

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5635200号
(P5635200)

(45) 発行日 平成26年12月3日(2014.12.3)

(24) 登録日 平成26年10月24日(2014.10.24)

(51) Int.Cl.	F I	
G06F 13/16	(2006.01)	G06F 13/16 510D
G06F 12/00	(2006.01)	G06F 12/00 597U
G06F 12/06	(2006.01)	G06F 12/06 515H
G06F 12/16	(2006.01)	G06F 12/16 320L
G06F 3/06	(2006.01)	G06F 3/06 301A

請求項の数 9 (全 26 頁) 最終頁に続く

(21) 出願番号	特願2013-535800 (P2013-535800)	(73) 特許権者	000005108
(86) (22) 出願日	平成23年9月30日 (2011.9.30)		株式会社日立製作所
(86) 国際出願番号	PCT/JP2011/072653		東京都千代田区丸の内一丁目6番6号
(87) 国際公開番号	W02013/046463	(74) 代理人	110000062
(87) 国際公開日	平成25年4月4日 (2013.4.4)		特許業務法人第一国際特許事務所
審査請求日	平成25年10月16日 (2013.10.16)	(72) 発明者	石川 篤
			日本国東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
		(72) 発明者	園田 浩二
			日本国東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
		(72) 発明者	上原 剛
			日本国東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内

最終頁に続く

(54) 【発明の名称】 不揮発半導体記憶システム

(57) 【特許請求の範囲】

【請求項1】

フラッシュメモリコントローラと、
前記フラッシュメモリコントローラに接続された、第1データバスと、
前記フラッシュメモリコントローラに接続された、第2データバスと、
複数の第3データバスと、
複数の第4データバスと、
それぞれが前記複数の第3データバスの1つに接続された、複数の第1フラッシュメモリチップと、
それぞれが前記複数の第4データバスの1つに接続された、複数の第2フラッシュメモリチップと、
前記第1データバスに接続され、前記複数の第3データバスのいずれか1つに選択的に接続される、第1スイッチと、
前記第2データバスに接続され、前記複数の第4データバスのいずれか1つに選択的に接続される、第2スイッチと、
前記フラッシュメモリコントローラと前記第1スイッチとに接続された切替信号線と、
を備え、
前記第1スイッチは、前記切替信号線を用いて切替信号を送信するため、前記第2スイッチに直接接続されるよう構成され、これにより前記フラッシュメモリコントローラは、前記切替信号線を用いて前記第1スイッチ及び前記第2スイッチの両方に1つの切替信号

10

20

を供給することによって、前記複数の第 1 フラッシュメモリチップの 1 つ及び前記複数の第 2 フラッシュメモリチップの 1 つに並行してアクセスする、フラッシュメモリパッケージ。

【請求項 2】

前記フラッシュメモリコントローラは、前記第 1 データバスを前記複数の第 3 データバスの 1 つに接続するための前記第 1 スイッチと、前記第 2 データバスを前記複数の第 4 データバスの 1 つに接続するための前記第 2 スイッチとを制御するよう構成される、請求項 1 記載のフラッシュメモリパッケージ。

【請求項 3】

前記フラッシュメモリコントローラは、ライトデータを複数のデータ要素に分割し、前記複数のデータ要素の第 1 データ要素を前記複数の第 1 フラッシュメモリチップの 1 つへ、及び前記複数のデータ要素の第 2 データ要素を前記複数の第 2 フラッシュメモリチップの 1 つへ、それぞれ並行して送信するよう構成される、請求項 2 記載のフラッシュメモリパッケージ。

【請求項 4】

前記フラッシュメモリコントローラに接続された複数のチップイネーブル信号線を更に含み、前記複数のチップイネーブル信号線のそれぞれは、複数の第 1 フラッシュメモリチップの 1 つと複数の第 2 フラッシュメモリチップの 1 つとに接続される、請求項 3 記載のフラッシュメモリパッケージ。

【請求項 5】

前記複数のチップイネーブル信号線は、第 1 及び第 2 のチップイネーブル信号線を含み、

(A) 前記フラッシュメモリコントローラは、前記第 1 のチップイネーブル信号線を介してチップイネーブル信号を送信し、且つ、前記異なるスイッチのそれぞれを、前記第 1 のチップイネーブル信号線を介してチップイネーブル信号を受ける前記フラッシュメモリチップが接続先となるよう制御し、

(B) 前記フラッシュメモリコントローラは、前記複数のデータ要素のうちの連続する 2 以上のデータ要素を、前記第 1 のチップイネーブル信号線を共通とする 2 以上のフラッシュメモリチップに並行して送信し、

(C) 前記フラッシュメモリコントローラは、前記 (B) でデータ要素を受けた前記フラッシュメモリチップに接続されているスイッチの接続先を、前記第 2 のチップイネーブル信号線を介してチップイネーブル信号を受けることになる前記フラッシュメモリチップに切り替え、

(D) 前記フラッシュメモリコントローラは、前記第 2 のチップイネーブル信号線を介して前記チップイネーブル信号を送信し、

(E) 前記フラッシュメモリコントローラは、前記複数のデータ要素のうちの残りのうちの連続する 2 以上のデータ要素を、前記第 2 のチップイネーブル信号線を共通とする 2 以上の前記フラッシュメモリチップに並行して送信する、

請求項 4 記載のフラッシュメモリパッケージ。

【請求項 6】

前記フラッシュメモリコントローラが、前記チップイネーブル信号線毎に、ブロックに対して行われた消去処理の回数の合計を管理し、前記複数のデータ要素の転送先として選択される前記フラッシュメモリチップは、前記チップイネーブル信号を共通にする複数の前記フラッシュメモリチップのうち、前記消去処理の回数が最も少ないフラッシュメモリチップであり、前記フラッシュメモリチップにおいてデータ要素の書込み先となる前記ブロックは、消去回数が最も少ないブロックである、請求項 5 記載のフラッシュメモリパッケージ。

