(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第3717212号

(P3717212)

(45)発行日　平成17年11月16日(2005.11.16)　(24)登録日　平成17年9月9日(2005.9.9)

(51)Int.Cl.[7]　F I

G06F 12/08　G06F 12/08　505B

G06F 12/08　509Z

G06F 12/08　561

G06F 12/08　563

請求項の数 4　(全 13 頁)

(21)出願番号　特願平7-280836

(22)出願日　平成7年10月27日(1995.10.27)

(65)公開番号　特開平9-128293

(43)公開日　平成9年5月16日(1997.5.16)

審査請求日　平成14年10月24日(2002.10.24)

(73)特許権者　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号

(74)復代理人　100084940
弁理士　丸山　清

(74)復代理人　100087354
弁理士　市村　裕宏

(74)復代理人　100093492
弁理士　鈴木　市郎

(74)復代理人　100099520
弁理士　小林　一夫

(74)復代理人　100106758
弁理士　橘　昭成

(74)代理人　100078134
弁理士　武　顕次郎

最終頁に続く

(54)【発明の名称】情報処理装置及び情報処理ユニット

(57)【特許請求の範囲】

【請求項１】

主記憶と階層化された複数のキャッシュを有し、演算で用いるオペランドデータを該演算の実行に先立って予め主記憶からキャッシュに転送するソフトウェアプリフェッチ命令を実行する情報処理装置において、前記ソフトウェアプリフェッチ命令の命令コードにプリフェッチするデータの転送先のキャッシュ階層を指定する指示ビットを設け、プリフェッチ命令実行時、前記指示ビットの内容に応じて指示された階層のキャッシュにデータの転送を行うことを特徴とする情報処理装置。

【請求項２】

前記指示ビットにより指示された内容が、プリフェッチしたオペランドデータを前記階層化された複数のキャッシュの全ての階層のキャッシュに転送する指示であることを特徴とする請求項１記載の情報処理装置。

【請求項３】

前記指示ビットにより指示される内容が、プリフェッチしたオペランドデータを前記階層化された複数のキャッシュの１次キャッシュ以外の全ての階層のキャッシュに転送する指示であることを特徴とする請求項１記載の情報処理装置。

【請求項４】

１次キャッシュと２次キャッシュとからなる２階層キャッシュを有し、演算で用いるオペランドデータを該演算の実行に先立って予め外部主記憶からキャッシュに転送するソフトウェアプリフェッチ命令を実行する情報処理ユニットにおいて、前記ソフトウェアプリ

フェッチ命令の命令コードにプリフェッチするデータの転送先のキャッシュ階層を指定する指示ビットを設け、プリフェッチ命令実行時、前記指示ビットの内容に応じて指定された階層のキャッシュにオペランドデータを転送することを特徴とする情報処理ユニット。

【発明の詳細な説明】

【０００１】

【発明が属する技術分野】

本発明は、階層構成のキャッシュを有する情報処理装置及び情報処理ユニットに係り、特に、演算に使用するオペランドデータを予め主記憶からキャッシュに転送しておくことにより、キャッシュミスによるオーバヘッドを隠すことができるようにしたソフトウェアプリフェッチ命令を備えた情報処理装置及び情報処理ユニットに関する。

【０００２】

【従来の技術】

一般に、キャッシュを有する情報処理装置は、命令により参照したオペランドがキャッシュに存在しない場合、すなわち、キャッシュミスが生じた場合、前記オペランドを主記憶から読み出して使用している。通常、この読み出しには、キャッシュアクセスの数倍から数十倍の時間を要する。このため、この種の情報処理装置は、キャッシュミスが生じると、オペランドが主記憶から読み出されるまでの間、後続命令の実行を待たせることになり、実行時間が長くなってその性能が抑えられてしまうという問題点を有している。

【０００３】

前述の問題点を解決することのできる技術として、予め将来用いるオペランドデータを主記憶からキャッシュに転送し、そのオペランドデータを用いるときにはキャッシュヒットとなるようにして、キャッシュミスのペナルティを抑えるようにした技術が知られており、これを実現するためのソフトウェアプリフェッチ命令に関する研究がなされ、様々な情報処理装置で使用されている。

【０００４】

ソフトウェアプリフェッチ命令に関する従来技術として、例えば、論文　Callahan,D.,Kennedy,K.Porterfield,A.,“Software Prefetching," Proceedings of the 4th International Conference on Architectural Support for Programming Languages and Operating Systems,April 1991,pp.40-52等に記載された技術が知られている。

【０００５】

以下、従来技術による情報処理装置によるソフトウェアプリフェッチ命令の動作を図面により説明する。

【０００６】

図５は従来技術による情報処理装置の構成例を示すブロック図、図６はソフトウェアプリフェッチの有無による処理動作を説明するタイムチャートである。図５において、２１はＣＰＵ（中央処理装置）、２２は１次キャッシュ、２４はＳＣＵ（記憶制御装置）、２５は主記憶である。

