

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5859185号
(P5859185)

(45) 発行日 平成28年2月10日(2016.2.10)

(24) 登録日 平成27年12月25日(2015.12.25)

(51) Int.Cl.		F I			
G06F	12/08	(2016.01)	G06F	12/08	505B
G06F	9/38	(2006.01)	G06F	12/08	511B
			G06F	9/38	310A

請求項の数 20 (全 14 頁)

(21) 出願番号	特願2009-540408 (P2009-540408)	(73) 特許権者	595020643
(86) (22) 出願日	平成19年12月3日(2007.12.3)		クアルコム・インコーポレイテッド
(65) 公表番号	特表2010-512580 (P2010-512580A)		QUALCOMM INCORPORATED
(43) 公表日	平成22年4月22日(2010.4.22)		アメリカ合衆国、カリフォルニア州 92
(86) 国際出願番号	PCT/US2007/086254		121-1714、サン・ディエゴ、モア
(87) 国際公開番号	W02008/073741		ハウス・ドライブ 5775
(87) 国際公開日	平成20年6月19日(2008.6.19)	(74) 代理人	100108855
審査請求日	平成21年7月8日(2009.7.8)		弁理士 蔵田 昌俊
(31) 優先権主張番号	11/608,309	(74) 代理人	100109830
(32) 優先日	平成18年12月8日(2006.12.8)		弁理士 福原 淑弘
(33) 優先権主張国	米国 (US)	(74) 代理人	100158805
前置審査			弁理士 井関 守三
		(74) 代理人	100194814
			弁理士 奥村 元宏

最終頁に続く

(54) 【発明の名称】 低複雑性命令プリフェッチシステム

(57) 【特許請求の範囲】

【請求項 1】

プリフェッチのための方法であって、

第1のキャッシュ内の第1の命令の第1のフェッチアドレスのミスを示すインジケーションを受け取れることを備え、前記インジケーションは、第1の動作の後に受け取られ、

前記方法はさらに、

前記インジケーションに応答して、

前記第1のフェッチアドレスに基づいて、第2のキャッシュからの第1の命令のラインをフェッチし、前記第1の動作が非連続的であることを条件として、前記第2のキャッシュからの第2の命令のラインをプリフェッチすることと、ここで、前記第2の命令のラインは、前記第1の命令のラインの次の連続キャッシュライン、または、分岐履歴テーブルもしくは分岐予測ロジックからの情報にしたがって非連続的であるキャッシュラインである、

前記第1の動作が連続的であることに応答して、前記第2のキャッシュからの前記第2の命令のラインをプリフェッチすることなく、前記第1のフェッチアドレスに基づいて、前記第2のキャッシュからの前記第1の命令のラインをフェッチすることとを備える、方法。

【請求項 2】

前記第1の動作が非連続的であることを条件として前記プリフェッチすることは、前記第1の命令のラインをフェッチすることと、前記第2の命令のラインをプリフェッチする

10

20

ことを結合することによって長データ転送を可能にすることを備える請求項 1 に記載の方法。

【請求項 3】

前記第 1 の動作は、前記第 1 のキャッシュに結合されたプロセッサ内のロジックによって非連続的であると識別される非連続的な動作を備える請求項 1 に記載の方法。

【請求項 4】

前記第 1 の動作は、前記第 1 のフェッチアドレスを前の命令の前のフェッチアドレスと比較することによって、非連続的であると識別される非連続的な動作を備える請求項 1 に記載の方法。

【請求項 5】

前記第 1 の動作は、分岐動作を含む非連続的な動作を備える請求項 1 に記載の方法。

【請求項 6】

前記第 2 の命令のラインは、前記第 1 の命令のライン内の既定の閾値位置に係する前記第 1 の命令のライン内の前記第 1 の命令の位置をさらに条件としてプリフェッチされる請求項 1 に記載の方法。

【請求項 7】

前記第 2 の命令のラインは、前記第 1 の命令のラインに非連続的である請求項 1 に記載の方法。

【請求項 8】

前記第 1 の動作は、サブルーチンコール動作を含む非連続的な動作を備える請求項 1 に記載の方法。

【請求項 9】

前記第 1 の動作は、復帰動作を含む非連続的な動作を備える請求項 1 に記載の方法。

【請求項 10】

システムであって、

第 1 の命令の第 1 のフェッチアドレスが命令キャッシュ内でミスしたという通知を生成するように構成された命令キャッシュを備え、インジケーションは、第 1 の動作の後に受け取られ、

前記システムはさらに、

前記第 1 のフェッチアドレスに基づいて、第 2 のキャッシュからの第 1 の命令のラインをフェッチするように構成されたフェッチロジックと、

前記通知に応答して、

前記第 1 の動作が非連続的であることを条件として、前記第 2 のキャッシュからの第 2 の命令のラインをプリフェッチし、ここで、前記第 2 の命令のラインは、前記第 1 の命令のラインの次の連続キャッシュライン、または、分岐履歴テーブルもしくは分岐予測ロジックからの情報にしたがって非連続的であるキャッシュラインである、

前記第 1 の動作が連続的であることに応答して、前記第 2 のキャッシュからの前記第 2 の命令のラインをプリフェッチすることを控えるように構成される、

プリフェッチ判定ロジックと

を備えるシステム。

【請求項 11】

特有の第 1 の動作が非連続的であるかどうかを判定するように構成されたキャッシュ制御ユニットをさらに備える請求項 10 に記載のシステム。

【請求項 12】

前記プリフェッチされた第 2 の命令のラインは、前記命令キャッシュへの長データ転送において提供される請求項 10 に記載のシステム。

【請求項 13】

前記第 2 の命令のラインは、前記第 1 の命令のラインに連続的である請求項 10 に記載のシステム。

【請求項 14】

前記第 2 の命令のラインは、前記第 1 の命令のラインに非連続的である請求項 10 に記載のシステム。

【請求項 15】

第 1 の命令の第 1 のフェッチアドレスが第 1 のキャッシュ内でミスしたという通知を受け取ることであって、インジケーションは、第 1 の動作の後に受け取られ、

