

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4904802号
(P4904802)

(45) 発行日 平成24年3月28日 (2012.3.28)

(24) 登録日 平成24年1月20日 (2012.1.20)

(51) Int.Cl.	F I	
G06F 12/08 (2006.01)	G06F 12/08	509F
G06F 9/46 (2006.01)	G06F 12/08	531B
	G06F 12/08	503C
	G06F 12/08	503F
	G06F 12/08	511E

請求項の数 7 (全 19 頁) 最終頁に続く

(21) 出願番号	特願2005-366569 (P2005-366569)	(73) 特許権者	000002369
(22) 出願日	平成17年12月20日 (2005.12.20)		セイコーエプソン株式会社
(65) 公開番号	特開2006-244460 (P2006-244460A)		東京都新宿区西新宿2丁目4番1号
(43) 公開日	平成18年9月14日 (2006.9.14)	(74) 代理人	100066980
審査請求日	平成20年11月20日 (2008.11.20)		弁理士 森 哲也
(31) 優先権主張番号	特願2005-24976 (P2005-24976)	(74) 代理人	100075579
(32) 優先日	平成17年2月1日 (2005.2.1)		弁理士 内藤 嘉昭
(33) 優先権主張国	日本国(JP)	(74) 代理人	100103850
			弁理士 田中 秀▲てつ▼
		(72) 発明者	轟 晃成
			長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内
		審査官	野田 佳邦

最終頁に続く

(54) 【発明の名称】 キャッシュメモリ及びプロセッサ

(57) 【特許請求の範囲】

【請求項1】

複数のプロセッサによって記憶装置から読み出されたデータの少なくとも一部をキャッシュしておき、キャッシュされたデータの少なくとも一部を前記プロセッサに供給するキャッシュメモリであって、

複数の前記プロセッサによってキャッシュされたデータが保存されるデータ保存手段と、

前記データ保存手段に記憶されているデータのアドレスを、前記データ保存手段が記憶しているデータ全体について一括して管理するアドレス管理手段と、

前記プロセッサによって供給を要求されたデータのアドレスを前記アドレス管理手段によって管理されているアドレスと照合し、供給を要求されたデータが前記データ保存手段から読み出し可能であるか否か検出するヒット検出手段と、

前記ヒット検出手段によってデータが読み出し可能であることが検出された場合、検出されたデータを前記プロセッサに供給するデータ供給手段と、

前記データ保存手段から読み出されたデータのみを、該データが前記プロセッサに出力されるまで一時的に保存するためのリード・バッファと、

前記プロセッサから送出されたデータのみを、該データが前記データ保存手段に書き込まれるまで一時的に保存するためのライト・バッファと、

前記リード・バッファに保存されているデータと前記ライト・バッファに保存されているデータとを比較し、前記リード・バッファに保存されているデータと前記ライト・バッ

ファに保存されているデータとの間で本来同じデータであるべきデータの不一致が検出された場合、前記リード・バッファに保存されているデータを前記データ保存手段に書き込まれたデータに更新する、または、前記リード・バッファに保存されているデータをいったん無効にし、前記ライト・バッファからデータを前記データ保存手段に書き込んだ後、書き込まれたデータを読み出すことにより、前記リード・バッファに保存されているデータと前記ライト・バッファに保存されているデータとを一致させるキャッシュ制御部と、を含むことを特徴とするキャッシュメモリ。

【請求項 2】

前記データ供給手段は、前記ヒット検出手段によってデータが読出し可能であることが検出された場合、検出されたデータを前記プロセッサに供給すると共に、前記プロセッサに供給されたデータと連続するデータを含むデータを、前記データ保存手段から読み出されたデータを一時的に保存する前記バッファにも供給することを特徴とする請求項 1 に記載のキャッシュメモリ。

10

【請求項 3】

前記プロセッサによって供給が要求されることが予想されるデータをキャッシュしておく先読みデータ保存手段をさらに備えることを特徴とする請求項 1 に記載のキャッシュメモリ。

【請求項 4】

前記データ管理手段は、前記データ保存手段のアドレスを複数のウェイとして管理すると共に、データ保存手段に保存されているデータを前記データ保存手段に保持する際の優先度である保存優先度を前記ウェイごとに付し、かつ、各ウェイに付される保存優先度を該ウェイで管理されているデータに対するアクセスの状態に基づいて決定することを特徴とする請求項 1 から 3 のいずれか 1 項に記載のキャッシュメモリ。

20

【請求項 5】

前記データ保存手段、前記データ管理手段の少なくとも一方がマルチポートメモリであることを特徴とする請求項 1 から 4 のいずれか 1 項に記載のキャッシュメモリ。

【請求項 6】

複数のプロセッサによって記憶装置から読み出されたデータの少なくとも一部をキャッシュしておき、キャッシュされたデータの少なくとも一部を前記プロセッサに供給するキャッシュメモリを備えたプロセッサであって、

30

前記キャッシュメモリは、

複数の前記プロセッサによってキャッシュされたデータが保存されるデータ保存手段と、

前記データ保存手段に記憶されているデータのアドレスを、前記データ保存手段が記憶しているデータ全体について一括して管理するアドレス管理手段と、

前記プロセッサによって供給を要求されたデータのアドレスを前記アドレス管理手段によって管理されているアドレスと照合し、供給を要求されたデータが前記データ保存手段から読出し可能であるか否か検出するヒット検出手段と、

前記ヒット検出手段によってデータが読出し可能であることが検出された場合、検出されたデータを前記プロセッサに供給するデータ供給手段と、

40

前記データ保存手段から読み出されたデータのみを、該データが前記プロセッサに出力されるまで一時的に保存するためのリード・バッファと、

前記プロセッサから送出されたデータのみを、該データが前記データ保存手段に書き込まれるまで一時的に保存するためのライト・バッファと、

前記リード・バッファに保存されているデータと前記ライト・バッファに保存されているデータとを比較し、前記リード・バッファに保存されているデータと前記ライト・バッファに保存されているデータとの間で本来同じデータであるべきデータの不一致が検出された場合、前記リード・バッファに保存されているデータを前記データ保存手段に書き込まれたデータに更新する、または、前記リード・バッファに保存されているデータをいったん無効にし、前記ライト・バッファからデータを前記データ保存手段に書き込んだ後、

50

書き込まれたデータを読み出すことにより、前記リード・バッファに保存されているデータと前記ライト・バッファに保存されているデータとを一致させるキャッシュ制御部と、を備えることを特徴とするプロセッサ。

【請求項 7】

複数の前記プロセッサの各々が、スレッドごとに処理を実行すると共に、処理の実行中に実行されているスレッドを他のスレッドと変更し得ることを特徴とする請求項 6 に記載のプロセッサ。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、キャッシュメモリ及びプロセッサに係り、特にマルチスレッドプロセッサ等の複数の処理を並列に実行するプロセッサに備えられるキャッシュメモリ及びこのようなキャッシュメモリを備えたプロセッサに関する。