【０００７】

図５に示す従来技術は、１階層キャッシュを持つ情報処理装置の例であり、この情報処理装置は、情報の処理を行うＣＰＵ２１と、１次キャッシャ２２と、主記憶２５と、主記憶２５に対する情報の書き込み、読み出しを制御するＳＣＵ２４とにより構成されている。この情報処理装置において、ＣＰＵ２１は、オペランドデータを参照する際１次キャッシュ２２を検索し、キャッシュミスの場合ＳＣＵ２４に対して、リクエスト線２０１、アドレス線２０２を介して該オペランドデータの転送要求を発行する。ＳＣＵ２４は、主記憶２５から該オペランドデータを読み出し、そのデータをデータ線２０４を介してＣＰＵ２１に転送する。ＣＰＵ２１は、受け取ったオペランドデータを１次キャッシュ２２に格納すると共に演算のために使用する。

【０００８】

次に、ＣＰＵ２１による命令の処理が、ＩＦ（命令フェッチ）、Ｄ（デコード）、Ｅ（実行）、Ａ（オペランドアクセス）、Ｗ（レジスタ書き込み）の５段のパイプラインステー

ジにより処理されるものとして、図５に示す情報処理装置の動作を説明する。

【０００９】

前述のように構成される情報処理装置において、命令０～命令３が順次パイプラインに投入されて処理が行われ、ソフトウェアプリフェッチを行わない場合のタイムチャートが図６（ａ）に示されている。

【００１０】

図６（ａ）に示す例は、ロード命令である命令１のオペランドアクセスでキャッシュミスが生じたとしている。この場合、命令１の処理は、オペランドアクセスを行うＡステージで、主記憶からオペランドが読み出されるまでの間待たされることになり、命令１のＷステージの処理は、それまで実行されずに待たされることになる。これに伴い、命令２、３のＡ、Ｅステージ及び以降の命令処理が待たされることになるため、キャッシュミスにより主記憶からオペランドを読み出すために要する時間の全てがペナルティとして表れてしまう。

【００１１】

図６（ｂ）に示すタイムチャートは、ソフトウェアプリフェッチを行う場合の例であり、この場合、ロード命令である命令１の実行に対して、オペランドを主記憶から転送するに要する時間だけ先立って、命令１’として示すソフトウェアプリフェッチ命令を実行する。この結果、命令１’によるソフトウェアプリフェッチによってオペランドが主記憶から転送される間、パイプラインステージは、処理を中断することなく先に進み、命令１’に続く命令２、命令３の処理が実行される。そして、命令１がオペランドデータをアクセスする時点では、この命令１が必要とするオペランドデータは、ソフトウェアプリフェッチ命令である命令１’により１次キャッシュ２２に格納されていることになり、キャッシャヒットとなる。これにより、キャッシュミスによるペナルティを隠すことができる。

【００１２】

図７は従来技術による情報処理装置の他の構成例を示すブロック図、図８はソフトウェアプリフェッチが行われる場合の処理動作を説明するタイムチャートである。図７において、２３は２次キャッシュであり、他の符号は図５の場合と同一である。

【００１３】

図７に示す従来技術は、２階層キャッシュを持つ情報処理装置の構成例であり、この情報処理装置は、ＣＰＵ２１内に１次キャッシュ２２を内蔵し、２次キャッシュ２３を備えて構成される点で図５に示す従来技術と相違し、その他の点で同一である。

【００１４】

図７に示す従来技術において、ＣＰＵ２１は、オペランドデータを参照する際、まず、１次キャッシュ２２を検索し、１次キャッシュ２２がキャッシュミスの場合、２次キャッシュ２３を検索する。２次キャッシュ２３がヒットであれば、２次キャッシュ２３から１次キャッシュ２２へオペランドデータが転送される。以降の説明において、この転送をブロック転送、転送されるオペランドデータをブロックと呼ぶこととする。

【００１５】

２次キャッシュ２３がキャッシュミスとなると、ＳＣＵ２４は、主記憶２５からそのオペランドデータを読み出してＣＰＵ２１に転送する。以降の説明において、この転送をライン転送、転送されるデータをラインと呼ぶこととする。通常、ブロックのデータ量はラインのデータ量より小さく、ライン転送の際に２次キャッシュ２３にはラインのデータが格納され、１次キャッシュ２２には参照された１ブロックのデータのみが格納される。

【００１６】

次に、前述した場合と同様に、ＣＰＵ２１による命令の処理が、ＩＦ（命令フェッチ）、Ｄ（デコード）、Ｅ（実行）、Ａ（オペランドアクセス）、Ｗ（レジスタ書き込み）の５段のパイプラインステージにより処理されるものとして、図７に示す情報処理装置の動作を説明する。

