

(19)日本国特許庁(JP)

(12)特許公報(B1)

(11)特許番号
特許第7595229号
(P7595229)

(45)発行日 令和6年12月5日(2024.12.5)

(24)登録日 令和6年11月27日(2024.11.27)

(51)国際特許分類

F I

G 0 6 F 12/06 5 2 1 C

G 0 6 F 12/06 5 2 1 G

G 0 6 F 12/06 5 2 5 A

請求項の数 15 (全19頁)

(21)出願番号	特願2024-534636(P2024-534636)	(73)特許権者	591016172
(86)(22)出願日	令和4年12月7日(2022.12.7)		アドバンスト・マイクロ・デバイス
(86)国際出願番号	PCT/US2022/052164		・インコーポレイテッド
(87)国際公開番号	WO2023/114071		ADVANCED MICRO DEVI
(87)国際公開日	令和5年6月22日(2023.6.22)		CES INCORPORATED
審査請求日	令和6年7月8日(2024.7.8)		アメリカ合衆国 9 5 0 5 4 カリフォル
(31)優先権主張番号	17/549,359		ニア州、 サンタ クララ、 オーガスティ
(32)優先日	令和3年12月13日(2021.12.13)		ン ドライブ 2 4 8 5
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	100108833
早期審査対象出願			弁理士 早川 裕司
		(74)代理人	100111615
			弁理士 佐野 良太
		(74)代理人	100162156
			弁理士 村雨 圭介
		(72)発明者	スリセシャン スリカーント
			最終頁に続く

(54)【発明の名称】 ファイングレイン D R A Mにおける通常アクセス性能を回復するための方法及び装置

(57)【特許請求の範囲】

【請求項 1】

ファイングレインダイナミックランダムアクセスメモリ（ D R A M ）であって、
複数のグレインに分割されたメモリアレイを含む第 1 のメモリバンクであって、各グレインは行バッファ及び入力 / 出力（ I / O ）回路を含む、第 1 のメモリバンクと、
第 2 のメモリバンクと、
前記第 1 のメモリバンク内の各グレインの前記 I / O 回路に結合されたデュアルモード I / O 回路であって、第 1 のデータ幅を有するコマンドが各グレインにルーティングされ、各グレインにおいて個別に実行される第 1 のモードと、前記第 1 のデータ幅とは異なる第 2 のデータ幅を有するコマンドが前記グレインのうち少なくとも 2 つによって並列に実行される第 2 のモードと、で動作するデュアルモード I / O 回路と、を備える、
ファイングレイン D R A M。

【請求項 2】

前記デュアルモード I / O 回路は、マルチキャスト列アドレスストロープ（ C A S ）コマンドに応じて、前記第 2 のモードでコマンドを実行する、
請求項 1 のファイングレイン D R A M。

【請求項 3】

前記デュアルモード I / O 回路は、前記グレインのうち少なくとも 2 つにおいて選択された列に対して列アドレスストロープを並列にアサートさせることによって、前記マルチキャスト C A S コマンドに応答する、

請求項 2 のファイングレイン D R A M。

【請求項 4】

前記デュアルモード I / O 回路は、前記第 1 のメモリバンク内の各グレインの前記 I / O 回路内の列デコーダに結合されたマルチプレクサを含む、

請求項 3 のファイングレイン D R A M。

【請求項 5】

前記デュアルモード I / O 回路は、前記第 2 のモードにおいて、前記マルチキャスト C A S コマンド内の初期グレイン識別子ビットを、異なるグレインに関連付けられた新しい値に置き換えることによって、前記マルチキャスト C A S コマンドのローカライズされたリレーを前記グレインのうち少なくとも 2 つのグレインの少なくとも 1 つに転送するように実行する、

10

請求項 3 のファイングレイン D R A M。

【請求項 6】

前記デュアルモード I / O 回路は、ルックアップテーブル及び加算器のうち何れかを使用して前記新しい値を識別する、

請求項 5 のファイングレイン D R A M。

【請求項 7】

揮発性メモリを動作させる方法であって、

メモリコントローラから前記揮発性メモリ上の第 1 のバンクの第 1 のグレインに列アドレスストロブ (C A S) コマンドを送信することと、

20

前記揮発性メモリ上のデュアルモード I / O 回路にマルチキャスト C A S コマンドを送信することと、

前記揮発性メモリにおいて、前記マルチキャスト C A S コマンドに応じて、前記第 1 のグレインを含む複数のグレインに C A S 信号を送信することと、を含む、

方法。

【請求項 8】

前記 C A S コマンドに応じて、第 1 のデータ幅を有するデータを読み取ることと、前記 C A S 信号に応じて、前記第 1 のデータ幅とは異なる第 2 のデータ幅を有する第 2 のデータを前記複数のグレインから並列に読み取ることと、を含む、

請求項 7 の方法。

30

【請求項 9】

前記デュアルモード I / O 回路において、前記マルチキャスト C A S コマンドを前記複数のグレインに分配することを含む、

請求項 7 の方法。

【請求項 10】

前記デュアルモード I / O 回路において、前記マルチキャスト C A S コマンド内の初期グレイン識別子ビットを、異なるグレインに関連付けられた新しい値で置き換えることによって、マルチキャスト C A S のローカライズされたリレーを前記複数のグレインのうち少なくとも 1 つに転送することを含む、

請求項 7 の方法。

40

【請求項 11】

前記デュアルモード I / O 回路は、ルックアップテーブル及び加算器のうち何れかを使用して前記新しい値を識別する、

請求項 10 の方法。

【請求項 12】

ジャストインタイムコンパイラを使用して、元のメモリアクセス要求に基づいて前記マルチキャスト C A S コマンドを生成することを含む、

請求項 7 の方法。

【請求項 13】

コンパイラを使用して、元のメモリアクセス要求に基づいて前記マルチキャスト C A S

50

コマンドを生成することを含む、

請求項 7 の方法。

【請求項 1 4】

メモリコントローラにおいて、所定のデータ幅に関連付けられた命令セットアーキテクチャ (ISA) コマンドを認識することに基づいて前記マルチキャスト CAS コマンドを生成することを含む、

請求項 7 の方法。

【請求項 1 5】

メモリコントローラにおいて、前記メモリコントローラによって処理されるメモリコマンドに関連付けられたメモリ領域に基づいて前記マルチキャスト CAS コマンドを生成することを含む、

10

請求項 7 の方法。

【発明の詳細な説明】

【背景技術】

【0001】

高帯域幅ダイナミックランダムアクセスメモリ (dynamic random-access memory、DRAM) は、グラフィックス処理ユニット (Graphics Processing Unit、GPU) 及び他のスループット指向の並列プロセッサによって使用される。次世代の GPU 専用 DRAM は、主に、先ず、グラフィックスダブルデータレートメモリとともに使用されるような高周波オフチップシグナリングを使用することによって、より最近では、プロセッサダイと、高帯域幅メモリ (High Bandwidth Memory、HBM / HBM2) モジュールにおいて使用されるような DRAM チップのスタックへの広い高帯域幅インターフェースとのオンパッケージ統合を通して、帯域幅を最大化するように最適化されている。

20

【0002】

そのようなシステムのための多くの重要なワークロードクラスは、不規則なファイングレインメモリアクセス (fine-grained memory accesses) を経験する。そのようなワークロードとしては、グラフトラバースル (graph traversal) (ソーシャルネットワーキング、検索、e コマース、クラウド、及びギグエコノミーにおいて広く使用される)、キーバリューストアアクセス (key-value store access)、及び、高性能コンピューティング (high performance computing、HPC) が挙げられる。次世代のグレインベースのスタック型 DRAM は、エネルギー効率が良く、高帯域幅のファイングレインで不規則なアクセスを可能にするのに重要である。

30

【0003】

ファイングレイン DRAM (Fine-grained DRAM、FG-DRAM) は、実用的な熱設計電力 (Thermal Design Power、TDP) の限界を超えずに、HBM 等のスタック型 DRAM メモリがメモリ帯域幅をスケールアップすることを可能にする。FG-DRAM は、DRAM ダイを、コマンドアドレス (command-address、CA) ピンではなく専用の DQ ピンを各々が有するより小さい独立したユニット (グレインと呼ばれる) にパーティション分割することによって、そのようなスケールアップを行う。例えば、従来の DRAM バンクを 2 つに分割して、元のバンクと比較して各々が半分の数の列を有する 2 つのグレインを実現することができる。その結果、メモリレベルの並列性の著しい向上が得られる。更に、活性化エネルギーの大幅な低減が実現される。

