

(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)

G06F 12/0811 (2016.01)

G O 6 F 12/0811

5 B 2 0 5

G O 6 F 12/0895 (2016.01)

G O 6 F 12/0895 1 1 2

G O 6 F 12/0864 (2016.01)

G O 6 F 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)

(22) 出願日 平成30年6月26日 (2018. 6. 26)

(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

(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

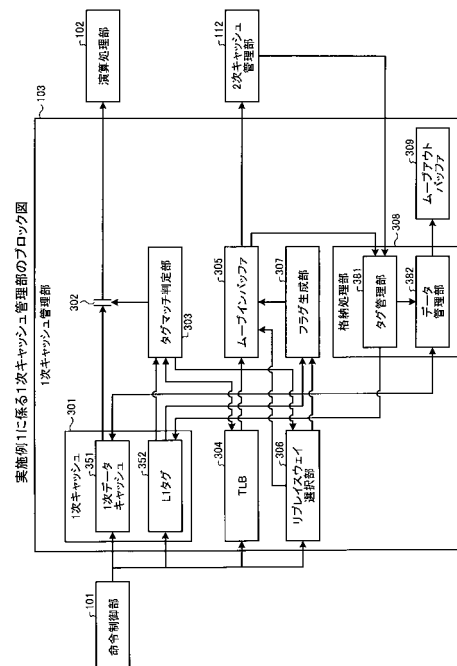
(54) 【発明の名称】 情報処理装置、演算処理装置及び情報処理装置の制御方法

(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
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かどうかを判定する判定部と、

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

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かどうかを判定し、

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、

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】

【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（L2（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（L1（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

10

20

30

40

50

間において、データのコヒーレンシ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シを維持する。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シ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】

10

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】

20

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。

30

【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。

40

【0008】

なお、キャッシュのコヒーレンシ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

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タを確認する。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へのアクセス時にはステータ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ステータの場合、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ステータ又はSステータの場合、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ステータ時の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ステータ時の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

【0015】

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】

【0016】

【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L1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L1タグの遷移を表す図である。

【図10】図10は、ストアバッファを用いない場合の競合時のL1タグの遷移を表す図である。

【発明を実施するための形態】

【0017】

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】

【0018】

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10、I/O（Input Output）ユニット20、ディスクユニット30を有する。

【0019】

システムボード10は、複数のCPU11及びメモリ12が搭載される。システムボード10は、バスによりI/Oユニット20と接続される。システムボード10上のCPU11及びメモリ12は、システムボード10とI/Oユニット20とを繋ぐバスを介してI/Oユニット20と通信する。

30

【0020】

演算処理装置であるCPU11は、I/Oユニット20を介してディスクユニット30とデータの送受信を行う。また、CPU11は、バスを介して主記憶装置であるメモリ12と接続される。CPU11は、ディスクユニット30に格納された各種プログラムを読み出し、メモリ12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CPU11を例に図示したが、CPU11の数に特に制限は無い。

40

【0021】

I/Oユニット20は、ディスクユニット30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30は、補助記憶装置である。ディスクユニット30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。

【0022】

図2は、CPUのブロック図である。CPU11は、コア100及び2次キャッシュ管理部112を有する。

【0023】

コア100は、命令制御部101、演算処理部102及び1次キャッシュ管理部103を有する。命令制御部101、演算処理部102及び1次キャッシュ管理部103は、そ

50

れぞれ互いに接続され、相互に通信を行う。

【0024】

命令制御部101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101は、演算処理の実行命令を演算処理部102へ出力する。その後、命令制御部101は、出力した実行命令の応答を演算処理部102から取得する。

【0025】

また、命令制御部101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103に送信する。その後、命令制御部101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103から取得する。

【0026】

演算処理部102は、演算処理の実行命令の入力を命令制御部101から受ける。また、演算処理部102は、1次キャッシュ管理部103からデータを取得する。そして、演算処理部102は、1次キャッシュ管理部103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。

