

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6205418号
(P6205418)

(45) 発行日 平成29年9月27日 (2017.9.27)

(24) 登録日 平成29年9月8日 (2017.9.8)

(51) Int. Cl.

F I

G 0 6 F 12/08 (2016.01)

G 0 6 F 12/08 5 0 5 B

G 0 6 F 12/08 5 0 9 C

請求項の数 9 (全 19 頁)

(21) 出願番号	特願2015-527632 (P2015-527632)	(73) 特許権者	591016172
(86) (22) 出願日	平成25年8月15日 (2013.8.15)		アドバンスト・マイクロ・ディバイズ・
(65) 公表番号	特表2015-529356 (P2015-529356A)		インコーポレイテッド
(43) 公表日	平成27年10月5日 (2015.10.5)		ADVANCED MICRO DEVI
(86) 国際出願番号	PCT/US2013/055119		CES INCORPORATED
(87) 国際公開番号	W02014/028724		アメリカ合衆国、94088-3453
(87) 国際公開日	平成26年2月20日 (2014.2.20)		カリフォルニア州、サニibel、ピー・
審査請求日	平成28年8月15日 (2016.8.15)		オウ・ボックス・3453、ワン・エイ・
(31) 優先権主張番号	13/588,622		エム・ディ・プレイス、メイル・ストップ
(32) 優先日	平成24年8月17日 (2012.8.17)		・68 (番地なし)
(33) 優先権主張国	米国 (US)	(74) 代理人	100108833
早期審査対象出願			弁理士 早川 裕司
前置審査		(74) 代理人	100111615
			弁理士 佐野 良太

最終頁に続く

(54) 【発明の名称】 データキャッシュプリフェッチのヒント

(57) 【特許請求の範囲】

【請求項 1】

第1のキャッシュに関連付けられた第1のプリフェッチャで、前記第1のプリフェッチャによって決定されるメモリアドレスのシーケンス内のいくつかの数のメモリアドレスからデータをプリフェッチするための要求を発行するのをバイパスすることであって、前記数は、第2のキャッシュに関連付けられた第2のプリフェッチャから受け取った要求に示されており、前記第2のプリフェッチャで未決のプリフェッチ要求の数に基づいている、ことと、

前記バイパスされたメモリアドレスの後のメモリアドレスからデータをプリフェッチするための要求を、前記第1のプリフェッチャから発行することと、
を含む方法。

【請求項 2】

前記第1のプリフェッチャで、少なくとも1つの第1のアドレスに対する少なくとも1つのキャッシュミスを検出することに応じて、データを前記第1のキャッシュのラインにプリフェッチするためのメモリアドレスの前記シーケンスを決定することを含み、メモリアドレスの前記シーケンスは、前記少なくとも1つの第1のアドレスを基準にして決定される、請求項1に記載の方法。

【請求項 3】

前記第2のプリフェッチャで未決のプリフェッチ要求の数を示す情報を記憶することを含む、請求項1に記載の方法。

【請求項 4】

前記未決のプリフェッチ要求の数を示す情報を記憶することは、前記第 2 のプリフェッチャが、プリフェッチ要求を発行することと、未決のプリフェッチ要求の数を前記第 2 のプリフェッチャによって発行された前記プリフェッチ要求に関連付けられたストリームエントリに追加することと、に応じて、未決のプリフェッチ要求の数をカウントすることを含む、請求項 3 に記載の方法。

【請求項 5】

前記未決のプリフェッチ要求の数を示す情報を、前記第 2 のプリフェッチャから前記第 1 のキャッシュに送られるプリフェッチ要求とともに前記第 1 のキャッシュに送ることを含む、請求項 4 に記載の方法。

10

【請求項 6】

メモリアドレスの前記シーケンスは、対応する複数のフラグによって示されており、前記いくつかのメモリアドレスからデータをプリフェッチするための要求を発行するのをバイパスすることは、前記バイパスされたメモリアドレスの数に対応する数の前記フラグをセットしていない状態とすることと、少なくとも 1 つの以後のフラグをセットして、前記対応するメモリアドレスに対する未決のプリフェッチ要求を示すことと、を含む、請求項 1 に記載の方法。

【請求項 7】

前記データをプリフェッチするための要求を発行することは、セットされている前記少なくとも 1 つの以後のフラグに対応する前記メモリアドレスからデータをプリフェッチするための要求を発行することを含む、請求項 6 に記載の方法。

20

【請求項 8】

第 1 のレベルの第 1 のキャッシュと、第 2 のレベルの第 2 のキャッシュと、を備えるメモリであって、前記第 1 のレベルは前記第 2 のレベルよりも高い、メモリと、

第 1 のプリフェッチャと、
第 2 のプリフェッチャと、
カウンタと、を備え、

前記第 2 のプリフェッチャは、プリフェッチ要求を発行することによって、前記第 2 のキャッシュ内にデータをプリフェッチするためのものであり、

前記第 1 のプリフェッチャは、前記第 1 のプリフェッチャの 1 つ以上のストリームエンジン
のプリフェッチストリームエントリによって決定される前記メモリからのデータを前
記第 1 のキャッシュ内にプリフェッチするためのものであって、メモリアドレスのシー
ケンス内のいくつかの数のメモリアドレスからデータをプリフェッチするための要求を発
行するのをバイパスし、前記いくつかの数のメモリアドレスは、前記第 2 のキャッシュに
関連付けられた前記第 2 のプリフェッチャからのプリフェッチ要求に追加されたヒントに
示されており、

30

前記カウンタは、前記第 2 のプリフェッチャに関連付けられており、前記第 2 のプリフ
ェッチャが前記プリフェッチ要求を発行したことに応じて、前記第 2 のプリフェッチャの
プリフェッチ要求に関連付けられた未決のプリフェッチ要求の数をカウントして、前記第
2 のプリフェッチャの前記プリフェッチ要求に追加する前記ヒントを決定する、

40

プロセッサベースのシステム。

【請求項 9】

前記第 2 のプリフェッチャは、前記ヒントを、前記第 2 のプリフェッチャから前記第 1 のキャッシュに送られるプリフェッチ要求とともに前記第 1 のキャッシュに送る、請求項 8 に記載のプロセッサベースのシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本願は、概してプロセッサベースのシステムに関し、より詳細には、プロセッサベースのシステムにおけるデータキャッシュプリフェッチの間にヒントを提供することに関する

50

。

【背景技術】

【0002】

多くの処理装置は、メモリに記憶されている情報にアクセスするのに要する平均時間を短縮するために、キャッシュを利用している。キャッシュは、相対的に頻繁に使用されることが予想される命令またはデータのコピーを記憶する、より小型かつより高速なメモリである。例えば、中央演算処理装置（CPU）は、概してキャッシュまたはキャッシュメモリ素子の階層と関連付けられている。グラフィック処理装置または加速処理装置等の他のプロセッサもキャッシュシステムを実装できる。CPUによって使用されることが予想される命令またはデータは、（相対的に大型かつ低速の）メインメモリからキャッシュ内に移される。CPUがメインメモリ内のロケーションを読み取るまたは書き込む必要がある場合には、CPUは、最初に、所望のメモリロケーションのコピーがキャッシュメモリに含まれているか否かを確かめるためのチェックを行う。このロケーションがキャッシュに含まれている場合（キャッシュヒット）には、CPUは、次に、キャッシュメモリロケーション内のコピーに対して読取り動作または書き込み動作を実行できる。このロケーションがキャッシュに含まれていない場合（キャッシュミス）には、CPUは、次に、メインメモリに記憶されている情報にアクセスする必要がある、いくつかの場合、情報をメインメモリからコピーし、キャッシュに追加することができる。キャッシュを適切に設定して操作することは、メモリアクセスの平均呼出し時間を、メインメモリ呼出し時間未満の値であって、キャッシュアクセス呼出し時間に近い値に低減し得る。

10

20

【発明の概要】

【発明が解決しようとする課題】

【0003】

プリフェッチャは、ライン内の情報がキャッシュから要求される前に、当該ラインをキャッシュ内に読み込むために使用できる。プリフェッチャは、CPUで実行中のアプリケーションと関連付けられたメモリ要求を監視することができ、CPUがメインメモリ内のメモリアドレスの特定のシーケンスにアクセスする可能性があると判断または予測するために、監視されている要求を使用できる。例えば、プリフェッチャは、以前のキャッシュミスのアドレスを記憶するミスアドレスバッファを監視することによって、CPUによるシーケンシャルメモリアクセスを検出してもよい。プリフェッチャは、次に、ミスアドレスバッファ内のシーケンシャルメモリアクセスによって決定されるシーケンス（および方向）のメインメモリ内のロケーションから情報をフェッチし、CPUによって情報が要求される前に当該情報を入手できるように、当該情報をキャッシュに記憶する。プリフェッチャは、複数のストリームを追跡し続けることができ、異なるストリームのために独立してデータをプリフェッチできる。

