

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5872560号
(P5872560)

(45) 発行日 平成28年3月1日 (2016.3.1)

(24) 登録日 平成28年1月22日 (2016.1.22)

(51) Int. Cl.

F I

G 0 6 F 13/16 (2006.01)
 G 0 6 F 13/36 (2006.01)
 G 0 6 F 12/00 (2006.01)
 G 0 6 F 3/08 (2006.01)

G 0 6 F 13/16 5 1 0 A
 G 0 6 F 13/36 5 3 0 A
 G 0 6 F 12/00 5 9 7 U
 G 0 6 F 3/08 H

請求項の数 7 (全 34 頁)

(21) 出願番号 特願2013-530377 (P2013-530377)
 (86) (22) 出願日 平成23年9月23日 (2011.9.23)
 (65) 公表番号 特表2014-501950 (P2014-501950A)
 (43) 公表日 平成26年1月23日 (2014.1.23)
 (86) 国際出願番号 PCT/US2011/053129
 (87) 国際公開番号 W02012/040649
 (87) 国際公開日 平成24年3月29日 (2012.3.29)
 審査請求日 平成26年9月10日 (2014.9.10)
 (31) 優先権主張番号 61/386, 237
 (32) 優先日 平成22年9月24日 (2010.9.24)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 513070613
 テキサス・メモリー・システムズ・インコ
 ーポレイテッド
 アメリカ合衆国、テキサス・77042、
 ヒューストン、ウエストヘイマー・ロード
 ・10777、スイート・600
 (74) 代理人 110001173
 特許業務法人川口国際特許事務所
 (72) 発明者 フロスト、ホロウェイ・エイチ
 アメリカ合衆国、テキサス・77082、
 ヒューストン、ベルサイユ・レイクス・レ
 イン・11615

最終頁に続く

(54) 【発明の名称】 高速メモリ・システム

(57) 【特許請求の範囲】

【請求項 1】

高速シリアル通信を有するフラッシュ・ベースのメモリ・システムであって：
 複数の入出力モジュールであり、それぞれの入出力モジュールは1つ以上の外部通信リン
 クを通じて、外部デバイスと通信するように構成されている複数の入出力モジュール；
 複数のフラッシュ・ベースのメモリ・カードであって、それぞれのフラッシュ・ベースの
 メモリ・カードは複数のフラッシュ・メモリ素子を含み、それぞれのフラッシュ・メモリ
 素子は、ブロックに分割された物理メモリ・スペースを有し、それぞれのブロックは更に
 ページに分割されており、それぞれのページは、その上でメモリ操作が遂行される、個別
 にアドレス可能なメモリロケーションを表し、多数のそのようなメモリロケーションは、
 1ブロックのグルーピング内で、同時に消去可能である、それぞれの複数のフラッシュ・
 ベースのメモリ・カード；並びに
 複数のクロスバー・スイッチング素子であって、それぞれのクロスバー・スイッチング素
 子は、フラッシュ・ベースのメモリ・カードのそれぞれの1つと接続され、及び前記入出
 力モジュールのそれぞれの1つが、フラッシュ・ベースのメモリ・カードのそれぞれの1
 つと通信することが可能なように構成されている、複数のクロスバー・スイッチング素子
 とを含み、
 それぞれの入出力モジュールは、それぞれのクロスバー・スイッチング素子と、高速シリ
 アル通信リンクにより接続され、それぞれの高速シリアル通信リンクは、それぞれの入出
 力モジュールが、それぞれのクロスバー・スイッチング素子への、及びクロスバー・スイ

10

20

ッティング素子からのコマンド、命令、及び／又はデータを表すビットを送受信することを可能にし、並びにそれぞれのクロスバー・スイッチング素子は、フラッシュ・ベースのメモリ・カードのそれぞれの１つに、複数の並列通信リンクにより接続され、それぞれの並列通信リンクは、１つのクロスバー・スイッチング素子を、フラッシュ・ベースのメモリ・カードのそれぞれの１つのフラッシュ・メモリ素子の１つと接続している、フラッシュ・ベースのメモリ・システム。

【請求項２】

入出力モジュールから送信されるコマンド、命令、及び／又はデータを表す前記ビットが、読み出し・修正・書き込みを含むDMA要求を含む直接メモリ・アクセス(DMA)要求の一部である、請求項１に記載のフラッシュ・ベースのメモリ・システム。

10

【請求項３】

前記入出力モジュールが、以下のプロトコル：ファイバー・チャンネル(Fibre Channel)、インフィニバンド(InfiniBand)、イーサネット(Ethernet)、及び前面パネル・データ・ポート(Front Panel Data Port)のうちの１つを含む高速通信プロトコルを用いる１つ以上の外部通信リンクを通じて、外部デバイスと通信するために構成されている請求項１に記載のフラッシュ・ベースのメモリ・システム。

【請求項４】

それぞれの前記入出力モジュールが、高速通信プロトコルを用いる１つ以上の外部通信リンクを通じて、前記入出力モジュールが外部デバイスと通信可能とするために構成された高速インターフェースを含む、請求項３に記載のフラッシュ・ベースのメモリ・システム。

20

【請求項５】

それぞれの前記入出力モジュールが、更に複数のマルチ・ギガビット・トランシーバーを有する高速制御装置を含み、１つの前記マルチ・ギガビット・トランシーバーは前記入出力モジュールに接続されたそれぞれの高速シリアル通信リンクのためであり、それぞれの前記入出力モジュールは、前記マルチ・ギガビット・トランシーバーを用いて、それぞれの前記クロスバー・スイッチング素子との間のコマンド、命令及び／又はデータを表すビットを送受信するために構成される、請求項４に記載のフラッシュ・ベースのメモリ・システム。

30

【請求項６】

それぞれの前記入出力モジュールが、更にCPU及び共有スイッチング素子を含み、この共有スイッチング素子は、CPU、１つ以上の外部通信リンクを介する外部デバイス、及び前記マルチ・ギガビット・トランシーバーの間の通信を規制するために構成されている、請求項５に記載のフラッシュ・ベースのメモリ・システム。

【請求項７】

それぞれの前記クロスバー・スイッチング素子が、複数のマルチ・ギガビット・トランシーバーを有する高速制御装置を含み、それぞれの高速シリアル通信リンクのための１つの前記マルチ・ギガビット・トランシーバーは前記クロスバー・スイッチング素子に接続され、それぞれの前記クロスバー・スイッチング素子は、前記マルチ・ギガビット・トランシーバーを用いて、それぞれの前記入出力モジュールとの間のコマンド、命令及び／又はデータを表すビットを送受信するために構成される、請求項１に記載のフラッシュ・ベースのメモリ・システム。

40

【発明の詳細な説明】

【技術分野】

【０００１】

本出願は、２０１０年９月２４日に出願された、“High-Speed Memory System”と題された米国特許仮出願第６１／３８６，２３７号による優先権を主張し、同仮出願は、引用によりその全体が本願に組み込まれる。

【０００２】

50

本開示は一般的に、外部ホスト装置及び／又は外部通信装置によりアクセスされる記憶システムに関する。

【背景技術】

【0003】

記憶システムは、しばしば、入来するデータ及びデータ要求を受け取り、並びにそれらの要求をデータの記憶又は取り出しのために処理するために、さまざまな方法及びデバイスを用いる。しばしば、そのような記憶システムは、限られた数の入出力ポートしか持たないため、及び／又は、そのシステム内に存在するボトルネックのために、その帯域幅が限定される。そのようなボトルネックは、比較的遅いデータパスの使用と同様に、複雑なスイッチング又は伝送構造、及び／又は伝送プロトコルの使用により生じ得る。

10

【発明の概要】

【発明が解決しようとする課題】

【0004】

より効果的な、効率的な、及び最適なメモリ・システムが要求される。

【課題を解決するための手段】

【0005】

開示される実施態様は、より効果的な、効率的な、及び最適な高速メモリ・システムを提供するための方法及び装置に関する。一態様においては、本願で開示される実施態様は、一般に、高速度のシリアル通信を有する、フラッシュ・ベースのメモリ・モジュールに関する。このフラッシュ・ベースのメモリ・モジュールは、複数の入出力（I/O）モジュールであって、それぞれの入出力モジュールは1つ以上の外部通信リンクを通じて、外部デバイスと通信するように構成されている、複数の入出力モジュールを含み、複数のフラッシュ・ベースのメモリ・カードであって、それぞれのフラッシュ・ベースのメモリ・カードは複数のフラッシュ・メモリ素子を含み、それぞれのフラッシュ・メモリ素子は、ブロックに分割された物理メモリ・スペースを有し、それぞれのブロックは更にページに分割されており、それぞれのページは、その上でメモリ操作が遂行される、個別にアドレス可能なメモリロケーションを表し、多数のそのようなメモリロケーションは、1ブロックのグルーピング内で、同時に消去可能である、それぞれの複数のフラッシュ・ベースのメモリ・カードを含み、並びに複数のクロスバー・スイッチング素子であって、それぞれのクロスバー・スイッチング素子は、フラッシュ・ベースのメモリ・カードのそれぞれの1つと接続され、及び前記入出力モジュールのそれぞれの1つが、フラッシュ・ベースのメモリ・カードのそれぞれの1つと通信することが可能なように構成されている、複数のクロスバー・スイッチング素子とを、特に含む。それぞれの入出力モジュールは、それぞれのクロスバー・スイッチング素子と、高速シリアル通信リンクにより接続され、それぞれの高速シリアル通信リンクは、それぞれの入出力モジュールが、それぞれのクロスバー・スイッチング素子への、及びクロスバー・スイッチング素子からのコマンド、命令、及び／又はデータを表すビットを送受信することを可能にし、並びにそれぞれのクロスバー・スイッチング素子は、フラッシュ・ベースのメモリ・カードのそれぞれの1つに、複数の並列通信リンクにより接続され、それぞれの並列通信リンクは、1つのクロスバー・スイッチング素子を、フラッシュ・ベースのメモリ・カードのそれぞれの1つのフラッシュ・メモリ素子と接続している。

20

30

40

【0006】

一般に、別の態様では、開示される実施態様は、拡張可能な高速メモリに関する。この拡張可能な高速メモリは、印刷回路板（PCB）、PCB上に取り付けられ、及び高速メモリ・カードが、1つ以上の高速シリアル通信リンクを通じた1つ以上の外部デバイスからのコマンド、命令、及び／又はデータをあらわすビットを受信可能になるように構成されたインターフェース回路を含み、PCB上に取り付けられた複数のメモリ素子であって、それぞれのメモリ素子はその上でメモリ操作が遂行される物理的メモリ空間を有する複数のメモリ素子を含み、並びにPCB上に取り付けられた制御装置であって、インターフェース回路及び複数のメモリ素子に接続され、前記制御装置は、メモリ操作を遂行するた

50

めにインターフェース回路及びそれぞれのメモリ素子の間の通信を制御するように構成されている、制御装置とを、特に含む。前記インターフェース回路は前記制御装置に、複数の高速シリアル通信回線により接続され、それぞれの高速シリアル通信回線は、高速シリアル通信リンクの1つに対応し、及び前記制御装置は、前記複数のメモリ素子に、所定の数の並列通信回線により接続されており、前記制御装置は、高速シリアル通信リンクからのコマンド、命令及び／又はデータを表すビットを、シリアル・フォーマットから、並列フォーマットに変換するように構成されている。

【0007】

一般に、更に別の態様では、開示される実施態様は、高速シリアル通信を有するメモリ・モジュールに関する。前記メモリ・モジュールは、第一の複数の入力処理ブロック及び第二の複数の入力処理ブロックを含み、それぞれの入力処理ブロックは、シリアル・フォーマットに従ったコマンド、命令及び／又はデータを表すビットを受け取り、及びそのコマンド、命令及び／又はデータを表すビットを、並列フォーマットに従うように再配列するために構成されており、複数のメモリ素子であって、それぞれのメモリ素子は、その上でメモリ操作が遂行される物理的メモリ空間を有する複数のメモリ素子を含み、及び第一及び第二の複数の入力処理ブロックと前記メモリ素子を接続する制御装置であって、及び前記制御装置はメモリ操作を遂行するために、第一及び第二の複数の入力処理ブロック及びそれぞれのメモリ素子の間の通信を制御するために構成されている、制御装置を特に含む。

前記制御装置は：

(a) 第一及び第二の複数の入力処理ブロックにそれぞれ接続された第一の多重チャンネル・バッファ及び第二の多重チャンネル・バッファであって、それぞれの多重チャンネル・バッファは、並列フォーマットのコマンド、命令及び／又はデータを表すビットを前記第一及び第二の複数の入力処理ブロックからそれぞれ受け取り、及びコマンド、命令及び／又はデータを表すビットから所定の数のワードを構築するために構成されており、それぞれのワードは所定の数のビットから成っている、第一の多重チャンネル・バッファ及び第二の多重チャンネル・バッファを含み、

(b) 第一のエラー修正及びデータ保護回路、及び第二のエラー修正及びデータ保護回路であって、前記第一及び第二の多重チャンネル・バッファにそれぞれ接続され、前記第一及び第二のエラー修正及びデータ保護回路は、前記第一及び第二の多重チャンネル・バッファからのワードを、それぞれ受け取り、前記ワードを用いる1つ以上のエラー修正コードのビットを生成させ、それぞれのワードのための前記エラー修正コードのビットを前記ワードに加え、及びそれぞれのワードを、そのワードに加えられたエラー修正コードのビットとともに出力するために構成されている、第一のエラー修正及びデータ保護回路、及び第二のエラー修正及びデータ保護回路を含み；

(c) 第一及び第二のエラー修正及びデータ保護回路にそれぞれ接続された、第一の出力バッファ及び第二の出力バッファであって、前記第一及び第二の出力バッファは、第一及び第二のエラー修正及びデータ保護回路の1つからの第一のワードが、第一及び第二の出力バッファの1つに提供され、及び第一及び第二のエラー修正及びデータ保護回路の別の1つからの次のワードが、第一及び第二の出力バッファの別の1つに提供されることができるよう、順々に第一及び第二のエラー修正及びデータ保護回路からの、前記エラー修正コードビットが加えられたワードを受け取るために構成されている、第一の出力バッファ及び第二の出力バッファを含み、並びに

(d) メモリ・バッファであって、前記エラー修正コードのビットが加えられたワードを、前記第一及び第二の出力バッファから受け取り、及び前記所定の数のワードを所定の方式によって、スーパーワードを形成するために組み合わせるために構成されている、メモリ・バッファとを特に含む。

【図面の簡単な説明】

【0008】

【図1】本開示の特定の教示に従って構築された高速記憶システムを図示している。

10

20

30

40

50

【図 1 A】本開示の高速記憶システムにおいて行われ得る D M A 読み出し (R E A D) - 修正 (M O D I F Y) - 書き込み (W R I T E) 操作の例を図示している。