前記通知に応答して、

前記第 1 のフェッチアドレスに基づいて、第 2 のキャッシュからの第 1 の命令のラインをフェッチし、前記第 1 の動作が非連続的であることと、フェッチアドレスキャッシュライン位置が閾値キャッシュライン位置を越えていることとを条件として、前記第 2 のキャッシュからの第 2 の命令のラインをプリフェッチすることとであって、ここで、前記第 2 の命令のラインは、前記第 1 の命令のラインの次の連続キャッシュライン、または、分岐履歴テーブルもしくは分岐予測ロジックからの情報にしたがって非連続的であるキャッシュラインである、前記フェッチアドレスキャッシュライン位置は、前記第 1 のフェッチアドレスに関連付けられ、前記閾値キャッシュライン位置は、第 1 の命令のラインに関連付けられていることと、

前記第 1 の動作が非連続的であることと、前記フェッチアドレスキャッシュライン位置が前記閾値キャッシュライン位置を越えないこととに併答して、前記第 2 の命令のラインをプリフェッチすることなく、前記第 1 のフェッチアドレスに基づいて、前記第 2 のキャッシュからの前記第 1 の命令のラインをプリフェッチすることと

を備える方法。

【請求項 16】

前記第 2 の命令のラインは、前記第 1 の命令のラインに非連続的である請求項 15 に記載の方法。

【請求項 17】

前記第 1 のフェッチアドレスは、特有の第 1 の動作が連続的である場合、前のフェッチアドレスに対して連続的であるメモリアクセスのターゲットであり、前記第 1 のフェッチアドレスは、前記特有の第 1 の動作が非連続的である場合、前記前のフェッチアドレスに対して非連続的であるメモリアクセスの前記ターゲットである、請求項 1 に記載の方法。

【請求項 18】

前記第 1 のフェッチアドレスが、前の命令の第 2 のフェッチアドレスに関連付けられて、非連続的であるか、または、連続的であるかどうかを決定することをさらに備え、

前記第 1 のフェッチアドレスが非連続的であると決定することを条件として、前記第 2 の命令のラインをプリフェッチする、請求項 1 に記載の方法。

【請求項 19】

前記第 1 のフェッチアドレスが、前の命令の第 2 のフェッチアドレスに関連付けられて、非連続的であるか、または、連続的であるかどうかを決定するように構成されているキャッシュ制御ユニットをさらに備え、

前記プリフェッチ判定ロジックは、キャッシュ制御ユニットによって前記第 1 のフェッチアドレスが非連続的であるという決定に基づいて、前記第 2 の命令のラインをプリフェッチするように構成されている、請求項 10 に記載のシステム。

【請求項 20】

前記第 1 のフェッチアドレスが、前の命令の第 2 のフェッチアドレスに関連して非連続的であると決定することに基づいて、前記第 2 の命令のラインをプリフェッチし、前記第 1 のフェッチアドレスが、前記前の命令の前記第 2 のフェッチアドレスに関連して連続的であると決定することに基づいて、前記第 2 の命令のラインをプリフェッチしない、請求項 15 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般に命令キャッシュの分野に関し、特に、命令キャッシュ内のミスにおけ

10

20

30

40

50

る命令プリフェッチに関する。

【背景技術】

【0002】

例えば携帯電話、ラップトップコンピュータ、携帯情報端末（PDA）等のような多くのポータブル製品は、例えば通信プログラム及びマルチメディアプログラムのようなプログラムを実行するプロセッサを用いる。このような製品の処理システムは、命令及びデータを格納するメモリ及びプロセッサコンプレックスを含む。大容量メインメモリは通常、プロセッササイクル時間と比較して緩慢なアクセス時間を有する。その結果、プロセッサと密接に位置づけられた高性能低容量のキャッシュを用いると、従来、キャッシュメモリの性能及び容量に基づいて、階層内にメモリコンプレックスが構成された。例えば、レベル1命令キャッシュ及びレベル1データキャッシュは一般に、プロセッサに直接接続されるであろう。一方、レベル2統一キャッシュは、レベル1（L1）命令キャッシュ及びL1データキャッシュに接続される。更に、システムメモリは、レベル2（L2）統一キャッシュに接続される。レベル1命令キャッシュは通常、プロセッサ速度で動作し、レベル2統一キャッシュ動作は、レベル1キャッシュよりも緩慢に動作するが、システムメモリより高速なアクセス時間を有する。代替メモリ構成は、例えば、L1及びL2キャッシュに加えてレベル3キャッシュを有するメモリ階層を有する。別のメモリ構成は、レベル1キャッシュ及びシステムメモリのみを用いる。

10

【0003】

なぜ命令キャッシュにメモリ階層を用いることができるかの根拠となる原理の1つは、命令は、メモリ内の連続位置からアクセスされる傾向があるということである。コードの最後に用いられたセクションを保持するキャッシュを有することによって、プロセッサは、高い性能レベルで実行することができる。プログラムはまた、分岐、コール、及び復帰型命令を含み、例えば割込みのような非連続動作をサポートするので、連続局所性の原理は、コードの比較的短いセクションのみでしか維持されない。このような非連続動作によって、命令キャッシュへの命令フェッチはミスを起こし、命令フェッチは、高いメモリ容量と緩慢なアクセス時間で動作する、次の上位メモリレベルに適用される。ミスは、プロセッサに、命令を待機する機能を停止させる恐れがある。プロセッサ性能を高く保つために、ミスレートは低くなければならない。

