

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4703189号  
(P4703189)

(45) 発行日 平成23年6月15日(2011.6.15)

(24) 登録日 平成23年3月18日(2011.3.18)

(51) Int.Cl.

F I

G O 6 F 13/16 (2006.01)

G O 6 F 12/00 (2006.01)

G O 6 F 12/06 (2006.01)

G O 6 F 13/16 5 1 0 A

G O 6 F 12/00 5 5 0 C

G O 6 F 12/00 5 6 0 B

G O 6 F 12/00 5 9 7 R

G O 6 F 12/06 5 2 5 A

請求項の数 17 (全 14 頁)

(21) 出願番号 特願2004-566539 (P2004-566539)  
 (86) (22) 出願日 平成15年12月12日(2003.12.12)  
 (65) 公表番号 特表2006-513492 (P2006-513492A)  
 (43) 公表日 平成18年4月20日(2006.4.20)  
 (86) 国際出願番号 PCT/US2003/039609  
 (87) 国際公開番号 W02004/064413  
 (87) 国際公開日 平成16年7月29日(2004.7.29)  
 審査請求日 平成18年12月6日(2006.12.6)  
 (31) 優先権主張番号 10/340,390  
 (32) 優先日 平成15年1月10日(2003.1.10)  
 (33) 優先権主張国 米国 (US)

(73) 特許権者 500267099  
 エス・アール・シィ・コンピューターズ・  
 インコーポレイテッド  
 アメリカ合衆国、80907 コロラド州  
 、コロラド・スプリングス、ノース・ネバ  
 ダ・アベニュー、4240  
 (74) 代理人 100064746  
 弁理士 深見 久郎  
 (74) 代理人 100085132  
 弁理士 森田 俊雄  
 (74) 代理人 100083703  
 弁理士 仲村 義平  
 (74) 代理人 100096781  
 弁理士 堀井 豊

最終頁に続く

(54) 【発明の名称】 インタリーブされたメモリコントローラとともに用いるために再構成可能な処理要素を1つ以上のマイクロプロセッサに結合するスイッチ/ネットワークアダプタポート

(57) 【特許請求の範囲】

【請求項1】

コンピュータシステムであって、  
 少なくとも1つのプロセッサと、  
 前記少なくとも1つのプロセッサを制御ブロックおよびメモリバスに結合するためのコントローラと、  
 前記メモリバスに結合された複数のメモリモジュールスロットと、  
 前記複数のメモリモジュールスロットのサブセットに関連付けられるアダプタポートと、  
 、  
 前記アダプタポートに結合されるクラスタ相互接続ファブリックとを含み、  
 前記複数のメモリモジュールスロットと直接接続することによって制御情報が前記アダプタポートとの間でやりとりされ、  
前記クラスタ相互接続ファブリックに結合された再構成可能なプロセッサ要素と、  
前記メモリバス上において前記コントローラから受信したデータを変更するよう動作する、前記クラスタ相互接続ファブリックに結合されたプロセッサ要素と、  
変更されたデータを前記メモリバス上に配置する前に、外部ソースから受信したデータを変更するよう動作する、前記クラスタ相互接続ファブリックに結合されたプロセッサ要素と、  
前記クラスタ相互接続ファブリックに結合される制御ブロックと、のうちの少なくとも1つをさらに含む、コンピュータシステム。

10

20

**【請求項 2】**

前記コントローラはインタリーブされたメモリコントローラを含む、請求項 1 に記載のコンピュータシステム。

**【請求項 3】**

前記複数のメモリモジュールスロットは D I M M メモリモジュールスロットを含む、請求項 1 に記載のコンピュータシステム。

**【請求項 4】**

前記アダプタポートは、前記 D I M M メモリモジュールスロットのうちの 1 つの内部における保持のための D I M M 物理フォーマットを含む、請求項 3 に記載のコンピュータシステム。

10

**【請求項 5】**

前記複数のメモリモジュールスロットはランバスフォーマットメモリモジュールスロットを含む、請求項 1 に記載のコンピュータシステム。

**【請求項 6】**

前記アダプタポートは、前記 R I M M メモリモジュールスロットのうちの 1 つの内部における保持のためのランバス物理フォーマットを含む、請求項 5 に記載のコンピュータシステム。

**【請求項 7】**

前記制御ブロックは、  
周辺バス制御ブロックと、  
グラフィックス制御ブロックと、  
システムメンテナンス制御ブロックと、のうちの少なくとも 1 つを含む、請求項 1 に記載のコンピュータシステム。

20

**【請求項 8】**

前記アダプタポートは、前記メモリバスからデータ情報を受信するよう適合されたメモリバスデータマルチプレクサを含む、請求項 1 に記載のコンピュータシステム。

**【請求項 9】**

前記アダプタポートはさらに、  
前記メモリバスからアドレス情報を受信するよう適合されたアドレスおよびコマンドデコードと、  
前記クラスタ相互接続ファブリックに結合されたプロセッサ要素に前記メモリバスから受信した前記データ情報を結合するための書込バッファ、および、前記プロセッサ要素によって供給されるデータを前記メモリバスに結合するための読出バッファと、  
のうちの少なくとも 1 つを含む、請求項 8 に記載のコンピュータシステム。

30

**【請求項 10】**

書込バッファを含む場合、前記システムは、  
前記書込バッファを前記プロセッサ要素に結合するためのパケット化装置と、  
前記プロセッサ要素を前記読出バッファに結合するための非パケット化装置とをさらに含む、請求項 9 に記載のコンピュータシステム。

40

**【請求項 11】**

前記メモリバスデータマルチプレクサに結合された複数のレジスタをさらに含む、請求項 8 に記載のコンピュータシステム。

**【請求項 12】**

前記複数のレジスタは、  
構成レジスタと、  
状態レジスタと、のうちの少なくとも 1 つを含む、請求項 11 に記載のコンピュータシステム。

**【請求項 13】**

前記クラスタ相互接続ファブリックに結合されたプロセッサ要素に前記アドレスおよびコマンドデコードを結合するコマンドフォーマットをさらに含む、請求項 9 に記載のコン

