

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3709235号  
(P3709235)

(45) 発行日 平成17年10月26日(2005.10.26)

(24) 登録日 平成17年8月12日(2005.8.12)

(51) Int.Cl.<sup>7</sup>

G06F 12/08

F I

G06F 12/08 507Z

G06F 12/08 511Z

G06F 12/08 515Z

G06F 12/08 551Z

G06F 12/08 553Z

請求項の数 1 (全 16 頁)

(21) 出願番号	特願平8-58025	(73) 特許権者	000005223
(22) 出願日	平成8年3月14日(1996.3.14)		富士通株式会社
(65) 公開番号	特開平8-272688		神奈川県川崎市中原区上小田中4丁目1番1号
(43) 公開日	平成8年10月18日(1996.10.18)	(74) 代理人	100070150
審査請求日	平成14年8月19日(2002.8.19)		弁理士 伊東 忠彦
(31) 優先権主張番号	406305	(72) 発明者	モーゲンス ローリッツェン
(32) 優先日	平成7年3月17日(1995.3.17)		アメリカ合衆国 カリフォルニア州 94024 ロス・アルトス オークハースト・アヴェニュー 1310番
(33) 優先権主張国	米国(US)		
早期審査対象出願			
前置審査		審査官	清木 泰
			最終頁に続く

(54) 【発明の名称】 階層化された記憶項目とキャッシュタグを単一キャッシュアレイ構造に格納するキャッシュメモリ装置及び方法

## (57) 【特許請求の範囲】

## 【請求項1】

外部メモリから論理ユニットまで階層化された記憶項目を提供するキャッシュメモリ装置であって：

記憶容量が固定である一次キャッシュと；

上記階層化された記憶項目を格納する記憶項目部と、一つのキャッシュエントリーを識別するキャッシュタグを格納するキャッシュタグ部とからなる複数のキャッシュエントリーを有する少なくとも一つの単一キャッシュアレイ構造により構成され、上記単一キャッシュアレイの数を決定することにより記憶容量を変えることが可能であり、第二アドレス信号に応答し少なくとも一つの上記キャッシュエントリーの内容を読み出す二次キャッシュと；

10

上記キャッシュタグと比較するための第一アドレス信号と、上記複数のキャッシュエントリーから1つのキャッシュエントリーを選択するための第二アドレス信号と、上記1つのキャッシュエントリー内の一部を選択するための第三アドレス信号を発生し、上記キャッシュタグを上記第一アドレス信号と比較する手段を有する論理回路と；

上記キャッシュアレイと上記論理回路を相互接続し、上記第二アドレス信号を上記キャッシュアレイに伝達するアドレスバスと；

上記キャッシュアレイと上記論理回路を相互接続し、上記キャッシュアレイと上記論理回路の間で上記キャッシュタグと上記記憶項目とを伝達するデータバスと；

を備え、上記キャッシュエントリーは、連続的なキャッシュエントリーの系列からなる

20

複数のキャッシュラインに分類され、上記キャッシュアレイは上記キャッシュラインの内容を上記データベースを介して順次に送出させる手段を更に有し、

キャッシュラインがキャッシュタグにより識別され、キャッシュラインのキャッシュデータは複数のキャッシュエントリーにわたって格納され、各キャッシュエントリーは単一キャッシュアレイ構造内にあり、各キャッシュエントリーは記憶項目部とキャッシュタグ部から成り、キャッシュタグは複数のセグメントに分割され、分割された各セグメントはそれぞれ異なるキャッシュエントリーのキャッシュタグ部に格納される

ことを特徴とするキャッシュメモリ装置。

【発明の詳細な説明】

【0001】

10

【発明の属する技術分野】

本発明は、一般的にキャッシュメモリ装置と方法に係り、特に、階層化された記憶項目とキャッシュタグを単一キャッシュアレイ構造内に格納するキャッシュメモリ装置と方法に関する。

【0002】

【従来の技術】

現在のコンピュータシステムは、アクセス時間と記憶容量に関し記憶装置を順序付けする階層的なメモリ設計を採用している。参考のため引用したストーン(H.S. Stone)の“高性能コンピュータアーキテクチャ(High-Performance Computer Architecture)”、第2版(1990年)、アディソン・ウェスレイ(Addison-Wesley)発行、29-87ページに記載されているように、最高速のアクセス時間と、最小の記憶容量が得られるキャッシュメモリ装置は階層の最上位に置かれる。キャッシュメモリ装置は、コンピュータのプログラムとデータの中で最も頻繁に使用されるセグメントを格納するため使用される。キャッシュメモリよりも下位の階層には、アクセス時間が低下し、記憶容量が増加する順序で、主メモリと2次記憶装置が置かれている。メモリキャッシュは、主メモリから階層化された2種類の記憶項目：読み出し専用として扱われるプログラム命令と、データを格納するため使用される。

20

【0003】

キャッシュメモリ装置は：中央処理装置(CPU)の一体的な部分としてオンチップに設けられた1次又はレベル-1キャッシュと、CPUの外部にある2次又はレベル-2キャッシュの二つのタイプに分類される。1次キャッシュは、より高速のアクセス時間が得られるが、寸法、電力消費、及びCPUに関連する他のエンジニアリング上の得失によって1次キャッシュの設計は制限され得るので、記憶容量は固定されている。

30

【0004】

一方、2次キャッシュはCPUから分離した物理的構成部品のため、上記得失による影響を受けないので、付加的なメモリチップを追加することにより容量を増加させることが可能である。典型的に、スタティックランダムアクセスメモリ(SRAM)チップは、通常のダイナミックランダムアクセスメモリ(DRAM)チップよりも高速のアクセス時間が得られるので使用される。

【0005】

40

2次メモリキャッシュの容量は、一般的に主メモリの容量よりも著しく小さいので、各キャッシュエントリーが対応する主メモリの位置を識別するためにキャッシュタグを使用することが必要である。従って、二つの別個のメモリチップの実体が使用され、これにより、実体は少なくとも一つの離散的なメモリチップである可能性があり、一方の実体は階層化された記憶項目を格納し、もう一方の実体がキャッシュタグを格納する。階層化された記憶項目とキャッシュタグを格納するため別個のメモリチップの実体を利用する2次メモリキャッシュの例は、クロフォード(Crawford)他の米国特許第5,210,845号明細書と、リー(Lee)他の米国特許第5,339,399号明細書と、マックウィリアムズ(MacWilliams)他の米国特許第5,239,603号及び第5,228,134号明細書に記載されている。

【0006】

50

**【発明が解決しようとする課題】**

階層化された記憶項目とキャッシュタグを格納するための別個のメモリアレイ構造の使用によって生じる問題は、付加的な構成部品のためにアーキテクチャ上の複雑さが増大し、製造コストが増加し、動作速度が低下することである。

本発明は、上記従来技術の問題点を解決するため、階層化された記憶項目（プログラムの命令又はデータ）とキャッシュタグを格納する単一メモリアレイを使用するキャッシュメモリ装置及び方法の提供を目的とする。

