

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2007-207240

(P2007-207240A)

(43) 公開日 平成19年8月16日(2007.8.16)

(51) Int. Cl.	F I	テーマコード (参考)
G06F 12/08 (2006.01)	G06F 12/08 505B	5B005
	G06F 12/08 509Z	
	G06F 12/08 511B	
	G06F 12/08 515Z	

審査請求 未請求 請求項の数 27 O L (全 27 頁)

(21) 出願番号 特願2007-19613 (P2007-19613)
 (22) 出願日 平成19年1月30日 (2007.1.30)
 (31) 優先権主張番号 11/347414
 (32) 優先日 平成18年2月3日 (2006.2.3)
 (33) 優先権主張国 米国 (US)

(71) 出願人 390009531
 インターナショナル・ビジネス・マシー
 ズ・コーポレーション
 INTERNATIONAL BUSIN
 ESS MASCHINES CORPO
 RATION
 アメリカ合衆国10504 ニューヨーク
 州 アーモンク ニュー オーチャード
 ロード
 (74) 代理人 100108501
 弁理士 上野 剛史
 (74) 代理人 100112690
 弁理士 太佐 種一
 (74) 代理人 100091568
 弁理士 市位 嘉宏

最終頁に続く

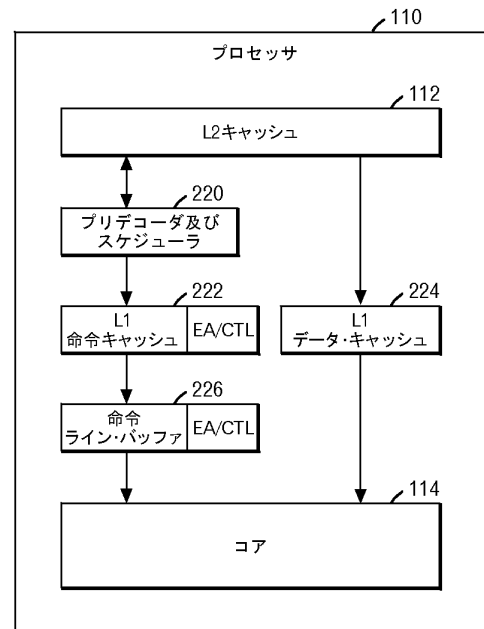
(54) 【発明の名称】 データ・ラインのための自己プリフェッチL2キャッシュ機構

(57) 【要約】

【課題】 キャッシュされたメモリを用いるプロセッサにおいて命令及びデータを取り出すことについて改善された方法を提供すること。

【解決手段】 方法は、レベル2キャッシュから第1命令ラインをフェッチするステップと、第1の命令ラインから、第1の命令ラインまたは異なる命令ラインに含まれているデータ・アクセス命令のターゲットとされたデータを含む第1のデータ・ラインの識別アドレスを抽出するステップと、抽出されたアドレスを用いて、レベル2キャッシュから第1のデータ・ラインをプリフェッチするステップを含む。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

データ・ラインをプリフェッチする方法であって、

(a) レベル 2 キャッシュから第 1 の命令ラインをフェッチするステップと、

(b) 前記第 1 の命令ラインから、当該第 1 の命令ラインまたは異なる命令ラインに含まれているデータ・アクセス命令のターゲットとされたデータを含む第 1 のデータ・ラインの識別アドレスを抽出するステップと、

(c) 前記抽出されたアドレスを用いて、前記レベル 2 キャッシュから第 1 のデータ・ラインをプリフェッチするステップと、

を含む方法。

10

【請求項 2】

前記第 1 の命令ラインにおいて、当該第 1 の命令ラインの外部にある命令をターゲットとする分岐命令を識別するステップと、

前記識別された分岐命令に対応する出口アドレスを抽出するステップと、

前記抽出された出口アドレスを用いて、前記レベル 2 キャッシュから、前記ターゲットとされた命令を含む第 2 の命令ラインをプリフェッチするステップと、

をさらに含む、請求項 1 に記載の方法。

【請求項 3】

前記第 2 の命令ラインについて (a) から (c) の前記ステップを繰り返し、第 2 のデータ・アクセス命令のターゲットとされた第 2 のデータを含む第 2 のデータ・ラインをプリフェッチするステップをさらに含む、請求項 2 の記載の方法。

20

【請求項 4】

前記第 2 のデータ・アクセス命令が前記第 2 の命令ラインに存在する、請求項 3 に記載の方法。

【請求項 5】

前記第 2 のデータ・アクセス命令が前記第 1 の命令ラインに存在する、請求項 3 に記載の方法。

【請求項 6】

所定の数のデータ・ラインがプリフェッチされるまで (a) から (c) の前記ステップを繰り返すことをさらに含む、請求項 1 に記載の方法。

30

【請求項 7】

第 1 の命令ラインにおいて、第 2 のデータをターゲットとする第 2 のデータ・アクセス命令を識別するステップと、

前記識別された第 2 のデータ・アクセス命令から第 2 のアドレスを抽出するステップと、

前記抽出された第 2 のアドレスを用いて、前記レベル 2 キャッシュから、前記ターゲットとされた第 2 のデータを含む第 2 のデータ・ラインをプリフェッチするステップと、

をさらに含む、請求項 1 に記載の方法。

【請求項 8】

前記抽出されたアドレスが、命令ラインに含まれる有効アドレスとして格納される、請求項 1 に記載の方法。

40

【請求項 9】

前記命令ラインが第 1 の命令ラインである、請求項 8 に記載の方法。

【請求項 10】

前記有効アドレスが、前の識別された分岐命令の実行の間に算出される、請求項 8 に記載の方法。

【請求項 11】

前記第 1 の命令ラインが、2 以上のデータをターゲットとする 2 以上のデータ・アクセス命令を含み、前記第 1 の命令ラインに格納されているデータ・アクセス履歴値が、識別されたデータ・アクセス命令がキャッシュ・ミスを引き起こすと予測されることを示す、

50

請求項 1 に記載の方法。

【請求項 1 2】

前記第 1 の命令ラインにおいて、第 1 のデータをターゲットとするデータ・アクセス命令を識別することをさらに含む、請求項 1 に記載の方法。

【請求項 1 3】

レベル 2 キャッシュと、

各々が 1 以上の命令を含む命令ラインを前記レベル 2 キャッシュから受信するように構成されたレベル 1 キャッシュと、

前記レベル 1 キャッシュから取り出された命令を実行するように構成されたプロセッサ・コアと、

(a) レベル 2 キャッシュから第 1 の命令ラインをフェッチし、

(b) 前記第 1 の命令ラインにおいて、前記第 1 の命令ラインまたは異なる命令ラインに含まれているデータ・アクセス命令のターゲットとされたデータを含む第 1 のデータ・ラインの識別アドレスを識別し、

(c) 前記抽出されたアドレスを用いて、前記レベル 2 キャッシュから第 1 のデータ・ラインをプリフェッチするように構成された回路と、

を備える、プロセッサ。

10

【請求項 1 4】

前記抽出されたアドレスが、命令ラインに含まれる有効アドレスとして格納される、請求項 1 3 に記載のプロセッサ。

20

【請求項 1 5】

前記有効アドレスが、前の識別された分岐命令の実行の間に算出される、請求項 1 4 に記載のプロセッサ。

【請求項 1 6】

前記有効アドレスが、トレーニング・フェーズの間に算出される、請求項 1 5 に記載のプロセッサ。

【請求項 1 7】

データ・ターゲット・アドレスを命令ラインに格納する方法であって、

命令ラインにある 1 以上の命令を実行するステップと、

前記 1 以上の命令がデータ・ラインの中のデータにアクセスし、キャッシュ・ミスをもたらしかどうかを判定するステップと、

30

キャッシュ・ミスをもたらしと判定された場合に、前記データ・ラインに対応するデータ・ターゲット・アドレスを、プリフェッチ機構によってアクセス可能な位置に格納するステップと、

を含む方法。

【請求項 1 8】

前記プリフェッチ機構によってアクセス可能な位置が前記命令ラインである、請求項 1 7 に記載の方法。

【請求項 1 9】

前記命令ラインおよび前記ターゲット・データ・アドレスを、レベル 2 キャッシュに書戻すことをさらに含む、請求項 1 8 に記載の方法。

40

【請求項 2 0】

前記データ・ターゲット・アドレスを伴う命令ラインをレベル 2 キャッシュに格納するステップと、

前記レベル 2 キャッシュから、前記データ・ターゲット・アドレスを伴う命令ラインをフェッチし、当該命令ラインをレベル 1 キャッシュに配置するステップと、

前記格納されたデータ・ターゲット・アドレスを用いて、前記データ・ラインをプリフェッチするステップと、

をさらに含む、請求項 1 8 に記載の方法。

【請求項 2 1】

50

前記プリフェッチ機構によってアクセス可能な位置はシャドウ・キャッシュである、請求項 17 に記載の方法。

【請求項 22】

前記プリフェッチ機構によってアクセス可能な位置に、前記 1 以上の命令に対応するデータ・アクセス履歴情報を格納するステップをさらに含む、請求項 17 に記載の方法。

【請求項 23】

前記命令ラインにある 1 以上の命令のその後の実行の間に、前記命令ラインにある第 2 の 1 以上の命令を実行するステップと、

前記第 2 の 1 以上の命令が第 2 のデータ・ラインの中のデータにアクセスし、当該アクセスが第 2 のキャッシュ・ミスをもたらす場合に、前記 1 以上の命令に対応する前記データ・アクセス履歴情報が、前記キャッシュ・ミスが予測可能であることを示すかどうかを判定するステップと、

前記キャッシュ・ミスが予測可能ではなかった場合には、前記第 2 のデータ・ラインに対応する前記命令ラインに、第 2 のデータ・ターゲット・アドレスを付加するステップと、

をさらに含む、請求項 22 に記載の方法。

【請求項 24】

前記データ・アクセス・アドレスを格納するステップは、複数の命令ラインが繰り返し実行される最初の実行フェーズの間に行われる、請求項 17 に記載の方法。

【請求項 25】

前記命令ラインにある第 2 の 1 以上の命令を実行するステップと、

前記第 2 の 1 以上の命令が、別の命令ラインにある命令へ分岐するかどうかを判定するステップと、

分岐すると判定された場合に、他の命令ラインに対応する出口アドレスを前記位置に格納するステップと、

をさらに含む、請求項 17 に記載の方法。

【請求項 26】

前記データ・アクセス・アドレスは、前記 1 以上の命令のうちの 1 つの実行の間に算出された有効アドレスである、請求項 17 に記載の方法。

【請求項 27】

命令ラインにある 1 以上の命令を実行するプロセッサ・コアと、

前記 1 以上の命令がデータ・ラインの中のデータにアクセスし、キャッシュ・ミスをもたらすかどうかを判定し、キャッシュ・ミスをもたらすと判定された場合に、前記データ・ラインに対応するデータ・ターゲット・アドレスを、プリフェッチ機構によってアクセス可能な位置に格納する回路と、

を備える、プロセッサ。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般に、コンピュータ・プロセッサの分野に関する。より具体的には、本発明は、コンピュータ・プロセッサによって用いられるキャッシュ機構に関する。

【背景技術】

【0002】

現代のコンピュータ・システムは、一般的に、コンピュータ・システム内の情報を処理するために使用することができるプロセッサを含む、数個の集積回路 (IC) を備える。プロセッサによって処理されたデータは、プロセッサによって実行されるコンピュータ命令を含み、さらに、そのコンピュータ命令を用いるプロセッサによって取り扱われるデータを含むことができる。一般的に、コンピュータ命令およびデータは、コンピュータ・システムのメイン・メモリに格納される。