50

ピュータシステム。

【請求項 1 4】

前記クラスタ相互接続ファブリックに結合される制御ブロックを含む場合、前記制御ブロックはさらに、

前記アダプタポート上における識別されたアルゴリズムと、前記アダプタポートによって供給されるオペランドとを実行するよう構成可能な少なくとも 1 つのフィールドプログラマブルゲートアレイを含む、請求項 1 に記載のコンピュータシステム。

【請求項 1 5】

前記アダプタポートに結合された制御ブロックを前記少なくとも 1 つのフィールドプログラマブルゲートアレイに結合するデュアルポートメモリブロックをさらに含む、請求項 1 4 に記載のコンピュータシステム。

10

【請求項 1 6】

前記クラスタ相互接続ファブリックに結合されるチェーンポートをさらに含む、請求項 1 に記載のコンピュータシステム。

【請求項 1 7】

前記制御ブロックは、それに構成情報を供給するための読出専用メモリをさらに含む、請求項 1 6 に記載のコンピュータシステム。

【発明の詳細な説明】

【技術分野】

【0001】

20

関連する出願の相互参照

この発明は、2003 年 1 月 10 日に出願され本願と同じタイトルを有する米国特許出願連続番号第 10 / 340 , 390 号の優先権を主張する。これは、「デュアル・インライン・メモリモジュールフォーマットの多重適応プロセッサのチェーンを用いるクラスタ化されたコンピュータのためのスイッチ/ネットワークアダプタポート ( “ Switch/Network Adapter Port for Clustered Computers Employing a Chain of Multi-Adaptive Processors in a Dual In-Line Memory Module Format ” ) 」と題され 2001 年 8 月 17 日に出願された米国特許出願連続番号第 09 / 932 , 330 号と、「汎用の再構成可能な演算ハードウェアおよびソフトウェア ( “ General Purpose Reconfigurable Computing Hardware and Software ” ) 」と題され 2002 年 10 月 31 日に出願された米国仮特許出願連続番号第 60 / 422 , 722 号とに関し、そのすべてがエス・アール・シー・コンピュータズ・インコーポレイテッド ( SRC Computers, Inc. ) ( コロラド州 ( Colorado ) 、コロラドスプリングス ( Colorado Springs ) ) に譲渡され、それらの開示の全体が引用によりこの明細書中に具体的に援用されている。

30

【0002】

発明の背景

発明の分野

この発明は、一般に、1 つ以上の再構成可能な ( reconfigurable ) 処理要素を組み込んだコンピュータシステムおよび方法の分野に関する。より特定的には、この発明は、データ転送レートを上げるためにインタリーブされたメモリコントローラとともに用いるための多重適応処理要素 ( 「 MAP 」 ( 登録商標 ) ) を採用する計算システムのためのデュアル・インライン・メモリモジュール ( 「 DIMM 」 ) またはランバス ( 「 RIMM 」 ) フォーマットのスイッチ/ネットワークアダプタポート ( 「 SNAP<sup>TM</sup> 」 ( 登録商標 ) 、共に SRC Computers, Inc. の登録商標 ) に関する。

40

【背景技術】

【0003】

関連する背景技術

プロセッサカウントが大きく経済的なコンピュータをもたらす現在最も有望な方法の中には、パーソナルコンピュータ ( PC ) で一般的に見出されるような比較的低価格のマイクロプロセッサベースのボードをいくつかまとめてクラスタリングすることに関するもの

50

がある。次に、利用可能なクラスタリングソフトウェアを用いてこれらさまざまなボードを動作させることにより、1つ以上の大きな問題を解くためにこれらを一斉に実行させることが可能となる。このような問題を解くプロセスの最中には、プロセッサボード間で中間演算結果が共有されることが多い。

#### 【0004】

現在利用可能な技術を利用した場合、この共有は周辺構成要素相互接続（P C I）バスを介してなされなければならないが、これは現在のP Cで一般的に見られる最も高性能の外部インターフェイスバスである。このバスにはさまざまな利用可能なバージョンがあるが、これらはすべて1 G B / s e c未満の帯域幅に制限され、また、これらはプロセッサバスよりも数レベル下のチップに位置するため、いずれも極めて長い待ち時間を呈する。低価格のP Cにおいては、このバスは典型的には2 5 6 M B / s e cのオーダの帯域幅を提供するに過ぎない。

#### 【0005】

これらの要因は、個々としても全体としても、クラスタの全体的な有効性を大幅に制限するおそれがあり、より高速のインターフェイスを見出すことができればクラスタが大きな問題を解く能力は大幅に向上するだろう。残念ながら、このようなポートを与え得る新たな専用のチップセットを設計することは非常に費用がかかるだけでなく、遭遇するクラスタリング相互接続の種類ごとに特化する必要もあるだろう。当然、これによってチップセットのいずれか1つのバージョンについても潜在的な売上量が比較的少なくなり、非経済的になるだろう。

#### 【0006】

プロセッサの速度がますます速くなりつつあるために、高性能のメモリサブシステムの必要性が引き続き高まってきている。上述の米国特許出願連続番号第09/932,330号に開示されるスイッチ/ネットワークアダプタポートシステムの開発以来、パーソナルコンピュータ（「P C」）の市場に対する高性能のメモリサブシステムのための技術には、インタリーブされたメモリが用いられるようになってきた。

#### 【0007】

インタリーブされたメモリシステムにおいては、2つ以上のデュアル・インライン・メモリモジュール（「D I M M」）スロットがメモリコントローラによって同時にアクセスされる。二重インタリーブ方式が用いられる場合、データバスの幅が事実上2倍にされるので、メモリに対して得られる帯域幅も2倍になる。4つのD I M Mスロットを用いる四重または4方向にインタリーブされたシステムを形成するために類似の構成が確立可能である。このメモリコントローラの形は、現在、よりハイエンドのサーバシステムに見出されるより一般的な高性能メモリのうちの1つであり、より主流の製品において急速に利用可能になりつつある。