【図 1 B】本開示の高速記憶システムにおいて行われ得る D M A 読み出し - 修正 - 書き込み操作の例を図示している。

【図 1 C】本開示の高速記憶システムにおいて行われ得る D M A 読み出し - 修正 - 書き込み操作の例を図示している。

【図 2】本開示の高速記憶システムにおいて用い得る入出力モジュールの例示的な実施態様を図示している。

【図 3】本開示の高速記憶システムにおいて用い得る入出力モジュールの代替の例示的な実施態様を図示している。

【図 4】図 1 に関連して上述された種類のクロスバー・モジュールを実行するための例示的なアプローチの 1 つを図示している。

【図 5】本開示のシステムにおいて用い得るメモリ・カードの例示的な実施態様の 1 つを図示している。

【図 6】本開示の高速記憶システムにおいて用い得るメモリ・カードの代替的な実施態様の 1 つを図示している。

【図 7】入出力モジュールが、クロスバー・モジュールなしで、直接メモリ・ボードと通信する代替的な高速記憶システムを図示している。

【図 8】図 7 の高速記憶システムにおいて用い得る高速メモリ・ボードの例示的な実施態様を図示している。

【図 9】例示的な高速メモリボードの一部分の一般的な詳細を図示している。

【図 10 A】高速メモリ・ボードのシリアライザ/デシリアライザ及びパッキング・モジュールの例示的な実施態様を図示している。

【図 10 B】高速メモリ・ボードのための先入れ先出し (F I F O) 記憶バッファの例示的な実施態様を図示している。

【図 10 C】高速メモリ・ボードのための F I F O 記憶バッファの例示的な実施態様を図示している。

【図 10 D】高速メモリ・ボードのための F I F O 記憶バッファの例示的な実施態様を図示している。

【図 11】高速メモリ・ボードのための多重チャンネル F I F O バッファの例示的な実施態様を図示している。

【図 12】エラー修正及びデータ保護回路から高速メモリ・ボードの出力 F I F O バッファへのデータ移動の例示的なプロセスを図示している。

【図 12 A】エラー修正及びデータ保護回路から高速メモリ・ボードの出力 F I F O バッファへのデータ移動の例示的なプロセスを図示している。

【図 13 A】多重の個別高速シリアル通信リンクを通じたデータ移動の例示的なプロセスを図示している。

【図 13 B】多重の個別高速シリアル通信リンクを通じたデータ移動の例示的なプロセスを図示している。

【図 14 A】データが、高速シリアル通信リンクを通じて移動されるときに、異なる W R I T E 操作から不適切にデータを「混ぜること」 (i n t e r m i n g l i n g) を防止するための例示的なプロセスを図示している。

【図 14 B】データが、高速シリアル通信リンクを通じて移動されるときに、異なる W R I T E 操作から不適切にデータを「混ぜること」 (i n t e r m i n g l i n g) を防止するための例示的なプロセスを図示している。

【図 14 C】データが、高速シリアル通信リンクを通じて移動されるときに、異なる W R I T E 操作から不適切にデータを「混ぜること」 (i n t e r m i n g l i n g) を防止するための例示的なプロセスを図示している。

【図 14 D】データが、高速シリアル通信リンクを通じて移動されるときに、異なる W R I T E 操作から不適切にデータを「混ぜること」 (i n t e r m i n g l i n g) を防止

10

20

30

40

50

するための例示的なプロセスを図示している。

【図 1 5】本開示の高速記憶システムの高速シリアル通信リンクを通じて移動されるデータのための R E A D 及び W R I T E アクセスの間の競合制御のための例示的なプロセスを図示している。

【図 1 6】例示的な制御装置及び本開示の高速記憶システムにおける制御装置及び物理 R A M メモリの間の接続を図示している。

【図 1 7】図 1 7 は、本開示の高速記憶システムにおける 2 つの基本的なメモリ周期を通じた図 1 6 の制御装置の例示的な操作を図示している。

【発明を実施するための形態】

【 0 0 0 9 】

上述の図並びに下記の具体的な構造及び機能の記載は、出願人の発明の範囲、又は添付の請求範囲を限定するために提示されるのではない。むしろ、上述の図及び明細書は、任意の当業者が、特許による保護が求められている発明を作成し、使用することを教示するために提供されている。当業者は、明瞭性及び理解のために、本発明の工業的な実施態様の全ての特徴が記載されるか、又は示されているのではないことを認識するであろう。当技術分野の熟練者は、本発明の態様を組み込んでいる実際の工業的な実施態様の開発が、工業的な実施態様のための開発者の究極の目的を達成するために、莫大な数の、実施に特異的な決断を必要とするであろう事を認識するであろう。実施に特異的な決断としては、おそらく限定はされないが、システム関連、ビジネス関連、政府関連、及び他の制限に対する遵守であって、それらは、具体的な実施、場所により、及び時により変化するであろう。開発者の努力は、絶対的な感覚で、複雑であり、多大な時間を要するものであることが、そのような努力は、しかしながら、本開示の利益を有する当業者にとっては、日常的な業務であろう。本明細書において開示された発明、及び教示 (t a u g h t) は、莫大な数の、並びにさまざまな修正及び代替形態を許容するものであることを理解されたい。最後に、限定はされないが、“ a ” などの、単数形の使用は、その品目の数を限定することを意図していない。また、限定はされないが、「頂点」、「底部」、「左」、「右」、「上の」、「下の」、「下へ」、「上へ」、「側面」などの関係語の使用は、明細書において、図への特定の参照において明瞭性のために用いられ、及び発明又は添付の請求の範囲を限定することを意図していない。

【 0 0 1 0 】

図、特に図 1 に移るが、ここには、本開示の特定の教示に従い構築される高速記憶システム 1 0 0 が図示されている。一般にこの高速記憶システム 1 0 0 は、R E A D 及び W R I T E 要求などのデータ関連要求を外部ホスト装置から受け取り、及び物理メモリへの / 物理メモリからのデータの記憶及び / 又は取り出しのためにそれらの要求を処理する。

【 0 0 1 1 】

例示的な記憶システム 1 0 0 は、接続可能な高帯域記憶システムを提供するために、複数の入出力モジュール、高速シリアル通信チャンネル、設定変更可能なクロスバー・モジュール、並列メモリ・バス、及び物理メモリ・カードを包含して利用する。システム及びその多くの新規な態様の追加的な詳細は、以下で提供される。

【 0 0 1 2 】

図 1 を参照するが、上記システム 1 0 0 は、複数の入出力 (I / O) インターフェース・モジュール 1 0 2、1 0 4、1 0 6、1 0 8 及び 1 1 0 を包含する。この例には、5 つの入出力モジュールが図示されているが、入出力モジュールの数は致命的に重要ではなく、変更の対象となる。それぞれの入出力モジュールは、1 つ以上の外部ホスト装置と、ホスト通信チャンネル (それぞれ、1 0 2 a、1 0 4 a、1 0 6 a、1 0 8 a 及び 1 1 0 a) を介して連結される。これらのチャンネル 1 0 2 a、1 0 4 a、1 0 6 a、1 0 8 a 及び 1 1 0 a は、サーバーや他の装置などの、ホストとして稼動可能な外部デバイスが、入出力インターフェース・モジュール 1 0 2、1 0 4、1 0 6、1 0 8 及び 1 1 0 から、コマンド、命令及びデータを受け取ること、及び / 又はそれらへ、コマンド、命令及びデータを提供することを可能にする。

【 0 0 1 3 】

このホスト通信チャンネルは、さまざまな形態をとることができ、及び限定はされないが、ファイバー・チャンネル、インフィニバンド、イーサネット（登録商標）、及び前面パネル・データ・ポート（F P D P）などのさまざまなプロトコルに従う。この通信リンクの正確な物理及び論理構成はこの開示されたシステムにとり致命的に重要ではないが、開示されたシステムにより提供される利益を最大限に利用するためには、ホスト通信リンクが、高帯域データ移転をサポートする能力があることが望ましい。

【 0 0 1 4 】

図 1 の例においては、入出力モジュール 1 0 2、1 0 4、1 0 6、1 0 8 及び 1 1 0 のそれぞれは、複数の高速シリアル通信チャンネル（図 1 に 1 1 5 として集合的に標識されている）により、4 つの設定変更可能なクロスバー・モジュール 1 2 0、1 2 2、1 2 4 及び 1 2 6 に連結されている。以下でより詳細に記載するが、これらのチャンネルは、入出力モジュールがクロスバー・モジュールとの間でコマンド、命令及び / 又はデータを通信することを可能にする。

10

【 0 0 1 5 】

図示された例においては、4 つのクロスバー・モジュール 1 2 0、1 2 2、1 2 4 及び 1 2 6 があるが、クロスバー・モジュールの数は致命的に重要ではなく、変化できることが理解されるであろう。図 1 では、4 つのクロスバー・モジュールのそれぞれが、5 つの入出力モジュールのそれぞれと、個別の（*d i s t i n c t*）直列システム通信チャンネルを介して連結される。代替的な実施態様は、クロスバー・モジュールの 1 つ以上が入出力モジュールの全てよりも少ない数に連結されるものに想定される。

20

【 0 0 1 6 】

図 1 に反映されているように、例示的なシステムにおいては、それぞれのクロスバー・モジュールは、複数の並列メモリ・バスにより、メモリ・カードの集団に連結されている。例えば、図示された例においては、クロスバー・モジュール 1 2 0 は、メモリ・カードの第一のグループ 1 4 0 に、並列メモリ・バスの第一のグループ 1 3 0 により連結されている。同様の方法で、クロスバー・モジュール 1 2 2、1 2 4 及び 1 2 6 は、メモリ・カード・グループ 1 4 2、1 4 4 及び 1 4 6 に、それぞれ並列メモリ・バス・グループ 1 3 2、1 3 4 及び 1 3 6 を介して連結されている。図示された例においては、それぞれのクロスバー・モジュールは、5 つのメモリ・カードのグループに、それぞれがデータ転送と同様に制御、コマンド、モニタリング情報を通信するために使用できる、5 つの多重送信並列バスにより連結されている。メモリ・カードの数及びメモリ・インターフェース・バスの種類は、本開示の教示を逸脱することなく変更できることが認識されるであろう。

30

【 0 0 1 7 】

図示された例においては、それぞれのメモリ・カード、及び従ってメモリ・カードのグループのそれぞれは、一般的にメモリロケーションの特定の範囲に対応する。この特定の範囲は、実際の物理アドレス範囲（例えば、それぞれのアドレスが、前記メモリ・カードに関連付けられている特定の物理メモリロケーションに対応しているアドレスの範囲）、又は論理アドレス範囲（例えば、それぞれの論理アドレスメモリ・カード上の制御装置又は外部制御装置により、メモリ・カードに付随する物理メモリロケーションに対してマップされることができる、論理アドレス範囲）であることができる。

40

【 0 0 1 8 】

図 1 の例においては、全体のシステム 1 0 0 は、一連の利用可能な記憶アドレスを外部ホストに提供し、及びそれぞれのメモリ・カードは、一般的にシステム 1 0 0 により提供されるアドレスの特定の範囲に対応する。加えて、メモリ・カードのそれぞれのグループがアドレスの特定の範囲に関連付けられているために、それぞれのクロスバーも、それが接続されているメモリ・カードのグループに対応するアドレスの範囲に関連付けられる。

【 0 0 1 9 】

図 1 のシステム 1 0 0 においては、システム制御装置 1 5 0 は、入出力モジュール 1 0 2、1 0 4、1 0 6、1 0 8 及び 1 1 0 のそれぞれ、及びクロスバー・モジュール 1 2 0

50

、 1 2 2、 1 2 4 及び 1 2 6 のそれぞれに連結されている。システム制御装置は図 1 の 5 つの入出力モジュールの全てに接続されているが、そのような接続のうちの 1 つだけが図示されている。同様に、システム制御装置、及びクロスバー・モジュールの間の 4 つの接続の内の 1 つだけが示されている。一般にシステム制御装置 1 5 0 は、タイミング及び同期信号を提供するが、それは、監視、報告及び他の監督タスクの遂行のためにも用い得る。

【 0 0 2 0 】

一般的な操作においては、入出力モジュールのそれぞれは、入出力モジュールが接続されている外部ホスト装置から、外部データ関連要求を、そのホスト通信リンクを通じて受け取る。それぞれの入出力モジュールは、記憶システム 1 0 0 への、又は記憶システム 1 0 0 からのデータの、記憶及び / 又は取り出しを可能にするために、これらのデータ関連要求を処理する。図 1 の実施態様においては、このデータ要求が受け取られ、及び個別の通信「パケット」の形態の応答が返されるが、代替的な通信プロトコルを用いることもできる。

【 0 0 2 1 】

正確なプロトコルの詳細は用途に依存するが、外部プロトコルは、高いデータのスループットを促進するために低い通信オーバーヘッドに対して選択されるべきである。図 1 のシステムにおいては、前記入出力モジュールにより受け取られて、処理される通信パケットは、典型的にはヘッダー、及びそのパケットがデータを移転するものであれば、データ・ペイロードを含む。ヘッダーの正確な内容及びフォーマットは、用途ごとに異なるが、このヘッダーは典型的には以下のものを含む：(a) 要求の種類を識別する識別子（例えば、その要求が、システム内にデータを記憶するための W R I T E 要求であるか、又はそのシステム内に以前に記憶されたデータを取り出すための R E A D 要求であるのか）；(b) 要求に関連付けられた特定のアドレスを示す識別子（例えば、W R I T E 要求については、一般的にその要求に関連付けられたデータの記憶が開始されるロケーションを識別するか、又は R E A D 要求については、その要求に関連付けられたデータの取り出しが開始されるべきロケーションを識別する、特定の論理又は物理“アドレス”であって、これらのアドレスのそれぞれは、時に「T A R G E T A D D R E S S」と称される）；並びにいくつかのシステムについては、(c) データ関連要求のソースの識別（時に「S O U R C E A D D R E S S」と称される）；及び(d) 要求に関連付けられたデータの量を示す識別子（例えば、その要求の一部としてのシステムへの保存、又はシステムからの取り出しのためのバイト数、ページ数、又は他のデータ量の表示）。いくつかの通信プロトコルについては、全ての外部データ関連要求は、特定のサイズにされたデータのページ、又は特定の数のデータ・ワードなどの、データの特定の量に関連付けられていると推測される。このようなプロトコルでは、要求が、特定のプロトコルのためのデータの標準量に關することが暗黙的であるために、特定の外部要求に関連付けられたデータの量を明確に指定することは必要ない。

【 0 0 2 2 】

外部デバイスがデータ関連要求を、それらが接続された入出力モジュールに発行するちょうどそのときに、入出力モジュールも D M A（直接メモリ・アクセス）要求、又は単に D M A と呼ばれる要求を、それらが接続されたクロスバー・モジュール及びそれらが取り付けられたメモリ・カードに発行する。それぞれの外部要求は、多重のメモリ・カード、及びクロスバー・モジュールに及ぶシステム・アドレスの範囲にアクセスを要求することができる。たとえ、特定の外部要求が、ただ 1 つのメモリ・カードにアクセスを要求する場合でも、外部要求に関連付けられたデータの量は、単独の D M A 要求に関連付けられたデータの最大量を超える可能性がある。この理由のために、入出力モジュールは、それが受け取るそれぞれの外部要求のために 2 つ以上の D M A 要求を発行することができる。この場合には、それらの D M A 要求は、データ記憶又は取り出しが、そのシステム内の正しい物理メモリロケーションで行われるようにメモリ・カードに向けられる。外部 R E A D 要求の場合には、例えば、この入出力モジュールは、要求されたデータを、それが保存さ