【背景技術】

【0002】

近年、複数のスレッドやタスク（実施形態 1 ではスレッドに統一して記す）を並列に実行するマルチプロセッサ（マルチコア）、あるいはマルチスレッドプロセッサが注目されている。このようなプロセッサは、総称してマルチプロセッサシステムとも呼ばれる。マルチプロセッサシステムでは、データ等が蓄積されている外部メモリへのアクセスを効率化するため、外部メモリからいったん読み出されたデータのうちの処理に使用される可能性のあるデータを保存しておくキャッシュメモリを備えるマルチプロセッサシステムがある。このような構成の従来技術として、例えば、特許文献 1 が掲げられる。

【0003】

また、特許文献 1 に示したようなキャッシュメモリを備えたマルチプロセッサでは、複数のプロセッサ間で処理に使用されるデータの一致性（コヒーレンシー）を保つことが必要である。従来プロセッサでは、データのコヒーレンシーを保つためにバス・スヌープが多く採用されている。バス・スヌープとは、各プロセッサ間で共有されるメモリインターフェイスバス上のトランザクションを観察し、自身に割り当てられているキャッシュメモリ上にあるデータにかかるトランザクションが発生したか否かを検出する機能である。

【0004】

自身に割り当てられたキャッシュメモリ上にあるデータのトランザクションが発生した場合、プロセッサは、キャッシュメモリの該当するエントリを更新し、マルチプロセッサシステムにおける各プロセッサのキャッシュメモリ領域等に保存されているデータの内容を統一する。バス・スヌープには多くの実装法があり、例えばライトワンスやバークレイプロトコルがある。

【特許文献 1】特開 2004 - 178571

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかしながら、上記した特許文献 1 の発明は、キャッシュメモリの記憶領域を独立の領域に分割し、マルチプロセッサシステムで同時に実行されるスレッドごとに割り当てている。このような特許文献 1 によれば、キャッシュメモリに必要とされるデータが保存されていて、このデータに対するアクセスの成功（ヒット）の割合（ヒット率）が低下するという不具合がある。

【0006】

さらに、特許文献 1 に記した構成でバス・スヌープの機能を用いコヒーレンシーを保持する場合、バスをモニタする回路によってマルチプロセッサシステムのハードウェア構成が大規模化する。また、バスを常時モニタするために消費電力が高まり、キャッシュメモリの独立した各領域に各々アクセスしてデータを書き換えるため、コヒーレンシー保持の動作効率が低いという課題が生じる。

10

20

30

40

50

【0007】

本発明は、上記した点に鑑みてなされたものであって、マルチプロセッサシステムにあって装置構成を大きくする、あるいは消費電力を高めることがなく、プロセッサ間でデータのコピーレシーを動作効率よく保つことができるキャッシュメモリ及びこのキャッシュメモリを備えたプロセッサを提供することを目的とする。

【課題を解決するための手段】

【0008】

以上の課題を解決するため、本発明のキャッシュメモリは、複数のプロセッサによって記憶装置から読み出されたデータの少なくとも一部をキャッシュしておき、キャッシュされたデータの少なくとも一部を前記プロセッサに供給するキャッシュメモリであって、前記記憶装置から読み出されたデータが保存されるデータ保存手段と、前記データ保存手段に記憶されているデータのアドレスを、前記データ保存手段が記憶しているデータ全体について一括して管理するアドレス管理手段と、前記プロセッサによって供給を要求されたデータのアドレスを前記アドレス管理手段によって管理されているアドレスと照合し、供給を要求されたデータが前記データ保存手段から読出し可能であるか否か検出するヒット検出手段と、前記ヒット検出手段によってデータが読出し可能であることが検出された場合、検出されたデータを前記プロセッサに供給するデータ供給手段と、を備えることを特徴とする。

10

【0009】

このような発明によれば、複数のプロセッサによってキャッシュされたデータのデータ保存手段におけるアドレスを一括して管理することができるので、複数のプロセッサでキャッシュされたデータを実質的に1つのデータ保存手段に保存することになり、データ保存手段内におけるデータの不統一をなくすることができる。このため、プロセッサ間でデータのコピーレシーを動作効率よく保つことができるキャッシュメモリを提供することができる。さらに、データのコピーレシーを保つにあたり、別途回路等を追加する必要がないため、キャッシュメモリの装置構成を大きくする、あるいは消費電力を高めることがない。

20

【0010】

また、本発明のキャッシュメモリは、前記データ保存手段から読み出されたデータ、前記データ保存手段に書き込まれるデータの少なくとも一方を一時的に保存するバッファ手段をさらに備えることを特徴とする。

30

このような発明によれば、データ管理手段及びデータ保存手段に対するアクセス回数が低減し、キャッシュメモリへのアクセス速度を高め、キャッシュメモリを採用したプロセッサの処理速度を向上することができる。

【0011】

また、本発明のキャッシュメモリは、前記データ供給手段が、前記ヒット検出手段によってデータが読出し可能であることが検出された場合、検出されたデータを前記プロセッサに供給すると共に、前記プロセッサに供給されたデータと連続するデータを含むデータを、前記データ保存手段から読み出されたデータを一時的に保存する前記バッファ手段にも供給することを特徴とする。

40

【0012】

このような発明によれば、プロセッサに供給されたデータに続くデータを予めバッファに保存することができ、次回にこのデータが要求されたときデータ管理手段及びデータ保存手段にアクセスする必要がなくなる。このため、データ管理手段及びデータ保存手段に対するアクセス回数が低減し、キャッシュメモリへのアクセス速度を高め、キャッシュメモリを採用したプロセッサの処理速度を向上することができる。

【0013】

また、本発明のキャッシュメモリは、前記プロセッサによって供給が要求されることが予想されるデータをキャッシュしておく先読みデータ保存手段をさらに備えることを特徴とする。

50

このような発明によれば、データ管理手段及びデータ保存手段に対するアクセス回数が低減し、キャッシュメモリへのアクセス速度を高め、キャッシュメモリを採用したプロセッサの処理速度を向上することができる。

【0014】

また、本発明のキャッシュメモリは、前記データ管理手段が、前記データ保存手段のアドレスを複数のウェイとして管理すると共に、データ保存手段に保存されているデータを前記データ保存手段にデータを保持する際の優先度である保存優先度を前記ウェイごとに付し、かつ、各ウェイに付される保存優先度を該ウェイで管理されているデータに対するアクセスの状態に基づいて決定することを特徴とする。

【0015】

このような発明によれば、データを複数のウェイで管理する場合にもLRU方式を採用してキャッシュのヒット率を高めることができる。

また、本発明のキャッシュメモリは、前記データ保存手段、前記データ管理手段の少なくとも一方がマルチポートメモリであることを特徴とする。

このような発明によれば、複数のプロセッサがデータメモリ、タグメモリに高速にアクセスすることが可能になってマルチプロセッサの処理能力を向上させることができる。

【0016】

また、本発明のプロセッサは、複数のプロセッサによって記憶装置から読み出されたデータの少なくとも一部をキャッシュしておき、キャッシュされたデータの少なくとも一部を前記プロセッサに供給するキャッシュメモリを備えたプロセッサであって、前記キャッシュメモリは、前記記憶装置から読み出されたデータが保存されるデータ保存手段と、前記データ保存手段に記憶されているデータのアドレスを、前記データ保存手段が記憶しているデータ全体について一括して管理するアドレス管理手段と、前記プロセッサによって供給を要求されたデータのアドレスを前記アドレス管理手段によって管理されているアドレスと照合し、供給を要求されたデータが前記データ保存手段から読み出し可能か否かを検出するヒット検出手段と、前記ヒット検出手段によってデータが読み出し可能であることが検出された場合、検出されたデータを前記プロセッサに供給するデータ供給手段と、を備えることを特徴とする。