**【0007】****【課題を解決するための手段】**

上記目的を達成するため、本発明の階層化された記憶項目（プログラムの命令又はデータ）とキャッシュタグを格納する単一メモリアレイにおいて、キャッシュタグは複数のメモリチップ上に広がる連続的なキャッシュエントリに亘って分散され、タグは2次メモリアレイ構造への連続的な読みを行なうことにより再構築される。

**【0008】**

本発明の一実施例によれば、外部メモリから論理ユニットまで階層化された記憶項目を提供するキャッシュメモリ装置が得られる。当該装置は、記憶容量が固定である一次キャッシュと；上記階層化された記憶項目を格納する記憶項目部と、一つのキャッシュエントリを識別するキャッシュタグを格納するキャッシュタグ部とからなる複数のキャッシュエントリを有する少なくとも一つの単一キャッシュアレイ構造により構成され、上記単一キャッシュアレイの数を決定することにより記憶容量を変えることが可能であり、アドレス信号に応答し少なくとも一つの上記キャッシュエントリの内容を読み出す二次キャッシュと；上記キャッシュタグと比較するための第一アドレス信号と、上記複数のキャッシュエントリから1つのキャッシュエントリを選択するための第二アドレス信号と、上記1つのキャッシュエントリ内の一部を選択するための第三アドレス信号を発生し、上記キャッシュタグを上記第一アドレス信号と比較する手段を有する論理回路と；上記キャッシュアレイと上記論理回路を相互接続し、上記アドレス信号を上記キャッシュアレイに伝達するアドレスバスと；上記キャッシュアレイと上記論理回路を相互接続し、上記キャッシュアレイと上記論理回路の間で上記キャッシュタグと上記記憶項目とを伝達するデータバスとからなるキャッシュメモリ装置である。

**【0009】**

本発明の他の実施例によれば、階層化された記憶項目を格納する部分と、上記階層化された記憶項目を識別するキャッシュタグを格納する部分とからなる単一キャッシュアレイ構造内に上記階層化された記憶項目と上記キャッシュタグとを格納するキャッシュメモリ装置が得られる。当該装置は：実行ユニットからのメモリ要求に応じて論理回路を用いてアドレス信号を発生する手段、及び、上記論理回路と上記キャッシュアレイ構造を相互接続するアドレスバスに沿って上記アドレス信号をアサートする手段と；上記構造の連続的な部分にあり、複数の上記階層化された記憶項目とキャッシュラインに対応する一つのキャッシュタグとからなるキャッシュラインを、上記アドレス信号に応じて選択する手段と；上記複数の階層化された記憶項目の各々と、上記対応するキャッシュタグを、上記論理回路と上記キャッシュアレイ構造を相互接続するデータバスに沿って上記論理回路に順次に送る手段と；上記キャッシュタグと上記アドレス信号の間の一致よりなる初期キャッシュヒット、又は、上記キャッシュタグと上記アドレス信号の間の不一致よりなる初期キャッシュミスを検出する手段とからなるキャッシュメモリ装置である。

**【0010】**

本発明を実施するため考え得るその例に限定されることのない最良の形態の例による本発明の実施例だけが以下の詳細な説明に示され説明されている。当業者は、かかる以下の詳細な説明に基づいて本発明の他の実施例を容易に知ることができる。本発明の精神と範囲から逸れることなく、本発明の他の種々の実施例を形成することが可能であり、かつ、本発明の幾つかの細部は、種々の明らかな点で変更し得ることが分かる。従って、添付図面及び以下の説明は本質的に例示であり、その例に限定されるものではないと見なす必要が

10

20

30

40

50

ある。

【 0 0 1 1 】

【 発明の実施の形態 】

本発明は 2 個の別々の S R A M チップ上に分散した 4 個の 3 ビットセグメントを結合することによりキャッシュタグが形成される 2 次メモリキャッシュに関する。本発明の一実施例は、パイプライン化された実行ユニットと共に縮小命令セットコンピュータ ( R I S C ) 設計を利用する S P A R C ( 登録商標 ) ( スケーラブル プロセッサ アーキテクチャー ) マイクロプロセッサ上で実現されている。 S P A R C は、カリフォルニア州、マウンテン ビューのサンマイクロシステムズ社の登録商標である。

【 0 0 1 2 】

本発明は、更に、2重キャッシュタグが、多重の記憶場所への読出しに完全なキャッシュタグを得ることが必要とされる環境で2次キャッシュの多重の場所に格納された部分に分割され、これにより、冗長なキャッシュタグを利用してキャッシュラインの多重記憶場所の全部が読まれる前にキャッシュヒット又はキャッシュミスを検出することが可能になる2次メモリキャッシュに関する。

【 0 0 1 3 】

図 1 には本発明の一実施例のキャッシュメモリ装置の概略図が示されている。キャッシュメモリ装置 1 は、キャッシュアドレスバス 4 と、キャッシュ制御信号 5 の組を用いて共有されたキャッシュデータとキャッシュタグを含む 2 次キャッシュ 3 に接続されたマイクロプロセッサ 2 のような論理ユニットからなる。マイクロプロセッサ 2 は、主メモリアドレスバス 8 と、主メモリ制御信号 9 の組を用いて主メモリ 7 にも接続されている。2次キャッシュ 3 と主メモリ 7 は共に、共有データバス 6 を用いて、マイクロプロセッサとの間で記憶項目を転送する。4 - 段の先入れ先出し ( F I F O ) 双方向記憶バッファ 1 0 は、マイクロプロセッサから主メモリに書き込まれる記憶項目を緩衝するため使用される。

【 0 0 1 4 】

マイクロプロセッサ内の実行ユニット ( 図示しない ) は、28ビットの物理的主メモリアドレスからなる読出し要求を発生する。全部で32バイトのキャッシュラインが要求される場合、2次キャッシュのルックアップ ( 検査 ) は主メモリのルックアップと並行して始められる。要求された記憶項目が2次キャッシュで検出された場合、主メモリへのルックアップは取り消される。要求された記憶項目が2次キャッシュで検出されなかった場合、要求された記憶項目は主メモリから取り出され、2次キャッシュ内の適当な場所は、取り出された主メモリの項目が利用可能になると同時に上記取り出された主メモリの項目で更新される。

【 0 0 1 5 】

実行ユニットは更にメモリ書込み要求を発生する。キャッシュメモリ装置は、ライトスルー キャッシュ コヒーレンシー プロトコル ( Write-through cache coherency protocol ) を使用する。全てのキャッシュラインが書き込まれたとき、新しいデータは2次メモリキャッシュと主メモリとに同時に書き込まれ、2次キャッシュタグは新しいタグの物理アドレスを使用して更新される。

【 0 0 1 6 】

2 次キャッシュ 3 は直接アドレスマッピングを介してアクセスされる。直接アドレスマッピングとは、アドレスバス 4 を介して 2 次キャッシュ 3 に送られたアドレスがキャッシュ内の固有の場所を直接参照することを意味する。マイクロプロセッサ 2 は、主メモリ 7 内の場所の 2 8 ビット物理アドレスの形でメモリアクセスの要求を形成する。

