

(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2020-3959  
(P2020-3959A)

(43) 公開日 令和2年1月9日(2020.1.9)

(51) Int. Cl.	F 1	テーマコード(参考)
<b>G06F 12/0811 (2016.01)</b>	G06F 12/0811	5B205
<b>G06F 12/0895 (2016.01)</b>	G06F 12/0895 112	
<b>G06F 12/0864 (2016.01)</b>	G06F 12/0864	

審査請求 未請求 請求項の数 9 O L (全 20 頁)

(21) 出願番号	特願2018-121355 (P2018-121355)	(71) 出願人	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
(22) 出願日	平成30年6月26日(2018.6.26)	(74) 代理人	110002147 特許業務法人酒井国際特許事務所
		(72) 発明者	平野 孝仁 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
		Fターム(参考)	5B205 JJ11 KK13 KK14 MM05 NN46 UU32

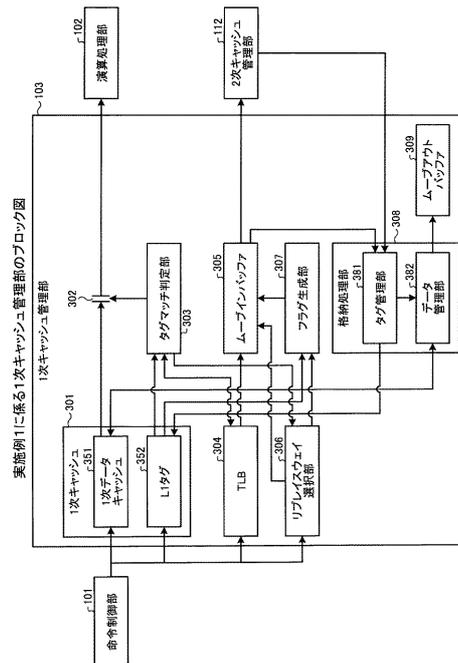
(54) 【発明の名称】 情報処理装置、演算処理装置及び情報処理装置の制御方法

(57) 【要約】

【課題】 処理性能を向上させる情報処理装置、演算処理装置及び情報処理装置の制御方法を提供する。

【解決手段】 タグマッチ判定部303は、所定データの取得要求を受けて、1次キャッシュ301に所定データが存在するか否かを判定する。ムーブインバッファ305は、1次キャッシュ301に所定データが存在しない場合、2次キャッシュ管理部112又は記憶装置へ所定データの取得要求を出力するとともに、所定データを格納する所定領域の状態情報に基づく判定用情報を保持する。格納処理部308は、取得要求に対する応答を2次キャッシュ管理部112又は記憶装置から取得し、取得した応答が所定の種類の場合、ムーブインバッファ305が保持する判定用情報を基に、1次キャッシュ301に格納された状態情報を取得するか否かを判定し、状態情報を取得しないと判定した場合、所定領域を無効化し、所定領域に応答に含まれる所定データを格納する。

【選択図】 図3



**【特許請求の範囲】****【請求項 1】**

複数の演算処理装置を有する情報処理装置であって、  
前記演算処理装置は、  
データ及びデータを格納する領域の状態情報を保持するキャッシュメモリと、  
所定データの取得要求を受けて、前記キャッシュメモリに前記所定データが存在するか否かを判定する判定部と、

前記キャッシュメモリに前記所定データが存在しない場合、前記キャッシュメモリよりも下位のキャッシュメモリ又は記憶装置へ前記所定データの取得要求を出力するとともに、前記所定データを格納する所定領域の状態情報に基づく判定用情報を保持する制御部と、

前記取得要求に対する応答を前記下位のキャッシュメモリ又は記憶装置から取得し、取得した前記応答が所定の種類の場合、前記制御部が保持する前記判定用情報を基に、前記キャッシュメモリに格納された前記状態情報を取得するか否かを判定し、前記状態情報を取得しないと判定した場合、前記所定領域を無効化し、前記所定領域に前記応答に含まれる前記所定データを格納する格納処理部と

を備えたことを特徴とする情報処理装置。

**【請求項 2】**

前記格納処理部は、前記状態情報を取得すると判定した場合、前記キャッシュメモリから前記状態情報を取得し、取得した前記状態情報が第 1 状態の場合、前記所定領域が保持する既存データを退避させ、前記所定領域を無効化し、前記所定領域に前記所定データを格納し、取得した前記状態情報が第 2 状態の場合、前記所定領域を無効化し、前記所定領域に前記所定データを格納することを特徴とする請求項 1 に記載の情報処理装置。

**【請求項 3】**

前記格納処理部は、前記応答が所定の種類でない場合、前記キャッシュメモリから前記状態情報を取得して確認し、前記所定領域に前記所定データを格納することを特徴とする請求項 1 又は 2 に記載の情報処理装置。

**【請求項 4】**

前記キャッシュメモリは、データ格納部、並びに、前記データの検索情報及び前記状態情報を保持するタグ格納部を有し、

前記判定部は、前記タグ格納部が保持する前記検索情報を基に前記データ格納部に前記所定データが存在するか否かを判定し、

前記制御部は、前記データ格納部に前記所定データが存在しない場合、前記下位のキャッシュメモリ又は記憶装置へ前記所定データの取得要求を出力するとともに、前記データ格納部における前記所定データの格納先領域の前記状態情報を前記タグ格納部から取得して前記判定用情報を生成して保持し、

前記格納処理部は、

前記状態情報を取得しないと判定した場合、前記格納先領域に関する情報を保持する前記タグ格納部の対応領域を無効化し、前記対応領域に前記所定データの検索情報及び前記格納先領域の状態情報を格納するタグ管理部と、

前記格納先領域に前記所定データを格納するデータ管理部とを有することを特徴とする請求項 1 ~ 3 のいずれか一つに記載の情報処理装置。

**【請求項 5】**

前記格納処理部は、前記応答がデータ移動を要求する種類の応答の場合、前記制御部が保持する前記判定用情報を基に、前記キャッシュメモリに格納された前記状態情報を取得するか否かを判定することを特徴とする請求項 1 ~ 4 のいずれか一つに記載の情報処理装置。

**【請求項 6】**

前記格納処理部は、前記制御部が保持する前記判定用情報が、前記所定データが前記下位のキャッシュメモリ又は記憶装置に存在する状態を表す場合、前記キャッシュメモリに

10

20

30

40

50

格納された前記状態情報を取得しないと判定することを特徴とする請求項 1 ~ 5 のいずれか一つに記載の情報処理装置。

【請求項 7】

前記制御部は、前記キャッシュメモリが保持する前記所定データの格納先の所定領域の状態情報及び前記所定領域に対するデータの更新状況に基づく判定用情報を保持することを特徴とする請求項 1 ~ 6 のいずれか一つに記載の情報処理装置。

【請求項 8】

データ及びデータを格納する領域の状態情報を保持するキャッシュメモリと、  
所定データの取得要求を受けて、前記キャッシュメモリに前記所定データが存在する  
否かを判定する判定部と、

10

前記キャッシュメモリが保持する前記所定データの格納先の所定領域の状態情報を基に  
判定用情報を生成する生成部と、

前記キャッシュメモリに前記所定データが存在しない場合、前記キャッシュメモリより  
も下位のキャッシュメモリ又は記憶装置へ前記所定データの取得要求を出力するとともに  
、前記生成部が生成した前記判定用情報を保持する制御部と、

前記取得要求に対する応答を前記下位のキャッシュメモリ又は記憶装置から取得し、取  
得した前記応答が所定の種類の場合、前記制御部が保持する前記判定用情報を基に、前記  
キャッシュメモリに格納された前記状態情報を取得するか否かを判定し、前記状態情報を  
取得しないと判定した場合、前記所定領域を無効化し、前記所定領域に前記応答に含まれ  
る前記所定データを格納する格納処理部と

20

を備えたことを特徴とする演算処理装置。

【請求項 9】

データ及びデータを格納する領域の状態情報を保持するキャッシュメモリが搭載された  
複数の演算処理装置を有する情報処理装置の制御方法であって、

所定データの取得要求を受けて、前記キャッシュメモリに前記所定データが存在する  
否かを判定し、

前記キャッシュメモリに前記所定データが存在しない場合、前記キャッシュメモリより  
も下位のキャッシュメモリ又は記憶装置へ前記所定データの取得要求を出力し、且つ、前  
記所定データの格納先の所定領域の状態情報に基づく判定用情報を保持し、

30

前記取得要求に対する応答を前記下位のキャッシュメモリ又は記憶装置から取得し、  
取得した前記応答が所定の種類の場合、前記判定用情報を基に前記キャッシュメモリに  
格納された前記状態情報を取得するか否かを判定し、

前記状態情報を取得しないと判定した場合、前記所定領域を無効化し、前記所定領域に  
前記応答に含まれる前記所定データを格納する

処理を各前記演算処理装置に行わせることを特徴とする情報処理装置の制御方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置、演算処理装置及び情報処理装置の制御方法に関する。

【背景技術】

40

【0002】

プロセッサ内部のキャッシュメモリは、主記憶メモリへのアクセスレイテンシを隠蔽す  
るために用いられる。プロセッサは、主記憶メモリに近い側に比較的大容量の 2 次キャッ  
シュ (L 2 (Level-2) キャッシュ) とコア内部に高速アクセスできる 1 次キャッシュ (L 1 (Level-1) キャッシュ) とを有する。このように、容量の異なる複数のキャッシュ  
を有する構成は、階層キャッシュと呼ばれる。

