

(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.C1.⁷

F 1

G06F 12/08

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

(22) 出願日

平成8年3月14日(1996.3.14)

(65) 公開番号

特開平8-272688

(43) 公開日

平成8年10月18日(1996.10.18)

審査請求日

平成14年8月19日(2002.8.19)

(31) 優先権主張番号

406305

(32) 優先日

平成7年3月17日(1995.3.17)

(33) 優先権主張国

米国(US)

早期審査対象出願

前置審査

(73) 特許権者 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 100070150

弁理士 伊東 忠彦

(72) 発明者 モーゲンス ローリッツエン

アメリカ合衆国 カリフォルニア州 94
024 ロス・アルトス オークハースト
・アヴェニュー 1310番

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单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つて：

記憶容量が固定である一次キャッシュ；

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单一キャッシュアレイ構造により構成され、上記单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；

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

10

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バ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单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单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ー(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、ステイ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ー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ーに亘って分散され、タグは2次メモリキャッシュ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。 10

【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单一キャッシュアレイ構造により構成され、上記单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ー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。 20

【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单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。 40

【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 50

ある。

【0011】

【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 アークテクチャー）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。

【0012】

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み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みら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。

【0013】

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(FIFO)双方向記憶バッファ10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。

【0014】

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み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。

【0015】

実行ユニットは更にメモリ書き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。

【0016】

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28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。

【0017】

上記28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、单一キャッシュエントリーに対応する64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15ビットA D D R [17:3]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 [14:0]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 ビットアドレスに応答し、2 次キャッシュ 3 はキャッシュエンタ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タリーからなる 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タ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、ストロー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き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 × 36M B 82VPO36 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タリーからなる单一キャッシュアレイ構造をなす。各 36 ビットのエンタリーは、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に、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と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。4050

【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ーに格納されている。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ーからなり、4個のキャッシュエントリーの中の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ー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ーを送出し続ける。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ーは、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が分かる前に読まれる。それにも係わらず、上記エントリーの両方のデータは、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ーが受けられる。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ーからのデータは、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

【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はなく、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ー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ー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ーを送出する。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エントリーで、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ーを送り続ける。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ー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ー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-エントリー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ーアドレスをアサートするだけでよい。

30

【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

40

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。

【0034】

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PA[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位5ビットのPA[22:18]及び上位5ビットのPA[27:23]の二つの5ビットセグメントに分割される。

【0035】

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(XNOR)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。XNOR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。SEL_LOW_TA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。

【0036】

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、AND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IOCLK信号97上でアサートされたとき、AND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AND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(ENB_CANCEL)が線94にアサートされた場合、AND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(DATA_RDY)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。

【0037】

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DATA_RDY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XOR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、XOR論理ゲート106及び107を用いて上記セグメントに関しparityビットと比較される。その結果はOR論理ゲート108によって比較され、レジスタ109に格納される。レジスタは、CPU_CLK信号110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INEPBLE(ENB_READY)がアサートされ、AND論理ゲートによって定められたようなparityエラーが存在しない場合、データ準備完了(DATA_READY)信号112がアサートされ、有効キャッシュエントリーがparityエラー無しで受け取られたことを示す。

【0038】

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。全てのメモリ書き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)書き込みはキャッシュタグエントリーを抹消し、一方、ブロック書き込みはキャッシュタグエントリーを更新する。読み出し動作の実行時に、非ブロック読み出しは2次キャッシュメモリを回避し、一方、ブロック書き込みにより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。

【0039】

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parityエラーが検査される。parityエラーが最初の二つのデータエントリーに生じた場合、parity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。parity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、parityエラーによってparityエラーが通知される。外部主メモリのブロック読み出しのparity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。

【0040】

図11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。書き込みがブロック書き込みを試みる場合(ステップ40)、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41)。ブロック書き込み以外の場合、非ブロック書き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き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。