10

20

30

40

50

れているさまざまなロケーションから「収集する」ために、1つ以上のDMA READ要求を発行する。外部WRITE要求の場合には、この入出力モジュールは、全ての受け取られたデータを、それが保存されるべきさまざまなロケーションに送達するために、1つ以上のDMA WRITE要求を一般的に発行する。

【0023】

一般にDMA WRITE要求は、その要求及び同一のパケット内に存在する関連付けられたデータとともに、「パケット」に編成される。一方、DMA READ要求は、一般的に要求情報（アドレス、データ量など）のみを含む。これらの要求タイプのそれぞれに対する応答は、おおむね相補的である。入出力モジュールにより発行されたそれぞれのDMA WRITE要求に対して、クロスバーは、WRITE操作が成功したか失敗したかを示す状況情報を最小限含むDMA WRITE応答パケットを返す。入出力モジュールから発行されたそれぞれのDMA READ要求に対して、クロスバーは、要求されたデータ及び、最小限返されたデータの正当性を示す状況情報の両方を含むDMA READ応答パケットを返す。

【0024】

システム内のメモリ・カードそれぞれがフラッシュ・メモリを用いる一実施態様では、それぞれのDMA要求は、データの固定量の移転を、その固定量が単一のフラッシュ・メモリ・「ページ」に記憶された等しいものとして特定する。この例では、1ページは、一般的に個別的にフラッシュ・メモリに書き込まれることが可能な情報の最小量である。

【0025】

図1のシステム100では、一般的に、DMA要求は固定量のデータ（4キロバイト）を、固定されたアドレス境界上に整列されたアドレス（4キロバイト）から、及びそのアドレスへと、移転することであると定義される。いくつかの場合には、外部WRITE要求により広げられるアドレス範囲が、これらの固定DMA要求アドレス境界上に整列しないことも可能である。図1A～1Cに示されるように、外部WRITE要求は、DMA要求の許可された開始アドレスに対応しないアドレスから開始されることができる。同様に、DMA WRITE要求は、DMA要求の終了アドレスに対応しないアドレスで終了できる。この「不均衡（misalignment）」が起きるとき、前記入出力カードは、1つ以上の読み出し - 修正 - 書き込み（RMW）操作を遂行できる。読み出し - 修正 - 書き込み操作では、前記入出力モジュールは、外部WRITE要求の開始及び/又は終了にまたがるアドレス範囲からのシステム・メモリからデータを取り出すためにDMA READ要求を発行する。次いで、前記入出力モジュールは、この取り出されたデータの一部を、整列されていないWRITE要求からの外部データによって修正（置換）し、及び次いで、相補的なDMA WRITE要求を発行することにより、この修正されたデータを、システムメモリに返す。このような方法で、前記入出力モジュールは、個別のDMA要求の整列制限に従わないアドレス範囲のシステムメモリ内でのデータの保存を可能にする。

【0026】

図1Aは、外部WRITE要求が1ページに完全に含まれるが、開始アドレスがDMAアドレス境界に整列していないために、読み出し - 修正 - 書き込み操作が遂行されることを引き起こす、データのアドレス及び量160を特定する、「中に含まれる（contained within）」場合の例を示している。この場合には、1つのDMA READ要求のみが必要である。新しい（修正された）データは、DMA READ要求と同じ（整列された）アドレスを使用する発行されたDMA WRITE要求とともに、以前に保存されたデータのロケーションの上に重ねられる。図1Bは、外部WRITE要求を2つのフラッシュ・ページにまたがる、特定されたアドレス及びデータ量170とともに示している。この「またがる（straddle）」ケースは、データの取り出しのために2つのDMA READ要求を生成させ、及び続いて新しいデータ・イメージを戻すために2つのDMA WRITE要求を生成させる。図1Cは、その特定されたアドレス及びデータ量180が3ページにわたっている外部WRITE要求を示している。この「広

がる (spanning)」場合では、入出力カードは、第一のページに対し1つのDMA READ要求を発行し、新しいデータを重ね、及びこの変化に責任を持つためにDMA WRITE要求を発行する。次のページは完全に新しいデータにより上書きされるために、入出力カードは、このページで読み出し - 修正 - 書き込み操作を行う必要はなく、及び整列されたアドレスへ、単にDMA WRITE要求を発行する。しかしながら、新しいデータ・セットの残りは、第3のページを完全には満たさない。従って、入出力カードは、以前に記載した方式でこのページのために読み出し - 修正 - 書き込みを遂行する。これらの3つの場合は、外部WRITE要求操作の種類の、遂行できるサブセットに過ぎず、及び単に読み出し - 修正 - 書き込み操作を図示することが意図されていることが理解されるであろう。

10

【0027】

外部要求の構造は一般的にプロトコル依存性である一方、内部DMA要求/応答プロトコルは、一般的に特定のシステムに対して固定されており、及びシステム性能を最大化するために設計されている。図1のシステムでは、DMA要求及びDMA応答はヘッダー、及びもしパケットがデータ転送のためのものである場合には、データ・ペイロードを含むパケットとして構造化される。

【0028】

DMA要求は典型的には以下のものを含む：(a) DMA要求の種類を示す識別子（すなわち、DMA要求が、データをシステム内に保存するためのWRITE要求であるか、又はシステム内に以前に保存されたデータを取り出すためのREAD要求であるか）；(b) そのDMA要求に関連付けられた特定のアドレスを示す識別子（例えば、DMA WRITE要求に対しては、一般的に、要求に関連付けられたデータの保存が開始されるべきロケーションを特定するか、又はDMA READ要求に対しては要求に関連付けられたデータの取り出しが開始されるべきロケーションを特定する、特定の論理又は物理「アドレス」）；(c) DMA要求がそこから由来する入出力モジュールを示す識別子；及び(d) それぞれの入出力モジュールに対しては、そのモジュールから由来する特定のDMA要求を一意的に識別する「タグ(tag)」。DMA WRITE要求に対しては、この要求は、その要求に対してシステムメモリ中に保存されるべきデータも含む。図1のシステム100においては、それぞれのDMA要求に関連付けられたデータの量は固定されている。データ量が可変であるシステムに対しては、ヘッダーは更に(e) その要求に関連付けられたデータの量を示す識別子も含む。

20

30

【0029】

DMA応答は典型的に以下のものを含む：(a) DMA応答の種類を示す識別子（すなわち、応答がDMA WRITE要求か、又はDMA READ要求に対応するものか）、(b) それぞれの入出力モジュールに対しては、入出力モジュールにより、それぞれのDMA応答を、それに対応するDMA要求に一意的に関連付けるために使用できる「タグ」、及び(c) DMA WRITE応答に対しては、対応するWRITE操作が成功であるか、失敗であるかを表示し、又はDMA READ要求に対しては、取り出されたデータの正当性を表示する、状況インジケータ。DMA READ応答に対しては、その応答は、さらにオリジナルのDMA READ要求により要求されたデータも含む。

40

【0030】

図1のシステム100においては、入出力モジュールは、外部要求を満たすために対応するDMA要求が向けられるべきクロスバー（及び、可能であればそのメモリボード）を決定するために、それがREAD要求であるか、又はWRITE要求であるかを決定するために、特に、その要求を最小限度処理することにより、外部データ関連要求に対して応答する。

【0031】

上述のように、入出力モジュールにより外部データ関連要求が最小限処理された後に、システムの構成要素は要求を満たすために作動する。例えば、入出力モジュールは、その種類（READ又はWRITE）を識別するためと同時に、外部要求を満たすために発行

50

されるべき特定のDMA要求を決定するために、最初にそれぞれの外部データ関連要求を処理する。この入出力モジュールは、次いで、更なる処理のために、必要なDMA要求を適切なクロスバー・モジュールに伝達する。クロスバー・モジュールは、順々に、(a)それぞれの受け取られたDMA要求が向けられるべきメモリ・カードを決定し、(b)それぞれの受け取られたDMA要求を、シリアル・フォーマットからメモリ・カードに提供するのに適切な並列フォーマットに変換し、及び次いで、(c)それぞれのDMA要求の全部又は一部を、そのアドレス範囲が前記DMA要求に関連付けられているアドレスに広がっている、メモリ・カードに転送する。以下でより詳細に記載するように、メモリ・カード内の電気回路は、提供されたデータを物理メモリに保存するためか(DMA WRITE要求に対して)、又は物理メモリからデータを取り出すか(DMA READ要求に対して)の、いずれかのために、更にそのDMA要求を処理する。一般的に、それぞれのDMA WRITE要求に関連付けられたデータ・ペイロードが、その要求の一部として送達されることに注目することが重要である。いくつかの実施態様では、データが成功裡にメモリ・カードに伝送されたことを確認するために、肯定応答(ACK)信号が、クロスバー・モジュール及びメモリ・カードの間に存在することができる。

10

【0032】

入出力モジュールのための高速シリアル・システム通信リンク、及びクロスバー・モジュール、設定変更可能なクロスバー・モジュール、及びクロスバー・モジュールのための高速並列メモリ通信リンク、及びメモリ・カードの組み合わせの使用により、記憶システム100は、極度に高帯域のシステムを提供する。

20

【0033】

上記で開示された実施例において、入出力インターフェース・モジュールをクロスバー・モジュールに連結させる高速シリアル通信リンクは、コマンド及び制御情報(例えば、DMA要求のヘッダーに見出される種類の情報)、並びにデータ(例えば、データ・ペイロード中で提供される情報)の両方を運搬するために用いられる。データ・ペイロード中で受け取られた種類のデータの移転を含まない通信を提供するために、いくつかの追加的な最小限のシリアル通信リンクが用いられている代替的な実施態様が想起される。例えば、比較的低帯域、低速のシリアル通信リンクが、通信の特定の 카테고리 に用いることができるであろうし、従って高速リンクから一定量のトラフィックをオフロード(offloading)し、及びこれらのリンクの性能の中で、対応するブーストを達成する実施態様が想起される。それぞれのDMA READ要求及びDMA WRITE応答は、例えば、この方式で伝送でき、その両方もが、それぞれのパケットの中で、データを移送するDMA WRITE要求及びDMA READ応答に対して高速シリアル・リンクを保存する。DMA READ要求、及びDMA WRITE応答はいかなるデータ・ペイロードも転送しないため、それらは一般的にサイズにおいてはるかに小さく、及びより遅い通信リンクを、顕著な性能の低下なしで利用することができるであろう。代替的な実施態様では、DMA READ要求及びDMA WRITE応答は、代替的な低帯域シリアル・チャンネルを通じて転送されるか、全てのクロスバーへ、又は全てのクロスバーから別個に共有される高速リンクを通じるブロードキャストし得る。

30

【0034】

高速シリアル・リンクは、データと同様の信号速度において、このサイドバンド・データを提供するために、マルチドロップ・バスを用いて、全てのクロスバーを全ての入出力モジュールに接続するためにも使用し得るであろう。代替的に、前記サイドバンド・リンクは、環状構造において実行し得るであろう。両方の場合に、このことは、追加される競合制御論理(arbitration logic)及びハードウェアを費やして、システム内のポイントツーポイント接続の数を減らすであろうが、より高いデータ処理能力及び帯域幅の追加的利益を伴う。

40

【0035】

上記で説明された特定の素子に関する追加的詳細、及びその素子の代替的な実施態様及び/又は全体のシステムは以下で提供される。

50

【0036】

入出力モジュール

図1の入出力モジュールは、多くの形態をとることができる。好適な入出力モジュールの1つの例示的な実施態様が図2に提供されている。

【0037】

図2を参照すると、例示的な入出力モジュール200が図示されている。この例では、入出力モジュール200は、2つの別個のファイバー・チャンネル(FC)ポート(201及び203)を有するインターフェースを提供し、それぞれが、特定の入出力モジュール200がファイバー・チャンネル・ホスト・バス・アダプター(図示されていない)と接続することを可能にし、それによって、外部ホストがSCSI装置としてのシステム1000にアクセスすることを可能にする。それぞれのポート(201又は203)は、単独のトランスミット・リンク及び単独のレシーブ・リンクを含み、それぞれのリンクは2.125ギガビット/秒(Gb/sec)、又は4.25 Gb/sec のいずれかで作動する。それぞれのポートは、ポイントツーポイント及び競合制御型ループ・ファイバー・チャンネル・プロトコルをサポートできる。この物理インターフェース、及び前記SCSI、及びファイバー・チャンネル(FC)プロトコルは、例示的であって、他の物理インターフェース及びプロトコルを用い得ることが理解されるであろう。

【0038】

図2を参照し、前記例示的な入出力モジュール200は、情報を光と電気フォーマットの間で変換するための光電変換器202a及び202bを含む。一般にこれらの装置のそれぞれ(202a及び202b)は、その外部光レシーブ・リンク(201又は203)で受け取った信号を電気信号に変換し、その電気信号は次いで高速制御装置204に転送される。同様に、これらの装置のそれぞれは、高速制御装置204で受け取られた電気信号を光信号に変換し、この光信号はその外部光トランスミット・リンクを通じて伝送される。周知の光電変換器を用いることができる。

【0039】

上記で示唆されたように、前記光電変換器は、高速制御装置204に連結される。図2の例では、前記高速制御装置204は、例えばXilinx Virtex-4 FPGA装置のような、構成されたフィールド・プログラマブル・ゲート・アレイ(FPGA)である。この制御装置204は、通信リンクにより、Freescal MPC8547プロセッサ206などの、プログラム可能なマイクロプロセッサ(CPU)206及びメモリ208に連結されている。図示された例においては、メモリ208は、DDRメモリ成分から形成されている。

【0040】

見られるように、前記制御装置204は4つの高速シリアル・システム通信チャンネル210、212、214及び216を提供する。前記プロセッサ206及び制御装置204は、前記外部ファイバー・チャンネル・ポート(201及び203)及び前記4つのシリアル・システム通信チャンネル210、212、214及び216の間でのプロトコル変換機能とともに遂行する。図示された例においては、シリアル・システム通信チャンネルのそれぞれは、トランスミット(TX)サブチャンネル及びレシーブ(RX)サブチャンネルの両方を含む全二重物理層を含む。それぞれのサブチャンネルは、ともに結合されて、10 Gb/s データのデータ移転速度の能力がある単独のサブチャンネルを形成する、2つの個別の5 Gb/s シリアル通信リンクを更に含む。この例では、データは、シリアル・システム通信チャンネルを通じて伝送される前に、8 B/10 Bで符号化され、8 Gb/sec (又は毎秒あたり1 Gバイト(GB/sec))のデータ転送速度をもたらす。代替的な符号化のスキーム(例えば、64 B/66 B符号化)も用い得るであろう。図2の例では、それぞれのシリアル通信チャンネルは、Xilinx Virtex-4 FPGA装置内で利用可能なマルチ・ギガビット・トランシーバー(MGT)モジュールを用いて実行されるが、他の同様な実行方法が、本開示の教示を逸脱することなく使用できることが理解されるであろう。FPGA装置のXilinx Virtex-4ファミ

