

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関  
国際事務局



(10) 国際公開番号

WO 2011/049051 A1

PCT

(43) 国際公開日  
2011年4月28日(28.04.2011)

- (51) 国際特許分類:  
G06F 12/08 (2006.01) G11C 15/04 (2006.01)  
G06F 12/12 (2006.01)
- (21) 国際出願番号: PCT/JP2010/068298
- (22) 国際出願日: 2010年10月19日(19.10.2010)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:  
特願 2009-241446 2009年10月20日(20.10.2009) JP
- (71) 出願人 (米国を除く全ての指定国について): 国立大学法人電気通信大学(The University of Electro-Communications) [JP/JP]; 〒1828585 東京都調布市調布ヶ丘一丁目5番地1 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 岡部 翔 (OKABE Sho) [JP/JP]; 〒1828585 東京都調布市調

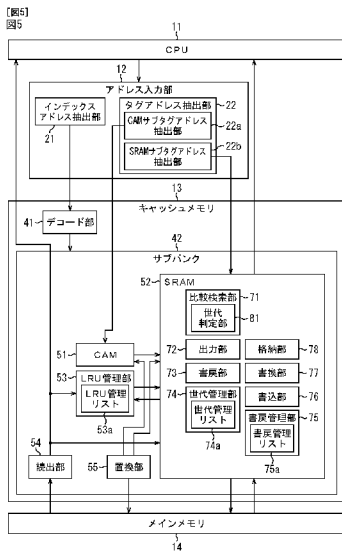
布ヶ丘一丁目5番地1 国立大学法人電気通信大学内 Tokyo (JP). 阿部 公輝(ABE Koki) [JP/JP]; 〒1828585 東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内 Tokyo (JP).

- (74) 代理人: 稲本 義雄, 外(INAMOTO Yoshio et al.); 〒1600023 東京都新宿区西新宿7丁目11番18号 711ビルディング4階 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[続葉有]

(54) Title: CACHE MEMORY AND CONTROL METHOD THEREOF

(54) 発明の名称: キャッシュメモリおよびその制御方法



- 12 ADDRESS INPUT UNIT
- 13 CACHE MEMORY
- 14 MAIN MEMORY
- 21 INDEX ADDRESS EXTRACTION UNIT
- 22 TAG ADDRESS EXTRACTION UNIT
- 22a CAM SUB-TAG ADDRESS EXTRACTION UNIT
- 22b SRAM SUB-TAG ADDRESS EXTRACTION UNIT
- 41 DECODE UNIT
- 42 SUB-BANK
- 51 LRU MANAGEMENT UNIT
- 52 LRU MANAGEMENT LIST
- 54 READ-OUT UNIT
- 55 REPLACEMENT UNIT
- 71 COMPARISON RETRIEVAL UNIT
- 72 OUTPUT UNIT
- 73 WRITE-BACK UNIT
- 74 GENERATION MANAGEMENT UNIT
- 75 WRITE-BACK MANAGEMENT LIST
- 76 WRITE-IN UNIT
- 78 STORAGE UNIT
- 81 GENERATION DETERMINATION UNIT

(57) Abstract: Provided are a cache memory and a control method thereof, by which electric power consumption can be reduced without reducing a cache hit ratio. A CAM sub-tag address extraction unit (22a) extracts a CAM sub-tag address from a tag address. A SRAM sub-tag address extraction unit (22b) extracts a SRAM sub-tag address from the tag address. A CAM (51) retrieves data by comparing CAM sub-tag addresses. A comparison retrieval unit (71) retrieves data by comparing the extracted SRAM sub-tag address and the SRAM sub-tag address stored in the SRAM, with respect to recently read out first generation data among the data of the retrieved tag address. An output unit (72) outputs, as a response to a request, the first generation data which is retrieved by the comparison retrieval unit (71) and stored in association with the SRAM sub-tag address. The method can be applied to a cache memory.

(57) 要約: 本発明は、キャッシュヒット率を低減させることなく、電力消費を低減させることができるキャッシュメモリおよびその制御方法に関する。CAMサブタグアドレス抽出部22aは、タグアドレスよりCAMサブタグアドレスを抽出する。SRAMサブタグアドレス抽出部22bは、タグアドレスよりSRAMサブタグアドレスを抽出する。CAM51は、CAMサブタグアドレス同士の比較によりデータを検索する。比較検索部71は、検索されたタグアドレスのデータのうち、最近読み出された第1世代のデータについて、抽出されたSRAMサブタグアドレスと、SRAMに格納されたSRAMサブタグアドレスとの比較によりデータを検索する。出力部72は、比較検索部71により検索された、SRAMサブタグアドレスに対応付けて記憶されている第1世代のデータを要求に対する応答として出力する。本発明は、キャッシュメモリに適用することができる。

WO 2011/049051 A1



(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI

(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

## 明 細 書

### 発明の名称： キャッシュメモリおよびその制御方法

#### 技術分野

[0001] 本発明は、キャッシュメモリおよびその制御方法に関し、特に、性能を低減させること無く、低電力で動作できるようにしたキャッシュメモリおよびその制御方法に関する。

#### 背景技術

[0002] 近年、パーソナルコンピュータなどの情報処理装置において、CPU（Central Processing Unit）とメインメモリとの性能差を吸収し、処理を円滑に実行させるものとしてキャッシュメモリが欠かせない存在となっている。

[0003] キャッシュメモリとは、CPUなど処理装置がデータや命令などの情報を取得または更新する際にメインメモリやバスなどの遅延または低帯域を隠蔽化させ、処理装置と記憶装置の性能差を埋めるために用いる高速小容量メモリである。

[0004] コンピュータは以前から記憶装置の性能が処理装置の性能に追いつけず、この差が全体性能に対するボトルネックとされてきた（ノイマンズ・ボトルネック）。そして、ムーアの法則に基づく処理装置の加速度的な高性能化により現在ではますますこの差が拡大されている。キャッシュメモリは、記憶階層の観点からこれを解消しようとするものであり、一般に、主記憶装置（メインメモリ）とCPUなど処理装置との間に構成される。

[0005] 従来のキャッシュメモリの構成として、図1で示される4ウェイセットアソシアティブ方式のものが知られている。

[0006] 図1の場合、キャッシュメモリは、4ウェイ（連想度4）のSRAM（Static Random Access Memory）から構成されており、各ウェイ毎に、インデックスアドレスが設定され、そのインデックスアドレス毎にタグアドレスが設けられており、このタグに対応付けてデータが記憶されている。図1においては、各ウェイは、00、01、10、11の2ビットの識別情報で区別されて

管理されている。また、アドレスは、図1の上部で示されるように、タグアドレスとインデックスアドレスから構成されている。すなわち、図1においては、1個のインデックスアドレスが特定されると、4個のタグが特定される。ここで、ウェイとは、同一のインデックスアドレスにより指定できるタグアドレスの数のことであり、連想度とも言う。

[0007] 図1のキャッシュメモリの場合、GPUがアドレスを指定して、データの読み出しを要求すると、まず、アドレスのうち、インデックスアドレスに基づいてデータが特定される。図1の場合、4ウェイで構成されているため、インデックスアドレスが特定されることにより、同一のインデックスアドレスに対して、4個の各ウェイのタグアドレスが指定されるため、4個の読み出しデータの候補が特定される。

[0008] 次に、GPUにより指定されたアドレスのうちのタグアドレスの情報と、各ウェイのタグアドレスとを比較し、比較結果を出力すると共に、タグアドレスで指定される4個のデータが同時に読み出される。この結果、一致するタグアドレスが存在した場合、キャッシュメモリには、GPUから要求されたデータが存在する（キャッシュヒットする）ことになる。そこで、出力された4個のデータのうち、同一のタグアドレスで管理されていたデータのみがGPUに供給され、その他のデータは破棄される。一方、同一のタグアドレスがない場合、キャッシュメモリには、GPUから要求されたデータが存在しない（キャッシュミスする）ことになるので、キャッシュメモリは、要求されたアドレスのデータをメインメモリより読み出して、GPUに供給すると共に、最も前のタイミングで読み出されたデータに上書き更新する。

[0009] しかしながら、この手法では、アドレスのうちインデックスアドレスで特定できるのはウェイ数分の候補だけであり、さらに、この中で必要とされるデータは0または1であることから、インデックスアドレスで特定されたデータが、仮に用いられても、（ウェイ数-1）個分のミスが発生することとなる。SRAMは、読み出しの際に消費される電力が大きいいため、ミスを低減させない限り消費電力の低減を期待できない。このため、インデックスアドレ

スで特定されるデータの個数、つまり、ウェイ数を減らさない限り消費電力を低減させることができない。しかしながら、ウェイ数が減れば、キャッシュヒット率は低減することとなるため、性能を低減させる一因となってしまう。

[0010] そこで、このように1のインデックスアドレスで特定されるデータ数に起因するミスを抑減させるために、CAM (Content Addressable Memory) を用いた高連想度キャッシュメモリが提案されている。

[0011] 図2は、この高連想度キャッシュメモリの構成を示している。図2の高連想度キャッシュメモリは、ラインサイズが32 bytesであり、容量が8KBとなる32ウェイの高連想度キャッシュメモリである。図2の高連想度キャッシュメモリにおいては、上述したインデックスアドレスに対応する8個のサブバンク (各1KB) に区切ることにより、1回のGPUからの要求 (以下キャッシュアクセスとも称する) に対して、1のサブバンクだけをアクティブにし、他のサブバンクの電力消費を抑減することで、電力削減が図られている。

[0012] すなわち、アドレスのうちインデックスアドレスの情報がデコーダ (decoder) によりデコードされて、キャッシュアクセスすべきサブバンクに供給される。これにより、特定されたサブバンクのみがアクティブにされ、アクティブにされたバンクにおける、タグアドレスを管理するCAMにタグアドレスの情報を供給する。CAMは、供給されたタグアドレスに基づいて、全ウェイについて検索する。そして、アクティブな1個のサブバンクにおいては、CAMにより全ウェイに対して並列にタグアドレスとの比較が行われ、一致したタグアドレスに対応して格納されているデータだけがSRAMより出力される。

[0013] しかしながら、CAMを駆動させる場合、各ウェイについて全てのタグアドレスとの比較が実行されるため、1回のGPUからの読み出しに対して、連想度分のCAMが駆動されることにより電力消費が大きいという問題があった。

[0014] そこで、この問題を解決させるために、いくつかの手法が提案されているが、その中にLPHAC (Low Power Highly Associative Cache) 法と呼ばれるものがある (非特許文献1参照)。

- [0015] LPHAC法は、タグアドレスが24ビットで構成される場合、図3で示されるように、タグアドレス（以下、単にタグとも称する）を上位ビットと下位ビットの2つのサブタグアドレス（以下、単にサブタグとも称する）に分ける手法である。従来の高連想度キャッシュでは、図3のタグアドレス構成（a）で示されるように、全てCAMで管理するタグアドレス（以下、CAMタグとも称する）で構成されている。これに対し、LPHAC法では、図3のタグアドレス構成（b）で示されるように、タグアドレスの下位sビットのサブタグアドレスをCAMで管理し（以下、CAMサブタグアドレスまたはCAMサブタグとも称する）、上位ビット（この例では24-sビット）のサブタグアドレスをSRAMで管理する（以下、SRAMサブタグアドレスまたはSRAMサブタグとも称する）。例えば、ウェイ数が32の場合、各ラインを区別するには $s \geq 5$ ビット以上必要となる。
- [0016] 動作は、はじめにCAMサブタグアドレスとの部分的な比較検索が行われ、このときに検索できなければ（ミスすれば）キャッシュミスとなる。非特許文献1によれば、 $s = 8$ のとき、CAMサブタグアドレスとの部分的な比較検索のみでキャッシュミスの85%が判明するといわれている。そして、CAMサブタグアドレスでヒットした場合は、ヒットしたラインに対してSRAMサブタグアドレスとの比較検索が行われる。より詳細には、最初の半クロックでCAMサブタグアドレスとの部分的な比較検索が行われ、後半の半クロックで、CAMサブタグアドレスで特定されたラインについて、SRAMサブタグアドレスによる部分的な比較検索が行われ、同時に、データが読み出される。
- [0017] ここで、図4を参照して、具体的な比較検索例を参照して説明する。尚、図4においては、説明の便宜上、アドレスが、6ビットからなる場合の比較検索例について説明する。
- [0018] まず、図4の比較例aで示されるように、アドレスが「101000」であって、CAMサブタグが「1000」、およびSRAMサブタグが「10」である場合、図示せぬキャッシュメモリにCAMサブタグとして、「1111」、「0101」、および「1000」が登録されており、それらに対応付けて、SRA

Mサブタグが「11」、「10」、および「10」として登録されているものとする。すなわち、比較例aにおいては、アドレス「111111」、「100101」、「101000」に対応するデータが図示せぬキャッシュメモリに記憶されている。

[0019] 比較例aの場合、入力されたアドレス（入力アドレス）のCAMサブタグが用いられて、キャッシュメモリ内のCAMサブタグが部分的に比較検索されることにより、図中丸印で囲まれて示されるように、3段目の「1000」が一致するものとして検索されることになるので、キャッシュヒットとなる。このため、CAMサブタグ「1000」に対応付けて登録されているSRAMサブタグ「10」と、入力アドレスのSRAMサブタグと比較検索されるが一致するので、同時に読み出されたデータがCPUに読み出されることになる。

[0020] また、図4の比較例bで示されるように、アドレスが「100000」であって、CAMサブタグが「0000」、およびSRAMサブタグが「10」である場合、図示せぬキャッシュメモリには、CAMサブタグとして、「1111」、「0101」、および「1000」が登録されており、それらに対応付けて、SRAMサブタグが「11」、「10」、および「10」として登録されているものとする。すなわち、比較例bにおいては、アドレス「111111」、「100101」、「101000」に対応するデータが図示せぬキャッシュメモリに記憶されている。

[0021] 比較例bの場合、まず、入力アドレスのCAMサブタグが用いられて、キャッシュメモリ内のCAMサブタグが比較検索されることにより、「0000」が検索されるが一致するCAMサブタグがない。すなわち、この場合、キャッシュミスとなる。ところが、CAMサブタグは、4ウェイであり、図4の比較例bにおいては、1ウェイ分の空きがある。そこで、メインメモリよりアドレス「100000」に対応するデータが読み出されて、CPUに供給されると共に、図4の比較例cで示されるように、キャッシュメモリには、空いている最下段の1ウェイにCAMサブタグ「0000」が登録され、さらに、対応付けてSRAMサブタグ「10」が登録されて、今メインメモリから読み出されたデータが

登録される。

- [0022] さらに、図4の比較例dで示されるように、入力アドレスが「001000」であって、CAMサブタグが「1000」、およびSRAMサブタグが「00」である場合、CAMサブタグとして、「1111」、「0101」、および「1000」が登録されており、それらに対応付けて、SRAMサブタグが「11」、「10」、および「00」として登録されているものとする。すなわち、比較例dにおいては、アドレス「111111」、「100101」、「101000」に対応するデータが図示せぬキャッシュメモリに記憶されている。
- [0023] このとき、まず、入力アドレスのCAMサブタグが用いられて、キャッシュメモリ内のCAMサブタグが比較検索されることにより、「1000」が検索されて、図中丸印で囲んで示されるように、同一のものが検索される。次に、入力アドレスのSRAMサブタグと、検索されたSRAMサブタグ「10」とが比較検索されるが、図中の×印で示されるように一致しない。すなわち、この場合、キャッシュミスとなる。しかしながら、既に、CAMサブタグとして「1000」が登録されていることから、1ウェイ分の空きがあっても、メインメモリから新たにアドレスが「001000」に対応するデータが読み出されて登録されると、CAMサブタグ「1000」に対して重複して、SRAMサブタグ「00」が登録されてしまう。すなわち、同一のCAMサブタグ「1000」が重複して登録されて、それぞれSRAMサブタグ「10」、および「00」が登録されてしまう。
- [0024] ところが、LPHAC法においては、同一のCAMサブタグの重複が回避されるように登録が管理される。このため、メインメモリよりアドレス「001000」に対応するデータが読み出されて、CPUに供給されると共に、図4の比較例eで示されるように、登録済みのCAMサブタグ「1000」に対応付けて、新たに読み出されたSRAMサブタグ「00」が上書きされて、図示せぬ今メインメモリから読み出されたデータが登録される。すなわち、この場合、登録されていたアドレス「101000」に対応するデータは、破棄され、CAMサ



ブタグ「1000」に対応付けて登録されるSRAMサブタグが1に維持される。

## 先行技術文献

## 非特許文献

- [0025] 非特許文献1 : Zhang, C. : A Low Power Highly Associative Cache for Embedded Systems, Proc. IEEE ICCD, pp.31-36 (2006).

