

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4796580号
(P4796580)

(45) 発行日 平成23年10月19日(2011.10.19)

(24) 登録日 平成23年8月5日(2011.8.5)

(51) Int. Cl. F I
G 0 6 F 12/08 (2006.01)
 G 0 6 F 12/08 5 0 1 D
 G 0 6 F 12/08 5 0 5 B
 G 0 6 F 12/08 5 1 9 Z

請求項の数 6 (全 15 頁)

(21) 出願番号	特願2007-534139 (P2007-534139)	(73) 特許権者	504199127
(86) (22) 出願日	平成17年9月21日 (2005.9.21)		フリースケール セミコンダクター イン
(65) 公表番号	特表2008-515075 (P2008-515075A)		コーポレイテッド
(43) 公表日	平成20年5月8日 (2008.5.8)		アメリカ合衆国 78735 テキサス州
(86) 国際出願番号	PCT/IB2005/053109		オースティン ウィリアム キャノン
(87) 国際公開番号	W02006/035370		ドライブ ウェスト 6501
(87) 国際公開日	平成18年4月6日 (2006.4.6)	(74) 代理人	100089705
審査請求日	平成20年4月10日 (2008.4.10)		弁理士 社本 一夫
(31) 優先権主張番号	10/955,220	(74) 代理人	100140109
(32) 優先日	平成16年9月30日 (2004.9.30)		弁理士 小野 新次郎
(33) 優先権主張国	米国 (US)	(74) 代理人	100075270
			弁理士 小林 泰
		(74) 代理人	100080137
			弁理士 千葉 昭男

最終頁に続く

(54) 【発明の名称】 フェッチ・バーストを用いて情報をキャッシュ・モジュールに与える装置及び方法

(57) 【特許請求の範囲】

【請求項1】

情報をキャッシュ・モジュールに与える装置であって、
 キャッシュ・ミスに繋がる2つのデータ要求の同時受信に応じて、第1及び第2の必須のデータ単位の検索を要求する指示を発生するよう適合されたキャッシュ論理と、
 前記のキャッシュ論理に結合された制御器であって、前記指示を受け取り、それに対応して、

第1の必須のデータ単位及び第2の必須のデータ単位のフェッチングを含む少なくとも1つのフェッチ・バーストを開始する前に、当該第1の必須のデータ単位及び第2の必須のデータ単位が、単一のフェッチ・バースト中にフェッチされ得るかを決定し、

単一のフェッチ・バースト中に検索可能なメモリ空間が前記第1及び第2の必須のデータ単位を備える場合、前記単一のフェッチ・バーストを開始し、

単一のフェッチ・バースト中に検索可能なメモリ空間が前記第1及び第2の必須のデータ単位を備えていない場合、複数のフェッチ・バーストを開始するよう適合されている前記制御器とを備える装置。

【請求項2】

フェッチ・バースト中に検索可能なメモリ空間の少なくとも1つの特性が、前記メモリ空間の場所に応答する請求項1記載の装置。

【請求項3】

10

20

前記制御器が更に、投機型フェッチ・オペレーションを開始するよう適合されている請求項 1 記載の装置。

【請求項 4】

情報をキャッシュ・モジュールに与える装置であって、

前記キャッシュ・モジュールに結合され、前記キャッシュ・モジュールから第 1 及び第 2 のデータ単位を検索する第 1 及び第 2 の要求を開始する少なくとも 1 つのプロセッサと

、
前記要求を受け取り、キャッシュ・ミスに繋がる 2 つのデータ要求の同時受信に応じて
前記第 1 及び第 2 のデータ単位が必須のデータ単位であるかどうかを決定するよう適合
された論理と、

10

前記キャッシュ・モジュールに結合された制御器であって、

第 1 の必須のデータ単位及び第 2 の必須のデータ単位のフェッチングを含む少なくと
も 1 つのフェッチ・バーストを開始する前に、当該第 1 の必須のデータ単位及び第 2 の必
須のデータ単位が、単一のフェッチ・バースト中にフェッチされ得るかを決定するよう適
合され、

単一のフェッチ・バースト中に検索可能なメモリ空間が前記第 1 及び第 2 の必須のデ
ータ単位を備える場合前記単一のフェッチ・バーストを開始するよう適合され、

且つ単一のフェッチ・バースト中に検索可能なメモリ空間が前記第 1 及び第 2 の必須
のデータ単位を備えていない場合複数のフェッチ・バーストを開始するよう適合されてい
る前記制御器と、

20

を備える装置。

【請求項 5】

キャッシュ・モジュールを更に備える請求項 4 記載の装置。

【請求項 6】

フェッチ・バースト中に検索可能なメモリ空間の少なくとも 1 つの特性が、前記メモリ
空間の場所に応答する請求項 4 記載の装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報をキャッシュ・モジュールに与える装置及び方法に関し、特に、複数の
データ・バスに接続されたマルチポート・キャッシュ・モジュールに対してデータを検索
する方法及び装置に関する。

30

【背景技術】

【0002】

キャッシュ・モジュールは、データの高速検索を容易にする高速メモリである。典型的
には、キャッシュ・モジュールは、比較的高価であり、そして特に外部メモリと比較して
大きさが小さいことにより特徴付けられる。

【0003】

最新のプロセッサベースのシステムの性能は、通常、キャッシュ・モジュール性能に依
存し、そして特にキャッシュ・ヒットとキャッシュ・ミスとの関係に依存する。キャッシ
ュ・ヒットは、キャッシュ・モジュール・メモリに存在するデータが要求されるとき起
こる。キャッシュ・ミスは、要求されたデータがキャッシュ・モジュールに存在しないで、
別の（通常外部の）メモリからフェッチされねばならないとき起こる。

40

【0004】

様々なキャッシュ・モジュール及びプロセッサ・アーキテクチャ、並びにデータ検索ス
キームは、ますます増大する性能の要求を満たすよう長年にわたって開発されたきた。こ
れらのアーキテクチャは、マルチポート・キャッシュ・モジュール、マルチレベル・キャ
ッシュ・モジュール・アーキテクチャ、スーパー・スケーラ・タイプ・プロセッサ及び類
似のものを含む。

