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

特許第4574712号
(P4574712)

(45)発行日　平成22年11月4日(2010.11.4)　(24)登録日　平成22年8月27日(2010.8.27)

(51)Int.Cl.　F I
G06F 12/08 (2006.01)　G06F 12/08 505B
G06F 12/08 509B

請求項の数 9　(全 20 頁)

(21)出願番号　特願2008-502585 (P2008-502585)
(86)(22)出願日　平成18年2月28日(2006.2.28)
(86)国際出願番号　PCT/JP2006/303743
(87)国際公開番号　WO2007/099598
(87)国際公開日　平成19年9月7日(2007.9.7)
審査請求日　平成20年5月2日(2008.5.2)

前置審査

(73)特許権者　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　100074099
弁理士　大菅　義之
(74)代理人　100133570
弁理士　▲徳▼永　民雄
(72)発明者　本藤　幹雄
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

審査官　▲高▼橋　正▲徳▼

最終頁に続く

(54)【発明の名称】演算処理装置、情報処理装置及び制御方法

(57)【特許請求の範囲】
【請求項１】
データを保持する記憶装置に接続される演算処理装置において、
処理を行う演算に用いるデータを要求する演算処理部と、
第１のラインサイズの第１のキャッシュラインを複数有するとともに、前記演算処理部が要求したデータを、前記複数の第１のキャッシュラインのいずれにも保持しない場合、キャッシュミス通知を送信する第１のキャッシュメモリと、
前記第１のラインサイズと異なる第２のラインサイズの第２のキャッシュラインを複数有するとともに、受信したプリフェッチ要求に対応して前記記憶装置から入力するデータを保持する第２のキャッシュメモリと、
前記キャッシュミス通知を受信した場合、前記第２のキャッシュライン毎に、前記第２のラインサイズのデータ量を、前記記憶装置から前記第２のキャッシュメモリにプリフェッチするプリフェッチ要求を前記キャッシュミスが発生したアドレスから前記第２のラインサイズ分先のアドレスに対して発行するプリフェッチ制御部を有することを特徴とする演算処理装置。
【請求項２】
前記演算処理装置において、
前記プリフェッチ制御部は、
前記第２のラインサイズ毎に、前記プリフェッチ要求を１回発行することを特徴とする請求項１記載の演算処理装置。

【請求項３】
前記演算処理装置において、
前記プリフェッチ制御部は、
前記第２のキャッシュメモリにおいて、同一のアドレスから開始される前記第２のラインサイズのデータに対して、前記プリフェッチ要求を複数回発行することにより、前記第２のキャッシュメモリが有するプリフェッチ要求記憶部に前記プリフェッチ要求を複数登録することを特徴とする請求項１記載の演算処理装置。
【請求項４】
前記演算処理装置において、
前記プリフェッチ制御部はさらに、
プリフェッチ要求のデータ量の設定を行うムーブインサイズ設定レジスタと、
前記ムーブイン設定レジスタに基づいて、前記第１のラインサイズ毎に前記プリフェッチ要求を発行するか又は前記第２のラインサイズ毎に前記プリフェッチ要求を発行するかを切替える切換部を有することを特徴とする請求項１～３のいずれか１項に記載の演算処理装置。
【請求項５】
データを保持する記憶装置と、
処理を行う演算に用いるデータを要求する演算処理部と、
第１のラインサイズの第１のキャッシュラインを複数有するとともに、前記演算処理部が要求したデータを、前記複数の第１のキャッシュラインのいずれにも保持しない場合、キャッシュミス通知を送信する第１のキャッシュメモリと、
前記第１のラインサイズと異なる第２のラインサイズの第２のキャッシュラインを複数有するとともに、受信したプリフェッチ要求に対応して前記記憶装置から入力するデータを保持する第２のキャッシュメモリと、
前記キャッシュミス通知を受信した場合、前記第２のキャッシュライン毎に、前記第２のラインサイズのデータ量を、前記記憶装置から前記第２のキャッシュメモリにプリフェッチするプリフェッチ要求を前記キャッシュミスが発生したアドレスから前記第２のラインサイズ分先のアドレスに対して発行するプリフェッチ制御部を有することを特徴とする情報処理装置。
【請求項６】
前記情報処理装置において、
前記プリフェッチ制御部は、
前記第２のラインサイズ毎に、前記プリフェッチ要求を１回発行することを特徴とする請求項５記載の情報処理装置。
【請求項７】
前記情報処理装置において、
前記プリフェッチ制御部は、
前記第２のキャッシュメモリにおいて、同一のアドレスから開始される前記第２のラインサイズのデータに対して、前記プリフェッチ要求を複数回発行することにより、前記第２のキャッシュメモリが有するプリフェッチ要求記憶部に前記プリフェッチ要求を複数登録することを特徴とする請求項５記載の情報処理装置。
【請求項８】
前記情報処理装置において、
前記プリフェッチ制御部はさらに、
プリフェッチ要求のデータ量の設定を行うムーブインサイズ設定レジスタと、
前記ムーブイン設定レジスタに基づいて、前記第１のラインサイズ毎に前記プリフェッチ要求を発行するか又は前記第２のラインサイズ毎に前記プリフェッチ要求を発行するかを切替える切換部を有することを特徴とする請求項５～７のいずれか１項に記載の情報処理装置。
【請求項９】

データを保持する記憶装置に接続される演算処理装置の制御方法において、
前記演算処理装置が有する演算処理部が、処理を行う演算に用いるデータを要求するステップと、
第１のラインサイズの第１のキャッシュラインを複数有する第１のキャッシュメモリが、前記演算処理部が要求したデータを、前記複数の第１のキャッシュラインのいずれにも保持しない場合、キャッシュミス通知を送信するステップと、
前記演算処理装置が有するプリフェッチ制御部が、前記キャッシュミス通知を受信した場合、前記第１のラインサイズと異なる第２のラインサイズの第２のキャッシュライン毎に、前記第２のラインサイズのデータ量を、前記記憶装置から前記第２のキャッシュラインを複数有する第２のキャッシュメモリにプリフェッチするプリフェッチ要求を前記キャッシュミスが発生したアドレスから前記第２のラインサイズ分先のアドレスに対して発行するステップと、
前記第２のキャッシュメモリが、受信したプリフェッチ要求に対応して前記記憶装置から入力するデータを保持するステップとを有することを特徴とする制御方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、お互いにラインサイズが異なる２階層以上のキャッシュメモリを備えるプロセッサに関し、更に詳しくはキャッシュメモリに対するプリフェッチ機能を有するプロセッサについての技術に関する。
【背景技術】
【０００２】
従来からＨＰＣ（Ｈｉｇｈ　Ｐｅｒｆｏｒｍａｎｃｅ　Ｃｏｍｐｕｔｉｎｇ）など、科学技術計算等で用いられるメモリへの連続アクセスを行うコンピュータでは、キャッシュレジスタに対してプリフェッチの技術が適用されている。
【０００３】
プリフェッチは、近い将来必要とされている命令もしくはデータを予め予測してキャッシュメモリ等に読み込んでおく手法で、キャッシュメモリのキャッシュミスを減らすことが出来る。
【０００４】
特許文献１は、プリフェッチ機能を備えたキャッシュシステムについて開示がある。特許文献１のシステムでは、メモリデータへの連続アクセスする場合において、キャッシュミスを起因として、連続アクセスにおいて次にアクセスするラインサイズ先の予測アドレスをキューに登録しておき、実際にアクセスアドレスがキューにヒットし、予測が当たった場合に、連続アクセスであると判断し、ラインサイズ先の次にアクセスするアドレスに対して、プリフェッチを発行している。
【０００５】
複数階層のキャッシュメモリにおいて、上位階層のキャッシュメモリと下位階層のキャッシュメモリとでラインサイズが異なる場合、最下位階層のキャッシュミスによって、ムーブインされるデータサイズは、上位階層のラインサイズから下位階層のラインサイズまでのいずれかのサイズである。そして、ハードウエアプリフェッチが機能する連続アクセスのケースでは、もっともデータサイズの大きい下位階層のラインサイズであるケースがもっとも性能が高くなるため、上記ケースでは、ムーブインされるデータサイズは、下位階層のキャッシュのラインサイズである可能性が高い。
【０００６】
たとえば、Ｃｏｌｕｍｂｕｓ２メモリシステムでは、キャッシュミスによりムーブインされるデータサイズは、メモリアクセスのケースでは下位階層のキャッシュのラインサイズであるが、コピーバックのケースでは上位階層のキャッシュのラインサイズである。
【０００７】
連続アクセスの多いＨＰＣ系ＪＯＢでは、コピーバック率が低いため、上記の連続アク