## 発明の概要

### 発明が解決しようとする課題

- [0026] ところで、上述したLPHAC法では、CAMサブタグのミスによる置換はLRU (Least Recently Used) 方式で行うことが前提とされている。LRU方式とは、最も古くアクセスされたデータを、最近読み出されたデータで上書きする方式である。すなわち、時間的局所性に鑑みれば、過去最もアクセスのなかったデータは将来にわたってもアクセスされる可能性は少ないと言える。従って、この方法はヒット率を向上させるための方法としてよく採用されている。
- [0027] 従って、CAMサブタグでヒットした場合、一般には複数のSRAMサブタグに対応するCAMサブタグと一致することになるが、LPHAC法ではCAMサブタグとの部分的な比較検索において、ヒットするデータが上述したように1個に絞られる。
- [0028] すなわち、CAMサブタグでヒットし、SRAMサブタグでミスした場合、ミスしたデータが置換される対象のデータとされる（新たに読み出されたデータにより置換されて、消滅してしまうデータとされる）。これにより、CAMサブタグのビット数  $s \geq 5$  ビットであればCAMサブタグの部分的な比較検索により複数のデータと一致しないようにできる。
- [0029] しかしながら、LPHAC法では、CAMサブタグでヒットし、SRAMサブタグでミスした場合、ミスしたデータが置換される対象データとされることによりヒットするデータが1個に絞られる。この場合、LRU方式とは異なる置換方式を取らざるを得ず、比較的最近アクセスがあったデータであっても置換により

消滅してしまう可能性があり、結果として、ヒット率が低下してしまう原因となっていた。

[0030] また、CAMサブタグのビット数  $s$  を小さく設定すると、このようなケースが増加し、LRU方式とは異なる置換方式を取る頻度が上がる。これにより、CAMサブタグのビット数  $s$  を小さく設定することができないため、LPHAC法では、消費電力の大きなCAM部分の縮小に限界があり、消費電力の低減には限界があった。

[0031] 本発明はこのような状況に鑑みてなされたものであり、特に、キャッシュメモリのヒット率を低減させること無く、消費電力を低減できるようにするものである。

### 課題を解決するための手段

[0032] 本発明の第1の側面のキャッシュメモリは、連想度  $n$  ( $n$ は自然数)のCAM (Content Addressable Memory) とSRAM (Static Random Access Memory) とを有し、接続されるCPU (Central Processing Unit) より第1サブタグアドレスおよび第2サブタグアドレスとからなるタグアドレスを指定され、対応するデータを格納または読み出すキャッシュメモリであって、前記データを、読み出し要求された時期に応じて、少なくとも、最近読み出しが要求された第1世代と、前記最近ではない時期に読み出しが要求された第2世代とに分類し、前記CAMにて、前記第1サブタグアドレスを管理し、前記SRAMにて、前記第2サブタグアドレスを管理し、同一の前記第1サブタグアドレスに対し、複数の前記第2サブタグアドレスが対応付けられることを容認することを特徴とする。

[0033] 本発明の第2の側面のキャッシュメモリは、連想度  $n$  ( $n$ は自然数)のCAM (Content Addressable Memory) とSRAM (Static Random Access Memory) とを有し、接続されるCPU (Central Processing Unit) より第1サブタグアドレスおよび第2サブタグアドレスとからなるタグアドレスを指定され、対応するデータを格納または読み出すキャッシュメモリであって、前記データを、読み出し要求された時期に応じて、少なくとも、最近読み出しが要求され

た第1世代と、前記最近ではない時期に読み出しが要求された第2世代とに分類する世代管理手段と、前記CAMにて、前記第1サブタグアドレスを管理し、前記CPUより指定される前記タグアドレスの第1サブタグアドレスとの比較により、対応する第1サブタグアドレスを検索する第1検索手段と、前記SRAMにて、前記第2サブタグアドレスを管理し、前記第1検索手段により検索された第1サブタグアドレスを含み、かつ、前記第1世代のデータに対応する前記タグアドレスの第2サブタグアドレスとの比較により、対応する第2サブタグアドレスを検索する第2検索手段と、前記第2検索手段により検索された、前記第2サブタグアドレスに対応付けて記憶されている前記データを出力する出力手段とを含み、同一の前記第1サブタグアドレスに対し、複数の前記第2サブタグアドレスを対応付けることを容認することを特徴とする。

- [0034] 前記第2検索手段には、前記第1世代のデータに対応する前記第1サブタグアドレスの第2サブタグアドレスとの比較により、対応する第2サブタグアドレスを検索できないとき、前記第2世代のデータに対応する前記第1サブタグアドレスの前記第2サブタグアドレスとの比較により、対応する第2サブタグアドレスを検索させるようにすることができる。
- [0035] 前記第1検索手段による前記第1サブタグアドレスの比較によりデータが検索されないとき、または前記第2検索手段により前記第2サブタグアドレスの比較により第2世代のデータが検索されないとき、接続するメインメモリより前記読み出しが要求されたタグアドレスに対応するデータを読み出し、前記世代管理手段にて管理される、前記第2世代のデータのうち、最も古く読み出されたデータを、前記メインメモリより読み出されたデータで置換し、併せて前記第1サブタグアドレス、および第2サブタグアドレスを置換する置換手段をさらに含ませるようにすることができる。
- [0036] 前記置換手段には、一の前記第1サブタグアドレスに対し、複数の前記第2サブタグアドレスが対応付けられて置換されることを許容するようにさせることができる。

- [0037] 前記世代管理手段には、前記第2のサブタグアドレスにより管理される各データが読み出された時期に基づいて、前記データが第1世代であるか、または、前記第2世代であるかを世代管理リストにより管理し、前記置換手段により前記メインメモリより読み出されたデータで、前記最も古く読み出されたデータが置換されるとき、前記第2サブタグアドレスにより管理されるデータが読み出された時期が、最も古く読み出されたデータを削除し、前記メインメモリより読み出されたデータを、最近読み出された時期として世代管理リストを更新させるようにすることができる。
- [0038] 前記世代管理手段には、前記世代管理リストに基づいて、前記第2サブタグアドレスにより管理される各データについて、最近読み出されたデータを第1世代のデータとし、最近ではない時期に読み出されたデータを第2世代のデータとして、世代管理させるようにすることができる。
- [0039] 前記連想度  $n$  が連想度 3 2 であって、前記タグアドレスが 2 4 ビットの場合、第1サブタグアドレスが 2 ビットとし、第2サブタグアドレスが 2 2 ビットとすることができる。
- [0040] 前記連想度  $n$  が連想度 3 2 の前記第1サブタグアドレスで管理されるデータのうち、前記第1世代のデータは、読み出された時期が最近に近い順に上位 2 位までのデータとすることができる。
- [0041] 本発明のプロセッサは、複数の請求項 1 乃至 8 のいずれかに記載のキャッシュメモリと、前記CPUと、前記CPUより前記タグアドレスで指定されるデータの読み出し、または書き込みが要求された場合、前記タグアドレスより前記第1のサブタグアドレスを抽出する第1のサブタグアドレス抽出手段と、前記CPUより前記タグアドレスで指定されるデータの読み出し、または書き込みが要求された場合、前記タグアドレスより前記第2のサブタグアドレスを抽出する第2のサブタグアドレス抽出手段とを含むことを特徴とする。
- [0042] 本発明の第2の側面のキャッシュメモリの制御方法は、連想度  $n$  ( $n$  は自然数) のCAM (Content Addressable Memory) とSRAM (Static Random Access Memory) とを有し、接続されるCPU (Central Processing Unit) より第1サ

ブタグアドレスおよび第2サブタグアドレスとからなるタグアドレスを指定され、対応するデータを格納または読み出すキャッシュメモリの制御方法であって、前記データを、読み出し要求された時期に応じて、少なくとも、最近読み出しが要求された第1世代と、前記最近ではない時期に読み出しが要求された第2世代とに分類する世代管理ステップと、前記CAMにて、前記第1サブタグアドレスを管理し、前記CPUより指定される前記タグアドレスの第1サブタグアドレスとの比較により、対応する第1サブタグアドレスを検索する第1検索ステップと、前記SRAMにて、前記第2サブタグアドレスを管理し、前記第1検索ステップの処理により検索された第1サブタグアドレスを含み、かつ、前記第1世代のデータに対応する前記タグアドレスの第2サブタグアドレスとの比較により、対応する第2サブタグアドレスを検索する第2検索ステップと、前記第2検索ステップの処理により検索された、前記第2サブタグアドレスに対応付けて記憶されている前記データを出力する出力ステップとを含み、同一の前記第1サブタグアドレスに対し、複数の前記第2サブタグアドレスが対応付けることを容認することを特徴とする。

[0043] 本発明の第1の側面においては、連想度  $n$  ( $n$ は自然数)のCAM (Content Addressable Memory) とSRAM (Static Random Access Memory) とを有し、接続されるCPU (Central Processing Unit) より第1サブタグアドレスおよび第2サブタグアドレスとからなるタグアドレスを指定され、対応するデータが格納または読み出され、前記データが、読み出し要求された時期に応じて、少なくとも、最近読み出しが要求された第1世代と、前記最近ではない時期に読み出しが要求された第2世代とに分類され、前記CAMにて、前記第1サブタグアドレスが管理され、前記SRAMにて、前記第2サブタグアドレスが管理され、同一の前記第1サブタグアドレスに対し、複数の前記第2サブタグアドレスを対応付けることが容認される。

[0044] 本発明の第2の側面においては、連想度  $n$  ( $n$ は自然数)のCAM (Content Addressable Memory) とSRAM (Static Random Access Memory) とを有し、接続されるCPU (Central Processing Unit) より第1サブタグアドレスおよび

第2サブタグアドレスとからなるタグアドレスを指定され、対応するデータを格納または読み出すキャッシュメモリであって、前記データが、読み出し要求された時期に応じて、少なくとも、最近読み出しが要求された第1世代と、前記最近ではない時期に読み出しが要求された第2世代とに分類され、前記CAMにて、前記第1サブタグアドレスが管理され、前記CPUより指定される前記タグアドレスの第1サブタグアドレスとの比較により、対応する第1サブタグアドレスが検索され、前記SRAMにて、前記第2サブタグアドレスが管理され、検索された第1サブタグアドレスが含まれ、かつ、前記第1世代のデータに対応する前記タグアドレスの第2サブタグアドレスとの比較により、対応する第2サブタグアドレスが検索され、検索された、前記第2サブタグアドレスに対応付けて記憶されている前記データが出力され、同一の前記第1サブタグアドレスに対し、複数の前記第2サブタグアドレスを対応付けることが容認される。

### 発明の効果

[0045] 本発明の一側面によれば、キャッシュメモリのヒット率を低減させることなく、消費電力を低減させることが可能となる。

### 図面の簡単な説明

- [0046] [図1]従来のキャッシュメモリの構成を示す図である。  
[図2]その他の従来のキャッシュメモリの構成を示す図である。  
[図3]LPHAC法を説明する図である。  
[図4]LPHAC法を説明する図である。  
[図5]本発明を適用した一実施の形態の構成例であるキャッシュメモリにより実現される機能を説明する機能ブロック図である。  
[図6]キャッシュメモリに格納されるデータの管理構造を説明する図である。  
[図7]プログラム実行処理を説明するフローチャートである。  
[図8]命令キャッシュ処理を説明するフローチャートである。  
[図9]キャッシュヒットミス判定処理を説明するフローチャートである。  
[図10]キャッシュヒットミス判定処理を説明する図である。

[図11] データキャッシュ（書き込み）処理を説明するフローチャートである。

[図12] データキャッシュ（読み出し）処理を説明するフローチャートである。

[図13] 本発明の効果を説明する図である。

[図14] 本発明の効果を説明する図である。

[図15] 本発明の効果を説明する図である。

[図16] 本発明の効果を説明する図である。

### 発明を実施するための最良の形態

[0047] [本発明を適用したキャッシュメモリの構成例]

図5は、本発明を適用したキャッシュメモリにより実現される機能の一実施の形態の構成例を示す機能ブロック図である。

[0048] CPU (Central Processing Unit) 11は、各種の処理を実行させるとき、プログラムカウンタより実行すべき命令のアドレスをアドレス入力部12に供給し、キャッシュメモリ13に対してデータを要求すると共に取得して実行する。

[0049] アドレス入力部12は、CPU11より命令を指定するアドレスが供給されると、インデックスアドレスをキャッシュメモリ13のデコード部41に供給すると共に、タグアドレスのうち、CAM (Content Addressable Memory) サブタグアドレスをキャッシュメモリ13のCAM51に供給すると共に、SRAM (Static Random Access Memory) サブタグアドレスをSRAM52に供給する。より詳細には、アドレス入力部12は、インデックスアドレス抽出部21、およびタグアドレス抽出部22を備えている。インデックスアドレス抽出部21は、入力されたアドレスよりインデックスアドレスを抽出し、抽出したインデックスアドレスをデコード部41に供給する。

[0050] タグアドレス抽出部22は、さらにCAMサブタグアドレス抽出部22aおよびSRAMサブタグアドレス抽出部22bを備えている。CAMサブタグアドレス抽出部22aは、入力アドレスに含まれるタグアドレスが、例えば、Xビットで

ある場合、その下位sビットをCAMサブタグアドレスとして抽出してキャッシュメモリ13内にある各サブバンク42のCAM51に供給する。また、SRAMサブタグアドレス抽出部22bは、入力アドレスに含まれるタグアドレスのうち、CAMサブタグアドレス以外の上位(X-s)ビットをSRAMサブタグとして抽出してキャッシュメモリ13内にある各サブバンク42のSRAM52に供給する。

- [0051] キャッシュメモリ13は、デコード部41、およびサブバンク42より構成されている。デコード部41は、アドレス入力部12より供給されてくるインデックスアドレスを解析し、いずれのサブバンク42を起動させるかを判定し、起動させるべきサブバンク42に対して起動を指示する信号を供給する。尚、図5においては、サブバンク42の構成を示すため、便宜上サブバンク42が1個のみ表示されているが、実際には、複数のサブバンク42が設けられている。すなわち、デコード部41とサブバンク42の関係については、図2で示される構成と同様の関係となっており、デコード部41より起動を指示する信号が供給されたサブバンクのみがオンの状態とされ、オンの状態となったサブバンク42のみが、タグアドレスの情報に基づいて処理を実行する。
- [0052] サブバンク42には、CAM51、SRAM52、LRU管理部53、読出部54、および置換部55を備えている。CAM51は、デコード部41より供給されてくる起動を指示する信号に基づいて動作を開始し、CAMサブタグアドレス抽出部22aより供給されてくるCAMサブタグアドレスに基づいて、CPU11で指定したアドレスのデータの候補を検索し、検索結果をヒットフラグとしてSRAM52に供給する。
- [0053] SRAM52は、CAM51からのヒットフラグの情報、およびSRAMサブタグアドレスに基づいて、CPU11がアドレスで指定したデータを検索し、データが検索できた場合、CPU11に出力する。また、SRAM52は、CPU11がアドレスで指定したデータが検索できなかった場合、メインメモリ14にアクセスし、そのデータを読み出して、CPU11に供給すると共に、最も古いデータと置



換して新たに記憶する。

- [0054] より詳細には、SRAM 5 2は、比較検索部 7 1、出力部 7 2、書戻部 7 3、置換部 5 5 世代管理部 7 4、書戻管理部 7 5、書込部 7 6、書換部 7 7、および格納部 7 8を備えている。比較検索部 7 1は、世代判定部 8 1を制御して、CAM 5 1からのヒットフラグの情報により特定されている、検索されるデータの候補のうち、まず、最初の処理で第 1 世代のもののみを特定させる。
- [0055] ここでいう世代とは、世代管理部 7 4により管理されるものであり、SRAM 5 2の格納部 7 8に格納されている各データが最近アクセスされた日時により設定される順序のうち、例えば、上位 2 位までのデータが第 1 世代に設定され、それ以下の順位のデータが第 2 世代に設定される。つまり、最近アクセスされた、読み出される可能性の比較的高いデータが第 1 世代に設定され、それ以外のデータが第 2 世代に設定される。尚、上位 2 位までが第 1 世代に設定される例について説明したが、それ以外の順位までを第 1 世代とし、それ以外のデータを第 2 世代としてもよいものである。
- [0056] 比較検索部 7 1は、最初の処理で世代判定部 8 1を制御して、CPU 1 1によりアドレスで特定される読み出しが要求されたデータの候補のうち、最新のものとして分類されている第 1 世代のデータを特定させる。そして、比較検索部 7 1は、第 1 世代のものとして特定されたCAMサブタグアドレスで検索されたアドレスにより指定されたデータの候補のそれぞれのSRAMサブタグアドレスと、SRAMサブタグアドレス抽出部 2 2 bより供給されてきたSRAMサブタグアドレスとを比較して検索し、一致するものが検索できた場合、検索されたデータを出力部 7 2に供給する。
- [0057] 出力部 7 2は、検索されたデータを取得すると、CPU 1 1によりアドレスで指定されている要求されたデータとして供給する。また、第 1 世代のものとして特定されたCAMサブタグで検索されたアドレスにより指定されたデータの候補のSRAMサブタグアドレスのうち、抽出されたSRAMサブタグアドレスと一致するものが検索できなかった場合、比較検索部 7 1は、世代判定部 8 1により第 1 世代と判定されなかった、すなわち、第 2 世代のデータのSRAMサブ