【0005】

50

以下の米国特許及び特許出願は、全て本明細書に援用されるが、これらは、キャッシュ・モジュール及びデータ・フェッチ方法の幾らかの現状技術水準の概要を与える。以下の以下の米国特許及び特許出願とは、発明の名称が「異なるプロセッサに関する別々のディレクトリ及びプリフェッチを用いて単一のキャッシュ制御をデュアル・チャネルの中へ仮想化する論理を有するバス・エキスパンダ (Bus expander with logic for virtualizing single cache control into dual channels with separate directories and prefetch for different processors)」であるJohnson他の米国特許No. 48538746、発明の名称が「固有のトランザクション衝突回避を有するパイプライン非ブロッキング・レベル2キャッシュ・システム (Pipelines non-blocking level two cache system with inherent transaction collision-avoidance)」であるRichardson他の米国特許出願No. 20020069326、発明の名称が「マルチウェイ・マルチポート・キャッシュを有する並列プロセッサ・システムにおける同一で同時のアクセスに関する検出回路 (Detection circuit for identical and simultaneous access in a parallel processor system with a multi-way multi-port cache)」であるKawasakiの米国特許No. 5,742,790、発明の名称が「マルチポート化された非ブロッキング・キャッシュにおけるインライン・バンク競合検出及び解法 (In-line bank conflict detection and resolution in a multi-ported non-blocking cache)」であるHetherington他の米国特許No. 6,081,873、及び発明の名称が「デュアルポート化されパイプライン化された2レベル・キャッシュ・システム (Dual-ported, pipelined, two level cache system)」である米国特許No. 6,272,597である。

【発明の開示】

【発明が解決しようとする課題】

【0006】

典型的には、データをキャッシュ・モジュールから要求するプロセッサは、それが要求されたデータを受け取るまで機能停止又は停止される。複数のキャッシュ・ミス事象が同時に生じるとき、要求するプロセッサは、複数のフェッチ・オペレーションと関連した待ち時間に起因して、延長された期間停止されることができる。

【0007】

従って、複数のキャッシュ・ミス事象を効率的に処理することができるシステム及び方法を与える必要性が存在する。

【課題を解決するための手段】

【0008】

本発明は、フェッチ・バーストを実行し、そしてフェッチ・バースト量、並びにフェッチ・バーストの大きさを決定することが可能である装置及び方法を提供する。

【0009】

本発明は、キャッシュ・ミス事象を同時に起こすデータ単位の場所に応答してフェッチ・バースト量を決定する装置及び方法を提供する。

【0010】

本発明は、キャッシュ・ミス事象を同時に起こした2つ又はそれより多いデータ単位が異なるアドレスを有するときでさえフェッチ・バースト量を低減することができる装置及び方法を提供する。

【0011】

本発明は、キャッシュ・ミス事象を起こした第1及び第2のデータ単位が事前定義された大きさのメモリ空間に属する場合フェッチ・バースト量を低減することができる装置及

10

20

30

40

50

び方法を提供する。メモリ空間の大きさは、1又はそれより多いフェッチ・バースト中にフェッチされることができるデータ量に等しくすることができる。

【0012】

本発明は、情報をキャッシュ・モジュールに与える装置であって、(i)前記キャッシュ・モジュールに結合され、前記キャッシュ・モジュールから第1及び第2のデータ単位を検索する第1及び第2の要求を開始する少なくとも1つのプロセッサと、(ii)前記要求を受け取り、前記第1及び第2のデータ単位が必須のデータ単位であるかどうかを決定するよう適合された論理と、(iii)前記キャッシュ・モジュールに結合された制御器であって、事前定義された大きさのメモリ空間が前記第1及び第2の必須のデータ単位を含む場合前記単一のフェッチ・バーストを開始するよう適合され、且つそのようなメモリ空間が前記第1及び第2の必須のデータ単位を含まない場合複数のフェッチ・バーストを開始するよう適合されている前記制御器とを含む前記装置を提供する。

10

【0013】

事前定義された大きさのメモリ空間は、1又はそれより多いフェッチ・バースト中に検索可能なメモリ空間であることが都合良い。

【0014】

本発明は、情報をキャッシュ・モジュールに与える方法であって、(i)複数の情報単位をキャッシュ・モジュールに対して検索する複数の要求を受け取るステップと、(ii)フェッチ・バースト中に検索可能なメモリ空間と複数の情報単位の場所との関係を決定するステップと、(iii)前記の決定に応答して少なくとも1つのフェッチ・バーストを実行するステップとを含む前記方法を提供する。

20

【0015】

本発明は、情報をキャッシュ・モジュールに与える方法であって、(i)第1及び第2の必須のデータ単位の検索を要求する指示を受け取るステップと、(ii)事前定義された大きさのメモリ空間が第1及び第2の必須のデータ単位を備える場合単一のフェッチ・バーストを実行するステップと、(iii)その事前定義された大きさのメモリ空間が第1及び第2の必須のデータ単位を含まない場合、複数のフェッチ・バーストを実行するステップとを含む前記方法を提供する。

【0016】

本発明は、情報をキャッシュ・モジュールに与える装置であって、(i)第1及び第2の必須のデータ単位の検索を要求する指示を発生するよう適合された論理と、(ii)前記のキャッシュ論理に結合された制御器であって、前記指示を受け取り、それに応答して、事前定義された大きさのメモリ空間が前記第1及び第2の必須のデータ単位を含む場合前記単一のフェッチ・バーストを開始し、且つそのメモリ空間が前記第1及び第2の必須のデータ単位を含まない場合複数のフェッチ・バーストを開始するよう適合されている前記制御器とを含む前記装置を提供する。

30

【0017】

本発明は、添付図面と関係した以下の詳細な説明からより完全に理解され、認められるであろう。

【発明を実施するための最良の形態】

40

【0018】

以下の説明は、複数のデータ・フェッチ・オペレーション及びデータ・キャッシュ・モジュールに関連する。当業者は、開示されたシステム及び方法が命令検索、命令キャッシュ・モジュール、そしてデータ検索及び命令検索の組み合わせにさえ、更に命令及びデータの両方を格納するキャッシュ・モジュールに、必要に応じて変更を加えて、適用することができることを認めるであろう。

【0019】

本装置及び方法は、キャッシュ・ミス事象を起こした第1及び第2のデータ単位の場所に応答してフェッチ・バースト量を決定する。両方が事前定義された大きさのメモリ空間内に配置されている場合、フェッチ・バースト量が、低減されることができる。事前定義