【0017】

このような発明によれば、複数のプロセッサによってキャッシュされたデータのデータ保存手段におけるアドレスを一括して管理することができるので、複数のプロセッサでキャッシュされたデータを実質的に1つのデータ保存手段に保存することになり、データ保存手段内におけるデータの不統一をなくすることができる。このため、プロセッサ間でデータのコピーレンシーを動作効率よく保つことができるプロセッサを提供することができる。さらに、データのコピーレンシーを保つにあたり、別途回路等を追加する必要がないため、プロセッサにあって装置構成を大きくする、あるいは消費電力を高めることがない。

【0018】

また、本発明のキャッシュメモリは、複数の前記プロセッサの各々が、スレッドごとに処理を実行すると共に、処理の実行中に実行されているスレッドを他のスレッドと変更し得ることを特徴とする。

このような発明によれば、1つのプロセッサが共通するデータにアクセスする可能性が高いマルチスレッドプロセッサにあってもプロセッサ間でデータのコピーレンシーを動作効率よく保つことができる。

【発明を実施するための最良の形態】

【0019】

以下、図を参照して本発明に係るキャッシュメモリ及びこのキャッシュメモリを備えたプロセッサの実施の形態1、実施形態2を説明する。

【0020】

(実施形態1)

図1は、本発明の実施形態1、実施形態2に共通のキャッシュメモリを備えたマルチス

10

20

30

40

50

レッドプロセッサ101を示した図である。マルチスレッドプロセッサ101は、外部メモリ105からデータを読み出す、あるいは実行されたスレッドの結果等を外部メモリ105に書込んでいる。

【0021】

また、マルチスレッドプロセッサ101は、キャッシュメモリ109を備え、外部メモリ105に対し、キャッシュメモリ109を介してデータを読み出す、あるいは書込んでいる。キャッシュメモリ109は、マルチスレッドプロセッサ101に含まれる複数のプロセッサによってキャッシュされたデータを保存しておき、キャッシュされたデータの少なくとも一部を複数のプロセッサのいずれかに供給する。このため、マルチスレッドプロセッサ101は、外部メモリ105にアクセスすることなくデータの多くの供給を受けることができる。

10

【0022】

キャッシュメモリ109は、一般的に外部メモリ105よりもプロセッサが高速にアクセス可能な構成を有している。このため、キャッシュメモリ109にアクセスしてデータを取得する構成は、マルチスレッドプロセッサ101のデータの読出し及び書込みを高速化すると共に外部メモリ105に対するアクセス回数を低減し、マルチスレッドプロセッサ101の処理速度及び処理効率を高めることができる。

【0023】

また、マルチスレッドプロセッサ101は、複数のスレッドと、スレッドの実行に使用されるプロセッサとを一对一に割り付けるものでなく、複数のプロセッサの各々が、スレッドごとに処理を実行すると共に、処理の実行中に実行されているスレッドを他のスレッドと変更し得る。このようなマルチスレッドプロセッサ101は、マルチスレッドOSによって動作するものである。

20

【0024】

すなわち、マルチスレッドプロセッサ101では、複数のプロセッサが、スレッドの優先順位に応じて実行されるスレッドを動的に変更する。このような動作を図2に例示する。図2に示した例では、マルチスレッドプロセッサ101がプロセッサ0からプロセッサ3の4つのプロセッサを備えている。そして、いずれのプロセッサもスレッドの実行中により優先度の高いスレッドの割り込みが入り、割り込んだスレッドに処理を切り換えている。

30

【0025】

スレッドが切り替えられる場合、マルチスレッドプロセッサ101は、切り替え直前のスレッドの状態や結果(コンテキスト)を退避させると共に、次に実行されるスレッドのコンテキストをプロセッサ設定する。

マルチスレッドプロセッサ101では、プロセッサ0からプロセッサ3のいずれでもOSが動作し、他のプロセッサを制御し得る。このようなマルチスレッドプロセッサは、複数のプロセッサが同等な立場で処理を分担するものであって、対称型マルチプロセッサ(SMP: Symmetric Multi Processor)とも呼ばれている。

【0026】

また、キャッシュメモリ109は、データの保存に使用されるキャッシュメモリ部107と、キャッシュメモリ部107に対するデータの保存を制御するキャッシュ制御部103とを備えている。キャッシュメモリ部107は、後に図示するように、データに付されたアドレスや状態を管理するタグメモリと、データ本体を保存するためのデータメモリとを含んでいる。

40

【0027】

図3は、キャッシュメモリ109の構成をより詳細に示した図である。なお、実施形態1では、キャッシュメモリ109がプロセッサ0からプロセッサ3までの4つのプロセッサと接続し、4つのプロセッサからデータの要求を受けると共に、4つのプロセッサから受け取ったデータを書き込むことが可能である。なお、各プロセッサがキャッシュメモリ109に対して行うデータの要求を、実施形態1では以降読出し命令と記す。

50

【0028】

キャッシュメモリ109は、図1に示したように、キャッシュ制御部103と、キャッシュメモリ部107と、ヒット検出部208とを備えている。キャッシュメモリ部107は、タグメモリ206とデータメモリ207とを有し、データメモリ207は、外部メモリ105から読み出されたデータが保存される構成である。また、タグメモリ206は、外部メモリ105から読み出されたデータを保存するデータメモリ207や後述するリード・バッファにおけるアドレスを一括して管理する構成である。

【0029】

タグメモリ206は、データの外部メモリ105におけるデータと、このデータが現在保存されているアドレスとを対応付けるデータを例えばテーブル等によって保存するメモリである。データは、外部メモリ105の他、データメモリ207にも保存されている可能性があるから、データが現在保存されているアドレスは、データメモリ207のアドレスをもとり得るものである。

【0030】

なお、実施形態1では、タグメモリ206が、データのアドレスの他、状態(ステータス)をも管理する。ここでいうステータスとは、データの有効、無効やダーティ(外部メモリ105から読み出された後に変更された)であるか否か等を示す情報である。

また、キャッシュ制御部103は、アドレス制御部201、バッファ管理部202、ライト・バッファ203、リード・バッファ205を備えている。

【0031】

アドレス制御部201は、プロセッサから入力された読出し命令から要求されたデータのアドレスを取得し、タグメモリ206及びデータメモリ207をアクセスする際のアドレスに変換してタグメモリ206に出力する。あるいは、データメモリ207にキャッシュされているデータを読み出す際、タグメモリ206およびデータメモリ207に読み出されるデータのアドレスを出力する。さらに、データメモリ207にキャッシュされていないデータを外部メモリ105から読み出す際、外部メモリ105のアドレスを生成し、外部メモリ105に出力する。

