ８０４に記憶される（簡約化のため、図１０ではＤＲＡＭコントローラ４１０も省略したことに注意されたい）。

【００５５】

図１０に示すように、ボディ部分と処理済みヘッダ部分は、イングレスＤＭＡによって受信された順番にパケット・メモリに記憶される。例えば、第１のパケット１００２のヘッダ部分とボディ部分を受信し分離する際、第１のボディ部分１００２ｂはメモリ８０４に記憶され、一方、第１のヘッダ部分１００２ａはパケット・プロセッサ・ユニット４０８による処理のために転送される。第１のヘッダ部分１００２ａが処理されている間、イングレスＤＭＡ４０４は第２のパケット１００４を受信し、第２のパケットのヘッダ部分とボディ部分とを分離する。次いで第２のボディ部分１００４ｂはメモリに記憶され、第２のヘッダ部分１００４ａは処理のために送られる。この時点で、第１のヘッダ部分１００２ａの処理は完了しており、処理済みの第１のヘッダ部分は処理から転送されてメモリ８０４に記憶される。

【００５６】

次に、第２のヘッダ部分１００４ａが処理されている間、イングレスＤＭＡは第３のパケット１００６を受信し、第３のパケットのヘッダ部分とボディ部分とを分離する。次いで第３のボディ部分１００６ｂはメモリに記憶され、一方、第３のヘッダ部分１００６ａは処理のために送られる。この時点までに、第２のヘッダ部分１００４ａの処理は完了しており、その処理済みヘッダ部分は処理から転送されて、メモリに記憶される。第３のヘッダ部分１００６ａが処理されている間、イングレスＤＭＡは第４のパケット１００８を受信し、第４のパケットのヘッダ部分とボディ部分とを分離する。次いで第４のボディ部分１００８ｂはパケット・メモリに記憶され、一方、第４のヘッダ部分１００８ａは処理のために送られる。この時点までに、第３のヘッダ部分１００６ａの処理は完了しており、その処理済みの第３のヘッダ部分は処理から転送されて、メモリに記憶される。最後に、第４のヘッダ部分１００８ａの処理が完了すると、これもまた処理から転送されて、メモリに記憶される。

【００５７】
図１１は、本発明の一実施形態による、パケットのボディ部分からヘッダ部分を分離するプロセス７０４の流れ図である。オペレーション１１０２では、入力パケットは受信された際にセルに分割される。入力パケットの第１のセルにはヘッダ部分が含まれる。このヘッダ部分は、１つまたは複数のＭＡＣ／ＩＰ／ＴＣＰヘッダならびにアプリケーション・データを含むことができる。好ましい一実施形態では、各セルは１２８バイトのサイズを有する。オペレーション１１０４では、ヘッダ部分（すなわち、第１のセル）のためにメモリに空間が予約される。次いで、オペレーション１１０６でメモリ内の予約した空間を指すポインタが生成される。次いでこのヘッダ部分は、オペレーション１１０８の処理に送られる。
【００５８】
図１２は、本発明の一実施形態による、パケットのボディ部分をヘッダ部分からの分離後にメモリに記憶するプロセス７０６を示す流れ図である。オペレーション１２０２では、パケットのボディ部分の第１の受信されたセル（すなわち、入力パケットの第２のセル）がメモリに記憶される。オペレーション１２０４では、そのセルをヘッダ部分のために予約された空間（すなわち、パケットの第１のセル用に予約された空間）にリンクするためのポインタが生成される。入力パケットの後続の各入力セルも、記憶されているセルのそれぞれを総合的にリンクするために、新たに記憶されたセルから前回記憶されたセルへのポインタと共にメモリに記憶される（判断１２０６およびオペレーション１２０８および１２１０を参照のこと）。
【００５９】
図１３は、本発明の一実施形態による、メモリにパケットの処理済みヘッダ部分を記憶するプロセス７１２を示す流れ図である。オペレーション１３０２で、処理済みヘッダ部分が受信される。オペレーション１３０４では、処理済みヘッダ部分のためにメモリに予約された空間が、その予約済み空間を指すポインタを用いて突き止める（オペレーション１２０４を参照のこと）。次いで処理済みヘッダは、オペレーション１３０６で予約済みの空間に記憶される。
【００６０】
図１４は、本発明の一実施形態による、入力パケットのセルのメモリへの記憶を示す概略図である。入力パケット１４００は、イングレスＤＭＡ４０４によって受信された際に複数のセル（C_0 １４０２、C_1 １４０４、C_2 １４０６、C_3 １４０８、C_4 １４１０）に分割される（図１１のオペレーション１１０２を参照のこと）。図１４に示すように、第１のセルC_0 １４０２はパケットのヘッダ部分を含んでおり、一方、残りのセルC_1 １４０４、C_2 １４０６、C_3 １４０８、C_4 １４１０はパケットのボディ部分を含んでいる。
【００６１】
パケット・メモリ１４１２では、空間１４１４がそのヘッダを含んでいるので第１のセルC_0 １４０２のために予約されており、その予約済み空間を指すポインタが生成される（図１１のオペレーション１１０４および１１０６を参照のこと）。一実施形態では、予約済み空間のサイズは、処理中にそのヘッダ部分に追加される可能性のある任意の追加データに対して十分な空間を提供するために、予約済みの第１のセルC_0 １４０２のサイズよりも大きい（図７の７０８を参照のこと）。次いで予約済み空間１４１４を指すポインタ１４１６が生成され、その結果、現在処理されているヘッダ部分を含んでいる第１のセルC_0 １４０２を予約済み空間に引き続いて記憶することができる（図１１のオペレーション１１０６を参照のこと）。
【００６２】
次に、入力パケット１４００のボディ部分の第１のセグメントを含んでいる第２の受信されたセルC_1 １４０４が、パケット・メモリ１４１２に記憶され（図１２のオペレーション１２０２を参照のこと）、記憶されているセルC_1 １４０４を予約済み空間にリンクするためのポインタ１４１８が生成される（図１２のオペレーション１２０４を参照のこと

）。第３のセル$C_2$１４０６が受信されると、これもまたパケット・メモリ１４１２に記憶することができ、記憶されているセル$C_2$１４０６を前回記憶されたセル$C_1$１４０４にリンクするためのポインタ１４２０が生成される（図１２のオペレーション１２０６、１２０８、１２１０を参照のこと）。同じ要領で、第４と第５のセル$C_3$１４０８、$C_4$１４１０が受信される際に、これらはそれぞれパケット・メモリ１４１２に記憶され、ポインタ１４２２、１４２４を生成させて、セルC_3をC_2に、また、セルC_4をセルC_3にそれぞれリンクする。図１２に示すプロセスによれば、第５のセル$C_4$１４１０はパケット１４００の最終セルなので、第５のセルC_4はポインタを含んでいる（１４２６を参照のこと）。

【００６３】

図１５は、本発明の一実施形態による、ポート・コントローラからスイッチ・ファブリック・システムのスイッチ・ファブリックにパケットを転送するプロセス１５００の流れ図である。オペレーション１５０２で、パケットがポート・コントローラで既に処理済みであることを示す通知が受信される。この処理済みパケットはオペレーション１５０４および１５０６で出力優先順位キューに割り当てられ、リンクされる。その後、オペレーション１５０８でそのパケットはリンクを介して取り出され、オペレーション１５１０でポート・コントローラからスイッチ・ファブリックに転送される。

【００６４】

本発明の一態様では、パケットは、リンクされたリスト内の割り当てられた出力優先順位キューにリンクすることができる。

【００６５】

本発明の一実施形態では、後続の処理されるパケットを同じ出力優先順位キューに割り当て、リンクすることができる。任意選択で、ポインタは、そのパケットを後続の処理されるパケットにリンクすることができる。本発明の別の実施形態では、そのパケットはメモリに記憶し、また、メモリから取り出すことができる。さらなる任意選択として、パケットを複数のリンクされたセルでメモリに記憶することができる。

【００６６】

本発明の別の態様では、処理済みパケットは、スイッチ・ファブリックを介してアクセス可能なパケットに対する１つまたは複数の宛先を示す宛先ベクトルを含むことができる。

【００６７】

本発明は、ポート・コントローラのメモリにパケットを記憶するためのデータ構造を含むこともできる。このデータ構造は、複数のキューから構成することができ、各キューはそれ自体に関連付けられたヘッド／開始アドレスを有する。各パケットをそのキューのうちの１つに割り当てて、１つまたは複数のパケットをメモリに記憶することができる。キューごとに、開始アドレスはそのキューの第１のパケットへのリンクを有することができ、そのキュー内の各パケットは、そのキュー内の次の後続パケットへのリンクを有する。一態様では、各パケットは、複数のリンクされたセルでメモリに記憶することができる。好ましい一実施形態では、ポインタは、パケットのセルをそのパケットの後続セルにリンクし、パケットの最終セルはヌル・ポインタを有する。

【００６８】

図１６は、本発明の一実施形態による、イグレスに対してキューに入れられたパケットのリンクされたリストを生成するプロセス１６００の流れ図である。オペレーション１６０２では、イングレスＤＭＡ４０４は、出力向けスケジューラ４１２にパケットのヘッダ部分が既に処理済みであることを通知する。その後、イングレスＤＭＡは、オペレーション１６０４で、出力向けスケジューラから戻されてきた、パケットがどの出力キュー（ポート・コントローラからのイグレス用）に割り当てられているかに関する情報を受信する。オペレーション１６０６で、ポインタが生成され、パケット・メモリに現在記憶されているヘッダ部分を割り当てられた出力キューにリンクする。割り当てられたキューが、既にそれに割り当てられている別のパケットを有しない場合、ポインタは、処理されたヘッダ部分を割り当てられたキューの開始アドレスにリンクする。