50

された大きさは、単一のフェッチ・バースト中に検索されることができるデータ量に対応することが都合良い。これは、必ずしもそうであるわけではなく、そしてこの事前定義された大きさは、より大きくすることができる。例えば、この大きさは、キャッシュ・モジュール・ラインの大きさに対応させることができる。説明の便宜のため、この大きさは、単一のフェッチ・バースト中に検索されることができるデータ量に対応すると仮定し、そしてメモリ空間は、単一のフェッチ・バースト中に検索可能なメモリ空間と呼ばれる。

【0020】

図1は、本発明の一実施形態に従った装置100を示す。装置100は、プロセッサ110、データ・チャンネル130、メモリ管理ユニット(MMU)300、命令チャンネル340、レベル1RAMメモリ370、並びにインターフェース・ユニット380を含む。装置100は、高レベル・メモリ・モジュール50のような追加のメモリ・モジュールにシステム・バス60を介して接続される。

10

【0021】

プロセッサ110及び命令チャンネル340は、単一のプログラム・バス120に接続される。プロセッサ110は、第1のデータ・ポート116及び第2のデータ・ポート118を有する。第1のデータ・ポート116は、第1のデータ・バス(XA)122を介してデータ・チャンネル130の第1のポート132、MMU300、レベル1RAMメモリ370に接続される。第2のデータ・ポート118は、第2のデータ・バス(XB)124を介してデータ・チャンネル130の第2のポート134、MMU300及びレベル1RAMメモリ370に接続される。

20

【0022】

データ・チャンネル130は、データ・フェッチ・バス126を介してインターフェース380に接続され、当該インターフェース380は、次いで、高レベル・メモリ・モジュール50に接続される。追加のメモリが、マルチレベル・キャッシュ・アーキテクチャの一部であることができるのに対し、データ・キャッシュ・モジュール200が、第1レベル・キャッシュ・モジュールであり、そして追加のメモリが、レベル2キャッシュ・メモリであることに注目されたい。追加のメモリはまた、外部メモリの一部であることができ、この外部メモリはまた、メイン・メモリと呼ばれる。

【0023】

MMU300は、プログラム及びデータ・ハードウェア保護を与え、そして仮想アドレスの物理アドレスへの高速変換を実行するよう適合されている。MMU300はまた、様々なキャッシュ及びバス制御信号を与えることが可能である。仮想アドレスは、プロセッサ110により発生されるアドレスであり、そしてプロセッサ110により実行されるコードにより見られるようにされている。物理アドレスを用いて、様々なメモリ・バンクにアクセスする。

30

【0024】

データ・チャンネル130は、データ・キャッシュ・モジュール200、データ・フェッチ・ユニット(DFU)170のような複数の支援ユニット、複数の追加のバッファ(図示せず)及びデータ制御ユニット(DCU)150を含む。DFU170は、データのフェッチ及びプリフェッチを担当する。データ・フェッチ・オペレーションは、必須型フェッチ・オペレーション(mandatory fetching operations)及び投機型フェッチ・オペレーション(speculated fetching operations)を含むことができる。必須型フェッチ・オペレーションは、キャッシュ・ミスを起こしたデータ単位を検索することを含む。投機型フェッチ(これはまたプリフェッチと呼ばれる。)・オペレーションは、キャッシュ・ミスを起こさなかったデータ単位を検索することを含む。通常、この後者のタイプのデータは、プリフェッチ後直ぐに用いられることを期待される。この期待は、通常、多くのデータ要求が性質上順次的であるという仮定に基づく。

40

【0025】

各フェッチ・オペレーションが単一の基本データ単位(BDU)をフェッチすることを

50

含む。従って、必須型フェッチ・オペレーション中にフェッチされるBDUは、必須BDU (mandatory BDU) と呼ばれ、そして投機型フェッチ・オペレーション中にフェッチされるBDUは、投機BDU (speculated BDU) と呼ばれる。

【0026】

DCU150は、DFU170及び追加のバッファにより発生された、フェッチ・データ・バス126に対するデータ・アクセス要求を調停することを担当する。典型的には、DFU170から発信するデータ・フェッチ要求は、最高の優先順位を与えられる。

【0027】

プロセッサ110は、バスXA122及びXB124を介して2つのデータ要求を同時に発行することが可能である。データ・チャンネル130は、これらの要求を処理して、1又はそれより多いキャッシュ・ヒットが生じたかどうかを決定する。基本的には、データ・チャンネル130は、2つのデータ要求がキャッシュ・ヒットをもたらしたことと、両方の要求がキャッシュ・ミスをもたらしたことと、1つの要求がキャッシュ・ヒットをもたらす一方、他方の要求がキャッシュ・ミスをもたらしたことのいずれかであることを決定することができる。

【0028】

本発明の一実施形態に従って、プロセッサ110は、それが受け取る全てのデータがそれが要求した全てのデータであるまで停止されるが、しかしこれは、必ずしもそうであるわけではない。例えば、本発明の別の実施形態に従って、プロセッサのほんの一部のみが、停止される。

【0029】

停止するステップを開始し、終了する様々な方法がある。キャッシュ・ミスは、そのようなステップへの入ることをトリガすることができる。プロセッサ110は、ひとたびそれがキャッシュ・ミス指示をデータ・チャンネル130から受け取ると、停止されるステップに入ると仮定する。プロセッサ110は、ひとたびそれが要求されたデータが使用可能である指示をデータ・チャンネル130から受け取ると、停止ステップを出る。プロセッサ110とデータ・チャンネル130間を接続するライン302は、プロセッサ110が停止ステップに入りそしてそのようなステップを出るようにする停止信号を搬送する。プロセッサ110が、データ要求のどれがキャッシュ・ミスを起こしたかのような追加の情報を受け取ることができることに注目されたい。

【0030】

図2は、本発明の一実施形態に従ったデータ・キャッシュ・モジュール200の概略図である。データ・キャッシュ・モジュール200は、キャッシュ論理210のような論理、及びキャッシュ・メモリ・バンク250を含む。キャッシュ・メモリ・バンク250は、256個のライン250(0) - 250(255)を含み、それぞれのラインは、16個の128ビット長基本データ単位)を含む。これらの基本データ単位(BDU)は、252(0,0) - 252(255,15)により表示される。キャッシュ・ヒット又はキャッシュ・ミスは、BDUベースで決定される。上記論理は、キャッシュ・モジュールの外側に配置されることができ、これは必ずしもそうであるわけではないことに注目されたい。