30

【0004】

開示されている主題は、上述した1つ以上の問題の影響に対処することに向けられている。以下に、開示されている主題のいくつかの態様の基本的な理解を提供するために、開示されている主題の簡略化された概要を提示する。この概要は、開示されている主題の網羅的なものではない。この概要は、開示されている主題の重要もしくは重大な要素を識別すること、または、開示されている主題の範囲を詳細に叙述することを目的としていない。この概要の唯一の目的は、後述するより詳細な説明に対する前置きとして、簡略化された形でいくつかの概念を提示することである。

40

【課題を解決するための手段】

【0005】

一実施形態では、データキャッシュプリフェッチヒンティングのための方法が提供される。この方法の1つの例示的な実施形態は、第1のキャッシュと関連付けられた第1のプリフェッチャで、第1のプリフェッチャによって決定されたメモリアドレスのシーケンス内のいくつかの数のメモリアドレスからデータをプリフェッチするための要求を発行するのをバイパス（回避）することを含む。メモリアドレスの数は、第2のキャッシュに関連

50

付けられた第2のプリフェッチャから受け取った要求に示されている。この方法の実施形態は、バイパスされたメモリアドレスの後のメモリアドレスからデータをプリフェッチするための要求を、第1のプリフェッチャから発行することを含む。

【0006】

別の実施形態では、データキャッシュプリフェッチヒンティングのための装置が提供される。装置の1つの例示的な実施形態は、第1のプリフェッチャによって決定されたメモリアドレスのシーケンス内のいくつかの数のメモリアドレスからデータをプリフェッチするための要求を発行するのをバイパスするように構成可能な第1のプリフェッチャを含む。メモリアドレスの数は、第2のキャッシュに関連付けられた第2のプリフェッチャから受け取った要求において示されている。第1のプリフェッチャの例示的な実施形態は、バイパスされたメモリアドレスの後のメモリアドレスからデータをプリフェッチするための要求を発行するように構成可能である。

10

【0007】

さらに別の実施形態では、データキャッシュプリフェッチヒンティングのためのプロセッサベースのシステムが提供される。プロセッサベースシステムの1つの例示的な実施形態は、メモリと、メモリと関連付けられた第1のキャッシュと、第1のキャッシュ内にデータをプリフェッチするための第1のプリフェッチャと、第1のキャッシュに関連付けられた第2のキャッシュと、第2のキャッシュ内にデータをプリフェッチするための第2のプリフェッチャと、を含む。第1のプリフェッチャは、第1のプリフェッチャによって決定されたメモリアドレスのシーケンス内のいくつかの数のメモリアドレスからデータをプリフェッチするための要求を発行するのをバイパスするように構成可能である。メモリアドレスの数は、第2のプリフェッチャから受け取った要求において示されている。第1のプリフェッチャは、バイパスされたメモリアドレスの後のメモリアドレスからデータをプリフェッチするための要求を発行するように構成可能である。

20

【0008】

さらなる実施形態では、第1のプリフェッチャによって決定されたメモリアドレスのシーケンス内のいくつかの数のメモリアドレスからデータをプリフェッチするための要求を発行するのをバイパスするように構成可能な第1のプリフェッチャを含む集積回路を備える半導体素子の製造に用いられる製造プロセスを、実行されたときに構成可能な命令を含むコンピュータ可読媒体が提供される。メモリアドレスの数は、第2のキャッシュに関連付けられた第2のプリフェッチャから受け取った要求において示されている。第1のプリフェッチャは、バイパスされたメモリアドレスの後のメモリアドレスからデータをプリフェッチするための要求を発行するように構成可能である。

30

【0009】

開示される主題は、同様の参照数字が同様の要素を識別する添付の図面と併せて、以下の説明を参照することにより理解され得る。

【図面の簡単な説明】

【0010】

【図1】半導体ウェハ内または半導体ウェハ上に形成され得る半導体デバイスの第1の例示的な実施形態を概念的に示す図である。

40

【図2】図1に示されるL1プリフェッチャまたはL2プリフェッチャ等のプリフェッチャの1つの例示的な実施形態を概念的に示す図である。

【図3】プリフェッチストリームを割り当て、プリフェッチフラグを使用してルックアヘッドウィンドウを識別するための方法の1つの例示的な実施形態を概念的に示す図である。

【図4】フラグが付けられ、プリフェッチされ得るアドレスのシーケンスを含むプリフェッチストリームの第1の例示的な実施形態を概念的に示す図である。

【図5】プリフェッチストリームに関連付けられたアドレスシーケンスの1つの例示的な実施形態を概念的に示す図である。

【図6】プリフェッチストリームエントリの1つの例示的な実施形態を概念的に示す図で

50

ある。

【図 7】プリフェッチストリームを割り当て、プリフェッチフラグを使用してルックアヘッドウィンドウを識別するための方法の 1 つの例示的な実施形態を概念的に示す図である。

【図 8 A】プリフェッチストリームに関連付けられたアドレスシーケンスの 1 つの例示的な実施形態を概念的に示す図である。

【図 8 B】図 8 A のプリフェッチストリームに関連付けられたアドレスシーケンスの別の例示的な実施形態を概念的に示す図である。

【 0 0 1 1 】

開示される主題は、修正されてもよいし、代替の形態をとってもよいが、開示される主題の特定の実施形態は、図面に一例として示され、本明細書に詳しく説明されている。しかしながら、特定の実施形態の本明細書における説明は、開示されている主題を、開示されている特定の形態に制限することを目的とするのではなく、逆に、添付の特許請求の範囲に含まれる全ての変更形態、均等物および改変形態を対象とすることを意図している。

【 発明を実施するための形態 】

【 0 0 1 2 】

以下に具体例が説明される。明確にするために、実際の実装例の全ての特徴が本明細書で説明されるわけではない。言うまでもなく、係る実施形態の開発では、例えば、実装ごとに変わる、システムおよび事業に関係した制約とのコンプライアンス等のように、開発者の特定の目標を達成するために、多数の実装に特定の決定を下す必要があることが理解される。さらに、係る開発作業は、複雑かつ多大な時間を必要とする場合があるが、それにも関わらず、本開示の利点を有する当業者にとっては日常的な作業となるであろうことが理解される。本説明および図面は、単に主張されている主題の原則を示すにすぎない。したがって、当業者が、ここに明示的に説明または図示されていなくても、ここに説明される原理を実現し、主張されている主題の範囲に含まれ得る多様な構造を創作できる可能性があることが理解されるべきである。さらに、ここに列挙される全ての例は、技術を促進させるために本発明者によって寄与された、請求された主題の原理および概念を理解する上での、読者を手助けするための教育上の目的となることを意図しており、明確に列挙されている係る例および条件に対して制限されていないと解釈されるべきである。

【 0 0 1 3 】

開示されている主題は、添付の図を参照して説明される。多様な構造、システムおよび装置は、説明のためだけに、および、当業者に周知である詳細により本発明を分かりにくくすることがないように、図面で概略的に示される。それでも、添付の図面は、開示されている主題の実例となる例を記述し、説明するために含まれている。本明細書で用いられる単語および語句は、これらの単語および語句の当業者による理解と一貫した意味を有すると理解され、解釈されるべきである。用語または語句の特別な定義、つまり当業者によって理解される通常および慣習となっている意味と異なる定義は、本明細書の用語または語句の一貫した使用によって暗示されることを意図していない。用語または語句が特別な意味、つまり当業者によって理解される意味以外の意味を有することを意図する程度まで、係る特別な定義は、用語または語句に特別な定義を直接的かつ明白に与える定義的な方法で明細書にて明示的に述べられる。さらに、本書に使用される用語「または」は、別段の指示がない限り、排他的ではない「または」を指す（例えば、「あるいは」または「または別の方法では」）。また、本明細書で説明される多様な実施形態は、いくつかの実施形態を新たな実施形態を構成するために 1 つまたは複数の他の実施形態と組み合わせることができるので、必ずしも相互に排他的ではない。