10

20

30

40

50

【請求項 7】

前記各フラッシュメモリチップは、複数のブロックを有しており、
前記フラッシュメモリコントローラは、リクラメーション処理を行うように構成されて
おり、

前記リクラメーション処理において、前記フラッシュメモリコントローラは、

- (r 1) 前記複数のブロックから移動元のブロックを選択し、
- (r 2) 前記複数のブロックから移動先のブロックを決定し、
- (r 3) 前記移動先ブロックに、前記移動元ブロックのページ内の有効データを移動し、
- (r 4) 前記移動元ブロックからデータを消去する消去処理を行い、

前記 (r 2) において、前記移動先ブロックは、前記移動元ブロックを有するフラッシュメモリチップとデータ通信可能な媒体インタフェースとデータ通信可能な前記フラッシュメモリチップの中から決定される、

請求項 6 記載のフラッシュメモリパッケージ。

10

【請求項 8】

前記媒体インタフェースには、2 以上のデータバスを介して 2 以上の前記スイッチが接続され、

前記移動元ブロックと前記移動先ブロックは、同一の前記データバスを介してデータ通信可能な 1 以上の前記フラッシュメモリチップに存在する、

請求項 7 記載のフラッシュメモリパッケージ。

20

【請求項 9】

前記 (r 2) で、前記フラッシュメモリコントローラは、前記移動元ブロックを有するフラッシュメモリチップとデータ通信するためのデータバスと同じデータバスを介してデータ通信可能な 1 以上のフラッシュメモリチップに空きのブロックが所定数以上あるか否か判断し、

その判断の結果が肯定的であれば、前記 1 以上のフラッシュメモリチップからいずれかの空きのブロックを移動先ブロックとして決定し、

その判断の結果が否定的であれば、前記移動元ブロックを有するフラッシュメモリチップとデータ通信可能な前記媒体インタフェースと異なる媒体インタフェースとデータ通信可能な 2 以上の前記フラッシュメモリチップの中から空きのブロックを移動先ブロックとして決定する、

30

請求項 7 記載のフラッシュメモリパッケージ。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、複数の不揮発半導体記憶媒体を有する記憶システムに関する。

【背景技術】

【0002】

ストレージシステムは、一般に、複数の記憶デバイスで構成された R A I D (R e d u n d a n t A r r a y o f I n d e p e n d e n t D i s k s) グループに基づいて作成された論理ボリュームを、上位装置 (例えばホストコンピュータ) へ提供する。近年、記憶デバイスとしては、HDD (H a r d D i s k D r i v e) に加えて又は代えて、複数の不揮発チップを有する不揮発半導体記憶装置が採用されている。不揮発半導体記憶デバイスとして、例えば、複数のフラッシュメモリチップ (以下、F M チップ) を有するフラッシュメモリが採用される (例えば特許文献 1) 。

40

【先行技術文献】

【特許文献】

【0003】

【特許文献 1】特開 2010 - 3161 号公報

【発明の概要】

【発明が解決しようとする課題】

50

【0004】

不揮発半導体記憶デバイスにおいても、記憶容量の増加が要請されている。記憶容量を増加させるためには、搭載する不揮発チップを増加させることが必要であるが、より多くの不揮発チップを搭載するようにすると、不揮発チップに接続されるインタフェースデバイスを含んだ回路（以下、制御回路、例えばASIC（Application Specific Integrated Circuit））には、複数の不揮発チップに各種信号を送信するためのバスを接続するためのピンを用意しておく必要がある。

【0005】

このため、不揮発半導体記憶デバイスの記憶容量を増加させようとする、制御回路のサイズが大きくなってしまふ。一方、制御回路のサイズを小さくすると、接続可能な不揮発チップの数が少なくなり、記憶容量を増加させることが困難である。また、多数の不揮発チップを接続すると効率よくデータ転送をすることが困難である。

【0006】

本発明の目的は、制御回路のサイズを抑えつつ、搭載できる不揮発チップの数を増加させ、効率よくデータ転送することのできる技術を提供することにある。

【課題を解決するための手段】

【0007】

不揮発半導体記憶システムが、（a）複数の不揮発半導体記憶媒体と、（b）複数の不揮発半導体記憶媒体に接続された媒体インタフェース群（1以上のインタフェースデバイス）を有する制御回路と、（c）複数のスイッチとを有する。媒体インタフェース群と複数のスイッチとがデータバスを介して接続され、各スイッチと、各2以上の不揮発チップとが、データバスを介して接続される。スイッチは、媒体インタフェース群に接続されているデータバスとそのスイッチに接続されている複数の不揮発チップのいずれかに接続されているデータバスとの接続を切り替えるよう構成されている。制御回路は、ライト対象のデータを複数のデータ要素に分割し、複数のスイッチを制御することにより接続を切り替えて、複数のデータ要素を複数の不揮発チップに分散して送信する。

【0008】

不揮発半導体記憶システムは、上記（a）、（b）及び（c）の要素を有する記憶媒体グループ（例えば、後述のフラッシュメモリPKG10）であっても良いし、そのような記憶媒体グループを複数個有する記憶装置（例えば後述のフラッシュメモリデバイス400）であっても良いし、そのような記憶装置を複数個とそれらの記憶装置に接続されたコントローラとを有するシステム（例えば、後述のストレージシステム1）であっても良い。

【0009】

媒体インタフェース群は、複数の媒体インタフェースを有し、媒体インタフェースが、N個の不揮発半導体記憶媒体（Nは1以上の整数）毎に存在しても良い。1つの不揮発半導体記憶媒体を構成する複数の不揮発チップが、例えば1つのDIMMに存在して良い。DIMMが、それら複数の不揮発チップに接続される1以上のスイッチを有して良い。

