

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3872342号

(P3872342)

(45) 発行日 平成19年1月24日(2007. 1. 24)

(24) 登録日 平成18年10月27日(2006. 10. 27)

(51) Int. Cl. F I  
**HO 4 L 12/56 (2006. 01)** HO 4 L 12/56 E

請求項の数 6 (全 41 頁)

|               |                               |           |  |
|---------------|-------------------------------|-----------|--|
| (21) 出願番号     | 特願2001-521005 (P2001-521005)  | (73) 特許権者 | 390009531  |
| (86) (22) 出願日 | 平成12年8月24日 (2000. 8. 24)      |           | インターナショナル・ビジネス・マシー<br>ズ・コーポレーション                     |
| (65) 公表番号     | 特表2003-508967 (P2003-508967A) |           | INTERNATIONAL BUSIN<br>ESS MASCHINES CORPO<br>RATION |
| (43) 公表日      | 平成15年3月4日 (2003. 3. 4)        |           | アメリカ合衆国10504 ニューヨーク<br>州 アーモンク ニュー オーチャード<br>ロード     |
| (86) 国際出願番号   | PCT/US2000/020796             |           |  |
| (87) 国際公開番号   | W02001/017179                 | (74) 代理人  | 100086243  |
| (87) 国際公開日    | 平成13年3月8日 (2001. 3. 8)        |           | 弁理士 坂口 博   |
| 審査請求日         | 平成14年3月13日 (2002. 3. 13)      | (74) 代理人  | 100091568  |
| (31) 優先権主張番号  | 09/384, 747                   |           | 弁理士 市位 嘉宏  |
| (32) 優先日      | 平成11年8月27日 (1999. 8. 27)      |           |  |
| (33) 優先権主張国   | 米国 (US)                       |           |  |

最終頁に続く

(54) 【発明の名称】 ネットワークのための装置およびスケーラブル・ネットワーク・プロセッサ

## (57) 【特許請求の範囲】

## 【請求項 1】

制御点プロセッサと、  
 前記制御点プロセッサに動作的に接続されるインタフェース装置と  
 を含み、前記インタフェース装置が、  
 半導体基板と、  
 前記基板上に形成される少なくとも5個のインタフェース・プロセッサと、  
 前記インタフェース・プロセッサによりアクセスされる命令を記憶する、前記基板上に  
 形成され、前記インタフェース・プロセッサにより共用される共通内部命令メモリと、  
 前記インタフェース装置を通過し、前記インタフェース・プロセッサによりアクセスさ 10  
 れるデータを記憶する、前記基板上に形成される内部データ・メモリと、  
 前記基板上に形成される複数の入出力ポートと  
 を含み、  
 前記入出力ポートの少なくとも1つが、前記データ・メモリを外部データ・メモリに接  
 続し、  
 他の少なくとも2つの前記入出力ポートが、前記制御点プロセッサから伝送されるガイ  
 ド・フレームを使用して前記インタフェース・プロセッサの指示の下で、前記インタフェ  
 ース装置を通過するデータを外部ネットワークと交換し、  
 前記制御点プロセッサが前記インタフェース装置と協働して、前記インタフェース・プ  
 ロセッサにより実行される命令を前記共通内部命令メモリにロードし、前記命令が実行さ 20

れて、前記入出力ポート間でのデータの交換、及び前記データ・メモリを介するデータの流れを指示し、

前記インタフェース装置が2つの高速相互接続ポートを有し、前記高速相互接続ポートの各々が、

Nビット並列データとして表されるデータ・ストリームを、各々がnビット(但しnはNの分数)を有する複数の部分に構文解析する論理回路と、

前記論理回路と通信して、前記データ・ストリームの各解析部分を直列化するシリアライザと、

前記シリアライザと通信して、データ・ストリームを転送する、各々が前記データ・ストリームの前記解析部分の各々に関連付けられる、前記解析部分の数に等しい複数のチャネルと、

10

前記チャネルと通信し、該チャネルを介して転送される直列データ・ストリームを受信し、前記データ・ストリームをNビット並列データとして復元するデシリアライザを含む装置。

#### 【請求項2】

前記高速相互接続ポートの各々がお互いに接続され、前記インタフェース装置を通じるデータ・フローが、1つの前記インタフェース・プロセッサから、前記相互接続ポートを介して、別の前記インタフェース・プロセッサに至る、請求項1記載の装置。

#### 【請求項3】

識別可能なアドレスから装置に送られて来るデータを、前記装置から識別アドレスに転送するように指示する自己ルーティング・スイッチング・ファブリック装置を含み、前記高速相互接続ポートの各々が前記スイッチング・ファブリック装置に接続され、前記装置を通過するデータ・フローが、1つの前記インタフェース・プロセッサから、前記スイッチング・ファブリック装置にインバウンド方向に送られ、前記スイッチング・ファブリック装置から、別の前記インタフェース・プロセッサを通じてアウトバウンド報告に送られる、請求項2記載の装置。

20

#### 【請求項4】

請求項1に記載のインタフェース装置に使用され制御点プロセッサと協働して動作するスケラブル・スイッチング・プロセッサであって、

N個( $N > 1$ )のプロセッサを含むプロセッサ・コンプレックスと、

30

複数の入出力ポートと、

前記入力ポートから受信されるイングレス・データを処理する少なくとも1つのアップサイド・メディア・アクセス制御(MAC)モジュールと、

アップパス内に配置され、前記プロセッサ・コンプレックス及びアップサイドMACに動作的に接続されるアップサイド・エンキュー/デキュー・スケジューラと、

ダウンパス内に配置され、前記出力ポートに接続されて、イーグレス・データを処理し、処理された前記データを前記出力ポートに転送する少なくとも1つのダウンサイドMACと、

前記ダウンパス内に配置され、前記プロセッサ・コンプレックス及びダウンサイドMACに動作的に接続されるダウンサイド・エンキュー/デキュー・スケジューラと、

40

前記アップサイド・エンキュー/デキュー・スケジューラ及びダウンサイド・エンキュー/デキュー・スケジューラに動作的に接続されるコネクタ・システムと

を含み、単一のネットワーク・プロセッサまたはM個( $M \geq 2$ )のネットワーク・プロセッサ内の前記入出力ポートが動作的に相互接続され、前記制御点プロセッサから伝送されるガイド・フレームの指令を使用してデータが前記単一のネットワーク・プロセッサまたはM個のネットワーク・プロセッサ内の前記入出力ポート間で交換される、スケラブル・ネットワーク・プロセッサ。

#### 【請求項5】

前記コネクタ・システムが、

並列入力及び直列出力を有し、アップサイドで接続される第1の変換モジュール及び第

50

2の変換モジュールと、

直列入力及び並列出力を有し、ダウンサイドで接続される第3の変換モジュール及び第4の変換モジュールと

を含む、請求項4記載のスケラブル・ネットワーク・プロセッサ。

【請求項6】

単一ネットワーク・プロセッサの入出力ポートが結合されるとき、前記第1または第2の変換モジュールの前記直列出力の少なくとも1つを、前記第3または第4の変換モジュールの前記直列入力の少なくとも1つに相互接続する伝送回線を含む、請求項5記載のスケラブル・ネットワーク・プロセッサ。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、様々なタイプ及び能力の情報処理システムまたはコンピュータをリンクするために使用される通信ネットワーク装置、及びこうした装置のコンポーネントに関する。特に、本発明は、こうした装置をアSEMBLするのに有用なスケラブル・スイッチ装置及びコンポーネントに関する。本発明はまた、改善された多機能インタフェース装置、並びに、こうした装置とメディア速度ネットワーク・スイッチを提供する他の要素との組み合わせに関する。本発明は更に、ネットワーク・スイッチのデータ・フロー処理能力を改善する、こうした装置の操作方法に関する。

【0002】

【従来の技術】

以下の説明は、ネットワーク・データ通信、並びにこうした通信ネットワークで使用されるスイッチ及びルータに関する知識を前提とする。特に、ネットワーク動作をレイヤに分割するネットワーク・アーキテクチャのISOモデルに精通していることを前提とする。ISOモデルにもとづく典型的なアーキテクチャは、物理パスまたはメディアであるレイヤ1(ときに"L1"と示される)から始まり、信号は、上流のレイヤ2、3、4などを通じてレイヤ7に至る。レイヤ7は、ネットワークにリンクされるコンピュータ・システム上で実行されるアプリケーション・プログラミング・レイヤである。本明細書では、L1、L2などは、ネットワーク・アーキテクチャの対応レイヤを指し示すものとする。本開示はまた、こうしたネットワーク通信においてパケット及びフレームとして知られる、ビット・ストリングの基本的理解を前提とする。

【0003】

今日のネットワーク化された世界では、帯域幅が重要な資源である。インターネットや他の新たなアプリケーションの登場による、ネットワーク・トラフィックの増加が、ネットワーク・インフラストラクチャの能力を圧迫している。こうした流れに追随するために、様々な組織や機関が、トラフィックの増加や音声とデータのコンバージェンスをサポートするための、より優れた技術や方法を検討している。

【0004】

ネットワーク・トラフィックの今日の劇的な増加は、インターネットの普及、情報へのリモート・アクセスの必要性の増加、及び新たなアプリケーションなどによる。インターネットは電子商取引の爆発的な成長により、ときにネットワーク・バックボーンにサポート不能な負荷を課する。データ・トラフィック量の増加の最も重要な1原因は、それが初めて音声トラフィックを超えることにある。電子メール、データベース・アクセス、及びファイル転送などを含む、リモート・アクセス・アプリケーションの要求の増加が、更にネットワークを圧迫している。

【0005】

音声とデータのコンバージェンスは、将来のネットワーク環境を定義する上で重要な役割を果たす。現在、インターネット・プロトコル(IP)ネットワークを介するデータの伝送は無料である。音声伝送は自然に最廉価の経路を通るので、音声は必然的にデータとコンバージされる。ボイス・オーバーIP(VOIP)、ボイス・オーバーATM(VOATM)

10

20

30

40

50

)、ボイス・オーバ・フレーム・リレー (V o F R) などの技術は、この変化の速い市場においてコスト効率の高い技術と言える。しかしながら、これらの技術への移行を可能にするために、業界は音声のサービス品質 (Q o S) を保証し、データ回線上での音声転送のためにどのように課金すべきかを決定しなければならない。1996年のTelecommunication Deregulation Actが、更にこの環境を複雑化する。この法規は、選択音声プロトコルすなわちA T Mと、選択データ・プロトコルすなわちI Pとの間の共生関係を強化する。

【0006】

新たな製品及び機能が使用可能になると、レガシー・システム (システム資産) を統合することが、組織にとって重大な問題となる。既存の設備及びソフトウェアへの投資を保護するために、組織は彼らの現在の運用を混乱させることなく、新たな技術への移行を可能にする解決策を要求する。

10

【0007】

サービス・プロバイダにとって、ネットワーク障害を取り除くことが先決である。ルータはしばしば、これらの障害の原因となる。しかしながら、ネットワーク輻輳は一般に、しばしば帯域幅問題と誤診され、より高い帯域幅を求めることにより取り込まれる。今日、メーカはこの難題を認識しつつある。彼らはネットワーク・プロセッサ技術に移行することにより、帯域幅資源をより効率的に管理し、ルータ及びネットワーク・アプリケーション・サーバで一般に見いだされる高度データ・サービスを有線速度で提供しつつある。これらのサービスにはロード・バランシング、Q o S、ゲートウェイ、ファイアウォール、セキュリティ、及びウェブ (W e b) キャッシングが含まれる。

20

【0008】

リモート・アクセス・アプリケーションでは、性能、要求時帯域幅、セキュリティ及び認証が最優先の部類に入る。Q o S及びC o Sの統合、統合音声処理、及びより複雑なセキュリティ・ソリューションが、将来のリモート・アクセス・ネットワーク・スイッチの設計を方向付けるであろう。更に、リモート・アクセスは、I S D N、T 1、E 1、O C - 3乃至O C - 4 8、ケーブル、及びx D S Lモデムなど、より多くの物理メディアに対応しなければならない。

【0009】

業界コンサルタントはネットワーク・プロセッサ (ここでは"N P"とも称す) を、次に挙げる機能の1つ以上を実行できるプログラマブル通信集積回路として定義した。すなわち

30

- 1) パケット分類: アドレスやプロトコルなどの既知の特性にもとづく、パケットの識別。
- 2) パケット変更: I P、A T M、または他のプロトコルに従う、パケットの変更 (例えばI Pのヘッダ内の存続時間 (time-to-live) フィールドの更新)。
- 3) キュー/ポリシ管理: 特定のアプリケーションにおけるパケットのパケット・キューイング、デキューイング、及びスケジューリングの設計方針の反映。
- 4) パケット転送: スイッチ・ファブリックを介するデータの伝送及び受信、及び適切なアドレスへのパケットの転送または経路指定。

【0010】

この定義は早期N Pの基本フィーチャの正確な記述であるが、N Pの完全な潜在的能力及び利点はまだ実現されていない。ネットワーク・プロセッサは、従来ソフトウェアで処理されたネットワーク・タスクをハードウェアで実行することにより、帯域幅を増加し、広範囲のアプリケーションにおける待ち時間問題を解決することができる。更に、N Pは並列分散処理及びパイプライン処理設計などのアーキテクチャを通じて、速度の改善を提供できる。これらの機能は効率的な検索エンジンを可能にし、スループットを向上させ、複雑なタスクの迅速な実行を提供する。

40

【0011】

ネットワーク・プロセッサは、P CにおけるC P Uのように、ネットワークの基本ネットワーク構築ブロックになるものと期待される。N Pにより提供される一般的な機能は、リ

50

アルタイム処理、セキュリティ、蓄積交換、スイッチ・ファブリック、及びIPパケット処理及び学習機能である。NPはISOレイヤ2乃至5をターゲットとし、ネットワーク特定タスクを最適化するように設計される。

【0012】

プロセッサ・モデルNPは、複数の汎用プロセッサ及び特殊論理を組み込む。提供者はこの設計に傾倒することにより、変更に適宜に且つ低廉に対応できる、スケーラブルで柔軟性のあるソリューションを提供しようとしている。プロセッサ・モデルNPは、低レベルの統合において分散処理を可能にし、より高度なスループット、柔軟性及び制御を提供する。プログラマビリティが、新たなASIC設計を要求することなく、新たなプロトコル及び技術への容易な移行を可能にする。プロセッサ・モデルNPにより、NEVは払戻し不能なエンジニアリング・コストの低減、及び製品化までの時間の短縮の利益を甘受することができる。

10

【0013】

【発明が解決しようとする課題】

本発明の第1の目的は、転送されるデータの処理速度を向上する一方で、サポート機能を各種の潜在的な要求にサイズ変更可能な、データ通信ネットワークで使用されるスケーラブル・スイッチ・アーキテクチャを提供することである。

【0014】

本発明の別の目的は、単一基板上に集積化され、レイヤ2、レイヤ3、レイヤ4及びレイヤ5を含むフレームのメディア速度切替えを提供するように協働する、複数のサブアセンブリを含むインタフェース装置またはネットワーク・プロセッサ(これらの用語は互換性をもって使用される)を提供することである。

20

【0015】

【課題を解決するための手段】

前記第1の目的は、関連処理ユニットの作業負荷から、従来に比較してより多くのデータ処理を削除するコンポーネント、及びコンポーネントの集合を提供することにより実現される。

【0016】

インタフェース装置は、ワーク・グループ・スイッチとして、第1レベルの機能を提供する独立型のソリューションとして使用されるか、またはより高度な機能のワーク・グループ・スイッチを提供する相互接続ソリューションとして使用されるか、或いは、スイッチング・ファブリック装置との協働により、更に機能向上に向けてスケールアップされる。

30

【0017】

【発明の実施の形態】

【数1】

$$Y/\overline{X}$$

40

は、本明細書ではY/Xバーと記載する。

【0018】

本発明は、本発明の好適な実施例を示す添付の図面を参照して、以下で詳述されるが、説明の冒頭に当たり、当業者であれば、本発明の好適な結果を達成するように、ここで述べる発明を変更することができよう。従って、以下で述べる説明は本発明を制限するものではなく、当業者に対する広範囲の教示の開示として理解されるべきである。

【0019】

ここで開示される装置はスケーラブルで、デスクトップまたはワークグループ・スイッチを相互接続し、こうしたスイッチをネットワーク・バックボーンに統合し、バックボーン・スイッチング・サービスを提供するように機能することができる。この装置はレイヤ2

50

、レイヤ3、及びレイヤ4 + 転送をハードウェアでサポートできる。特定の形態の装置は、デスクトップまたはワークグループ・スイッチ統合のために設計され、また他のものはコア・バックボーン・スイッチとして設計される。

#### 【0020】

装置のために使用されるアーキテクチャは、インタフェース装置またはネットワーク・プロセッサ・ハードウェア・システムと、制御点上で実行されるソフトウェア・ライブラリとにもとづき、これについては本明細書の中で詳述される。インタフェース装置またはネットワーク・プロセッサ・サブシステムは、L2、L3及びL4 + プロトコル・ヘッダの構文解析及び変換のために設計される高性能フレーム転送エンジンである。これはプロトコルがハードウェアの使用により、より高速にスイッチされることを可能にする。インタフェース装置またはネットワーク・プロセッサ・サブシステムは、ボックスを通じて高速パスを提供する一方、ソフトウェア・ライブラリ及び制御点プロセッサは、高速パスを維持するために必要な管理及びルート発見機能を提供する。制御点プロセッサ及びそれ上で動作するソフトウェア・ライブラリは、一緒にシステムの制御点(CP)を定義する。CPでは、透過型ブリッジング及びOSPFなどの、実際のブリッジング及びルーティング・プロトコルが実行される。これはシステムの低速パスとも呼ばれる。