【 0 0 1 4 】

メモリアクセスは、プロセッサコアによってキャッシュから要求されることが予想されるキャッシュラインをプリフェッチできる。例えば、プロセッサコアからの 2 つの要求がキャッシュ内の連続したアドレスをミスする場合には、プリフェッチャは、プロセッサコアが、ミスによって示されるパターンで追加アドレスを要求すると予測することがある

10

20

30

40

50

。したがって、プリフェッチャは、プロセッサコアがキャッシュラインを要求する場合に、これらのラインがキャッシュで利用可能となるように、メモリまたは別のキャッシュからキャッシュラインをプリフェッチしてよい。階層キャッシュシステムでは、プロセッサコアは、（例えばL1データまたは命令キャッシュ等の）より低いレベルのキャッシュからキャッシュラインを要求してよく、より低いレベルのキャッシュは、（例えばL2キャッシュ等の）より高いレベルのキャッシュからキャッシュラインを要求してよく、より高いレベルのキャッシュは、メインメモリまたは（例えばL3キャッシュ等の）さらにより高いレベルのキャッシュからラインを要求してよい。

【0015】

通常、より高いレベルのキャッシュは、より低いレベルのキャッシュと比べてより多い呼出し時間を有する。つまり、より高いレベルのキャッシュの方が、次に高いメモリまたはキャッシュレベルから情報を取り出すのにより多くの時間がかかる。例えば、より長いプリフェッチ距離を有する等のように、より多い呼出し時間を有するキャッシュは、通常、ミスに回答してより多くのプリフェッチ要求を発行する。例えば、L1データキャッシュは、3つの未決のプリフェッチ要求を有し、L2キャッシュは、12の未決のプリフェッチ要求を有することがある。従来のL1キャッシュおよびL2キャッシュは、互いの未決のプリフェッチ要求に気付いていない。その結果、より高いレベルのキャッシュは、同じラインを2回（例えば、より低いレベルのキャッシュからのデマンド要求に回答して一回、および、より高いレベルのプリフェッチャからのプリフェッチ要求に回答して一回）要求して、電力、時間および他のシステムリソースを不必要に消費する、重複する作業を生じさせることがある。また、重複する要求を発行することは、プリフェッチャが別のキャッシュラインをプリフェッチする機会を失わせる。

【0016】

本明細書で説明されるプリフェッチャの実施形態は、より低いレベルのキャッシュプリフェッチストリームに対する未決のプリフェッチ要求の数を示す情報を記憶することによって、従来技術の上述した問題点に対処するように構成されてよい。一実施形態では、プリフェッチヒントと呼ばれることがあるプリフェッチ要求情報が、ストリームエントリにアタッチされてよい。プリフェッチヒントは、より低いレベルのプリフェッチャがより高いレベルのキャッシュに対してデマンド要求を発行することに対応して、より高いレベルのキャッシュに転送され得る。より高いレベルのキャッシュのためのプリフェッチャは、プリフェッチヒントおよびデマンド要求のアドレスを使用して、より低いレベルのプリフェッチャで未決であるプリフェッチ要求に対応するアドレスを無視（skip over）して、以後のアドレスに対する要求を発行するようにより高いレベルのプリフェッチストリームをトレーニングできる。より高いレベルのキャッシュのためのプリフェッチャは、次に、より低いレベルのキャッシュをミスするデマンド要求またはプリフェッチ要求に対応して、プリフェッチ要求を発行してよい。例えば、より低いレベルのキャッシュは、第1のミスを使用してプリフェッチストリームを割り当て、第2のミスを使用してプリフェッチストリームをトレーニングしてよい。より低いレベルのプリフェッチャからより高いレベルのキャッシュへの第1の（および後の）要求は、プリフェッチヒント情報を含む。より高いレベルのキャッシュが要求された情報を有しないと仮定すると、より高いレベルのキャッシュは、第1のミスを使用してプリフェッチストリームを割り当て、第2のミスを使用してプリフェッチストリームをトレーニングし、第1のプリフェッチ要求を（ヒントとともに）使用してその次のプリフェッチ要求のアドレスを選択する。

【0017】

図1は、半導体ウェハ（または半導体ダイ）内または半導体ウェハ（または半導体ダイ）上に形成されてよい半導体デバイス100の第1の例示的な実施形態を概念的に示す図である。半導体デバイス100は、溶着、成長、フォトリソグラフィ、エッチング、平坦化、研磨、アニール等の周知のプロセスを使用して半導体ウェハ内または半導体ウェハ上に形成されてよい。図示されている実施形態では、デバイス100は、メインメモリ110に記憶されている命令またはデータにアクセスするように構成された中央演算処理装

10

20

30

40

50

置(CPU)105を含む。図示されている実施形態では、CPU105は、命令の実行またはデータの操作のために使用される少なくとも1つのCPUコア115を含む。また、CPU105は、選択された命令またはデータをキャッシュに記憶することによって命令またはデータへのアクセスを加速するために使用される階層(またはマルチレベル)キャッシュシステムを実装する。ただし、本開示の利点を有する当業者は、デバイス100の代替実施形態が、例えば外部キャッシュを使用する構成等のように、CPU105と異なる構成を実装し得ることを理解するべきである。さらに、本願で説明される技法は、例えば、グラフィック処理装置(GPU)、加速処理装置(APU)等の他のプロセッサに適用されてもよい。

【0018】

図示されているキャッシュシステムは、メインメモリ110に記憶された命令またはデータのコピーを記憶するためのレベル2(L2)キャッシュ120を含む。図示されている実施形態では、L2キャッシュ120は、メインメモリ110内の各ラインを、L2キャッシュ120の16個の特定のライン(従来「ウェイ」と呼ばれる)との間で潜在的にコピーできるように、メインメモリ110に結合的(associative)な16ウェイである。ただし、本開示の利点を有する当業者は、メインメモリ110またはL2キャッシュ120の代替実施形態が、任意の結合規則を使用して実装できることを理解するべきである。メインメモリ110と比べて、L2キャッシュ120は、より小型でより高速のメモリ素子を使用して実装され得る。また、L2キャッシュ120は、CPUコア115とL2キャッシュ120との間でより迅速に、または、より少ない呼出し時間で情報が交換され得るように、(メインメモリ110と比べて)CPUコア115に対して論理的または物理的により近くに配置されてもよい。

【0019】

また、図示されているキャッシュシステムは、メインメモリ110またはL2キャッシュ120に記憶されている命令またはデータのコピーを記憶するためのL1キャッシュ125を含む。L1キャッシュ125は、L2キャッシュ120と比べて、L1キャッシュ125のラインに記憶されている情報がCPU105によって迅速に取り出すことができるように、より小型でより高速なメモリ素子を使用して実装されてよい。また、L1キャッシュ125は、(メインメモリ110およびL2キャッシュ120の通信と比べて)CPUコア115とL1キャッシュ125との間でより迅速にまたは少ない呼出し時間で情報が交換され得るように、(メインメモリ110およびL2キャッシュ120と比べて)CPUコア115に対して論理的または物理的により近くに配置されてもよい。本開示の利点を有する当業者は、L1キャッシュ125およびL2キャッシュ120が、マルチレベル階層キャッシュメモリシステムの1つの例示的な実施形態を表していることを理解するべきである。代替実施形態は、L0キャッシュ、L1キャッシュ、L2キャッシュ、L3キャッシュ等の要素を含む異なるマルチレベルキャッシュを使用してよい。いくつかの実施形態では、より高いレベルのキャッシュは、より低いレベルのキャッシュのラインが、包含的なより高いレベルのキャッシュに記憶されるように、1つまたは複数のより低いレベルのキャッシュを含んでよい。キャッシュは、通常、スタティックランダムアクセスメモリ(SRAM)で実装されるが、ダイナミックランダムアクセスメモリ(DRAM)等の他のタイプのメモリに実装されてもよい。

【0020】

図示されている実施形態では、L1キャッシュ125は、命令およびデータを記憶するためのレベル1(L1)キャッシュであって、L1-Iキャッシュ130およびL1-Dキャッシュ135と呼ばれるL1キャッシュに分けられる。L1キャッシュ125を、命令を記憶するためのL1-Iキャッシュ130と、データを記憶するためのL1-Dキャッシュ135とに分けるまたは区切ることは、これらのキャッシュの各々を、命令またはデータを要求する可能性のあるエンティティのより近くに配置し得る。結果的に、この配置は、衝突、ワイヤ遅延を低減し、概して命令およびデータに関連付けられる呼出し時間を短縮し得る。一実施形態では、L1-Iキャッシュ130のラインがL2キャッシュ1