【図面の簡単な説明】

【0010】

【図1】図1は、実施形態に係る計算機システムの構成例を示す。

【図2】図2は、実施形態に係るフラッシュメモリデバイスの構成例を示す。

【図3】図3は、実施形態に係るフラッシュメモリパッケージの構成例を示す。

【図4】図4は、実施形態に係るフラッシュメモリパッケージの一部の詳細な構成例を示す。

【図5】図5は、実施形態に係る書込み処理の第1の例を説明する図である。

【図6】図6は、実施形態に係る書込み処理の第2の例を説明する図である。

【図7】図7は、実施形態に係るフラッシュメモリデバイスの正面上方からの斜視図の一例である。

【図8】図8は、実施形態に係るフラッシュメモリPKGの上面側からの斜視図の一例で

10

20

30

40

50

ある。

【図 9】図 9 は、実施形態に係るフラッシュメモリ P K G の下面側からの斜視図の一例である。

【図 10】図 10 は、実施形態に係る D I M M の概略構成の一例を示す。

【図 11】図 11 は、実施形態に係るチップリード処理のフローチャートの一例である。

【図 12】図 12 は、実施形態に係るチップライト処理のフローチャートの一例である。

【図 13】図 13 は、実施形態に係るチップ多重ライト処理のフローチャートの一例である。

【図 14】図 14 は、実施形態の一変形例に係るフラッシュメモリパッケージの一部の詳細な構成例を示す。

【図 15】図 15 は、実施形態に係る論理アドレス層と物理層との関係の一例を示す。

【図 16】図 16 は、実施形態に係るユーザアドレス空間とフラッシュストレージ論理空間との関係の一例を示す。

【図 17】図 17 は、実施形態に係る論理物理変換情報の構成例を示す。

【図 18】図 18 は、実施形態に係るリクレーション処理のフローチャートの一例を示す。

【図 19】図 19 は、実施形態に係る消去管理情報の構成例を示す。

【発明を実施するための形態】

【0011】

以下、一実施形態を、図面を参照して説明する。

【0012】

なお、以下の説明では、要素（例えば、ページ、フラッシュメモリチップ（F M チップ）、スイッチ（S W ））を特定するために番号を含む識別情報が使用されるが、識別情報として、番号を含まない情報が使用されても良い。

【0013】

また、以下の説明では、同種の要素を区別して説明する場合、要素名と参照符号との組合せに代えて、要素名と識別情報との組合せが使用されることがある。例えば、識別情報（識別番号）「0」のスイッチを、「スイッチ # 0」と表記することがある。

【0014】

また、以下の説明では、インタフェースデバイスを「I / F」と略記することがある。

【0015】

また、以下の説明では、不揮発半導体記憶媒体は、フラッシュメモリ（F M）であるとす。そのフラッシュメモリは、ブロック単位で消去が行われ、ページ単位でアクセスが行われる種類のフラッシュメモリ、典型的には N A N D 型のフラッシュメモリであるとす。しかし、フラッシュメモリは、N A N D 型に代えて他種のフラッシュメモリ（例えば N O R 型）でも良い。また、フラッシュメモリに代えて、他種の不揮発半導体記憶媒体、例えば相変化メモリが採用されても良い。

【0016】

また、以下の説明では、不揮発半導体記憶媒体は、前述したように、N A N D 型フラッシュメモリである。このため、ページとブロックという用語が使用される。また、或る論理領域（この段落において「対象論理領域」と言う）がライト先であり、且つ、対象論理領域に既にページ（この段落において「第 1 のページ」と言う）が割り当てられていて第 1 のページにデータが格納されている場合、対象論理領域には、第 1 のページに代えて、空きのページ（この段落において「第 2 のページ」と言う）が割り当てられ、第 2 のページにデータが書き込まれることになる。第 2 のページに書き込まれたデータが、対象論理領域にとって最新のデータであり、第 1 のページに格納されているデータは、対象論理領域にとって古いデータとなる。以下、各論理領域について、最新のデータを「有効データ」と言い、古いデータを「無効データ」と言うことがある。また、有効データを格納しているページを「有効ページ」と言い、無効データを格納しているページを「無効ページ」と言うことがある。

10

20

30

40

50

【 0 0 1 7 】

さて、まず、本実施形態の概要を説明する。

【 0 0 1 8 】

フラッシュメモリデバイス 4 0 0 は、図 2 に示すように、例えば、1 以上のフラッシュメモリパッケージ（フラッシュメモリ P K G）1 0 を備える。フラッシュメモリ P K G 1 0 は、図 3 に示すように、複数のフラッシュメモリチップ（F M チップ）3 2 を有する。

【 0 0 1 9 】

フラッシュメモリ P K G 1 0 の媒体インタフェースの一例である F M I / F 制御部 2 4 は、F M チップ 3 2 に対して、チップイネーブル信号（C E 信号）、及び、その F M チップ 3 2 に書き込むデータ、そのデータのライト先とするアドレスを出力する。本実施形態においては、図 4 に示すように、F M I / F 制御部 2 4 から出力される C E 信号用の信号線 2 7 は、複数の F M チップ 3 2 に接続されるように配されている。F M I / F 制御部 2 4 では、一つの C E 信号線 2 7 に対して、一つの出力端子（ピン）があればよいので、F M チップ 3 2 の個数よりも少ない数のピンがあればよい。このため、F M I / F 制御部 2 4 を含む A S I C 等の回路におけるピンの配置に必要な領域を低減することができる。

10

【 0 0 2 0 】

また、本実施形態においては、図 4 に示すように、F M I / F 制御部 2 4 から出力されるデータ、アドレス等の信号（C E 信号以外の信号）が流れるバス（バス：制御線が含まれてないがバスということとする）2 5 は、それぞれスイッチ 3 1 に接続されている。また、スイッチ 3 1 には、M 個（M は 2 以上の整数、例えば、M = 4）の F M チップ 3 2 がバス 2 8 を介して接続されている。スイッチ 3 1 は、バス 2 5 と、いずれかのバス 2 8 との接続を切り替えることができるようになっている。複数の F M チップ 3 2 に対するリードとライトにおいては、バス 2 5 を介してデータ、アドレス等がやり取りされるので、F M I / F 制御部 2 4 では、バス 2 5 が接続されるピンが確保されていればよい。したがって、F M I / F 制御部 2 4 を含む A S I C 等の回路におけるピンの配置に必要な領域を低減することができる。また、バス 2 5 と、複数のバス 2 8 のいずれかをスイッチ 3 1 により接続させるので、複数のバス 2 8 が電気的に接続されている状態とはならない。このため、バス 2 5 を介して F M チップ 3 2 に接続される全体の配線における負荷容量を抑えることができ、F M チップ 3 2 との間のデータ交換の品質を比較的高くすることができる。