10

#### 【0021】

ここで開示される装置は、マルチレイヤ転送をハードウェアでサポートするが、装置はL2専用スイッチとしても動作でき、開示される最も単純な形態では、これがそのデフォルト動作モードである。各ポートは単一のドメインに配置され、任意の装置が任意の他の装置と通信することを可能にする。装置はL2において構成可能であり、システム管理者に次のようなフィーチャ、すなわち、ポートを別々のドメインまたはトランクにグループ化したり、仮想LAN(VLAN)セグメントや、ブロードキャスト及びマルチキャスト・トラフィックを制御するフィルタを構成するなどの能力を提供する。

20

#### 【0022】

このスケーラブル装置は多くの利点を有する。第1に、これはシステム管理者に、L2で使用されるハードウェアと同一のものを使用して、同一の速度で、IP及びIPXトラフィックのL3転送及びルーティングを構成する能力を提供する。第2に、構内建物を相互接続するために、外部ルータを使用する必要性を排除する一方で、同時に性能を向上させる。第3に、L2/L3サービスの管理を単純化または結合することにより、これらを単一の制御点に組み込む。最後に、スケーラブル装置は付加価値機能をL4 + 機能と共に提供し、これはサーバ間での負荷平準化を目的として、主幹業務のアプリケーション及びネットワーク・ディスパッチャをサポートするために、システム管理者が異なるトラフィック分類を割当てて能力を提供する。

30

#### 【0023】

スケーラブル装置は、インタフェース装置またはネットワーク・プロセッサを使用するモジュラ・ユニットとして、または制御点(CP)として、或いはその基本ビルディング・ブロックとしての、オプションのスイッチング・ファブリック装置として設計される。インタフェース装置は好適には、L2/L3/L4 + 高速パス転送サービスを提供し、CPは高速パスを維持するために必要な管理及びルート発見機能を提供する。3つ以上のインタフェース装置サブシステムと一緒に結合される場合には、オプションのスイッチング・ファブリック装置が使用される。オプションのスイッチング・ファブリック装置は、1991年4月16日発行の米国特許第5008878号"High Speed Modular Switching Apparatus for Circuit and Packet Switched Traffic"で開示される。

40

#### 【0024】

装置は、プリント回路基板要素を用いて組み立てられるものと見込まれる。これはここでは"ブレード"とも称される。プリント回路基板要素は回路要素を実装され、装置ハウジング内に設けられるコネクタに収容される。同様の装置は"オプション・カード"としても知られる。装置は、適切なコネクタ及びバックプレーン電気接続が設けられる条件の下で、ブレードが様々なシャーシまたはハウジングの間で交換されることを考慮する。全てのブ

50

レード上で見いだされる基本コンポーネントは、キャリア・サブシステムである。キャリア・サブシステムから始まり、3つのタイプのブレードが生成される。第1のタイプはCP専用ブレードであり、これはキャリア・サブシステム及びCPサブシステムから成る。CP専用ブレードは、冗長性が一番の関心事である製品に対して主に使用される。第2のタイプはCP+メディア・ブレードであり、これはキャリア・サブシステム、CPサブシステム、及び1対3メディア・サブシステムから成る。CP+メディア・ブレードは、冗長性よりもポート密度が重要視される製品に対して、主に使用される。第3のタイプはメディア・ブレードであり、これはキャリア・サブシステムと1対4メディア・サブシステムとから成る。メディア・ブレードは任意のシャーシ内で使用され、使用されるメディア・サブシステムのタイプは構成可能である。

10

**【0025】**

ブレード管理は障害検出、電力管理、新たな装置の検出、初期化及び構成を含む。この管理は様々なレジスタ、入出力信号、及びCPとキャリア・サブシステムとの間で通信するために使用されるガイド・セル・インタフェースを用いて行われる。しかしながら、シャーシと異なり、全てのブレード上には、プログラマブル装置及びメモリが存在する。プログラマビリティは、ブレードのタイプに依存する。CPサブシステムがブレード上に存在する場合、CP及びキャリア・サブシステムの両方がプログラマブルである。メディア・サブシステムもプログラマブルであるが、キャリア・サブシステムを通じて間接的にプログラマブルなだけである。

**【0026】**

20

高機能製品では更に、スイッチング・ファブリック装置サブシステムを含むスイッチ・ブレードが存在する。このブレードの管理は、故障検出、電力管理、新たな装置の検出、及び初期化を含む。この管理は、CPサブシステム内にマップされる様々なレジスタ及び入出力信号を用いて行われる。

**【0027】**

最も単純な形態では、本発明により考慮されるスイッチ装置は制御点プロセッサと、制御点プロセッサに動作的に接続されるインタフェース装置とを有する。好適には、ここで開示されるように、インタフェース装置(ネットワーク・プロセッサとして知られる)は、半導体基板と、基板上に形成される複数のインタフェース・プロセッサと、インタフェース・プロセッサによりアクセスされる命令を記憶する、前記基板上に形成される内部命令メモリと、インタフェース装置を通過し、インタフェース・プロセッサによりアクセスされるデータを記憶する、前記基板上に形成される内部データ・メモリと、複数の入出力ポートとを有する、単一の超大規模集積回路(VLSI)装置またはチップである。インタフェース・プロセッサはここではときに、ピコプロセッサまたは処理ユニットと称される。提供されるポートは、内部データ・メモリを外部データ・メモリに接続する少なくとも1つのポートと、インタフェース・プロセッサの指示の下で、インタフェース装置を通過するデータを外部ネットワークと交換する少なくとも2つの他のポートとを含む。制御点プロセッサはインタフェース装置と協働して、インタフェース・プロセッサにより実行される命令を命令メモリにロードする。そして、これらの命令が実行されて、入出力ポート間でのデータの交換、及びデータ・メモリを介するデータの流れを指示する。

30

40

**【0028】**

ここで開示されるネットワーク・プロセッサは、それが組み込まれるスイッチ・アセンブリとは切り離して、本発明に関わる。更に、ここで開示されるネットワーク・プロセッサは、ここで述べられるその要素内に、ここでは詳細に述べられない他の発明を有するものとみなされる。

**【0029】**

図1は、基板10と、基板上に集積化される複数のサブアセンブリとを含むインタフェース素子チップのブロック図である。サブアセンブリは、アップサイドすなわち上流側構成と、ダウンサイドすなわち下流側構成とに編成される。本明細書では、"アップサイド"は、ネットワークから本明細書で開示される装置に送られるデータ・フローを指し示し、"

50

ダウンサイド"は、逆に装置から、この装置によりサービスされるネットワーク・サービスに送られるデータを指し示す。データ・フローは反復構成に従う。結果的に、アップサイド・データ・フロー及びダウンサイド・データ・フローが存在する。アップサイドのサブアセンブリには、エンキュー・デキュー・スケジューリング・アップ(EDS-UP)論理16、多重化MACアップ(PPM-UP)14、スイッチ・データ・ムーバ・アップ(SDM-UP)18、システム・インタフェース(SIF)20、データ・アライン・シリアル・リンクA(DASLA)22、及びデータ・アライン・シリアル・リンクB(DASLB)24が含まれる。データ・アライン・シリアル・リンクは、1999年6月11日出願の米国特許出願第09/330968号"High Speed Parallel/Serial Link for Data Communication"で詳細に述べられている。ここで開示される本発明の装置の好適な形態はDASLリンクを使用するが、本発明は、特にデータ・フローがVLSI構造内にあるように制限される場合などのように、かなり高いデータ・フロー・レートを達成するために、他の形態のリンクが使用されることも考慮する。

#### 【0030】

ダウンサイドのサブアセンブリには、DASLA26、DASLB28、SIF30、SDM-DN32、EDS-DN34、及びPPM-DN36が含まれる。チップは更に複数の内部SRAM、トラフィック管理スケジューラ40、及び組み込みプロセッサ・コンプレックス(EPC)12を含む。インタフェース装置38は、イーサネット(R)フィジカル(ENETPHY)またはATMフレームなどの、任意の好適なL1回路である。インタフェースのタイプは、部分的に、チップが接続されるネットワーク・メディアにより決定される。複数の外部DRAM及びSRAMが設けられ、チップにより使用される。

#### 【0031】

ここでは特に、関連スイッチング及びルーティング装置の外部の一般データ・フローが、建物内に導入された配線やケーブルなどの電気導体を通過するネットワークが開示されるが、本発明は、ここで開示されるネットワーク・スイッチ及びコンポーネントが、無線環境において使用されることも同様に考慮する。例として、ここで述べられるメディア・アクセス制御(MAC)要素は、ここで述べられる要素を直接ワイヤレス・ネットワークにリンクする機能を有する、好適な無線周波数要素により置換されてもよく、例えば、既知のシリコン・ゲルマニウム技術などが使用される。こうした技術が適切に使用される場合には、無線周波数要素は当業者により、ここで開示されるVLSI構造内に統合される。或いは、無線周波数、または赤外線応答装置などの他のワイヤレス応答装置が、ワイヤレス・ネットワーク・システム用のスイッチ装置を実現する他の要素と一緒に、ブレード上に実装されてもよい。

#### 【0032】

図中の矢印は、インタフェース装置内のデータの一般的な流れを示す。イーサネット(R)MACから受信されるフレームは、EDS-UPにより、内部データ・ストア・バッファに配置される。これらのフレームは、通常データ・フレームまたはシステム制御ガイド・フレーム(Guided Frame)として示され、EPC(図1)にエンキューされる。EPCは、最大Nフレーム( $N > 1$ )を並列に処理できるN個のプロトコル・プロセッサを含む。10プロトコル・プロセッサ(図14)の実施例では、10個のプロトコル・プロセッサの内の2個が特殊化される。すなわち、1つはガイド・フレームを処理するように(汎用中央ハンドラ(GCH))、また他は制御メモリ内にルックアップ・データを作成するように(汎用ツリー・ハンドラ(GTH))、特殊化される。図13に示されるように、EPCは、新たなフレームをアイドル・プロセッサに突き合わせるディスパッチャと、フレーム・シーケンスを保持する完了(completion)ユニットと、10個の全プロセッサにより共用される共通命令メモリと、フレーム分類を決定する分類子ハードウェア補助機構及びフレームの開始命令アドレスを決定する支援をするコプロセッサと、フレーム・バッファの読み出し及び書き込み操作を制御する入口側(以降イングレス(ingress)と称す)及び出口側(以降イーグレス(egress)と称す)データ・ストア・インタフェースと、10

10

20

30

40

50



個のプロセッサが制御メモリを共用することを可能にする制御メモリ・アービタと、ウェブ(Web)制御と、内部インタフェース装置データ構造へのデバッグ・アクセスを可能にするアービタ及びインタフェースと、その他のハードウェア構造を含む。

【0033】

ガイド・フレームは、GCHプロセッサが使用可能になると、ディスパッチャによりGCHプロセッサに送信される。レジスタ書込み、カウンタ読出し、イーサネット(R)MAC構成変更など、ガイド・フレーム内でエンコードされる操作が実行される。MACまたはIPエントリの追加など、ルックアップ・テーブル変更が、メモリ読出しや書込みなどの制御メモリ操作のために、ルックアップ・データ・プロセッサに渡される。MIBカウンタ読出しなどの幾つかのコマンドは、応答フレームの作成、及び適切なインタフェース装置上の適切なポートへの転送を要求する。ときとして、ガイド・フレームはインタフェース装置のイーグレス側に合わせてエンコードされる。これらのフレームは、照会されるインタフェース装置のイーグレス側に転送され、これが次にエンコード操作を実行し、適切な応答フレームを作成する。

10

【0034】

データ・フレームは、次の使用可能なプロトコル・プロセッサにディスパッチされ、そこでフレーム参照が実行される。フレーム・データは、分類子ハードウェア補助機構(CHA)エンジンからの結果と一緒に、プロトコル・プロセッサに渡される。CHAはIPまたはIPXを構文解析する。結果がツリー構造探索アルゴリズム及び開始共通命令アドレス(CIA)を決定する。サポートされるツリー構造探索アルゴリズムは、固定マッチ・ツリー(レイヤ2イーサネット(R)MACテーブルなど、正確なマッチを要求する固定サイズ・パターン)、最長プレフィックス・マッチ・ツリー(サブネットIP転送など、可変長マッチを要求する可変長パターン)、及びソフトウェア管理ツリー(フィルタ規則で使用されるような、範囲またはビット・マスク・セットのいずれかを定義する2つのパターン)を含む。

20

【0035】

ルックアップ(探索)は、各プロトコル・プロセッサの一部である、ツリー構造探索エンジン(TSE)コプロセッサの支援により実行される。TSEコプロセッサは制御メモリ・アクセスを実行し、プロトコル・プロセッサを解放して実行を継続する。制御メモリは全てのテーブル、カウンタ、及びピココードにより必要とされる他のデータを記憶する。制御メモリ操作は、10個のプロセッサ・コンプレックスの間のメモリ・アクセスを調停する制御メモリ・アービタにより管理される。

30

【0036】

フレーム・データはデータ・ストア・コプロセッサを通じてアクセスされる。データ・ストア・コプロセッサは、基本データ・バッファ(フレーム・データの最大8個の16バイト・セグメント)と、スクラッチ・パッド・データ・バッファ(フレーム・データの最大8個の16バイト・セグメント)と、データ・ストア操作のための幾つかの制御レジスタとを含む。一旦マッチが見いだされると、イングレス・フレーム変更がVLANヘッダ挿入またはオーバーレイを含み得る。この変更は、インタフェース装置プロセッサ・コンプレックスにより実行されるのではなく、ハードウェア・フラグが導出され、他のイングレス・スイッチ・インタフェース・ハードウェアが変更を実行する。他のフレーム変更はピココード及びデータ・ストア・コプロセッサにより、イングレス・データ・ストア内に保持されるフレーム内容を変更することにより達成される。

40

【0037】

他のデータは、フレームをスイッチ・ファブリック装置に送信する前に収集され、スイッチ・ヘッダ及びフレーム・ヘッダを作成するために使用される。制御データには、フレームの宛先ブレードなどのスイッチ情報や、イーグレス・インタフェース装置の情報が含まれ、宛先ポートのフレーム探索、マルチキャストまたはユニキャスト操作、及びイーグレス・フレーム変更を推進する。

【0038】

50

完了時、エンキュー・コプロセッサが、フレームをスイッチ・ファブリックにエンキューするのに必要なフォーマットを作成し、それらを完了ユニットに送信する。完了ユニットは、10個のプロトコル・プロセッサからスイッチ・ファブリック・キューへのフレーム順序を保証する。スイッチ・ファブリック・キューからのフレームは、64バイト・セルにセグメント化され、それらがPrizma-Eスイッチに伝送されるとき、フレーム・ヘッダ・バイト及びスイッチ・ヘッダ・バイトが挿入される。

**【0039】**

スイッチ・ファブリックから受信されたフレームは、イーグレスEDS34により、イーグレス・データ・ストア（イーグレスDS）バッファに配置され、EPCにエンキューされる。フレームの一部はディスパッチャにより、アイドル・プロトコル・プロセッサに送信され、フレーム探索が実行される。フレーム・データは、分類子ハードウェア補助機構からのデータと一緒に、プロトコル・プロセッサにディスパッチされる。分類子ハードウェア補助機構は、イーグレス・インタフェース装置により作成されるフレーム制御データを使用して、開始コード命令アドレス（CIA）の決定を支援する。

10

**【0040】**

イーグレス・ツリー探索は、イーグレス探索においてサポートされるのと同じのアルゴリズムをサポートする。ルックアップはTSEコプロセッサにより実行され、プロトコル・プロセッサを解放して実行を継続する。全ての制御メモリ操作は、制御メモリ・アービタにより管理され、アービタは10個のプロセッサ・コンプレックスの間で、メモリ・アクセスを割り当てる。

20

**【0041】**

イーグレス・フレーム・データは、データ・ストア・コプロセッサを通じてアクセスされる。データ・ストア・コプロセッサは基本データ・バッファ（フレーム・データの最大8個の16バイト・セグメント）と、スクラッチ・パッド・データ・バッファ（フレーム・データの最大8個の16バイト・セグメント）と、データ・ストア操作のための幾つかの制御レジスタとを含む。探索の成功結果は転送情報を含み、ときとしてフレーム変更情報を含む。フレーム変更はVLANヘッダ検出、存続時間増分（IPX）または減分（IP）、IPヘッダ・チェックサム再計算、イーサネット（R）フレームCRCオーバーレイまたは挿入、及びMAC DA/SAオーバーレイまたは挿入を含む。IPヘッダ・チェックサムはチェックサム・コプロセッサにより用意される。変更はインタフェース装置プロセッサ・コンプレックスにより実行されるのではなく、ハードウェア・フラグが作成され、PMMイーグレス・ハードウェアが変更を実行する。完了時に、エンキュー・コプロセッサが、フレームをEDSイーグレス・キューにエンキューし、それらを完了ユニットに送信するために必要なフォーマットを作成する。完了ユニットは、10個のプロトコル・プロセッサから、EDSイーグレス・キューへのフレーム順序を保証して、イーグレス・イーサネット（R）MAC36に供給する。

30

**【0042】**

完了フレームは最終的に、PMMイーグレス・ハードウェアによりイーサネット（R）MACに送信され、イーサネット（R）ポートから出力される。

**【0043】**

ウェブと呼ばれる内部バスは、内部レジスタ、カウンタ及びメモリへのアクセスを可能にする。ウェブ（Web）はまた、命令ステップや、デバッグ及び診断のための割込み制御を制御する外部インタフェースを含む。

40

**【0044】**

ツリー構造探索エンジン・コプロセッサは、メモリ範囲チェック及び不当メモリ・アクセス通知を提供し、ツリー構造探索命令（メモリ読み出し、書き込み、またはread-add-writeなど）を、プロトコル・プロセッサ実行と並列に実行する。