【００６９】
一方、割り当てられたキューが、それに割り当てられた１つまたは複数の別のパケットを既に有している場合、ポインタは、処理済みヘッダ部分をそのキュー内のそれの直前に配置されているもう１つのパケットの記憶されたヘッダ部分にリンクする。オペレーション１６０２、１６０４および１６０６が、ポート・コントローラによって処理される各ヘッダ部分に対して反復される（オペレーション１６０８を参照のこと）。
【００７０】
図１７は、本発明の一実施形態により、パケット・メモリ８０４に記憶されているパケットをそれらの割り当てられている出力優先順位キューにリンクすることを示す概略図である。出力向けスケジューラ４１２は、例示の実施形態では４つのキューから成る、複数の出力優先順位キュー１７０２、１７０４、１７０６、１７０８を有する。好ましい一実施形態では、出力向けスケジューラとイングレスＤＭＡのどちらも各キューの開始アドレスを認識している。
【００７１】
各優先順位キューにはパケット・メモリに記憶されている複数のパケットが割り当てられており、特定キューに割り当てられている複数のパケットは、それらが割り当てられているキューの下に示される。各キューは先入れ先出し（ＦＩＦＯ）なので、記憶されている各パケットは、特定キュー内の１つのパケットから同じ割り当てられたキュー内の次に続くパケットの開始へのリンクを提供する次のパケット・ポインタ（「ＮＰＰ」）を有する。補足的な矢印は、次のセルが保存されているアドレスを各セルがどのように指すかを示している。例えば、図１７のキュー０　１７０２を参照すると、キュー０　１７０２に割り当てられたパケット・メモリに記憶されている第１のパケット１７１０は、キュー内の次の記憶されているパケット１７１４の開始を指すポインタ１７１２を有し、パケット１７１４は、そのキュー内の第３の記憶されているパケット１７１８の開始を指すポインタ１７１６を有する、という具合である。
【００７２】
図１８は、本発明の一実施形態による、図１７に示されるパケット・メモリに記憶される例示的パケット１７１０を示す概略図である。例示的パケット１７１０は、第１のセル１８０２がパケットの処理済みヘッダ部分を含み、残りのセル１８０４、１８０６、１８０８、１８１０がパケットのボディを含むように、複数のセルに分割される。セル１８０２、１８０４、１８０６、１８０８は、それぞれに、次に続くセルの開始を指す次のセル・ポインタ（「ＮＣＰ」）１８１２、１８１４、１８１６、１８１８を有する。各ＮＣＰは、パケットの最終セルの最後の１つを除いて、次のセルを指す。最後のＮＣＰは、いかなる値をも有することができるが、これは出力向けスケジューラによって無効であるとされる。何故ならば、出力向けスケジューラは、内部タグ情報からパケットのサイズを認識しているからである。
【００７３】
一実施形態では、イングレスＤＭＡ４０４は、リンク／ポインタの生成に応答可能である。好ましい一実施形態では、ポインタのサイズは１６ビット（２バイト）であってよい。
【００７４】
上記の通り、第１のセル１８０２は、パケットのサイズと、そのパケットがスイッチ・ファブリック・システムから出力される宛先ポートを示すことのできる内部タグを含む処理済みヘッダ部分を有する。また、第１のセル１８０２は、キュー０　１７０２内の次に記憶されているパケット１７１４を指す次のパケット・ポインタ１７１２を含む。
【００７５】
各キューは先入れ先出し（ＦＩＦＯ）であってよい。出力向けスケジューラ４１２が、そのＦＩＦＯキューのうちの１つに割り当てられたパケットをパケット・メモリから取り出す準備ができると、出力向けスケジューラは、パケットの次のセルを探すために各セルのＮＣＰを読み取ることができる。一実施形態では、最終セルのＮＣＰは、出力向けスケジューラが取り出しているパケットのサイズの知識に基づいてその出力向けスケジューラか

ら無視される場合がある。
【００７６】
図１９は、本発明の一実施形態による、ポート・コントローラ４００を介してスイッチ・システムから抜けようとしているパケットを処理するプロセスを示す概略的な流れ図である。このプロセスでは、パケットは、ポート・コントローラ４００によって受信される前に既に処理が完了している。抜けようとしているパケットがインタフェース受信器４２０によって最初にポート・コントローラ内に受信される。インタフェース受信器４２０は、そのパケットをイグレス・コントローラ４２２に送り、次いでイグレス・コントローラ４２２は、そのパケットをイグレス・バッファ４２４に転送することができる。イグレス・コントローラが、抜けようとしているパケットをポート・コントローラから送出する準備ができると、イグレス・コントローラ４２２は、イグレス・バッファ４２４からパケットを取り出し、そのパケットを、ポート・コントローラから送出するためにＭＡＣインターフェイス４０２に転送する（矢印１９０４を参照のこと）。
【００７７】
（スイッチ・ファブリック）
図２０は、本発明の一実施形態によるスイッチ・ファブリック構成要素２０００のアーキテクチャを示す概略的ブロック図である。スイッチ・ファブリック構成要素２０００には、インタフェース受信器２００２、スイッチ・マトリックス２００４、インターフェース送信器２００６、中央コントローラ２００８、フリー・セル・マネージャ２０１０、キュー・マネージャ２０１２、および外部インターフェイス２０１４が含まれる。
【００７８】
より明瞭かつ容易に理解できるように、図２０には１つのインタフェース受信器２００２と１つのインターフェース送信器２００６しか示さない。しかし、スイッチ・ファブリック構成要素が複数のインタフェース受信器と複数のインターフェース送信器を含むことができるということを理解されたい。例えば、図１に示されている８×８のスイッチ・ファブリック構成要素（例えば１０４ａ）では、各スイッチ・ファブリック構成要素は８つのインタフェース受信器と８つのインターフェース送信器とを有する。同様に、図２に示されているスイッチ・ファブリック構成要素２０４は、４つのインタフェース受信器と４つのインターフェース送信器とを含んでいる。
【００７９】
インタフェース受信器２００２は、パケットを受信するように働く。このインタフェース受信器は、別のスイッチ・ファブリック構成要素のインターフェース送信器、またはポート・コントローラのインターフェース送信器４０１４への１つの接続用外部Ｉ／Ｏを有する。このインタフェース受信器は、受信したパケットをインタフェース受信器２００２からスイッチ・マトリックス２００４に転送するために、スイッチ・マトリックス２００４に結合されている。インタフェース受信器２００２の機能は、ポート・コントローラのインタフェース受信器４２０の機能と類似である。インタフェース受信器２００２は、外部Ｉ／Ｏによってリンクを構成し、制御のために中央コントローラ２００８と対話する。
【００８０】
スイッチ・マトリックス２００４は、スイッチ・ファブリック構成要素内のパケットへのデータ・パスを提供する。スイッチ・マトリックスは、１つまたは複数のインタフェース受信器２００２と１つまたは複数のインターフェース送信器２００６とに結合されている。スイッチ・マトリックスは、インタフェース受信器からデータ（すなわち、パケット）を受信する。各入力パケットは、スイッチ・ファブリック構成要素からの送出のためにパケットをインターフェース送信器２００６に転送できるようになるまでは、スイッチ・マトリックスのメモリ内に記憶される。好ましい一実施形態では、スイッチ・マトリックスは６４Ｋバイトのメモリを有する。
【００８１】
上記の通り、インターフェース送信器２００６は、スイッチ・マトリックス２００４に結合されている。インターフェース送信器は、別のスイッチ・ファブリック構成要素のイン

タフェース受信器または外部Ｉ／Ｏを介したポート・コントローラのインタフェース受信器に接続可能である。インターフェース送信器２００６は、スイッチ・マトリックスからのパケットをスイッチ・ファブリック構成要素から送出する。各インターフェース送信器２００６は、別のスイッチ・ファブリック構成要素のインタフェース受信器またはポート・コントローラに外部的に接続され、外部Ｉ／Ｏによってリンクを構成することができる。各インターフェース送信器２００６は、制御のために中央コントローラ２００８と対話する。

【００８２】

中央コントローラ２００８は、スイッチ・ファブリック構成要素の全てのユニットに接続されており、それらを制御する。中央コントローラ２００８は、イングレス・ユニット２００８ａ、マトリックス・ユニット２００８ｂ、およびイグレス・ユニット２００８ｃの３つの構成要素を含むことができる。中央コントローラ２００８は、フリー・セル・マネージャ２０１０によって提供されるフリー・セル・アドレスに従って、スイッチング・マトリックスのメモリ内へ入力パケットを記憶すること管理する。中央コントローラ２００８は、また、パケットの宛先ベクトルから宛先ポートを読み取り、そのパケットをキュー・マネージャ２０１２によってキューに記憶する。さらに、中央コントローラは、キュー・マネージャ２０１２の情報に従ってパケットをスイッチ・マトリックスからインターフェース送信器２００６に転送するための信号を生成することもできる。

【００８３】

フリー・セル・マネージャ２０１０は、中央コントローラとキュー・マネージャ２０１２に結合されている。フリー・セル・マネージャ２０１０は、中央コントローラ２００８を介して提供された情報に基づいてスイッチング・マトリックスのメモリ内のフリー・セルを管理する。具体的には、フリー・セル・マネージャ２０１０は、スイッチ・マトリックス内の占有されたセルとフリー・セルに関する情報を表形式で維持し、この情報を中央コントローラ２００８とキュー・マネージャ２０１２に提供する。

【００８４】

キュー・マネージャ２０１２は、中央コントローラ２００８とフリー・セル・マネージャ２０１０とに結合されている。キュー・マネージャ２０１２は、入力パケットに関する情報を受信し、出力ポートに従ってＦＩＦＯでリンクされるリストとしてその情報を記憶する。キュー・マネージャ２０１２は、抜けようとしているパケットのキューも管理する。

【００８５】

外部インターフェイス２０１４は、中央コントローラとスイッチ・マトリックスに接続されている。外部インターフェイスの機能は、ポート・コントローラのネットワーク・インターフェイスと同様である。外部インターフェイスは、制御パケットを外部Ｉ／Ｏを介して内部チャネルと交換し、次いでスイッチ・ファブリック構成要素の内部レジスタを維持するためにこれらのパケットを分析する。

【００８６】

図２１は、本発明の一実施形態による、スイッチ・ファブリックを介してデータグラムを送信するプロセス２１００の流れ図である。受信したデータグラムから宛先ベクトルが抽出される（オペレーション２１０２および２１０４を参照のこと）。宛先ベクトルは、データグラムの１つまたは複数の宛先（すなわち、単一キャストまたはマルチキャスト）を示す。オペレーション２１０６で、宛先ベクトルに基づいてデータグラムに対してポート・ビットマップが生成される。オペレーション２１０８では、このデータグラムは引き続いてポート・ビットマップを利用して宛先に向けて送信される。

【００８７】

本発明の一態様では、宛先ベクトルは３２ビットのデータからなる。本発明の別の態様では、ポート・ビットマップは８ビットのデータでもよい。

【００８８】

本発明の一実施形態では、ポート・ビットマップを生成することは、宛先ベクトルからデータグラムの全ての宛先を獲得すること、および、次いでそのデータグラムが宛先に到達

するにはそのスイッチ・ファブリック構成要素のどのスイッチ・ポートを介してルーティングすることができるかを決定するためにスイッチング・テーブルを利用することを含む。次いでデータグラムがその対象とする宛先に到達するために、そこを介するようにルーティングすることができる全ての決定されたスイッチ・ポートを示すためにポート・ビットマップを生成する。

【００８９】

本発明の別の実施形態では、データグラムを転送することは、ポート・ビットマップに基づいてマスクを作成すること、および、データグラムの送信に際してマスクを用いてポート・ビットマップを変更することを含むことができる。このような一実施形態では、マスクを用いてポート・ビットマップを変更するステップは、データグラムがそこに向けて送信されている宛先に関連付けられたデータのポート・ビットマップをクリアすることも含む。

【００９０】

本発明のさらなる実施形態では、データグラムが宛先に到達するためにそこを介してデータグラムをルーティングすることができるスイッチ・ポートの可用性を判定することもできる。次いでデータグラムは使用可能と判定されたスイッチ・ポートを介して送信することができる。