40

【0004】

しかしながら、そのようなグレインベースのアーキテクチャでは、各グレインが DRAM チャンネルに対してより狭いインターフェースを有するので、開いた行 (row) への列 (column) アクセスコマンドを介した DRAM アトムへのアクセスレイテンシは、バンクごとのグレインの数に比例して増加する。そのような増加は、レイテンシの影響を受けやすい通常アクセスのワークロードにとって有害である。

【図面の簡単な説明】

【0005】

50

【図 1】先行技術で知られている加速処理ユニット (accelerated processing unit、A P U) 及びメモリシステムのブロック図である。

【図 2】いくつかの実施形態による、図 1 と同様の A P U で用いるのに適したメモリコントローラのブロック図である。

【図 3】グレインの使用を示す 2 つの D R A M の 2 つのメモリバンクを比較するブロック図である。

【図 4】グレインに依存しないフォーマットでインターリーブされたファイングレイン D R A M メモリのためのアドレス線の配置を示すブロック図である。

【図 5】グレインストライプフォーマットでインターリーブされたファイングレイン D R A M メモリのためのアドレス線の配置を示すブロック図である。

10

【図 6】先行技術による、ファイングレインメモリバンクを示すブロック図である。

【図 7】いくつかの実施形態による、ファイングレインメモリバンクを示すブロック図である。

【図 8】いくつかの追加の実施形態による、ファイングレインメモリバンクを示すブロック図である。

【図 9】いくつかの実施形態による、ファイングレインメモリを動作させるためのプロセスのフロー図である。

【図 10】いくつかの実施形態による、ファイングレインメモリを動作させるためのライブラリ命令を含むデータ処理システムを示すブロック図である。

【図 11】いくつかの実施形態による、ファイングレインメモリを動作させるためのプロセスのフロー図である。

20

【発明を実施するための形態】

【0006】

以下の説明において、異なる図面における同一の符号の使用は、同様のアイテム又は同一のアイテムを示す。別段の言及がなければ、「結合される (coupled)」という単語及びその関連する動詞形は、当該技術分野で周知の手段による直接接続及び間接電気接続の両方を含み、また、別段の言及がなければ、直接接続の任意の記述は、好適な形態の間接電気接続を使用する代替の実施形態も同様に意味する。

【0007】

ファイングレインダイナミックランダムアクセスメモリ (D R A M) は、第 1 のメモリバンクと、第 2 のメモリバンクと、デュアルモード I / O 回路と、を含む。第 1 のメモリバンクは、複数のグレインに分割されたメモリアレイを含み、各グレインは、行バッファ及び入力 / 出力 (I / O) 回路を含む。デュアルモード I / O 回路は、第 1 のメモリバンク内の各グレインの I / O 回路に結合され、第 1 のデータ幅を有するコマンドが各グレインにルーティングされ、各グレインにおいて個別に実行される第 1 のモードと、第 1 のデータ幅とは異なる第 2 のデータ幅を有するコマンドがグレインのうち少なくとも 2 つによって並列に実行される第 2 のモードと、で動作する。

30

【0008】

揮発性メモリを動作させる方法は、メモリコントローラから当該揮発性メモリ上の第 1 のバンク内の第 1 のグレインに列アドレスストロブ (column address strobe、C A S) コマンドを送信することを含む。本方法は、当該揮発性メモリ上のデュアルモード I / O 回路にマルチキャスト C A S コマンドを送信することを含む。当該揮発性メモリにおいて、マルチキャスト C A S コマンドに応じて、本方法は、当該第 1 のグレインを含む複数のグレインに C A S 信号を送信することを含む。

40

【0009】

データ処理システムは、複数の処理ユニットと、複数の処理ユニットに結合されたデータファブリックと、処理ユニットからのメモリアクセス要求に対応するためにデータファブリックに結合されたメモリコントローラと、メモリコントローラと通信するファイングレインダイナミックランダムアクセスメモリ (D R A M) と、を含む。ファイングレイン D R A M は、第 1 のメモリバンクと、第 2 のメモリバンクと、デュアルモード I / O 回路

50

と、を含む。第 1 のメモリバンクは、複数のグレインに分割されたメモリアレイを含み、各グレインは、行バッファ及び入力／出力（I／O）回路を含む。デュアルモード I／O 回路は、第 1 のメモリバンク内の各グレインの I／O 回路に結合され、デュアルモード I／O 回路は、第 1 のデータ幅を有するコマンドが各グレインにルーティングされ、各グレインにおいて個別に実行される第 1 のモードと、第 1 のデータ幅とは異なる第 2 のデータ幅を有するコマンドがグレインのうち少なくとも 2 つによって並列に実行される第 2 のモードと、で動作する。

【 0 0 1 0 】

図 1 は、いくつかの実施形態による、加速処理ユニット（A P U）1 0 0 のブロック図である。A P U 1 0 0 は、様々なホストデータ処理プラットフォームの一部であり得るシステムオンチップ（System-on-Chip、S o C）として実装される。この実施形態では A P U が示されているが、中央処理ユニット（central processing unit、C P U）又はグラフィックス処理ユニット（graphics processing unit、G P U）等の他のデータ処理プラットフォームが使用されてもよい。例えば、いくつかの実施形態では、本明細書のファイングレインメモリアクセス技術は、グラフィックスカード又は他のグラフィックス処理モジュールにおいて採用される G P U チップにおいて具現化される。他の実施形態では、インテリジェンス処理ユニット（intelligence processing unit、I P U）等の専用プロセッサコアが使用されてもよい。この実施形態では、A P U 1 0 0 は、概して、C P U コア複合体 1 1 0、グラフィックスコア 1 2 0、ディスプレイエンジンのセット 1 3 0、メモリ管理ハブ 1 4 0、データファブリック 1 5 0、周辺コントローラのセット 1 6 0、周辺バスコントローラのセット 1 7 0、システム管理ユニット（system management unit、S M U）1 8 0、フラッシュメモリ 2 0 5、及び、F G _ D R A M メモリコントローラのセット 1 9 0 を含む。

【 0 0 1 1 】

C P U コア複合体 1 1 0 は、C P U コア 1 1 2 及び C P U コア 1 1 4 を含む。この例では、C P U コア複合体 1 1 0 が 2 つの C P U コアを含むが、他の実施形態では、C P U コア複合体 1 1 0 が任意の数の C P U コアを含むことができる。C P U コア 1 1 2 及び 1 1 4 の各々は、制御ファブリックを形成するシステム管理ネットワーク（system management network、S M N）1 4 5 及びデータファブリック 1 5 0 に双方向に接続され、データファブリック 1 5 0 にメモリアクセス要求を提供することができる。C P U コア 1 1 2 及び 1 1 4 の各々は、単体のコアであってもよいし、更にキャッシュ等の特定のリソースを共有する 2 つ以上の単体のコアを有するコア複合体であってもよい。

【 0 0 1 2 】

グラフィックスコア 1 2 0 の各々は、頂点処理、フラグメント処理、シェーディング、テクスチャレンディング等のグラフィックス動作を高度に統合化された並列方式で実行することができる高性能グラフィックス処理ユニット（G P U）である。各グラフィックスコア 1 2 0 は、S M N 1 4 5 及びデータファブリック 1 5 0 に双方向に接続され、メモリアクセス要求をデータファブリック 1 5 0 に提供することができる。これに関して、A P U 1 0 0 は、C P U コア複合体 1 1 0 とグラフィックスコア 1 2 0 とが同じメモリ空間を共有する統合メモリアーキテクチャ、又は、C P U コア複合体 1 1 0 とグラフィックスコア 1 2 0 とがメモリ空間の一部を共有する一方でグラフィックスコア 1 2 0 が C P U コア複合体 1 1 0 によりアクセスできないプライベートグラフィックスメモリも使用するメモリアーキテクチャの何れかをサポートすることができる。