タグアドレスと、SRAMサブタグアドレス抽出部22bより供給されてきたSRAMサブタグアドレスとを比較検索し、一致するものが検索された場合、検索されたデータをCPU11に出力部72に供給する。

[0058] 一方、比較検索部71は、いずれの世代においてもCPU11によりアドレスで特定されたデータが検索できない場合、CAMサブタグアドレスの情報、およびSRAMサブタグアドレスの情報と共に、対応するデータをメインメモリ14より読み出すように読出部54に通知する。読出部54は、CAMサブタグアドレス、およびSRAMサブタグアドレスの情報に基づいて、タグアドレスを特定し、メインメモリ14にアクセスして、そのタグアドレスに対応するデータを読み出し、CPU11に供給すると共に、CAM51、SRAM52、および置換部55に対して読み出したデータのCAMサブタグ、SRAMサブタグ、およびデータそのものを供給する。

[0059] 書戻部73は、CPU11からの命令が書き込み処理である場合、サブバンク42により管理すべきデータが更新される時、SRAM52の格納部78に格納されているデータを置換して更新する前に、メインメモリ14に対応するデータを書き戻す。

[0060] 置換部55は、CPU11からの命令により指定されたタグアドレスの情報が検出できず、メインメモリ14からデータが読み出された場合、新たなデータとタグアドレスとで、それまで記憶していたデータとタグアドレスとを置換して更新する。

[0061] 書込部76は、CPU11からの命令が書き込み処理の場合、格納部78に格納されているデータをCPU11からの処理結果であるデータで書き換える。

[0062] 書換部77は、CPU11からの命令が書き込み処理の場合、書込部76が格納部78に格納されているデータをCPU11からの処理結果であるデータで書き換える前に、メインメモリにアクセスし、対応するデータを書き換える。

[0063] 尚、LRU管理部53、世代管理部74、および書戻管理部75の詳細については、後述するデータ管理構造と併せて説明するものとする。

[0064] [データ管理構造]

次に、図6を参照して、格納部78に格納されるデータの管理構造について説明する。図6においては、垂直方向に各データ単位のラインが形成されており、ラインL1乃至L32までの合計32データが記録されていることが示されている。尚、このラインL1乃至L32のそれぞれを特に区別する必要がない場合、単にラインLと称するものとする。

- [0065] 図6の各ラインLには、左からCAMサブタグアドレス、SRAMサブタグアドレス、データ、アクセス日時情報、世代情報、および書戻フラグが含まれており、データ単位で管理される情報が示されている。このラインLの数は、CAM51およびSRAM52の物理的な構成により決定されるウェイ数に相当しており、図6の例では、32ウェイの例であることが示されている。ただし、このウェイ数は、ハードウェアの構成により決まるものであるため、ハードウェアの構成により、ウェイ数は様々な数をとることが可能であり、例えば、2、4、8、16、32などでもよいし、これ以外のウェイ数であってもよい。
- [0066] 尚、図6においては、各データを特定するタグアドレスは、図6の左側の2列で示されている。すなわち、CAMサブタグアドレスとSRAMサブタグアドレスとが特定されることにより、CPU11により指定されたタグアドレスのデータが検索されることになる。
- [0067] アクセス日時情報は、LRU (Least Recently Used) 管理部53のLRU管理リスト53aによりデータ単位（ライン単位）で管理されている情報である。すなわち、LRU (Least Recently Used) 管理部53は、LRU管理リスト53aを備えており、LRU管理リスト53a上にデータ単位で（ライン単位で）、図6中のデータ列の右側に表示されている最近アクセスされたアクセス日時情報をデータ単位（ライン単位）で管理している。すなわち、LRU管理部53は、SRAM52のアクセスがある度に、LRU管理リスト53aを更新して、図6におけるデータ毎（ライン毎）のアクセス日時の情報を管理している。また、図6には図示されていないが、LRU管理部53は、LRU管理リスト53aの情報に基づいて、各ラインを最近アクセスされた日時順に並び替えて、ライン

毎の順位を求める。

[0068] 世代情報は、世代管理部74の世代管理リスト74aによりデータ単位（ライン単位）で管理されている情報である。すなわち、世代管理部74は、世代管理リスト74aを備えており、LRU管理部53のLRU管理リスト53aにアクセスし、最近アクセスされたライン毎の日時順の情報に基づいて、世代管理リスト74aに、ライン単位で最近アクセスされた上位2位までのデータに第1世代データであることを示すフラグを立て、それ以外のデータについては第2世代データであることを示すため、第1世代であることを示すフラグを立てないようにして管理する。

[0069] 書戻フラグは、書戻管理部75の書戻管理リスト75aによりデータ単位（ライン単位）で管理されている情報である。すなわち、書戻管理部75は、書戻管理リスト75aを備えており、GPU11からの指令が書き込み処理である場合、データが更新される度に、SRAM52のデータを更新する前に、先にメインメモリ14のデータを書き戻して更新する必要があるので、書戻が必要となるデータにフラグを立てて管理する。すなわち、書戻管理部75は、書き込み処理であるか否かにより、書戻管理リスト75aを更新し、図6におけるデータ毎（ライン毎）の書戻フラグを管理している。

[0070] [プログラム実行処理]

次に、図7のフローチャートを参照して、プログラム実行処理について説明する。

[0071] ステップS1において、GPU11は、図示せぬプログラムカウンタより実行すべき命令のアドレスを取得し、アドレス入力部12に供給する。この処理により、アドレス入力部12は、GPU11よりアドレスの入力を受け付ける。

[0072] インデックスアドレス抽出部21は、入力されたアドレスよりインデックスアドレスを抽出する。また、タグアドレス抽出部22は、入力されたアドレスよりタグアドレスを抽出し、CAMサブタグアドレス抽出部22aおよびSRAMサブタグアドレス抽出部22bに供給する。CAMサブタグアドレス抽出部22aは、タグアドレスの、下位sビットをCAMサブタグアドレスとして抽出し

てキャッシュメモリ 13 内にある各サブバンク 42 のCAM 51 に供給する。SRAM サブタグアドレス抽出部 22 b は、タグアドレスのうち、CAM サブタグアドレス以外の上位 (X-s) ビットを SRAM サブタグアドレスとして抽出してキャッシュメモリ 13 内にある各サブバンク 42 の SRAM 52 に供給する。

[0073] ステップ S2 において、キャッシュメモリ 13 は、命令キャッシュ処理を実行し、CPU 11 に入力されたアドレスに対応するデータを読み出し、CPU 11 に供給する。

[0074] [命令キャッシュ処理]

ここで、図 8 のフローチャートを参照して、命令キャッシュ処理について説明する。

[0075] ステップ S21 において、キャッシュメモリ 13 は、いずれかのサブバンク 42 を指定してキャッシュヒットミス判定処理を実行させる。

[0076] [キャッシュヒットミス判定処理]

ここで、図 9 のフローチャートを参照して、キャッシュヒットミス判定処理について説明する。

[0077] ステップ S41 において、インデックスアドレス抽出部 21 は、抽出したインデックスアドレスをデコード部 41 に供給する。デコード部 41 は、インデックスアドレスを解析し、解析結果からキャッシュヒットミス処理を実行させるサブバンク 42 を特定し、特定したサブバンク 42 に対して起動を指示する信号を供給する。この処理により、起動が指示されたサブバンク 42 は、電力供給を受けて起動を開始する。一方、起動が指示されていないサブバンク 42 は起動されず、結果として、処理が必要とされていないサブバンク 42 の電力消費が低減される。

[0078] ステップ S42 において、CAM 51 は、タグアドレス抽出部 22 の CAM サブタグアドレス抽出部 22 a より供給されてくる CAM タグアドレスと、SRAM 52 内に管理されるデータに対応付けて登録されている CAM タグアドレスとを比較し、一致するものが検索されるか否かを判定する。

[0079] すなわち、例えば、図 10 の左部で示されるように、ライン L0 の CAM タグ

アドレスが「0001」であり、ラインL1が「0101」であり、ラインL2が「1001」であり、ラインL3が「0000」であり、ラインL4が「1001」であり、ラインL5が「0001」であり、ラインL6が「0010」であり、・・・ラインL28が「0001」であり、ラインL29が「0111」であり、ラインL30が「1001」であり、ラインL31が「1000」である場合、入力されたアドレスから抽出されたCAMサブタグアドレスが「1001」であるとき、CAM51は、全てのラインLのCAMサブアドレスとを比較して検索すると、図中の矢印で示されているラインL2, L4, L30のCAMサブタグアドレスがヒットすることになる。そこで、このような場合、ステップS42において、CAM51は、タグアドレス抽出部22のCAMサブタグアドレス抽出部22aより供給されてくるCAMタグアドレスと、SRAM52内に管理されるデータに対応付けて登録されているCAMタグアドレスとに一致するものがあると判定し、処理は、ステップS43に進む。

[0080] 尚、図10においては、垂直方向にラインLごとに左からCAMサブタグアドレス、SRAMサブタグアドレス、データ、および世代情報が対応付けて示されている。また、図10の世代情報においては、ラインL2, L3, L29のみが第1世代(1st)であり、それ以外が第2世代(2nd)であることが示されている。

[0081] ステップS43において、CAM51は、一致したCAMタグアドレス、および一致したラインLを示す情報として、一致したラインLに対してヒットフラグを立ててSRAM52に供給する。この処理により、SRAM52の比較検索部71は、一致したCAMタグアドレスとラインLの情報を取得する。すなわち、図10の左部においては、ラインL2, L4, L30についてヒットフラグが付されることになる。

[0082] そして、比較検索部71は、世代判定部81を制御して、CAMタグアドレスが一致したラインLのうち、第1世代のラインLの情報を比較対象のラインLとして検索させる。例えば、図10の左部の場合、ラインL2のデータのみが第1世代に設定されているので、ラインL2のみが検索される。さらに

、比較検索部 7 1 は、検索されたCAMタグアドレスが一致したラインLのうち、第 1 世代の、比較対象となったラインL 2 のSRAMサブタグアドレスと、SRAMサブタグアドレス抽出部 2 2 b より供給されてきたSRAMサブタグアドレスとを比較すると共に、比較対象となったラインL 2 で管理されているデータを格納部 7 8 より比較検出部 7 1 に対して出力させる。

[0083] ステップS 4 4において、比較検索部 7 1 は、比較対象となった第 1 世代のラインLのSRAMサブタグアドレスと、SRAMサブタグアドレス抽出部 2 2 b より供給されてきたSRAMサブタグアドレスとを比較して、一致するものがあるか否かを検索する。例えば、図 1 0 の左部の場合、検索されたのはラインL 2 のデータのみであるので、ラインL 2 のSRAMサブタグデータと、SRAMサブタグアドレス抽出部 2 2 b より供給されてきたSRAMサブタグアドレスとが比較される。

[0084] ステップS 4 4において、例えば、図 1 0 の左部で「hit」と示されるように、SRAMサブタグアドレスの両者が一致した場合、ステップS 4 5において、比較検索部 7 1 は、キャッシュヒットしたことを認識し、一致したSRAMサブタグアドレスに対応して格納部 7 8 より出力されてきたデータを出力部 7 2 に供給する。

[0085] 一方、ステップS 4 4において、一致するSRAMサブタグアドレスがない場合、ステップS 4 6において、比較検索部 7 1 は、世代判定部 8 1 を制御して、CAMタグアドレスが一致したラインLのうち、第 2 世代のラインLの情報を比較対象のラインLとして検索させる。例えば、図 1 0 の右部で示されるように、ラインL 4, L 3 0 のデータのみが第 2 世代のデータ（ラインL）として検索される。さらに、比較検索部 7 1 は、検索されたCAMタグアドレスが一致したラインLのうち、第 2 世代の、すなわち、比較対象となったラインLのSRAMサブタグアドレスと、SRAMサブタグアドレス抽出部 2 2 b より供給されてきたSRAMサブタグアドレスとを比較すると共に、比較対象となったラインLで管理されているデータを格納部 7 8 より比較検出部 7 1 に対して出力させる。

- [0086] ステップS 4 7において、比較検索部 7 1は、比較対象となった第 2 世代のライン L のSRAMサブタグアドレスと、SRAMサブタグアドレス抽出部 2 2 bより供給されてきたSRAMサブタグアドレスとを比較して、一致するものがあるか否かを検索する。例えば、図 1 0の右部の場合、検索されたのはライン L 4, L 3 0のデータであるので、ライン L 4, L 3 0のSRAMサブタグデータと、SRAMサブタグアドレス抽出部 2 2 bより供給されてきたSRAMサブタグアドレスとが比較される。
- [0087] ステップS 4 7において、例えば、図 1 0の右部でライン L 3 0において「hit」と示されるように、両者が一致した場合、ステップS 4 5において、比較検索部 7 1は、キャッシュヒットしたことを認識し、一致したSRAMサブタグアドレスに対応して格納部 7 8より出力されてきたデータを出力部 7 2に供給する。すなわち、この場合、ライン L 4, L 3 0の両方のデータが格納部 7 8から比較検索部 7 1に出力されている。しかしながら、SRAMサブタグアドレスが一致したのは、図中 (hit) と示されたライン L 3 0のものであるので、比較検索部 7 1は、ライン L 3 0のデータのみを出力部 7 2に供給すると共に、ライン L 4のデータを破棄する。このように、比較検索部 7 1は、ヒットしたデータのみを出力部 7 2に供給し、キャッシュミスとなったデータを破棄する。
- [0088] 一方、ステップS 4 2において、CAMサブタグアドレスに一致するものがない場合、または、ステップS 4 7において、例えば、比較対象となるSRAMタグアドレスとSRAMサブタグアドレス抽出部 2 2 bより供給されてきたSRAMサブタグアドレスが一致しない場合、ステップS 4 8において、比較検索部 7 1は、格納部 7 8には、CPU 1 1より供給されてきたアドレスに対応するデータが格納されていない、すなわち、キャッシュミスであるものと認識し、処理は終了する。
- [0089] すなわち、まず、CAMサブタグアドレスで比較対象となるラインを検索し、一致するものがあった場合、最初は、CAMサブタグアドレスが一致するラインのうち、第 1 世代のものだけを比較対象とするラインとして、SRAMサブタグ



アドレスで検索し、一致すればキャッシュヒットであるものと判定する。また、CAMサブタグアドレスが一致するラインのうち、第1世代の比較対象とするラインを、SRAMサブタグアドレスで検索しても一致するものがないとき、CAMサブタグアドレスが一致するラインのうち、第2世代の比較対象とするラインを、SRAMサブタグアドレスで検索するようにした。

[0090] このため、SRAMサブタグアドレスが一致する可能性の高い最近のデータを含む第1世代となる一部のラインのみを比較対象とすることができるので、SRAMサブタグアドレスでの一致の可能性の低いラインを排除することで、CAMサブタグアドレスで一致する全てのラインに対してSRAMサブタグアドレスによる検索をしないようにすることができる。

[0091] 結果として、全てのラインを比較対照としないため、キャッシュミスとなるラインが低減されてSRAM 52の消費電力としてインパクトの大きな格納部78からの出力動作を低減させることができ、全体として消費電力を低減させることが可能となる。また、SRAMサブタグアドレスが一致する可能性の高い最近のデータを含む第1世代となる一部のラインのみを比較対象とすることができるので、ヒット率を低減させずに、消費電力を低減させることが可能となる。

[0092] ここで、図8のフローチャートの説明に戻る。

[0093] ステップS22において、比較検索部71は、キャッシュヒットであるか否かを判定する。

[0094] ステップS22において、例えば、上述したステップS45の処理により、キャッシュヒットである場合、処理は、ステップS23に進む。

[0095] ステップS23において、出力部72は、比較検索部71より供給されてきたデータをCPU11に供給する。

[0096] ステップS24において、比較検出部71は、出力部72に供給したデータ（ライン）のCAMサブタグアドレスおよびSRAMサブタグアドレスを供給し、対応するデータにアクセスがあったことをLRU管理部53に通知する。LRU管理部53は、アクセスがあったことを示す情報に基づいて、LRU管理リスト5

3 a を更新する。より詳細には、LRU管理部 5 3 は、LRU管理リスト 5 3 a の情報のうち、アクセスがあったとして通知されてきたCAMサブタグアドレスおよびSRAMサブタグアドレスのデータの最近読み出された日時の情報を現在時刻に更新すると共に、読み出された日時の順位を第 1 位に更新する。尚、日時の順位を管理するに当たり、日時の情報をそのまま用いて管理すると、膨大なデータを管理する必要が生じる恐れがある。そこで、日時の順位管理には、日時の情報をそのまま扱うのではなく、例えば、3 2 ウェイであれば、5 ビットのカウンタビットを用いるようにして、データの順位のみを管理するようにしてもよい。