リーにより可能になるような通信リンクを結合するために利用可能な手順を、前記通信チャンネルを形成するために用いることができる。代替的に、他のシリアル化/脱シリアル化 (SerDes) プロトコルに関連付けられた結合スキームも用い得る。

【0041】

図2の例では、前記入出力インターフェース・モジュール200は、4つのシリアル・システム通信チャンネル210、212、214及び216を提供し、それぞれは、異なるクロスバー・モジュールに接続されている。4つの全二重シリアル・システム通信チャンネルの使用は、それぞれの入出力モジュールに対して、潜在的な4GB/secのリード帯域及び4GB/secのライト帯域を集合的に提供する。もし、図1のシステム100が5個の入出力モジュールを使うならば、例示的な図示されたシステムは、合計20GB/secのリード帯域及び合計20GB/secのライト帯域を有するであろう。

10

【0042】

一般に前記高速制御装置204は、FCインターフェース201及び203を通じて受け取られる外部データ関連要求を受け取り、及び処理する。上述したように、それぞれの外部データ関連要求は、典型的には以下を示すヘッダーを含み：(a)要求がREAD又はWRITE要求のどちらであるのか、及び(b)TARGET ADDRESSを示す。このヘッダーは随意に、(c)SOURCE ADDRESS、及び(d)要求に関連付けられたデータの量の表示、を示すことができる。WRITE要求に対しては、前記要求は、更にデータ・ペイロードを含む。前記高速制御装置204は、前記要求を処理し、及びヘッダー情報の大半(しかしデータ・ペイロードではなく)を、プロセッサ又はCPU206に提供できる。このCPU206は、FCプロトコルの少なくともいくつかの部分を管理し、及びシステムインターフェース全体の管理を支援する。

20

【0043】

前に説明したように、前記高速制御装置204及びCPU206は、外部ファイバー・チャンネル・リンクを管理するために、ともに作動する。いくつかの最小のCPU支援に加えて、ほとんどの実施態様では、前記高速制御装置204は、外部データ関連要求のそれぞれを満たすために要求される内部DMA要求を発行することに単独で責任を負う。特に、前記高速制御装置204は、それぞれの外部要求を満たすために、発行すべき特定のDMA要求、及びどのクロスバーにそのDMA要求が発行すべきかを決定する。高速制御装置が、外部要求を満たすために、要求されるクロスバー・モジュール及びDMA要求を特定するとすぐに、それは適切なクロスバーに対するDMA要求の発行を開始する。一般に、入出力モジュール200により発行されることのできる、未処理のDMA要求(それらに対する応答が未だ受け取られていない要求)の数は、DMAヘッダー内のタグ・フィールドのサイズにより限定される。比較的小さい移転については、要求されるDMA要求の数は少なく、及び前記高速制御装置204は、応答の受領を待つことなくそれらの全てを発行できる。大きな移転に関しては、要求されるDMA要求の数は大きく、及び前記高速制御装置204は、DMA要求の発行を遅延させるように強いられる。8ビットのタグ値の場合、例えば、前記高速制御装置は、タグ値を再使用することなく、最初に256個のDMA要求を発行できるであろう。それぞれの未処理DMA要求は、一意的にタグ付けされ、前記高速制御装置がそれぞれのDMA応答を対応するDMA要求に関連付けることを可能にする。DMA要求への応答が返されたときに、前記高速制御装置は、完了されたDMA要求からのタグ(それに対するDMA応答が受け取られたもの)の再使用を開始できるであろうし、それにより、その後のDMA要求の発行を可能にする。

30

40

【0044】

一般にDMA要求及びそれらに対応するDMA応答は、シリアル・システム通信チャンネル210、212、214、及び216を通じて移転される。それぞれの入出力モジュールは、4つのシリアル通信チャンネルに提供するために、それは、同時に、又はほぼ同時に、独立して、4つの異なるクロスバー・モジュールへの、又は4つの異なるクロスバー・モジュールからのデータを独立して伝送及び/又は受け取る。

【0045】

50

以前に説明したように、前記高速制御装置 204 は、それぞれの外部データ関連要求を満たすために、READ 及び WRITE DMA 要求を適正なクロスバー・モジュールに送達することに責任を有する。一実施態様では、前記高速制御装置 204 は、前記高速制御装置メモリ 208 中に保持される、ルックアップ・テーブルを使用することにより、それぞれの DMA 要求をどのクロスバー・モジュールが受け取るべきかを決定する。この実施態様では、前記制御装置 204 は、特定のクロスバー・モジュールを有するシステムにより提供されるアドレスの全体範囲の中で、アドレスの特定の範囲にに関する情報を含むルックアップ・テーブルにアクセスし、及び維持する。このアプローチの 1 つの利益は、それが、設定における、ある程度の柔軟性及びシステムの再配置を提供することである。より更なる実施態様では、メモリ 208 内部の前記ルックアップ・テーブルは、受け取られたアドレスが、予定のクロスバー・モジュール、及び所定のクロスバー・モジュールに連結された、特定のメモリ・カードと関連させることができるように、提示されたアドレスの特定の範囲を、所定のメモリ・カードと、及びさまざまなメモリ・カードをクロスバー・モジュールと更に関連付ける。

10

【0046】

代替的な実施態様では、所定のクロスバー及び受け取られたアドレスの間の関係は所定のアドレス範囲が、常に特定のクロスバー・モジュールに対応できるように「組み込まれ (hardwired)」ている。

【0047】

もっと別の実施態様では、所定のクロスバー及び受け取られたアドレスの間の関係は動的であり、それにより不合格の、又はさもなければ使用不可能のメモリ・カード又はクロスバーを別のメモリ・カード又はクロスバーに「再マップする」ことを可能にする。

20

【0048】

図 2 の入出力インターフェース・モジュール 200 などの、入出力モジュールにより発行される DMA 移転は、固定、又は可変のデータ量のいずれかの移転を特定できる。DMA 移転サイズが固定されている実施態様については、DMA 要求が、移転されるべきデータ量に関する情報を提供することは不必要であり得る。メモリ・カードがフラッシュ・メモリを用いる実施態様では、前記 DMA 移転サイズは、単独のフラッシュ・メモリ・ページに記憶されるデータ量に固定されている。

【0049】

DMA 操作は、可変量のデータを移転できる。そのような実施態様では、それぞれの DMA 要求は、典型的には、移転されるべきデータの量 (しばしばデータの最小量の倍数として) の表示を含む。

30

【0050】

DMA WRITE 操作の間、DMA 移転が固定又は可変移転サイズを用いて遂行されるかどうかに関係なく、前記高速制御装置 204 は、それが取り付けられた外部ホスト装置から受け取るデータをキュー待ちさせる。多重の DMA WRITE 要求から成る、より大きなデータ転送については、前記入出力インターフェースは、DMA WRITE 要求の発行を開始する前に、全ての外部データの受け取りを待つ必要はない。その代わり、前記高速制御装置は、DMA WRITE 要求を発行開始し、及びそれぞれの DMA W 40
RITE 要求をサポートするために十分なデータが利用可能になるとすぐに、DMA WRITE 要求を発行し続ける。

【0051】

DMA READ 操作の間、前記高速制御装置は、データ流を反対方向にキュー待ちさせる (システムメモリの外へ)。この場合には、前記高速制御装置は、多数の DMA READ 要求を発行し、及び次いで DMA READ 応答が、個別のクロスバー・モジュールから返されることを待つ。これらの応答は、それらの対応する DMA READ 要求とは違う順序で受け取られることができるため、前記高速制御装置は、前記クロスバーから受け取られたデータを、それらのデータが、その正しい順序で送達されることができま 50
で、キュー待ちさせなければならない。外部データ要求が、256 個の DMA READ

要求の発行を要求するシナリオを想定すると、最後のDMA READ 応答が、第一のDMA READ 要求に対応することが可能であり、それにより256個のDMA 移転からのデータをキュー待ちさせることを、前記入出力インターフェースに要求する。前記入出力インターフェースは、データが連続して送達される場合にのみ、外部要求者（ホスト装置）へのデータをアンロードできる。

【0052】

上述された方式によって、前記入出力インターフェース・モジュール200は、前記要求を満たし、及び要求されたデータを外部のホスト装置に提供するために、外部ホストからデータ関連要求を受け取ることができ、システム100内のメモリにアクセスできる。

【0053】

図2の前記入出力インターフェース・モジュールは、本開示のシステムにおいて使用可能な入出力インターフェース・モジュールの一例でしかないと理解されたい。代替的な入出力インターフェースを用いることができる。代替的なインターフェースの一例は図3に提供されている。

【0054】

図3は、複数の高速インフィニバンド・ポート（集散的に301と称されている）を介したデータ関連要求の受け取り能力のある、代替的な入出力インターフェース・モジュール300を図示している。

【0055】

図3を参照するが、前記例示的な代替的な入出力モジュール300は、複数のインフィニバンド（InfiniBand）ポート301を通じて、データ関連要求及び応答を受け取り、伝送する能力のあるインフィニバンド・インターフェース・チップ302を含む。このインフィニバンド・インターフェース・チップは、Mellanoxから入手できるような、標準インフィニバンド・チップであってよい。

【0056】

一般に前記インフィニバンド・インターフェース・チップ302は、データ関連要求をインフィニバンド・ポート301を経由して受け取り、及び少なくともコマンド・ヘッダーを、及び外部WRITE 要求については、少なくともデータ・ペイロード識別する程度に受け取った要求を最小限復号する。インターフェース・チップ302は、次いで、図3の例ではPCI Express（PCIe）スイッチである、スイッチング素子、ルーティング素子、又は共有メモリ素子305の使用を介して、制御CPU306が利用可能なコマンド・ヘッダー情報を作成する。スイッチング/ルーティング/メモリ素子305を使用して、インターフェース・チップ302は、高速制御装置304（構成されたFPGAであることができる）が利用可能な受け取られたデータも作成する。前記高速制御装置304、多重シリアル・システム通信チャンネル310、312、314及び316を、クロスバー・モジュールに提供し、及び図2の高速制御装置200に関連して上記に記載されていると同様の方式で、クロスバー・モジュールへの、及びクロスバー・モジュールからの、情報及びデータを受け取り、及び提供するために、受け取られたデータ関連要求を操作する。

【0057】

上記で説明された、前記入出力インターフェース・モジュール200及び300は、単に例示的であり、及び他の種類の入出力インターフェース・モジュール、異なる物理層、異なる通信プロトコル、及び個別素子（構成されたFPGAsとは対照的に）を用いるものも含めて、本開示を逸脱することなく、使用できるであろうことが、理解されるであろう。

【0058】

一般に入出力モジュールの主な要件は、外部から提供されたデータ関連要求を受け取り、及び応答することができること、並びにそれらの要求（及びその要求に応答するために情報及びデータを受け取ること）を複数の高速システム通信チャンネルを通じて多重クロスバー・モジュールに伝送できることである。

10

20

30

40

50

【 0 0 5 9 】

クロスバー・モジュール

上述したように、本開示のシステム中のそれぞれのクロスバー・モジュールの全体的な一般的機能は：（ a ）前記入出力モジュールから D M A 要求を受け取り、及びそれらの D M A 要求（加えて、付随するデータ・ペイロード）を、クロスバー・モジュールに接続された、適切なメモリ・カードに提供すること、並びに（ b ）取り出されたデータをクロスバー・モジュールに接続されたメモリ・カードから受け取り、及び取り出されたデータ（及び潜在的ないくつかの情報）を、取り出しデータに関連付けられえた前記入出力モジュールに提供することである。

【 0 0 6 0 】

開示されたシステムのクロスバー・モジュールを実行するためには、さまざまなアプローチを用い得る。図 4 は、図 1 に関連して上記に記載されたクロスバー・モジュールを実行するための 1 つの例示的なアプローチを図示している。

【 0 0 6 1 】

図 4 を参照するが、例示的なクロスバー・モジュール 4 0 0 が図示されている。クロスバー・モジュールはディスクリット電気回路、特定用途向け集積回路（ A S I C s ）、又はこれらの 2 つの混合により実行できるが、図 4 の実施態様では、前記クロスバー・モジュール 4 0 0 は、 X i l i n x V i r t e x 4 F P G A のような構成された F P G A の使用により実行される。

【 0 0 6 2 】

この例では、構成された F P G A は、 5 つのシリアル・システム通信ポート 4 0 2、 4 0 4、 4 0 6、 4 0 8 及び 4 1 0 を提供するために構成されている。それぞれのシリアル通信ポートは、高速、双方向通信チャンネルを、それ自身と、入出力モジュールの前記高速通信ポートの 1 つとの間に提供するために構成され、及び接続されている。従って、この例では、クロスバー・モジュール 4 0 0 が、 5 つの全二重通信ポートを提供しているために、それは同時に及び独立してシステム 1 0 0 の前記入出力モジュールのそれぞれと通信できる。従って、シリアル通信が、前記入出力モジュールのそれぞれ及びクロスバー・モジュール 4 0 0 の間に存在するように、通信ポート 4 0 2 は、入出力インターフェース・モジュール 1 0 2 のシリアル通信ポートの 1 つに連結でき、ポート 4 0 4 は、入出力インターフェース・モジュール 1 0 4 のポートに連結でき、その他も同様である。

【 0 0 6 3 】

上述の 5 つのシリアル通信ポートを提供することに加えて、図 4 のクロスバー・モジュール 4 0 0 は、クロスバー・モジュール 4 0 0 及び複数のメモリ・カードの間の通信を可能にするために用いられる、多重並列通信バスも提供する。図 4 の例では、 5 つのメモリ・カード 4 1 2、 4 1 4、 4 1 6、 4 1 8 及び 4 2 0 があり、及びクロスバー・モジュール 4 0 0 は、並列バスを用いてメモリ・カードと通信する。図 4 の例では、クロスバー・モジュール 4 0 0、及び個別のメモリ・カードのそれぞれの間の通信は、この例では、それぞれのメモリ・カードに対して 1 6 ビット長のデータ・バス及び 5 ビット長の制御バスを含む、並列バスを用いて達成される。

【 0 0 6 4 】

一般的な操作において、クロスバー・モジュール 4 0 0 は、第一にシリアル・システム・リンクの 1 つを経由して、 D M A 要求を受け取り、 D M A 要求が向けられるべき特定のメモリ・カードを特定し、受け取られた D M A 要求をシリアル・フォーマットから並列フォーマットに変換し、及びその並列フォーマット化された D M A 要求を適切なメモリ・カードに提供する手段として機能する。加えて、クロスバー・モジュール 4 0 0 は、典型的にはメモリ・カードにより提供されたデータを受け取り、そのデータが、どの入出力モジュールを意図しているか決定し、及び受け取られたデータを、シリアル・システム通信リンクの 1 つを通じて、 D M A 応答として適切な入出力モジュールに転送する。

