

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5417879号  
(P5417879)

(45) 発行日 平成26年2月19日(2014.2.19)

(24) 登録日 平成25年11月29日(2013.11.29)

(51) Int.Cl. F I  
**G 0 6 F 12/08 (2006.01)** G O 6 F 12/08 5 0 3 C  
 G O 6 F 12/08 5 1 5 P

請求項の数 9 (全 27 頁)

<p>(21) 出願番号 特願2009-34588 (P2009-34588)                  (22) 出願日 平成21年2月17日(2009.2.17)                  (65) 公開番号 特開2010-191638 (P2010-191638A)                  (43) 公開日 平成22年9月2日(2010.9.2)                  審査請求日 平成23年11月2日(2011.11.2)</p> <p>前置審査</p>	<p>(73) 特許権者 308014341                  富士通セミコンダクター株式会社                  神奈川県横浜市港北区新横浜二丁目10番                  23</p> <p>(74) 代理人 100105957                  弁理士 恩田 誠</p> <p>(74) 代理人 100068755                  弁理士 恩田 博宣</p> <p>(72) 発明者 月城 玄                  愛知県春日井市高蔵寺町二丁目1844番                  2 富士通VLSI株式会社内</p> <p>審査官 野田 佳邦</p>
---	--

最終頁に続く

(54) 【発明の名称】 キャッシュ装置

(57) 【特許請求の範囲】

【請求項1】

記憶装置に格納されたデータの一部を格納するキャッシュメモリと、  
 第1バッファ回路と、

前記記憶装置に対する処理装置からのデータ格納要求に基づいて、前記データ格納要求に含まれるアドレスに対応するデータが前記キャッシュメモリに格納されているか否かを判定し、格納されていないと判定した場合には、前記アドレスに対応するデータを前記記憶装置から読み出して前記キャッシュメモリへ格納する読み出し制御信号を前記記憶装置へ出力すると共に前記処理装置からの出力データの前記第1バッファ回路への格納を許可し、前記許可に応じて前記第1バッファ回路へ格納された前記出力データを前記キャッシュメモリへ格納する制御回路とを有し、

前記制御回路は、前記読み出し制御信号に応じて前記データをキャッシュメモリへ格納する処理と、前記処理装置からの出力データを前記第1バッファ回路へ格納する処理とを並行して実行させる

ことを特徴とするキャッシュ装置。

【請求項2】

前記制御回路は、前記アドレスに対応するデータを前記記憶装置から読み出して前記キャッシュメモリへ格納した後、前記第1バッファ回路へ格納された前記出力データを前記キャッシュメモリへ格納する

ことを特徴とする請求項 1 に記載のキャッシュ装置。

【請求項 3】

前記制御回路は、前記アドレスに対応するデータが前記キャッシュメモリに格納されていると判定した場合には、前記処理装置からの出力データの前記第 1 バッファ回路への格納を許可し、前記許可に応じて前記第 1 バッファ回路へ格納された前記出力データを前記キャッシュメモリへ格納する

ことを特徴とする請求項 1 又は 2 に記載のキャッシュ装置。

【請求項 4】

第 2 バッファ回路を更に有し、

前記制御回路は、前記処理装置からの前記データ格納要求の次のデータ格納要求に基づいて、前記次のデータ格納要求に含まれるアドレスに対応するデータが前記キャッシュメモリに格納されているか否かを判定し、格納されていないと判定した場合には、前記処理装置からの出力データの前記第 2 バッファ回路への格納を許可する

ことを特徴とする請求項 1 ~ 3 のうちの何れか 1 項に記載のキャッシュ装置。

【請求項 5】

第 3 バッファ回路を更に有し、

前記制御回路は、前記記憶装置に対する前記処理装置からのデータ読み出し要求に基づいて、前記データ読み出し要求に含まれるアドレスに対応するデータが前記キャッシュメモリに格納されているか否かを判定し、格納されていると判定した場合には、該アドレスに対応するデータを前記キャッシュメモリから前記第 3 バッファ回路へ格納した後で前記

ことを特徴とする請求項 1 ~ 3 のうちの何れか 1 項に記載のキャッシュ装置。

【請求項 6】

第 4 バッファ回路を更に有し、

前記制御回路は、前記格納されていないと判定した場合には、前記キャッシュメモリに格納されているデータの中で最も長く使用されていないデータを前記第 4 バッファへ格納する

ことを特徴とする請求項 1 ~ 3 のうちの何れか 1 項に記載のキャッシュ装置。

【請求項 7】

前記キャッシュメモリは複数のラインを含み、複数の前記ラインが順次前記記憶装置のデータの格納に用いられる

ことを特徴とする請求項 1 ~ 6 のうちの何れか 1 項に記載のキャッシュ装置。

【請求項 8】

前記制御回路は、前記最も長く使用されていないデータの格納位置を示す第 1 ポインタを有する

ことを特徴とする請求項 6 に記載のキャッシュ装置。

【請求項 9】

前記制御回路は、前記キャッシュメモリに格納されているデータの中で最後に更新されたデータの格納位置を示す第 2 ポインタを有する

ことを特徴とする請求項 8 に記載のキャッシュ装置。

【発明の詳細な説明】

【技術分野】

【0001】

キャッシュ装置に関する。

【背景技術】

【0002】

従来、コンピュータシステムでは、主記憶装置のアクセス時間を短縮するために、主記憶装置と、その主記憶装置の情報（データ）を読み出す処理装置との間に、キャッシュ装置が接続されている（例えば、特許文献 1 参照）。

【0003】

10

20

30

40

50

キャッシュ装置は、主記憶装置と比べてアクセス速度、即ちデータの書き込み・読み出し速度が高速な記憶領域（キャッシュメモリ）を有している。キャッシュ装置は、このキャッシュメモリに処理装置が主記憶装置をアクセスした内容（データ）を記憶しておき、同じデータにアクセスしたときにはキャッシュメモリから取り出すことにより、主記憶装置に対するアクセス速度を見かけ上高速化している。

【0004】

また、連続した複数のアドレスに記憶されたデータは、連続的にアクセスされる可能性が高いことから、キャッシュ装置は、所定数（例えば32バイト）のデータを1つのブロックとし、複数のブロックをキャッシュメモリに格納する。そして、キャッシュ装置は、ブロックの格納位置（ライン番号）と主記憶装置のアドレスとを対応付けて記憶する管理領域を有している。

10

【0005】

要求されたデータがキャッシュメモリに記憶されている場合をキャッシュヒットといい、データがキャッシュメモリに記憶されていない場合をキャッシュミスという。従来のキャッシュ装置は、処理装置から主記憶装置内のデータに対するアクセスが要求されると、データのアドレスと管理領域に記憶されたアドレスとを比較してキャッシュヒットかキャッシュミスかを判断する。キャッシュヒットの場合、キャッシュ装置は、管理領域を参照して、データが含まれるブロックのライン番号を検索し、そのライン番号のブロックに対してアクセスに対応する処理を実行する。例えば、アクセス要求が書き込み要求の場合、キャッシュ装置は、処理装置から出力されるデータを該当するブロックに格納し、アクセス要求が読み出し要求の場合には該当するブロックのデータを処理装置に出力する。キャッシュミスの場合、キャッシュ装置は、主記憶装置からデータが含まれる1ブロック分の記憶領域のデータを読み出してキャッシュメモリに格納し、その格納したブロックに対してアクセスに対応する処理を実行する。

20

【0006】

主記憶装置のブロックとキャッシュメモリのブロックとを対応付けるマッピング方式の代表的なものとして、セットアソシアティブ方式、フルアソシアティブ方式、セクタ方式がある。フルアソシアティブ方式は、他の方式と比べてランダムアクセスに適しており、キャッシュヒット率が高い。

【先行技術文献】

30

【特許文献】

【0007】

【特許文献1】特開平7-152566号公報

【発明の概要】

【発明が解決しようとする課題】

【0008】

しかし、キャッシュ装置は、キャッシュヒットであっても、アクセスを終了するまでに時間、所謂レイテンシが発生する。このため、アクセスを終了するまでの時間短縮、即ち高速なアクセスが求められている。

【0009】

40

このキャッシュ装置で、高速なアクセスを可能とすることを目的とする。

【課題を解決するための手段】

【0010】

開示のキャッシュ装置は、記憶装置に格納されたデータの一部を格納するキャッシュメモリと、第1バッファ回路と、前記記憶装置に対する処理装置からのデータ格納要求に基づいて、前記データ格納要求に含まれるアドレスに対応するデータが前記キャッシュメモリに格納されているか否かを判定し、格納されていないと判定した場合には、前記アドレスに対応するデータを前記記憶装置から読み出して前記キャッシュメモリへ格納する読み出し制御信号を前記記憶装置へ出力すると共に前記処理装置からの出力データの前記第1バッファ回路への格納を許可し、前記許可に応じて前記第1バッファ回路へ格納された前

50

記出力データを前記キャッシュメモリへ格納する制御回路とを有し、前記制御回路は、前記読み出し制御信号に応じて前記データをキャッシュメモリへ格納する処理と、前記処理装置からの出力データを前記第1バッファ回路へ格納する処理とを並行して実行させる。

【発明の効果】

【0011】

開示のキャッシュ装置は、高速なアクセスを可能とすることができるという効果を奏する。

【図面の簡単な説明】

【0012】

【図1】コンピュータシステムの概略構成図である。

10

【図2】キャッシュ装置の概略ブロック図である。

【図3】キャッシュメモリ及び各バッファの説明図である。

【図4】(a)(b)はデータライト時の動作説明図である。

【図5】(a)(b)はデータライト時のタイミング図である。

【図6】データライト時の動作説明図である。

【図7】(a)~(c)はデータライト時のタイミング図である。

【図8】(a)(b)はデータリード時の動作説明図である。

【図9】(a)(b)はデータリード時のタイミング図である。

【図10】ライトバック時の動作説明図である。

【図11】ライトバック時のタイミング図である。

20

【図12】ライトバック時のタイミング図である。

【図13】(a)(b)はライトバック時の動作説明図である。

【図14】ライトバック時のタイミング図である。

【図15】(a)~(c)はライトバック時の動作説明図である。

【図16】データライト時の動作説明図である。

【図17】データライト時の動作説明図である。

【図18】ライトバック時のタイミング図である。

【図19】(a)~(c)はキャッシュメモリのライン設定の説明図である。

【発明を実施するための形態】

【0013】

30

以下、一実施形態を図1~図12に従って説明する。

図1に示すように、コンピュータシステムは、処理装置としてのCPU11、キャッシュ装置12、メモリコントローラ13、主記憶装置としてのメインメモリ14を含む。CPU11とキャッシュ装置12はバスB1を介して接続され、キャッシュ装置12とメモリコントローラ13はバスB2を介して接続され、メモリコントローラ13とメインメモリ14はバスB3を介して接続されている。バスB1は、CPU11とキャッシュ装置12との間でデータの授受を行うために必要な1又は複数の制御信号を伝達する配線からなる制御バス、複数ビットのアドレス信号を伝達する配線からなるアドレスバス、複数ビットのデータを伝達する配線からなるデータバスを含む。同様に、バスB2、B3は、それぞれ、制御バス、アドレスバス、データバスを含む。以下、CPU11とキャッシュ装置12との間のバスをCPUバスといい、キャッシュ装置12とメインメモリ14との間のバスをメモリバスという。