#### 【発明の開示】

#### 【課題を解決するための手段】

#### 【0008】

#### 発明の概要

この明細書中の代表的な実施例に開示されるように、SRC Computers, Inc.の独自仕様のS N A P<sup>TM</sup>技術は、2つ以上のD I M M（またはR I M M）スロットからの信号が共通の制御チップに経路指定されるように向上されてきた。物理的には、二重構成においては、2つのD I M Mフォームファクタ・スイッチ/ネットワークアダプタボードは、リッジフレックス（ridged flex）回路構成を用いて共に結合されて単一のアセンブリを形成し得る。D I M Mボードのうちの1つはまた、制御フィールドプログラマブルゲートアレイ（「F P G A」）に装着されてもよく、これには両方のD I M Mスロットからの信号が経路指定されてもよい。次いで、制御チップは、スロットの独立したクロックを用いて両方のスロットからデータを抽出する。両方のスロットからのデータを用いてデータパケットを形成し、これがシステムの他の部分に送信される。この明細書中に開示される技術が、同様の態様で、3つ以上のD I M Mスロットと共に用いられてもよく、たとえば4つの

D I M Mスロットは4方向にインタリーブされたシステムとなる。

【0009】

この発明に従うと、インタリーブされたメモリシステムは、ロックステップ(lock-step)において実行する2つ以上のメモリチャネルを用いる。D I M Mスロットのうちの1つへの接続がなされ、得られた信号がもとの組のS N A P<sup>TM</sup>ボード信号と共に用いられる。これにより、動作の際に、メモリに対するデータバスの幅が有効に2倍(以上)にされる。この技術は、インタリーブされたメモリをサポートするメモリおよび入出力(「I/O」)コントローラ(「ノースブリッジ(North Bridge)」)チップを適切に選択することと共に実現可能である。

【0010】

この明細書中に特に開示されるコンピュータシステムは、少なくとも1つのプロセッサと、プロセッサを制御ブロックおよびメモリバスに結合するためのコントローラと、メモリバスに結合された複数のメモリモジュールスロットと、複数のメモリモジュールスロットのサブセットに関連付けられるアダプタポートと、アダプタポートに結合されるプロセッサ要素とを含む。好ましい実施例においては、当該アダプタポートはD I M MまたはR I M Mフォームファクタに便宜的に設けられてもよい。

【0011】

この明細書中に開示されるコンピュータシステムは、また、少なくとも1つのプロセッサと、プロセッサを制御ブロックおよびメモリバスに結合するためのインタリーブされたコントローラと、メモリバスに結合される複数のメモリスロットと、メモリスロットのうちの少なくとも2つに関連付けられるアダプタポートと、アダプタポートの各々に結合されるプロセッサ要素とを含む。

【0012】

この発明の上述および他の特徴および目的、ならびにこれらを達成する態様がより明らかとなり、この発明自体は、添付の図面と関連して好ましい実施例の以下の記載を参照することにより最適に理解されるだろう。

【発明を実施するための最良の形態】

【0013】

代表的な実施例の説明

図1を参照すると、コンピュータシステム100の具体的な実施例の機能ブロック図が示され、当該コンピュータシステム100は、周辺構成要素相互接続(「P C I」)バスから得られる以上のデータ転送レートを大幅に向上させるためにD I M Mフォーマットの多重適応プロセッサのチェーンを用いるクラスタ化されたコンピュータのためのスイッチ/ネットワークアダプタポートを含む。

【0014】

図示される特定の実施例においては、コンピュータシステム100は1つ以上のプロセッサ102<sub>0</sub>および102<sub>1</sub>を含み、当該1つ以上のプロセッサ102<sub>0</sub>および102<sub>1</sub>は関連するP CメモリおよびI/Oコントローラ104に結合される。動作の際に、コントローラ104は、P C I制御ブロック106からの制御情報を送受信する。なお、この発明の代替的な実現例においては、制御ブロック106はまたアクセラレーテッド・グラフィックス・ポート(「A G P」)またはシステムメンテナンス(「S M」)制御ブロックであってもよい。P C I制御ブロック106は、帯域幅が比較的低いP C Iバス110によって1つ以上のP C Iカードスロット108に結合されることにより、実質的に256 M B / s e cの速度でデータを転送することが可能となる。この発明の上述の代替的な実施例においては、カードスロット108は代替的にはアクセラレーテッド・グラフィックス・ポート(「A G P」)またはシステムメンテナンス(「S M」)バス接続を含み得る。

【0015】

コントローラ104はまた、実質的に2.1 G B / s e c以上のデータ転送レートを可能にする帯域幅のはるかに高いD I M Mバス116によっていくつかのD I M Mスロット114に便宜的に結合される。この発明の特定の實現例に従うと、D I M M M A P要素

10

20

30

40

50

112は、DIMMスロット114のうちの1つに関連付けられるかまたは物理的にその内部に位置する。PCIバス110とDIMM MAP要素112とを相互接続する接続118によってDIMM MAP要素112との間で制御情報がやりとりされる。次いで、DIMM MAP要素112が、MAPチェーンポートに接続されるクラスタ相互接続ファブリック接続120によって別のクラスタ化されたコンピュータMAP要素に結合され得る。これは後により十分に記載されるとおりである。DIMM MAP要素112がランバスDIMM(「RIMM」)MAP要素をも含み得ることに留意されたい。

【0016】

DIMMスロット114内に位置するDIMMメモリは、PCマイクロプロセッサ102<sub>0</sub>、102<sub>1</sub>に対する主要な記憶位置を含むので、プロセッサバスに電氣的に非常に「近く」なるよう設計され、こうして待ち時間が非常に短くなる。先述のとおり、DIMMに関連付けられる待ち時間がPCIバス110の待ち時間の僅か25%のオーダーとなることは珍しくない。本質的には、この帯域幅をコンピュータシステム100間の相互接続として利用することにより、クラスタ性能の大幅な向上が実現され得る。