【0003】

10

20

30

40

50

プロセッサは、一般的に、一連の小ステップに分けて命令を実行することによって命令を処理する。プロセッサによって処理される命令の数を増やす（そしてプロセッサの速度を上げる）ために、プロセッサをパイプラインすることができる場合がある。ここで、「パイプラインすること」とは、各々が1つの命令を実行するために必要な小ステップの1以上を行う独立した段を、プロセッサに提供することをいう。パイプラインは（他の回路に加えて）、プロセッサ・コアと呼ばれるプロセッサの一部に配置することができる。プロセッサは、複数のプロセッサ・コアを有することができる場合がある。

【0004】

パイプラインにおける命令実行の1つの例として、第1の命令が受信されたときに、第1のパイプライン段はその命令の小部分を処理することができる。第1のパイプライン段がその命令の小部分の処理を終了したときに第2のパイプライン段が第1の命令の別の小部分の処理を開始することができ、その間に第1のパイプライン段は第2の命令を受信し、その小部分の処理を開始する。このようにして、プロセッサは、2以上の命令を同時に（並行して）処理することができる。

10

【0005】

データおよび命令へのより高速のアクセス、および、プロセッサのより良好な活用を提供するために、プロセッサは、数個のキャッシュを有することができる。キャッシュとは、一般的に、メイン・メモリより小さく、プロセッサと同じダイ（すなわちチップ）上に製造されるメモリである。現代のプロセッサは、一般的に、複数のレベルのキャッシュを有する。プロセッサ・コアの最も近辺に設置される最も高速のキャッシュは、レベル1キャッシュ（L1キャッシュ）と呼ばれる。プロセッサは、一般的に、L1キャッシュに加えて、レベル2キャッシュ（L2キャッシュ）と呼ばれる、第2のより大型のキャッシュを有する。プロセッサは、他の付加的なレベルのキャッシュを有することもある（例えば、L3キャッシュおよびL4キャッシュ）。

20

【0006】

プロセッサにプロセッサのパイプラインの各段を満たすのに十分な命令を与えるために、プロセッサは、L2キャッシュから命令ラインと呼ばれる複数の命令を含んだグループの形で命令を取り出すことができる。取り出された命令ラインはL1命令キャッシュ（I-キャッシュ）に配置することができ、プロセッサ・コアは命令ラインにある命令にアクセスすることができる。同様に、プロセッサによって処理されるべきデータ・ブロックは、L2キャッシュから取り出し、L1キャッシュ・データ・キャッシュ（D-キャッシュ）に配置することができる。

30

【0007】

より上位レベルのキャッシュから情報を取り出し、その情報をより下位レベルのキャッシュに配置するプロセスは「フェッチする」と称され、一般的に、ある程度の時間（待ち時間）を要する。例えば、プロセッサ・コアが情報を要求し、その情報がL1キャッシュに存在しなかった場合（これはキャッシュ・ミスと呼ばれる）、その情報をL2キャッシュからフェッチすることができる。要求された情報のために次のレベルのキャッシュ/メモリが検索されることから、キャッシュ・ミスは付加的な待ち時間をもたらす。例えば、要求された情報がL2キャッシュに存在しなかった場合には、プロセッサは、L3キャッシュやメイン・メモリの中にその情報を探すであろう。

40

【0008】

プロセッサは、ある命令およびデータを、キャッシュおよび/またはメモリから取り出される命令およびデータよりも速く処理することができる場合がある。例えば、命令ラインが処理された後で、処理されるべき次の命令ラインにアクセスするために時間がかかる場合がある（例えば、次の命令を含む命令ラインを求めてL1キャッシュが検索された時に、キャッシュ・ミスがあった場合など）。プロセッサがより上位レベルのキャッシュまたはメモリから次の命令ラインを取り出している間に、パイプライン段は前の命令の処理を終了し、それ以上処理する命令が残っていない場合がある（これは、パイプライン停止と呼ばれる）。パイプラインが停止したときには、プロセッサは活用されておらず、パイ

50

ブラインされたプロセッサ・コアによって得られる利益を失うことになる。

【0009】

命令（そして命令ライン）は、一般的に、逐次的に処理されることから、プロセッサは、逐次的にアドレスされた命令ラインのブロックをフェッチすることによって、パイプライン停止を防ぐことを試みる場合がある。逐次的にアドレスされた命令ラインのブロックをフェッチすることにより、次の命令ラインは必要とされた時に既にL1キャッシュにおいて利用可能となることができ、したがって、プロセッサ・コアは、現行の命令ラインにある命令の処理を終了したときに、次の命令ラインにある命令に容易にアクセスすることができる。

【0010】

逐次的にアドレスされた命令ラインのブロックをフェッチしても、パイプライン停止を防ぐことができない場合がある。例えば、出口分岐（exit branch）命令と呼ばれる命令は、プロセッサを、逐次的にアドレスされた命令ラインのブロックの外部にある命令（ターゲット命令と呼ばれる）へ分岐させる場合がある。ここで、出口分岐命令は、現行の命令ラインに存在しないかまたは次の逐次的にアドレスされ既にフェッチされた命令ラインに存在するターゲット命令へ分岐する場合がある。したがって、分岐が行われるとプロセッサが判定したときに、出口分岐のターゲット命令を含む次の命令ラインは、L1キャッシュにおいて利用可能とならない場合がある。その結果として、パイプラインが停止し、プロセッサが非効率的に動作する場合がある。

【0011】

データをフェッチすることに関して、命令がデータにアクセスする場合に、プロセッサが、L1キャッシュにあるデータを含むデータ・ラインを見つけることを試みる場合がある。データ・ラインがL1キャッシュにおいて見つけれなかった場合には、所望のデータ・ラインを求めてL2キャッシュ、および、上位のメモリが検索されている間、プロセッサが停止する場合がある。所望のデータのアドレスは命令が実行されるまで知られないことから、プロセッサは、その命令が実行されるまで所望のデータ・ラインを検索することができない場合がある。プロセッサが実際にデータ・ラインを検索したときに、キャッシュ・ミスが起こってパイプラインが停止することがある。

【0012】

プロセッサは、現行のアクセスされているデータ・アドレスに近い（連続する）データ・アドレスを含むデータ・ラインのブロックをフェッチすることによって、こうしたキャッシュ・ミスを防ぐようにされる場合がある。近傍のデータ・ラインをフェッチすることは、あるデータ・ラインに存在するデータ・アドレスがアクセスされたときに、その近傍のデータ・アドレスもまたアクセスされる可能性が高いという仮定に依拠している（この仮定は、一般に、「参照の局所性」と呼ばれる）。しかし、この仮定が正しくない場合もあり、そのような場合、命令により現行のデータ・ラインの近傍に位置していないデータ・ラインにあるデータがアクセスされることによって、キャッシュ・ミスおよびプロセッサの非効率性がもたらされる。

【発明の開示】

【発明が解決しようとする課題】

【0013】

したがって、キャッシュされたメモリを用いるプロセッサにおいて命令およびデータを取り出すことについて改善された方法に対する必要性が存在する。

【課題を解決するための手段】

【0014】

本発明の実施形態は、データ・ラインをプリフェッチするための方法および装置を提供する。方法は、レベル2キャッシュから第1の命令ラインをフェッチするステップと、第1の命令ラインから、第1の命令ラインまたは異なる命令ラインに含まれているデータ・アクセス命令のターゲットとされたデータを含む第1のデータ・ラインの識別アドレスを抽出するステップと、抽出されたアドレスを用いて、レベル2キャッシュから第1のデー

10

20

30

40

50

タ・ラインをプリフェッチするステップとを含む。

【0015】

ある実施形態においては、プロセッサが提供される。プロセッサは、レベル1キャッシュと、レベル2キャッシュと、プロセッサ・コアと、回路とを備える。レベル1キャッシュは、レベル2キャッシュから、1以上の命令を含む命令ラインを受信するように構成される。プロセッサ・コアは、レベル1キャッシュから取り出された命令を実行するように構成される。回路は、レベル2キャッシュから第1の命令ラインをフェッチし、第1の命令ラインまたは異なる命令ラインに含まれているデータ・アクセス命令のターゲットとされたデータを含む第1のデータ・ラインの識別アドレスをその第1の命令ラインにおいて識別し、抽出されたアドレスを用いて、レベル2キャッシュから第1のデータ・ラインをプリフェッチするように構成される。

10

【0016】

ある実施形態においては、命令ラインにデータ・ターゲット・アドレスを格納する方法が提供される。方法は、命令ラインにある1以上の命令を実行するステップと、1以上の命令がデータ・ラインの中のデータにアクセスし、キャッシュ・ミスをもたらすかどうかを判定するステップと、キャッシュ・ミスをもたらすと判定された場合に、データ・ラインに対応するデータ・ターゲット・アドレスを、プリフェッチ機構によってアクセス可能な位置に格納するステップとを含む。

【0017】

ある実施形態においては、プロセッサが提供される。プロセッサは、命令ラインにある1以上の命令を実行するプロセッサ・コアと、1以上の命令がデータ・ラインの中のデータにアクセスし、キャッシュ・ミスをもたらすかどうかを判定し、キャッシュ・ミスをもたらすと判定された場合に、データ・ラインに対応するデータ・ターゲット・アドレスを、プリフェッチ機構によってアクセス可能な位置に格納する回路を備える。

20

【発明を実施するための最良の形態】

【0018】

本発明の上記の特徴、利点、および目的が得られる手法が詳細に理解できるように、上記で要約された本発明のより具体的な説明が、添付の図面に示されている本発明の実施形態を参照することによってなされる。

【0019】

しかし、添付の図面は本発明の典型的な実施形態のみを図示するものであって、本発明の範囲を限定するものと見なされるべきではないことに留意すべきである。

30

【0020】

本発明の実施形態は、データ・ラインをプリフェッチするための方法および装置を提供する。本発明の実施形態においては、データ・ラインにあるデータをターゲットにするデータ・アクセス命令（例えばロード命令または格納命令）について、フェッチされている命令ラインを検査することができる。このようなデータ・アクセス命令のターゲット・データ・アドレスを抽出して、ターゲットにされたデータを含むデータ・ラインをL2キャッシュからプリフェッチするために用いることができる。その結果として、データ・ターゲティング命令が実行された場合に、ターゲットにされたデータ・ラインは、既にL1データ・キャッシュ（「D-キャッシュ」）に存在するようにすることができ、そうすることによって、コストのかかるデータ・キャッシュ・ミスをなくして全体の性能を向上させることができる。

40

【0021】

本発明の実施形態においては、プリフェッチ・データ（例えば、ターゲット・アドレス）は、プリフェッチ・データが属する（例えば命令ラインまたはデータ・ラインに付加された）対応する情報ブロックの形で従来のキャッシュ・メモリに格納することができる。例えば、対応する情報のラインがキャッシュ・メモリからフェッチされる際に、そこに含まれているプリフェッチ・データを検査して、他の関連する情報のラインをプリフェッチするために用いることができる。次いで、各々の他のプリフェッチされた情報のラインに

50

格納されたプリフェッチ・データを用いて、同様のプリフェッチを行うことができる。フェッチされた命令ライン内の情報を用いて、命令ラインにある命令によってターゲットとされたデータを含むデータ・ラインをプリフェッチすることにより、フェッチされた情報ブロックに関連付けられたキャッシュ・ミスを防ぐことができる。