【００９１】

図２２は、本発明の実施形態によるスイッチ・ファブリックを介してパケットをルーティングするためのスイッチング・テーブルの実施を示す８行８列のスイッチ・ファブリック構成要素を含むスイッチ・ファブリック２２００の概略ブロック図である。この実施形態の下で議論される基礎となっている原理は、本明細書によるいかなるスイッチ・ファブリックにおいても利用することが可能である。この議論のため、スイッチ・ファブリックおよびスイッチ・ファブリック構成要素は、すべて双方向であることに留意することが重要である。

【００９２】

図２２に示すとおり、スイッチ・ファブリック２２００は、ＳＦ１　２２０２ａ、ＳＦ２　２２０２ｂ、ＳＦ３　２２０２ｃ、ＳＦ４　２２０２ｄとして示される８行８列のスイッチ・ファブリック構成要素を含む。各スイッチ・ファブリック構成要素は、８つのポート（「スイッチ・ポート」）を有する。例えば、図２２では、ＳＦ１　２２０２ａの８つのスイッチ・ポートが、１／１、２／１、３／１、４／１、５／１、６／１、７／１、８／１（「１の１、１の２、１の３．．．」等）として示される。ＳＦ２、ＳＦ３、ＳＦ４の８つのスイッチ・ポートも同様に、図２２で示される。例えば、ＳＦ２　２２０２ｂのポートは、ｎ／２（ただし、ｎは、１から８までの整数）を使用して示され、ＳＦ３　２２０２ｃのポートは、ｎ／３を使用して示され、またＳＦ４　２２０２ｄのポートは、ｎ／４（ただし、ｎは、１から８までの整数）を使用して示される。

【００９３】

図２２の示されているスイッチ・ファブリックでは、各スイッチ・ファブリック構成要素が、２つの隣接するスイッチ・ファブリック構成要素に結合された２つのポートを有している。例えば、ＳＦ１　２２０２ａのポート５／１および６／１は、それぞれ、ＳＦ２　２２０２ｂのポート２／２および１／２に結合され、またＳＦ１　２２０２ａのポート７／１および８／１は、それぞれ、ＳＦ３　２２０２ｃの４／３および３／３に結合される。

【００９４】

図２２の参照を続けると、スイッチ・ファブリック２２００は、複数のデバイス・ポートも含む。この議論では、デバイス・ポートは、外部デバイス（例えば、ポート・コントローラなど）に結合することができ、スイッチ・ファブリック内の別のスイッチ・ファブリック構成要素に結合されていないスイッチ・ファブリック構成要素のポートと定義することができる。図２２に示す例としての実施形態では、スイッチ・ファブリックは、十六（１６）のデバイス・ポート（これらには、図２２で１ないし１６の番号が付けられている

）を有し、各スイッチ・ファブリック構成要素は、４つのデバイス・ポートを有している。例えば、ＳＦ１　２２０２ａは、デバイス・ポート１、２、３、４を有し、ＳＦ２　２２０２ｂは、デバイス・ポート５、６、７、８を有し、ＳＦ３　２２０２ｃは、デバイス・ポート９、１０、１１、１２を有し、またＳＦ４　２２０２ｄは、デバイス・ポート１３、１４、１５、１６を有する。

【００９５】

（スイッチング・テーブル）

本発明の実施形態では、スイッチング・テーブルを利用して、スイッチ・ファブリックの柔軟性を高めるのを促進することができる。スイッチング・テーブルを各スイッチ・ファブリック構成要素に関連付けて、スイッチ・ファブリック構成要素の各ポートを介してどのデバイスに到達できるかを記録することができる。言い換えれば、各スイッチング・テーブルは、特定のスイッチ・ファブリック構成要素のどのポートを使用して（すなわち、アクセス可能であって）当該の特定のスイッチ・ファブリック構成要素からパケットを伝送することができるかに関する情報を提供する。

【００９６】

図２３、２４、２５、２６は、本発明の実施形態による図２２に描いたスイッチ・ファブリック２２００に関する例としてのスイッチング・テーブルを示す概略図である。詳細には、図２３に示す例としてのスイッチング・テーブル２３００は、スイッチ・ファブリック構成要素ＳＦ１　２２０２ａに対するものであり、図２４に示す例としてのスイッチング・テーブル２４００は、スイッチ・ファブリック構成要素ＳＦ２　２２０２ｂに関するものであり、図２５に示す例としてのスイッチング・テーブル２５００は、スイッチ・ファブリック構成要素ＳＦ３　２２０２ｃに対するものであり、また図２６に示す例としてのスイッチング・テーブル２６００は、スイッチ・ファブリック構成要素ＳＦ４　２２０２ｄに対するものである。

【００９７】

各スイッチング・テーブルは、スイッチ・ファブリックのデバイス・ポートに対する複数の列２３０２、および特定のスイッチ・ファブリック構成要素のポートに対する複数の行２３０４を有する格子（例えば、複数のビットマップ・アレイ）として表すことができる。列と行は一緒に、スイッチング・テーブルのビットマップ・アレイを形成する。この格子の表現では、ビットマップ・アレイは、スイッチング・テーブル格子の列と行に形成されたボックスを含む。各格子の列と行によって形成されたボックス内には、１および０がある。ボックス内の「１」は、特定のデバイス・ポートが、特定の行のポートを介してアクセス可能であることを示し、一方、ボックス内の「０」は、特定のデバイス・ポートが、特定の行のポートを介してアクセス可能ではないことを示す。そのような表現を使用して、特定のポートに関連付けられたアレイに基づいてビットマップ、スイッチ・ファブリック構成要素のポートを生成することができる。

【００９８】

例えば、スイッチ・ファブリック構成要素ＳＦ１　２２０２ａに対するスイッチング・テーブル２３００を表す図２３を参照すると、ポート１／１に対して指定された一番上の行を見ることにより、ポート１／１に対するビットマップ・アレイを確認することができる。このアレイでは、デバイス・ポート１に対する列のボックス内に「１」が存在し、また残りの列のボックスのそれぞれの中に「０」が存在する。このことは、デバイス・ポート１だけが、ＳＦ１　２２０２ａのポート１／１を介してアクセス可能であることを示す（すなわち、デバイス・ポート１を通る伝送を介してその宛先がアクセス可能であるＳＦ１のパケットは、ポート１／１を介するようにルーティングしなければならない）。同様に、ポート５／１に対して指定された行（すなわち、アレイ）を参照すると、デバイス・ポート５、６、７、８、９、１０の列のボックスのそれぞれの中に「１」が存在し、またデバイス・ポート１、２、３、４、１１、１２、１３、１４、１５、１６の列のボックスのそれぞれの中に「０」が存在する。このことは、デバイス・ポート５、６、７、８、９、１０が、ＳＦ１　２２０２ａのポート５／１を介するＳＦ１内のパケットに関してアクセ

ス可能であることを示す。
【００９９】
図２２および２３－２６に示す例としての実施形態では、スイッチング・テーブルが、スイッチ・ファブリック構成要素のどのポートからのアクセスも、６つのデバイス・ポート未満に制限するように生成されていることに留意されたい。これは、図２２のスイッチ・ファブリック２２００を介してパケットの伝達の理解を簡単にするのを助けるために行っている。図２２のスイッチ・ファブリック２２００の場合、スイッチ・ファブリック構成要素のポートは、最高で１２のポートに対するアクセスが可能である（例えば、ポート５／１を介して伝送されるＳＦ１内のパケットは、スイッチング・テーブルがそのように生成された場合、デバイス・ポート５ないし１６に対するアクセスが可能である）ことを理解されたい。
【０１００】
また、図２３－２６に示すスイッチング・テーブルのそれぞれは、最高で３２デバイス・ポートに対する列が含まれる好ましい実施形態に従って生成されていることにも留意されたい。スイッチング・テーブルのこの好ましい実施形態は、３２のポート・コントローラが接続された３２のデバイス・ポートを備えたスイッチ・ファブリックを有する図１に示すスイッチ・ファブリックの好ましい実施形態におけるスイッチング・テーブルの使用を助けることである。図２２に示す例としてのスイッチ・ファブリック２２００は、１６だけのデバイス・ポートを有する（例えば、１６のポート・コントローラをそこに接続して）ので、列１７ないし３２内のボックスはすべて、４つのスイッチング・テーブル２３－２６のそれぞれの中に「０」を含む。
【０１０１】
図２７は、本発明の実施形態によるスイッチ・ファブリック構成要素を介するパケットを処理するための処理２７００に関する流れ図である。オペレーション２７０２で、パケットがスイッチ・ファブリック構成要素によって受信される。オペレーション２７０４で、宛先ベクトルが、受信されたパケットから抽出される。次に、オペレーション２７０６で、スイッチ・ファブリック構成要素のスイッチング・テーブルに基づき、抽出された宛先ベクトルからポート・ビットマップが生成される。オペレーション２７０８で、スイッチ・ファブリック構成要素の論理が、ポート・ビットマップ内に示されたスイッチ・ファブリック構成要素のポートのどれかがアイドルであるかどうかを検査して判定する。ポート・ビットマップによって示されるポートのどれかが、アイドルであると判明した場合には、オペレーション２７１０で、ポート・ビットマップに基づき、識別されたアイドルなポートのためのマスクが作成される。オペレーション２７１２で、次に、スイッチ・ファブリックの識別されたアイドルなポートを介して発信パケットが伝送され、パケットがそれを介して伝送されたばかりのポートに関連する情報が、ポート・ビットマップからクリアされる。クリア・プロセスは、識別されたポートにおける発信パケットの宛先ベクトルに対しても実行されて、発信パケットの宛先ベクトルが変更される。判定２７１４で、ポート・ビットマップ内に別の宛先が含まれると判定された場合には、追加の宛先のそれぞれに対してオペレーション２７１０および２７１２が反復される。
【０１０２】
パケットがスイッチ・ファブリック構成要素に入った際、パケットの宛先ベクトルおよびスイッチ・ファブリック構成要素のスイッチング・テーブルを使用して、スイッチ・ファブリック構成要素のどのポートを介してパケットが出るべきかが判定される。一実施形態では、判定結果をリンクされたリストとして記憶することが可能である。一態様では、判定結果は、ポート・ビットマップを含むことが可能である。別の態様では、判定結果は、宛先ベクトルを含むことが可能である。
【０１０３】
図２８は、本発明の実施形態による、判定結果がポート・ビットマップを含むスイッチ・ファブリック構成要素を介してパケットを処理するためのプロセス２８００を示す概略図である。スイッチ・ファブリック構成要素内へのイングレスに先立ち、ポート・コントロ

ーラ２８０４におけるパケットの処理中にパケット２８０２内にタグが挿入される。タグは、パケットが宛先とする１つまたは複数のデバイス・ポートを示す宛先ベクトル２８０６を含む。好ましい実施形態では、宛先ベクトル２８０６は、３２ビットの長さ（すなわち、「３２ビットの宛先ベクトル」）を有することが可能である。ポート・コントローラは、スイッチ・ファブリック構成要素２８０８のデバイス・ポートに結合されて、挿入されたタグを有するパケットをスイッチ・ファブリック構成要素に転送させる。

