

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3816530号

(P3816530)

(45) 発行日 平成18年8月30日(2006.8.30)

(24) 登録日 平成18年6月16日(2006.6.16)

(51) Int. Cl.	F I
HO 4 L 12/56 (2006.01)	HO 4 L 12/56
HO 4 L 12/28 (2006.01)	HO 4 L 12/28
HO 4 L 12/46 (2006.01)	HO 4 L 12/46
HO 4 L 12/66 (2006.01)	HO 4 L 12/66

請求項の数 10 (全 16 頁)

(21) 出願番号	特願平9-530403	(73) 特許権者	富士通株式会社
(86) (22) 出願日	平成9年2月19日(1997.2.19)		神奈川県川崎市中原区上小田中4丁目1番1号
(65) 公表番号	特表平11-504496	(74) 代理人	弁理士 石田 敬
(43) 公表日	平成11年4月20日(1999.4.20)	(74) 代理人	弁理士 土屋 繁
(86) 国際出願番号	PCT/US1997/002938	(74) 代理人	弁理士 戸田 利雄
(87) 国際公開番号	W01997/031462	(74) 代理人	弁理士 西山 雅也
(87) 国際公開日	平成9年8月28日(1997.8.28)		
審査請求日	平成16年1月19日(2004.1.19)		
(31) 優先権主張番号	08/603,926		
(32) 優先日	平成8年2月22日(1996.2.22)		
(33) 優先権主張国	米国(US)		

最終頁に続く

(54) 【発明の名称】 低い待ち時間、高いクロック周波数、プレジオ非同期 パケット・ベースクロスバー・スイッチング・チップ・システム及び方法

(57) 【特許請求の範囲】

【請求項1】

各フレームが複数のデータワードを含む、データのフレームを受けるための複数の入力ポートと、

第1の出力ポートと前記第1の出力ポートとは異なる第2の出力ポートを含む、データのフレームを送るための複数の出力ポートと、

少なくとも1つの入力ポートがデータワードを少なくとも2つの入力バッファへ送るように結合され、入力ポートで受けたデータワードを格納するための複数の入力バッファと、複数の入力と出力を有しており、各入力バッファは前記入力バッファ専用のクロスバースイッチの入力に結合され、各出力ポートは前記出力ポート専用のクロスバースイッチの出力に結合されており、スイッチ可能に如何なる入力バッファと出力ポートをも結合させるクロスバースイッチと、

少なくとも第1のアービタサブシステムと、前記第1のアービタサブシステムとは異なる第2のアービタサブシステムを含む、複数のアービタサブシステムを具備し、前記第1のアービタサブシステムはクロスバースイッチを通して入力バッファの1つから前記第1の出力ポートへの第1の接続パスを選択するように設けられており、前記第2のアービタサブシステムはクロスバースイッチを通して入力バッファの1つから前記第2の出力ポートへの第2の接続パスを選択するように設けられている、ことを特徴とする近隣の回路と通信するためのルータ回路。

【請求項2】

10

20

入力ポートから半ワードを受信して全ワードを入力バッファへ送るように結合され、1つのクロックサイクル内で2度提供された半ワードを、サイクルごとに1度クロックされた全ワードに変換するデマルチプレクサをさらに具備する請求項1に記載のルータ回路。

【請求項3】

前記アービタ・サブシステムは、前記第1の接続パスを1クロックサイクル以内で選択し、前記第2の接続パスを1クロックサイクル以内で選択する請求項1に記載のルータ回路。

【請求項4】

同期回路をさらに具備する請求項1に記載のルータ回路。

【請求項5】

前記同期回路は前記入力バッファからの読み出し及び書き込みを行うリード及びライト・ポインタを含む請求項4に記載のルータ回路。

【請求項6】

前記リード及びライト・ポインタは同じ入力バッファを特定する請求項5に記載のルータ回路。

【請求項7】

リード・ポインタはローカルクロックドメイン上で動作し、ライト・ポインタは入来クロックドメイン上で動作する、請求項5に記載のルータ回路。

【請求項8】

ルータ回路における複数の入力バッファの1つから複数の出力ポートの1つへ情報を送信する方法であって、前記複数の入力バッファと、前記複数の出力ポートの少なくとも第1及び第2の出力ポートと、入力バッファを複数の出力ポートへ結合するクロスバースイッチと、少なくとも第1のアービタサブシステムと第2のアービタサブシステムを含む複数のアービタサブシステムを具備し、前記方法は、

入力バッファの1つで情報を受信すること、

第1の出力ポートと、前記第1の出力ポートとは異なり宛先出力ポートとして情報を受信するための第2の出力ポートの一つを決定すること、

第1の出力ポートが宛先出力ポートの場合には、第1のアービタサブシステムに前記入力バッファのアドレスを提供し、そして第2の出力ポートが宛先出力ポートの場合には、前記第1のアービタサブシステムとは異なる第2のアービタサブシステムに前記入力バッファのアドレスを提供すること、

提供されたアドレスに応じて、第1の出力ポートが宛先出力ポートの場合には、第1のアービタサブシステムがクロスバースイッチを通して前記入力バッファから第1の出力ポートへの第1の接続パスを選択するようにし、そして第2の出力ポートが宛先出力ポートの場合には、第2のアービタサブシステムがクロスバースイッチを通して前記入力バッファから第2の出力ポートへの第2の接続パスを選択するようにすること、

第1の出力ポートが宛先出力ポートの場合には、アドレス入力バッファを第1のアービタサブシステムに対応する第1の出力ポートと接続するようにクロスバースイッチを構成し、そして第2の出力ポートが宛先出力ポートの場合には、アドレス入力バッファを第2のアービタサブシステムに対応する第2の出力ポートと接続するようにクロスバースイッチを構成すること、

から成ることを特徴とする方法。

【請求項9】

さらに、第1の出力ポートが宛先出力ポートの場合には、前記入力バッファで受信した情報の優先度を第1の出力ポート専用の第1のアービタサブシステムに提供し、そして第2の出力ポートが宛先出力ポートの場合には、前記入力バッファで受信した情報の優先度を第2の出力ポート専用の第2のアービタサブシステムに提供することを含む、請求項8記載の方法。

【請求項10】

さらに、第1のアービタサブシステムに第1の接続パスを選択させ、そして第2のアービ

10

20

30

40

50

タサブシステムに第2の接続パスを選択させること、は1クロック以内で行われる、請求項9記載の方法。

【発明の詳細な説明】

発明者

Thomas M. Wicki, Jeffrey D. Larson, Albert Mu および Raghunath Sastri

関連出願の相互参照

本願発明の主題は、下記に掲げる出願の主題と関連している。

米国特許出願No. 08/605,677、“非同期パケット交換”の名称で、Thomas M. Wicki, Patrick J. Helland, Takeshi Shimizu, Wolf-Dietrich Weber および Winfried W. Wilcke によって1996年2月22日に出願され、米国特許No. 5,959,995として1999年9月28日に発行、

10

米国特許出願No. 08/605,676、“ダイナミックなネットワーク・トポロジー探査のシステムおよび方法”の名称で、Thomas M. Wicki, Patrick J. Helland, Wolf-Dietrich Weber および Winfried W. Wilcke によって1996年2月22日に出願され、米国特許No. 5,740,346として1998年4月14日に発行、

米国特許出願No. 08/603,880、“パケットスイッチングネットワーク内の経路選択素子出力アクセス用調整方法および装置”の名称で、Jeffrey D. Larson, Albert Mu および Thomas M. Wicki によって1996年2月22日に出願され、米国特許No. 5,892,766として1999年4月6日に発行、