【0003】

ここで、1 次キャッシュでキャッシュミスが発生すると、2 次キャッシュに対してムー  
ブイン (M I :Move In) 要求が行われる。そして、2 次キャッシュからキャッシュミス  
したデータが 1 次キャッシュへ登録される。この時、ライトバック方式の階層キャッシュ

50

間において、データのコヒーレンスを維持することが要求される。

【0004】

ライトバック方式の階層キャッシュを有するプロセッサは、例えば、1次キャッシュ及び2次キャッシュにおいてMESI (Modified Exclusive Shared Invalid) プロトコルを採用し、階層間でのデータのコヒーレンスを維持する。MESIプロトコルとは、キャッシュステートと呼ばれるキャッシュメモリにおけるデータの状態を用いてコヒーレンスを維持するためのプロトコルである。キャッシュステートは、TAG - RAM (Random Access Memory) で管理されている。プロセッサは、MESIプロトコルを実行する場合、TAG - RAMにアクセスすることでキャッシュステートを監視する。

【0005】

MESIプロトコルにおいて、キャッシュは以下の4つの状態のいずれかにある。M (Modify) ステートは、ストア命令などによりキャッシュが更新されたことを示しており、更新されたデータについてライトバックが要求される状態である。E (Exclusive) ステートは、ストア命令を実行するためにキャッシュが排他権を有しており、ストア実行前であればライトバックは行わなくてよい状態である。ただし、ストア実行後であれば、キャッシュがEステートからMステートに遷移して、ライトバックを要求する状態となる。S (Share) ステートは、共有型でデータを保持しておりライトバックを行わなくてよい状態である。I (Invalid) ステートは、キャッシュのデータが無効であることを示しており、ライトバックを行わなくてよい状態である。

【0006】

1次キャッシュ管理部は、1次キャッシュでキャッシュミスが発生すると、データ登録に使用する物理アドレス及びリプレースWAY情報を2次キャッシュ管理部に通知する。2次キャッシュ管理部は、1次キャッシュのキャッシュステートがIステートであるか、又は、Mステート、SステートもしくはEステートのいずれかであるかを管理する。2次キャッシュ管理部は、1次キャッシュがIステートであれば、元データの移動が付随しないデータ格納要求を1次キャッシュ管理部に発行する。また、1次キャッシュがMステート、SステートもしくはEステートのいずれかであれば、2次キャッシュ管理部は、元データの移動が付随するデータ格納要求を1次キャッシュ管理部に対して発行する。この場合、データ格納要求には、物理アドレス、リプレースWAY情報及びコード情報が含まれる。そして、2次キャッシュ管理部は、データを読み出し、1次キャッシュへの新たなデータの転送を行う。

【0007】

1次キャッシュ管理部は、元データの移動が付随しないデータ格納要求を受信した場合、1次キャッシュのリプレース対象部録の状態はIステートであるので、無効化処理を行わずにプロトコルチェックを行う。これに対して、元データの移動が付随するデータ格納要求を受信した場合、1次キャッシュのリプレース対象ブロックは有効状態であるので、1次キャッシュ管理部は、1次キャッシュのTAG - RAMを1度読み出してキャッシュステートを確認する。1次キャッシュがMステートであれば、1次キャッシュ管理部は、ライトバックのために該当データを読み出して、MOB (Move Out Buffer) にデータを退避させつつ、リプレース対象ブロックの無効化を行う。これに対して、1次キャッシュがEステート又はSステートの場合、1次キャッシュ管理部は、ライトバックを行わずに、リプレース対象ブロックの無効化を行う。その後、1次キャッシュ管理部は、リプレース対象の無効化が行われた後に、2次キャッシュから転送された新たに格納するデータを1次キャッシュに登録してリプレース処理を終える。

【0008】

なお、キャッシュのコヒーレンスを維持する技術として、以下の従来技術がある。上位キャッシュのラインに下位キャッシュにおけるデータの存在の有無を示すフラグを設け、下位キャッシュでラインの置換えが発生した場合に上位キャッシュに通知し、ラインが上位キャッシュに存在する場合、フラグをセットする従来技術がある。また、ダーティフラグがセットされたキャッシュエントリのラインデータをメモリへライトバックする従来技

10

20

30

40

50

術がある。

【先行技術文献】

【特許文献】

【0009】

【特許文献1】特開平10-55309号公報

【特許文献2】国際公開第2005/050454号

【発明の概要】

【発明が解決しようとする課題】

【0010】

しかしながら、2次キャッシュ管理部から元データの移動が付随するデータ格納要求を受信した場合、1次キャッシュ管理部は、1回目のTAG-RAMへのアクセスで、TAG-RAMを読み出して、1次キャッシュのステートを確認する。次に、1次キャッシュ管理部は、2回目のTAG-RAMへのアクセスで、TAG-RAMの無効化を行う。このように、1次キャッシュ管理部は、2次キャッシュ管理部から元データの移動が付随するデータ格納要求を受信した場合、2回のTAG-RAMへのアクセスを行う。TAG-RAMへのアクセスはパイプライン処理であり、且つTAG-RAMへのアクセス時間にはステートの確認するためのサイクルが含まれる。そのため、TAG-RAMへアクセスする処理は時間が掛かる処理であり、TAG-RAMへのアクセス時間が、1次キャッシュミスの時の演算処理装置の処理性能を低減させるおそれがある。

10

【0011】

1次キャッシュがMステートの場合、ライトバックが行われることから、データの退避のためにTAG-RAMに対する2回のアクセスが行われる。ただし、1次キャッシュがEステート又はSステートの場合、最終的にTAG-RAMを無効化することができればよく、TAG-RAMへの2回のアクセスは無駄な遅延を発生させているといえ、情報処理装置の処理性能を低下させているといえる。

20

【0012】

ここで、上位キャッシュのラインに下位キャッシュにおけるデータの存在の有無を示すフラグを設ける従来技術を用いても、Mステート時のTAG-RAMへのアクセス回数には影響せず、情報処理装置の性能を向上させることは困難である。また、ダーティフラグを用いてメモリへのライトバックを制御する従来技術であっても、Mステート時のTAG-RAMへのアクセス回数は考慮されておらず、情報処理装置の処理性能を向上させることは困難である。

30

【0013】

開示の技術は、上記に鑑みてなされたものであって、処理性能を向上させる情報処理装置、演算処理装置及び情報処理装置の制御方法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】

【0014】

本願の開示する情報処理装置、演算処理装置及び情報処理装置の制御方法の一つの態様において、情報処理装置は、以下の各部を備える複数の演算処理装置を有する。キャッシュメモリは、データ及びデータを格納する領域の状態情報を保持する。判定部は、所定データの取得要求を受けて、前記キャッシュメモリに前記所定データが存在するか否かを判定する。制御部は、前記キャッシュメモリに前記所定データが存在しない場合、前記キャッシュメモリよりも下位のキャッシュメモリ又は記憶装置へ前記所定データの取得要求を出力するとともに、前記所定データを格納する所定領域の状態情報に基づく判定用情報を保持する。格納処理部は、前記取得要求に対する応答を前記下位のキャッシュメモリ又は記憶装置から取得し、取得した前記応答が所定の種類の場合、前記制御部が保持する前記判定用情報を基に、前記キャッシュメモリに格納された前記状態情報を取得するか否かを判定し、前記状態情報を取得しないと判定した場合、前記所定領域を無効化し、前記所定領域に前記応答に含まれる前記所定データを格納する。

40

【発明の効果】

50

## 【 0 0 1 5 】

1つの側面では、本発明は、情報処理装置の処理性能を向上させることができる。

## 【 図面の簡単な説明 】

## 【 0 0 1 6 】

【 図 1 】 図 1 は、情報処理装置のハードウェア構成図である。

【 図 2 】 図 2 は、CPUのブロック図である。

【 図 3 】 図 3 は、実施例 1 に係る 1 次キャッシュ管理部のブロック図である。

【 図 4 】 図 4 は、1 次キャッシュ管理部によるロード命令の処理のフローチャートである。

【 図 5 】 図 5 は、移動不要格納要求の場合の無効化処理のフローチャートである。

10

【 図 6 】 図 6 は、移動付随格納要求の場合の無効化処理のフローチャートである。

【 図 7 】 図 7 は、実施例 2 に係る 1 次キャッシュ管理部のブロック図である。

【 図 8 】 図 8 は、ストア命令が先行する状態での競合時の L 1 タグの遷移を表す図である。

【 図 9 】 図 9 は、ロード命令が先行する状態での競合時の L 1 タグの遷移を表す図である。

【 図 1 0 】 図 1 0 は、ストアバッファを用いない場合の競合時の L 1 タグの遷移を表す図である。

## 【 発明を実施するための形態 】

## 【 0 0 1 7 】

20

以下に、本願の開示する情報処理装置、演算処理装置及び情報処理装置の制御方法の実施例を図面に基づいて詳細に説明する。なお、以下の実施例により本願の開示する情報処理装置、演算処理装置及び情報処理装置の制御方法が限定されるものではない。

