

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2014-179084

(P2014-179084A)

(43) 公開日 平成26年9月25日(2014.9.25)

(51) Int.Cl.		F I			テーマコード (参考)
G06F 3/06 (2006.01)		G06F 3/06	3 O 1 K		5 B O 6 O
G06F 3/08 (2006.01)		G06F 3/08	H		
G06F 12/02 (2006.01)		G06F 3/06	3 O 6 K		
G06F 12/00 (2006.01)		G06F 12/02	5 3 O E		
		G06F 12/00	5 9 7 U		
審査請求 未請求 請求項の数 10 O L (全 21 頁)					

(21) 出願番号 特願2014-48594 (P2014-48594)
 (22) 出願日 平成26年3月12日 (2014.3.12)
 (31) 優先権主張番号 61/783,555
 (32) 優先日 平成25年3月14日 (2013.3.14)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 13/963,074
 (32) 優先日 平成25年8月9日 (2013.8.9)
 (33) 優先権主張国 米国 (US)

(特許庁注：以下のものは登録商標)

1. イーサネット
2. ETHERNET
3. フロッピー

(71) 出願人 508243639
 エルエスアイ コーポレーション
 アメリカ合衆国カリフォルニア州95131, サンノゼ, リッター・パーク・ドライブ 1320
 (74) 代理人 100087642
 弁理士 古谷 聡
 (74) 代理人 100082946
 弁理士 大西 昭広
 (74) 代理人 100121061
 弁理士 西山 清春
 (74) 代理人 100195693
 弁理士 細井 玲
 (74) 代理人 100094112
 弁理士 岡部 譲

最終頁に続く

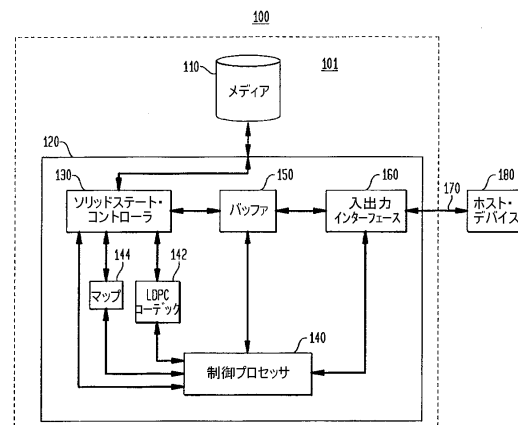
(54) 【発明の名称】 ソリッドステート・メディアにおいてマルチレベル・マッピングを使用する機構

(57) 【要約】

【課題】論理アドレスと、アドレス範囲とを含む要求を受け取るメディア・コントローラを提供する。

【解決手段】この要求に回答して、メディア・コントローラは、受け取られた要求が無効にする要求であるかどうかを判定する。受け取られた要求タイプが無効にする要求である場合、メディア・コントローラは、マップを使用して、その論理アドレスおよび範囲に関連付けられたマップの1つまたは複数のエントリを特定する。これらのマップ・エントリのそれぞれに関連付けられたマップの中の標識が、これらのマップ・エントリが無効にされるべきことを示すように設定される。メディア・コントローラは、無効にする要求が完了したという確認応答をホスト・デバイスに対して行い、メディア・コントローラのアイドル・モードにおいて、無効にされるべきマップ・エントリに基づいて空きスペース・カウントを更新する。無効にされたマップ・エントリに関連付けられた物理アドレスが、ホスト・デバイスからのその後の要求に関して再使用されるように利用可能にされる。

【選択図】図1



【特許請求の範囲】**【請求項 1】**

ソリッドステート・メディアおよびホスト・デバイスと通信状態にあるメディア・コントローラによって受け取られたホスト要求を処理する方法であって、

前記メディア・コントローラにより、前記ホスト・デバイスからの要求を受け取り、前記要求は、少なくとも 1 つの論理アドレスと、アドレス範囲とを含むこと、

前記要求を受け取ったことに応答して、

前記受け取られた要求が無効にする要求であるかどうかを判定し、前記受け取られた要求タイプが無効にする要求である場合、

前記メディア・コントローラのマップを使用して、前記ソリッドステート・メディアの前記少なくとも 1 つの論理アドレスおよび前記アドレス範囲に関連付けられた前記マップの 1 つまたは複数のエントリを特定すること、

前記マップの中で、前記マップ・エントリのそれぞれに関連付けられた標識に印を付け、前記標識は、前記 1 つまたは複数のマップ・エントリが無効にされるべきことを示すこと、

前記ホスト・デバイスに対して、前記無効にする要求が完了したという確認応答を行うこと、

前記メディア・コントローラのアイドル・モードにおいて、無効にされるべき前記 1 つまたは複数のマップ・エントリに基づいて前記メディア・コントローラの空きスペース・カウントを更新すること、および

前記無効にされたマップ・エントリに関連付けられた物理アドレスを、前記ホスト・デバイスからのその後の要求に関して再使用されるように利用可能にすることを備える方法。

【請求項 2】

前記マップは、第 1 レベル・マップと、第 2 レベル・マップとを備えるマルチレベル・マップである方法であって、

複数の第 2 レベル・マップ・ページの複数のエントリのそれぞれを、前記ソリッドステート・メディアの物理アドレスに関連付けること、および

前記第 1 レベル・マップの複数のエントリのそれぞれを或る第 2 レベル・マップ・ページに関連付け、その結果、前記少なくとも 1 つの論理アドレス、および前記アドレス範囲を前記第 2 レベル・マップ・エントリのうちの少なくとも 1 つに関連付けることをさらに備える、請求項 1 に記載の方法。

【請求項 3】

各第 2 レベル・マップ・エントリは、有効指示を備え、

各第 1 レベル・マップ・エントリは、前記第 2 レベル・マップ・ページのうちの対応する 1 ページのアドレスと、前記第 1 レベル・マップ・エントリに関連付けられた前記第 2 レベル・マップの 1 つまたは複数のエントリに対応する有効指示と、前記第 2 レベル・マップのエントリが無効であるが、書き込むのに利用できない場合にそのことを示すように構成された処理されるべき (T B P) 標識とを備える方法であって、

前記要求の前記アドレス範囲内に完全に含まれる前記第 2 レベル・マップ・ページのうちのページに関連付けられた前記第 1 レベル・マップの中の前記 T B P 標識のうちの特定の T B P 標識を設定することによって、前記第 2 レベル・マップのエントリを無効にすることをさらに備える、請求項 2 に記載の方法。

【請求項 4】

前記要求の前記アドレス範囲内に部分的にだけ含まれる前記第 2 レベル・マップ・ページのうちのページのエントリを直接にトリミングすることによって、前記第 2 レベル・マップのエントリを無効にすることをさらに備える、請求項 3 に記載の方法。

【請求項 5】

前記 T B P 標識が所与の第 2 レベル・マップに関して設定されている場合、前記所与の第 2 レベル・マップに関する後続の無効にする要求が受け取られるまで、前記所与の第 2

10

20

30

40

50

レベル・マップに関する前記無効にする要求を処理することを遅延させて、その結果、前記無効にする要求に関する処理時間を短くし、前記所与の第2レベル・マップを更新する前記ソリッドステート・メディアに対する書き込み動作を少なくすることをさらに備える、請求項3に記載の方法。

【請求項6】

前記第2レベル・マップの前記無効にされたエントリのそれぞれを、前記ホスト・デバイスからのその後の要求に関して再使用されるように利用可能にする前記ステップは、設定されている場合、前記関連付けられたTBP標識をクリアすること、および設定されている場合、関連付けられた有効標識をクリアすることを備える、請求項3に記載の方法。

10

【請求項7】

前記メディア・コントローラのマップ・メモリの中に前記第1レベル・マップを格納すること、
前記ソリッドステート・メディアの中に前記第2レベル・マップ・ページのすべてを格納すること、および
前記メディア・コントローラの制御プロセッサに結合されたマップ・キャッシュの中に前記第2レベル・マップ・ページの少なくともサブセットを一時的に格納することをさらに備える、請求項3に記載の方法。

【請求項8】

前記メディア・コントローラの前記空きスペース・カウントは、複数のエントリを有するブロック使用されるスペース・テーブルを備える方法であって、
前記ブロック使用されるスペース・テーブルの各エントリを前記ソリッドステート・メディアの複数の物理領域のうちの1つに関連付けることをさらに備える、請求項1に記載の方法。

20

【請求項9】

前記無効にする要求は、シリアル・アドバンスト・テクノロジー・アタッチメント(SATA)TRIMコマンド、スモール・コンピュータ・システム・インターフェース(SCSI)UNMAPコマンド、マルチメディアカード(MMC)ERASEコマンド、およびセキュア・デジタル(SD)カードERASEコマンドのうちの1つである、請求項1に記載の方法。

30

【請求項10】

各第1レベル・マップ・エントリは、前記要求の前記範囲内に含まれる前記関連付けられた第2レベル・マップのいくつかの部分を追跡するように構成された少なくとも1つのTRIMアドレス範囲標識を備える請求項3に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明はソリッドステート・メディアにおけるマルチレベル・マッピングに関する。

【0002】

(関連出願の相互参照)

40

本出願は、参照によって教示の全体が本明細書に組み込まれている、2012年8月8日に出願した国際特許出願PCT/US2012/049905号の一部継続出願であり、PCT/US2012/049905号の出願日の利益を主張するものである。

【0003】

本出願は、参照によって教示の全体が本明細書に組み込まれている、2013年3月14日に出願した米国特許仮出願第61/783,555号の出願日の利益を主張するものである。

【0004】

本出願の主題は、参照によって教示の全体が本明細書に組み込まれている、2012年5月4日に出願した米国特許出願第13/464,433号、2012年8月4日に出願

50

した米国特許出願第 13 / 567, 025 号、2012 年 8 月 31 日に出願した米国特許出願第 13 / 600, 464 号、2012 年 12 月 28 日に出願した米国特許出願第 13 / 729, 966 号、および 2013 年 1 月 23 日に出願した米国特許出願第 13 / 748, 260 号と関係する。