20

米国特許出願No. 08/604,920、“電圧の揺れを少なくし、かつ、内部のブロック化データパスを生じさせないクロスバー交換機およびその方法”の名称で、Albert Mu および Jeffrey D. Larson によって1996年2月22日に出願され、米国特許No. 5,991,296として1999年11月23日に発行、

米国特許出願No. 08/603,913、“ネットワークエレメント間のデータ転送を制御するシステムおよび方法”の名称で、Thomas M. Wicki, Patrick J. Helland, Jeffrey D. Larson, Albert Mu, Raghunath Sastri および Richard L. Schober, Jr. によって1996年2月22日に出願され、米国特許No. 6,003,064として1999年12月14日に発行、

30

米国特許出願No. 08/603,911、“相互接続の障害検出およびその位置特定方法および装置”の名称で、Raghunath Sastri, Jeffrey D. Larson, Albert Mu, John R. Slice, Richard L. Schober, Jr. および Thomas M. Wicki によって1996年2月22日に出願され、米国特許No. 5,768,300として1998年6月16日に発行、

米国特許出願No. 08/603,923、“多重ワード通信におけるエラー検出方法および装置”の名称で、Thomas M. Wicki, Patrick J. Helland および Takeshi Shimizu によって1996年2月22日に出願され、米国特許No. 5,931,967として1999年8月3日に発行、

40

米国特許出願No. 08/603,882、“正のソース帰還をそなえたクロック動作されるセンス増幅器”の名称で、Albert Mu によって1996年2月22日に出願され、米国特許No. 5,615,161として1997年3月25日に発行、

参考として、上記の出願の全てを本願発明の全体に亘って取り入れている。

発明の技術分野

この発明は、非同期及びプレジオ・クロナス(plesiochronous)パケット・ベース、クロスバー・スイッチング・システム及び方法に関する。

背景

50

コンピュータ・メッシュにおいてノードに接続する特定な従来のルータ・システムは、パケットの第1の部分の出力送信の開始前に、全体のパケットの存在を考慮するための各ルータ・エレメントへの要求により、待ち時間で制限される。多くのこのようなルータ・システムは、選択された入力バッファから特定の出力ポートへのアクセスのアービトレーションの複雑性に起因して、ルータ出力ポートにおける輻輳により、さらに厄介なものとなる。さらに、多くのルータ・システムは、データ、パケット、フレームの伝送処理能力を減ずる同期システムを生じる複数の待ち時間を含む。

ヘッダ情報を含むデータの packets を伝送する半導体ルータ・チップにおいて、待ち時間、ルータ・チップのネットワークを持つシステムにおける単一チップを経てデータの伝送に要求される時間、を最小にすることは望ましいことである。

10

さらに、単一の半導体チップ上に存在するルータ・システムを設けることは望ましいことである。

発明の要約

本発明によれば、ルータ・システムは、クロスバー・スイッチを経て所定の出力ポートとともにアービトレーションにより選択された入力バッファの相互接続を確立する。本発明の実施形態によれば、各出力ポートは、その出力ポートに接続するための最も高いプライオリティの入力バッファを選択するために専用アービタを有する。本発明の方法によれば、フレーム・ワードは、クロスバー・スイッチを経て選択された入力バッファ及び出力ポートを経て独立に流れる。本発明によるストリームは、フレーム・ヘッダの解析により確立されたパスの限定とともに開始する。並列フレーム・エンベロープ信号は、特定フレーム

20

への所定の出力ポートのデディケーションを終える。本発明によるフレーム伝送は、プレジオ・クロナス、即ち、同期に近いもの、である。本発明の一つの実施形態によれば、フレーム伝送は非同期である。簡単にいえば、本発明のルータ・システムは同期ではない。フレームは1つのクロックとともに到達し、他のクロックとともにクロックアウトされる。従って、伝送は、「プレジオ非同期」、即ち、同期又は非同期に近いもの、と考えることができる。

本発明によれば、フレーム伝送はルータ内の分配同期アーキテクチャとともに完成され、ルータでは、入力バッファはリード及びライトポイントの読出し及び不一致のためにチェックされ、出力ポートの識別及びプライオリティ・レベルの情報は多段レジスタに格納され、そしてバッファ・ステータス・メッセージはORアキュミュレートされる。

30

本発明によれば、半導体ルータ・チップ・システムは、データの伝送において高いクロック周波数と低い待ち時間を持つストリームラインのアーキテクチャを含む。本発明によれば、低い待ち時間は、誤り訂正により伴われなかった選択されたエラーチェックを除去することを含んで、各パケット上で行われる動作を減少させることにより達成される。本発明は、待ち時間を最小にする同期方法を含む。本発明によれば、内部チップのデータ処理は、ルータ・チップ間の外部通信として2倍幅のデータパス上で完成される。本発明のデータパスは、内部ブロッキング無しで完全に接続されたクロスバー・スイッチを含み、入力バッファの各バッファ・エレメントに対するクロスバーへの専用の接続を設ける。これは互いから出力ポートに対してアービトレーション・ユニットと減結合する。他の特徴は、クリティカル・パスの外で扱われるフロー・コントロールプロトコルを使用し、本発明

40

によるバーチャル・カッター・ルーチンをサポートする。さらに、本発明によれば、データパスは、データパケットの開始と終端をマークするために帯域外シグナリングを含み、それにより、データストリームのデコードを顕著に減少する。加えて、本発明によれば、各パイプライン・ステージにおけるパスの長さは、実行した動作を減少することにより最小化され、それにより高周波数におけるチップ動作を可能にする。

本発明の一つの実施形態によるルータ・システムは単一の半導体チップ上に存在する。

【図面の簡単な説明】

図1は、本発明によるルータ・チップのブロック図である。
図2は、1サイクル・アービトレーションを含む本発明による多段ステージ・パイプライン方法の図である。

50

図 3 A は、本発明による入力バッファ・エレメントの同期アーキテクチャの図である。

図 3 B は、本発明による制御同期及びフレーム・ヘッダ・デコーダ・アーキテクチャの図である。

図 3 C は、本発明によるバッファ・ステータス及びアキュミュレーション回路の同期アーキテクチャの図である。

図 3 D は、本発明によるルータ・システムの動作フローチャートである。

図 4 A は、スイッチング・バスさらにメタル及びクロスポイント・スイッチを含む、本発明によるクロスバー・スイッチの回路図である。

図 4 B は、本発明によるデータバスのブロック図である。

図 5 A は、本発明によるフロー・コントロールバスの第 1 部分のブロック図である。

10

図 5 B は、本発明によるフロー・コントロールバスの第 2 部分のブロック図である。

図 5 C は、本発明によるフロー・コントロールバスの第 1 部分の流れ図である。

図 5 D は、本発明によるフロー・コントロールバスの第 2 部分の流れ図である。

図 6 は、本発明によるルータ・チップ・システムのブロック図である。

図 7 A は、本発明による伝送フレームのブロック図である。

図 7 B は、本発明による方法のフローチャートである。

図 7 C は、本発明によるアービタ機能のフローチャートである。

図 7 D は、本発明によるフレーム・エンベロープ信号のタイミング図である。

図 7 E は、本発明によるパケット・バーチャル・カッターのタイミング図であり、ヘッダ及びデータ情報の送信は情報の独立のワードでずらされる。

20

発明の詳細な説明

図 1 は、本発明の一実施形態による単一の半導体チップ上に存在するルータ・システム 3 のブロック図である。ルータ・システム 3 は、その 1 つを図 1 に示す複数の入力ポート 104 を含むバーチャル・カッター・パケットスイッチを含む。