【０１０４】

スイッチ・ファブリック構成要素２８０８内へのパケット（挿入されたタグを有する）２８０２のイングレスの後、タグが抽出されて宛先ベクトル２８０６が獲得される。次に、宛先ベクトル２８０６およびスイッチ・ファブリック構成要素２８０８のスイッチング・テーブル２８１２に基づき、ポート・ビットマップ２８１０が生成される。好ましい実施形態では、ポート・ビットマップは、８ビットの長さ（すなわち、「８ビットのポート・ビットマップ」）を有する。次に、パケットが、スイッチ・ファブリック構成要素のスイッチ・マトリックスのメモリ内のセルの中に記憶される。生成されたポート・ビットマップ２８１０を使用して、リンクされたリスト２８１４の中でパケットが記憶されたメモリのセルがリンクされる。また、ポート・ビットマップ２８１０も、スイッチ・ファブリック構成要素２８０８内のリンクされたリスト２８１４の中に記憶される。

【０１０５】

適切な出力ポートが、利用可能であると判定されたとき、パケットは、出力ポートを介してスイッチ・ファブリック構成要素を出ることができる。適切な出力ポートが利用可能であると判定された後、適切なセルに関する情報（ポート・ビットマップを含む）が、リンクされたリスト２８１４から読み取られてポート・ビットマップ２８１０が獲得される。次に、スイッチング・テーブル２８１２を使用して、イグレスのポートのためのマスク２８１６が生成される。好ましい実施形態では、マスク２８１６は、８ビットの長さ（すなわち、「８ビットのビット・マスク」）を有する。次に、マスク２８１６を使用してポート・ビットマップ２８１０が変更され、出力ポートを介するパケットの送信に関連する情報が、ポート・ビットマップから除去されるようにする。次に、変更されたポート・ビットマップ１０１８が、リンクされたリストの中に記憶される。変更されたポート・ビットマップ２８１８が、すべて「０」のビット（すなわち、「１」が全くない）を含む場合には、スイッチ・マトリックス内の対応するメモリ・セルが解放されている。そのような実施形態の好ましい態様では、リンクされたリストは、ポート・ビットマップを記憶するための８ビット／ワードで２５６ワードのメモリでよい。そのような実施形態の別の好ましい態様では、スイッチング・テーブルは、マスクを生成する論理（すなわち、回路）を含むことが可能である。この実施形態のさらに別の好ましい態様では、スイッチング・テーブルの行は、互いに排他的であることが可能である。

【０１０６】

図２８に示すプロセス２８００に対する好ましいオプションとして、パケットがスイッチ・ファブリック構成要素の発信ポートを出る際にパケットから宛先ベクトルをクリアすることができる。そのようなオプションを実施することの利点は、図２２に描いたスイッチ・ファブリック２２００を使用する以下の実施形態で見ることができる。この実施形態では、スイッチ・ファブリック２２００に入るパケットが、ＳＦ１　２２０２ａ内のデバイス・ポート２およびＳＦ２　２２０２ｂ内のデバイス・ポート７という２つの宛先を有するものと想定する。図３０は、本実施形態に従ってデバイス・ポート１からＳＦ１　２２０２に入り、ＳＦ１　２２０２ａ内のデバイス・ポート２およびＳＦ２　２２０２ｂ内のデバイス・ポート７という宛先を有するパケットに関する宛先ベクトル３０００（特に、３２ビットの宛先ベクトルとする）を示す概略図である。

【０１０７】

（ポート・ビットマップ）

図２９は、本発明の実施形態によるスイッチ・ファブリック構成要素内のポート・ビットマップ（詳細には、８ビットのポート・ビットマップ）の例としての利用を示す概略図で

ある。この例は、図２２に示す例としてのスイッチ・ファブリック２２００に基づき、スイッチ・ファブリック構成要素ＳＦ１　２２０２ａ内で生じる。この例では、パケットが、デバイス・ポート１に結合されたポート・コントローラを介してＳＦ１によって受信され、（１）デバイス・ポート２（ＳＦ１　２２０２ａにある）および（２）デバイス・ポート９（ＳＦ４　２２０２ｄにある）という２つの宛先を有する。

【０１０８】

ＳＦ１　２２０２ａ内で、パケットから宛先ベクトルが抽出される。図２９に示すとおり、抽出された宛先ベクトル２９０２は、パケットが目指す宛先ポート（デバイス・ポート２および９）を示すビット２　２９０４内およびビット９　２９０６内に「１」を有する。抽出された宛先ベクトル２９０２の情報に基づき、スイッチ・ファブリック構成要素は、スイッチング・テーブルを使用して、パケットが目指す宛先に到着するのに、そのスイッチ・ポートのどれを介するようにパケットをルーティングすることができるかを判定する。図２３に示すスイッチ・ファブリック構成要素ＳＦ１に対するスイッチング・テーブル２３００を参照すると、デバイス・ポート２は、ＳＦ１のスイッチ・ポート２／１を介してアクセスすることができ（行２／１参照）、またデバイス・ポート９は、ＳＦ１のスイッチ・ポート５／１および６／１を介してアクセスできる（行５／１および６／１参照）ことが分かる。

【０１０９】

ポート・ビットマップは、スイッチ・ファブリック構成要素によって受信されたパケットの抽出された宛先ベクトルからの情報およびスイッチ・ファブリック構成要素のスイッチング・テーブルの情報に基づいてスイッチ・ファブリック構成要素によって生成されるビットマップである。好ましくは、ポート・ビットマップは、スイッチ・ファブリック構成要素の中央コントローラによって生成される。ポート・ビットマップは、パケットが目指す宛先（すなわち、デバイス・ポート）に到達するため、スイッチ・ファブリック構成要素のスイッチ・ポートのどれを介するようにルーティングされる必要があるかを示す。ポート・ビットマップは、複数のビットを含み、各ビットは、関連スイッチ・ファブリック構成要素の対応するスイッチ・ポートに関連付けられている。例えば、４つのスイッチ・ポートを有するスイッチ・ファブリック構成要素では、そのスイッチ・ファブリックに対するポート・ビットマップは、（少なくとも）４ビットを有し、各ビットは、スイッチ・ファブリック構成要素の４つのスイッチ・ポートのどれかに関連付けられている。別の例として、図２２に示す実施形態では、スイッチ・ファブリック構成要素のそれぞれが、８つのスイッチ・ポートを有する。したがって、各スイッチ・ファブリック構成要素に対して生成されるポート・ビットマップは、少なくとも８ビットを有し、８ビットのそれぞれは、それぞれのスイッチ・ファブリック構成要素の対応するスイッチ・ポートに関連付けられている。好ましい実施形態では、ポート・ビットマップは、八（８）ビットを含む。

【０１１０】

ポート・ビットマップのビットは、それぞれ、「１」または「０」を含む。一実施形態では、ポート・ビットマップのあるビットが「１」を含む場合には、そのことは、その特定のビットに関連するスイッチ・ポートが、スイッチ・ファブリック構成要素からパケットを出力するためのポートとして働くことができるのを示す。反対に、ポート・ビットマップのあるビットが「０」を含む場合には、そのことは、その特定のビットに関連するスイッチ・ポートが、スイッチ・ファブリック構成要素からパケットを出力するためのポートとして働くことができないということを示す。

【０１１１】

再び図２９を参照すると、受信されたパケットのスイッチ・ファブリック要素ＳＦ１により、抽出された宛先ベクトル２９０２およびスイッチ・ファブリック要素ＳＦ１のスイッチング・テーブル２３００に基づいて生成することのできるポート・ビットマップ２９０８が示されている。このポート・ビットマップ２９０８は、それぞれ、ＳＦ１のスイッチ・ポート２／１、５／１、および６／１に関連するビット２、５、および６　２９１０、２９１２、２９１４に「１」を含む。

【０１１２】
この例では、スイッチ・ポート５／１または６／１がスイッチ・ポート２／１よりも前に利用可能になる（すなわち、アイドル状態になる）と判定されるものと仮定する。この場合、この例ではまず、パケットがスイッチ・ポート５／１または６／１を介して装置ポートの所期の宛先に向かって（ＳＦ２を介し、次いでＳＦ４を介して）出力される。図２７に示されたフローチャートに示されているように、引き続き図２９を参照するとわかるように、第１のマスク２９１６は、ポート・ビットマップおよび出力スイッチ・ポートに基づいて生成される。この例では、この第１のマスクは、パケットがスイッチ・ポート５／１または６／１を介して出力されるときにポート・ビットマップにおける処理されるビットを示すビット５および６　２９１８、２９２０に「１」を含む。
【０１１３】
引き続きこの例で説明する。次に、第１のマスク２９１６を使用して、処理されるビットがポート・ビットマップからクリアされる。一般に、クリアは、ポート・ビットマップ内の、「１」を含むビットがクリアされ、この場合、マスクの対応するビットも「１」を含む。言い換えれば、クリアは、ポート・ビットマップ内の、「１」を含むビットが「１」から「０」に変化し、この場合、マスクの対応するビットも「１」を含む。たとえば、図２９には、処理されるビットが、第１のマスクを使用してクリアされた後の（修正された）ポート・ビットマップ２９０８’のビットが示されている。図示のように、修正されたポート・ビットマップにおいて、「１」はビット２　２９１０にのみ残っており、「０」はビット６　２９１２、２９１４にあり、ビット２はクリアされておらず（すなわち、スイッチ・ポート２／１はまだ処理されていない）、ビット５および６がクリアされている（すなわち、スイッチ・ポート５／１または６／１が処理されている）ことがわかる。
【０１１４】
次に、スイッチ・ポート２／１が利用可能である（すなわち、アイドル状態である）と判定されると、スイッチ・ポート２／１の処理に基づいてポート・ビットマップ用の第２のマスク２９１６が生成される。このマスクはビット２　２９２２に「１」を有し、スイッチ・ポート２／１を介したパケットの出力を示している。この場合、スイッチ・ポート２／１（すなわち、装置ポート２）からパケットが出力されるとき、第２のマスク２９１６を使用してポート・ビットマップをクリアすることができる。図２９には、第２のマスク２９１６によってクリアされた後の修正されたポート・ビットマップ２９０８”が示されている。図２９に示されているように、修正されたポート・ビットマップ２９０６”のすべてのビットが「０」を含み、それにより、ポート・ビットマップには、パケットを出力する宛先が残っていないことがわかる。
【０１１５】
図３０は、本発明の実施形態による例示的な宛先ベクトル３０００、宛先ベクトルの発信マスク３００２、および修正された宛先ベクトル３００４の概略図である。図３０に示されているように、この例では、装置ポート２と７　３００６、３００８のビットは「１」を含み、それにより、関連するパケットについての装置２と７の宛先を示している。ＳＦ１において、パケットは（装置ポート７に到達するように）２／１（すなわち、装置ポート２）を介し、また５／１または６／１を介してＳＦ２に送信される。次に、ＳＦ２に送信されたパケットが、宛先ベクトル３０００を修正せずに送信された場合、ＳＦ２は、修正されていない宛先ベクトル３０００を抽出し、ＳＦ２の宛先の１つが装置ポート２である場合と同様にパケットを処理し、したがって、パケットのコピーを、装置ポート２を通して送信できるようにＳＦ１に送り返す。
【０１１６】
この問題を解消するには、スイッチ・ファブリック要素の各出力ポートの端部にクリア・プロセスを挿入することができる。クリア・プロセスでは、発信ポートを通して到達できる装置に対応するビットを除いて、発信パケットの宛先ベクトルのすべてのビットがクリアされる。このクリア・プロセスでは、発信マスク３００２が生成され、次いで発信マスク３００２を使用して発信パケットの宛先ベクトルが修正される。発信マスク３００２は