40

【0014】

CPU11は、キャッシュ装置12及びメモリコントローラ13を介してメインメモリ14をアクセスする。メインメモリ14は例えばSDRAMである。メインメモリ14には、CPU11が実行する制御プログラム、制御プログラムの実行に用いられるデータ、制御プログラムを実行するCPU11により生成されるデータが格納される。なお、制御プログラムやデータは、CPU11や他の処理装置が実行するブートプログラムにより、図示しないROMやディスク装置等の補助記憶装置から読み出されてメインメモリ14に格納される。

50

## 【 0 0 1 5 】

C P U 1 1 は、制御プログラムやデータをメインメモリ 1 4 から読み出すためのアクセス要求（リードリクエスト）、変更又は生成したデータをメインメモリ 1 4 に書き込むためのアクセス要求（ライトリクエスト）を出力する。

## 【 0 0 1 6 】

キャッシュ装置 1 2 は、メインメモリ 1 4 と比べて小容量かつ高速なメモリ（以下、キャッシュメモリという）を有している。キャッシュメモリは、例えばスタティック R A M である。キャッシュ装置 1 2 は、過去に C P U 1 1 がメインメモリ 1 4 をアクセスしたデータをキャッシュメモリに記憶する。次に C P U 1 1 から出力されるアクセス要求に対応するデータがキャッシュメモリに記憶されている場合をキャッシュヒットといい、データがキャッシュメモリに記憶されていない場合をキャッシュミスという。キャッシュ装置 1 2 は、C P U 1 1 から出力されるアクセス要求に回答して、キャッシュヒットのリード動作の場合にはキャッシュメモリに記憶したデータを C P U 1 1 に出力し、キャッシュヒットのライト動作の場合にはキャッシュメモリに書き込みを行う。即ち、キャッシュヒットの場合、C P U 1 1 のアクセス要求は、C P U 1 1 とキャッシュ装置 1 2 との間のデータの送受信だけで終了する。従って、C P U 1 1 におけるデータ入出力のためのアクセスを高速化することができる。

10

## 【 0 0 1 7 】

キャッシュ装置 1 2 は、キャッシュミスの場合にメインメモリ 1 4 をアクセスするための要求をメモリコントローラ 1 3 に出力する。この時、キャッシュ装置 1 2 は、要求されるデータのアドレスから連続したアドレスによって指定される複数バイト（例えば 3 2 バイト）のデータをメインメモリ 1 4 から読み出すようにアクセス要求を出力する。この複数バイトのデータを格納する領域をラインと呼ぶ。つまり、1 つのラインには複数バイトのデータが格納され、各データのアドレスは連続している。C P U 1 1 が実行する制御プログラムは、複数の命令語（オペコード）とデータ（オペランド）から構成され、逐次実行する命令語及びその命令語の実行に必要なデータはメインメモリ 1 4 上の連続したアドレスにより指定される領域に格納されていることが多い。従って、C P U 1 1 のアクセス要求は、ある時点の前後でメインメモリ 1 4 の限られた領域に集中する。これを局所性と呼ぶ。このように、アクセス要求が集中する領域のデータを、キャッシュ装置 1 2 のキャッシュメモリに記憶することで、アクセスの高速化を図るとともにメインメモリ 1 4 に対するアクセス回数を低減する。

20

30

## 【 0 0 1 8 】

メモリコントローラ 1 3 は、キャッシュ装置 1 2 から出力される要求に応じて、メインメモリ 1 4 をアクセスするための 1 又は複数の信号をメインメモリ 1 4 に出力する。メインメモリ 1 4 は、その信号に回答してリード処理又はライト処理を実行し、メモリセルからデータの読み出し、又はメモリセルへのデータの書き込みを行う。そして、メモリコントローラ 1 3 は、メインメモリ 1 4 が出力するデータを受け取り、その受け取ったデータをキャッシュ装置 1 2 に出力する。

## 【 0 0 1 9 】

次に、キャッシュ装置 1 2 の構成を説明する。

40

図 2 に示すように、キャッシュ装置 1 2 は、キャッシュメモリ 2 1、複数（本実施形態では 2 つ）のライトバッファ（W B ）2 2 a , 2 2 b、リードバッファ（R B ）2 3、ライトバックバッファ（W B B ）2 4、タグメモリ（タグ）2 5、制御回路 2 6 を有している。なお、カッコ内は図中の表記を示す。

## 【 0 0 2 0 】

キャッシュメモリ 2 1 は、複数（例えば、1 2 8）のラインに区画可能な記憶容量を持つ。

キャッシュメモリ 2 1 と C P U 1 1 との間には、ライトバッファ 2 2 a , 2 2 b 及びリードバッファ 2 3 が接続されている。ライトバッファ 2 2 a , 2 2 b 及びリードバッファ 2 3 は、1 ラインのデータを記憶可能な容量（例えば、3 2 バイト）のバッファメモリで

50

ある。ライトバッファ22a, 22b及びリードバッファ23は、内部バスIB1及びCPUバスB1を介してCPU11と接続されている。また、ライトバッファ22a, 22b及びリードバッファ23は、内部バスIB2を介してキャッシュメモリ21と接続されている。

#### 【0021】

この内部バスIB1のバス幅は、CPUバスB1のバス幅と同じ値に設定されている。内部バスIB2のバス幅は、内部バスIB1のバス幅、即ちCPUバスB1のバス幅の正数倍（本実施形態では1倍）に設定されている。CPU11は、1回のアクセス要求に対して複数バイト（例えば4バイト）のデータを入出力可能に構成されている。従って、CPUバスB1のバス幅は、この複数バイト（4バイト）のデータ転送が可能なバス幅に設定されている。そして、内部バスIB1, IB2のバス幅は、CPUバスB1のバス幅と同じ値（4バイト）に設定されている。

10

#### 【0022】

尚、内部バスIB1, IB2は、ライトバッファ22a, 22bとCPU11との間のバス幅と、ライトバッファ22a, 22bとキャッシュメモリ21との間のバス幅を示すためのものであり、両内部バスが電氣的に接続された1つのバスにより構成されている。即ち、キャッシュメモリ21は、シングルポート構成のスタティックRAMである。

#### 【0023】

CPU11から出力された4バイトのライトデータ（出力データ）WDcは、CPUバスB1及び内部バスIB1を介してライトバッファ22a, 22bに供給され、ライトバッファ22a又はライトバッファ22bは供給されたデータWDcを記憶する。ライトバッファ22a, 22bから読み出されたデータは内部バスIB2を介してキャッシュメモリ21に供給され、キャッシュメモリ21は供給されたデータを記憶する。

20

#### 【0024】

キャッシュメモリ21から読み出されたデータは内部バスIB2を介してリードバッファ23に供給され、そのリードバッファ23はそのデータを記憶する。リードバッファ23から読み出されたデータは、内部バスIB1及びCPUバスB1を介してCPU11にリードデータ（入力データ）RDcとして供給される。

#### 【0025】

また、リードバッファ23は内部バスIB1に接続されている。従って、CPU11から出力されたライトデータWDcを記憶可能に構成されている。また、本実施形態のキャッシュ装置12は、キャッシュメモリ21から読み出されたデータを直接的にCPU11に供給する経路（データバス）P1を有している。従って、キャッシュメモリ21から読み出したデータをリードデータRDcとして直接的にCPU11に供給可能に構成されている。

30

#### 【0026】

キャッシュメモリ21とメモリコントローラ13の間には、ライトバックバッファ24が接続されている。ライトバックバッファ24は、1ラインのデータを記憶可能な容量（32バイト）のバッファである。キャッシュメモリ21から読み出された1ライン分のデータは、ライトバックバッファ24に記憶される。そして、ライトバックバッファ24から読み出されたデータは、ライトデータWDmとしてメモリバスB2を介してメモリコントローラ13に出力される。一方、メモリコントローラ13からメモリバスB2に出力されたリードデータRDmは、直接的にキャッシュメモリ21に供給され、キャッシュメモリ21は供給された1ライン分のデータを記憶する。

40

#### 【0027】

タグメモリ25は、キャッシュメモリ21の使用状況、キャッシュメモリ21に記憶されたラインと、メインメモリ14に記憶されたラインとを対応付けるための情報（タグ情報）、等を記憶するためのメモリである。制御回路26は、タグメモリ25に格納した情報に基づいて、キャッシュメモリ21、各バッファ22a, 22b, 23, 24における書き込み、読み出しを制御する。制御回路26は、CPUバスB1を介してCPU11と

50

接続されるとともに、メモリバス B 2 を介してメモリコントローラ 1 3 と接続されている。制御回路 2 6 には、CPU 1 1 から出力されるアクセス要求のための制御信号 S c c 及びアドレス信号 A d c が供給される。尚、制御信号 S c c は、制御回路 2 6 が CPU 1 1 に出力する応答信号を含む。

**【 0 0 2 8 】**

制御回路 2 6 は、アドレス信号 A d c に基づいて、CPU 1 1 がアクセスするアドレスのデータがキャッシュメモリ 2 1 に格納されているか否か、即ちキャッシュヒットかキャッシュミスかを判断する。上記したように、タグメモリ 2 5 には、キャッシュメモリ 2 1 に記憶されたラインと、メインメモリ 1 4 に記憶されたラインとを対応付けたタグ情報が記憶されている。制御回路 2 6 は、タグメモリ 2 5 を検索する。そして、制御回路 2 6 は、アクセス要求のアドレスがタグ情報に含まれている場合にはキャッシュヒットと判断し、アクセス要求に対応する処理を行う。また、制御回路 2 6 は、アクセス要求のアドレスがタグ情報に含まれていない場合にはキャッシュミスと判断し、アクセス要求に対応する処理を行う。

10

**【 0 0 2 9 】**

キャッシュメモリ 2 1 とメインメモリ 1 4 との対応関係、タグメモリ 2 5 に記憶する情報を、図 3 に従って説明する。

図 3 に示すように、タグメモリ 2 5 は、管理領域 3 1 とタグ領域 3 2 を含む。管理領域 3 1 には、総ライン数、使用ライン数、先頭ポインタ、末尾ポインタが格納されている。尚、カッコ内の文字は、図 3 におけるキャッシュメモリ 2 1 の状態を示す値である。図 3 には、説明を理解し易くするために、10 個のライン L 0 ~ L 9 を記憶可能なキャッシュメモリ 2 1 と、28 個のライン M 0 ~ M 2 7 に区分されたメインメモリ 1 4 を示している。メインメモリ 1 4 において、ハッチングを付したライン M 3 , M 7 , M 9 , M 1 2 , M 1 5 は、アクセス要求に回答してデータを読み出したラインを示している。そして、図中の矢印は、キャッシュメモリ 2 1 のライン L 0 , L 1 , L 2 , L 3 , L 4 , L 5 に、メインメモリ 1 4 のライン M 3 , M 1 2 , M 7 , M 1 5 , M 9 のデータが格納されていることを示す。

