

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4164452号  
(P4164452)

(45) 発行日 平成20年10月15日(2008.10.15)

(24) 登録日 平成20年8月1日(2008.8.1)

(51) Int.Cl.	F 1
<b>G 0 6 F 12/08 (2006.01)</b>	G 0 6 F 12/08 5 2 3 C
	G 0 6 F 12/08 5 1 7 B
	G 0 6 F 12/08 5 3 1 C
	G 0 6 F 12/08 5 4 1 Z

請求項の数 9 (全 10 頁)

(21) 出願番号	特願2004-25349 (P2004-25349)	(73) 特許権者	000001007 キヤノン株式会社 東京都大田区下丸子3丁目30番2号
(22) 出願日	平成16年2月2日(2004.2.2)	(74) 代理人	100076428 弁理士 大塚 康德
(65) 公開番号	特開2005-216220 (P2005-216220A)	(74) 代理人	100112508 弁理士 高柳 司郎
(43) 公開日	平成17年8月11日(2005.8.11)	(74) 代理人	100115071 弁理士 大塚 康弘
審査請求日	平成18年11月29日(2006.11.29)	(74) 代理人	100116894 弁理士 木村 秀二
		(72) 発明者	小川 武志 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

最終頁に続く

(54) 【発明の名称】 情報処理方法及び装置

(57) 【特許請求の範囲】

【請求項1】

メモリの有する同一のメモリエリアを、複数の異なるアドレスセットでCPUによりアクセス可能に構成したメモリ回路と、

前記複数の異なるアドレスセットの一方のアドレスセットをキャッシュابل用のアドレスセットとし、他方をアンキャッシュابل用のメモリアドレスセットとしてキャッシュ制御するキャッシュ手段と、

前記メモリに所定サイズのエリアが要求されると、当該要求された所定サイズにキャッシュ単位であるラインサイズの2倍のサイズを付加して、前記メモリの前記アンキャッシュابل用のメモリアドレスセットによりアクセス可能なメモリエリアに割当てる割当て手段と、

を有することを特徴とする情報処理装置。

【請求項2】

更に、前記メモリに確保されたエリアに対してキャッシュのフラッシュ及びクリアを行う手段を有することを特徴とする請求項1に記載の情報処理装置。

【請求項3】

前記所定サイズのエリアの前後それぞれに、前記キャッシュ単位であるラインサイズのデータエリアが付加されることを特徴とする請求項1に記載の情報処理装置。

【請求項4】

前記メモリの前記所定サイズのエリアのデータをDMA転送するDMA制御手段を更に

有することを特徴とする請求項 1 に記載の情報処理装置。

【請求項 5】

C P U、メモリ、キャッシュメモリ及びキャッシュコントローラを具備する情報処理装置で実行される情報処理方法であって、

前記メモリに所定サイズのエリアが要求されると、当該要求された所定サイズにキャッシュ単位であるラインサイズの少なくとも 2 倍のサイズを付加して前記メモリに前記エリアを確保する工程と、

前記エリアを前記メモリのアンキャッシュابل用のメモリエリアに割当てる工程と、を有することを特徴とする情報処理方法。

【請求項 6】

前記確保する工程では、前記所定サイズのエリアの前後それぞれに、前記キャッシュ単位であるラインサイズのデータエリアが付加されることを特徴とする請求項 5 に記載の情報処理方法。

【請求項 7】

更に、前記メモリに確保されたエリアに対してキャッシュのフラッシュ及びクリアを行う工程を有することを特徴とする請求項 5 に記載の情報処理方法。

【請求項 8】

前記メモリの前記所定サイズのエリアのデータを D M A 転送する D M A 制御工程を更に有することを特徴とする請求項 5 に記載の情報処理方法。

【請求項 9】

請求項 5 乃至 8 のいずれか 1 項に記載の情報処理方法を、コンピュータに実行させることを特徴とするプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、キャッシュ機能を備えた情報処理装置及びその方法に関するものである。