【0031】

図3は、本発明の一実施形態に従ったキャッシュ論理210の概略図である。キャッシュ論理210は、2つのデータ要求を同時に管理することが可能であり、そして2つの同一の部分212及び214を含み、各部分は、単一のキャッシュ・ヒット又はキャッシュ・ミスのいずれが生じたかを決定することが可能である。説明を簡単にするため、キャッシュ論理210の第1の部分212が、詳細に示されている。キャッシュ論理210は、ウェイ0 - ウェイ7 220(0) - 220(7)で参照される8個のウェイを含む。各ウェイは、32個のラインと関連したアドレス及びステータス情報を格納する。アドレス情報は、タグ・アドレスを含み、そしてステータス情報は、BDU有効度及び更新情報を含む。説明を簡単にするため、ウェイ1 220(0)のみが詳細に示され、一方他のウェイ

10

20

30

40

50

は、ボックス 220(1) - 220(7) により表される。

【0032】

各ラインは、拡張されたタグ値、及びそのライン内の各 BDU の有効度を表す 16 個の有効度ビットと関連付けられる。ウェイ 0 220 は、16 個の拡張されたタグ・アドレス 220(0) - 220(15)、並びに 1 セットが 16 個の BDU 有効度フラグから成る 16 セット 220(0) - 220(15, 15) を格納する。

【0033】

各 BDU はまた、BDU がより高いレベル・メモリにおいて更新されること無しに変更されたかどうかを指示するダーティ・ビットと関連付けられることができる。

【0034】

ひとたびプロセッサ 110 が 32 ビット・アドレス 400 を第 1 のデータ・バス XA122 を介して与えると、キャッシュ論理 210 の第 1 の部分 212 は、このアドレスを処理して、要求されたデータがキャッシュ・モジュールに格納された(キャッシュ・ヒット)かされない(キャッシュ・ミス)かを決定する。キャッシュ・ヒットが生じた場合、要求されたデータは、XA122 又は XB124 からの適切なデータ・バスを介してプロセッサ 110 に送られる。さもなければ、DFU170 は、キャッシュ・ミスについて通知される。

【0035】

32 ビット・アドレス 400 は、アドレス 400 の 12 個の最上位ビットを含む 12 ビット・タグ・アドレス 402、4 ビット・ライン・インデックス 404、BDU オフセット 405 及び 4 ビット・バイト・オフセット 408 に仕切られる。4 ビット・バイト・オフセット 408 は、キャッシュ・メモリ・バンク 250 からのデータ検索に用いられる。

【0036】

ウェイ 220(0) 内に格納された 16 個のタグ・アドレス 220(0) - 220(15) のそれぞれは、タグ・アドレス 402 と並列に比較される。当業者は、そのような比較があらゆる方法で並列に行われることを認めるであろう。

【0037】

その上、BDU オフセット 405 及び 4 ビット・ライン・インデックス 404 を用いて、要求された BDU に対応する有効度フラグを検索する。4 ビット・ライン・インデックス 404 は、1 セットの BDU 有効度フラグを 16 セットのウェイ 0 220(0) から選択するのに用いられ、一方 4 ビット BDU オフセット 405 は、有効度フラグをその選択されたセットの BDU 有効度フラグから選択するのに用いられる。

【0038】

キャッシュ・ヒットは、格納されたタグ・アドレスのうちの 1 つとタグ・アドレスとの間にマッチ(一致)がある場合で、且つ選択された BDU が有効である場合生じる。

【0039】

DFU170 はキャッシュ・ヒット及びキャッシュ・ミスの指示を受け取る。両方のデータ要求がキャッシュ・ヒットをもたらした場合、DFU170 は、必須のフェッチを実行するよう要求されない。データ要求のうちの唯 1 つの要求がキャッシュ・ミスをもたらした場合、DFU170 は、単一の必須のフェッチを実行するよう要求される。両方のデータ要求がキャッシュ・ミスをもたらした場合、DFU170 は、1 又はそれより多い必須のフェッチを実行するよう要求される。

【0040】

本発明の一実施形態に従って、DFU170 は、必須型フェッチ・オペレーションに影響を与え得るフェッチ特性を受け取る。このフェッチ特性は、フェッチ・プロセスに影響を与え得る事前定義された大きさのメモリ空間(例えば、1 又はそれより多いフェッチ・バーストで検索されることが出来るデータ量、フェッチ・バーストの大きさ、及び類似のもの)を含む。

【0041】

これらの特性は、その上でデータ転送が生じるバスの大きさ、外部メモリ・バンクの構

10

20

30

40

50

成及び類似のものを表すことができる。例えば、DRAMメモリは、典型的には、DRAM行に配列される。DRAM行全体のコンテンツは、単一のフェッチ・オペレーションによりフェッチされることができ、従って1ライン全体のコンテンツをフェッチすることは、有効であることができる。装置100は、各バスがそれ自身のフェッチ特性を有する複数のバスを介して複数のメモリ・バンクに接続されることができると注目されたい。

【0042】

データ・フェッチ・バス126は、1フェッチ・オペレーションにつき単一のBDUをフェッチすることを可能にする。典型的なフェッチ・バーストは、4つの連続したフェッチ・オペレーションを含み、従って4つのBDUの合計は、単一のフェッチ・バースト中に検索されることができ、1フェッチ・バースト当たりのフェッチ・オペレーションの大きさは、4以外でも可能である。

10

【0043】

典型的には、フェッチ・バーストを実行するよう適合されているメモリ・モジュールは、固定の大きさのデータ単位セットに仕切られている。或る一定のデータ単位を受け取る要求を含むフェッチ・バーストは、そのセットの検索になる。フェッチされたデータ単位のオーダは、特定の要求されたデータ・セットに依存する。

【0044】

通常、単一のフェッチ・バーストは、たとえメモリ空間がラップアラウンド(循環)オペレーションを適用することにより与えられても、連続したメモリ空間を検索することができる。2つの必須のBDUが単一の連続したメモリ空間内に含まれる場合、2つのフェッチ・バーストがより多くの時間を消費するので、2つのフェッチ・バーストではなく単一のフェッチ・バーストを実行することは非常に有益である。