10

20

30

40

50

20からの命令で置換され、L1-Dキャッシュ135のラインがL2キャッシュ120からのデータで置換される、という置換方法が決まっている。ただし、当業者は、L1キャッシュ125の代替実施形態が、別々の命令専用キャッシュおよびデータ専用キャッシュ130, 135に区切られていなくてもよいことを理解するべきである。

【0021】

また、CPU105は、L1キャッシュ125, 130, 135のうち1つ以上でラインを読み込むために使用できるL1プリフェッチャ140を含む。一実施形態では、L1プリフェッチャ140は、データキャッシュ135内にラインをプリフェッチするように構成されてよく、その場合、L1プリフェッチャ140は、L1データキャッシュプリフェッチャ140と呼ばれることがある。L1プリフェッチャ140は、図示されている実施形態では、CPU105内の別個の論理素子として示されている。ただし、本開示の利点を有する当業者は、L1プリフェッチャ140が代わりに他の論理素子の一部として実装され得ることも理解するべきである。たとえば、L1プリフェッチャ140は、L1-Dキャッシュ135のロジックの一部として実装され得る。一実施形態では、L1プリフェッチャ140は、CPUコア115で実行中のアプリケーションと関連付けられたメモリ要求を監視できる。例えば、L1プリフェッチャ140は、L1ミスアドレスバッファ145に記録され得るキャッシュヒットまたはキャッシュミスにつながるメモリ要求を監視できる。L1プリフェッチャ140は、CPUコア115が、メインメモリ110においてメモリアドレスの特定のシーケンスにアクセスする可能性があると判断または予測してよい。例えば、L1プリフェッチャ140は、CPUコア115によってアクセスされる2つ以上のシーケンシャルメモリアccessを検出してよい。シーケンスの方向は、シーケンシャルメモリアccessの時間的なシーケンスに基づいて決定することができ、CPUコア115は、現在または以前のシーケンシャルメモリアccessに基づいて推定することによって将来のメモリアccessを予測するために、この方向を使用できる。次に、L1プリフェッチャ140は、予測されたロケーションの情報をL2キャッシュ120またはメインメモリ110からフェッチし、当該情報がCPUコア115によって要求される前に当該情報を利用できるように、当該情報を適切なキャッシュに記憶できる。

【0022】

また、CPU105は、L2キャッシュ120でラインを読み込むために使用できるL2プリフェッチャ150も含んでよい。L2プリフェッチャ150は、図示されている実施形態では、CPU105内の別個の論理素子として示されている。ただし、本開示の利点を有する当業者は、L2プリフェッチャ150が代わりに他の論理素子の一部として実装され得ることも理解するべきである。例えば、L2プリフェッチャ150は、L2キャッシュ120のロジックの一部として実装され得る。一実施形態では、L2プリフェッチャ150は、CPUコア115で実行中のアプリケーションに関連付けられたメモリ要求を監視できる。例えば、L2プリフェッチャ150は、L2ミスアドレスバッファ155に記録され得るキャッシュヒットまたはキャッシュミスにつながるメモリ要求を監視できる。L2プリフェッチャ150は、CPUコア115またはL1プリフェッチャ140が、メインメモリ110のメモリアドレスの特定のシーケンスにアクセスする可能性があると判断または予測してよい。例えば、L2プリフェッチャ150は、CPUコア115、キャッシュ125, 130, 135またはL1プリフェッチャ140によってアクセスされる2つ以上のシーケンシャルメモリアccessを検出してもよい。シーケンスの方向は、シーケンシャルメモリアccessの時間的なシーケンスに基づいて決定することができ、L2プリフェッチャ150は、現在または以前のシーケンシャルメモリアccessに基づいて推定することによって将来のメモリアccessを予測するために、この方向を使用できる。次に、L2プリフェッチャ150は、予測されたロケーションの情報をメインメモリ110からフェッチし、当該情報がCPUコア115、キャッシュ125, 130, 135またはL1プリフェッチャ140によって要求される前に当該情報を利用できるように、当該情報を適切なキャッシュに記憶できる。

【0023】

図 2 は、図 1 に示された L 1 プリフェッチャ 1 4 0 または L 2 プリフェッチャ 1 5 0 等のプリフェッチャ 2 0 0 の 1 つの例示的な実施形態を示す図である。例示的な実施形態では、プリフェッチャ 2 0 0 は、例えばロード命令と関連付けられたヒットまたはミス、記憶命令と関連付けられたヒットまたはミス等のメモリアクセス要求に関係する特定のイベントを示す信号を受信する。例えばロードまたは記憶のためのヒットイベントまたはミスイベント等のミスアドレスバッファ (M A B) イベントは、プリフェッチャ 2 0 0 の他のステージ (stage) に渡されることになるイベントを選択するのに用いられるイベントセクタブロック 2 0 5 によって受け取られ、または、アクセスされてよい。例えば、最高の優先順位のイベントが、例えば以降のクロックサイクルの間、1 つ以上のストリームエンジン 2 1 5 およびストリーム割当て部 2 2 0 に渡されるまで、最高の優先順位のイベントは、レジスタ 2 1 0 に記憶されてよい。イベントの優先順位は、例えば、最高の優先順位をロードミスに与え、次いで、順次低くなる優先順位を記憶ミス、ロードヒットおよび記憶ヒットに割り当てる等のような階層 (hierarchy) を用いて、決定され得る。ただし、本開示の利点を有する当業者は、イベントの優先順位を確立するために、代替の階層を使用できることも理解するべきである。

【 0 0 2 4 】

図示されている実施形態では、プリフェッチャ 2 0 0 は、別々のプリフェッチストリームの管理に使用できる 1 つ以上のストリームエンジン 2 1 5 を含む。ストリームエンジン 2 1 5 は、カレントイベントが、ストリームエンジン 2 1 5 によって管理されるストリームをヒットした、または、ミスしたの何れかを示すために、ストリーム割当て部 2 2 0 に信号を提供してよい。既存のストリームの何れも M A B ミスイベントに対するヒットを示さない場合には、ストリーム割当て部 2 2 0 は、次いで、カレントイベント情報を使用して、新たなストリームを異なるストリームエンジン 2 1 5 に割り当てることができる。ストリームが最初に割り当てられると、ストリームエンジン 2 1 5 は、ページアドレスおよびオフセット値を、カレントイベントキャッシュラインアドレスに設定する。次に、ストリームエンジン 2 1 5 は、何れかの方向でカレントイベントキャッシュラインアドレスに隣接するアドレスにてイベントを検出するために、さらなる M A B イベントを監視できる。例えば、カレントイベントキャッシュラインアドレスが A に設定される場合には、ストリームエンジン 2 1 5 は、次いで、例えばアドレス A + 1 または A - 1 等の、カレントイベントキャッシュラインアドレスに関するアドレスにて、イベントを探す。ストリームエンジン 2 1 5 が 1 つのアドレスを確認する場合には、ストリームエンジン 2 1 5 は、適切な方向 (A + 1 の場合には正の方向、A - 1 の場合には負の方向) でストリームを定義し、新たなプリフェッチストリームをトレーニングする。一実施形態では、ストリームエンジン 2 1 5 は、本明細書で説明されるように、カレントストリームアドレスに対する潜在的なプリフェッチを示すプリフェッチフラグのセットを維持する。プリフェッチフラグは、新たなプリフェッチストリームがトレーニングされるときに設定されてよい。

【 0 0 2 5 】

また、プリフェッチャ 2 0 0 は、ストリームエンジン 2 1 5 からのプリフェッチ要求の調停 (arbitrate) に用いられる要求アービタ 2 2 5 を含んでもよい。一実施形態では、要求アービタ 2 2 5 は、回転優先順位アービタ (rotating priority arbiter) である。ただし、本開示の利点を有する当業者は、他のタイプの要求アービタ 2 2 5 が、代替でプリフェッチャ 2 0 0 に実装され得ることを理解するべきである。要求は、例えば、以後のクロックサイクルの間、要求情報がプリフェッチ要求インタフェース 2 3 5 に提供されるように、要求アービタ 2 2 5 からレジスタ 2 3 0 に転送され得る。プリフェッチ要求インタフェース 2 3 5 は、ストリームエンジン 2 1 5 からの未決の要求間の選択または調停に用いられるフィードバックを、要求アービタ 2 2 5 に提供できる。