【0027】

1次キャッシュ管理部103は、1次キャッシュ301を有する。1次キャッシュ管理部103は、キャッシュのロード命令の入力を命令制御部101から受ける。そして、1次キャッシュ管理部103は、ロード命令で指定されたデータが1次キャッシュ301に存在するか否かを判定する。1次キャッシュ301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103は、指定されたデータを1次キャッシュ301から読み出して演算処理部102へ出力する。その後、1次キャッシュ管理部103は、キャッシュのロード命令の処理完了を命令制御部101へ通知する。

【0028】

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301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103は、ムーブイン要求を2次キャッシュ管理部112に送信する。その後、1次キャッシュ管理部103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112から受ける。そして、1次キャッシュ管理部103は、データ格納要求にしたがい1次キャッシュ301に対する処理を行い、受信したデータを1次キャッシュ301に格納する。その後、1次キャッシュ管理部103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103は、キャッシュのロード命令の処理完了を命令制御部101へ通知する。

【0029】

また、1次キャッシュ管理部103は、キャッシュのストア命令の入力を命令制御部101から受ける。そして、1次キャッシュ管理部103は、ストア命令で指定されたデータが1次キャッシュ301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301に存在しない場合、1次キャッシュ管理部103は、データを新たに1次キャッシュ301に登録する。さらに、1次キャッシュ管理部103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121へのデータの書き込みを2次キャッシュ管理部112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301に存在する場合、1次キャッシュ管理部103は、1次キャッシュ301上のデータを更新する。さらに、1次キャッシュ管理部103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121へのデータの書き込みを2次キャッシュ管理部112へ依頼する。

【0030】

2次キャッシュ管理部112は、2次キャッシュ121を有する。2次キャッシュ121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121は、1次キャッシュ301の各領域のキャッシュステートがIステートであるか否かの情報を保持する。

【0031】

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、セレクト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、リプ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

【0039】

ここで、L1タグ352が保持するキャッシュステートは、MESI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。

【0040】

1次キャッシュ301の1次データキャッシュ351及びL1タグ352は、命令制御部101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。L1タグ352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303及びフラグ生成部307へ出力する。

10

【0041】

1次データキャッシュ351は、後述するタグマッチ判定部303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クタ302を介して演算処理部102へ出力する。

【0042】

TLB304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。TLB304は、命令制御部101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、TLB304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、TLB304は、取得した物理アドレスをタグマッチ判定部303へ出力する。その後、タグマッチ判定部303からタグミスマッチの通知を受けた場合、TLB304は、ムーブインバッファ305へ取得した物理アドレスを出力する。

20

【0043】

タグマッチ判定部303は、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TLB304から受ける。そして、タグマッチ判定部303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351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303は、タグマッチしたタグの情報をセクタ302に送る。

30

【0044】

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351に存在しない場合、タグマッチ判定部303は、タグミスマッチと判定する。そして、タグマッチ判定部303は、タグミスマッチをTLB304及びリプレイスウェイ選択部306に通知する。このタグマッチ判定部303が、「判定部」の一例にあたる。

【0045】

セクタ302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303から受ける。そして、セクタ302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の中から選択して演算処理部102へ出力する。

40

【0046】

リプレイスウェイ選択部306は、命令制御部101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306は、タグマッチ判定部303からタグミスマッチの通知を受けると、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306は、1次キャッシュ301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306は、選択したウェイの情報をムーブインバッファ305及びフラグ生成部307へ出力する。以下では、このリプレイスウェイ選択部306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

【0047】

フラグ生成部307は、格納先ウェイの情報をリプレイスウェイ選択部306から取得する。そして、フラグ生成部307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L1タグ352から取得する。

【0048】

次に、フラグ生成部307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、L1タグ352に対するキャッシュステートの確認の実行を表す。

【0049】

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307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、L1タグ352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307は、格納先ウェイの判定用フラグをムーブインバッファ305へ出力する。

【0050】

ムーブインバッファ305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TLB304から受ける。また、ムーブインバッファ305は、格納先ウェイの情報の入力をリプレイスウェイ選択部306から受ける。さらに、ムーブインバッファ305は、格納先ウェイの判定用フラグの入力をフラグ生成部307から受ける。そして、ムーブインバッファ305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112へ出力する。このムーブインバッファ305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。

【0051】

格納処理部308は、タグ管理部381及びデータ管理部382を有する。タグ管理部381は、ムーブインバッファ305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112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308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。