【背景技術】

【0002】

マイクロコンピュータの主記憶として、安価で大容量な D R A M を利用する場合、D R A M を含むバスよりも C P U チップ内のバスを高速に設計することが可能であり、C P U 内部バスのクロックを高くして設計することでコンピュータの処理能力を高めることができる。このような C P U を備えたコンピュータシステムにおいて、C P U から外部バスへのメモリアクセスを効率的に行うために、キャッシュメモリのハードウェアを内部バスと外部バスの間に介在させるのが一般的である。

【0003】

このようなキャッシュメモリは、C P U の内部バスと外部バスの間に存在しているので、C P U 側から見ると透過的だが、例えば D M A 等において外部バスを介して外部メモリの読み書きを行った場合は、キャッシュ内にあるデータと主記憶内にあるデータとが一致しない場合が発生する。このような不具合を避けるために、D M A 転送を行う前後に、その D M A 転送対象のメモリエリアに対してキャッシュメモリのクリアやフラッシュ（キャッシュ内容をメモリに書き戻す）を行う方法と、D M A 転送の対象となるメモリエリアをアンキャッシュابلにしておく方法がある（例えば、特許文献 1）。本願発明は、後者のアンキャッシュابل領域を割り当てる方法に関する。

【0004】

D M A 転送に使用するメモリ領域を動的に割り当てる A P I を持った組み込み用の O S が広く利用されている。このような O S はメモリマネージャからメモリエリアを取得し、キャッシュコントローラを制御して、その取得したメモリエリアをアンキャッシュابلに設定できる。一般にキャッシュコントローラはメモリエリアの設定機能を備えており、その機能を用いて、メモリエリアに対してキャッシュابلエリアとアンキャッシュابلエリアを複数指定できる。しかしながら、このようなエリアを指定するためのレジスタの数が

10

20

30

40

50

有限であり、また指定できるエリアが多いほど回路が冗長になるため、無制限にこれらエリアを指定できるわけではない。

【 0 0 0 5 】

一方、DMAを利用するアプリケーションソフトウェアでは、1つのDMAチャンネルに対して複数のバッファを用意して処理の効率化を図っている。例えば一つ目のバッファへのDMA転送中には、二つ目のバッファに対するDMAの準備や後処理を行っている。

【特許文献1】特開2000-181796号公報

【発明の開示】

【発明が解決しようとする課題】

【 0 0 0 6 】

上記事情に鑑みると、システムで利用できるDMAのチャンネル数よりも、指定できるアンキャッシュابلエリアの数が数倍あることが望ましいが、DMAのチャンネル数が多いと、このようなエリアコントロールのためのレジスタの数が増えて回路規模が膨大になってしまう。

【 0 0 0 7 】

別の方法として、予め大きめのアンキャッシュابلエリアを用意しておき、そのエリアを管理するメモリマネージャによってアンキャッシュابلエリアを割り当てる方法もある。その場合、アンキャッシュابلメモリのエリアを静的に決めておく必要があり、メモリ全体の利用効率を低下させる。

【 0 0 0 8 】

本発明は上記問題点に鑑みてなされたもので、キャッシュのエリア制御を容易にした情報処理装置及びその方法を提供する。

【課題を解決するための手段】

【 0 0 0 9 】

本発明の情報処理装置の特徴は、メモリの有する同一のメモリエリアを、複数の異なるアドレスセットでCPUによりアクセス可能に構成したメモリ回路と、前記複数の異なるアドレスセットの一方のアドレスセットをキャッシュابل用のアドレスセットとし、他方をアンキャッシュابل用のメモリアドレスセットとしてキャッシュ制御するキャッシュ手段と、前記メモリに所定サイズのエリアが要求されると、当該要求された所定サイズにキャッシュ単位であるラインサイズの2倍のサイズを付加して、前記メモリの前記アンキャッシュابل用のメモリアドレスセットによりアクセス可能なメモリエリアに割り当てる割当て手段とを有することにある。