本発明の一実施形態によれば、各入力ポート 104 は 34 ビット幅である。データは、立上がり及び立下がり入力クロック・エッジ上で入力ポートにて受信される。各ポート 104 は対称的である。各ポート 104 は、レジスタ 107 と第 1 及び第 2 並列レジスタ 108 a , 108 b を含むデマルチプレクサ 400 を経て、多段エントリ・入力バッファ・ユニット 135 に接続される。レジスタ 107 は 1 クロックサイクルの第 1 部分にて半ワードを受け、1 クロックサイクルの次の部分の間でレジスタ 107 の内容が第 1 の並列レジスタ 108 a にエンターされ、残りの半ワードはポート 104 から第 2 並列レジスタ 108 b に直接エンターされる。並列レジスタ 108 a , 108 b の内容は、適切なバッファ・エレメント 135 に挿入するためにデマルチプレクサ 130 ' にエンターされる。ルータ・システム 3 は、キュー・アービタ・ユニット 146 及び第 1 及び第 2 フロー・コントロールシステム 147 a , 147 b を含むフロー・コントロールユニット 147 を含む。

30

ルータ・システム 3 は、フレーム・エンベロープ (F _ E N V) 入力ライン 102 及びデータ・ノット・ステータス (D / N S) 入力ライン 103 に接続された入力レジスタ 105 をさらに含む。入力レジスタ 105 は、同期及びフレーム・ヘッダ・デコーダ回路 128 を制御するためにフレーム・エンベロープ及び D / N S 信号を提供する入力制御回路 106 に接続され、また、バッファステータス及びアキュミュレータ回路 129 に接続され

40

、さらに入力バッファ制御回路 130 に接続される。フロー・コントロールレシーバ回路 147 は、以下に詳しく説明するように、バッファステータス及びアキュミュレータ回路 147 からのバッファステータス信号及び情報を受けるためにバッファステータス及びアキュミュレータ回路 147 a に接続される。フロー・コントロールレシーバ回路 147 a は、そのステータスに関してアービタ 146 から情報を受け、フレーム・ヘッダ・モディファイア回路 150 及びアービタ 146 に情報を提供する。入力制御回路 106 は、入力バッファ・エレメント 135 に関して書込み動作を制御する書込み制御回路 131 に信号を提供する。アービタ 146 はクロスバー・スイッチ 144 ' を制御し、入力バッファ・エレメント 135 のフレーム部分の読出しをレジスタ 138 に許可するように読出し制御回路 136 を制御する。レジスタ 149 はクロスバー・スイッチ 144 ' からフレーム部

50

分を受け、それらをフレームヘッダ・モディファイア150に提供する。レジスタ138は、ルータ・チップ及びノードのメッシュの隣接ルータ又はノードにフレーム・エンベロープ(FENV)及び(DNS)信号を生じる出力制御回路148に情報を追加提供する。マルチプレクサ151は、フレームヘッダ・モディファイア150から及びフロー制御トランスミッタ回路147bからモディファイドヘッダとともにフレームを受信する。マルチプレクサ151は、各クロックサイクルごとにワード幅フレーム部分を受け、立上がり及び立下がりクロックエッジにて半ワード幅フレーム部分を生じる。ルータ・システム3は、さらに入力ポート当たりのメッシュ・リンククロック範囲110及びローカルクロック範囲109を含む。メッシュ・リンククロック範囲110は入力クロック103'に従ってクロックされる。ローカルクロック範囲109はローカルクロック203に従ってクロックされる。

10

本発明の一実施形態によれば、ルータ・システム3はクロック周波数200MHzで動作することができ、マルチプロセッサを越えて相互接続する。待ち時間許容差のリンクレベルフロー制御及びフレキシブル・マルチプル・クロック範囲のプレジオクロナス通信と共に、ルータ・システム3は、数メータのオーダの距離で分離された処理ノードの相互接続をサポートする。複数のルータ・チップ3は計測可能に相互接続することができる。

図2は、1サイクルのアービトレーションを含む本発明による多段ステージ・パイプライン方法の図である。図2に示すように、本発明の方法はメッシュ・リンククロック範囲110からローカルクロック範囲109への情報の同期を含む。加えて、本発明の方法は、アドレス伝搬を処理しデコードすることを含む。入力ポートにて受けたルート情報はキュー・アービタユニット106に一斉送信される。本発明の方法はさらにアービトレーションを含む。本発明の方法はさらにクロスバー・スイッチ144'を経たデータ伝搬を含む。本発明の方法はさらにフレームヘッダ・モディフィケーションを含む。

20

本発明によれば、データ、クロック及び帯域外制御信号は、n個の入力ポートの1つに到達する。データは、リンク帯域幅を最適化するために到達クロックの両方のエッジにラッチされる。データはスイッチ・ユニットの内部でデータパス幅にマルチプレックスされる。拡張されたデータパス幅はクロックレートの半分で内部論理動作を許容する。データは、m個の入力バッファ・エレメントの中の選択された1つを経たストリームにてヘッダ及びデータ部分を含むフレームにて通信される。送り側は、送信に先立って受けるフロー制御情報に基づいて、m入力バッファ・エレメントのなかの特定の1つを選択する。各バッファ・エレメントはクロスバー・マトリクスにて専用パスに接続され、それにより特定の出力ポートに向けられた1つの入力ポートから、これらの出力ポートがアイドルである場合に、フレームのブロックを除去する。各バッファ・エレメントは、競合が生じた場合に最大サイズのフレームを格納するために十分に大きい。スイッチング・ユニットはパーティシャル・カットスルー・ルーチンをサポートする。入力バッファは、入力クロック信号に従って入力データを受けることにより及び出力クロック信号でデータを生じることにより、入来からローカルクロック範囲へのデータに同期する。各フレームヘッダにおけるルーチン情報はフレームヘッダ・デコーダ128に提供される。ルーチン情報の評価の後、入力バッファ・エレメントのアドレスは、n出力ポートのなかの選択された1つに対応するアービトレーション・ユニット146に格納される。各バッファ・エレメントは、クロスバー・マトリクスへの専用パスを持ち、選択された出力ポートへスイッチされるフレームを許可する。フレーム・ヘッダ・モディフィケーションは、現在のルーチン・ステップを除去し、次の下流近隣に対する新規な入力バッファ・エレメントのアドレスを挿入することにより完成される。入力可能な適切な入力バッファ・エレメント135は次の下流近隣から到達するフロー制御情報から抽出される。データは、2倍幅の内部バスから外部バス幅にマルチプレックスされる。クロック及び外帯域制御信号は加えられ、並列にデータとともに送信される。

30

40

本発明によるフロー制御方法は、ここで相互参照した特許出願において詳しく述べられているように、ルータ・システム3のクレジット基準のフロー制御動作に基づく。ルータ・チップ3は、空のバッファが情報フレームの受容のために受け入れ可能なときのみフレー

50

ムを送信する。従って受信ルータ・チップ又はノードは、バッファ・ステータスメッセージを送信することにより、いくつかのバッファが受入れ可能かを送信ルータ・チップに通知する。これらのメッセージはレギュラー・データフレームでインターリーブされ、アイドルリンク上のデフォルト・トラフィックとなる。フロー制御は、過渡的なエラーに対して強く、自己治癒力のあるプロトコル保証に追従する。本発明によるルータ・システム 3 は、伝送エラーに起因する永続的損失から入力バッファを回復するためにタイムアウト機構を含む。バッファ・ステータス・メッセージの停止を防止するために、バッファ・ステータス・メッセージの送信は、参照した特許出願にて詳細にしたように、もし同期はずれの条件が存在するならば、データフレーム上でプライオリティを受ける。