【0052】

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キャッシュステートをL1タグ352から取得する。そして、タグ管理部381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328にデータの格納を依頼する。その後、タグ管理部381は、1次データキャッシュ351に新たに格納され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。

【0053】

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381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305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381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ので、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L1タグ352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381は、データの退避をデータ管理部382に指示する。

【0054】

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。

【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。

【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】

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 T L B 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）。

【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）。

10

20

30

40

50

【0065】

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）。

【0066】

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）。

【0067】

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がある無効化処理を実行する（ステップS8）。 10

【0068】

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。

【0069】

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。

【0070】

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

【0071】

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）。

【0072】

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）。

【0073】

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

【0074】

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）。

【0075】

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）。

【0076】

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

【0077】

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）。

【0078】

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）。

【0079】 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】

TLB 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、TLB 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 311 へ出力する。

【0087】

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 TLB 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。

40

【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 310 は、タグミスの通知をタグマッチ判定部 303 から受ける。この場合、ストア実行部 310 は、1 次データキャッシュ 351 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 310 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 L1 タグ 352 に登録する。

【0090】

一方、タグマッチの場合、ストア実行部 310 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 303 から受ける。そして、ストア実行部 310 は、ムーブインバッファ 305 に同じ物理アドレス及びウェイの情報を有するエン

10

【0091】

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 310 は、ストアバッファ 311 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 310 は、1 次データキャッシュ 351 及び L1 タグ 352 に対してデータのストアを行う。

【0092】

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 310 は、ストアバッファ 311 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 310 は、1 次データキャッシュ 351 及び L1 タグ 352 に対してデータのストアを行う。

20

【0093】

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 310 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 310 は、ストア処理の実行を待機する。その後、ストア実行部 310 は、ムーブインバッファ 305 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

【0094】

ストアバッファ 311 は、物理アドレスの情報の入力を TLB 304 から受ける。また、ストアバッファ 311 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 303 から受ける。さらに、ストアバッファ 311 は、ストア実行部 310 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。

【0095】

すなわち、ストアバッファ 311 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 311 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

【0096】

判定部 312 は、ロード命令が命令制御部 101 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 305 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 312 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 311 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 311 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 ステートであり、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 ステートから M ステートに遷移する。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 ステート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 E ステートだったブロックが M ステート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 ステートである。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 で表されるプロ

50

ックはIステートに遷移する。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ステートは、状態213のようにMステートに遷移する。

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ステートとしてEステートを保持する。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】

20

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ステートに遷移する。

【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ステート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ステートのブロックに格納されたデータの退避が行われないため、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ステート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

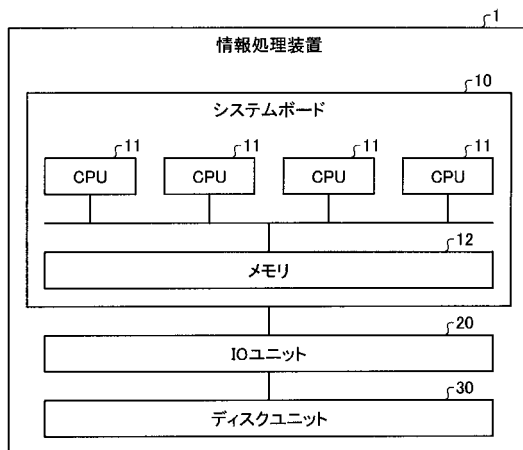
1 0 1	命令制御部
1 0 2	演算処理部
1 0 3	1次キャッシュ管理部
1 1 2	2次キャッシュ管理部
1 2 1	2次キャッシュ
3 0 1	1次キャッシュ
3 0 2	セレクタ
3 0 3	タグマッチ判定部
3 0 4	T L B
3 0 5	ムーブインバッファ
3 0 6	リプレイスウェイ選択部
3 0 7	フラグ生成部
3 0 8	格納処理部
3 0 9	ムーブアウトバッファ
3 1 0	ストア実行部
3 1 1	ストアバッファ
3 1 2	判定部
3 5 1	1次データキャッシュ
3 5 2	L 1 タグ
3 8 1	タグ管理部
3 8 2	データ管理部