【0022】

本発明の実施形態によれば、プリフェッチ・データを命令ラインの一部としてキャッシュに格納することにより、プリフェッチ・データおよび予測データを排他的に格納する特別なキャッシュまたはメモリを不要とすることができる。しかし、上述されたように、かかる情報は、そうした履歴情報の格納を専門とする特別なキャッシュまたはメモリを含むいかなる位置に格納してもよい。また、異なるキャッシュ（およびキャッシュ・ライン）、バッファ、専用キャッシュ、および他の位置の組み合わせを用いて、ここに説明された履歴情報を格納することができる場合がある。

10

【0023】

以下は、添付の図面に図示された本発明の実施形態の詳細な説明である。本発明の実施形態はあくまで例示であり、本発明を明確に教示するため、詳細なものとなっている。しかし、開示された詳細な説明は、実施形態の予期される変形物を排除することを意図しておらず、逆に、添付の特許請求の範囲によって定められる本発明の精神および範囲内にある全ての修正物、均等物、および代替物を包含することにあることを意図している。

【0024】

本発明の実施形態は、例えばコンピュータ・システムなどのシステムとともに用いることができ、また、そうしたシステムに関して説明される。本明細書において用いられるシステムという用語は、パーソナル・コンピュータ、インターネット機器、デジタル・メディア機器、携帯情報端末（PDA）、携帯型音楽/映像再生機、およびビデオゲーム機を含めた、プロセッサおよびキャッシュ・メモリを用いるいかなるシステムも含むことができる。キャッシュ・メモリは、そのキャッシュ・メモリを使用するプロセッサと同じダイに設置することができるが、プロセッサとキャッシュ・メモリは、異なるダイ（例えば、独立したモジュール内の独立したチップ、または単一のモジュール内の独立したチップ）に設置することもできる。

20

【0025】

本発明の実施形態は、各々がパイプラインを用いて命令を実行する複数のプロセッサ・コアと複数のL1キャッシュとを有するプロセッサに関して以下で説明されるが、単一の処理コアを有するプロセッサおよび/または命令の実行にパイプラインを用いないプロセッサを含む、キャッシュを用いるいずれかのプロセッサとともに用いることもできる。一般的に、本発明の実施形態は、いかなるプロセッサとも用いることができ、特定の構成に限定されない。

30

【0026】

本発明の実施形態は、以下で、L1命令キャッシュ（L1 I - キャッシュ222）とL1データ・キャッシュ（L1 D - キャッシュ224）に分割されたL1キャッシュを有するプロセッサに関して説明されるが、統合されたL1キャッシュが用いられるような構成において用いることができる。さらに、本発明の実施形態は、以下で、命令ラインおよびデータ・ラインをL2キャッシュからプリフェッチすること、および、プリフェッチされたラインをL1キャッシュに配置することに関して説明されるが、いずれかのキャッシュまたはメモリ・レベルから別のいずれかのキャッシュまたはメモリ・レベルへ命令ラインおよびデータ・ラインをプリフェッチするために用いることもできる。

40

【0027】

（例示的なシステムの概要）

図1は、本発明の実施形態によるシステム100を図示するブロック図である。システム100は、命令およびデータを格納するためのシステム・メモリ102、グラフィックス処理のためのグラフィックス処理装置104、外部デバイスと通信するためのI/Oインターフェース、命令およびデータを長期格納するためのストレージ・デバイス108、

50

および、命令およびデータを処理するためのプロセッサ 110 を含むことができる。

【0028】

本発明の実施形態により、プロセッサ 110 は、L2 キャッシュ 112、複数の L1 キャッシュ 116 を有することができ、L1 キャッシュ 116 の各々は、複数のプロセッサ・コア 114 の 1 つによって用いられる。ある実施形態によれば、プロセッサ・コア 114 の各々はパイプラインすることができ、命令の各々は一連の小ステップの形で実施され、ステップの各々は異なるパイプライン段によって実施される。

【0029】

図 2 は、本発明の実施形態によるプロセッサ 110 を図示するブロック図である。簡略化のために、図 2 は、プロセッサ 110 の単一のコア 114 を図示し、これに関して説明 10 される。ある実施形態においては、コア 114 の各々は、同一のものとしてすることができる（例えば、同一のパイプライン段を伴う同一のパイプラインを含む）。別の実施形態においては、コア 114 の各々は異なるものとしてすることができる（例えば、異なる段を伴う異なるパイプラインを含む）。

【0030】

本発明の実施形態においては、L2 キャッシュは、プロセッサ 110 によって使用される命令およびデータの一部を格納することができる。プロセッサ 110 は、L2 キャッシュ 112 に格納されていない命令およびデータを要求する場合がある。要求された命令およびデータが L2 キャッシュ 112 に格納されていない場合には、要求された命令およびデータを（上位のレベルのキャッシュまたはシステム・メモリ 102 から）取り出して、 20 L2 キャッシュに配置することができる。プロセッサ・コア 114 が L2 キャッシュ 112 から命令を要求したときに、その命令を、まずプリデコーダおよびスケジューラ 220 で処理することができる（以下でより詳細に説明する）。

【0031】

本発明の実施形態においては、図 1 に図示された L1 キャッシュ 116 は、2 つの部分、すなわち、命令ラインを格納するための L1 命令キャッシュ 222（L1 I - キャッシュ）222、および、データ・ラインを格納するための L1 データ・キャッシュ（L1 D - キャッシュ）224 に分割することができる。L2 キャッシュ 112 から取り出された命令ラインがプリデコーダおよびスケジューラ 220 によって処理された後で、命令ラインを命令キャッシュ 222 に配置することができる。同様にして、L2 キャッシュ 1 30 112 からフェッチされたデータ・ラインは、データ・キャッシュ 224 に配置することができる。L2 キャッシュ 112 にある情報のラインが命令ラインであるかデータ・ラインであるかを追跡するために、命令ラインおよびデータ・ラインの各々の 1 ビットを使うことができる。

【0032】

本発明の実施形態においては、命令は、命令ラインと呼ばれるグループとして L2 キャッシュ 112 および命令キャッシュ 222 からフェッチして、プロセッサ・コア 114 が命令ラインにある命令にアクセスすることができる命令ライン・バッファ 226 に配置することができる。同様にして、データは、データ・ラインと呼ばれるグループとして L2 40 キャッシュ 112 およびデータ・キャッシュ 224 からフェッチすることができる。ある実施形態においては、命令キャッシュ 222 および命令ライン・バッファ 226 の一部は、有効アドレスおよび制御ビット（E A / C T L）を格納するために用いることができ、それらは、命令ラインの各々を処理して、例えば以下で説明されるデータ・プリフェッチ機構を実行するために、コア 114 および / またはプリデコーダおよびスケジューラ 220 によって用いられることができる。

【0033】

（L2 キャッシュからのデータ・ラインのプリフェッチ）

図 3 は、本発明の実施形態による、データ・ラインにあるデータ（D 4₁）をターゲットとするデータ・アクセス命令（I 5₁）を格納している例示的な命令ラインを示す図である。ある実施形態においては、命令ライン（命令ライン 1）は、複数の命令（例えば、 50

I₁₁、I₂₁、I₃₁、など)、並びに、有効アドレスおよび制御ビットなどの制御情報を格納することができる。同様にして、データ・ライン(データ・ライン1)は、複数のデータ・ワード(例えば、D₁₁、D₂₁、D₃₁、など)を格納することができる。ある規模においては、命令ラインの各々にある命令は順番に実行され、それにより、命令I₁₁が第一に実行され、I₂₁が第二に実行され、以下続くことになる。命令が順番に実行されるので、一般的に命令ラインもまた順番に実行される。したがって、命令ラインがL2キャッシュ112から命令キャッシュ222に移動されるたびに、プリデコーダおよびスケジューラ220は、命令ライン(例えば命令ライン1)を検査して、次の逐次的な命令ライン(例えば命令ライン2)をプリフェッチし、そのことにより、次の命令ラインが命令キャッシュ222に配置され、プロセッサ・コア114によりアクセスできるようにすることができる場合がある。

10

【0034】

プロセッサ・コア114によって実行される命令ラインは、命令I₅₁のようなデータ・アクセス命令(例えば、ロード命令または格納命令)を含むことができる。データ・アクセス命令は、あるアドレス(例えばD₄₁)に存在するデータをターゲットとして(例えばロードまたは格納の)動作を行う。データ・アクセス命令は、データ・アドレスを、(例えばデータ・レジスタに格納されたアドレスなどの)何らかの他のアドレスからのオフセットとして要求することがあり、それにより、データ・アドレスは、データ・アクセス命令が実行された時に算出される。

【0035】

命令I₅₁がプロセッサ・コア114によって実行されるときに、プロセッサ・コア114は、その命令によりデータD₄₁がアクセスされるかどうかを判定することができる。プロセッサ・コア114は、データD₄₁を含んだデータ・ライン(データ・ライン1)を、データ・キャッシュ224からフェッチすることを試みることができる。いくつかの場合においては、データ・ライン1は、データ・キャッシュ224に存在せず、それによりキャッシュ・ミスを引き起こす場合がある。データ・キャッシュにおいてキャッシュ・ミスが検出されたときに、L2キャッシュ112に対して、データ・ライン1を求めるフェッチ要求を発行することができる。フェッチ要求がL2キャッシュ112によって処理されている間、コア114のプロセッサ・パイプラインが停止し、それにより、プロセッサ・コア114による命令の処理を中止させる場合がある。データ・ライン1がL2キャッシュ112に存在しない場合には、データ・ラインが上位のレベルのキャッシュおよび/またはメモリからフェッチされる間、プロセッサ・パイプラインがより長期間停止する場合がある。

20

30

【0036】

本発明の実施形態によれば、現在フェッチされている命令ラインから抽出されたデータ・ターゲット・アドレスによってデータ・ラインをプリフェッチすることにより、データ・キャッシュ・ミスの数を減少させることができる。

【0037】

図4は、本発明の実施形態による、データ・キャッシュ・ミスを減少させるためのプロセス400を図示する流れ図である。プロセス400は、L2キャッシュ112から命令ラインがフェッチされるステップ404で開始することができる。ステップ406において、データ・アクセス命令を識別することができる。ステップ408において、データ・アクセス命令のターゲットとされたデータのアドレス(データ・ターゲット・アドレスと呼ばれる)を抽出することができる。次いで、ステップ410において、データ・ターゲット・アドレスを用いて、L2キャッシュ112からターゲットとされたデータを含むデータ・ラインをプリフェッチすることができる。ターゲットとされたデータを含むデータ・ラインをプリフェッチし、プリフェッチされたデータをデータ・キャッシュ224に配置することにより、データ・アクセス命令が実行された場合のキャッシュ・ミスを防ぐことができる。データ・ターゲット・アドレスは、実際にデータ・キャッシュ・ミスまたはデータ・キャッシュ・ミスの履歴が存在する場合にのみ格納することができる場合がある。

40

50

【0038】

ある実施形態においては、図5に図示されたように、データ・ターゲット・アドレスは、直接、命令ラインに格納する（付加する）ことができる。格納されるデータ・ターゲット・アドレスEA1は、有効アドレスまたは有効アドレスの一部（例えば、有効アドレスの高位の32ビット）とすることができる。図示されたように、データ・ターゲット・アドレスEA1は、データ・アクセス命令I5₁のターゲットとされたデータD4₁のアドレスを含むデータ・ラインを識別することができる。