セスのケースでは、ムーブインされるデータサイズは、下位階層のキャッシュのラインサイズである可能性が高い。

【０００８】

上位階層のキャッシュメモリと下位階層のキャッシュメモリとでラインサイズが異なるキャッシュメモリシステムにおいてプリフェッチを行うと、以下の問題点が生じる。

下位階層のキャッシュミスでムーブインされるデータサイズが、下位階層のキャッシュのラインサイズである場合、上位階層のキャッシュから下位階層のキャッシュに発行されるハードウエアプリフェッチの要求（下位階層のキャッシュへのムーブイン要求）は、下位階層のキャッシュのラインサイズにつき１回でよい。しかし、従来のキャッシュシステムでは上位階層のキャッシュのラインサイズごとに発行してしまい、無駄な下位階層キャッシュアクセスパイプラインを消費することとなる。

【０００９】

下位階層のキャッシュミスでムーブインされるデータサイズが、下位階層のキャッシュのラインサイズである場合、下位階層のキャッシュに発行されるハードウエアプリフェッチの要求は、下位階層のキャッシュのラインサイズにつき１回でよい。しかし、ハードウエアプリフェッチは、実装上の制約により、プリフェッチ要求をロストしてしまうケースが時々あり、ハードウエアプリフェッチがロストした場合には、プリフェッチ要求の発行が１回のみだと、下位階層のキャッシュへのメモリデータのムーブイン要求が発行されなくなってしまう。

【００１０】

下位階層のキャッシュレジスタで生じたキャッシュミスでムーブインされるデータサイズが、下位階層のキャッシュメモリのラインサイズである場合、下位階層のキャッシュメモリに対して発行されるプリフェッチ要求は、下位階層のキャッシュメモリのラインサイズにつき１回でよい。よって、上位階層のキャッシュレジスタでミスしたアドレスに対して、上位階層のキャッシュメモリのラインサイズ分先のアドレスをプリフェッチ要求のプリフェッチアドレスの初期値としてしまうと、下位階層のキャッシュにとって同一ラインである可能性があるため、無駄なプリフェッチ要求で下位階層キャッシュアクセスパイプラインを消費する。

【００１１】

プリフェッチが機能するメモリへの連続アクセスの場合では、下位階層のキャッシュメモリへムーブインされるデータサイズは、下位階層のキャッシュメモリのラインサイズである可能性が高いが、場合によっては下位階層のキャッシュのラインサイズとは異なるデータサイズであることもある。

【特許文献１】特開２００４－３８３４５号公報

【発明の開示】

【００１２】

本発明の課題は、上記問題点を解決したプリフェッチ機能を有するプロセッサを提供することである。

上記課題を解決するため、本発明によるプリフェッチ機能を有するプロセッサは、第１の階層のキャッシュメモリ、第２の階層のキャッシュメモリ、及びプリフェッチ制御部を備える。

【００１３】

第１の階層のキャッシュメモリは、第１のラインサイズを持つ。

第２の階層のキャッシュメモリは、当該第１の階層のキャッシュメモリの下位階層で、前記第１のラインサイズとは異なる大きさの第２のラインサイズを持つ。

【００１４】

プリフェッチ制御部は、前記第２のラインサイズ毎に、前記第１のラインサイズ分のブロックをプリフェッチするように、前記第１の階層のキャッシュメモリから前記第２の階層のキャッシュに対するプリフェッチ要求を発行する。

【００１５】

この構成により、不必要なプリフェッチ要求が発行されるのを防ぐことが出来る。
前記プリフェッチ制御部は、前記第２のラインサイズ毎に１回乃至複数回前記プリフェッチ要求を発行する構成とすることも出来る。

【００１６】

また前記プリフェッチ制御部は、前記第１のラインサイズの２倍以上のブロックをプリフェッチするように、前記プリフェッチ要求を発行する構成とすることも出来る。
この構成により、実装上の制約によって、プリフェッチ要求がロストする場合にも対処することが出来る。

【００１７】

更に前記プリフェッチ制御部は、前記プリフェッチ要求を行うプリフェッチ先のアドレスを、前記第１の階層のキャッシュメモリでミスしたアドレスから前記第２のラインサイズ分先のアドレスとする構成とすることも出来る。

また前記プリフェッチ制御部は、前記第１のラインサイズ毎に前記プリフェッチ要求を発行するのと、前記第２のラインサイズ毎に前記プリフェッチ要求を発行するのとを、ムーブインしたデータの大きさに基づいて切り換える切換部を更に備える構成とすることも出来る。

【００１８】

この構成により、コピーバック等第２のラインの大きさ以外のムーブインにも対処することが出来る。
本発明によれば、第１の階層のキャッシュメモリではなく、第２の階層のキャッシュメモリのラインサイズである第２のラインサイズ毎にプリフェッチ要求が発行されるので、無駄な発行によって、第２の階層のキャッシュメモリのアクセスパイプラインが消費されるのを抑制することが出来る。

【００１９】

また、実装上の制約によって、プリフェッチ要求がロストしても、第２の階層のキャッシュメモリへのメモリデータのムーブイン要求が発行される可能性が高くなるようにすることで、性能向上が図れる。

【００２０】

更に、第１の階層のキャッシュメモリにミスしたアドレスに対して、第１のラインサイズ分先のアドレスではなく、第２の階層のキャッシュメモリのラインサイズ分先のアドレスをハードウエアプリフェッチのプリフェッチアドレスの初期値とすることによって、無駄な要求によって第２の階層のキャッシュアクセスパイプラインが消費されるのを抑制することが出来る。

【００２１】

また最終的にムーブインしたデータサイズに応じて、ハードウエアプリフェッチ要求を発行することによって、第２の階層のキャッシュメモリへムーブインされるデータサイズが、第２の階層のキャッシュメモリのラインサイズと異なる場合でも、必要な要求がもれることなく、正しく要求が発行されるようになる。

【図面の簡単な説明】

【００２２】

【図１】本実施形態におけるコンピュータシステムのプロセッサ及びその周辺構成の概略図である。
【図２】本実施形態におけるプロセッサのメモリ管理部分を中心に描いた図である。
【図３】第１の実施形態のプリフェッチキュー（ＰＦＱ）の構成例を示す図である。
【図４】加算器の出力アドレスと比較器の出力の関係を示す図である。
【図５】手順８、９、１０における各状態を示した図である。
【図６】第１の実施形態のプリフェッチキュー（ＰＦＱ）の動作を示すフローチャートである。
【図７】第２の実施形態のプリフェッチキュー（ＰＦＱ）の構成例を示す図である。