図 3 A は、本発明による、フレームを保持する入力バッファ・エレメント回路 1 3 5 の同期アーキテクチャの図である。特に、入力バッファ・エレメント回路 1 3 5 は、ローカルクロック範囲で動作するグレイ (Gray) ・リード・ポインタ 1 3 5 a と、入来クロック範囲で動作するグレイ・ライト・ポインタ 1 3 5 b と、フレーム又はフレーム部分を保持する特定の F I F O バッファ・エレメントが空きか否かを決定するためにチェックする比較エレメントと、を含む。同期は、安定したデータが既に読み出されライト・ポインタが進んでいるときに満たされる条件である。ポインタに対してポインタ 1 3 5 a , 1 3 5 b のグレイ・コードを使用することにより、圧縮エラーが回避される。バッファ・エレメントは、グレイ・リード・ポインタ及びグレイ・ライト・ポインタの両方が等しい時、本発明によれば空きと考慮され、特定のバッファに関してリード及びライト動作が完了したことを示し、結果的にリード動作が停止される。リード及びライト動作の完了は一連のデータ同期を示す。

図 3 B は、制御同期の同期アーキテクチャ及びフレーム・ヘッダ・デコーダ回路 1 2 8 の図である。特に、制御同期及びフレーム・ヘッダ・デコーダ回路 1 2 8 は、単一の F I F O を構成する第 1 及び第 2 出力ポート情報及びプライオリティレベル情報 F I F O 1 2 8 a , 1 2 8 b を含む。もしこれらのレジスタの 1 つが全ての条件であれば、ライトポインタ 1 2 8 c 及びリードポインタ 1 2 8 d は、リード又はライト動作を完了するために一方又は他方のレジスタをポイントすることができる。同期は、バッファが安定したリード及びライト動作の獲得に固執するので発生する。即ち、アキュミュレータは、読み出され書き込まれるバッファ・ステータス・メッセージに対して連続的に開放している。

図 3 C は、バッファ・ステータス及びアキュミュレーション回路 1 2 9 のアーキテクチャの図である。特に、ステータス及びアキュミュレーション回路 1 2 9 はアキュミュレーティング O R バッファを含み、アキュミュレーティング O R バッファは、ビット位置にて拡張され、O R 機能の同期にてメッセージをアキュミュレートすることにより選択ビット位置におけるメッセージ又はメッセージ指示を受けることができる。

図 3 D は、本発明によるルータ・システム 3 の動作フローチャートである。特に、本発明による各ルータ・システム 3 は、データフレームとともに到達するクロック信号により駆動される 6 つの独立入来クロック範囲とインタフェースする。データ、クロック及び制御信号は n 入力ポートにおいて受信される (3 0 0) 。各ルータ・システム 3 は、グローバル・ロジック及び複数の出力ポートの間で共用されるローカルクロック範囲を持つ。F I F O 同期はフレームデータに対して各ルータ・チップにてインプリメントされる。相互接続されたルータ・チップは各々独立のクロック源をもつ。従って、本発明による相互接続されたルータは受信及び送信クロックの間の僅かな周波数ダイバシティに適合する。データはデマルチプレックスされ (3 0 1) 、各々がクロスバー・マトリクスに対して専用のパスをもつ m 入力バッファ・エレメントの選択された 1 つに与えられる。入力及びローカルクロック範囲の間の同期は、フレームデータの同期を実行し出力ポートの競合のためのメモリを提供する入力バッファ・エレメントにて完了する。入力バッファ・エレメント・フラグの同期は、準安定で固定したラッチにより非同期にサンプルされたグレイ・コード・ポインタと比較することにより達成される。加えて、本発明によれば、ヘッダは出力先ポートを示す情報を含む。特に、フレーム・ヘッダ・カレント・ルーティン情報は、入力バッファ・エレメント・アドレスを決めるためにフレーム・ヘッダ・デコーダに提供され

10

20

30

40

50

(302)、特定の入力バッファ・エレメント及び含まれる情報を選択するために、n出力ポートの中の所望の1つに対応する適切なアービタを可能とする。ヘッダはさらにプライオリティを示す情報を提供する。本発明によれば、ヘッダは2入力FIFOに同期される。ヘッダは連続的なサイクルでバック・ツー・バックに到達することができず、FIFOのオーバーフローは、データが確実に読まれる前に起こすことはできない。さらに、バッファ・ステータス・メッセージのみが、FIFOがオーバーフローするような十分なレートで到達することができるので、本発明によれば、バッファ・ステータス・メッセージは、FIFOの使用を避けるローカルクロック範囲でサンプリングするためにアキュムレートされる。本発明によれば、ルータ・システム3はクロスバー構成を含み、格納された各パケットは入力ポートをアクセスするために専用化される。クロスバーに対して入力での競合はない。本発明の結果として、内部ブロッキングは除去され、アービタレーションは1つのクロックサイクルで完了される。特に、アービタは入力バッファの内容のプライオリティに基づいて許可信号を発行する(303)。3つのキュークラス、即ち、プライオリティに対する3つのFIFOとともに出力ポート当たりの1つのアービタ146は、所定のプライオリティ内でのスタベーション(starvation)を避けるために使用される。本発明によるFIFOの順番は、比較的満たされたバッファからのフレームに好都合であるようにインプリメントされる。

10

本発明によるルータ・システム3は、多段プライオリティレベルにてフレーム・トラフィックに適合し、送信又は受信ノードにて無限のバッファ資源無しでプロトコル・デッドロックを防止する戦略を可能にする。本発明によるルータ・システム3は、リンク上の伝送エラー又はソフトRAMエラーのような困難に適合し打破する。各アービタ146はリクエストを格納するためのRAMを使用し、RAMエラーは、欠陥のあるリクエストを発生し、実際のリクエストでの損失を生じ、正しくないリクエストを許可することを引き起こす。従って、自己矛盾のないチェック機構が本発明によるルータ・システム3に含まれる。特に、本発明によれば、ルータ・システム3は、アービタ・ユニット146が誤りの出力ポートにフレームを送信するならばリセットされる。本発明の一つの実施形態によれば、ルータ・システム3は、フレームが過度の時間長で入力バッファに止まるときにはリセットされる。

20

図4Aは、スイッチングバスさらに金属及び差動論理クロスポイント・スイッチを含む本発明によるクロスバー・スイッチ144'の部分回路図である。特に、図4Aは、第1及び第2のバス・サイド11a, 11bを含み、さらに、本発明によれば、10mm金属及び36クロスポイント・スイッチを含むスイッチング・バス11を具備するクロスバー・スイッチ部分を示す。クロスバースイッチ11は、VDDに接続されたプリチャージ・トランジスタ41, 42のゲートに与えられる信号PC(プリチャージ)によりプリチャージされる。スイッチングは、トランジスタ43aで受信されたGRANT信号に 응답してトランジスタ回路43による受信DATAに関して完了される。GRANT信号はトランジスタ43a, 44aのゲートに同時に与えられる。GRANT信号により一度イネイブルされると、DATAはトランジスタ43bを経て与えられ、DATAの補足はトランジスタ44bに付随して与えられる。インバータ45はDATAをその補足に変える。クロスバー・スイッチ11はさらに、信号伝搬をスピードアップする差動及び減少スイング回路を含む。本発明の一実施形態によれば、ビットライン・スイングは500mVのオーダでセットされる。クロスバー・スイッチ11は、本発明の実施形態にて“0.2ns”センスタイムを持つように設計されたセンス増幅器46を含む。