## 【 実施例 1 】

## 【 0 0 1 8 】

図 1 は、情報処理装置のハードウェア構成図である。サーバなどの情報処理装置 1 は、システムボード 1 0、I O ( Input Output ) ユニット 2 0、ディスクユニット 3 0 を有する。

## 【 0 0 1 9 】

システムボード 1 0 は、複数の CPU 1 1 及びメモリ 1 2 が搭載される。システムボード 1 0 は、バスにより I O ユニット 2 0 と接続される。システムボード 1 0 上の CPU 1 1 及びメモリ 1 2 は、システムボード 1 0 と I O ユニット 2 0 とを繋ぐバスを介して I O ユニット 2 0 と通信する。

30

## 【 0 0 2 0 】

演算処理装置である CPU 1 1 は、I O ユニット 2 0 を介してディスクユニット 3 0 とデータの送受信を行う。また、CPU 1 1 は、バスを介して主記憶装置であるメモリ 1 2 と接続される。CPU 1 1 は、ディスクユニット 3 0 に格納された各種プログラムを読み出し、メモリ 1 2 上に展開して実行することで、OS ( Operating System ) やアプリケーションを動作させる。ここで、本実施例では、4つの CPU 1 1 を例に図示したが、CPU 1 1 の数に特に制限は無い。

40

## 【 0 0 2 1 】

I O ユニット 2 0 は、ディスクユニット 3 0 に対するデータの読み書きの制御を行う。ディスクユニット 3 0 は、補助記憶装置である。ディスクユニット 3 0 は、OS やアプリケーションなどの各種プログラムを格納する。

## 【 0 0 2 2 】

図 2 は、CPUのブロック図である。CPU 1 1 は、コア 1 0 0 及び 2 次キャッシュ管理部 1 1 2 を有する。

## 【 0 0 2 3 】

コア 1 0 0 は、命令制御部 1 0 1、演算処理部 1 0 2 及び 1 次キャッシュ管理部 1 0 3 を有する。命令制御部 1 0 1、演算処理部 1 0 2 及び 1 次キャッシュ管理部 1 0 3 は、そ

50

れぞれ互いに接続され、相互に通信を行う。

【 0 0 2 4 】

命令制御部 1 0 1 は、プログラムを実行する際に、各種命令を生成する。そして、命令制御部 1 0 1 は、演算処理の実行命令を演算処理部 1 0 2 へ出力する。その後、命令制御部 1 0 1 は、出力した実行命令の応答を演算処理部 1 0 2 から取得する。

【 0 0 2 5 】

また、命令制御部 1 0 1 は、キャッシュのロード命令又はストア命令を 1 次キャッシュ管理部 1 0 3 に送信する。その後、命令制御部 1 0 1 は、送信したロード命令又はストア命令に対する応答を 1 次キャッシュ管理部 1 0 3 から取得する。

【 0 0 2 6 】

演算処理部 1 0 2 は、演算処理の実行命令の入力を命令制御部 1 0 1 から受ける。また、演算処理部 1 0 2 は、1 次キャッシュ管理部 1 0 3 からデータを取得する。そして、演算処理部 1 0 2 は、1 次キャッシュ管理部 1 0 3 から取得したデータなどを用いて実行命令にしたがって演算を実行する。

【 0 0 2 7 】

1 次キャッシュ管理部 1 0 3 は、1 次キャッシュ 3 0 1 を有する。1 次キャッシュ管理部 1 0 3 は、キャッシュのロード命令の入力を命令制御部 1 0 1 から受ける。そして、1 次キャッシュ管理部 1 0 3 は、ロード命令で指定されたデータが 1 次キャッシュ 3 0 1 に存在するか否かを判定する。1 次キャッシュ 3 0 1 に指定されたデータが存在するキャッシュヒットの場合、1 次キャッシュ管理部 1 0 3 は、指定されたデータを 1 次キャッシュ 3 0 1 から読み出して演算処理部 1 0 2 へ出力する。その後、1 次キャッシュ管理部 1 0 3 は、キャッシュのロード命令の処理完了を命令制御部 1 0 1 へ通知する。

【 0 0 2 8 】

これに対して、ロード命令で指定されたデータが 1 次キャッシュ 3 0 1 に存在しないキャッシュミスの場合、1 次キャッシュ管理部 1 0 3 は、ムーブイン要求を 2 次キャッシュ管理部 1 1 2 に送信する。その後、1 次キャッシュ管理部 1 0 3 は、ロード命令で指定されたデータ及びデータ格納要求の入力を 2 次キャッシュ管理部 1 1 2 から受ける。そして、1 次キャッシュ管理部 1 0 3 は、データ格納要求にしたがい 1 次キャッシュ 3 0 1 に対する処理を行い、受信したデータを 1 次キャッシュ 3 0 1 に格納する。その後、1 次キャッシュ管理部 1 0 3 は、キャッシュのロード命令を再度実行し、キャッシュヒットとなった場合の処理を実行する。その後、1 次キャッシュ管理部 1 0 3 は、キャッシュのロード命令の処理完了を命令制御部 1 0 1 へ通知する。

【 0 0 2 9 】

また、1 次キャッシュ管理部 1 0 3 は、キャッシュのストア命令の入力を命令制御部 1 0 1 から受ける。そして、1 次キャッシュ管理部 1 0 3 は、ストア命令で指定されたデータが 1 次キャッシュ 3 0 1 に存在するか否かを判定する。ストア命令で指定されたデータが 1 次キャッシュ 3 0 1 に存在しない場合、1 次キャッシュ管理部 1 0 3 は、データを新たに 1 次キャッシュ 3 0 1 に登録する。さらに、1 次キャッシュ管理部 1 0 3 は、登録したデータを保持しておき、予め決められた条件を満たすと、2 次キャッシュ 1 2 1 へのデータの書き込みを 2 次キャッシュ管理部 1 1 2 へ依頼する。また、ストア命令で指定されたデータが 1 次キャッシュ 3 0 1 に存在する場合、1 次キャッシュ管理部 1 0 3 は、1 次キャッシュ 3 0 1 上のデータを更新する。さらに、1 次キャッシュ管理部 1 0 3 は、更新したデータを保持しておき、予め決められた条件を満たすと、2 次キャッシュ 1 2 1 へのデータの書き込みを 2 次キャッシュ管理部 1 1 2 へ依頼する。

【 0 0 3 0 】

2 次キャッシュ管理部 1 1 2 は、2 次キャッシュ 1 2 1 を有する。2 次キャッシュ 1 2 1 は、1 次キャッシュよりも動作速度が低速であるが、容量が大きいキャッシュである。2 次キャッシュ 1 2 1 は、1 次キャッシュ 3 0 1 の各領域のキャッシュステートがイステートであるか否かの情報を保持する。

【 0 0 3 1 】

10

20

30

40

50

2次キャッシュ管理部112は、ムーブイン要求の入力を1次キャッシュ管理部103から受ける。そして、2次キャッシュ管理部112は、ムーブイン要求で指定されたデータが2次キャッシュ121に存在するか否かを判定する。

【0032】

指定されたデータが2次キャッシュに存在する場合、2次キャッシュ管理部112は、1次キャッシュ301の指定されたデータの格納先の領域のキャッシュステートを確認する。キャッシュステートがIステートである場合、2次キャッシュ管理部112は、元データの移動を行わないデータ格納要求を生成する。以下では、データ登録における元データの移動を行わないデータ格納要求を「移動不要格納要求」と言う。その後、2次キャッシュ管理部112は、移動不要格納要求とともにムーブイン要求で指定されたデータを1次キャッシュ管理部103へ出力する。

10

【0033】

これに対して、キャッシュステートがIステート以外であれば、2次キャッシュ管理部112は、元データの移動を伴うデータ格納要求を生成する。以下では、元データの移動を伴うデータ格納要求を「移動付随格納要求」という。その後、2次キャッシュ管理部112は、移動付随格納要求とともにムーブイン要求で指定されたデータを1次キャッシュ管理部103へ出力する。

【0034】

これに対して、指定されたデータが2次キャッシュに存在しない場合、2次キャッシュ管理部112は、その指定されたデータをメモリ12から取得する。そして、2次キャッシュ管理部112は、取得したデータを2次キャッシュ121に格納する。その後、2次キャッシュ管理部112は、命令で指定されたデータが2次キャッシュに存在する場合と同様の処理を実行する。この2次キャッシュ管理部112が、「下位記憶装置」の一例にあたる。

20

【0035】

次に、図3を参照して、1次キャッシュ管理部103の詳細について説明する。図3は、実施例1に係る1次キャッシュ管理部のブロック図である。図3に示すように、1次キャッシュ管理部103は、1次キャッシュ301、セクタ302、タグマッチ判定部303、TLB(Translation Lookaside Buffer)304を有する。さらに、1次キャッシュ管理部103は、キャッシュの書き換えを管理するムーブインバッファ305、リブレイスウェイ選択部306、フラグ生成部307、格納処理部308及びムーブアウトバッファ309を有する。