- [0097] ステップ S 2 5 において、世代管理部 7 4 は、LRU管理部 5 3 に対してLRU管理リスト 5 3 a を要求し、取得する。そして、世代管理部 7 4 は、LRU管理リスト 5 3 a における上位 2 位までのラインについて、世代管理リスト 7 4 a の世代情報を第 1 世代に更新し、併せて、それ以外のラインについて、世代情報を第 2 世代に更新する。
- [0098] 一方、ステップ S 2 2 において、キャッシュヒットではない、すなわち、キャッシュミスであったと判定された場合、処理は、ステップ S 2 6 に進む。
- [0099] ステップ S 2 6 において、比較検索部 7 1 は、読出部 5 4 に対して、CPU 1 1 よりアドレスにより指定されたデータが存在しないため、メインメモリ 1 4 より読み出すように指示すると共に、CAMサブタグアドレスおよびSRAMサブタグアドレスを供給する。読出部 5 4 は、この指示に応じて、CAMサブタグアドレスおよびSRAMサブタグアドレスに基づいてタグアドレスを特定し、メインメモリ 1 4 にアクセスして、データを読み出すと共に、CAM 5 1、比較検索部 7 1 および置換部 5 5 に供給する。
- [0100] ステップ S 2 7 において、置換部 5 5 は、LRU管理部 5 3 に対してLRU管理リスト 5 3 a を要求し、取得する。そして、置換部 5 5 は、LRU管理リスト 5 3 a において、最下位となるライン、すなわち、最も古いデータとタグアドレスとを検索し、読出部 5 4 より供給されてきたデータとタグアドレスとで

置換してCAM 5 1 およびSRAM 5 2 の格納部 7 8 の情報を更新する。

- [0101] 以上のように命令キャッシュ処理により、GPU 1 1により命令の読み出しが指示されたデータについて、LRU管理リスト 5 3 aのアクセスされた日時が更新されて、さらに、日時に応じて設定される順位が順次更新することが可能となる。また、LRU管理リスト 5 3 aの更新に応じて、世代管理部 7 4における世代管理リスト 7 4 aも併せて更新することが可能となり、最近アクセスされた日時に応じて世代情報を更新させることが可能となる。
- [0102] また、以上の処理においては、置換部 5 5は、同一のCAMサブタグアドレスに対して、異なる複数のSRAMサブタグアドレスが格納部 7 8に重複するようなデータが置換されることも認めるように処理する。この結果、格納部 7 8に格納されるデータは、同一のCAMサブタグアドレスに対して、SRAMサブタグアドレスが1に絞られるようなことがなく、LRU法に則って、最近アクセスされたデータが確実に置換されて格納されることとなるので、LPHAC法で見られたようなキャッシュヒット率の低下を抑制することが可能となる。
- [0103] ここで、図 7 のフローチャートの説明に戻る。
- [0104] ステップ S 2において、命令キャッシュ処理が終了すると、ステップ S 3において、GPU 1 1は、使用するレジスタオペランドを指定する。
- [0105] ステップ S 4において、GPU 1 1は、キャッシュメモリ 1 3より供給されてきたデータからなる命令に対応する計算、演算、または比較等の処理を実行する。
- [0106] ステップ S 5において、GPU 1 1は、実行している命令がストア命令であるか否かを判定する。ステップ S 5において、例えば、実行している命令がストア命令である場合、処理は、ステップ S 6に進む。
- [0107] ステップ S 6において、GPU 1 1は、キャッシュメモリ 1 3に対してステップ S 1と同様にアドレスを指定してデータキャッシュ処理を指示する。これに応じてアドレス入力部 1 2は、入力されたアドレスにより、インデックスアドレス、CAMサブタグアドレス、およびSRAMサブタグアドレスをキャッシュメモリ 1 3に供給する。キャッシュメモリ 1 3は、このインデックスアドレ

ス、CAMサブタグアドレス、およびSRAMサブタグアドレスに基づいて、データキャッシュ（書き込み）処理を実行する。

[0108] [データキャッシュ（書き込み）処理]

ここで、図11のフローチャートを参照して、データキャッシュ（書き込み）処理について説明する。尚、図11のフローチャートにおいて、ステップS61、S62、S65、S66の処理については、図8のフローチャートを参照して説明したステップS21、S22、S24、S25の処理と同様であるので、その説明は省略するものとする。

[0109] すなわち、ステップS62において、例えば、キャッシュヒットであった場合、処理は、ステップS63に進む。

[0110] ステップS63において、比較検索部71は、キャッシュヒットしたものととして検索されたラインの情報（CAMサブタグアドレス、およびSRAMサブタグアドレスの情報）を書戻管理部75、および書込部76に供給する。そして、書込部76は、CPU11の書き込み指示に従って、検索されたCAMサブタグアドレス、およびSRAMサブタグアドレスで特定されるラインに対応付けて、演算結果のデータを格納部78に書き込む。

[0111] ステップS64において、書戻管理部75は、CAMサブタグアドレス、およびSRAMサブタグアドレスで特定されるラインの書き戻しフラグをオンの状態にして、書戻管理リスト75aを更新し、以降の処理において、そのラインが書き戻し対象であることを記録する。

[0112] 一方、ステップS62において、キャッシュヒットではない、すなわち、キャッシュミスであった場合、処理は、ステップS67に進む。

[0113] ステップS67において、比較検索部71は、キャッシュミスしたことを示す情報を書換部77に供給する。書換部77は、メインメモリ14にアクセスし、CAMサブタグアドレス、およびSRAMサブタグアドレスで特定されるタグアドレスのデータを、CPU11より供給されてくるデータで書き換える。

[0114] ステップS68において、書戻部73は、LRU管理部53および書戻管理部75に対して、それぞれLRU管理リスト53aおよび書戻管理リスト75aを

要求し、取得する。そして、書戻部 73 は、LRU 管理リスト 53 a において、最下位となるライン、すなわち、最も古いラインを検索し、さらに、その最も古いラインについて、書戻管理リスト 75 a で照合し、書戻フラグがオンの状態であるか否かを判定する。ステップ S 68 において、例えば、最下位となるライン、すなわち、最も古いデータであって置換されるラインの書戻フラグがオンであり、書き戻し対象となっている場合、処理は、ステップ S 69 に進む。

[0115] ステップ S 69 において、書戻部 73 は、メインメモリ 14 にアクセスし、最下位となるライン、すなわち、最も古いデータであって置換される対象となっているラインの書戻フラグがオンとなっているデータをメインメモリ 14 に書き戻す。

[0116] すなわち、この処理により、新たに GPU 11 より供給されてくるデータで書き換えられて、キャッシュメモリ 13 内から消滅する前に、データがメインメモリ 14 に書き戻されることになる。

[0117] 一方、ステップ S 68 において、最も古いデータであって置換されるラインの書戻フラグがオンではなく、オフとされており、書き戻し対象のラインではない場合、ステップ S 69 の処理は、スキップされる。

[0118] ステップ S 70 において、置換部 55 は、LRU 管理部 53 に対して LRU 管理リスト 53 a を要求し、取得する。そして、置換部 55 は、LRU 管理リスト 53 a において、最下位となるライン、すなわち、最も古いデータとタグアドレスとを検索し、GPU 11 より供給されてきたデータとタグアドレスとで置換して CAM 51 および SRAM 52 の格納部 78 の情報を更新する。

[0119] 以上の処理により、書き込みデータのデータキャッシュが管理されるので、アクセス頻度の高い書き込みデータのみがキャッシュメモリ 13 内で管理され、それ以外のデータについては、順次メインメモリ 14 に書き戻されていく。

[0120] ここで、図 7 のフローチャートの説明に戻る。

[0121] ステップ S 6 の処理により、データキャッシュ（書き込み）処理が完了す

ると、プログラム実行処理は終了する。

[0122] 一方、ステップS 5において、実行している命令がストア命令ではないと判定された場合、処理は、ステップS 7に進む。ステップS 7において、CPU 11は、実行している命令がロード命令であるか否かを判定する。ステップS 7において、例えば、実行している命令がロード命令である場合、処理は、ステップS 8に進む。

[0123] ステップS 8において、CPU 11は、キャッシュメモリ13に対してステップS 1と同様にアドレスを指定してデータキャッシュ処理を指示する。これに応じてアドレス入力部12は、入力されたアドレスにより、インデックスアドレス、CAMサブタグアドレス、およびSRAMサブタグアドレスをキャッシュメモリ13に供給する。キャッシュメモリ13は、このインデックスアドレス、CAMサブタグアドレス、およびSRAMサブタグアドレスに基づいて、データキャッシュ（読み出し）処理を実行する。

[0124] [データキャッシュ（読み出し）処理]

ここで、図12のフローチャートを参照して、データキャッシュ（読み出し）処理について説明する。尚、図12のフローチャートにおけるステップS 81乃至S 89の処理は、図11のフローチャートにおけるステップS 61、S 62、図8のフローチャートにおけるステップS 23乃至S 26、並びに、図11のフローチャートにおけるステップS 68乃至S 70の処理と同様であるので、その説明は省略する。

[0125] すなわち、図12のフローチャートにおいては、図11のフローチャートを参照して説明した処理における書き込みを対象とした処理が省略され、メインメモリ14からの読み出し処理のみとなる。

[0126] ここで、図7のフローチャートの説明に戻る。

[0127] ステップS 8において、データキャッシュ（読み出し）処理が終了すると、ステップS 9において、CPU 11は、図示せぬレジスタファイルに処理結果を書き込み、処理を終了する。

[0128] また、ステップS 7において、命令がロード命令ではない場合、ステップ

S 1 0において、GPU 1 1は、図示せぬレジスタファイルに処理結果を書き込み、さらに、図示せぬプログラムカウンタを更新して、処理を終了する。

[0129] 以上の一連の処理のうち、特に、タグアドレスの一部をSRAMサブタグアドレスとして設定し、SRAM 5 2により管理させるようにし、さらに、最近アクセスされた第1世代のデータを優先的に検索させるようにした。このため、CAMサブタグアドレスにより検索されたラインであって、第1世代のラインのみを、SRAMサブタグアドレスで検索するようにしたので、対応付けて格納されているデータのうち、SRAM 5 2の検索に伴って同時に読み出されるデータ数を絞り込むことが可能となり、結果として、SRAM 5 2の消費電力として大きな出力動作を制限させることができ、消費電力を低減させることが可能となる。また、最初の処理で、アクセスされる頻度が高い第1世代のラインを優先的に検索させることにより、キャッシュヒット率の低減を抑制することが可能となる。

[0130] さらに、1のCAMサブタグアドレスに対して、複数のSRAMサブタグアドレスであっても、重複してデータを格納させることを認めて、LRU法に則って、アクセスされる可能性の高いデータを格納できるようにしたので、1のCAMサブタグアドレスに対して、1のSRAMサブタグアドレスに絞られることによりキャッシュヒット率が低減してしまうことを抑制することが可能となる。

[0131] このように、1のCAMサブタグアドレスに重複して複数のSRAMサブタグアドレスとなるデータの格納を認めて、かつ、最近アクセスされた第1世代のデータを優先的に検索させることにより、キャッシュヒット率を低減させることなく、消費電力を低減させることが可能となる。

[0132] [ベンチマークプログラムによるミス率の比較]

次に、図13を参照して、本発明を適用したキャッシュメモリを用いた場合のミス率と、その平均値について説明する。尚、図13においては、上段に命令キャッシュによる結果が示されており、下段にデータキャッシュによる結果が示されている。また、図13における横軸に対応する番号1乃至21のベンチマークプログラムは、1から順に、bitcount, qsort, susan\_e, s

usan\_c, susan\_s, lame, tiff2bw, tiff2rgba, dijkstra, patricia, ispell, stringsearch, blow\_d, blow\_e, rijndael\_d, rijndael\_e, sha, crc, rawaudio(adpcm), rawaudio(adpcm), untoast(gsm)であり、番号22は全体の平均値 (average) である。

- [0133] また、各プログラム毎のミス率は、左から従来型の4ウェイのSRAM、8ウェイのSRAM、32ウェイのCAM、タグアドレスを24ビットとしたときのCAMサブタグアドレスのビット数 $s$ を、 $s=5, 8, 9, 11$ としたLPHAC法、および本発明の例 ( $s=4$ ) (黒で表記されている) である。
- [0134] 図13の上段で示されるように命令キャッシュのミス率は、番号22で示される平均値から、LPHAC法は従来型のCAMよりも高い値であることがわかる。特に、CAMサブタグアドレスのビット数 $s=5$ の場合、ミス率が際立って高い。これは、CAMサブタグアドレスの比較検索で検索できる (ヒットする) が、SRAMタグアドレスの比較検索で検索できない (ミスする) 場合が多いためと思われる。また、CAMサブタグアドレスのビット数 $s=8, 9$ の場合、多くのベンチマークプログラムでのミス率は、従来型のCAMに近い値となるがrijndael\_dやqsortでは、ミス率は大きい。CAMサブタグアドレスのビット数 $s=11$ では全てのベンチマークプログラムにおいて従来型のCAMとほぼ同等となる。一方、本発明のキャッシュメモリにおいては、CAMサブタグアドレスのビット数 $s=4$ としても、ミス率には影響を与えないことがわかる。
- [0135] すなわち、平均値をみると、LPHAC法ではCAMサブタグアドレスのビット数 $s=5, 8, 9, 11$ のとき、それぞれ0.75%, 0.43%, 0.43%, 0.42%となるのに対し、本発明のキャッシュメモリにおいては、CAMサブタグアドレスのビット数 $s=4$ において、従来型のCAMと同じ0.42%である。
- [0136] また、図13の下段で示されるように、データキャッシュについても命令キャッシュと同様の傾向が見られる。すなわち、LPHAC法では、CAMサブタグアドレスのビット数 $s=8$ の場合、rijndael\_dやrijndael\_eはミス率が高く、patriciaやcrcではCAMサブタグアドレスのビット数 $s=9$ でもミス率が高い。
- [0137] 一方、本発明のキャッシュメモリの場合、命令キャッシュと同様、従来型



のCAMと同様のミス率となっており、CAMサブタグアドレスのビット数 $s=4$ でも影響を受けない。番号22で示される平均値をみると、LPHAC法が $s=5, 8, 9, 11$ のとき、それぞれ1.95%, 0.91%, 0.80%, 0.67%となるのに対し、本発明のキャッシュメモリにおいてはCAMサブタグアドレスのビット数 $s=4$ において従来型のCAMと同じ0.66%である。

[0138] このように、本発明のキャッシュメモリにおいてはCAMサブタグアドレスとの部分的な比較検索において、ヒットするキャッシュラインを1つに絞らないため、常にLRU方式を採用できるので、ヒット率の低下を抑制することができる。また、LPHAC法のようにCAMサブタグアドレスのビット数 $s$ の大きさによって、特定のベンチマークのミス率が大きな影響を受けることはない。

[0139] なお、従来型のCAMと本発明のキャッシュメモリは、4, 8ウェイの従来型のSRAMと比較し、命令キャッシュでそれぞれ10.7%, 8.6%となり、データキャッシュにおいてそれぞれ11.6%, 5.0%平均ミス率が小さい。

[0140] [消費電力量の比較]

次に、図14乃至図16を参照して、消費電力量の評価について説明する。本発明の消費電力量の評価は、上述したベンチマークプログラムを実行したときの平均値に基づいて行うものとし、以下の式(1)で表される消費電力量 $E$ で表現されるものとする。

$$E = E_{tag} + E_{data} \quad \dots (1)$$

[0141] ここで、 $E_{tag}$ は、以下の式(2)で表される。

$$E_{tag} = E_{CAM} + E_{SRAM} \quad \dots (2)$$

[0142] ここで、式(2)における $E_{CAM}$ および $E_{SRAM}$ は以下の式(3)、式(4)で表される。

$$E_{CAM} = \alpha \times w \times s \times NC \quad \dots (3)$$

$$E_{SRAM} = (t - s) \times (NS1 + NS2) \quad \dots (4)$$

[0143] ここで、 $\alpha$ はCAM51の1ビット検出に要する消費電力量を、 $w$ はウェイ数を、 $s$ はCAMサブタグアドレスのビット数を、 $t$ はタグアドレスのビット数を、 $NC$ はCAMへのアクセス数を、 $NS1$ は第1世代として管理されているSRAMのデータ

へのアクセス数を、NS2は第2世代として管理されているSRAMのデータへのアクセス数をそれぞれ示している。

[0144] さらに、Edataは、以下の式(5)で表される。

$$Edata = \beta \times l \times (NS1 + NS2) \quad \dots (5)$$

[0145] ここで、 $\beta$ はSRAMのデータ1ビットの出力に要する消費電力量を、 $l$ は1ライン当たりのデータビット長をそれぞれ表している。

[0146] まず、図14を参照して、タグアドレスの検出に要する消費電力量Etagおよびデータ出力に要する消費電力量Edataと、それぞれのCAMサブタグアドレスのビット長 $s$ との関係について説明する。