30

40

図4Bは、本発明によるデータ・パスブロック図である。特に、ルータ・システム3は、入力ポート3a、出力ポート3b及びクロスバー・マトリクス144'を含む。入力ポート3aは、デマルチプレクサ400、マルチプレクサ130'、フレーム・ヘッダ・デコーダ128及びバッファ・エレメント5a~5mを含む入力バッファ・エレメント135を含む。出力ポート3bはアービタ146、ヘッダ・モディファイア150及びマルチプレクサ151を含む。

入力ポイント104で受信されたデータは、デマルチプレクサ400及び130'を経て

50

入力バッファ5 a ~ 5 mのうちで選択された1つに送られる。各入力バッファは、クロスバー・マトリクス144'への入力バッファ135のいずれか1つの接続を選択するための如何なるアービトレーションも必要とすることなく、クロスバー・マトリクス144'への直接の専用ラインを持つ。アービタ146は、関連する出力ポートへの接続のために入力バッファ5 a ~ 5 mの1つを選択するために、クロスバー・マトリクス144'の各出力ポートに設けられる。制御信号は、フレーム・ヘッダの存在を示すためにデマルチプレクサ400を経てフレーム・ヘッダ・デコーダ128へ入力ポート103, 103にて提供される。これは、アービタ146のコントロールの下で、クロスバー144'を経て入力バッファアドレスの抽出を可能にする。クロスバー144'からの出力伝送の後に、フレーム・ヘッダは、前方の宛先位置にシフトし、現在のルート・システム3の宛先位置を削除するためにモディファイされる。フレーム・ヘッダ・モディファイア150からの出力の後、フレームはN幅ワードから交換するためにマルチプレクサ151によりマルチプレックスされ、N幅ワードは、立上がり及び立下がりクロックエッジにてトリガされた各半クロックサイクルのN/2幅の半ワードを提供するための各クロックサイクルで送信される。

10

図5 Aは、本発明による、隣接スイッチング・ユニットからのデータ、クロック及び制御信号を受信するための第1フローコントロール・パスシステム490のブロック図である。特に、選択されたスイッチング・ユニットにおけるフロー制御は、所定のビット数(n)に対してn/2幅であるパス上でのフロー制御情報を受けることにより達成される。フロー制御情報は、スイッチング・ユニットにて到達したデータと同じパス上で、タイム・マルチプレックスされる。図5 Aに示すように、フロー制御情報は、nビット幅の内部二重幅パスにデマルチプレクサ501 (Demux)によりデマルチプレックスされる。入力クロック範囲からの入来情報の同期のために、フロー制御情報は、同期中の情報損失を避けるためにフローコントロール(FC)・アキュムレータ502 (FC Accum)にてアキュムレートされる。第1のフローコントロール・パスシステム490は、さらに受信した情報を評価しアービトレーション・ユニット504に制御信号を進めるためのフローコントロール・レシーバ503を含む。加えて、フローコントロール・レシーバ503はフレーム・ヘッダ・モディファイア505へ入手可能なバッファ・エレメント・アドレスを進める。

20

図5 Bは、第1のフローコントロール・パスシステム490の動作フローチャートである。特に、フロー制御情報は受信され(510)、デマルチプレックスされ(511)、アキュムレートされ(512)(それにより同期され)、評価され(513)、そして特定の出力ポートに関係したアービトレーション・ユニット504に提供される(514)。加えて、情報はパケット・ヘッダ・モディファイア505に提供される。

30

図5 Cは、本発明による第2のフローコントロール・パスシステム491のブロック図である。特に、第2のフローコントロール・パスシステム491は、入力バッファ・ステータス決定エレメント506と、フローコントロール・トランスミッタ507と、マルチプレクサ508を含む。

図5 Dは、本発明によるフローコントロール・パスの第2の部分のフローチャートである。特に、図5 Dは第2のフローコントロール・パスシステム491の動作フローチャートである。第2のフローコントロール・パスシステム491は、入力バッファ・ステータスを決定し(520)、フローコントロール・トランスミッタ507からの入力バッファ・ステータス情報を送信し(521)、内部Nビット幅パスから外部N/2ビット幅パスの次ノードへのN-ビット通信のための入力バッファ・ステータス情報とともにフローコントロール・メッセージをマルチプレックスする(522)。

40

図6は本発明によるルータ・チップ・システム3の部分ブロック図である。本発明によるルータ・チップ・システム3は、その1つが図1に示される複数の入力ポート604を含むバーチャル・カッター・パケットスイッチを含む。各入力ポート604 iは、複数のバッファ・エレメント605 iと関係しており、その各々は出力ライン616 a ~ 616 fのなかの選択された1つに接続可能な対応するクロスバー入力ライン615 i (j)

50

に關係している。各入力ポート604は対称的であり、中央クロスバー・スイッチ144'と關係している。各ポート604はK入力ラインを含み、多段エントリ入力バッファ・ユニット605及びアービタ606に接続される。

アービタ606は、クロスバー出力ユニット610a~610fに關係した複数のアービタ・ユニット606a~606fを含む。各クロスバー出力ユニット610iは、複数のヘッダ・モディフィケーション・ユニット607iを含むヘッダ・モディフィケーション回路607に接続された関連出力ライン616iを含む。ヘッダ・モディフィケーション・ユニット607は、複数のドライバー・エレメント608iを含むドライバー回路608に接続される。各ドライバー・エレメント608iは対応するヘッダ・モディフィケーション・ユニット607iからのデータを受ける。出力データは、従ってクロック606に従って出力ライン609上に生じる。

10

図7Aは、本発明による送信されたフレーム670のブロック図である。特に、フレーム670はフレーム・ヘッダ671とパケット672を含む。フレーム・ヘッダ671を出力ポート数のフィールド、プライオリティ・フィールド及び入力バッファ・アドレスを含む。パケット672はさらにパケット・ヘッダ673、エラー検出コード(EDC)674、及びデータ675を含む。

図7Bは、本発明による方法のフローチャートである。図7Bに示す方法によれば、フレーム・ヘッダ671はルータ・システム3により受信される(700)。加えて、ルータ・システム3は、フレーム670の到達とさらに完全な伝送の係属を指示するために、フレーム670と並列にフレーム・エンベローブ信号を受ける。本発明による一つの実施形態によれば、フレーム・エンベローブ信号は、以下の図7Dで詳しく説明するように、ノッチを含む。本発明によれば、フレーム・ヘッダ出力ポート数、プライオリティ・フィールド及び入力バッファアドレスはコピーされる(701)。次に、フレーム670は、入力クロックの立上がりで選択された入力バッファに流れる(702)。出力ポートは、プレジオクロナスに個々にクロックされる。加えて、マルチプレクサ128'は、ライン128"の一つに沿って、フレーム・ヘッダにて選択された出力ポートに従って専用のアービタ146に接続され、接続の選択上でアービトレートする専用のアービタ146の正しいプライオリティのFIFOに入力バッファアドレスのエントリを可能にし、出力ポートの接続に特定の入力バッファを選択し、そして「許可」信号を送る。さらに、クロスバー接続は、選択された出力ポートと選択された入力ポートの間で確立される(704)。さらに、フレーム・ヘッダ671、EDC674、パケット・ヘッダ673、データ675は、出力又はローカルクロックの立上がりエッジにおいて、選択された出力ポートを経て、ルータ・システム3から流れ、本発明によれば、正確な同期入力及び出力クロックを要求しない。