**【0045】**

共通命令メモリは、1つの1024×128RAMと、2セットのデュアル512×128RAMとから構成される。デュアルRAMの各セットは、同一のピココードの2つのコ

50

ピーを提供し、プロセッサによる同一アドレス範囲内の命令への独立したアクセスを可能にする。各128ビット・ワードは4つの32ビット命令を含み、合計8192命令を提供する。

【0046】

ディスパッチャは、10個のプロトコル・プロセッサへのフレームの受け渡しを制御し、割込み及びタイマを管理する。

【0047】

完了ユニットは、プロセッサ・コンプレックスから、スイッチ・ファブリック及びターゲット・ポート・キューへのフレーム順序を保証する。豊富な命令セットには、条件付き実行、(入力ハッシュ・キーの)パッキング、条件付き分岐、符号付き及び符号無し演算、先行0のカウントなどが含まれる。

10

【0048】

分類子ハードウェア補助機構エンジンは、各フレームのレイヤ2及びレイヤ3プロトコル・ヘッダを構文解析し、フレームがプロトコル・プロセッサにディスパッチされるとき、この情報をフレームに提供する。

【0049】

制御メモリ・アービタは、内部及び外部メモリの両方へのプロセッサ・アクセスを制御する。

【0050】

外部制御メモリ・オプションは、5個乃至7個のDDR DRAMサブシステムを含む。各サブシステムは、1対の2M×16ビット×4バンクDDR DRAM、または1対の4M×16ビット×4バンクDDR DRAMをサポートする。DDR DRAMインターフェースは、133MHzクロック・レート及び266MHzデータ・ストロークで動作し、構成可能なCAS待ち時間及びドライブ強度をサポートする。オプションの133MHz ZBT SRAMが、128K×36、2×256K×18、または2×512K×18構成のいずれかで追加される。

20

【0051】

イーグレス・フレームは1つの外部データ・バッファ(例えばDS0)、または2つの外部データ・バッファ(DS0及びDS1)のいずれかに記憶される。各バッファは、1対の2M×16ビット×4バンクDDR DRAM(最大256Kの64バイト・フレームを記憶可能)、または1対の4M×16ビット×4バンクDDR DRAM(最大512Kの64バイト・フレームを記憶可能)から構成される。2.28Mbpsの単一の外部データ・バッファ(例えばDS0)を選択するか、第2のバッファ(例えばDS1)を追加し、4.57Mbpsのレイヤ2及びレイヤ3スイッチングをサポートする。第2のバッファの追加は性能を改善するが、フレーム容量を増加しない。外部データ・バッファ・インターフェースは、133MHzクロック・レート、及び266MHzのデータ・ストロークで動作し、構成可能なCAS待ち時間及びドライブ強度をサポートする。

30

【0052】

内部制御メモリは、2つの512×128ビットRAMと、2つの1024×36ビットRAMと、1つの1024×64ビットRAMとを含む。

40

【0053】

内部データ記憶は、イーグレス方向(UP)に、最大2048個の64バイト・フレームをバッファリングすることができる。

【0054】

固定フレーム変更は、イーグレス方向のVLANタグ挿入と、VLANタグ削除と、持続時間増分/減分(IP、IPX)と、イーグレス(DOWN)方向のイーサネット(R)CRCオーバーレイ/挿入及びMAC DA/SAオーバーレイ/挿入とを含む。

【0055】

ポート・ミラーリングは、プロトコル・プロセッサ資源を用いることなく、1つの受信ポート及び1つの送信ポートを、システム指定の観測ポートにコピーすることを可能にする

50

。ミラーリングされたインタフェース装置ポートは、フレーム及びスイッチ制御データを追加するように構成される。別々のデータ・パスが、インGRESS・スイッチ・インタフェースへの直接フレーム・エンキューイングを可能にする。

【0056】

インタフェース装置は、4つのイーサネット(R)マクロを統合する。各マクロは、1ギガビットまたは10/100高速イーサネット(R)モードのいずれかで動作するように、個々に構成される。各イーサネット(R)マクロは、最大10個の10/100Mbps MACが、4つのマクロの各々に対して、1つの1000Mbps MACをサポートする。

【0057】

図2は、MACコアのブロック図を示す。各マクロは、3つのイーサネット(R)コア設計、すなわちマルチポート10/100Mbps MACコア(FENET)、1000Mbps MACコア(GENET)、及び100Mbps物理コーディング・サブレイヤ・コア(PCSC)を含む。

【0058】

マルチポート・イーサネット(R)10/100MACフィーチャ:

- 1) 物理レイヤとの10個のシリアル・メディア独立インタフェースをサポートする。
- 2) 10Mbpsまたは100Mbpsメディア速度の10個のポートを、任意に混在させて処理できる。
- 3) 単一のMACが時分割多重インタフェースにより、10個の全てのポートをサービスする。
- 4) 全てのポート上で全2重/半2重動作をメディア速度でサポートする。
- 5) IEEE802.3バイナリ指数バックオフをサポートする。

【0059】

1000Mbpsイーサネット(R)MACコア・フィーチャ:

- 1) 物理PCSCレイヤとの、または直接的に物理レイヤとのギガビット・メディア独立インタフェース(GMII)をサポートする。
- 2) PCSCコアにより、完全なTBI(8b/10b)ソリューションをサポートする。
- 3) 全2重Point-to-Point接続をメディア速度でサポートする。
- 4) IBM PCSCコア有効バイト信号方式をサポートする。

【0060】

1000Mbpsイーサネット(R)物理コーディング・サブレイヤ・コア・フィーチャ:

- 1) 8b/10bエンコード及びデコードを実行する。
- 2) IEEE802.3zで定義されるPMA(10ビット)サービス・インタフェースをサポートする。このインタフェースは、IEEE802.3zに準拠する任意のPMAに接続する。
- 3) PMAから受信されるデータ(2フェーズ・クロック)を、MAC(1フェーズ)クロックに同期させる。
- 4) 次の2ページを含むオートネゴシエーションをサポートする。
- 5) 規格で定義された2フェーズ・クロック・システムを、1フェーズ・クロックに変換する。
- 6) 新たなデータを含むクロック・サイクルを示す信号を、MACに提供する。
- 7) 受信コード・グループ(10ビット)内のCOMMAをチェックし、ワード同期を確立する。
- 8) 8b/10b実行中ディスペリティを計算及びチェックする。

【0061】

図3の(A)乃至(D)は、インタフェース素子チップの異なる構成を示す。これらの構成はDASLと、スイッチング・ファブリック装置への接続とによりサポートされる。各DASLは2つのチャンネル、すなわち送信チャンネルと受信チャンネルとを含む。

10

20

30

40

50

## 【 0 0 6 2 】

図 3 の ( A ) は、単一インタフェース装置のラップ ( wrap ) 構成を示す。この構成では、送信チャンネルが受信チャンネルにラップ、すなわち折り返される。

## 【 0 0 6 3 】

図 3 の ( B ) は、2 つのインタフェース素子チップが接続される構成を示す。各インタフェース素子チップは、少なくとも 2 つの D A S L を提供される。この構成では、1 つのチップ上の 1 D A S L 上のチャンネルが、他のチップ上のマッピング D A S L のチャンネルに、動作的に接続される。各チップ上の他の D A S L は、ラップされる。

## 【 0 0 6 4 】

図 3 の ( C ) は、複数のインタフェース装置がスイッチ・ファブリックに接続される構成を示す。両頭矢印は両方向の伝送を示す。 10

## 【 0 0 6 5 】

図 3 の ( D ) は、メイン・スイッチ及びバックアップ・スイッチが複数のインタフェース装置に接続される構成を示す。メイン・スイッチが故障すると、バックアップ・スイッチが使用可能になる。

## 【 0 0 6 6 】

制御点 ( C P ) はシステム・プロセッサを含み、これは各構成に接続される。とりわけ、C P のシステム・プロセッサは、チップに対して、初期化及び構成サービスを提供する。C P は 3 つの位置、すなわちインタフェース素子チップ内か、チップが実装されるブレード上か、或いはブレードの外部のいずれかに配置される。ブレードの外部の場合、C P はリモートとなり、どこか別の場所に内蔵され、インタフェース装置及び C P が接続されるネットワークを介して通信する。C P の要素が図 2 0 に示され、メモリ素子 ( キャッシュ、フラッシュ及び S D R A M )、メモリ制御装置、P C I バス、及びバックプレーン及び L 1 ネットワーク・メディア用のコネクタを含む。 20

## 【 0 0 6 7 】

図 2 1 は、単一チップ・ネットワーク・プロセッサと、E D S - U P、トラフィック管理 ( M G T ) スケジューラ、及び E D S - D O W N ( D N ) により提供される機能とを示す。U 字形アイコンはキューを表し、キュー内の内容を追跡する制御ブロック ( C B ) は、矩形アイコンにより表される。

## 【 0 0 6 8 】

各要素及びそれらの機能及び相互作用は、次の通りである。 30

## 【 0 0 6 9 】

P M M : これは M A C ( F E n e t、P O S、G E n e t ) を含み、外部 P H Y 装置に接続されるネットワーク・プロセッサの一部である。

## 【 0 0 7 0 】

U P - P M M : この論理は P H Y からバイトを受け取り、それを F I S H ( 1 6 バイト ) にフォーマットし、U P - E D S に受け渡す。P M M 内には 4 つの D M U が存在し、各々は 1 G E n e t または 1 0 F E n e t 装置と連携することができる。

## 【 0 0 7 1 】

U P - E D S : この論理は U P - P M M から F I S H を受け取り、それらを U P - D A T A ストア ( 内部 R A M ) に記憶する。これは 1 度に 4 0 フレームを処理することができ、適切な数のバイトが受信されると、フレームを E P C にエンキューする。E P C がフレーム処理を完了すると、U P - E D S がフレームを適切なターゲット・ポート・キューにエンキューし、U P - S D M にフレームの送信を開始する。U P - E D S は全てのバッファ及びフレーム管理の責任を負い、U P - S D M への転送が完了するとき、バッファ/フレームを空きプールに戻す。 40

## 【 0 0 7 2 】

E P C : この論理はピコプロセッサを含み、組み込み P o w e r P C を含み得る。この論理はフレーム・ヘッダを突き止め、フレームをどのように処理すべきかを決定する ( 転送、変更、フィルタリングなど )。E P C は幾つかのルックアップ・テーブルをアクセスで 50

き、ピコプロセッサがネットワーク・プロセッサの高帯域幅要求に応じるためのハードウェア補助機構を有する。

【 0 0 7 3 】

UP - SDM : この論理はフレームを受け取り、それらをスイッチ・ファブリックへの伝送のために、Prizmaセルにフォーマットする。この論理はVLANヘッダをフレームに挿入することができる。

【 0 0 7 4 】

UP - SIF : この論理はUP - DASLマクロを含み、外部スイッチ入出力(I/O)に接続する。

【 0 0 7 5 】

DN - SIF : この論理はDN - DASLマクロを含み、外部スイッチ入出力(I/O)からPrizmaセルを受信する。

【 0 0 7 6 】

DN - SDM : この論理はPrizmaセルを受信し、フレーム再組み立てのために、それらを事前処理する。

【 0 0 7 7 】

DN - EDS : この論理は各セルを受け取り、それらを再度フレームに組み立てる。セルは外部データ・ストアに記憶され、バッファと一緒にリンクされて、フレームを形成する。全フレームを受信されると、フレームはEPCにエンキューされる。EPCがフレーム処理を終えると、フレームはスケジューラ・キュー(但し存在する場合)またはターゲット・ポート・キューにエンキューされる。DN - EDSは次に、フレーム、変更情報、及び制御情報などをDN - PMMに送信することにより、フレームを適切なポートに送信する。

【 0 0 7 8 】

DN - PMM : DN - EDSから情報を受け取り、フレームをイーサネット(R)、POSなどにフォーマットし、フレームを外部PHYに送信する。

【 0 0 7 9 】

SPM : この論理は、ネットワーク・プロセッサが外部装置(PHY、LED、FLASHなど)とインタフェースするために使用されるが、3つの入出力(I/O)を要求するだけである。ネットワーク・プロセッサはシリアル・インタフェースを用いて、SPMと通信し、次にSPMがこれらの外部装置を管理するために必要な機能を実行する。

【 0 0 8 0 】

アップサイド・フロー :

- 1) フレームがPHYに到来する。
- 2) バイトがUP - PMMにより受信される。
- 3) UP - PMMがFISHをUP - EDSに送信する(FISHはフレームの一部を意味する)。
- 4) UP - EDSがFISHをUP - DSに記憶する。
- 5) UP - EDSがヘッダをEPC送信する。
- 6) EPCがヘッダを処理し、エンキュー情報をUP - EDSに返送する。
- 7) UP - EDSがフレームの残りをUP - PMMから受信し続ける。
- 8) スイッチへの適切なデータの送信準備が整うと、UP - EDSが情報をUP - SDMに送信する。
- 9) UP - SDMがフレーム・データを読み出し、それをPrizmaセルにフォーマットする。
- 10) UP - SDMがセルをUP - SIFに送信する。
- 11) UP - SIFがDASLシリアル・リンクを介して、セルをPrizmaに転送する。
- 12) 全てのデータが受け取られると、UP - EDSがバッファ/フレームを解放する。

【 0 0 8 1 】

10

20

30

40

50

ダウンサイド・フロー：

- 1) DN - S I F が P r i z m a セルを受信する。
- 2) DN - S D M がセルを記憶し、それらを再組み立て情報として事前処理する。
- 3) DN - E D S がセル・データ及び再組み立て情報を受信し、セルをダウンサイド側の新たなフレームにリンクする。
- 4) DN - E D S がセルを DN - D S に記憶する。
- 5) 全てのデータが受信されると、DN - E D S がフレームを E P C にエンキューする。
- 6) E P C がヘッダを処理し、エンキュー情報を DN - E D S に返送する。
- 7) DN - E D S がフレームをスケジューラ・キュー（但し存在する場合）またはターゲット・ポート・キューにエンキューする。
- 8) DN - E D S がキューをサービスし、フレーム情報を P C B に送信する。
- 9) DN - E D S が P C B を用いてフレームを"解体"（unravel）し、適切なデータを読み出し、そのデータを DN - P M M に送信する。
- 10) DN - P M M がデータをフォーマットし（必要に応じて変更を加える）、フレームを外部 P H Y に送信する。
- 11) DN - P M M が DN - E D S に、バッファがもはや必要とされないことを通知し、DN - E D S がこれらの資源を解放する。

【 0 0 8 2 】

フレーム制御フロー：

- 1) ヘッダが U P - D S または DN - D S から E P C に送信される。
- 2) E P C がルックアップ・テーブルでヘッダ情報を調査し、フレーム・エンキュー情報を受信する。
- 3) E P C がエンキュー情報を E D S に返送し、フレームが適切なキューにエンキューされる。
- 4) セル・ヘッダ及びフレーム・ヘッダがフレーム・データと一緒に送信され、再組み立て及びフレーム転送を支援する。

【 0 0 8 3 】

C P 制御フロー：

- 1) 制御点がガイド・フレームをフォーマットし、それをネットワーク・プロセッサに送信する。
- 2) ネットワーク・プロセッサがガイド・フレームを、G C H ピコプロセッサにエンキューする。
- 3) G C H がガイド・フレームを処理し、Rainierの要求領域を読み書きする。
- 4) G C H がテーブル更新要求を G T H に渡す。
- 5) G T H が適切なテーブルを、ガイド・フレームからの情報により更新する。
- 6) 肯定応答ガイド・フレームが C P に返送される。

【 0 0 8 4 】

ネットワーク・プロセッサ制御フロー：

- 1) ピコプロセッサがガイド・フレームを作成し、情報を別のRainierまたは制御点に送信する。
- 2) ガイド・フレームが適切な位置に送信され処理される。

【 0 0 8 5 】

単一インタフェース装置は、最大 40 の高速イーサネット（R）ポート（図 3 の（A）参照）のための、メディア速度切替えを提供する。I B M のデータ整合同期リンク（D A S L : Data Aligned Synchronous Link）技術を用いて、2つのインタフェース装置が相互接続される場合、80個の高速イーサネット（R）ポートがサポートされる（図 3 の（B）参照）。各 D A S L 差動対は、440 M b p s のデータを伝搬する。従って、8対の2つのセットが、3.5 G b p s 全2重接続（各方向において440 M b p s の8倍）を提供する。図 3 の（C）及び（D）に示されるように、複数のインタフェース装置を、I B M の P r i z m a - E スイッチなどのスイッチに相互接続することにより、より大規模な

10

20

30

40

50

システムが構成される。インタフェース装置は、2つの3.5 Gbps全2重DASL接続を提供し、これらの一方が基本接続で他が補助接続であり、後者は、ローカル・フレーム・トラフィックのためのラップ・バックパス(wrap-backpath)を提供するか(2つのインタフェース装置が直接接続される場合(図3の(B))、或いは、冗長スイッチ・ファブリックへの接続を提供する(図3の(D)、バックアップ・スイッチ)。以上から、単一ネットワーク・プロセッサ・チップは、1つのチップがローエンド・システム(比較的低いポート密度(例えば40)を有する)から、ハイエンド・システム(比較的高いポート密度(例えば80)を有する)まで提供するために使用されるといふ点でスケラブルである。

**【0086】**

システム内の1つのインタフェース装置が、最大10個の10/100Mbps高速イーサネット(R)ポートか、単一の1000Mbpsイーサネット(R)ポートを介して、システム・プロセッサに接続される。システム・プロセッサへのイーサネット(R)構成は、インタフェース装置に接続されるEEPROM内に記憶され、初期化の間にロードされる。システム・プロセッサは、イーサネット(R)フレームとしてカプセル化される特殊ガイド・フレームを作成することにより、システム内の全てのインタフェース装置と通信する(図3参照)。カプセル化されたガイド・フレームは、DASLリンクを介して他の装置に転送され、システム内の全てのインタフェース装置が、単一点から制御されることになる。