【 0 0 6 5 】

例示的なケースでは、クロスバー・モジュール 4 0 0 は、それぞれの入出力モジュール

10

20

30

40

50

と、シリアル・インターフェース・バスを用いて通信し、及びそれぞれのメモリ・カードとは、並列インターフェース・バスを用いて通信する。前記入出力インターフェース・バスの実行は、前記入出力モジュールに依存し、及びメモリインターフェース・バスは、メモリ・カードに依存することに注意されたい。理論的には、そのバスが要求されるデータ速度で実行するという前提で、どちらのバスも、シリアル又は並列インターフェースのいずれかにより実行される。インターフェース・バスは、用途依存性であり、及び特定のバスの選択に影響する設計決定は、当業者にとっては自明である。例として、コネクタ密度を最小化するために、前記メモリ・カード・インターフェースは、前記入出力モジュールに本質的に同様である高速シリアル通信として実行されることができ、クロスバーとメモリ間のインターフェースを単純化できるであろう。

10

【0066】

クロスバーポートを入出力モジュールに接続するバスのそれぞれは、ポイントツーポイント (point-to-point) であるために、競合制御 (arbitration) は最小であり、及びオーバーランを防止することが必要な場合には、クロスバーが前記入出力バスをホールドオフする (hold off) ることを可能にするために設計されたメカニズムにより規制される。

【0067】

これらのタスクを遂行するためのFPGAの構成は、本開示の利益を受ける当業者にとり自明である。

【0068】

20

メモリ・カード

開示されたシステムで用い得るメモリ・カードは多くの形態をとり得る。さまざまな種類のランダムアクセスメモリ (DDR RAM、DDR2 RAM、DDR3 RAMなどの) 及びさまざまな種類のフラッシュ・メモリ (MLCフラッシュ・メモリ及びSLCフラッシュ・メモリなどを含む) を含むさまざまなメモリを使うことができる。メモリ・カードの構造も、部分的には使用されるメモリに応じて変更できる。

【0069】

図5は、フラッシュ・メモリを用い、及び本開示のシステムにおいて用い得るメモリ・カード500の例示的な実施態様を図示している。図5の例では、このフラッシュ・メモリは、SLCフラッシュ・メモリであるが、MLCフラッシュ・メモリが用いられる実施態様も予測される。

30

【0070】

図5を参照するが、このフラッシュ・メモリ・カード500は、システム制御装置をクロスバー・モジュール (図5には図示されていない) の1つと連結させている、並列バス504を通じて情報を受け取るシステム制御装置502を含む。上述したように、並列バス504は、16ビット長のデータ・バス及び5ビット長の制御バスを含む。クロスバー・モジュールから提供されるDMA要求は、システム制御装置502により受け取られて、及び処理される。システム制御装置502は、メモリ・カード及びクロスバー・モジュールの間の通信プロトコルを取り扱い、及びバス・エラー管理のためのエラー修正などの機能を実行できる。加えて、システム制御装置502は、メモリ・カード上のどの特定のメモリ素子が、TARGET ADDRESSと関連付けられているかを決定するために、受け取られたTARGET ADDRESSアドレスを部分的に処理することもできる。

40

【0071】

例示的なシステムでは、システム制御装置502は、多数の個別のフラッシュ制御装置506、508、510及び512と、直接接続を介して通信する。システム制御装置502は、個別のフラッシュ制御装置との通信もできるCPU503とも通信する。この通信は、図5に示されるように、フラッシュ制御装置への直接接続を介するか、又はCPU503がフラッシュ制御装置にアクセスするために、システム制御装置502と通信する「通過 (pass-through)」接続を介することができる。これは、典型的には

50

バスのファンアウト (f a n - o u t) を減少させるために行われ、及び当業者には明白な設計検討である。

【 0 0 7 2 】

図示された例においては、システム制御装置 5 0 2、及びフラッシュ制御装置 5 0 6、5 0 8、5 1 0 及び 5 1 2 との間の通信は、独立した 1 6 ビット長の並列バスを用いて達成され、そのような独立した並列バスは、システム制御装置 5 0 2 及び個別のフラッシュ制御装置 5 0 6、5 0 8、5 1 0、及び 5 1 2 のそれぞれの間に連結されている。同様の独立した 1 6 ビット長の並列バスは、C P U 5 0 3 がそれぞれのフラッシュ制御装置 5 0 6、5 0 8、5 1 0 及び 5 1 2 と直接通信する実施態様において使用できる。

【 0 0 7 3 】

個別のフラッシュ制御装置のそれぞれは、物理フラッシュ・メモリ空間 5 1 3、5 1 4、5 1 5、及び 5 1 6、制御装置メモリ 5 1 7、5 1 8、5 1 9、及び 5 2 0 (例えば、D D R R A Mメモリであってよい) とそれぞれ連結されている。図示された例においては、それぞれの物理フラッシュ・メモリ空間 (5 1 3 ~ 5 1 6) は、1 0 個の独立したフラッシュ・メモリ・チップから形成されている。操作においては、フラッシュ制御装置 5 0 2 は、提供されたデータを特定の物理フラッシュ・メモリに保存し、及び要求されたデータを取り出し、及びそれをシステム制御装置 5 0 2 に提供するために、D M A 要求を処理する。多数のさまざまな種類のフラッシュ制御装置を用い得る。好適な制御装置及びその操作の一つは、2 0 0 9 年 9 月 5 日に出願された同時係属の米国特許出願第 1 2 / 5 5 4 , 8 8 8 号、1 2 / 5 5 4 , 8 9 1 号、及び 1 2 / 5 5 4 , 8 9 2 号により詳細が記載されており、参照により本願に組み込まれる。

【 0 0 7 4 】

特に、本開示のシステム 1 0 0 は、前記入出力インターフェース・モジュールにより受け取られた T A R G E T A D D R E S S が、メモリ・カードに提供されるアドレスであるように作動するため、本願のシステムは、フラッシュ・メモリ (究極的に受け取られた T A R G E T A D D R E S S の物理フラッシュ・アドレスロケーションへの移行を必要とする) 及び R A Mメモリ (又は論理から物理アドレス変換を必要としない他のメモリ) の両方と容易に使用できる。更に、T A R G E T A D D R E S S から物理フラッシュ・アドレスのいかなる変換も、D M A 要求がメモリ・カードに提供された後に行われるために、本願のシステムは、フラッシュ・メモリ・カード及び R A Mメモリ・カードの両方の使用を可能にする。このことは、受け取られた T A R G E T A D D R E S S を特定の物理フラッシュ・アドレスにマッピングするフラッシュ制御装置の操作が、クロスバー・モジュールをメモリ・カードに連結させるインターフェース・バスにとって、大部分トランスペアレント (t r a n s p a r e n t) であるためである。

【 0 0 7 5 】

図 6 は、本開示のシステムにおいて用い得る R A Mメモリを用いるメモリ・カード 6 0 0 の代替的な実施態様を図示している。この実施態様においてはクロスバー・モジュールによりメモリ・カードに提供されるアドレスは、アクセスする R A Mメモリに対して物理アドレスとして使用される。図を参照するが、メモリ・カード 6 0 0 は単独の多重層印刷回路板に配置され添付されることのできる多数の素子を含む。

【 0 0 7 6 】

一般に、メモリ・カード 6 0 0 は、システム制御装置 6 0 2 を含み、このシステム制御装置 6 0 2 は、図 5 からの前記制御装置 5 0 2 のように、D M A 要求をクロスバー・モジュールから受け取り、及びその要求を、そのシステム (エラー修正を含むか、又はそれと結合されることのできる) のために採用されたプロトコルを用いて処理する。システム制御装置は、次いで要求及びデータ・ペイロードがあれば、それを受け取られたデータの T A R G E T A D D R E S S に対応する、適切な物理アドレスへの急速 D M A 移転を手配できる (W R I T E 要求に対して) メモリ制御装置 6 0 4 に渡すか、又は D M A 移転を用いて、データを R A Mメモリ 6 0 6 (D D R、D D R 2、D D R 3 又は任意の他の高速 R A Mメモリであってよい) から取り出し、及びそのデータをシステム制御装置 6 0 2 に提

供する（READ要求に対して）。図6に示される矢印の方向性は、WRITE操作の例を描写することのみを意味していることが理解されるであろう。実行の目的で、制御装置602及び制御装置604の両方が1つのFPGA装置に結合されることも理解されるであろう。

【0077】

代替的なシステムの実施態様：

上述のシステム100では、クロスバー・モジュールは、シリアル形式で伝送されたデータ関連要求を受け取り、及びそれらの要求を、メモリ・カードに提供される並列形式の要求に変換するために用いられる。クロスバー・モジュールが除去され、及び前期データ関連要求が、シリアル形式で直接前記入出力モジュールからメモリ・カードに伝送される代替的な実施態様が想起される。

10

【0078】

図7は、入出力モジュール又は装置が、クロスバー・モジュールを使わずに高速シリアル通信リンクを用いて直接的にメモリ・ボードと通信できる代替的な記憶システム700を図示している。

【0079】

図7を参照するが、代替的な記憶システム700は、複数の入出力モジュール701、702、703、704、705、706、707、708、709、710、711、及び712を含み、それらのそれぞれは、1つ以上の外部ホストと、1つ以上の通信チャンネルを経由して連結されている。図示された例においては、それぞれの入出力モジュールは、前記入出力モジュールのそれぞれが、複数のメモリ・ボード740、742、744及び746のそれぞれ1つと通信することを可能にする、複数の高速双方向性全二重シリアル通信チャンネルを提供する。図7の例では、それぞれの入出力モジュール701～712は、8個のシリアル通信ポートを提供し、及びそれぞれの入出力モジュール及びメモリ・ボードの接続回線は、2つの個別の全二重通信チャンネルを反映することが意図されている。

20

【0080】

前記入出力モジュール701～712は、メモリ・ボード740、742、744及び746に、極度に高帯域通信が可能になる方法で連結されている。例えば、第一の入出力モジュール701は、8個のシリアル・データ・ポート（メモリ・ボードのそれぞれに対して2個）を提供する。従って、前記入出力モジュール701が提供できる最大データ速度は、シリアル通信チャンネルの最大帯域幅の8倍である。

30

【0081】

操作においては、前記入出力モジュール701～712は、図2及び3に関連して上記で記載された、前記入出力モジュールと同様の操作を行う。しかしながら、図7の入出力モジュールは、より多数のシリアル通信チャンネルを提供するため、それらは、同時で、及び独立した、より多くのデータ関連要求又は応答通信を可能にする。このことは、一般的に記憶システム700に、並びに特に前記入出力モジュール701～712とメモリ・ボード740、742、744及び746の間に、より高い帯域幅をもたらす。例えば、前記入出力モジュール701は、8個のシリアル通信ポート（メモリ・ボードのそれぞれに対して2個）を提供するため、同時（又はほぼ同時）に、及び独立して2つのWRITE要求を、メモリ・ボード740、742、744及び746のそれぞれに対し発行し、合計8個の同時（又はほぼ同時）に処理されたWRITE要求を発行できるであろう。元のシステム100において言及されたように、クロスバーなしでは、同じ競合制御（arbitration）及びホールドオフメカニズムは、いくつかの実施態様ではそれぞれのメモリ・ボードに配置される。

40

【0082】

図7の例では、それぞれのシリアル・システム通信リンクは、前記入出力モジュール又は通信装置及びメモリ・ボードへのデータ及び制御情報の両方を、又はそれらからのデータ及び制御情報の両方を、渡させるために用いられる。とりわけ、それぞれのメモリ・ボ

50

ードは、デジタル・データを受け取り、及びWRITE要求に応答し（メモリ・ボード上の物理的メモリ空間の中の特定のメモリロケーションの中にデジタル・データを記憶させるために）、及びREAD要求に応答する（ボード上の物理メモリの特定のメモリロケーションの中に記憶されたデジタル・データを取り出し及び提供するために）。それぞれの高速シリアル通信リンクは、非常に高速（この実施例では625MB/sec）で、シリアル・データを受け取るか、又は伝送するため、図7のシステムは、非常に高いデータ速度でデータの記憶、又は取り出しができる。

【0083】

図8は、図7のシステム700で使用するための、例示的なメモリ・ボード800の構造に関する詳細を図示している。

10

【0084】

図8に転じて、メモリ・ボード800は、記憶制御ユニット802及び複数の個別のメモリ・チップから形成された物理的メモリ空間804を含む。このメモリ・ボードは、メモリ・ボードが複数のシリアル通信リンクを受け取ることを可能にするために、更に適切なインターフェース回路806を含む。図8では、インターフェース回路806は、メモリ・ボード800が、24個のシリアル通信リンクからの入力を受け取ることを可能にする。図8の例では、さまざまな要素全てが同一の印刷回路板800の上に配置されている。

【0085】

記憶制御ユニット802は、Xilinxから入手可能な333MHzで稼動するVirtex-6FPGA(XC6VLX240T-2FFG1156CES)などの、フィールド・プログラマブル・ゲート・アレイ(FPGA)の使用により形成できる。しかしながら、記憶制御ユニット802は、代替的に他の種類のFPGA装置、ディスクリート回路、プログラムされたマイクロプロセッサ、又はそれらの任意の、又は全ての組み合わせによっても実行できるであろうことが理解されるであろう。

20

【0086】

図示された例においては、物理的メモリ空間804は、Micron Technologies, Incから入手可能な333MHzで稼動するDDR3-800(MT4J128M8BY-25E)のような複数のダブルデータレートダイナミックRAM・チップ、DDR、を用いて構築される。しかしながら、代替的なDDRデバイス、フラッシュ及び他の種類のメモリなどの他の形態のメモリが、本開示の教示を逸脱することなく使用できることが理解されるであろう。

30

【0087】

この実施例では、記憶制御ユニット802は、物理的メモリ空間804への288本の外部並列データ回線（図9によりよく示される）を駆動する。メモリ空間は、333MHzでクロックされるDDRメモリを使用するため、図8の前記高速メモリ・ボードは、それぞれのクロック・サイクルに対して物理的メモリ空間内に、最大576ビットのデジタル情報（512ビットのデータ及び64ビットのECC）を記憶する能力があり、従って全体でデータ及びECCについては約24000MB/sec、データ単独では20.83GB/secの記憶速度を提供する。

40

【0088】

図9は、例示的な高速メモリ・ボードの構造の一部分をより詳細に図示している。特に、図9は、シリアル通信リンクからのデジタル・データを受け取り、物理的メモリ空間804内への記憶のために、そのようなデータを操作するために使うことができる、図8の記憶制御ユニット802内の構造を図示している。