【0039】

ある実施形態により、命令ラインはまた、他の有効アドレス（例えばEA2）および制御ビット（例えばCTL）を格納することもできる。以下で説明されるように、他の有効アドレスは、命令ラインまたは付加的なデータ・ラインにある分岐命令のターゲットとされた命令を含む命令ラインをプリフェッチするために用いることができる。制御ビットCTLは、データ・アクセス命令の履歴（DAH）を示す1以上のビット、および、データ・アクセス命令の位置（LOC）を示す1以上のビットを含むことができる。命令ラインに格納されるそのような情報の用途もまた以下で説明される。

【0040】

本発明の実施形態においては、ここで説明される有効アドレス・ビットおよび制御ビットは、他の形では使用されない命令ラインのビットに格納することができる。例えば、L2キャッシュ112にある情報のラインの各々は、異なるキャッシュ・レベルの間で転送されるデータのエラー訂正に用いることができる追加のデータ・ビット（例えば、転送されたデータが破損されていないことを保証し、実際に起こったいずれかの破損を修復するために用いられるエラー訂正コードECCなど）を有することができる。各々のレベルのキャッシュ（例えばL2キャッシュ112および命令キャッシュ222）は、各々の命令ラインの同一コピーを格納することができる。各々のレベルのキャッシュが所与の命令ラインの同一コピーを格納する場合には、ECCは、用いなくともよい。代わって、例えば、キャッシュ間で命令ラインが適切に転送されたかどうかを判定するために、例えばパリティ・ビットを用いることができる。キャッシュ間で命令ラインが不適切に転送されたことをパリティ・ビットが示した場合には、（キャッシュにそのラインが含まれているため）、エラー確認を行う代わりに転送キャッシュからその命令ラインを再フェッチすることができる。

【0041】

アドレスおよび制御情報を、他の形では用いられない命令ラインのビットに格納することの例として、格納された2ワードごとにエラー訂正用に11ビットを用いるエラー訂正プロトコルを考える。1つの命令ラインにおいて、2つの命令ごとにパリティ・ビットを格納するために、11ビットのうちの一つを用いることができる（1ワードごとに1つの命令が格納される場合）。残りの命令1つにつき5ビットを、各命令の制御ビットおよび/またはアドレス・ビットを格納するために用いることができる。例えば、5ビットのうち4ビットを、その命令についての履歴情報（例えば、その命令が前に取られた分岐命令であるか否か、または、その命令が前にデータ・キャッシュ・ミスを引き起こしたデータ・アクセス命令であるか否か）といった、命令の制御ビット（履歴ビットなど）を格納するために用いることができる。命令ラインが32の命令を含む場合には、残りの32ビット（命令の各々につき1ビット）は、例えばデータ・ターゲット・アドレスまたは分岐出口アドレスの全てまたは一部を格納するために用いることができる。

【0042】

（例示的なプリフェッチ回路）

図6は、本発明の実施形態による、命令およびデータ・ラインをプリフェッチするための回路を図示するブロック図である。本発明の実施形態においては、回路は、データ・ラインのみをプリフェッチすることができる。本発明の別の実施形態においては、回路は、命令ラインとデータ・ラインの両方をプリフェッチすることができる。

【0043】

命令ラインまたはデータ・ラインが L 2 キャッシュ 1 1 2 からフェッチされて命令キャッシュ 2 2 2 またはデータ・キャッシュ 2 2 4 に配置されるたびに、命令 / データ (I / D) によって制御される選択回路 6 2 0 は、フェッチされた命令ラインまたはデータ・ラインを、適切なキャッシュに送ることができる。

【 0 0 4 4 】

プリデコードおよびスケジューラ 2 2 0 は、L 2 キャッシュ 1 1 2 によって出力された情報を検査することができる。複数のプロセッサ・コア 1 1 4 が用いられるある実施形態においては、複数のプロセッサ・コアの間で単一のプリデコードおよびスケジューラ 2 2 0 を共有することができる。別の実施形態においては、プリデコードおよびスケジューラ 2 2 0 は、プロセッサ・コア 1 1 4 の各々に個別に提供される。

10

【 0 0 4 5 】

ある実施形態においては、プリデコードおよびスケジューラ 2 2 0 は、L 2 キャッシュ 1 1 2 によって出力された情報が命令ラインであるかデータ・ラインであるかを判定するプリデコード制御回路 6 1 0 を有することができる。例えば、L 2 キャッシュ 1 1 2 は、L 2 キャッシュ 1 1 2 に格納された情報ブロックの各々に特定のビットを設定することができ、プリデコード制御回路 6 1 0 はその特定のビットを検査して、L 2 キャッシュ 1 1 2 によって出力された情報ブロックが命令ラインであるかデータ・ラインであるかを判定することができる。

【 0 0 4 6 】

プリデコード制御回路 6 1 0 が、L 2 キャッシュ 1 1 2 によって出力された情報が命令ラインであると判定した場合、プリデコード制御回路 6 1 0 は、命令ライン・アドレス選択回路 6 0 4 およびデータ・ライン・アドレス選択回路 6 0 6 を使って、命令ラインに格納されるいずれかの適切な有効アドレス (例えば E A 1、または E A 2) を選択することができる。次いで、選択 (S E L) 信号を用い、選択回路 6 0 8 によって、有効アドレスを選択することができる。次いで、選択された有効アドレスを、例えば対応する命令ラインまたはデータ・ラインを L 2 キャッシュ 1 1 2 からプリフェッチするために使われる 3 2 ビットのプリフェッチ・アドレスとして、プリフェッチ回路 6 0 2 に出力することができる。

20

【 0 0 4 7 】

上述されたように、第 1 の命令ラインにあるデータ・ターゲット・アドレスは、第 1 のデータ・ラインをプリフェッチするために用いることができる。第 1 のフェッチされた命令ラインは、また、第 2 の命令ラインにあるターゲット命令へ分岐する分岐命令 (出口分岐命令と呼ばれる) を格納することができる場合がある。ある実施形態においては、第 2 の命令ラインに対応するアドレス (出口アドレスと呼ばれる) はまた、第 1 のフェッチされた命令ラインに格納することもできる。第 1 の命令ラインがフェッチされたときに、格納された出口アドレスを用いて、第 2 の命令ラインをプリフェッチすることができる。第 1 の命令ラインにある分岐が続行され、第 2 の命令ラインにあるターゲット命令が命令キャッシュから要求された場合には、第 2 の命令ラインをプリフェッチすることによって命令キャッシュ・ミスをなくすことができる。

30

【 0 0 4 8 】

したがって、フェッチされた単一の命令ラインに基づいて、命令ラインおよびデータ・ラインのグループ (チェーン) を命令キャッシュ 2 2 2 およびデータ・キャッシュ 2 2 4 へプリフェッチすることができる場合があり、そのことにより、フェッチされた命令ラインまたはプリフェッチされた命令ラインにある出口分岐命令またはデータ・アクセス命令が命令キャッシュ・ミスまたはデータ・キャッシュ・ミスを引き起こす機会を減少させることができる。

40

【 0 0 4 9 】

出口アドレスによって示された第 2 の命令ラインが L 2 キャッシュ 1 1 2 からプリフェッチされたときに、第 2 の命令ラインを検査して、その第 2 の命令ラインが、第 2 の命令ライン内のデータ・アクセス命令によってアクセスされた第 2 のデータ・ラインに対応す

50

るデータ・ターゲット・アドレスを格納するか否かを判定することができる。プリフェッチされた命令ラインが第2のデータ・ラインに対応するデータ・ターゲット・アドレスを格納する場合には、第2のデータ・ラインをプリフェッチすることもできる。

【0050】

ある実施形態においては、プリフェッチされた第2の命令ラインは、やはりプリフェッチすることができる第3の命令ラインの有効アドレスを格納することができる。ここでも、第3の命令ラインは、プリフェッチすることができるターゲットデータ・ラインの有効アドレスを格納することができる。命令ラインおよび対応するデータ・ラインをプリフェッチするプロセスは、繰り返すことができる。プリフェッチされた命令ラインの各々は、メイン・メモリからプリフェッチされるべき複数の命令ラインおよび/または複数のデータ・ラインのどちらについての有効アドレスも格納することができる。

10

【0051】

例示として、ある実施形態においては、データ・キャッシュ224は、L2キャッシュ112から同時に2つのデータ・ラインがフェッチされ、2ポート・データ・キャッシュに配置されるように、2ポート・キャッシュとすることができる。そのような構成が用いられる場合、2つのデータ・ラインに対応する2つの有効アドレスを命令ラインの各々に格納することができ、その命令ラインがL2キャッシュ112からフェッチされた場合には、その有効アドレスを用いて両方のデータ・ラインを同時にL2キャッシュ112からプリフェッチし、データ・キャッシュ224に配置して、恐らくはデータ・キャッシュ・ミスをなくすことができる。

20

【0052】

したがって、フェッチされた単一の命令ラインに基づいて、命令ラインおよびデータ・ラインのグループ(チェーン)を、命令キャッシュ222およびデータ・キャッシュ224へプリフェッチすることができる場合があり、そのことにより、フェッチされた命令ラインまたはプリフェッチされた命令ラインにある出口分岐命令またはデータ・アクセス命令が命令キャッシュ・ミスまたはデータ・キャッシュ・ミスを引き起こす機会を減少させることができる。

【0053】

ある実施形態により、プリフェッチされた命令ラインがプリフェッチされるべき複数の有効アドレスを格納している場合には、プリフェッチ回路602に有効アドレスの各々が送信されている間、(例えば、プリデコーダ制御回路610、命令ライン・アドレス選択回路604、または何らかの他のバッファに)アドレスを一時的に格納することができる。別の実施形態においては、プリフェッチ・アドレスは、プリフェッチ回路602および/またはL2キャッシュ112に並列に送信することができる。

30

【0054】

プリフェッチ回路602は、要求された有効アドレスがL2キャッシュ112にあるか否かを判定することができる。例えば、プリフェッチ回路602は、要求された有効アドレスがL2キャッシュ112にあるか否かを判定することができる変換索引バッファ(TLB)のような内容アドレス可能メモリ(content addressable memory)(CAM)を格納することができる。要求された有効アドレスがL2キャッシュ112にあった場合には、プリフェッチ回路602は、要求された有効アドレスに対応する実アドレスをフェッチする要求を、L2キャッシュに発行することができる。次いで、実アドレスに対応する情報ブロックを選択回路620に出力し、適切なL1キャッシュ(例えば命令キャッシュ222、またはデータ・キャッシュ224)に向けることができる。プリフェッチ回路602が、要求された有効アドレスはL2キャッシュ112に存在しないと判定した場合には、プリフェッチ回路は、より上位レベルのキャッシュおよび/またはメモリに信号を送信することができる。例えば、プリフェッチ回路602はL3キャッシュにアドレスのプリフェッチ要求を送信することができ、次いでL3キャッシュは要求されたアドレスを検索される。