【図８】第３の実施形態のプリフェッチキュー（ＰＦＱ）の構成例を示す図である。
【図９】第４の実施形態のプリフェッチキュー（ＰＦＱ）の構成例を示す図である。
【発明を実施するための最良の形態】
【００２３】
以下に本発明の一実施形態を図面を参照しながら説明する。
図１は本実施形態におけるコンピュータシステムのプロセッサ及びその周辺構成の概略図である。
【００２４】
図１の構成では、プロセッサユニット１、プリフェッチ制御装置２、１次キャッシュ３、２次キャッシュ４及び主記憶装置５を有している。
プロセッサユニット１は、ＡＬＵ、レジスタ等を含み、実際の計算やデータ処理を司る部分である。また同図の構成では、分岐予測等もプロセッサユニット１内で行われ、予測結果に基いたリクエストを１次キャッシュ３に行う。プリフェッチ制御装置２は、プリフェッチ処理の制御全般を受け持つ装置で、プロセッサユニット１から１次キャッシュ３へのリクエストアドレスを監視しながら、２次キャッシュ４にプリフェッチを要求する。１次キャッシュ３は、１次キャッシュシステムで、アクセス速度の早いメモリと１次キャッシュ制御装置から構成されている。２次キャッシュ４は、２次キャッシュシステムで、主記憶装置５よりはアクセス速度が早く１次キャッシュ３より容量の大きなメモリと２次キャッシュ制御装置から構成されている。また本実施形態では、プリフェッチされたデータはこの２次キャッシュ４に保持される。主記憶装置５は、ＤＲＡＭ等によって構成されるメモリである。
【００２５】
プロセッサユニット１が、主記憶装置５上のデータにアクセスする際は、要求アドレスをリクエストアドレス６から指定し、読み出し時にはフェッチデータ７を読み出し、書き込み時にはストアデータ８として１次キャッシュ３に出力する。
【００２６】
１次キャッシュ３は、プロセッサユニット１からの読み出し要求に対して、要求アドレスのデータを自己が保持していれば、そのデータをフェッチデータ７としてプロセッサユニット１に出力し、保持していない場合には、リクエストバス１１からそのデータを含む１ライン分のデータを２次キャッシュ４に対して要求すると共にキャッシュミス９としてプリフェッチ制御装置２に通知する。そして、フェッチデータ１２を受け取ると、プロセッサユニット１に要求されたデータをフェッチデータ７として出力する。また１次キャッシュ３は、自己が保持しているキャッシュデータが更新された場合、適当なタイミングでデータバス１３からそのデータを２次キャッシュ４にライトバックする。
【００２７】
２次キャッシュ４は、１次キャッシュ３からのデータの要求に対して、そのデータを保持していれば、そのデータを含む１ライン分のデータをフェッチデータ７として１次キャッシュ３に出力し、保持していない場合には、リクエストバス１４からそのデータを含む１ライン分のデータを主記憶装置５に対して要求する。そして、フェッチデータ１５を受け取ると、１ライン分のデータを１次キャッシュ３に出力する。また２次キャッシュ４は、１次キャッシュ３と同様、自己が保持しているキャッシュデータが更新されると、適当なタイミングでデータバス１６からそのデータを主記憶装置５にライトバックする。
【００２８】
プロセッサユニット１が１次キャッシュ３に対してデータを要求する時、アドレスバス６でアドレスを指定するが、このアドレス値をプリフェッチ制御装置２は監視し、自己が備えているプリフェッチアドレスキューをこのアドレス値によって検索する。そしてこのアドレスが、プリフェッチアドレスキューに存在するアドレスを先頭とする１ブロック中にある（以下、ヒットするという）場合、プリフェッチアドレスバス１０から２次キャッシュ４にプリフェッチ要求アドレスを出力してプリフェッチ要求を行うと共にアドレスをプリフェッチアドレスキュー２５に登録し、またプリフェッチアドレスキュー内に存在し

ない場合はプリフェッチを要求しない。

【００２９】

なお本実施形態では、１次キャッシュ３と２次キャッシュ４はお互いに異なるラインサイズを持つキャッシュメモリであり、以下の説明では、１次キャッシュ３のラインサイズは６４バイト（以下Ｂと記す）、２次キャッシュ４のラインサイズは２５６Ｂであるとする。

【００３０】

図２は本実施形態におけるプロセッサのメモリ管理部分を中心に描いた図である。

同図において、プロセッサは、メモリ管理用の構成要素として、フェッチポート（ＦＰ）２１、ストアポート（ＳＰ）２２、１次キャッシュアクセスパイプライン２３及び１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４を１次キャッシュ３内に備え、プリフェッチキュー２５をプリフェッチ制御装置２内に備え、２次キャッシュムーブインポート（Ｌ２＄ＭＩＰ）２６、２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）、２次キャッシュアクセスパイプライン２８及び２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＰ）２９を２次キャッシュ４内に備え、システムコントローラムーブインポート（ＳＣＭＩＰ）３０を主記憶装置５内に備えている。

【００３１】

フェッチポート（ＦＰ）２１は、プロセッサユニット１からのｌｏａｄ命令やｓｔｏｒｅ命令等を受け付けるポートである。またストアポート（ＳＰ）２２は、ストアコミットしたｓｔｏｒｅ命令が、キャッシュにデータを書き込むためのポートである。また２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＰ）２９及びシステムコントローラムーブインポート（ＳＣＭＩＰ）３０は、それぞれ２次キャッシュ４及び主記憶装置５に対するムーブイン要求を受け付けるポートである。

【００３２】

１次キャッシュアクセスパイプライン２３及び２次キャッシュアクセスパイプライン２８は、１次キャッシュ３及び２次キャッシュ４に対するアクセス要求を受け付けるパイプラインである。１次キャッシュアクセスパイプライン２３は、Ｐ、Ｔ、Ｍ、Ｂ及びＲの５つのステージを持ち、Ｐステージではアドレスを選択してそのアドレスを転送し、Ｔステージでは転送されたアドレスで１次キャッシュのタグとＴＬＢ（トランスレーションルックアップテーブル）を参照し、ＭステージではＴステージの参照結果として得られたデータの比較（マッチング）を行ない、Ｂステージでは比較結果に基づいて、１次キャッシュのデータを選択して転送し、Ｒステージでは１次キャッシュミスやＴＬＢミスなどに対して転送したデータが有効かあるいは無効かを示すフラグを計算して送る。２次キャッシュアクセスパイプライン２８は、PR1,XP0-14のステージを持ち、各ステージでは、ポートの選択、L2＄タグ検索、アドレス比較、L2＄ミス時にL2$MIBに登録、L2＄ヒット時にL2＄データの読み出し、L2＄データのL1$MIBへの転送などを行っている。

【００３３】

１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４及び２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９は、１次キャッシュ３及び２次キャッシュ４に対して生じたムーブイン命令をバッファリングするものである。

【００３４】

プリフェッチキュー（ＰＦＱ）２５は、以前プリフェッチを行ったアドレスの１ライン分先のアドレスを登録しており、１次キャッシュ３でキャッシュミスが生じると、キャッシュミスが生じたアドレスとプリフェッチキュー（ＰＦＱ）２５内に登録されているアドレスをマッチングし、プリフェッチキュー（ＰＦＱ）２５に一致するアドレスが登録されていれば、２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７に、プリフェッチ要求を発行する。２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７は、プリフェッチキュー（ＰＦＱ）２５からのプリフェッチ要求を受け付けるものである。

【００３５】

同図における動作を以下に説明する。

プロセッサユニット１で、ｌｏａｄ命令等をデコードし、メモリの読み出し要求が発行されると、この要求は、フェッチポート（ＦＰ）２５から１次キャッシュアクセスパイプライン２３に入力される。読む出し要求に対して１次キャッシュ２がヒットすればそのままデータをフェッチポート（ＦＰ）２５から要求を発行したプロセッサユニット１に返して、データをレジスタ３１に書きこむ。

【００３６】

１次キャッシュ２がミスしたときは、２次キャッシュ３からデータを持ってこなければいけないので、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４に要求を入れる。１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４は、２次キャッシュ３に対して読み出し要求を出す。これは２次キャッシュ３のリクエストを受け取る２次キャッシュムーブインポート（Ｌ２＄ＭＩＰ）２６を介して２次キャッシュアクセスパイプライン２８に入る。