【背景技術】

【0005】

フラッシュ・メモリは、或る特定のタイプの電氣的に消去可能なプログラマブル読み出し専用メモリ (EEPROM) である不揮発性メモリ (NVM) である。1 つの一般的に使用されているタイプのフラッシュ・メモリ技術が、NAND フラッシュ・メモリである。NAND フラッシュ・メモリは、1 セル当たり要求するチップ面積が小さく、通常、1 つまたは複数のバンクまたはプレーンに分割される。各バンクは、ブロックに分割され、各ブロックは、ページに分割される。各ページは、ユーザ・データまたは誤り訂正コード (ECC) 情報、あるいはユーザ・データと ECC 情報の両方を格納するための数バイトを含む。

【0006】

NAND デバイスに関して以下の 3 つの基本的な動作、すなわち、読み出し、書き込み、および消去が存在する。読み出し動作および書き込み動作は、ページごとに実行される。ページ・サイズは、一般に、 2^N バイトのユーザ・データ (加えて、ECC 情報のためのさらなる数バイト) であり、ただし、N は、整数であり、典型的なユーザ・データ・ページ・サイズは、例えば、1 ページ当たり 2,048 バイト (2KB)、4,096 バイト (4KB)、8,192 バイト (8KB)、またはそれ以上である。「読み出し単位」は、最少量のデータ、および NVM から読み取られ、ECC によって訂正されることが可能な対応する ECC 情報であり、通常、4K ビットから 32K ビットまでの範囲内であることが可能である (例えば、一般に、1 ページ当たり整数の読み出し単位が存在する)。ページは、通常、ブロックとして構成され、消去動作は、ブロックごとに実行される。典型的なブロック・サイズは、例えば、1 ブロック当たり 64 ページ、128 ページ、またはそれ以上である。ページは、通常、ブロック内の下位のアドレスから上位のアドレスに向って順次書き込まなければならない。より下位のアドレスには、そのブロックが消去されるまで再書き込みすることができない。各ページに関連付けられているのが、ECC 情報および / またはメモリ管理のために使用される他のメタデータの格納のために一般に使用される予備の領域 (通常、100 ~ 640 バイト) である。ECC 情報は、一般に、ページの中に格納されたユーザ・データの誤りを検出し、訂正するのに使用され、メタデータは、物理アドレスに論理アドレスをマップするため、および物理アドレスから論理アドレスにマップするために使用されることが可能である。複数のバンクを有する NAND フラッシュ・チップにおいて、各バンクからの複数のページが実質的に並行にアクセスされることを可能にするマルチバンク動作がサポートされることが可能である。

【0007】

NAND フラッシュ・メモリは、フローティング・ゲート・トランジスタから作られたメモリ・セルのアレイの中に情報を格納する。これらのトランジスタは、電荷とも呼ばれる電圧レベルを、外部電源が供給されることなしに、数ヶ月または数年のオーダの長期間にわたって保つ。シングル・レベル・セル (SLC) フラッシュ・メモリにおいて、各セルは、1 ビットの情報を格納する。マルチレベル・セル (MLC) フラッシュ・メモリにおいて、各セルは、フラッシュ・メモリの複数のセルのフローティング・ゲートに印加すべき複数のレベルの電荷の間で選択を行うことによって、1 セル当たり複数のビットを格納することができる。MLC NAND フラッシュ・メモリは、同一の数のトランジスタを使用して、より多くのビットが格納されることを可能にするようにシリアルで結合されたトランジスタ構成で 1 セル当たり複数の電圧レベルを使用する。このため、個々に考慮すると、各セルは、そのセルの中に格納された論理ビット値 (複数可) に対応する特定のプログラミングされた電荷を有し (例えば、SLC フラッシュの場合、0 または 1、MLC フラッシュの場合、00、01、10、11)、それらのセルは、各セルに関する 1 つ

または複数のしきい値電圧に基づいて読み取られる。しかし、1セル当たりのビット数を増やすことは、セル間干渉および保持ノイズを増加させて、読み出し誤りの尤度を高め、このため、システムのビット誤り率(BER)を高める。さらに、各セルの読み出ししきい値電圧は、例えば、読み出しディスタース、書き込みディスタース、保持損失、セルの経年変化、ならびにプロセス、電圧、および温度(PVT)の変動に起因して、NVMの動作時間につれて変化して、やはり、BERを高める。

【0008】

前述したとおり、通常のNVMは、新たなデータがブロックに書き込まれることが可能であるにはまず、そのブロックが消去されることを要求する。このため、1つまたは複数のNVMチップを使用するソリッドステート・ディスク(SSD)などのNVMシステムは、通常、「陳腐な」、つまり、古くなったデータを消去して、実現されるフラッシュ・メモリ容量を小さくする、大部分が古くなったデータでフラッシュ・メモリが一杯になることを防止する「ガーベジ・コレクション」プロセスを周期的に開始する。しかし、NVMブロックは、デバイス障害が生じるまで、限られた回数しか消去することができない。例えば、SLCフラッシュは、約100,000回しか消去できない可能性があり、MLCフラッシュは、約10,000回しか消去できない可能性がある。したがって、NVMの動作寿命につれて(例えば、NANDフラッシュの場合、定格のプログラミング/消去(P/E)サイクル数につれて)、NVMは消耗し、フラッシュ・メモリのブロックは、障害を生じて、使用不能になる。NVMにおけるブロック障害は、ハードディスク・ドライブ(HDD)におけるセクタ障害に類似する。通常のNVMシステムは、NVMのすべてのブロックにわたって可能な限り均一にP/Eサイクルを分散させる消耗平準化(wear-leveling)を実行することも可能である。このため、NVMシステムの寿命にわたって、不良なブロックの数が増加し、および/またはシステム・データ要件(例えば、論理-物理変換テーブル、ログ、メタデータ、ECCなど)のために使用されるストレージの量が増加するにつれ、全体的なストレージ容量が小さくなる可能性がある。このため、ガーベジ・コレクション・プロセス中にNVMに書き込まれるデータの量を減らすことが重要であり得る。

【0009】

ガーベジ・コレクション・プロセス中、依然として有効であるブロック内のユーザ・データが、背景プロセスにおいて記憶媒体上の新たなロケーションに移動される。「有効な」ユーザ・データは、ホスト・デバイスがこのデータをもはや使用していない場合でも、少なくとも1回、書き込まれたことのある任意のアドレスである可能性がある。ガーベジ・コレクション中に再書き込みされる「有効」ではあるが、もはや必要とされないデータの量を減らすのに、いくつかのストレージ・プロトコルは、NVMが、以前に保存されたデータのブロックを不要または無効と指定できるようにして、それらのブロックがガーベジ・コレクション中に移動されないようにし、それらのブロックが新たなデータを格納するのに利用可能にされ得るようにするコマンドをサポートする。そのようなコマンドの例が、SATA TRIM(データ・セット管理)コマンド、SCSI UNMAPコマンド、マルチメディアカード(MultiMediaCard)(MMC)ERASEコマンド、およびセキュア・デジタル(SD)カードERASEコマンドである。一般に、そのようなコマンドは、完全にトリミングされたNVMが、同一のタイプの新規に製造された(すなわち、空の)NVMの性能に近い性能を有するようにNVM性能を向上させる。しかし、多数のブロックに関してこれらのコマンドを一度に実行することは、時間がかかり、NVMの動作効率を低下させる可能性がある。

【発明の概要】

【課題を解決するための手段】

【0010】

この概要は、詳細な説明において後段でさらに説明される選定された概念を、簡略化された形態で概説するために与えられる。この概要は、主張される主題の重要な特徴、または不可欠な特徴を特定することは意図しておらず、主張される主題の範囲を限定するのに

10

20

30

40

50

使用されることも意図していない。

【0011】

説明される実施形態は、ソリッドステート・メディアのためのメディア・コントローラを提供する。このメディア・コントローラは、少なくとも1つの論理アドレスと、アドレス範囲とを含む要求をホスト・デバイスから受け取る制御プロセッサを含む。この要求に応答して、制御プロセッサは、受け取られた要求が無効にする要求 (invalidating request) であるかどうかを判定する。受け取られた要求タイプが無効にする要求である場合、制御プロセッサは、メディア・コントローラのマップを使用して、その論理アドレスおよび範囲に関連付けられたマップの1つまたは複数のエントリを特定する。これらのマップ・エントリのそれぞれに関連付けられたマップの中の標識が、これらのマップ・エントリが無効にされるべきことを示すように設定される。制御プロセッサは、無効にする要求が完了したという確認応答をホスト・デバイスに対して行い、メディア・コントローラのアイドル・モードにおいて、無効にされるべきマップ・エントリに基づいて空きスペース・カウントを更新する。無効にされたマップ・エントリに関連付けられた物理アドレスが、ホスト・デバイスからのその後の要求に関して再使用されるように利用可能にされる。

10

説明される実施形態の他の態様、特徴、および利点が、後段の詳細な説明、添付の特許請求の範囲、および添付の図面から、より完全に明白となる。図面では、同様の参照符号は、同様の、または同一の要素を識別する。

【図面の簡単な説明】

20

【0012】

【図1】例示的な実施形態によるフラッシュ・メモリ・ストレージ・システムを示すブロック図である。

【図2】単一の標準のフラッシュ・メモリ・セルを示す例示的な機能ブロック図である。

【図3】例示的な実施形態による例示的なNAND MLCフラッシュ・メモリ・セルを示す図である。

【図4】図1のフラッシュ・メモリ・ストレージ・システムのフラッシュ・メモリの例示的な構成を示すブロック図である。

【図5】図1のフラッシュ・メモリ・ストレージ・システムの論理ブロック図番号 (LBA) の論理ページ番号 (LPN) 部分の例示的なマッピングを示すブロック図である。