[0147] 図14においては、左部が命令キャッシュ処理であり、右部がデータキャッシュ処理である。また、いずれも実線および1点鎖線がタグアドレスの検出に要する消費電力量Etagを、点線および2点鎖線がデータ出力に要する消費電力量Edataをそれぞれ表している。タグアドレスの検出に要する消費電力量Etagは、SRAMタグアドレス1ビット検出に要する電力量を単位とする。データ出力に要する消費電力量Edataは、SRAMデータ1ビットの出力に要する電力量を単位とする。尚、1ラインは256ビットとしている。また、CAM51の1ビット検出に要する消費電力量 $\alpha=5$ とし、CAMサブタグアドレスのビット数 $s=0$ 乃至8、第1世代ウェイ数 $r=0, 4$ について比較している。比較のために、CAMサブタグアドレスのビット数 $s=8$ におけるLPHAC法のタグアドレスの検出に要する消費電力量Etag(×印)およびデータ出力に要する消費電力量Edata(十字マーク)も示している。

[0148] LPHAC法では、CAMサブタグアドレスのビット数 $s$ を小さくするとミス率が大きくなるため、CAMサブタグアドレスのビット数 $s \geq 8$ でなければならないのに対し、本発明のキャッシュメモリでは、CAMサブタグアドレスのビット数 $s < 8$ とすることが可能である。

[0149] タグアドレスの検出に要する消費電力量Etagは、CAMサブタグアドレスのビット数 $s$ を小さくすると線形に減少する。CAMサブタグアドレスのビット数 $s=8$ に対して、CAMサブタグアドレスのビット数 $s=4$ では、タグアドレス検出に要

する消費電力量はほぼ1/2となる。

- [0150] データ出力に要する消費電力量 $E_{data}$ は、CAMサブタグアドレスのビット数 $s$ を小さくするとある値以下では急激に増加する。急激に増加するときのCAMサブタグアドレスのビット数 $s$ の値は、第1世代SRAMのウェイ数 $r$ によって異なる。ウェイ数 $r=0$ のときはCAMサブタグアドレスのビット数 $s$ は $s \leq 6$ で急激に増加するのに対し、ウェイ数 $r=4$ とすれば、CAMサブタグアドレスのビット数 $s=4$ まであまり変化しない。LPHAC法は、タグアドレスの検出に要する消費電力量 $E_{tag}$ およびデータ出力に要する消費電力量 $E_{data}$ 共に本発明のキャッシュメモリと同程度の値をとるが、CAMサブタグアドレスのビット数 $s$ を小さくすることが難しい。
- [0151] 本発明のキャッシュメモリは、ウェイ数 $r=0$ のときは世代を分けていないため、CAMサブタグアドレスのビット数 $s$ を小さくすることが難しいが、図15の波形図a, bで示されるように、ウェイ数 $r=4$ ならば、世代を分けた効果が見られる。
- [0152] ここで、図15の波形図a, bは、それぞれ命令キャッシュとデータキャッシュについて、CAMサブタグアドレスのビット数 $s$ を $s=4$ とし、ウェイ数 $r$ を $r=0, 1, 2, 4, 8, 16, 32$ と変えたとき、点線で示されるタグアドレスの検出に要する消費電力量 $E_{tag}$ および実線で示されるデータ出力に要する消費電力量 $E_{data}$ のウェイ数 $r$ との関係が示されている。
- [0153] すなわち、命令キャッシュ、およびデータキャッシュのどちらの場合も、ウェイ数 $r=2$ 乃至4の近傍で消費電力量が最少となる。ウェイ数 $r$ の値により消費電力量に最小値が現れるのは、ウェイ数 $r$ が小さいと第2世代に設定されたデータとの比較検索回数が増加し、ウェイ数 $r$ が大きいと第1世代に設定されたデータで複数比較検索の頻度が上がるためである。
- [0154] ところで、SRAMにおいて、デコーダや比較器のような論理回路の電力が、ワードライン駆動時に流れるビットラインキャパシタンスの充放電に要する電力と比較して十分小さいという原理が知られている (Itoh, K., Sasaki, K. and Nakagome, Y.: Trends in low-power RAM circuit technologies, Pro

ceedings of the IEEE, Vol. 83, No. 4, pp. 524-543 (1995). 参照)。そこで、本発明のキャッシュメモリの消費電力量 $E$ について考えるに当たっては、その原理に則り、SRAMサブタグアドレス1ビットの検出に要する電力量は、SRAMデータ1ビット出力に要する電力量に等しいとする近似を用いる。すなわち、式(5)においてSRAMのデータ1ビットの出力に要する消費電力量 $\beta=1$ とする。

[0155] 図15の波形図c, dは、それぞれ命令キャッシュとデータキャッシュについて、CAMサブタグアドレスのビット数 $s$ を変えたときの命令キャッシュとデータキャッシュの消費電力量 $E$ を示しており、点線はウェイ数 $r$ が $r=0$ のときであり、実線はウェイ数 $r$ が $r=4$ のときの波形である。また、図15の波形図e, dは、それぞれウェイ数 $r=4$ とし、第1世代のウェイ数 $r$ を変えたときの命令キャッシュとデータキャッシュの消費電力量 $E$ を示している。これらの波形図から、本発明のキャッシュメモリにおいては、ウェイ数 $r=4$ の場合、CAMサブタグアドレスのビット数 $s$ が $s=2$ のとき、消費電力量が最小となる。

[0156] 次に、図16を参照して、従来のSRAM、CAM、およびLPHAC法、並びに本発明のキャッシュメモリの消費電力量を比較する。比較対象は、図16の左から、4, 8ウェイの従来型(Traditional) SRAM、32ウェイの従来型CAM、CAMサブタグアドレスのビット数 $s$ が $s=8$ のLPHAC法、並びに、ウェイ数 $r$ が $r=4$ であつて、かつ、CAMサブタグアドレスのビット数 $s$ が $s=2$ の本発明のキャッシュメモリ(proposed)である。なお、いずれの場合もラインサイズが32bytesであり、キャッシュ容量が8KBである。また、従来型のCAMとLPHAC法、および本発明のキャッシュメモリについては、図16の左からCAMの消費電力量 $\alpha=5, 7, 10$ と値を変えて比較を行った命令キャッシュ(図16上段)、およびデータキャッシュ(図16下段)の消費電力量の比較結果が示されている。

[0157] 命令キャッシュ、およびデータキャッシュの消費電力量は、共に従来型のCAMが、いずれの場合と比較しても消費電力量が一番高いことが示されている。LPHAC法は従来型のCAMに比べると消費電力量を抑えることができている。しかしながら、従来型のSRAMと比べると、CAMの消費電力量 $\alpha \leq 7$ のとき、ウェイ数 $r$ が $r=8$ のSRAMの消費電力量より小さくなるが、ウェイ数 $r$ が $r=4$ のSRAM

の消費電力量と比べるといずれも大きくなる。これに対して、本発明のキャッシュメモリでは $\alpha$ によらず従来型CAMとLPHAC法のいずれよりも小さくなる。CAMの消費電力量 $\alpha=10$ において、従来型のウェイ数 $r$ が $r=4$ のSRAMと同程度の消費電力量となっており、CAMの消費電力量 $\alpha$ によらず、略同等の消費電力量になっている。

[0158] このように、本発明のキャッシュメモリはキャッシュミス率の低下を抑制しつつ、消費電力を低減させることが可能となっている。

[0159] 尚、本明細書において、プログラムを記述するステップは、記載された順序に沿って時系列的に行われる処理は、もちろん、必ずしも時系列的に処理されなくとも、並列的あるいは個別に実行される処理を含むものである。

### 符号の説明

[0160] 1 1 CPU, 1 2 アドレス入力部, 1 3 キャッシュメモリ, 1 4 メインメモリ, 2 1 インデックスアドレス抽出部, 2 2 タグアドレス抽出部, 2 2 a CAMサブタグアドレス抽出部, 2 2 b SRAMサブタグアドレス抽出部, 4 1 デコード部, 4 2 サブバンク, 5 1 CAM, 5 2 SRAM, 5 3 LRU管理部, 5 3 a LRU管理リスト, 5 4 読出部, 5 5 置換部, 7 1 比較検索部, 7 2 出力部, 7 3 書戻部, 7 4 世代管理部, 7 4 a 世代管理リスト, 7 5 書戻管理部, 7 6 書込部, 7 7 書換部, 7 8 格納部, 8 1 世代判定部

## 請求の範囲

[請求項1] 連想度  $n$  ( $n$ は自然数) のCAM (Content Addressable Memory) とSRAM (Static Random Access Memory) とを有し、接続されるCPU (Central Processing Unit) より第1サブタグアドレスおよび第2サブタグアドレスとからなるタグアドレスを指定され、対応するデータを格納または読み出すキャッシュメモリであって、

前記データを、読み出し要求された時期に応じて、少なくとも、最近読み出しが要求された第1世代と、前記最近ではない時期に読み出しが要求された第2世代とに分類し、

前記CAMにて、前記第1サブタグアドレスを管理し、

前記SRAMにて、前記第2サブタグアドレスを管理し、

同一の前記第1サブタグアドレスに対し、複数の前記第2サブタグアドレスを対応付けることを容認する

ことを特徴とするキャッシュメモリ。

[請求項2] 連想度  $n$  ( $n$ は自然数) のCAM (Content Addressable Memory) とSRAM (Static Random Access Memory) とを有し、接続されるCPU (Central Processing Unit) より第1サブタグアドレスおよび第2サブタグアドレスとからなるタグアドレスを指定され、対応するデータを格納または読み出すキャッシュメモリであって、

前記データを、読み出し要求された時期に応じて、少なくとも、最近読み出しが要求された第1世代と、前記最近ではない時期に読み出しが要求された第2世代とに分類する世代管理手段と、

前記CAMにて、前記第1サブタグアドレスを管理し、前記CPUより指定される前記タグアドレスの第1サブタグアドレスとの比較により、対応する第1サブタグアドレスを検索する第1検索手段と、

前記SRAMにて、前記第2サブタグアドレスを管理し、前記第1検索手段により検索された第1サブタグアドレスを含み、かつ、前記第1世代のデータに対応する前記タグアドレスの第2サブタグアドレスと

の比較により、対応する第2サブタグアドレスを検索する第2検索手段と、

前記第2検索手段により検索された、前記第2サブタグアドレスに対応付けて記憶されている前記データを出力する出力手段とを含み、

同一の前記第1サブタグアドレスに対し、複数の前記第2サブタグアドレスを対応付けることを容認する

を含むことを特徴とするキャッシュメモリ。

[請求項3]

前記第2検索手段は、前記第1世代のデータに対応する前記第2サブタグアドレスの第2サブタグアドレスとの比較により、対応する第2サブタグアドレスを検索できないとき、前記第2世代のデータに対応する前記第2サブタグアドレスの第2サブタグアドレスとの比較により、対応する第2サブタグアドレスを検索する

ことを特徴とする請求項2に記載のキャッシュメモリ。

[請求項4]

前記第1検索手段による前記第1サブタグアドレスの比較によりデータが検索されないとき、または前記第2検索手段により前記第2サブタグアドレスの比較により第2世代のデータが検索されないとき、接続するメインメモリより前記読み出しが要求されたタグアドレスに対応するデータを読み出し、前記世代管理手段にて管理される、前記第2世代のデータのうち、最も古く読み出されたデータを、前記メインメモリより読み出されたデータで置換し、併せて前記第1サブタグアドレス、および第2サブタグアドレスを置換する置換手段をさらに含む

ことを特徴とする請求項2に記載のキャッシュメモリ。

[請求項5]

前記置換手段は、一の前記第1サブタグアドレスに対し、複数の前記第2サブタグアドレスが対応付けられて置換されることを許容する

ことを特徴とする請求項4に記載のキャッシュメモリ。

[請求項6]

前記世代管理手段は、前記第2のサブタグアドレスにより管理される各データが読み出された時期に基づいて、前記データが第1世代

であるか、または、前記第2世代であるかを世代管理リストにより管理し、前記置換手段により前記メインメモリより読み出されたデータで、前記最も古く読み出されたデータが置換されるとき、前記第2サブタグアドレスにより管理されるデータが読み出された時期が、最も古く読み出されたデータを削除し、前記メインメモリより読み出されたデータを、最近読み出された時期として世代管理リストを更新することを特徴とする請求項4に記載のキャッシュメモリ。

[請求項7] 前記世代管理手段は、前記世代管理リストに基づいて、前記第2サブタグアドレスにより管理される各データについて、最近読み出されたデータを第1世代のデータとし、最近ではない時期に読み出されたデータを第2世代のデータとして、世代管理することを特徴とする請求項6に記載のキャッシュメモリ。

[請求項8] 前記連想度 $n$ が連想度32であって、前記タグアドレスが24ビットの場合、第1サブタグアドレスが2ビットであり、第2サブタグアドレスが22ビットであることを特徴とする請求項2に記載のキャッシュメモリ。

[請求項9] 前記連想度 $n$ が連想度32の前記第1サブタグアドレスで管理されるデータのうち、前記第1世代のデータは、読み出された時期が最近に近い順に上位2位までのデータであることを特徴とする請求項8に記載のキャッシュメモリ。

[請求項10] 複数の請求項2乃至9のいずれかに記載のキャッシュメモリと、前記CPUと、前記CPUより前記タグアドレスで指定されるデータの読み出し、または書き込みが要求された場合、前記タグアドレスより前記第1のサブタグアドレスを抽出する第1のサブタグアドレス抽出手段と、前記CPUより前記タグアドレスで指定されるデータの読み出し、または書き込みが要求された場合、前記タグアドレスより前記第2のサブタグアドレスを抽出する第2のサブタグアドレス抽出手段とを含む



ことを特徴とするプロセッサ。

[請求項11]

連想度  $n$  ( $n$  は自然数) のCAM (Content Addressable Memory) とSRAM (Static Random Access Memory) とを有し、接続されるCPU (Central Processing Unit) より第1サブタグアドレスおよび第2サブタグアドレスとからなるタグアドレスを指定され、対応するデータを格納または読み出すキャッシュメモリの制御方法であって、

前記データを、読み出し要求された時期に応じて、少なくとも、最近読み出しが要求された第1世代と、前記最近ではない時期に読み出しが要求された第2世代とに分類する世代管理ステップと、

前記CAMにて、前記第1サブタグアドレスを管理し、前記CPUより指定される前記タグアドレスの第1サブタグアドレスとの比較により、対応する第1サブタグアドレスを検索する第1検索ステップと、

前記SRAMにて、前記第2サブタグアドレスを管理し、前記第1検索ステップの処理により検索された第1サブタグアドレスを含み、かつ、前記第1世代のデータに対応する前記タグアドレスの第2サブタグアドレスとの比較により、対応する第2サブタグアドレスを検索する第2検索ステップと、