【００１７】

前述のように構成される情報処理装置において、命令０～命令３が順次パイプラインに投

入されて処理が行われる場合の処理が図８に示されている。図８（ａ）は１次キャッシュミスを防止するためのソフトウェアプリフェッチ命令を実行する場合のタイムチャート、図８（ｂ）は１次キャッシュミス、２次キャッシュミスの両方を防止するためのソフトウェアプリフェッチ命令を実行する場合のタイムチャートである。

【００１８】

通常、１次キャッシュミス、２次キャッシュミスによるペナルティサイクル数は、それらの間に相当に大きな差異がある。例えば、前者が４～５サイクル、後者が３０～４０サイクル程度である。従って、前述した２階層キャッシュ構成の情報処理装置において、オペランドデータの１次キャッシュミスを避けるためのソフトウェアプリフェッチ命令の実行は、実際にオペランドデータを参照する命令の少なくとも５サイクル以上前に行われればよいが、２次キャッシュミスを避けるためのソフトウェアプリフェッチ命令の実行は、実際にオペランドデータを参照する命令の少なくとも４０サイクル以上前に行う必要がある。すなわち、前述した従来技術は、１次キャッシュミスを避けるか、２次キャッシュミスを避けるかによって、ソフトウェアプリフェッチ命令と実際にオペランドデータを参照する命令との間隔を分けなければならない。

【００１９】

【発明が解決しようとする課題】

前述した２階層以上のキャッシュ構成を有する従来技術による情報処理装置は、オペランドデータがどの階層のキャッシュにヒットするかによりソフトウェアプリフェッチ命令を実行する時期を変える必要がある。このため、前述の従来技術は、コンパイラがソフトウェアプリフェッチ命令を生成しようとする際の制御が複雑になり、ソフトウェアプリフェッチ命令の効果を十分に生かすことが困難になるという問題点が生じる。

【００２０】

図９はソフトウェアプリフェッチ動作の対象となる連続領域に配置された配列データのアドレス関係を示す図、図１０はソフトウェアプリフェッチ命令を用いて図９に示すように配列されているオペランドデータを転送する場合の動作を説明するタイムチャートである。

【００２１】

以下、図７に示す装置においてブロックサイズが３２Ｂであり、ラインサイズが１２８Ｂであり、また、ブロック転送には４サイクル、ライン転送には４０サイクルを要するものとして、前述の従来技術の問題点を説明する。

【００２２】

ソフトウェアプリフェッチ命令が使用される例として、最も一般的な連続領域に配置されている配列Ａ［ｉ］（ｉ＝０、１、２、……）を順番に参照する場合を考える。配列Ａ［ｉ］の個々のデータサイズを８Ｂとすると、ライン、ブロックのアドレス関係は図９に示すようになる。

【００２３】

初期状態において、配列Ａ［ｉ］のデータは、全て１次、２次キャッシュともにキャッシュミスとなったとする。この場合、まず、Ａ［０］のデータを主記憶からライン転送するため、Ａ［０］のデータを参照する命令の４０サイクル前にソフトウェアプリフェッチ命令が実行される。この実行により、１ブロック相当の３２Ｂのデータ、すなわち、Ａ［０］からＡ［３］までのデータが１次キャッシュヒットになる。同様に、ライン相当の１２８Ｂのデータ、すなわち、Ａ［０］からＡ［１５］までのデータが２次キャッシュヒットになる。

【００２４】

従って、その後の命令の処理におけるＡ［１］からＡ［３］のデータの参照は、すでにこれらのデータが１次キャッシュ２２に存在するため、ソフトウェアプリフェッチ命令を必要としない。しかし、Ａ［４］のデータを必要とする場合、Ａ［４］のデータは、２次キャッシュにのみ存在するので、２次キャッシュからブロック転送を行うため、Ａ［４］のデータを参照する命令の４サイクル前にソフトウェアプリフェッチ命令が実行される。こ

の実行により、１ブロック相当の３２Ｂのデータ、すなわち、Ａ［４］からＡ［７］までのデータが１次キャッシュヒットになる。Ａ［８］以降のデータの参照は、前述と同様なソフトウェアプリフェッチ命令の発行を繰り返えすことにより可能となる。前述の動作のタイムチャートを示したのが図１０である。

【００２５】

図１０から判るように、また、図８によりすでに説明したように、ソフトウェアプリフェッチ命令の実行間隔は一定でなく、不規則なものになっている。この理由は、主記憶からライン転送を行うためのソフトウェアプリフェッチ命令と、２次キャッシュからブロック転送を行うためのソフトウェアプリフェッチ命令との実行間隔が一致していないためである。