、特定の発信ポートを通してアクセスできるすべてのビットに「１」を含む（残りのビットは「０」である）。引き続きこの例で説明すると、ＳＦ１におけるパケットは、ＳＦ１のいずれかのポート５／１または６／１から装置ポート７に到達することができる（図２３のＳＦ１のスイッチング・テーブルを参照されたい）。図２３に示されているＳＦ１のスイッチング・テーブルを参照すると、ポート５／１または５／６を介して到達できる装置ポートは装置ポート５～１０であることがわかる。したがって、生成された発信マスク３００２は、図３０に示されているように、ビット５から１０に「１」を含み、残りのビットに「０」を含む。次いで、この発信マスクを使用し、発信マスクの「１」でカバーされていないビット内の任意の「１」の宛先ベクトルをクリアすることにより、ポート５／１または６／１からＳＦ２に向かうパケットの宛先を修正することができる。したがって、図３０に示されているこの例の修正された宛先ベクトル３００４は、装置ポート２用のビットの「１」が発信マスクの「１」によってマスクされていないので装置ポート７　３０１０用のビットに１つの「１」を有するに過ぎない。このようなクリアプロセスを用いた場合、修正された宛先ベクトルは、ＳＦ２に、パケットのコピーを、装置ポート２に送信されるようにＳＦ１に送り返させる情報を含まない。好ましい実施形態では、スイッチ・ファブリック要素の各出力ポートの発信マスクが、それぞれの各ポートに対応するスイッチング・テーブルの関連する行と同じであってよいことに留意されたい。このような好ましい実施形態の一態様では、スイッチ・ファブリック要素の各出力ポートは、それぞれの切換え要素のスイッチング・テーブルの関連する行と同じ内容を含むレジスタを有してよい。

【０１１７】

図３１は、本発明に実施形態によって判定結果が宛先ベクトルを含むスイッチ・ファブリック要素を通してパケットを処理する他のプロセス３１００を示す概略図である。図２８に示されている実施形態と同様に、パケット３１０２が、スイッチ・ファブリック要素に到着する前に、スイッチ・ファブリック要素に結合されたポート・コントローラで処理されている間に、宛先ベクトル（好ましくは、３２ビット宛先ベクトル）を示すタグがパケット３１０２に挿入される。パケット３１０２がスイッチ・ファブリック要素に到着すると、タグが抽出され、宛先ベクトル３１０４が得られる。図２８に示されたプロセス２８００と同様に、次いで、宛先ベクトル３１０４およびスイッチ・ファブリック要素３１００のスイッチング・テーブル３１０８に基づいてポート・ビットマップ３１０６が生成される。好ましい実施形態では、ポート・ビットマップ３１０６は８ビット・ポート・ビットマップからなる。生成されたポート・ビットマップ３１０６は、リンク済みリスト３１１０内で、パケットが記憶されているメモリのセルをリンクするために使用される。しかし、８ビット・ポート・ビットマップ３１０６は、リンク済みリスト３１１０には格納されない。その代わり、（３２ビット）宛先ベクトル３１０４がリンク済みリスト３１１０に格納される。

【０１１８】

引き続き図３１を参照するとわかるように、パケットの発信にあたって適切な出力ポートが利用可能であることが判定されると、リンク済みリスト３１１０からセル情報が読み取られ、宛先ベクトル３１０４が得られる。スイッチング・テーブルを使用して、発信ポートの宛先マスク３１１２が生成される。宛先マスクの長さは３２ビットである（すなわち、「３２ビット宛先マスク」）ことが好ましい。一態様では、３２ビット宛先マスクは、発信ポートに対応するスイッチング・テーブルの行を含む。次に、宛先マスク３１１２を使用して宛先ベクトル３１１４が修正され、宛先ベクトル３１１４はその後、リンク済みリスト３１１０に格納される。図２８の修正されたポート・マスク２８１８と同様に、修正された宛先ベクトル３１１４がすべての「０」ビットを含む場合、対応するメモリ・セルが解放される。

【０１１９】

このような実施形態３１００の好ましい態様では、リンク済みリストは、宛先ベクトル３１０４を格納するための３２ビット／ワード×２５６ワードメモリを含んでよい。このよ

うな実施形態の他の好ましい態様では、図２８に示されている実施形態のスイッチング・テーブル２８１２とは異なり、宛先マスクがスイッチング・テーブル３１０８の行に過ぎないので、スイッチング・テーブル３１０８が、宛先マスクを生成する論理（すなわち、回路）を含む必要はない。この実施形態の他の好ましい態様では、スイッチング・テーブルの行が互いに排他的である必要はない。

【０１２０】

（後方伝搬）

図３２は、本発明の実施形態によるスイッチ・ファブリック中のスイッチング・テーブルを更新するプロセス３２００のフローチャートである。オペレーション２０２において、送信側のソースと受信側のスイッチ・ファブリック要素との間のハンドシェーキングなしで、スイッチ・ファブリック要素によって１つまたは複数のパケットが受信される。各ステータス・パケットは、出力ポートのステータスに関する情報を含む。次いで、オペレーション２０４において、受信されたステータス・パケットのステータス情報に基づいてスイッチング・テーブルが更新される。

【０１２１】

本発明の態様では、スイッチ・ファブリック・システムによるイーサネット・パケットの送信専用のパスとは別の、ステータス・パケットの送信専用のパス／チャネル／コンジットを介してステータス・パケットを受信することができる。本発明の他の態様では、ステータス・パケットを定期的に受信することができ、スイッチング・テーブルをリアルタイムで更新することができる。

【０１２２】

本発明の実施形態では、受信されたステータス・パケットの組み合わされたステータス情報に基づいて、組み合わされたステータス・パケットを生成することができる。

【０１２３】

次いで、組み合わされたステータス・パケットを他のスイッチ・ファブリック要素に送信することができる。このような実施形態の一態様では、組み合わされたステータス・パケットを外側層スイッチ・ファブリック要素から中間層スイッチ・ファブリック要素に送信することができる。

【０１２４】

本発明の一態様では、受信されるステータス・パケットをポート・コントローラから生成することができる。本発明の他の態様では、受信されるステータス・パケットをスイッチ・ファブリック要素から生成することができる。本発明の他の態様では、ステータス・パケットは、各ビットがスイッチ・ファブリック要素のポートに関連する３２ビットのデータを含んでよい。

【０１２５】

一実施形態において、スイッチ・ファブリック・ポート・コントローラを備えるスイッチ・システムは、パケットがスイッチ・ファブリックを通って宛先ポート・コントローラに至るパスを、ハンドシェーキング・プロトコルを使用して判定する。しかし、ハンドシェーキング・プロトコルを使用すると、（１）各着信パケットから、スイッチ・ファブリックを通るパスを判定しなければならず、（２）ハンドシェーキング・プロトコルが、パケットがスイッチ・ファブリックを通って伝搬するのと同じチャネルまたはコンジットを使用して実行されるため、スイッチ・システムの帯域幅が小さくなる可能性がある。

【０１２６】

図３３は、本発明の実施形態によるスイッチ・ファブリックにおけるステータス情報および制御情報の後方伝搬の例示的な構成３３００を示す概略図である。図３３に示されている例示的なスイッチ・ファブリック３３００は、図１に示されているマルチポート・スイッチ・アーキテクチャ１００のＣｌｏｓネットワーク・トポロジに配置することが好ましい１２個のスイッチ・ファブリック要素（ＳＦ１からＳＦ１２）を備えている。例示的なスイッチ・ファブリック３３００は、３２個のポート・コントローラ（ＰＣ０１からＰＣ３２）と、外側層スイッチ・ファブリック要素（ＳＦ１～４およびＳＦ９～１２）のそれ

ぞれに結合された４つのポート・コントローラとをさらに有している。
【０１２７】
後方伝搬機構３３００において、各ポート・コントローラは、それぞれのポート・コントローラを、関連する外側層スイッチ・ファブリック要素に結合する専用コンジット（たとえば、３３０２）を有している。同様に、各外側層スイッチ・ファブリック要素（ＳＦ１～４およびＳＦ９～１２）は、それぞれの外側層スイッチ・ファブリック要素を各中間層スイッチ・ファブリック要素に結合する専用コンジット（たとえば、３３０４）を有している。ＳＦ１が各ポート・コントローラからステータス情報をどのように受信するかの観点から後方伝搬プロセスを説明するうえで図を明確にするために、図３３には、外側層スイッチ・ファブリック要素ＳＦ２～４およびＳＦ９～１２からの専用コンジットしか示されていない。外側スイッチ・ファブリック要素（すなわち、ＳＦ１～４およびＳＦ９～１２）がそれぞれ、各中間スイッチ・ファブリック要素（すなわち、ＳＦ５～８）の専用コンジットを有することに留意されたい。各中間スイッチ・ファブリック要素（すなわち、ＳＦ５～８）は、それぞれの中間スイッチ・ファブリック要素を各外側スイッチ・ファブリック要素（すなわち、ＳＦ１～４およびＳＦ９～１２）に結合する専用コンジット（たとえば、３３０６）を有している。この場合も、ＳＦ１が各ポート・コントローラからステータス情報をどのように受信するかに関して後方伝搬プロセスを説明するうえで図を明確にするために、中間スイッチ・ファブリック要素ＳＦ５から外側層スイッチ・ファブリック要素ＳＦ１への専用コンジット３３０６しか示されていない。しかし、外側層スイッチ・ファブリック要素が、図３３の例で説明したのと同様に各中間層スイッチ・ファブリック要素からステータス情報を受信できることを認識されたい。
【０１２８】
引き続き図３３を参照するとわかるように、専用コンジット（たとえば、３３０２、３３０４、３３０６）を介すことにより、パケット（および他の制御情報）を受信するポート・コントローラの可用性に関するステータス情報をスイッチ・ファブリックを通して伝搬させることができる。専用コンジット３３０２、３３０４、３３０６の矢印は、スイッチ・システム全体からスイッチ・ファブリック要素ＳＦ１へのステータス情報の伝搬を示している。好ましい実施形態では、専用パスはそれぞれ、１サイクル当たり２ビットの情報を送信できるようにする帯域幅を有してよい。このような好ましい実施形態では、帯域幅の１ビットを制御情報の送信に使用し、他のビットを、スイッチ・システムのポート・コントローラの可用性に関するステータス情報の送信に使用してよい。
【０１２９】
図３３には、ＳＦ５を介してステータス情報を受信するように中間層スイッチ・ファブリック要素ＳＦ５に結合されたスイッチ・ファブリック要素ＳＦ１が示されているが、中間層スイッチ・ファブリック要素ＳＦ６、ＳＦ７、およびＳＦ８はそれぞれ、ステータス情報をＳＦ１に与えるようにＳＦ１に結合された専用コンジットを有してよい。この冗長性は、ＳＦ５がステータス情報をＳＦ１に送信できない場合でもポート・コントローラに関する情報を引き続き受信することができるので有利である。
【０１３０】
一般に、出力ポート・ステータスは、スイッチ・ファブリックを通して伝播するときにスイッチ・ファブリックのローカル出力ポート・ステータスと併合され、それによってスイッチ・ファブリックの全体的なパケット・ステータスに関する与えられた情報を蓄積することができる。したがって、このような機構を用いた場合、スイッチ・システム内の各要素はポートのリアルタイムに更新されたパケット・ステータスを維持し、このローカル情報を使用して、スイッチ・システムへのパケットのイングレス点とスイッチ・システム内のパケットの宛先との間のハンドシェーキング・プロセスなしにパケットの流れを制御することができる。
【０１３１】
図３４は、本発明の実施形態による、スイッチ・ファブリック・システム内のスイッチング・テーブルを更新するプロセス３４００のフローチャートである。好ましい実施形態で