【0041】

図12には、読み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。非ブロック読み出しが行なわれる場合(ステップ50)、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51)。非ブロック読み出し以外の場合、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52)、データparityエラーが検出された場合(ステップ53)、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54)。データparity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55)、データparityが検出された場合(ステップ56)、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57)。

【0042】

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parity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58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58)、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59)。クリアされていない場合、キャッシュタグparityビットが検

10

20

30

40

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。読み出し要求は時点T₀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T₁でアサートされる。時点T₂で2次キャッシュは読み出し動作を開始し、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T₃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み出し準備完了（S C R e a d R e a d y）信号をアサートし、一方、2次キャッシュは、時点T₄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T₅で受けられ、キャッシュヒットの初期検出が時点T₅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T₆及びT₇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。読み出し要求は時点T₀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T₁でアサートされる。時点T₂で2次キャッシュは読み出し動作を開始し、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T₃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み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T₄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T₅で受けられ、キャッシュミスの初期検出が時点T₅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、ローレベルの読み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、主メモリ読み出しを伴う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。読み出し要求は時点T₀で行なわれる。キャッシュラインのアドレスは時点T₁でアサートされる。時点T₂で2次キャッシュは読み出し動作を開始し、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T₃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み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T₄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T₅で受けられ、キャッシュミスの初期検出が時点T₅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、ローレベルの読み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T₂₃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。

【0047】

【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が得られる。

【0048】

10

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。

20

【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ー検査回路を表わす論理図である。

【図11】書き込み動作用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】読み出し動作用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。