【0032】

キャッシュ制御部103は、アドレス制御部201で生成されたアドレスに基づいてタグメモリ206とデータメモリ207とに対するアクセスを制御する。また、キャッシュ制御部103は、タグメモリ206およびデータメモリ207へのアクセスが実際に生じるときだけ、メモリアクセスのクロックを供給するなどの方法によって消費電力を低減することも可能である。

【0033】

ライト・バッファ203は、データメモリ207に書き込まれるデータを一時的に保存(バッファリング)するバッファであって、プロセッサ0からプロセッサ3の各々に対応するバッファ203a、203b、203c、203dを備えている。また、リード・バッファ205は、データメモリ207から読み出されたデータをバッファリングするバッファであって、ライト・バッファ203と同様に、プロセッサ0からプロセッサ3の各々に対応するバッファ205a、205b、205c、205dを備えている。ライト・バッファ203、リード・バッファ205は、データメモリ207に対する書込みあるいは読出しのタイミングを調整するために設けられた構成である。

【0034】

さらに、キャッシュ制御部103は、バッファ管理部202を備えている。バッファ管理部202は、ライト・バッファ203とリード・バッファ205との間のデータの整合性をとるための構成である。すなわち、バッファ管理部202は、ライト・バッファ203に保存されているデータとリード・バッファ205に保存されているデータとを比較し、本来同じデータであるべきデータの不一致が検出された場合には、例えばリード・バッファ205に保存されている側のデータを更新、あるいは削除することによって両者を一致させている。

10

20

30

40

50

【0035】

ヒット検出部208は、プロセッサによって供給を要求されたデータのアドレスをタグメモリ206で管理されているアドレスと照合し、データメモリ207に対して供給を要求されたデータがあるか否か検出する。さらに、実施形態1では、供給を要求されたデータが検出された場合、ヒット検出部208が検出されたデータをマルチスレッドプロセッサ101に供給している。

【0036】

次に、以上述べた構成の動作を、キャッシュメモリ109に対するリードとライトとについて説明する。

【0037】

(リード動作)

マルチスレッドプロセッサ101は、複数のプロセッサのうちの例えばプロセッサ0から読出し命令を出力する。読出し命令は、供給が要求されるデータの外部メモリ105におけるアドレス(読出しアドレス)と、読出しを指示する信号(読出し制御信号)とを含んでいる。アドレス制御部201は、リード・バッファ205のプロセッサ0に対応するバッファ205aに保存されているデータを検出する。

【0038】

そして、バッファ205aにデータが保存されている場合、保存されているデータに付されたタグアドレスと、読出しアドレスとを照合する。そして、読出しアドレスがリード・バッファ205のバッファ205aに保存されているデータのタグアドレスと一致した場合、バッファ205aに保存されているデータをプロセッサ0に出力し、読出し処理を完了させる。

【0039】

読出しアドレスに対応するタグアドレスが付されたデータがリード・バッファ205に保存されていない場合、アドレス制御部201は、タグメモリ206にアクセスして読出しアドレスを照合する。なお、プロセッサは、読出しアドレスとして、外部メモリ105におけるアドレスを使用する。実施形態1では、アドレス制御部201が外部メモリ105におけるアドレスをデータメモリ207にアクセスする際のアドレスに変換する。

【0040】

タグメモリ206は、読出しの対象となるデータのデータメモリ207におけるアドレスとステータスとを対応付けて保存している。タグメモリ206における照合の結果、読出しアドレスによって指定されたデータがデータメモリ207にあった場合、タグメモリ206は、ヒット検出部208にデータのステータスを出力する。また、タグメモリ206における照合結果はデータメモリ207にも出力され、データメモリ207からヒット検出部208へプロセッサ0によって読出されるデータが出力される。ヒット検出部208は、データのステータスに基づいてデータが読出し可能である、つまりキャッシュヒットしたと判断した場合、ヒットしたデータをプロセッサ0へ出力する。

【0041】

また、ヒット検出部208は、データの出力と同時に出力されたデータにかかる1エントリ分のデータをリード・バッファ205に転送する。このような処理により、次のアクセス時に転送されたデータが読み出される場合、キャッシュ制御部103がタグメモリ206とデータメモリ207とにアクセスすることなくデータの読出しが可能になる。

一方、読出しアドレスに対応するデータがタグメモリにない、あるいはデータのステータスが無効である等の理由によってデータが読み出せない(キャッシュミスヒットした場合、キャッシュ制御部103は、外部メモリ105からデータメモリに読み込むと共にリード・バッファ205に転送した後にプロセッサ0へ出力する。また、キャッシュヒットした場合と同様に1エントリ分のデータをリード・バッファ205に転送する。

【0042】

なお、以上述べた実施形態1では、リード・バッファ205が複数のプロセッサの各々に対応するバッファ205a~205dを備えている。しかし、実施形態1は、このよう

10

20

30

40

50

な構成に限定されるものでなく、プロセッサ 0 による読出し命令の読出しアドレスと一致するタグアドレスが付されたデータがバッファ 205 a 以外のバッファに保存されていた場合にもバッファ 205 a 以外のバッファに保存されているデータを読み出す機能を追加することもできる。

【0043】

複数のプロセッサが動的にスレッドを切り換えて処理を実行するマルチスレッドプロセッサは、複数の異なるプロセッサによって同一のデータが繰り返し使用される可能性がある。このため、上記したように、読出し命令を出力したプロセッサがリード・バッファ 205 にあるバッファ 205 a ~ 205 d のいずれにもアクセス可能とすれば、例えばプロセッサ 0 が他のプロセッサの処理時にリード・バッファ 205 に転送されたデータを読み出すことができる。

10

【0044】

このような処理によれば、キャッシュ制御部 103 がタグメモリ 206、データメモリ 207 にアクセスする回数が低減し、マルチスレッドプロセッサ 101 のデータ読出しに係る処理効率が向上する。

また、以上述べた実施形態 1 は、プロセッサ間でデータのコヒーレンシーを動作効率よく保つことができる。すなわち、例えばプロセッサごとにタグメモリやデータメモリを備える構成では、同じタグアドレスを持つデータが複数の異なるデータメモリに保存されることになる。そして、複数のデータメモリに存在するデータのうちの一部だけが更新される等の理由によってプロセッサ間におけるデータの不一致が発生する。

20

【0045】

しかし、実施形態 1 によれば、外部メモリ 105 から読み出されたデータのデータメモリ 207 におけるアドレスをタグメモリ 206 が一括して管理するため、実質的に唯一のデータメモリによって読出されたデータを保存することになり、キャッシュメモリにおけるデータの不一致をなくすることができる。また、このために実施形態 1 は、バスを監視する必要がなく、バスを監視するための回路や消費電量が不要である。したがって、プロセッサ間でデータのコヒーレンシーを保つためにマルチプロセッサシステムの装置構成を大きくする、あるいは消費電力を高めることがない。

【0046】

さらに、実施形態 1 のマルチスレッドプロセッサは、図 2 に示したように、1つのプロセッサが複数のスレッドを動的に切り替えて実行するマルチスレッド OS のように、データの不一致が生じやすい構成に適用した場合に特に有利である。