20

**【 0 0 3 0 】**

この状態において、キャッシュメモリ 2 1 の総ライン数は「10」であり、使用ライン数は「5」である。また、先頭ポインタは、最新のラインが書き込まれた位置を示し、末尾ポインタは、最古のラインが書き込まれた位置を示す。

30

**【 0 0 3 1 】**

タグ領域 3 2 に格納された情報は、ライン番号 3 2 a と、アドレス 3 2 b と、変更ビット 3 2 c と、有効ビット 3 2 d を含む。ライン番号 3 2 a とアドレス 3 2 b は、データが格納された領域を示すとともに、キャッシュメモリ 2 1 とメインメモリ 1 4 との対応関係（関連付け）を示す。即ち、アドレス 3 2 b にて示される領域のデータがメインメモリ 1 4 から読み出され、そのデータがキャッシュメモリ 2 1 のライン番号 3 2 a にて示されるラインに格納されている。また、ライン番号 3 2 a にて示されるラインのデータは、アドレス 3 2 b にて示されるメインメモリ 1 4 の領域に書き戻される。

**【 0 0 3 2 】**

変更ビット 3 2 c は、対応するラインのデータが CPU 1 1 により書き換えられているか否かを示すビットであり、例えば「1」の場合には書き換えられていることを示し、「0」の場合は書き換えられていないことを示す。有効ビット 3 2 d は、キャッシュメモリ 2 1 のラインが使用可能か否か、つまりデータの書き込みが可能か否かを示すビットであり、例えば「1」は使用不能を示し、「0」は使用可能を示す。

40

**【 0 0 3 3 】**

図 3 に示す状態において、リードアクセスによりキャッシュミスが発生した場合、制御回路 2 6 は、該当するアドレスのデータをメインメモリ 1 4 から読み出す。そして、制御回路 2 6 は、管理領域 3 1 の情報に基づいて、メインメモリ 1 4 から読み出したデータをライン L 5 に格納するとともに、そのライン L 5 を示す情報と、読み出したデータのアド

50

レスとを関連付けてタグ領域 3 2 に格納する。更に、データを読み出したのみであり、そのデータは書き換えられていないため、制御回路 2 6 は、変更ビット 3 2 c に「0」をセットするとともに、ライン L 5 のデータが有効であることを示すため、有効ビット 3 2 d に「1」をセットする。

**【 0 0 3 4 】**

尚、図 2 に示す制御回路 2 6 は、キャッシュメモリ 2 1 をラウンドロビン方式にて管理する。即ち、制御回路 2 6 は、メインメモリ 1 4 から読み出したデータをライン L 0 からライン L 9 に向かって順次格納する。そして、データをライン L 9 に格納した後、制御回路 2 6 は、次にメインメモリ 1 4 から読み出したデータをライン L 0 に格納する。つまり、制御回路 2 6 は、キャッシュメモリ 2 1 のラインに循環的に使用し、メインメモリ 1 4 から読み出したデータをラインに書き込む。

10

**【 0 0 3 5 】**

次に、制御回路 2 6 が実行する処理を、アクセス要求に応じて説明する。

先ず、アクセス要求がライトアクセスの場合を説明する。

制御回路 2 6 は、制御信号 S c c に基づいて、C P U 1 1 からのアクセス要求がライトアクセスと判断すると、2つのライトバッファ 2 2 a , 2 2 b のうちの何れか一方を選択し、C P U 1 1 に対してデータライトを許可する。その許可に回答して C P U 1 1 は、ライトするデータを出力する。例えば、ライトバッファ 2 2 a を選択した場合、図 4 ( a ) に示すように、ライトバッファ 2 2 a は C P U 1 1 から出力されたライトデータを格納する。これにより、C P U 1 1 における書き込み要求及びライトデータの書き込みが終了する。つまり、ライトバッファ 2 2 a , 2 2 b にライトデータを書き込むことにより、C P U 1 1 に対するライトアクセスの処理が終了する。従って、ライトアクセスを出力してからライトバッファ 2 2 a にライトデータを格納するまでの時間が、C P U 1 1 のライトアクセスに対する応答時間となる。

20

**【 0 0 3 6 】**

ライトバッファ 2 2 a に格納されたデータはキャッシュメモリ 2 1 に格納しなければならない。このため、制御回路 2 6 は、C P U 1 1 からのアクセス要求がライトアクセスと判断すると、アドレス信号 A d c に基づいて、C P U 1 1 がアクセスするアドレスのデータがキャッシュメモリ 2 1 に格納されているか否か、即ちキャッシュヒットかキャッシュミスかを判断する。上記したように、タグメモリ 2 5 には、キャッシュメモリ 2 1 に記憶されたラインと、メインメモリ 1 4 に記憶されたラインとを対応付けたタグ情報が記憶されている。制御回路 2 6 は、タグメモリ 2 5 を検索する。そして、制御回路 2 6 は、アクセス要求のアドレスがタグ情報に含まれている場合にはキャッシュヒットと判断し、アクセス要求のアドレスがタグ情報に含まれていない場合にはキャッシュミスと判断する。

30

**【 0 0 3 7 】**

キャッシュヒットの場合、制御回路 2 6 は、図 4 ( a ) に示すように、ライトバッファ 2 2 a のデータをキャッシュメモリ 2 1 のライン L c に格納する。そして、制御回路 2 6 は、このライン L c に対応する変更ビット 3 2 c を「1」にセットする。

**【 0 0 3 8 】**

一方、キャッシュミスの場合、制御回路 2 6 は、アドレスのデータを含むラインのデータを読み出すためのアクセス要求をメモリコントローラ 1 3 に出力する。このアクセス要求は、読み出しを示す信号を含む制御信号 S c m と、ラインのアドレスを指定するためのアドレス信号 A d m を含む。このアクセス要求に従って図 1 に示すメインメモリ 1 4 から読み出されたラインは、メモリコントローラ 1 3 を介してリードデータ R D m としてキャッシュ装置 1 2 に供給される。図 4 ( b ) に示すように、制御回路 2 6 は、このリードデータ R D m をキャッシュメモリ 2 1 のライン L c に格納するとともに、そのラインの番号とリードデータ R D m を読み出したアドレスとを対応付けて図 3 に示すタグメモリ 2 5 に格納する。そして、制御回路 2 6 は、キャッシュヒットの時と同様に、ライトバッファ 2 2 a のデータをキャッシュメモリ 2 1 のライン L c に格納し、このライン L c に対応する変更ビット 3 2 c を「1」にセットする。

40

50

## 【 0 0 3 9 】

キャッシュミスした場合における動作、即ち上記のライトバッファ 2 2 a ( 2 2 b ) にデータを書き込む動作と、メインメモリ 1 4 からデータを読み出してキャッシュメモリ 2 1 に書き込む動作は、データ転送経路が異なる。従って、制御回路 2 6 は、上記のライトバッファ 2 2 a ( 2 2 b ) に対するデータの書き込み処理と、メインメモリ 1 4 からデータを読み出してキャッシュメモリ 2 1 に書き込む処理とを、並行して行う。その動作を図 5 ( a ) に示す。図中、メインメモリ 1 4 を「 R A M 」、キャッシュメモリ 2 1 を「 C M 」、ライトバッファ 2 2 a を「 W B 0 」として記す。「書込要求」に対応して図面右側に向って延びる矢印は、 C P U 1 1 がライトアクセスを出力してからアクセスが許可されてデータを出力するまでの動作期間を示す。「 C P U W B 0 」は C P U 1 1 が出力するデータをライトバッファ 2 2 a に格納するまでの動作期間を示す。「 R A M C M 」は、制御回路 2 6 がメモリコントローラ 1 3 にリード要求を出力してからメインメモリ 1 4 から読み出した 1 ライン分のデータをキャッシュメモリ 2 1 に格納するまでの動作期間を示す。「 W B 0 C M 」は、ライトバッファ 2 2 a のデータを読み出してキャッシュメモリ 2 1 に格納する、即ちライトバッファ 2 2 a のデータをキャッシュメモリ 2 1 に転送する動作期間を示す。尚、上記の図 5 ( a ) ( b ) 及び以下の説明に対するタイミング図において、矢印は、動作の順番とその動作に要する期間を概念的に示すものであって、実際の動作期間を示していない場合がある。

10

## 【 0 0 4 0 】

図 5 ( a ) に示すように、 C P U 1 1 がライトバッファ 2 2 a にデータを書き込む処理 ( C P U W B 0 ) は、メインメモリ 1 4 からキャッシュメモリ 2 1 にデータを転送する処理 ( R A M C M ) の一部と、並行に実行される。そして、キャッシュメモリ 2 1 にデータが格納されると、制御回路 2 6 は、ライトバッファ 2 2 a からキャッシュメモリ 2 1 にデータを転送する ( W B 0 C M ) 。

20

## 【 0 0 4 1 】

図 5 ( b ) は、通常のライトアロケート動作を行うキャッシュ装置におけるライト動作を示す。このキャッシュ装置では、書込要求によってキャッシュミスが発生すると、メインメモリからキャッシュメモリにデータを転送 ( R A M C M ) した後に C P U に対してアクセスを許可し、 C P U からキャッシュメモリにデータが書き込まれる ( C P U C M ) 。図中、「書込要求」について実線の矢印はキャッシュミスが発生するまでを示し、二点線の矢印はアクセスが許可されるまで、即ち要求が保留される期間を示している。従って、 C P U のライトアクセスに対する応答時間は、キャッシュメモリにデータが書き込まれるまでとなる。

30

## 【 0 0 4 2 】

つまり、本実施形態のキャッシュ装置 1 2 は、ライトアロケート動作におけるキャッシュミスに対するメインメモリ 1 4 のリード動作と、 C P U 1 1 のデータ書き込み動作とを並行して行う。従って、本実施形態のキャッシュ装置 1 2 は、従来の構成に比べて応答速度が短くなる、即ち従来構成と比べて高速なアクセスを実現することができる。

## 【 0 0 4 3 】

尚、図 5 ( b ) に示す動作を行うキャッシュ装置では、キャッシュヒットの場合、書き込み要求の後に C P U からキャッシュメモリにデータが書き込まれる。従って、キャッシュミスの場合とキャッシュヒットの場合とで応答時間が異なることになる。その点、本実施形態のキャッシュ装置 1 2 では、 C P U 1 1 からライトバッファ 2 2 a ( 2 2 b ) にライトデータを書き込むため、キャッシュヒットの場合の応答速度と、キャッシュミスの場合の応答速度が同じになる。つまり、ライトアクセスにおいて常に同じ応答速度で処理を行うことができる。

40

## 【 0 0 4 4 】

尚、図 5 ( a ) の動作説明では、ライトバッファ 2 2 a の状態がデータを書き込み可能である場合について説明し。しかし、ライトバッファ 2 2 a にデータが保持され、そのデータがキャッシュメモリ 2 1 に囲まれていない場合がある。このような状態は、 1 ライン