30

【図15】主メモリ読み出しを伴う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

40

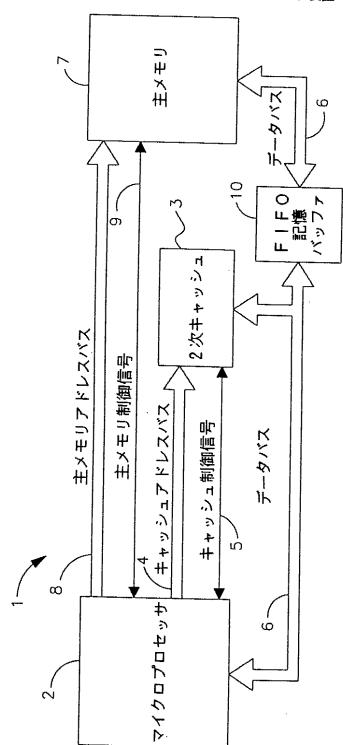
8 主メモリアドレスバス

9 主メモリ制御信号

10 FIFO記憶バッファ

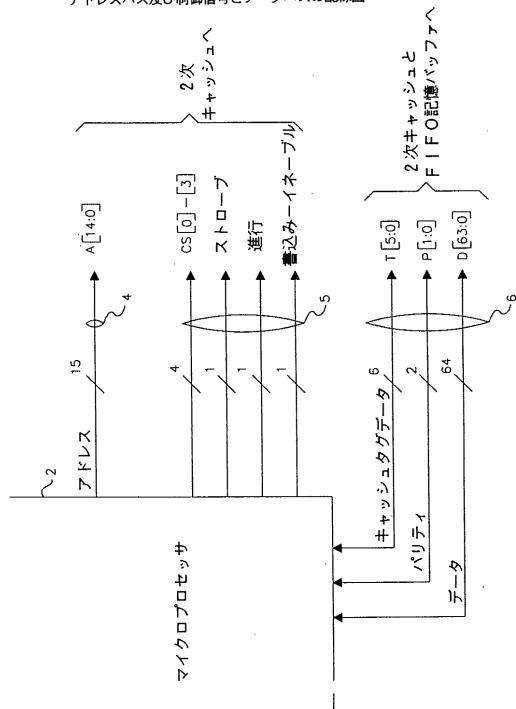
【図1】

本発明の一実施例のキャッシュメモリ装置の概略図



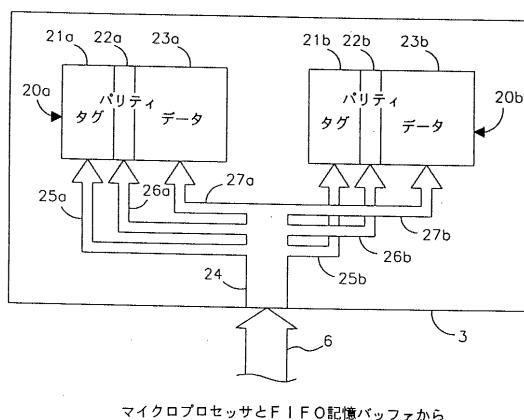
【図2】

アドレスバス及び制御信号とデータバスの配線図



【図3】

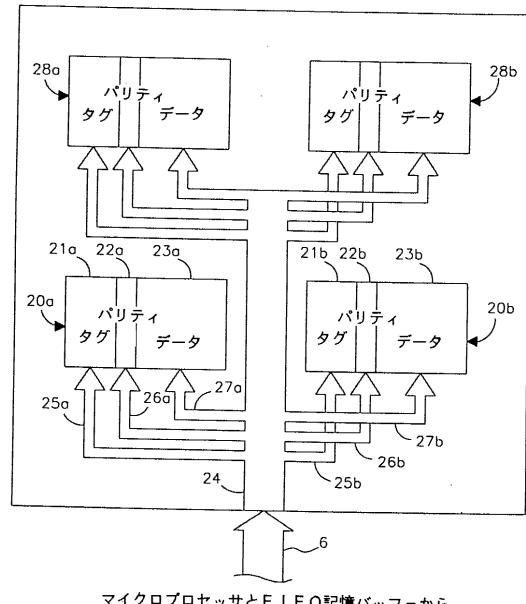
2次キャッシュの概略構成図



マイクロプロセッサとFIFO記憶バッファから

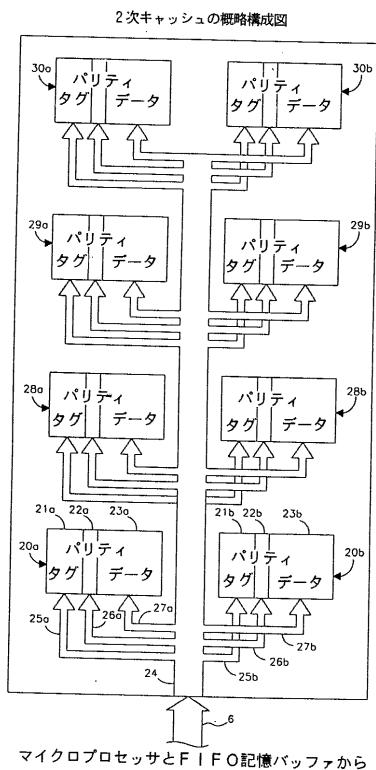