**【0087】**

ガイド・フレームは、制御点(CP)と組み込みプロセッサ・コンプレックスとの間で、及びインタフェース装置内で制御情報を伝達するために使用される。ここでの議論を明らかにするガイド・セルの従来の開示が、1998年3月3日発行の米国特許第5724348号"Efficient Hardware/Software Interface for a Data Switch"で述べられている。

**【0088】**

CPから発信されるガイド・フレーム・トラフィックでは、CPがそのローカル・メモリ内のデータ・バッファ内に、ガイド・フレームを作成する。CPのデバイス・ドライバがガイド・フレームを、ネットワーク・プロセッサのメディア・インタフェースの1つに送信する。メディア・アクセス制御(MAC)ハードウェアがガイド・フレームを回復し、それを内部データ・ストア(UDS)メモリに記憶する。ガイド・フレームは適切なブレードに経路指定され、処理され、必要に応じて再度CPに経路指定される。外部CPとインタフェース装置との間でやり取りされるガイド・フレームは、外部ネットワークのプロトコルに適應するようにカプセル化される。その結果、外部ネットワークがイーサネット(R)を含む場合、ガイド・フレームはイーサネット(R)フレームとしてカプセル化される。

**【0089】**

イーサネット(R)カプセル化は、CPとインタフェース装置との間のガイド・トラフィックのトランスポート手段を提供する。インタフェース装置のイーサネット(R)MAC(Enet MAC)は、フレームの受信時に、宛先アドレス(DA)またはソース・アドレス(SA)を分析しない。この分析はEPCピココードにより実行される。ガイド・トラフィックは、インタフェース装置が構成されておらず、DA及びSAがEPCピココードにより分析されないものと仮定する。従って、これらのフレームは元来、自己ルーティングである。しかしながら、イーサネット(R)MACは、イーサネット(R)タイプ・フィールドを分析し、ガイド・トラフィックとデータ・トラフィックとを区別する。ガイド・フレームのこのイーサネット(R)タイプ値の値は、E\_Type\_Cレジスタにロードされる値と合致しなければならない。このレジスタは、インタフェース装置のブート・ピココードにより、フラッシュ・メモリからロードされる。

**【0090】**

CPはそのローカル・メモリ内のデータ・バッファ内に、ガイド・フレームを構成する。

10

20

30

40

50



CPプロセッサ内の32ビット・レジスタの内容が、図4に示されるように、ビッグ・エンディアン形式でローカル・メモリ内に記憶される。ガイド・フレームが構成されると、CPのデバイス・ドライバがイーサネット(R)フレームを送信する。このフレームは、特定ガイド・セル・ハンドラ(GCH)のDA、CPの大域MACアドレスまたは特定インタフェースのMACアドレスに対応するSA、ガイド・フレームを示す特殊イーサネット(R)タイプ・フィールド、及びガイド・フレーム・データを含む。ポートに到来する全てのイーサネット(R)フレームは、イーサネット(R)MACにより受信され、分析される。E\_Type\_Cレジスタの内容に合致するイーサネット(R)タイプ値を有するフレームの場合、イーサネット(R)MACがDA、SA及びイーサネット(R)タイプ・フィールドを剥ぎ取り、ガイド・フレーム・データをU\_DSメモリに記憶する。バイトが1つずつ、イーサネット(R)MACにより、FISHと呼ばれる16バイトのブロック内に収集される。これらのバイトはビッグ・エンディアン形式で記憶され、ガイド・フレームの第1バイトが、FISHの最上位バイト位置(バイト0)に記憶される。続くバイトは、FISH内の続くバイト位置(バイト1, バイト2, . . . , バイト15)に記憶される。これらの16バイトは次に、U\_DS内のバッファに、FISH0位置から記憶される。続くFISHは、バッファ内の連続するFISH位置(FISH1、FISH2、FISH3など)に記憶される。ガイド・フレームの残りを記憶するために、必要に応じて、空きプールから追加のバッファが獲得される。

10

## 【0091】

インタフェース装置10内のガイド・トラフィックのフローが、図5に示される。インタフェース装置のイーサネット(R)MAC機能が、フレーム・ヘッダ情報を調査し、フレームがガイド・フレームであることを判断する。イーサネット(R)MACはガイド・フレームからフレーム・ヘッダを除去し、その残りの内容をインタフェース装置の内部U\_DSメモリにバッファリングする。イーサネット(R)MACは、フレームがGCHにより処理されるために、汎用制御(GC)キューにエンキューされるべきことを示す。ガイド・フレームの終わりに達すると、エンキュー、デキュー及びスケジュール(EDS)論理が、フレームをGCキューにエンキューする。

20

## 【0092】

CPに局所的に接続されるブレード上のGCHピココードは、フレーム制御情報(図7参照)を調査し、ガイド・フレームがシステム内の他のブレードに向けられるか、及びガイド・フレームがインタフェース装置のダインサイド側で実行されるべきかを判断する。フレームが、局所的に接続されるブレード以外のブレード、またはそれに加えたブレードに向けられる場合、GCHピココードがフレーム制御ブロック(FCB)内のTB値を、ガイド・フレームのフレーム制御情報からのTB値により更新し、EDSにフレームを、マルチキャスト・ターゲット・ブレード・フレーム開始(TB\_SOF)キューにエンキューするように命令する。性能上の理由から、示された宛先ブレードの数とは無関係に、全てのガイド・トラフィックがマルチキャストTB\_SOFキューにエンキューされる。

30

## 【0093】

フレームが局所的に接続されたブレードだけに向けられる場合、GCHピココードがフレーム制御情報のUP/DOWNバー・フィールド(以降、'バー'は'0'bで有効なことを意味する)(図7参照)を調査し、ガイド・フレームがインタフェース装置のアップサイドまたはダウンサイドのどちらで実行されるべきかを判断する。ガイド・フレームがインタフェース装置のダウンサイドで実行される場合、GCHピココードがFCB内のTB値を、ガイド・フレームのフレーム制御情報からのTB値により更新し、EDSにフレームをマルチキャストTB\_SOFキューにエンキューするように命令する。フレーム制御情報が、ガイド・フレームがアップサイドで実行されるべきことを示す場合、GCHピココードがガイド・フレームを分析し、そこに含まれるガイド・コマンドにより示される操作を実行する。

40

## 【0094】

ガイド・コマンドを処理する前に、ピココードはフレーム制御情報のACK/NOACK

50

バー・フィールドの値をチェックする、この値が'0'bの場合、ガイド・フレームが処理に続き廃棄される。但し、ガイド読出しコマンドは、この範疇ではない。

【0095】

A C K / N O A C Kバー・フィールドが'1'bで、E A R L Y / L A T Eバー・フィールドが'1'bの場合、ガイド・フレーム内の任意のガイド・コマンドを処理する前に、ピココードが早期A C Kガイド・フレームを作成する。このとき、フレーム制御のT Bフィールドの値は、早期A C Kガイド・フレームの内容に等しく、フレーム制御のT Bフィールドの値は、M y \_\_ T Bレジスタの内容に等しい。ピココードは、フレームのF C B内のT B値を、L A N制御点アドレス(L A N\_C P\_A d d r)レジスタのT Bフィールドに含まれる値により更新し、E D SにフレームをマルチキャストT B \_\_ S O Fキューにエンキューする  
10  
ように命令することにより、早期A C Kガイド・フレームをC Pに返送する。ピココードは次に、ガイド・フレームのガイド・コマンドを処理し、ガイド・フレームを廃棄する。但し、ガイド読出しコマンドはこの範疇ではない。

【0096】

他方、A C K / N O A C Kバー・フィールドの値が'1'bで、E A R L Y / L A T Eバー・フィールドの値が'0'bの場合、ピココードはフレーム制御情報のR E S P / R E Qバー・フィールドを'1'bに変更して、ガイド・フレーム応答を示し、T BフィールドをM y \_\_ T Bレジスタの内容により更新し、ガイド・フレーム内の各ガイド・コマンドを処理する。ガイド・コマンドの処理の間、ピココードは次のガイド・コマンドの完了コード・フィールドを、現ガイド・コマンドの完了ステータス・コード値により更新する。ピコ  
20  
コードは、F C B内のT B値をソース・ブレードに対応する値(C PのL A N\_C P\_A d d rレジスタ値)により更新し、E D SにフレームをマルチキャストT B \_\_ S O Fキューにエンキューするように命令することにより応答をソースに返送する。

【0097】

T B \_\_ S O Fキュー内に存在するフレームは、E D Sにより、転送をスケジュールされる。スイッチ・データ・ムーバ(S D M)が、F C Bに含まれる情報から、スイッチング・ファブリック・セル・ヘッダ及びインタフェース装置フレーム・ヘッダを作成する。これらのセルはスイッチング・ファブリック装置を通過し、ターゲット・ブレードに到達し、そこでセルがD - D Sメモリ内で、フレームに再組み立てされる。ダウンサイドのS D Mは、フレームがガイド・フレームであることを認識し、E D SにそれをG Cキューにエン  
30  
キューするように伝える。

【0098】

G CキューまたはG Tキューからの圧力により、ピココードはガイド・フレームをアクセスし分析する。ダウンサイドに到来する全てのガイド・フレームは、最初にG Cキューにエンキューされる。これらのフレームにおけるフレーム制御情報のG T H / G C Hバー値が、G C Hピココードにより調査される。G T H / G C Hバー値が'0'bの場合、ガイド・フレームがG Tキューにエンキューされる。それ以外では、G C Hピココードが、フレーム制御情報のR E S P / R E Qバー・フィールドを調査し、ガイド・フレームが既に実行されたか否かを判断する。R E S P / R E Qバー・フィールドが値'1'bを有する場合、ガイド・フレームは既に実行されており、C Pに経路指定される。C P接続に対応する  
40  
ターゲット・ポート値は、E P Cピココードにより保持される。これらのターゲット・ポート・キューからのフレームが、インタフェース装置からC Pに返送される。

【0099】

R E S P / R E Qバー・フィールドが値'0'bを有する場合、ブレードはC Pに対してローカルまたはリモートである。これはL A N\_C P\_A d d rレジスタのT Bフィールドの値を、マイ・ターゲット・ブレード(M y \_\_ T B)レジスタの内容と比較することにより解明される。これらが一致する場合、ブレードはC Pにとってローカルであり、それ以外では、ブレードはC Pにとってリモートである。いずれの場合にも、ピココードはフレーム制御情報のU P / D O W Nバー値を調査する。U P / D O W Nバー値が'1'bの場合、フレームはラップT Pキューにエンキューされ、U \_\_ D Sへ転送されて、アップサイドのG C Hに  
50

より処理される。それ以外では、ピココード（GCHまたはGTH）が、ガイド・フレームに含まれるガイド・コマンドにより指示される操作を実行する。ガイド・コマンドの処理に先立ち、ピココードがフレーム制御情報のACK/NOACKバー・フィールドの値をチェックする。この値が'0'bの場合、ガイド・フレームが処理に続き廃棄される。但し、ガイド読出しコマンドはこの範疇ではない。

【0100】

ACK/NOACKバー・フィールドの値が'1'bで、EARLY/LATEバー・フィールドの値が'1'bの場合、ガイド・フレーム内の任意のガイド・コマンドを処理する前に、ピココードが早期ACKガイド・フレームを作成し、このとき、フレーム制御情報のTBフィールドの値は、My\_\_TBレジスタの内容に等しい。ブレードがCPにとってリモートの場合、ピココードは早期ACKガイド・フレームをラップ・ポートに経路指定する。それ以外では、ブレードはCPにとってローカルであり、フレームはCPに対応するポート・キューに経路指定される。ピココードはガイド・コマンドを処理する一方、ラップ・ポートが早期ACKガイド・フレームをD\_DSからU\_DSに転送し、フレームをアップサイドのGCキューにエンキューするか、フレームがポート・キューからCPに返送される。U\_DSにラップバック（wrap back）されるフレームに対して、GCHピココードは再度このフレームを調査するが、RESP/REQバー・フィールドは値'1'bを有する。GCHピココードは、FCB内のTBフィールドを、LAN\_CP\_AddrレジスタのTBフィールドに含まれる値により更新し、EDSにフレームをマルチキャストTB\_\_SOFキューにエンキューするように命令することにより、フレームをCPに返送する。TB\_\_SOFキュー内に存在するフレームは、EDSにより、転送をスケジュールされる。スイッチ・データ・ムバ（SDM）が、FCBに含まれる情報から、Prizmaセル・ヘッダ及びインタフェース装置フレーム・ヘッダを作成する。このフレームからのセルはPrizmaを通過し、CPのローカル・ブレード上で、フレームに再組み立てされる。ダウンサイドのSDMは、フレームがガイド・フレームであることを認識し、EDSにそれをGCキューにエンキューするように伝える。GCHピココードがフレームを分析するとき、RESP/REQバー・フィールドは値'1'bを有する。このことは、このブレードがCPに局所的に接続されることを意味し、従ってガイド・フレームは、CPに対応するポート・キューに経路指定される。このキュー内のフレームは、インタフェース装置からCPに返送される。

【0101】

他方、ACK/NOACKバー・フィールドの値が'1'bで、EARLY/LATEバー・フィールドの値が'0'bの場合、ピココードはRESP/REQバー・フィールドを'1'bに変更し、ガイド・フレーム応答であることを示し、TBフィールドをMy\_\_TBレジスタの内容により置換し、次にガイド・フレーム内の各ガイド・コマンドを処理する。ガイド・コマンドの処理の間、ピココードは次のガイド・コマンドの完了コード・フィールドを、現ガイド・コマンドの完了ステータス・コード値により更新する。ブレードがCPにとってリモートの場合、ピココードはガイド・フレームをラップ・ポートに経路指定する。それ以外では、ブレードはCPにとってローカルであり、フレームはCPに対応するポート・キューに経路指定される。ラップ・ポートがガイド・フレームをD\_DSからU\_DSに転送し、フレームをアップサイドのGCキューにエンキューするか、フレームがポート・キューからCPに返送される。U\_DSにラップバックされるフレームに対して、GCHピココードは再度このフレームを調査するが、RESP/REQバー・フィールドは値'1'bを有する。GCHピココードは、FCB内のTBフィールドを、LAN\_CP\_AddrレジスタのTBフィールドに含まれる値により更新し、EDSにフレームをマルチキャストTB\_\_SOFキューにエンキューするように命令することにより、フレームをCPに返送する。TB\_\_SOFキュー内に存在するフレームは、EDSにより、転送をスケジュールされる。SDMが、FCBに含まれる情報から、Prizmaセル・ヘッダ及びインタフェース装置フレーム・ヘッダを作成する。このフレームからのセルはPrizmaを通過し、CPのローカル・ブレード上のダウンサイドでフレームに再組み立てされる

10

20

30

40

50

。ダウンサイドのSDMは、フレームがガイド・フレームであることを認識し、EDSにそれをGCキューにエンキューするように伝える。GCHピココードがD\_DSからのフレームを分析するとき、RESP/REQバー・フィールドは値'1'bを有する。このことは、このブレードがCPに局所的に接続されることを意味し、ガイド・フレームが、CPに対応するポート・キューに経路指定される。このキュー内のフレームは、インタフェース装置からCPに返送される。

#### 【0102】

何らかの理由により、GCHピココードが、フレーム制御情報のTBフィールドが'0000'hに等しいガイド・フレームに遭遇する場合、GCHピココードはこのフレームを、このブレードだけに向けられるものと解釈し、それにもとづき処理を行う。このアクションは、My\_TBレジスタの値が全てのブレードに対して'0000'hである初期化の間に要求される。CPは、フレーム制御情報が'0000'hのTB値を有するガイド・フレーム内の、書込みガイド・コマンドを送信することにより、局所的に接続されるブレードのMy\_TBレジスタを初期化する。

10

#### 【0103】

EPC内の任意のピコプロセッサがガイド・フレームを生成できる。このフレームは非送信請求ガイド・フレームか、他の形式のガイド・フレームである。このタイプの内部的に生成されるフレームは、肯定応答を許可しないように生成される(すなわちACK/NOACKバー='0'b)。これらのフレームは、同一のEPC内の2つのピコプロセッサの一方(GCHまたはGTH)に送信されるか、他のブレードのGCHまたはGTHに送信される。

20

#### 【0104】

非送信請求ガイド・フレームはCPにも送信され得る。同一のEPCに向けられるガイド・フレームは、D\_DS内のデータ・バッファを用いて構成される。これらのフレームは次に処理のために、GCまたはGTキューにエンキューされる。これらのフレームは次に通常通り処理され、廃棄される。局所的に接続されるCPに向けられる非送信請求ガイド・フレームは、D\_DS内のデータ・バッファを用いて構成される。これらのフレームは、それらがEPCにより実行されたことを示すように構成される(すなわち、RESP/REQバー='1'b、TB=My\_TB)。これらのフレームは、CPに対応するポート・キューにエンキューされる。このキューからフレームは、CPに返送される。

30

#### 【0105】

別のブレードに向けられるガイド・フレームは、D\_DSまたはU\_DS内のデータ・バッファを用いて構成される。CPに向けられる非送信請求ガイド・フレームは、それらがEPCにより実行されたことを示すように構成される(すなわち、RESP/REQバー='1'b、TB=My\_TB)。D\_DS内のデータ・バッファを用いて構成されたフレームは、ラップ・ポートにエンキューされる。これらのフレームはU\_DSに転送され、アップサイドのGCキューにエンキューされる。'1'bのRESP/REQバー値を有する非送信請求ガイド・フレームは、LAN\_CP\_Addrレジスタ内のTB値を用いてCPに経路指定される。それ以外では、GCHピココードが、ガイド・フレームのフレーム制御情報のTB値を用いて、これらのフレームを経路指定する。受信ブレードにおいて、フレームがダウンサイドのGCキューにエンキューされる。このブレードのGCHは、フレームを実行し廃棄するか(RESP/REQバー='0'b、GTH/GCHバー='1')、フレームをGTキューにエンキューするか(RESP/REQバー='0'b、GTH/GCHバー='0')、フレームをCPに対応するポート・キューにエンキューする(RESP/REQバー='1'b)。U\_DS内のデータ・バッファを用いて構成されたフレームは、アップサイドのGCキューに直接エンキューされる。この点から以降、これらのフレームは同一の経路に従い転送され、D\_DSデータ・バッファを用いて構成されたフレームと同様に処理される。図6は、ガイド・フレームの汎用フォーマットを示す。