【 0 0 1 3 】

ディスプレイエンジン 1 3 0 は、モニタ上に表示するために、グラフィックスコア 1 2 0 によって生成されたオブジェクトをレンダリング及びラスターライズする。グラフィックスコア 1 2 0 及びディスプレイエンジン 1 3 0 は、メモリ内の適切なアドレスへの均一な変換のために共通のメモリ管理ハブ 1 4 0 に双方向で接続され、メモリ管理ハブ 1 4 0 は、そのようなメモリアクセスを生成してメモリシステムから戻される読み取りデータを受信するためにデータファブリック 1 5 0 に双方向で接続される。

10

20

30

40

50

【 0 0 1 4 】

データファブリック 1 5 0 は、任意のメモリアクセスエージェントとメモリコントローラ 1 9 0 との間でメモリアクセス要求及びメモリ応答をルーティングするためのクロスバースイッチを含む。また、データファブリックは、システム構成に基づいてメモリアクセスの送信先を判定するための、基本入出力システム (basic input/output system、B I O S) によって規定されるシステムメモリマップ、及び、各仮想接続のためのバッファも含む。

【 0 0 1 5 】

周辺コントローラ 1 6 0 は、U S B コントローラ 1 6 2 及びシリアルアドバンストテクノロジーアタッチメント (serial advanced technology attachment、S A T A) インターフェースコントローラ 1 6 4 を含み、その各々は、システムハブ 1 6 6 及び S M N 1 4 5 に双方向に接続される。これら 2 つのコントローラは、A P U 1 0 0 において使用され得る周辺コントローラの単なる例示である。

10

【 0 0 1 6 】

周辺バスコントローラ 1 7 0 は、システムコントローラハブ 1 7 2 及び周辺コントローラハブ 1 7 4 を含み、その各々は、入力 / 出力 (I / O) ハブ 1 7 6 及び S M N 1 4 5 に双方向に接続される。システムコントローラハブ 1 7 2 は、適切な通信リンクを介してフラッシュメモリ 2 0 5 に接続する。I / O ハブ 1 7 6 は、システムハブ 1 6 6 及びデータファブリック 1 5 0 に対して双方向に接続される。したがって、例えば、C P U コアは、データファブリック 1 5 0 が I / O ハブ 1 7 6 を介してルーティングするアクセスにより、U S B コントローラ 1 6 2、S A T A インターフェースコントローラ 1 6 4、システムコントローラハブ 1 7 2、又は、周辺コントローラハブ 1 7 4 内のレジスタをプログラムすることができる。

20

【 0 0 1 7 】

S M U 1 8 0 は、A P U 1 0 0 上のリソースの動作を制御してそれらの間の通信を同期させるローカルコントローラである。S M U 1 8 0 は、A P U 1 0 0 上の様々なプロセッサのパワーアップシーケンシングを管理し、リセット、イネーブル及び他の信号を介して複数のオフチップデバイスを制御する。また、S M U 1 8 0 は、様々なプロセッサ及び他の機能ブロックの電力を管理する。

【 0 0 1 8 】

S o C 実施形態が示されているが、これは限定的なものではなく、他のコンピューティングプラットフォームも、本明細書で説明される技術から利益を得ることができる。

30

【 0 0 1 9 】

図 2 は、図 1 のような A P U で使用するのに好適なメモリコントローラ 2 0 0 のブロック図である。メモリコントローラ 2 0 0 は、概して、インターフェース 2 1 2、メモリアンターフェースキュー 2 1 4、コマンドキュー 2 2 0、アドレス生成器 2 2 2、コンテンツアドレスابلメモリ (content addressable memory、C A M) 2 2 4、リプレイキュー 2 3 0 を含むリプレイ制御ロジック 2 3 1、リフレッシュ制御ロジック 2 3 2、リフレッシュ制御ロジック 2 3 2、タイミングブロック 2 3 4、ページテーブル 2 3 6、アービタ 2 3 8、エラー訂正コード (error correction code、E C C) チェック回路 2 4 2、E C C 生成ブロック 2 4 4、及び、データバッファ 2 4 6 を含む。

40

【 0 0 2 0 】

インターフェース 2 1 2 は、外部バスを介したデータファブリックに対する第 1 の双方向接続を有し、出力を有する。メモリコントローラ 2 0 0 において、この外部バスは、「A X I 4」として知られている英国ケンブリッジの A R M H o l d i n g s , P L C によって指定された高度拡張可能インターフェースバージョン 4 と適合するが、他の実施形態では他のタイプのインターフェースであり得る。インターフェース 2 1 2 は、メモリアクセス要求を、F C L K (又は M E M C L K) ドメインとして知られている第 1 のクロックドメインから、U C L K ドメインとして知られているメモリコントローラ 2 0 0 の内部の第 2 のクロックドメインに変換する。同様に、メモリアンターフェースキュー 2 1 4 は

50

、UCLKドメインからDFIインターフェースと関連付けられるDFICLKドメインへのメモリアクセスを与える。

【0021】

アドレス生成器222は、AXI4バスを介してデータファブリックから受信されるメモリアクセス要求のアドレスを復号する。メモリアクセス要求は、正規化フォーマットで表された物理アドレス空間内のアクセスアドレスを含む。アドレス生成器222は、正規化されたアドレスを、メモリシステム内の実際のメモリデバイスをアドレス指定するために、及び、関連するアクセスを効率的にスケジュールするために使用され得るフォーマットに変換する。このフォーマットは、メモリアクセス要求を特定のランク、行アドレス、列アドレス、バンクアドレス、及び、バンクグループと関連付ける領域識別子を含む。起動時に、システムBIOSは、メモリシステム内のメモリデバイスに問い合わせるそれらのサイズ及び構成を判定し、アドレス生成器222と関連付けられた構成レジスタのセットをプログラムする。アドレス生成器222は、構成レジスタに記憶された構成を使用して、正規化されたアドレスを適切なフォーマットに変換する。コマンドキュー220は、CPUコア112、114及びグラフィックスコア120等のAPU100内のメモリアクセスエージェントから受信されるメモリアクセス要求のキューである。コマンドキュー220は、アドレス生成器222によってデコードされたアドレスフィールド、及び、アービタ238がアクセスタイプ及びサービス品質（quality of service、QoS）識別子を含むメモリアクセスを効率的に選択することを可能にする他のアドレス情報を記憶する。CAM224は、ライトアフターライト（write after write、WAW）及びリードアフターライト（read after write、RAW）順序規則等の順序規則を実施するための情報を含む。コマンドキュー220は、それぞれが複数のコマンドエントリを含む複数のエントリスタックを含むスタックされたコマンドキューであり、この実施形態では、以下で更に説明するように、それぞれが4つのエントリからなる32のエントリスタックである。

【0022】

誤り訂正コード（ECC）生成ブロック244は、メモリに送られる書き込みデータのECCを判定する。次いで、このECCデータは、データバッファ246内の書き込みデータに追加される。ECCチェック回路242は、受信されたECCを着信ECCと照合してチェックする。

【0023】

リプレイキュー230は、アドレス及びコマンドパリティ応答等の応答を待っているアービタ238によって選択されたメモリアクセスを記憶するための一時キューである。リプレイ制御ロジック231は、ECCチェック回路242にアクセスして、戻されたECCが正しいか又はエラーを示すかを判定する。リプレイ制御ロジック231は、これらのサイクルのうち何れかのパリティ又はECCエラーの場合にアクセスがリプレイされるリプレイシーケンスを開始して制御する。リプレイされたコマンドは、メモリインターフェースキュー214に配置される。