50

分のアドレス範囲を超えるアドレスの第1データと第2データが同一のライトバッファに対して書き込まれる場合に発生する。このような場合、制御回路26は、第2データに対するアクセス要求に対するキャッシュ判定においてキャッシュミスと判定すると、その第2データメインメモリ14に対して第2データに対応するリード要求の発行以降に、第1データをキャッシュメモリ21に書き込む動作を開始する。そして、制御回路26は、キャッシュメモリ21に対する第1データの書き込みを終了すると、第2データをライトバッファ22aに格納する。従って、この場合における応答時間は、CPU11が第2データにおけるアクセス要求を出力してから第2データをライトバッファ22aに格納するまでの時間が、CPU11のライトアクセスに対する応答時間となる。

【0045】

第1データをキャッシュメモリ21に書き込む処理に要する時間と、第2データをライトバッファ22aに格納する処理に要する時間の合計値は、図5(b)に示す応答時間と比べて十分に短い。従って、本実施形態のキャッシュ装置12は、従来の構成に比べて応答速度が短くなる、即ち従来構成と比べて高速なアクセスを実現することができる。

【0046】

更に、本実施形態のキャッシュ装置12は、2つのライトバッファ22a, 22bを有している。制御回路26は、CPU11が書き込むデータのアドレスに対応して2つのライトバッファ22a, 22bを交互に使用することで、複数の連続したライトアクセスに対する応答時間を短縮することができる。

【0047】

即ち、図1に示すCPU11は、図6に示すように、1つのラインに格納されたデータのアドレスの範囲を超えるアドレスのデータD1, D2を連続的にライトアクセスする。図2に示す制御回路26は、先ず第1のライトバッファ22aを選択してこのライトバッファ22aにデータD1を格納する。次に、制御回路26は、第2のライトバッファ22bを選択してこのライトバッファ22bにデータD2を格納する。データD1, D2の書き込みが終了すると、CPU11が出力したライトアクセスに対する見かけ上の処理が終了するため、CPU11は他の処理を実行することができる。

【0048】

制御回路26は、両データD1, D2のライトアクセスがそれぞれキャッシュミスであると判断すると、それぞれのアドレスに対応するラインLa, Lbのデータをメインメモリ14から順次読み込み、キャッシュメモリ21のラインLa, Lbに書き込む。そして、制御回路26は、各ライトバッファ22a, 22bのデータを、キャッシュメモリ21のラインLa, Lbにそれぞれ書き込む。これにより、データD1, D2をキャッシュメモリ21に格納する処理を終了する。

【0049】

同様に、3つの連続したライトアクセスがキャッシュミスの場合における動作を、図7に従って説明する。

図7(a)~(c)は、ライトアクセスする3つのデータD1, D2, D3のアドレスが、1ラインのアドレス範囲を超える場合の動作を示す。データD1に対する書き込み動作を実線細線矢印で示し、データD2に対する書き込み動作を実線太線矢印で示し、データD3に対する書き込み動作を一点鎖線矢印で示している。

【0050】

図7(a)は、データD1~D3のアドレスが、1ラインのアドレス範囲を相互に超えている場合の動作波形を示す。

先ず、図1に示すCPU11は、データD1のライトアクセスを出力する。図2に示す制御回路26は、第1のライトバッファ22a(WB0)を選択してデータD1に対するアクセスを許可するとともに、そのデータD1のアドレスに対応するラインのデータのアクセス要求をメモリコントローラ13に出力する。従って、データD1をライトバッファ22aに書き込む処理(CPU WB0)と、メインメモリ14をアクセスしてデータをキャッシュメモリ21に書き込む処理(RAM CM)とを並列に実行する。

10

20

30

40

50

## 【0051】

次に、CPU 11は、アクセス許可に 응답してデータD1を出力した後、データD2のライトアクセスを出力する。制御回路26は、第2のライトバッファ22b(WB1)を選択してデータD2に対するアクセスを許可する。CPU 11は、アクセス許可に 응답してデータD2を出力した後、データD3のライトアクセスを出力する。

## 【0052】

制御回路26は、データD1に対応するメインメモリ14のデータがキャッシュメモリ21に格納される(RAM CM)と、データD2のアドレスに対応するラインのデータのアクセス要求をメモリコントローラ13に出力するとともに、第1のライトバッファ22aに格納されたデータD1をキャッシュメモリ21に転送する(WB0 CM)。

10

## 【0053】

すると、第1のライトバッファ22aに対する書き込みが可能となるため、制御回路26は、第1のライトバッファ22a(WB0)を選択してデータD3に対するアクセスを許可し、CPU 11はアクセス許可に 응답してデータD3を出力する。

## 【0054】

制御回路26は、データD2に対応するメインメモリ14のデータがキャッシュメモリ21に格納される(RAM CM)と、データD3のアドレスに対応するラインのデータのアクセス要求をメモリコントローラ13に出力するとともに、第2のライトバッファ22bに格納されたデータD2をキャッシュメモリ21に転送する(WB1 CM)。

20

## 【0055】

そして、制御回路26は、データD3に対応するメインメモリ14のデータがキャッシュメモリ21に格納される(RAM CM)と、第1のライトバッファ22aに格納されたデータD3をキャッシュメモリ21に転送する(WB0 CM)。

## 【0056】

図7(b)は、データD1, D2のアドレスが1ラインのアドレス範囲内であり、データD3のアドレス範囲がデータD1, D2のアドレスに対して1ラインのアドレス範囲を超える場合の動作波形を示す。

## 【0057】

先ず、図1に示すCPU 11は、データD1のライトアクセスを出力する。図2に示す制御回路26は、第1のライトバッファ22a(WB0)を選択してデータD1に対するアクセスを許可するとともに、そのデータD1のアドレスに対応するラインのデータのアクセス要求をメモリコントローラ13に出力する。更に、制御回路26は、出力したアクセス要求に対応するアドレス、つまりデータD1のアドレス値を含み、ラインの容量に対応するアドレスの範囲の値を図示しないレジスタに記憶する。従って、データD1をライトバッファ22aに書き込む処理(CPU WB0)と、メインメモリ14をアクセスしてデータをキャッシュメモリ21に書き込む処理(RAM CM)とを並列に実行する。

30

## 【0058】

次に、CPU 11は、アクセス許可に 응답してデータD1を出力した後、データD2のライトアクセスを出力する。制御回路26は、データD2のアドレスとレジスタに記憶した値と比較し、そのアドレスが上記のアドレス範囲内か否かを判断する。データD2のアドレスは、データD1を格納するラインのアドレスの範囲内であるため、制御回路26は、選択を変更することなくデータD2に対するアクセスを許可する。即ち、本実施形態の制御回路26は、アクセス要求に対して、そのアクセス要求に伴うアドレスを、図3に示すタグ領域32に格納したアドレスと比較するとともに、レジスタに記憶したアドレス、つまりメインメモリ14から読み出しを行っている途中のデータのアドレス範囲と比較する。そして、制御回路26は、その比較結果に基づいて、メインメモリ14に対するアクセス要求、ライトバッファ22a, 22bの選択、等の動作を行う。

40

## 【0059】

CPU 11は、アクセス許可に 응답してデータD2を出力した後、データD3のライトアクセスを出力する。制御回路26は、第2のライトバッファ22b(WB1)を選択し

50

てデータD3に対するアクセスを許可する。CPU11は、アクセス許可にตอบสนองしてデータD3を出力する。

【0060】

次に、制御回路26は、データD1に対応するメインメモリ14のデータがキャッシュメモリ21に格納される(RAM CM)と、第1のライトバッファ22aに格納されたデータD1, D2をキャッシュメモリ21に転送する(WB0 CM)。

【0061】

制御回路26は、データD3に対応するメインメモリ14のデータがキャッシュメモリ21に格納される(RAM CM)と、第2のライトバッファ22bに格納されたデータD3をキャッシュメモリ21に転送する(WB1 CM)。

10

【0062】

図7(c)は従来のライトアロケート動作を行うキャッシュ装置の動作波形を示す。

このキャッシュ装置は、データD1に対するライトアクセスにตอบสนองしてメインメモリからキャッシュメモリに該当するラインのデータを転送した後、CPUにアクセスを許可することで、CPUからキャッシュメモリにデータD1が書き込まれる。データD2, D3に対しても同様の動作を行う。

【0063】

次に、アクセス要求がリードアクセスの場合を説明する。

図2に示すリードバッファ23は、連続したリードアクセスのアドレスが、キャッシュメモリ21の1つのラインに格納したデータのアドレス範囲内のときにアクセス速度を向上する。

20

【0064】

詳述すると、制御回路26は、制御信号Sc cに基づいて、CPU11からのアクセス要求がリードアクセスと判断すると、アドレス信号Ad cに基づいて、CPU11がアクセスするアドレスのデータがキャッシュメモリ21に格納されているか否か、即ちキャッシュヒットかキャッシュミスかを判断する。

【0065】

キャッシュヒットの場合、制御回路26は、リードアクセスに伴うアドレスのデータが格納されたラインのデータをリードバッファ23に格納する。例えば、図8(a)に示すように、リードアクセスに伴うアドレスのデータがキャッシュメモリ21のラインL2に格納されている場合、制御回路26は、キャッシュメモリ21のラインL2からデータを出力させてリードバッファ23に格納する、即ちデータをキャッシュメモリ21からリードバッファ23に転送する。そして、制御回路26は、リードバッファ23からリードアクセスに対応するリードデータを出力させるとともに、このリードデータのアドレスを図示しないレジスタに記憶する。このリードデータをCPU11が受け取ることで、リードアクセスに対応する処理が終了する。従って、CPU11がリードアクセスを出力してからデータを受け取るまでの時間が、リードアクセスの応答時間となる。

30

【0066】

次に、CPU11がリードアクセスを出力すると、制御回路26は、そのリードアクセスに伴うアドレスとレジスタに格納したアドレスとを比較する。そして、制御回路26は、CPU11から新たに出力されたアドレスが、レジスタに格納したアドレスに対して、ラインに格納したデータのアドレス範囲内、即ちリードバッファ23に格納された複数のデータを示すアドレスのうちの一つである場合、そのアドレスに対応するリードデータをリードバッファ23から出力される。

40

【0067】

上記例示したように、リードバッファ23の容量は32バイトであり、CPU11が1回のアクセスにより要求するデータのサイズ(バス幅)は4バイトである。従って、制御回路26のレジスタには、リードバッファ23に格納された複数のデータの先頭アドレスと末尾アドレスが記憶される。制御回路26は、この2つのアドレスとCPU11から出力されたアドレスとを比較することで、CPU11から要求されているデータがリードバ

50

ッファ23内に格納されているか否かを判断することができる。この判断に要する時間は、キャッシュメモリ21に要求されているデータが格納されているか否か、即ちキャッシュヒットかキャッシュミスかを判断する時間に比べて短い。従って、制御回路26は、CPU11が要求するデータを、キャッシュヒット/キャッシュミス判断するよりも早く出力する。これにより、CPU11のリードアクセスに対する応答時間が短くなる、即ちアクセス速度が従来例よりも高くなる。