【0017】

この目的のために、DIMM MAP要素112をPCのDIMMスロット114のうちの1つに配置することにより、(以下により十分に記載される)その制御チップが通常のメモリ「読出」および「書込」トランザクションを受取り、これらを、相互接続スイッチまたはネットワークによって用いられるフォーマットに変換し得る。後により十分に記載されるように、各MAP要素112は、当該MAP要素112を他のMAP要素112に結合することを可能にするチェーンポートを含む。チェーンポートを用いて接続120を介して外部のクラスタリングファブリックに接続することにより、データパケットがリモートノードに送信され、ここで、同一のボードによって受信され得る。この特定の応用例においては、DIMM MAP要素112はパケットからデータを抽出し、受信プロセッサ102が必要とするまでこれを記憶するだろう。

【0018】

この技術により、PCIバス110などの現在利用可能ないかなるPCインターフェイスよりも数倍高いデータ転送レートがもたらされる。しかしながら、DIMMの電氣的プロトコルは、データが受信機に到着しても、DIMMスロット114内のDIMMモジュールが当該データの到着をマイクロプロセッサ102に信号で伝える可能性がないようなものであり、この能力がないので、データが到着したかどうかを判断するためにDIMM MAP要素112の連続したポーリングを用いることによってプロセッサ102の活動を同期させなければならない。このような技術はマイクロプロセッサ102とそのバス帯域幅の大部分とを完全に消費するため、他のすべてのバスエージェントを機能停止させることとなる。

【0019】

この状況を避けるために、DIMM MAP要素112はさらに接続118を備えてもよく、これにより既存のPCIバス110と通信することが可能となり、通信パケットが生成され、これらがPCIバス110を介してプロセッサ102に送信され得る。移動させる全データのうちこれらのパケットが占める割合はほんの僅かにすぎないので、PCIバス110の低帯域幅の影響が最小限にされ、従来のPCI割込信号を用いて、データが到着したことをプロセッサ102に通知することもできる。この発明の別の実現例に従うと、プロセッサ102に信号を送るのにシステムメンテナンス(「SM」)バス(図示せず)がまた用いられてもよい。SMバスはシリアル電流モードバスであり、従来、プロセッサボード上のさまざまな装置がプロセッサ102に割込むことを可能にする。代替的な実施例においては、プロセッサ102に信号を送るのにアクセラレーテッド・グラフィックス・ポート(「AGP」)が用いられてもよい。

【0020】

DIMMスロット114全体たり得るものに関連付けられるDIMM MAP要素112を用いる場合、PCは、DIMM MAP要素112が用いるための、典型的には1 G

10

20

30

40

50

Bのオーダのアドレスの大きなブロックを割当ててもらう。これらのうちいくつかはコマンドとしてデコードされ得るが、ストレージとして用いられ得るものが依然として多い。周辺装置からデータを転送するのに用いられる通常の入出力(「I/O」)ブロックサイズと少なくとも同じ数のアドレス位置を有することにより、(コントローラ104を含む)大抵のPCで用いられる従来のインテル(Intel)(登録商標)チップセットは、DIMM MAP要素112への直接的なI/O転送を可能にするだろう。これにより、データがたとえばディスクから到着し、直接DIMM MAP要素112に渡ることが可能となる。これは、所望される如何なる態様にも変更されてもよく、パケット化され、接続120を介してリモートノードに伝送されてもよい。ディスクのPCIバス110ならびにDIMM MAP要素112およびDIMMスロット114が共にPCメモリコントローラ104によって制御されるので、プロセッサバス帯域幅はこの転送によっては消費されない。

10

#### 【0021】

或るPCにおいては、DIMMスロット114内におけるいくつかのDIMMがインタリーブされて、より幅広いメモリアクセス能力がもたされることによりメモリ帯域幅が増大し得ることに留意されたい。これらのシステムにおいては、前述の技術はまた、いくつかのDIMMスロット114において同時に利用され得る。しかしながら、選択された特定の実現例にも拘らず、最終結果として、PCIバス110への1つ以上の接続と接続120を介する外部のスイッチまたはネットワークとを有するDIMMベースのMAP要素112ということになり、これにより、PCIベースの接続の単独での性能や、データ

20

#### 【0022】

加えて、図2Aを参照すると、この発明に従ったスイッチ/ネットワークアダプタポート200Aの具体的な実施例の機能ブロック図が示される。上述の図に関連して開示されるものと同じ構造および機能には同じ参照符号が付けられ、これに対しては上述の説明で十分であるだろう。スイッチ/ネットワークアダプタポート200Aは、共通の制御要素202に結合されるインタリーブされたDIMMスロットフォームファクタのSNAP要素204Aおよび204Bの二重構成で示される。この実施例においては、コントローラ104は、チャンネルA 216AおよびチャンネルB 216BによってDIMMスロット114およびSNAP要素204A、204Bに双方向に結合されるインタリーブされたメモリコントローラである。

30

#### 【0023】

加えて図2Bを参照すると、この発明に従ったスイッチ/ネットワークアダプタポート200Bの別の具体的な実施例の機能ブロック図が示される。また、上述の図に関連して開示されるものと同じ構造および機能には同じ番号が付けられ、これに対しては上述の説明で十分であるだろう。スイッチ/ネットワークアダプタポート200Bは、共通の制御要素202に結合されるインタリーブされたDIMMスロットフォームファクタのSNAP要素204A~204Dの四重構成で示される。この実施例においては、コントローラ104はまた、それぞれのチャンネルA 216A、チャンネルB 216B、チャンネルC 216CおよびチャンネルD 216DによってDIMMスロット114ならびにSNAP要素204A、204B、204Cおよび204Dに双方向に結合されるインタリーブされたメモリコントローラである。

40

#### 【0024】

さらに図3を参照すると、SRC Computers, Inc.から入手可能なSRC-6Eコンピュータシステム300の高レベルの機能ブロック図が示される。コンピュータシステム300は、図2Aおよび図2Bのスイッチ/ネットワークアダプタポート要素204を組込んでもよく、いくつかのMAP要素302<sub>0</sub>~302<sub>N</sub>を含んでおり、各々は、それぞれのスイッチ/ネットワークアダプタポート304<sub>0</sub>~304<sub>N</sub>によって対応するマイクロプロセッサボード306<sub>0</sub>~306<sub>N</sub>に結合されている。帯域幅が1200MB/sec以上であるいくつかのチェーンポート308は、さまざまなMAP要素302<sub>0</sub>~302<sub>N</sub>をMAP