【 0 0 1 7 】

上記 2 8 ビット物理アドレスは、三つの構成要素に分割される。最初の 3 ビット A D D R [ 2 : 0 ] は、単一キャッシュエントリーに対応する 6 4 ビットのダブルワード内の 1 バイトを選択するため実行ユニットによって内部的に使用される。次の 1 5 ビット A D D R [ 1 7 : 3 ] は、キャッシュエントリーを選択するため、キャッシュアドレスバス 4 、即ち、図 2 の A [ 1 4 : 0 ] を介して送られる。2 次キャッシュ 3 は、キャッシュアドレス

10

20

30

40

50

バス4上でアサートされた15ビットアドレスに应答し、2次キャッシュ3はキャッシュエントリーを読み出し始める。2次キャッシュへのルックアップは、4個の64ビットのキャッシュエントリーからなる32バイトのキャッシュラインを使用して行なわれる。残りの10ビットADDR〔27:18〕は、上記10ビットを2次キャッシュタグと比較することによりキャッシュヒット又はキャッシュミスを検出するためマイクロプロセッサによって内部的に使用される。キャッシュタグとビットADDR〔27:18〕の間の一致はキャッシュヒットであり、キャッシュタグとビットADDR〔27:18〕の間の不一致はキャッシュミスである。キャッシュヒットとキャッシュミスの両方は、実行ユニットへの信号によってマイクロプロセッサ内で示される。

【0018】

10

図2には、マイクロプロセッサと2次キャッシュの間で接続されたアドレスバス及び制御信号と、キャッシュメモリ装置用のデータバスが示されている。キャッシュアドレスバス4は、アドレスデータA〔14:0〕の15ビットを伝達する一組のアドレス線によって構成される。上記15ビットは物理アドレスADDR〔17:3〕から得られる。このアドレスはキャッシュエントリーを識別するため2次キャッシュで絶対アドレスとして使用される。キャッシュ制御信号5は、アクセスするためのメモリチップを選択する3本のチップ選択信号CS〔0〕-〔3〕の一組と、ストロブ信号を伝送するための4本の1ビット信号線と、進行信号と、2次キャッシュへの書込みイネーブル信号とを含む4組の信号からなる。

【0019】

20

データバス6は72ビットのデータを伝達するデータ線の三つのグループからなる。第1のグループはキャッシュタグデータT〔5:0〕の6ビットを伝達する。第2のグループはパリティP〔1:0〕の2ビットを伝達する。第3のグループはキャッシュデータD〔63:0〕の64ビットを伝達する。キャッシュタグデータ線は、マイクロプロセッサ2と2次キャッシュ3とによって排他的に使用される。

【0020】

図3乃至5には2次キャッシュ3の3通りの構成の概略図が示されている。2次キャッシュ3は、例えば、図3の256KB（キロバイト）キャッシュ、図4の512KBキャッシュ、又は、図5の1MB（メガバイト）キャッシュを含むあらゆる数の記憶容量で構成することが可能であり、これにより、記憶容量は使用されるメモリチップの数で決まる。256KBの構成は1対のメモリチップ（20a及び20b）を使用する。512KBの構成は2対のメモリチップ（20a及び20bと、28a及び28b）を使用する。1MBの構成は、4対のメモリチップ（20a及び20bと、28a及び28bと、29a及び29bと、30a及び30b）を使用する。

30

【0021】

図2に示されたチップ選択信号CS〔0〕-〔3〕は、所望のメモリチップの対を選択するため使用され、これにより、CS〔0〕にアサートされた信号はチップ対20a及び20bを選択し、CS〔1〕はチップ対28a及び28bを選択し、CS〔2〕はチップ対29a及び29bを選択し、CS〔3〕はチップ対30a及び30bを選択する。

【0022】

40

本発明の一実施例によれば、各メモリチップは、36ビットのデータの32KBのSRAMチップにより構成される。SRAMチップの一例は、富士通製の32K×36MB82VP036形のSRAMメモリチップである。以下の説明は256KBの構成だけに基づいているが、512KB及び1MBの構成にも同様に適用される。

【0023】

図3には1対のメモリチップ20a及び20bからなる256KBキャッシュとして構成された2次メモリが示されている。各メモリチップは、32,768個の36ビットの個別にアドレス指定可能なエントリーからなる単一キャッシュアレイ構造をなす。各36ビットのエントリーは、32ビット（又は4バイト）のデータと、1パリティビットと、3

50

ビットのキャッシュタグデータを含む。キャッシュエントリーは、常に対の形で取り出されるので、72ビットの各対がキャッシュエントリーとして参照される。キャッシュラインは、32バイトのアドレス境界から始まる4個の連続的に格納されたキャッシュエントリーからなり、かつ、32バイトのデータ(1キャッシュエントリー当たり8バイト×1キャッシュライン当たり4エントリー)を含む。

#### 【0024】

従来技術の場合、典型的に、少なくとも一つのキャッシュアレイ構造は、階層化された記憶項目(データ)と、対応するパリティビットとを格納するため使用され、少なくとも一つのキャッシュアレイ構造がキャッシュタグを格納するため使用される。しかし、本発明によれば、少なくとも一つのデータキャッシュアレイ構造内の余分な場所はキャッシュタグを格納するため使用されるので、キャッシュタグを別の専用の構造に格納する必要は取り除かれる。

10

#### 【0025】

各単一キャッシュアレイ構造は、キャッシュデータと、パリティビットと、キャッシュタグを格納する3列に論理的に分割される。図3において、キャッシュタグは、メモリチップ20a及び20bの夫々の列21a及び21bに格納されている。同様に、パリティビットは列22a及び22bに格納され、キャッシュデータは列23a及び23bに格納される。

#### 【0026】

72ビットのデータは、1回のアクセスにつき2次キャッシュメモリとマイクロプロセッサの間で伝送される。データバス6は2次キャッシュ3の内部データバス24に接続する。内部データバスは、キャッシュタグセグメントバス25a及び25bと、パリティセグメントバス26a及び26bと、キャッシュデータセグメントバス27a及び27bを用いて3ビットのキャッシュデータと、1ビットのパリティと、32ビットのキャッシュデータを得るため、各キャッシュアレイ構造で枝分かれする。

20

#### 【0027】

図6には2次キャッシュの256KB構成で使用される内部バスの配線図が示されている。同様の内部バスは2次キャッシュの512KB又は1MBの構成で使用される。マイクロプロセッサと2次キャッシュの間でキャッシュタグデータを伝達する6本の線のグループは主メモリには延在しないにも係わらず、データバス6と内部データバス24は同一データを伝達する。キャッシュタグセグメントバス25a及び25bは、キャッシュタグを二つの3ビットセグメントのT〔2:0〕とT〔5:3〕に分割する。パリティビットセグメントバス26a及び26bは、パリティビットを二つの1ビットセグメントのP〔0〕とP〔1〕に分割する。キャッシュデータセグメントバス27a及び27bは、キャッシュデータを二つの32ビットセグメントのD〔31:0〕とD〔64:32〕に分割する。