【００２６】

一方、前述したような配列を持つデータに対するアクセスおよび演算は、極めて規則的な繰り返しであるため、通常、分岐命令によるループで構成される命令列により実行される。ソフトウェアプリフェッチ命令の実行間隔が不規則であると、これらのソフトウェアプリフェッチ命令をループに組み込むことが困難となる。また、ライン転送かブロック転送かを判断するための命令が必要になり、このため、命令数の増加を招きソフトウェアプリフェッチ命令による処理装置の性能の向上が妨げられることになる。さらに、１ループでライン全てのデータを扱うようにした場合でも、同様に命令数の増加になりソフトウェアプリフェッチ命令による処理装置の性能の向上が妨げられる。

【００２７】

本発明の目的は、前述した従来技術の問題点を解決し、２階層以上のキャッシュ構成を有する情報処理装置において、コンパイラが命令列を生成しやすいソフトウェアプリフェッチ命令を提供し、キャッシュミスによる性能の低下を効果的に抑えることができ、これにより、性能の向上が図られた情報処理装置及び情報処理ユニットを提供することにある。

【００２８】

【課題を解決するための手段】

本発明によれば前記目的は、主記憶と階層化された複数のキャッシュを有し、演算で用いるオペランドデータを該演算の実行に先立って予め主記憶からキャッシュに転送するソフトウェアプリフェッチ命令を実行する情報処理装置において、前記ソフトウェアプリフェッチ命令の命令コードにプリフェッチするデータの転送先のキャッシュ階層を指定する指示ビットを設け、プリフェッチ命令実行時、前記指示ビットの内容に応じて指示された階層のキャッシュにデータの転送を行うことにより達成される。

【００２９】

また、前記目的は、前記情報処理装置が２階層のキャッシュを有する情報処理装置であり、前記指示ビットにより指示される内容が、プリフェッチするキャッシュの階層の指定、プリフェッチするオペランドのデータ量の指定、または、その両方であり、あるいは、主記憶から２次キャッシュへある量のオペランドデータの転送を行うとき、該オペランドデータを同じ量だけ１次キャッシュへもプリフェッチする指示、または、ソフトウェアプリフェッチでない通常のキャッシュアクセスを行う命令により転送されるデータ量の整数倍を転送する指示とすることにより達成される。

【００３０】

本発明は、２階層以上のキャッシュ構成を有する情報処理装置において、コンパイラが命令を生成するとき、前記指示ビットを用いて、転送を行うキャッシュの階層または転送するデータサイズを明示的に指定することができるので、規則的にソフトウェアプリフェッチ命令を生成することが可能となる。このため、本発明は、命令列が、配列データのアクセス、配列データの演算のように命令のループで構成される場合に、アドレス関係を判定する命令等を生成する必要をなくすことができる。

【００３１】

また、本発明は、アドレスが連続するオペランドデータに対するアクセスを行う場合、１個のソフトウェアプリフェッチ命令により、一度に複数ラインのデータを転送することが

できるので、ソフトウェアプリフェッチ命令の数を減らすことが可能となる。

【００３２】

さらに、本発明によるセットアーキテクチャと従来技術のセットアーキテクチャとの互換性を容易に実現することができる。例えば、ブロックサイズが３２バイトの場合、ソフトウェアプリフェッチ命令のオペランドアドレスの下位５ビットは、アドレスとして用いられることはない。この５ビットを使用して前述した指示を行うようにすれば、命令セットのアーキテクチャを拡張することなく、互換性を保つことができる。

【００３３】

本発明は、以上のように、ソフトウェアプリフェッチ命令を有効に用いることができ、これにより、情報処理装置の性能の向上を図ることができる。

【００３４】

【発明の実施の形態】

以下、本発明による情報処理装置及び情報処理ユニットの一実施形態を図面により詳細に説明する。

【００３５】

図１は本発明の一実施形態で使用するソフトウェアプリフェッチ命令のオペランドアドレスのフォーマットを示す図、図２は本発明の一実施形態による情報処理装置の構成例を示すブロック図、図３はソフトウェアプリフェッチ命令のオペランドアドレス中の指示ビットにより指示される動作を説明する図、図４は本発明の一実施形態によるソフトウェアプリフェッチ命令動作を説明するタイムチャートである。図２における符号は図７の場合と同一である。

【００３６】

以下に説明する本発明の一実施形態において、２次キャッシュから１次キャッシュに転送するデータのブロックサイズは３２Ｂであり、主記憶から２次キャッシュに転送するデータのラインサイズは１２８Ｂであるとする。図１に示すソフトウェアプリフェッチ命令のオペランドアドレスのフォーマットにおいて、オペランドアドレスは、ブロックサイズが３２Ｂであることより、オペランドアドレスの下位５ビットがソフトウェアプリフェッチ命令では不要である。このため、本発明の一実施形態では、この下位５ビットを、プリフェッチ動作の内容を指示するビット（以下、ＰＦビットという）として用いる。