20

30

【 0 0 2 1 】

次に、本実施形態を詳細に説明する。

【 0 0 2 2 】

図 1 は、本実施形態に係る計算機システムの構成例を示す。

【 0 0 2 3 】

計算機システムは、ストレージシステム 1 と、ホストコンピュータ（ホストともいう）2 0 0 とを有する。ストレージシステム 1、ホスト 2 0 0 の数は、それぞれ、1 以上とすることができる。ストレージシステム 1 と、ホスト 2 0 0 とは、通信ネットワーク（例えば、S A N（Storage Area Network））を介して相互に接続されている。ストレージシステム 1 は、ホスト 2 0 0 で利用されるデータを記憶する。ホスト 2 0 0 は、各種処理を実行し、ストレージシステム 1 からデータを読み出したり、ストレージシステム 1 へデータを書き込んだりする。

40

【 0 0 2 4 】

ストレージシステム 1 は、複数の記憶デバイスと、それら複数の記憶デバイスに接続された R A I D（Redundant Array of Independent（or Inexpensive）Disks の略）コントローラデバイス 3 0 0 とを有する。

【 0 0 2 5 】

複数の記憶デバイスは、複数種類の記憶デバイスを含む。少なくとも 1 種類の記憶デバ

50

イスは、1以上存在して良い。記憶デバイスとして、例えば、フラッシュメモリデバイス400、SSD(Solid State Drive)デバイス500、HDD(Hard Disk Drive)デバイス(SAS:Serial Attached SCSI)600及びHDDデバイス(SATA:Serial ATA)700がある。

【0026】

RAIDコントローラデバイス300は、複数のRAIDコントローラ301を有する。各RAIDコントローラ301は、フラッシュメモリデバイス400、SSDデバイス500、HDDデバイス(SAS)600及びHDDデバイス(SATA)700と内部バスを介して接続されている。

【0027】

なお、RAIDコントローラ301は、フラッシュメモリデバイス400、SSDデバイス500、HDDデバイス(SAS)600及びHDDデバイス(SATA)700にとっての上位装置の一例である。RAIDコントローラ301は、RAIDコントローラ301にとっての上位装置(例えば、ホスト200)からI/Oコマンドを受け、そのI/Oコマンドに従い、フラッシュメモリデバイス400、SSDデバイス500、HDDデバイス(SAS)600又はHDDデバイス(SATA)700へのアクセス制御を行う。RAIDコントローラ301は、フラッシュメモリデバイス400、SSDデバイス500、HDDデバイス(SAS)600、HDDデバイス(SATA)700のそれぞれの記憶領域をそれぞれ異なる記憶階層として管理し、データのライト先の論理領域に対して、いずれかの記憶階層の記憶領域を割当てて処理を行うようにしてもよい。

【0028】

ここで、SSDデバイス500の方がフラッシュメモリデバイス400よりもフラッシュメモリの書き込み可能回数が多い、一方、読み出し速度及びコスト面では、フラッシュメモリデバイス400の方が優れているという特徴があるとする。このため、RAIDコントローラ301は、リードが比較的頻繁に行われるデータを、フラッシュメモリデバイス400に格納し、ライトが比較的頻繁に行われるデータを、SSDデバイス500に格納してもよい。

【0029】

図2は、本実施形態に係るフラッシュメモリデバイスの構成例を示す。

【0030】

フラッシュメモリデバイス400は、1以上の上位I/Fスイッチ(上位I/F Switch)401と、1以上のフラッシュメモリパッケージ(PKG)10とを有する。上位I/Fスイッチ401は、RAIDコントローラ301と、複数のフラッシュメモリPKG10との間のデータの中継を行う。

【0031】

図3は、本実施形態に係るフラッシュメモリパッケージの構成例を示す。

【0032】

フラッシュメモリPKG10は、主記憶メモリの一例としてDRAM(Dynamic Random Access Memory)11を有し、また、FMコントローラ20と、複数(又は1つ)のDIMM(Dual Inline Memory Module)30とを有する。DRAM11は、FMコントローラ20で使用されるデータ等を記憶する。DRAM11は、FMコントローラ20に搭載されていても良いし、FMコントローラ20とは別の部材に搭載されていても良い。

【0033】

FMコントローラ20は、例えば、1つのASIC(Application Specific Integrated Circuit)で構成されており、CPU21と、内部バス22と、上位I/F(インタフェース)23と、複数(又は1つ)のFM I/F制御部24とを有する。内部バス22は、CPU21と、上位I/F23と、DRAM11と、FM I/F制御部24とを通信可能に接続する。

【0034】

上位I/F23は、上位I/F Switch401に接続され、上位装置との通信を

10

20

30

40

50

仲介する。上位 I / F 2 3 は、例えば、S A S の I / F である。F M I / F 制御部 2 4 は、複数の F M チップ 3 2 とのデータのやり取りを仲介する。本実施形態では、F M I / F 制御部 2 4 は、F M チップ 3 2 とのやり取りを実行するバス（データバス等）を複数組有し、複数のバスを用いて、複数の F M チップ 3 2 とのデータのやり取りを仲介する。本実施形態では、D I M M 3 0 毎に F M I / F 制御部 2 4 が設けられ、F M I / F 制御部 2 4 は、その制御部 2 4 に接続された D I M M 3 0 が有する複数の F M チップ 3 2 との通信を仲介する。なお、F M I / F 制御部 2 4 が担当する D I M M 3 0 の枚数は、2 以上であってもよい。C P U 2 1 は、D R A M 1 1（又は図示しない他の記憶領域）に記憶されるプログラムを実行することによって、各種処理を実行することができる。C P U 2 1 は複数あってもよく、複数の C P U 2 1 が各種処理を分担してもよい。C P U 2 1 による具体的な処理については、後述する。

10

【 0 0 3 5 】

D I M M 3 0 は、1 以上の S W 3 1 と、複数の F M チップ 3 2 とを有する。F M チップ 3 2 は、例えば、M L C (Multi Level Cell) 型の N A N D フラッシュメモリチップである。M L C 型の F M チップは、S L C 型の F M チップと比べて書き換え可能な回数が劣るが、1 セルあたりの記憶容量が多いという特徴を有している。