20

【0004】

命令キャッシュは一般に、命令キャッシュ内の単一アドレスに位置付けられた複数の命令によって構成される。この複数の命令は一般に、キャッシュライン又は単にラインと呼ばれる。ミスは、キャッシュライン内の任意の場所への命令アクセスにおいて発生しうる。ミスが発生すると、必要な命令をフェッチするだけでなく、ミスが起こった命令からキャッシュラインの末尾までのキャッシュラインの残りもフェッチされることがある。いくつかのシステムにおいて、このプリフェッチ技術は更に、ミスが起こると、常にキャッシュラインの残り及び次のキャッシュラインをプリフェッチすることに及ぶ。次のキャッシュラインを常にプリフェッチする従来の技術は、次のキャッシュラインが、まもなく必要になる命令を含むという仮定に基づいている。次のキャッシュライン内の命令を用いるこの仮定は、例えば、ライン内に条件付分岐があり、この条件によって、分岐が次の連続命令まで失敗する場合にも有効である。常に次のキャッシュラインをプリフェッチすることにより、ミスを低減することができる。

30

40

【0005】

命令の連続アクセスの局所性原理は、やはりプログラム内のいくつかのポイントで失敗し、分岐、コール、及び復帰等による非連続動作によってミスが発生する。連続アクセスによるミスは、例えば、命令がキャッシュラインの末尾でフェッチされ、次の連続命令キャッシュライン内に存在すべき次の連続命令がキャッシュ内に存在しない場合、発生しうる。非連続アクセスによるミスは、例えば、分岐命令があり、分岐がプログラムアドレスを新たな位置に変更し、新たな位置の命令がキャッシュ内に存在しない場合、発生しうる。次のキャッシュラインを常にプリフェッチする従来の技術は、用いない命令をフェッチ

50

し、それによって、メモリアクセス帯域幅の不必要な損失、電力使用の増加、及び低プロセッサ性能をもたらす。

【発明の概要】

【0006】

本開示は、ミスが発生した場合、ミスの割合と、メモリアクセス帯域幅と、電力消費とを低減するプリフェッチ技術へのニーズが存在することを認識する。本発明の1つの実施形態は、命令プリフェッチのための方法を提供する。ミスが発生すると、フェッチアドレスが命令キャッシュ内でミスしたという通知が受け取られる。ミスの原因になったフェッチアドレスの属性が判定され、その属性に基づいて、命令のラインがプリフェッチされる。

10

【0007】

本発明の別の実施形態は、命令キャッシュ及びプリフェッチシステムを扱う。命令キャッシュは、フェッチアドレスがミスしたという通知を生成するように構成される。プリフェッチ判定ロジックは、ミスの原因になったフェッチアドレスの属性を判定するように構成され、プリフェッチロジックは、属性に基づいて命令のラインをプリフェッチするように構成される。

【0008】

本発明の別の実施形態は、メモリシステムを扱う。メモリシステムは、フェッチアドレスがミスしたという通知を生成するように構成された命令キャッシュを用いる。プリフェッチ判定ロジックは、ミスの原因になったフェッチアドレスの属性を判定し、ミスしたフェッチアドレスと属性とに基づいてミス情報を生成するように構成される。システムメモリ階層は、ミスしたフェッチアドレスに関連する命令によって、ミス情報に応答するように構成される。

20

【0009】

本発明の様々な実施形態が例示によって示され説明される以下の発明を実施する形態から、当業者には、本発明の他の実施形態が容易に明らかになるであろうことが理解される。理解されるように、本発明は、他の実施形態及び異なる実施形態も可能であり、いくつかの詳細は、全て本発明から逸脱することなく様々な他の観点における変更が可能である。従って、図面及び発明を実施する形態は、本質的に例示的であり、限定的ではないと見なされる。

30

【図面の簡単な説明】

【0010】

【図1】図1は、無線通信システムを示す。

【図2】図2は、命令キャッシュラインのプリフェッチが、メモリ効率の良い動作のためのヒューリスティックに基づいて適用されうるプロセッサ及びメモリコンプレックスの機能ブロック図である。

【図3】図3は、低複雑性プリフェッチ処理を示すフローチャートである。

【発明を実施する形態】

【0011】

添付図面と関連して以下で説明される発明を実施する形態は、本発明の様々な典型的な実施形態の説明として意図されており、本発明が実施されうる実施形態のみを表すことは意図されていない。発明を実施する形態は、本発明の完全な理解を提供する目的のために、具体的な詳細を含む。しかし、本発明はこれら具体的な詳細なしでも実現されうる。当業者には明らかであるだろう。いくつかの例において、周知の構成及びデバイスは、本発明の概念を不明確にすることを避けるためにブロック図形式で示される。

40

【0012】

図1は、本開示の実施形態が有利に用いられうる典型的な無線通信システム100を示す。例示目的のために、図1は、3つの遠隔ユニット120、130、及び150と、2つの基地局140とを示す。典型的な無線通信システムは、更に多くの遠隔ユニット及び基地局を有することができる。遠隔ユニット120、130、

50

及び 150 は、以下で更に説明されるような、本発明を具現化するために適した構成要素 125A、125C、及び 125B によって表されるハードウェア部品、ソフトウェア構成要素、又はその両方をそれぞれ含む。図 1 は、基地局 140 から遠隔ユニット 120、130、及び 150 への順方向リンク信号 180 と、遠隔ユニット 120、130、及び 150 から基地局 140 への逆方向リンク信号 190 とを示す。

【0013】

図 1 では、遠隔ユニット 120 は、無線ローカル・ループ・システムにおけるモバイル電話として示され、遠隔ユニット 130 は、無線ローカル・ループ・システムにおけるポータブル・コンピュータとして示され、遠隔ユニット 150 は、無線ローカル・ループ・システムにおける位置固定式遠隔ユニットとして示される。例えば遠隔ユニットは、代わりに、携帯電話、ページャ、ウォークー・トーカー、ハンドヘルドパーソナル通信システム (PCS) ユニット、例えば携帯情報端末のようなポータブル・データ・ユニット、あるいは例えば計測装置のような位置固定式データユニットであることができる。図 1 は本発明の教示に従う典型的な遠隔ユニットを示すが、本発明は、示されたこれら典型的なユニットに限定されない。本発明は、命令キャッシュ及びシステムメモリを備えるプロセッサを有する任意のデバイスによって適切に用いることができる。