【0089】

図9を参照するが、前記高速シリアル通信リンクの1つ、シリアル・リンク900が、記憶制御ユニット802と連結しているところが図示されている。前記高速シリアル通信リンク900は、記憶制御回路802内に配置されるシリアライザ/デシリアライザ及びパッキング・モジュール912に提供される。このシリアライザ/デシリアライザ及びパ

50

ッキング・モジュール 912 は、図 10A に更に説明されており、シリアル・リンクが記憶制御ユニット 802 内に配置されるマルチ・ギガビット・トランシーバー (MGT) 1010 へ入力されるところを描写している。MGT 1010 は、シリアル通信リンクからシリアル・データを受け取り、及び定期的に、受け取られたシリアル・データを並列データに変換する。図 10A の実施例に進んで、MGT 1010 はシリアル・データを受け取り、及び 156.25 MHz で並列 32 ビット長のデータを提供する。

【0090】

図 10A の例では、MGT 1010 からの 32 ビット長の並列データは、256 ビット長の多重ワード・ディープ先入れ先出し (FIFO) 記憶バッファ 1012 に、MGT 1010 から受け取られた 32 ビット長のデータが 256 ビット長のデータ・ワードにパックされる方式で提供される。このプロセスは図 10B ~ 10D に一般的に図示されている。

10

【0091】

図 10B を参照するが、256 ビット長の FIFO 記憶装置 1012 が、256 ビット長のワード A1 ~ A8 とパックされるところが模式的に図示されている。256 ビット長のワードを形成するために、図示されたシステムで MGT 1010 からの 32 ビット長のデータをパックしている方法は以下のものである。第一に、この例では A1 である、最初の 32 ビット長のワードが MGT 1010 から受け取られる。その最初の 32 ビット長のワード A1 は、図 10B に反映されているように最初のデータメモリロケーションに記憶される。MGT 1010 から、この例では A2 である、次の 32 ビット長のワードが受け取られて、第二のメモリロケーションに記憶され、第三のワード A3 が 3 番目のメモリロケーションに記憶され、図 10B に示されているように、256 ビット長のワード A1 ~ A8 を形成するために 8 個のワード A1 ~ A8 が受け取られ、FIFO 1012 に記憶されるまで、以下同様となる。

20

【0092】

256 ビット長のワード A1 ~ A8 が、上述のプロセスで形成され、及びこの例では B1 である、別の 32 ビット長のワードが、MGT 1010 から受け取られると、最初のワード A1 ~ A8 は、FIFO 1012 内の別のメモリロケーションにシフト「下げ」され、及び新しく受け取られた 32 ビット長のワードは、図 10C に反映されているように、最初のメモリロケーションに記憶される。図 10C に反映されているように、このプロセスは、第二の完全な 256 ビット長のワード B1 ~ B8 が構築されるまで繰り返される。MGT 1010 から 32 ビットのワードを受け取り、受け取られた 32 ビットのワード 256 ビットのワードにパックし、及び 256 ビット長のワードを FIFO 1012 を通じてシフトさせるプロセスは、FIFO 1012 内の 256 ビット長のメモリロケーションが、完全に占められるまで継続する。より完全に占められている FIFO 1012 の例は、図 10D に図示されている。

30

【0093】

スタッキング FIFO 1012 の「深さ (depth)」は実行に応じて変化するが、図 9 の例、及び 10A ~ 10D においては、FIFO 1012 は、少なくとも 32 個の 256 ビット長ワードの深さである。このことは、この例では、32 個の 256 ビット長のワードのグループが、物理的メモリ空間 804 への、又は物理的メモリ空間 804 からの、データ転送において使用される基本単位であるためである。そのようなものとして、本願では、32 個の 256 ビット長のワードのそれぞれのグループを、基本メモリ周期 (Basic Memory Cycle: BMC) 単位と称する。この例では、それぞれの BMC 単位は 8 個の、192 データビット (32 x 256) 又は 1 KB のデータを含む。

40

【0094】

図 9 に戻って、パッキング FIFO 912 からのデータは、多重チャンネル FIFO バッファ 914 に提供されるが、図 9 では、それらのうちの 1 つのチャンネルだけが図示されている。図示された例においては、多重チャンネル FIFO バッファ 914 は 256 ビット長であり、データを 256 ビット / 秒の速度でデータを受け取ることができ、及

50

び27.7MHzの速度でクロックされている。データは、シリアライザ/デシリアライザ及びパッキング・モジュール912から、多重チャンネルFIFOバッファ914に、データの完全なBMC単位が、32クロック周期を通じて転送される、バースト転送を用いて提供される。

【0095】

データの完全なBMC単位が、シリアライザ/デシリアライザ及びパッキング・モジュール912から、多重チャンネルFIFOバッファ914に転送されると、多重チャンネルFIFOバッファ914は、データの完全なBMC単位を、32個の連続的な256ビットの「ワード」の形態で含むようになる。図示された例においては、多重チャンネルFIFOバッファ914の深さは、それがデータの多重BMC単位を記憶することができるようなものである。このことは、多重チャンネルFIFOバッファ914の内容が3つのデータのBMC単位を含むものとして描写されている(1140a、1140b及び1140cのブロックにより表されている)図11に一般的に反映されている。

【0096】

図9に戻り参照すると、多重チャンネルFIFOバッファ914に記憶されたデータは、333MHzで稼動する256ビット長の並列データ・バスを經由してエラー修正及びデータ保護回路916に提供される。エラー修正及びデータ保護回路916は、受け取られたデータを1つ以上のエラー修正コード(ECC)・ビットを導入するために処理し、及びそのデータの破損に対する強化された保護のためにデータを処理する。ECC処理の例は、当業者にとって周知であり、及びいかなる適切なECC処理機構をも使用できる。加えて、前記エラー修正及びデータ保護回路は、他のデータ保護技法又はチップキル(Chipkill)などの変換、及び強化されたデータ保護のために通常用いられる技法も実行できる。当業者が理解し得るチップキルは、IBMにより開発された進化したECCの形態であり、コンピュータ・メモリ・システムを、いかなる単独のメモリ・チップの不具合からと同時に、単独のメモリ・チップのいかなる部分のマルチビット・エラーからも保護する。チップキルに関する追加的情報は、以下の文献に見出すことができ、これは参照により本願に組み込まれる: Timothy J. Dell, A White Paper on the Benefits of Chipkill - Correct ECC for PC Server Main Memory, (1997), IBM Microelectronics Division。

【0097】

エラー修正及びデータ保護回路916の操作の結果として、回路に提供されるそれぞれの256ビット「ワード」にビットが加えられ、及び結果として生成する回路の出力は、この実施例ではそれぞれの256ビットの入力に対して288ビットである。この288ビットの出力はECC及びデータ保護回路916に提供される入力データは、回路916により加えられる保護及びECCビットに対応する。

【0098】

このデータ保護回路916からの288ビット長の出力は、288ビット長の並列バスを經由して、2つの出力FIFOバッファ918及び920に提供される。この実施例では、この並列バスは、333MHzでクロックされている。データ保護回路からのデータは、ピンポン方式で出力FIFOバッファ918及び920に提供される回路からの第一の288ビット「ワード」が出力FIFOバッファ(例えば、出力FIFOバッファ918)の1つに提供され、一方次の288ビット「ワード」は、次のクロック周期で出力FIFOバッファ(例えば、FIFOバッファ920)の他の1つに提供される。この方式で、完全なBMC単位に対応するデータは、データの半分がそれぞれの出力FIFOバッファに存在するように、2つの出力FIFOバッファ918及び920に配置される。この実施例では、それぞれの出力FIFOバッファ918及び920の深さは、それぞれのバッファが、多重BMC単位に対応するデータを保存できるようなものである。

【0099】

回路 916 から、出力 F I F O バッファ 918 及び 920 へのデータ移動のこのプロセスは、図 12 及び 12A に一般的に図示されており、ここでは出力 F I F O バッファ 918 及び 920 の例示的な内容は、B M C 単位の移転後が図示されている。この図に反映されているように、例示的な B M C 単位のデータは、両方のバッファに含まれ、及び完全な B M C 単位を作り上げるデータは、2つの出力 F I F O バッファ 918 及び 920 の間に「インターリーブ」される。

【0100】

図 9 に戻って参照するが、2つの出力 F I F O バッファ 918 及び 920 からのデータは高速の 576 ビット長の出力メモリ・バッファ 922 に提供される。図 9 の例では、それぞれの F I F O 出力バッファ 918 及び 920 は、288 ビットのデータ及びエラー保護情報を、それぞれのクロック周期で、前記高速出力メモリ・バッファ 922 に提供し、及び前記高速メモリ・バッファ 922 への入力、この実施例では、333 MHz でクロックされている。このスピード及びデータ転送速度では、データ (32 ワード) の完全な B M C 単位は、16 クロック周期に変換される。ここで、16 ワードが F I F O 出力バッファ 918 及び 920 のそれぞれに記憶され、及び従って B M C 全体の移転には 16 周期が必要である。

【0101】

出力 F I F O バッファ 918 及び 920 と、前記高速出力メモリ・バッファ 922 の間の連結は、出力 F I F O バッファ 918 及び 920 に記憶されていた、かつて、インターリーブされた (interleaved) データが、単一の適正に順序付けられた 576 ビットの“スーパー・ワード”を形成するために組み合わせられるようなものであり、このスーパー・ワードの中では、前記高速シリアル通信リンク 900 を通じてデータがもともと受け取られたときの順序を反映している。

【0102】

図 9 に戻って参照するが、前記高速出力バッファからのデータは、333 MHz D D R でクロックされている 288 ビット長の並列バスを通じて、物理メモリ 804 (図示されていない) に提供される。この D D R クロックは、データを両方のクロック・エッジで移転し、出力メモリ・バッファ 922 からのデータは、効果的に 667 MHz の速度で移転される。結果として図 9 の回路のメモリへの効果的なデータ転送速度は 20.83 GB/sec である。

【0103】

説明の目的のために、図 9 の実施例は、この実施例では (図 7 を参照のこと) シリアル入力 701 である、単独の高速シリアル入力のみとの関連において本システムの操作を図示及び記述している。完全な例示的なシステムにおいては、最大 24 個の個別の高速シリアル通信リンクが提供され得る。そのような多重シリアル通信リンクを有するシステムからのデータが、本開示のシステムにより処理される方式は、図 13A ~ 13B に反映されている。

【0104】

最初に図 13A を参照するが、図 9 との関係で上記で説明されたものに酷似したシステムが図示されている。しかしながら、図 13A の例では、例示的な高速入力についての入力 M G T 1010 及びパッキング F I F O バッファ 1012 は、入力処理ブロック 1370 を形成するために組み合わせられている。加えて、それぞれが対応する入力処理ブロックを有する、11 個の他のシリアル入力 1371 ~ 1381 が示されている。これらの入力処理ブロックのそれぞれは、データの完全な B M C 単位を、バーストモード方式で多重チャンネル F I F O バッファ 914 に提供することを可能にするために、図 9 及び 10A ~ 10D との関係で上記に記載された回路のように作動する。

【0105】

図 13A のシステムの設計のために、ピーク作動状態の間には、W R I T E 操作からのデータは、ほとんど常に多重チャンネル F I F O バッファ 914 に提供され、及び多重チャンネル F I F O バッファ 914 からのデータは、ほとんど常にエラー修正及びデー

10

20

30

40

50

タ保護回路 9 1 6 に提供されるように、このシステムは、データを非常な高速で受け取ることができる。

【 0 1 0 6 】

しかしながら、図 1 3 A のシステムは、好適には、記載されたシステムにより提供される前記高速シリアル通信リンクの半分からのデータを処理する。図 1 3 B に示されるように、この例示的なシステムは、残りの 1 2 個のシリアル通信リンクからのデータを処理し、及びそのデータを第二のエラー修正及びデータ保護回路 9 1 6 ' に提供する、別の多重チャンネル F I F O バッファ 9 1 4 ' を含む、上記に記載されたものと同様の、別の一式の回路を含む。

【 0 1 0 7 】

図 1 3 B に反映されているように、第二のエラー修正及びデータ保護回路 9 1 6 ' は、“ピンポン”方式で、その出力を出力 F I F O バッファ 9 1 8 及び 9 2 0 に提供する。しかしながら、この例示的なシステムでは、第二のエラー修正及びデータ保護回路 9 1 6 ' からのデータを移転する手順は、第一のデータ保護回路 9 1 6 がデータを出力 F I F O バッファの 1 つ（例えば、出力 F I F O バッファ 9 1 8 ）に出力している間、第二のエラー修正及びデータ保護回路 9 1 6 ' は、データを他の出力 F I F O バッファ（例えば、出力 F I F O バッファ 9 2 0 ）に移転するようなやり方で、第一のエラー修正及びデータ保護回路 9 1 6 とは、「位相ずれ（out - of - phase）」で作動する。次のクロック周期の間、この移転は「スイッチ」する。この方式により、それぞれのクロック周期の間に、データは常に両方の出力 F I F O バッファ 9 1 8 及び 9 2 0 に送達される。

【 0 1 0 8 】

異なる W R I T E 操作からのデータを不適切に「混合する」ことを防止するために、第一及び第二のエラー修正及びデータ保護回路 9 1 6 及び 9 1 6 ' により提供されたデータは、単一のシリアル通信リンクを通じて受け取られたデータに見えるように再結合できる方式で、好適には、出力 F I F O バッファ 9 1 8 及び 9 2 0 へ移転される。これを達成するための 1 つのアプローチは、図 1 4 A 及び 1 4 B に反映されている。

【 0 1 0 9 】

図 1 4 A 及び 1 4 B を参照するが、第一のデータ転送操作の間、第一のエラー修正及びデータ保護回路 9 1 6 （例えば、ワード 0、a 0）から移転される準備ができていない B M C 単位の第一の偶数のワードは、第一の回路 9 1 6 により、第一の回路 9 1 6 のために予約されている出力 F I F O バッファ 9 1 8 の一部分に移転される。同時に、第二のエラー修正及びデータ保護回路 9 1 6 ' から移転される準備ができていない B M C 単位の第一の偶数ワード（ワード 0、b 0）は、第二の回路 9 1 6 ' からのデータの記憶のために予約されている出力 F I F O バッファ 9 2 0 の一部分に移転される。次のクロック周期の間、第一の回路 9 1 6 中の B M C からの第一の奇数ワード（例えば、ワード 1、a 1）は、出力 F I F O バッファ 9 2 0 の予約された空間に記憶され、及び F I F O 9 1 6 ' 中の B M C からの第一の偶数ワード（例えば、ワード 1、b 1）は、出力 F I F O バッファ 9 1 8 内の予約された空間に記憶される。この方法により、最大帯域幅が保持されるように、データは常に出力 F I F O バッファ 9 1 8 及び 9 2 0 の両方に移転される。