前記第2検索ステップの処理により検索された、前記第2サブタグアドレスに対応付けて記憶されている前記データを出力する出力ステップとを含み、

同一の前記第1サブタグアドレスに対し、複数の前記第2サブタグアドレスを対応付けることを容認する

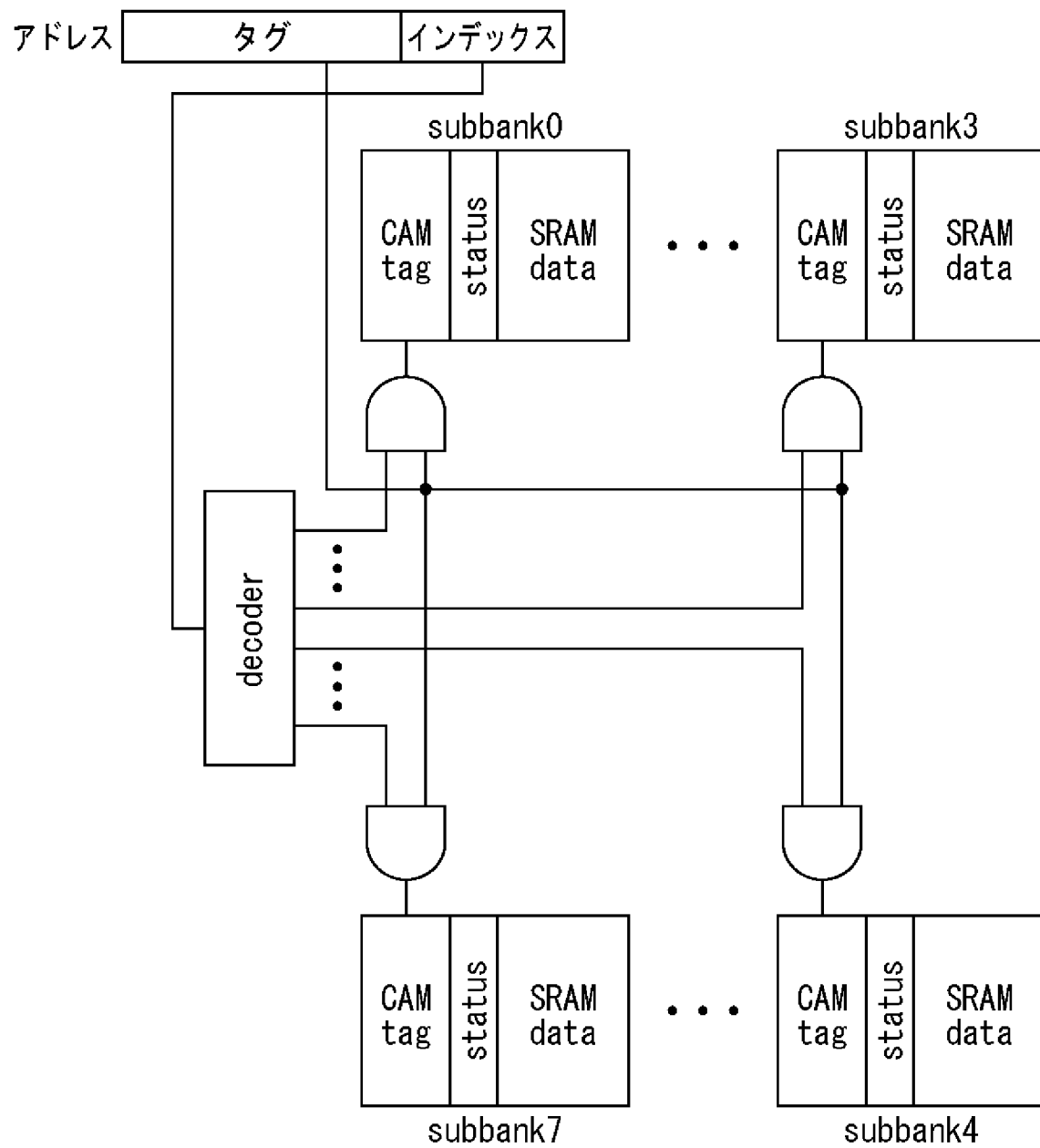
ことを特徴とする

を含むキャッシュメモリの制御方法。



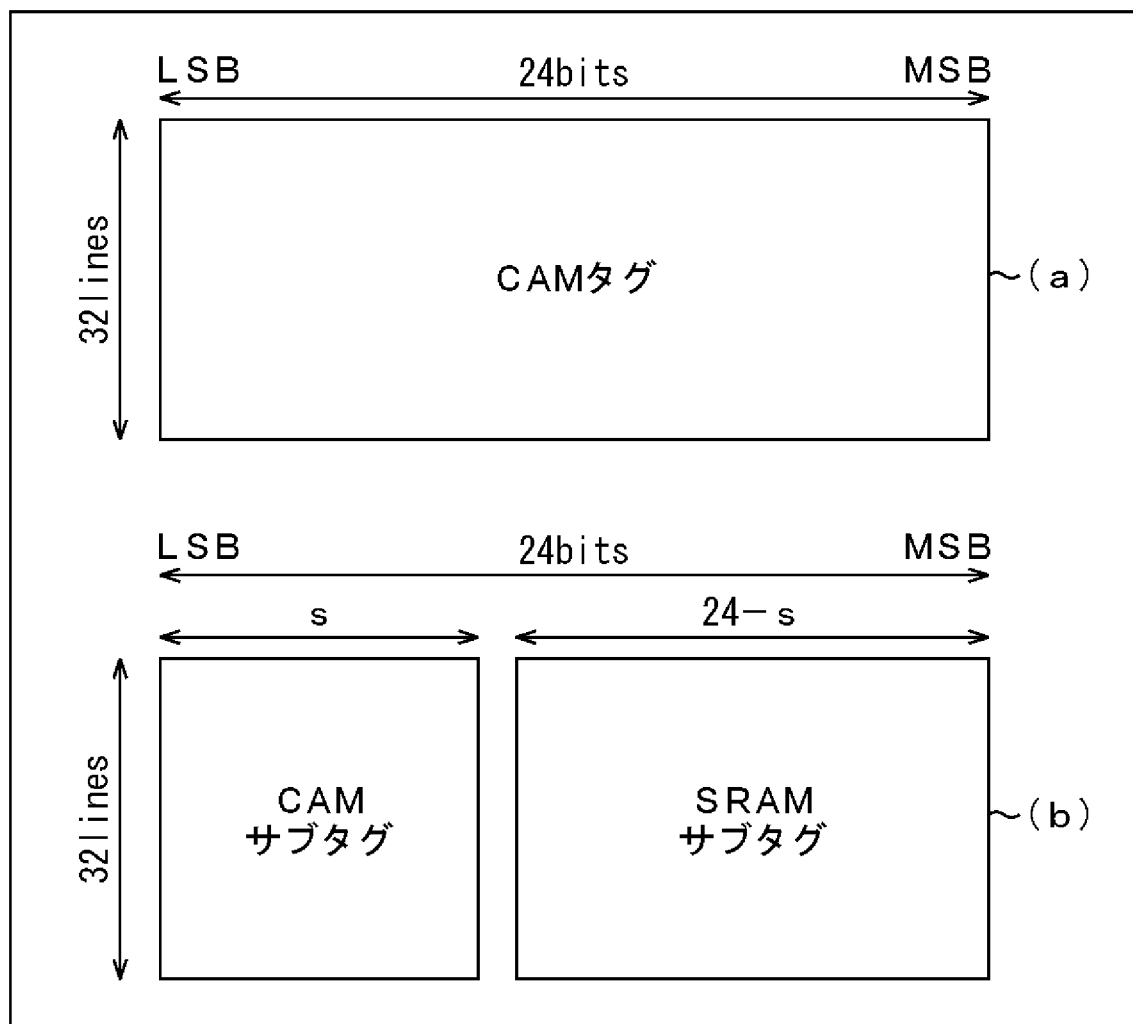
[図2]

図2



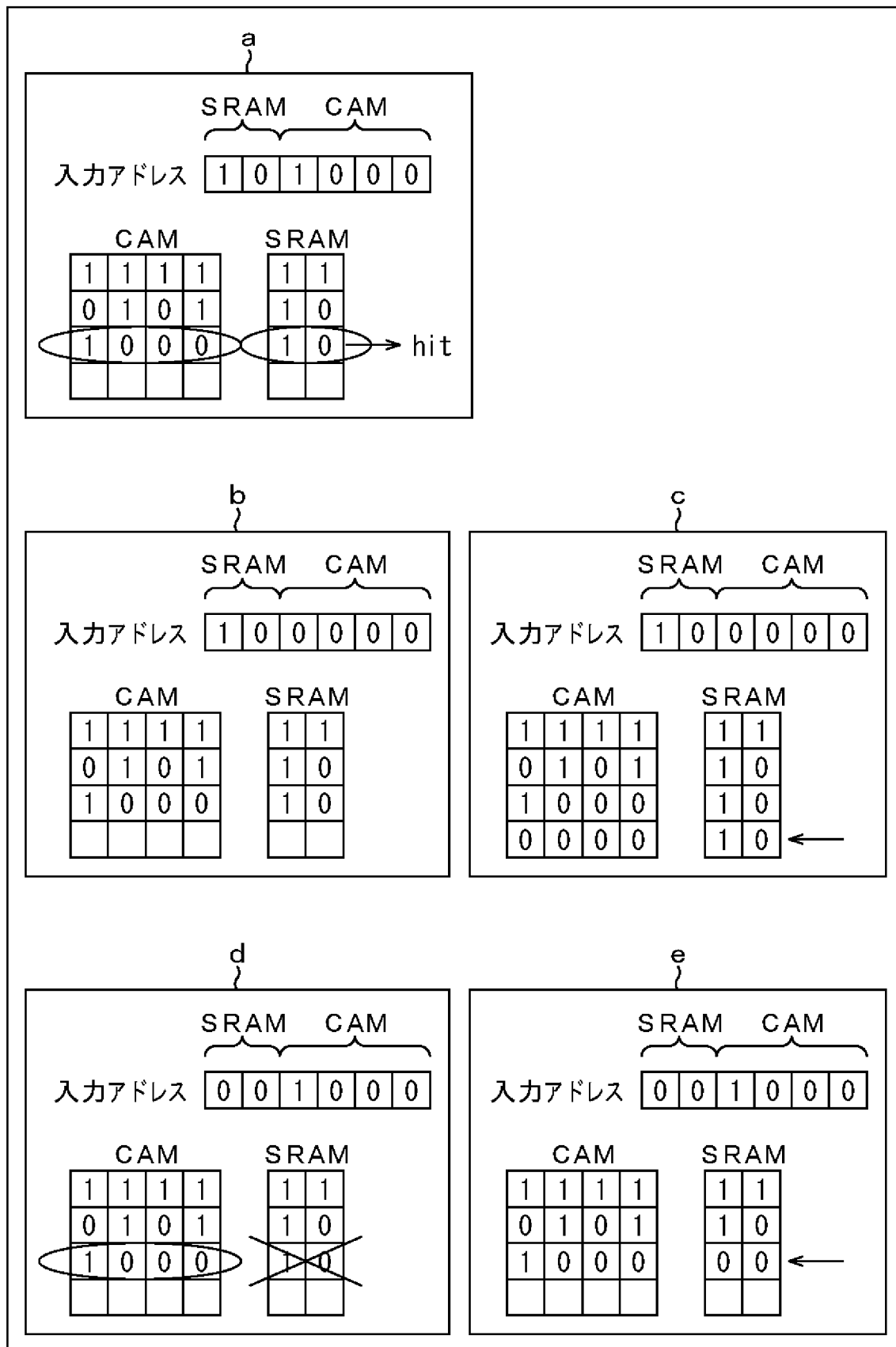
[図3]

図3

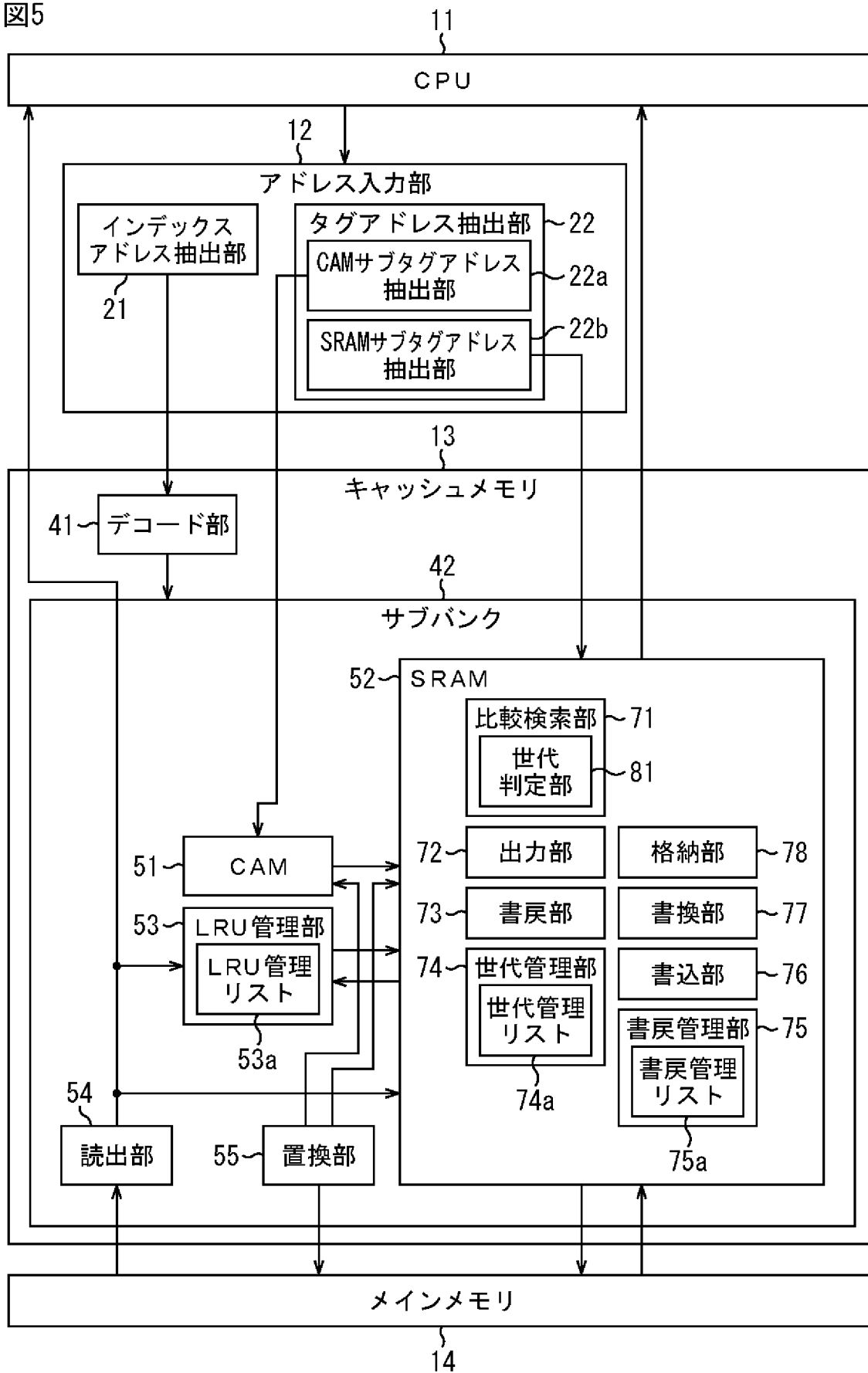


[図4]