【００３７】

この下位５ビットのＰＦビットの値とソフトウェアプリフェッチ動作との対応の例が図３に示されており、以下、これについて説明する。

【００３８】

ＰＦビットの値が０のとき、何の指示もないことを示す。このとき、ソフトウェアプリフェッチは、従来技術の場合と同様に実行される。この指示は従来技術との互換性を保ちたい場合に有効である。

【００３９】

ＰＦビットの値が１のとき、オペランドアドレスで示されるラインの全てを１次キャッシュ２２にプリフェッチすることを示す。２次キャッシュヒットの場合、１ライン分１２８Ｂ、すなわち、４ブロックを１次キャッシュ２２に転送する。２次キャッシュミスの場合、主記憶２５から当該ラインをライン転送し、１次および２次キャッシュに格納することを示す。

【００４０】

ＰＦビットの値が２のとき、オペランドアドレスで示されるブロックを２次キャッシュ２３から１次キャッシュ２２へブロック転送することを示す。２次キャッシュミスの場合、主記憶から当該ラインのライン転送は行わない。

【００４１】

ＰＦビットの値が３のとき、オペランドアドレスで示されるラインを主記憶２５から２次キャッシュ２３のみへ転送することを示す。すなわち、２次キャッシュミスの場合、主記憶２５から当該ラインを２次キャッシュにライン転送する。１次キャッシュ２２へは転送

しない。
【００４２】
ＰＦビットの値が４のとき、オペランドアドレスで示されるラインを含むアラインされた５１２Ｂ、すなわち、４ラインを主記憶から２次キャッシュのみへ転送することを示す。
【００４３】
次に、図２に示す本発明の一実施形態による情報処理装置におけるソフトウェアプリフェッチ命令の動作を説明する。図示本発明が適用される情報処理装置は、２階層のキャッシュ構成を有する情報処理装置であり、図７に示す従来技術の場合と同一の構成を有する。
【００４４】
図２において、いま、ソフトウェアプリフェッチ命令が実行され、該命令のＰＦビットの値が１であるとする。この場合、ＣＰＵ２１は、オペランドアドレスにより２次キャッシュ２３を検索し、２次キャッシュヒットであれば、２次キャッシュ２３よりオペランドアドレスで示される１ライン１２８Ｂのデータを１次キャッシュ２２へ転送する。また、２次キャッシュ２３がミスであれば、ＣＰＵ２１は、リクエスト線２０１、アドレス線２０２、ＰＦビット線２０３を介してライン転送要求をＳＣＵ２４へ送出する。ＳＣＵ２４は、主記憶２５より１ライン１２８Ｂのデータを読み出し、そのデータをデータ線２０４を介してＣＰＵ２１へ転送する。ＣＰＵ２１は、受け取ったオペランドデータ１２８Ｂを２次キャッシュへ格納すると共に、１次キャッシュ２２へも１２８Ｂの全てのデータを格納する。
【００４５】
この動作により、１ライン１２８Ｂのデータの全てを１つのソフトウェアプリフェッチ命令により１次キャッシュ２２にプリフェッチすることができる。
【００４６】
ＰＦビットの値が２であるとする。この場合、ＣＰＵ２１は、オペランドアドレスにより２次キャッシュ２３を検索し、２次キャッシュヒットであれば、２次キャッシュ２３よりオペランドアドレスで示される１ブロック３２Ｂのデータを１次キャッシュ２２へ転送する。また、２次キャッシュ２３がミスであれば、ＣＰＵ２１は、何も動作を起こさない。すなわち、この場合、ＣＰＵ２１は、ＳＣＵ２４へはライン転送要求を送らない。
【００４７】
この動作により、予め２次キャッシュ２３へプリフェッチされているオペランドデータを１次キャッシュ２２へブロック転送することができる。
【００４８】
ＰＦビットの値が３であるとする。この場合、ＣＰＵ２１は、オペランドアドレスにより２次キャッシュ２３を検索し、２次キャッシュヒットであれば、ＣＰＵ２１は何も動作を起こさない。すなわち、この場合、２次キャッシュ２３から１次キャッシュ２２へのデータの転送は行わない。また、２次キャッシュ２３がミスであれば、ＣＰＵ２１は、リクエスト線２０１、アドレス線２０２、ＰＦビット線２０４を介してライン転送要求をＳＣＵ２４へ送出する。ＳＣＵ２４は、これにより、主記憶２５より１ライン１２８Ｂのデータを読み出し、このデータをデータ線２０４を介してＣＰＵ２１へ転送する。ＣＰＵ２１は、受け取ったオペランドデータ１２８Ｂを２次キャッシュへ格納する。この場合、１次キャッシュへの格納は行わない。
【００４９】
この動作により、１次キャッシュに影響を与えることなく、主記憶２５から２次キャッシュ２３へデータをプリフェッチすることができる。
【００５０】
ＰＦビットの値が４であるとする。この場合、ＣＰＵ２１は、リクエスト線２０１、アドレス線２０２、ＰＦビット線２０３を介しライン転送要求をＳＣＵ２４へ送出する。ＳＣＵ２４は、これにより、主記憶２５より４ライン分の合計５１２Ｂのデータを読み出し、このデータをデータ線２０４を介してＣＰＵ２１へ転送する。ＣＰＵ２１は、受け取ったオペランドデータ５１２Ｂを２次キャッシュ２３へ格納する。この場合、１次キャッシュ