30

【0036】

1次キャッシュ301は、1次データキャッシュ351及びL1タグ352を有する。1次データキャッシュ351は、データを格納する格納領域である。この1次データキャッシュ351が、「データ格納部」の一例にあたる。また、L1タグ352は、1次データキャッシュ351に格納されたデータに対応するアドレス情報であるタグを格納する格納領域である。このL1タグ352が、「タグ格納部」の一例にあたる。そして、タグが、「検索情報」の一例にあたる。1次キャッシュ301に格納されたデータとそのデータに対応するタグのペアが、キャッシュラインである。

40

【0037】

本実施例では、1次キャッシュ301は、セットアソシアティブ方式を採用し、インデックスごとにブロックにまとめられており、ブロックを格納できる領域数をウェイという。1次キャッシュ301では、4ウェイのデータが同一インデックスに格納できる。以下では、1次データキャッシュ351における各データを格納した領域をブロックとして説明する。

【0038】

また、L1タグ352は、キャッシュラインに有効なデータが入っているかを示すvalid情報及び1次データキャッシュ351の各ブロックのデータ保持の状態を表すキャッシュステートなどの制御情報を保持する。

50

## 【 0 0 3 9 】

ここで、L 1 タグ 3 5 2 が保持するキャッシュステートは、M E S I プロトコルでキャッシュコヒーレンシの制御に用いられる情報である。キャッシュステートには、M ステート、E ステート、S ステート及びI ステートが存在する。M ステートは、変更型でブロックを保持する状態を表す。また、E ステートは、排他型でブロックを保持する状態を表す。S ステートは、共有型でブロックを保持する状態を表す。I ステートは、ブロックを保持していない状態を表す。このキャッシュステートが、「状態情報」の一例にあたる。

## 【 0 0 4 0 】

1 次キャッシュ 3 0 1 の 1 次データキャッシュ 3 5 1 及び L 1 タグ 3 5 2 は、命令制御部 1 0 1 から出力されたキャッシュのロード命令の入力を受ける。L 1 タグ 3 5 2 は、キャッシュのロード命令の入力を受けて、ロード命令で指定されたデータを格納するウェイを表すタグをタグマッチ判定部 3 0 3 及びフラグ生成部 3 0 7 へ出力する。

10

## 【 0 0 4 1 】

1 次データキャッシュ 3 5 1 は、後述するタグマッチ判定部 3 0 3 によりタグマッチが確認された場合、マッチしたタグに対応するデータをセレクト 3 0 2 を介して演算処理部 1 0 2 へ出力する。

## 【 0 0 4 2 】

T L B 3 0 4 は、例えば、論理アドレスから物理アドレスを取得するための連想メモリを有する。T L B 3 0 4 は、命令制御部 1 0 1 から出力されたキャッシュのロード命令の中から、ロードするデータの格納先の論理アドレスを取得する。そして、T L B 3 0 4 は、取得した論理アドレスに対応する物理アドレスを取得する。次に、T L B 3 0 4 は、取得した物理アドレスをタグマッチ判定部 3 0 3 へ出力する。その後、タグマッチ判定部 3 0 3 からタグ mismatches の通知を受けた場合、T L B 3 0 4 は、ムーブインバッファ 3 0 5 へ取得した物理アドレスを出力する。

20

## 【 0 0 4 3 】

タグマッチ判定部 3 0 3 は、タグの情報の入力を L 1 タグ 3 5 2 から受ける。また、タグマッチ判定部 3 0 3 は、物理アドレスの入力を T L B 3 0 4 から受ける。そして、タグマッチ判定部 3 0 3 は、タグの情報及び物理アドレスの情報を用いて、ロード命令で指定されたデータが 1 次データキャッシュ 3 5 1 に存在するか否かを確認し、存在する場合にはタグマッチと判定する。タグマッチの場合、タグマッチ判定部 3 0 3 は、タグマッチしたタグの情報をセレクト 3 0 2 に送る。

30

## 【 0 0 4 4 】

また、ロード命令で指定されたデータが 1 次データキャッシュ 3 5 1 に存在しない場合、タグマッチ判定部 3 0 3 は、タグ mismatches と判定する。そして、タグマッチ判定部 3 0 3 は、タグ mismatches を T L B 3 0 4 及びリプレイスウェイ選択部 3 0 6 に通知する。このタグマッチ判定部 3 0 3 が、「判定部」の一例にあたる。

## 【 0 0 4 5 】

セレクト 3 0 2 は、出力するデータのタグの情報の入力をタグマッチ判定部 3 0 3 から受ける。そして、セレクト 3 0 2 は、取得したタグに対応するデータを 1 次データキャッシュ 3 5 1 の中から選択して演算処理部 1 0 2 へ出力する。

40

## 【 0 0 4 6 】

リプレイスウェイ選択部 3 0 6 は、命令制御部 1 0 1 から発行されたキャッシュのロード命令の入力を受ける。リプレイスウェイ選択部 3 0 6 は、タグマッチ判定部 3 0 3 からタグ mismatches の通知を受けると、予め決められたウェイの選択アルゴリズムにしたがって新たなデータを格納するウェイを選択する。例えば、リプレイスウェイ選択部 3 0 6 は、1 次キャッシュ 3 0 1 の中でデータが格納されたタイミングが最も古いウェイをリプレイスウェイとして選択する。そして、リプレイスウェイ選択部 3 0 6 は、選択したウェイの情報をムーブインバッファ 3 0 5 及びフラグ生成部 3 0 7 へ出力する。以下では、このリプレイスウェイ選択部 3 0 6 が選択したロード命令で指定されたデータの格納先とするウェイを「格納先ウェイ」と言う。

50

## 【 0 0 4 7 】

フラグ生成部 3 0 7 は、格納先ウェイの情報をリブレイスウェイ選択部 3 0 6 から取得する。そして、フラグ生成部 3 0 7 は、取得した格納先ウェイの情報をを用いて格納先ウェイに含まれるブロックのキャッシュステートを L 1 タグ 3 5 2 から取得する。

## 【 0 0 4 8 】

次に、フラグ生成部 3 0 7 は、取得したキャッシュステートが M ステートである場合、そのウェイに対応する判定用フラグの値として M ステートを表す 1 を設定する。判定用フラグは、値が 1 の場合、L 1 タグ 3 5 2 に対するキャッシュステートの確認の実行を表す。

## 【 0 0 4 9 】

これに対して、取得したキャッシュステートが M ステート以外の場合、フラグ生成部 3 0 7 はそのウェイに対応する判定用フラグの値として M ステート以外を表す 0 を設定する。判定用フラグは、値が 0 の場合、L 1 タグ 3 5 2 に対するキャッシュステートの確認を行わないことを表す。その後、フラグ生成部 3 0 7 は、格納先ウェイの判定用フラグをムーブインバッファ 3 0 5 へ出力する。

## 【 0 0 5 0 】

ムーブインバッファ 3 0 5 は、キャッシュのロード命令で指定されたデータの物理アドレスの入力を T L B 3 0 4 から受ける。また、ムーブインバッファ 3 0 5 は、格納先ウェイの情報の入力をリブレイスウェイ選択部 3 0 6 から受ける。さらに、ムーブインバッファ 3 0 5 は、格納先ウェイの判定用フラグの入力をフラグ生成部 3 0 7 から受ける。そして、ムーブインバッファ 3 0 5 は、キャッシュのロード命令で指定されたデータの物理アドレス及び格納先ウェイの情報を含むムーブイン要求を 2 次キャッシュ管理部 1 1 2 へ出力する。このムーブインバッファ 3 0 5 が、「制御部」の一例にあたる。そして、判定用フラグが、「判定用情報」の一例にあたる。さらに、ムーブイン要求が、「取得要求」の一例にあたる。

## 【 0 0 5 1 】

格納処理部 3 0 8 は、タグ管理部 3 8 1 及びデータ管理部 3 8 2 を有する。タグ管理部 3 8 1 は、ムーブインバッファ 3 0 5 から出力されたムーブイン要求に対する応答であるデータ格納要求及びロード命令で指定されたデータの入力を 2 次キャッシュ管理部 1 1 2 から受ける。以下では、ロード命令で指定されたデータを「対象データ」という。そして、格納処理部 3 0 8 は、取得したデータ格納要求が移動不要格納要求か移動付随格納要求かを判定する。

## 【 0 0 5 2 】

取得したデータ格納要求が移動不要格納要求の場合、タグ管理部 3 8 1 は、格納先ウェイにおける対象データを格納するブロックのキャッシュステートを L 1 タグ 3 5 2 から取得する。そして、タグ管理部 3 8 1 は、キャッシュステートが I ステートであることを確認してプロトコル処理を実行した後、データ管理部 3 2 8 にデータの格納を依頼する。その後、タグ管理部 3 8 1 は、1 次データキャッシュ 3 5 1 に新たに格納された対象データを表すタグを生成して L 1 タグ 3 5 2 に登録し、且つ、対象データを格納したウェイのブロックのキャッシュステートとして S ステートを登録する。

## 【 0 0 5 3 】

一方、取得したデータ格納要求が移動付随格納要求の場合、タグ管理部 3 8 1 は、格納先ウェイの判定用フラグの値をムーブインバッファ 3 0 5 から取得する。この移動付随格納要求が、「所定の種類の応答」及び「データ移動を要求する種類の応答」の一例にあたる。そして、タグ管理部 3 8 1 は、判定用フラグの値が 1 の場合、キャッシュステートが M ステートである可能性があるため、格納先ウェイにおける対象データを格納するブロックのキャッシュステートを L 1 タグ 3 5 2 から取得する。そして、キャッシュステートが M ステートである場合、タグ管理部 3 8 1 は、データの退避をデータ管理部 3 8 2 に指示する。