【0024】

リフレッシュ制御ロジック232は、メモリアクセスエージェントから受信した通常の読み取り及び書き込みメモリアクセス要求とは別に生成される様々な電源断、リフレッシュ及び終端抵抗（ZQ）較正サイクルのためのステートマシンを含む。例えば、メモリバンクがプリチャージパワーダウンにある場合、リフレッシュロジックは、リフレッシュサイクルを実行するために定期的に起動されなければならない。リフレッシュ制御ロジック232は、DRAMチップ内のメモリセルの蓄積コンデンサからの電荷の漏れによって引き起こされるデータエラーを防止するために、定期的に、定められた条件に応じて、リフレッシュコマンドを生成する。リフレッシュ制御ロジック232は、アクティブ化カウンタ248を含み、この実施形態において、アクティブ化カウンタ248は、メモリチャネルを介してメモリ領域に送信されるアクティブ化コマンドのローリング数をカウントするカウンタをメモリ領域ごとに有する。メモリ領域は、以下で更に説明するように、いくつかの実施形態ではメモリバンクであり、他の実施形態ではメモリサブバンクである。更に

10

20

30

40

50

、リフレッシュ制御ロジック 232 は、システム内の熱変化に起因するオンダイ終端抵抗の不一致を防止するために ZQ を定期的に較正する。

【0025】

アービタ 238 は、コマンドキュー 220 に双方向に接続され、メモリコントローラ 200 の心臓部であり、メモリバスの使用を改善するためにアクセスのインテリジェントスケジューリングを実行する。この実施形態では、アービタ 238 は、以下で更に説明するように、複数のモードにおけるファイングレインメモリアccessのためのコマンド及びアドレスを変換するためのグレインアドレス変換ブロック 240 を含む。いくつかの実施形態では、そのような機能はアドレス生成ブロック 222 によって実行され得る。アービタ 238 は、タイミングブロック 234 を使用して、コマンドキュー 220 内の特定のアクセスが DRAM タイミングパラメータに基づいて発行に適格であるかどうかを判定することによって、適切なタイミング関係を実施する。例えば、各 DRAM は、「t_{RC}」として知られるアクティブ化コマンド間の最小指定時間を有する。タイミングブロック 234 は、JEDC 仕様で定められたこのタイミングパラメータ及び他のタイミングパラメータに基づいて適格性を判定するカウンタのセットを維持し、リプレイキュー 230 に対して双方向で接続される。ページテーブル 236 は、アービタ 238 のためのメモリチャネルの各バンク及びバンクにおけるアクティブページに関する状態情報を維持し、リプレイキュー 230 に対して双方向で接続される。アービタ 238 は、コマンドキュー 220 の各エントリスタックに対する単一のコマンド入力を含み、そこからコマンドを選択して、メモリインターフェースキュー 214 を介して DRAM チャネルにディスパッチするようにスケジュールする。

10

20

【0026】

インターフェース 212 から受信した書き込みメモリアccess要求に応じて、ECC 生成ブロック 244 は、書き込みデータに従って ECC を計算する。データバッファ 246 は、受信したメモリアccess要求に関する書き込みデータ及び ECC を記憶する。データバッファは、アービタ 238 がメモリチャネルへのディスパッチのために対応する書き込みアクセスを選択すると、組み合わされた書き込みデータ/ECC をメモリインターフェースキュー 214 に出力する。

【0027】

メモリコントローラ 200 は、関連するメモリチャネルへのディスパッチのためのメモリアccessを選択することを可能にする回路を含む。所望のアービトレーション決定を行うために、アドレス生成器 222 は、アドレス情報を、メモリシステム内のバンク、行アドレス、列アドレス、バンクアドレス、及び、バンクグループを含むプリデコードされた情報にデコードし、コマンドキュー 220 がプリデコードされた情報を記憶する。構成レジスタ（図示せず）は、アドレス生成器 222 が受信したアドレス情報をどのようにデコードするかを判定するための構成情報を記憶する。アービタ 238 は、デコードされたアドレス情報、タイミングブロック 234 によって示されるタイミング適格性情報、及び、ページテーブル 236 によって示されるアクティブページ情報を使用して、サービス品質（QoS）要件等の他の基準を遵守しながら、メモリアccessを効率的にスケジュールする。例えば、アービタ 238 は、メモリページを変更するために必要なプリチャージコマンド及びアクティブ化コマンドのオーバーヘッドを回避するために、オープンページへのアクセスの優先度を実装し、あるバンクへのオーバーヘッドアクセスを別のバンクへの読み取り及び書き込みアクセスとインターリーブすることによって隠す。特に、通常動作中、アービタ 238 は、通常、ページを、これらのページが異なるページを選択する前にプリチャージされる必要があるまで、異なるバンクで開いたままにする。

30

40

【0028】

図 3 は、グレインの使用を示す 2 つの DRAM の 2 つのメモリバンク 302 及び 304 を比較するブロック図 300 である。DRAM バンク 302 はグレインなしで構成され、一方、DRAM バンク 304 は 2 つのグレイン 306 及び 308 を含む。DRAM バンク 304 は、複数の DRAM バンクを含むより大きな DRAM 集積回路の一部である。DR

50

A Mバンク 3 0 4 は、概して、複数のグレイン、この場合は 2 つのグレインに分割されたメモリアレイを含み、各グレインは、行バッファ及び入力 / 出力 (I / O) 回路を含む。

【 0 0 2 9 】

D R A Mバンク 3 0 2 の従来の非グレインアーキテクチャは、D R A Mバンク 3 0 2 に割り当てられた擬似チャネル (pseudo-channel、p C H) 上で 8 バースト又は「ビート (beats) 」でアクセスされる 3 2 本のデータ線 D Q [0 : 3 1] を介して 3 2 バイト (3 2 B) メモリアクセスを実行することができる。バンク内の開いた行に列コマンドを発行することにより、8 バーストで 3 2 バイト (3 2 B) のデータが提供されるが、これは、バンクが擬似チャネルの 3 2 個全ての D Q ピンにアクセスしているためであり、ピンが分割されていないため、D R A Mバンク 3 0 4 の 2 グレインアーキテクチャと比較して 2 倍高速のデータ読み出しが提供される。その結果、非グレインアーキテクチャにおいて、開いた行から 6 4 B キャッシュ線を読み出すための時間は、 $2 * 8 = 16$ バーストに 1 つの追加の列から列への長い遅延時間 (t C C D L) を加え、オーバーヘッド時間のデータバス上のバースト持続時間 (t B U R S T) を引いた時間を要する。

10

【 0 0 3 0 】

本明細書の技術なしにファイングレインアーキテクチャを使用する場合、D R A Mバンク 3 0 4 は、p C H のデータ線の 2 つの指定されたセット D Q [0 : 1 5] 及び D Q [1 6 : 3 1] を介して 1 6 B アクセスでアクセスされる。グレイン専用の擬似チャネルの 1 6 個の D Q ピンを有するこの例示的な構成では、グレイン内の開いた行は、8 バーストで 1 6 B のデータを提供することによって列コマンドに応答する。したがって、開いた行から 6 4 B キャッシュ線を読み出すための時間は、 $4 * 8 = 32$ バーストに追加の 3 (t C C D L - t B U R S T) サイクルのオーバーヘッドを加えた時間を要する。したがって、グレインアーキテクチャは、個々の 3 2 B アクセスに対して 3 3 % のレイテンシペナルティを被る。

20

【 0 0 3 1 】

図 4 は、グレインに依存しないフォーマットでインターリーブされたファイングレイン D R A Mメモリのためのアドレス線の配置を示す図である。図示した配置では、アドレス線番号は、各アドレス線番号の下にラベル付けされたアドレス線を使用して上部に示されている。「R O」は行を表し、「C O」は列を表し、「G R」はグレイン番号を表し、「B K」はバンク番号を表し、「B G」はバンクグループ番号を表し、「P C H」は擬似チャネルを表し、「X」はアドレス線が割り当てられていないことを示す。グレインに依存しないアドレスインターリーブでは、イレギュラースループット指向の 1 6 B アクセスを可能にするためにストライピング (striping) は行われず、アドレス線 1 2 のグレインアドレスビットはバンクアドレスビットと同様に使用される。

30

【 0 0 3 2 】

図 5 は、グレインストライプフォーマットでインターリーブされたファイングレイン D R A Mメモリのためのアドレス線の配置を示す図である。図示された実施形態は、グレインストライプアドレスインターリーブングを用いてグレインのストライプの 3 2 B 配置を実現するために、メモリコントローラにおいて修正されたアドレスインターリーブングを展開する。3 2 B のデータは、それぞれ 1 6 B の 2 つの相関グレインに分割される。G R アドレスビットは、最下位アドレス線 (線 4) に存在し、隣接するグレインがこのビットによってアドレス指定されることを意味する。この手法は、例えば、4 つのグレイン、8 つのグレイン又は 1 6 のグレイン等のように、バンクごとに増加した数のグレインをサポートするように拡張され得ることに留意されたい (理論的には、バンク内の列の数によってのみ制限される) 。