20

【0045】

これは、特に、フェッチ・バーストに割り込むことを可能にしないシステムで真である。この効率的なフェッチング・スキームは、特にプロセッサ110がそれが両方の必須のBDUを受け取るまで停止されるので、プロセッサ停止期間を低減することができる。

【0046】

ひとたび2つのキャッシュ・ミスが同時に生じると、DFU170は、両方の必須のBDUが単一のフェッチ・バースト中にフェッチされることができかどうかを決定しなければならない。その答えが肯定である場合、そのようなフェッチ・バーストが開始される。ひとたびフェッチ・バーストが終わると、必須のBDUがプロセッサ110に送られ、プロセッサ110は、停止状態を出ることができる。

30

【0047】

本発明の一実施形態に従って、フェッチ・バーストは、或る一定のデータ単位セットに属する4つの隣接データ単位の4つのフェッチ・オペレーションを含む。換言すると、単一のフェッチ・バーストは、連続したメモリ空間を検索するのに利用することができる。従って、DFU170は、両方の必須のBDUが単一の連続したメモリ空間内に含まれるかどうかを決定しなければならない。

【0048】

データ・キャッシュ・モジュール200、及び特にキャッシュ論理210は、DFU170のようなコントローラに接続されて、信号CACHE__A__HIT/MISS201及びCACHE__B__HIT/MISS203により2つのキャッシュ事象についての指示を与える。次いで、DFU170は、どのフェッチ・オペレーションを実行すべきか及び類似のことを決定し得る。DFU170の要求、並びに、ライトバック(書き戻し)・オペレーションを完了するためのWBB180のような様々な支援ユニットからの要求は、様々な要求間を調停するDFU170に送られる。様々な構成要素は、フェッチ要求及びフェッチ確認(acknowledgement)信号を交換する。CACHE__A__HIT/MISS201信号は、データを第1のデータ・バスXA122を介して検索する要求と関連したキャッシュ・ミス事象の発生にตอบสนองして、アサートされる。この信号は、対応するキャッシュ・ヒット事象が発生したときニゲートされる。CACHE__B__H

40

50

IT/MISS203信号は、データを第2のデータ・バスXB124を介して検索する要求と関連したキャッシュ・ミス事象の発生にตอบสนองしてアサートされる。この信号は、対応するキャッシュ・ヒット事象が発生したときニゲートされる。

【0049】

データ・キャッシュ・モジュール200はまた、第1のデータ・バスXA122、第2のデータ・バスXB124及び/又はデータ・フェッチ・バス126に接続されたバッファリング手段を含み得る。

【0050】

図4は、本発明の一実施形態に従ったデータ検索の方法500のフロー・チャートである。

10

【0051】

方法500は、第1のデータ要求及び第2のデータ要求を受け取るステップ510により開始する。第1のデータ要求は、第1のアドレスと関連した第1のデータ単位を受け取る要求である。第2のデータ要求は、第2のアドレスと関連した第2のデータ単位を受け取る要求である。図3に示される例を参照すると、プロセッサ110は、第1のデータ要求を第1のデータ・バスXA122を介して発行し、そして第2のデータ要求を第2のデータ・バスXB124を介して発行する。

【0052】

ステップ510には、第1のデータ単位がデータ・キャッシュ・モジュール内に配置されているかどうかを決定し、そして第2のデータ単位が当該データ・キャッシュ・モジュール内に配置されているかどうかを決定するステップ520が続く。図3に示される例を参照すると、キャッシュ論理210は、第1のデータ単位がキャッシュ・メモリ・バンク250に格納されているかどうか、及び第2のデータ単位がキャッシュ・メモリ・バンク250に格納されているかどうかを決定する。

20

【0053】

両方のデータ単位がデータ・キャッシュ・モジュール内に配置されている場合、ステップ520には、第1及び第2のデータ単位をデータ・キャッシュ・モジュールから与えるステップ530が続く。図3に示される例を参照すると、第1のデータ単位(これは通常BDUの一部である。)は、キャッシュ・メモリ・バンク250からデータ・チャンネル130の第1のポート132を通じ、そして第1のデータ・バスXA122を介してプロセッサ110の第1のデータ・ポート116に送られる。第2のデータ単位(これは通常BDUの一部である。)は、キャッシュ・メモリ・バンク250からデータ・チャンネル130の第2のポート134を通じ、そして第2のデータ・バスXB124を介してプロセッサ110の第2のデータ・ポート118に送られる。

30

【0054】

第1のデータ単位がデータ・キャッシュ・モジュールに配置されていないで、そして第2のデータ単位がデータ・キャッシュ・モジュールに配置されている場合、ステップ520には、第2のデータ単位をデータ・キャッシュ・モジュールから検索し、そして必須型フェッチ・オペレーションを実行して第1のデータ単位を別のメモリから検索するステップ540が続く。図3に示される例を参照すると、データ・キャッシュ・モジュール200が第2のデータ単位を第2のデータ・バスXB124に与える一方、DFU170は、第2のデータ単位を含む第2のBDUの必須のフェッチを含むフェッチ・バーストを開始する。

40

【0055】

第2のデータ単位がデータ・キャッシュ・モジュールに配置されていないで、そして第1のデータ単位がデータ・キャッシュ・モジュールに配置されている場合、ステップ520には、第1のデータ単位をデータ・キャッシュ・モジュールから検索し、そして必須型フェッチ・オペレーションを実行して第2のデータ単位を別のメモリから検索するステップ545が続く。図3に示される例を参照すると、データ・キャッシュ・モジュールが第1のデータ単位を第1のデータ・バスXA122に与える一方、DFU170は、第1の

50

データ単位を含む第1のBDUの必須のフェッチを含むフェッチ・バーストを開始する。

【0056】

第1及び第2の両方のデータ単位がデータ・キャッシュ・モジュール内に配置されていない場合、ステップ520には、少なくとも1つのフェッチ特性にตอบสนองして、且つ複数の情報単位と関連したアドレス情報にตอบสนองして、フェッチ・スキームを決定するステップ550が続く。少なくとも1つのフェッチ特性は、フェッチされるデータ単位の大きさ、フェッチされる単位の最適な大きさ、単一のフェッチ・バーストを形成することができるフェッチ・オペレーションの量、フェッチ・オペレーションの優先順位、フェッチ・バーストに割り込むことができるかどうかということ、及び類似のことを表すことができる。フェッチ・スキームは、等しい大きさのフェッチ・バースト又は異なる大きさのフェッチ・バーストを含むことができる。バースト量はまた、フェッチ特性に従って変わることができる。