【 0 0 3 6 】

S W 3 1 は、F M I / F 制御部 2 4 と、データバスを含むバス 2 5 を介して接続されている。本実施形態では、S W 3 1 は、F M I / F 制御部 2 4 に接続されるデータバスを含む一組のバス 2 5 に対して、一つ対応するように設けられている。また、S W 3 1 は、複数の F M チップ 3 2 とデータバスを含むバス 2 8 を介して接続されている。S W 3 1 は、F M I / F 制御部 2 4 からのバス 2 5 と、いずれかの F M チップ 3 2 のバス 2 8 とを選択的に切り替えて接続できるようになっている。ここで、D I M M 3 0 に、S W 3 1 と、複数の F M チップ 3 2 とが設けられ、配線がされているので、これらを接続するためのコネクタを別に用意せずともよく、必要なコネクタ数を低減することが期待できる。

20

【 0 0 3 7 】

なお、図 3 によれば、F M チップ 3 2 が別の F M チップ 3 2 を介することなく S W 3 1 に接続されているが、F M チップ 3 2 が別の F M チップ 3 2 を介して S W 3 1 に接続されても良い。すなわち、S W 3 1 に、直列になった 2 以上の F M チップ 3 2 が接続されても良い。

30

【 0 0 3 8 】

図 4 は、本実施形態に係るフラッシュメモリパッケージの一部の詳細な構成例を示す。

【 0 0 3 9 】

F M I / F 制御部 2 4 は、E C C (Error Correcting Code) 回路 2 4 1 と、制御レジスタ 2 4 2 と、F M / S W 制御部 2 4 3 と、バッファ 2 4 4 と、F M バスプロトコル制御部（図では「プロトコル 1、プロトコル 2」のように表記）2 4 6 と、D M A (Direct Memory Access) 部 2 4 7 とを有する。本実施形態では、バッファ 2 4 4、F M バスプロトコル制御部 2 4 6、及び D M A 部 2 4 7 の組を、F M I / F 制御部 2 4 が担当するデータバスの数分（例えば、2 組）備えている。

【 0 0 4 0 】

D R A M 1 1 では、後述するように、ライト対象のデータが複数のデータ要素に分割される。E C C 回路 2 4 1 は、D R A M 1 1 からライト対象のデータ要素を読み出し、ライト対象のデータ要素に対応する（例えば付加される）誤り訂正符号を生成する誤り訂正処理を実行し、ライト対象のデータ要素と、そのデータ要素に対応する誤り訂正符号とをバッファ 2 4 4 に書き込む。

40

【 0 0 4 1 】

また、E C C 回路 2 4 1 は、バッファ 2 4 4 からリード対象のデータ要素とそのデータ要素に対応する誤り訂正符号とを含むデータを読み出し、リード対象データ要素に誤りが発生しているか否かをそのデータ要素に対応する誤り訂正符号を用いて判断する。その判断の結果が肯定の場合、E C C 回路 2 4 1 は、リード対象データ要素の誤りを訂正する誤

50

り訂正処理を実行する。ECC回路241は、DRAM11に、リード対象のデータを格納する。

【0042】

本実施形態では、ECC回路241は、複数のデータバスに接続される複数のFMチップ32に対する誤り訂正符号生成処理、誤り訂正処理を担当する。なお、誤り訂正符号生成処理を実行する符号生成回路部、誤り訂正処理を実行する誤り訂正回路部は、一つであってもよく、複数であってもよい。符号生成回路部及び/又は誤り訂正回路部の数を抑えると、FM I/F制御部24の大きさを抑えることができる。なお、符号生成回路部については、回路規模が比較的小さいので、符号生成回路部は複数存在してもよい。いずれのケースでも、少なくとも誤り訂正回路部が、複数のデータバスについて共通なので、回路規模の軽減を期待することができる。

10

【0043】

制御レジスタ242は、FMチップ32に対するアクセスを制御するために必要な情報を記憶する。アクセスを制御するために必要な情報は、例えば、CPU21の制御、FMバスプロトコル制御部246により設定される。

【0044】

FM/SW制御部243は、制御レジスタ242の設定に応じて、DIMM30の複数のSW26を切り替える信号(切替信号)と、アクセス対象のFMチップ32を選択するチップイネーブル信号(CE信号)とを出力する。本実施形態では、FM/SW制御部243には、複数のCE信号線27(27-1、27-2、27-3)と、それら複数のCE信号線27に共通の、切替信号用の信号線26(切替信号線)が接続されている。各CE信号線27は、異なるSW31の配下にある異なるFMチップ32に接続される。

20

【0045】

FM/SW制御部243に接続された切替信号線26は、DIMM30の複数のSW31(SW1、SW2)に接続される。これにより、複数のSW31には、同じ切替信号が供給されることとなる。例えば、同一のDIMM30における複数のSW31の全てについて、同一番号のピンには、同一のCE信号線27に接続されたFMチップ32が接続されているとする。このケースでは、それら複数のSW31が同じ切替信号を受信すれば、各SW31の接続先を、同一のCE信号線27に接続されたFMチップ32とすることができる。故に、複数のデータ要素の書込み先を、同一のCE信号線27に接続された複数のFMチップ32にすることができ、以って、それら複数のデータ要素を並行して書き込むことが期待できる。

30

【0046】

FM/SW制御部243に接続されたCE信号線27は、複数のSW31が担当する複数のFMチップ32に対して接続されている。本実施形態では、CE信号線27-1は、FM#1-1と、FM#2-1とに接続され、CE信号線27-2は、FM#1-2と、FM#2-2とに接続され、同様に、CE信号線27-Nは、FM#1-Nと、FM#2-Nとに接続されている。このような構成により、同一のCE信号線27に接続された複数のFMチップ32には、CE信号が、ほぼ同時に供給されることとなる。このため、これらFMチップ32は、略同時に並行して動作することができる。本実施形態においては、切替信号が供給されると、SW#1及び#2では、それぞれ同一のCE信号線27に接続されているFMチップ32のデータバスを含むバスが、FM I/F制御部24に接続されるように切り替えられる。

40

【0047】

バッファ244は、FMチップ32へのライト対象データ要素及びその誤り訂正符号を一時的に記憶する。また、バッファ244は、FMチップ32から読み出したリード対象データ要素及びその誤り訂正符号を一時的に記憶する。