40

【0055】

50

プリデコーダおよびスケジューラ 220 が命令ラインまたはデータ・ラインを L2 キャッシュ 112 からプリフェッチすることを試みる前に、プリデコーダおよびスケジューラ 220 (または随意的にプリフェッチ回路 602) は、要求されプリフェッチされる命令ラインまたはデータ・ラインが既に命令キャッシュ 222 またはデータ・キャッシュ 224 に格納されているか否かを判定することができ、または、要求された命令ラインまたはデータ・ラインのプリフェッチ要求が既に発行済みであるか否かを判定することができる場合がある。例えば、最近フェッチされまたはプリフェッチされた命令ラインまたはデータ・ラインのアドレスの履歴を格納する小型キャッシュを用いて、命令ラインまたはデータ・ラインに関するプリフェッチ要求が既に発行済みであるか否かを判定することができ、または、要求された命令ラインまたはデータ・ラインが既に命令キャッシュ 222 またはデータ・キャッシュ 224 にあるか否かを判定することができる。

10

【0056】

要求された命令ラインまたはデータ・ラインが既に命令キャッシュ 222 またはデータ・キャッシュ 224 に設置されていた場合には、L2 キャッシュ・プリフェッチは必ずしも必要ではなく、したがって、実施されない場合もある。前のプリフェッチ要求によって第 2 のプリフェッチ要求が不必要にされる場合においては、現行の有効アドレスを命令ラインに格納することは、やはり必ずしも必要ではなく、命令ラインに他の有効アドレスが格納することが可能となる (以下で説明する)。

【0057】

本発明の実施形態においては、プリデコーダおよびスケジューラ 220 は、所定の数 (閾値数) の命令ラインおよび / またはデータ・ラインがフェッチされるまで、命令ライン (およびデータ・ライン) のプリフェッチを続行することができる。閾値は、いずれかの適切な方法で選択することができる。例えば、閾値は、命令キャッシュに配置することができる命令ラインの数および / またはデータ・キャッシュに配置することができるデータ・ラインの数に基づいて選択することができる。命令キャッシュおよび / またはデータ・キャッシュがより大きな容量を有する場合には大きなプリフェッチ所定の数を選択することができるが、命令キャッシュおよび / またはデータ・キャッシュがより小さな容量を有する場合には小さなプリフェッチ所定の数を選択することができる。

20

【0058】

別の例として、命令ライン・プリフェッチの所定の数は、フェッチされた命令ライン内の条件付き分岐命令の予測可能性に基づいて選択することができる。いくつかの場合においては、条件付き分岐命令の結果 (その分岐が取られるか否か) が予測可能となることがあり、したがって、プリフェッチすべき適正な命令ラインが予測可能となることがある。しかし、命令ラインの間の分岐予測の数が増大するに従い、全体の予測精度は小さくなることがあり、そのことにより、与えられた命令ラインがアクセスされることになる可能性は小さくなる場合がある。予測不可能性のレベルは、予測不可能な分岐命令を用いるプリフェッチの数が増大するにしたがって、増大する場合がある。したがって、ある実施形態においては、命令ライン・プリフェッチの所定の数は、予測された、プリフェッチされた命令ラインにアクセスする可能性が、所与の率を下回らないように選び出される。また、予測不可能な分岐 (例えば、その分岐に関する予測可能性の値が予測可能性の閾値を下回るような分岐) が到達されるようないくつかの場合においては、分岐命令の両方の経路に関する (例えば、予測される分岐経路と予測されない分岐経路の両方に関する) 命令ラインをフェッチすることができる。

30

40

【0059】

別の例として、フェッチされたデータ・ライン内のデータ・アクセスの予測可能性に基づいて、所定の数のデータ・ライン・プリフェッチを行うことができる。ある実施形態においては、前に実行された時にデータ・キャッシュ・ミスをもたらしたデータ・アクセス命令のターゲットとされているデータを格納しているデータ・ラインに関して、データ・ライン・プリフェッチを発行することができる。データ・キャッシュ・ミスを引き起こすデータ・アクセス命令に関して、予測可能性データを格納することもできる。予測可能性

50

データが格納される場合には、プリフェッチされたデータ・ラインに関して起こるデータ・キャッシュ・ミスの相対的な予測可能性に基づいて、所定の数のプリフェッチを行うことができる。

【0060】

命令ラインおよびデータ・ラインのプリフェッチに関して選出された閾値は、サンプル命令のテスト・ランにより選択された固定の数とすることができる。テスト・ランおよび閾値の選択は設計時に行うことができ、閾値は、プロセッサ110に予めプログラムすることができる。随意的に、テスト・ランは、プログラム実行の最初の「トレーニング」フェーズの間に行うことができる（以下で詳細に説明する）。別の実施形態においては、プロセッサ110は、予測不可能な分岐命令および/または予測不可能なデータ・アクセスを格納しているプリフェッチされた命令ラインおよびデータ・ラインの数を追跡して、予測不可能な分岐命令または予測不可能なデータ・アクセス命令を格納している所与の数の命令ラインおよびデータ・ラインがフェッチされた後にのみ命令ラインおよびデータ・ラインのプリフェッチを止め、そのことにより、プリフェッチされた命令ラインの所定の数が命令ラインの実行履歴に基づいて動的に変化するようにすることができる。

【0061】

本発明の実施形態においては、命令ラインにある命令のためのデータ・ターゲット・アドレスは、異なる命令ラインに格納することができる。図7は、本発明の実施形態による、複数の命令ラインに格納された単一の命令ラインにあるデータ・アクセス命令のための複数のデータ・ターゲット・アドレスを図示するブロック図である。図示されたように、命令ライン1は、3つの独立したデータ・ライン（実線の曲線で図示されたデータ・ライン1、データ・ライン2、データ・ライン3）にあるデータ・ターゲット・アドレス D_{2_1} 、 D_{4_2} 、 D_{5_3} にアクセスする3つのデータ・アクセス命令（ I_{4_1} 、 I_{5_1} 、 I_{6_1} ）を格納することができる。本発明の実施形態においては、1以上のデータ・アクセス命令のターゲット・アドレスに対応するアドレスは、フェッチ・シーケンスにおいてソース命令ライン（命令ライン1）と隣接する命令ライン（命令ライン0または命令ライン2）に格納することができる。

【0062】

データ・アクセス命令 I_{4_1} 、 I_{5_1} 、 I_{6_1} が命令ライン1において以下で説明されるように検出されたときに、データ・ライン1に対応するデータ・ターゲット・アドレスを命令ライン0の位置EA2に、データ・ライン2に対応するデータ・ターゲット・アドレスを命令ライン1の位置EA2に、データ・ライン3に対応するデータ・ターゲット・アドレスを命令ライン2の位置EA2に、格納することもできる（破線の曲線で図示される）。データ・アクセス命令 I_{4_1} 、 I_{5_1} 、 I_{6_1} によるターゲット・データのターゲット・アドレス D_{2_1} 、 D_{4_2} 、 D_{5_3} へのアクセスを追跡する目的で、各々の命令ラインにおいて、例えばその命令ラインに付加される位置（LOC）制御ビットに、そのデータ・ターゲット情報のソース（例えば命令ライン1）を示す位置情報を格納することができる。

【0063】

このようにして、データ・ライン1および命令ライン1の有効アドレスを命令ライン0に格納し、データ・ライン2および命令ライン2の有効アドレスを命令ライン1に格納し、データ・ライン3の有効アドレスを命令ライン2に格納することができる。命令ライン0がフェッチされたときに、命令ライン0に格納された有効アドレスを用いて命令ライン1をプリフェッチすることができ、命令ライン1に格納された有効アドレスを用いて命令ライン2をプリフェッチすることができる。データ・ライン1にアクセスするデータ・アクセス命令を命令ライン0が格納していない場合であっても、命令ライン0に格納された有効アドレスを用いてデータ・ライン1をプリフェッチすることができ、そのことにより、命令ライン2にある命令 I_{4_1} がデータ・ライン1にあるデータ D_{2_1} へのアクセスを試みた場合/時のデータ・キャッシュ・ミスをなくすることができる。同様にして、命令ライン1および2がプリフェッチされたときに、データ・ライン2およびデータ・ライン3

10

20

30

40

50

といったデータ・ラインをプリフェッチすることができ、そのことにより、命令ライン 1 にある命令 I 5₁ がデータ位置 D 4₂ にアクセスを試みた場合 / 時のデータ・キャッシュ・ミスがなくすることができ、命令ライン 1 にある命令 I 6₁ がデータ位置 D 5₃ にアクセスを試みた場合 / 時のデータ・キャッシュ・ミスなくすることができる。

【 0 0 6 4 】

ある命令ラインに存在する命令のためのデータ・ターゲット・アドレスを、異なる命令ラインに格納することは、格納されるデータ・ターゲット・アドレスを必ずしも全ての命令ラインが格納していないようないくつかの場合には有用となる。例えば、データ・ターゲット・アドレスが格納される場合において、そのターゲット・アドレスのデータへのアクセスがデータ・キャッシュ・ミスを引き起こすときに、1つの命令ラインがデータ・キャッシュ・ミスを引き起こす数個のデータ・アクセス命令（例えば3つの命令）を格納し、一方、他の命令ラインはデータ・キャッシュ・ミスを引き起こすようないかなるデータ・アクセス命令も格納しないようにすることができる。したがって、1つの命令ラインにおいてデータ・キャッシュ・ミスを引き起こすようなデータ・アクセス命令のための1以上のデータ・ターゲット・アドレスを他の命令ラインに格納し、そのことにより、データ・ターゲット・アドレスのストレージを他の命令ラインに分散することができる（例えば、3つのデータ・ターゲット・アドレスのうち2つを、2つの他の命令ラインに、それぞれ格納することができる）。

【 0 0 6 5 】

（命令ラインのためのデータ・ライン・プリフェッチ・アドレスの格納）

本発明の実施形態により、データ・アクセス命令を実行してデータ・ターゲット・アドレスを格納するデータ・ラインを要求することがデータ・キャッシュ・ミスにつながる時に、そのデータ・アクセス命令のデータ・ターゲット・アドレスを抽出して命令ラインに格納することができる。

【 0 0 6 6 】

図 8 は、本発明の実施形態による、データ・アクセス命令に対応するデータ・ターゲット・アドレスを格納するためのプロセス 8 0 0 を図示する流れ図である。プロセス 8 0 0 は、例えば命令キャッシュ 2 2 2 から命令ラインがフェッチされるステップ 8 0 2 で開始することができる。ステップ 8 0 4 において、フェッチされた命令ラインの中のデータ・アクセス命令を実行することができる。ステップ 8 0 6 において、データ・アクセス命令のターゲットとされたデータを格納しているデータ・ラインがデータ・キャッシュ 2 2 4 に位置するか否かの判定を行うことができる。データ・アクセス命令のターゲットとされたデータを格納しているデータ・ラインがデータ・キャッシュ 2 2 4 に存在しない場合には、ステップ 8 0 8 において、ターゲットとされたデータの有効アドレスがデータ・ターゲット・アドレスとして格納される。ターゲットとされたデータに対応するデータ・ターゲット・アドレスを記録することにより、次に L 2 キャッシュ 1 1 2 から命令ラインがフェッチされたときに、L 2 キャッシュ 1 1 2 から、ターゲットとされたデータを格納しているデータ・ラインをプリフェッチすることができる。データ・ラインをプリフェッチすることにより、データ・アクセス命令が実行された場合 / 時に、他の場合であれば起こり得るデータ・キャッシュ・ミスを、いくつかの場合においては防ぐことができる。

【 0 0 6 7 】

別の選択肢として、データ・アクセス命令のためのデータ・ターゲット・アドレスを、実行時に判定し、そのデータ・アクセス命令がデータ・キャッシュ・ミスを引き起こすか否かに拘わりなく命令ラインに格納することができる。例えば、データ・アクセス命令の各々のためのデータ・ターゲット・アドレスを抽出し、命令ラインに格納することができる。随意的に、最も頻繁に実行されるデータ・アクセス命令（単数または複数）のためのデータ・ターゲット・アドレスを抽出し、命令ラインに格納することができる。データ・ターゲット・アドレスを判定し格納する他の方法は、以下でより詳細に論じられる。