20

30

最後に、フレーム・テール・ノッチが検出され(706)、出力ポートからのデータ伝送を停止し、特定の入力バッファ及び出力ポートの間のクロスバー接続を廃止する。

図7Cは、本発明によるアービタ機能のフローチャート(770)である。特に、図7Cのフローチャートは本発明によるアービタ146の特徴を示す。本発明による方法は、フィリング(filling)771を含み、アービタ146の選択された1つが關係する出力ポートが接続された選択された入力バッファのアドレスをもったアービタ146の特定の先入れ先出し(FIFO)バッファを含む。次に、特定のFIFOは、クロスポイント・スイッチ144'の特定のクロスポイントを選択するために読み出され(772)、関連出力ポートと選択された入力バッファの間の接続を行う。

40

図7Dは、本発明によるフレーム・エンベローブ信号102(FENV)のタイミング図である。フレーム・エンベローブ信号102は、図1に示すようにデータと並列にバッファ105に提供される。図7Dは第1及び第2の論理状態、1及び0、を持つフレーム・エンベローブ信号102を示す。本発明の一つの実施形態によれば、フレーム・エンベローブ信号102は、並列データが並列に提供されないときは0に設定される。フレーム・エンベローブ信号102は、並列データが提供されたときは時刻t0において論理状態1に立ち上がる。並列データは時刻t3にて提供されて終了し、その時間にてフレーム・

50

エンベロープ信号 102 の論理状態は 0 に戻る。本発明によれば、ゼロに戻ることは、所定の時刻 t_1 及び t_2 の間でフレーム・エンベロープ信号 102 の信号レベルにてノッチにより予期され、その期間の間、フレーム・エンベロープ信号 102 の論理レベルはゼロ論理レベルにて持続する。このような予期は、選択された入力バッファを特定の出力ポートに接続する専用状態からクロスバー・スイッチ 144' のタイムリーな開放を容易にする。

図 7 E は、本発明によるフレーム・カット・スルーのタイミング図であり、ヘッダ及びデータ情報の伝送は情報の独立なワードにずらされる。図 7 D に示すように、フレーム・エンベロープ信号 102 の時刻 t_0 及び t_3 の間は、ヘッダ及びデータ情報は選択されたルート・チップを経て伝送される。特に、データのワード及びヘッダ情報は独立のワードで伝送され、伝送は、全てのデータが以前に選択された入力ポート・バッファにより受けられたか否かをチェックすることなく、選択された入力バッファにてヘッダに関して直ちに開始する。従って、特定フレームのヘッダは、特定フレームの全てのワードがいずれか指定された入力バッファにて受けられたか否かをチェックすることを抑止せずに、既に次のルータ・チップに伝送される。