## 【 0 0 5 4 】

その後データの退避が完了すると、タグ管理部 381 は、格納先ウェイにおける対象データを格納するブロックの L1 タグ 352 を無効化する。次に、タグ管理部 381 は、2 次キャッシュ管理部 112 から取得した対象データを格納先ウェイにおける格納先のブロックに格納する。さらに、タグ管理部 381 は、格納した対象データを表すタグを生成して L1 タグ 352 に登録し、且つ、対象データを格納したウェイのブロックのキャッシュステートとして S ステートを登録する。

【0055】

これに対して、判定用フラグの値が 0 の場合にはキャッシュステートが M ステートではないので、タグ管理部 381 は、格納先ウェイにおける対象データを格納するブロックの L1 タグ 352 を無効化する。この判定用フラグの値が 0 の場合が、「判定用情報が、所定データが下位のキャッシュメモリ又は記憶装置に存在する状態を表す場合」の一例にあたる。そして、タグ管理部 381 は、データ管理部 328 にデータの格納を依頼する。その後、タグ管理部 381 は、格納した対象データを表すタグを生成して L1 タグ 352 に登録し、且つ、対象データを格納したウェイのブロックのキャッシュステートとして S ステートを登録する。

10

【0056】

データ管理部 382 は、データの退避の指示をタグ管理部 381 から受ける。そして、データ管理部 382 は、データ格納要求で指定された対象データの格納先のブロックの既存のデータを 1 次データキャッシュ 351 から読み出してムーブアウトバッファ 309 に退避させる。

20

【0057】

また、データ管理部 382 は、データの格納の依頼をタグ管理部 381 から受ける。そして、データ管理部 382 は、2 次キャッシュ管理部 112 から取得した対象データを格納先ウェイに格納する。

【0058】

ムーブアウトバッファ 309 は、格納先ウェイにおける対象データを格納するブロックに存在した既存データの入力をデータ管理部 382 から受ける。その後、ムーブアウトバッファ 309 は、所定条件が満たされ 1 次キャッシュ管理部 103 により 2 次キャッシュ 121 へのライトバックが行われるまで、取得した既存データを保持する。

【0059】

30

次に、図 4 を参照して、本実施例に係る 1 次キャッシュ管理部 103 によるロード命令の処理の流れについて説明する。図 4 は、1 次キャッシュ管理部によるロード命令の処理のフローチャートである。

【0060】

1 次キャッシュ管理部 103 は、命令制御部 101 から発行されたロード命令を受信する（ステップ S1）。

【0061】

1 次キャッシュ管理部 103 は、L1 タグ 352 に格納されたタグの情報及び TLB 304 により特定された物理アドレスを用いてキャッシュヒットか否かを判定する（ステップ S2）。

40

【0062】

キャッシュヒットの場合（ステップ S2：肯定）、1 次キャッシュ管理部 103 は、取得したタグに対応するデータを 1 次データキャッシュ 351 から選択し、選択したデータを演算処理部 102 へ出力する（ステップ S3）。

【0063】

これに対して、キャッシュミスの場合（ステップ S2：否定）、1 次キャッシュ管理部 103 は、ムーブイン要求を 2 次キャッシュ管理部 112 へ送信する（ステップ S4）。

【0064】

その後、1 次キャッシュ管理部 103 は、データ格納要求を 2 次キャッシュ管理部 112 から受信する（ステップ S5）。

50

## 【 0 0 6 5 】

次に、1次キャッシュ管理部103は、受信したデータ格納要求が移動付随格納要求か否かを判定する(ステップS6)。

## 【 0 0 6 6 】

受信したデータ格納要求が移動付随格納要求でない、すなわち、移動不要格納要求の場合(ステップS6:否定)、1次キャッシュ管理部103は、データ置換なしの無効化処理を実行する(ステップS7)。

## 【 0 0 6 7 】

これに対して、移動付随格納要求の場合(ステップS6:肯定)、1次キャッシュ管理部103は、データ置換の可能性のある無効化処理を実行する(ステップS8)。

10

## 【 0 0 6 8 】

その後、1次キャッシュ管理部103は、データ格納要求に含まれるデータを1次キャッシュ301に登録し(ステップS9)、ステップS1へ戻る。

## 【 0 0 6 9 】

次に、図5を参照して、移動不要格納要求の場合の無効化処理の流れについて説明する。図5は、移動不要格納要求の場合の無効化処理のフローチャートである。図5に示すフローチャートは、図4におけるステップS7で実行される処理の一例にあたる。

## 【 0 0 7 0 】

格納処理部308のタグ管理部381は、2次キャッシュ管理部112から発行された移動不要格納要求を受理する(ステップS101)。

20

## 【 0 0 7 1 】

次に、タグ管理部381は、移動不要格納要求から対象データの格納先ウェイを取得する。そして、タグ管理部381は、格納先ウェイにおけるデータを格納するブロックのキャッシュステートをL1タグ352から読み出す(ステップS102)。

## 【 0 0 7 2 】

タグ管理部381は、読み出したキャッシュステートがIステートであることを確認してプロトコルチェックを行う。その後、タグ管理部381は、無効化処理の完了を格納処理部308のデータ管理部382に通知する(ステップS103)。

## 【 0 0 7 3 】

次に、図6を参照して、移動付随格納要求の場合の無効化処理の流れについて説明する。図6は、移動付随格納要求の場合の無効化処理のフローチャートである。図6に示すフローチャートは、図4におけるステップS8で実行される処理の一例にあたる。

30

## 【 0 0 7 4 】

格納処理部308のタグ管理部381は、2次キャッシュ管理部112から発行された移動付随格納要求を受理する(ステップS201)。

## 【 0 0 7 5 】

次に、タグ管理部381は、移動付随格納要求で指定された格納先ウェイに対応する判定用フラグをムーブインバッファ305から読み出す(ステップS202)。

## 【 0 0 7 6 】

そして、タグ管理部381は、読み出した判定用フラグの値が1か否かを判定する(ステップS203)。判定用フラグが0の場合(ステップS203:否定)、格納処理部308は、処理をステップS208へ進ませる。

40

## 【 0 0 7 7 】

これに対して、判定用フラグが1の場合(ステップS203:肯定)、タグ管理部381は、移動不要格納要求で指定された格納先ウェイにおけるデータを格納するブロックのキャッシュステートをL1タグ352から読み出す(ステップS204)。

## 【 0 0 7 8 】

次に、タグ管理部381は、読み出したキャッシュステートを用いてMESIのプロトコルチェックを行う(ステップS205)。

## 【 0 0 7 9 】

50

そして、タグ管理部 381 は、読み出したキャッシュステートが M ステートか否かを判定する (ステップ S 206)。読み出したキャッシュステートが M ステートでない場合 (ステップ S 206: 否定)、格納処理部 308 は、処理をステップ S 208 へ進ませる。

【0080】

これに対して、読み出したキャッシュステートが M ステートの場合 (ステップ S 206: 肯定)、タグ管理部 381 は、格納先ウェイの対象データを格納するブロックに存在するリプレースデータをムーブアウトバッファ 309 へ退避させる (ステップ S 207)。

【0081】

その後、タグ管理部 381 は、L1 タグ 352 の中の格納先ウェイの対象データを格納するブロックに対応するタグを無効化する (ステップ S 208)。

10

【0082】

そして、タグ管理部 381 は、無効化処理の完了をデータ管理部 382 へ通知する (ステップ S 209)。

【0083】

以上に説明したように、本実施例に係る 1 次キャッシュ管理部は、キャッシュミスが発生した場合に、2 次キャッシュから送られたデータの格納先となるブロックのキャッシュステートが M ステートか否かを表す判定用フラグを設定して、ムーブイン要求を行う。その後、1 次キャッシュ管理部は、ムーブイン要求に対する応答であるデータ格納要求を受信したときに、保持した判定用フラグを確認する。そして、判定用フラグが M ステート以外を表していれば、1 次キャッシュ管理部は、L1 タグに格納されたキャッシュステートの確認を行わずに、データを格納するブロックを無効化する。これにより、1 次キャッシュ管理部は、キャッシュミスの際に L1 タグに対してアクセスする回数を低減することができ、応答処理にかかる時間を短縮することができる。したがって、情報処理装置における 1 次キャッシュミス時の処理性能を向上させることができる。

20

【実施例 2】

【0084】

図 7 は、実施例 2 に係る 1 次キャッシュ管理部のブロック図である。本実施例に係る 1 次キャッシュ管理部 103 は、図 2 に示した実施例 1 の各部に加えて、ストア実行部 310、ストアバッファ 311 及び判定部 312 を有する。本実施例に係る 1 次キャッシュ管理部 103 は、ストア命令の実行状態を考量して判定用フラグを設定することが実施例 1 と異なる。以下の説明では、実施例 1 と同様の各部の機能については説明を省略する。

30

【0085】

L1 タグ 352 は、キャッシュのストア命令を命令制御部 101 から受ける。そして、L1 タグ 352 は、ストア命令で指定されたストアするデータのタグの情報をタグマッチ判定部 303 へ出力する。