【００３７】

そしてこの読み出し要求が２次キャッシュ３でヒットすれば、そのデータを１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４に入れ、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４は１次キャッシュラインアクセスパイプラインを獲得して１次キャッシュキャッシュ２にデータを書きにゆく（１次キャッシュミス２次キャッシュヒットの場合）。

【００３８】

次にハードウエアプリフェッチを行う場合について説明する。

１次キャッシュ２でミスして、プリフェッチキュー（ＰＦＱ）２５にハードウエアプリフェッチとして動作すべきアドレスが登録されていない場合、そのアドレスを一旦プリフェッチキュー（ＰＦＱ）２７に登録する。このとき特許文献１に示してあるように、６４Ｂ分先のアドレスを登録する、次に６４バイト先のアクセスしに行ったときには１次キャッシュがミスすると同時にプリフェッチキュー（ＰＦＱ）２５はヒットする。このときプリフェッチキュー（ＰＦＱ）２５は、更に６４Ｂを足して＋１２８Ｂのアドレスのプリフェッチのリクエストをプリフェッチポート（Ｌ２＄ＰＦＰ）２７に出す。

【００３９】

１次キャッシュミスは、２次キャッシュムーブインポート（Ｌ２＄ＭＩＰ）２６と２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７に登録され、２次キャッシュにアクセスしてヒットすればデータを返す。また、ミスすれば２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９に登録してシステムコントローラムーブインポート（ＳＣＭＩＰ）３０に出力して主記憶装置５にリクエストを出す。そして主記憶装置５データが帰ってきたら、それを２次キャッシュアクセスパイプライン２８を介して２次キャッシュ３に書きこみ、同時にバイパスで１次キャッシュアクセスパイプライン２３に返しこれを１次キャッシュ２に書き込む。

【００４０】

図３は、第１の実施形態のプリフェッチキュー（ＰＦＱ）２５の構成例を示す図である。

同図のプリフェッチ３、選択回路４４、加算器４５、選択回路４６及び４７、及び加算器４８を備え、各エントリ４１－１～４１－ｎはそれぞれ、エントリ４１に登録されるアドレス値等がセットされるレジスタ４９、レジスタ４９内のアドレスとリクエストアドレスを比較する比較器５０及び比較器５０の比較結果と後述するレジスタ４９内の有効ビットとのＡＮＤを求めるＡＮＤ回路５１を有している。

【００４１】

レジスタ４９は、アドレス値の他に状態フラグとして働く有効ビット、待機ビット及びＬ２＄ＰＦＰ登録許可フラグを記録している。

レジスタ４９内の有効ビットは、レジスタ４８にセットされているアドレス値が有効かどうかを示すもので、アドレス値がレジスタ４８に登録される時セットされ、このエントリ４１からアドレス値が読み出された時にリセットされる。待機ビットは、有効ビットが

セットされているエントリ４１において、１次キャッシュアクセスパイプライン２３からのリクエストアドレスがレジスタ４８内に登録されているアドレス値にマッチした場合セットされる。プリフェッチアドレスキュー（ＰＦＱ）２５は、この待機ビットの状態から読み出しを行うエントリ４１－１～４１－ｎを決定する。Ｌ２＄ＰＦＰ登録許可フラグは、１次キャッシュアクセスパイプライン２３からのリクエストアドレスと、このエントリ４１に登録されているアドレスがマッチ（ヒット）したときに次の２５６Ｂの連続アドレスを２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７に登録するかどうかの判断に用いられるもので、Ｌ２＄ＰＦＰ登録許可フラグに１がセットされていれば２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）に登録を行ない、０がセットされていれば２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）に登録を行わない。

【００４２】

１次キャッシュアクセスパイプライン２３からリクエストアドレスが入力され、これが新規登録される場合、有効ビットには１、待機ビットには０、Ｌ２＄ＰＦＰ登録許可フラグには１、及びアドレスにはリクエストアドレスに加算器４５で６４を加えた値が入力される。

【００４３】

またレジスタ４９の登録アドレスが更新される場合には、Ｌ２＄ＰＦＰ登録許可フラグには、比較器４２による加算器４５の出力のビット［７：６］が０かどうかの比較結果と新規登録かどうかのＯＲをＯＲ回路４３で求めた結果が入力される。このＬ２＄ＰＦＰ登録許可フラグの内容は、選択回路４６及び４７の選択信号となっており、Ｌ２＄ＰＦＰ登録許可フラグが１のときＰＦＰリクエスト信号として１が出力され、またＰＦＰリクエストアドレスとしてそのエントリ４１のレジスタ４９に登録されているレジスタ値に加算器４８によって２５６Ｂ加算した値が出力される。そしてこれらの出力によって、２次キャッシュＰＦＰ（Ｌ２＄ＰＦＰ）２７にＰＦＰリクエストアドレスが登録される。

【００４４】

図４は、加算器４５の出力アドレスと比較器４２の出力の関係を示す図である。

加算器４５によって、レジスタ４９にセットされているアドレス値を６４Ｂインクリメントしてレジスタ４９に登録すると、その出力アドレスのビット［７：６］は、４回に一回０となり、よって比較器４２からは４回に１回１が出力され、これがＯＲ回路４３を介してＬ２＄ＰＦＰ登録許可フラグにセットされる。なおレジスタ４９への登録が新規登録の場合、Ｌ２＄ＰＦＰ登録許可フラグには１がセットされる。よって、アドレスの新規登録からアドレスが４回更新されるごとに１回Ｌ２＄ＰＦＰ登録許可フラグに１がセットされ、２次キャッシュＰＦＰ（Ｌ２＄ＰＦＰ）２７にＰＦＰリクエストアドレスが登録される。

【００４５】

このＬ２＄ＰＦＰ登録許可フラグは、プリフェッチキュー（ＰＦＱ）２５に新規登録を行うとき及び２５６Ｂ境界の先頭６４Ｂアドレスをプリフェッチキュー（ＰＦＱ）２５に更新登録するときにセットされる。また２５６Ｂ境界の先頭６４Ｂ以外のアドレスをPFQに更新登録するときリセットされる。

【００４６】

このプリフェッチキュー（ＰＦＱ）２５は、１次キャッシュアクセスパイプライン２３からプリフェッチのリクエストアドレスが入力されると、このアドレス値はレジスタ４９内のアドレス値と比較器５０によって比較され、この比較結果と有効ビットとのＡＮＤをＡＮＤ回路５１で取り、結果をＰＦＱにヒットしたかどうかを示すＰＦＱヒット信号として１次キャッシュアクセスパイプライン２３に出力する。したがって、リクエストアドレスとレジスタ４９が一致し、且つ有効ビットが１のときＰＦＱヒット信号が１になる。

【００４７】

また１次キャッシュ２のラインの大きさが６４Ｂに対して２次キャッシュ３のラインの大きさが２５６Ｂと、上位と下位のキャッシュメモリのラインサイズが異なるときであっても、２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７へのアドレス値の登録は

４回に１回（２５６Ｂ／６４Ｂ）にすることが出来、プリフェッチ要求は２次キャッシュ３のラインサイズに付き１回となる。よって無駄な下位階層キャッシュアクセスパイプラインの消費を抑止し、性能向上を図ることが出来る。

【００４８】

次に、図２、図３を用い、プリフェッチ動作を含む、プロセッサのメモリアクセス命令に対する処理の詳細手順を以下に示す。

以下の説明では、アドレスＡ、Ａ＋８、Ａ＋１６、．．．、Ａ＋５６に対するｌｏａｄ命令がプロセッサユニットでデコードされた場合を例として示す。

１：ｌｏａｄ命令が、フェッチポート（ＦＰ）２１を介して１次キャッシュアクセスパイプライン２３を獲得。

２：１次キャッシュアクセスパイプライン２３において、１次キャッシュに対してアドレスＡでアクセス。

３：１次キャッシュアクセスパイプライン２３において、２の結果、１次キャッシュミスを検出。

４：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４にミスアドレスを登録。

【００４９】

４．１：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４は、２次キャッシュ３から１次キャッシュ２へのムーブイン要求を、２次キャッシュムーブインポート（Ｌ２＄ＭＩＰ）２６に対して発行。