30

#### 【0028】

2次キャッシュ内の4個の連続的なキャッシュエントリーの4回の連続的な読出しには、一本のキャッシュラインを読むことが要求される。しかし、完全なキャッシュタグは、キャッシュラインを構成する4個のキャッシュエントリーの中の2個のキャッシュエントリーだけに格納することが可能である。図7にはキャッシュタグの構成図が示されている。キャッシュタグは、三つのフィールドを含む12ビット長からなる。ビット0乃至9よりなる物理アドレスフィールドは、上記キャッシュラインに格納された階層化された記憶項目の物理アドレスの上位10ビットを含む。ビット10によって構成される有効ビットフィールドVは、キャッシュラインに有効データが含まれているかどうかを示す。ビット11からなるパリティビットフィールドPは、2次キャッシュタグのパリティ検査のため使用される。動作中、キャッシュタグ有効ビットがクリアされている場合、キャッシュタグパリティビットは検査されない。更に、キャッシュタグのパリティエラーにより2次キャッシュのミスが発生し、キャッシュタグのパリティエラーはシステムソフトウェアでは分からない。

40

50

## 【0029】

2次キャッシュタグは四つの3ビットセグメント内の二つの連続的なキャッシュエントリに格納されている。図8にはキャッシュタグセグメントの記憶スキームの概略が示されている。キャッシュラインは4個のキャッシュエントリからなり、4個のキャッシュエントリの中の2個だけがキャッシュタグを格納するため必要とされるので、キャッシュタグを冗長的に格納し、キャッシュヒット又はキャッシュミスの初期検出を行なうため使用できる利点がある。

## 【0030】

キャッシュタグを2回格納することにより、キャッシュラインを構成する4個のキャッシュエントリの中の二つだけがマイクロプロセッサによって受け取られた後に、キャッシュヒット又はキャッシュミスを検出することが可能である。キャッシュヒットは初期に検出されるので、ラップアラウンドの特徴（以下に説明する）を使用して、2次キャッシュは、更なる遅延を受けることなく即座に利用可能になる連続的なキャッシュエントリを送出し続ける。逆に、キャッシュミスは既に受信されたデータの取消を要求するだけである。

10

## 【0031】

本発明の一実施例において、マイクロプロセッサは、多段の命令パイプライン（図示しない）と共に実行ユニット（図示しない）を有する。好ましくは、マイクロプロセッサは、データと共にロードされ、概念的なパイプで同時に実行されるプログラム命令の重複を使用する命令の同時又は並列処理を行なうパイプラインプロセッサである。キャッシュラインを構成する最初の二つのキャッシュエントリは、キャッシュヒット又はキャッシュミスが発生したかどうか分かる前に読まれる。それにも係わらず、上記エントリの両方のデータは、受取と同時に即座に実行ユニットに進められ、命令パイプラインの中に置かれる。1クロックサイクル当たり一つのキャッシュエントリが受けられる。同様に、パイプライン中の1段は、1クロックサイクルで実行される。かくして、最初の二つのエントリからのデータは、キャッシュヒット又はキャッシュミスが検出されたとき、命令パイプラインの中を2段進められている。キャッシュミスが発生したとき、マイクロプロセッサはデータを取り消し、実行ユニットによって使用される前にパイプラインから実際に取り除く。キャッシュヒットが発生した場合、データはパイプライン中で2段の深さにあり、2クロックサイクルが節約される利点がある。

20

30

## 【0032】

より一般的に言うと、パイプライン化された実行ユニットを有するマイクロプロセッサに適用可能なだけでなく、キャッシュタグの冗長な格納によって、マイクロプロセッサは、完全なキャッシュラインが受け取られたときキャッシュヒット又はキャッシュミスが発生したかどうかを判定することができる。

2次キャッシュメモリはラップアラウンドの特徴を利用し、これによって、一つのキャッシュエントリは連続的な各クロックサイクルで出力される。もう一度図2を参照すると、ストローブ制御信号でローレベルをアサートすることにより、キャッシュエントリアドレスがキャッシュアドレスバス4上でアサートされたことが示されている。応答として、2次キャッシュはアドレスバス4上でアサートされたアドレスに格納されたキャッシュエントリを送出する。進行制御信号上で同様にローレベルがアサートされている場合、2次キャッシュは、1クロック当たり1エントリで、2次キャッシュ内の連続的な記憶場所にあるキャッシュエントリを送り続ける。2次キャッシュは、キャッシュラインを構成する4個のキャッシュエントリが送られる前にキャッシュラインの最後に達したとき、キャッシュをキャッシュラインの先頭にラップアラウンドさせるモジュロ4のカウンタを使用する。ラップアラウンドの特徴を使用することにより、マイクロプロセッサは、完全な4 - エントリのキャッシュラインの32バイト境界に対応する必要がない一つのキャッシュエントリアドレスをアサートするだけでよい。

40

## 【0033】

図9には、キャッシュヒット又はキャッシュミスの検出回路を表わす論理図が示されてい

50

る。上記回路の目的は、連続的なクロックサイクルで受けられたキャッシュエントリーに関してキャッシュタグの上側と下側のセグメントを比較することによりキャッシュヒット又はキャッシュミスを検出することである。最初に、キャッシュタグの一方の半分が比較され、その結果が格納される。次のクロックサイクル中に、キャッシュタグのもう一方の半分が比較され、上側及び下側の両方のセグメントが一致する結果を有するかどうかを判定するため、その結果と前の比較の結果が併せて比較される。上側及び下側の両方のセグメントが一致する結果を有するならば、キャッシュヒットが発生している。上側及び下側の両方のセグメントが一致しない場合、キャッシュミスが発生している。

#### 【 0 0 3 4 】

上記回路はマイクロプロセッサの内側にある。キャッシュタグの物理アドレスフィールドには、キャッシュラインに格納された階層化された記憶項目の物理アドレスの上位 10 ビットが含まれている点に注意する必要がある。各キャッシュタグは、二つの連続的なキャッシュエントリーに格納されているので、2 次キャッシュメモリからデータバスを介して二つの 6 ビットセグメント T〔5 : 0〕に受け取られる。2 次キャッシュから受けられたような各セグメントは、内部キャッシュタグバス 80 に沿って一時記憶装置用のレジスタ 81 に進められる。要求された記憶項目の上位 10 ビット P A〔27 : 18〕は、内部物理アドレスバス 82 上でアサートされ、内部バス 83 及び 84 で夫々アサートされた下部 5 ビットの P A〔22 : 18〕及び上位 5 ビットの P A〔27 : 23〕の二つの 5 ビットセグメントに分割される。

#### 【 0 0 3 5 】

レジスタ 81 に格納されたキャッシュタグセグメントは、排他的否定論理和 (X N O R) 論理ゲート 85 及び 86 を用いて両方の 5 ビット物理アドレスセグメントと比較される。X N O R 演算から得られた結果は、上側又は下側の何れが比較されているかに基づいて適当な結果を選択するためマルチプレクサ 87 を用いて選択される。S E L \_ L O W \_ T A G 信号は、上側又は下側のキャッシュタグセグメントの比較結果を選択するためマルチプレクサ 87 への入力として制御信号 88 でアサートされる。