【0014】

図 2 は、メモリ効率の良い動作のためのヒューリスティックスに基づいて、命令キャッシュラインのプリフェッチが適用されうるプロセッサ及びメモリコンプレックス 200 の機能ブロック図である。プロセッサ及びメモリコンプレックス 200 は、プロセッサ 202 と、直接マップされたレベル 1 (L1) 命令キャッシュ (Iキャッシュ) 203 と、システムメモリ階層 208 とを含む。L1 命令 I キャッシュ 203 は、L1 I キャッシュラインアレイ 204 と、L1 I キャッシュ制御ユニット 206 とを備える。L1 I キャッシュ制御ユニット 206 は、例えばセット連想キャッシュのような様々なタイプのキャッシュで用いられうるような、命令タグマッチングのための命令コンテンツアドレス指定可能メモリと、プリフェッチロジックとを含むことができる。プロセッサコンプレックスに接続することができる周辺デバイスは、説明の明確性のために図示しない。プロセッサ及びメモリコンプレックス 200 は、L1 I キャッシュ 203 及びシステムメモリ階層 208 に格納されたプログラムコードを実行するために、構成要素 125A 乃至 125C によって適切に用いられうる。

【0015】

L1 I キャッシュラインアレイ 204 は、例えばキャッシュライン 210 及び 211 のような複数のラインを含むことができる。各ラインはそれぞれ、複数の命令 (I0、I1、・・・、I7) 212 乃至 219 から成る。タグ 200 と、例えば有効ビットのようなライン状態 (LS) 221 とが各ラインに関連付けられる。キャッシュライン 210 及び 211 はそれぞれ、ラインアドレス 222 及び 223 の L1 I キャッシュラインアレイ 204 内に存在する。L1 I キャッシュ制御ユニット 206 は、キャッシュラインにアクセスするために命令アドレス (IA) インタフェース 225 を介して受け取られる命令アドレス (IA) 224 に応答するアドレス制御ロジックを含む。IA 224 は、タグ 226、ラインアドレス 228、命令「I」フィールド 232、及びバイト「B」フィールド 234 から成ることができる。ラインアドレス 228 は、「ライン+」フィールド 229 及び偶数「E」ビット 230 から成る。

【0016】

プロセッサ及びメモリコンプレックス 200 内の命令をフェッチするために、プロセッサ 202 は、フェッチされることが望まれる命令の命令アドレス (IA) 224 を生成し、そのフェッチアドレスを L1 I キャッシュ制御ユニット 206 へ送る。受信した IA 224 に基づいて、L1 I キャッシュ制御ユニット 206 は、命令が L1 I キャッシュラインアレイ 204 内に存在するかをチェックし調べる。このチェックは、例えば、IA 224 によって選択されたようなライン 210 内の一致するタグ 236 をチェックする比較ロジックを用いることにより達成される。例えば有効ビット 238 のようなライン状

10

20

30

40

50

態ビット (L S) 2 2 1 を確認することによって、ラインが有効命令を含むかに対しても判定が行われる。命令が存在する場合、一致すなわちヒットが生じ、L 1 I キャッシュ制御ユニット 2 0 6 は、L 1 I キャッシュ 2 0 3 内に命令が存在することを示す。命令が存在しない場合、不一致すなわちミスが発見され、L 1 I キャッシュ制御ユニット 2 0 6 は、命令が I キャッシュ 2 0 3 内に存在しないことを示すミスインジケーションを提供する。

【 0 0 1 7 】

命令が存在する場合、その命令フェッチアドレスの命令が、L 1 I キャッシュラインアレイ 2 0 4 から選択される。命令はその後、命令出力バス 2 4 0 でプロセッサ 2 0 2 へ送られる。

10

【 0 0 1 8 】

命令が命令キャッシュ内に存在しない場合、ミス情報が、ミス信号経路 2 4 2 を介してメモリ階層 2 0 8 へ提供され、ミスが発生したことを示す。レベル 1 I キャッシュ 2 0 3 内のミスを検出すると、プロセッサ及びメモリコンプレックス 2 0 0 は、望まれる命令をフェッチしようと試み、以下で更に詳しく説明するように、システムメモリの上位レベルから追加の命令をプリフェッチすることができる。例えば命令は、システムメモリ階層 2 0 8 に位置するレベル 2 (L 2) 統一キャッシュからプリフェッチされうる。ミスによって所望となった命令及び任意のプリフェッチされた命令は、メモリデータバスインタフェース 2 4 4 でシステムメモリ階層 2 0 8 から提供される。

【 0 0 1 9 】

20

ミスが発生すると、ミスの原因となったフェッチアドレスに関連する 1 つ又は複数の属性が判定されうる。例えばフェッチアドレスは、連続アクセス又は非連続アクセスの何れかのターゲットとして分類されうる。フェッチアドレスは更に、例えばそのアドレスがキャッシュラインのうちの X % を越えるかによっても分類されうる。フェッチアドレスはまた、そのアドレスが命令キャッシュ内の偶数アドレスであるかに基づいて分類されることもある。これらの分類及びその他の分類は、いつプリフェッチするかをインテリジェントに決定するヒューリスティックの形式で併用されうる属性を定義する。インテリジェントにプリフェッチすることによって、電力が節約され、メモリ帯域幅が改善され、ミスの割合が低減され、一般にプロセッサ及びシステムの性能は改善されうる。