50

要素 302<sub>0</sub> ~ 302<sub>N</sub>のうちの別のものに結合する。動作の際に、コンピュータシステム 300は、データ共有が少なく、MAP要素 302の入出力(「I/O」)の利用が多く、プロセッサ間通信が非常に少ないプロセッサ利用に等しいMAP要素 302を提供する。

#### 【0025】

加えて図4を参照すると、上述の図のSRC-6Eコンピュータシステム300の一部のより詳細な機能ブロック図が計算システム400として示され、さらに、MAPボード302の半分とともにプロセッサボード306の半分が示される。図示のとおり、プロセッサボード306は1対のマイクロプロセッサ402<sub>0</sub>および402<sub>1</sub>を含み、各々は関連するレベル2(「L2」)のキャッシュ404とメモリおよびI/Oコントローラ(たとえば「ノースブリッジ」)406とに結合されている。コントローラ406は、たとえば1つ以上の周辺構成要素相互接続(「PCI」)スロット408および専用のメモリ410へのアクセスを提供する。専用のメモリ410は、たとえば、関連するスイッチ/ネットワークアダプタポート304を有するインタリーブされたDIMMまたはRIMMスロットを含み得る。これは以下により十分に記載されるとおりである。

#### 【0026】

MAP要素302は制御ブロックまたは制御チップ412を含み、これが、プロセッサボード306におけるスイッチ/ネットワークアダプタポート304に結合されて、これを制御する。制御チップ412はフィールドプログラマブルゲートアレイ(「FPGA」)を含み得るが、オンボードメモリ414に結合されており、当該オンボードメモリ414は、たとえばデュアルポートのスタティックランダムアクセスメモリ(「SRAM」)または他の種類のメモリ素子を含み得る。オンボードメモリ414はまた、(FPGAとしても設けられ得る)1つ以上のユーザチップ416<sub>0</sub>および416<sub>1</sub>に結合される。

#### 【0027】

加えて図5を参照すると、上述の図のMAP要素302のより詳細なさらなる機能ブロック図が示される。この図においては、制御FPGA412、オンボードメモリ414、ユーザFPGA416<sub>0</sub>および416<sub>1</sub>ならびにそれらの間のさまざまな接続の帯域幅をより詳細に示す機能ブロック500が示される。また、制御FPGA412の動作を特定するためのマイクロコード読出専用メモリ(「ROM」)502および構成ROM504が示され、さらに、ジョイント・テスト・アクション・グループ(「JTAG」)の米国電気電子学会(IEEE 1149.1)規格の境界スキャン接続がいくつか示される。計算システム全体においてさまざまなMAP要素302を直接結合するためのいくつかのチェーンポート308が示される。

#### 【0028】

好ましい実施例においては、制御FPGA412は、ザイリンクス(Xilinx, Inc.)から入手可能なVirtex II XC2V6000デバイスを含み得る。これは、ユーザFPGA416を制御するためのダイレクトメモリアクセス(「DMA」)機能で実現される600万個のゲートを含み得る。同様に、ユーザFPGA416はまた、多重周波数に対し100MHzクロックの位相ロックループ(「PLL」)で実現されるXC2V6000デバイスを含み得る。オンボードメモリ414は、64ビットのデータ経路を備え100MHzで動作可能な24Mバイトの6個のデュアルポートメモリバンクを含み得る。

#### 【0029】

加えて図6を参照すると、図4および図5に示されたMAP要素302の制御ブロック412のより詳細なさらなる機能ブロック図が機能ブロック600として示される。上述のとおり、制御ブロック412は、プロセッサボード306のスイッチ/ネットワークアダプタポート304と、オンボードメモリ414およびユーザチップ416とに結合される。

#### 【0030】

制御ブロック412は、コマンドリスト(「ComList」)を処理し、かつDMA

10

20

30

40

50



要求を発行するためのコマンドプロセッサ602を含む。オンボードメモリ604に対するDMAエンジンはコマンドプロセッサ602に結合され、当該DMAエンジンを用いてオンボードメモリ414の6個のメモリバンクを線形ストライドでアドレス指定し、16エントリの不規則なアクセスをもたらす。外部アクセス608に対する別のDMAエンジンがまたコマンドプロセッサ602に結合され、MAP要素302の外部のメモリに対するDMA動作を実行する。これもストライドの能力を有し、順序がばらばらな応答のトラックを維持する。

#### 【0031】

DMAエンジン608はパケット化装置606に結合され、当該パケット化装置606は伝送すべき32バイトのパケットを形成し、誤り訂正符号(「ECC」)を生成して、送受信されるデータの妥当性を確実にする。非パケット化装置610はまたDMAエンジン608に関連付けられ、データパケットの妥当性をチェックするよう機能し、かつECCデータをチェックして訂正するよう機能する。データレジスタ612およびフラグレジスタ614はコマンドプロセッサ602とユーザチップ416との間に結合される。前者は16組のデータを保持し、各組は32個の64ビット値を含み、後者は32個の1ビット値を保持する。

#### 【0032】

制御ブロック412はまた、いくつかの記憶位置を含み、図示される具体的な実施例においては、コマンドプロセッサ602とDMAエンジン608との間に結合される記憶0ブロック616および記憶1ブロック618を含む。記憶0ブロック616は、第1のComListと制御ブロック412に対するユーザ構成データの第1の位置とを保持するのに用いられる。記憶1ブロック618は、第2のComListとユーザ構成データの第2の位置とを保持する。ユーザチップコンフィギュレータ620はコマンドプロセッサ602およびユーザチップ416に結合されてユーザ論理ビットストリームを制御し、ユーザチップ416<sub>0</sub>および416<sub>1</sub>の各々を構成する。