【 0 0 2 6 】

図示されている実施形態では、ストリームエンジン 2 1 5 は、ストリームエンジン 2 1 5 に関連付けられたプリフェッチストリームの状態を識別する情報を含むストリームエントリを維持してもよい。一実施形態では、ストリームエントリは、エントリのデータベー

10

20

30

40

50

スに記憶されてよい。ただし、本開示の利点を有する当業者は、ストリームエントリが、任意のレジスタ、メモリまたはキャッシュロケーションに記憶されてもよいことを理解すべきである。ここで説明されるように、各ストリームエントリは、例えば、ストリームエントリのためのページアドレスもしくはラインオフセット、プリフェッチされ得るアドレスに関連付けられた1つ以上のプリフェッチフラグ等のアドレス指定情報、または、他のアドレス指定情報を含んでよい。また、ストリームエントリは、ストリームエントリに関連付けられた未決プリフェッチ要求の数を示す「ヒント」情報を含んでもよい。

【0027】

図3は、プリフェッチストリームを割り当て、プリフェッチフラグを使用してルックアヘッドウィンドウを識別するための方法300の1つの例示的な実施形態を概念的に示す図である。方法300の実施形態は、図1に示すL1データキャッシュプリフェッチャ140で実装されてよい。ただし、本開示の利点を有する当業者は、方法300の代替実施形態が、プリフェッチを実装する別のキャッシュからラインを要求可能な任意のプリフェッチャで実装され得ることを理解すべきである。図示されている実施形態では、例えば、ミスアドレスバッファからのMABイベントについての情報を用いて第1のキャッシュミスが検出され（ブロック305）、ストリームがプリフェッチャストリームエンジンに割り当てられる。プリフェッチャストリームエンジンは、第1のキャッシュミスとともに（正方向または負方向の）シーケンス内の第2のミスを検出するまで（ブロック310）待機する。シーケンシャルアドレスに対するキャッシュミスの検出は、図示されている実施形態では、プリフェッチストリームの生成をトリガするために用いられているが、当業者は、代替実施形態において、プリフェッチストリームの生成をトリガするために異なるキャッシュミスパターンを使用してもよいことを理解すべきである。例えば、プリフェッチストリームは、所定のストライド値または何らかの他のパターンによって区切られたアドレスに対するキャッシュミスを検出することに応じて、割り当てられてよい。

【0028】

図示されている実施形態では、ストリームエンジンは、第1のキャッシュミスおよび第2のキャッシュミスのアドレスを使用して、シーケンス方向を決定し得る（ブロック315）。プリフェッチフラグは、プリフェッチストリーム内の選択された数のアドレスに割り当てられてよい（ブロック320）。フラグが割り当てられるアドレスは、第1のキャッシュミスおよび第2のキャッシュミスによって確立されたシーケンスまたはパターンに従う。例えば、フラグは、プリフェッチストリームに対して確立された方向での第2のミスのアドレスに従う選択された数のアドレス（例えば、9つのアドレス）に割り当てられ得る（ブロック320）。次に、選択された数のフラグは、プリフェッチャが以後のクロックサイクルでこれらのアドレスから情報を取得する必要があることを示すために、セットされ得る（ブロック325）。例えば、サブセットのアドレス（例えば、9つのフラグ付きアドレスのうち4つのアドレス）のフラグは、これらのアドレスがメモリからキャッシュに取得可能となるように、セットされ得る。

【0029】

次に、プリフェッチャは、未決のプリフェッチ要求の数を示すのに用いられ得るヒントを決定してよい（ブロック330）。一実施形態では、ヒントは、フラグシーケンスの初期にセットされた別のプリフェッチフラグによって示された情報に対する要求をプリフェッチャが発行した後に、セットされたプリフェッチフラグの数をカウントすることによって決定されてよい（ブロック330）。ヒント情報は、本明細書で説明されるように、ヒント情報を別のプリフェッチャに提供できるようにストリームエントリに追加されてもよいし、記憶されてよい。

【0030】

図4は、フラグが付され得るアドレスであって、プリフェッチ可能なアドレスのアドレスシーケンス400を含むプリフェッチストリームの第1の例示的な実施形態を概念的に示す図である。アドレスシーケンス400（1～4）は、連続時間間隔またはクロックサイクルにおける同一セットのアドレスを示している。図示されている実施形態では、アド

レスシーケンス 400 (1~4) は、連続クロックサイクルで示されている。ただし、本開示の利点を有する当業者は、アドレスシーケンス 400 (1~4) が、厳密に言えばシーケンシャルでなくてもよく、いくつかの場合には、他のイベント、時間間隔、クロックサイクル等によって分けられ得ることを理解するべきである。図示されている実施形態では、プリフェッチストリームエンジンは、矢印 405 で示されるベースアドレスの先の 9 つのアドレスにフラグを割り当てている。また、ストリームエンジンは、塗りつぶされたフラグで示されるように、ベースアドレス 405 の先のアドレスのうち 4 つのアドレスのフラグをセットして、メモリのこれらのアドレスの情報がキャッシュ内に取得される必要があることを示す。4 つのセットされたプリフェッチフラグは、このプリフェッチストリームに関連付けられた 4 つの未決の要求があることを示している。本開示の利点を有する当業者は、割り当てられたフラグの数、または、セットされ、割り当てられたフラグのサブセット内のフラグの数が、設計の選択の問題であって、異なる設計考慮事項を満たすために変更されてよいことを理解するべきである。

10

【0031】

フラグは、プリフェッチフラグのルックアヘッドウィンドウの外部にあるキャッシュミスに応じてセットされてよい。図示されている実施形態では、連続キャッシュミス 410 (1~2) に応じて、プリフェッチストリームが割り当てられ、プリフェッチウィンドウのアドレスにフラグが付けられる。フラグが付けられたアドレスは、ベースアドレス 405 で始まり、アドレスは、キャッシュミス 410 (1~2) またはベースアドレス 405 を基準にして定義される。図示されている実施形態では、ベースアドレス 405 は、1 という選択されたアドレスオフセット値により、キャッシュミス 410 (2) からオフセットされる。例えば、第 1 のキャッシュミス 410 (1) のアドレスが A として定義され、プリフェッチウィンドウの相対アドレスが (± 2 、 ± 3 、 ± 4 、 ± 5 、...) として定義される場合には、フラグアドレスは、シーケンスの方向に応じて、第 1 のキャッシュミスアドレス 410 (1) を基準にして ($A \pm 2$ 、 $A \pm 3$ 、 $A \pm 4$ 、 $A \pm 5$ 、...) として定義されてよい。本開示の利点を有する当業者は、アドレスオフセット値が設計の選択の問題であって、異なる設計考慮事項を満たすために変更されてよいことを理解するべきである。

20

【0032】

アドレスシーケンス 400 (1) は、ベースアドレス 405 の先のアドレスにセットされたフラグを有する。このため、プリフェッチャは、このアドレスのデータをキャッシュ内にフェッチしてよい。データが取得されると、ベースアドレス 405 は、次のアドレスに進められてよく、取得されたばかりのアドレスは、ストリームの履歴の一部となり、このアドレスのフラグは、アドレスシーケンス 400 (2) で縞を付けたフラグ記号で示される履歴フラグとなる。フェッチの後、3 つのプリフェッチフラグがセットされたままとなっているため、このプリフェッチストリームにおいて 3 つの未決のプリフェッチ要求が存在している。ベースアドレス 405 の先のフラグの数が同じままとなるように、追加のフラグが、確立されたシーケンスの方向でシーケンスの最後のアドレスに向かうアドレスに割り当てられてよい。図示されている実施形態では、プリフェッチャは、セットされたフラグを有するアドレスをフェッチし続け、セットされたフラグを有するアドレスの全てが取得されるまでベースアドレス 405 を進め続ける。それぞれのセットされたフラグに対応するアドレスが取得されるにつれ、このアドレスに関連付けられたフラグは、アドレスシーケンス 400 (3~4) に示される履歴フラグに変更される。追加のフラグは、ベースアドレス 405 の先のフラグの数を維持するために、次のシーケンシャルアドレスに割り当てられてよい。また、履歴フラグは、ベースアドレス 405 の後続く履歴フラグのセット数を維持するために、外されてもよい。