40

#### 【0106】

図示のフォーマットは、左側に最上位バイトを有し、右側に最下位バイトを有する論理表

50

現である。4バイト・ワードが先頭のワード0から開始し、ページの終わりに向けて増加する。

【0107】

インタフェース装置がCPにより構成される前に、ガイド・フレームは経路指定され、処理されなければならないので、これらのフレームは自己ルーティングでなければならない。探索及び分類により通常獲得される結果が、ガイド・フレームのこのフレーム制御情報フィールド内に含まれ、チップが探索操作を実行することなく、FCBをこの情報により更新することを可能にする。ガイド・フレームに含まれるターゲット・ブレード情報は、ガイド・フレーム・ハンドラにより、FCBのリーフ・ページ・フィールドを用意するために使用される。CPがターゲット・ブレード情報を提供する一方、GCHピココードがFCB内の他のフィールドを記入する。このFCB情報はSDMにより、セル・ヘッダ及びフレーム・ヘッダを用意するために使用される。ガイド・フレームのフレーム制御情報フィールドのフォーマットが図7に示される。

10

【0108】

次に、図7の各ビット位置の略語について説明する。

【0109】

RESP/REQバー：応答及び非要求バー標識値。このフィールドは、要求（未処理）ガイド・フレームと応答ガイド・フレームとを区別するために使用される。

0：要求

1：応答

20

【0110】

ACK/NOACKバー：肯定応答または無肯定応答制御値。このフィールドは、GCHピココードがガイド・フレームを肯定応答するか否かを制御するために使用される（肯定応答する場合ACK、そうでない場合NOACK）。ガイド・フレームが肯定応答されない場合、読出しを実行するいずれの形式のガイド・コマンドも含まれない。

0：無肯定応答

1：肯定応答

【0111】

EARLY/LATEバー：早期及び遅延肯定応答制御値。このフィールドは、要求される肯定応答（ACK/NOACKバー='1'b）が、ガイド・フレームが処理される前に発生するか（EARLY）、または後に発生するか（LATE）を制御するために使用される。ACK/NOACKバー='0'の場合、このフィールドは無視される。

30

0：ガイド・フレーム処理後の肯定応答

1：ガイド・フレーム処理前の肯定応答

【0112】

NEG/ALLバー：否定応答または全肯定応答制御値。このフィールドは、ガイド・コマンドが成功裡に完了しない場合を除き、ACK/NOACKバー・フィールドが値'0'bを有するとき無視される。

0：ACK/NOACKバー・フィールドが'1'bの場合、全てのガイド・フレームを肯定応答する。早期または遅延肯定応答は、EARLY/LATEバーの値により決定される。

40

1：成功裡に完了しないガイド・フレームだけを肯定応答する。この肯定応答は、ACK/NOACKバー及びEARLY/LATEバーの値に関係なく発生し、もちろん遅延肯定応答である。

【0113】

UP/DOWNバー：アップまたはダウン制御値。この値は、フレームがアップサイドまたはダウンサイドのどちらで処理されるかを制御するために使用される。このフィールドは、RESP/REQバーが'1'bのとき無視される。全てのマルチキャスト・ガイド・フレームは、'0'bのUP/DOWNバー値を有する。更に、GTHハードウェア補助機構命令の使用を要求するガイド・コマンドは、'0'bのUP/DOWNバー値を有する。

50

0 : ダウンサイド処理

1 : アップサイド処理

【 0 1 1 4 】

G T H / G C H 値 : 汎用ツリー・ハンドラまたはガイド・セル・ハンドラ制御値。この値は、ガイド・フレームを適切なピコプロセッサに転送するために使用される。

0 : G C H ピコプロセッサ

1 : G T H ピコプロセッサ

【 0 1 1 5 】

T B : ターゲット・ブレード値。R E S P / R E Q 値が ' 0 ' b のとき、このフィールドは P r i z m a により使用されるルーティング情報を含む。各ビット位置はターゲット・ブレードに対応する。この値が ' 0 0 0 0 ' h のとき、ガイド・フレームがこのブレードに当てはまるとみなされ、従って実行される。T B フィールドの 1 つ以上のビット位置の ' 1 ' b の値は、セルが対応するターゲット・ブレードに経路指定されることを示す。R E S P / R E Q 値が ' 1 ' b のとき、このフィールドは応答ブレードの M y \_ T B 値を含む。

10

【 0 1 1 6 】

ガイド・フレームのワード 1 は、相関関係子の値を含む ( 図 8 ) 。この値は C P ソフトウェアにより割当てられ、ガイド・フレーム応答をそれらの要求に相関付ける。相関関係子は、機能を割当てられた複数のビットを含む。

【 0 1 1 7 】

あらゆるガイド・コマンドは、コマンド制御情報フィールドで開始する。このコマンド制御は、G C H ピココードがガイド・フレームを処理するのを支援する情報を含む。この情報のフォーマットが図 9 に示される。

20

【 0 1 1 8 】

レングス値 : この値は、制御情報に含まれる 3 2 ビット・ワードの総数 ( コマンド・ワード 0 ) 、アドレス情報 ( コマンド・ワード 1 ) 、及びガイド・フレームのオペランド ( コマンド・ワード 2 + ) 部分を含む。

【 0 1 1 9 】

完了コード値 : このフィールドは C P により初期化され、ガイド・コマンドを処理するとき、G C H ピココードにより変更される。G C H ピココードはこのフィールドを、コマンド・リスト内の先行ガイド・コマンドの完了ステータスとして使用する。全てのガイド・コマンド・リストは、終了区切り文字ガイド・コマンドで終了するので、最後のコマンドの完了ステータスが、終了区切り文字の完了コード・フィールド内に含まれる。

30

【 0 1 2 0 】

ガイド・コマンド ( G C ) タイプ値 ( シンボル名 ) :

【 表 1 】

| シンボル名               | タイプ値      | タイプ記述                      |    |
|---------------------|-----------|----------------------------|----|
| End_Delimiter       | 0000      | ガイド・フレーム・シーケンスの終わりをマーク     |    |
| Build_TSE_Free_List | 0001      | フリー・リストを作成                 |    |
| Software_Action     | 0010      | ソフトウェア・アクションを実行            |    |
| Unsolicited         | 0011      | EPCピココードにより開始されるフレーム       |    |
| Block_Write         | 0100      | データ・ブロックを連続アドレスに書込む        | 10 |
| Duplicate_Write     | 0101      | 重複データをレジスタまたはメモリに書込む       |    |
| Read register       | 0110      | レジスタまたはメモリ・データの読出しの要求または応答 |    |
|                     | 0111      | 予約済み                       |    |
| Insert_Leaf         | 1000      | リーフを探索ツリーに挿入               |    |
| Update_Leaf         | 1001      | 探索ツリーのリーフを更新               |    |
| Read_Leaf           | 1010      | リーフの読出しの要求及び応答             | 20 |
| Leaf                |           |                            |    |
|                     | 1011      | 予約済み                       |    |
| Delete_Leaf         | 1100      | 探索ツリーのリーフを削除               |    |
|                     | 1101-1111 | 予約済み                       |    |

## 【 0 1 2 1 】

ガイド・フレームに含まれるアドレッシング情報は、ネットワーク・プロセッサのアドレッシング機構内の要素を識別する。アドレス情報フィールドの汎用形式が、図 1 0 に示される。

30

## 【 0 1 2 2 】

インタフェース装置は 3 2 ビット・アドレッシング方式を採用する。このアドレッシング方式は、アドレス値を、インタフェース装置のあらゆるアクセス可能な構造に割り当てる。これらの構造はプロセッサの内部に存在するか、プロセッサの制御に従い、インタフェースに接続される。これらの構造のあるものは、組み込みプロセッサ・コンプレックス ( E P C ) により、ウェブ・インタフェースと呼ばれる内部インタフェースを介してアクセスされる。残りの構造は、メモリ制御装置インタフェースを介してアクセスされる。全ての場合において、アドレスの汎用形式が図 1 1 に示される。

## 【 0 1 2 3 】

ネットワーク制御装置は、主要チップ・アイランドに細分化される。各アイランドは固有のアイランド ID 値を与えられる。この 5 ビット・アイランド ID 値は、そのチップ・アイランドにより制御される構造のアドレスの、最上位 5 ビットを形成する。エンコード化されたアイランド ID 値とチップ・アイランド名との対応が、図 1 2 に示される。ウェブ・アドレスの第 2 の部分は、次の最上位 2 3 ビットを含む。このアドレス・フィールドは、構造アドレス部分と、エレメント・アドレス部分とに区別化される。各セグメントに対して使用されるビット数は、アイランド毎に異なる。一部のアイランドは数個の大きな構造だけを含むのに対して、他のアイランドは多くの小さな構造を含む。そうした理由から、これらのアドレス・セグメントには固定サイズが存在しない。構造アドレス部分は、アイランド内のアレイをアドレス指定するために使用されるのに対して、エレメント・アドレス部分は、アレイ内の要素をアドレス指定するために使用される。アドレスの残りの部

40

50

分は、ウェブ・インタフェースの32ビット・データ・バス制限を調整する。この4ビット・ワード・アドレスは、アドレス指定されるエレメントの32ビット・セグメントを選択するために使用される。これは32ビットより広い構造エレメントを、ネットワーク制御装置のウェブ・データ・バスを介して、移動するために必要である。ワード・アドレス値'0'hは、構造エレメントの最上位32ビットを指し示すのに対して、順次ワード・アドレス値は、構造エレメントの連続的な下位セグメントに対応する。アドレスのワード・アドレス部分は、ウェブ・インタフェースを介してアクセスされない構造に対しては、要求されない。この理由から、アップ・データ・ストア、制御メモリ、及びダウン・データ・ストアは、アドレスの最下位27ビット全てを使用し、構造エレメントをアクセスする。このフォーマットの別の例外は、SPMインタフェースのアドレスである。この場合、アドレスの27ビット全てが使用され、32ビットよりも大きな幅のエレメントは存在しない。

10

#### 【0124】

組み込みプロセッサ・コンプレックス(EPIC)は、インタフェース素子チップのプログラマビリティを提供し制御する。これは次のコンポーネントを含む(図13参照)。

#### 【0125】

N個の処理ユニット(G×Hと呼ばれる): G×Hは、共通命令メモリに記憶されるピココードを同時に実行する。各G×Hは処理ユニット・コア(CLPと呼ばれる)を含み、これは3ステージ・パイプライン、16GPR、及びALUを含む。各G×Hはまた、例えばツリー構造探索エンジンのように、幾つかのコプロセッサを含む。

20

#### 【0126】

命令メモリ: 初期化の間にロードされ、フレームを転送し、システムを管理するピココードを含む。

#### 【0127】

ディスパッチャ: アップ及びダウン・ディスパッチャ・キューからフレーム・アドレスをデキューする。デキューの後、ディスパッチャがアップまたはダウン・データ・ストア(DS)から、フレーム・ヘッダの一部をプリフェッチし、これを内部メモリに記憶する。G×Hがアイドルになると、ディスパッチャが直ちに、コード命令アドレス(CIA)のような適切な制御情報を有するフレーム・ヘッダを、G×Hに渡す。ディスパッチャはまた、タイマ及び割込みを処理する。

30

#### 【0128】

ツリー構造探索メモリ(TSM)アービタ: 各G×Hが使用可能な多数の共用内部メモリ位置及び外部メモリ位置が存在する。このメモリは共用されるので、メモリへのアクセスを制御するために、アービタが使用される。TSMはピココードにより直接アクセスされ、例えばエージング・テーブルを記憶するために使用される。TSMはまた、ツリー構造探索の間に、TSEによりアクセスされる。

#### 【0129】

完了ユニット(CU): 完了ユニットは2つの機能を実行する。第1に、これはN個の処理ユニットを、アップ及びダウンEDS(エンキュー、デキュー、及びスケジュール・アイランド)とインタフェースする。EDSはエンキュー・アクションを実行する。すなわち、フレーム・アドレスが、FCBページと呼ばれる適切なパラメータと一緒に、伝送キュー、廃棄キュー、またはディスパッチャ・キューのいずれかに、待ち行列化される。第2に、完了ユニットはフレーム・シーケンスを保証する。複数のG×Hが、同一のフローに属するフレームを処理している可能性があるため、これらのフレームがアップまたはダウン伝送キューに正しい順序でエンキューされるように、事前注意が払われなければならない。完了ユニットは、フレーム・ディスパッチ時に分類子ハードウェア補助機構により生成されるラベルを使用する。

40

#### 【0130】

分類子ハードウェア補助機構: アップ・フレームでは、分類子ハードウェア補助機構は、フレーム・フォーマットの周知のケースを分類する。分類結果はフレーム・ディスパッチ

50



の間に、C I A及び1つ以上のレジスタの内容に関して、G x Hに渡される。ダウン・フレームでは、分類子ハードウェア補助機構は、フレーム・ヘッダに応じて、C I Aを決定する。アップ及びダウン・フレーム・ディスパッチの両方のために、分類子ハードウェア補助機構は、フレーム・シーケンスを維持するために、完了ユニットにより使用されるラベルを生成する。

【0131】

アップ/ダウン・データ・ストア・インタフェース及びアービタ：各G x Hはアップ及びダウン・データ・ストアをアクセスできる。すなわち、"より多くのF I S H"を讀出すとき、讀出しアクセスが提供され、F I S Hプールの内容をデータ・ストアに書戻すとき、書込みアクセスが提供される。N個の処理ユニットが存在し、1度にそれらの1つだけがアップ・データ・ストアをアクセスでき、1度に1つだけがダウン・データ・ストアをアクセスできるので、各データ・ストアに対して、1つのアービタが必要とされる。

10

【0132】

ウェブ・アービタ及びウェブウォッチ・インタフェース：ウェブ・アービタはG x Hの間で、ウェブへのアクセスを調停する。全てのG x Hはウェブをアクセスでき、このことは、インタフェース装置内の全てのメモリ及びレジスタ機能をアクセスすることを可能にする。これにより、G x Hは全ての構成領域を変更または讀出すことができる。ウェブはインタフェース装置のメモリ・マップとみなされる。ウェブウォッチ・インタフェースは、3つのチップ入出力ピンを用いて、チップの外部からウェブ全体をアクセスできるようにする。

20

【0133】

デバッグ、割込み及び単一ステップ制御：ウェブはG C Hまたはウェブウォッチャが、必要に応じて、チップ上の各G x Hを制御することを可能にする。例えば、ウェブはG C Hまたはウェブウォッチャにより、G x H上で命令を単一ステップ操作するために使用される。

【0134】

Power PCなどの組み込み汎用プロセッサ：  
4つのタイプのG x Hが存在する（図14参照）。

【0135】

G D H（汎用データ・ハンドラ）：8個のG D Hが存在する。各G D Hは完全なC L Pと、5個のコプロセッサ（次のセクションで述べられる）を有する。G D Hは主に、フレームを転送するために使用される。

30

【0136】

G C H（ガイド・セル・ハンドラ）：G C Hは、G D Hと正に同一のハードウェアを有する。しかしながら、ガイド・フレームだけがG C Hにより処理される。G C Hがデータ・フレームについても処理できるか否かは（この場合、G C HはG D Hの役割をする）、ウェブ上でプログラマブルである（C L P \_ E n aレジスタ）。G C HはG D Hに比較して、追加のハードウェア、すなわち、ツリー挿入及び削除を実行するハードウェア補助機構を有する。G C Hは、ガイド・セル関連ピココードを実行したり、エージングなどの、チップ及びツリー管理関連ピココードを実行したり、或いは制御情報をC Pや別のG C Hと交換するために使用される。実行すべきこうしたタスクが存在しない場合、G C Hはフレーム転送関連ピココードを実行し、この場合、丁度G D Hのように動作することになる。

40

【0137】

G T H（汎用ツリー・ハンドラ）：G T Hはツリー挿入、ツリー削除、及びロープ管理を実行する追加のハードウェア補助機構を有する。G P Q内に（ツリー管理コマンドを含む）フレームが存在しない場合、G T Hはデータ・フレームを処理する。

【0138】

G P H（汎用Power PCハンドラ）：G P HはG D H及びG T Hに比較して、追加のハードウェアを有する。G P Hはメールボックス・インタフェース（I / F）を介して、汎用プロセッサとインタフェースする。

50

## 【 0 1 3 9 】

G × H の数 ( 1 0 個 ) は " 最善推量 " ( best-guess ) である。性能評価により、実際に要求される G × H の個数が決定される。アーキテクチャ及び構造は、より多くの G × H に向けて完全にスケラブルであり、唯一の制限はシリコン面積である ( より大きなアービタ及び命令メモリを含むはずである ) 。

## 【 0 1 4 0 】

各 G × H は、図 1 5 に示されるように構造化される。汎用レジスタ ( G R P ) 及び演算論理ユニット ( A L U ) を有する C L P に加え、各 G × H は次の 5 つのコプロセッサを含む。

## 【 0 1 4 1 】

( D S ) コプロセッサ・インタフェース：ディスパッチャ、並びにアップ及びダウン・データ・ストアへの読み出し及び書き込みアクセスを提供するサブアイランドとインタフェースする。D S インタフェースは、いわゆる F I S H プールを含む。