【 0 0 1 0 】

また本発明の情報処理方法の特徴は、CPU、メモリ、キャッシュメモリ及びキャッシュコントローラを具備する情報処理装置で実行される情報処理方法であって、

前記メモリに所定サイズのエリアが要求されると、当該要求された所定サイズにキャッシュ単位であるラインサイズの少なくとも2倍のサイズを付加して前記メモリに前記エリアを確保する工程と、前記エリアを前記メモリのアンキャッシュابل用のメモリエリアに割り当てる工程とを有することを特徴とする。

【発明の効果】

【 0 0 1 1 】

本発明によれば、キャッシュのエリア制御を容易にできる。

【発明を実施するための最良の形態】

【 0 0 1 2 】

以下、添付図面を参照して本発明の好適な実施の形態を詳しく説明する。

【 0 0 1 3 】

図1は、本発明の実施の形態に係る情報処理装置におけるCPU周りのハードウェア構成を説明する図で、入出力部や表示部などの構成を省略して示している。

【 0 0 1 4 】

図1において、101はCPUで、メインメモリ104に記憶されたOSやアプリケー

10

20

30

40

50

ションプログラムに従って各種動作を制御している。キャッシュコントローラ102は、キャッシュメモリ103を使用したキャッシュ制御を実行している。メインメモリ104は、CPU101により実行されるプログラムやデータ等を記憶している。105はDMAコントローラ(DMAC)で、CPU101を介することなく、例えばハードディスクコントローラ(HDC)106とメインメモリ104との間でのデータ転送を行うことができる。ハードディスク(HD)は大容量の外部メモリ装置で、ここにインストールされているOSや各種アプリケーションがメインメモリ104にロードされてCPU101により実行される。ここでキャッシュメモリ103は高速のSRAMを使用し、キャッシュコントローラ102の制御の下に大量のデータを蓄積し、古くて使用頻度の低いデータアドレスを破棄しながらキャッシュメモリの有効利用を図っている。

10

## 【0015】

ここでキャッシュメモリ103とCPU101はワードやバイトといった単位でデータ転送を行うのに対して、キャッシュメモリ103とメインメモリ104との間でのデータ転送はライン単位で行う。即ち、1ビットのデータが必要な場合でも、ライン全体のデータがやり取りされる(ラインフィル)。このためにキャッシュメモリ103を備える装置では、キャッシュメモリ103とメインメモリ104間のデータ転送はラインサイズ(例えばバス幅が64ビットの場合には256ビット)を最小単位として行われる。

## 【0016】

<シャドウメモリ>

図2は、CPU101とメインメモリ104との間のアドレス配線を説明する図である。

20

## 【0017】

CPU101から出ているアドレス信号の内、上位の数ビットを配線しないことにより、メモリ空間を小さく抑えることが可能である。図において、CPU101は、アドレスビットA0~A22で定義されるメモリ空間、即ち、アドレス0番地~7FFFFFF番地までのメモリ空間を有しており、メインメモリ104は、20ビットのアドレス(A0~A19)、即ち、アドレス0番地~FFFFFF番地までのメモリ空間を有している。ここでCPU101のメモリアドレスの内、上位アドレス(アドレスA21, A22)のアドレス線をメモリ104のアドレスに接続しておらず、またアドレスA20は、メインメモリ104のCE(チップイネーブル)に接続されている。

30

## 【0018】

図3は、図2の配線の場合のCPU101のメモリマップを説明する図である。

## 【0019】

CPU101から見たとき、メモリ104の0番地から0xFFFF番地までエリア301の内容と同じものが、例えばアドレス0x200000番地から0x2FFFFFF番地までのメモリエリア302でもみえることになる。この場合、アドレス0x200000番地~0x2FFFFFF番地は、0番地から0xFFFF番地のシャドウという。