10

【0057】

フェッチ・スキームは、データ・フェッチ・オペレーションの量及び更に大きさと、各フェッチ・オペレーション中にフェッチされるべきBDUと、フェッチ・バースト量と、類似のものを含み得る。単一のフェッチ・バースト中に第1及び第2の両方の必須のBDUが検索されることができる場合、フェッチ・スキームがその単一のフェッチを含むことが都合よい。

【0058】

本発明の一実施形態に従って、各フェッチ・バーストは、或る一定の連続したメモリ空間を検索することができ、従って、ステップ550は、両方の必須のBDUが或る一定の連続したメモリ空間内に含まれるかどうかを検査するステップを含み得る。従って、たとえば第1のアドレスが第2のアドレスと異なっても、フェッチ・オペレーションの量を低減することができ、そしてプロセッサ停止期間を減少させることができる。

20

【0059】

本発明の一実施形態に従って、フェッチ・オペレーションの順序は、必須のBDUの場所に対応する。例えば、単一のフェッチ・バースト中に検索可能なメモリ空間が4つのBDUを含み、そして第2及び第3のBDUが必須のBDUである場合、フェッチ・バーストは、第2のBDUをフェッチすることにより開始するであろう。プロセッサは、フェッチ・バーストが終わる前にこれらの必須のBDUを受け取ることができる。

30

【0060】

図1に示される例を参照すると、DFU170は、フェッチ・オペレーション及び/又は第1及び第2のデータ単位を検索するよう要求されたフェッチ・バーストの量を決定する。第1及び第2の両方のデータ単位が単一のBDU内に含まれる場合、単一の必須型フェッチ・オペレーションが要求される。DFU170がフェッチ・バースト・ベースで動作する場合、そのようなフェッチ・バーストは、投機型フェッチが続く必須のフェッチを含む。第1及び第2のデータ単位が単一のフェッチ・バースト中に検索されることができる単一のメモリ空間内に含まれる場合、フェッチ・バーストは、2つの必須のフェッチ及び2つの投機型フェッチを含む。

【0061】

40

本発明の一実施形態に従って、フェッチ・スキームはまた、投機型フェッチを含む。2つのキャッシュ・ミス事象が発生した場合、投機型フェッチ・オペレーションは、インターレースされた要領で順序付けられることができる。

【0062】

ステップ550には、情報フェッチ・スキームを適用するステップ560が続く。上記の適用の成果は、キャッシュ・モジュール又は要求するエンティティに対する第1及び第2のデータ単位の検索である。

【0063】

本発明の一実施形態に従って、投機型フェッチ・オペレーションは、必須型フェッチ・オペレーションと、必須のBDUを含むラインの状態(ステータス)とにตอบสนองする。この

50

状態は、そのライン内で有効であるBDUの量、及び非有効なBDUの場所を含む。

【0064】

投機型フェッチ・オペレーションは、必須のBDUと同じラインに属する投機的BDUをフェッチすることに向けられている。投機型フェッチ・オペレーションは、ラインが充填されるまで、又は必須型フェッチ・オペレーションが要求されるまで繰り返される。

【0065】

本発明者には、投機型フェッチ・オペレーションをインターリーブすることが極めて有効であることが分かった。このインターリーブすることは、両方のキャッシュ・ミス事象に実質的に同じ処置を与えることを促進する。換言すると、それは、或る一定の大きさの公平性をフェッチング・スキームに導入する。従って、両方のラインが充填される前に必須型フェッチ・オペレーションが要求される場合、両方のラインは、実質的に同じ量の投機的BDUを受け取るよう管理する。

【0066】

本発明の一実施形態に従って、投機型フェッチ・オペレーションは、各ライン内の非有効なBDUの量、或る一定のラインに割り当てられる優先順位、及び類似のことを含む追加のパラメータに応答する。

【0067】

図5は、本発明の一実施形態に従った第1のライン252(1)及び第2のライン252(2)の概略的説明図である。BDU252(1,121) - 252(1,15)が有効であり、そしてBDU252(1,0) - 252(1,11)が有効でない。BDU252(2,12) - 252(2,15)が有効であり、そしてBDU252(2,0) - 252(2,11)が有効でない。更に、第1のラインと関連したフェッチ・バースト・サイズは、4個のBDUであり、一方第2のラインと関連したフェッチ・バースト・サイズは、2個のBDUであると仮定する。更に、フェッチ・オペレーションは、プロセッサ110がBDU252(1,4)及びBDU252(2,8)を要求した後で開始し、そしてBDUを異なるラインからフェッチすることが別々のフェッチ・バーストを要求すると仮定する。

【0068】

表1及び表2は、フェッチ・バーストの様々な代替的例示シーケンスを示す。図5は、表1のフェッチ・バーストの例証を含む。フェッチ・バーストは、その図5において参照符号FB1 - FB9で示されている。

【0069】

【表1】

	必須型フェッチ・オペレーション	第1の投機型フェッチ・オペレーション	第2の投機型フェッチ・オペレーション	第3の投機型フェッチ・オペレーション
1	BDU 252 (1,4)	BDU 252 (1,5)	BDU 252 (1,6)	BDU 252 (1,7)
2	BDU 252 (2,8)	BDU 252 (2,9)		
3	BDU 252 (1,0)	BDU 252 (1,1)	BDU 252 (1,2)	BDU 252 (1,3)
4	BDU 252 (2,10)	BDU 252 (1,11)		
5	BDU 252 (1,8)	BDU 252 (1,9)	BDU 252 (1,10)	BDU 252 (1,11)
6	BDU 252 (2,4)	BDU 252 (2,5)		
7	BDU 252 (2,6)	BDU 252 (2,7)		
8	BDU 252 (2,0)	BDU 252 (2,1)		
9	BDU 252 (2,2)	BDU 252 (2,3)		