【0048】

DMA247は、バッファ244に記憶されたライト対象データ要素及びその誤り訂正符号を読み出して、それらをFMチップ32に書き込む。また、DMA247は、FM

50

チップ32からリード対象データ要素及びその誤り訂正符号を読み出して、それらをバッファ244に書き込む。

【0049】

FMバスプロトコル制御部246は、制御レジスタ242の設定に従って、FMチップ32に、コマンド(リードコマンド、プログラムコマンド)を発行(出力)する。また、FMバスプロトコル制御部246は、コマンドに対するFMチップ32の動作結果(ステータス)を確認し、動作結果を制御レジスタ242に設定する。

【0050】

DMA247には、データバスが接続され、FMバスプロトコル制御部246には、コマンド用の信号線が接続され、データバス、コマンド用の信号線等を含むバス25が、SW31に接続される。

10

【0051】

SW31には、切替信号線26が接続されるとともに、データバスを含むバス25が接続される。本実施形態では、SW#1及び#2には、同一の切替信号線26が接続される。また、SW31には、複数のFMチップ32に繋がるデータバスを含むバス28が接続されている。SW31は、切替信号線26により供給される切替信号に基づいて、複数のバス28のいずれか1つを選択的にバス25に接続する。このSW31によると、1つのバス25により複数のFMチップ32に対するアクセスを実行することができようになる。また、SW31は、複数のバス28のいずれか1つを選択的にバス25に接続するので、信号を送信する際におけるバスの負荷容量を抑えることができ、信号の品質を高品質に維持することができる。本実施形態においては、切替信号が供給されると、SW#1及び#2では、それぞれ同一のCE信号線27に接続されているFMチップ32のデータバスを含むバス28が、FM I/F制御部24に繋がるバス25に接続されるように切り替えられる。

20

【0052】

図5は、本実施形態に係る書込み処理の第1の例を説明する図である。

【0053】

FMコントローラ20が、ライト対象のデータ(ライトデータ)を、DRAM11へ格納し、ライトデータを、複数のデータ要素(#0~#6等)に分割し、それら複数のデータ要素を複数のFMチップ32へ転送する。ここで言う「ライトデータ」は、典型的には、ホスト200からRAIDコントローラ301が受けたデータの一部又は全部である。また、データ要素のサイズは、FMチップ32のページサイズと、ECCのサイズとに基づくサイズである。データ要素が圧縮されてページに格納される場合には、データ要素のサイズはページサイズ以上であっても良い。ページには、データ要素とECCが格納される。

30

【0054】

以下、書込み処理の具体例を詳細に説明する。

【0055】

まず、FMコントローラ20(例えばCPU21)は、SW#1を切り替えて、FMチップ#1-1とバス25-1とを接続し、バス25-1を介して、FMチップ#1-1へデータ要素#0を転送する。データ要素#0は、FMチップ#1-1のページ#000に書き込まれる。また、FMコントローラ20(例えばCPU21)は、SW#2を切り替えて、FMチップ#2-1とバス25-2とを接続し、バス25-2を介してFMチップ#2-1へデータ要素#1を転送する。データ要素#1はFMチップ#2-1のページ#100へ書き込まれる。なお、FMコントローラ20は、FMチップ#1-1及び#2-1に接続されたCE信号線27-1を介してCE信号を送信して良い。これにより、FMチップ#1-1及び#2-1にデータ要素#0及び#1を並行して書き込むことができる。

40

【0056】

同様にして、FMコントローラ20は、SW#3を切り替えて、FMチップ#3-1へ

50

データ要素# 2を転送し、SW# 4を切り替えて、FMチップ# 4 - 1ヘデータ要素# 3を転送する。なお、FMコントローラ20は、FMチップ# 3 - 1及び# 4 - 1が接続されたCE信号線27を介してCE信号を送信して良い。これにより、FMチップ# 3 - 1及び# 4 - 1にデータ要素# 2及び# 3を並行して書き込むことができる。

【0057】

続いて、FMコントローラ20は、バス25 - 1を介して、FMチップ# 1 - 1ヘデータ要素# 4を転送する。データ要素# 4はFMチップ# 1 - 1のページ001に書き込まれる。同様に、FMコントローラ20は、FMチップ# 2 - 1ヘデータ要素# 5を転送し、FM# 3 - 1ヘデータ要素# 6を転送する。

【0058】

ここで、FMコントローラ20がFMチップ# 1 - 1ヘデータ要素# 0を転送すると、バス25 - 1がビジー状態となり、ビジー状態の間はバス25 - 1を介してデータを転送することはできない。また、FMチップ# 1 - 1へ転送されたデータ要素# 0は、FM# 1 - 1内のバッファ(図示せず)へ格納された後に、ページ# 000へ書き込まれる。FM# 1 - 1は、バッファに格納されたデータ要素# 0の書き込みが完了するまで、ビジー状態となる。一般に、ライト処理の場合、FMチップ32のビジー状態の時間は、バス25のビジー状態の時間よりも長い。このため、FMコントローラ20がデータ要素# 4をFMチップ# 1 - 1に転送する際、FMチップ# 1 - 1がビジー状態である場合もあるが、この場合には、FMコントローラ20は、FMチップ# 1 - 1のビジー状態が解除されてからデータ要素# 4を転送する。

【0059】

なお、上記の流れにおいて、DRAM11上のデータ要素(例えば# 0)は、そのデータ要素の転送先のFMチップ32(例えば# 1 - 1)が接続されているSW31(例えば# 1)に接続されているバッファ244(例えば# 1)に格納される。CPU21からの指令により、プロトコル制御部246(例えば# 1)が、DMA247(例えば# 1)に起動をかける。起動したDMA247が、バッファ244(例えば# 1)内のデータ要素(例えば# 0)を、そのデータ要素の格納先のFMチップ32(例えば# 1 - 1)に転送する。データ要素(例えば# 0)がFMチップ32(例えば# 1 - 1)に書き込まれれば、そのFMチップ32(例えば# 1 - 1)から完了ステータスがプロトコル制御部246(例えば# 1)に送られる。プロトコル制御部246(例えば# 1)は、その完了ステータスを受けたことを表す情報を、制御レジスタ242に書き込んで良い。CPU21は、制御レジスタ242を参照することで、FMチップ32(例えば# 1 - 1)にデータ要素(例えば# 0)が書き込まれたことを知ることができる。