**【0068】**

更に、本実施形態のキャッシュ装置12は、図2に示すように、キャッシュメモリ21から読み出されたデータを直接的にCPU11に供給する経路(データバス)P1を有している。制御回路26は、図8(b)に示すように、リードアクセスに伴うアドレスのデータがキャッシュメモリ21のラインL2に格納されている場合、そのラインL2のデータをリードバッファ23に格納するとともに、リードアクセスに伴うアドレスに対応するデータを、経路P1を介して出力する。なお、図8(b)では、ラインL2に格納されている複数のデータのうち、先頭アドレスのデータ(ラインL2の左端のデータ)が要求されている場合を示す。従って、ラインL2のデータを一旦リードバッファ23に格納してからCPU11に出力する場合と比べて、CPU11がデータを受け取るまでに要する時間、即ち応答時間が短くなる。

**【0069】**

また、本実施形態のキャッシュ装置12は、図2に示すように、リードバッファ23が内部バスIB1に接続され、CPU11から出力されたライトデータWdcをリードバッファ23に記憶可能に構成されている。この構成は、同一アドレスに対するリードアクセスとライトアクセスとが連続する場合のアクセス速度を高速化する。即ち、リードバッファ23から出力されたデータを図1に示すCPU11が処理した後、処理後のデータをリードアクセスの時のアドレスと同一のアドレスにてライトアクセスする。このライトアクセスにより、CPU11から出力されるデータは、ライトバッファ22a, 22bの何れかに書き込まれることになる。例えば、ライトバッファ22aに書き込まれた場合、このライトバッファ22a内のデータと、リードバッファ23内のデータとが相違する、即ちコヒーレンシが保てなくなる。すると、次に同一のアドレスに対してリードアクセスする場合には、図9(b)に示すように、ライトバッファ22aのデータをキャッシュメモリ21に格納した後、再びキャッシュメモリ21からリードバッファ23にデータを転送しなくてはならない。しかし、本実施形態のキャッシュ装置12の制御回路26は、図9(a)に示すように、ライトアクセス(W)に回答してCPU11から出力されるデータをライトバッファ22a(WB0)に格納するとともにリードバッファ23(RB)に格納する。この動作により、このライトバッファ22a内のデータと、リードバッファ23内のデータとのコヒーレンシが保てるため、次のリードアクセス(R)において、リードバッファ23から同一アドレスのデータをCPU11に出力することができるため、応答時間が短くなる、即ち高速なアクセスを実現することができる。

**【0070】**

次に、ライトバック動作を説明する。

キャッシュメモリ21の各ラインに格納されたデータは、CPU11が実行する処理に使用される。そして、ラインに格納されたデータは、CPU11が実行する処理に応じて、書き換えられる場合がある。この場合、キャッシュメモリ21に格納されたラインのデータと、メインメモリ14に格納されたデータが相違する。このように、データが相違するラインのデータをダーティなデータといい、このラインをダーティライン(又はダーティエントリ)と呼ぶ。メインメモリ14のデータと、キャッシュメモリ21のデータとが一致するように、キャッシュメモリ21のデータをメインメモリ14の該当するアドレスに書き込む必要がある。このキャッシュメモリ21のデータをメインメモリ14に格納する動作をライトバックという。

**【0071】**

図10に示すように、キャッシュ装置12はライトバックバッファ24を有し、このラ

10

20

30

40

50

イトバックバッファ24は、キャッシュメモリ21とメモリコントローラ13との間に接続されている。キャッシュメモリ21のすべてのラインLa~Ldには、データが格納されている。図2に示す制御回路26は、アクセス要求がキャッシュミスと判断すると、ライトバックするラインLbのデータをライトバックバッファ24に転送する。そして、制御回路26は、メモリコントローラ13に対してライト要求を出力した後、ライトバックバッファ24に格納したデータをメモリコントローラ13を介してメインメモリ14に格納する。

#### 【0072】

このように、ライトバックバッファ24を利用することにより、CPU11に対する応答時間を短縮することができる。

10

即ち、図11に示すように、CPU11からアクセス要求(例えばライトアクセス)が出力され、キャッシュミスが発生すると、制御回路26は、アクセス要求のアドレスに基づいてリード要求をメモリコントローラ13に出力する。そして、制御回路26は、キャッシュメモリ21のラインLb(図10参照)のデータをライトバックバッファ24に転送する(CM WBB)。つまり、キャッシュ装置12は、図11に示すように、メインメモリ14からデータを読み出してキャッシュメモリ21に格納する動作の一部と、キャッシュメモリ21のデータをライトバックバッファ24に転送する動作を同時に行う。そして、ラインLbのデータをライトバックバッファ24に転送することにより、ラインLbにデータを書き込むことが可能となる。

#### 【0073】

20

やがて、メインメモリ14から読み出したデータがメモリコントローラ13から出力されると、制御回路26は、そのデータをキャッシュメモリ21に格納する(RAM CM)。そして、制御回路26は、CPU11にアクセスを許可し、CPU11から出力されるデータをキャッシュメモリ21(ラインLb)に格納する(CPU CM)。このデータ格納により、CPU11のアクセス要求に対する処理を終了する。つまり、CPU11がアクセス要求と出力してから、CPU11が出力するデータをキャッシュメモリ21に格納するまでの時間が、CPU11に対する応答時間となる。

#### 【0074】

上記動作に対し、ライトバックバッファを備えていないキャッシュ装置(従来例)の場合、アクセス要求によりキャッシュミスが発生すると、ラインのデータをキャッシュメモリからメインメモリにライトバックする(CM RAM)。次に、メインメモリをアクセスしてデータをキャッシュメモリに格納する(RAM CM)。そして、CPUから出力されるデータをキャッシュメモリに格納する(CPU CM)。このように、従来例は、ライトバックの期間と、メインメモリからキャッシュメモリにデータを転送する時間、CPUの要求が待たされることになる。その点、本実施形態のキャッシュ装置12は、CPU11はメインメモリ14からキャッシュメモリ21にデータを転送する時間のみアクセス要求が待たされることになり、応答時間が短くなる。そして、上記したように、図2に示すライトバッファ22a, 22bを用いることにより、更に応答時間が短くなる。

30

#### 【0075】

また、制御回路26は、ライトバックバッファ24にデータを格納した時点で、メモリコントローラ13に対してそのライトバックバッファ24のデータをメインメモリ14に格納するためのライト要求を出力する。従って、制御回路26は、キャッシュミスに対してメモリコントローラ13に出力したリード要求に続いてライト要求を出力することができる。

40

#### 【0076】

そして、メインメモリ14に書き込むデータは、ライトバックバッファ24から出力され、キャッシュメモリ21に書き込むデータは、メモリコントローラ13から直接キャッシュメモリ21に入力される。そして、効率的なメモリコントローラ13において、図12に示すように、リードデータ転送と、ライトデータ転送とが重なることがある。従って、キャッシュ装置12とメモリコントローラ13との間、メモリコントローラ13とメイ

50

ンメモリ14との間のバスにおいて、リードデータとライトデータが競合しないように、バスを分離することや、メモリコントローラ13及びメインメモリ14の動作タイミング（又はデータの出力タイミング）を調整することにより、図11に示すように、ライトバックバッファ24からデータを出力する動作を、他の動作と並行して実施することができるようになり、動作効率を高めることができる。

【0077】

以上記述したように、本実施形態によれば、以下の効果を奏する。

(1) キャッシュ装置12は、CPU11とキャッシュメモリ21との間にライトバッファ22a, 22bを有する。制御回路26は、CPU11から出力されるデータをライトバッファ22a, 22bに格納するライト動作と、キャッシュミスによりメインメモリ14からデータを読み出してキャッシュメモリ21に書き込むメモリリード動作とを同時に行う。その結果、CPU11のライトアクセスは、ライトバッファ22a, 22bにデータを格納することで終了するため、CPU11に対する応答時間が短くなり、高速なアクセスを可能とすることができる。

10

【0078】

(2) キャッシュ装置12は、2つのライトバッファ22a, 22bを有している。制御回路26は、CPU11が書き込むデータのアドレスに対応して2つのライトバッファ22a, 22bを交互に使用することで、複数の連続したライトアクセスに対する応答時間を短縮することができる。

【0079】

20

(3) キャッシュ装置12は、CPU11とキャッシュメモリ21との間にリードバッファ23を有する。制御回路26は、CPU11のアクセス要求に応答してキャッシュメモリ21から1ライン分のデータをリードバッファ23に転送し、そのリードバッファ23からリードデータを出力するようにした。その結果、リードバッファ23に格納されたデータのアドレスをアクセスする要求に対して、リードバッファ23からリードデータを出力させることで対応することができるため、応答時間が短くなり、高速なアクセスを可能とすることができる。

【0080】

(4) キャッシュ装置12は、キャッシュメモリ21から読み出されたデータを直接的にCPU11に供給する経路（データバス）P1を有している。従って、ラインL2のデータを一旦リードバッファ23に格納してからCPU11に出力する場合と比べて、CPU11がデータを受け取るまでに要する時間、即ち応答時間が短くなり、高速なアクセスを可能とすることができる。

30

【0081】

(5) キャッシュ装置12は、CPU11から出力されたライトデータWDcをリードバッファ23に記憶可能に構成されている。制御回路26は、ライトバッファ22a, 22bに格納したデータのアドレスと、リードバッファ23に格納したデータのアドレスが同じ場合に、CPU11から出力されるライトデータをライトバッファ22a, 22bに格納するとともに、リードバッファ23にも格納する。従って、同じアドレスに対するリードアクセスの場合に、ライトバッファ22a内のデータと、リードバッファ23内のデータとのコヒーレンシが保てるため、次のリードアクセス（R）において、リードバッファ23から同一アドレスのデータをCPU11に出力することができるため、応答時間が短くなる、即ち高速なアクセスを実現することができる。

40

【0082】

(6) キャッシュ装置12はライトバックバッファ24を有している。制御回路26は、アクセス要求がキャッシュミスと判断すると、ライトバックするラインLbのデータをライトバックバッファ24に転送する。そして、制御回路26は、メモリコントローラ13に対してライト要求を出力した後、ライトバックバッファ24に格納したデータをメモリコントローラ13を介してメインメモリ14に格納する。この結果、データをメインメモリ14に格納する前に、アクセス要求に対応するデータをメインメモリ14から読み出

50

すことができるため、CPU 11 に対する応答時間を短縮して高速なアクセスを実現することができる。

【0083】

尚、上記各実施形態は、以下の態様で実施してもよい。

・上記実施形態では、ライトバックバッファ 24 により、ライトバック動作におけるアクセス速度を高速化するようにしたが、その他の構成により、アクセス速度を高速化するようにしてもよい。

【0084】

例えば、ライトバックするタイミングを変更する。

上記したように、図 2 に示す制御回路 26 は、キャッシュメモリ 21 の各ラインを循環的に使用している。制御回路 26 は、図 3 に示す管理領域 31 に、キャッシュメモリ 21 の総ライン数、有効なデータを格納したライン数を示す使用ライン数、最新のデータを格納したラインを示す先頭ポインタ、現時点で最も使用していないデータを格納したラインを示す末尾ポインタを格納している。また、制御回路 26 は、図 3 に示すタグ領域 32 に、各ラインに対応する変更ビット 32c と有効ビット 32d を格納している。制御回路 26 は、これらの情報に基づいて、使用可能なラインが無くなったときに、ライトバック動作を実行する。