【 0 0 6 8 】

本発明の実施形態においては、データ・ターゲット・アドレスは、そのデータ・ターゲ

10

20

30

40

50

ット・アドレスにアクセスするデータ・アクセス命令が実行されるまで、算出されない場合がある。例えば、データ・アクセス命令は、そのデータ・アクセスが行われるはずのアドレス・レジスタに格納されているアドレスからのオフセット値を指定することができる。データ・アクセス命令が実行されたときに、ターゲット・データの有効アドレスを算出し、データ・ターゲット・アドレスとして格納することができる。全部の有効アドレスを格納することができる場合もある。しかし、他の場合においては、一部の有効アドレスのみを格納することができる。例えば、有効アドレスのより高位の32ビットのみを用いて、データ・アクセス命令のターゲット・データを格納しているキャッシュされたデータ・ラインを見つけることができる場合には、データ・ラインをプリフェッチする目的のために、これら32ビットのみをデータ・ターゲット・アドレスとして保管することができる。

10

【0069】

(データ・ラインのアクセス履歴の追跡および記録)

本発明の別の実施形態においては、データ・アクセス命令を実行することなしに、データ・ターゲット・アドレスを判定することができる。例えば、L2キャッシュ112からデータ・ラインがフェッチされる際に、フェッチされたデータ・ラインにあるデータ・アクセス命令から、データ・ターゲット・アドレスを抽出することができる。

【0070】

本発明の実施形態においては、種々の量のデータ・アクセス履歴情報を格納することができる。データ・アクセス履歴は、命令ラインに存在するどのデータ・アクセス命令が実行されることになるか(または実行される可能性があるか)を示すことができる。随意的に、データ・アクセス履歴は、どのデータ・アクセス命令がデータ・キャッシュ・ミスを引き起こすことになるか(または既に引き起こしたか)を示すことができる。どのデータ・ターゲット・アドレスが命令ラインに格納されるのか(および/またはどのデータ・ラインがプリフェッチされるのか)は、リアルタイム実行の間に、または実行前の「トレーニング」期間の間に生成され格納されたデータ・アクセス履歴情報に基づいて判定することができる。

20

【0071】

上記で説明されたように、ある実施形態により、命令ラインにおいて最後に実行されたデータ・アクセス命令に対応するデータ・ターゲット・アドレスのみを格納することができる。命令ラインにおいて最後に実行されたデータ・アクセス命令に対応するデータ・ターゲット・アドレスのみを格納することは、後でその命令ラインがフェッチされたときに同じデータがアクセスされることを効果的に予測する。よって、前に実行されたデータ・アクセス命令のためのターゲット・データを格納しているデータ・ラインをプリフェッチすることができる。

30

【0072】

いくつかの場合においては、1以上のビットを用いてデータ・アクセス命令の履歴を記録することができる。ビットは、どのデータ・ラインが最も頻繁にアクセスされるか、またはどのデータ・ラインがアクセスされたときにデータ・キャッシュ・ミスを引き起こすかを判定するために用いることができる。例えば、図5に図示されたように、命令ライン(命令ライン1)に格納されている制御ビットCTLは、その命令ラインに存在するどのデータ・アクセス命令が前に実行されたかまたは前にデータ・キャッシュ・ミスを引き起こしたかを示す情報(LOC)を格納することができる。命令ラインはまた、データ・アクセス命令がいつ実行されたかまたはいつキャッシュ・ミスを引き起こしたかの履歴(DAH)(例えば、監視された実行回数のうちの何回その命令が実行されたか、またはある回数前の実行のうちの何回キャッシュ・ミスを引き起こしたか)を格納することができる。

40

【0073】

データ・アクセス命令位置LOCおよびデータ・アクセス履歴DAHをどのようにして用いることができるかの例として、L2キャッシュ112にあって、まだL1キャッシュ

50

222にフェッチされていない命令ラインを考える。命令ラインがL1キャッシュ222にフェッチされたときに、プリデコーダおよびスケジューラ220は、最初に、その命令ラインがいかなるデータ・ターゲット・アドレスも持たないと判定することができ、したがって、別のデータ・ラインをプリフェッチしないであろう。

【0074】

フェッチされた命令ラインにある命令がトレーニングの間に行われる際に、プロセッサ・コア114は、その命令ライン内のデータ・アクセス命令が実行されているか否かを判定することができる。データ・アクセス命令が検出された場合には、データ・ターゲット・アドレスをEA1に格納することに加えて、LOCにそのデータ・アクセス命令の命令ライン内の位置を格納することができる。命令ラインの各々が32の命令を格納する場合には、LOCに(可能性ある命令位置の各々に対応する)0-31の番号を格納して出口分岐命令を示すことができるように、LOCを5ビットのバイナリ数とすることができる。随意的に、(単一の命令ラインの有効アドレスを複数の命令ラインに格納することに関して上述されたように)LOCがソース命令およびソース命令ラインを示す場合には、LOCは、命令ライン内の位置と、そのデータ・アクセス命令がどの命令ラインに隣接して位置しているかの両方を示す、追加のビットを格納することができる。

【0075】

ある実施形態においては、DAHに、LOCに設置されたデータ・アクセス命令が実行されたことまたはデータ・キャッシュ・ミスを引き起こしたことを示す値を書込むことができる。例えば、DAHが単一ビットであった場合には、データ・アクセス命令が実行されたときに、命令ラインにある命令の最初の実行の間に、DAHにその命令に関して0を書込むことができる。DAHに格納された0は、LOCに設置されたそのデータ・アクセス命令が、その後その命令ラインに格納されている命令を実行する間に実行されることの弱い予測を示す。随意的に、DAHに格納された0は、LOCに設置されたそのデータ・アクセス命令が、その後その命令ラインに格納されている命令を実行する間にデータ・キャッシュ・ミスを引き起こすことの弱い予測を示す。

【0076】

命令ラインにある命令のその後の実行の間に、LOCに設置されたデータ・アクセス命令が再び実行された(またはデータ・キャッシュ・ミスを引き起こした)場合には、DAHを1に設定することができる。DAHに格納された1は、LOCに設置されたそのデータ・アクセス命令が再び実行されること、または再びデータ・キャッシュ・ミスを引き起こすことの強い予測を示す。

【0077】

しかし、同一の命令ラインが再びフェッチされ(DAH=1)、かつ、異なる分岐出口命令が取られる場合には、LOCおよびEA1の値は同じままとすることができるが、DAHは、命令ラインに格納されている命令のその後の実行の間に、前に取られた分岐が取られることの弱い予測を示す0にクリアされる。

【0078】

DAHが0であり、かつ、LOCに示されているデータ・アクセス命令以外のデータ・アクセス命令が実行される(または実行されてデータ・キャッシュ・ミスを引き起こす)場合には、データ・ターゲット・アドレスEA1を、そのデータ・アクセス命令のデータ・ターゲット・アドレスで上書きすることができ、LOCを、その命令ラインにある実行されるデータ・アクセス命令(またはデータ・キャッシュ・ミスを引き起こすデータ・アクセス命令)に対応する値に変更することができる。

【0079】

このようにして、データ・アクセス履歴ビットが用いられる場合においては、命令ラインは、データ・ターゲット・アドレスに対応する格納されたデータ・ターゲット・アドレスを格納することができる。このような定期的に行われるデータ・アクセス命令またはデータ・キャッシュ・ミスを引き起こすアクセス命令は、それほど頻繁に行われないデータ・アクセス命令またはそれほど頻繁にデータ・キャッシュ・ミスを引き起こさないデ

10

20

30

40

50

ータ・アクセス命令よりも優先される場合がある。しかし、データ・アクセス命令が弱く予測され、かつ別のデータ・アクセス命令が実行されまたはデータ・キャッシュ・ミスを引き起こした場合には、データ・ターゲット・アドレスを、そのデータ・アクセス命令に対応するアドレスに変更し、そのことにより、他のデータ・アクセス命令が定期的に行われるか、または随意的に、定期的にキャッシュ・ミスを引き起こす時に、弱く予測されたデータ・アクセス命令が優先されないようにすることができる。

【0080】

ある実施形態においては、DAHは、LOCによって示されるデータ・アクセス命令のより長い履歴を格納することができるように、複数の履歴ビットを格納することができる。例えば、DAHが2つのバイナリ・ビットである場合には、00は非常に弱い予測に対応することができる（その場合には他のデータ・アクセス命令を実行すること、または他のデータ・アクセス命令がデータ・キャッシュ・ミスを引き起こすと判定することは、データ・ターゲット・アドレスおよびLOCを上書きすることになる）、一方、01は弱い予測、10は強い予測、11は非常に強い予測に対応することができる（その場合には他のデータ・アクセス命令を実行すること、または他のデータ・キャッシュ・ミスを検出することは、必ずしもデータ・ターゲット・アドレスまたはLOCを上書きするとは限らない）。例として、データ・キャッシュ・ミスの強い予測に対応するデータ・ターゲット・アドレスを置換するためには、プロセッサ構成100は、3つの他のデータ・アクセス命令が命令ラインにある命令の3回の連続実行においてデータ・キャッシュ・ミスを引き起こすことを必要とする場合がある。

【0081】

さらに、ある実施形態においては、データ・ターゲット・アドレスに対応するデータ・ラインは、いくつかの場合においては、（例えば、プロセッサ・コア114がデータ・ラインへのアクセスを試みた時に）データ・キャッシュ・ミスが非常に強く予測されることをDAHビットが示した場合においてのみプリフェッチすることができる。随意的に、データ・ラインをプリフェッチする前提条件として、異なるレベルの予測可能性（例えば、非常に強い予測可能性に対して強い予測可能性）を選択することができる。

【0082】

本発明の実施形態においては、複数のデータ・アクセス履歴（例えばDAH1、DAH2など）、複数のデータ・アクセス命令位置（例えばLOC1、LOC2など）、および/または複数の有効アドレスを用いることができる。例えば、ある実施形態においては、DAH1、DAH2などを用いて複数のデータ・アクセス履歴を追跡することができるが、DAH1、DAH2などのうち最も予測可能性が高いデータ・アクセスおよび/または予測されたデータ・キャッシュ・ミスに対応する1つのデータ・ターゲット・アドレスのみを、EA1に格納することができる。随意的に、複数のデータ・アクセス履歴および複数のデータ・ターゲット・アドレスを単一の命令ラインに格納することができる。ある実施形態においては、LOCによって示される所与のデータ・アクセス命令が予測可能である（例えば、実行される、および/またはデータ・キャッシュ・ミスを引き起こす）ことをデータ・アクセス履歴が示した場合においてのみ、データ・ターゲット・アドレスを用いてデータ・ラインをプリフェッチすることができる。随意的に、いくつかの格納されたアドレスの中から最も予測可能性が高いデータ・ターゲット・アドレスに対応するデータ・ラインのみを、プリデコーダおよびスケジューラ220によってプリフェッチすることができる。

【0083】

上述されたように、ある実施形態においては、データ・アクセス命令がデータ・キャッシュ・ミスを引き起こすかどうかを用いて、データ・ターゲット・アドレスを格納するかどうかを判定することができる。例えば、与えられたデータ・アクセス命令が稀にしかデータ・キャッシュ・ミスを引き起こさない場合には、そのデータ・アクセス命令が命令ラインにある他のデータ・アクセス命令よりも頻繁に行われるような場合であっても、そのデータ・アクセス命令に対応するデータ・ターゲット・アドレスを格納しなくてもよい。