【００５０】

４．２：２次キャッシュムーブインポート（Ｌ２＄ＭＩＰ）２６は、２次キャッシュアクセスパイプライン２８を獲得し、２次キャッシュに対してアドレスＡでアクセス。

４．３：２次キャッシュアクセスパイプライン２８において、手順４．２の結果、２次キャッシュミスを検出。

【００５１】

４．４：２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９にミスアドレスを登録。

４．５：２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９は、主記憶装置５から２次キャッシュ３へのムーブイン要求を、システムコントローラムーブインポート（ＳＣＭＩＰ）３０に対して発行。

【００５２】

４．６：システムコントローラムーブインポート（ＳＣＭＩＰ）３０は、主記憶装置５からミスアドレスＡから２５６Ｂ分のデータを取り出し、２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９にムーブイン。

【００５３】

４．７：２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９は、２次キャッシュアクセスパイプライン２８を獲得し、２次キャッシュ４に２５６Ｂのムーブインデータを書き込む。

【００５４】

４．８：２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９は、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４に６４Ｂのムーブインデータをバイパス転送。

４．９：アドレスＡで１次キャッシュミスしたｌｏａｄ命令は、１次キャッシュアクセスパイプライン２３を獲得し、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４に転送されたムーブインデータをバイパスして読み出し、プロセッサユニット1内のレジスタ３１にデータを書き込む。

【００５５】

４．１０：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４は、１次キャッシュアクセスパイプライン２３を獲得し、１次キャッシュ２に６４Ｂのムーブインデータを書き込む。

５：プリフェッチキュー（ＰＦＱ）２５がミスを検出。

６：次の連続アドレス（Ａ＋６４）をプリフェッチキュー（ＰＦＱ）２５に新規登録。レジスタ４９内のＬ２＄ＰＦＰ登録許可フラグをセット。
７：同様に、連続アドレス（Ａ＋８、Ａ＋１６、．．．、Ａ＋５６）にアクセスするｌｏａｄ命令が、１次キャッシュアクセスパイプライン２３を獲得。
８：このとき２次キャッシュ３からのムーブインデータが到着していなければ、１次キャッシュＭＩＢヒット、データミスを検出し、１次キャッシュアクセスパイプライン２３をアボート。アボートされた要求は、フェッチポート（ＦＰ）２１に戻る。
９：また２次キャッシュ３からのムーブインデータが到着していて、１次キャッシュ２にデータが書き込まれていなければ、１次キャッシュＭＩＢヒット、データヒットを検出し、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４のデータをバイパスして読み出し、レジスタ３１にデータを書き込む。
１０：また２次キャッシュ３からのムーブインデータが到着していて、１次キャッシュにデータが書き込まれていれば、１次キャッシュヒットを検出し、１次キャッシュ２からデータを読み出し、レジスタ３１にデータを書き込む。

【００５６】

図５は、この手順８、９、１０における各状態を示したものである。
手順８の状態では、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４からアドレスは到着しているがデータは到着しておらず、また１次キャッシュ２にはデータは書き込まれていないので、１次キャッシュアクセスパイプライン２３をアボートする。

【００５７】

また手順９の状態では、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４からアドレス及びデータが到着しているが、１次キャッシュ２にはデータは書き込まれていないので、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４からデータを読み出してレジスタ３１にデータを書き込む。

【００５８】

また手順１０の状態では、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４からアドレス及びデータが到着しており、また１次キャッシュ２にはデータが書き込まれているので、１次キャッシュ２からデータを読み出してレジスタ３１にデータを書き込む。

【００５９】

以下に続けて連続アドレス（Ａ＋６４）にアクセスするｌｏａｄ命令についての処理を説明する。
１１：手順１と同様に、連続アドレス（Ａ＋６４）にアクセスするｌｏａｄ命令が、１次キャッシュアクセスパイプライン２３を獲得。
１２：手順１１の結果、１次キャッシュミスを検出。

【００６０】

１２．１：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４にミスアドレスを登録し、２次キャッシュ３にアクセス。
１２．２：２次キャッシュヒットを検出し、２次キャッシュから６４Ｂのデータを読み出し、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４に転送。

【００６１】

１２．３：アドレス（Ａ＋６４）で、１次キャッシュミスしたｌｏａｄ命令は、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４のデータをバイパスして読み出し、レジスタ３１にデータを書き込む。
１２．４：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４は、１次キャッシュ２に６４Ｂデータを書き込む。
１３：プリフェッチキュー（ＰＦＱ）２５のヒットを検出。レジスタ４９内の待機ビットをセット。
１４：次の連続アドレス（Ａ＋１２８）をプリフェッチキュー（ＰＦＱ）２５に登録。レジスタ４９内のＬ２＄ＰＦＰ登録許可フラグをリセット。
１５：手順１４でリセットされるまでＬ２＄ＰＦＰ登録許可フラグがセットされていたの

で、次の２５６Ｂ連続アドレス（Ａ＋６４＋２５６）を２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７に登録。

【００６２】

１５．１：２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７は、２次キャッシュアクセスパイプライン２８を獲得し、２次キャッシュ３に対して、アドレス（Ａ＋６４＋２５６）でアクセス。

【００６３】

１５．２：手順１５．１の結果、２次キャッシュミスを検出。

１５．３：２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９にキャッシュミスアドレスを登録。

【００６４】

１５．４：２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９は、主記憶装置５から２次キャッシュ３へのムーブイン要求を、システムコントローラムーブインポート（ＳＣＭＩＰ）３０に発行。

【００６５】

１５．５：システムコントローラムーブインポート（ＳＣＭＩＰ）３０は、主記憶装置５からミスアドレス（Ａ＋６４＋２５６）から２５６Ｂ分のデータを取り出し、それを２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９にムーブイン。

【００６６】

１５．６：２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９は、２次キャッシュアクセスパイプライン２８を獲得し、２次キャッシュ３に２５６Ｂのムーブインデータを書き込む。

１６：同様に、連続アドレス（Ａ＋６４＋８、Ａ＋６４＋１６、．．．、Ａ＋６４＋５６）にアクセスするｌｏａｄ命令が、１次キャッシュアクセスパイプライン２３を獲得。

１７：２次キャッシュ４からのムーブインデータが到着していなければ、１次キャッシュＭＩＢヒット、データミスを検出し１次キャッシュアクセスパイプライン２３をアボート。アボートした要求は、フェッチポート（ＦＰ）２１に戻る。

１８：２次キャッシュ３からのムーブインデータが到着していて、１次キャッシュ２にデータが書き込まれていなければ、１次キャッシュＭＩＢヒット、データヒットを検出し、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４のデータをバイパスして読み出し、それをレジスタ３１に書き込む。

１９：２次キャッシュ３からのムーブインデータが到着していて、１次キャッシュ２にデータが書き込まれていれば、１次キャッシュヒットを検出し、１次キャッシュ２からデータを読み出し、それをレジスタ４１に書き込む。

【００６７】

以下に続けて連続アドレス（Ａ＋１２８）にアクセスするｌｏａｄ命令についての処理を説明する。

２０：手順１、１１と同様に、連続アドレス（Ａ＋１２８）にアクセスするｌｏａｄ命令が、１次キャッシュアクセスパイプライン２３を獲得。

２１：手順２０の結果、１次キャッシュミスを検出。

【００６８】

２１．１：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４にキャッシュミスアドレスを登録し、２次キャッシュ３にアクセス。