30

また、実施形態 1 は、ヒットしたデータの読出しと共に、このデータにかかる 1 エントリ分のデータをリード・バッファ 205 に転送しておく。このため、プロセッサがタグメモリ 206 やデータメモリ 207 にアクセスする回数を低減し、データの読出しにかかるマルチスレッドプロセッサの負荷を軽減することができる。

【0047】

つまり、読出しの対象となるデータを保存するデータメモリは、一般にバッファに比べてアクセスタイムが遅く、プロセッサのパフォーマンス向上のボトルネックになりやすい。データメモリの前段にリード・バッファやライト・バッファを設けることによってメモリアccessの遅延を見かけ上隠すことが可能になり、結果としてプロセッサの性能向上を図ることができる。

40

【0048】

(ライト動作)

次に、実施形態 1 のマルチスレッドプロセッサによるライト動作について説明する。なお、以下に述べるライト動作は、ライトバックを例にしているが、ライトスルーにも適用することができる。

マルチスレッドプロセッサ 101 は、複数のプロセッサのうちの例えばプロセッサ 0 からデータを書き込むよう指示する命令(書込み命令)を出力する。書込み命令は、書込みが要求されるデータの外部メモリ 105 におけるアドレス(書込みアドレス)と、書込み

50

を指示する信号（書込み制御信号）とを含んでいる。また、書込み動作にあつては、書込み命令と共に書込まれるデータ（書込みデータ）もがプロセッサから送出される。

【0049】

書込みデータは、キャッシュメモリ109において、先ず、書込みアドレスと共にライト・バッファ203に保存される。ライト・バッファ203は、FIFOメモリ（First In First Out memory）であり、書込まれたデータを書込みの順にデータメモリ207に書き込んでいる。

キャッシュ制御部103は、データをデータメモリ207に書き込むため、先ず、タグメモリ206のタグアドレスに書込みアドレスを照合すると共に、データのステータスを検出する。この結果、データメモリ207にデータの書込みが可能である、つまりキャ

10

【0050】

ッシュヒットしたと判断された場合、書込みデータをデータメモリ207に書込む。また、タグメモリに保存されているデータのステータスを示すフラグを「ダーティ」にする。

また、キャッシュ制御部103は、データメモリ207にデータの書込みが不可能である、つまりキャッシュミスヒットしたと判断した場合、外部メモリ105から書込みデータに該当するデータをデータメモリ207に読み出す。そして、ライト・バッファ203に書き込む、タグメモリ206を更新する。

実施形態1ではライト・バッファ203にFIFOメモリを採用したため、プロセッサによるライト・バッファ203への書込みが、ライト・バッファ203が一杯になるまで次々に行われる。また、ライト・バッファ203に書き込まれたデータは、タグメモリ2

20

【0051】

なお、バッファ管理部202は、上記のリード動作におけるリード・バッファ205への書込み、ライト動作におけるライト・バッファ203への書込みや読出しを調停する。また、ライト・バッファ203、リード・バッファ205間でデータ一緻性（コヒーレンシー）を保障する。

ライト・バッファ203、リード・バッファ205間のデータ一緻性は、ライト・バッファ203に書込まれたデータがデータメモリ207書き込まれる前、同一のデータであつてリード・バッファ205にあるものにリードアクセスが生じたときに問題となる。こ

30

【0052】

なお、処理効率の観点からは、リード・バッファ205の内容をライトされたデータに更新することが望ましい。

また、このようなリード・バッファ205の更新にあつては、プロセッサ0がリード・バッファ205の対応するリード・バッファ部分にアクセスした場合にも、他のプロセッサに対応するリード・バッファ205も更新対象にする必要がある。すなわち、同一の

40

【0053】

以上述べた実施形態1によれば、書込みデータが、タグメモリ206によってアドレスが一括して管理されるデータメモリ207に書き込まれるため実質的に唯一のデータメモリによって読出されたデータを保存することになり、キャッシュメモリにおけるデータの不一致をなくすることができる。また、このために実施形態1は、バスを監視する必要がなく、バスを監視するための回路や消費電量が不要である。したがって、プロセッサ間でデータの

50

あるいは消費電力を高めることがない。

【0054】

図4は、以上述べた実施形態1のキャッシュメモリにおいて実行されるデータの読出しあるいは書込みの動作を説明するためのフローチャートである。また、図5は、図4と比較するため、従来のキャッシュメモリで実行されるデータの読出しの動作を説明するためのフローチャートである。

図4に示したように、実施形態1のキャッシュメモリは、複数のプロセッサのうちの一つ(プロセッサkとする)が、キャッシュメモリ109にアクセスを要求した場合、タグメモリ206においてキャッシュヒットを検出する処理を行う(S401)。この結果、キャッシュヒットが検出された場合(S402: Yes)、データメモリ207にアクセスし、読出しの対象となるデータを読み出す(S406)。また、タグメモリ206におけるデータのステータスを更新する(S407)。

10

【0055】

また、キャッシュメモリ109は、ステップS402において、キャッシュミスヒットを検出した場合(S402: No)、データメモリ207に保存されているデータのうち入れ替えられるデータを決定する(S403)。そして、データメモリ207からデータ・データを書き出し(S404)、外部メモリ105に保存されている新規なデータをデータメモリ207に読み出す(S405)。

【0056】

図5に処理を示した従来のキャッシュメモリは、k個のプロセッサの各々が独立したタグメモリ及びデータメモリを備える点で実施形態1のキャッシュメモリと相違する。このため、ステップS502の判断において、従来のキャッシュメモリは、アクセスを要求したプロセッサkに対応するデータメモリに対してアクセスし(S506)、プロセッサkに対応するタグメモリを更新する(S507)。

20

【0057】

さらに、プロセッサkのアクセスが書込みであるか否か判断し(S508)、書込であった場合には(S508: Yes)、プロセッサk以外のプロセッサに対応するデータメモリ及びタグメモリのデータをも更新し、データメモリ間におけるデータの coherence を調整している。

また、以上述べた本実施形のキャッシュメモリは、以上述べた構成に限定されるものでなく、データの先読み機能を付加した構成とすることも可能である。図6は、実施形態1のキャッシュメモリを命令キャッシュにも適用し、先読みキャッシュとして構成したものである。

30

【0058】

図6に示した構成は、プロセッサによって供給が要求されることが予想されるデータ(命令)をキャッシュしておく先読みデータ保存手段である先読みバッファをさらに備えている。キャッシュメモリと接続する複数のプロセッサは、各々独立に別のプログラムにアクセスするため、先読みバッファは、プロセッサの個数に対応した数必要になる。先読みキャッシュのアドレス制御部601は、先読みのためアドレスの連続性を検出するため、あるいはタグメモリおよびデータメモリをキャッシュのエントリ単位で行うためにエントリの境界をまたがる度に次のタグメモリのアドレスを生成する。

40

【0059】

なお、先読みバッファの機能は、従来はタグメモリおよびデータメモリに対するアクセス回数を低減して低消費電力化することを目的にしていた。しかし、先読みの機能をマルチスレッドプロセッサに適用する場合、図3に示したリード・バッファ205、ライト・バッファ203と同様に、メモリデータにアクセス回数を低減してメモリアクセスのボトルネックを解消し、処理速度の向上を図ることができる。