【0085】

制御回路 26 は、CPU 11 のアクセス要求によってキャッシュミスが発生すると、そのアクセス要求に必要なデータをメインメモリ 14 から読み出すべくリード要求をメモリコントローラ 13 に出力する。そして、制御回路 26 は、メインメモリ 14 から読み出したデータを、無効なラインに格納するとともに、そのラインを有効化する。制御回路 26 は、ラインの有効・無効を、そのラインに格納したデータが有効か無効かにより判断する。

【0086】

有効なデータは、ラインに格納されたデータをキャッシュメモリ 21 又はメインメモリ 14 に保存しなければならないデータであり、無効なデータは、保存する必要のないデータである。

【0087】

1 つの有効なデータは、CPU 11 により書き込まれたデータである。キャッシュメモリ 21 のラインには、メインメモリ 14 から読み出されたデータが格納される。このデータは、CPU 11 が実行する処理に応じて、書き換えられる場合がある。また、CPU 11 が実行する処理に応じて生成したデータをキャッシュメモリ 21 のラインに書き込む。このようなデータは、メインメモリ 14 のデータと相違するため、メインメモリ 14 に書き込まなければならないので、キャッシュメモリ 21 に保存する必要がある。

【0088】

別の有効なデータは、最近、メインメモリ 14 から読み出してキャッシュメモリ 21 のラインに格納したデータである。キャッシュ装置 12 は、データの局所性、即ち、アクセスしたデータ、又はそのデータのアドレスと連続するアドレスのデータは近い将来に高い確率でアクセスされる。従って、このようなデータは、キャッシュメモリ 21 に保存する必要がある。

【0089】

1 つの無効なデータは、ライトバックしたデータである。別の無効なデータは、最も長い間使用されなかったデータである。

制御回路 26 は、ラウンドロビン方式によりキャッシュメモリ 21 のラインを使用し、使用状況を図 3 に示す管理領域 31 の情報により管理する。上記したように、制御回路 26 は、図 13 (a) に示すように、メインメモリ 14 から読み出してライン Lc にデータを書き込み、そのライン Lc を有効化する。従って、全てのラインにデータを書き込むと、全てのライン La ~ Ld が有効となる。この時、管理領域 31 の情報、即ち末尾ポインタが示すライン Ld に格納したデータが、最も長い間使用されなかったデータである。

10

20

30

40

50

## 【 0 0 9 0 】

次に、制御回路 2 6 は、このライン L d に対応する変更ビット 3 2 c により、このライン L d のデータが変更されているか否かを判断する。データが変更されている場合、制御回路 2 6 は、このライン L d のデータをライトバックする。そして、制御回路 2 6 は、この末尾ポインタが示すライン L d に対応するタグ領域 3 2 の有効ビット 3 2 d を「 0 」にしてこのライン L d を無効化するとともに、末尾ポインタを次のライン L a を示す値に変更する。

## 【 0 0 9 1 】

つまり、制御回路 2 6 は、キャッシュミスによりメインメモリ 1 4 から読み出したデータをキャッシュメモリ 2 1 に格納した時に、ライトバックの要否を判断し、必要な場合、即ちラインのデータが変更されている場合にライトバックを実行し、そのラインを無効化する。

10

## 【 0 0 9 2 】

尚、キャッシュメモリ 2 1 の各ラインは、CPU 1 1 によりデータを書き込む、メインメモリ 1 4 から読み出したデータを格納することにより、データが上書きされる。無効化されたラインは、ライトバックを実行するまでは別のデータによる上書きを禁止し、別のデータにより上書きするまで元のデータを保持している。従って、この無効化したラインに対してアクセス要求が発生した場合、制御回路 2 6 は、無効化されていたラインを有効とする、即ちラインに対応する有効ビット 3 2 d に「 1 」をセットすることもできる。この場合も、メインメモリ 1 4 からデータを読み出す場合と比べて、応答時間が短くなる、即ち高速なアクセスを行うことができる。

20

## 【 0 0 9 3 】

このように、無効化されていたラインを有効とすると、使用可能なラインが無くなる。そして、この有効化したラインは、最近使用したデータを含むラインである。従って、制御回路 2 6 は、この有効化したラインのライン番号を、先頭ポインタに設定する。すると、使用可能なラインが無くなる。このため、制御回路 2 6 は、上記と同様に、末尾ポインタが示すラインについて、ライトバックの要否を判断して必要な処理を行った後、このラインを無効化する。

## 【 0 0 9 4 】

そして、無効化したラインには、キャッシュミスにより発生したメモリアクセス、即ちメインメモリ 1 4 をアクセスして読み出したデータを格納することができる。つまり、制御回路 2 6 は、CPU 1 1 のアクセス要求によりキャッシュミスが発生した場合、メモリコントローラ 1 3 にリード要求を出力する。そして、制御回路 2 6 は、メインメモリ 1 4 から読み出したデータを、上記の処理によって無効化したラインに格納する。言い換えると、キャッシュミスに対応するデータを書き込むラインを、そのキャッシュミスより先に実行するメインメモリ 1 4 のアクセスに応じて無効化する。また、キャッシュミスに対応するデータを書き込むラインのデータを、そのキャッシュミスより先に実行するメインメモリ 1 4 のアクセスに応じてライトバックする、ということもできる。つまり、制御回路 2 6 は、キャッシュミスによりデータを書き込むラインのデータ、そのキャッシュミスより先行してライトバックする。従って、キャッシュミスが発生した時には常に無効なラインが存在することになるため、キャッシュミスの後にライトバックする方式と比べて応答時間を短くすることができる。

30

40

## 【 0 0 9 5 】

なお、図 1 3 ( a ) を用いた説明においては、無効化したラインとは、キャッシュミスがあった場合に、メインメモリからキャッシュメモリに読み出してくるデータを格納する次のラインと読み替えてもよい。

## 【 0 0 9 6 】

図 1 3 ( b ) は、上記のようにライトバックを行わないキャッシュ装置における動作を示す。尚、符号は本実施形態と同じ符号を付す。キャッシュメモリ 2 1 の全てのライン L a ~ L d が使用されている。この時、キャッシュメモリ 2 1 のライン L c のデータをメイ

50

ンメモリ 14 にライトバックした後、メインメモリ 14 から読み出したデータをライン Lc に格納する。

【0097】

図 14 は、ライトアクセスに対するキャッシュ装置 12 と、従来例、つまり先行ライトバックを行わないキャッシュ装置の動作を示す。

本実施形態のキャッシュ装置 12 は、アクセス要求によりキャッシュミスが発生すると、メインメモリ 14 をアクセスしてデータをキャッシュメモリ 21 に格納する (RAM CM)。そして、CPU 11 から出力されるデータをキャッシュメモリ 21 に格納し (CPU CM)、次のラインのデータをキャッシュメモリ 21 からメインメモリ 14 にライトバックする (CM RAM)。CPU 11 に対する応答時間は、CPU 11 がアクセス要求を出力してから、CPU 11 から出力されるデータをキャッシュメモリ 21 に格納するまでの期間となる。

10

【0098】

従来例のキャッシュ装置は、図 14 の下段に示すように、アクセス要求によりキャッシュミスが発生すると、ラインのデータをキャッシュメモリからメインメモリにライトバックする (CM RAM)。次に、メインメモリをアクセスしてデータをキャッシュメモリに格納する (RAM CM)。そして、CPU から出力されるデータをキャッシュメモリに格納する (CPU CM)。このように、従来例は、ライトバックの期間と、メインメモリからキャッシュメモリにデータを転送する時間、CPU の要求が待たされることになる。その点、図 14 の上段に示す動作では、CPU はメインメモリからキャッシュメモリにデータを転送する時間のみアクセス要求が待たされることになり、応答時間が短くなる。そして、上記したように、図 2 に示すライトバッファ 22 a, 22 b を用いることにより、更に応答時間が短くなる。

20

【0099】

尚、上記の説明では、ライトバックしたラインの有効ビット 32 d に「0」をセットして無効化した。有効ビット 32 d を「1」のままとしてもよい。即ち、図 15 (a) に示すように、キャッシュメモリ 21 のライン L0 ~ L8 までデータを格納した。従って、管理領域 31 の先頭ポインタは L8 を示し、末尾ポインタは L0 を示す。この状態において、キャッシュミスが発生すると、制御回路 26 は、図 15 (b) に示すように、ライン L9 にデータを格納し、先頭ポインタに L9 をセットするとともに、有効ビット 32 d に「1」をセットする。そして、制御回路 26 は、図 15 (a) にて末尾ポインタが示すライン L0 のデータをライトバックした後、図 15 (b) に示すように、末尾ポインタに L1 をセットする。

30

【0100】

次に、キャッシュミスが発生した場合、上記と同様に、図 15 (c) に示すように、ライン L0 にデータを格納し、先頭ポインタに L0 をセットする。そして、制御回路 26 は、図 15 (b) にて末尾ポインタが示すライン L1 のデータをライトバックした後、図 15 (c) に示すように、末尾ポインタに L2 をセットする。

【0101】

このように、ライトバックしたラインの有効ビット 32 d を「1」のままとすることで、有効ライン数は、減少しない。つまり、キャッシュヒットとなるライン数が減少しないため、アクセス要求の総数に対するキャッシュヒットの数の比率、即ちヒット率を高くすることができるようになる。

40

【0102】

・上記実施形態では、CPU 11 とライトバッファ 22 a, 22 b との間のバス幅と、ライトバッファ 22 a, 22 b とキャッシュメモリ 21 との間のバス幅と同じ値に設定したが、バス幅が相違するようによい。

【0103】

例えば、図 16 に示すように、ライトバッファ 22 a, 22 b とキャッシュメモリ 21 との間のバス幅を、CPU 11 とライトバッファ 22 a, 22 b の間のバス幅の n 倍 (n

50

= 2, 4, 8, ...) の構成とする。例えば、2倍とすると、図18に示すように、ライトバッファ22aに対して2回のアクセスにより書き込んだデータを、1回でキャッシュメモリ21に転送することができる。従って、ライトバッファ22a, 22bからキャッシュメモリ21へデータ転送する転送サイクル数を削減可能となる。このため、ライトバッファ22a, 22bのデータをキャッシュメモリ21に転送する動作が、CPU11からライトバッファ22a, 22bにデータを書き込む動作に対する影響を低減することができる。つまりライトバッファ22a, 22bからキャッシュメモリ21にデータを転送している間に、CPU11がライトアクセスを出力すると、そのライトアクセスは、ライトバッファ22a, 22bからキャッシュメモリ21へのデータ転送を終了するまで待たされることになる。このため、ライトバッファ22a, 22bからキャッシュメモリ21へのデータ転送にかかる時間を短くすることで、このデータ転送とCPU11のライトアクセスが重なり難くなる、またはライトアクセスが許容されるまでの時間が短くなる、即ちアクセス要求に対するレイテンシを低減することができる。