## 【0020】

本実施の形態に係る情報処理装置では、このシャドウ側のメモリエリア302をアンキャッシュابلに設定して利用する。例えば、0番地から0xFFFF番地までのエリア301をキャッシュابلとし、0x200000番地から0x2FFFFFF番地までのエリア302をアンキャッシュابلにする。このアンキャッシュابلエリアの設定がキャッシュコントローラ102に対して与えられると、これ以降キャッシュコントローラ102は、そのメモリエリアへのアクセス時には、キャッシュ動作を実行しないように制御する。

40

## 【0021】

<汎用メモリマネージャの利用>

本実施の形態に係る情報処理装置では、アンキャッシュابلメモリ領域302の割り当てに、汎用のメモリマネージャを利用する。この汎用のメモリマネージャとは、例えばC言語の標準関数であるmalloc関数として提供される。このmalloc関数によって確保したメモリエリアは、そのままではキャッシュابلエリア301に設定される。従って、確保し

50

たアドレスの内のデコードされていないビットを変更して、アンキャッシュابلメモリエリア 3 0 2 のアドレスとして、それをクライアントプログラム、即ち、malloc関数を実行したプログラムに返却する。例えば、malloc関数によって、キャッシュابلメモリ領域 3 0 1 のアドレス0x01000番地からアドレス0x014FF番地に確保されたメモリエリアを、CPU 1 0 1 からみたアドレスビット A 2 1 を「1」にすることにより、アンキャッシュابلメモリエリア 3 0 2 のアドレス0x201000番地からアドレス0x2014FF番地に確保する。

【 0 0 2 2 】

次に、このアンキャッシュابلメモリエリア 3 0 2 を解放する場合には、クライアントプログラムは逆に、メインメモリ 1 0 4 のアドレスの内、デコードされていないビット（上述の例ではアドレスビット A 2 1 ）を変更し、キャッシュابلエリア 3 0 1 のエリア（アドレス0x01000番地～アドレス0x01500番地）に変更する。そして次に、free関数を使って、メモリマネージャに、この確保しているエリアの解放を指示する。

【 0 0 2 3 】

このようにシャドウメモリを利用することによって、ハードウェアのエリア指定レジスタを使わずに無制限にキャッシュابلエリアを生成することが可能となる。

【 0 0 2 4 】

<クリア>

上記のようにしてアンキャッシュابلメモリとして割り当てられたメモリエリア（上述の例では、アドレス0x01000番地～アドレス0x014FF番地）は、アンキャッシュابلエリアに割り当てられる直前までキャッシュابلメモリとして利用されていた可能性がある。その場合、アンキャッシュابلメモリとして使用し始めてからキャッシュの吐き出しが起こると、それまでキャッシュされていた内容により、そのメモリエリアの内容が破壊されてしまう。従って、少なくとも、クライアントにアンキャッシュابلメモリとして割り当てる前に、そのメモリエリアに対してキャッシュされている情報をクリアする必要がある。

【 0 0 2 5 】

<ラインサイズ>

上記のように割り当てたアンキャッシュابلメモリは、そのメモリエリア周辺のアドレスがキャッシュابلメモリとして利用されている。一般にキャッシュは、ライン単位（キャッシングの単位で、例えば1ラインは256バイト）でLRU（Least Recent Used）のアルゴリズムで管理されているが、汎用のメモリマネージャは、キャッシュのラインサイズとは無関係に、メモリエリアの割り当てを行う。

【 0 0 2 6 】

図4を参照して、より詳しく説明する。

【 0 0 2 7 】

図4において、400はクライアントプログラムに割り当てられたメモリエリアを示している。しかし、汎用のメモリマネージャを利用しているため、401で示すラインの一部がクライアント領域の外側と共有されている。クライアントプログラムへメモリエリアを引き渡したときに、402で示す部分は、まだクライアントプログラムからの書き込みが行われていない。その後、403で示す部分、即ち、クライアント領域とラインを共有している外側の領域を他のクライアントプログラムが書き換えると、クライアント使用領域401に対してダーティビットが立つ。そして次に、クライアントプログラムがクライアント使用領域400全体のデータ404をDMA転送する。この時点ではメモリは破壊されていない。そしてシステムのプログラムの実行が進みLRUアルゴリズムによって、401で示すラインがキャッシュから吐き出される時、405で示すように、ライン401全体が書き出される。これにより領域406が、DMA転送前のデータで上書きされ、そのラインの内容が破壊されてしまう。