【0060】

(実施形態2)

次に、本発明の実施形態2について説明する。

50

実施形態2のキャッシュメモリは、実施形態1で説明した図3の構成を有している。このため、実施形態2では、キャッシュメモリの構成の図示及び説明の一部を省くものとする。実施形態2のキャッシュメモリは、データメモリ207、タグメモリ206少なくとも一方がマルチポートメモリである。データメモリ207をマルチポートメモリとする場合、このマルチポートメモリは、プロセッサの数にウェイの数を乗じた数のポートが必要になる。また、タグメモリ206をマルチポートメモリとする場合、このマルチポートメモリは、プロセッサの数のポートが必要になる。

【0061】

実施形態2のプロセッサは、実施形態2で説明したキャッシュメモリのデータ書出し（データをデータメモリ207から外部メモリ105に書き出す動作）に関する構成を説明するものである。

10

実施形態2では、読み込み及び書き出しの操作に対し、いわゆるLRU（Least Recently Used algorithm）方式を採用している。LRUとは、キャッシュされているデータのうち、プロセッサがアクセスした後最も長い時間が経過したものをキャッシュメモリから取り除く方法によりデータに対するプロセッサの供給要求の状況に基づいて決定している。このような方式によれば、常にプロセッサの要求頻度が多いデータをデータメモリ207にキャッシュしておくことができ、キャッシュメモリを採用した構成の処理効率を高めることができる。

【0062】

また、キャッシュメモリの方式には種々のものがあるが、実施形態2では、キャッシュメモリ109が、2ウェイ（ウェイA、B）のセット・アソシアティブ方式のキャッシュメモリであるものとする。なお、セット・アソシアティブ方式とは、キャッシュメモリを複数の領域（ウェイ）に分割し、それぞれのウェイに、メモリデバイス上の異なるアドレスのデータを格納しておくことにより、ヒット率を向上させることができる方式である。

20

【0063】

実施形態2のキャッシュメモリは、タグメモリ206が、データメモリ207のアドレスを複数のウェイとして管理する。そして、データメモリ207に保存されているデータを前記データ保存手段にデータを保持する際の優先度である保存優先度を前記ウェイごとに付し、かつ、各ウェイに付される保存優先度を該ウェイで管理されているデータに対する

なお、実施形態2では、保存優先度を決定する基となるアクセスの状態を、現在に比較

30

的に近い所定の期間のアクセス回数とする。このような実施形態2によれば、

先ず、実施形態2のタグメモリ206及びデータメモリ207の構成を詳細に説明する。図7（a）、（b）、（c）は、タグメモリ206、データメモリ207に記憶されるデータの構造を説明するための図である。（a）は、タグメモリ206によって管理されるステータスのフラグを示している。

【0064】

フラグは、プロセッサ0～3の各々ごとにタグメモリ206に保存されていて、実施形態2では、データのステータスをValid flag、Dirty flag、Used flagの3つのフラグによって示している。Valid flagは、データの有効性を示すフラグである。Dirty flagは、キャッシュされているデータが読み込んだ値から交信されている状態（ダーティデータ）であることを示し、Used flagは、書出しの優先度（書出し優先度）を示している。

40

なお、実施形態2では、各ウェイの書出し優先度をUsedフラグに基づいてテーブル（Usedテーブル）で管理する。Usedテーブルについては、図8に示して説明する。

【0065】

図7（b）は、タグメモリ206のデータ構造を説明するための図である。また、図7（c）は、データメモリ207のデータ構造を説明するための図である。2ウェイのセット・アソシアティブ方式を採用した実施形態2では、プロセッサ0～プロセッサ3の各々について2つのウェイを持っていて、タグメモリ206は、データメモリ207を合計8

50

個のウェイ(W0～U7)として管理している。

【0066】

タグメモリ206に保存されるデータ(タグ情報)は、データのヒット、ミスヒットを検出するためのデータであって、アクセスされるデータのアドレスの16ビット分が保存されている。また、タグ情報に基づいて読み出されるデータは、1ワードが32ビットのデータとしてデータメモリにキャッシュされている。

図8(a)、(b)は、実施形態2のLRUの処理を説明するための図であって、データの読出しの前後のUsedテーブルを示している。実施形態2では、Usedテーブルをタグメモリ206に保存し、キャッシュ制御部103によって更新するものとした。

【0067】

プロセッサ0～3によって読み出されたデータは、データメモリ207のウェイU0～U7のいずれかにおいてキャッシュされる。ウェイU0～U7においてキャッシュ可能な数のデータがキャッシュされた後、さらに他のデータをキャッシュする必要がある場合、キャッシュ制御部103は、現在ウェイU0～U7にキャッシュされているデータの1つを外部メモリ105に書き出す。そして、新たにキャッシュされた他のデータを書き出されたデータがキャッシュされていた領域に保存する。

【0068】

合計8つのウェイのキャッシュメモリを持つ実施形態2では、ウェイU0～U7にキャッシュされているデータのいずれを書き出すかを、ウェイに対するプロセッサの供給要求の状態に基づいて決定している。そして、決定したウェイにあって最もアクセス回数が少なかったデータを書き出すものとする。

このような実施形態2において、アクセス回数の判断を短い時間に限定すれば、直前のデータアクセスの有無によってデータ書出しの対象となるウェイを決定することも可能である。

【0069】

書出し優先順位は、図8(a)、(b)の表中にLRU順位としてUsedテーブルに記録されている。実施形態2でいうLRU順位は、0～7の数値によって表されていて、0は最高のLRU順位を示し、1は最低のLRU順位を示す。そして、LRU順位7が付されたウェイのデータは、次に起こるキャッシュミスヒット時に他のウェイのデータに優先して外部メモリ105に書き出される。

【0070】

図8(a)は、プロセッサによって要求されたデータがデータメモリ207にミスヒットした場合のLRU順位の決定について説明するための図である。プロセッサが要求したデータがタグメモリ206に管理されるデータにないため、プロセッサは、外部メモリ105にアクセスしてデータを読み出し、データメモリ207にキャッシュする。この際、キャッシュ制御部103は、Usedテーブルを参照し、ウェイU0～U7のLRU順位を参照する。

【0071】

図8(a)の場合、ウェイU6のLRU順位が最低の7であるから、キャッシュ制御部103は、ウェイU6にキャッシュされていて、キャッシュされたウェイ6のLRU順位を0にし、他のウェイUのLRUデータに対するプロセッサの供給要求の状態に基づいて決定している。プロセッサがアクセスした後最も長い時間が経過したものをキャッシュメモリから取り除く方法によりデータを書き出す。そして、最新のデータが順位をそれぞれ1つずつ低下させる。

【0072】

また、図8(b)は、プロセッサが要求したデータがデータメモリ207で管理されているデータにヒットした場合のLRU順位の決定について説明するための図である。プロセッサが要求したデータがウェイU4のデータにヒットした場合、キャッシュ制御部103は、このデータを読み出してプロセッサに供給する。このとき、データの書出しは必要ないが、直前にヒットしたウェイU4のLRU順位を0に更新し、この更新に伴ってヒッ