30

【図6】図1のフラッシュ・メモリ・ストレージ・システムの例示的な2レベル・マッピング構造を示すブロック図である。

【図7】図1のフラッシュ・メモリ・ストレージ・システムによって使用される例示的なマップ・ページ・ヘッダを示すブロック図である。

【図8】図1のフラッシュ・メモリ・ストレージ・システムによって使用されるメガTRIM動作を示す例示的な流れ図である。

【発明を実施するための形態】

【0013】

説明される実施形態は、ソリッドステート・メディアに関するメディア・コントローラを提供する。このメディア・コントローラは、少なくとも1つの論理アドレスと、アドレス範囲とを含む要求をホスト・デバイスから受け取る制御プロセッサを含む。この要求に応答して、制御プロセッサは、受け取られた要求が無効にする要求であるかどうかを判定する。受け取られた要求タイプが無効にする要求である場合、制御プロセッサは、メディア・コントローラのマップを使用して、その論理アドレスおよび範囲に関連付けられたマップの1つまたは複数のエントリを特定する。これらのマップ・エントリのそれぞれに関連付けられたマップの中の標識が、これらのマップ・エントリが無効にされるべきことを示すように設定される。制御プロセッサは、無効にする要求が完了したという確認応答をホスト・デバイスに対して行い、メディア・コントローラのアイドル・モードにおいて、無効にされるべきマップ・エントリに基づいて空きスペース・カウントを更新する。無効にされたマップ・エントリに関連付けられた物理アドレスが、ホスト・デバイスからのそ

40

50

の後の要求に関して再使用されるように利用可能にされる。

【 0 0 1 4 】

表 1 が、説明される実施形態を理解する助けとして本明細書全体で使用される頭字語のリストを定義する。

【表 1】

表 1			
BER	ビット誤り率	BUS	ブロック使用されるスペース
CAM	連想メモリ	ECC	誤り訂正コード
eDRAM	組み込みダイナミック・ランダム・アクセス・メモリ	EEPROM	電氣的に消去可能なプログラマブル読み出し専用メモリ
FLM	第1レベル・マップ	HDD	ハードディスク・ドライブ
IC	集積回路	I/O	入出力
LBA	論理ブロック・アドレス	LDPC	低密度パリティ検査
LLR	対数尤度比	LPN	論理ページ番号
LSB	最下位ビット	LRU	未使用時間が最長の
MLC	マルチレベル・セル	MLM	マルチレベル・マップ
MMC	マルチメディアカード	MSB	最下位ビット
NVM	不揮発性メモリ	OOS	スペース不足
OP	オーバ・プロビジョニング	PCI-E	ペリフェラル・コンポーネント・インターコネクト・エクスプレス
P/E	プログラミング／消去	PVT	プロセス、電圧、温度
SAS	シリアル・アタッチトSCSI	SATA	シリアル・アドバンスト・テクノロジー・アタッチメント
SCSI	スモール・コンピュータ・システム・インターフェース	SD	セキュア・デジタル
SLC	シングル・レベル・セル	SLM	第2レベル・マップ
SoC	システム・オン・チップ	SRAM	スタティック・ランダム・アクセス・メモリ
SRIO	シリアル・ラビッド入出力	SSD	ソリッドステート・ディスク
TBP	処理されるべき	USB	ユニバーサル・シリアル・バス

【 0 0 1 5 】

図 1 は、不揮発性メモリ（NVM）ストレージ・システム 100 のブロック図を示す。NVM ストレージ・システム 100 は、メディア・コントローラ 120 に結合されたメディア 110 を含む。メディア 110 は、NAND フラッシュ・ソリッドステート・ディスク（SSD）として、ハードディスク・ドライブ（HDD）などの磁気記憶媒体として、またはハイブリッド・ソリッドステート・磁気システムとして実装されることが可能である。図 1 には示されないものの、メディア 110 は、複数のフラッシュ・チップなどの 1 つまたは複数の物理メモリ（例えば、不揮発性メモリ、NVM）を、通常、含むことが可能である。図 1 に示されるとおり、メディア 110 とメディア・コントローラ 120 とがひとまとまりとして SSD 101 である。メディア・コントローラ 120 は、ソリッドステート・コントローラ 130 と、制御プロセッサ 140 と、バッファ 150 と、入出力インターフェース 160 とを含む。メディア・コントローラ 120 は、メディア 110 と、通信リンク 170 に結合されたホスト・デバイス 180 との間のデータの転送を制御する。メディア・コントローラ 120 は、システム・オン・チップ（SoC）として、または

他の集積回路（ＩＣ）として実装されることが可能である。ソリッドステート・コントローラ１３０は、メディア１１０内のメモリ・ロケーションにアクセスするのに使用されることが可能であり、通常、低レベルのデバイス固有の動作を実装して、メディア１１０とのインターフェースをとる。バッファ１５０は、制御プロセッサ１４０のためのキャッシュの役割、および／またはソリッドステート・メディア１１０とホスト・デバイス１８０との間の動作のための読み出し／書き込みバッファの役割をするように使用されるＲＡＭバッファであることが可能である。例えば、データが、一般に、入出力インターフェース１６０とリンク１７０とを介した、ソリッドステート・メディア１１０とホスト・デバイス１８０との間の転送中にバッファ１５０の中に一時的に格納されることが可能である。バッファ１５０は、通信リンク１７０のデータ転送サイズとメディア１１０の格納単位サイズ（例えば、読み出し単位サイズ、ページ・サイズ、セクタ・サイズ、またはマップされた単位サイズ）との差を勘案するようにデータをグループ化する、または分割するのに使用されることが可能である。バッファ１５０は、スタティック・ランダム・アクセス・メモリ（ＳＲＡＭ）として、またはメディア・コントローラ１２０内部の組み込みダイナミック・ランダム・アクセス・メモリ（ｅＤＲＡＭ）として実装されることが可能であるが、バッファ１５０は、ダブル・データ・レート（例えば、ＤＤＲ－３）ＤＲＡＭとして通常、実装されることが可能な、メディア・コントローラ１２０外部のメモリ（図示せず）を含むことも可能である。

10

【００１６】

制御プロセッサ１４０が、ソリッドステート・コントローラ１３０と通信して、メディア１１０の中のデータに対するデータ・アクセス（例えば、読み出し動作または書き込み動作）を制御する。制御プロセッサ１４０は、１つまたは複数のＰｅｎｔｉｕｍ（登録商標）プロセッサ、Ｐｏｗｅｒ ＰＣ（登録商標）プロセッサ、Ｔｅｎｓｉｌｉｃａ（登録商標）プロセッサ、またはＡＲＭプロセッサとして、あるいは様々なプロセッサ・タイプの組み合わせとして実装されることが可能である（Ｐｅｎｔｉｕｍ（登録商標）は、インテル・コーポレーション（Ｉｎｔｅｌ Ｃｏｒｐｏｒａｔｉｏｎ）の登録商標であり、Ｔｅｎｓｉｌｉｃａ（登録商標）は、テンシリカ社（Ｔｅｎｓｉｌｉｃａ Ｉｎｃ．）の商標であり、ＡＲＭプロセッサは、ＡＲＭホールディングス（ＡＲＭ Ｈｏｌｄｉｎｇｓ， ｐｌｃ）によるものであり、Ｐｏｗｅｒ ＰＣ（登録商標）は、ＩＢＭ社の登録商標である）。図１では単一のプロセッサとして示されるものの、制御プロセッサ１４０は、複数のプロセッサ（図示せず）によって実装されるとともに、説明される実施形態により、しきい値最適化された（ｔｈｒｅｓｈｏｌｄ ｏｐｔｉｍｉｚｅｄ）動作を実行することを含む動作のために必要とされるソフトウェア／ファームウェアを含むことも可能である。制御プロセッサ１４０は、メディア１１０に書き込まれるデータに関するＬＤＰＣ符号化、およびメディア１１０から読み出されるデータに関する復号化を実行する低密度パリティ検査（ＬＤＰＣ）符号器／復号器（コーデック）１４２と通信状態にある。また、制御プロセッサ１４０は、ホスト動作の論理アドレス（例えば、読み出し／書き込み動作のための論理ブロック・アドレス（ＬＢＡ）など）とメディア１１０上の物理アドレスとの間の変換を行うのに使用されるマップ１４４とも通信状態にある。本明細書で使用されるように、ＬＢＡという用語は、ＨＰＡ（ホスト・ページ・アドレス）と同義である。

20

30

40

【００１７】

通信リンク１７０が、ＮＶＭシステム１００とインターフェースをとるコンピュータ・システムであることが可能なホスト・デバイス１８０と通信するのに使用される。通信リンク１７０は、カスタム通信リンクであることが可能であり、あるいは、例えば、スモール・コンピュータ・システム・インターフェース（「ＳＣＳＩ」）プロトコル・バス、シリアル・アタッチト（Ｓｅｒｉａｌ Ａｔｔａｃｈｅｄ）ＳＣＳＩ（「ＳＡＳ」）プロトコル・バス、シリアル・アドバンスト・テクノロジー・アタッチメント（Ｓｅｒｉａｌ Ａｄｖａｎｃｅｄ Ｔｅｃｈｎｏｌｏｇｙ Ａｔｔａｃｈｍｅｎｔ）（「ＳＡＴＡ」）プロトコル・バス、ユニバーサル・シリアル・バス（Ｕｎｉｖｅｒｓａｌ Ｓｅｒｉａｌ Ｂｕｓ）（「ＵＳＢ」）、イーサネット（Ｅｔｈｅｒｎｅｔ）リンク、ＩＥＥＥ ８０２．