への格納は行わない。

【００５１】

この動作により、通常のキャッシュアクセスを行う命令により転送されるデータ量の４倍（４倍に限らず整数倍であればよい）の４ライン分の５１２Ｂのデータの全てを、１つのソフトウェアプリフェッチ命令により１次キャッシュに影響を与えることなく、主記憶２５から２次キャッシュ２３へプリフェッチすることができる。

【００５２】

前述したＰＦビットの値が４の場合において、ＣＰＵ２１がＳＣＵ２４に対し４回ライン転送要求を発行するようにして、５１２Ｂのデータを２次キャッシュ２３へ格納することもできる。これにより、ＳＣＵ２４の設計を全く変更することなく、本発明によるソフトウェアプリフェッチを実現することができる。

【００５３】

前述した本発明の一実施形態によるソフトウェアプリフェッチの動作のタイムチャートが図４に示されている。この例では、ブロック転送に４サイクル、ライン転送に４０サイクルを要するものとしている。

【００５４】

ソフトウェアプリフェッチ命令が使用される例として、従来技術において説明した最も一般的な連続領域に配置されている配列Ａ［ｉ］（ｉ＝０、１、２、……）を順番に参照する場合を例とし、配列Ａ［ｉ］の個々のデータサイズを８Ｂとすると、ライン、ブロックのアドレス関係は、すでに説明した図９に示すようになる。

【００５５】

図４に示す例は、ソフトウェアプリフェッチ命令内のＰＦビットの値を１として、１ライン１２８Ｂのデータを１次キャッシュ２２にプリフェッチする場合のタイムチャートである。この場合、１２８Ｂ、すなわち配列データを１６個アクセスする度にその４０サイクル前にソフトウェアプリフェッチ命令を発行すればよいことになる。

【００５６】

前述した本発明一実施形態は、ソフトウェアプリフェッチ命令内のＰＦビットを命令のオペランドアドレスの使用しないビットに設けるとし、また、ＰＦビットによる指示内容をその５種類について例示して説明したが、本発明は、ＰＦビットを命令のオペランドコードの使用しないビットに設けるようにすることもでき、また、指示内容をさらに多種類とすることもできる。

【００５７】

また、前述した本発明一実施形態は、２層のキャッシュを備える情報処理装置に本発明を適用したものとして説明したが、本発明は、さらに多層のキャッシュを備える情報処理装置に対しても適用することができる。

【００５８】

【発明の効果】

以上説明したように本発明によれば、ソフトウェアプリフェッチ命令の中で、データを転送するキャッシュの階層を明示的に指示することができ、また、ソフトウェアプリフェッチ命令を規則的な間隔で発行することができるので、コンパイラにとって命令コードを生成しやすく、アドレスを判定するための余分な分岐命令も不要とすることができる。

【００５９】

また、本発明によれば、ソフトウェアプリフェッチ命令の中で、転送するデータ量をライン、ブロックの整数倍に指定することができるので、大量のデータを少ない数のソフトウェアプリフェッチ命令により転送することができる。さらに、本発明によれば、ソフトウェアプリフェッチ命令の中の従来のオペランドコードで未使用であった部分、または、オペランドアドレスの下位の未使用である部分に、前述の指示ビットを設けることができるので、従来技術の場合とのアーキテクチャ上の互換性を保つことができる。

【００６０】

本発明によれば、前述により、ソフトウェアプリフェッチ命令による情報処理装置の性能

の向上を効果的に図ることができる。

【図面の簡単な説明】

【図１】本発明の一実施形態で使用するソフトウェアプリフェッチ命令のオペランドアドレスのフォーマットを示す図である。

【図２】本発明の一実施形態による情報処理装置の構成例を示すブロック図である。

【図３】ソフトウェアプリフェッチ命令のオペランドアドレス中の指示ビットにより指示される動作を説明する図である。

【図４】本発明の一実施形態によるソフトウェアプリフェッチ命令動作を説明するタイムチャートである。

【図５】従来技術による情報処理装置の構成例を示すブロック図である。

【図６】ソフトウェアプリフェッチの有無による処理動作を説明するタイムチャートである。

【図７】従来技術による情報処理装置の他の構成例を示すブロック図である。

【図８】ソフトウェアプリフェッチが有る場合の処理動作を説明するタイムチャートである。

【図９】ソフトウェアプリフェッチ動作の対象となる連続領域に配置された配列データのアドレス関係を示す図である。

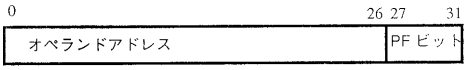
【図１０】ソフトウェアプリフェッチ命令を用いて図９に示すオペランドデータを転送する場合の動作を説明するタイムチャートである。