30

40

【0033】

図 5 は、プリフェッチストリームに関連付けられたアドレスシーケンス 500 の 1 つの例示的な実施形態を概念的に示す図である。図示されている実施形態では、プリフェッチストリームエンジンは、矢印 505 によって示されるベースアドレスの先の 9 つのアドレ

50

スにフラグを割り当てている。ここで説明されるように、ベースアドレス 505 およびアドレスシーケンスの方向は、アドレス 510 に対するミスに基づいて決定されてよい。ストリームエンジンは、塗りつぶされたフラグで示されるように、ベースアドレス 505 の先のアドレスのうち 4 つのアドレスのフラグをセットして、メモリのこれらのアドレスの情報がキャッシュ内に取得される必要があることを示す。4 つのセットされたプリフェッチフラグは、このプリフェッチストリームに関連付けられた 4 つの未決の要求があることを示している。

【0034】

図示されている実施形態では、プリフェッチストリームエンジンは、第 1 のセットフラグに関連付けられたアドレスによって示されるキャッシュラインに対するプリフェッチ要求 515 を発行する。プリフェッチ要求 515 を発行することに応じて、カウンタ 520 は、プリフェッチストリームに関連付けられた未決のプリフェッチ要求の数をカウントするために使用されてよい。例えば、プリフェッチ要求 515 の後、アドレスシーケンス 500 は、3 つの未決プリフェッチ要求に対するセットフラグを含んでいる。したがって、カウンタ 520 は、3 つの未決プリフェッチ要求をカウントしてよい。次に、カウンタによって決定された値は、ヒント 525 として記憶されてよい。図示されている実施形態では、ヒント 525 は、未決のプリフェッチ要求が 0 ~ 3 つある可能性があることを示す情報を記憶するには十分である 2 ビット [4:3] によって表されている。ただし、本開示の利点を有する当業者は、代替実施形態が、異なる数の潜在的な未決のプリフェッチ要求を表すために、ヒント 525 に異なる数のビットを含んでよいことを理解するべきである。

【0035】

図 6 は、プリフェッチストリームエン트리 600 の 1 つの例示的な実施形態を概念的に示す図である。図示されている実施形態では、プリフェッチストリームエン트리 600 は、プリフェッチャ内の対応するストリームエンジンによって生成されてもよいし、維持されてもよいし、記憶されてもよい。プリフェッチストリームエン트리 600 は、ストリームに関連付けられたデータのページに対応するページアドレス 605 と、ストリームに関連付けられたページの内部のカレントラインオフセットを示すラインオフセット 610 と、ページ内で関連付けられたラインからデータをプリフェッチするための要求を示すためにセットできるプリフェッチフラグ 615 と、を示す情報を含む。ストリームエン트리 600 は、ストリームが昇順アドレスに対応するのか、または、降順アドレスに対応するのかを示す方向ビット (DIR) を含んでもよい。また、ストリームエン트리 600 は、対応するプリフェッチストリームに関連付けられた未決プリフェッチ要求の数を示す情報を含むプリフェッチヒント情報 620 を含む。一実施形態では、プリフェッチストリームに関連付けられたプリフェッチストリームエンジンは、プリフェッチヒント 620 の値を、例えばプリフェッチ要求を発行することに応じて変更できる。例えば、プリフェッチヒント 620 が 3 という値を初期に有しており、プリフェッチ要求を発行することで未決のプリフェッチ要求の数を 2 に低減する場合には、プリフェッチストリームエンジンは、2 という値を有するようにプリフェッチヒント 620 の値を変更してよい。

【0036】

図 7 は、プリフェッチストリームを割り当て、プリフェッチフラグを使用してルックアヘッドウィンドウを識別するための方法 700 の 1 つの例示的な実施形態を概念的に示す図である。方法 700 の実施形態は、図 1 に示す L2 データキャッシュプリフェッチャ 150 で実装されてよい。ただし、本開示の利点を有する当業者は、方法 700 の代替実施形態が、プリフェッチを実装する別のキャッシュからデマンド要求またはプリフェッチ要求を受け取ることができる任意のプリフェッチャで実装され得ることを理解するべきである。図示されている実施形態では、例えば、ミスアドレスバッファからの MAB イベントについての情報を用いて第 1 の L2 キャッシュミスが検出され (ブロック 705)、ストリームがプリフェッチャストリームエンジンに割り当てられる。次に、プリフェッチャストリームエンジンは、第 1 のキャッシュミスとともに (正方向または負方向の) シーケン

10

20

30

40

50

ス内の第2のミスを検出するまで(ブロック710)待機する。図示されている実施形態では、シーケンシャルアドレスに対するキャッシュミスの検出は、プリフェッチストリームの生成をトリガするために用いられているが、当業者は、代替実施形態において、プリフェッチストリームの生成をトリガするために異なるキャッシュミスパターンを使用してよいことを理解するべきである。例えば、プリフェッチストリームは、所定のストライド値または何らかの他のパターンで区切られるアドレスに対するキャッシュミスを検出することに応じて、割り当てられてもよい。

【0037】

図示されている実施形態では、ストリームエンジンは、シーケンス方向を決定するために(ブロック715)、第1のキャッシュミスおよび第2のキャッシュミスのアドレスを使用できる。プリフェッチフラグは、プリフェッチストリーム内の選択された数のアドレスに割り当てられてよい(ブロック720)。フラグが割り当てられるアドレスは、第1のキャッシュミスおよび第2のキャッシュミスによって確立されるシーケンスまたはパターンに従う。例えば、フラグは、プリフェッチストリームに対して確立された方向での第2のミスのアドレスに従う選択された数のアドレス(例えば、12以上のアドレス)に割り当てられ得る(ブロック720)。次に、プリフェッチストリームのためのストリームエンジンは、プリフェッチヒント情報が、例えば関連付けられたキャッシュからのデマンド要求またはプリフェッチ要求とともに、関連付けられたキャッシュから受け取られたか否かを判断してもよい(ブロック725)。

【0038】

次に、割り当てられたフラグのサブセットは、受け取られたデマンド要求またはプリフェッチ要求に基づいてセットされてよい。一実施形態では、ヒント情報が受け取られない場合には、以後のクロックサイクルにおいて、フラグが付けられたアドレスからプリフェッチャが情報を取得する必要があることを示すために、所定のオフセットで開始する、選択された数のフラグが、セットされ得る(ブロック730)。例えば、アドレスのサブセット(例えば、12のフラグ付きのアドレス)のフラグは、これらのアドレスがメモリからキャッシュ内に取得可能となるようにセットされ得る(ブロック730)。ただし、ストリームエンジンが、関連付けられたキャッシュでの未決のプリフェッチ要求の数を示すヒント情報が受け取られたと判断した場合には(ブロック725)、プリフェッチフラグは、所定のオフセットとは異なるオフセットでセットされてよい(ブロック735)。図示されている実施形態では、プリフェッチフラグは、ヒント情報によって示される未決のプリフェッチ要求の数を加えた所定のオフセットと等しいオフセットで始まるアドレスにセットされてよい(ブロック735)。

【0039】

このようにしてアドレスを設定すること(ブロック735)は、プリフェッチャが、他の関連付けられたキャッシュによって既に要求されていた可能性があるラインに対する要求を発行することをバイパスする、または、無視することができるようにする。例えば、L2キャッシュは、L1キャッシュに関連付けられた1つ以上のプリフェッチストリームで要求されたか、または、未決であるかの何れかであるラインに関連付けられたアドレスに先立つアドレスに関連付けられたフラグをセットできる(ブロック735)。これらのアドレスをバイパスまたは無視することは、L2レベルで実行される冗長なプリフェッチ要求の数を低減させ、L2キャッシュ内に情報をプリフェッチするのに使用可能な機会の数を増加させる。

【0040】

図8Aは、プリフェッチストリームに関連付けられたアドレスシーケンス800の1つの例示的な実施形態を概念的に示す図である。図示されている実施形態では、プリフェッチストリームエンジンは、矢印805で示されるベースアドレスの先の12以上のアドレスにフラグを割り当てている。ここで説明されるように、ベースアドレス805およびアドレスシーケンスの方向は、アドレス810に対するミスに基づいて決定されてよい。また、ストリームエンジンは、塗りつぶされたフラグで示されるベースアドレス805の先