【 0 0 2 0 】

30

ミスの原因となったフェッチアドレスに関連する属性は、フェッチアドレスから動的に決定することができる。例えば、フェッチアドレスをモニタすることによって、連続フェッチアドレスか非連続フェッチアドレスかが判定されうる。フェッチアドレス及びキャッシュラインの長さに基づいて、キャッシュライン内のアクセスの位置は、動的に決定することができる。例えば、決定された位置が閾値 X % に等しいか、又はそれを越えた場合、この属性は関連性を有する。また、フェッチアドレス及びキャッシュラインの長さに基づいて、フェッチアドレス内のビットは、アクセスが偶数キャッシュラインへのアクセスであるか奇数キャッシュラインへのアクセスであるかを判定するために用いられうる。これらの属性は、以下で更に詳しく説明される。

【 0 0 2 1 】

40

1 つの実施形態において、命令キャッシュ内のミスに、ミス属性のうちの 1 つ又は複数に関連している場合、メモリ階層の次のレベルから次の連続キャッシュラインがプリフェッチされる。1 つ又は複数のミス属性が関連していない場合、次の連続キャッシュラインはプリフェッチされない。一般に、連続していないキャッシュラインはプリフェッチされることが留意される。プリフェッチするキャッシュラインの決定は、ミスの属性、プログラムの属性、又はそのような属性の組み合わせに基づくことができる。例えば、分岐履歴テーブル又は分岐予測ロジックからの情報は、命令の連続ラインをプリフェッチするか命令の非連続ラインをプリフェッチするかを判定するために用いられうるプログラムの実行のフローの属性を提供するために用いられるであろう。

【 0 0 2 2 】

50

別の実施形態において、命令キャッシュ内のミスが、例えば命令キャッシュ内に存在しない命令のアドレスへの分岐、サブルーチンコール等のような非連続ミスによる場合、メモリ階層の次のレベルからの命令のラインがプリフェッチされる。命令キャッシュ内のミスが、例えば命令キャッシュ内に存在しない命令への連続するアクセスのような連続ミスによる場合、プリフェッチ動作は行われない。プリフェッチ動作のアドレスは、命令の連続ライン又は非連続ラインに對しうることも留意される。

【 0 0 2 3 】

別の実施形態において、命令キャッシュ内のミスが、命令キャッシュ内の偶数のアドレスで発生した場合、メモリ階層の次のレベルからの命令のラインがプリフェッチされる。命令キャッシュ内のミスが、命令キャッシュ内の奇数のアドレスで発生した場合、プリフェッチは行われない。

10

【 0 0 2 4 】

別の実施形態において、命令キャッシュ内のミスが、キャッシュラインのうちの X % に等しいアドレス又はそれを越えるアドレスでの非連続ミスによる場合、メモリ階層の次のレベルからの命令のラインがプリフェッチされる。命令キャッシュ内のミスが、キャッシュラインのうちの X % に満たないアドレスでのミス、あるいは連続ミスによる場合、プリフェッチは行われない。例えばキャッシュラインのうちの 75 % のように、キャッシュラインの末尾付近のフェッチアドレスでミスが発生した場合、メモリ階層の次のレベルからの命令のラインがプリフェッチされる。キャッシュラインのうちの割合 X は、試験的手段によって選ばれた値に定めることができる。又は X は、X % が、動作中のプログラムに適合できるようにプログラム可能でありうる。例えば、8 つの命令から成る命令ライン 210 を備えた図 2 の典型的な例示の場合、X % が 62.5 % であれば、ラインの 62.5 % は、I4 216、ラインのうちの 5 つの命令に関連する。ミスが、I4 216 のフェッチアドレス又はそれを越えるフェッチアドレスで発生した場合、プリフェッチ動作が可能である。ミスが、I3 215 のフェッチアドレス又はそれに満たないフェッチアドレスで発生した場合、プリフェッチ動作は可能ではない。代替実施形態において、命令キャッシュ内のミスが、キャッシュラインのうちの X % を越えたアドレスでのミスによる場合、プリフェッチが行われる。X % = 62.5 % である上記例の場合、I5 217 のフェッチアドレス又はそれを越えるフェッチアドレスでミスが発生した場合、プリフェッチが可能である。ミスが、キャッシュラインのうちの I4 216 に等しいかそれに満たないアドレスである場合、プリフェッチは行われない。

20

30

【 0 0 2 5 】

別の実施形態において、ミス属性は、プリフェッチするか否かを判定するために併用されうる。例えば、命令キャッシュ内のミスが、例えば命令キャッシュ内に存在しない命令への分岐、サブルーチンコール等のような非連続ミスによる場合、あるいは命令キャッシュ内のミスが、命令キャッシュ内の偶数のアドレスで発生した場合、メモリ階層の次のレベルからの命令のラインがプリフェッチされる。そうでない場合、ミスが発生しても、プリフェッチは行われない。

【 0 0 2 6 】

プリフェッチするかの判定は、アドレス指定された命令がキャッシュ内に存在するか否かを判定するために、キャッシュアクセスプローブと並行して動作することができるプリフェッチ判定ロジック 246 によって達成される。プリフェッチ判定ロジック 246 は、ミスインジケーション 248、例えば分岐、コール、又は復帰動作のような非連続動作がミスの原因となったことを示す非連続インジケーション 250、及び X % 値 252 を入力として選択し、例えばフェッチされた命令の I A 224 を含むことができる。ミスインジケーション 248、非連続インジケーション 250、及び X % 値 252 は、L1 I キャッシュ制御ユニット 206 によって提供される。ミスが、非連続フェッチ動作のフェッチアドレスに関連しているという判定は、プロセッサ 202 からの制御 254 から得ることができる。制御 254 は更に、例えば分岐、コール、復帰等のような命令に関連する非連続動作 256 のためのロジックから得ることができる。あるいは、L1 I キャッシュ制

40

50

御ユニット 206 は、フェッチが非連続動作に関連しているかを判定するために、現在のフェッチ命令アドレスと以前のフェッチ命令アドレスとを比較することができる。プリフェッチ判定ロジック 246 は、必要に応じて接続された信号によって命令キャッシュ外に位置しうることが留意される。例えばプリフェッチ判定ロジックは、命令アドレス 224 を生成するプロセッサ 202 に、より密接に関連付けられうる。