【0086】

T L B 304 は、キャッシュのストア命令を命令制御部 101 から受ける。そして、T L B 304 は、ストア命令で指定された論理アドレスを物理アドレスに変換してタグマッチ判定部 303 及びストアバッファ 31 へ出力する。

【0087】

40

タグマッチ判定部 303 は、ストアするデータのタグの情報の入力を L1 タグ 352 から受ける。また、タグマッチ判定部 303 は、物理アドレスの入力を T L B 304 から受ける。そして、タグマッチ判定部 303 は、取得したタグ及び物理アドレスを用いてタグマッチの判定を行う。

【0088】

タグマッチの場合、タグマッチ判定部 303 は、タグマッチしたブロックのウェイの情報及びキャッシュヒットしたことを表す情報をストアバッファ 311 へ出力する。さらに、タグマッチ判定部 303 は、ストア実行部 310 にタグマッチを通知する。

【0089】

ストア実行部 310 は、ストア命令の入力を命令制御部 101 から受ける。そして、タ

50

グがマッチしなかった場合、ストア実行部 3 1 0 は、タグミスの通知をタグマッチ判定部 3 0 3 から受ける。この場合、ストア実行部 3 1 0 は、1 次データキャッシュ 3 5 1 からデータを格納するブロックを選択して、選択したブロックにストア命令で指定されたデータを格納する。また、ストア実行部 3 1 0 は、格納したデータを表すタグ、有効を表す `valid` 情報及びデータの格納先のブロックのキャッシュステートを L 1 タグ 3 5 2 に登録する。

**【 0 0 9 0 】**

一方、タグマッチの場合、ストア実行部 3 1 0 は、タグマッチの通知とともにタグマッチしたウェイの情報の入力をタグマッチ判定部 3 0 3 から受ける。そして、ストア実行部 3 1 0 は、ムーブインバッファ 3 0 5 に同じ物理アドレス及びウェイの情報を有するエントリが存在し、そのエントリの判定用フラグが 0 か否かを確認する。

10

**【 0 0 9 1 】**

同じ物理アドレス及びウェイの情報を有するエントリが存在しない場合、ストア実行部 3 1 0 は、ストアバッファ 3 1 1 に格納されたその物理アドレス及びウェイの情報に対応するキャッシュヒットフラグを 1 に設定する。そして、ストア実行部 3 1 0 は、1 次データキャッシュ 3 5 1 及び L 1 タグ 3 5 2 に対してデータのストアを行う。

**【 0 0 9 2 】**

また、同じ物理アドレス及びウェイの情報を有するエントリが存在するが、判定用フラグが 1 の場合も同様に、ストア実行部 3 1 0 は、ストアバッファ 3 1 1 に格納されたその物理アドレス及びウェイの情報に対応するキャッシュヒットフラグを 1 に設定する。そして、ストア実行部 3 1 0 は、1 次データキャッシュ 3 5 1 及び L 1 タグ 3 5 2 に対してデータのストアを行う。

20

**【 0 0 9 3 】**

これに対して、同じ物理アドレス及びウェイの情報を有するエントリが存在し、且つ、判定用フラグが 0 の場合、ストア実行部 3 1 0 は、ストアバッファが格納するその物理アドレス及びウェイの情報に対応するキャッシュヒットフラグを 0 に設定する。そして、ストア実行部 3 1 0 は、ストア処理の実行を待機する。その後、ストア実行部 3 1 0 は、ムーブインバッファ 3 0 5 の確認を繰り返えし、同じ物理アドレス及びウェイの情報を有するエントリが存在しない、または、判定用フラグが 0 となった後にストア処理を実行する。キャッシュヒットフラグが、「所定領域に対するデータの更新状況」を表す情報の一例にあたる。

30

**【 0 0 9 4 】**

ストアバッファ 3 1 1 は、物理アドレスの情報の入力を T L B 3 0 4 から受ける。また、ストアバッファ 3 1 1 は、タグマッチしたデータのウェイの情報及びキャッシュヒットを表す情報をタグマッチ判定部 3 0 3 から受ける。さらに、ストアバッファ 3 1 1 は、ストア実行部 3 1 0 からの指示にしたがい、アドレス情報及びウェイの情報に対応するキャッシュヒットフラグを設定する。

**【 0 0 9 5 】**

すなわち、ストアバッファ 3 1 1 のキャッシュヒットフラグは、既に同じ物理アドレス及びウェイに対するムーブイン要求が発行されており、且つ、判定用フラグが 0 であれば、0 に設定される。また、ストアバッファ 3 1 1 のキャッシュヒットフラグは、同じ物理アドレス及びウェイに対するムーブイン要求が発行されていない場合、又は、同じ物理アドレス及びウェイに対するムーブイン要求が発行されているが判定用フラグが 1 であれば、1 に設定される。

40

**【 0 0 9 6 】**

判定部 3 1 2 は、ロード命令が命令制御部 1 0 1 から出力されタグミスとなった場合、ムーブインバッファ 3 0 5 に格納された物理アドレス及び格納先ウェイの情報を取得する。次に、判定部 3 1 2 は、取得した物理アドレス及び格納先ウェイと同一の物理アドレス及び格納先ウェイのエントリがストアバッファ 3 1 1 に格納されているか否かを判定する。同一の物理アドレス及び格納先ウェイのエントリがストアバッファ 3 1 1 に格納されて

50

いない場合、判定部 3 1 2 は、フラグ設定可能の通知をフラグ生成部 3 0 7 に行う。

【 0 0 9 7 】

これに対して、同一の物理アドレス及び格納先ウェイのエントリがストアバッファ 3 1 1 に格納されている場合、判定部 3 1 2 は、そのエントリのキャッシュヒットフラグを確認する。キャッシュヒットフラグが 1 であれば、判定部 3 1 2 は、フラグ設定禁止の通知をフラグ生成部 3 0 7 に行う。

【 0 0 9 8 】

フラグ生成部 3 0 7 は、格納先ウェイの情報をリプレイスウェイ選択部 3 0 6 から取得する。そして、フラグ生成部 3 0 7 は、取得した格納先ウェイの情報をを用いて格納先ウェイに含まれるブロックのキャッシュステートを L 1 タグ 3 5 2 から取得する。

10

【 0 0 9 9 】

取得したキャッシュステートが M ステートである場合、フラグ生成部 3 0 7 は、格納先ウェイに対応する判定用フラグを 1 に設定する。

【 0 1 0 0 】

一方、取得したキャッシュステートが M ステート以外の場合、フラグ生成部 3 0 7 は、判定部 3 1 1 からフラグ設定可能の通知を受けている場合、そのウェイに対応する判定用フラグを 0 に設定する。これに対して、判定部 3 1 1 からフラグ設定禁止の通知を受けている場合、フラグ生成部 3 0 7 は、そのウェイに対応する判定用フラグを 1 に設定する。その後、フラグ生成部 3 0 7 は、判定用フラグに対応するウェイの情報とともにムーブインバッファ 3 0 5 へ出力する。

20

【 0 1 0 1 】

格納処理部 3 0 8 は、実施例 1 と同様の処理を行う。具体的には、取得したデータ格納要求が移動付随格納要求の場合、タグ管理部 3 8 1 は、格納先ウェイの判定用フラグの値をムーブインバッファ 3 0 5 から取得する。そして、タグ管理部 3 8 1 は、判定用フラグの値が 1 の場合、格納先ウェイにおける対象データを格納するブロックのキャッシュステートを L 1 タグ 3 5 2 から取得する。そして、キャッシュステートが M ステートである場合、タグ管理部 3 8 1 は、データの退避をデータ管理部 3 8 2 に指示する。データ管理部 3 8 2 によるデータの退避完了後、タグ管理部 3 8 1 は、格納先ウェイにおける対象データを格納するブロックの L 1 タグ 3 5 2 を無効化する。次に、タグ管理部 3 8 1 は、2 次キャッシュ管理部 1 1 2 から取得した対象データの格納をデータ管理部 3 8 2 に指示する。さらに、タグ管理部 3 8 1 は、1 次データキャッシュ 3 5 1 に格納された対象データを表すタグを生成して L 1 タグ 3 5 2 に登録し、且つ、対象データを格納したウェイのブロックのキャッシュステートとして S ステートを登録する。

30

【 0 1 0 2 】

これに対して、判定用フラグの値が 0 の場合にはキャッシュステートが M ステートではないので、タグ管理部 3 8 1 は、格納先ウェイにおける対象データを格納するブロックの L 1 タグ 3 5 2 を無効化する。その後、タグ管理部 3 8 1 は、2 次キャッシュ管理部 1 1 2 から取得した対象データの格納をデータ管理部 3 8 2 に指示する。さらに、タグ管理部 3 8 1 は、格納した対象データを表すタグを生成して L 1 タグ 3 5 2 に登録し、且つ、対象データを格納したウェイのブロックのキャッシュステートとして S ステートを登録する。