【0060】

以上のように、FMコントローラ20は、ライトデータを複数のデータ要素に分割し、連続した2以上のデータ要素を順次異なるFMチップ32に転送する。このため、バス25及びFMチップ32がビジー状態の時間を利用して、他のバス及び他のFMチップ32に対してデータを転送することができるので、効率的にデータを転送することが可能となる。

【0061】

なお、さらに別のライトデータがDRAM11に格納された場合には、FMコントローラ20は、直前回のライトデータの末端のデータ要素が格納されるページを含んだFMチップ32(例えば# 1 - 1)の次のFMチップ32(例えば# 2 - 1)から、その別のライトデータを格納していく。別のライトデータも、複数のデータ要素に分割され、それら複数のデータ要素が並行して書き込まれる。そして、FMチップ# 1 - 1、# 2 - 1、# 3 - 1及び# 4 - 1(つまり1段目のFMチップ群)の末端ページまでデータ要素が書き込まれた場合には、FMコントローラ20は、各SW31の接続先を、1段目のFMチップ群に属するFMチップ32から別の段のFMチップ群に属するFMチップ(例えば、2段目のFMチップ群に属するFMチップ# 1 - 2、# 2 - 2、# 3 - 2、4 - 2)に切り替え、そのFMチップ32にデータ要素を転送する。この後、格納されたライトデータに

10

20

30

40

50

対する上書きライト（フラッシュメモリデバイス400が提供する論理アドレス空間の同一の論理領域をライト先としたライト）が発生した場合には、FMコントローラ20は、SW31を切り替えて、n段目のFMチップ群、(n+1)段目のFMチップ群、...にデータを分散して転送することとなる（nは1以上の整数）。この処理において、FMコントローラ20は、n段目のFMチップ群のFMチップ32がビジー状態であるときに、(n+1)段目のFMチップ群のFMチップ32にデータ要素を転送して良い。

【0062】

本実施形態では、前述したように、CE信号線27が、複数のFMチップ32（正確には、異なるSW31に接続された複数のFMチップ32）で共有されている。FMコントローラ20は、SW#1を切り替えてFMチップ#1-1とバス25-1を接続し、SW#2を切り替えてFMチップ#2-1とバス25-2とを接続することで、FM#1-1とFM#2-1に並行して連続した2つのデータ要素#0及び#1を転送することができる。さらに、FMコントローラ20は、データ要素#0をFMチップ#1-1へ、データ要素#1をFMチップ#2-1へ転送した後、CE信号線27-1を介しCE信号を送信することで、FMチップ#1-1とFMチップ#2-1を同時に（並行して）起動する。これにより、FMチップ#1-1（#2-2）は、受信したデータ要素#0（#1）を書き込む。

【0063】

つまり、FMコントローラ20は、ライトデータを複数のデータ要素に分割し、SWを順次に選択し、SWの接続先を同一のCE信号線に接続されたFMチップ32とし、連続した2以上のデータ要素を、同一のCE信号に接続された2以上のFMチップ32にそれぞれ並行して転送し、その同一のCE信号線を介してCE信号を送信する。これにより、より効率良くライトデータを書き込むことができる。FMコントローラ20が複数のデータ要素を転送する際、各SWを、それぞれ独立して切り替えるよう制御してもよいし、同期して切り替えるよう制御してもよい。

【0064】

図15は、本実施形態に係る論理アドレス層と物理層との関係の一例を示す。

【0065】

論理アドレス層1401は、フラッシュメモリデバイス400が上位装置（例えば、RAIDコントローラ301又はホスト200）に提供する論理アドレス空間である。ここで、論理アドレスとは、例えば、LBA（Logical Block Address）で良い。論理アドレス空間1401は、複数の論理領域1411に分割されて管理される。

【0066】

物理層1405は、複数のFMチップ32が有する記憶空間であり、複数のブロック1452で構成されている。各ブロック1452は、複数のページ1453で構成されている。

【0067】

論理領域1411は、物理ページ1453に関連付けられる。どの論理領域1411にどのページ1453が対応しているかを表す論理物理変換情報は、例えば、FMコントローラ20が有する記憶領域（例えばDRAM11）に記憶される。その情報は、1以上のFMチップ32にバックアップされても良い。

【0068】

例えば、図15においては、LBA0x00から0x07までの論理領域1411は、FMチップ#1-1のブロック#00のページ#000に割り当てられ、LBA0x08から0x0Fまでの論理領域1411は、FMチップ#2-1のブロック#10のページ#100に割り当てられている。このように、フラッシュメモリデバイス20にとっての上位装置の1つであるRAIDコントローラ301から、LBA0x00から0x07までのいずれかのLBAを指定したリード要求が発行された場合は、FMコントローラ20は、そのリード要求を受け、そのリード要求に従い、論理物理変換情報に基づいて、ページ#000からデータ要素をリードし、そのリードしたデータ要素を上位装置に返す。

10

20

30

40

50

【 0 0 6 9 】

図 1 7 は、論理物理変換情報の構成例を示す。

【 0 0 7 0 】

論理物理変換情報 T 6 0 1 は、ページ毎に、ページを有するブロックの番号、ページの番号、ページの属性（有効ページ、無効ページ又は空きページ）、及び、ページの割当て先の論理領域の論理アドレス（例えば先頭アドレス）を含む。CPU 2 1 は、この情報 T 6 0 1 を参照することにより、どの FM チップに空きブロックがあるか、どの FM チップ群におけるブロックが最も有効ページが少ないブロックであるか、及び、どの論理領域にどのページが割り当てられているか等を特定することができる。なお、この段落で言う「FM チップ群」は、1 以上の FM チップ 3 2 であり、例えば、同一の FM I / F 制御部 2 4 に接続されている複数の FM チップ 3 2、同一の SW 3 1 に接続されている複数の FM チップ 3 2、或いは、特定の FM チップ 3 2 である。

10

【 0 0 7 1 】

ところで、RAID コントローラ 3 0 1 がホスト 2 0 0 から指定される論理アドレスと、RAID コントローラ 3 0 1 がフラッシュメモリデバイス 4 0 0 に指定する論理アドレスは同一であっても良いが、本実施形態では、それらは異なる。