10

20

30

40

50

ト前のウェイU4のLRU順位(4)より高位のLRU順位を更新する。

【0073】

図9は、以上述べた実施形態2のうち、キャッシュの制御を説明するためのフローチャートである。キャッシュ制御部103は、プロセッサからデータアクセスの要求を受け、タグメモリ206に要求されたデータがヒットするか否かを検出する(S701)。プロセッサに対応して管理されているデータにヒットしたか否か判断し(S702)、ヒットが検出された場合には(S702:Yes)、アクセスがデータの書出しを要求するものか否か判断する(S707)。データの書出しが要求された場合(S707:Yes)、データメモリのタグに対応するウェイ(Way(n))へデータを書き込む(S710)。

10

【0074】

また、ステップS707において、アクセスがデータの書出しを目的とするものでないと判断された場合(S707:No)、Way(n)にキャッシュされているデータを読み出し、アクセスしたプロセッサに供給する(S708)。そして、キャッシュ制御部103は、LRUにしたがってこのデータのアクセス履歴等を示す情報をタグメモリ206において更新する(S709)。

【0075】

一方、ステップS702において、データがヒットしないと判断された場合(S702:No)、キャッシュ制御部103は、アクセス回数が最も少ないウェイ(Way(n))を検出し、さらにウェイ(Way(n))においてプロセッサがアクセスした後最も長い時間が経過したデータをLRUのアルゴリズムによって検出する(S703)。そして、検出されたデータがダーティデータであるか否か判断する(S704)。データがダーティデータである場合(S704:Yes)、このデータをデータメモリ207から書出し(S705)、書き出された領域に外部メモリ105からデータを読み出す(S706)。

20

【0076】

図10は、図9に示した処理のうち、UsedテーブルのLRU順位を変更するための処理を説明するためのフローチャートである。キャッシュ制御部103は、プロセッサによって要求されたデータをタグメモリ206に照会し、要求されたデータがいずれかのウェイにヒットしたか否か判断する(S801)。ヒットしたと判断された場合(S801:Yes)、ヒットしたウェイのLRU順位を0に更新する(S806)。

30

【0077】

次に、キャッシュ制御部103は、変数sを0に設定し(S807)、複数のプロセッサの各ウェイのLRU順位sを順次s+1に更新する(S808)。この更新は、更新後のLRU順位が、ヒットしたプロセッサのウェイのヒット直前のLRU順位に達するまで行われる(S809)。

一方、プロセッサkが要求したデータがキャッシュにヒットしないとき(S801:No)、キャッシュ制御部103は、LRU順位が最も低いウェイを検出する。そして、このウェイのLRU順位を0に更新する(S802)。なお、この際、外部メモリ105から読み出されたデータは、今回LRU順位が0に更新されたウェイにキャッシュされる。

40

【0078】

そして、キャッシュ制御部103は、変数sを0に設定し(S808)、複数のプロセッサの各ウェイのLRU順位sを順次s+1に更新する(S809)。この更新は、全てのウェイについて行われる。

以上述べた実施形態2によれば、ウェイを複数備えたマルチプロセッサに適したLRU方式を実現し、キャッシュヒット率を高めることができる。また、データメモリやタグメモリをマルチポートメモリとしたことによって複数のプロセッサがキャッシュメモリに同時にアクセスすることができる。このため、実施形態2は、マルチプロセッサの処理能力を向上させることができる。

【0079】

50

なお、複数のプロセッサが同時にキャッシュメモリにアクセスした場合のLRU順位の更新は、例えば、プロセッサに対して予め優先順位を付しておき、この優先順位にしたがう順序で更新するようにすることも可能である。

【図面の簡単な説明】

【0080】

【図1】本発明の実施形態1、実施形態2のキャッシュメモリを備えたマルチスレッドプロセッサを示した図である。

【図2】複数のプロセッサが優先順位に応じて実行されるスレッドを動的に変更する動作を説明するための図である。

【図3】図1に示したキャッシュメモリの構成をより詳細に示した図である。

10

【図4】実施形態1のキャッシュメモリにおいて実行されるデータの読出しあるいは書込みの動作を説明するためのフローチャートである。

【図5】図4と比較するため、従来のキャッシュメモリで実行されるデータの読出しの動作を説明するためのフローチャートである。

【図6】実施形態1のキャッシュメモリを命令キャッシュにも適用し、先読みキャッシュとして構成したものである。

【図7】本発明の実施形態2において、タグメモリ、データメモリに記憶されるデータの構造を説明するための図である。

【図8】実施形態2のLRUの処理を説明するための図である。

【図9】実施形態2のうちキャッシュの制御を説明するためのフローチャートである。

20

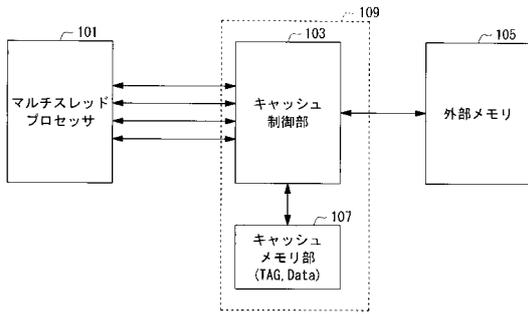
【図10】実施形態2のうちキャッシュの制御のうちLRUの更新を説明するためのフローチャートである。

【符号の説明】

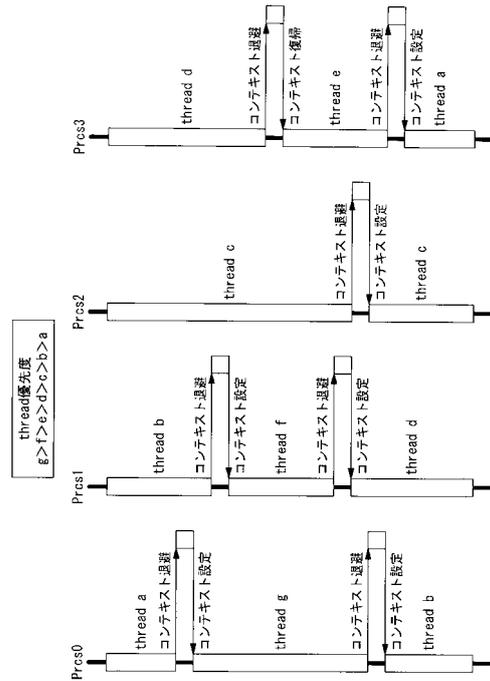
【0081】

101 マルチスレッドプロセッサ、103 キャッシュ制御部、105 外部メモリ、
107 キャッシュメモリ部、109 キャッシュメモリ、201 アドレス制御部、2
02 バッファ管理部、203 ライト・バッファ、205 リード・バッファ、206
タグメモリ、207 データメモリ、208 ヒット検出部