10

20

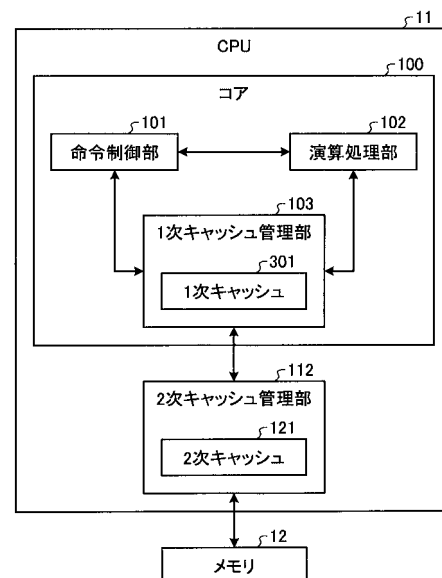
【図 1】

情報処理装置のハードウェア構成図

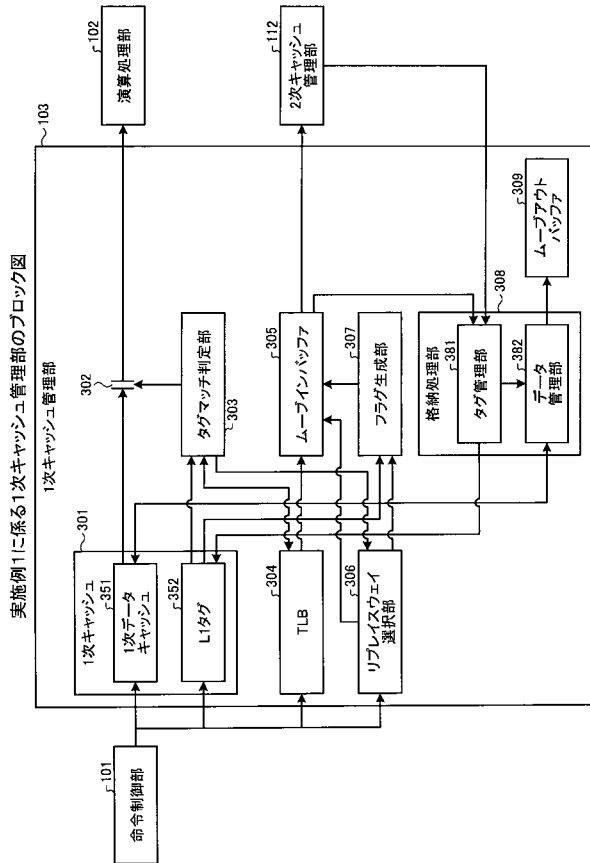


【図 2】

CPUのブロック図

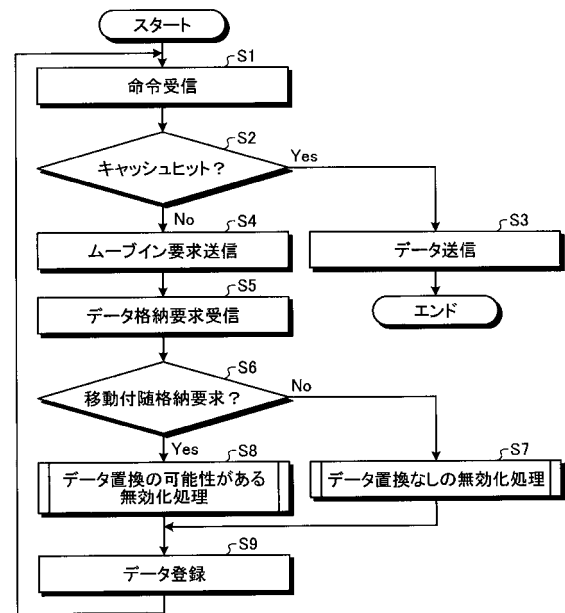


【図 3】



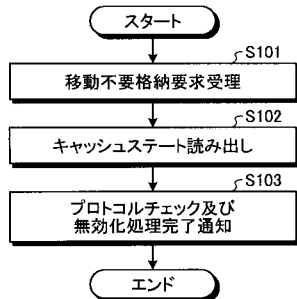
【図 4】

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