#### 【0033】

さらに図7を参照すると、代表的なスイッチ/ネットワークアダプタポート304のより詳細な追加の機能ブロック図が、図3および図4に示されたコンピュータシステムと共に用いることのできる機能ブロック700として示される。スイッチ/ネットワークアダプタポート304は便宜的にフォームファクタに設けられてマイクロプロセッサボード306(図3)のDIMMまたはRIMMスロットに差込まれてもよく、MAP要素302を1つ以上のマイクロプロセッサ402、たとえば図1、図2Aおよび図2Bのマイクロプロセッサ102に相互接続させる。この点に関して、スイッチ/ネットワークアダプタポート304は、データおよびアドレス情報を、たとえばダブルデータレート(「DDR」)メモリアンターフェイス702と1つ以上のマイクロプロセッサ402との間で交換するよう結合されてもよい。

#### 【0034】

スイッチ/ネットワークアダプタポート304はパケット化装置704を含み、当該パケット化装置704はMAP要素302に結合されて、DMAヘッダパケット応答を与え、機能性ならびにECCチェックバイト生成を再試行する。非パケット化装置706がまたMAP要素302に結合され、当該非パケット化装置706は、スイッチ/ネットワークアダプタポート304が受信したデータ上でDMAヘッダパケットソート、データ要求パケットソートならびにECCチェックおよび訂正機能を実行する。読出バッファ708は非パケット化装置706に結合され、図示される具体的な実施例においては、「読出」データの1024本のキャッシュラインを保持するための32Kバイト以上の能力を備える。対応する書込バッファ710はパケット化装置704に結合される。当該対応する書込バッファ710はまた、「書込」データの1024本のキャッシュラインを保持するための32Kバイト以上の能力を有し得る。

#### 【0035】

メモリバスデータマルチプレクサ712は、DDRメモリアンターフェイス702のデ

10

20

30

40

50

ータバスと読出バッファ708および書込バッファ710との間に結合される。メモリバスデータマルチプレクサ712はこれらの要素間でデータマルチプレクサとして機能するが、アドレスおよびコマンドデコードブロック714はDDRメモリインターフェイス702のアドレスバスとマルチプレクサ712との間に結合されて、メモリバスデータマルチプレクサ712のデータ読出/書込機能を制御して、情報を状態および構成レジスタ718に維持し、ダイレクトコマンドをMAP要素302へと送信する。この点に関して、シリアルダイレクトコマンドフォーマッタ716がアドレスおよびコマンドデコードブロック714に結合されて、MAP要素302にダイレクトコマンドサポートを与える。

【0036】

加えて図8を参照すると、MAP制御ブロック412とスイッチ/ネットワークアダプタポート304と(たとえば図4の)1つ以上のマイクロプロセッサ402との間の相互作用を示す図が、この発明に従ったスイッチ/ネットワークアダプタポートを利用するコンピュータシステムの動作の一例として提供される。

【0037】

この代表的な「読出」動作の際に、MAP制御ブロック412はまずComListを処理し、要求されるデータのアドレスおよび長さを含む「読出」要求を発行する。この要求は、関連するスイッチ/ネットワークアダプタポート304に転送され、そこで処理される。スイッチ/ネットワークアダプタポート304はその制御レジスタを更新し、要求されるデータのアドレスおよび長さをマイクロプロセッサ402に送信する。マイクロプロセッサ402は、「読出」または「書込」要求を待つDMAスレッドを有する。次いで、これがデータ要求を処理し、要求しているスイッチ/ネットワークアダプタポート304に対し要求されたデータを転送する。スイッチ/ネットワークアダプタポート304は受信したデータのデータ書込を処理し、MAP制御ブロック412にデータを転送するために先入れ先出しレジスタを充填し始める。MAP制御ブロック412のDMAエンジン608(図6)が当該データを受信し、当該データはDMAエンジン604によってオンボードメモリ414に書込まれる。図示される具体的な実施例においては、この動作全体に対する総待ち時間は、MAP制御ブロック412が要求を発行した時間からデータの第1のストリームが受信されるまで1.0ナノ秒のオーダーである。

【0038】

特定のシステム実現例および技術との関連でこの発明の原理を説明してきたが、上述の記載が例としてのみなされ、この発明の範囲を限定するものではないことが明確に理解されるだろう。特に、上述の開示の教示は当業者に他の変形例を示唆するであろうことが認められる。このような変形例は、それ自体が既に公知であって、ここに既に記載された特徴の代わりにまたはこれに加えて使用され得る他の特徴をも含み得る。この出願では特許請求の範囲は特徴の特定の組合せについて作成されているが、ここにおける開示の範囲は、明示的または暗示的に開示されたあらゆる新規の特徴もしくはあらゆる新規の特徴の組合せ、または当業者に明らかであろうその一般例もしくは変形例をも含み、これは、そのようなものがいずれかの請求項で現在請求されているものと同じ発明に関するかどうかにかかわらず、かつこれがこの発明が直面するものと同じ技術的問題のいずれかまたはすべてを軽減するか否かにかかわらないことを理解されたい。出願人は、この出願またはこれより発生する出願すべての審査手続期間において、そのような特徴および/またはそのような特徴の組合せに対して新たな請求項を作成する権利をこれにより留保する。

【図面の簡単な説明】

【0039】

【図1】周辺構成要素相互接続(「PCI」)バスから得られる以上のデータ転送レートを大幅に向上させるために、DIMMフォーマットの多重適応プロセッサのチェーンを用いるクラスタ化されたコンピュータのためのスイッチ/ネットワークアダプタポートの具体的な実施例を示す機能ブロック図である。

【図2A】共通の制御要素に結合されるインターリーブされたDIMMスロットフォームファクタのSNAP要素の二重構成を示す、この発明に従ったスイッチ/ネットワークアダ

10

20

30

40

50

プタポートの具体的な実施例を示す機能ブロック図である。

【図2B】共通の制御要素に結合されるインタリーブされたDIMMスロットフォームファクタのSNAP要素の四重構成を示す、この発明に従ったスイッチ/ネットワークアダプタポートの別の具体的な実施例を示す更なる機能ブロック図である。

【図3】前述の図のスイッチ/ネットワークアダプタポート要素を組み込み得、各プロセッサボードがそれ自体のMAP要素に結合されており、SRC Computers, Inc.から入手可能なSRC-6Eコンピュータシステムを示す高レベルの機能ブロック図である。