40

【 0 0 3 3 】

しかしながら、図 5 のグレインストライプアドレスインターリーブングを単独で単に使用するだけでは、同じ物理バンクへの列コマンド間に存在する t C C D L 制約のために、許容できない追加の 7 5 % オーバーヘッドに悩まされる。このオーバーヘッドは、「パートナー」グレイン (隣接するアドレスで機能するグレイン) を異なるバンクグループに分

50

離することによって、いくつかの実施形態では t C C D S のオーバーヘッドまで下げることができる。しかしながら、そのような実装においてであっても、オーバーヘッドは依然として無視できない。更に、パートナーグレインの異なるバンクグループへのそのような分離は、2つのグレイン間の非同期レイテンシを回避するために、メモリコントローラに追加のスケジューリング及びデータ応答再構成の複雑さを課す。

【0034】

図6は、先行技術による、ファイングレインメモリバンク600のブロック図であり、図7は、いくつかの実施形態による、ファイングレインメモリバンク700のブロック図である。

【0035】

メモリバンク600は、「グレイン0」及び「グレイン1」とラベル付けされた2つのグレインと、行デコーダ602と、2つの列デコーダ604及び606と、「Gr. Sel」とラベル付けされたグレイン選択回路608と、コマンドバス620と、アドレスバス622と、を含む。動作中、メモリバンク600は、メモリコントローラからコマンドバス620を介して送信される図示された列アドレスストローブ(CAS)コマンド等のコマンドでアクセスされる。図示されるように、アドレスバス622は、図4のアドレス方式に従って、列アドレスビット(CO)を列デコーダ604及び606に提供し、グレインビット(GR)をグレインセレクト回路608に提供する。メモリアクセスは、1つのグレインの幅であり、この例では16Bである。

【0036】

図7を参照すると、メモリバンク700は、「グレイン0」及び「グレイン1」とラベル付けされた2つのグレインと、行デコーダ702、2つの列デコーダ704及び706、各グレイン内の行バッファ(別々に図示せず)、「mCAS supp」(mCASサポート)とラベル付けされたデュアルモード入力/出力(I/O)回路708、720とラベル付けされたコマンドバス、並びに、アドレスバス722を含むI/O回路と、を含む。この実施形態では、デュアルモードI/O回路708は、メモリバンク内の各グレインのI/O回路に接続される。

【0037】

デュアルモードI/O回路708は、「mCASイネーブル」とラベル付けされた第1の入力と、「グレインビット」とラベル付けされた第2の入力と、「グレイン0イネーブル」とラベル付けされた第1の出力と、「グレイン1イネーブル」とラベル付けされた第2の出力と、を含む。第1の入力は、メモリコントローラからコマンドバス720を介してマルチキャストCASイネーブル信号を受信し、第2の入力は、グレインアドレスビット(GR、図5)を受信する。第1の出力は列デコーダ704に接続され、第2の出力は列デコーダ706に接続される。

【0038】

デュアルモードI/O回路708は、概して、第1のデータ幅(16B)を有するコマンドが各グレインにルーティングされ、各グレインにおいて個別に実行される第1のモードと、第1のデータ幅とは異なる第2のデータ幅(32B)を有するコマンドがグレインのうち少なくとも2つによって並列に実行される第2のモードと、で動作する。そのような構成は、この例では両方ともバンク700のグレインである2つ以上のパートナーとなるグレイン(partnered grain)に対するマルチキャスト列アドレスストローブ(mCAS)コマンドを使用することによって、上記で説明したtCCDL制約によって課されるオーバーヘッドを克服する。mCASコマンドを構成するアドレスビットは、グレイン識別子ビットを除いて、両方のグレインと同一である。したがって、図示した実施形態は、mCASコマンドのサポートを用いて共有コマンド及びアドレス(CA)バス/論理を増強する。

【0039】

図示した2グレインの例では、デュアルモードI/O回路708は、mCASイネーブル信号で増強されたマルチプレクサを用いて実装されるが、他の実施形態では他の論理実

10

20

30

40

50

装が採用される。デュアルモード I / O 回路 708 は、図の右側により詳細に示されており、この実施形態では、マルチプレクサ 710 及び 2 つの OR ゲート 712 を含む。マルチプレクサ 710 は、グレインアドレスビットを受信する入力と、OR ゲート 712 のそれぞれの入力に接続された 2 つの出力と、を有する。各 OR ゲート 712 は、マルチプレクサ 710 の出力に接続された第 1 の入力と、マルチキャスト C A S イネーブル信号を受信する第 2 の入力と、を有する。

【0040】

動作中、現在の C A S コマンドがマルチキャスト C A S コマンドであることを知らせるためにマルチキャスト C A S イネーブル信号が H I G H である場合、OR ゲート 712 の出力は H I G H になり、バンク 700 の両方のグレイン内の選択された列に対して C A S をアサートさせることによってマルチキャスト C A S コマンドをイネーブルする。次いで、グレインは、C A S に並列に応答して、より高いデータ幅を提供する。このモードでは、両方のグレインが m C A S コマンドを受信し、両方が、アドレスバス上に示されたそれぞれのアドレスから読み取られた 16 B のデータを用いて並列に応答する。マルチキャスト C A S イネーブル信号がローである場合、OR ゲート 712 の一度に 1 つだけがハイ出力を有し、一度に 1 つのグレインだけが C A S コマンドを受信することを可能にし、したがって通常のファイングレイン動作を提供する。

【0041】

この実施形態におけるバンク 700 は 2 つのグレインを有するが、デュアルモード I / O 回路 708 内のマルチレベルマルチプレクサツリーを含むいくつかの実施形態では、より高いグレインカウントが提供される。例えば、バンク 700 内に 8 つのグレインを有する 1 つの例示的な実施形態では、各グレインは、第 1 のモードにおいて 4 B の幅を有するファイングレインメモリアクセスを提供することができ、一方、第 2 のモードにおいて並列に動作する 8 つのグレイン全てが 32 B の幅を有するアクセスを提供する。そのような実施形態におけるデュアルモード I / O 回路 708 は、3 つのグレインビットを受信する 3 レベルマルチプレクサツリーを含む。第 1 のモードでは、マルチプレクサツリーは、アドレス指定されている適切なグレインに C A S コマンドをルーティングし、8 つの列デコードのうち 1 つのみをアクティブ化させる。第 2 のモードでは、デュアルモード I / O 回路は、8 つ全ての列デコードをイネーブルすることによって、マルチキャスト C A S コマンドに応じて、C A S コマンドを各グレインにおいてアサートさせる。

【0042】

図 8 は、いくつかの追加の実施形態による、ファイングレインメモリバンク 800 のブロック図である。メモリバンク 800 は、「グレイン 0」及び「グレイン 1」とラベル付けされた 2 つのグレインと、行デコード 802、2 つの列デコード 804 及び 806、各グレイン内の行バッファ（別々に図示せず）、「G r . S e l」とラベル付けされたグレイン選択回路 808、「m C A S リレー」とラベル付けされたマルチキャスト C A S リレー回路 810、コマンドバス 820 並びにアドレスバス 822 を含む I / O 回路と、を含む。

【0043】

この実施形態では、デュアルモード I / O 回路は、図 7 のようなマルチプレクサツリー実装ではなく、マルチキャスト C A S リレー回路 810 として具現化される。バンクごとに多数のグレイン（例えば、8 又は 16 等のグレインの数「G」）が存在する場合、マルチプレクサツリー実装の追加の深さは、クリティカルパスコマンドレイテンシの増加を引き起こす。マルチプレクサツリーを使用する代わりに、図示した実施形態は、マルチキャスト C A S リレー回路 810 を用いて、C A S 信号の局所化された非同期反復（リレー）を実行する。マルチキャスト C A S リレー回路 810 は、グレイン 0 の列デコード 804 と同じ場所に配置され、m C A S イネーブル信号を受信する第 1 の入力と、グレイン 1 の列デコードに接続された出力と、を含む。3 つ以上のグレインを有する実施形態では、m C A S を他のグレインに転送するために、複数のグレインのための I / O 回路内にマルチキャスト C A S リレー回路 810 が存在する。マルチキャスト C A S リレー回路 810 の