は、このプロセスは、明細書の、図３３に関する部分に記載された後方処理専用コンジットを有するスイッチ・システムで実行される。図３４に示されているように、プロセスの一部は、スイッチ・ファブリック・システムのポート・コントローラで実行され、他の部分は、外側層スイッチ・ファブリック要素および中間層スイッチ・ファブリック要素で実行される。

【０１３２】

オペレーション３４０２において、ポート・コントローラでステータス・パケットが実行される。ステータス・パケットは、発信パケットの受信に関するポート・コントローラの現在のステータス／可用性を示す情報を含む。生成されたステータス・パケットは、オペレーション３４０４で、ポート・コントローラに関連する外側層スイッチ・ファブリック要素に送信される。たとえば、図３３を参照するとわかるように、ポート・コントローラＰ３２－０８はスイッチ・ファブリック要素ＳＦ２に関連付けされている。したがって、ポート・コントローラＰ３２－０８は、生成したステータス・パケットをスイッチ・ファブリック要素ＳＦ２に送信する。好ましい実施形態では、ポート・コントローラは、ステータス・パケットを、関連する外側層スイッチ・ファブリック要素に、該ポート・コントローラと該スイッチ・ファブリック要素を連結する専用コンジット（たとえば、専用コンジット３３０２）を介して送信する。

【０１３３】

オペレーション３４０６で、外側層スイッチ・ファブリック要素は、関連するポート・コントローラから着信ステータス・パケットを受信し収集する。受信されたステータス・パケットからの情報は、次いでオペレーション３４０８で、外側層スイッチ・ファブリック要素によって生成されたステータス・パケットにおいて組み合わされる。オペレーション３４１０で組み合わされたステータス・パケットは、次いで中間層スイッチ・ファブリック要素に送信される。外側層スイッチ・ファブリック要素は、組み合わされたステータス・パケットを中間層スイッチ・ファブリック要素に、これらのスイッチ・ファブリック要素を連結する専用コンジット（たとえば、ＳＦ２とＳＦ５を連結する専用コンジット３３０４）を介して送信することが好ましい。外側層スイッチ・ファブリック要素はさらに、ポート・コントローラから受信したステータス・パケットの情報を使用して、それ自体の切換えステータスを更新する（オペレーション３４１２参照）。

【０１３４】

中間層スイッチ・ファブリック要素は、オペレーション３４１４で、外側層スイッチ・ファブリック要素によって送信されたステータス・パケットを受信し収集する。たとえば、図３３に示されている例示的な機構を参照する。オペレーション３４１６で、中間層スイッチ・ファブリック要素は、外側層スイッチ・ファブリック要素から受信されたステータス・パケットに含まれるステータス情報を組み合わせたステータス・パケットを生成する。たとえば、図３３に示されている例示的な実施形態では、中間層スイッチ・ファブリック要素ＳＦ５は、各パケットがそれぞれの各スイッチ・ファブリック要素に関連するポート・コントローラについてのステータス情報を持つステータス・パケットを外側層スイッチ・ファブリック要素ＳＦ２～ＳＦ１２から受信する（たとえば、スイッチ・ファブリック要素ＳＦ２からのステータス・パケットは、ポート・コントローラＰ３２－０８に関する情報を有し、他のパケットについても同様である）。

【０１３５】

オペレーション３４１８で、中間層スイッチ・ファブリック要素によって生成されたステータス・パケットは、受信側の外側層ファブリック要素（たとえば、図３３に示されている例では、スイッチ・ファブリック要素ＳＦ１）に送信される。中間層スイッチ・ファブリック要素は、それが生成した組み合わされたステータス・パケットを受信側の外側層スイッチ・ファブリック要素に、これらのスイッチ・ファブリック要素を連結する専用コンジット（ＳＦ５とＳＦ１を連結する専用コンジット３３０６）を介して送信するのが好ましい。さらに、中間層スイッチ・ファブリック要素は、（ポート・コントローラのステータスに関する）外側層スイッチ・ファブリック要素から受信したステータス・パケットの

情報を使用して、それ自体のスイッチング・テーブルを更新する（オペレーション３４２０参照）。
【０１３６】
オペレーション３４２２で、中間層スイッチ・ファブリック要素からステータス・パケットを受信した後、受信側の外側層スイッチ・ファブリック要素（たとえば、図３３に示されている例ではスイッチ・ファブリック要素ＳＦ１）は、受信されたステータス・パケットに含まれるステータス情報に基づいて、オペレーション３４２４でスイッチング・テーブルを更新する。
【０１３７】
したがって、図３３および図３４に示されている後方伝搬機構および手順の下では、パケットの受信および出力に関するすべての装置ポートの可用性（すなわち、出力ステータス）を定期的かつ／または連続的に（好ましくはリアルタイムで）通知する専用コンジットを設けてよい。さらに、この後方伝搬機構は、ステータス情報をすべてのスイッチ・ファブリック要素に送信するにあたって必ずしもハンドシェーキング・プロトコルを使用しなくてもよい。
【０１３８】
図３５は、本発明の好ましい実施形態による、図３４に示されているプロセスで使用できるステータス・パケット３５００の概略図である。このステータス・パケットは複数のビットを含み、好ましい実施形態では３２ビットでよい。ステータス・パケットの各ビットは、スイッチ・システムの特定のポート・コントローラと関連付けることができる。たとえば、図３３に示されているスイッチ・システムに基づいて、ステータス・パケットのビット０１　３５０２をポート・コントローラＰＣ０１と関連付けし、ステータス・パケットのビット０２　３５０４をポート・コントローラＰ５と関連付けし、他のビットについても同様に関連付けすることができる。ステータス・パケットの各ビットは、スイッチ・システムを通して送信されるパケットの受信／出力に関する、関連するポート・コントローラの可用性（すなわち、ステータス）を示すために使用することができる。
【０１３９】
一例として、「１」を含むビットは、関連するポート・コントローラが、パケットの受信／出力に利用できる（すなわち、このポート・コントローラのステータスは「利用可能」である）ことを示すことができる。逆に、「０」を含むビットは、関連するポート・コントローラが、パケットの受信／出力に利用できない（すなわち、このポート・コントローラのステータスは「利用不能」である）ことを示すことができる。したがって、図３５に示されている例示的なステータス・パケット３５００において、ビット０１および０２　３５０２、３５０４に含まれる「１」は、ポート・コントローラＰＣ０１およびＰ５が利用可能であることを示し、一方、ビット０９、１６、および１７　３５０６、３５０８、３５１０に含まれる「０」は、ポート・コントローラＰＣ０９、ＰＣ１６、ＰＣ１７が利用不能であることを示す。
【０１４０】
（マルチバンク・バッファ）
図３６は、スイッチ・ファブリックのスイッチ・マトリクスに、入力データグラムを保存するためのプロセス３６００のフローチャートである。スイッチ・マトリクスは、各バッファが１組の部分を有する１組のバッファを有する。データグラムのデータはオペレーション３６０２に受信され、バッファ部分は、オペレーション３６０４においてデータで連続的に満たされる。オペレーション３６０６において、周期的に、データの転送は、バッファからスイッチ・マトリクスに可能になる。バッファ部分が満たされるシーケンスにおいて、データの転送が可能になる各周期において、バッファ部分の１つにおけるデータは、オペレーション３６０８において、スイッチ・マトリクスに転送されてもよい。
【０１４１】
本発明の１つの態様において、バッファ部分が、データで満たされている、または、データグラムの終端（すなわち、データグラムのテールまたは最終バイト）を含んでいる場合

、バッファ部分におけるデータは、スイッチ・マトリックスに転送されるだけでよい。本発明の他の態様において、スイッチ・マトリクスは、転送されたデータを保存するための複数のメモリ・バンクで構成することもできる。このような実施形態において、メモリ・バンクは、バッファ部分から転送されたデータを受信する際に交替することができる。

【０１４２】

本発明のさらなる態様において、バッファ部分は、受信したデータを保存するために同等の保存容量をそれぞれ有することができる。好ましい実施形態において、各バッファ部分は、入ってきたデータを保存するために、１６バイトの保存容量を有する。他の好ましい態様において、データの転送は、１６サイクル毎に可能としてもよい。

【０１４３】

図３７は、本発明の実施形態によるスイッチ・ファブリック・コンポーネント３７００内のパケット（すなわち、データグラム）３７０２の入力および保存を示す模式的ダイヤグラムである。上記に検討したように、スイッチ・ファブリック・コンポーネント３７００は、Ｖユニット１～９として示される８個のイングレス受信器（例えば、図１および２２に示すスイッチ・ファブリック・コンポーネントのような）を有する図示された実施形態を備えた、複数のインタフェース受信器３７０４を含むことができる。

【０１４４】

各イングレス受信器は、スイッチ・マトリクス３７０６に接続され、これにより、特定のイングレス受信器から、スイッチ・マトリクス３７０６への入力パケットの転送を可能にする。特に、イングレス受信器は、関連するセットの入力バッファに結合されている。一実施形態において、入力バッファの各セットは、２つの入力バッファ３７０８、３７１０から構成される。一実施形態において、各入力バッファは、少なくとも１６バイトの保存容量を有することができる。好ましい実施形態において、各入力バッファは３２バイトの保存容量を有する。好ましい選択肢として、各入力バッファは、２つの部分（例えば、入力バッファ３７１０の２つの部分３７１２、３７１４）に分割してもよい。このような選択肢について、各入力バッファの２つの部分３７１２、３７１４が、同等のデータ保持容量（例えば、３２バイト容量の好ましい実施形態におけるもの）を有することが好ましい。