表1

【0070】

【表 2】

	必須型フェッチ・オペレーション	第1の投機型フェッチ・オペレーション	第2の投機型フェッチ・オペレーション	第3の投機型フェッチ・オペレーション
1	BDU 252 (1, 4)	BDU 252 (1, 5)	BDU 252 (1, 6)	BDU 252 (1, 7)
2	BDU 252 (2, 8)	BDU 252 (2, 9)		
3	BDU 252 (2, 10)	BDU 252 (2, 11)		
4	BDU 252 (1, 0)	BDU 252 (1, 1)	BDU 252 (1, 2)	BDU 252 (1, 3)
5	BDU 252 (2, 0)	BDU 252 (2, 1)		
6	BDU 252 (2, 2)	BDU 252 (2, 3)		
7	BDU 252 (1, 8)	BDU 252 (1, 9)	BDU 252 (1, 10)	BDU 252 (1, 11)
8	BDU 252 (2, 4)	BDU 252 (2, 5)		
9	BDU 252 (2, 6)	BDU 252 (2, 7)		

表 2

図 6 は、本発明の一実施形態に従って情報をキャッシュ・モジュールに与える方法 600 のフロー・チャートである。方法 600 は、複数の情報単位をキャッシュ・モジュールに対して検索する複数の要求を受け取るステップ 610 により開始する。ステップ 610 には、フェッチ・バースト中に検索可能なメモリ空間と複数の情報単位の場所との間の関係を決定するステップ 620 が続く。ステップ 620 は、フェッチ・バースト中に検索可能なメモリ空間を複数の情報単位のそれぞれと関連付けるステップと、これらのメモリ空間間のオーバーラップを捜すステップとを含み得る。ひとたびオーバーラップが探し出されると、フェッチ・バーストの量を低減することができる。

【0071】

例えば、2つの情報単位が同じBDU内にある場合、単一のフェッチ・オペレーションが、両方を検索することができる。2つの情報単位が異なるBDUに配置されているが、しかしこれらの2つのBDUが、フェッチ・バースト中に検索可能な単一のメモリ空間内に含まれる場合、それらは、単一のフェッチ・バースト中に検索される。ひとたび複数のキャッシュ・ミス事象が起こると、上記の要求が、通常、受け取られる。

【0072】

本発明の一実施形態に従って、フェッチ・バースト中に検索可能なメモリ空間の少なくとも1つの特性は、上記メモリ空間の場所に応答する。例えば、そのようなメモリ空間の大きさは、それが属するメモリ・バンクに依存し得る。

【0073】

ステップ 620 には、その決定に応答して少なくとも1つのフェッチ・バーストを実行するステップ 630 が続く。

【0074】

通常、ステップ 630 は、必須型フェッチ・オペレーション、並びに投機型フェッチ・オペレーションを含む。表 1 及び表 2 に示されるように、異なる必須の情報単位と関連した投機的BDUは、インターレースされた要領でフェッチされる。

【0075】

本明細書で説明したものの变形、変更及び他の実施形態が、特許請求の範囲に記載された発明の趣旨及び範囲から逸脱することなしに当業者により行われるであろう。従って、本発明は、前述の例示的説明により定義されるものではなく、代わって添付の特許請求の範囲の趣旨及び範囲により定義されるものである。

【図面の簡単な説明】

【0076】

【図 1】図 1 は、本発明の一実施形態に従った装置の概略図である。

【図 2】図 2 は、本発明の一実施形態に従ったデータ・キャッシュ・モジュールの概略図である。

【図 3】図 3 は、本発明の一実施形態に従ったキャッシュ論理の概略図である。

10

20

30

40

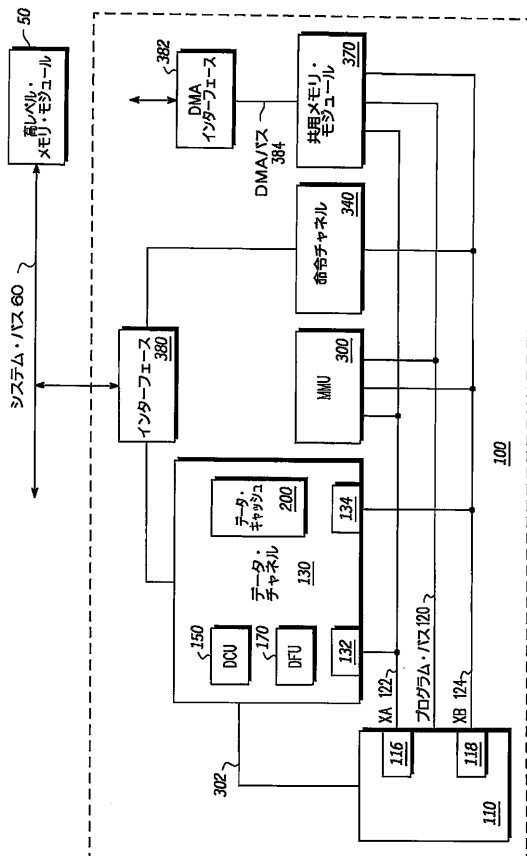
50

【図4】図4は、本発明の一実施形態に従ったデータ検索の方法のフロー・チャートである。

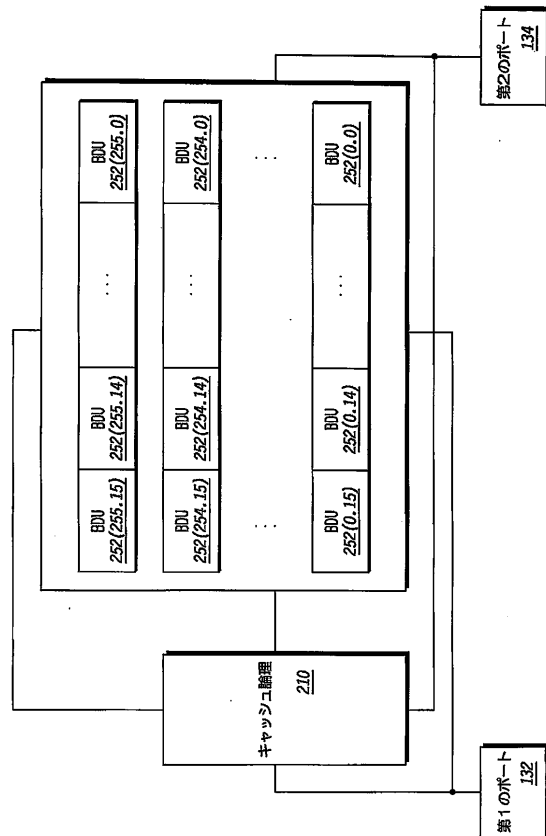
【図5】図5は、本発明の一実施形態に従った第1のライン及び第2のラインの概略的説明図である。