10

20

30

40

50

構造は、そのより短い電気導体に起因して、マルチプレクサツリ実装よりも低いキャパシタンスを有し、したがって、多くのグレインがある場合にコマンドレイテンシをあまり増加させない。

【 0 0 4 4 】

動作中、メモリバンク 8 0 0 は、メモリコントローラからコマンドバスを介して送信される図示された列アドレスストローブ (C A S) コマンド等のコマンドでアクセスされる。図示されるように、アドレスバスは、図 4 のアドレス方式に従って、列アドレスビット (図 4 の C O) を列デコーダ 8 0 4 及び 8 0 6 に運び、グレインビット (図 4 の G R) をグレインセレクト回路 8 0 8 に運ぶ。第 1 のモードでは、メモリアクセスは、1 つのグレインの幅であり、この例では 1 6 B である。第 2 のモードでは、メモリアクセスは、2 つ以上のグレイン (この例では 2 つ、合計 3 2 B) の幅である。第 2 のモードは、メモリコントローラからマルチキャスト C A S リレー回路 8 1 0 に供給される m C A S イネーブル信号によってアクティブにされる。第 2 のモードでは、マルチキャスト C A S リレー回路 8 1 0 は、マルチキャスト C A S コマンド内の初期グレイン識別子値を他のグレインを識別する新しい値に置き換えることによって、マルチキャスト C A S コマンドのローカライズされたリレーを、少なくとも 2 つのグレインの別の 1 つに転送するように第 2 のモードで実行する。いくつかの実施形態では、マルチキャスト C A S リレー回路 8 1 0 は、ルックアップテーブル又は加算器を使用して新しい値を識別する。

10

【 0 0 4 5 】

デュアルモード I / O 回路のための 2 つの異なる実施形態が図 7 及び図 8 に関して説明されるが、他の好適なデジタル論理方式が第 1 のモード及び第 2 のモードにおいて動作するステップを実行するために、メモリにおいて実装されてもよい。

20

【 0 0 4 6 】

図 9 は、いくつかの実施形態による、ファイングレインメモリを動作させるためのプロセスのフロー図 9 0 0 である。この実施形態では、プロセスは、キャッシュミスの場合にメインメモリへの 3 2 B 要求を使用する「 L D 」コマンドと、メインメモリからではなくオンチップスクラッチパッドからサービス可能な不規則なロードサイズ「 x 」を指定する「 L D _ x 」コマンドと、単一の 1 6 B 要求を介してメインメモリからサービス可能なサイズ x の不規則なロードを指定する「 L D _ x _ M E M 」コマンドと、を含む、右側に列挙されているメモリロード (読み取り) コマンドの 3 つの変形形態を使用する。

30

【 0 0 4 7 】

ブロック 9 0 2 に示されるように、プロセスは、選択された L D メモリアクセス要求が、メインメモリからの 1 6 B フェッチを引き起こさなければならない不規則な要求である場合、選択された L D メモリアクセス要求を L D _ x _ M E M 要求としてタグ付けすることを含む。ブロック 9 0 2 におけるタグ付けは、メモリアクセスコマンドが発生する元のソフトウェアを生成するプログラマによって、又は、指定されたメモリシステムのためのプログラミング命令を解釈するコンパイラ若しくはジャストインタイムコンパイラによって実行され得る。

【 0 0 4 8 】

ブロック 9 0 4 において、メモリコントローラは、例えば、プログラム又はライブラリがブロック 9 0 2 においてタグ付けされた命令を実行する場合に、タグ付けされたメモリアクセス要求に関連付けられたメモリアクセス要求を受信する。この時点での要求は、図示したように 1 6 B 等の第 1 のデータ幅、又は、図示したように 3 2 B 等の第 2 のデータ幅を有することができる。ブロック 9 0 6 において、プロセスは、要求が第 1 のデータ幅、例えば 1 6 B を有するかどうかを判定する。そうである場合、プロセスはブロック 9 0 8 に進み、そこで、1 6 B 幅のファイングレインメモリアクセスで要求に対応するために、要求にグレインに依存しないアドレスインターリーピングを適用する。次に、メモリコントローラは、C A S コマンドを 1 6 B 幅のメモリグレインに発行して、コマンドを実行する。上述したように、この要求は、第 1 のモードにおいて、それぞれのメモリバンク内のメモリの単一グレインへのアクセスによって対応される。ブロック 9 1 0 で 1 6 B C

40

50

A S に対してアクティブ化が発行されると、メモリコントローラは、16bの8バーストでメモリから応答データを受信する。

【0049】

再びブロック906を参照すると、要求が16B要求でない場合、プロセスはブロック912~914に進み、上述したように第2のモードで動作するデュアルモードI/O回路を使用して要求に対応する。ブロック912において、32B要求は、例えば、図5に示されるようなグレインのストライプ配置に従って互いに関連付けられた2つの16Bサブ要求に分割される。ブロック914において、そのようなグレインストライプ方式のためのアドレスインターリーブが2つの16Bコマンドに適用され、32BマルチキャストCASコマンドがメモリコントローラから発行されて、メモリからのコマンドの両方を実行する。マルチキャストCASコマンドに応じて、メモリは、少なくとも2つのグレイン内の選択された列に対して列アドレスストローブを並列にアサートさせることによって、要求に対応する。例えば、708(図7)のデュアルモードI/O回路、又はマルチキャストリレー回路808(図8)は、第2のモードで動作して、マルチキャストCASコマンドを複数のメモリグレインにおいてアサートさせる。得られたデータは、ブロック916に示すように、16B幅の2つの並列バーストでメモリコントローラに戻される。また、メモリコントローラは、受信されたデータの照合を実行して、元のメモリ要求に対する応答に適切な順序のデータを提供することができる。

【0050】

図10は、いくつかの実施形態による、ファイングレインメモリを動作させるためのライブラリ命令を含むデータ処理システム1000のブロック図である。データ処理システム1000は、オペレーティングシステム(operating system、OS)1002と、メモリコントローラ1008と、メモリ1010と、を含む。オペレーティングシステム1002は、データ処理システムの1つ以上のプロセッサ上で実行し、命令セットアーキテクチャ(instruction set architecture、ISA)コマンドライブラリ1004及びジャストインタイム(just-in-time、JIT)コンパイラ1006を含む。

【0051】

いくつかの実施形態では、拡張されたISAコマンドライブラリ1004は、ISAコマンドライブラリ1004内の省略記号によって示されるような一般的に使用されるISAライブラリ内の標準LDコマンド及び他のコマンドに加えて、図9に関して上述したようなロードコマンドLD__x及びLD__x__MEMを含む。追加のISAコマンドは、本明細書で説明する2つのモードでファイングレインメモリにアクセスすることをサポートするために、異なる方法で使用され得る。追加のコマンドは、オペレーティングシステム1002に公開され、オペレーティングシステム1002において実行中のプログラムによる使用のために利用可能である。オペレーティングシステム1002は、オプションでJITコンパイラ1006を含み、2つのモードでファイングレインメモリアクセスを可能にするために、拡張されたISAコマンドセットよりもむしろJITコンパイラ1006が使用され得る。

【0052】

異なるモード、例えば、16B及び32Bアクセスの上述した2つのモードでOS支援メモリ(OS-assisted memory)ファイングレインメモリアクセスを提供するために、プログラムは、通常のLDコマンド又はLD__x__MEM及びLD__xコマンドを呼び出すプログラムをコンパイルすることができる。他の実施形態では、新しいISA命令の代わりに、デュアルモードファイングレインアクセスがOSに公開され、その結果、プログラムは、不規則なアクセスを被ることになるメモリ領域に注釈を付けることができる。そのような実施形態では、プログラムは、新しい「malloc」(メモリ割振り)命令を使用してグレインに好適な割振りを要求する。そのようなmalloc命令に応じて、OSは、対応する領域のページテーブルエントリにタグを付けて、アクセス粒度、すなわち、領域が16Bアクセスをサポートするか32Bアクセスをサポートするかを示す。プロセッサのロード及びストアは、メモリにアクセスする前に、プロセッサの変換索引バッファ