【 0 0 2 8 】

そこで本実施の形態では、クライアントプログラムの使用領域414の両端に、それぞれラインサイズ分の不使用領域を設ける。これにより、汎用のメモリマネージャを利用しても、メモリの内容が破壊されないように防止する。

## 【 0 0 2 9 】

これを説明したのが図4の410で示す部分で、411, 412は、このようなデータ破壊を防ぐために両端に付加された不使用領域を示し、それぞれ1ライン分に相当するデータエリアである。

## 【 0 0 3 0 】

このように本実施の形態では、クライアントプログラムからアンキャッシュブル領域を要求されると、その要求された領域のサイズに2ライン分のサイズを上乗せしたサイズ(413で示す)で汎用メモリマネージャからメモリエリアを取得し、クライアントプログラムに、414で示す安全なエリア情報だけを使用領域として渡す。この場合のクライアントの使用可能なメモリエリアは414で示す領域となる。

10

## 【 0 0 3 1 】

図5は、本発明の実施の形態1に係る情報処理装置における処理を説明するフローチャートである。本実施の形態では、アンキャッシュブルなエリアをダイナミックに割り付けるメモリマネージメントの例で説明する。このプログラムは例えば、「cacheDmaMalloc( )」といった名前のAPIで、ユーザプログラムから呼び出されてアンキャッシュブルなエリアを割り付ける。このプログラムは、HD107にインストールされているOSに含まれており、OSのロードにメインメモリ104にロードされて実行される。このプログラムはまた既存のOSに対してインストール可能な実行コマンドとしても供給される。

## 【 0 0 3 2 】

まずステップS1で、APIを通じてキャッシュブルエリアの確保を開始する。通常、パラメータとしてユーザが要求するメモリサイズを受け取る。次にステップS2で、ユーザプログラムが要求したメモリサイズに対して、キャッシュの2ライン分のサイズを付加したサイズを計算し、そのサイズで通常メモリマネージャにメモリエリアを要求する。ここで、利用するメモリマネージャは、例えばC言語のインターフェースで、malloc( )関数に相当するものであり、特別なメモリマネージャではない。次にステップS3で、ステップS2での結果を検査し、メモリの確保に成功したかどうかをみる。失敗した場合はステップS4に進み、このサービスを終了する。この失敗する例としては、例えば、割り当てられるメモリエリアが無くなってしまった場合等が該当している。malloc( )関数の場合、メモリエリアの割り当てに失敗するとNULLコードを返却するので、ステップS4でも同様に失敗であることを示す、例えばNULLなどの値を返却する。

20

30

## 【 0 0 3 3 】

一方ステップS3で、メモリエリアの取得に成功するとステップS5に進み、その確保したメモリエリアに対してキャッシュコントローラ102を制御して、前述したフラッシュとクリアを行う。この時、確保したメモリエリアはメモリマネージャからみると未使用だった領域を割り当てているが、直前までユーザプログラムによって使用されていた場合もある。このため、使用中だった時に書き込んだ値がまだキャッシュ内部に残留していた場合には、後でキャッシュの吐き出しによって有効なデータを上書きして破壊する可能性があるため、キャッシュをクリアする。このクリアの前にフラッシュを行う理由は、確保したエリアとラインを共有している周辺エリアがキャッシュ内に未書き込みな有効なデータを持っている可能性があるからである。