【図4】MAPボードの半分とともにプロセッサボードの半分を示す、前述の図のSRC-6Eコンピュータシステムの一部を示すより詳細な機能ブロック図である。

【図5】制御FPGA、オンボードメモリおよびユーザFPGAを特に示し、前述の図のMAP要素を示すより詳細なさらなる機能ブロック図である。

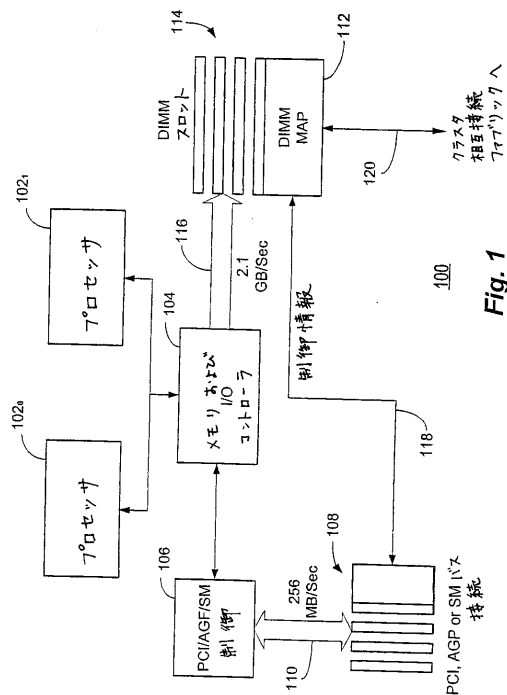
【図6】図4および図5に示されたMAPボードの制御ブロックまたはチップを示すより詳細なさらなる機能ブロック図である。

【図7】図3および図4に示されるコンピュータシステムと共に用いるための代表的なスイッチ/ネットワークアダプタポートを示すより詳細な追加の機能ブロック図である。

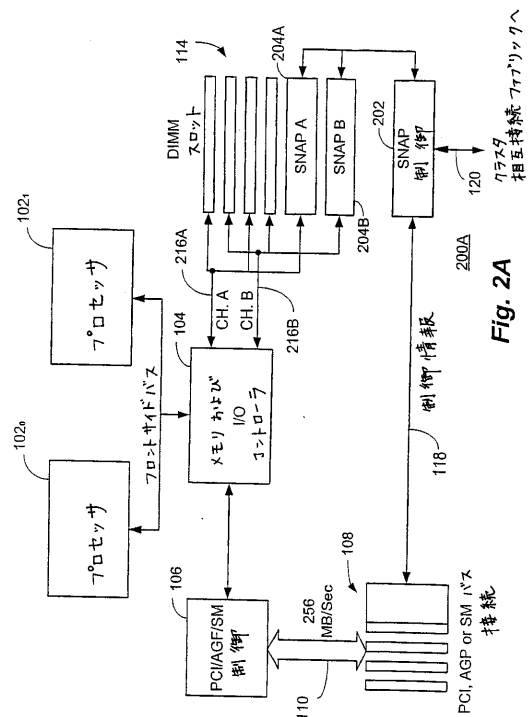
【図8】この発明に従ったスイッチ/ネットワークアダプタポートを利用するコンピュータシステムの動作におけるMAP制御ブロックとスイッチ/ネットワークアダプタポートとマイクロプロセッサとの間の相互作用を示す図である。