【図4】

2次キャッシュの概略構成図

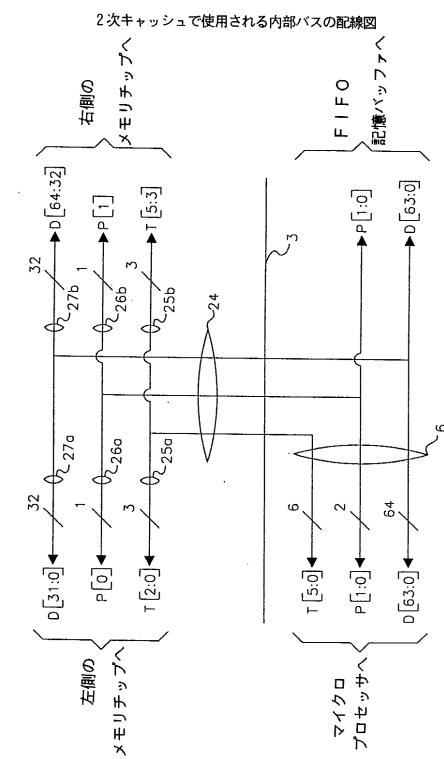


マイクロプロセッサとFIFO記憶バッファから

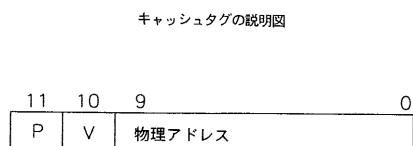
【図5】



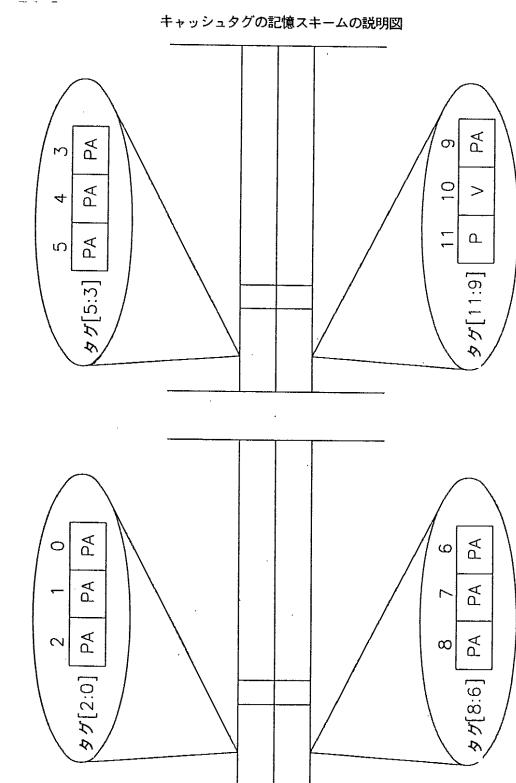
【図6】



【図7】

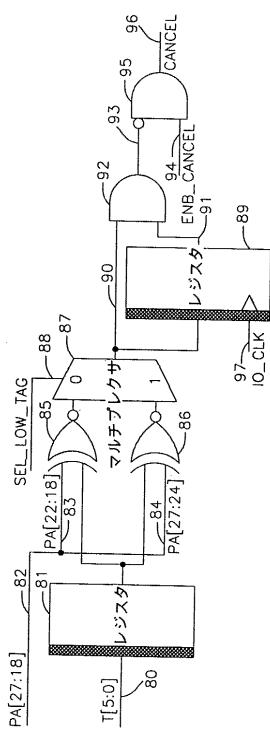


【図8】



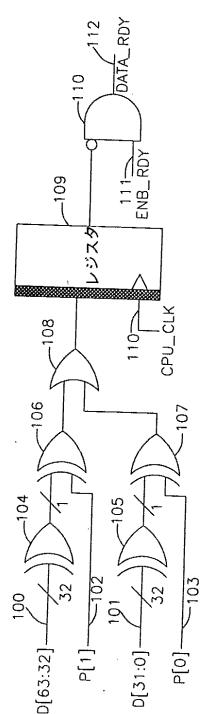
【図9】

キャッシュヒット又はミス検出回路の論理図



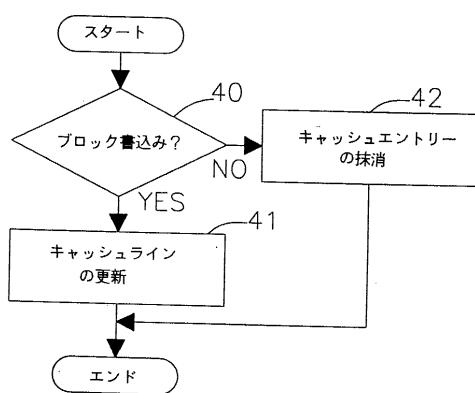
【図10】

キャッシュエントリーのバリティ一検査回路の論理図



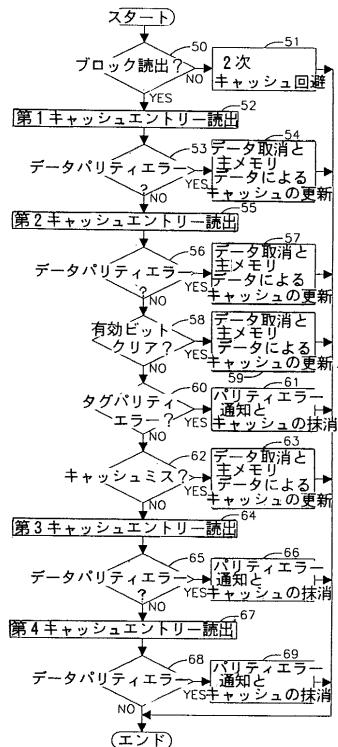