２１．２：手順２１．１の結果、２次キャッシュヒットを検出し、２次キャッシュ３から６４Ｂのデータを読み出し、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４に転送。

【００６９】

２１．３：アドレス（Ａ＋１２８）で１次キャッシュミスしたｌｏａｄ命令は、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４からデータをバイパスして読み出し、これをレジスタ３１に書き込む。

【００７０】
２１．４：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４は、１次キャッシュ２に６４Ｂデータを書き込む。
２２：プリフェッチキュー（ＰＦＱ）２５のヒットを検出。レジスタ４９の待機ビットをセット。
２３：次の連続アドレス（Ａ＋１９２）をプリフェッチキュー（ＰＦＱ）２５に登録。Ｌ２＄ＰＦＰ登録許可フラグをリセット。
２４：（手順２３でリセットされるまでＬ２＄ＰＦＰ登録許可フラグがセットされていたので、次の２５６Ｂ連続アドレス（Ａ＋１２８＋２５６）は、２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７に登録しない。）
２５：同様に、連続アドレス（Ａ＋１２８＋８、Ａ＋１２８＋１６、．．．、Ａ＋１２８＋５６）にアクセスするｌｏａｄ命令が、１次キャッシュアクセスパイプライン２３を獲得。
２６：２次キャッシュ３からのムーブインデータが到着していなければ、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４、データミスを検出し１次キャッシュアクセスパイプライン２３をアボート。アボートした要求は、フェッチポート（ＦＰ）２１に戻る。
２７：２次キャッシュ３からのムーブインデータが到着していて、１次キャッシュにデータが書き込まれていなければ、１次キャッシュＭＩＢヒット、データヒットを検出し、１次キャッシュＭＩＢのデータをバイパスして読み出し、レジスタ３１にデータを書き込む。
２８：２次キャッシュ３からのムーブインデータが到着していて、１次キャッシュ２にデータが書き込まれていれば、１次キャッシュヒットを検出し、１次キャッシュ２からデータを読み出し、レジスタ３１にデータを書き込む。
【００７１】
以下に続けて連続アドレス（Ａ＋１９２）にアクセスするｌｏａｄ命令についての処理を説明する
２９：手順１、１１，２０と同様に、連続アドレス（Ａ＋１９２）にアクセスするｌｏａｄ命令が、１次キャッシュアクセスパイプラインを獲得。
３０：手順２９の結果、１次キャッシュミスを検出。
【００７２】
３０．１：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４にキャッシュミスアドレスを登録し、２次キャッシュ３にアクセス。
３０．２：２次キャッシュヒットを検出し、２次キャッシュから６４Ｂのデータを読み出し、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４に転送。
【００７３】
３０．３：アドレス（Ａ＋１９２）で、１次キャッシュミスしたｌｏａｄ命令は、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４からデータをバイパスして読み出し、これをレジスタ３１を書き込む。
【００７４】
３０．４：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４は、１次キャッシュ２に６４Ｂデータを書き込む。
３１：プリフェッチキュー（ＰＦＱ）２５のヒットを検出。レジスタ４９の待機ビットをセット。
３２：次の連続アドレス（Ａ＋２５６）をプリフェッチキュー（ＰＦＱ）２５に登録。レジスタ４９のＬ２＄ＰＦＰ登録許可フラグをセット。
３３：（手順３２でセットされるまでＬ２＄ＰＦＰ登録許可フラグがリセットされていたので、次の２５６Ｂ連続アドレス（Ａ＋１９２＋２５６）は、２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７に登録しない。）
３４：同様に、連続アドレス（Ａ＋１９２＋８、Ａ＋１９２＋１６、．．．、Ａ＋１９２

＋５６）にアクセスするｌｏａｄ命令が、１次キャッシュアクセスパイプライン２３を獲得。
３５：２次キャッシュ３からのムーブインデータが到着していなければ、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４ヒット、データミスを検出し１次キャッシュアクセスパイプライン２３をアボート。アボートされた要求はフェッチポート（ＦＰ）２１に戻る。
３６：２次キャッシュ３からのムーブインデータが到着していて、１次キャッシュ２にデータが書き込まれていなければ、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４ヒット、データヒットを検出し、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４からデータをバイパスして読み出し、これをレジスタ３１に書き込む。
３７：２次キャッシュ３からのムーブインデータが到着していて、１次キャッシュ２にデータが書き込まれていれば、１次キャッシュヒットを検出し、１次キャッシュ２からデータを読み出し、これをレジスタ３１に書き込む。

【００７５】

以下に続けて連続アドレス（Ａ＋２５６）にアクセスするｌｏａｄ命令についての処理を説明する。
３８：同様に、連続アドレス（Ａ＋２５６）にアクセスするｌｏａｄ命令が、１次キャッシュアクセスパイプライン２３を獲得。
３９：手順３８の結果１次キャッシュ２ミスを検出。

【００７６】

３９．１：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４にミスアドレスを登録し、２次キャッシュ３にアクセス。
３９．２：２次キャッシュヒットを検出し、２次キャッシュ３から６４Ｂのデータを読み出し、これを１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４に転送。

【００７７】

３９．３：アドレス（Ａ＋２５６）で、１次キャッシュミスしたｌｏａｄ命令は、１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４からデータをバイパスして読み出し、これをレジスタ３１に書き込む。

【００７８】

３９．４：１次キャッシュＭＩＢは、１次キャッシュに６４Ｂデータを書き込む。
４０：プリフェッチキュー（ＰＦＱ）２５ヒットを検出。レジスタ４９内の待機ビットをセット。
４１：次の連続アドレス（Ａ＋３２０）をプリフェッチキュー（ＰＦＱ）２５に登録。レジスタ４９内のＬ２＄ＰＦＰ登録許可フラグをリセット。
４２：手順４１でリセットされるまでＬ２＄ＰＦＰ登録許可フラグがセットされていたので、次の２５６Ｂ連続アドレス（Ａ＋２５６＋２５６）を２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７に登録。

【００７９】

４２．１：２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７は、２次キャッシュアクセスパイプライン２８を獲得し、２次キャッシュ３に対してアドレス（Ａ＋２５６＋２５６）でアクセス。

【００８０】

４２．２：手順４２．１の結果、２次キャッシュミスを検出。
４２．３：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４にミスアドレスを登録。

【００８１】

４２．４：１次キャッシュムーブインバッファ（Ｌ１＄ＭＩＢ）２４は、主記憶装置５から２次キャッシュ３へのムーブイン要求を、システムコントローラムーブインポート（ＳＣＭＩＰ）３０に対して発行。

【００８２】

４２．５：システムコントローラムーブインポート（ＳＣＭＩＰ）３０は、主記憶装置５からキャッシュミスアドレス（Ａ＋２５６＋２５６）から２５６Ｂ分のデータを取り出し、このデータを２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９にムーブイン。

【００８３】

４２．６：２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９は、２次キャッシュアクセスパイプライン２８を獲得し、２次キャッシュ３に２５６Ｂのムーブインデータを書き込む。

【００８４】

以下、連続アドレス（Ａ＋３２０）、（Ａ＋３８４）、・・・にアクセスするｌｏａｄ命令について同様の処理を繰り返す。

図６は、図３に示した第１の実施形態のプリフェッチキュー（ＰＦＱ）２５の動作を示すフローチャートである。

【００８５】

ステップＳ１において、アドレスＡにて１次キャッシュ２にアクセスし、１次キャッシュ２がキャッシュミスし（ステップＳ２、Ｙ）、且つプリフェッチキュー（ＰＦＱ）２５もミスしたら（ステップＳ３、Ｙ）、ステップＳ４としてプリフェッチキュー（ＰＦＱ）２５に１次キャッシュ２の１ライン分先のアドレス（Ａ＋６４）を登録し、またプリフェッチキュー（ＰＦＱ）２５内のレジスタ４９のＬ２＄登録許可フラグをセットして、処理をステップＳ１に戻す。