図4

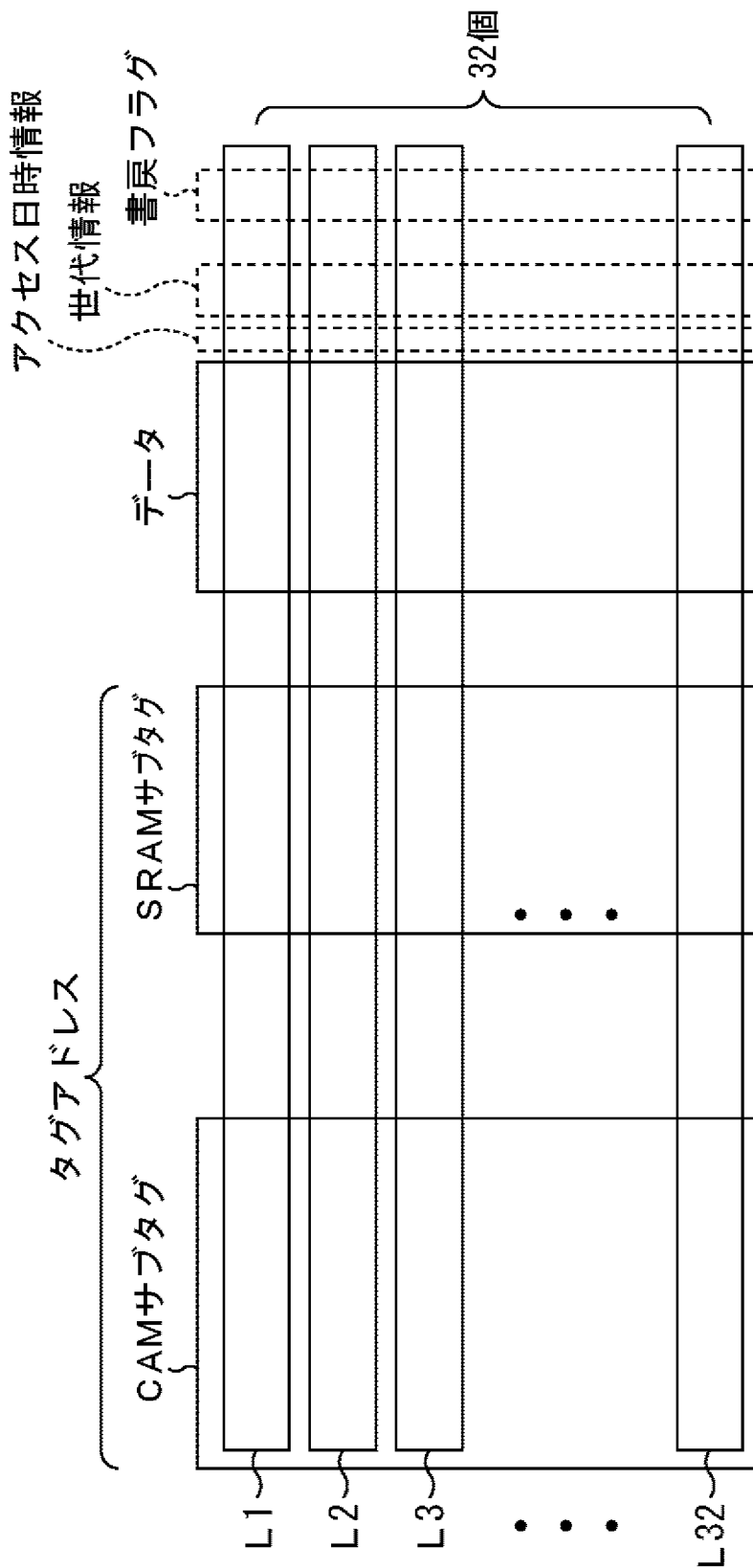


[図5]  
図5



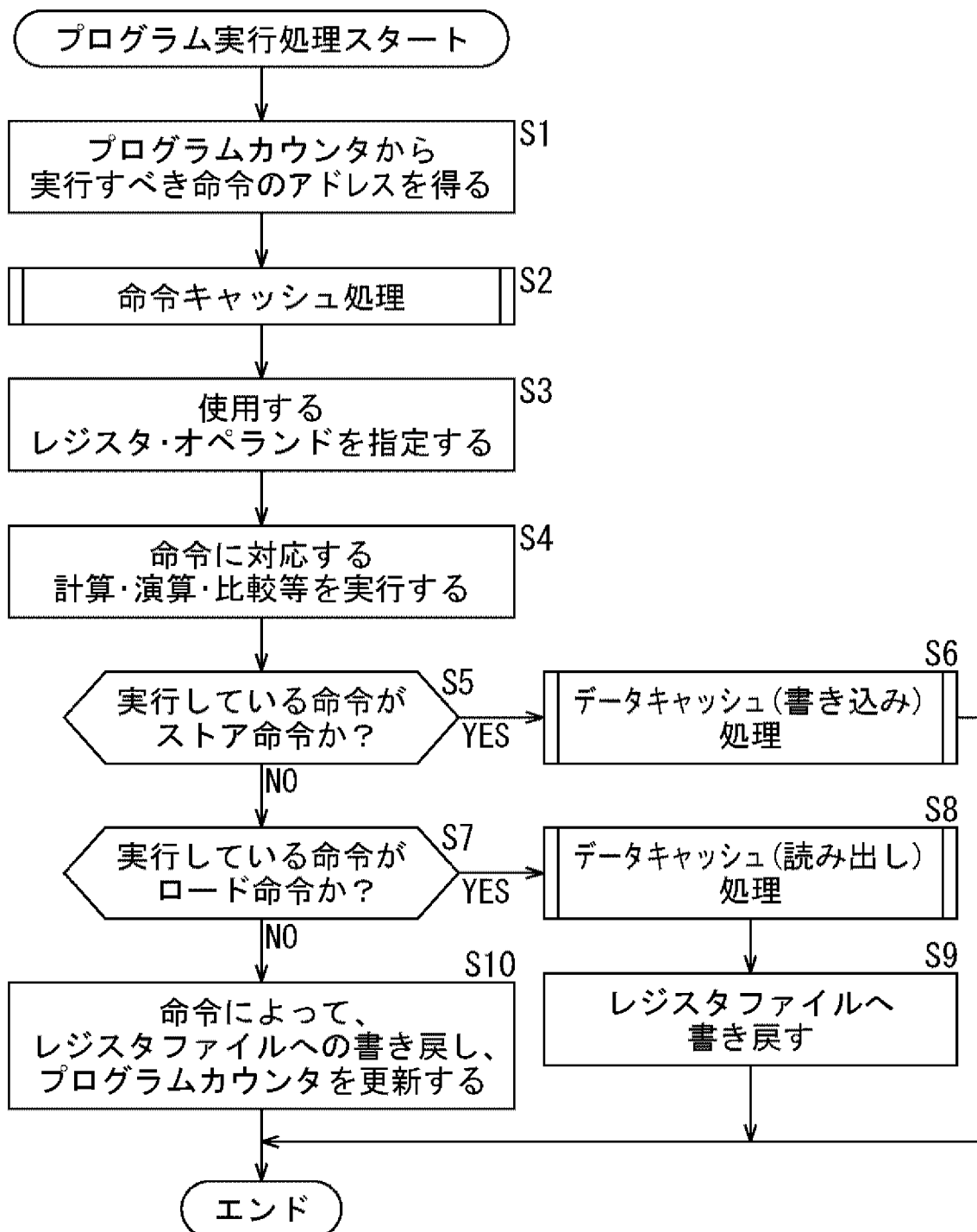
[図6]

図6



[図7]

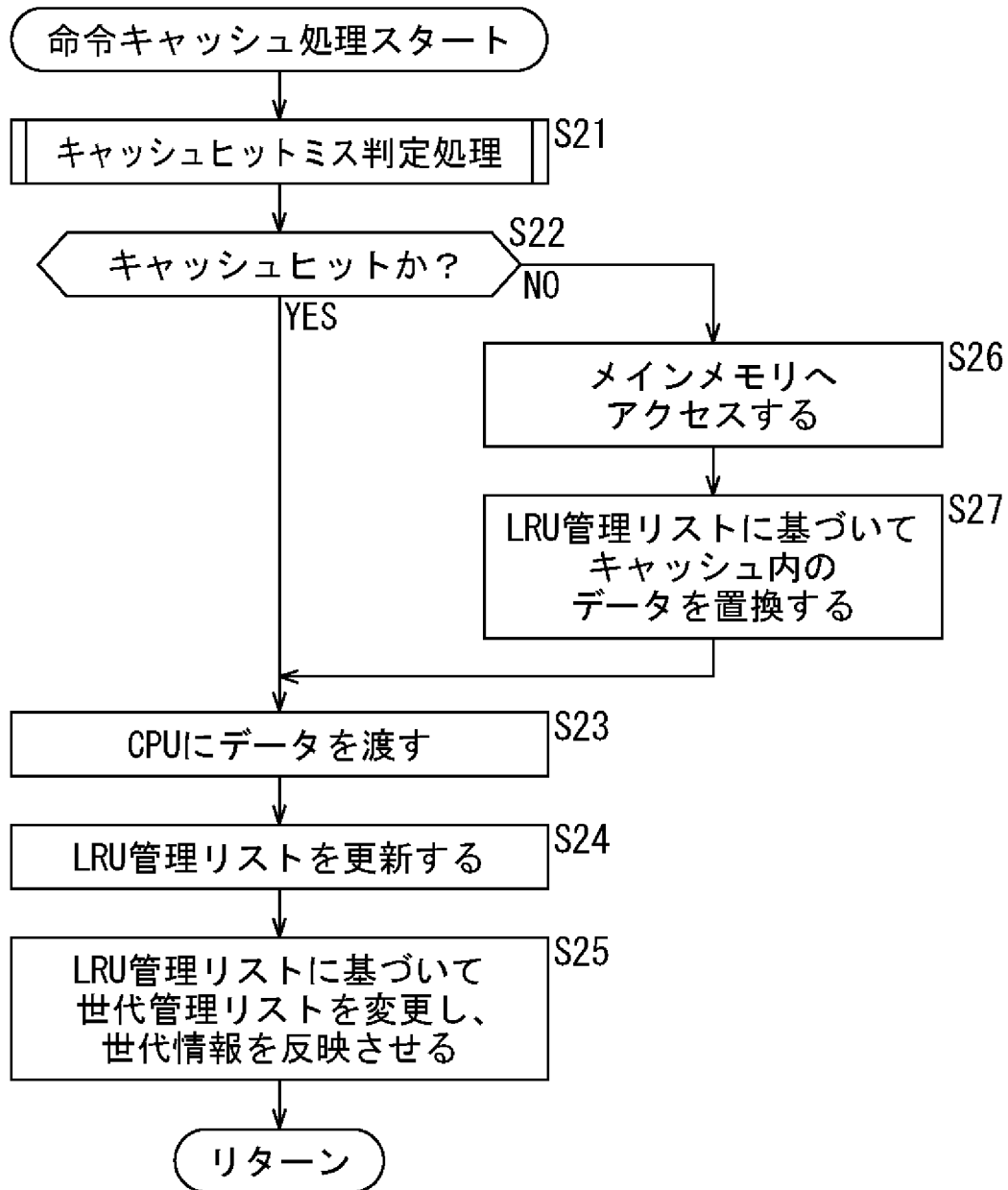
図7





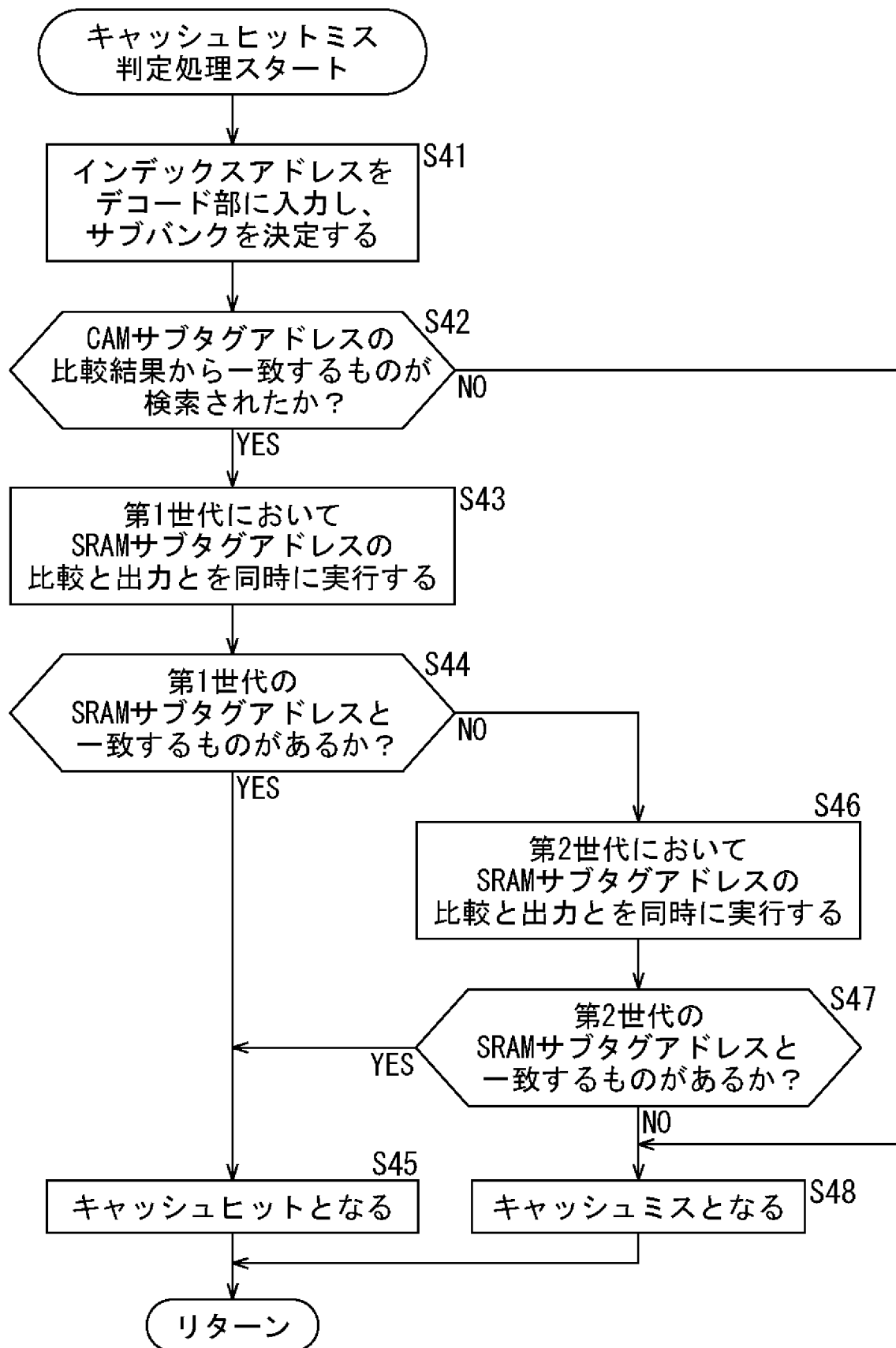
[図8]

図8

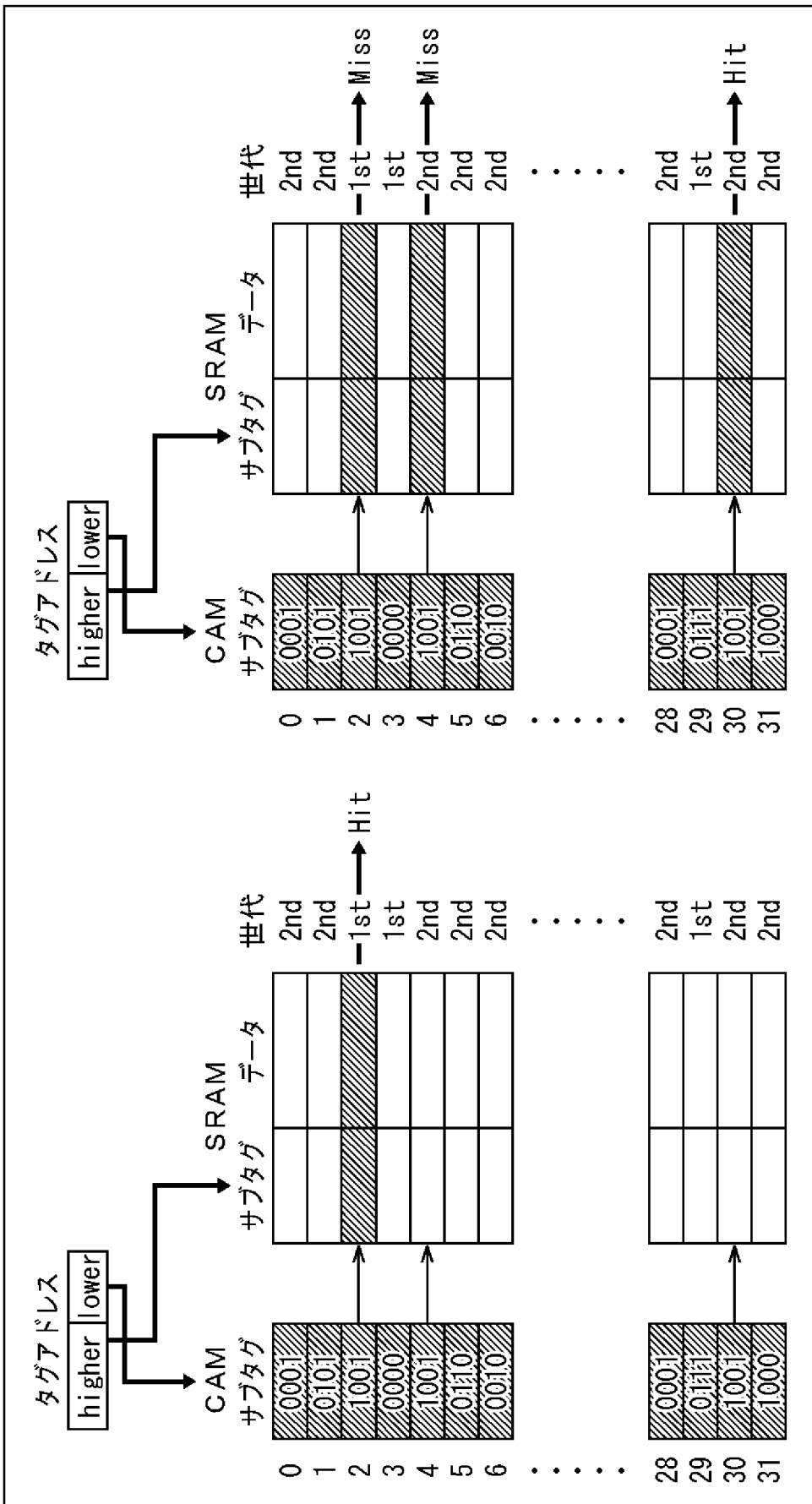


[図9]