【図6】図6は、本発明の一実施形態に従って情報をキャッシュ・モジュールに与える方法のフロー・チャートである。

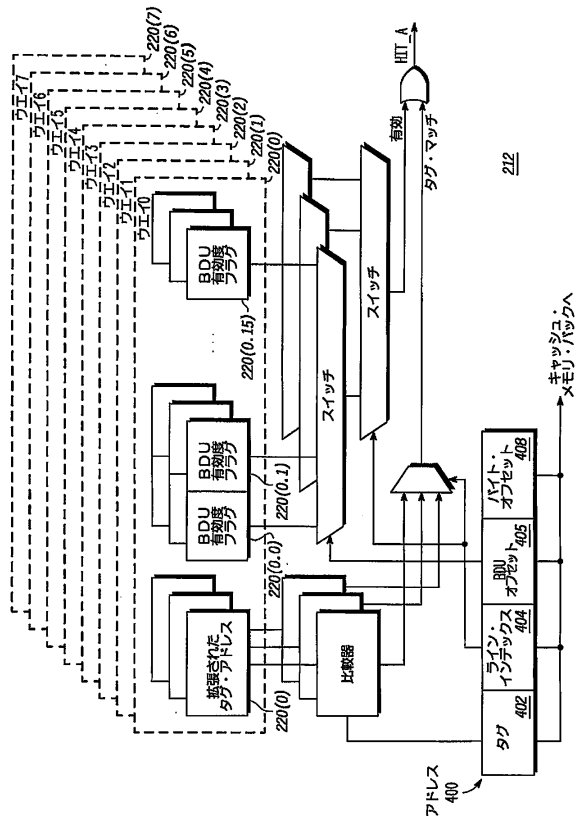
【図1】



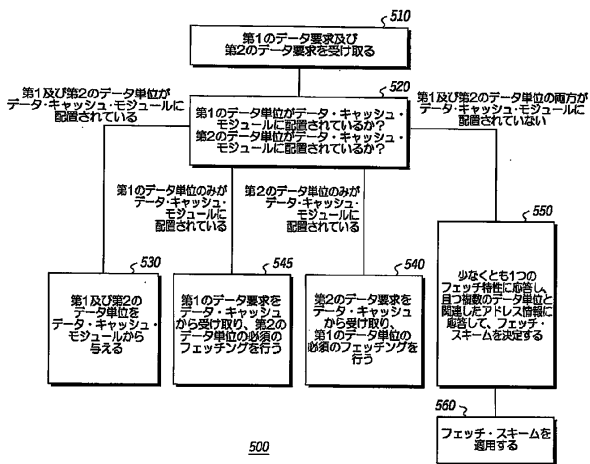
【図2】



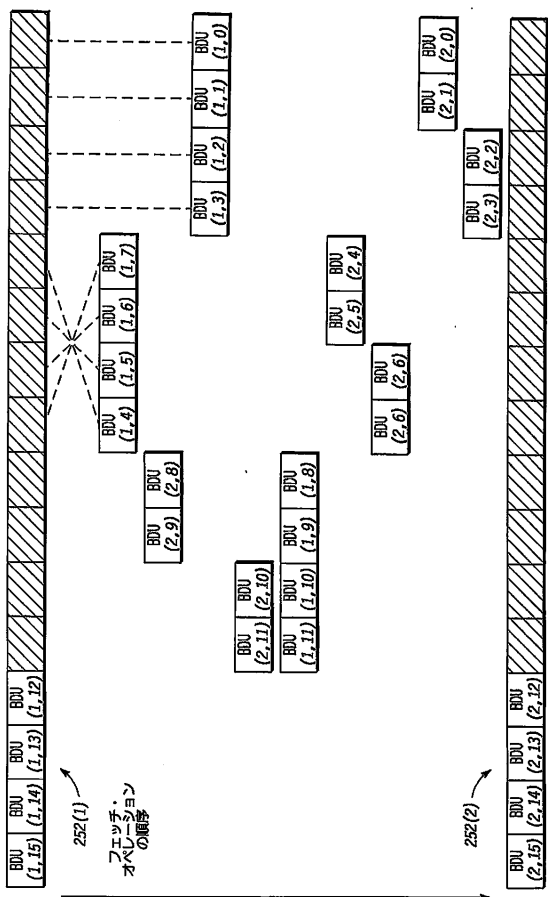
【図3】



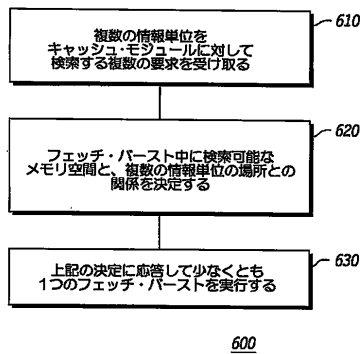
【図4】



【図5】



【図6】



フロントページの続き

(74)代理人 100096013

弁理士 富田 博行

(72)発明者 ゴディン, コスタンティン

イスラエル国 5 0 9 5 3 バット・ヤム, シェシェット・ハヤミム 1 0 / 1 6

(72)発明者 アンシエル, モシェ

イスラエル国 4 4 4 1 8 カフル・サベ, シャロム・ハリヘム 2 4

(72)発明者 エフラット, ヤコブ

イスラエル国 4 4 6 2 8 クファール・サバ, ピカート・ベイト・ハネットファ 1 3 エイ

(72)発明者 ロゼンシェイン, ズヴィカ

イスラエル国 4 4 4 4 4 クファール・サバ, ヘルツェル 4 3

(72)発明者 ザムスキー, ジブ

イスラエル国 4 3 2 5 6 ラアナナ, チャシュモナイム 3 7

審査官 浜岸 広明

(56)参考文献 米国特許出願公開第2004/0088490(US, A1)

米国特許第05809530(US, A)

米国特許出願公開第2005/0160239(US, A1)

(58)調査した分野(Int.Cl., DB名)

G06F 12/08-12/12