【0027】

例えば、フェッチアドレスが偶数アドレスキャッシュラインを示すかを判定するようなその他の属性は、偶数「E」ビット 230 に基づくことができる。ここで、偶数キャッシュラインの場合、E = 0 である。「E」ビット 230 アドレスビット位置又はビット値は、命令キャッシュラインアレイ 204 内のラインの長さに依存して変化しうる。図 2 に示すように、2 ビットの B フィールド 234 は、ワードにおいて最大 4 バイトを示し、3 ビットの I フィールド 232 は、ラインにおいて最大 8 ワードを示す。この構成によって、「E」ビット 230 は、32 バイトのキャッシュラインにアドレス指定された 8 ワード毎に状態を変化させる。

【0028】

更なる実施形態において、プリフェッチ判定ロジック 246 は、プリフェッチされるキャッシュラインが既に命令キャッシュ内に存在するかを、キャッシュのプロープに判定させることができる。プリフェッチターゲットに関連するキャッシュラインが既にキャッシュ内に存在する場合、プリフェッチ動作は停止されるか、開始を妨げられる。

【0029】

プリフェッチが、命令キャッシュに関連するプリフェッチロジックによって開始され、上位レベルのメモリ階層によって応答されると、命令の複数のラインが、例えばメモリデータバス 244 のようなメモリデータバスを介して L1 I キャッシュへ転送される。実現されるシグナリングプロトコル及びメモリデータバスインタフェース 244 の物理バス幅に依存して、データは、例えば 32 バイト、64 バイト、128 バイト、256 バイト、512 バイト等のような、定められた大きさの量又はパケットで転送されうる。一般に、短いデータパケットの複数の転送は通常、データを構成し送ることによるオーバーヘッドによって、大きいデータパケットサイズのバースト転送をするよりも効率が悪い。従って、別の実施形態において、プリフェッチする判定が下されると、例えば 2 つの個々のデータ転送において、命令の 2 つのラインをプリフェッチするのではなく、長いデータ転送が必要となる。プリフェッチにおける長いデータ転送によって、短い個別の転送で命令を送るのとは比べて、転送時間全体が最小化される。

【0030】

図 3 は、低複雑性プリフェッチ処理 300 を示すフローチャートである。図 3 のブロックと、図 2 の要素との間の典型的な関係は、プロセッサメモリコンプレックス 200 からの要素が方法 300 を実行するためにどのように適切に協働しうるかを説明することによって示される。

【0031】

ブロック 302 で、例えばプロセッサ 202 のようなプロセッサが命令をフェッチする。判定ブロック 304 で、フェッチされている命令が命令キャッシュ内に存在するかが判定される。例えば L1 I キャッシュ制御ユニット 206 が、命令が存在するかをチェックして調べる。命令が命令キャッシュ内に存在する場合、処理 300 は、ブロック 306 でキャッシュ内の命令をフェッチし、その後ブロック 302 へ戻り、次の命令フェッチを待つ。例えば、選択された命令は、命令出力バス 240 に置かれ、プロセッサ 202 内で受け取られる。

【0032】

命令が命令キャッシュ内に存在しない場合、ミス状況が検出されている。処理 300 は判定ブロック 308 へ進み、ここで、現在の命令アドレスに関連する次の連続ラインが既に命令キャッシュ内に存在するかが判定される。このような判定は、L1 I キャッシュラインアレイ 204 に直接アクセスせずに、L1 I キャッシュ制御ユニット 206 内で

達成することができる。次の連続ラインが L 1 I キャッシュ内に既に存在する場合、ブロック 310 で、現在の命令アドレスに対して、例えばシステムメモリ階層 208 のようなメモリ階層へフェッチが発行される。処理 300 はその後、ブロック 302 へ戻り、次の命令フェッチを待つ。並行して、ブロック 312 で、ミスの原因となったフェッチアドレスの属性が判定される。このような判定は、プリフェッチ判定ロジック 246 によって適切に達成されうる。

【0033】

次の連続ラインがキャッシュ内に存在しない場合、処理は、判定ブロック 314 へ進み、ここで、ブロック 312 で判定されたミス属性のうちの 1 つ又は複数に関連するかが判定される。複数の属性が判定されうるが、定められた又はプログラム可能な関数が、適用可能な属性全て又は適用可能な属性のサブセットを選択することができる。ミス属性が関連する場合、処理 300 はブロック 316 へ進み、命令及び決定されたプリフェッチキャッシュラインを次のレベルのメモリからフェッチし、ブロック 302 へ戻って次の命令フェッチを待つ。このように、判定ブロック 314 でプリフェッチする判定が下されると、命令の 2 つのラインをプリフェッチするのではなく、例えば 2 つの個々のデータ転送において、長いデータ転送が必要とされる。ミス属性が関連しない場合、ブロック 318 で、現在の命令アドレスに関して、例えばシステムメモリ階層 208 のようなメモリ階層へフェッチが発行される。処理 300 はその後、ブロック 320 へ戻り、次の命令フェッチを待つ。

【0034】

本明細書に開示された実施形態の典型的な説明は、直接マップされたキャッシュの実現形態を用いて示された。直接マップされたキャッシュは、例えば単一方向セット連想キャッシュのようなセット連想キャッシュの特別例と見なされうる。本明細書で説明された原理は、例えば 4 方向セット連想キャッシュ実現形態のような N 方向セット連想キャッシュにも適用可能である。同様に、本明細書で説明された原理は、完全連想キャッシュ実現形態にも適用可能である。