50

11リンク、IEEE802.15リンク、IEEE802.16リンク、ペリフェラル・コンポーネント・インターコネクト・エクスプレス(Peripheral Component Interconnect Express)(「PCI-E」)リンク、シリアル・ラピッド(Serial Rapid)入出力(「SRIO」)リンク、または周辺デバイスをコンピュータに接続するための他の任意の類似したインターフェース・リンクなどの標準通信プロトコルにより動作するバスであることが可能である。

【0018】

図2は、ソリッドステート・メディア110内に見られることが可能な単一のフラッシュ・メモリ・セルの例示的な機能ブロック図を示す。フラッシュ・メモリ・セル200は、2つのゲートを有するMOSFETである。ワード線制御ゲート230は、フローティング・ゲート240の上に配置される。フローティング・ゲート240は、ワード線制御ゲート230、およびMOSFETチャネルから絶縁層によって絶縁され、MOSFETチャネルは、Nチャネル250および260と、Pチャネル270とを含む。フローティング・ゲート240は、電氣的に絶縁されるため、フローティング・ゲート240に加えられた電荷は、留まり、通常、数ヶ月にもわたって大きく放電されない。フローティング・ゲート240が電荷を保持する場合、フローティング・ゲート240は、セルのしきい値電圧を変えるワード線制御ゲート230による電界を部分的にキャンセルする。しきい値電圧は、チャネルを導通させるように制御ゲート230に印加される電圧の量である。チャネルが導通することによって、例えば、フローティング・ゲート240上の電荷を検知することによって、セルの中に格納された値が特定される。

【0019】

図3は、ソリッドステート・メディア110内に見られることが可能である例示的なNANDMLCフラッシュ・メモリ・ストリング300を示す。図3に示されるとおり、フラッシュ・メモリ・ストリング300は、ドレインからソースまで直列に接続された、1つまたは複数のワード線トランジスタ200(2)、200(4)、200(6)、200(8)、200(10)、200(12)、200(14)、および200(16)(例えば、8つのフラッシュ・メモリ・セル)、ならびにビット線選択トランジスタ304を含むことが可能である。この直列接続は、ビット線322が完全に引き下げられるように対応するゲートを高に駆動することによって、接地選択トランジスタ302、ワード線トランジスタ200(2)、200(4)、200(6)、200(8)、200(10)、200(12)、200(14)、および200(16)、ならびにビット線選択トランジスタ304がすべて「オンにされる」ようになっている。オンにされるワード線トランジスタ200(2)、200(4)、200(6)、200(8)、200(10)、200(12)、200(14)、および200(16)の数(または、これらのトランジスタが線形領域で動作しているか、飽和領域で動作しているか)を変えることにより、MLCストリング300が複数の電圧レベルを実現することができるようになり得る。通常のMLC NANDフラッシュは、フローティング・ゲートを有する64のトランジスタの「NANDストリング」(例えば、図3に示されるとおり)を使用することが可能である。書き込み動作中、書き込まれるべきワード線位置でNANDストリングに高い電圧が印加される。読み出し動作中、所望される読み出しロケーションに対応するトランジスタを除いて、NANDストリングにおけるすべてのトランジスタのゲートに電圧が印加される。所望される読み出しロケーションは、フローティング・ゲートを有する。

【0020】

本明細書で説明されるとおり、SLCNANDフラッシュとMLCNANDフラッシュの両方において、各セルは、読み出ししきい値電圧レベルとの比較などによって検知され得る電荷電圧レベル(例えば、アナログ信号)を有する。メディア・コントローラは、電荷電圧レベルを読み出し、セルの対応するバイナリ値を検出するのに使用される所与の数の所定の電圧しきい値を有することが可能である。例えば、MLCNANDフラッシュに関して、3つのしきい値(0.1、0.2、0.3)が存在する場合、セル電圧レベルが0.0セル電圧<0.1である際、そのセルは、[00]という値を有すると検

出されることが可能である。セル電圧レベルが 0 . 1 セル電圧 < 0 . 2 である場合、値は、[1 0] であることが可能であり、以下同様である。このため、測定されたセル・レベルは、そのセル・レベルが、2 つのしきい値の間にあると判定され、およびそのセル・レベルが検出され得るまで、通常、1 つずつのしきい値と比較されることが可能である。このため、検出されたデータ値が、メモリ・コントローラ 1 2 0 の復号器に供給されて、検出された値（例えば、誤り訂正コードと一緒に）が復号されて、ホスト・デバイス 1 8 0 に供給されるべきデータにされる。

【 0 0 2 1 】

図 4 は、図 1 のソリッドステート・メディア 1 1 0 の例示的な構成のブロック図を示す。図 4 に示されるとおり、メディア 1 1 0 は、スペース不足 (O u t - o f - S p a c e) (O O S) 条件が生じるのを防止するようにオーバ・プロビジョニング (o v e r - p r o v i s i o n i n g) (O P) を伴って実装されることが可能である。図 4 に示されるとおり、O P は、3 つの様態で実現されることが可能である。第 1 に、S S D 製造業者が、通常、「G B」という用語を用いて 1 0 進数のギガバイトを表すが、1 0 進数のギガバイト (1 , 0 0 0 , 0 0 0 , 0 0 0 バイト、つまり、 10^9 バイト) と 2 進数のギガバイト (1 , 0 7 3 , 7 4 1 , 8 2 4 バイト、つまり、 2^{30} バイト) は等しくない。したがって、S S D の物理容量は、2 進数の G B に基づくので、S S D の論理容量が 1 0 進数の G B に基づく場合、S S D は、7 . 3 7 % (例えば、 $[(2^{30} - 10^9) / 10^9]$) の組み込まれた O P を有することが可能である。このことが、図 4 に「7 . 3 7 %」O P 4 0 2 として示される。しかし、この O P のいくらか、例えば、合計容量の 2 ~ 4 % は、N A N D フラッシュの不良なブロック (例えば、欠陥) のために失われる可能性がある。第 2 に、O P は、ホスト・デバイス 1 8 0 が利用できない、システムが使用するための特定の量の物理メモリを取って置くことによって実装されることも可能である。例えば、製造業者が、1 2 8 G B の合計物理容量に基づいて、1 0 0 G B、1 2 0 G B、または 1 2 8 G B の論理容量を有する製造業者の S S D に関する仕様を公開して、これにより、場合により、それぞれ 2 8 %、7 %、または 0 % の例示的な O P を実現することが可能である。このことが、図 4 に静的 O P (「0 から 2 8 + %」) 4 0 4 として示される。

【 0 0 2 2 】

第 3 に、いくつかのストレージ・プロトコル (例えば、S A T A) は、ホスト・デバイス 1 8 0 が、以前に保存されたデータのブロックを不要または無効と指定できるようにして、ガーベジ・コレクション中に N V M システム 1 0 0 がそれらのブロックを保存しないようにする「T R I M」コマンドをサポートする。T R I M コマンドに先立って、ホスト・デバイス 1 8 0 が或るファイルを消去した場合、そのファイルは、ホスト・デバイス・レコードから除去されているが、N V M システム 1 0 0 の実際の内容は、実際には消去されておらず、このことにより、N V M システム 1 0 0 が、ガーベジ・コレクション中に無効なデータを保持することになり、これにより、N V M 容量が小さくなる。T R I M コマンドを使用することによる効率的なガーベジ・コレクションによる O P が、図 4 に動的 O P 4 0 6 として示される。動的 O P 4 0 6 およびユーザ・データ 4 0 8 が、ホスト・デバイス 1 8 0 の活性のデータを含むメディア 1 1 0 の領域を形成する一方で、O P 領域 4 0 2 および 4 0 4 は、ホスト・デバイス 1 8 0 の活性のデータを含まない。T R I M コマンドは、オペレーティング・システムが、データのいずれのページが、ユーザまたはオペレーティング・システム自らによる消去のために今では無効であるかを S S D に通知することを可能にする。削除動作中、O S は、削除されたセクタに新たなデータのために空いているという印を付け、もはや有効でないという印が付けられるべき削除されたセクタに関連する S S D の論理ブロック・アドレス (L B A) の 1 つまたは複数の範囲を指定する T R I M コマンドを送る。

【 0 0 2 3 】

T R I M コマンドを実行した後、メディア・コントローラは、トリミングされた L B A からのデータを、ガーベジ・コレクション中に再配置せず、メディアに対する書き込み動作の回数を減らし、これにより、書き込み拡大を小さくするとともに、ドライブ寿命を増

10

20

30

40

50

加させる。T R I M コマンドは、一般に、T R I M コマンドが影響を及ぼすデータを不可逆に削除する。T R I M コマンドの例が、S A T A T R I M (データ・セット管理) コマンド、S C S I U N M A P コマンド、マルチメディアカード (M M C) E R A S E コマンド、およびセキュア・デジタル (S D) カード E R A S E コマンドである。一般に、T R I M は、完全にトリミングされた S S D が、同一のタイプの新規に製造された (すなわち、空の) S S D の性能に近い性能を有するように S S D 性能を向上させる。