【 0 1 1 0 】

代替的な方法では、第二のエラー修正及びデータ保護回路 9 1 6 ' からのデータの書き込みは、全ての偶数のワードが出力 F I F O バッファ 9 1 8 に記憶され、及び全ての奇数のワードが出力 F I F O バッファ 9 2 0 に記憶できるように、1 周期遅らせることができる（図 1 4 C 及び 1 4 D を参照のこと）。

【 0 1 1 1 】

上記で説明した図 9 及び他の図に反映されているように、データの受け取り及び処理のための電気回路のために提供することに加え、それが、前記高速シリアル通信チャンネルの 1 つを通じて、要求されたデータを次いで適切なホストに提供できる入出力インターフェース・モジュールに移転されるように、それぞれのメモリ・カードは、高速のデータ速

10

20

30

40

50

度でのRAMメモリからのデータの取り出し、及び受け取られたデータを解凍するための同様の構造（本質的に逆順に操作する）を含むことができる。

【0112】

図示された例においては、メモリ制御装置802を物理RAMメモリ804に連結するデータ及びアドレス回線は、任意の所定の時間に、READアクセスのみ、又はWRITEアクセスのみが遂行できるようになっている。従って、最適の成績を提供するためには、図15に一般的に示されるように、前記制御装置802は、何らかの形態の競合制御を行わなければならない。一実施態様では、この競合制御は、前記制御装置802内に、それぞれのデータ関連要求に対するコマンド・ヘッダー情報を受け取り、及びその情報を、その要求を提供した前記入出力インターフェース・モジュールに関連付けられたバッファ 10
ー中に記憶する競合制御モジュール1504を組み込むことで達成される。好適な実施態様では、この競合制御モジュール1504は、前記制御装置802が、入出力インターフェース・モジュール単位で、入出力インターフェース・モジュール上のバッファされた記憶装置の中で、それぞれの入出力インターフェース・モジュールから受け取られたREAD及びWRITE要求又はコマンドのリストを保持するように、READ操作のための個別のバッファ1500及び1502、並びにWRITE操作のための個別のバッファ1501及び1503を保持する。一般にこのリストは、タイム・スタンプ、又は時間 20
順の原理で保持されることができ、及びリストへのそれぞれのエントリーは、入出力インターフェース・モジュールから受け取られたそれぞれのデータ関連要求に関連付けられたコマンド情報、すなわち、TARGET ADDRESS、ワード・カウント（又は移転されるべきデータ量の他の表示）及び移転方向の表示（例えば、それがREAD又はWRITE操作のいずれであるか）を、含むことができる。

【0113】

性能を最適化するために、READ及びWRITE要求は以下のように処理されることができる：

一般に前記競合制御・モジュール1504は、READ要求を優先し、及び以下で説明する条件が満たされない限り、さまざまな入出力インターフェース・モジュールのために、READ要求バッファ中のREAD要求を先着順の原理で処理する。

【0114】

もし、競合制御・モジュール1504及び前記制御装置802が、未処理のREAD要求がないように、全てのREAD要求を処理すると、競合制御・モジュール1504及び前記制御装置802は、次いでWRITE要求バッファ中の未処理のWRITE要求を先着順の原理で処理する。 30

【0115】

多すぎる未処理のWRITE要求を持つことを避けるため、並びにアドレス待ち時間及び一貫性の問題を助力するために、競合制御・モジュール1504及び前記制御装置802は、未処理のWRITE要求の数が特定の閾値数を越えていると決定された場合には、WRITE要求を処理する。競合制御・モジュール1504は、この閾値を越えている決定を以下のことを考慮して行う：（i）前記入出力インターフェース・モジュールの合計からの全ての未処理のWRITE要求の合計数；（ii）所定の入出力インターフェース・モジュールからの未処理WRITE要求の合計数；又は（iii）上記のいくつかの組み合わせ。例えば、もし未処理WRITE要求の合計数、及び所定の入出力インターフェース・モジュールのためのWRITE要求の数が考慮される場合、競合制御・モジュール1504は、以下のどちらかである場合に、前記制御装置802にWRITE要求を処理させることができる：（i）未処理のWRITE要求の合計数が第一の閾値を越える（この場合には、未処理のWRITE要求は先着順の原理で処理される）又は（ii）所定の入出力インターフェース・モジュールのための未処理のWRITE要求が第二の閾値を越える場合（第一の閾値よりも低くてよい）、この場合には、競合制御・モジュール1504は、前記制御装置802に、先着順の原理で、閾値を越えるバッファに関連付けられる特定の入出力インターフェース・モジュールにWRITE要求を処理させる。 40
50

【 0 1 1 6 】

更に、競合制御部は、妨害又は絞り (t h r o t t l i n g) なしで、前記入出力インターフェース・モジュールが、データ及びデータ関連要求を、送信し続けることができるように、それがデータ受け取りに関連付けられた出力 F I F O バッファの 1 つの中に記憶されたデータが、出力 F I F O バッファの記憶能力の 1 / 2 を超える量などの特定の限界を越えたと判断した場合は、W R I T E 要求を処理でき、この場合にはその W R I T E 要求に関連付けられた出力 F I F O バッファが、処理される。

【 0 1 1 7 】

上述のことに加えて、競合制御・モジュール 1 5 0 4 は、データの一貫性が維持されるようにどのように R E A D 及び W R I T E 要求が処理されるかを制御することもできる。例えば、未処理の R E A D 要求が特定のアドレスに向けられること、及び時間的に早い未処理の W R I T E 要求があることが決定される場合、競合制御・モジュール 1 5 0 4 は、適切なデータを、W R I T E 要求への応答を確実に返すために、前記制御装置 8 0 2 に未処理の W R I T E 要求を処理させることができる。

【 0 1 1 8 】

メモリ・チップ・アクセス

メモリ制御装置 8 0 2 が物理 R A M メモリへ、又は物理 R A M メモリからデータを移転するために使用できるさまざまなアプローチ及び方法がある。一実施態様では、前記制御装置 8 0 2 は、競合制御・モジュール 1 5 0 4 からコマンド情報をデータ転送のタイミングを制御する開始インジケータとともに受け取る D D R 3 メモリ制御モジュールとして機能するか、又は D D R 3 メモリ制御モジュールを含むことができる。一般に競合制御・モジュール 1 5 0 4 から提供されるこのコマンド情報は、以下のものを含むことができる：競合制御の目的で用いられるバッファに記憶された同じ情報、すなわち、ターゲット・アドレス、及びより具体的には、移転のための開始アドレス；移転されるべきデータの量の識別子、これはワード・カウントであってよい；及び移転方向。メモリ制御モジュールとして作用する前記制御装置 8 0 2 は、次いで、T A R G E T A D D R E S S (開始アドレス) により指示されたときに、適切なメモリ・チップを活性化する。

【 0 1 1 9 】

移転の方向がメモリへの W R I T E を示す場合、前記制御装置 8 0 2 は、出力メモリ・バッファ 9 2 2 からのデータを転移する一方、メモリ・チップへのコマンド (例えば、タイミング及びアドレスコマンド) の発行を調整する。

【 0 1 2 0 】

移転の方向が R A M メモリへの R E A D を示す場合、前記制御装置 8 0 2 は、メモリ・チップからのデータを、データが向けられる前記入出力インターフェース・モジュールに関連付けられた出力 F I F O バッファに転移する一方、適切なアドレス及びタイミング・コマンドのメモリ・チップへの発行を調整する。

【 0 1 2 1 】

1 つの実施例では、前記制御装置 8 0 2 は、以前のコマンドに関連付けられたデータ転送が完了する前に、次のコマンド情報のセットを受け取り、従ってシステムが最大の帯域で作動することを可能にする。

【 0 1 2 2 】

物理 R A M メモリへの、又は物理 R A M からの、データ移転を制御する時の前記制御装置 8 0 2 の概略の作動は、図 1 6 及び 1 7 に一般的に反映されている。

【 0 1 2 3 】

最初に図 1 6 を参照するが、前記物理 R A M メモリの 1 つの特定の配置が示されている。図 1 6 の実施態様では、制御装置 8 0 2 が図示され、及び前記制御装置 8 0 2 及び物理 R A M メモリとの間の接続が示されている。この例では、前記制御装置 8 0 2 に連結された物理 R A M メモリは、4 つのセクション (1 6 0 2 、 1 6 0 4 、 1 6 0 6 、 及び 1 6 0 8) に分割された 7 2 個のメモリ・チップの形態をとり、それぞれのセクションが 1 8 個のメモリ・チップを有する。この例では、それぞれのメモリ・チップが、4 つのバンク (A

10

20

30

40

50

、B、C、D)を有する。メモリ・チップは、それらが以下のものを共有するように連結される：(a)アドレス及び制御回線(この実施例では22)、及び(b)データ回線(この実施例では72)。図16に示される連結は物理メモリ・チップを制御装置802に接続するただ1つの方法である。4つのセクションの全ては、個別の制御バスを持つにもかかわらず、調和して作動し、及び同時に同じバンクに対してコマンドが発行される。

【0124】

メモリ制御装置802の操作に関する追加的な詳細は、図17に見出すことができる。図17は、2つの基本メモリ周期(ここでは第二の基本メモリ周期は未完成として示されている)を通じた前記制御装置802の操作を図示している。一般にそれぞれのメモリ周期は32個の288ビットのワードのDDRメモリへの(又は、DDRメモリからの)移転を含む。この特定の実施例では、この移転は、16の実際のクロック周期を通じて起きる。333MHzでは、これは、それぞれのメモリ・カードに対して、48nsごとに1Kバイトのデータ、又は20.83Gバイト/秒(データ単独)の移転速度を提供する。図7に反映されているシステムが、多数のメモリ・カードを持つことができることを考慮すると、システムの合計の帯域幅は20.83Gバイト/秒よりも顕著に高くなることができ、及び5個以上のメモリ・カードが用いられる場合には、100Gバイト/秒を超えることができる。

10

【0125】

図17を参照するが、コマンド・バスを通じてそれぞれのセクションに付随するメモリのバンクAを活性化することにより、基本メモリ周期が開始される。これは、図17の例ではクロック周期1で起きる。4クロック周期後の、クロック周期5では、活性化コマンドが、バンクBに提供される。4周期後の、クロック周期9では、バンクCが活性化され、及びそれより4周期後の、クロック周期13では、バンクDが活性化される。さまざまなバンクを選択的に活性化するこのアプローチは、バンクAは、バンクAに対する次の活性化(activate)コマンドが発行される周期6及び19の間にプレチャージされるために、いかなる物理的プレチャージ時間(precharge time)に違反しないことを可能にする。

20

【0126】

さまざまなバンクのそれぞれに対して、バンクが活性化された5周期後に、データ転送が、READであるか、又はWRITEであるかを示すコマンドがその特定のバンクにアサートされる。従って、これは、バンクAに対してはクロック周期6で、バンクBに対してはクロック周期10で、バンクCに対してはクロック周期14で、及びバンクDに対してはクロック周期18で起きる。

30

【0127】

移転がREADであるか、又はWRITEであるかの指示の供給の5クロック周期後に、4つの連続周期の間、データがデータ回線に提供される。ダブルデータレート(DDR)でデータが移転されるために、それぞれのメモリクロック周期の間に2ワードが移転される。従ってそれぞれのバンクについて、基本メモリ周期全体にわたって8個の288ビットのワードのデータが移転される(図17中のデータ・バスに対する表記であるW0は、2つの288ビットのワードを包含する)。

40

【0128】

説明されている基本メモリ周期の間のデータ転送のコースの間に、クロック周期19でのバンクAに対する活性化コマンドのアサーションにより、次のメモリ周期に対する基本メモリ周期が開始される。次の基本メモリ周期の残りは、第一の基本メモリ周期に関して上記で以前に記述したことに従う。

【0129】

上記が示すように、完全な1キロバイトのデータが、基本メモリ周期ごとに移転される。一実施態様では、このシステムは、512バイトだけが移転される「ハーフライト」周期を提供できる。そのような半選択書き込み周期では、2つだけの連続したバンクの活性

50

化がアサートされる（例えば、バンク A 及び B、又はバンク C 及び D）。その他のバンクは、周期の間、アイドル状態にとどまる。移転についての T A R G E T A D D R E S S は、移転のために 4 つのバンクの内のどの 2 つを活性化するかを決定する。

【 0 1 3 0 】

上記の実施態様は説明的であり、及び限定的ではない。上述された本発明の 1 つ以上の態様を用いる、他の及び更なる開示される実施態様の趣旨を逸脱することなく、他の更なる実施態様を考案することができる。更に、本明細書に記載されたステップの順序は、特別に限定される場合を除き、さまざまな配列で起こることができる。本明細書に記載されたさまざまなステップは、他のステップと組み合わせること、公表されたステップの間に差し挟むこと、及び／又は複数のステップに分割することができる。同様に、素子は機能的に記載されており、及び個別の構成要素として具体化することができ、又は複数の機能を有する構成要素に組み合わせることができる。

【 0 1 3 1 】

開示されている実施態様は、好適な他の実施態様との関連で記載され、全ての実施態様は記載されていない。記載されている実施態様への明白な修正、及び変形は、当業者には利用可能である。開示されている、及び開示されていない実施態様は、本出願人により着想された、本発明の範囲及び適用性を限定、又は制限することを意図していない。本出願人は、以下の特許請求の範囲又は等価物の範囲内に含まれる、そのような修正及び改良の全てを完全に保護することを意図している。