10

## 【0104】

また、図17に示すように、ライトバッファ22a, 22bとキャッシュメモリ21との間のバス幅を、ライトバッファ22a, 22bの容量値と等しくする。この場合、ライトバッファ22a, 22bからキャッシュメモリ21へデータ転送する転送サイクル数を更に削減可能となり、レイテンシを低減することができる。更に、ライトバッファ22a, 22bに、ライトバッファ22a, 22bの容量値と等しい転送量であって、その後ライトバッファ22a, 22bからキャッシュメモリ21に書き込まれるデータが転送(例えば、バースト転送でもよい)された場合に、キャッシュ判定のヒット又はミスによらずとも、メインメモリ14へのリード要求をキャンセルする。その後、ライトバッファ22a, 22bに書き込まれたデータを、キャッシュメモリ21のラインに含まれる全てのデータに対して上書きするように書き込む。従って、メインメモリ14からキャッシュメモリ21にデータを転送する必要がなくなる。このため、CPU11のアクセス要求がライトアクセスの場合、キャッシュミスが発生しても、メインメモリ14からキャッシュメモリ21にデータを転送する処理を行わないようにする。これにより、ライトアクセスに対するレイテンシを削減することができ、ライトアクセスにおけるアクセスを高速化することができる。そして、ライトアクセスにおいてメインメモリ14からデータを読み出す処理を省略するため、メインメモリ14に対するアクセス数を低減することができる。

20

30

## 【0105】

・上記実施形態では、キャッシュメモリ21を1つとして示したが、複数のRAM(複数チップ又は複数ブロック)によりキャッシュメモリを構成してもよい。この構成により、複数のアクセス要求を連続して処理することができるようになる。

## 【0106】

即ち、図18に示すように、メインメモリ14に対するリード要求とライト要求が交互に2回発生すると、リード要求に対するリード動作と、ライトバックバッファ24へのライト動作が重なる場合がある。

## 【0107】

図19(a)に示すように、全てのラインL0~L9を含む1つのキャッシュメモリ21の場合、キャッシュミスによって使用するラインのデータをライトバッファに格納し、そのラインにメインメモリ14から読み出したデータをその格納する。そのアクセス要求の順序は以下ようになる。尚、以下の説明において、キャッシュメモリ21のラインからライトバックへの転送を「ラインnのリード」(nはライン番号)と表記し、メインメモリ14からラインmへの転送を「ラインmのライト」(mはライン番号)と表記する。

40

[使用するラインがラインL0の場合]

<1-1>ラインL0のリード、ラインL0のライト。

<1-2>ラインL1のリード、ラインL1のライト。

[使用するラインがラインL1の場合]

<2-1>ラインL1のリード、ラインL1のライト。

50

< 2 - 2 >ラインL 2のリード、ラインL 2のライト。

【 0 1 0 8 】

ここで、< 1 - 1 >のライト動作と、< 1 - 2 >のリード動作が重なる。ラインL 1から使用する場合も同様である。このため、これらの処理を同時に実行することができない。

【 0 1 0 9 】

このため、図19(b)に示すように、2つのキャッシュメモリ41, 42を含む構成とし、各キャッシュメモリ41, 42に対して、ライン番号を交互に割り当てる。すると、< 1 - 1 >におけるライト動作はキャッシュメモリ41に対して実行され、< 1 - 2 >のリード動作はキャッシュメモリ42に対して実行される。つまり、異なるキャッシュメモリに対する動作となるため、同時に実行することができる。

10

【 0 1 1 0 】

上記したように、先行ライトバックを行う場合、1つのキャッシュメモリ21に対するアクセス要求は、以下のようになる。

[使用するラインがラインL 0の場合]

< 3 - 1 >ラインL 1のリード、ラインL 0のライト。

< 3 - 2 >ラインL 2のリード、ラインL 1のライト。

[使用するラインがラインL 1の場合]

< 4 - 1 >ラインL 2のリード、ラインL 1のライト。

< 4 - 2 >ラインL 3のリード、ラインL 2のライト。

20

【 0 1 1 1 】

ここで、< 3 - 1 >のライト動作と、< 3 - 2 >のリード動作が重なる。ラインL 1から使用する場合も同様である。このため、これらの処理を同時に実行することができない。

【 0 1 1 2 】

このため、キャッシュ装置を2つのキャッシュメモリ51, 52を含む構成とする。そして、一方のキャッシュメモリにおいてライン番号4Nとライン番号4N+3が連続するように設定し、他方のキャッシュメモリにおいてライン番号4N+1とライン番号4N+2で連続するように設定し、それらの規則を両キャッシュメモリに交互に割り当てる(なお、Nは例えば0以上とする)。例えば、図19(c)に示すように、キャッシュ装置を2つのキャッシュメモリ51, 52を含む構成とする。そして、キャッシュメモリ51においてラインL 0とラインL 3が連続するように設定し、キャッシュメモリ52において、ラインL 1とラインL 2とが連続するように設定する。次に、キャッシュメモリ52において、ラインL 2とラインL 5が連続するように設定し、キャッシュメモリ51において、ラインL 3とラインL 4とが連続するように設定する。すると、< 3 - 1 >におけるライト動作はキャッシュメモリ51に対して実行され、< 3 - 2 >のリード動作はキャッシュメモリ52に対して実行される。つまり、異なるキャッシュメモリに対する動作となるため、同時に実行することができる。

30

【 0 1 1 3 】

・上記実施形態では、メインメモリ14のデータを保持するキャッシュ装置12について説明したが、キャッシュ装置12に保持するデータはメインメモリ14に限らず、磁気ディスク装置、光ディスク装置、光磁気ディスク装置等の媒体記憶装置におけるキャッシュ装置に適用してもよい。

40

【 0 1 1 4 】

・上記実施形態に基づき、ライトバッファとリードバッファとライトバックバッファのうちの少なくとも1つを含むキャッシュ装置に具体化してもよい。

・上記実施形態のキャッシュ装置12は、ライトバッファ22a, 22bとリードバッファ23とを含む構成としたが、リードライト可能な複数のバッファを含む構成としてもよい。また、キャッシュ装置が含む各バッファの数を適宜変更してもよい。

【 0 1 1 5 】

50

・上記各実施形態では、シングルポートのスタティックRAMをキャッシュメモリ21に用いたが、デュアルポートのRAMを用いてもよい。

上記各実施形態に関し、以下の付記を開示する。

(付記1)

処理装置と記憶装置との間に介在されるキャッシュ装置であって、  
前記記憶装置のデータを記憶するキャッシュメモリと、  
前記処理装置から出力される出力データを保持するバッファと、  
前記記憶装置に対するアクセス要求に基づいてキャッシュ判定をし、キャッシュミスの場合には前記アクセス要求に対応して前記出力データを前記バッファに保持し、前記アクセス要求により要求されたデータを含むラインのデータを前記記憶装置から読み出すためのリード要求を出力し、前記記憶装置の前記ラインから出力されたデータを前記キャッシュメモリに格納し、前記バッファに保持してある前記出力データを前記キャッシュメモリに格納させる制御回路と  
を有することを特徴とするキャッシュ装置。

10

(付記2)

前記キャッシュメモリは、複数のラインを含み、各ラインは前記出力データの容量のn倍(nは2以上の整数)に設定され、前記ラインに前記記憶装置のデータを記憶し、  
前記バッファは、少なくとも前記キャッシュメモリの1ライン分の記憶容量に設定されてなること、  
を特徴とする付記1記載のキャッシュ装置。

20

(付記3)

前記バッファは複数設けられ、  
前記出力データのアドレスが前記ラインのアドレス範囲を超える場合に前記データを格納する前記バッファに切り替えること  
を特徴とする付記1又は2記載のキャッシュ装置。

(付記4)

前記バッファから前記キャッシュメモリに前記出力データを転送する転送量は前記ラインの容量値と等しく設定され、  
前記アクセス要求がライトアクセスの場合に、そのライトアクセスにより発生するキャッシュミスに対する前記記憶装置からのデータリードをキャンセルすること、  
を特徴とする付記1～3のうちの何れか1項に記載のキャッシュ装置。

30

(付記5)

処理装置と記憶装置との間に介在されるキャッシュ装置であって、  
前記記憶装置のデータを記憶するキャッシュメモリと、  
前記キャッシュメモリから前記処理装置に出力される入力データを保持する第1バッファと、を有すること  
を特徴とするキャッシュ装置。

(付記6)

前記キャッシュメモリから前記処理装置に対して入力データを直接出力する経路を有し、  
前記キャッシュメモリから前記第1バッファに1ライン分のデータを転送する際に、前記アクセス要求に対応するデータを前記第1バッファに格納するとともに前記処理装置に出力させる、  
を特徴とする付記5記載のキャッシュ装置。

40

(付記7)

前記処理装置から出力される出力データを保持する第2バッファを有し、  
前記アクセス要求により前記第2バッファに前記ライトデータを格納する場合に、該出力データのアドレスが前記第1バッファのアドレス範囲内の場合には前記出力データを前記第2バッファに格納すること、  
を特徴とする付記5又は6記載のキャッシュ装置。

50

## (付記 8)

処理装置と記憶装置との間に介在されるキャッシュ装置であって、  
前記記憶装置のデータを記憶するキャッシュメモリと、  
前記キャッシュメモリから前記記憶装置にライトバックするデータを保持するバッファと、  
を有し、  
前記記憶装置に対するリード要求を出力し、前記キャッシュメモリからライトバックするデータを前記バッファに転送し、前記記憶装置から読み出したデータを前記キャッシュメモリに格納すること、  
を特徴とするキャッシュ装置。

10

## (付記 9)

処理装置と記憶装置との間に介在されるキャッシュ装置であって、  
前記記憶装置のデータを記憶するキャッシュメモリを有し、  
前記キャッシュメモリは複数のラインを含み、複数の前記ラインを順次使用して前記記憶装置のデータを格納し、  
前記記憶装置のデータを格納可能なラインを使用可能ラインとし、  
前記使用可能ラインが無くなったときに、次にキャッシュミスが発生した場合に使用するラインのデータを前記記憶装置にライトバックすること、  
を特徴とするキャッシュ装置。

20

## (付記 10)

各ラインのデータが有効か・無効かを管理し、前記ライトバックしたラインのデータを無効化すること、  
を特徴とする付記 9 記載のキャッシュ装置。

20

## (付記 11)

無効化した前記ラインに対してキャッシュヒットした場合には当該ラインを有効化するとともに、次にキャッシュミスが発生したときに使用するラインのデータを前記記憶装置にライトバックすること、  
を特徴とする付記 10 記載のキャッシュ装置。

## (付記 12)

各ラインのデータが有効か・無効かを管理し、前記ライトバックしたラインのデータを有効のままとすること、  
を特徴とする付記 9 記載のキャッシュ装置。

30

## (付記 13)

処理装置から出力される出力データを保持するバッファと、  
前記バッファが出力する前記出力データを記憶するキャッシュメモリと、  
前記キャッシュメモリの第 1 ラインに対応する第 1 データが前記バッファに格納されている場合に前記キャッシュメモリにおける前記第 1 ラインとは異なる第 2 ラインに対応する第 2 のデータのキャッシュ判定においてキャッシュミスと判定されると、記憶装置に前記第 2 データに対応するリード要求の発行以降に前記第 1 データの前記第 1 ラインへの書き込みを開始させる制御回路と、  
を有することを特徴とするキャッシュ装置。