【0024】

一般に、メディア・コントローラ 120 は、ホスト・デバイス 180 から受け取られたコマンドを実行する。これらのコマンドのうちの少なくともいくつかは、データがホスト・デバイス 180 から送られて、メディア 110 にデータを書き込み、またはメディア 110 からデータを読み出し、読み出されたデータをホスト・デバイス 180 に送る。メディア・コントローラ 120 は、1 つまたは複数のデータ構造を使用して、論理メモリ・アドレス (例えば、ホスト動作に含められた L B A) をメディアの物理アドレスにマップする。L B A が S S D の中に書き込まれる際、L B A は、一般に、毎回、異なる物理ロケーションに書き込まれ、各書き込みは、その L B A のデータが不揮発性メモリ (例えば、メディア 110) の中でどこに存在するかを記録するようにマップを更新する。例えば、2012 年 8 月 8 日に出願した国際特許出願 P C T / U S 2 0 1 2 / 0 4 9 9 0 5 号で説明されるようなシステムにおいて、メディア・コントローラ 120 は、リーフ・レベルと、より上位の 1 つまたは複数のレベルとを含むマルチレベル・マップ構造 (例えば、マップ 144) を使用する。リーフ・レベルは、1 つまたは複数のエントリをそれぞれが有するマップ・ページを含む。接続されたメディア (例えば、メディア 110) の L B A などの論理アドレスが、マルチレベル・マップ構造の中で調べられて、リーフ・レベル・ページのうちの特定の 1 つのページの中の複数のエントリのうちの対応する 1 つが特定される。L B A の対応するエントリは、その L B A に関連付けられたメディア 110 の物理アドレスなどの、その L B A に関連付けられた情報を含む。一部の実施例において、対応するエントリは、その対応するエントリが有効であるか、または無効であるかについて、および、オプションとして、その L B A に対して T R I M コマンドが実行され (「トリミングされ」) ているか、または全く書き込まれていないかについての指示をさらに備える。例えば、無効なエントリは、関連する L B A がトリミングされているかどうかなどの情報を、無効なエントリの物理ロケーション部分の中に符号化することができる。

【0025】

L B A を調べるのを迅速化するのに、リーフ・レベル・ページのうちの少なくともいくつかのページのキャッシュ (図示せず) が保持されることが可能である。一部の実施形態において、マップ・データ構造の少なくとも一部分が、ホスト・デバイス 180 からは見えないプライベート・ストレージのために使用される (例えば、メディア・コントローラ 120 のログ、統計、マッピング・データ、または他のプライベート / 制御データを格納するのに)。

【0026】

本明細書で説明されるとおり、マップ 144 は、ホスト・デバイス 180 によって使用される論理データ・アドレス指定と、メディア 110 によって使用される物理データ・アドレス指定の間で変換を行う。例えば、マップ 144 は、ホスト・デバイス 180 によって使用される L B A と、メディア 110 の 1 つまたは複数のフラッシュ・チップ (f l a s h d i e s) のブロック・アドレスおよび / またはページ・アドレスの間で変換を行う。例えば、マップ 144 は、論理アドレスと物理アドレスの間の変換を実行する、または調べる 1 つまたは複数のテーブルを含むことが可能である。

【0027】

各 L B A に関連するデータは、固定の、圧縮されていないサイズで、またはそれぞれの圧縮されたサイズで、メディア 110 の対応する物理アドレスにおいて格納される。本明細書で説明されるとおり、読み出し単位は、メディア 110 のページの一部分などの、独立に読み出し可能であるメディア 110 の最も細かい粒度である。読み出し単位は、誤り

10

20

30

40

50

訂正コード (ECC) によって保護されるすべてのデータとともに、ECC の検査ビットおよび / または冗長性データを含む (またはそのようなビットおよび / またはデータに対応する) ことが可能である。図 5 は、マップ 144 による LBA の LPN 部分をマップする実施形態の選択された詳細を示す。図 5 に示されるとおり、LBA 506 は、論理ページ番号 (LPN) 502 と、論理オフセット 504 とを含む。マップ 144 は、LPN 502 を、読み出し単位アドレス 508、および読み出し単位での長さ 510 (および、場合により、省略記号で示される他のマップ・データ) を含むマップ・データ 512 に変換する。マップ・データ 512 は、通常、マップ 144 のマップ・テーブルの中にマップ・エントリとして格納されることが可能である。マップ 144 は、通常、システム 100 が能動的に使用中である各 LPN につき 1 つのマップ・エントリを保持することが可能である。図示されるとおり、マップ・データ 512 は、読み出し単位アドレス 508 と、読み出し単位での長さ 510 とを含む。一部の実施形態において、長さおよび / またはスパンは、LPN に関連付けられたデータの長さを、読み出し単位での長さ 510 の全部 (または一部分) におけるスパンからのオフセットとして格納することなどによって、符号化されて格納される。スパン (または読み出し単位での長さ) は、LPN に関連付けられたデータを取得するのに読み出されるべき読み出し単位の数に指定するのに対して、長さ (LPN に関連付けられたデータの) は、SSD の各ブロックの中の使用されるスペースの量を追跡するブロック使用されるスペース (Block Used Space) (BUS) などの統計のために使用される。通常、長さは、スパンと比べて、より細かい粒度を有する。

10

20

【0028】

一部の実施形態において、第 1 の LPN が、第 1 のマップ・エントリに関連付けられ、第 2 の LPN (第 1 の LPN とは異なるが、第 1 の LPN によって参照される論理ページと同一サイズの論理ページを参照する) が、第 2 のマップ・エントリに関連付けられ、第 1 のマップ・エントリの読み出し単位でのそれぞれの長さは、第 2 のマップ・エントリの読み出し単位でのそれぞれの長さとは異なる。そのような実施形態において、同一の時点で、第 1 の LPN が、第 1 のマップ・エントリに関連付けられ、第 2 の LPN が、第 2 のマップ・エントリに関連付けられ、第 1 のマップ・エントリのそれぞれの読み出し単位アドレスは、第 2 のマップ・エントリのそれぞれの読み出し単位アドレスと同一であり、したがって、第 1 の LPN に関連付けられたデータと第 2 の LPN に関連付けられたデータはともに、メディア 110 の同一の物理読み出し単位の中に格納される。

30

【0029】

様々な実施形態において、マップ 144 は、1 レベル・マップ、第 1 レベル・マップ (FLM) と、ホスト・プロトコルの LBA をメディア 110 内の物理格納アドレスに関連付ける 1 つまたは複数の第 2 レベル (またはより下位レベルの) マップ (SLM) とを含む 2 レベル・マップのいずれかである。例えば、図 6 に示されるとおり、例えば、マップ 144 内で、FLM 610 が、メディア・コントローラ 120 内にオンチップで保持される。一部の実施形態において、FLM 610 の不揮発性 (ただし、わずかに古い) コピーもまた、メディア 110 上に格納される。FLM 610 の中の各エントリは、実質的に、SLM ページ (例えば、SLM 616 のうちの 1 つ) に対するポインタである。SLM 616 は、メディア 110 の中に格納され、一部の実施形態において、これらの SLM のうちのいくつかは、マップ 144 のオンチップ SLM キャッシュ (例えば、SLM キャッシュ 608) の中にキャッシュされる。FLM 610 の中のエントリは、対応する第 2 レベル・マップ・ページ (例えば、SLM キャッシュ 608 またはメディア 110 の中の) のアドレス (および、場合により、アドレスまたは他の情報のデータ長 / 範囲) を含む。図 6 に示されるとおり、マップ・モジュール 144 は、所与の LBA (例えば、LBA 602) の第 1 の関数 (例えば、その LBA を、第 2 レベル・マップ・ページのそれぞれの中に含まれたエントリの固定の数で割った際に得られる商) を、SLM 616 として示される複数の第 2 レベル・マップ (SLM) のうちの 1 つの中のそれぞれのアドレスに関連付ける第 1 レベル・マップ (FLM) 610 を有する 2 レベル・マップを含むことが

40

50

可能であり、各 S L M は、L B A の第 2 の関数（例えば、その L B A を、第 2 レベル・マップ・ページのそれぞれの中に含まれたエントリの固定の数で割った際に得られる剰余）を、その L B A に対応するメディア 1 1 0 内のそれぞれのアドレスに関連付ける。

【 0 0 3 0 】

例えば、図 6 に示されるとおり、トランスレータ 6 0 4 が、或るホスト動作に対応する L B A (L B A 6 0 2)（例えば、ホスト 1 8 0 からの、メディア 1 1 0 上の対応する L B A に対して読み出しまたは書き込みを行う要求）を受け取る。トランスレータ 6 0 4 は、例えば、L B A 6 0 2 を、対応する S L M ページ 6 1 6 の各ページの中のエントリの整数で割ることによって、L B A 6 0 2 を F L M インデックス 6 0 6 および S L M ページ・インデックス 6 1 4 に変換する。説明される実施形態において、F L M インデックス 6 0 6 は、この除算演算の商であり、S L M ページ・インデックス 6 1 4 は、この除算演算の剰余である。この除算演算を使用することは、S L M ページ 6 1 6 が、2 のべき乗でない数のエントリを含むことを可能にし、このことは、S L M ページ 6 1 6 のサイズが小さくされることを許して、S L M ページ 6 1 6 を更新する書き込み動作に起因するメディア 1 1 0 の書き込み拡大を小さくすることが可能である。F L M インデックス 6 0 6 は、F L M 6 1 0 の中のエントリを一意的に識別するのに使用され、このエントリは、S L M ページ 6 1 6 のうちの 1 つに対応する S L M ページ・インデックス (6 1 4) を含む。6 1 2 で示されるとおり、F L M エントリの S L M ページ・インデックスに対応する S L M ページが S L M キャッシュ 6 0 8 の中に格納される実施形態において、F L M 6 1 0 が、L B A 6 0 2 に対応するメディア 1 1 0 の物理アドレスを戻すことが可能である。S L M ページ・インデックス 6 1 4 は、S L M 6 1 6 の中のエントリを一意的に識別するのに使用され、このエントリは、6 1 8 で示されるとおり、L B A 6 0 2 に対応するメディア 1 1 0 の物理アドレスに対応する。S L M 6 1 6 のエントリは、読み出し単位アドレス（例えば、フラッシュ・ページの E C C 訂正可能な下位単位のアドレス）、および読み出し単位の長さとして符号化されることが可能である。