【００８６】

またステップＳ２において、１次キャッシュ２がヒットし（ステップＳ２、Ｎ）、プリフェッチキューがミスしたとき（ステップＳ３、Ｙ）、処理をステップＳ１に戻す。

ステップＳ２において、１次キャッシュ２がミスし（ステップＳ２、Ｙ）、プリフェッチキュー（ＰＦＱ）２５はヒットしたとき（ステップＳ３、Ｎ）、及び１次キャッシュ２がヒットし（ステップＳ２、Ｎ）、プリフェッチキュー（ＰＦＱ）２５はヒットしたとき（ステップＳ３、Ｎ）、処理をステップＳ６に移し、プリフェッチを行う。

【００８７】

ステップＳ６では、プリフェッチキュー（ＰＦＱ）２５のレジスタ４９内の待機ビットをセットする。そしてレジスタ４９内のＬ２＄登録許可フラグがセットされていたら（ステップＳ７、Ｙ）、２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７にＰＦＰリクエストアドレス（Ａ＋６４）でプリフェッチう要求を登録し、Ｌ２＄登録許可フラグがセットされていなければ（ステップＳ７、Ｎ）、２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７にリクエストを登録しない。

【００８８】

次にステップＳ９としてプリフェッチキュー（ＰＦＱ）２５の登録アドレスをＡ＋６４に更新した後、（Ａ／６４＋１）を４で割った余りが０ならばステップＳ１１としてレジスタ４９内のＬ２＄登録許可フラグをセットして、処理をステップＳ１に戻す。またＡ／６４＋１）を４で割った余りが０でないのならばステップＳ１２としてＬ２＄登録許可フラグをリセットして、処理をステップＳ１に戻す。

【００８９】

このように第１の実施形態では、下位層のキャッシュレジスタのラインの大きさ毎にプリフェッチ要求を発行することが出来るので、無駄なプリフェッチ要求がアクセスパイプラインを占めることが無く、性能向上を図ることが出来る。

【００９０】

次に、プリフェッチキュー（ＰＦＱ）２５の第２の構成例について説明する。

第１の実施形態のプリフェッチキュー（ＰＦＱ）２５が、下位階層のキャッシュの１ラインの大きさが上位階層のキャッシュの１ラインの大きさのｎ倍であったとき、ｎ回の連続アクセスに対して１回２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７に対してプリフェッチ要求を登録していたが、第２の実施形態のプリフェッチキュー（ＰＦＱ）

２５は、ｎ回に２回以上プリフェッチ要求を登録する。

【００９１】

下位階層のキャッシュのミスでムーブインされるデータサイズが、下位階層のキャッシュのラインサイズである場合、下位階層のキャッシュに発行されるハードウエアプリフェッチの要求は、第１の実施形態のプリフェッチキュー（ＰＦＱ）２５のように下位階層のキャッシュのラインサイズにつき１回でよい。

【００９２】

しかし、プリフェッチ要求は、ハードウエア実装上の制約によりロストしてしまうケースが時々あるため、プリフェッチ要求の発行が１回のみだと、ハードウエアプリフェッチがロストした場合に、下位階層のキャッシュへのメモリデータのムーブイン要求が発行されなくなってしまう。この実装上の制約とは、例えば２次キャッシュ３がミスすると２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９に登録されるが、２次キャッシュムーブインバッファ（Ｌ２＄ＭＩＢ）２９が一杯のときは、再登録を行わずにプリフェッチ要求がロストしてしまうことがある。

【００９３】

この点に対処し、第２の実施形態のプリフェッチキュー（ＰＦＱ）２５は、下位階層のキャッシュのラインサイズにつき複数回プリフェッチ要求を２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７に発行する。

【００９４】

図７は、第２の実施形態のプリフェッチキュー（ＰＦＱ）２５の構成例を示す図である。なお同図は、図３の第１の実施形態のプリフェッチキュー（ＰＦＱ）２５と対比する形で記載されている。

【００９５】

図７を図３の第１の実施形態のプリフェッチキュー（ＰＦＱ）２５と比較すると、比較器４２ａの入力が、加算器４５ａから出力されるアドレスのうちビット［６］のみになっている。よって第１の実施形態ではレジスタ４９のアドレスが４回更新されると１回Ｌ２＄ＰＦＰ登録許可フラグが１にセットされているが、第２の実施形態のプリフェッチキュー（ＰＦＱ）２５では、アドレスが４回更新されると２回Ｌ２＄ＰＦＰ登録許可フラグに１がセットされ、２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７にリクエストが登録される。

【００９６】

これにより第２の実施形態では、１つのプリフェッチ要求が、ハードウエアの実装上の問題でロストしても、対処することが出来る。

次に、プリフェッチキュー（ＰＦＱ）２５の第３の実施形態について説明する。

【００９７】

第３の実施形態のプリフェッチキュー（ＰＦＱ）２５も、第２の実施形態と同様、ハードウエアの実装上の問題で、プリフェッチ要求がロストする場合に対処したものである。

第３の実施形態では、上位階層のキャッシュのラインサイズの２倍以上のブロックをプリフェッチするように２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７にプリフェッチ要求を発行する。これによりプリフェッチ要求は、２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７で２倍以上に展開され、複数のプリフェッチ要求が発行される。

【００９８】

図８は、第３の実施形態のプリフェッチキュー（ＰＦＱ）２５の構成例を示す図である。なお同図も、図３の第１の実施形態のプリフェッチキュー（ＰＦＱ）２５と対比する形で記載されている。

【００９９】

図８の第３の実施形態のプリフェッチキュー（ＰＦＱ）２５の構成を図３の第１の実施形態の構成と比較すると、２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７に出力されるＰＦＰリクエストブロックサイズ６１が１２８Ｂと、１次キャッシュ２のライン

の２倍になっている。なお不図示であるが、図３の第１の実施形態ではこのＰＦＰリクエストブロックサイズは、１次キャッシュ２のラインサイズと同じ６４Ｂとなっている。

【０１００】

この構成により第３の実施形態のプリフェッチキュー（ＰＦＱ）２５では、１次キャッシュ２のラインサイズの２倍のサイズのブロックサイズを指定して、プリフェッチ要求を２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７に対して発行するので、２次キャッシュプリフェッチポート（Ｌ２＄ＰＦＰ）２７では２回のプリフェッチ要求が発行されることになる。

【０１０１】

これにより第３の実施形態でも、１つのプリフェッチ要求が、ハードウエアの実装上の問題でロストしても、対処することが出来る。

次に、プリフェッチキュー（ＰＦＱ）２５の第４の実施形態について説明する。

【０１０２】

第４の実施形態のプリフェッチキュー（ＰＦＱ）２５は、本実施形態で行われている下位層キャッシュのラインサイズ（２５６Ｂ）毎のムーブインと、従来のプロセッサで行われている上位層キャッシュのラインサイズ（６４Ｂ）毎のムーブインを切り換えることが出来るようしたものである。

【０１０３】

これにより、コピーバックの際に実行される上位層キャッシュのラインサイズ（６４Ｂ）毎のムーブインにも対処することが出来る。

図９は、第４の実施形態のプリフェッチキュー（ＰＦＱ）２５の構成例を示す図である。なお同図も、図３の第１の実施形態のプリフェッチキュー（ＰＦＱ）２５と対比する形で記載されている。

【０１０４】

同図の第４の実施形態のプリフェッチキュー（ＰＦＱ）２５と図３の第１の実施形態を比較すると、図９の構成では、レジスタ４９ｂにムーブイン（ＭＩ）データサイズが記憶されている。このＭＩデータサイズは、０がセットされるとプリフェッチキュー（ＰＦＱ）２５は２５６Ｂ毎のムーブインを行ない、１がセットされると６４Ｂ毎のムーブインを行う。