10

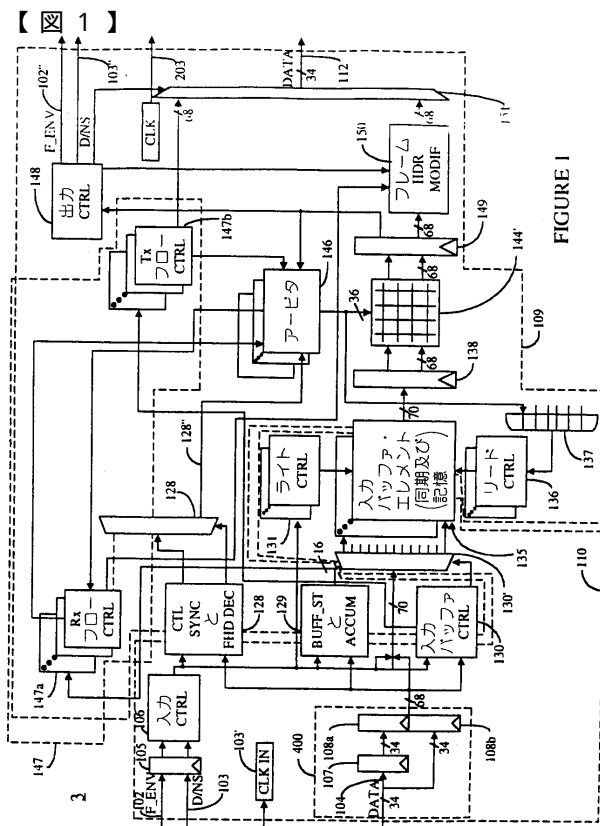


FIGURE 1

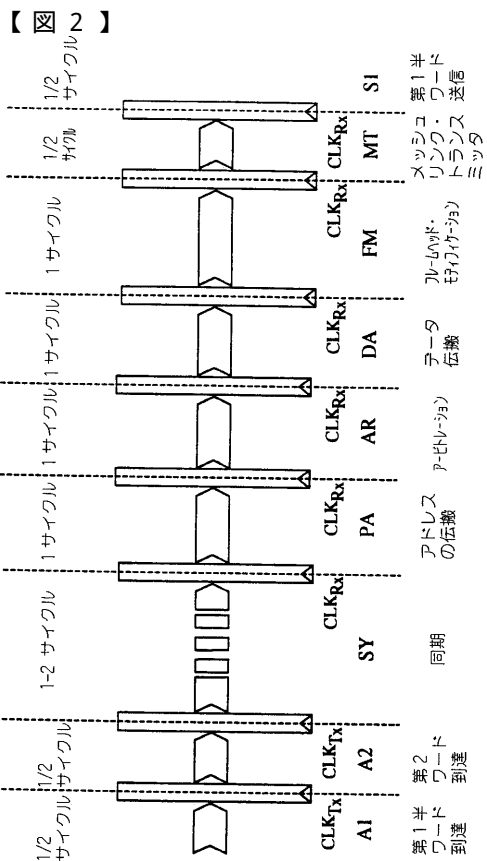


FIGURE 2

【図 3 A】

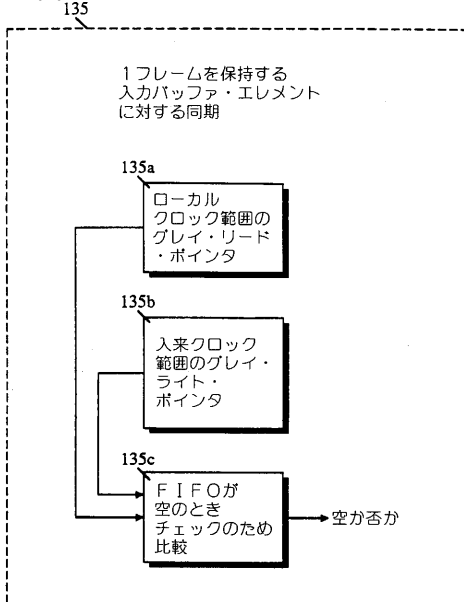


FIGURE 3A

【図 3 B】

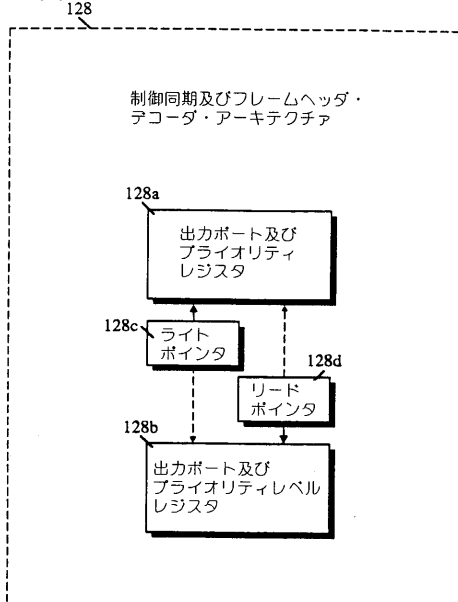


FIGURE 3B

【図 3 C】

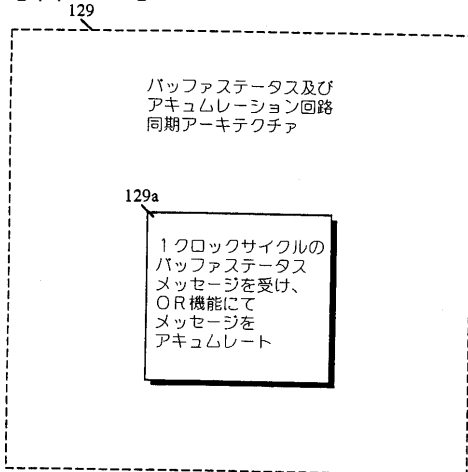


FIGURE 3C

【図 3 D】

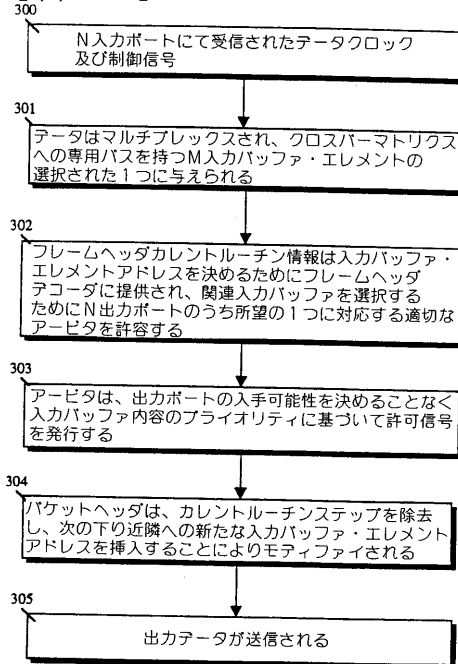


FIGURE 3D

【 図 4 A 】

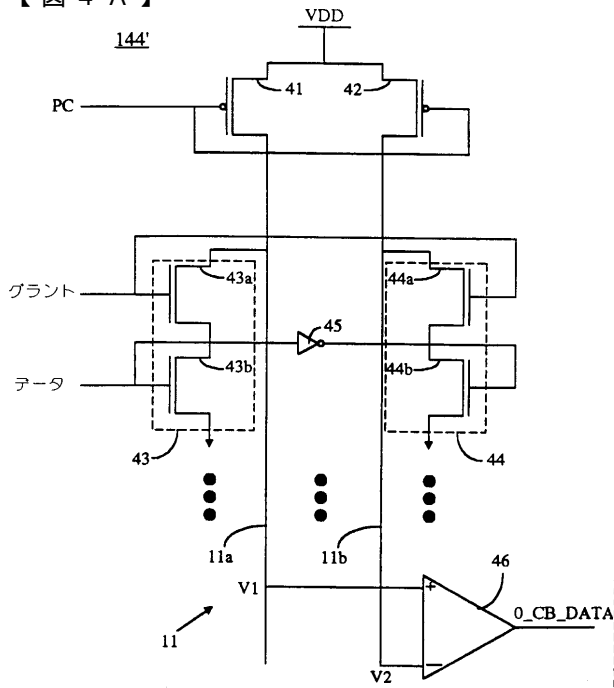


FIGURE 4A

【 図 4 B 】

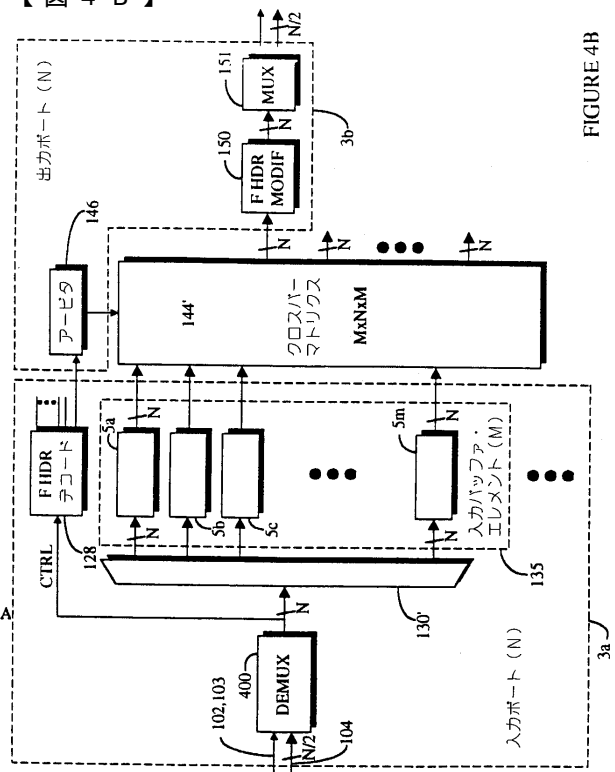


FIGURE 4B

【 図 5 A 】

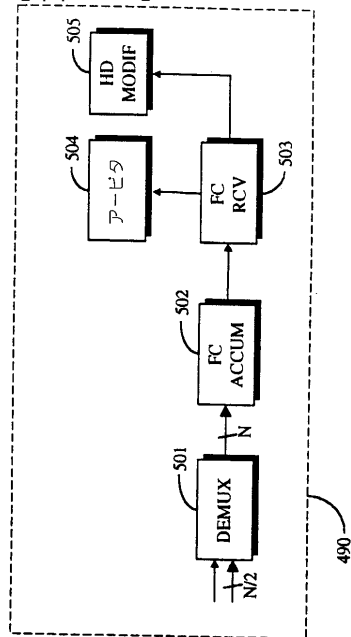


FIGURE 5A

【 図 5 B 】

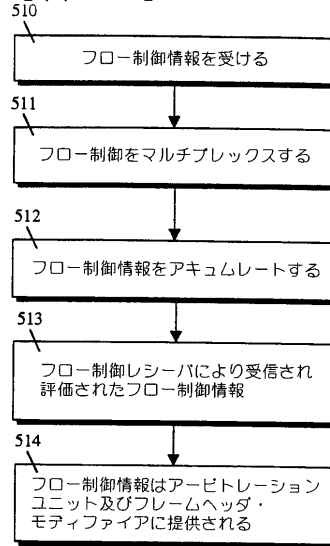


FIGURE 5B

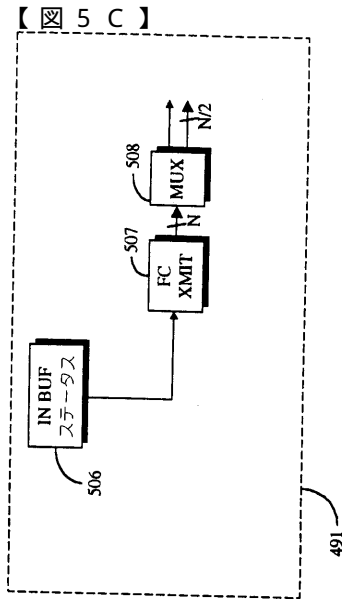


FIGURE 5C

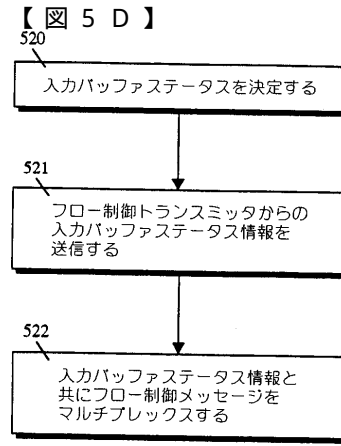


FIGURE 5D