の 1 2 のアドレスのフラグをセットして、メモリのこれらのアドレスの情報がキャッシュ内に取得される必要があることを示す。図示されている実施形態では、フラグは、アドレス 8 1 0 (3) で示されるラインのために関連付けられた (例えば、より低いレベルの) キャッシュからプリフェッチ要求を受け取ったことに応じて、セットされている。ストリームエンジンは、より低いレベルのプリフェッチャからヒントを受け取っていないため、プリフェッチ要求によって示されるアドレス 8 1 0 (3) から 1 という所定のオフセットで開始するフラグがセットされる。ただし、本書に説明されるように、所定のオフセットは設計の選択肢の問題であって、異なる設計または操作上の考慮事項を満たすために変更されてよい。

【 0 0 4 1 】

図 8 B は、プリフェッチストリームに関連付けられたアドレスシーケンス 8 0 0 の別の例示的な実施形態を概念的に示す図である。図示されている実施形態では、プリフェッチストリームエンジンは、矢印 8 0 5 で示されるベースアドレスの先の 1 2 以上のアドレスにフラグを割り当てている。ここで説明されるように、ベースアドレス 8 0 5 およびアドレスシーケンスの方向は、アドレス 8 1 0 に対するミスに基づいて決定されてもよい。ストリームエンジンは、塗りつぶされたフラグで示されるベースアドレス 8 0 5 の先のアドレスのフラグを設定して、メモリのこれらのアドレスでの情報がキャッシュ内に取得される必要があることを示す。図示されている実施形態では、フラグは、アドレス 8 1 0 (3) によって示されるラインのために関連付けられた (例えば、より低いレベルの) キャッシュからプリフェッチ要求を受け取ることに応じて、セットされている。図示されている実施形態のストリームエンジンは、関連付けられたより低いレベルのプリフェッチャでの 3 つの未決のプリフェッチ要求を示すヒントを、より低いレベルのプリフェッチャから受け取る。その結果として、ストリームエンジンは、プリフェッチ要求によって示されるアドレス 8 1 0 (3) から、未決のプリフェッチ要求の数 (3) が加えられた 1 という所定のオフセットで始まるフラグをセットする。したがって、より高いレベルのプリフェッチャは、バイパスまたは無視されたアドレスに対するプリフェッチ要求を発行することを、バイパスしてもよいし、無視してもよい。

【 0 0 4 2 】

(プロセッサシステム 1 0 0 等の) 本明細書で説明されるプリフェッチヒントを提供または活用することができるプロセッサシステムの実施形態は、多様なプロセッサ設計に従って半導体製造施設で製造できる。一実施形態では、プロセッサ設計は、コンピュータ可読媒体に記憶されたコードとして表すことができる。プロセッサ設計を定義するおよび / または表すために使用されてよい例示的なコードは、HDL、Verilog 等を含んでよい。コードは、エンジニアによって作成され、他の処理装置によって合成され、プロセッサ設計の中間表現 (例えばネットリスト、GDSII データ等) を生成するために使用されてよい。中間表現はコンピュータ可読媒体に記憶され、半導体製造施設で実行される製造 / 製作プロセスを構成し、制御するために使用できる。半導体製造施設は、半導体基板上にトランジスタおよび他の回路網を形成するために使用される蒸着、フォトリソグラフィ、エッチング、研磨 / 平坦化、計測学および他のプロセスを実行するための処理ツールを含んでよい。処理ツールは、例えば GDSII データから生成されるマスクワークを使用することによって、中間表現を使用して構成することができ、操作される。

【 0 0 4 3 】

開示されている主題および対応する詳細な説明の部分は、ソフトウェア、または、アルゴリズムおよびコンピュータメモリ内部のデータビットに対する動作の象徴表現に関して提示される。これらの説明および表現は、当業者が自分の作業の要旨を他の当業者に効果的に伝えることができるようにする説明および表現である。アルゴリズムは、用語がここで使用されるように、用語が一般に使用されるように、所望の結果につながるステップの首尾一貫したシーケンスとなるように考え出される。ステップは、物理的な量の物理的な操作を必要とするステップである。通常、必ずしもではないが、これらの量は、記憶され、転送され、結合され、比較され、それ以外の場合操作されることが可能な光学信号、電

10

20

30

40

50

気信号または磁気信号の形をとる。これらの信号をビット、値、要素、記号、文字、用語、数等として参照することは、主に一般的な使用の理由からときおり便利であることが判明している。

【 0 0 4 4 】

ただし、これらの用語および類似する用語の全てが適切な物理量と関連付けられなければならない、これらの量に適用される便利なラベルにすぎないことが念頭に入れられるべきである。特に明確な指定がない限り、または、説明から明らかであるように、「処理する」、「算定する」、「計算する」、「決定する」または「表示する」等の用語は、コンピュータシステムまたは類似する電子計算装置（コンピュータシステムのレジスタおよびメモリ内の物理電子量として表されるデータを操作し、または、当該データを、コンピュータシステムのメモリやレジスタ内の物理量、もしくは、他の係る情報記憶装置、情報送信装置、情報表示装置内の物理量として同様に表される他のデータに変換する装置）のアクションおよびプロセスを指す。

10

【 0 0 4 5 】

また、開示されている主題のソフトウェア実装態様が通常何らかの形のプログラム記憶媒体上で符号化され、または、何らかのタイプの送信媒体上で実装されることにも留意されたい。プログラム記憶媒体は、磁気（例えば、フロッピー（登録商標）ディスクもしくはハードドライブ）または、光学（例えば、コンパクトディスク読取り専用メモリ、つまり「CD ROM」）であってもよいし、読取り専用またはランダムアクセスであってもよい。同様に、送信媒体は、ツイストペア線、同軸ケーブル、光ファイバ、または、既知の何らかの他の適切な伝送媒体であってもよい。開示されている主題は、任意の所与の実装例の態様によって制限されない。

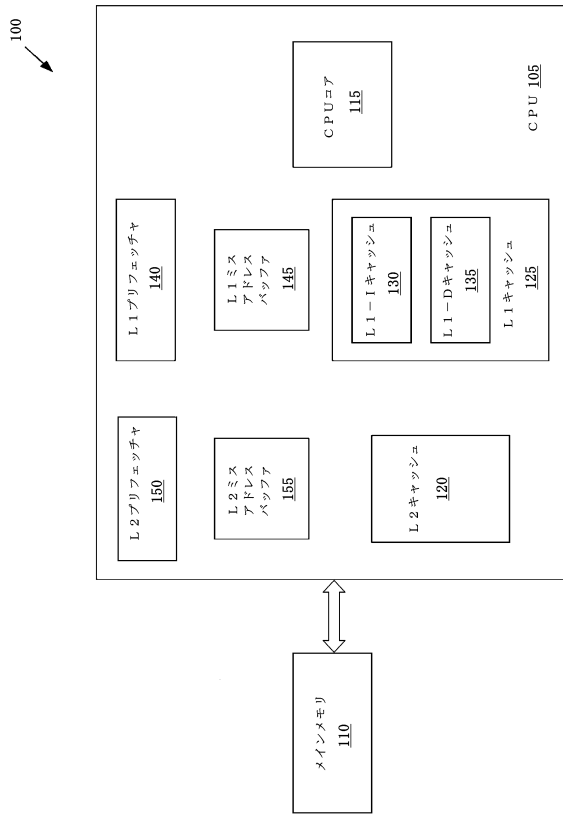
20

【 0 0 4 6 】

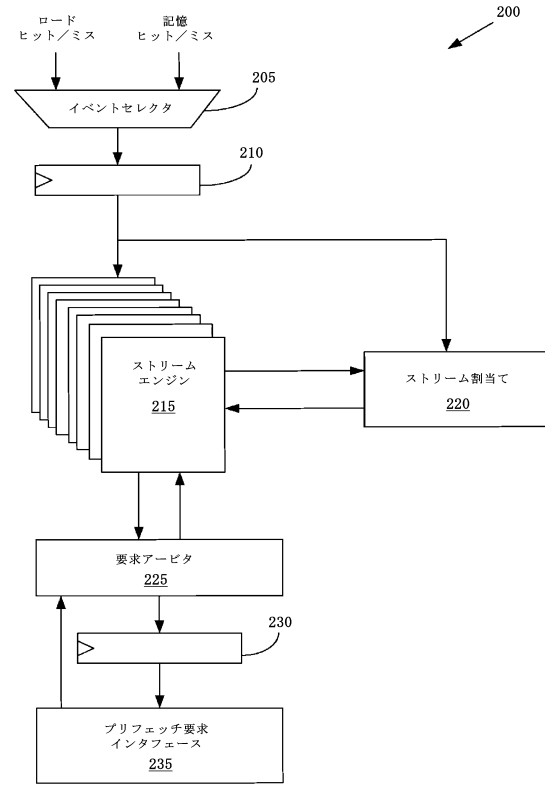
開示されている主題は、本書の教示の利点を有する当業者にとって明らかな方法、または、異なっているが同等の方法で修正され、実践されてよいので、上記に開示された特定の実施形態は実例的にすぎない。さらに、以下の特許請求の範囲に説明される以外、ここで示される構造または設計の詳細に対する制限は意図されていない。したがって、上記に開示された特定の実施形態が改変または修正されてよく、係る全ての変形態様が、開示されている主題の範囲内にあるとみなされることは明白である。したがって、ここで求められる保護は、以下の特許請求の範囲に述べられるとおりである。