【 0 0 3 1 】

S L M ページ 6 1 6（または、より低レベルのマルチレベル・マップ (M L M) 構造）はすべて、同一数のエントリを含むことが可能であり、または S L M ページ 6 1 6（または、より低レベルの M L M 構造）のそれぞれは、異なる数のエントリを含むことが可能である。さらに、S L M ページ 6 1 6（または、より低レベルの M L M 構造）のエントリは、同一の粒度であることが可能であり、または粒度は、S L M ページ 6 1 6（または、より低レベルの M L M 構造）のそれぞれに関して設定されることが可能である。例示的な実施形態において、F L M 6 1 0 は、1 エントリ当たり 4 K B の粒度を有し、S L M ページ 6 1 6（または、より低レベルの M L M 構造）のそれぞれは、1 エントリ当たり 8 K B の粒度を有する。このため、例えば、F L M 6 1 0 の中の各エントリは、5 1 2 B の L B A の整列された 8 セクタ (4 K B) 領域に関連付けられ、S L M ページ 6 1 6 のうちの 1 つのページの中の各エントリは、5 1 2 B の L B A の整列された 1 6 セクタ (8 K B) 領域に関連付けられる。

【 0 0 3 2 】

一部の実施形態において、F L M 6 1 0（または M L M 構造のより高レベルのマップ）のエントリは、対応するより低レベルのマップ・ページのフォーマット情報を含む。図 7 は、例示的な F L M 7 0 0 のブロック図を示す。図示されるとおり、F L M 7 0 0 の N 個のエントリ 7 0 1 のそれぞれが、対応するより低レベルのマップ・ページのフォーマット情報を含む。図示されるとおり、F L M 7 0 0 は、S L M ページ粒度 7 0 2 と、読み出し単位物理アドレス範囲 7 0 4 と、各 L B A に関するデータ・サイズ 7 0 6 と、データ無効標識 7 0 8 と、T R I M 動作進行中標識 7 1 0 と、T R I M L B A 範囲 7 1 2 と、処理されるべき (T o - B e - P r o c e s s e d) (T B P) 標識 7 1 4 とを含むことが可能である。また、他のメタデータ（図示せず）が含まれることも可能である。マップ・ページ粒度 7 0 2 は、F L M 7 0 0 のエントリに対応する S L M ページの粒度を示す。読み出し単位物理アドレス範囲 7 0 4 は、例えば、開始読み出し単位アドレスおよびスパン

として、FLM700のエントリに対応するSLMページの読み出し単位（複数可）の物理アドレス範囲を示す。各LBA706に関するデータ・サイズは、関連するLBAのデータを得るのに読み出されるべき読み出し単位の数、またはFLM700のエントリに対応するSLMページに関してメディア110の中に格納された関連するLBAのデータのサイズを示す。データ無効標識708は、関連するLBAのデータが既にトリミングされている、またはそれ以外で無効にされているなどのために、関連するLBAのデータが、メディア110の中に存在しないことを示す。代替の実施形態において、データ無効標識は、読み出し単位物理アドレス範囲704の一部として符号化されることが可能である。後段でより詳細に説明されるとおり、TRIM動作進行中標識710は、TRIM LBA範囲712によって示されるLBAに対してTRIM動作が進行中であることを示す。10
一部の実施形態において、TRIM動作進行中標識710は、TRIM LBA範囲712の一部として符号化されることが可能である。TBP標識714は、マップ・ページに関連するLBAが既に無効にされている（例えば、トリミングされているようにホスト180に見える）が、それらのLBAが、新たなデータを書き込むのにまだ利用できない場合、そのことを示す。より高レベルのマップ・エントリに無効の印を付けるのとは異なり、より高レベルのマップ・エントリのTBPビットを設定することは、そのより高レベルのマップ・エントリの中に格納されたより低レベルのマップ・ページの物理アドレスが無効であることは暗示せず、つまり、そのより低レベルのマップ・ページがBUS更新のために処理されるまで、その物理アドレスは、要求され、そのより低レベルのマップ・ページ自体、割り当て解除され得ない。このため、より低レベルのマップ・ページは、以下の20
3つの状態、すなわち、無効、有効、またはTBPのうちの1つに入っていることが可能である。

【0033】

本明細書で説明されるマルチレベル・マップ（MLM）構造を使用するSSDは、複数のリーフ・レベル・マップ単位にわたる改良されたTRIM動作を可能にする。このため、標準のTRIM動作の場合のように個々のLBAエントリを無効にする代わりに、改良されたTRIM動作は、MLM構造のより高いマップ・レベルにおけるリーフ単位全体を無効にすることができる。このことは、メディア・コントローラ120に結合されたホスト・デバイスから見てTRIM動作の待ち時間を短縮して、より高いシステム性能を有利に可能にする。しかし、リーフ・レベル・マップの中のトリミングされた個々のLBAエントリを単に破棄することは、トリミングされたLBAが依然としてブロック使用されるスペース（BUS）に寄与しているように見えるので、BUSアカウンティングの不正確さを被る可能性がある。BUSカウントは、所与のブロック、または所与のグループのブロック（例えば、最も少ないBUSを有するブロック）に対していつガーベジ・コレクションを実行すべきかを判定して、ガーベジ・コレクション書き込み拡張を小さくする1つの方法として、フラッシュ・ブロックごとに、またはグループのフラッシュ・ブロックごとになど、SSDの不揮発性メモリの各領域に関して、メディア・コントローラ120によってメディア110の中に保持される。このため、BUSが不正確であることは、不正確なガーベジ・コレクション、および/またはメディア110に対するより多くの回数の書き込みをもたらして、書き込み拡大を大きくするとともに、SSD寿命を短くする可能性30
がある。改良されたTRIM動作は、ホスト・デバイスに対してTRIM動作の確認応答を行った後に、背景でBUSカウントを更新することによってBUSの正確さも維持しながら、LBAの迅速なトリミングを実行することができる。

【0034】

説明される実施形態において、TRIM動作は、トリミングされたすべてのLBAに無効の印を付けるようにMLM構造を更新する。さらに、TRIM動作は、メディア110の対応する領域のBUSカウントから、トリミングされたLBAによって前に使用されていたフラッシュ・スペースを引いて、正確なガーベジ・コレクションをもたらす。このため、特定のLBAを適切にトリミングするのに、以下の2つのことが行われる。すなわち、その特定のLBAがMLM構造の中で無効にされ、BUSカウントが、その特定のLBA50

A がもはやフラッシュ・スペースを消費しないことを反映して更新される。しかし、大きいトリム領域（例えば、SSD全体）または複数の大きいトリム領域の場合、無効化およびBUS更新を実行するのに要求される時間は、大きくなり、システム性能に悪影響を与える可能性がある。

【0035】

本明細書で説明される、FLMの中に格納されたSLMページ情報は、対応するSLMページ内のLBAが既に無効にされている（例えば、トリミングされているようにホスト180に見える）が、TRIM動作のBUS更新部分がまだ完了していない場合、そのことを示す指示（例えば、処理されるべき(TBP)標識714)を含むことが可能である。より高レベルのマップ・エントリに無効の印を付けるのとは異なり、より高レベルのマップ・エントリのTBP標識を設定することは、そのより高レベルのマップ・エントリの中に格納されたより低レベルのマップ・ページの物理アドレスが無効であることは暗示せず、つまり、そのより低レベルのマップ・ページがBUS更新のために処理されるまで、その物理アドレスは、要求され、そのより低レベルのマップ・ページ自体、割り当て解除され得ない。しかし、そのより高レベルのマップ・エントリに関連するすべてのユーザ・データは、そのより高レベルのマップ・エントリに無効の印が付けられているとした場合と同じく、ホスト読み出し操作に関して無効である。

10

【0036】

メディア110の中に格納された関連するLBAのデータのサイズ（例えば、706）は、SSD101がTRIM動作を実行する際に対応する領域に関するBUS値を更新するのに使用される。例えば、これらのサイズ値が、対応する領域のBUSカウントから引かれる。MLM構造を使用する実施形態において、BUSカウントを更新することは、BUSカウントを更新することが、リーフ・レベル・マップ・エントリを1つずつ処理することを要求するので、時間がかかり得る。処理時間を改善するのに、説明される実施形態は、SSD101の背景動作モードでメディア110の対応する領域のBUSカウントを更新するメガTRIM動作を使用する。

20

【0037】

例えば、SSD101が、ホスト180からTRIMコマンドを受け取ると、メディア・コントローラ120が、TRIMコマンドに関連付けられたSLMページ（複数可）に対応するFLMエントリ（例えば、701）のそれぞれのTBP標識（例えば、714）を設定するメガTRIM動作を実行する。そのTRIM動作が、そのSLMページの中のSLMエントリの一部分だけにしか影響しない場合、一部の実施形態は、トリミングされるSLMエントリに無効の印を付けること、およびSLMページのトリミングされる部分を反映するようにBUSカウントを更新することによって、部分的SLMページの個々を更新することによって、部分的SLMページの個々のエントリを処理することが可能である。他の実施形態は、TBP標識（例えば、714）、TRIM動作進行中標識（例えば、710）、およびTRIM LBA範囲（例えば、712）を使用して、トリミングされたSLMエントリに無効の印を付けること、およびBUSカウントを更新することの遅延を許すことによって、部分的SLMページを更新することを遅らせることが可能である。次に、部分的にトリミングされたSLMページのその後の部分的TRIM動作が、オプションとして、および/または選択的に、部分的にトリミングされたSLMページに対する更新動作の一部またはすべてを即時に実行して、所与のTRIM LBA範囲（例えば、712）内で複数の部分範囲を追跡する必要性を回避する。しかし、代替の実施形態は、TRIM LBA範囲（例えば、712）内の複数の部分範囲を追跡して、トリミングされたSLMエントリに無効の印を付けること、およびBUSカウントを更新することのより長い遅延を許すことが可能である。

30

40

【0038】

メガTRIM動作が実行されると、関連するLBAを無効にした後、SSD101は、BUSカウントが更新されるのに先立って、ホスト180に対してTRIMコマンドの確認応答を行うことが可能である。次に、BUSカウントを更新することが、SSD101