## 【 0 1 4 2 】

ツリー構造探索エンジン・コプロセッサ ( T S E ) : T S E はツリー内の探索を実行し、ツリー構造探索メモリ ( T S M ) とインタフェースする。

## 【 0 1 4 3 】

エンキュー・コプロセッサ：完了ユニット・インタフェースとインタフェースし、F C B ページを含む。このコプロセッサは、ピココードがエンキュー・パラメータを含む F C B ページを作成するために使用する追加のハードウェア補助機構と共に、2 5 6 ビット・レジスタを含む。一旦 F C B ページが作成されると、ピコプロセッサがエンキュー命令を実行し、F C B ページが完了ユニットに転送される。

## 【 0 1 4 4 】

ウェブ・インタフェース・コプロセッサ：このコプロセッサはウェブ・アービタとのインタフェースを提供し、インタフェース装置への書き込み及び読み出しを可能にする。

## 【 0 1 4 5 】

チェックサム・コプロセッサ：F I S H プール ( 後述 ) に記憶されるフレーム上に、チェックサムを生成する。

## 【 0 1 4 6 】

処理ユニットは、イングレス処理とイーグレス処理との間で共用される。イングレス処理対イーグレス処理において、どれだけの帯域幅が確保されるかは、プログラマブルである。現インプリメンテーションでは、2 つのモデルが存在し、一方は 5 0 / 5 0 ( すなわち、イングレス及びイーグレスが同じ帯域幅を獲得する ) で、他は 6 6 対 3 4 ( すなわち、イングレスがイーグレスの 2 倍の帯域幅を獲得する ) である。

## 【 0 1 4 7 】

処理ユニットの動作はイベント・ドリブン方式である。すなわち、フレームの到来がイベントとして、またタイマまたは割込みのポッピングとして扱われる。ディスパッチャは異なるイベントを同様に扱う。但し、優先順位は存在する ( 第 1 が割込みで、第 2 がタイマ・イベントで、第 3 がフレーム到来イベント ) 。イベントが処理ユニットに渡されると、適切な情報が処理ユニットに与えられる。フレーム到来イベントでは、これはフレーム・ヘッダの一部と、ハードウェア分類子からの情報とを含む。タイマ及び割込みでは、これはコード・エントリ・ポイントと、イベントに関連する他の情報とを含む。

## 【 0 1 4 8 】

フレームがイングレス側に到来し、このフレームの受信バイト数がプログラマブルしきい値を超える場合、フレーム制御ブロックのアドレスが G Q に書込まれる。

## 【 0 1 4 9 】

完全なフレームがイーグレス側で再組み立てされた場合、フレーム・アドレスが G Q に書込まれる。4 つのタイプの G Q が存在する ( そして、各タイプに対して、図 1 4 に示されるように、イングレス・バージョン及びイーグレス・バージョンが存在する ) 。すなわち、

10

20

30

40

50

G C Q : G C Hにより処理されなければならないフレームを含む。

G T Q : G T Hにより処理されなければならないフレームを含む。

G P Q : G P Hにより処理されなければならないフレームを含む。

G D Q : 任意のG D H (またはG C H / G T Hがデータ・フレームを処理できる場合には、G C H / G T H)により処理され得るフレームを含む。G D Qについては、複数の優先順位が存在し、高い優先順位でエンキューされたフレームは、低い優先順位でエンキューされたフレームより先に処理される。

【 0 1 5 0 】

一部の処理ユニットは特殊化され得る。現インプリメンテーションでは、4つのタイプの処理ユニット (G x H) が存在する (図 1 4 参照)。すなわち、

10

G D H (汎用データ・ハンドラ) : G D Hは主に、フレームを転送するために使用される。

【 0 1 5 1 】

G C H (ガイド・セル・ハンドラ) : G C Hは、G D Hと正に同一のハードウェアを有する。しかしながら、ガイド・フレームだけがG C Hにより処理される。G C Hがデータ・フレームについても処理できるか否かは (この場合、G C HはG D Hの役割をする)、ウェブ上でプログラマブルである (C L P \_ E n aレジスタ)。

【 0 1 5 2 】

G T H (汎用ツリー・ハンドラ) : G T HはG D H及びG C Hに比較して、追加のハードウェア、すなわち、ツリー挿入、ツリー削除、及びロープ管理を実行するハードウェア補助機構を有する。G P Q内に (ツリー管理コマンドを含む) フレームが存在しない場合、G T Hはデータ・フレームを処理する。

20

【 0 1 5 3 】

G P H (汎用P o w e r P Cハンドラ) : G P HはG D H及びG T Hに比較して、追加のハードウェアを有する。G P Hはメールボックス・インタフェースを介して、組み込みP o w e r P Cとインタフェースする。

【 0 1 5 4 】

実際のインプリメンテーションでは、G C H、G T H及びG P Hの役割は、単一の処理ユニット上で実現される。例えば、1インプリメンテーションはG C H及びG P Hに対して、1つの処理ユニットを有する。類似のコメントがG C Q、G T O及びG P Qについても

30

適用できる。

【 0 1 5 5 】

データ・ストア・コプロセッサの目的は、メディアから受信されたフレームを含むアップ・データ・ストアと、及びP r i z m a A t l a n t i cから受信される再組み立て済みフレームを含むダウン・データ・ストアとインタフェースすることである。

【 0 1 5 6 】

データ・ストア・コプロセッサはまた、タイマ・イベントまたは割込みのディスパッチの間に、構成情報を受信する。

【 0 1 5 7 】

データ・ストア・コプロセッサは、フレーム上のチェックサムを計算できる。

40

【 0 1 5 8 】

データ・ストア・コプロセッサはF I S Hプール (8 F I S Hを保持できる) と、スクラッチ・メモリ (8 F I S Hを保持できる) と、アップまたはダウン・データベースからF I S Hプール内容を読み書きする幾つかの制御レジスタとを含む。F I S Hプールは、データ・ストアのためのある種の作業領域とみなされる。すなわち、データ・ストアを直接読み書きする代わりに、大量のフレーム・データがデータ・ストアからF I S Hプールに読出されるか、大量のデータがF I S Hプールからデータ・ストアに書込まれる。転送の単位はF I S Hすなわち16バイトである。

【 0 1 5 9 】

F I S Hプールは、8 F I S H、すなわち各々が128ビットの8ワード分を含むことが

50

できるメモリとみなされる。CLPプロセッサ・アーキテクチャでは、FISHプールは128バイトのレジスタ・アレイである。FISHプール内の各バイトは、7ビット・バイト・アドレス(0乃至127)を有し、アクセスは16ビットまたは32ビット・ベースで行われる。全てのレジスタ・アレイ同様、FISHプールは循環アドレス方式を有する。すなわち、FISHプール内のロケーション126で開始するワード(すなわち4バイト)のアドレス指定は、バイト126、127、0及び1を返却する。更に、データ・ストア・コプロセッサの観点から、FISHプール内のFISHロケーションは、3ビットのFISHアドレスを有する。

**【0160】**

フレーム・ディスパッチに際して、フレームの最初のN個のFISHが、ディスパッチャにより自動的にFISHプールにコピーされる。Nの値は、ポート構成メモリ(PortConfigMemory)内でプログラマブルである。一般に、アップ・フレーム・ディスパッチでは、Nは4に等しく、ダウン・ユニキャスト・フレーム・ディスパッチでは2で、ダウン・マルチキャスト・フレーム・ディスパッチでは4で、割込み及びタイマでは0である。

10

**【0161】**

ピココードはフレームからより多くのバイトを読み出すことができ、この場合、データ・ストア・コプロセッサが自動的に、フレーム・データをFISHプール内の次のFISHアドレスに読み出し、FISHプールの境界に達すると、自動的に0に循環する。また、ピココードはアップ/ダウン・データ・ストアを絶対アドレスにおいて読み書きすることもできる。

20

**【0162】**

ウェブ・コプロセッサはEPCウェブ・アービタとインタフェースする。EPCウェブ・アービタは10個のG×Hとウェブウォッチャとの間で、インタフェース装置のウェブ・インタフェース上でマスタになるものを調停する。これは全てのG×Hがウェブ上で読み書きすることを可能にする。

**【0163】**

インタフェース装置のメモリ・コンプレックスは、図13に示される組み込みプロセッサ・コンプレックス(EPC)の記憶機構を提供する。メモリ・コンプレックスはツリー構造探索メモリ(TSM)アービタと、複数のオンチップ及びオフチップ・メモリを含む。メモリはツリー構造、カウンタ、及びピココードによりメモリ・アクセスを要求される他のものを記憶する。更に、メモリはフリー・リストやキュー制御ブロックなど、ハードウェアにより使用されるデータ構造を記憶するために使用される。ツリーのために割当てられない、またはハードウェアにより使用されないメモリ・ロケーションは、デフォルトでは、ピココードがカウンタやエージング・テーブルとして使用することができる。

30

**【0164】**

図16は、メモリ・コンプレックスの詳細ブロック図を示す。ツリー構造探索メモリ(TSM)アービタが、組み込みプロセッサ(G×H)とメモリとの間の通信リンクを提供する。メモリは5個のオンチップSRAMと、1個のオフチップSRAMと、7個のオフチップDRAMとを含む。TSMアービタは、10個の要求制御ユニット(各々が組み込みプロセッサG×Hの1つに接続される)と、13個のメモリ・アービタ・ユニット(各々が各メモリに対応する)とを含む。各制御ユニット及びその接続されるG×Hが、全てのメモリをアクセスできるように、バス構造が要求制御ユニット及びアービタ・ユニットを相互接続する。

40

**【0165】**

制御ユニットは、組み込みプロセッサ(G×H)とアービタとの間で、データを方向付けするために必要なハードウェアを含む。

**【0166】**

SRAMアービタ・ユニットはとりわけ、組み込みプロセッサG×Hとオンチップ及びオフチップSRAMとの間の、データのフローを管理する。

**【0167】**

50

D R A Mアービタ・ユニットはとりわけ、組み込みプロセッサG x HとオフチップD R A M素子との間のデータのフローを管理する。

【 0 1 6 8 】

各メモリ・アービタは"バックドア" ( back-door ) アクセスを含み、これは一般にチップの他の部分により使用され、最も高いアクセス優先順位を有する。

【 0 1 6 9 】

D R A Mメモリは2つの動作モードで実行できる。すなわち、

1 ) T D Mモード : D R A M内の4つのバンクへのメモリ・アクセスが、読出しウィンドウ及び書込みウィンドウを交互することにより実行される。読出しウィンドウでは、4つのバンクのいずれかへのアクセスが読出し専用であり、書込みウィンドウでは、4つのバンクのいずれかへのアクセスが書込み専用である。複数のD R A Mに対してT D Mモードを使用することにより、D R A M間で幾つかの制御信号を共用することが可能になり、希少資源である幾つかのチップ入出力を節約できる。

【 0 1 7 0 】

2 ) 非T D Mモード : D R A M内の4つのバンクへのメモリ・アクセスが、読出しと書込みの組み合わせとなり得、これは特定の規則に従わねばならない。例えば、アクセス・ウィンドウ内で、バンクAでは読出しを、バンクCでは書込みを実行できる。

【 0 1 7 1 】

T S Mアービタは、N個のリクエストが同時にM個のメモリをアクセスすることを可能にする。複数のリクエストが同一のメモリをアクセスしたい場合、ラウンドロビン・アービトレーションが実行される。

【 0 1 7 2 】

M個のメモリは異なるプロパティを有することができる。現インプリメンテーションでは、3つのメモリ・タイプ、すなわち、内部S R A M、外部S R A M、及び外部D R A Mが存在する。

【 0 1 7 3 】

M個のメモリ及びNリクエストは同種 ( homogeneous ) であり、任意のリクエストが任意のメモリをアクセスできる。

【 0 1 7 4 】

一部のメモリは複数のサブメモリ ( D R A M内の4つのバンクなど ) に論理的に分割され、これらは論理的に同時にアクセスできる。

【 0 1 7 5 】

M個のメモリの一部は、内部的に使用されるデータ構造を含む制御メモリとして使用され、これらはピコプロセッサと比較して、高い優先アクセスを有する。このことはまたチップのでバッグを可能にする。なぜなら、ピコプロセッサは制御メモリの内容を読出すことができるからである。

【 0 1 7 6 】

アービタは読出しアクセス、書込みアクセス、及びread-add-writeをサポートする。それにより、Nビット整数がアトミック演算において、メモリの内容に追加される。

【 0 1 7 7 】

また、メモリ内のオブジェクトの物理ロケーションが透過的となるように、汎用アドレス方式が、M個のメモリをアクセスするために使用される。

【 0 1 7 8 】

ツリー概念は、ツリー構造探索エンジンにより、情報を記憶及び検索するために使用される。検索、すなわちツリー探索及び挿入並びに削除は、キーにもとづき実行される。ここでキーは、例えばM A Cソース・アドレスのようなビット・パターンであるか、I Pソース・アドレスとI P宛先アドレスの連結である。少なくともキーを含む情報が、リーフと呼ばれる制御ブロック内に記憶される ( 後述のように、記憶ビット・パターンは実際にはハッシュ・キーである ) 。リーフは更に、エージング情報などの追加の情報や、ターゲット・ブレード及びターゲット・ポート番号などの情報を転送するユーザ情報を含み得る

10

20

30

40

50

## 【 0 1 7 9 】

3つのタイプ ( F M、 L P M、 S M T )、及び関連ツリー・タイプ探索、すなわち、固定マッチ、ソフトウェア管理ツリー、及び最長プレフィックス・マッチが存在する。ツリー探索の間にリーフをチェックするオプションの追加の基準は、ベクトルマスクである。ローピング、エージング及びラッチは、探索性能を向上させるために使用される。

## 【 0 1 8 0 】

F Mツリーの探索アルゴリズムが、図 1 7 に示される。探索アルゴリズムはキーを含む入力パラメータに作用し、キーにハッシュを実行し、ダイレクト・テーブル ( D T ) をアクセスし、パターン探索制御ブロック ( P S C B ) を通じてツリーを探索し、リーフに行き着く ( 図 1 7 )。3つのタイプのツリーが存在し、各々は異なる規則に従いツリー探索を発生させる独自の探索アルゴリズムを有する。例えば、固定マッチ ( F M ) ツリーでは、データ構造がパトリシア・ツリーである。リーフが見いだされるとき、このリーフは入力キーに一致する唯一の可能な候補である。ソフトウェア管理ツリーでは、リンク・リスト内でチェーンされる複数のリーフが存在し得る。この場合、一致が見いだされるか、チェーンが尽きるまで、チェーン内の全てのリーフが入力キーと符合される。入力キーをリーフに記憶されるパターンと比較するいわゆる "最終比較" ( compare at the end ) 操作が、リーフが真に入力キーに一致するか否かを確認する。リーフが見いだされ、一致が発生するとき、探索結果は O K であり、他の全ての場合には K O である。

## 【 0 1 8 1 】

探索操作への入力は、次のパラメータを含む。すなわち、  
キー ( 1 2 8 ビット ) : キーは探索 ( または挿入 / 削除 ) の前に、特殊なピココード命令を用いて作成される。1つのキー・レジスタだけが存在する。しかしながら、ツリー構造探索が開始されると、キー・レジスタは T S E が探索を実行するのと並行して、ピココードにより、次の探索のキーを作成するために使用される。これは T S E がキーをバッシュ ( bash ) し、結果を内部ハッシュドキー・レジスタに記憶することによる ( 従って、実際には2つのキー・レジスタが存在する )。

## 【 0 1 8 2 】

キー・レングス ( 7 ビット ) : このレジスタはキーの長さを含み、ビットで表す。これはキーの作成の間に、ハードウェアにより自動的に更新される。

## 【 0 1 8 3 】

L U 定義指標 ( L U D e f i n d e x ) ( 8 ビット ) : これは探索が発生するツリーの完全な定義を含む L U 定義テーブル ( L U D e f T a b l e ) への指標である。L U 定義テーブルは以下で詳述される。

## 【 0 1 8 4 】

T S R N r ( 1 ビット ) : 探索結果はツリー探索結果領域 0 ( T S R 0 ) または T S R 1 に記憶される。これは T S R N r により指定される。T S E が探索を行っている間、ピココードは他の T S R をアクセスし、前の探索の結果を分析することができる。

## 【 0 1 8 5 】

ベクトル指標 ( 6 ビット ) : ベクトルマスクをイネーブルされるツリーでは ( L U 定義テーブル内で指定される )、ベクトル指標がベクトルマスク内のビットを示す。探索の終わりに、このビットの値が返却され、ピココードにより使用される。

## 【 0 1 8 6 】

図 1 7 に示されるように、入力キーはハッシュドキーにハッシュされる。使用可能な6つの固定ハッシュ・アルゴリズムが存在する ( 1つの "アルゴリズム" はハッシュ関数を実行しない )。どのアルゴリズムが使用されるかについては、L U 定義テーブル内で指定される。プログラマブル・ハッシュ関数が、柔軟性を追加するために使用され得る。

## 【 0 1 8 7 】

ハッシュ関数の出力は常に 1 2 8 ビット数であり、オリジナル入力キーとハッシュ関数の出力との間には、1対1対応が存在する。後述のように、この特性はダイレクト・テーブ

10

20

30

40

50

ルの後で開始するツリーの深さを最小化する。

【0188】

図17の場合のように、カラーがツリーに対して許可される場合、16ビット・カラー・レジスタが128ビット・ハッシュ関数出力に挿入される。挿入はダイレクト・テーブルの直後に発生する。すなわち、図示のように、ダイレクト・テーブルが $2^N$ 個のエントリを含む場合、16ビット・カラー値がビット位置Nに挿入される。ハッシュ関数の出力は、挿入されるカラー値（イネーブルされる場合）と一緒に、ハッシュドキー・レジスタに記憶される。

【0189】

ハッシュ関数は、その出力内のほとんどのエントロピが最上位ビット側に存在するように定義される。ハッシュドキー・レジスタの最上位Nビットは、ダイレクト・テーブル(DT)への指標を計算するために使用される。