30

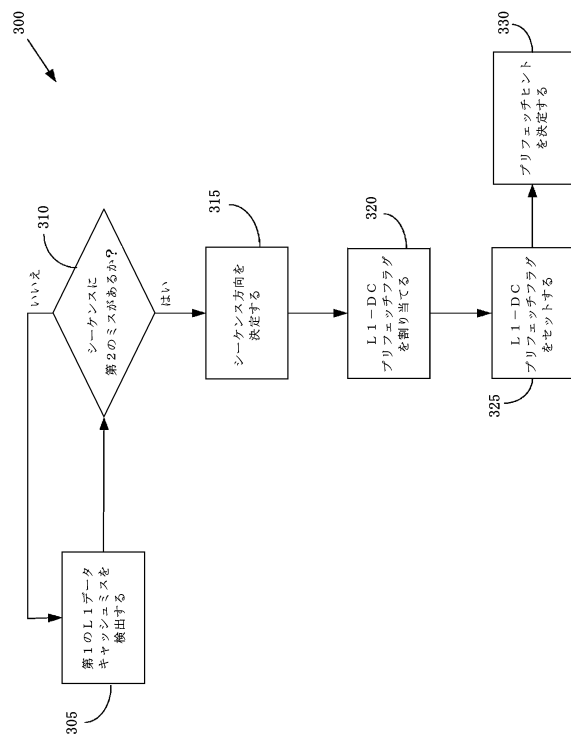
【図 1】



【図 2】



【図 3】



【図 4】

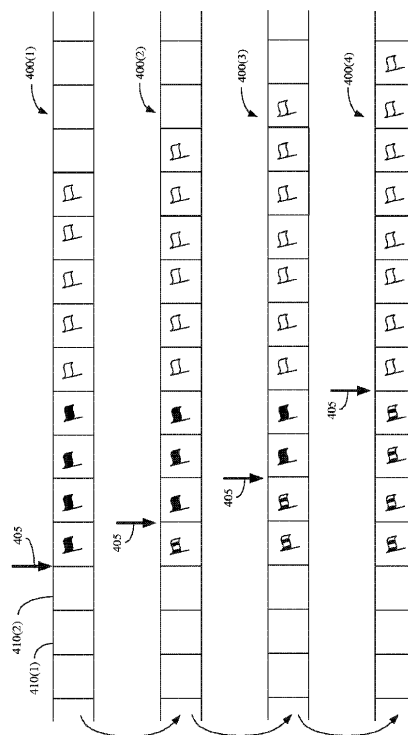
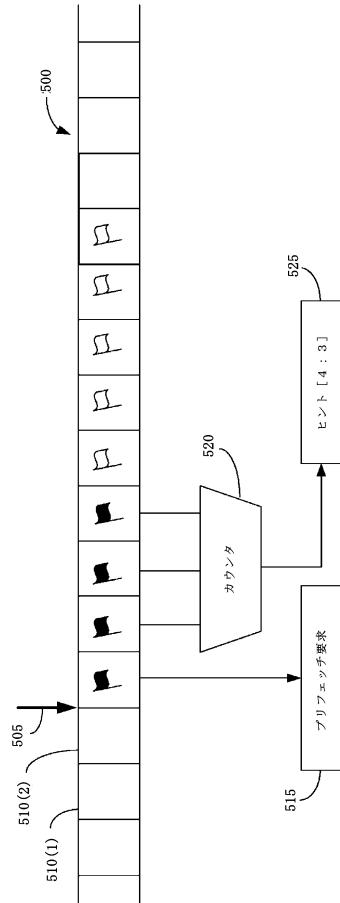
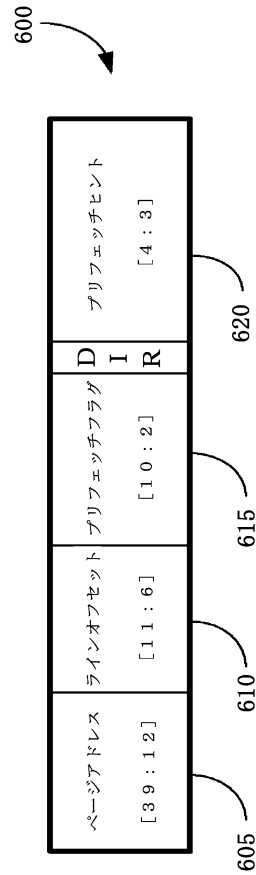


Figure 4

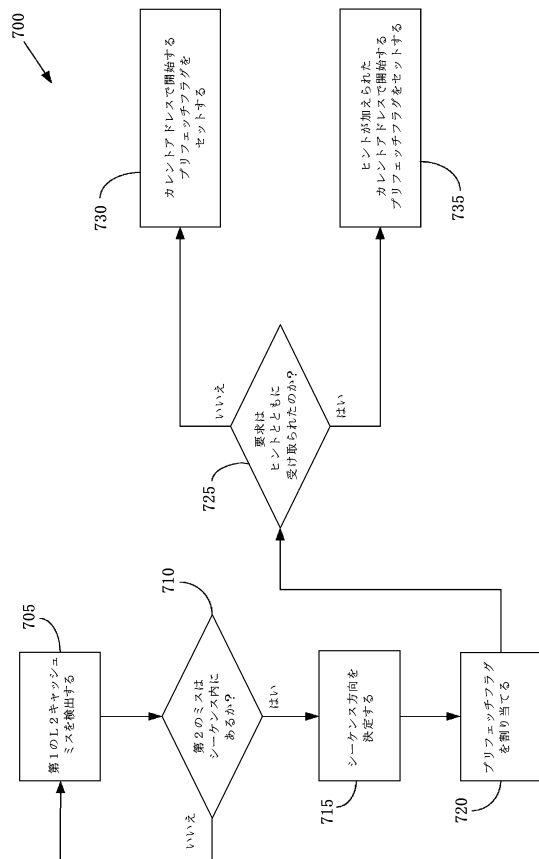
【図 5】



【図 6】



【図 7】



【図 8 A】

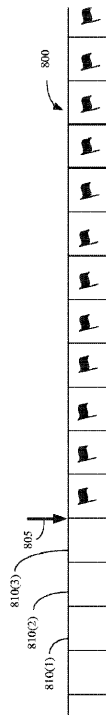


Figure 8A

【 図 8 B 】

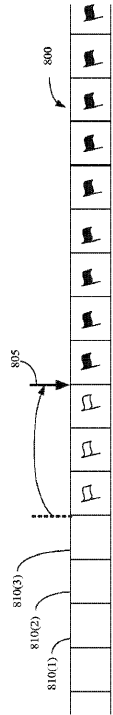


Figure 8B

フロントページの続き

(74)代理人 100162156

弁理士 村雨 圭介

(72)発明者 ドナルド ダブリュ . マッコーリー

アメリカ合衆国 7 8 7 3 4 テキサス州、レイクウェイ、エッジウォーター コープ 1 0 2

(72)発明者 ステファン ピー . トンプソン

アメリカ合衆国 8 0 5 0 3 コロラド州、ロングモント、プリンストン ドライブ 1 0 3 9

審査官 後藤 彰

(56)参考文献 米国特許第 0 6 4 4 6 1 6 7 (U S , B 1)

米国特許出願公開第 2 0 0 5 / 0 0 7 1 5 4 2 (U S , A 1)

米国特許出願公開第 2 0 1 2 / 0 0 8 4 5 1 1 (U S , A 1)

特開 2 0 0 0 - 3 3 9 1 5 7 (J P , A)

特開 2 0 0 0 - 2 4 2 5 5 7 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

G 0 6 F 1 2 / 0 8