【０１０５】

このＭＩデータサイズには、初期値として０がセットされ、キャッシュミスしたムーブインアドレスとレジスタ４９ｂにセットされているアドレスを比較器７１で比較した結果、両者が一致し、且つムーブインデータサイズが６４Ｂであったとき、ＡＮＤ回路７２の出力によって１がセットされる。またこのＭＩデータサイズは、ＯＲ回路７３によってＬ２＄ＰＦＰ登録許可フラグとのＯＲが取られた結果が選択回路４６ｂに入力される。よって、ＭＩデータサイズに０がセットされているときは２５６Ｂのムーブインを行ない、１がセットされているときは６４Ｂのムーブインを行なう、というようにムーブインの大きさを切り換えることができる。

【０１０６】

以上のように本実施形態によれば、上位層のキャッシュレジスタと下位層のキャッシュレジスタのラインの大きさが異なっても、下位層のキャッシュレジスタのラインの大きさ毎にプリフェッチ要求を発行することが出来るので、無駄なプリフェッチ要求でキャッシュアクセスパイプラインが消費されるのを抑止し、性能向上を図ることが出来る。

【０１０７】

また実装上の制約により、プリフェッチ要求がロストしてしまう点にも対処することが出来る。

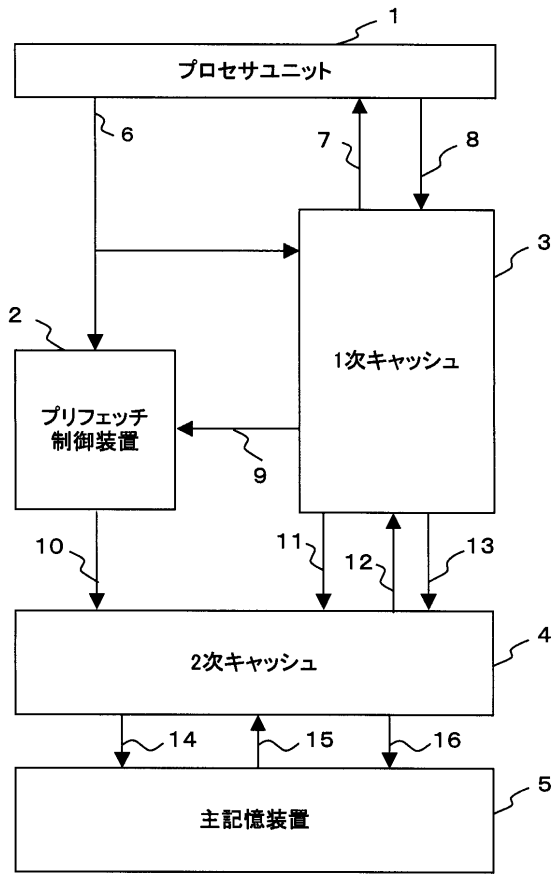
更には、コピーバックの際に実行される上位層キャッシュのラインサイズのムーブインにも対処することが出来る。

【０１０８】

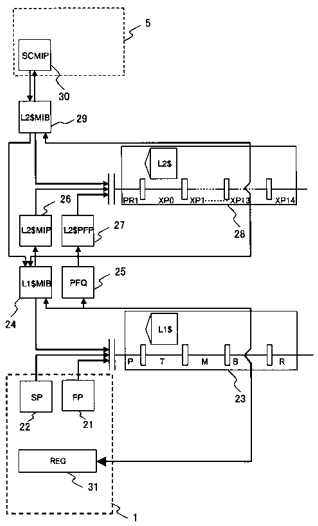
なお上記例では、本発明を２次キャッシュメモリと主記憶装置間のプリフェッチに適用

した場合を例として示したが、本発明はこれに限定されるものではなく、システムが３次キャッシュ以上のキャッシュメモリを備えている場合、２次キャッシュと３次キャッシュの間、３次キャッシュと主記憶装置の間等にも適用することが出来る。
また、上記例では、プリフェッチの連続アクセス方向が、昇順であるケースについて適用した場合を例として示したが、本発明はこれに限定されるものではなく、プリフェッチの連続アクセス方向が、降順であるケースについても、適用することができる。

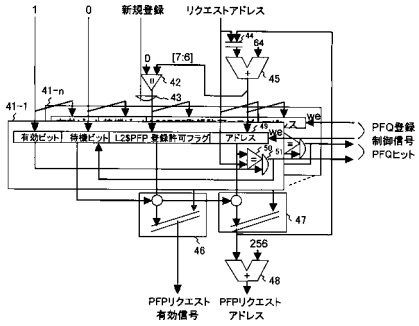
【図１】



【図２】



【図３】



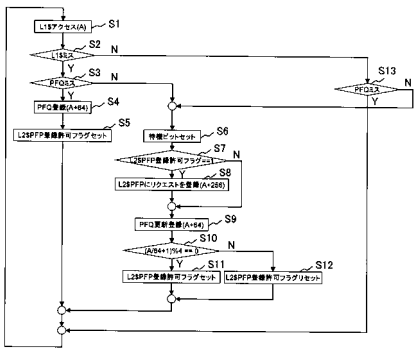
【図４】

比較器 42	7 6	5 4	3 2 1 0	
0	0 1	0 0	0 0 0 0	+64
0	1 0	0 0	0 0 0 0	+128
0	1 1	0 0	0 0 0 0	+192
1	0 0	0 0	0 0 0 0	+256

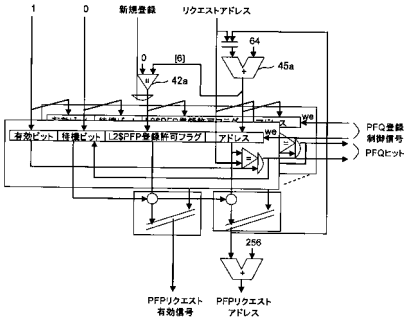
【図５】

	L1$ MIB Address	L1$ MIB Data	L1$ data	処理
手順 8	○	×	×	アボート
手順 9	○	○	×	L1$MIBから
手順 10	○	○	○	L1$ 直接

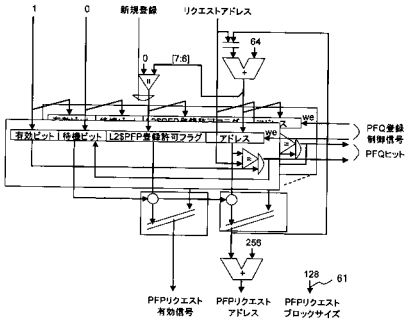
【図６】



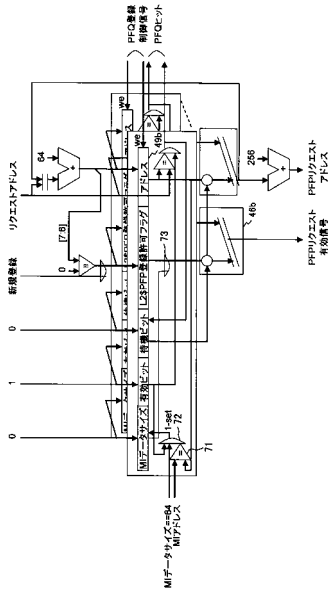
【図７】



【図８】



【図９】



フロントページの続き

(56)参考文献 特開平０２－３０１８４３（ＪＰ，Ａ）
特開平０９－１２８２９３（ＪＰ，Ａ）
特開２００４－０３８３４５（ＪＰ，Ａ）
特開２００６－０４０１４１（ＪＰ，Ａ）
特開昭６１－１６９９４９（ＪＰ，Ａ）
特開平０６－１４９６６９（ＪＰ，Ａ）
特開平０７－１２９４６４（ＪＰ，Ａ）
特開２０００－３３９１５７（ＪＰ，Ａ）
特開２００６－０４８１８１（ＪＰ，Ａ）
特開２００６－０４８１８２（ＪＰ，Ａ）

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