10

【0190】

探索はダイレクト・テーブルへのアクセスにより開始する。すなわち、ダイレクト・テーブル(DT)エントリがダイレクト・テーブルから読出される。DTエントリを読出すために使用されるアドレスは、ハッシュドキーの最上位Nビットから計算され、LU定義テーブル内で定義されるツリー属性に関しても同様である。これについては以下で詳述される。DTエントリはツリーのルートとみなされる。使用される特定のツリー・データ構造は、ツリー・タイプに依存する。この時点では、パトリシア・ツリー・データ構造がFMツリーとして、及びLPM及びSMTツリーのためのパトリシア・ツリーの拡張として使用されることを述べれば十分であろう。

20

【0191】

8エントリ・ダイレクト・テーブル(DT)の使用例が、図18に示される。DTを使用することにより、探索時間(すなわち、アクセスされなければならないPSCBの数)が低減される。従って、DTサイズを増加することにより、メモリ使用と探索性能との間でトレードオフが生じる

【0192】

図18から明らかのように、DTエントリは次の情報を含む。すなわち、

- 1) エンプティ: このDTエントリに接続されるリーフは存在しない。
- 2) リーフを指し示すポインタ: このDTエントリに接続される1つのリーフが存在する
- 3) PSCBを指し示すポインタ: このDTエントリに接続される2つ以上のリーフが存在する。DTエントリはツリーのルートを定義する。

30

【0193】

ソフトウェア管理ツリーの探索アルゴリズム、及びツリーを生成するアルゴリズムが、米国特許出願第09/312148号で述べられている。

【0194】

"選択ビット・アルゴリズム"と称されるアルゴリズムは、特定のメトリックを使用し、規則のセットまたは領域内の"ルール"と称されるアイテムから選択されるビットにもとづき、二分探索ツリーを作成する。ここで述べる全ての例は、インターネット・プロトコル(IP)ヘッダに関連して議論されるが、任意のタイプの固定フォーマット・ヘッダが代わりに使用され得る。

40

【0195】

IPでは、各ルールは次のサブセクション、すなわちソース・アドレス(SA)、宛先アドレス(DA)、ソース・ポート(SP)、宛先ポート(DP)、及びプロトコル(P)で作成される、特定のキーに関する。これらのデータはそれぞれ32、32、16、16及び8ビット長であり、従ってテストされるキーは104ビットから成る。選択ビット・アルゴリズムは、104ビットの内の特に有用な幾つかのビットを見いだす。実際に2、3のビットをテストすることは、1つのルールを除く全てを、または2、3のルールを除く全てを、可能なアプリケーションから削除する。一部のルールでは、単純な比較操作に

50

よる不等テストもまたふさわしい。ビット・テスト及び比較は、二分木において論理的に編成される。ツリーは、ビットを高速にテスト可能なハードウェア構造にマップされる。こうしたテストの結果、キーが適合する1つのルールまたは少数のルール(リーフ・チェーンと呼ばれる)が生成される。前者の場合、キーがルールにより詳細にテストされる。後者の場合、比較及び完全ルール・テストにより、キーがテストの枠内でテストされる。

【0196】

ルール・セット内の各ルールは、規則がキーに適合する最高優先順位の規則の場合に取られるアクションに関連付けられる。ルールは交差する(すなわち、1つのキーが2つまたはそれ以上のルールに適合する)。その場合、ルールは優先順位番号1、2、3、・・・を与えられ、任意の2つの交差するルールが異なる優先順位を有する(キーが2つ以上のルールに適合する場合、管理者はどのルールが上位になるかを宣言しなければならない)。従って、ビット・テスト及び比較後に、依然2つ以上のルールがテストされる場合、ルールは優先順位に従いテストされる。低い優先順位番号が、高い優先順位のルールを指定する。

10

【0197】

全く適合が見いだされない場合、デフォルト規定が指定される。

【0198】

最長プレフィックス・マッチング法の探索アルゴリズムが、米国特許第5787430号で述べられている。この方法は、前記データベースのノード(ルートノード)において入力するステップと、次の(子)ノードを識別するのに必要なエントリ部分だけを含む探索引数のセグメントと、第2のリンク情報とを、前記セグメントが消費されるか、前記第2のリンク情報を欠く(リーフ)ノードに達するまで連続的に処理することにより、ツリー状データベースを通じてあるノードから別のノードへの探索パスを決定するステップと、前記探索引数を、前記探索パスが終了するノードに記憶されるエントリと比較するステップと、前記現ノードにおいて、前記探索引数と前記エントリとの間に、少なくとも部分的な一致が見いだされない場合、前記現ノードの第1のリンク情報を処理することにより、前記探索パスを後戻りするステップと、前記少なくとも部分的な一致が見いだされるか、または前記ルート・ノードに達するまで、前の2つのステップを繰り返すステップとを含む。

20

【0199】

図19は、メイン・スイッチング・ファブリック装置の実施例を示す。好適には、各インタフェース素子チップは、少なくとも2つの集積化並列直列変換ポートを有し、これらは並列データを受信して、それを高速直列データ・ストリームに変換する。そして、直列データがシリアル・リンクを介して、スイッチング・ファブリック装置に転送される。高速シリアル・リンクを介してスイッチング・ファブリック装置から受信されるデータは、別のDASLにより、並列データに変換される。データ・アライン・シリアル・リンク(DASL)と称されるシリアライザ/デシリアライザの実施例が、ここでは述べられる。

30

【0200】

少なくとも1つのDASLが、スイッチング・ファブリック装置をシリアル・リンクにインタフェースする。シリアル・リンクからのデータは、並列データに変換され、これがスイッチング・ファブリック装置に転送する。同様に、スイッチング・ファブリック装置からの並列データは直列データに変換され、シリアル・リンクに転送される。シリアル・リンクはスループットを向上するために集約される。

40

【0201】

図19を更に参照すると、スイッチング・システムはスイッチ・ファブリック11と、スイッチ・ファブリック入力ポート15(15-1, ..., 15-k)に接続される入力スイッチ・アダプタ13(13-1, ..., 13-k)と、スイッチ・ファブリック出力ポート19(19-1, ..., 19-p)に接続される出力スイッチ・アダプタ17(17-1, ..., 17-p)とを含む。

【0202】

50



着信及び発信伝送リンク 21 (21-1, . . . , 21-g) 及び 23 (23-1, . . . , 23-r) は、それぞれ回線 (リンク) アダプタ 25 (25-1, 25-g) 及び 27 (27-1, . . . , 27-r) により、スイッチ・システムに接続される。伝送リンクは、ワークステーションや電話器などの接続ユニット (リンク指定 WS) に、或いはローカル・エリア・ネットワーク (リンク指定 LAN) に、更にサービス統合デジタル網 (ISDN) (リンク指定 ISDN) や他の通信システムに、回線交換またはパケット交換トラフィックを伝搬する。更に、プロセッサはスイッチ・アダプタ 13 及び 17 に直接接続される。回線アダプタ (LA) 及びスイッチ・アダプタ (SA) は、共通のインタフェースを有する。

#### 【0203】

入力スイッチ・アダプタでは、パケット交換及び回線交換インタフェースから様々なサービスが収集され、一様なミニパケット (幾つかの可能な固定長の 1 つを有する) に変換される。ルーティング情報を含むヘッダは、スイッチの要求出力ポート (及び出力リンク) を指定する。入力スイッチ・アダプタ内でのミニパケット・フォーマット及びミニパケット生成に関する詳細、及び出力スイッチ・アダプタ内でのパケット解除に関する詳細について、次に述べることにする。

#### 【0204】

スイッチ・ファブリックはミニパケットを、高速自己ルーティング相互接続ネットワークを介して、任意の入力ポートから任意の出力ポートに経路指定する。自己ルーティング・ネットワークの構造は、ミニパケットが競合無しに、内部的に同時に経路指定されること

#### 【0205】

スイッチング・システムの心臓部は、スイッチ・ファブリックである。2つの異なるインプリメンテーションが考慮され、これらについて次に述べる。1インプリメンテーションでは、スイッチ・ファブリックは、各入力ポートに対して、それぞれの入力ポートを全ての出力ポートに接続する自己ルーティング二分木を含み、(k個の入力ポートが提供される場合、) k個のこうした二分木が組み合わされて提供される。他のインプリメンテーションでは、出力RAMを有するバス構造が、各出力構造に対してスライスとして提供され、全ての入力ポートをそれぞれの出力ポートに接続する。(p個の出力ポートが提供される場合、) スwitch・ファブリックは、p個のこうしたスライスを組み合わせて提供される。

#### 【0206】

DASLについては、1999年6月11日出願の米国特許出願第09/330968号で述べられている。DASLインタフェースは、CMOSASICなどのパラレル・インタフェースからデータを受信し、パラレル・インタフェースからのビットを、少数の並列ビット・ストリームに区分化する。少数の並列ビット・ストリームが次に高速直列ストリームに変換され、これが伝送メディアを介して、他のモジュールの受信機に移送される。制御インピーダンスを有する差動ドライバが、データの直列ビット・ストリームを伝送メディアに駆動する。

#### 【0207】

DASLは、Nビット並列データとして表されるデータ・ストリームを、各々がnビット (但しnはNの分数) を有する複数の部分に構文解析し、データ・ストリームの各nビット部分を直列化し、直列化された各部分を複数の並列チャネルの対応するチャネルを介して転送し、データ・ストリームの各転送部分を非直列化し、データ・ストリームをNビット並列データとして復元する。

#### 【0208】

以上、図面を参照しながら、本発明の好適な実施例について述べてきた。説明の中で使用された特定の用語は総称的な意味を成すもので、限定的な意味で使用されるものではない。

#### 【図面の簡単な説明】

10

20

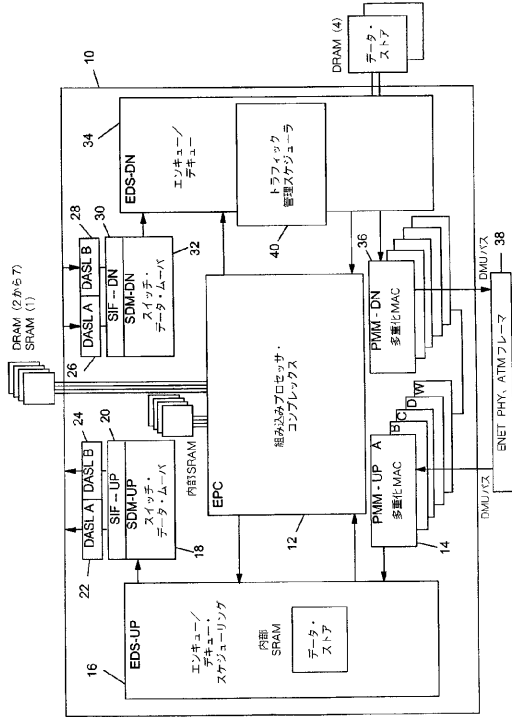
30

40

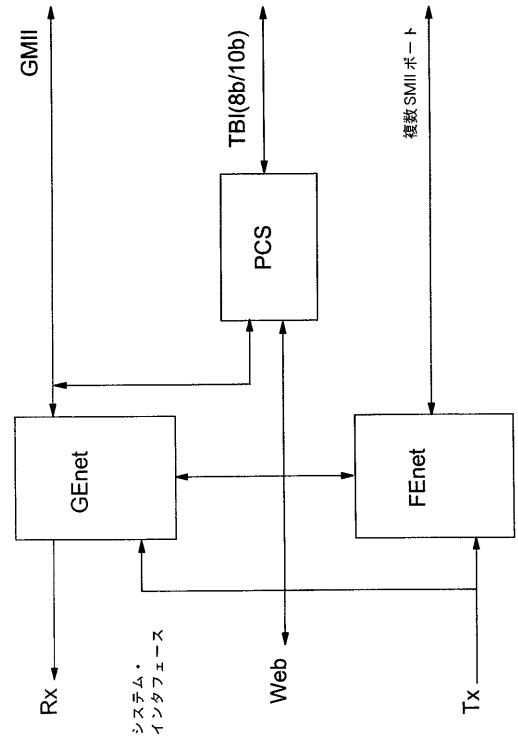
50

- 【図1】 本発明に従うインタフェース装置のブロック図である。
- 【図2】 MACのブロック図である。
- 【図3】 異なるシステム構成内の他のコンポーネントと相互接続されるインタフェース装置を示す図である。
- 【図4】 カプセル化されたガイド・フレームのフロー及び処理を示す図である。
- 【図5】 内部ガイド・フレームのフロー及び処理を示す図である。
- 【図6】 ガイド・セルの汎用フォーマットを示す図である。
- 【図7】 フレーム制御情報のフォーマットを示す図である。
- 【図8】 相関関係子のフォーマットを示す図である。
- 【図9】 コマンド制御情報フォーマットを示す図である。 10
- 【図10】 アドレッシング情報フォーマットを示す図である。
- 【図11】 構造アドレッシングの汎用形式を示す図である。
- 【図12】 アドレッシング、アイランド・エンコードを示す表である。
- 【図13】 組み込みプロセッサ・コンプレックスのブロック図である。
- 【図14】 組み込みプロセッサの概略図である。
- 【図15】 G×Hプロセッサの構造を示す図である。
- 【図16】 メモリ・コンプレックスのブロック図である。
- 【図17】 固定マッチ(FM)検索アルゴリズムのフローチャートである。
- 【図18】 ダイレクト・テーブルを使用する場合と、使用しない場合のデータ構造を示すフローである。 20
- 【図19】 Prismaなどのスイッチング・システムのブロック図である。
- 【図20】 CPのブロック図である。
- 【図21】 EDS-UP、EDS-DOWN及びEPCにおける、シングルチップ・ネットワーク・プロセッサの強調表示機能のブロック図である。
- 【符号の説明】
- 10、38 インタフェース装置
- 12 組み込みプロセッサ・コンプレックス(EPC)
- 14 多重化MACアップ(PPM-UP)、多重化MAC
- 16 エンキュー-デキュー-スケジューリング・アップ(EDS-UP)
- 18 スイッチ・データ・ムーバ・アップ(SDM-UP) 30
- 20、30 システム・インタフェース(SIF)
- 22 データ・アライン・シリアル・リンクA(DASLA)
- 24 データ・アライン・シリアル・リンクB(DASLB)
- 32 SDM-DN
- 34 EDS-DN
- 36 イーグレス・イーサネット(R)MAC
- 40 トラフィック管理スケジューラ

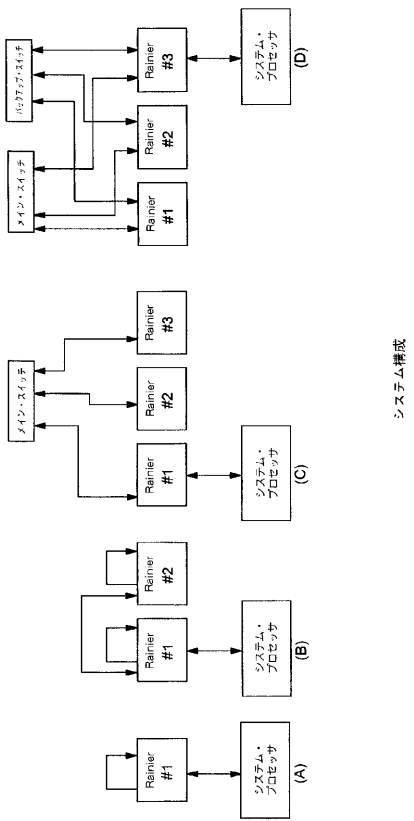
【 図 1 】



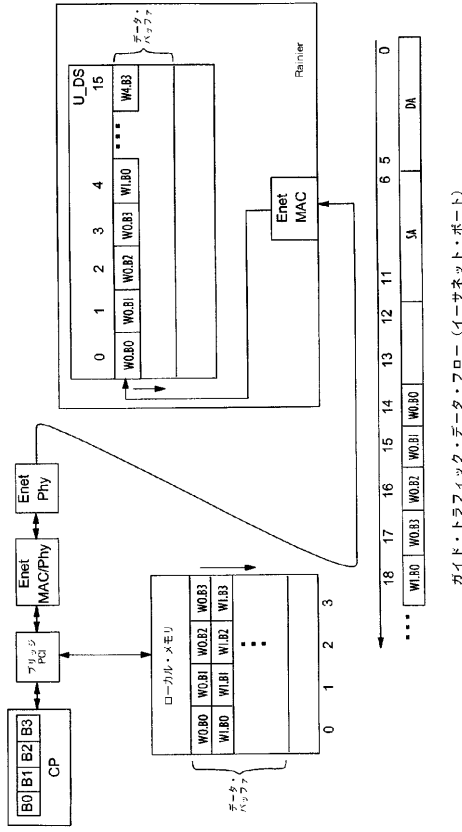
【 図 2 】



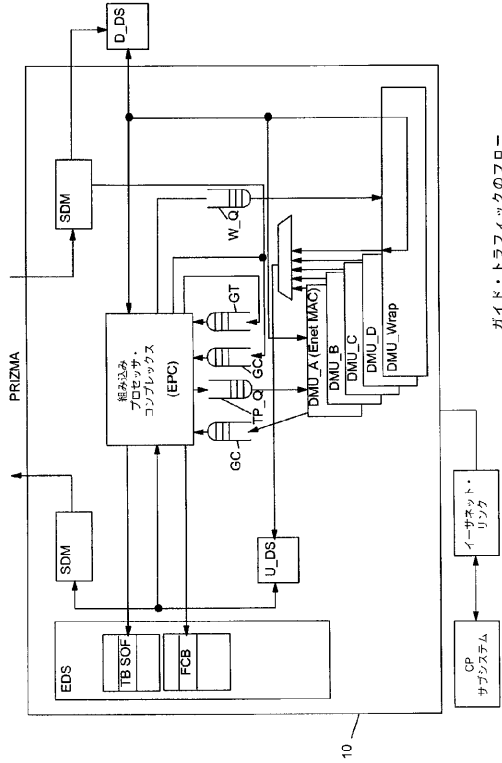
【 図 3 】



【 図 4 】



【 5 】



【 6 】

|            |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |            |    |    |    |    |    |   |   |   |   |   |   |   |   |   |   |
|------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|------------|----|----|----|----|----|---|---|---|---|---|---|---|---|---|---|
| 31         | 30 | 29 | 28 | 27 | 26 | 25 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15         | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| MSB, バイト 0 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | LSB, バイト 3 |    |    |    |    |    |   |   |   |   |   |   |   |   |   |   |