40

## 【 0 0 3 4 】

こうしてステップS5で、メモリの準備が完了するとステップS6に進み、その確保したエリアの先頭番地に1ライン分を足したアドレスを計算し、そのシャドウエリアの番地を計算する。これは図4で説明したように、実際にクライアントプログラムが使用する領域は、先頭と最後の各1ラインデータを除いた領域であるためである。例えばラインサイズが前述したように256バイトで、図3に示すようなメモリマップの場合で説明する。ここで確保したメモリのアドレスを「0x00004010」とすると、「0x00004010」+「0x100(ラインサイズ)」+「0x00200000」「0x00204110」がユーザに渡されるアドレスとなる。そしてステップS7で、ユーザプログラムに計算したメモリの番地を返却して終了する。

50

## 【 0 0 3 5 】

図 6 は、本実施の形態に係るクライアントプログラムで取得したアンキャッシュابلなメモリエリアを OS に返却する処理を示すフローチャートで、例えば「cacheDmaFree( )」といった名前の API で提供される。このプログラムも同様に、HD 107 にインストールされている OS に含まれており、OS のロードにメインメモリ 104 にロードされて実行される。このプログラムはまた既存の OS に対してインストール可能な実行コマンドとしても供給される。

## 【 0 0 3 6 】

まずステップ S 11 で、クライアントプログラムから呼び出される。このときパラメータとして、クライアントプログラムに割り当てられていたメモリエリアの先頭番地を受け取る。次にステップ S 12 で、指定された番地から 1 ライン分引いて、そのキャッシュابلエリア側の番地を計算する。例えば図 5 の例によれば、(0x00204110 & 0xFFFFFFFFFF) - 0x100 0x00004010 となる。ここで求めた値は、メモリマネージャから malloc( ) 関数で取得したアドレスであるため、ステップ S 13 で、free( ) 関数によってメモリマネージャへ返却する。そしてステップ S 14 で、このサービスを終了する。

## 【 0 0 3 7 】

以上説明したように本実施の形態によれば、アンキャッシュابلエリアの設定を、レジスタなどの数に制限されずに行うことができる。

## 【 0 0 3 8 】

また本実施の形態によれば、汎用のメモリマネージャを利用して、メモリエリアの確保及びキャッシュ制御を行うことができる。

## 【 0 0 3 9 】

## [ その他の実施の形態 ]

前述の実施の形態 1 では、汎用のメモリマネージャを使用していたが、キャッシュのラインサイズを意識したメモリマネージャを用意し、ユーザエリアとラインを共有する領域ができないように工夫をすることにより、前後 1 ライン分の不使用領域を省く方法もある。

## 【 0 0 4 0 】

また前述の実施の形態 1 では、フルアドレスデコードしない回路構成で説明しているが、MMU (Memory Management Unit) 等のしくみを使って、同一メモリを他のアドレスでアクセスするシャドウを構成してもかまわない。

## 【 0 0 4 1 】

本発明の目的は前述したように、本実施の形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体をシステム或は装置に提供し、そのシステム或は装置のコンピュータ (又は CPU や MPU) が記憶媒体に格納されたプログラムコードを読み出し実行することによっても達成される。この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。このようなプログラムコードを供給するための記憶媒体としては、例えば、フロッピーディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモリカード、ROM などを用いることができる。

## 【 0 0 4 2 】

また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施の形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼動している OS (オペレーティングシステム) などが実際の処理の一部又は全部を行い、その処理によって前述した実施の形態の機能が実現される場合も含まれている。

## 【 0 0 4 3 】

更に、記憶媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書きこまれた

10

20

30

40

50

後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部又は全部を行い、その処理によって前述した実施の形態の機能が実現される場合も含む。