50

の背景プロセスにおいて実行される（TRIM範囲、およびホスト180によって開始された活動の量に依存して、通常、数秒から数分までの範囲内に完了する）。関連するFLMエントリの中でTBP標識が設定されているSLMページのうちの1つが完全に処理される（例えば、トリミングされたSLMエントリに無効の印を付け、トリミングされたSLMページの中のすべてのSLMエントリに関するBUSカウントを更新して）たびに、関連するFLMエントリの中のTBP標識がクリアされる。SLMページのうちの1つのページのSLMエントリのすべてがトリミングされた場合、関連するFLMエントリには、トリミングされたという印が付けられて、新たな書き込みが、そのSLMページ内の少なくとも1つのエントリを有効にするまで、そのSLMページをさらに処理する必要性を除く。

10

【0039】

図8は、メガTRIM動作800の流れ図を示す。図8に示されるとおり、ステップ802で、TRIM動作要求が、SSD101によってホスト180から受け取られる。ステップ804で、SSD101が、TRIM動作の範囲（例えば、1つまたは複数の開始LBAおよび終了LBA）を特定する。SSD101は、TBP指標が設定されて、FLMのその部分が、BUSカウントを更新し、メディア110のメモリ・ブロックをホスト180が再び利用できるようにする背景動作を要求することを示す、FLMの部分を示すFLMの開始TBPインデックス（min__flm__index__tbt）および終了TBPインデックス（max__flm__index__tbt）を保持することが可能である。背景で（例えば、SSD101の別のアイドルの時間中に）、SSD101が、開始TBPインデックスにおけるFLMエントリを検査することと、そのFLMエントリ上にTBPが設定されている場合、関連するSLMページを読み出すことと、関連するSLMページの中の各エントリに応じてBUSカウントを更新し、そのFLMエントリの中のTBP標識をクリアし、そのFLMエントリにトリミングされたという印を付けて、そのSLMページ全体がトリミングされたことを示すことによって、そのSLMページ全体をトリミングすることが可能である。開始TBPインデックス（min__flm__index__tbt）が、そのエントリが処理されたことを示すように更新される。

20

【0040】

図8に示されるとおり、或るトリム範囲（例えば、SATAの場合、セクタごとに64のNCQトリム範囲（the 64 - per - sector NCQ trim range）のうちの1つ）を有するTRIMコマンドが処理されると、ステップ806で、SSD101が、そのTRIM範囲の最初のSLMページとそのTRIM範囲の最後のSLMページの少なくともいずれかが、部分的SLMページである（例えば、そのTRIM範囲が、そのSLMページの一部だけに適用される）かどうかを判定する。ステップ806で、その範囲の始め、または終わりに部分的SLMページが存在する場合、ステップ808で、SSD101は、その部分的SLMページがキャッシュ608の中に格納されているかどうかを判定する。ステップ808で、そのTRIM範囲の始め、または終わりの部分的SLMページがキャッシュ608の中に格納されている場合、プロセス800は、ステップ812に進む。ステップ808で、そのTRIM範囲の始め、または終わりの部分的SLMページがキャッシュ608の中に格納されていない場合、ステップ810で、SSD101は部分的SLMページをメディア110からキャッシュ608へフェッチし、プロセス800はステップ812に進む。ステップ812で、TRIM動作のその範囲内にある部分的SLMページのエントリに関してTRIM動作が実行される。例えば、その部分的SLMページの中のTRIM範囲内の任意のLBAに対応する、そのTRIM範囲内のSLMページ・エントリが更新される。SLMページの中のエントリを更新することは、データ無効標識を設定すること、およびBUSカウントを更新することを含む。プロセス800は、ステップ820に進む。

30

40

【0041】

ステップ806で、そのSLMページが部分的SLMページではなかった場合、ステップ814で、SSD101は、完全なSLMページがキャッシュ608の中に格納されて

50

いるかどうかを判定する。ステップ 814 で、完全な SLM ページがキャッシュ 608 の中に格納されている場合、プロセス 800 は、ステップ 816 に進む。ステップ 814 で、完全な SLM ページがキャッシュ 608 の中に格納されていなかった場合、ステップ 818 で、SSD 101 は、その SLM ページに対応する FL M 中の TBP 標識（例えば、714）を設定する。プロセス 800 は、ステップ 820 に進む。

【0042】

或る SLM ページがメディア 101 からフェッチされる必要がある際、関連する FL M エントリの中で TBP が設定されている場合、その SLM ページは、完全に無効にされている（その SLM ページ内のすべてのエントリは、ホスト・アクセスに関して無効として扱われる）が、その SLM ページは、BUS 更新の目的でまだ処理されていない。読み出しに関して、その SLM ページは、必要とされず（その SLM ページによって参照されるすべてのデータは、トリミングされている）、その SLM ページをフェッチすることは、要求されない。書き込みに関して、その SLM ページは、フェッチされ、その SLM ページの中のすべての LBA に関して BUS カウントが更新され、その SLM ページの中のすべてのエントリは、無効にされ、その後、書き込まれているその SLM ページ内の SLM エントリが更新される。ステップ 816 で、書き込みのための動作のサブセットが実行される。すなわち、その SLM ページの中のすべての LBA に関して BUS カウントが更新され、その SLM ページの中のすべてのエントリが無効にされる。

【0043】

ステップ 822 で、SSD 101 は、TBP 標識が設定されていて（例えば、min__flm__index__tbt および max__flm__index__tbt）、FL M のその部分が、BUS カウントを更新し、メディア 110 のメモリ・ブロックをホスト 180 が再び利用できるようにする背景動作を要求することを示す、FL M のエントリの範囲を特定する。ステップ 824 で、残りの TRIM 動作（例えば、BUS カウントを更新すること、およびメモリ・ブロックを、ホスト 180 によって使用可能であるものとして解放すること）が背景で行われる（例えば、SD 101 の別のアイドルな時間中に）。SSD 101 は、ステップ 816 でメモリ・ブロックがトリミングされると（例えば、メモリ・ブロックの BUS カウントが更新されると）更新される 1 つまたは複数のポイントを保持して、ブロックが処理されるにつれ新たな TRIM 範囲が記憶されることを確実にすることが可能である。例えば、SSD 101 が、開始 TBP インデックスにおける FL M エントリを検査することと、その FL M エントリ上で TBP が設定されている場合、関連する SLM ページを読み出すことと、BUS カウントを更新し、その FL M エントリの中の TBP 標識をクリアし、その FL M エントリにトリミングされたという印を付けて、その SLM ページ全体がトリミングされたことを示すことによってその SLM ページ全体をトリミングすることとが可能である。開始 TBP インデックス（min__flm__index__tbt）は、そのエントリが処理されたことを示すように更新される。ステップ 824 で背景 TRIM 動作が完了すると、ホスト 180 に対して TRIM 動作の確認応答が行われる。ステップ 826 で、プロセス 800 は完了する。

【0044】

このため、本明細書で説明されるとおり、説明される実施形態は、ソリッドステート・メディアのためのメディア・コントローラを提供する。このメディア・コントローラは、少なくとも 1 つの論理アドレスと、アドレス範囲とを含む要求をホスト・デバイスから受け取る制御プロセッサを含む。この要求に応答して、制御プロセッサは、受け取られた要求が無効にする要求であるかどうかを判定する。受け取られた要求タイプが無効にする要求である場合、制御プロセッサは、メディア・コントローラのマップを使用して、その論理アドレスおよび範囲に関連付けられたマップの 1 つまたは複数のエントリを特定する。これらのマップ・エントリのそれぞれに関連付けられたマップの中の標識が、これらのマップ・エントリが無効にされるべきことを示すように設定される。制御プロセッサは、無効にする要求が完了したという確認応答をホスト・デバイスに対して行い、メディア・コントローラのアイドル・モードにおいて、無効にされるべきマップ・エントリに基づいて

空きスペース・カウントを更新する。無効にされたマップ・エントリに関連付けられた物理アドレスが、ホスト・デバイスからのその後の要求に関して再使用されるように利用可能にされる。

【 0 0 4 5 】

本明細書で「 1 つの実施形態」または「或る実施形態」に言及することは、その実施形態に関連して説明される特定のフィーチャ、構造、または特徴が、少なくとも 1 つの実施形態に含まれ得ることを意味する。本明細書の様々な箇所における「一実施形態において」という句の出現は、必ずしもすべて同一の実施形態を指すわけではなく、別々の実施形態も、代替の実施形態も、必ずしも他の実施形態と相容れないわけでもない。同じことは、「実施例」という用語にも当てはまる。

10

【 0 0 4 6 】

本出願で使用されるように、「例示的」という語は、例、実例、または例示の役割をすることを意味するように本明細書で使用される。本明細書で「例示的」と説明される態様または設計は、必ずしも、他の態様または設計より好ましいとも、有利であるとも解釈されるべきではない。むしろ、例示的という語の使用は、具体的な方法で概念を提示することを意図している。

【 0 0 4 7 】

例示的な実施形態は、デジタル・シグナル・プロセッサ、マイクロコントローラ、または汎用コンピュータとしての可能な実装を含め、ソフトウェア・プログラムにおける処理ブロックに関連して説明されてきたが、説明される実施形態は、そのように限定されない。当業者には明白なとおり、ソフトウェアの様々な機能は、回路のプロセスとして実装されることも可能である。そのような回路は、例えば、単一の集積回路、マルチチップ・モジュール、単一のカード、またはマルチカード回路パックにおいて使用されることが可能である。