40

【 0 1 0 3 】

次に、図 8 及び 9 を参照して、本実施例に係る 1 次キャッシュ管理部 1 0 3 によるストア命令とロード命令とが競合した場合の L 1 タグ 3 5 2 の状態について説明する。図 8 は、ストア命令が先行する状態での競合時の L 1 タグの遷移を表す図である。図 9 は、ロード命令が先行する状態での競合時の L 1 タグの遷移を表す図である。図 8 及び 9 はいずれも図面に向かって下に進むにしたがい時間の経過を表す。処理 S はストア命令の処理を表す。また、処理 L はロード命令の処理を表す。

【 0 1 0 4 】

ストア命令が先行する場合、図 8 に示すように、時刻 T 0 で処理 S が開始される。この

50

場合、ムーブインバッファ305にはエントリが格納されていないので、ストアバッファ311には、物理アドレスを表すインデックスA及びウェイの情報としてw0が登録されるとともに、キャッシュヒットフラグは1に設定される。そして、状態201のL1タグ252においてインデックスA及びw0に対応するブロックはE状態であり、そのブロックに格納されたデータに対して処理Sによるストアが実行される。

**【0105】**

そして、処理Sが実行中の時刻T1において、ロード命令が発行され処理Lが開始される。このとき、ムーブインバッファ305には、インデックスA及びw0が登録される。そして、判定部312は、ストアバッファ311を確認して、ムーブインバッファ305に格納されたインデックスA及びw0を有するエントリを特定する。さらに、判定部312は、特定したエントリのキャッシュヒットフラグが1であることを確認して、フラグ設定禁止をフラグ生成部307に通知する。フラグ生成部307は、ムーブインバッファ305におけるインデックスA及びw0に対応する判定用フラグを1に設定する。

10

**【0106】**

処理Sが実行されることで、L1タグ252におけるインデックスA及びw0に対応するブロックに対してストアが行われ、そのブロックは状態202に示すようにE状態からM状態に遷移する。そして、時刻T2でストア命令が完了すると、ロード命令が開始される。この場合、ムーブインバッファ305に格納された判定用フラグは1であるので、タグ管理部381は、データ管理部382にインデックスA及びw0に対応するブロックのデータをムーブアウトバッファ309に退避させる。その後、タグ管理部381は、L1タグ352のインデックスA及びw0のタグを無効化する。これにより、L1タグ352は状態203となる。

20

**【0107】**

このように、本実施例に係る1次キャッシュ管理部103は、既にストア処理が実行されている状態で、ロード命令が発行された場合には、判定用フラグを1に設定して、M状態の場合のロード処理を実行する。これにより、ストア処理によりデータが更新されたE状態だったブロックがM状態に遷移した場合にも確実にデータの退避を行うことができる。

**【0108】**

次にロード命令が先行する場合について説明する。図9に示すように、L1タグ352は、状態211で示すように、インデックスA及びw0で表されるブロックはE状態である。そして、ストア命令が発行されていないので、ストアバッファ311は状態216のようにインデックスA及びw0のエントリを有さない。

30

**【0109】**

その後、時刻T3で処理Lが開始される。この場合、ムーブインバッファ305には、インデックスA及びw0が登録される。そして、判定部311は、インデックスA及びw0が登録されたエントリがストアバッファ311に存在しないことを確認する。そこで、判定部311は、フラグ生成部307にフラグ設定可能の通知を行う。フラグ生成部307は、フラグ設定可能の通知を受けて、状態214に示すようにムーブインバッファ305におけるインデックスA及びw0に対応する判定用フラグを1に設定する。

40

**【0110】**

その後、時刻T4で処理Sが開始される。この場合、ムーブインバッファ305には既にインデックスA及びw0が登録されているので、ストアバッファ311には、インデックスA及びw0が登録されるとともに、状態217に示すようにキャッシュヒットフラグは0に設定される。キャッシュヒットフラグが0であるため、ストア実行部310は、データのストアの実行を待機する。

**【0111】**

その後、処理LによりL1タグ352が有するキャッシュステートの確認を行わずにインデックスA及びw0に対応するタグが無効化される。そして、時刻T5で処理Lが完了し、状態212に示すように、L1タグ352のインデックスA及びw0で表されるブロ

50

ックはI状態に遷移する。さらに、ムーブインバッファ305からインデックスA及びw0のエントリが削除される。ムーブインバッファ305からインデックスA及びw0のエントリが削除されると、状態218に示すように、ストアバッファ311におけるインデックスA及びw0に対応するキャッシュヒットフラグが1に設定される。キャッシュヒットフラグが1になることで、ストア実行部310は、データのストアの実行を開始する。すなわち、時刻T4から時刻T5の間、ストア処理は待機させられる。

【0112】

その後、処理Sが完了すると、インデックスA及びw0で表されるブロックはデータが更新されるので、L1タグ352におけるインデックスA及びw0に対応するキャッシュ状態は、状態213のようにM状態に遷移する。

10

【0113】

ここで、図10を参照して、ストアバッファ311を用いない場合のL1タグ232の遷移について説明する。図10は、ストアバッファを用いない場合のL1タグの遷移を表す図である。

【0114】

この場合、図10に示すように、L1タグ352は、状態221で示すように、インデックスA及びw0で表されるブロックのキャッシュ状態としてE状態を保持する。そして、時刻T6で処理Lが開始される。この場合、ムーブインバッファ305には、インデックスA及びw0に対応する判定用フラグとして0が設定される。

【0115】

その後、時刻T7で処理Sが開始される。この場合、処理Sの待機は行われず、インデックスA及びw0で表されるブロックのデータが更新される。これにより、状態222に示すように、L1タグ352のインデックスA及びw0で表されるブロックはM状態に遷移する。

20

【0116】

しかし、ムーブインバッファ305の判定用フラグは0であるため、タグ管理部381は、L1タグ352が有するインデックスA及びw0で表されるブロックのキャッシュ状態を確認せずにインデックスA及びw0に対応するタグを無効化してしまう。この場合、M状態のブロックに格納されたデータの退避が行われなため、データ化けが発生してしまう。

30

【0117】

これに対して、本実施例に係る1次キャッシュ管理部103は、既にロード処理が実行されている状態で、ストア命令が発行された場合には、ロード処理が完了するまでストア命令を待機させ、その後ストア処理を実行する。これにより、ロード処理の実行中にストア処理によりブロックがM状態に遷移することを回避でき、データの退避を確実に行うことでデータ化けの発生を抑えることができる。

【0118】

さらに、以上の説明では、1次キャッシュ及び2次キャッシュを有する階層キャッシュにおいて、1次キャッシュでキャッシュミスが発生した場合の処理について説明した。ただし、各実施例で説明した機能は、階層キャッシュのどの階層に用いることも可能であり、また、階層キャッシュの階層の数も特に制限は無い。

40

【符号の説明】

【0119】

- 1 情報処理装置
- 10 システムボード
- 11 CPU
- 12 メモリ
- 20 IOユニット
- 30 ディスクユニット
- 100 コア

50

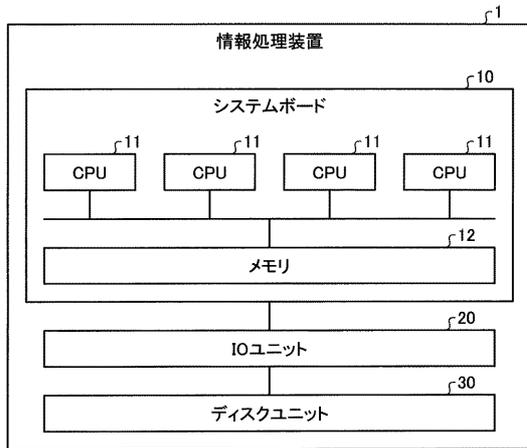
- 1 0 1 命令制御部
- 1 0 2 演算処理部
- 1 0 3 1次キャッシュ管理部
- 1 1 2 2次キャッシュ管理部
- 1 2 1 2次キャッシュ
- 3 0 1 1次キャッシュ
- 3 0 2 セレクタ
- 3 0 3 タグマッチ判定部
- 3 0 4 T L B
- 3 0 5 ムーブインバッファ
- 3 0 6 リプレイスウェイ選択部
- 3 0 7 フラグ生成部
- 3 0 8 格納処理部
- 3 0 9 ムーブアウトバッファ
- 3 1 0 ストア実行部
- 3 1 1 ストアバッファ
- 3 1 2 判定部
- 3 5 1 1次データキャッシュ
- 3 5 2 L 1 タグ
- 3 8 1 タグ管理部
- 3 8 2 データ管理部