10

【 図 1 】

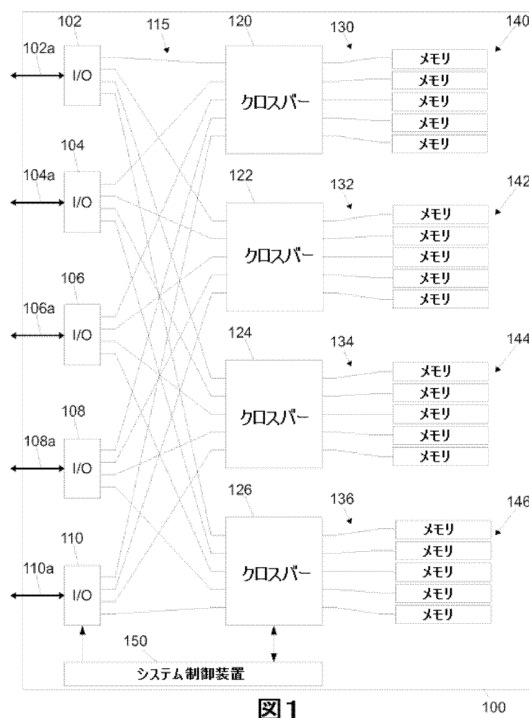


図 1

【 図 1 A 】

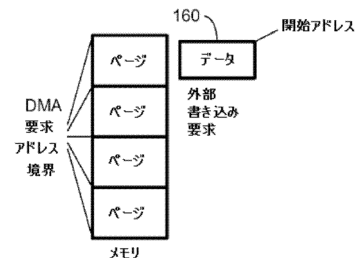


図 1A

【 図 1 B 】

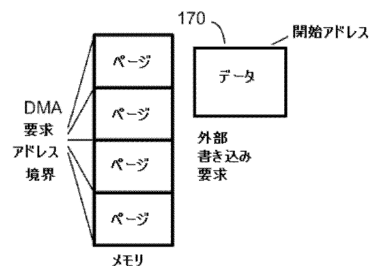


図 1B

【図 1 C】

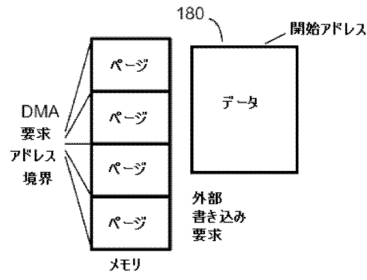


図1C

【図 2】

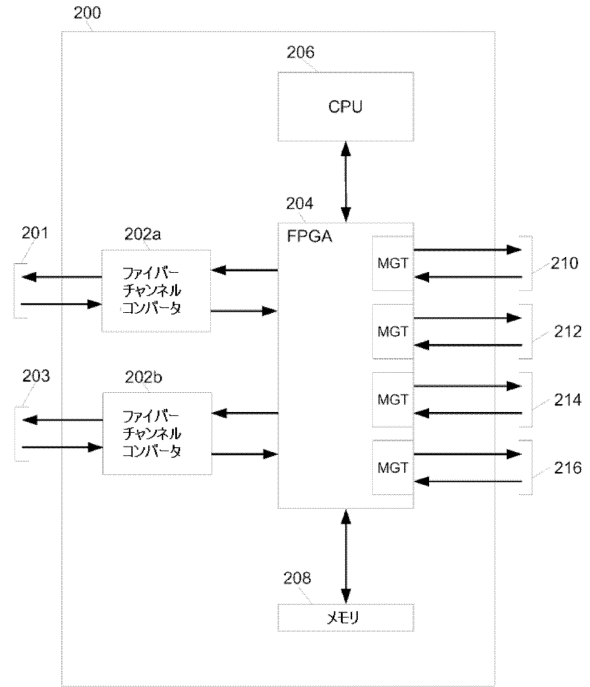


図2

【図 3】

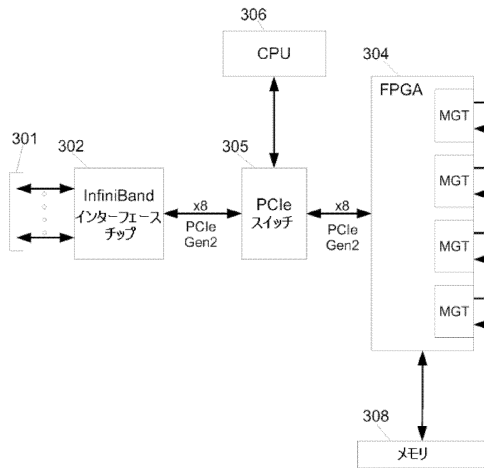


図3

【図 4】

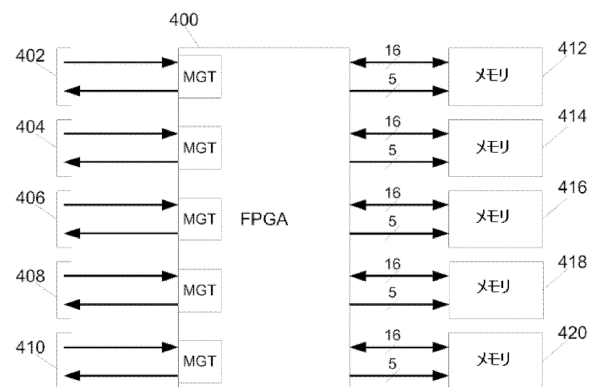


図4

【図 5】

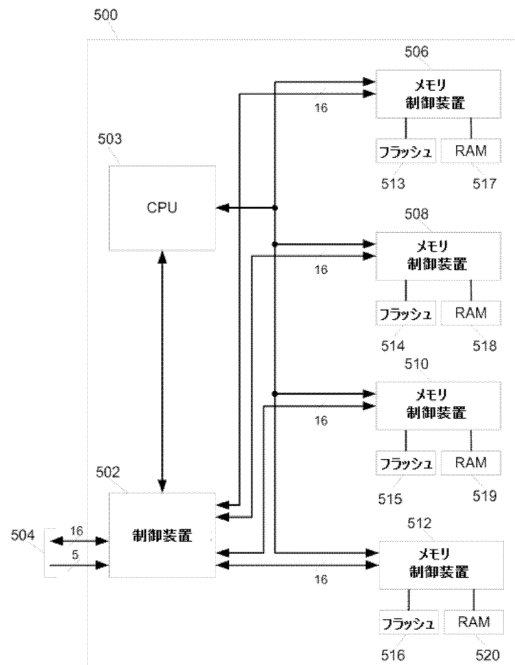


図5

【図 6】

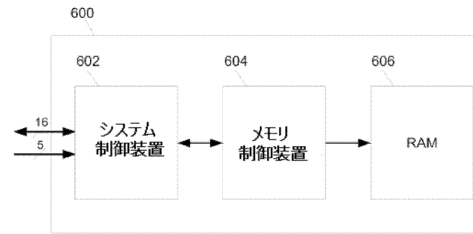


図6

【図 7】

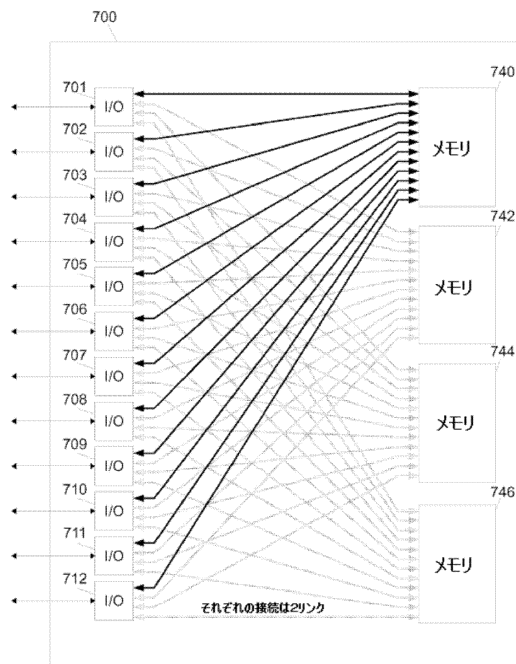


図7

【図 8】

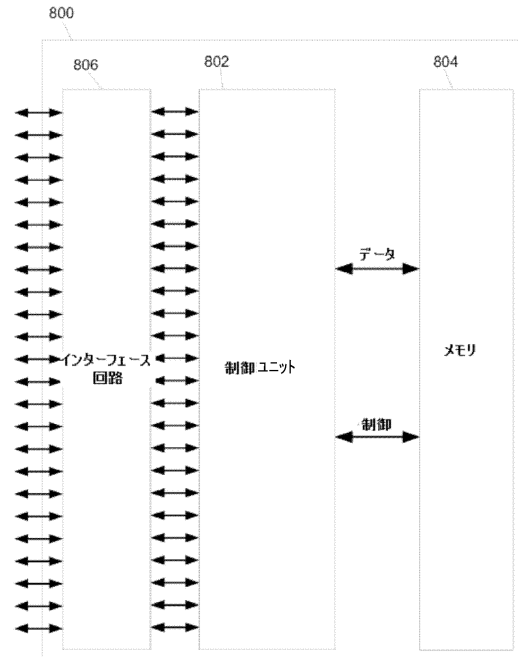


図8

【図 9】

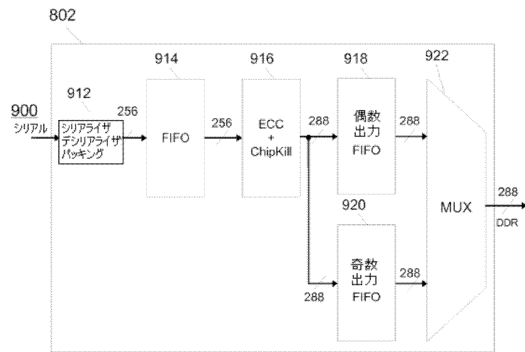


図9

【図 10 A】

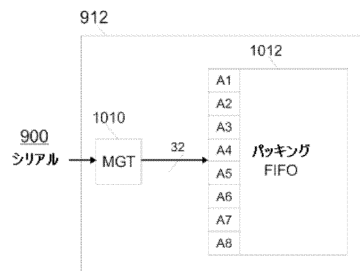


図10A

【図 10 B】

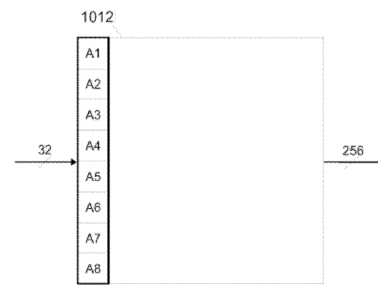


図10B

【図 10 C】

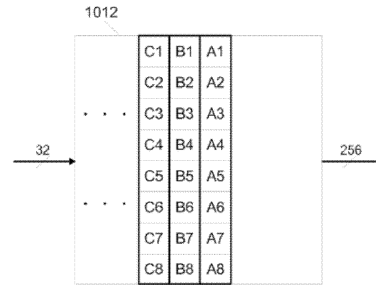


図10C

【図 10 D】

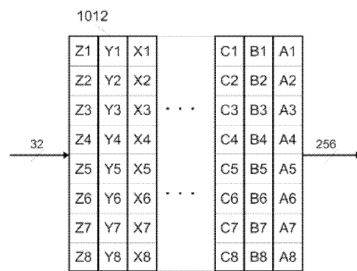


図10D

【図 1 2】

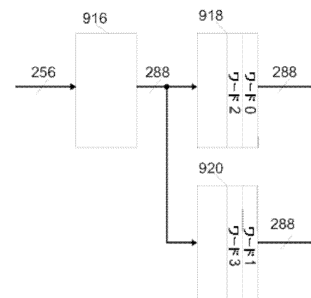


図12

【図 1 1】



図11

【図 1 2 A】

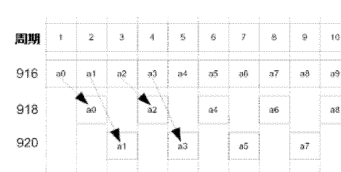


図12A

【図 13 A】

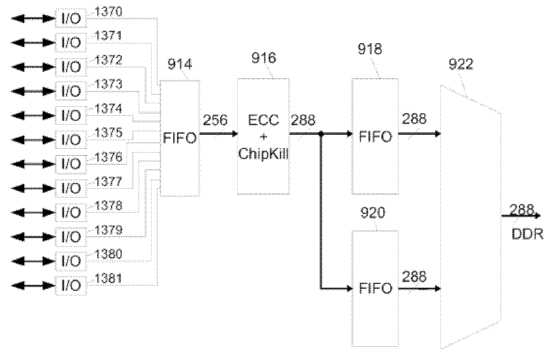


図 13A

【図 13 B】

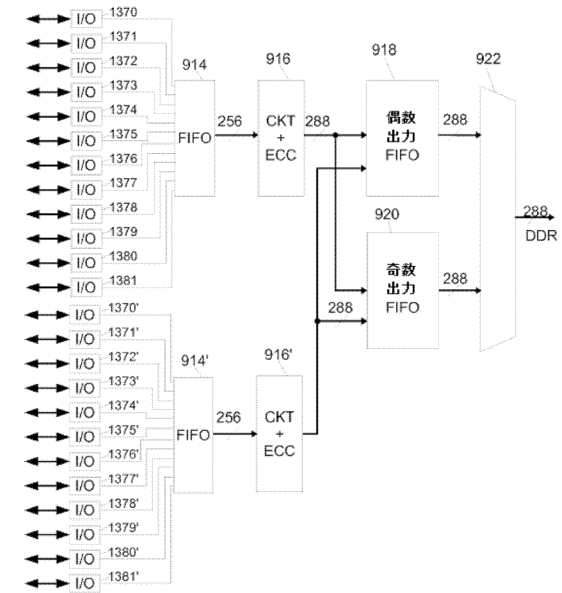


図 13B

【図 14 A】



図 14A

【図 14 C】



図 14C

【図 14 B】

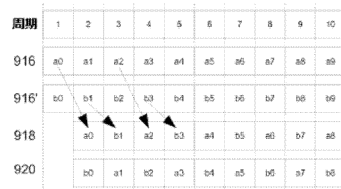


図 14B

【図 14 D】

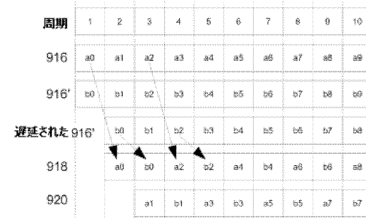


図 14D

【図 15】

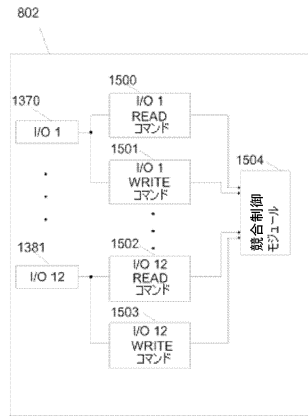


図15

【図 16】

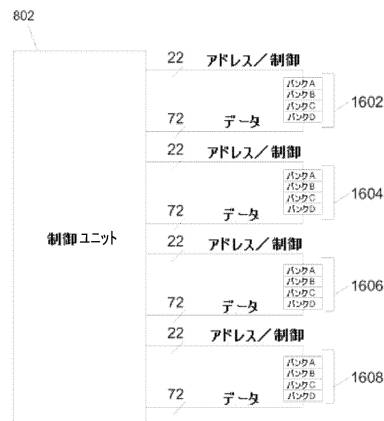


図16

【図 17】

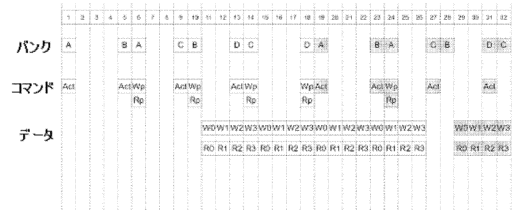


図17

フロントページの続き

(72)発明者 ハッツエル, レベツカ

アメリカ合衆国、テキサス・７７０７７、ヒューストン、リバービュー・ドライブ・１１４２６

審査官 酒井 恭信

(56)参考文献 特表２００９－５１０６５７（ＪＰ，Ａ）

米国特許出願公開第２００７／００７６４７９（ＵＳ，Ａ１）

特開２００８－３０５３１６（ＪＰ，Ａ）

特表２００７－５０４５３１（ＪＰ，Ａ）

米国特許出願公開第２００５／００５０２３７（ＵＳ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)

G 0 6 F 1 3 / 1 6 - 1 3 / 1 8

G 0 6 F 1 3 / 3 6

G 0 6 F 1 2 / 0 0 - 1 2 / 0 6

G 0 6 F 3 / 0 8