図9

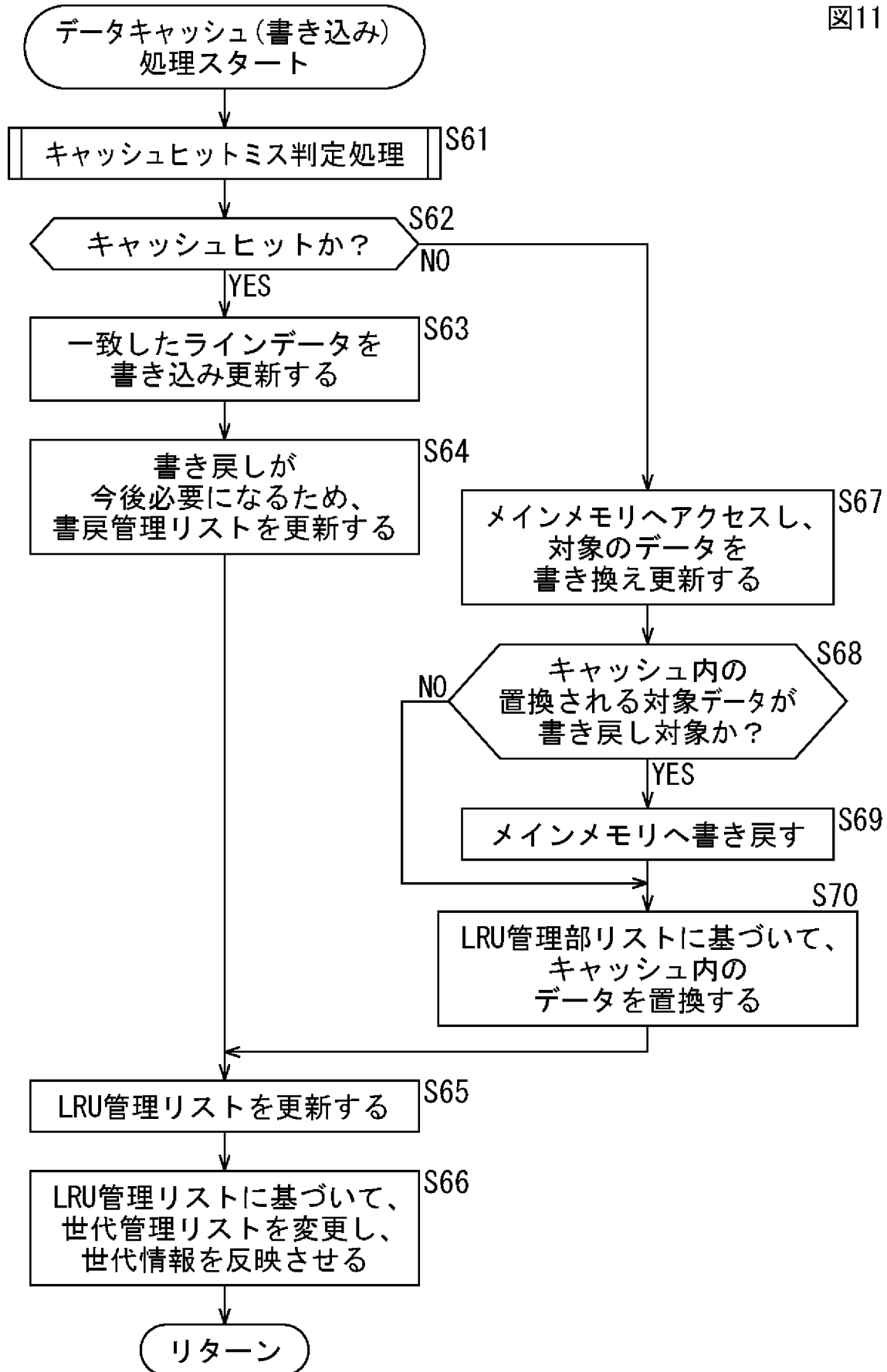


[図10]  
図10



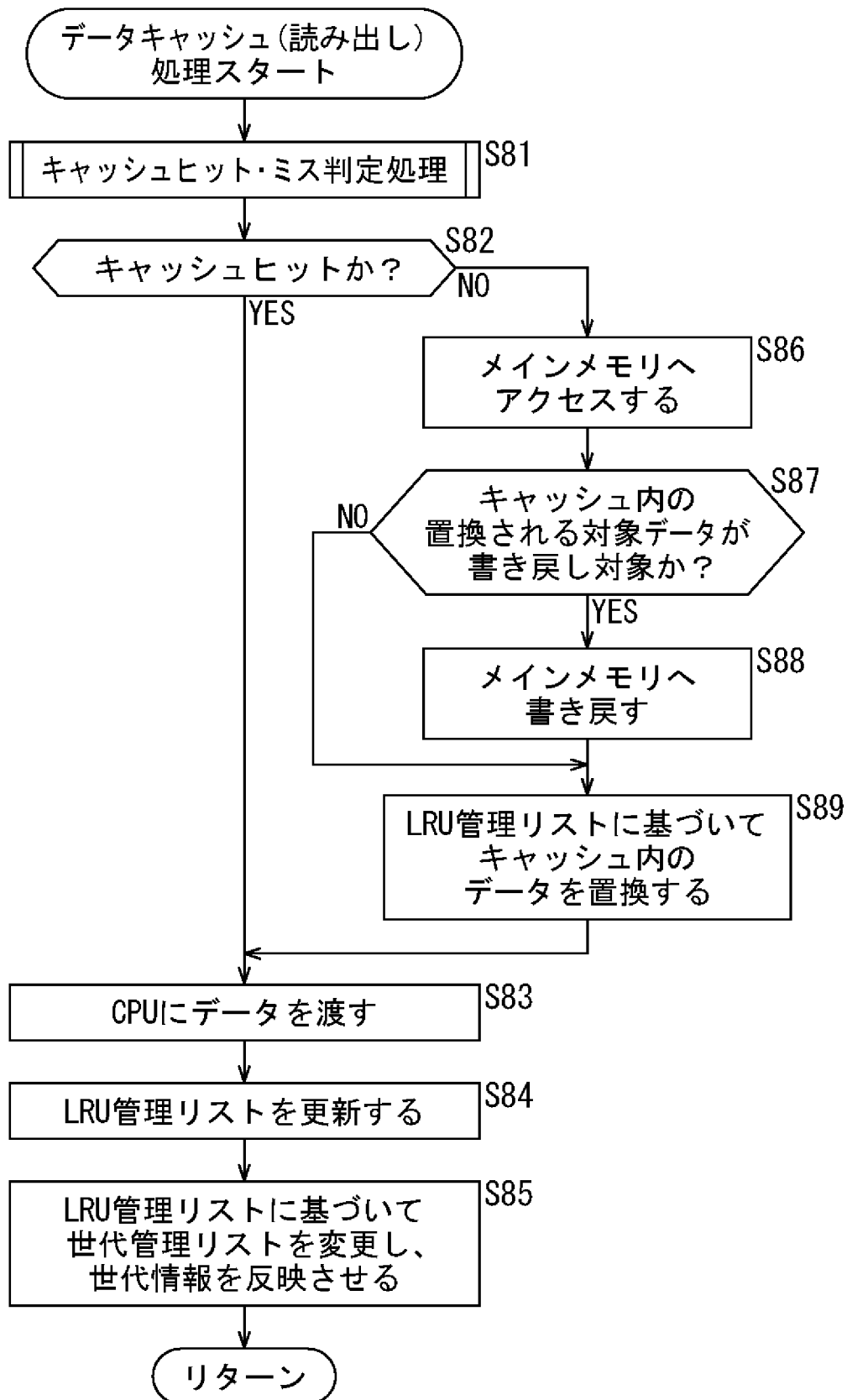
[図11]

図11



[図12]

図12

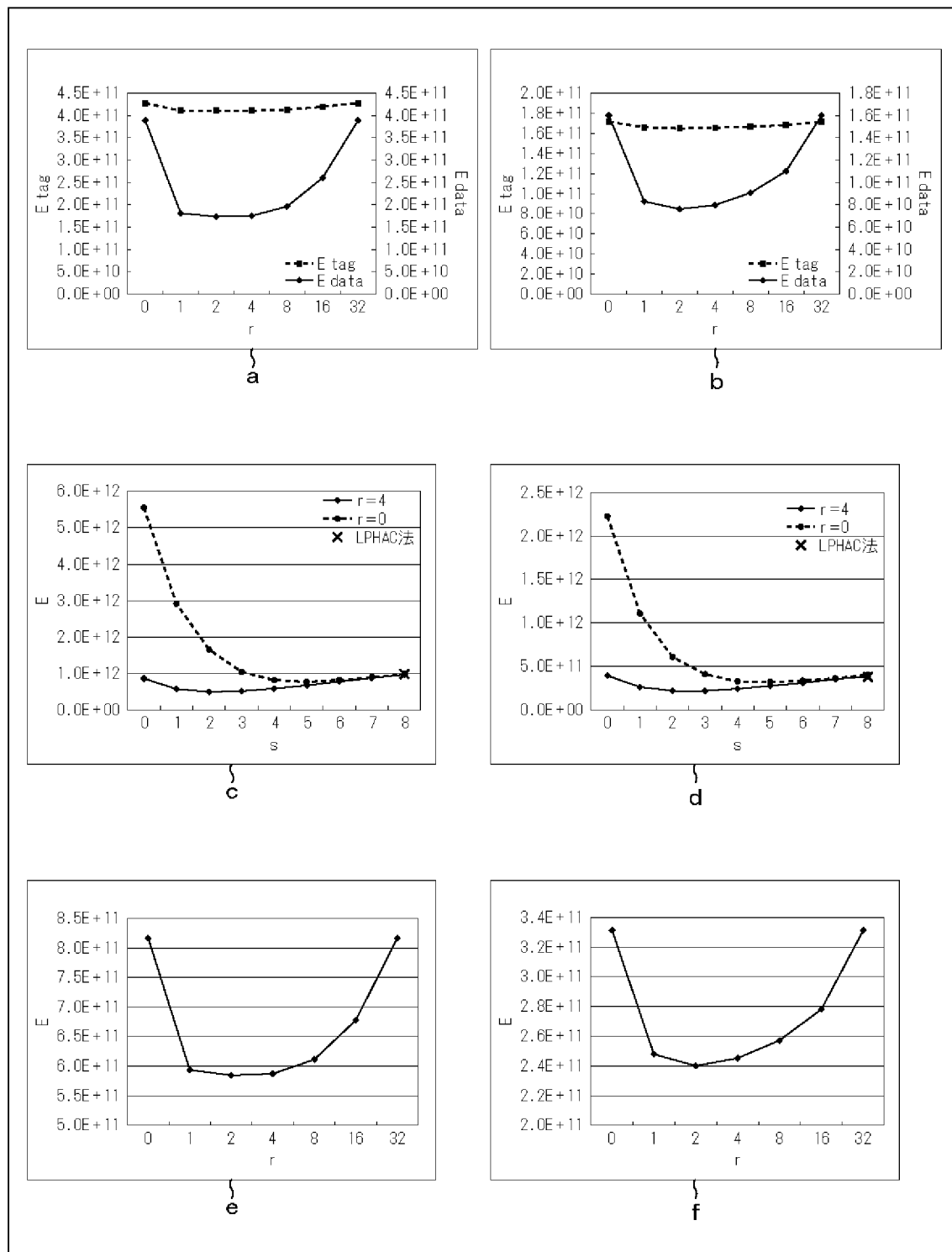






[図15]

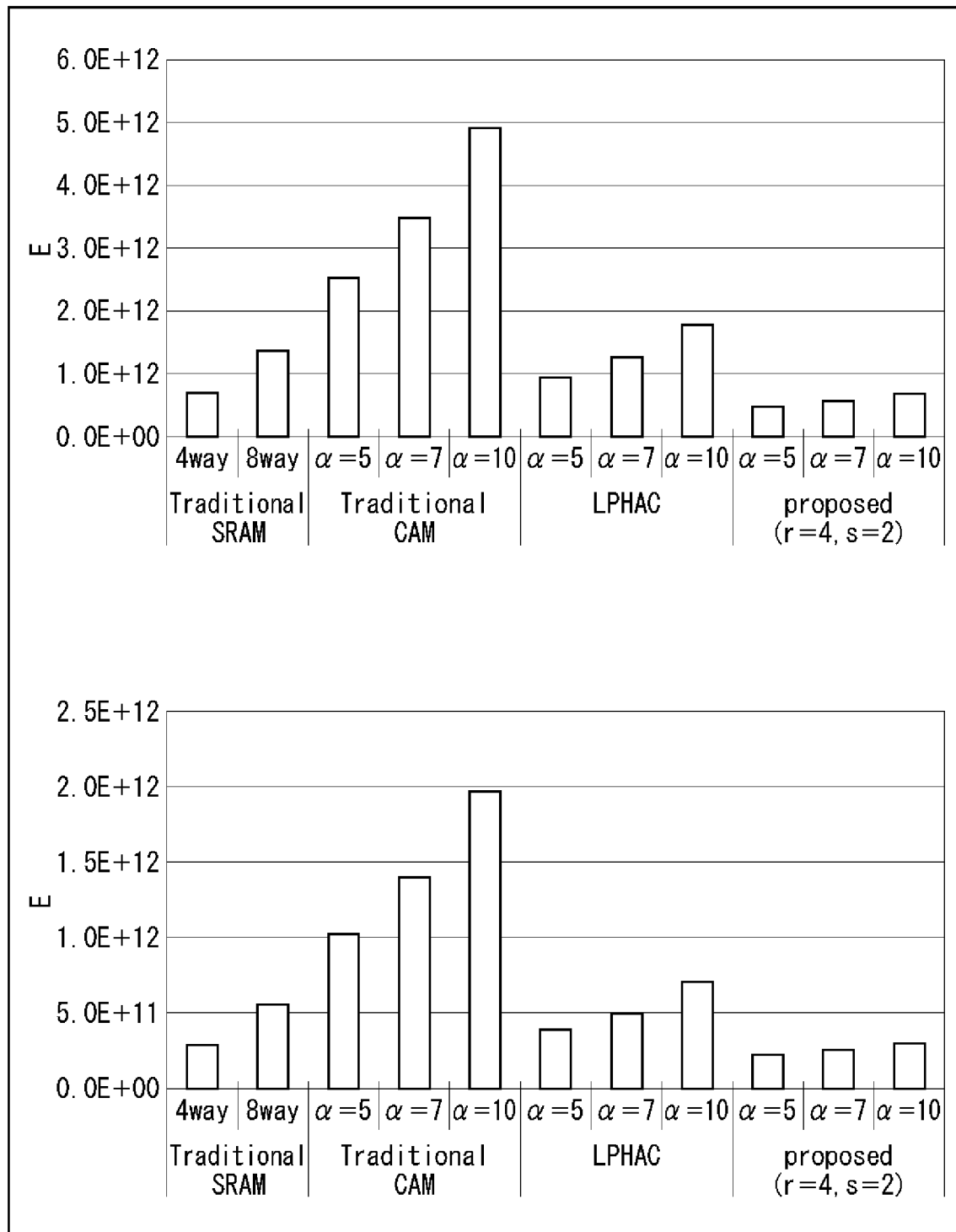
図15





[図16]

図16



**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2010/068298

**A. CLASSIFICATION OF SUBJECT MATTER**

G06F12/08(2006.01)i, G06F12/12(2006.01)i, G11C15/04(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G06F12/08-12/12, G11C15/04

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2010
Kokai Jitsuyo Shinan Koho	1971-2010	Toroku Jitsuyo Shinan Koho	1994-2010

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2003-519835 A (Advanced Micro Devices, Inc.), 24 June 2003 (24.06.2003), entire text; all drawings & US 6687789 B1 & EP 1244970 A1 & WO 2001/050272 A1 & CN 1415092 A	1-11
A	JP 62-293596 A (Matsushita Electric Industrial Co., Ltd.), 21 December 1987 (21.12.1987), entire text; all drawings (Family: none)	1-11
A	JP 6-131265 A (Inmos Ltd.), 13 May 1994 (13.05.1994), entire text; all drawings & US 5574875 A & EP 560598 A1	1-11

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search  
27 December, 2010 (27.12.10)

Date of mailing of the international search report  
11 January, 2011 (11.01.11)

Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2010/068298

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 5-035599 A (Matsushita Electric Industrial Co., Ltd.), 12 February 1993 (12.02.1993), entire text; all drawings (Family: none)	1-11
A	JP 3850669 B2 (Samsung Electronics Co., Ltd.), 29 November 2006 (29.11.2006), entire text; all drawings & US 2001/0026465 A1 & GB 2366046 A & FR 2807176 A & KR 10-2001-0091109 A	1-11
A	JP 8-263370 A (Toshiba Micro-Electronics Corp.), 11 October 1996 (11.10.1996), entire text; all drawings & US 5845309 A	1-11

A. 発明の属する分野の分類 (国際特許分類 (IPC))  
 Int.Cl. G06F12/08(2006.01)i, G06F12/12(2006.01)i, G11C15/04(2006.01)i

B. 調査を行った分野  
 調査を行った最小限資料 (国際特許分類 (IPC))  
 Int.Cl. G06F12/08-12/12, G11C15/04

最小限資料以外の資料で調査を行った分野に含まれるもの  
 日本国実用新案公報 1922-1996年  
 日本国公開実用新案公報 1971-2010年  
 日本国実用新案登録公報 1996-2010年  
 日本国登録実用新案公報 1994-2010年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	J P 2 0 0 3 - 5 1 9 8 3 5 A (アドバンスト・マイクロ・デバイス・インコーポレイテッド) 2003.06.24, 全文, 全図 & US 6687789 B1 & EP 1244970 A1 & WO 2001/050272 A1 & CN 1415092 A	1-11

C欄の続きにも文献が列挙されている。  パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー	の日の後に公表された文献
「A」特に関連のある文献ではなく、一般的技術水準を示すもの	「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)	「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「O」口頭による開示、使用、展示等に言及する文献	「&」同一パテントファミリー文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願	

国際調査を完了した日 27.12.2010	国際調査報告の発送日 11.01.2011
国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) ▲高▼橋 正▲徳▼ 電話番号 03-3581-1101 内線 3586

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 62-293596 A (松下電器産業株式会社) 1987. 12. 21, 全文, 全図 (ファミリーなし)	1-11
A	JP 6-131265 A (インモス、リミテッド) 1994. 05. 13, 全文, 全図 & US 5574875 A & EP 560598 A1	1-11
A	JP 5-035599 A (松下電器産業株式会社) 1993. 02. 12, 全文, 全図 (ファミリーなし)	1-11
A	JP 3850669 B2 (三星電子株式会社) 2006. 11. 29, 全文, 全図 & US 2001/0026465 A1 & GB 2366046 A & FR 2807176 A & KR 10-2001-0091109 A	1-11
A	JP 8-263370 A (東芝マイクロエレクトロニクス株式会社) 1996. 10. 11, 全文, 全図 & US 5845309 A	1-11