【０１４５】

中間コントローラ３７１６は、入力バッファに保存されたデータのスイッチ・マトリクスのメモリ３７１８への転送を制御する。一実施形態において、スイッチ・マトリクスのメモリは、内部の複数のメモリ・バンクから構成することができる。図３７に示すように、各入力バッファが２つの部分から構成されている好ましい実施形態において、スイッチ・マトリクスのメモリは、各メモリ・バンクが、各入力バッファの２つの部分の１つからデータを受信するための専用であること（例えば、メモリ・バンク３７２０が、入力バッファ部分３７１２の専用であり、メモリ・バンク３７２２が入力バッファ部分３７１４の専用であるなど）が好ましい、２つのメモリ・バンク３７２０、３７２２で構成することができる。

【０１４６】

図３８は、本発明の実施形態による、１組の２部分入力バッファ３７０８、３７１０、および、２メモリ・バンク・スイッチ・マトリクス３７０６を有するスイッチ・ファブリック・コンポーネント３７００内のパケットのイングレスと保存を示すシナリオの模式的ダイヤグラムである。２つのうちの第１の入力バッファ３７０８は、１組の部分（「部分Ａ」３８０２および「部分Ｂ」３８０４）から構成される一方、第２の入力バッファ３７１０は、他の１組の部分（「部分Ｃ」３７１２および「部分Ｄ」３７１４）から構成される。各入力バッファ３７０８、３７１０は、３２バイトの容量を有し、各部分３７１２、３７１４、３８０２、３８０４は、１６バイトの容量を有している。図３８を参照すると、バッファの各々における各バイトは、矩形で表される。黒い矩形は、特定のバイトがデータで満たされていることを表す。

【０１４７】

イングレス受信器３７０４は、各サイクルの入力バッファの１つのバイトに、それが受信中の入力パケットの１バイトを転送することができる。このデータ転送の間、イングレス受信器は、入力バッファのバイトを連続的に満たす。同様に、入力バッファの部分から、メモリへのデータの転送を可能にするために、中間コントローラは、周期的に、１６サイクル・カウント毎に（メモリ・バンク１　３７２０またはメモリ・バンク２　３７２２のいずれかの）メモリにアクセスすることを許可する。このシナリオにおいて、メモリのアクセスは、先ず、サイクル・カウント７で許可され、続いて、サイクル・カウント２３、３９、５５、７１、８７、１０３、１１９、１３５、１５１などで許可される。さらに、入力バッファからメモリ・バンクへのデータの転送は、入力バッファの部分がデータで完全に満たされた（すなわち、１６バイトのデータ）、または、パケットの終端バイト（すなわち、パケットの最終バイト）を含む場合にのみ行われる。

【０１４８】

さらに、メモリ・バンクは、入力バッファからのパケット・データの受信中に交替する。例えば、第１のメモリ・アクセスにおいて、データはメモリ・バンク１に転送してもよく、第２のメモリ・アクセスにおいて、データはメモリ・バンク２に転送してもよく、また、第３のメモリ・アクセスにおいて、データはメモリ・バンク３に転送してもよい、などである。

【０１４９】

図３８に示すシナリオにおいて、９７バイトの長さを有する第１の入力パケット３８０６は、スイッチ・ファブリック３７００のイングレス受信器３７０４に転送され、この後、４バイトのスペース３８０８が続き、その後、９７バイトの長さを有する第２の入力パケット３８１０が続く。サイクル・カウント１では、第１のパケットの第１のバイトは、イングレス受信器によって受信され、第１の入力バッファの部分Ａ３８０２の第１のバイト３８１２を満たすために転送される。

【０１５０】

サイクル７において、部分Ａ３８１２の１６バイトのうちの７つは、第１パケット３８０６のデータの最初の７バイトで満たされ、中間コントローラは、このシナリオにおいて初めてメモリ・バンクへのアクセスを許可する。しかし、入力バッファのどの部分も完全に満たされていず、それらのどれもパケットの終端バイトを含んでいないため、サイクル７において、入力バッファとメモリ・バンクとの間では、パケット・データの転送はない。

【０１５１】

サイクル・カウント１６において、第１の入力バッファの部分Ａ３８１２の全ての１６のバイトは、パケット１の最初の１６入力バイトで満たされている。しかし、この時点において、次のメモリ・アクセスは、サイクル・カウント２３まで行われない。

【０１５２】

図３９は、本発明の実施形態によるサイクル・カウント２３における図３８に示すシナリオの模式的ダイヤグラムである。サイクル・カウント２３までに、パケット１のさらに７つのバイトが、入力バッファ１の部分Ｂ３８０４に転送されている。同様に、図３９に示すように、サイクル・カウント２３において、メモリ・アクセスが中間コントローラによって可能となり、入力バッファ１の部分Ａが（サイクル・カウント１６以来）満たされているため、部分Ａ　３８０２に含まれるデータの１６バイトはメモリ・バンク１　３７２０に読み出され、書き込まれ（矢印３９０２を参照）、続いて、部分Ａ３８０２からクリアされる。

【０１５３】

サイクル・カウント３９における次のメモリ・アクセスの時までに、入力バッファ１の部分Ｂの全ては、パケット１からのデータで（サイクル・カウント３２以来）満たされており、パケット１のバイト３３～３９は、入力バッファ２　３７１０の部分Ｃ　３７１２に転送されている。したがって、サイクル・カウント３９において、部分Ｂのデータのバイトの全てがメモリ・バンク２　３７２２に転送される（すなわち、読み出しと書き込みの機能を実行する）。

【０１５４】
サイクル・カウント５５における次のメモリ・アクセスの時に、入力バッファ２の部分Ｃの全ては、パケット１からのデータで（サイクル・カウント４８以来）満たされており、パケット１のバイト４９～５５は、入力バッファ２　３７１０の部分Ｄ３７１４に転送されている。したがって、サイクル・カウント５５において、部分Ｃのデータのバイトの全ては、メモリ・バンク１　３７２０に（メモリ・バンクがデータの受信に対する利用可能性において交替するため）転送される（すなわち、読み出しと書き込みの機能を実行する）。
【０１５５】
サイクル・カウント７１における次のメモリ・アクセスの時に、入力バッファ２の部分Ｄの全ては、パケット１からのデータで（サイクル・カウント６４以来）満たされており、パケット１のバイト６５～７１は、入力バッファ１　３７０８の部分Ａ３８０２に転送されている。同様に、サイクル・カウント７１において、部分Ｄのデータのバイトの全ては、メモリ・バンク２　３７２２に転送される。
【０１５６】
次のメモリ・アクセスは、サイクル・カウント８７において行われ、この時、入力バッファ１の部分Ａの全ては、パケット１からのデータで（サイクル・カウント８０以来）満たされており、パケット１のバイト８１～８７は入力バッファ１　３７０８の部分Ｂ　３８０４に転送されている。同様に、部分Ａのデータのバイトの全ては、サイクル・カウント８７でメモリ・バンク１　３７２０に転送される。
【０１５７】
図４０は、本発明の実施形態によるサイクル・カウント１０３における図３８に示すシナリオの模式的ダイヤグラムである。図４０は、サイクル・カウント８８とサイクル・カウント１０３との間のシナリオにおいて、何が起こるかについての以下の検討のために使用される。サイクル・カウント９６において、入力バッファ１の部分Ｂの全ては、パケット１のバイト８１～９６で満たされている。サイクル・カウント９７において、パケット１の最終バイト（バイト９７）は、入力バッファＢの部分Ｃの第１のバイト４００２に転送される。パケット１のバイト９７はパケット１の終端バイトであるため、中間コントローラは、入力バッファ２をデータで満たすことを終える。
【０１５８】
サイクル９８、９９、１００、１０１の間、パケット１と２の間の４バイトのスペースのために、データはイングレス受信器によって受信されない。サイクル・カウント１０２において、パケット２の第１のバイトは、イングレス受信器によって受信され、入力バッファ１の部分Ａの第１のバイトに転送される。この時点で、入力バッファ１の部分Ｂがパケット１のバイト８１～９６を未だに含んでいることを念頭に置くことは重要である。入力バッファ１に転送された、サイクル・カウント１０２において受信されたパケット１の入力バイトは、同じく入力バッファ１に同時に保存されているパケット１からのデータのいずれをも上書きしない。これは、パケット２　４００２の第１の入力バイトが、入力バッファ１の部分Ａに保存される一方、パケット１の保存されたバイト（バイト８１～９６）は、入力バッファ１の部分Ｂに保存されるためである。
【０１５９】
サイクル・カウント１０３において、次のメモリ・アクセスが行われる。パケット１のバイト８１～９６は、メモリ・バンク２　３７２２に転送され、入力バッファ１の部分Ｂはクリアされる。同様に、サイクル・カウント１０３において、パケット２の第２の入力バイト４００６は、入力バッファ１の部分Ｂに転送される。パケット１　４００２のバイト９７は、入力バッファ２の部分Ｃ４００２に未だに残っている。サイクル・カウント１１９において、続く次のメモリ・アクセスが中間コントローラによって許可される。この時点において、バイト９７がパケットの終端バイトであるため、パケット１　４００２のバイト９７は、メモリ・バンク１に転送される。同様に、サイクル・カウント１１９においてメモリ・バンク１に転送されるものは、サイクル・カウント１１７において部分Ａを満

たすことを完了した、パケット２の最初の１６バイトである。
【０１６０】
図４１は、本発明の実施形態による図３８で使用されているデュアルの２部分バッファの代わりに、小さなデュアル入力バッファ４１０２、４１０４が使用されている図３８に示すシナリオに類似する流れ４１００の模式図である。同様に、このシナリオ４１００において、単一のバンク・メモリ４１０６が、デュアル・メモリ・バンクの代わりにスイッチ・マトリクスに設けられている。図３８に示すシナリオにおけるように、各入力バッファ４１０２、４１０４は、それぞれのバッファ内に矩形で示す各バイトを備えた３２バイトを保持するための容量を有する。黒い矩形は、特定のバイトがデータで満たされていることを示す。中間コントローラは、同様に、周期的に、入力バッファからメモリ・バンクへのデータの転送を可能にするために、（ちょうど図３８のシナリオにあるように）このシナリオ４１００において、１６サイクル・カウント毎にアクセスを許可する。この現在のシナリオにおいて、メモリ・アクセスも、同様に、先ず、サイクル・カウント７で可能となり、続いて、サイクル・カウント２３、３９、５５、７１、８７、１０３、１１９、１３５、１５１などで可能となる。入力バッファが、データ（すなわち、３２バイトのデータ）で完全に満たされているか、パケットの終端バイト（すなわち、パケットの最終バイト）を含んでいる場合にのみ、さらに、このシナリオ４１００において、入力バッファからメモリ・バッファへのデータの転送が行われる。
【０１６１】
さらに、図３８に示すシナリオにおけるように、このシナリオ４１００において、９７バイトの長さを有する第１の入力パケット３８０６は、イングレス受信器に転送され、次に、４バイトのスペースが続き、次に、９７バイトの長さを有する第２の入力パケットが続く。サイクル・カウント１で、第１パケットの第１バイトは、イングレス受信器によって受信され、第１入力バッファ４１０２を満たすために転送される。
【０１６２】
図３８に示す配置と比較する時、この配置に生じる問題としては、小さなデュアル入力バッファが、様々な大きさの入力パケットを扱えないことである。この問題は、サイクル・カウント１０２におけるシナリオ４１００の状況を示す図４１に示す。サイクル・カウント９６において、第１の入力バッファ４１０２は、第１の入力パケットのバイト６５～９６で完全に満たされている。第１入力バッファが満たされていず、パケットの終端バイトを含んでいなかったため、第１の入力バッファ４１０２のデータの転送は、サイクル・カウント７１および８７において、メモリ・アクセスで可能とされた。サイクル・カウント９７において、第１の入力パケットの最終バイト（バイト９７）４１０８が受信され、第２の入力バッファ４１０４の第１のバイトに転送される。しかし、次のメモリ・アクセスは、サイクル・カウント１０３において可能となる。したがって、第１の入力バッファの第１のパケットのデータは、サイクル・カウント１０３まで転送することができない。
【０１６３】
サイクル・カウント９８～１０１において、入力パケット間にスペースが発生し、そのため、これらのサイクルの間には入力バッファへの伝送されるデータはない。サイクル・カウント１０２において、第２のパケットの第１のバイト４１１０は、イングレス受信器によって受信され、（標準パケット転送プロトコルに従って）第１の入力バッファに転送され、したがって、第１の入力バッファに保存された第１のパケットのバイト４１１２（バイト６５）を上書きし、これによって、第１のパケットのデータを、これがメモリ・バンク４１０６に転送される前に、損傷する。第１のパケットのデータをそのような状態に損傷する問題を回避するための１つの選択肢は、第３の入力バッファを追加することである。他の選択肢は、図３８に示すデュアルの２部分入力バッファ配置を使用することである。デュアルの２部分入力バッファ配置を使用することによって、スイッチ・マトリクス／スイッチ・ファブリック・コンポーネントの全チップ面積のハードウェア面での削減を達成することができる。
【０１６４】