10

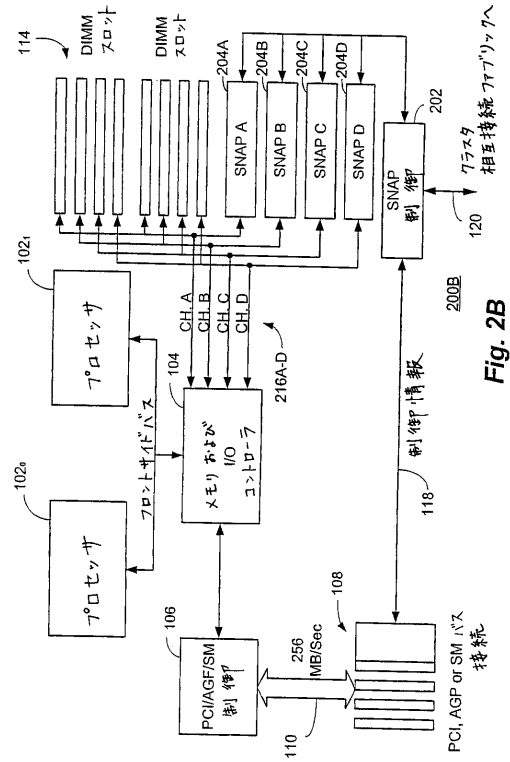
【図1】



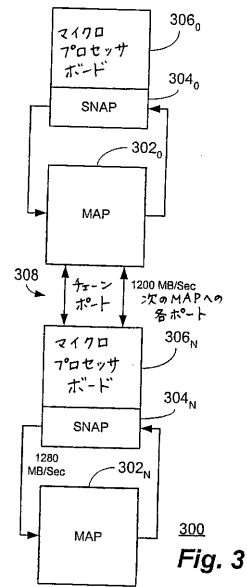
【図2A】



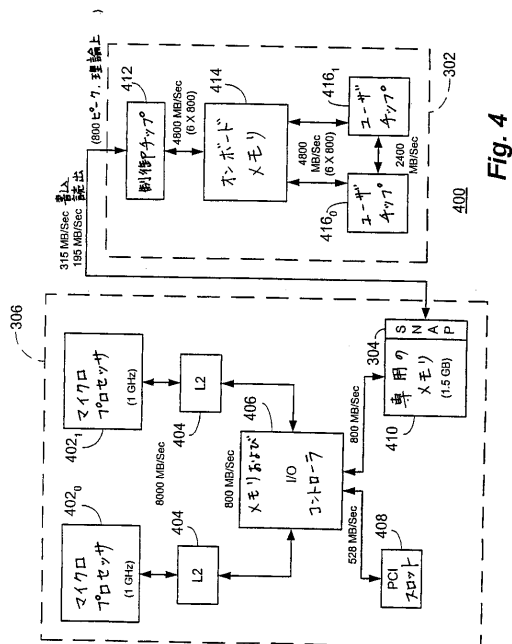
【図 2 B】



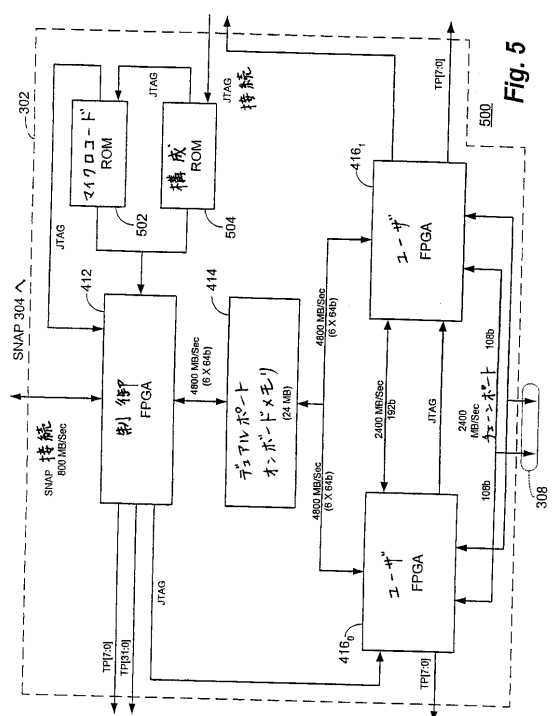
【図 3】



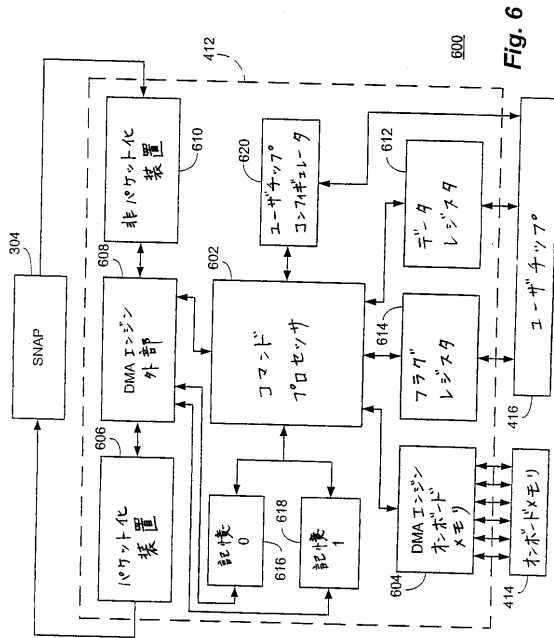
【図 4】



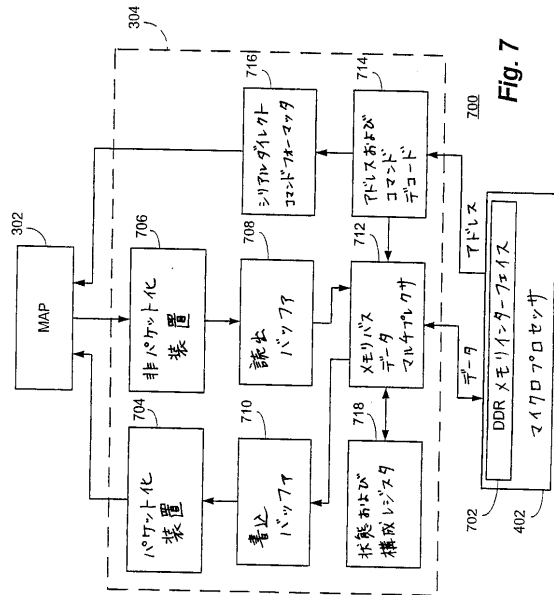
【図 5】



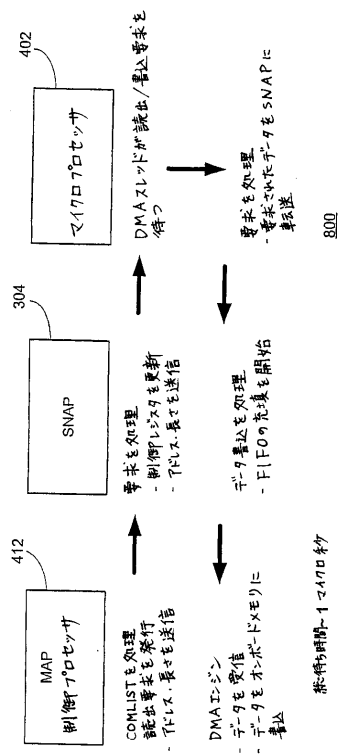
【 図 6 】



【 図 7 】



【圖 8】



**Fig. 8**

待ち時間~1マイクロ秒

---

フロントページの続き

(74)代理人 100098316

弁理士 野田 久登

(74)代理人 100109162

弁理士 酒井 将行

(72)発明者 フッペンサル, ジョン・エム

アメリカ合衆国、8 0 9 0 8 コロラド州、コロラド・スプリングス、バージェス・ロード、1 0  
0 1 5

(72)発明者 シーマン, トーマス・アール

アメリカ合衆国、8 0 9 1 9 コロラド州、コロラド・スプリングス、ハフマン・コート、5 7 6  
5

(72)発明者 バートン, リー・エイ

アメリカ合衆国、8 0 8 1 4 コロラド州、ディバイド、シーダー・マウンテン・ロード、9 6 6

審査官 堀江 義隆

(56)参考文献 米国特許出願公開第2 0 0 2 / 0 0 1 9 9 2 6 ( U S , A 1 )

特開平0 8 - 1 8 0 0 3 0 ( J P , A )

特表2 0 0 3 - 5 3 2 2 2 4 ( J P , A )

特開平0 2 - 1 6 3 8 4 8 ( J P , A )

特開平0 9 - 1 7 9 7 7 8 ( J P , A )

(58)調査した分野(Int.Cl. , D B 名)

G06F 12/00 - 12/06

G06F 13/16 - 13/18