【符号の説明】

２１　ＣＰＵ
２２　１次キャッシュ
２３　２次キャッシュ
２４　ＳＣＵ
２５　主記憶
２０１　リクエスト線
２０２　アドレス線
２０３　ＰＦビット線
２０４　データ線

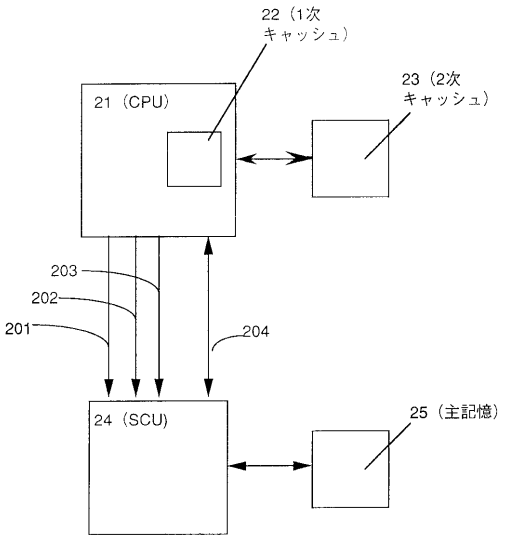
【図１】

【図１】



【図２】

【図２】



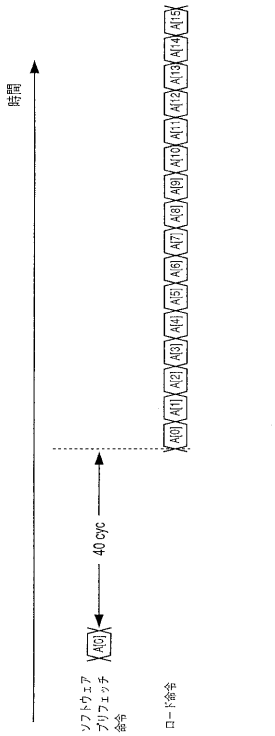
【図３】

【図３】

PFビット値	ソフトウェアプリフェッチ動作
0	無し（従来と同じ）
1	主記憶から1ライン128Bを1次/2次キャッシュ両方に転送する
2	2次キャッシュから1ブロック32Bを1次キャッシュへ転送する
3	主記憶から1ライン128Bを/2次キャッシュのみへ転送する
4	主記憶から4ライン512Bを2次キャッシュのみへ転送する
・ ・ ・	・ ・ ・