命令ラインにある別のデータ・アクセス命令が、実行される頻度は低い、全体的により多くのデータ・キャッシュ・ミスを引き起こす場合には、他のデータ・アクセス命令に対応するデータ・ターゲット・アドレスを、その命令ラインに格納することができる。上述されたように、1以上のデータ・キャッシュ「ミス」フラグのような履歴ビットを用いて、どのデータ・アクセス命令がデータ・キャッシュ・ミスを引き起こす可能性が最も高いかを判定することができる。

【0084】

命令ラインに格納されたビットを用い、データ・キャッシュ・ミスまたはプリフェッチに起因してデータ・キャッシュ224にデータ・ラインが配置されるか否かを示すことができる場合がある。ビットは、プロセッサ110により、キャッシュ・ミスを防ぐに当たって、プリフェッチの有効性を判定するために用いることができる。また、プリデコーダおよびスケジューラ220（または随意的に、プリフェッチ回路602）も、プリフェッチが不必要であると判定し、したがって命令ラインのビットを変更することができる場合もある。例えば、プリフェッチされた情報が既に命令キャッシュ222またはデータ・キャッシュ224に存在するなどの理由で、プリフェッチが不必要である場合においては、より多くの命令キャッシュ・ミスおよびデータ・キャッシュ・ミスを引き起こすアクセス命令に対応する他のデータ・ターゲット・アドレスを命令ラインに格納することができる。

10

【0085】

ある実施形態においては、データ・アクセス命令がデータ・キャッシュ・ミスを引き起こすかどうかを、データ・アクセス命令のためのデータ・ターゲット・アドレスを格納するか否かを判定するために用いられる唯一の要因とすることができる。別の実施形態においては、データ・アクセス命令を実行する予測可能性と、そのデータ・アクセス命令がデータ・キャッシュ・ミスを引き起こすかどうかの予測可能性の両方をともに用いて、データ・ターゲット・アドレスを格納するか否かを判定することができる。例えば、アクセス履歴およびミス履歴に対応する値を加算するか、乗算するか、または他の計算式の中で（例えば加重として）用いて、データ・ターゲット・アドレスを格納するか否か、および/または、そのデータ・ターゲット・アドレスに対応するデータ・ラインをプリフェッチするか否かを判定することができる。

20

【0086】

本発明の実施形態においては、データ・ターゲット・アドレス、データ・アクセス履歴、およびデータ・アクセス命令位置を実行時に継続的に追跡し、更新して、命令ラインに格納されたデータ・ターゲット・アドレスおよび他の値が、与えられた命令の組が実行されるに従い時間が経つにつれて変化するようにすることができる。このようにして、例えばプログラムが実行されるに従い、データ・ターゲット・アドレスおよびプリフェッチされたデータ・ラインを動的に修正することができる。

30

【0087】

本発明の別の実施形態においては、命令の組の最初の実行フェーズの間に（例えばプログラムが実行される最初の「トレーニング」期間の間に）、データ・ターゲット・アドレスを選択し、格納することができる。最初の実行フェーズは、初期化フェーズ、またはトレーニング・フェーズと呼ばれることもある。トレーニング・フェーズの間に、データ・アクセス履歴およびデータ・ターゲット・アドレスを追跡することができる。1以上のデータ・ターゲット・アドレスを（例えば、上述された基準によって）命令ラインに格納することができる。フェーズが完了したときに、格納されたデータ・ターゲット・アドレスは、引き続き、L2キャッシュ112からデータ・ラインをプリフェッチするために用いることができるが、しかし、フェッチされた命令ラインにあるデータ・ターゲット・アドレス（単数または複数）は、それ以上追跡されたり更新されたりしないであろう。

40

【0088】

ある実施形態においては、データ・ターゲット・アドレス（単数または複数）を格納している命令ラインの1以上のビットを用いて、そのデータ・ターゲット・アドレスが最初

50

の実行フェーズの間に更新されたか否かを示すことができる。例えば、トレーニング・フェーズの間に、1つのビットをクリアすることができる。ビットがクリアされている間に、命令ラインにある命令が実行されるに従い、データ・アクセス履歴を追跡することができる。データ・ターゲット・アドレス（単数または複数）を更新することができる。トレーニング・フェーズが完了したときに、ビットを設定することができる。ビットが設定されたときに、データ・ターゲット・アドレス（単数または複数）はそれ以上更新されない場合があり、最初の実行フェーズが完了となるであろう。

【0089】

ある実施形態においては、最初の実行フェーズは、特定の期間にわたって（例えばある数のクロック周期が経過するまで）続行することができる。ある実施形態においては、格納された最新のデータ・ターゲット・アドレスは、指定の期間が経過し、最初の実行フェーズが終了するまで、命令ラインに格納されたままでいることができる。別の実施形態においては、最も頻繁に実行されたデータ・アクセス命令に対応するデータ・ターゲット・アドレスか、または最も頻繁にデータ・キャッシュ・ミスを引き起こしたデータ・アクセス命令に対応するデータ・ターゲット・アドレスを命令ラインに格納して、その後のプリフェッチに用いることができる。

10

【0090】

本発明の別の実施形態においては、最初の実行フェーズは、1以上の出口基準が満たされるまで続行することができる。例えば、データ・アクセス履歴が格納されるような場合においては、最初の実行フェーズは、命令ラインにあるデータ・アクセス命令の1つが予測可能になる（または強く予測可能になる）までか、またはデータ・キャッシュ・ミスが予測可能になる（または強く予測可能になる）まで、続行することができる。与えられたデータ・アクセス命令が予測可能になったときに、最初のトレーニング・フェーズが完了したと、その命令ラインがL2キャッシュ112からフェッチされる時に行われるその後のデータ・ラインのプリフェッチの各々において、強く予測可能なデータ・アクセス命令のデータ・ターゲット・アドレスを用いることができることを示すロック・ビットを、命令ラインに設定することができる。

20

【0091】

本発明の別の実施形態においては、断続的なトレーニング・フェーズにおいて、命令ラインにあるデータ・ターゲット・アドレスを修正することができる。例えば、トレーニング・フェーズの各々について、頻度および期間の値を格納することができる。頻度に対応しているある数のクロック周期が経過する度ごとに、トレーニング・フェーズは初期化され、指定の期間値にわたって続行することができる。別の実施形態においては、頻度に対応しているある数のクロック周期が経過する度ごとに、トレーニング・フェーズは初期化され、指定の条件が満たされるまで（例えば、上述されたように、命令に関して、指定のレベルのデータ・アクセス予測可能性かまたは指定のレベルのキャッシュ・ミス予測可能性が到達されるまで）、続行することができる。

30

【0092】

本発明の実施形態においては、システム100において用いられる各々のレベルのキャッシュおよび/またはメモリは、命令ラインに格納されている情報のコピーを格納することができる。本発明の別の実施形態においては、指定のレベルのキャッシュおよび/またはメモリのみが、命令ラインに格納されている情報（例えば、データ・アクセス履歴およびデータ・ターゲット・アドレス）を格納することができる。ある実施形態においては、当業者には公知のキャッシュの一貫性原則を用いて、各々のレベルのキャッシュおよび/またはメモリにある命令ラインのコピーを更新することができる。

40

【0093】

命令キャッシュを用いる従来型のシステムにおいては、一般的には、命令はプロセッサ110によって修正されないことに注意されたい。よって、従来型のシステムにおいては、命令ラインは、ある程度の時間が経過した後、L2キャッシュ112に書き戻されるのではなく、命令キャッシュ222からエージアウトされることが一般的にである。しかし

50

、ここで説明されるように、本発明の実施形態においては、修正された命令ラインをL2 キャッシュ112に書戻すことができ、それにより、上位レベルのキャッシュおよび/またはメモリにおいてプリフェッチ・データが維持されることが可能になる。

【0094】

例として、命令ラインにある命令がプロセッサ・コアにより既に処理された(データ・ターゲット・アドレスおよび他の履歴情報が更新されることを引き起こす可能性がある)ときに、その命令ラインを、恐らくは命令キャッシュ222に格納されたより古いバージョンの命令ラインを上書きする形で、命令キャッシュ222に書込むことができる(書戻しと呼ばれる)。ある実施形態においては、命令ラインは、その命令ラインに格納された情報に対して変更が為された場合においてのみ、命令キャッシュ222に配置される。

10

【0095】

本発明の実施形態により、修正された命令ラインがL2 キャッシュ112に書戻されたときに、命令ラインを変更されたものとしてマークすることができる。命令ラインが命令キャッシュ222に書戻され、変更されたものとしてマークされる場合においては、命令ラインは異なった時間にわたって命令キャッシュに留まることができる。例えば、命令ラインがプロセッサ・コア114によって頻繁に使用される場合には、命令ラインは命令キャッシュ222へフェッチされ、戻されることが何度かあり、恐らくはその度に更新されるであろう。しかし、命令ラインが頻繁に使用されない場合には(エージングと呼ばれる)、その命令ラインを命令キャッシュ222からパージすることができる。命令ラインが命令キャッシュ222からパージされたときに、その命令ラインをL2 キャッシュ112に書戻すことができる。ある実施形態においては、命令ラインが修正されたものとしてマークされた場合にのみ、L2 キャッシュ112に書戻すことができる。別の実施形態においては、命令ラインは常にL2 キャッシュ112に書戻すことができる。ある実施形態においては、命令ラインは、随意的に、いくつかのキャッシュ・レベルに(例えばL2 キャッシュ112および命令キャッシュ222に)一度に書戻すか、または命令キャッシュ222以外のレベルに(例えばL2 キャッシュ112に直接)書戻すことができる。

20

【0096】

ある実施形態においては、命令ライン以外の位置にデータ・ターゲット・アドレス(単数または複数)を格納することができる。例えば、シャドウ・キャッシュにデータ・ターゲット・アドレスを格納することができる。図9は、本発明の実施形態による、命令およびデータ・ラインをプリフェッチするためのシャドウ・キャッシュ902を図示するブロック図である。

30

【0097】

本発明の実施形態においては、(例えば、データ・アクセス命令が頻繁に実行されるかまたは頻繁にデータ・キャッシュ・ミスを引き起こすことに起因して、および/または上記で列挙された基準のいずれかにより)命令ラインにあるデータ・アクセス命令のためのデータ・ターゲット・アドレスが格納されるべきときに、命令ラインに対応するアドレスまたはアドレスの一部(例えば、命令ラインの有効アドレス、または有効アドレスの高位の32ビット)、および、データ・ターゲット・アドレス(またはその一部)を、シャドウ・キャッシュ902の項目として格納することができる。いくつかの場合においてはシャドウ・キャッシュ902には、単一の命令ラインに関する複数のデータ・ターゲット・アドレス項目を格納することができる。随意的に、命令ラインに関する項目の各々は、複数のデータ・ターゲット・アドレスを格納することができる。

40

【0098】

L2 キャッシュ112から情報がフェッチされるときに、シャドウ・キャッシュ902(または、例えばプリデコード制御回路610のような、シャドウ・キャッシュ902を用いる他の制御回路)は、フェッチされた情報が命令ラインなのか否かを判定することができる。L2 キャッシュ112による出力が命令ラインであるとの判定が為された場合には、シャドウ・キャッシュ902を検索して(例えば、シャドウ・キャッシュ902は内容アドレス可能とすることができる)、フェッチされた命令ラインに対応する項目(また