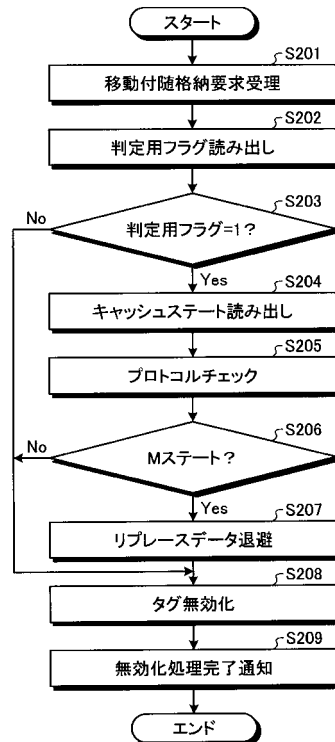
【図 5】

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

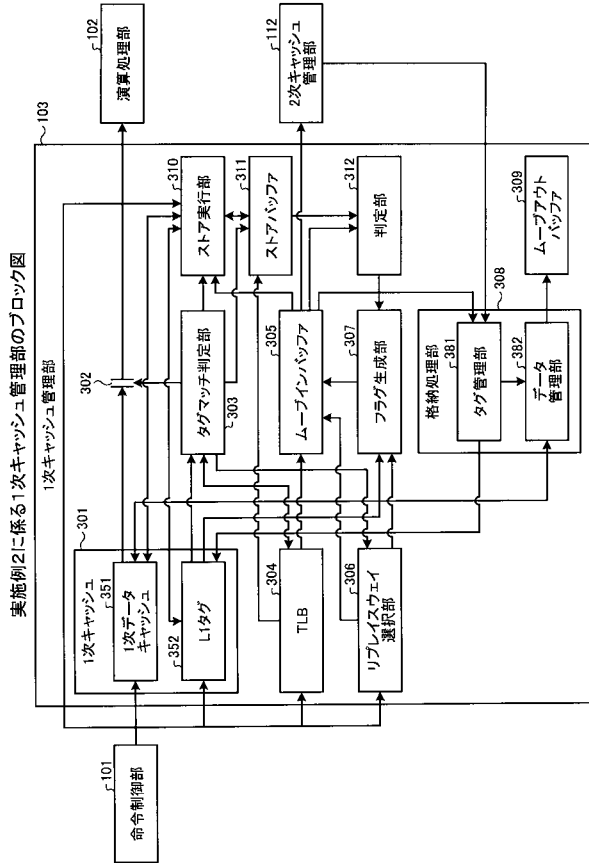


【図 6】

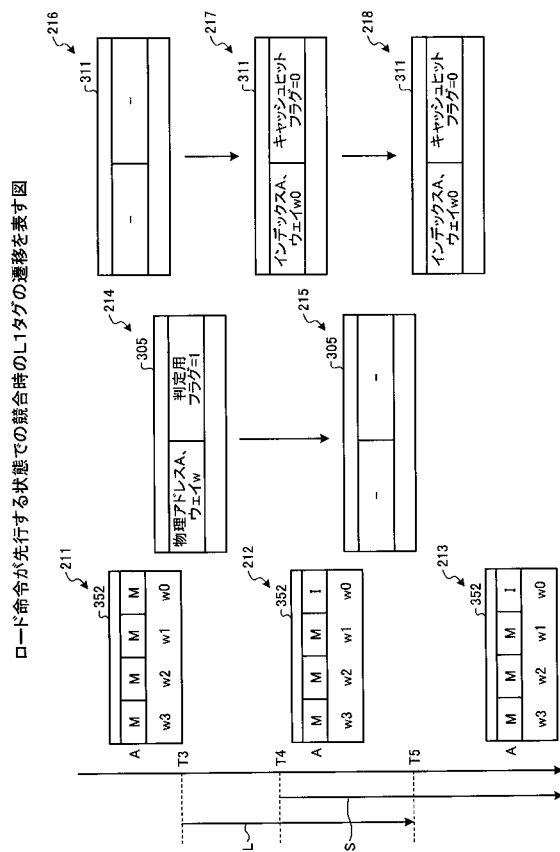
移動付随格納要求の場合の無効化処理のフローチャート



【図 7】



【図 9】



【図 8】