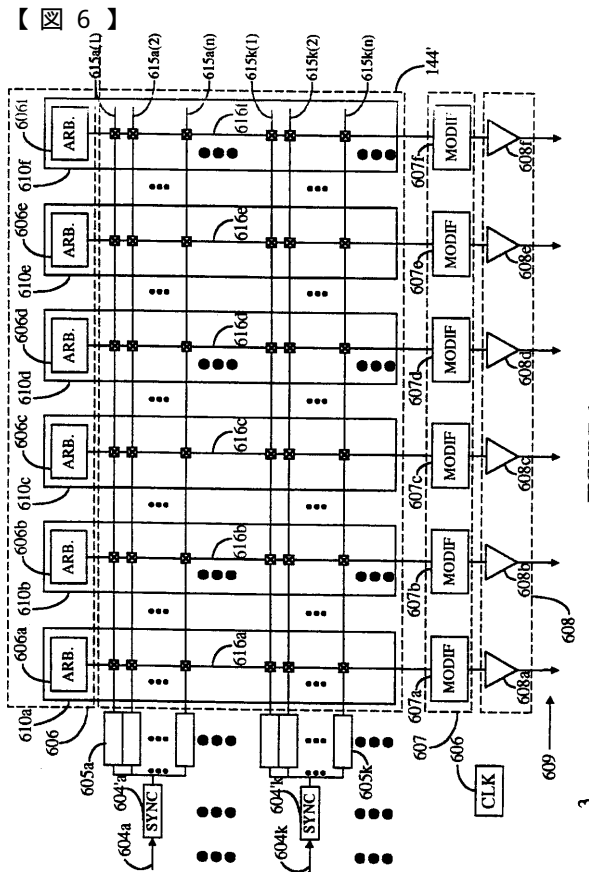


FIGURE 6

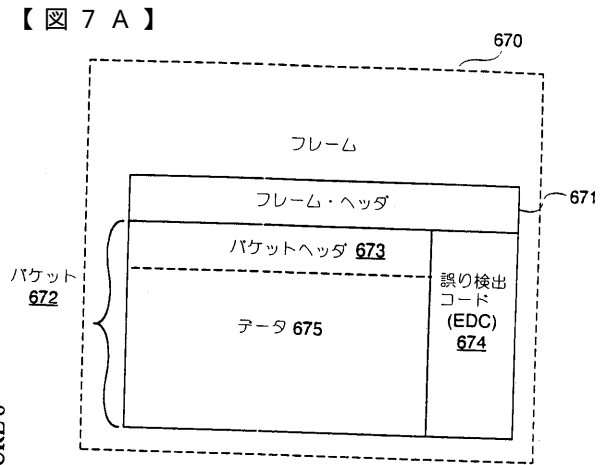


FIGURE 7A

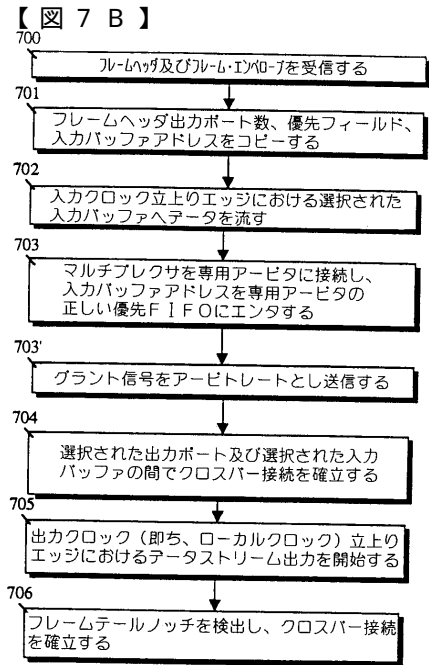


FIGURE 7B

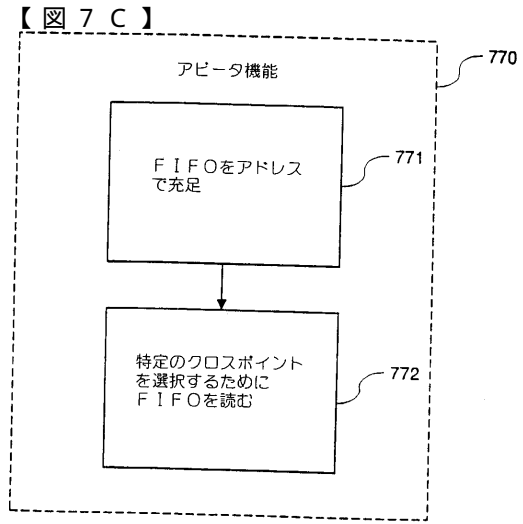


FIGURE 7C

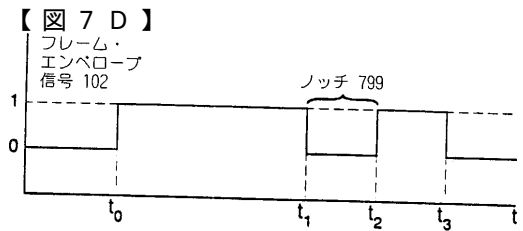


FIGURE 7D

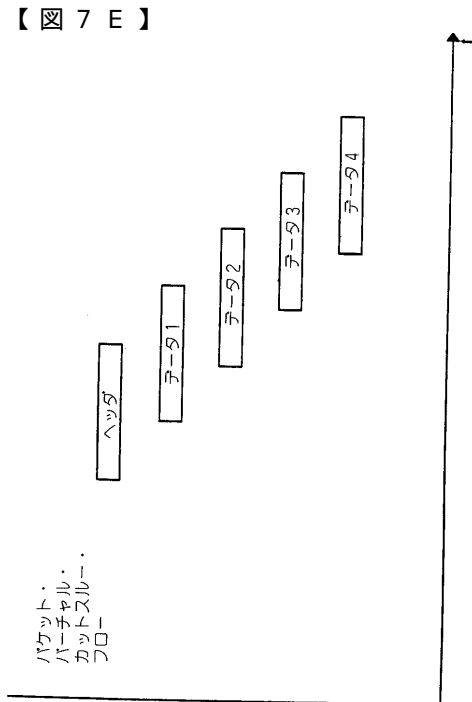


FIGURE 7E

フロントページの続き

- (72)発明者 ウィッキ, トーマス マーティン
アメリカ合衆国, カリフォルニア 94306, パロアルト, ベンチュラ アベニュー 275, ア
パートメント 7
- (72)発明者 ラーソン, ジェフリー ディー.
アメリカ合衆国, カリフォルニア 95125, サンジョゼ, ロブディ ウェイ 1029
- (72)発明者 ミュ, アルバート
アメリカ合衆国, カリフォルニア 95035, ミルピタス, サンドルウッド レーン 1051
- (72)発明者 サストリー, ラフ
アメリカ合衆国, カリフォルニア 95051, サンタクララ, キーリー ブールバード #79
1000

審査官 石井 研一

- (56)参考文献 特開平04-345242(JP, A)
米国特許第05267235(US, A)

(58)調査した分野(Int.Cl., DB名)

H04L 12/56
H04L 12/28
H04L 12/46
H04L 12/66