10

20

30

40

50

(T L B) 又は他のアドレス変換テーブルを調べるので、アクセス粒度タグは、各コマンドのアドレス変換から容易に取り出すことができる。このアクセス粒度タグは、好ましくは単一ビットとしてメモリコントローラに伝達され、関連するアクセスが有するのが第 1 の幅か第 2 の幅か、例えば 1 6 B か 3 2 B かを示す。

【 0 0 5 3 】

このような機構の重大な制限は、割り当てられたメモリ領域全体が、通常、同じアクセス粒度を有する必要があること、又は、メモリ領域が、実行全体を通して同じアクセス粒度を受ける必要があることである。これらの制限に対処するために、いくつかの実施形態は、プログラマが、コンパイラによって生成される L D 対 L D _ X _ M e m I S A 命令等の特定のアクセス粒度でオーバーライドされ得るアクセス粒度で、割り振られたメモリ領域をタグ付けするハイブリッド手法を採用する。例えば、アプリケーションは、3 2 B アクセス粒度領域として構造にタグ付けすることができるが、静的コンパイラ又は J I T コンパイラ 1 0 0 6 は、1 6 B アクセスをサポートすることができる L D _ X _ M e m I S A 命令を生成することができる。このような方式は、メモリ領域に対するアクセス粒度を、元のアプリケーションプログラミングによって割り振られたものから動的に切り替えることを可能にする。理解できるように、メモリ 1 0 1 0 は、メモリ領域のための I S A コマンド又は粒度情報を受信することに応じて、メモリコントローラ 1 0 0 8 からの命令に従って、上述したような 2 つのモードの何れかで動作され得るファイングレインメモリバンクを含む。

【 0 0 5 4 】

図 1 1 は、いくつかの実施形態による、ファイングレインメモリを動作させるためのプロセスのフロー図 1 1 0 0 である。ブロック 1 1 0 2 において、プログラムは、2 つ以上の粒度、例えば 1 6 B 及び 3 2 B の中からアクセス粒度を指定するメモリアクセスコマンドを用いて、説明したようにコンパイルされる。ブロック 1 1 0 4 では、コンパイルされたプログラムが O S によって実行される。ブロック 1 1 0 6 に示されるように、J I T コンパイラ 1 0 0 6 等のジャストインタイムコンパイラは、メモリにおいて利用可能なファイングレインアクセスに基づいて、それぞれの指定されたアクセス粒度のファイングレインコマンドを生成する。これらのコマンドは、メモリコントローラによって解釈されて、第 2 のアクセスモードが J I T コンパイラによって呼び出されるメモリアクセスを実施するためのマルチキャスト C A S コマンドを生成する。

【 0 0 5 5 】

図 2、図 3、図 7、図 8 の回路、又は、その任意の部分（アービタ 2 3 8 又はデュアルモード I / O 回路 7 0 8 等）は、データベースの形態のコンピュータアクセス可能なデータ構造又は集積回路を製造するためにプログラムによって読み取られ、直接的又は間接的に使用され得る他のデータ構造によって説明又は表示され得る。例えば、このデータ構造は、V e r i l o g 又は V H D L 等の高レベル設計言語（high level design language、H D L）におけるハードウェア機能の挙動レベル記述又はレジスタ転送レベル（register-transfer level、R T L）記述であり得る。記述は、合成ライブラリからゲートのリストを含むネットリストを生成するために記述を合成することができる合成ツールによって読み取ることができる。ネットリストは、集積回路を含むハードウェアの機能も表すゲートのセットを含む。ネットリストは、次に、マスクに適用される幾何学的形状を記述するデータセットを生成するために配置され、ルーティングされてもよい。次に、マスクを、様々な半導体製造工程で使用して、集積回路を製造してもよい。代替的に、コンピュータアクセス可能記憶媒体上のデータベースは、所望の場合、ネットリスト（合成ライブラリの有無にかかわらず）若しくはデータセット、又は、グラフィックデータシステム（Graphic Data System、G D S）I I データであってもよい。

【 0 0 5 6 】

特定の実施形態を説明してきたが、これらの実施形態に対する様々な修正が当業者には明らかであろう。例えば、メモリコントローラ 2 0 0 は、異なる実施形態において異なり得る。メモリコントローラ 2 0 0 は、高帯域幅メモリ（H B M）、R A M b u s D R A

10

20

30

40

50

M (R A M b u s D R A M、R D R A M) 等のように、D D R x 以外の他のタイプのメモリとインターフェースすることができる。図示した実施形態は、個別のD I M M 又はS I M M に対応するメモリの各ランクを示したが、他の実施形態では、各モジュールは複数のランクをサポートすることができる。更に他の実施形態は、ホストマザーボードに取り付けられたD R A M 等のように、特定のモジュールに含まれていない他のタイプのD R A M モジュール又はD R A M を含むことができる。したがって、添付の特許請求の範囲は、開示された実施形態の範囲に含まれる開示された実施形態の全ての変更を網羅することを意図している。

【要約】

ファイングレインダイナミックランダムアクセスメモリ (D R A M) は、第 1 のメモリバンクと、第 2 のメモリバンクと、デュアルモード I / O 回路と、を含む。第 1 のメモリバンクは、複数のグレインに分割されたメモリアレイを含み、各グレインは、行バッファ及び入力 / 出力 (I / O) 回路を含む。デュアルモード I / O 回路は、第 1 のメモリバンク内の各グレインの I / O 回路に結合され、第 1 のデータ幅を有するコマンドが各グレインにルーティングされ、各グレインにおいて個別に実行される第 1 のモードと、第 1 のデータ幅とは異なる第 2 のデータ幅を有するコマンドがグレインのうち少なくとも 2 つによって並列に実行される第 2 のモードと、で動作する。

【選択図】図 1

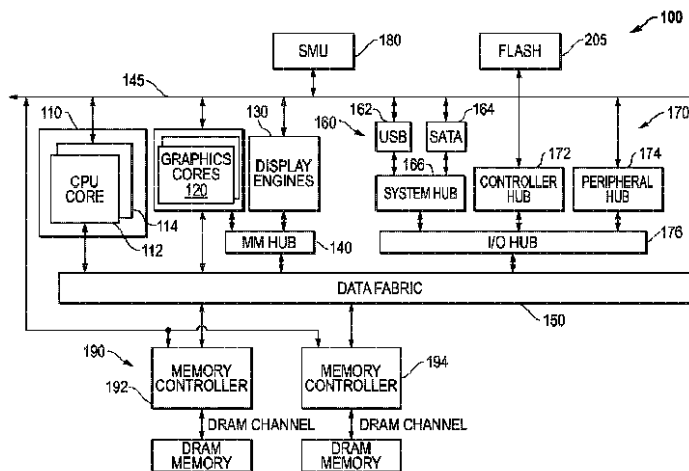
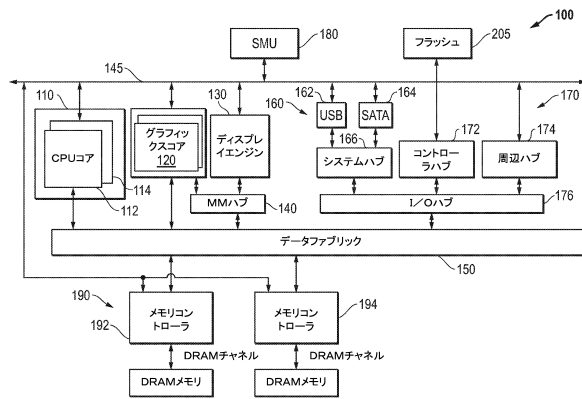