図４２は、本発明の実施形態によるスイッチ・マトリクス・メモリ・バンクのためのメモリ・アクセスのサイズを示す模式的ダイヤグラムである。入力バッファ４２０４への入力パケット（すなわち、データグラム）４２０２に対するデータ伝送速度（「Ｒｘ速度」）は、Ｒ（ビット／秒）と表すことができる。メモリ４２０６へのアクセスは、各入力バッファ（および、各出力バッファに対しても同じであってもよい）に対してＴ（秒）毎に許可される。本発明の実施形態によれば、メモリに対する最小メモリ・アクセス・サイズは、ＲＴ（ビット）である。好ましい実施形態において、様々な大きさのパケットを扱うために十分なメモリ・アクセス・サイズは、２ＲＴ（ビット）である。

【０１６５】

本発明によるシステムの一実施形態は、ＩＢＭ互換のパーソナル・コンピュータ、Ａｐｐｌｅ　Ｍａｃｉｎｔｏｓｈコンピュータ、または、ＵＮＩＸ（登録商標）ベースのワーク・ステーションなどのコンピュータのコンテキストにおいて実施することができる。代表的なハードウェア・コンフィギュレーションは、図４３に示し、この図は、マイクロプロセッサなどの中間処理ユニット４３１０、および、システム・バス４３１２を介して相互接続されたいくつかの他のユニットを有する実施形態によるワーク・ステーションの典型的なハードウェア・コンフィギュレーションを示す。このワーク・ステーションは、さらに、ランダム・アクセス・メモリ（ＲＡＭ）４３１４、リード・オンリー・メモリ（ＲＯＭ）４３１６、ディスク記憶ユニット４３２０などの周辺機器をバス４３１２に接続するためのＩ／Ｏアダプタ４３１８、キーボード４３２４を接続するためのユーザ・インターフェイス・アダプタ４３２２、マウス４３２６、スピーカ４３２８、マイクロフォン４３３２、および／または、バス４３１２へのタッチ・スクリーン（図示せず）などの他のユーザ・インターフェイス・デバイス、ワーク・ステーションを（例えば、データ処理ネットワークなどの）通信ネットワーク４３３５へ接続するための通信アダプタ４３３４、および、バス４３１２をディスプレー・デバイス４３３８に接続するためのディスプレー・アダプタ４３３６をさらに含む。ワーク・ステーションは、これに常駐で、Ｍｉｃｒｏｓｏｆｔ　Ｗｉｎｄｏｗｓ（登録商標）ＮＴまたはＷｉｎｄｏｗｓ／９５のオペレーティング・システム（ＯＳ）、ＩＢＭ　ＯＳ／２オペレーティング・システム、ＭＡＣ　ＯＳ、または、ＵＮＩＸオペレーティング・システムを典型的に有する。当業者には、本発明の実施形態が、上述したもの以外のプラットフォームおよびオペレーティング・システム上に実装することもできることを理解する。実施形態は、ＪＡＶＡ（登録商標）、Ｃ、および、Ｃ＋＋言語を使用して書くこともでき、オブジェクト志向のプログラミング方法論を利用することもできる。

【０１６６】

様々な実施形態が上記に説明された一方、それらが実施形態の方法でのみ提示され、限定とはならないことが理解される。したがって、好ましい実施形態の柔軟性および範囲は、上記に説明された実施形態のいずれによっても限定されるべきではないが、冒頭の特許請求の範囲およびそれに同等のものに従ってのみ定義されるべきである。

【図面の簡単な説明】

【図１】

本発明の実施形態による、マルチポート・スイッチ・アーキテクチャの概略的なブロック図である。

【図２】

本発明の実施形態による、他のマルチ・ポート・スイッチ・アーキテクチャの概略的なブロック図である。

【図３】

本発明の実施形態による、ギガビット・アップリンキングの概略的なブロック図である。

【図４】

本発明の実施形態による、ポート・コントローラのアーキテクチャの概略的なブロック図である。

【図５】

本発明の実施形態による、スイッチ・ファブリックを通る送信のデータグラムを準備するプロセスのフローチャートである。
【図６】
本発明の実施形態による、例示的なパケットの概略的な表示である。
【図７】
本発明の実施形態による、パケットを処理するプロセスのフローチャートである。
【図８】
本発明の実施形態による、ポート・コントローラにおいて図７に示したパケットを処理するプロセスを示す第１の概略的な流れ図である。
【図９】
本発明の実施形態による、ポート・コントローラにおいて図７に示したパケットを処理するプロセスを示す第２の概略的な流れ図である。
【図１０】
本発明の実施形態による、ポート・コントローラに入る複数の入力パケットの処理および記憶の例を表す概略図である。
【図１１】
本発明の実施形態による、パケットのボディ部分からヘッダ部分を分離するプロセスのフローチャートである。
【図１２】
本発明の実施形態による、ヘッダ部分から分離した後、パケットのボディ部分をメモリに記憶するプロセスのフローチャートである。
【図１３】
本発明の実施形態による、パケットの処理したヘッダ部分をメモリに記憶するプロセスのフローチャートである。
【図１４】
本発明の実施形態による、入力パケットのセルをメモリに記憶することを表す概略図である。
【図１５】
本発明の実施形態による、スイッチ・ファブリック・システムにおいて、ポート・コントローラからスイッチ・ファブリックまでパケットを転送するプロセスのフローチャートである。
【図１６】
本発明の実施形態による、イグレスのためにキューに入れたパケットの連結リストを生成するプロセスのフローチャートである。
【図１７】
本発明の実施形態による、パケット・メモリに記憶されたパケットを割り当てられた出力優先キューに連結することを示す概略図である。
【図１８】
本発明の実施形態による、図１７に示したパケット・メモリに記憶された例示的なパケットを示す概略図である。
【図１９】
本発明の実施形態による、ポート・コントローラを介してスイッチ・システムを出るパケットを処理するプロセスを示す概略的な流れ図である。
【図２０】
本発明の実施形態による、スイッチ・ファブリック構成要素のアーキテクチャの概略的なブロック図である。
【図２１】
本発明の実施形態による、スイッチ・ファブリックを通るデータグラムを送信するプロセスのフローチャートである。
【図２２】

本発明の実施形態による、スイッチ・ファブリックを通るパケットをルーティングするスイッチング・テーブルの実装を示すための、４つの８×８のスイッチ・ファブリック構成要素を備えるスイッチ・ファブリックの概略的なブロック図である。
【図２３】～
【図２６】
本発明の実施形態による、図２２に示したスイッチ・ファブリックのための例示的なスイッチング・テーブルの概略図である。
【図２７】
本発明の実施形態による、スイッチ・ファブリック構成要素を通るパケットを処理するプロセスのフローチャートである。
【図２８】
本発明の実施形態による、決定結果がポート・ビットマップを備えるスイッチ・ファブリック構成要素を通るパケットを処理するプロセスを示す概略図である。
【図２９】
本発明の実施形態による、スイッチ・ファブリック構成要素におけるポート・ビットマップ（特に、８ビットのポート・ビットマップ）の例示的な使用を示す概略図である。
【図３０】
本発明の実施形態による、例示的な宛先ベクトル、宛先ベクトルのためのイグレス・マスク、および変更した宛先ベクトルの概略図である。
【図３１】
本発明の実施形態による、決定結果が宛先ベクトルを備えるスイッチ・ファブリック構成要素を通るパケットを処理する他のプロセスを示す概略図である。
【図３２】
本発明の実施形態による、スイッチ・ファブリックのスイッチング・テーブルを更新するプロセスのフローチャートである。
【図３３】
本発明の実施形態による、スイッチ・ファブリックにおける状況と制御の情報の後方伝播のための例示的な構成を示す概略図である。
【図３４】
本発明の実施形態による、スイッチ・ファブリック・システムにおいてスイッチング・テーブルを更新するプロセスのフローチャートである。
【図３５】
本発明の好ましい実施形態による、図３４に記述したプロセスにおいて使用することが可能である状況パケットの概略図である。
【図３６】
本発明の実施形態による、入力データグラムをスイッチ・ファブリックのスイッチ・マトリックスに記憶するプロセスのフローチャートである。
【図３７】
本発明の実施形態による、スイッチ・ファブリック構成要素におけるパケットのイングレスと記憶（すなわち、データグラム）を示す概略図である。
【図３８】
本発明の実施形態による、２つの部分の入力バッファと２つのメモリ・バンクのスイッチ・マトリックスの対を有するスイッチ・ファブリック構成要素におけるパケットのイングレスと記憶を示すシナリオの概略図である。
【図３９】
本発明の実施形態による、サイクル・カウント２３における図３８に記述したシナリオの概略図である。
【図４０】
本発明の実施形態による、サイクル・カウント１０３における図３８に記述したシナリオの概略図である。

【図４１】
本発明の実施形態による、図３８で使用した二重の２つの部分のバッファの代わりにトリビアルな二重入力バッファを使用した、図３８に提示したシナリオに類似のシナリオの概略図である。
【図４２】
本発明の実施形態による、スイッチ・マトリックス・メモリ・バンクに対するメモリ・アクセス・サイズを示す概略図である。
【図４３】
本発明の一実施形態のハードウエア実装に関する概略図である。