#### 【 0 0 3 6 】

マルチプレクサ 87 からの結果は、共にレジスタ 89 に格納され、A N D 論理ゲート 90 に進められる。レジスタ 89 は前のクロックサイクルで判定された比較演算の結果を含み、次の連続的なクロックサイクルが I O \_ C L K 信号 97 上でアサートされたとき、A N D 論理ゲートに進められる。現在のタグセグメントのヒットは線 90 に沿って示される。前のタグセグメントのヒットは線 91 にそって示される。両方の線にアサートされた値は、線 93 でキャッシュヒット又はキャッシュミスを示す信号をアサートする A N D 論理ゲート 92 によって比較される。キャッシュミスが発生し、取消イネーブル信号 (E N B \_ C A N C E L) が線 94 にアサートされた場合、A N D 論理ゲート 95 は線 96 上に取消信号をアサートし、これにより、前のクロックサイクル中に実行ユニットに既に送られたデータ準備完了 (D A T A \_ R D Y) 信号を取り消す。上記取消によって、例えば、命令パイプライン中の記憶項目が取り消される。

#### 【 0 0 3 7 】

図 10 には、キャッシュエントリーパリティ検査回路を表わす論理図が示されている。この回路の目的は、キャッシュエントリーを構成する何れかの 36 ビットセグメントに関しパリティエラーが発生したかどうかを判定することである。上記回路はマイクロプロセッサの内部にある。パリティエラーがない場合を想定すると、内部 D A T A \_ R D Y 信号がアサートされ、これにより、2 次キャッシュからの有効データの存在が示される。2 次キャッシュからデータバス 6 を介して受けられた 64 ビットのキャッシュデータ及び二つのパリティビットは、二つの 32 ビットセグメントの D〔63 : 32〕及び D〔31 : 0〕と、2 個の別個のパリティビット P〔1〕及び P〔0〕とに分割される。キャッシュデータセグメントは、内部キャッシュデータバス 100 及び 101 でアサートされ、パリティビットはバス 102 及び 103 に沿ってアサートされる。キャッシュデータは、キャッシュデータセグメントのパリティを定めるため X O R 論理ゲート 104 及び 105 に入力さ

10

20

30

40

50



れる。次いで、その結果は、X O R 論理ゲート 1 0 6 及び 1 0 7 を用いて上記セグメントに関しパリティビットと比較される。その結果は O R 論理ゲート 1 0 8 によって比較され、レジスタ 1 0 9 に格納される。レジスタは、C P U \_ C L K 信号 1 1 0 上でアサートされるような次のクロックサイクルに上記結果を先に進める。準備完了信号イネーブル ( E N B \_ R E A D Y ) がアサートされ、A N D 論理ゲートによって定められたようなパリティエラーが存在しない場合、データ準備完了 ( D A T A \_ R D Y ) 信号 1 1 2 がアサートされ、有効キャッシュエントリーがパリティエラー無しで受け取られたことを示す。

#### 【 0 0 3 8 】

上記キャッシュメモリ装置は、ライトスルー キャッシュ コヒーレンシー プロトコルを利用する。書込み動作の実行時に、新しいデータが 2 次キャッシュメモリと主メモリとに同時に書き込まれる。全てのメモリ書込みは上記キャッシュタグを更新する。非ブロック ( キャッシュライン以外の ) 書込みはキャッシュタグエントリーを抹消し、一方、ブロック書込みはキャッシュタグエントリーを更新する。読出し動作の実行時に、非ブロック読出しは 2 次キャッシュメモリを回避し、一方、ブロック書込みにより 2 次キャッシュのルックアップが行なわれる。ルックアップの結果がキャッシュミスである場合、対応する主メモリのデータは、利用可能なときに 2 次キャッシュに書き込まれ、これに従ってキャッシュタグが更新される。

10

#### 【 0 0 3 9 】

2 次キャッシュの階層化された記憶項目は、マイクロプロセッサ 2 内の受取時にパリティエラーが検査される。パリティエラーが最初の二つのデータエントリーに生じた場合、パリティエラーはキャッシュミスと同様に処理され、既に読まれたデータは取り消され、キャッシュタグは更新される。パリティエラーが最後の二つのキャッシュデータエントリーに生じた場合、パリティエラーによってパリティエラーが通知される。外部主メモリのブロック読出しのパリティエラーは、2 次キャッシュミスがあるならば、指定された 2 次キャッシュラインを抹消する。

20

#### 【 0 0 4 0 】

図 1 1 には、書込み動作用のキャッシュ コヒーレンシー プロトコルのフローチャートが示されている。書込みがブロック書込みを試みる場合 ( ステップ 4 0 )、キャッシュは、キャッシュラインとそのタグを置き換えることにより更新される ( ステップ 4 1 )。ブロック書込み以外の場合、非ブロック書込みが行なわれるならば、キャッシュラインは、4 回の連続的なキャッシュエントリーの書込みでキャッシュタグを無効化、又は、ヌル値に設定することにより抹消される。

30

#### 【 0 0 4 1 】

図 1 2 には、読出し動作用のキャッシュ コヒーレンシー プロトコルのフローチャートが示されている。非ブロック読出しが行なわれる場合 ( ステップ 5 0 )、2 次キャッシュメモリは回避され、データは主メモリから直接読まれる ( ステップ 5 1 )。非ブロック読出し以外の場合、2 次キャッシュのルックアップが行なわれる。最初のキャッシュエントリーが読まれ ( ステップ 5 2 )、データパリティエラーが検出された場合 ( ステップ 5 3 )、データは取り消され、キャッシュラインは、利用可能になったとき主メモリのデータで更新される ( ステップ 5 4 )。データパリティが検出されなかった場合、2 次キャッシュメモリが読まれ ( ステップ 5 5 )、データパリティが検出された場合 ( ステップ 5 6 )、データは取り消され、キャッシュラインは、利用可能になったとき主メモリのデータで更新される ( ステップ 5 7 )。

40

#### 【 0 0 4 2 】

第 1 又は第 2 のキャッシュエントリーの何れにおいてもパリティエラーが発生していない場合を想定すると、完全なキャッシュタグが受け取られ、ステップ 5 8 から始めるキャッシュヒット又はキャッシュミスの初期の検出が行なわれる。キャッシュタグの有効ビットが検査され、クリアされている場合 ( ステップ 5 8 )、キャッシュラインは取り消される無効データを有し、キャッシュラインは、利用可能になったとき主メモリのデータで更新される ( ステップ 5 9 )。クリアされていない場合、キャッシュタグパリティビットが検

50

査され、パリティエラーが検出された場合（ステップ60）、パリティエラーが通知され、キャッシュラインは抹消される（ステップ61）。パリティエラーは検出されなかった場合、キャッシュタグが検査され、キャッシュミスが検出された場合（ステップ62）、データは取り消され、キャッシュラインは、利用可能になったとき主メモリのデータで更新される（ステップ63）。キャッシュヒットが検出された場合を想定すると、第3及び第4のキャッシュエントリーが読まれる（ステップ64及びステップ67）。データパリティエラーが何れかのキャッシュエントリーに関し検出された場合（ステップ65及び68）、パリティエラーが通知され、キャッシュラインは抹消される（ステップ66及び69）。データパリティエラーが発生しなかった場合、2次キャッシュのルックアップは成功する。