【0035】

本明細書に開示した実施形態に関連して説明された様々な例示的論理ブロック、モジュール、回路、要素、及び / 又は構成要素は、汎用プロセッサ、デジタル信号プロセッサ (DSP)、特定用途向け集積回路 (ASIC)、フィールド・プログラマブル・ゲート・アレイ (FPGA) あるいはその他のプログラマブル論理構成要素、ディスクリートゲートあるいはトランジスタロジック、ディスクリートハードウェア部品、又は上述された機能を実現するために設計された上記何れかの組合せを用いて実現又は実行されうる。汎用プロセッサとしてマイクロプロセッサを用いることが可能であるが、代わりに、従来技術によるプロセッサ、コントローラ、マイクロコントローラ、あるいは状態機器を用いることも可能である。プロセッサは、例えば DSP とマイクロプロセッサとの組み合わせ、複数のマイクロプロセッサ、DSP コアに接続された 1 つ又は複数のマイクロプロセッサ、又はこのような任意の構成である計算構成要素の組み合わせとして実現することも可能である。

【0036】

本明細書に開示された実施形態に関連して記述された方法は、ハードウェアによって直接、プロセッサによって実行されるソフトウェアモジュールによって、または、これらの組み合わせによって具現化されうる。ソフトウェアモジュールは、RAM メモリ、フラッシュメモリ、ROM メモリ、EPROM メモリ、EEPROM メモリ、レジスタ、ハードディスク、リムーバブルディスク、CD-ROM、あるいは当該技術分野で知られているその他の形式の記憶媒体に収納されうる。記憶媒体は、プロセッサがそこから情報を読み取り、またそこに情報を書き込むことができるようにプロセッサに結合されうる。または、記憶媒体はプロセッサに統合されうる。

【0037】

本発明は、実施形態に関して開示されたが、広範囲の様々な実現形態が、上記記載及び

10

20

30

40

50

以下の特許請求の範囲と矛盾せず当業者によって用いられうることが理解されるであろう。例えば実施形態は、主にレベル1命令キャッシュに関して開示された。これら実施形態は、レベル2命令キャッシュ、命令をプリフェッチするレベル2統一キャッシュ、又はプロセッサ・メモリ・コンプレックスに適用可能なその他のレベルのメモリ階層を用いて実現することもできる。命令キャッシュは、プロセッサコアに統合されるか、あるいはプロセッサコアに外付けであることができることも理解されるであろう。

以下に、本願出願の当初の特許請求の範囲に記載された発明を付記する。

[C 1]

命令プリフェッチのための方法であって、
フェッチアドレスが命令キャッシュ内でミスしたという通知を受け取ることと、
前記ミスの原因になったフェッチアドレスの属性を判定することと、
前記属性に基づいて、命令のラインをプリフェッチすることと
を備える方法。

10

[C 2]

前記属性は、前記フェッチアドレスが非連続動作のターゲットアドレスであることを示すC 1に記載の方法。

[C 3]

前記属性は、前記フェッチアドレスが非連続動作のターゲットアドレスであることと、
前記ターゲットアドレスは、キャッシュラインのうちのX %を越えることとを示すC 1に
記載の方法。

20

[C 4]

前記属性は、前記フェッチアドレスが、前記命令キャッシュ内の偶数アドレスであることを示すC 1に記載の方法。

[C 5]

前記属性は、前記フェッチアドレスが非連続動作のターゲットアドレスであること、又は前記フェッチアドレスが前記命令キャッシュ内の偶数アドレスであることを示すC 1に記載の方法。

[C 6]

前記プリフェッチすることは更に、
前記ミスした命令と、それに関連する命令のラインとを、前記プリフェッチする命令の
ラインと結合することによって、長データ転送を可能にすること
を更に備えるC 1に記載の方法。

30

[C 7]

前記プリフェッチすることは更に、
前記プリフェッチする命令のラインが前記命令キャッシュ内に存在するかを判定することと、
前記プリフェッチする命令のラインが前記命令キャッシュ内に存在しない場合、命令の
ラインをプリフェッチすることと
を更に備えるC 1に記載の方法。

40

[C 8]

前記プリフェッチすることは、
連続動作のターゲットアドレスで前記ミスが発生した原因になったフェッチアドレスを
前記属性として判定することと、
前記属性に基づいて、命令のラインをプリフェッチすることをディセーブルすることと
を更に備えるC 1に記載の方法。

[C 9]

前記非連続動作は、前記命令キャッシュに接続されたプロセッサ内のロジックから判定されるC 2に記載の方法。

[C 10]

前記非連続動作は、前記フェッチアドレスと前のフェッチアドレスとを比較することに

50

よって判定される C 2 に記載の方法。

[C 1 1]

フェッチアドレスがミスしたという通知を生成するように構成された命令キャッシュと

、
前記ミスの原因になったフェッチアドレスの属性を判定するように構成されたプリフェッチ判定ロジックと、

前記属性に基づいて命令のラインをプリフェッチするように構成されたプリフェッチロジックと

を備える命令キャッシュ及びプリフェッチシステム。

[C 1 2]

前記属性は、前記フェッチアドレスが、非連続動作のターゲットアドレスであることを示す C 1 1 に記載の命令キャッシュ及びプリフェッチシステム。

[C 1 3]

前記属性は、前記フェッチアドレスが非連続動作のターゲットアドレスであることと、前記ターゲットアドレスがキャッシュラインのうちの X % を越えることを示す C 1 1 に記載の命令キャッシュ及びプリフェッチシステム。

[C 1 4]

前記属性は、前記フェッチアドレスが、前記命令キャッシュ内の偶数アドレスであることを示す C 1 1 に記載の命令キャッシュ及びプリフェッチシステム。

[C 1 5]

前記属性は、前記フェッチアドレスが非連続動作のターゲットアドレスであること、又は前記フェッチアドレスが前記命令キャッシュ内の偶数アドレスであることを示す C 1 1 に記載の命令キャッシュ及びプリフェッチシステム。

[C 1 6]

前記命令キャッシュに接続されたプロセッサであって、前記フェッチアドレスを生成するように構成されたプロセッサと、

非連続分岐型動作を示す制御信号を生成するように構成された、前記プロセッサ内のロジックと

を更に備える C 1 1 に記載の命令キャッシュ及びプリフェッチシステム。

[C 1 7]