40

## 【符号の説明】

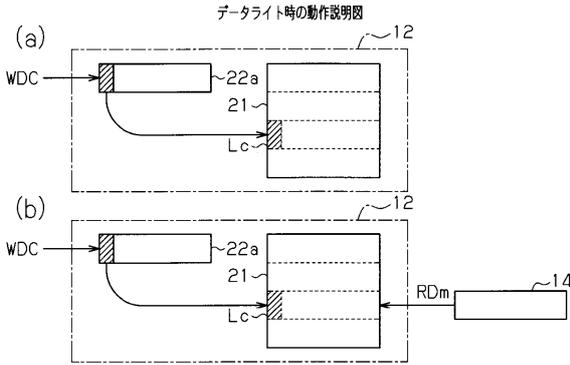
## 【 0 1 1 6 】

- 1 1 CPU (処理装置)
- 1 2 キャッシュ装置
- 1 4 メインメモリ (記憶装置)
- 2 1 キャッシュメモリ
- 2 2 a, 2 2 b ライトバッファ
- 2 3 リードバッファ
- 2 4 ライトバックバッファ

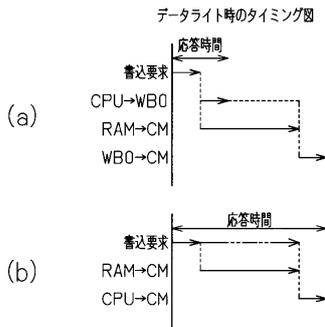
50



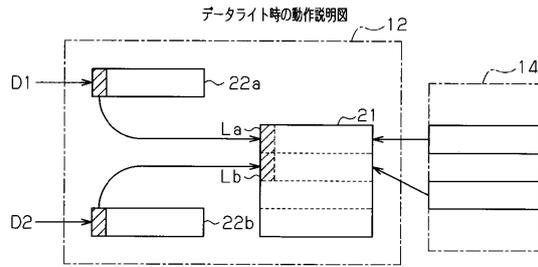
【図4】



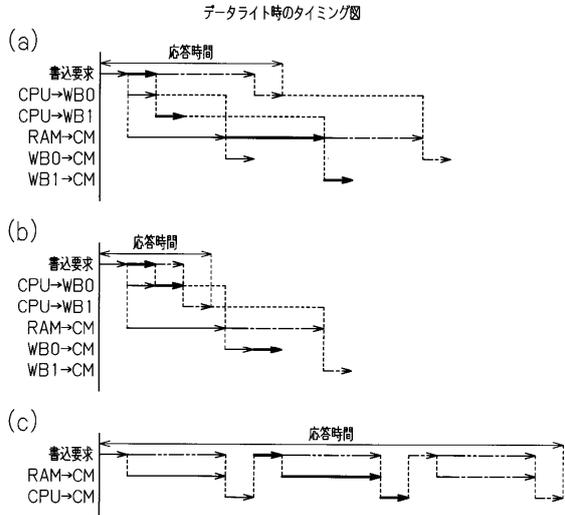
【図5】



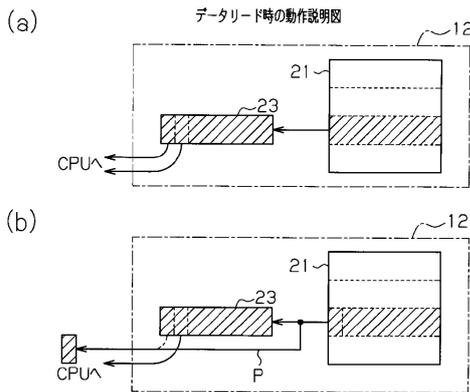
【図6】



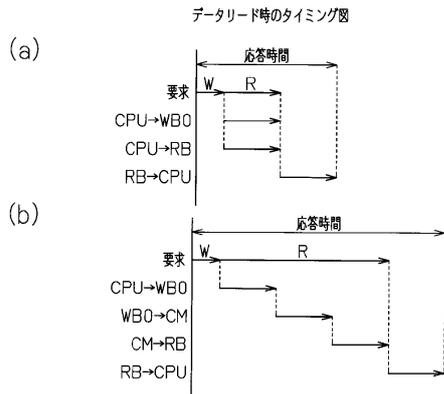
【図7】



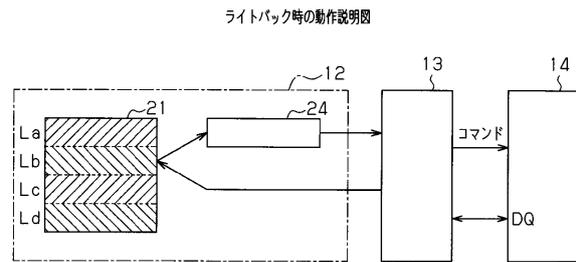
【図8】



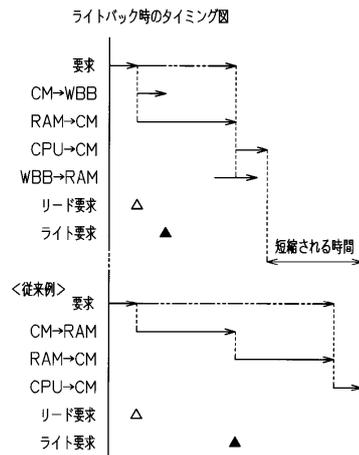
【図9】



【図10】

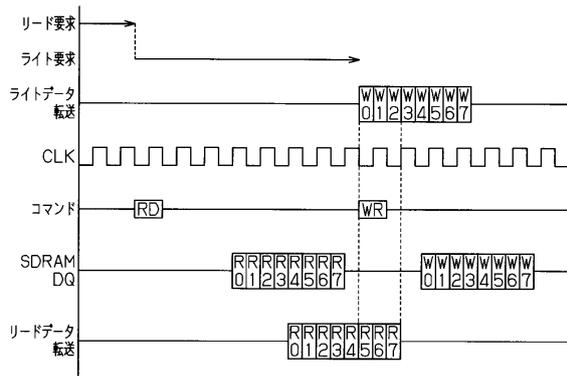


【図11】



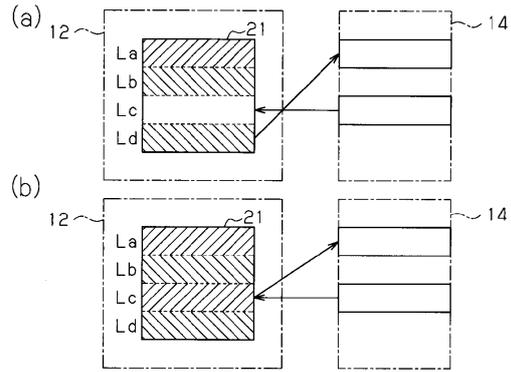
【図12】

ライトバック時のタイミング図



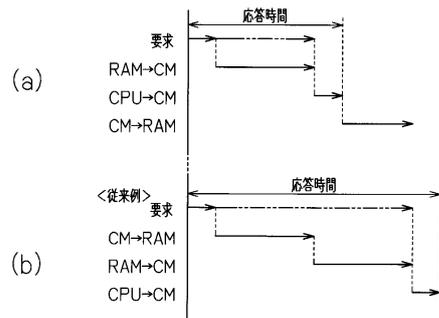
【図13】

ライトバック時の動作説明図



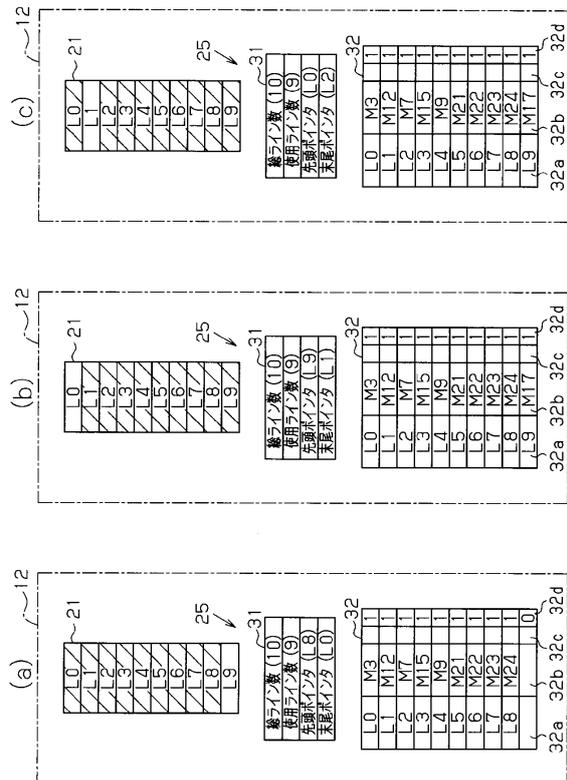
【図14】

ライトバック時のタイミング図



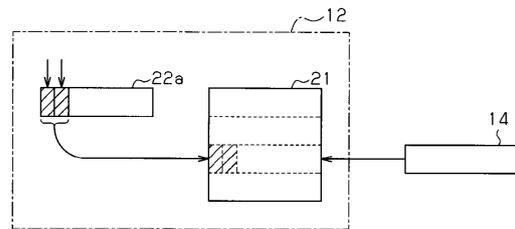
【図15】

ライトバック時の動作説明図



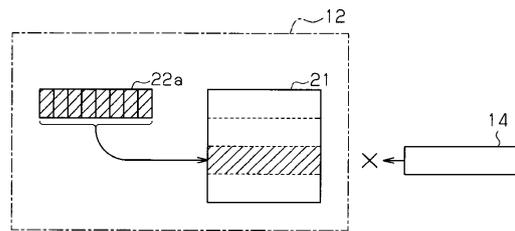
【図16】

データライト時の動作説明図

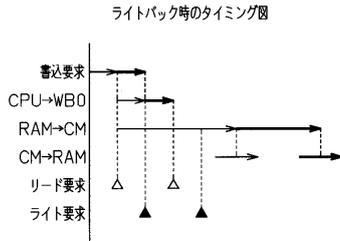


【図17】

データライト時の動作説明図

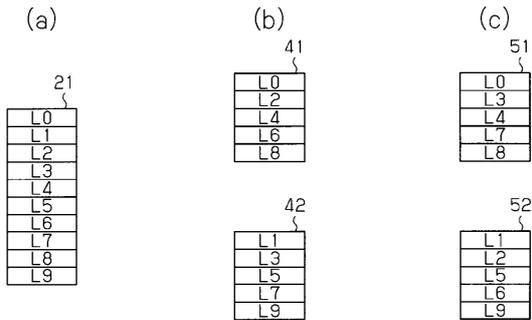


【 図 1 8 】



【 図 1 9 】

キャッシュメモリのライン設定の説明図



---

フロントページの続き

- (56)参考文献 特開昭53-148343(JP,A)  
特開平05-250258(JP,A)  
特開平07-152566(JP,A)  
特開平02-207351(JP,A)  
特開2006-244460(JP,A)  
特開平02-208763(JP,A)  
特開平01-314354(JP,A)  
特開平02-236748(JP,A)  
特開平04-367953(JP,A)  
特開平09-114734(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 12/08 - 12/12