10

#### 【0043】

図13には、2次キャッシュのヒットを表わすタイミングチャートが示されている。読出し要求は時点 $T_0$ で行なわれる。2次キャッシュのストローブ制御信号と、キャッシュラインのアドレスは、時点 $T_1$ でアサートされる。時点 $T_2$ で2次キャッシュは読出し動作を開始し、マイクロプロセッサは順次のキャッシュエントリーのルックアップを示す進行制御信号をアサートする。2次キャッシュは、時点 $T_3$ で第1のキャッシュエントリーを送り始める。マイクロプロセッサは、第1のキャッシュエントリーを受け、内部的にその受取を示すハイレベルの推測的な読出し準備完了（SC Read Ready）信号をアサートし、一方、2次キャッシュは、時点 $T_4$ で第2のキャッシュエントリーを送り始める。キャッシュタグのもう一方の半分を含む第2のキャッシュエントリーは、時点 $T_5$ で受けられ、キャッシュヒットの初期検出が時点 $T_5$ で行なわれる。キャッシュラインを構成する残りの二つのキャッシュエントリーは、時点 $T_6$ 及び $T_7$ で受けられる。

20

#### 【0044】

図14には、2次キャッシュのミスを表わすタイミングチャートが示されている。読出し要求は時点 $T_0$ で行なわれる。2次キャッシュのストローブ制御信号と、キャッシュラインのアドレスは、時点 $T_1$ でアサートされる。時点 $T_2$ で2次キャッシュは読出し動作を開始し、マイクロプロセッサは順次のキャッシュエントリーのルックアップを示す進行制御信号をアサートする。2次キャッシュは、時点 $T_3$ で第1のキャッシュエントリーを送り始める。マイクロプロセッサは、第1のキャッシュエントリーを受け、内部的にその受取を示すハイレベルの推測的な読出し準備完了信号をアサートし、一方、2次キャッシュは、時点 $T_4$ で第2のキャッシュエントリーを送り始める。キャッシュタグのもう一方の半分を含む第2のキャッシュエントリーは、時点 $T_5$ で受けられ、キャッシュミスの初期検出が時点 $T_5$ で行なわれる。マイクロプロセッサは、キャッシュミスの検出時に、ローレベルの読出し準備完了信号をアサートし、ハイレベルの取消信号をアサートする。

30

#### 【0045】

図15には、主メモリ読出しを伴う2次キャッシュミスを表わすタイミングチャートが示されている。読出し要求は時点 $T_0$ で行なわれる。キャッシュラインのアドレスは時点 $T_1$ でアサートされる。時点 $T_2$ で2次キャッシュは読出し動作を開始し、マイクロプロセッサは、順次のキャッシュエントリーのルックアップを示す進行制御信号をアサートする。2次キャッシュは、時点 $T_3$ で第1のキャッシュエントリーを送り始める。マイクロプロセッサは、第1のキャッシュエントリーを受け、内部的にその受取を示すハイレベルの推測的な読出し準備完了信号をアサートし、一方、2次キャッシュは、時点 $T_4$ で第2のキャッシュエントリーを送り始める。キャッシュタグのもう一方の半分を含む第2のキャッシュエントリーは、時点 $T_5$ で受けられ、キャッシュミスの初期検出が時点 $T_5$ で行なわれる。マイクロプロセッサは、キャッシュミスの検出時に、ローレベルの読出し準備完了信号をアサートし、ハイレベルの取消信号をアサートする。2次キャッシュラインの内容は、時点 $T_{23}$ から主メモリより受けられたデータを用いて更新され、新しいキャッシュエントリーは9クロック周期毎に受けられる。

40

#### 【0046】

上記本発明を、特に、その実施例を参照して示し、かつ、説明しているが、当業者は、本

50

発明の精神と範囲から逸れることなく、形態及び細部の点で上記実施例の上記及び他の変形をなし得ることが分かる。

【 0 0 4 7 】

【発明の効果】

本発明によれば、少なくとも一つのデータキャッシュアレイ構造内の余分な場所はキャッシュタグを格納するため使用されるので、キャッシュタグを別の専用の構造に格納する必要は取り除かれる利点がある。

本発明によれば、2次キャッシュが用いられるので、使用するメモリチップの数で記憶容量を決めることができる利点を得られる。

【 0 0 4 8 】

更に、キャッシュタグを冗長的に格納することができるので、キャッシュヒット及びキャッシュミスの初期検出を行なうため使用できる利点がある。

【図面の簡単な説明】

【図1】本発明の一実施例のキャッシュメモリ装置の概略図である。

【図2】マイクロプロセッサと2次キャッシュの間で接続されたアドレスバス及び制御信号と、キャッシュメモリ装置のデータバスの配線図である。

【図3】2次キャッシュの第1の構成の概略図である。

【図4】2次キャッシュの第2の構成の概略図である。

【図5】2次キャッシュの第3の構成の概略図である。

【図6】2次キャッシュで使用される内部バスの配線図である。

【図7】キャッシュタグの説明図である。

【図8】キャッシュタグセグメントの記憶スキームを示す図である。

【図9】キャッシュヒット又はミス検出回路を表わす論理図である。

【図10】キャッシュエントリーのパリティ検査回路を表わす論理図である。

【図11】書き込み動作のキャッシュコヒーレンシープロトコルを表わすフローチャートである。

【図12】読出し動作のキャッシュコヒーレンシープロトコルを表わすフローチャートである。

【図13】2次キャッシュヒットを表わすタイミングチャートである。

【図14】2次キャッシュミスを表わすタイミングチャートである。

【図15】主メモリ読出しを伴う2次キャッシュミスを表わすタイミングチャートである。

。

【符号の説明】

- 1      キャッシュメモリ装置
- 2      マイクロプロセッサ
- 3      2次キャッシュ
- 4      キャッシュアドレスバス
- 5      キャッシュ制御信号
- 6      キャッシュデータバス
- 7      主メモリ
- 8      主メモリアドレスバス
- 9      主メモリ制御信号
- 10     F I F O 記憶バッファ