【図11】

書き込み動作用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

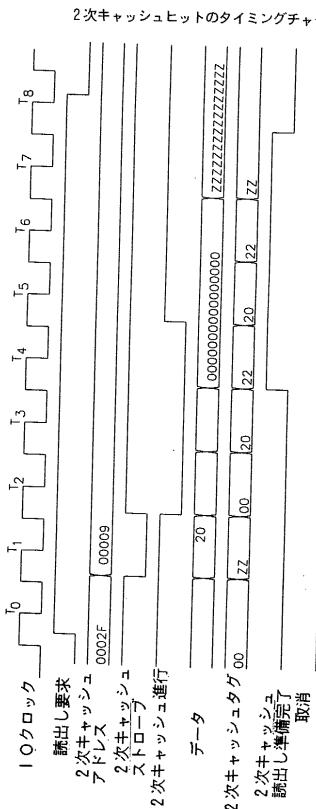


【図12】

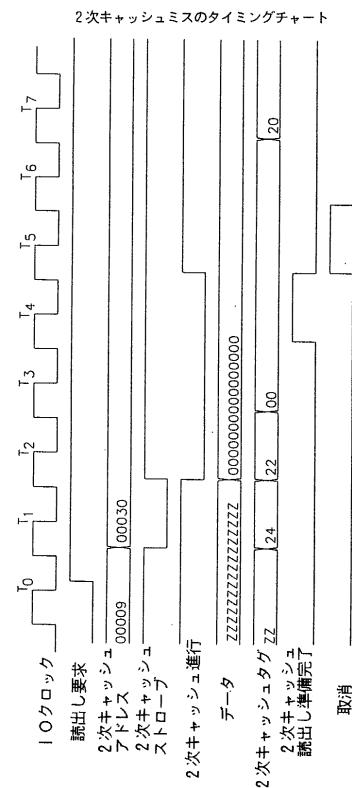
読み出し動作用キャッシュコヒーレンシープロトコルのフローチャート



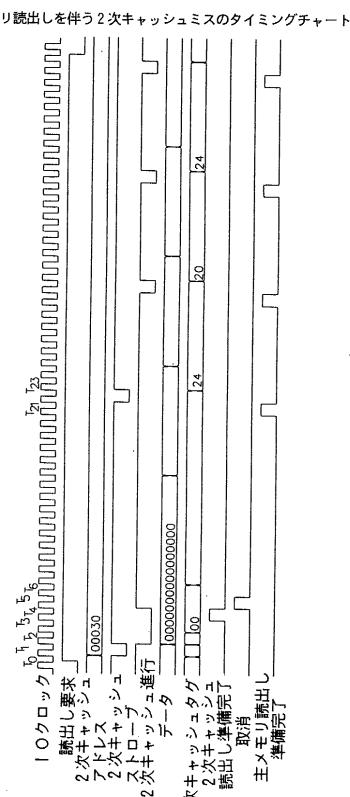
【 図 1 3 】



【 図 1 4 】



【 図 1 5 】



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, VR 3000 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.⁷ , DB名)

G06F12/08-12/12