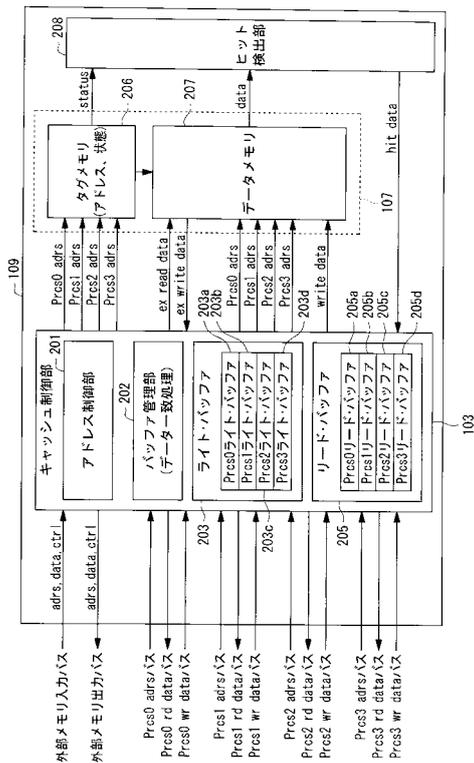
【図1】



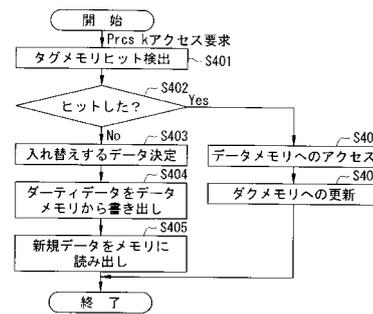
【図2】



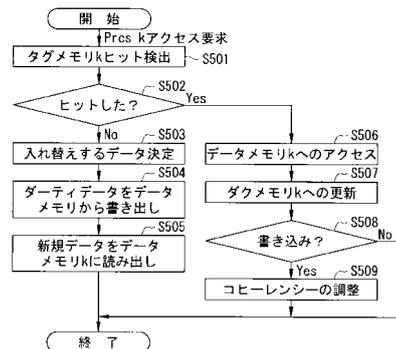
【図3】



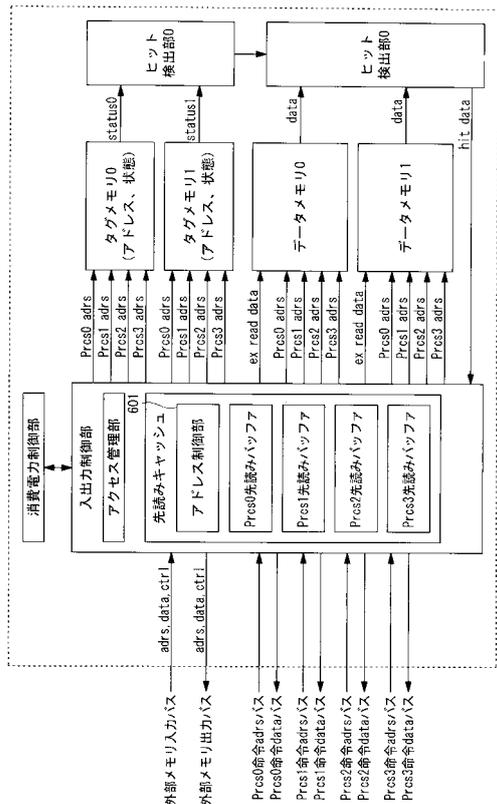
【図4】



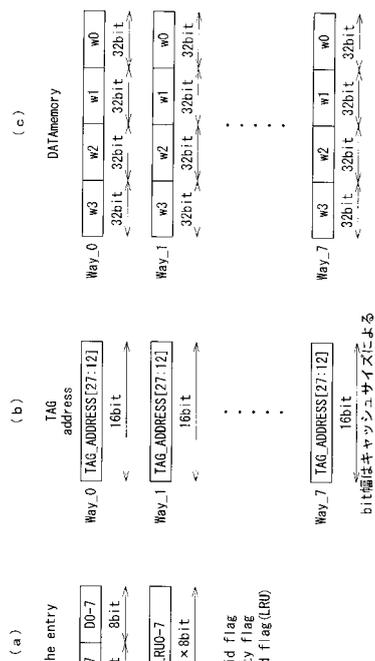
【図5】



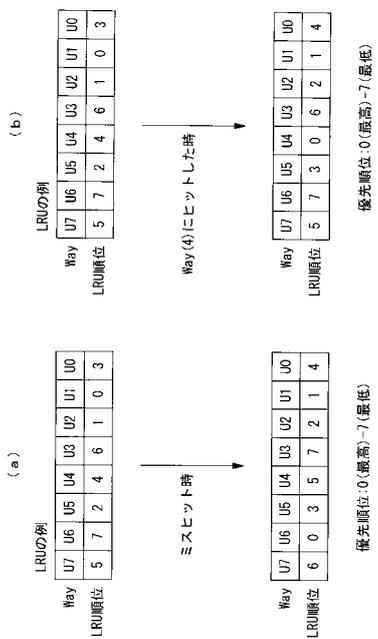
【図6】



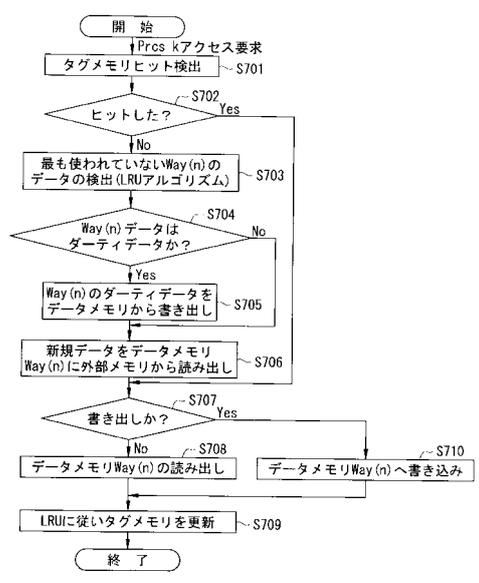
【図7】



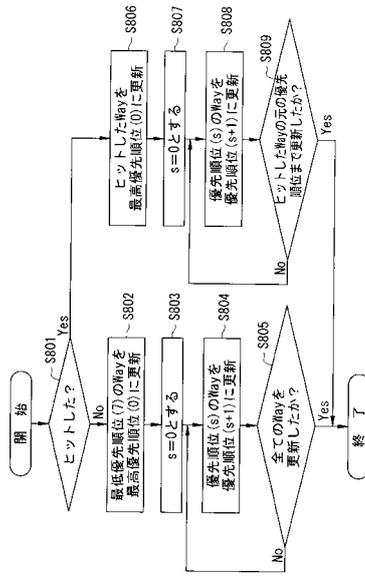
【図8】



【図9】



【図10】



フロントページの続き

(51)Int.Cl. F I
G 0 6 F 12/08 5 6 5
G 0 6 F 9/46 4 1 0

(56)参考文献 特開2002-140232(JP,A)
特開2001-043130(JP,A)
特開平02-224041(JP,A)
特開平02-204834(JP,A)

(58)調査した分野(Int.Cl., DB名)
G 0 6 F 1 2 / 0 8 - 1 2 / 1 2
G 0 6 F 9 / 4 6