10

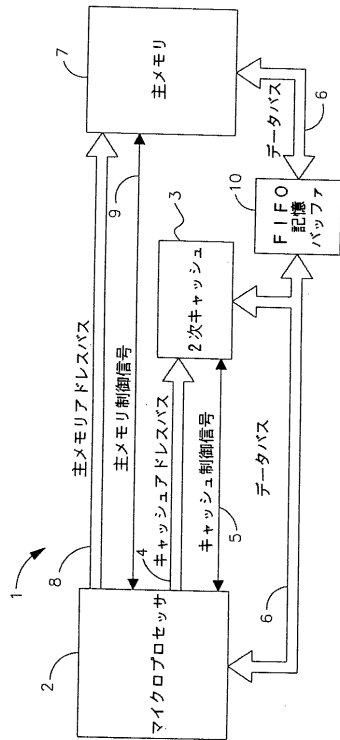
20

30

40

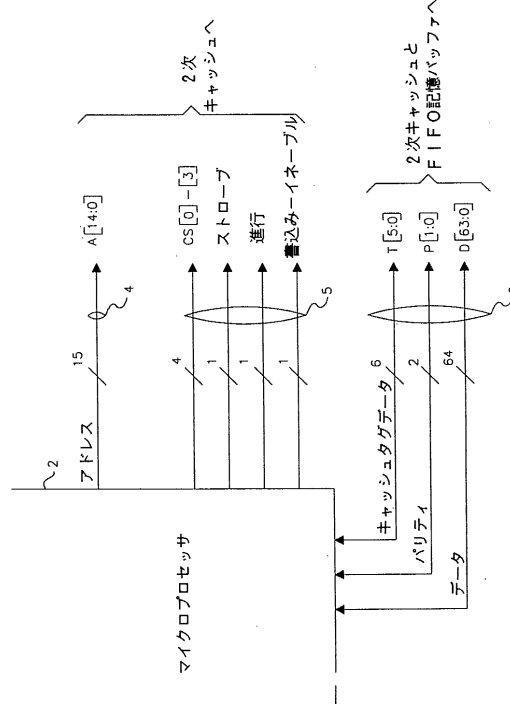
【図 1】

本発明の一実施例のキャッシュメモリ装置の概略図



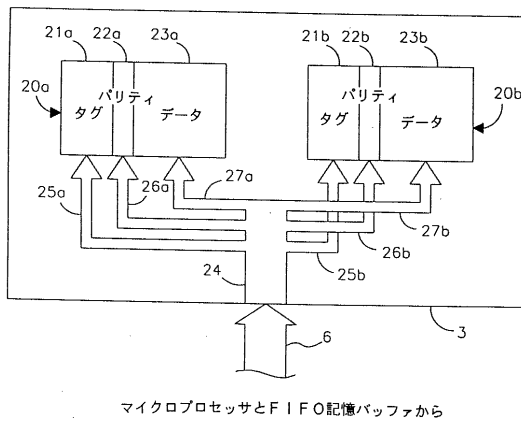
【図 2】

アドレスバス及び制御信号とデータバスの配線図



【図 3】

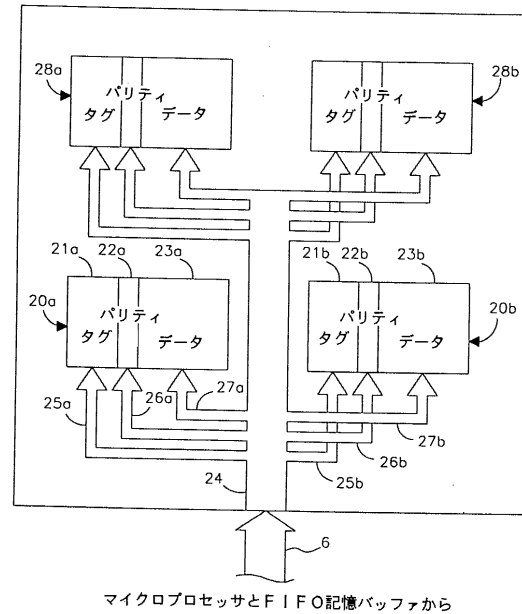
2次キャッシュの概略構成図



マイクロプロセッサとFIFO記憶バッファから

【図 4】

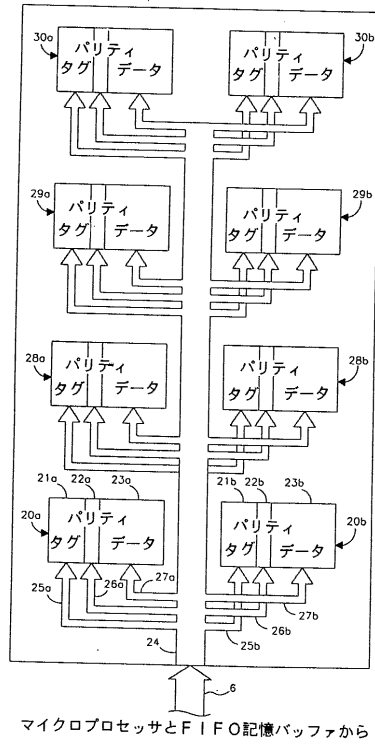
2次キャッシュの概略構成図



マイクロプロセッサとFIFO記憶バッファから

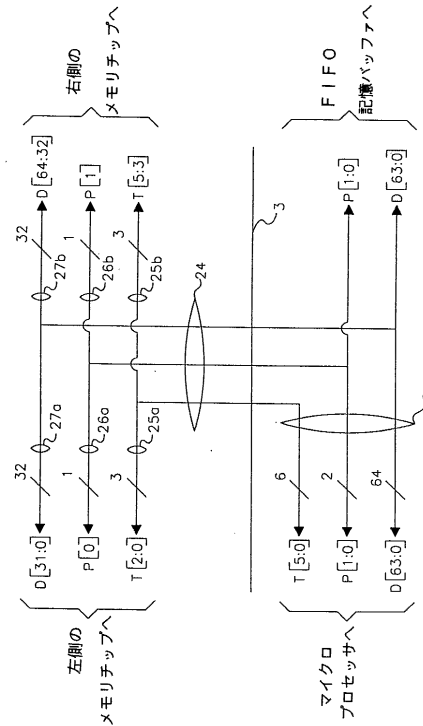
【図 5】

2次キャッシュの概略構成図



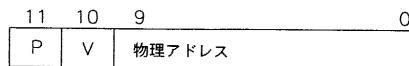
【図 6】

2次キャッシュで使用する内部バスの配線図



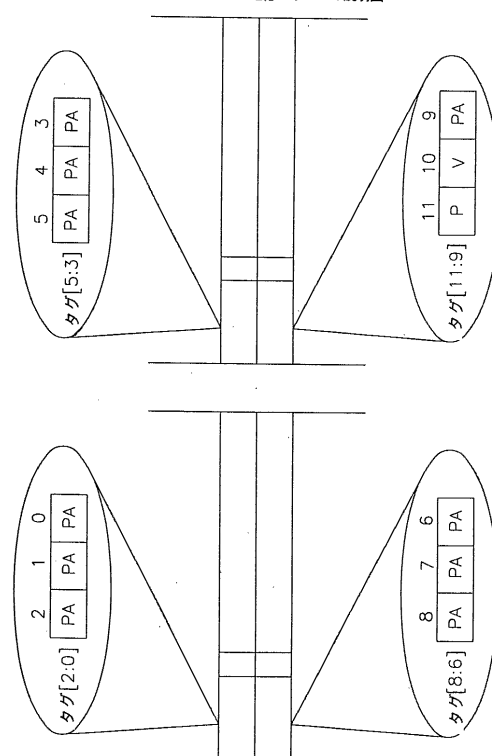
【図 7】

キャッシュタグの説明図



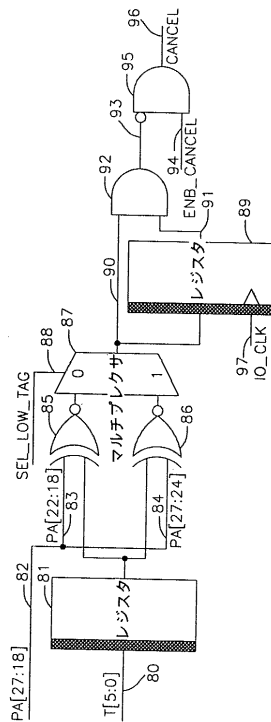
【図 8】

キャッシュタグの記憶スキームの説明図



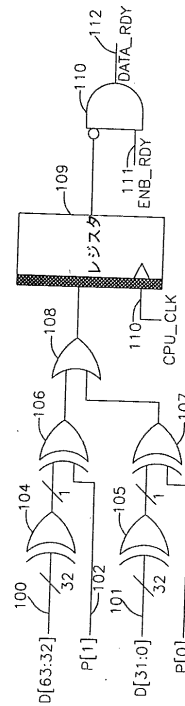
【図 9】

キャッシュヒット又はミス検出回路の論理図



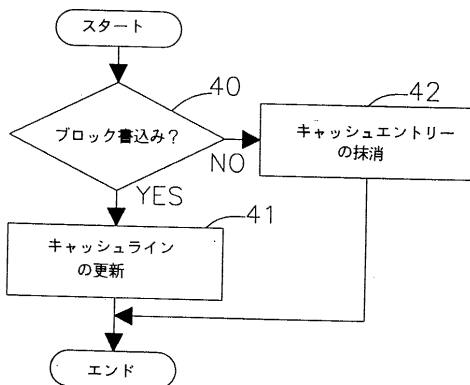
【図 10】

キャッシュエントリーのパリティ検査回路の論理図



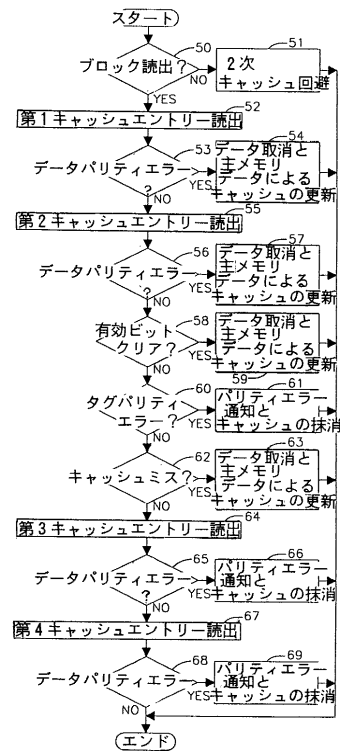
【図 11】

書き込み動作キャッシュコヒーレンシープロトコルのフローチャート



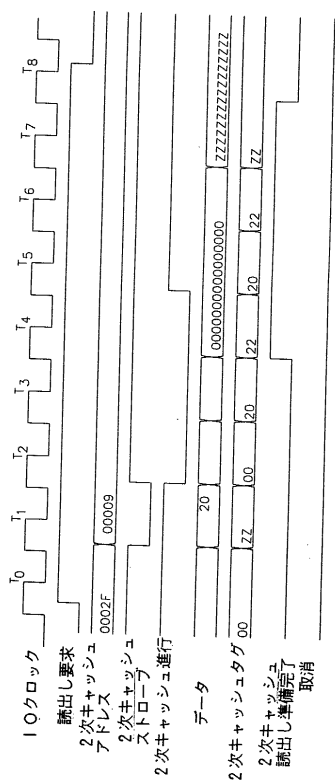
【図 12】