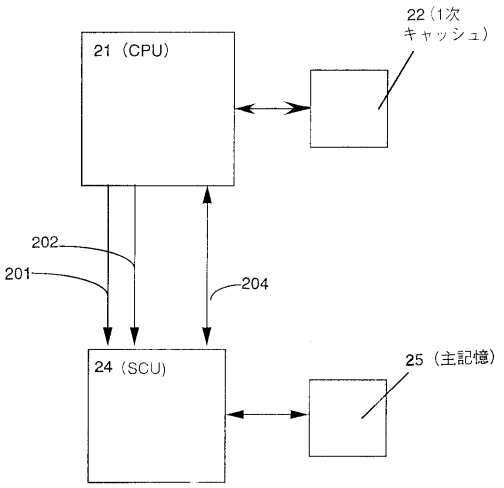
【図４】

【図４】



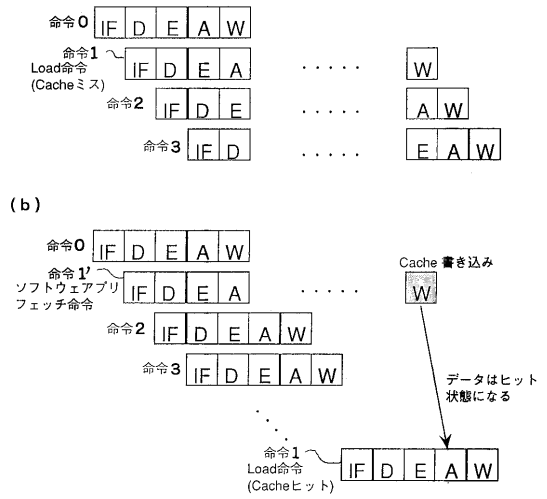
【図５】

【図５】



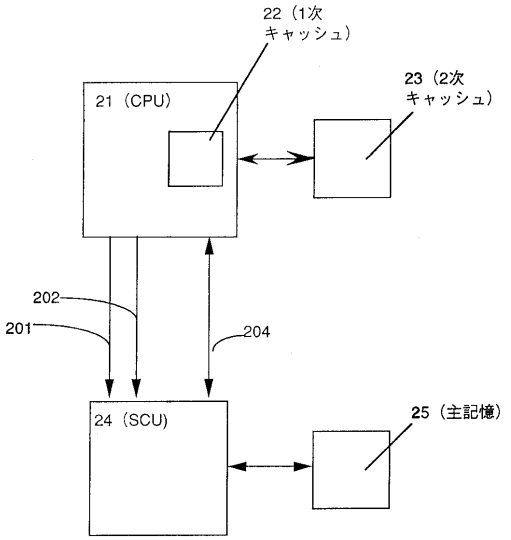
【図６】

【図６】



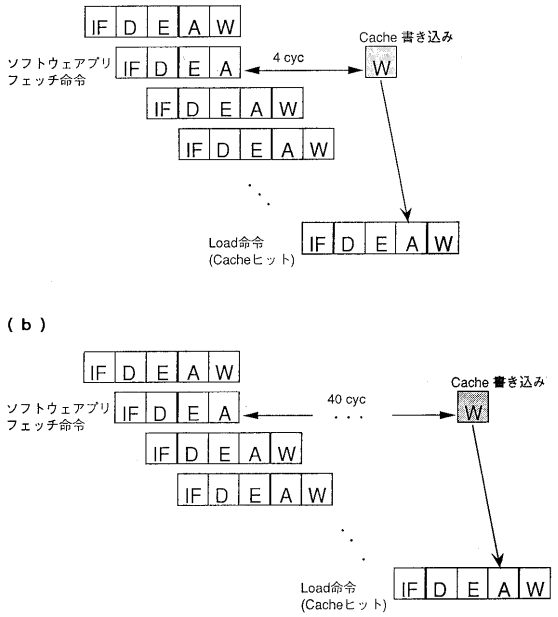
【図７】

【図７】



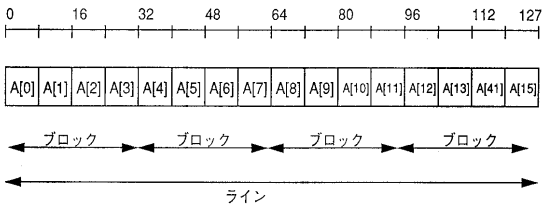
【図８】

【図８】



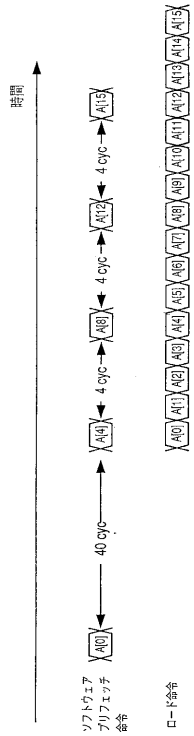
【図９】

【図９】



【図１０】

【図10】


時間
ソフトウェアプリフェッチ命令
A[0]
40 cyc
A[4]
4 cyc
A[8]
4 cyc
A[12]
4 cyc
A[16]
ロード命令
A[0] A[1] A[2] A[3] A[4] A[5] A[6] A[7] A[8] A[9] A[10] A[11] A[12] A[13] A[14] A[15]

フロントページの続き

(72)発明者 松原 健二
神奈川県秦野市堀山下１番地 株式会社 日立製作所 汎用コンピュータ事業部内
(72)発明者 栗原 俊彦
神奈川県秦野市堀山下１番地 株式会社 日立製作所 汎用コンピュータ事業部内
(72)発明者 位守 弘充
神奈川県秦野市堀山下１番地 株式会社 日立製作所 汎用コンピュータ事業部内

審査官 清木 泰

(56)参考文献 特開平７－９８６７３（ＪＰ，Ａ）
特開昭５４－４７５３０（ＪＰ，Ａ）
特開平５－２５７６８７（ＪＰ，Ａ）
特開平５－２８０２４（ＪＰ，Ａ）
特開平５－１４３４５１（ＪＰ，Ａ）
特開平９－１２８２９４（ＪＰ，Ａ）
特開平９－１２０３７２（ＪＰ，Ａ）
特開平７－１２９４６４（ＪＰ，Ａ）
特開平６－２６６６１６（ＪＰ，Ａ）
特開平６－１２３２７（ＪＰ，Ａ）
高澤嘉光訳，６４ビットRISC Alpha AXP アーキテクチャ概要，日本，共立出版株式会社，１９９３年 ８月１５日，初版１刷，p.176-177,251-252
GERRY KANE, JOE HEINRICH, MIPS RISC ARCHITECTURE, Prentice-Hall,Inc., １９９２年, pages:A-40～A-43
池井満，IA-64プロセッサ基本講座，日本，株式会社オーム社，２０００年 ８月２５日，p.108

(58)調査した分野(Int.Cl.[7]，ＤＢ名)
G06F12/08-12/12
G06F 9/30- 9/355
G06F 9/40- 9/42
G06F 9/44- 9/455