【図面の簡単な説明】

【0044】

【図1】本発明の実施の形態に係る情報処理装置のCPU周辺の構成を説明するブロック図である。

【図2】本実施の形態に係る情報処理装置におけるCPUとメモリとの接続を説明する図である。

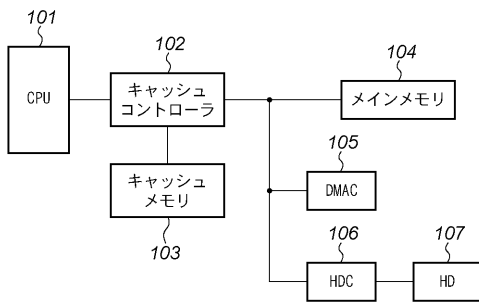
【図3】実施の形態に係るメインメモリのメモリマップを説明する図である。

【図4】従来のラインを共有することによるメモリにおけるデータの破壊と本実施の形態におけるメモリ使用領域の確保を説明する図である。

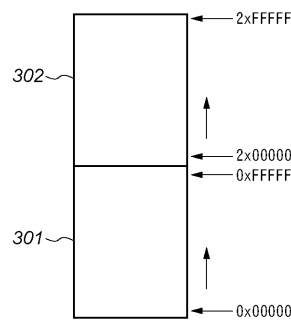
【図5】本発明の実施の形態に係る情報処理装置におけるアンキャッシュブルエリアの確保処理を説明するフローチャートである。

【図6】本発明の実施の形態に係る情報処理装置におけるアンキャッシュブルエリアの解放処理を説明するフローチャートである。

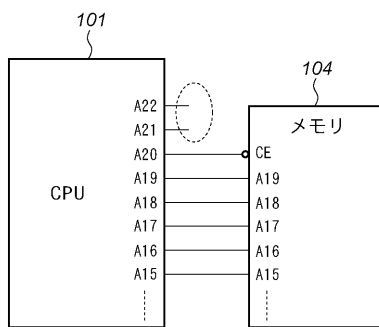
【図1】



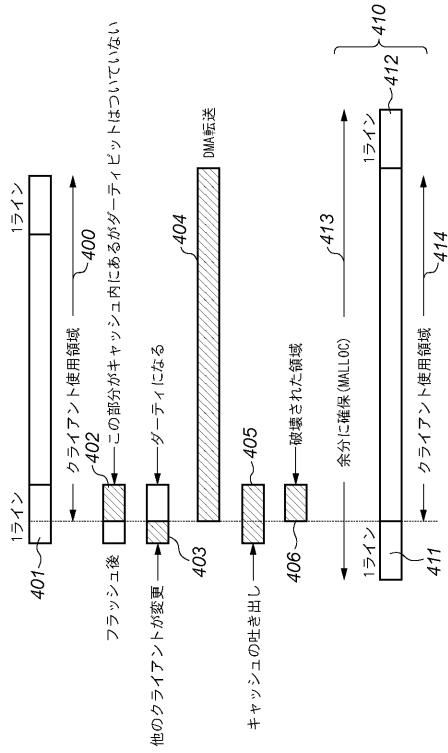
【図3】



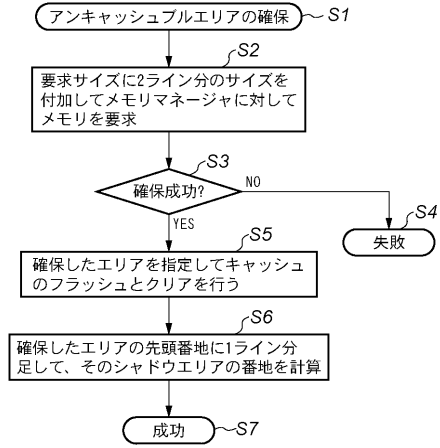
【図2】



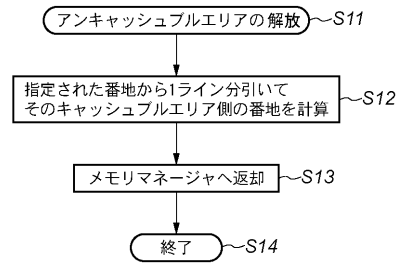
【図4】



【図5】



【図6】



---

フロントページの続き

(72)発明者 崎村 岳生

東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

審査官 清木 泰

(56)参考文献 特開2005-190161(JP,A)

特開平10-078916(JP,A)

特開平06-348593(JP,A)

特開平04-051373(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F12/08-12/12