読み出し動作キャッシュコヒーレンシープロトコルのフローチャート



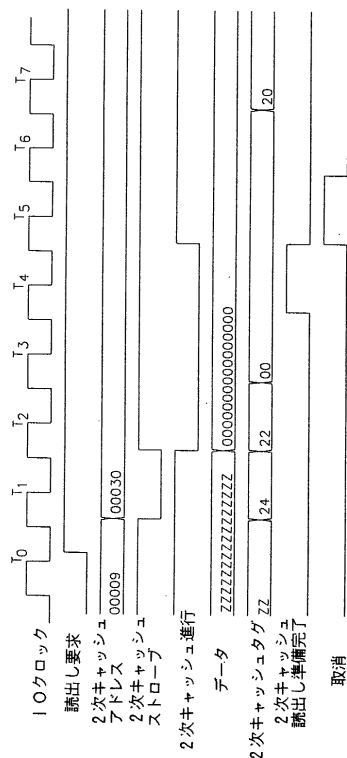
【 図 1 3 】

### 2次キャッシュヒットのタイミングチャート



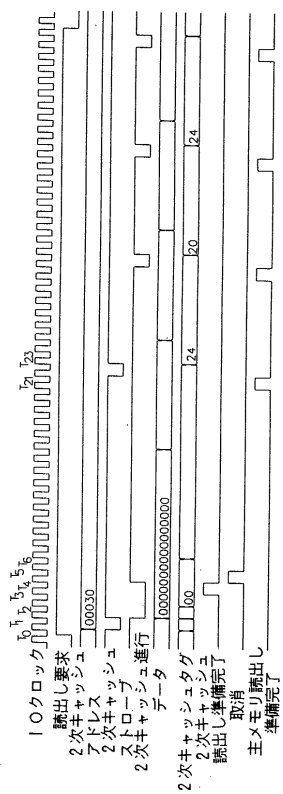
【 図 1 4 】

## 2次キャッシュミスのタイミングチャート



【 図 1 5 】

主メモリ読出しを伴う2次キャッシュミスのタイミングチャート



## フロントページの続き

(56)参考文献 特開昭62-184551(JP,A)

特開平5-158795(JP,A)

特開平2-66653(JP,A)

特開平1-280850(JP,A)

特開平8-227376(JP,A)

特開平8-153039(JP,A)

特開平6-259329(JP,A)

特開平6-259326(JP,A)

特開平5-108484(JP,A)

特開平4-211880(JP,A)

特開平4-209051(JP,A)

特開昭62-245448(JP,A)

特開昭62-135944(JP,A)

特開昭62-90740(JP,A)

特開昭62-55743(JP,A)

特開昭63-157249(JP,A)

特開昭62-187946(JP,A)

特開昭54-146932(JP,A)

実開昭61-189351(JP,U)

桑田明、山本満博,VR3000のハードウェア・インターフェース,インターフェース,日本, CQ出版株式会社, 1990年12月 1日, 第16巻, 第12号, p. 261-269

大貫徹, RISC入門(R2000/3000・SPARC・29000・88000), インターフェース, 日本, CQ出版株式会社, 1989年 1月 1日, 第15巻, 第1号, p. 227-242

Dick Pountain, スーパーコンピュータを指向するR8000チップ・セット, 日経バイト, 日本, 日経BP社, 1994年11月 1日, 第131号, p. 275-283

(58)調査した分野(Int.Cl.<sup>7</sup>, DB名)

G06F12/08-12/12