Word 0  
Word 1  
Word 2 End Word 0  
Word 3 End Word 1  
Word 4 End Word 2

【 7 】

|            |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |            |    |    |    |    |    |   |   |   |   |   |   |   |   |   |   |
|------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|------------|----|----|----|----|----|---|---|---|---|---|---|---|---|---|---|
| 31         | 30 | 29 | 28 | 27 | 26 | 25 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15         | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| MSB, バイト 0 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | LSB, バイト 3 |    |    |    |    |    |   |   |   |   |   |   |   |   |   |   |

Word 0

【 8 】

|            |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |            |    |    |    |    |    |   |   |   |   |   |   |   |   |   |   |
|------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|------------|----|----|----|----|----|---|---|---|---|---|---|---|---|---|---|
| 31         | 30 | 29 | 28 | 27 | 26 | 25 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15         | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| MSB, バイト 0 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | LSB, バイト 3 |    |    |    |    |    |   |   |   |   |   |   |   |   |   |   |

Word 1

【 9 】

|          |    |    |    |    |    |    |    |    |    |      |       |    |    |    |    |    |    |    |    |          |     |   |          |       |   |   |   |   |   |   |   |
|----------|----|----|----|----|----|----|----|----|----|------|-------|----|----|----|----|----|----|----|----|----------|-----|---|----------|-------|---|---|---|---|---|---|---|
| 1        | L  |    |    |    |    |    |    |    |    |      | 先7コード |    |    |    |    |    |    |    |    |          | res |   |          | GCタイプ |   |   |   |   |   |   |   |
| 31       | 30 | 29 | 28 | 27 | 26 | 25 | 24 | 23 | 22 | 21   | 20    | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11       | 10  | 9 | 8        | 7     | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| MSB,バイト0 |    |    |    |    |    |    |    |    |    | バイト1 |       |    |    |    |    |    |    |    |    | USB,バイト3 |     |   | USB,バイト3 |       |   |   |   |   |   |   |   |

コマンド制御情報フォーマット

Word 2 Cmd Word 0

【 1 1 】

|         |        |    |            |          |
|---------|--------|----|------------|----------|
| アイランドID | 構造アドレス | 23 | エレメント・アドレス | ワード・アドレス |
| 5       | 32     |    |            | 4        |

構造アドレスの汎用形式

【 1 0 】

|          |    |    |    |    |    |    |    |    |    |      |    |    |    |    |    |    |    |    |    |          |    |   |          |   |   |   |   |   |   |   |   |
|----------|----|----|----|----|----|----|----|----|----|------|----|----|----|----|----|----|----|----|----|----------|----|---|----------|---|---|---|---|---|---|---|---|
| アドレス     |    |    |    |    |    |    |    |    |    | アドレス |    |    |    |    |    |    |    |    |    | USB,バイト3 |    |   | USB,バイト3 |   |   |   |   |   |   |   |   |
| 31       | 30 | 29 | 28 | 27 | 26 | 25 | 24 | 23 | 22 | 21   | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11       | 10 | 9 | 8        | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| MSB,バイト0 |    |    |    |    |    |    |    |    |    | バイト1 |    |    |    |    |    |    |    |    |    | USB,バイト3 |    |   | USB,バイト3 |   |   |   |   |   |   |   |   |

アドレス情報フォーマット

Word 3 Cmd Word 1

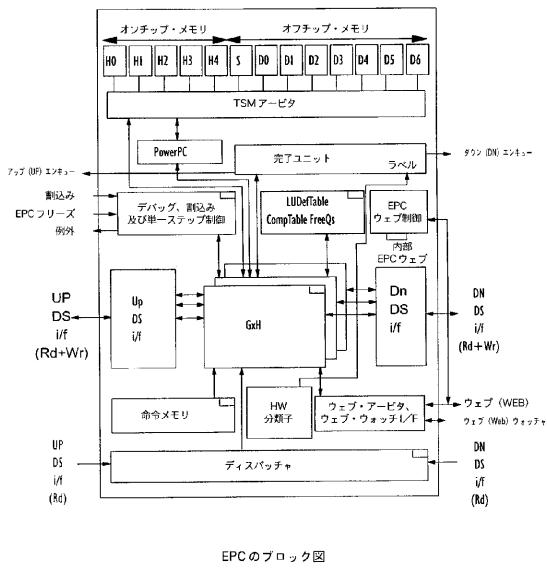
【 1 2 】

| アイランドID         | アイランド名            |
|-----------------|-------------------|
| '00000'b        | アップ・チャタ・ストア*      |
| '00001'b        | アップPMM            |
| '00010'b        | アップEDS            |
| '00011'b        | アップSDM            |
| '00100'b        | 組み込みプロセッサ・コンプレックス |
| '00101'b        | SPM               |
| '01000'b        | 予約済み              |
| '01001'-01111'b | 制御メモリ*            |
| '10000'b        | 予約済み              |
| '10001'b        | ダウン・チャタ・ストア*      |
| '10010'b        | ダウンPMM            |
| '10011'b        | ダウンEDS            |
| '10100'b        | ダウンSDM            |
| '10101'b        | 構成レジスタ            |
| '10110'b        | DASL              |
| '10111'b        | 予約済み              |

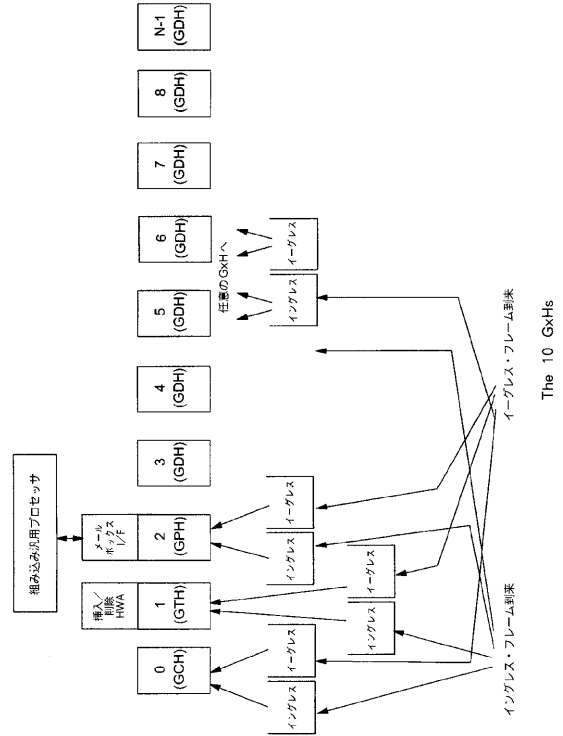
\*これらのアイランドはウェーブ・インタフェースを介してアクセスできない

アドレスリング、アイランド・エンコーディング

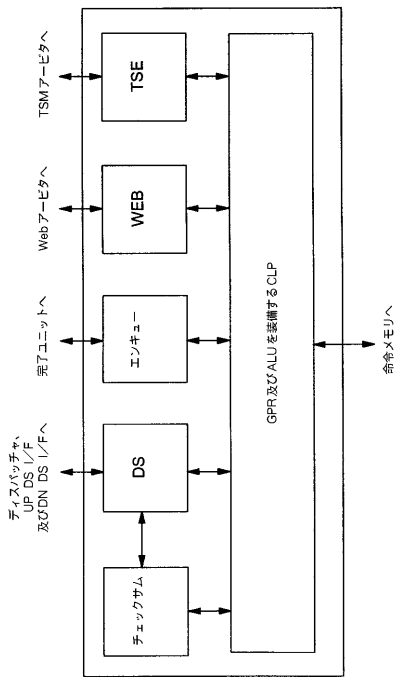
【 図 1 3 】



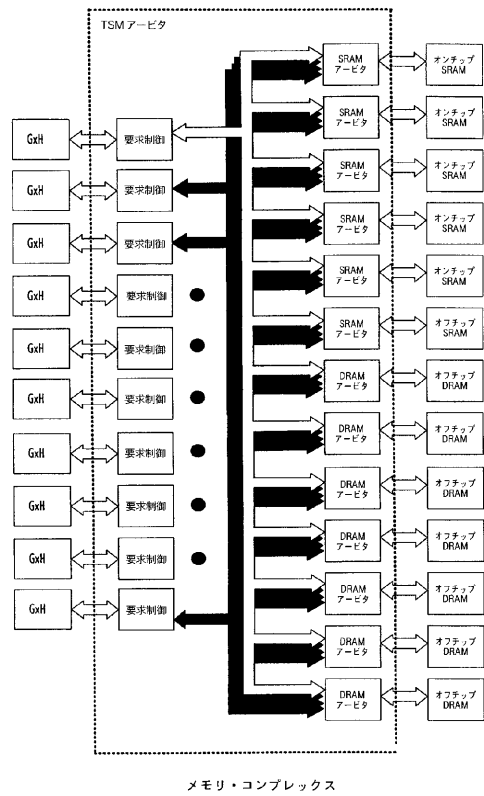
【 図 1 4 】



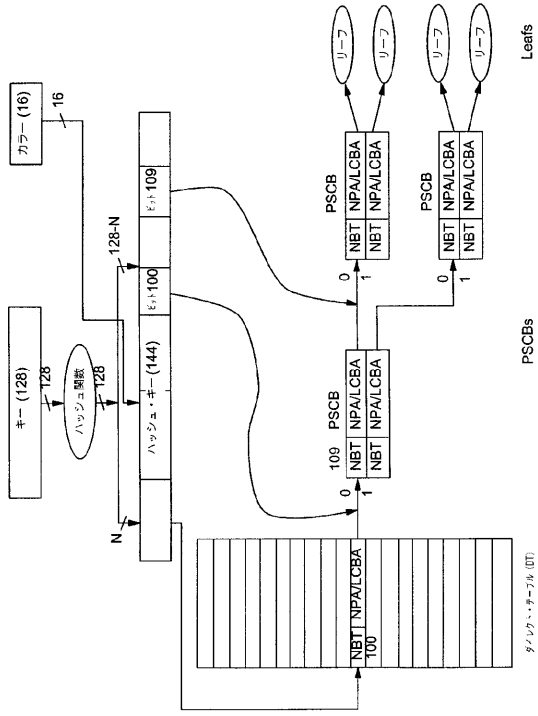
【 図 1 5 】



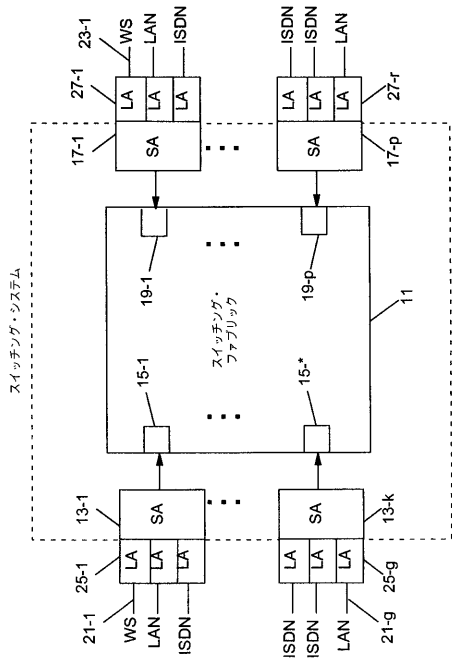
【 図 1 6 】



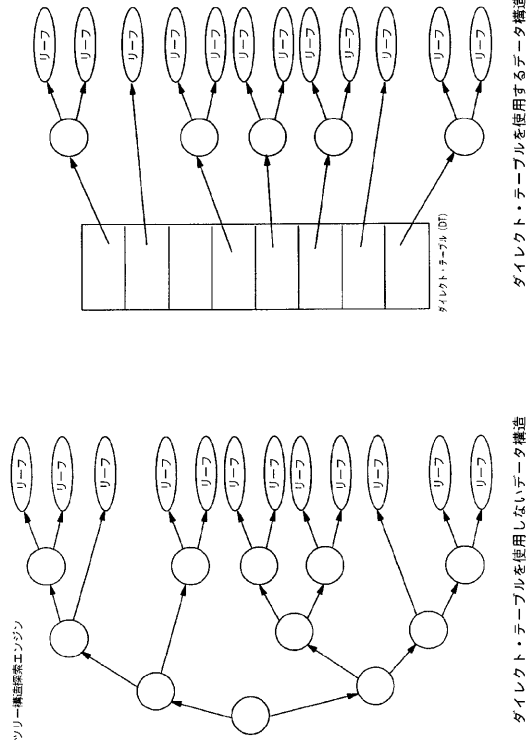
【 図 17 】



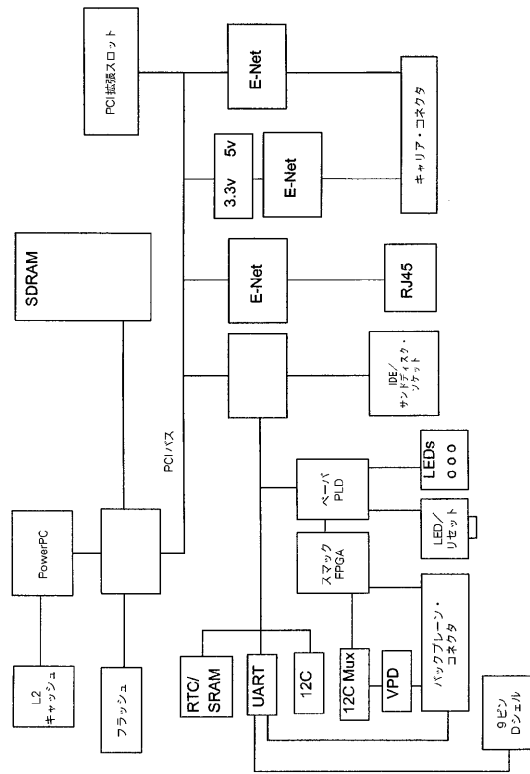
【 図 19 】



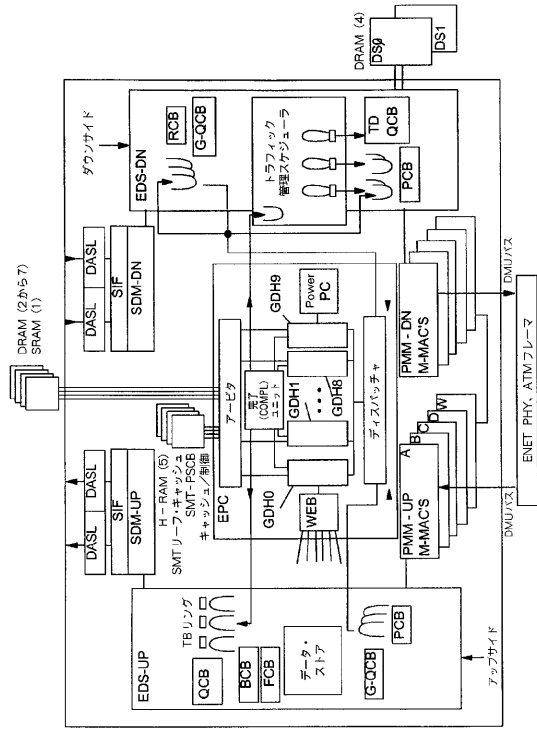
【 図 18 】



【 図 20 】



【 図 2 1 】





## フロントページの続き

- (72)発明者 アレン、ジェームス、ジュニア  
アメリカ合衆国27606、ノース・カロライナ州ラーレー、トバーモリー・レーン 2800
- (72)発明者 バス、ブライアン、ミッチェル  
アメリカ合衆国27502、ノース・カロライナ州アベックス、オールド・スターブリッジ・ドライブ 4021
- (72)発明者 カルピンナック、ジーン、ルイス  
アメリカ合衆国27511、ノース・カロライナ州カーリー、スプリング・ホロー・レーン 112
- (72)発明者 ガウアー、サントッシュ、プラサッド  
アメリカ合衆国27614、ノース・カロライナ州ラーレー、ピークトン・ドライブ 5516
- (72)発明者 ヘデス、マルコ、シイ  
アメリカ合衆国27612、ノース・カロライナ州ラーレー、グラント・メナー・コート 4109、ナンバー 308
- (72)発明者 シーゲル、マイケル、スティーブン  
アメリカ合衆国27613、ノース・カロライナ州ラーレー、ロワリー・ドライブ 10625
- (72)発明者 ヴァープランケン、ファブリス、ジーン  
フランス06610、ラ・ゴード、ルート・ド・カーニュ 9152

審査官 吉田 隆之

- (56)参考文献 特開平9 - 270804 (JP, A)  
特開昭62 - 115949 (JP, A)  
特表2000 - 513546 (JP, A)  
国際公開第99 / 37117 (WO, A1)  
特開平11 - 205339 (JP, A)  
特開平7 - 152591 (JP, A)  
Product Overview IBM Network Processor (IBM32NPR161EPXCAC133), 1999年11月 4日  
1998年信学通ソ大会 SB-7-3  
IBM Journal on Research and Development, Vol.47 No.2/3, p177-193  
IEEE Network, Vol.17 No.4, p12-20

- (58)調査した分野(Int.Cl., DB名)  
H04L 12/00