10

20

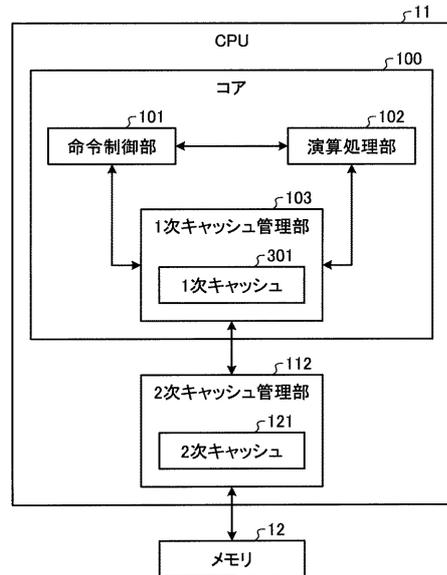
【 図 1 】

情報処理装置のハードウェア構成図

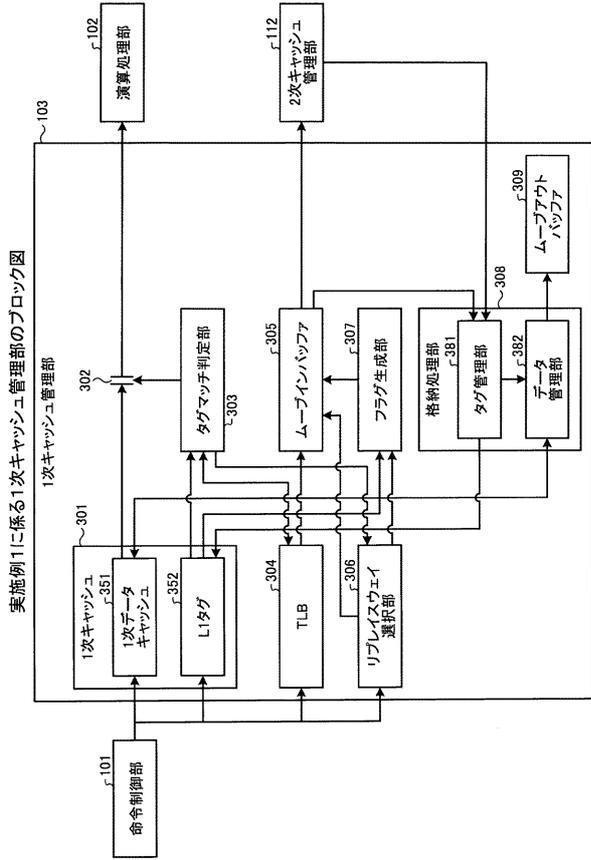


【 図 2 】

CPUのブロック図

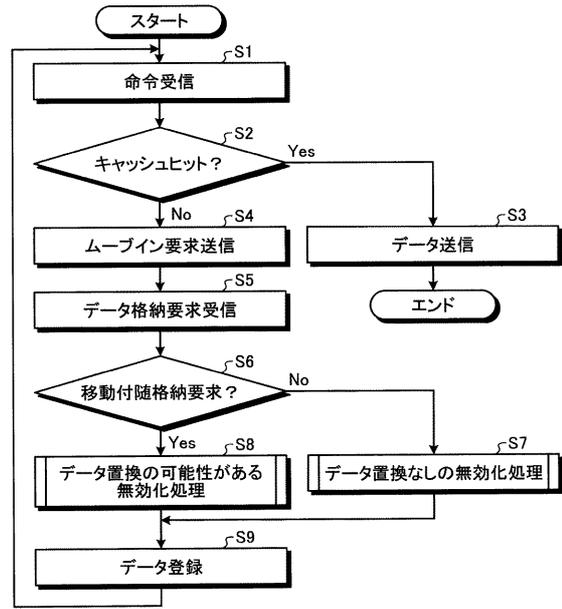


【 図 3 】



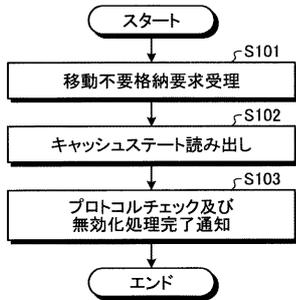
【 図 4 】

1次キャッシュ管理部によるロード命令の処理のフローチャート



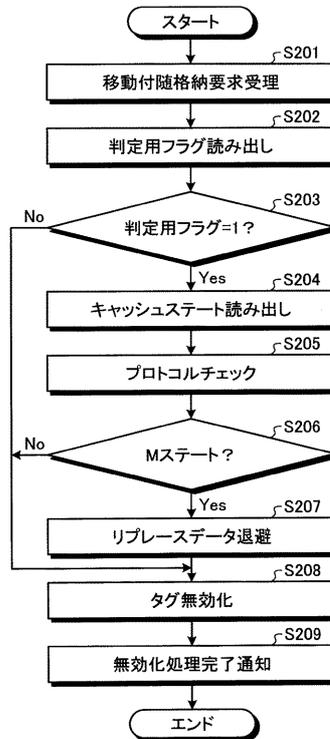
【 図 5 】

移動不要格納要求の場合の無効化処理のフローチャート

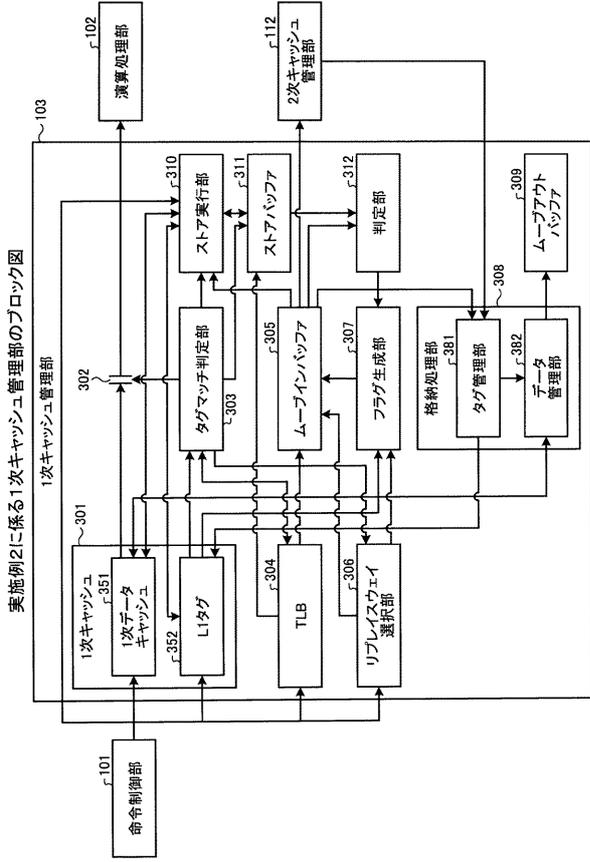


【 図 6 】

移動付随格納要求の場合の無効化処理のフローチャート

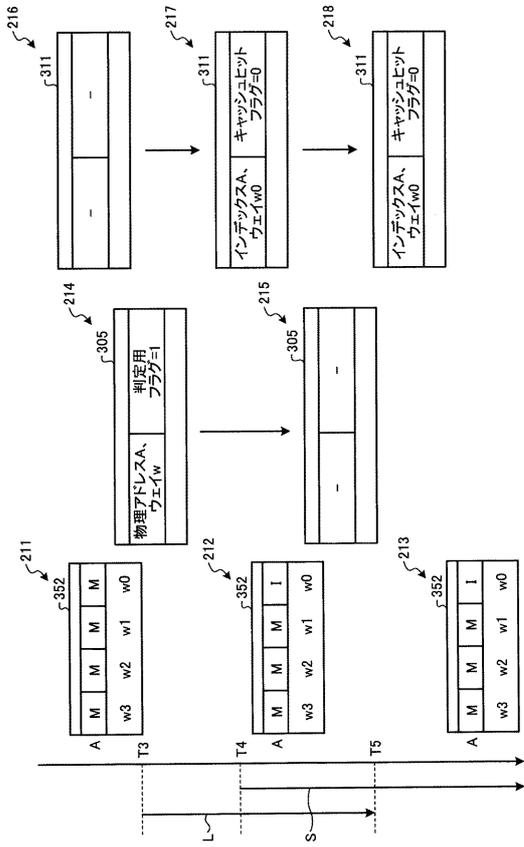


【 図 7 】



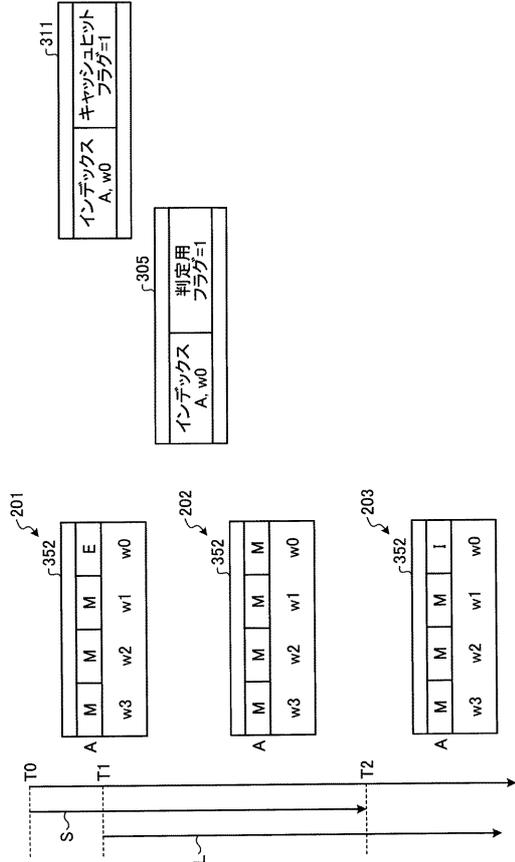
【 図 9 】

ロード命令が先行する状態での競合時のL1タグの遷移を表す図



【 図 8 】

ストア命令が先行する状態での競合時のL1タグの遷移を表す図



【 図 10 】

ストアバypassを用いない場合の競合時のL1タグの遷移を表す図