50

は複数の項目) (例えば、フェッチされた命令ラインと同一の有効アドレスを有する項目) を探すことができる。対応する項目が見つかった場合には、その項目に関連するデータ・ターゲット・アドレス(単数または複数)は、プリデコード制御回路610、プリデコードおよびスケジューラ220にある他の回路、および、プリフェッチ回路602により、シャドウ・キャッシュ902によって示されたデータ・ターゲット・アドレス(単数または複数)をプリフェッチするために用いられる。随意的に、シャドウ・キャッシュ902には、(専門に、またはデータ・ターゲット・アドレスとともに)分岐出口アドレスを格納することができる。上述されたように、シャドウ・キャッシュ902は、そこに格納された有効アドレス、および/または、フェッチされた命令ラインおよびプリフェッチされた命令ラインに格納された有効アドレスを使って、命令ラインおよびデータ・ラインのチェーン/グループをフェッチするために用いることができる。 10

【0099】

本発明の実施形態においては、シャドウ・キャッシュ902はまた、上述された制御ビット(例えば、履歴ビットおよび位置ビット)を格納することができる。随意的に、こうした制御ビットは、上述されたように、命令ラインに格納することができる。どちらの場合においても、ある実施形態においては、シャドウ・キャッシュ902にある項目は、どの項目が命令ラインに格納されるのかを判定することに関して上記で列挙されたいずれかの基準にしたがって管理することができる。各々がシャドウ・キャッシュ902を伴って実装することができる、上述された多くの技術のうちの一つの例として、強く予測されたデータ・キャッシュ・ミスを引き起こすデータ・アクセス命令のためのデータ・ターゲット・アドレスはシャドウ・キャッシュ902に格納し、一方、弱く予測されたデータ・キャッシュ・ミスに対応するデータ・ターゲット・アドレスは上書きすることができる。 20

【0100】

どの項目がシャドウ・キャッシュ902に格納されるのかを判定する上述された技術を用いることに加えて、ある実施形態においては、専門に、または上述された技術を含めて従来型のキャッシュ管理技術を用いて、シャドウ・キャッシュ902を管理することができる。例えば、シャドウ・キャッシュ902にある項目は、そのシャドウ・キャッシュ902にある項目がアクセスされた頻度を示すエージ・ビットを有することができる。与えられた項目が頻繁にアクセスされる場合には、エージ値は小さい(例えば、若い)まま留まることができる。しかし、項目が頻繁にアクセスされない場合には、エージ値は増大することができ、いくつかの場合においては、シャドウ・キャッシュ902から項目を廃棄することができる。 30

【0101】

上述したように、第1の命令ラインに格納されているデータ・アクセス命令のターゲットとされたデータのアドレスを格納し、ターゲットとされたデータを格納しているデータ・ラインを、L2キャッシュからプリフェッチするために用いることができる。その結果として、データ・キャッシュ・ミス、および対応する、データにアクセスする待ち時間の数を減少させることができ、プロセッサ性能を向上させることができる。

【0102】

以上、本発明の実施形態を一例として説明したが、本発明の他のさらなる実施形態をその基本的な範囲を逸脱することなく考案することができる。本発明の範囲は、添付の特許請求の範囲によって定められる。 40

【図面の簡単な説明】

【0103】

【図1】本発明の実施形態によるシステムを図示するブロック図である。

【図2】本発明の実施形態によるコンピュータ・プロセッサを図示するブロック図である。

【図3】本発明の実施形態による、データ・ラインにアクセスする命令ラインを図示する図である。

【図4】本発明の実施形態による、データ・キャッシュ・ミスを防ぐためのプロセスを図 50

示する流れ図である。

【図5】本発明の実施形態によるデータ・アクセス・アドレスを含んだ命令ラインを図示するブロック図である。

【図6】本発明の実施形態による命令およびデータ・ラインをプリフェッチするための回路を図示するブロック図である。

【図7】本発明の実施形態による、複数の命令ラインに格納された単一の命令ラインにあるデータ・アクセス命令のための複数のデータ・ターゲット・アドレスを図示するブロック図である。

【図8】本発明の実施形態による、データ・アクセス命令に対応するデータ・ターゲット・アドレスを格納するためのプロセスを図示する流れ図である。

【図9】本発明の実施形態による、命令およびデータ・ラインをプリフェッチするためのシャドウ・キャッシュを図示するブロック図である。

【符号の説明】

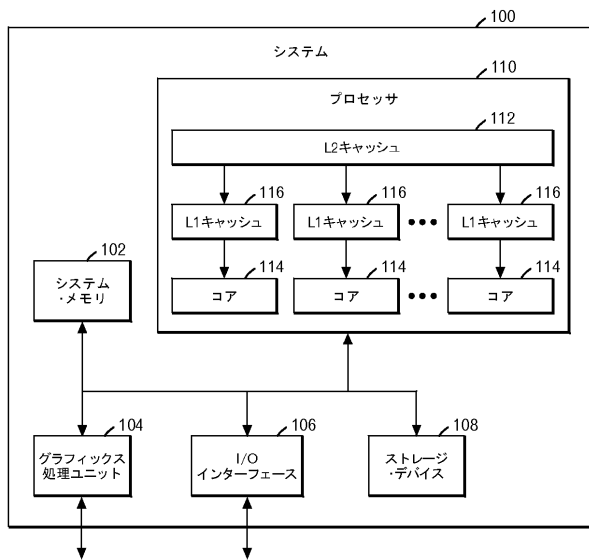
【0104】

- 110：プロセッサ
- 112：L2キャッシュ
- 220：プリデコーダおよびスケジューラ
- 222：L1命令キャッシュ
- 224：L1データキャッシュ
- 226：命令ラインバッファ
- 114：プロセッサ・コア

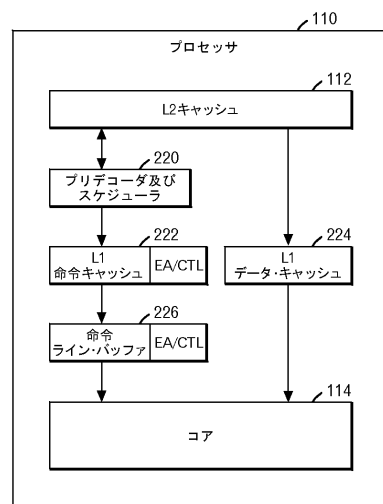
10

20

【図1】



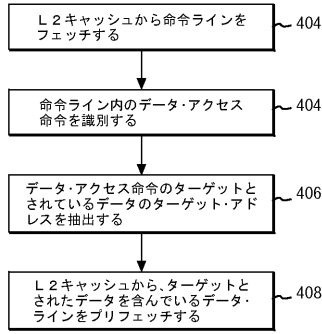
【図2】



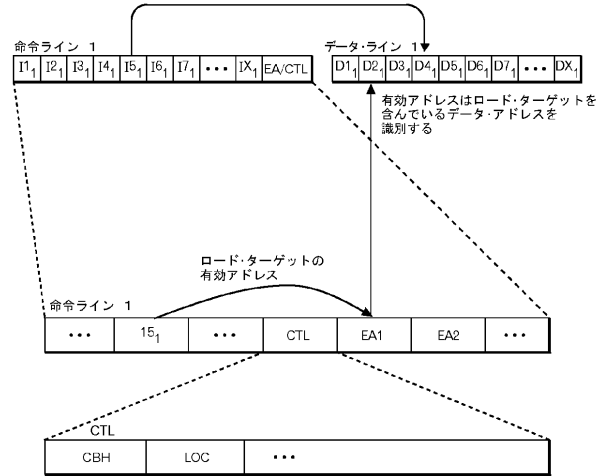
【図3】



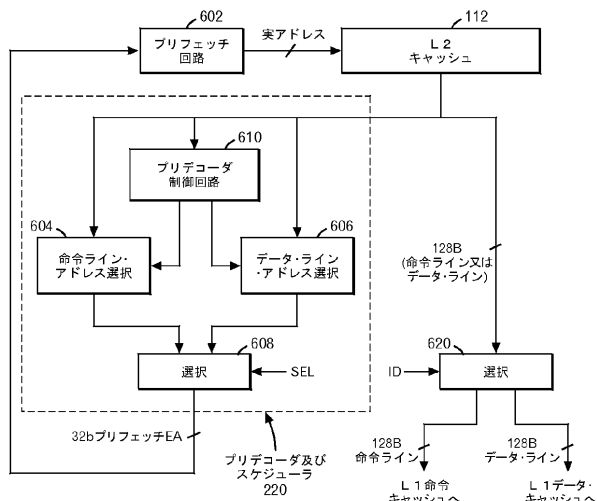
【 図 4 】



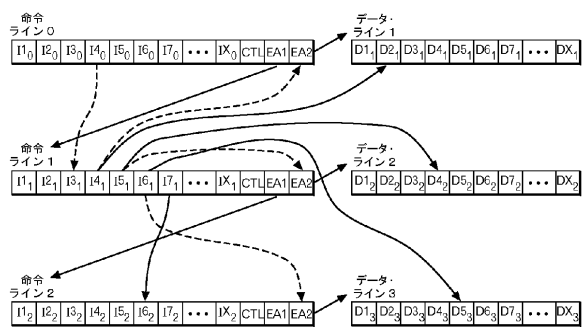
【 図 5 】



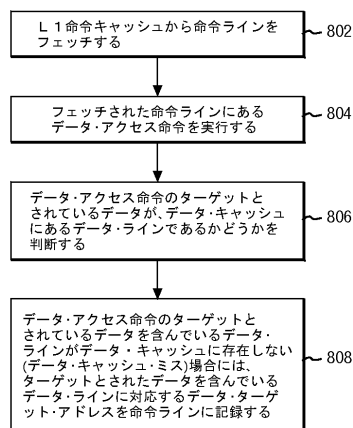
【 図 6 】



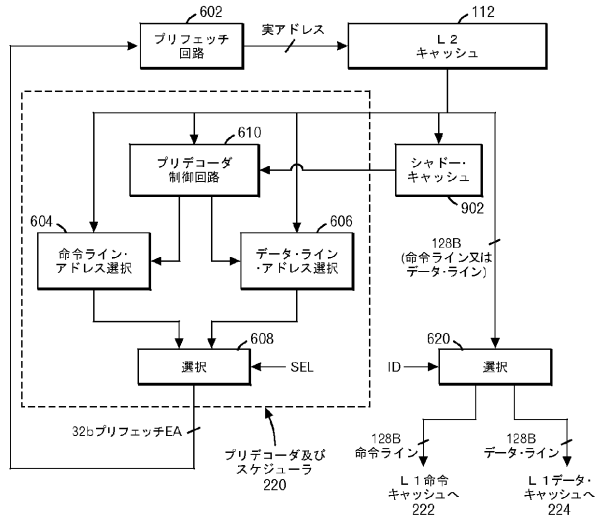
【 図 7 】



【 図 8 】



【 図 9 】



フロントページの続き

(74)代理人 100086243

弁理士 坂口 博

(72)発明者 デービッド・アーノルド・ルーイク

アメリカ合衆国 5 5 9 0 4 ミネソタ州 ロチェスター シンプソン・ロード サウス・イース
ト 3 5 6 3

Fターム(参考) 5B005 JJ13 KK23 LL01 LL11 MM05 NN22 UU32