フェッチアドレスがミスしたという通知を生成するように構成された命令キャッシュと

、
前記ミスの原因になったフェッチアドレスの属性を判定し、前記ミスしたフェッチアドレスと前記属性とに基づいてミス情報を生成するように構成されたプリフェッチ判定ロジックと、

前記ミスしたフェッチアドレスに関連する命令によって、前記ミス情報に応答するように構成されたシステムメモリ階層と

を備えるメモリシステム。

[C 1 8]

前記ミスしたフェッチアドレスに関連する命令は、プリフェッチされた命令を含む C 1 7 に記載のメモリシステム。

[C 1 9]

前記ミスしたフェッチアドレスに関連する命令は、バースト転送によって前記命令キャッシュへ提供される C 1 8 に記載のメモリシステム。

[C 2 0]

前記属性は、前記フェッチアドレスが非連続動作のターゲットアドレスであることを示す C 1 7 に記載のメモリシステム。

10

20

30

40

【図 1】

図 1

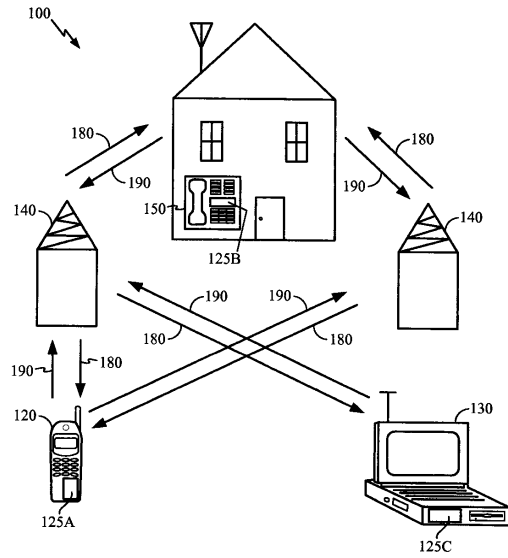


FIG. 1

【図 2】

図 2

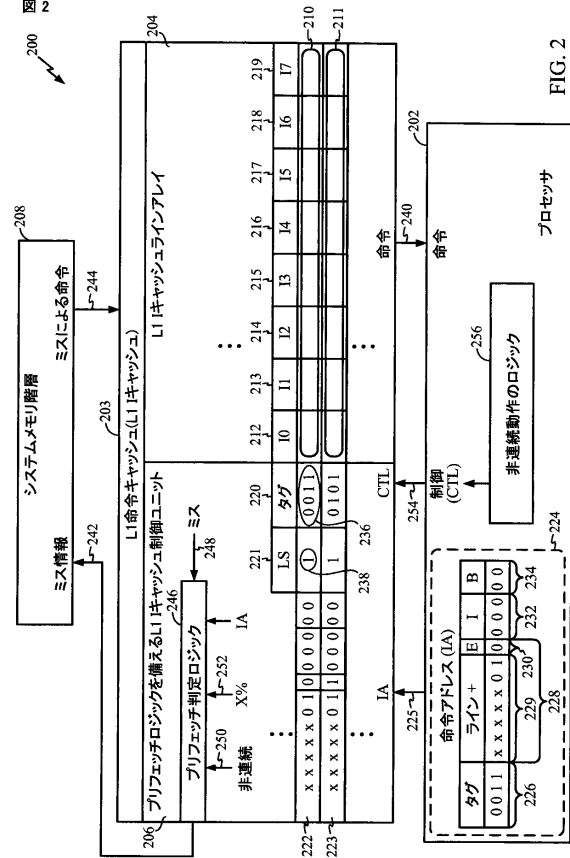


FIG. 2

【図 3】

図 3

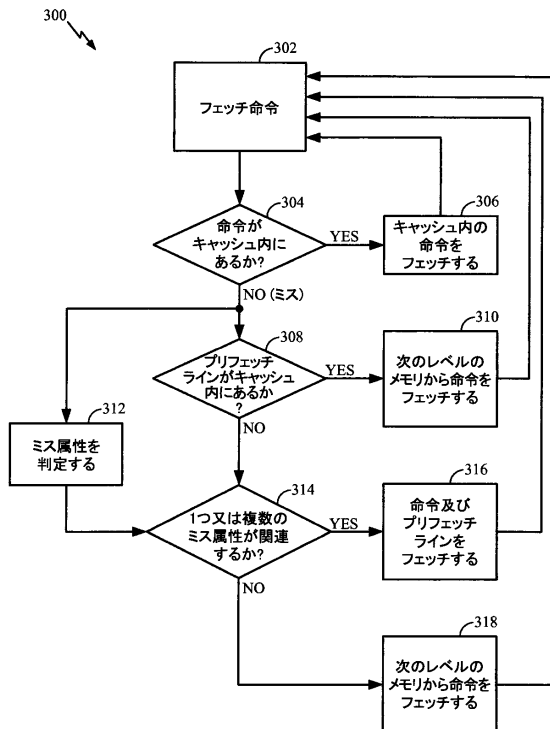


FIG. 3

フロントページの続き

- (72)発明者 モロー、マイケル・ウィリアム
アメリカ合衆国、カリフォルニア州 92121、サン・ディエゴ、モアハウス・ドライブ 57
75
- (72)発明者 ディーフェンダーファー、ジェームズ・ノリス
アメリカ合衆国、カリフォルニア州 92121、サン・ディエゴ、モアハウス・ドライブ 57
75

審査官 三坂 敏夫

- (56)参考文献 特開平10-232776(JP,A)
特開昭63-146143(JP,A)
特開平02-301843(JP,A)
特開2000-347934(JP,A)
特開昭61-005357(JP,A)
特開2001-142698(JP,A)
米国特許第06119222(US,A)
特開平08-123723(JP,A)
欧州特許第00772123(EP,B1)
特開平09-160827(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G06F 12/08
G06F 9/38