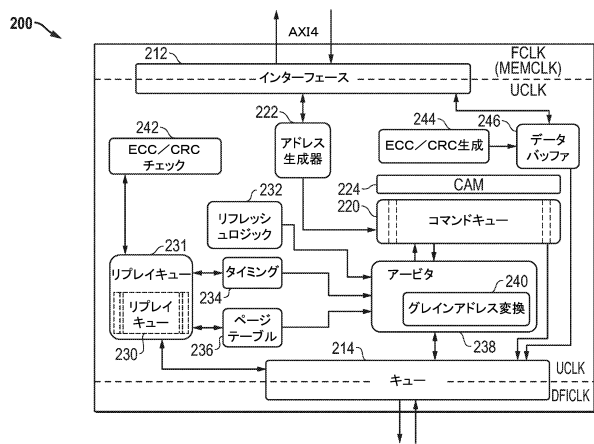
FIG. 1

【図面】

【 図 1 】

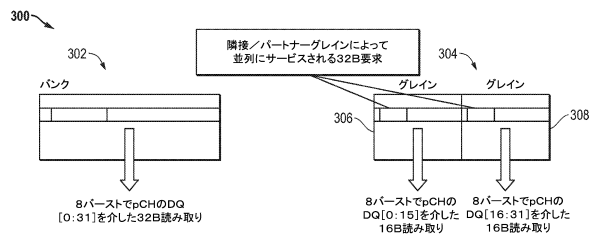


【圖 2】



10

【 図 3 】



【 図 4 】

29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	CO	CO	CO	<u>GR</u>	BK	BG	BG	BK	CO	PCH	CO	CO	X	X	X	X

20

30

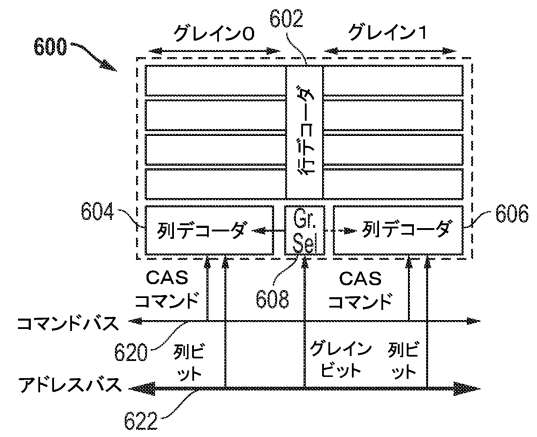
40

50

【 図 5 】

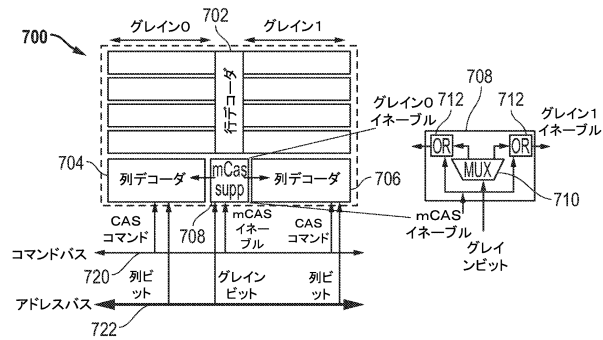
29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	CO	CO	CO	CO	BK	BG	BG	BK	CO	PCH	CO	<u>GR</u>	X	X	X	X

【 図 6 】

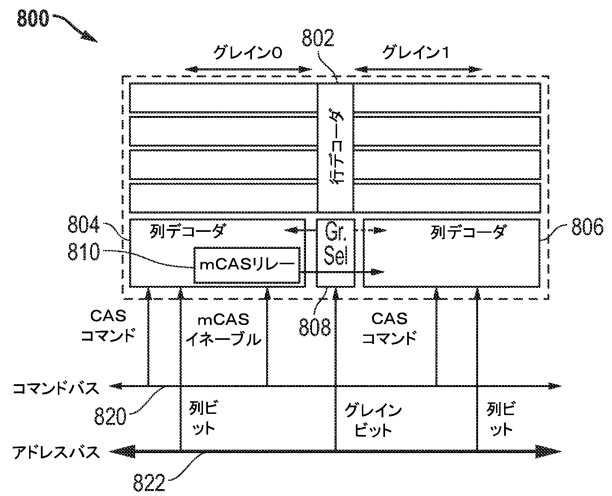


(先行技術)

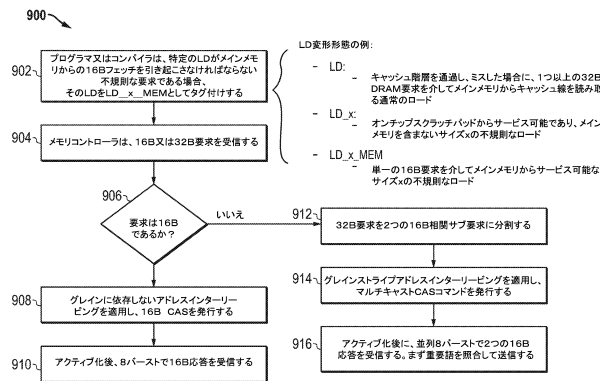
【圖 7】



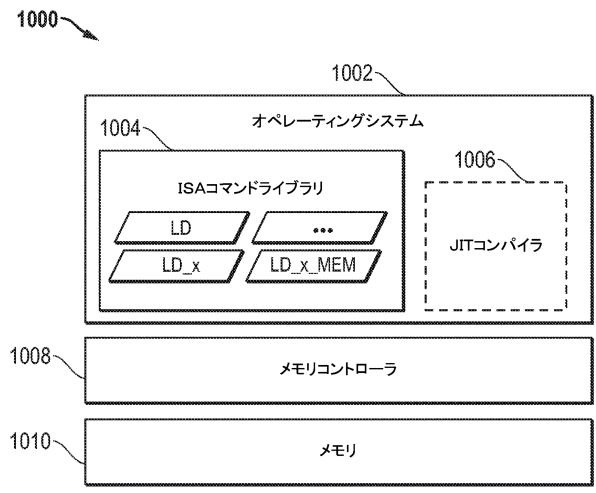
【 図 8 】



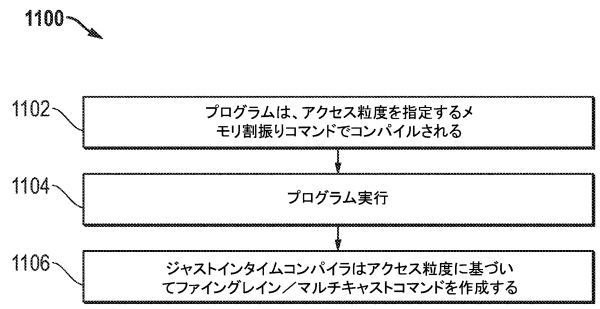
【図 9】



【図 10】



【図 11】



10

20

30

40

50

フロントページの続き

- アメリカ合衆国 7 8 7 3 5 テキサス州、オースティン、サウスウェスト パークウェイ 7 1 7
1、アドバンスト・マイクロ・ディバイシズ・インコーポレイテッド内
- (72)発明者 ヴィグネシュ アディナラヤナン
アメリカ合衆国 7 8 7 3 5 テキサス州、オースティン、サウスウェスト パークウェイ 7 1 7
1、アドバンスト・マイクロ・ディバイシズ・インコーポレイテッド内
- (72)発明者 ジャガディッシュ ビー . コトゥラ
アメリカ合衆国 7 8 7 3 5 テキサス州、オースティン、サウスウェスト パークウェイ 7 1 7
1、アドバンスト・マイクロ・ディバイシズ・インコーポレイテッド内
- (72)発明者 セルゲイ ブラゴドゥロフ
アメリカ合衆国 9 8 0 0 7 ワシントン州、ベルビュー、2 0 0 2 - 1 5 6 番アベニュー ノース
イースト、スイート 3 0 0、アドバンスト・マイクロ・ディバイシズ・インコーポレイテッド内
- 審査官 後藤 彰
- (56)参考文献 特開 2 0 2 0 - 1 8 7 7 4 7 (J P , A)
特開 2 0 2 0 - 8 7 4 7 7 (J P , A)
特表 2 0 0 9 - 5 3 5 7 4 8 (J P , A)
- (58)調査した分野 (Int.Cl. , D B 名)
G 0 6 F 1 2 / 0 6