20

【 0 0 4 8 】

また、説明される実施形態は、それらの方法を実施するための方法および装置の形態で実現されることも可能である。また、説明される実施形態は、磁気記録媒体、光記録媒体、ソリッドステート・メモリ、フロッピー・ディスク、CD-ROM、ハードドライブ、または他の任意の一時的でないマシン可読記憶媒体などの一時的でない実体のある媒体として実現されたプログラム・コードの形態で実現されることも可能であり、そのプログラム・コードが、コンピュータなどのマシンにロードされ、そのマシンによって実行されると、そのマシンが、説明される実施形態を実施するための装置となる。説明される実施形態は、例えば、一時的でないマシン可読記憶媒体の中に格納されて、マシンにロードされ、および/またはマシンによって実行されるか、あるいは電気配線もしくはケーブル配線を介して、光ファイバを介して、または電磁放射を介してなど、何らかの伝送媒体または搬送波を介して伝送されるかにかかわらず、プログラム・コードの形態で実現されることも可能であり、そのプログラム・コードが、コンピュータなどのマシンにロードされ、そのマシンによって実行されると、そのマシンが、説明される実施形態を実施するための装置となる。汎用プロセッサ上で実装されると、プログラム・コード・セグメントは、そのプロセッサと一緒に、特定の論理回路と同様に動作する独特のデバイスをもたらす。また、説明される実施形態は、説明される実施形態の方法および/もしくは装置を使用して生成された、媒体を介して電氣的に、もしくは光学的に伝送される信号値のビットストリームまたは他のシーケンス、磁気記録媒体の中の格納された磁界変化などの形態で実現されることも可能である。

30

40

【 0 0 4 9 】

本明細書で提示される例示的な方法のステップは、必ずしも、説明される順序で実行されることは必須ではないものと理解されるべきであり、そのような方法のステップの順序は、単に例示的であるものと理解されたい。同様に、そのような方法にさらなるステップが含まれることも可能であり、いくつかのステップが、説明される様々な実施形態と合致する方法において省略される、または組み合わされることも可能である。

50

【 0 0 5 0 】

要素および標準に関連して本明細書で使用される「互換性がある」という用語は、その要素が、その標準によって完全な状態で、または部分的に指定される状態で他の要素と通信し、他の要素によって、その標準によって指定される状態でそれらの他の要素と通信することが十分にできるものと認識されることを意味する。互換性のある要素は、内部では、その標準によって指定される状態で動作しなくてもよい。特に明記しない限り、各数値および各範囲は、「約」または「ほぼ」という語がその値または範囲の値に先行するかのよう、近似であるものと解釈されたい。

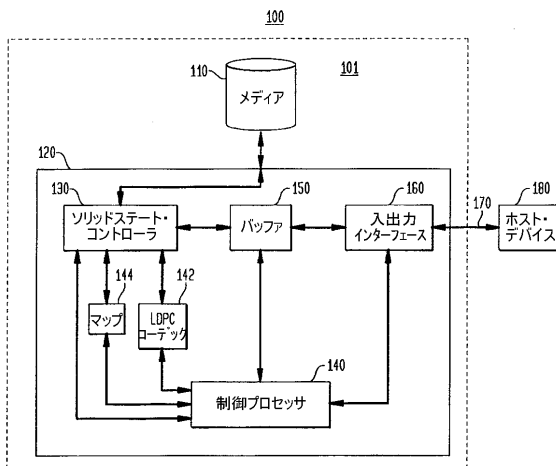
【 0 0 5 1 】

この説明に関して、「結合する」、「結合している」、「結合された」、「接続する」、「接続している」、または「接続された」という用語は、2つ以上の要素間でエネルギーが伝達されることが許される、当技術分野で知られている、または後に開発される任意の形態を指し、さらなる1つまたは複数の要素を介在させることが企図されるが、必須ではない。逆に、「直接に結合された」、「直接に接続された」などの言い方は、そのようなさらなる要素が無いことを暗示する。信号、および対応するノードもしくはポートは、同一の名前で呼ばれることが可能であるとともに、本明細書では互いに区別されない。

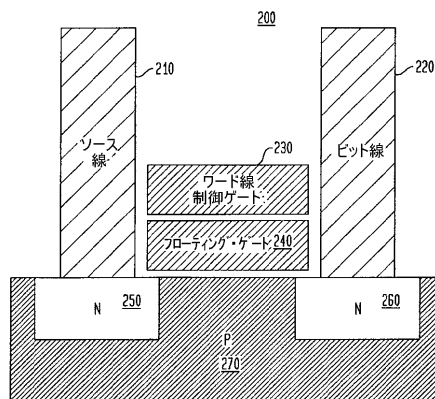
【 0 0 5 2 】

説明される実施形態の性質を説明するために説明され、例示されてきた部分の詳細、材料、および構成の様々な変更が、以下の特許請求の範囲に記載される範囲を逸脱することなく、当業者によって行われることが可能であることが、さらに理解されよう。

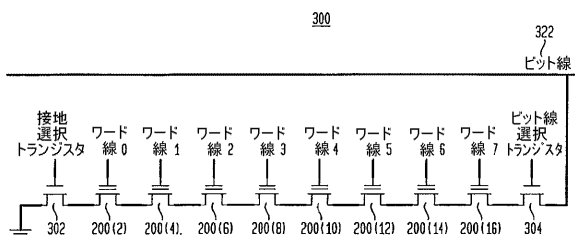
【 図 1 】



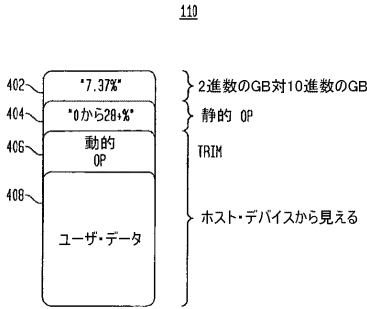
【 図 2 】



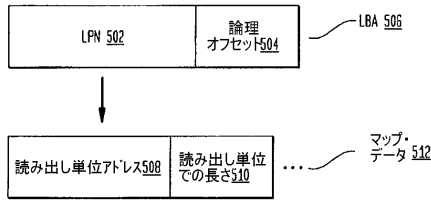
【 図 3 】



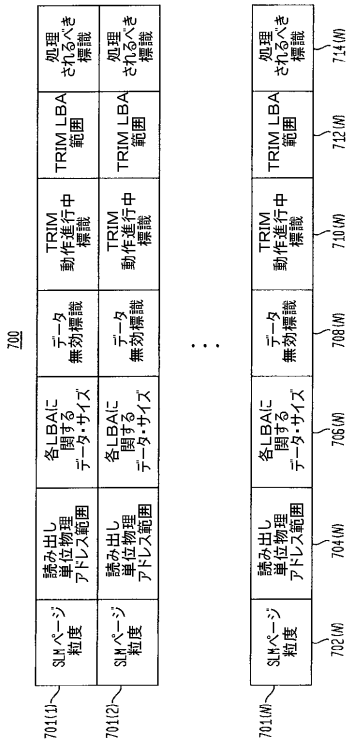
【図 4】



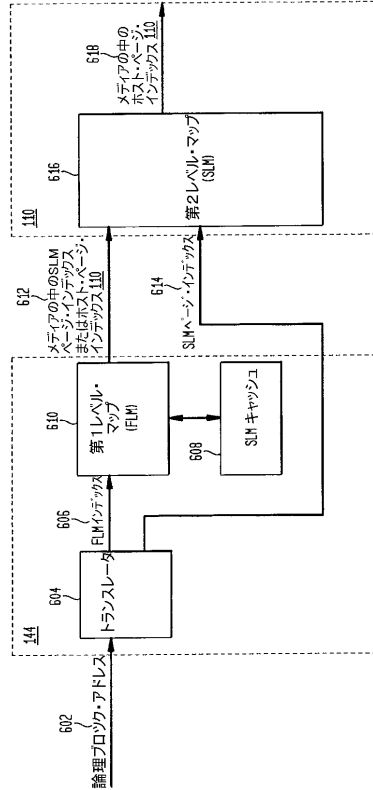
【図 5】



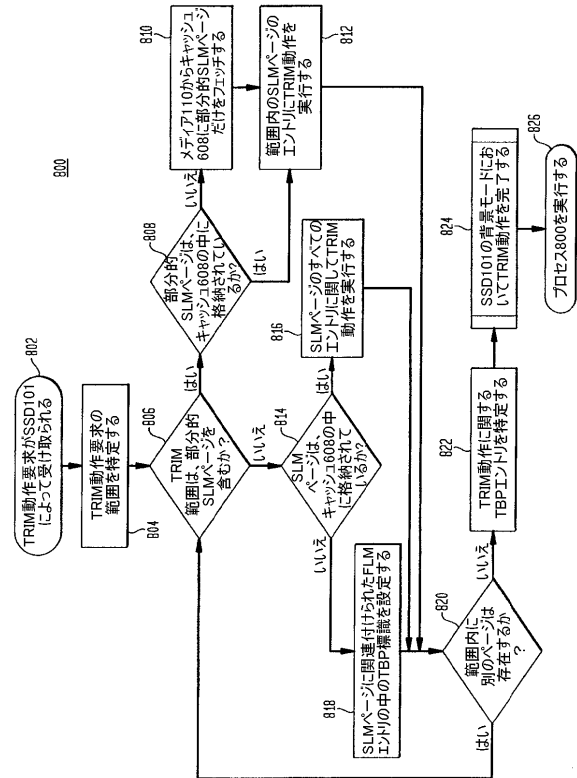
【図 7】



【図 6】



【図 8】



フロントページの続き

(74)代理人 100106183

弁理士 吉澤 弘司

(74)代理人 100170601

弁理士 川崎 孝

(74)代理人 100187964

弁理士 新井 剛

(72)発明者 アール ティー・コーエン

アメリカ合衆国 9 4 6 1 8 カリフォルニア, オークランド, エルモッサ アヴェニュー 7 8

(72)発明者 レオニード パリュディン

アメリカ合衆国 9 5 1 2 4 カリフォルニア, サンホセ, リー アヴェニュー 5 0 6 7

F ターム(参考) 5B060 AA09