【 0 0 7 2 】

以下、FM コントローラ 2 0 によるライトデータの分散ライトと、RAID コントローラ 3 0 1 による RAID のストライピングとの違いの一例を、図 1 6 を参照して説明する。

20

【 0 0 7 3 】

図 1 6 は、本実施形態に係るユーザアドレス空間と論理アドレス空間との関係の一例を示す。

【 0 0 7 4 】

ユーザアドレス空間 3 0 0 1 は、LU (Logical Unit) 番号とその論理アドレス (LBA) によって決定される。同図においては、LU 3 0 1 1 が複数あり、各 LU 3 0 1 1 は、複数の論理ブロック 3 0 2 1 で構成されている。論理ブロック 3 0 2 1 は、異なる複数のフラッシュストレージ論理空間 3 0 0 2 の同一論理アドレスの複数の論理ブロック 3 0 2 2 に割り当てられている。1 以上の論理ブロック 3 0 2 2 で前述した論理領域 1 4 1 1 (図 1 5 参照) が構成されている。フラッシュストレージ論理空間 3 0 0 2 は、典型的には、フラッシュメモリデバイス 4 0 0 が提供する論理アドレス空間 1 4 0 1 である。

30

【 0 0 7 5 】

同図によれば、ユーザアドレス空間 3 0 0 1 に関してのストライピングは、1 つの論理ブロック 3 0 2 1 が、異なる複数のフラッシュストレージ論理空間 3 0 0 2 に跨っていることを意味する。一方、フラッシュストレージ論理空間 3 0 0 2 (論理アドレス空間 1 4 0 1) に関してのストライピングは、図 1 5 によれば、アドレスの連続する 2 以上の論理領域 1 4 1 1 が、CE 信号線 2 7 を共通にする異なる 2 以上の FM チップ 3 2 に跨っていることを意味する。

【 0 0 7 6 】

なお、LU 3 0 1 1 は、Thin Provisioning に従う仮想的な LU (TP - LU) の領域に割り当てられるセグメントを含んだプール LU でも良い。プール LU は、容量プールを構成する LU であって、複数のセグメントに分割して管理される。TP - LU の領域に対してセグメントが割り当てられる。この場合、セグメントが、1 以上の論理ブロック 3 0 2 1 で構成されていて良い。

40

【 0 0 7 7 】

図 6 は、本実施形態に係る書込み処理の第 2 の例を説明する図である。

【 0 0 7 8 】

ホスト # A のデータ要素 # A (例えば、ページサイズ以下のライトデータ、又は、ライトデータにおける末尾のデータ要素) を書き込む処理と、ホスト # B のデータ要素 # B (

50

例えば、ページサイズ以下のライトデータ、又は、ライトデータにおける先頭のデータ要素)を書き込む処理とが発生した場合には、FMコントローラ20は、FMチップ32にデータ要素#Aを格納させ、そのFMチップ32と同一のCE信号線27に接続されている別のFMチップ32にデータ要素#Bを格納させるように決定し、データ要素#A及び#BをそれらのFMチップ32にほぼ同時に(並行して)書き込ませる。これにより、複数のホストからのデータを迅速に書き込むことができる。

【0079】

図7は、フラッシュメモリデバイス400の正面上方からの斜視図の一例である。

【0080】

フラッシュメモリデバイス400は、規格化された幅19インチのラックにマウントできる形状となっているフラッシュメモリデバイス400の高さは、例えば、2Uとなっている。フラッシュメモリデバイス400は、例えば、フラッシュメモリPKG10を12個(3列×4段)装填できるようになっている。

【0081】

図8は、フラッシュメモリPKG10の上面側からの斜視図の一例であり、図9は、フラッシュメモリPKG10の下面側からの斜視図の一例である。

【0082】

フラッシュメモリPKG10の上面側においては、ASICであるFMコントローラ20がそのPKG10の平面方向においてほぼ中央に配置され、その手前側及び奥側に、それぞれ2個ずつDIMM30が配置されている。また、フラッシュメモリPKG10の下面側においては、FMコントローラ20の下面の領域に対して、その手前側及び奥側にそれぞれ2個のDIMM30が配置されている。従って、フラッシュメモリPKG10には、8個のDIMMが配置されている。このように、FMコントローラ20をほぼ中央に配置しているため、FMコントローラ20から各DIMM30への配線の長さをほぼ均等にすることができる。

【0083】

図10は、DIMM30の概略構成の一例を示す。

【0084】

DIMM30は、例えば、8個のFMチップ32と、2個のSW31とを備える。1個のSW31が4個のFMチップ32へのバスの切替を行うようになっている。

【0085】

なお、DIMM30が有するFMチップ32の数及びSW31の数は、図10に示す数に限られない。

【0086】

また、DIMM30は、ECC回路34が備えてもよい。また、FM32毎にECC回路35が備えられてもよい。DIMM30または、FM32がECC回路を備える場合は、FMIF制御部24がECC回路241を備えなくてもよい。

【0087】

次に、フラッシュメモリPKG10における動作を説明する。以下では、FMIF制御部24がECC回路241を備える場合の処理を説明するが、DIMM30またはFM32がECC回路を備える場合は、DIMM30またはFM32のECC回路が誤り訂正処理を行う。

【0088】

図11は、チップリード処理のフローチャートの一例である。

【0089】

チップリード処理とは、FMチップ32からデータ要素を読み出す処理である。ここで、チップリード処理の前においては、FMコントローラ20のCPU21が、リード対象のFMチップ32を特定し、その特定したFMチップ32からデータ要素を読み出すための制御用の設定をFMIF制御部24の制御レジスタ242に対して行う。

【0090】

10

20

30

40

50

まず、FM/SW制御部243が、リード元のFMチップ32にバス25が接続されるように、SW31の切替信号を切替信号線26により送信する。これにより、SW31が接続を切り替えて、リード元のFMチップ32がバス25に接続されるようにする(ステップ1201)。

【0091】

FM/SW制御部243が、リード元のFMチップ32に繋がるCE信号線27を介してCE信号を駆動して、リード対象のFMチップ32をアクティブにする(ステップ1202)。次いで、FMバスプロトコル制御部246は、バス25を介してリードコマンドを発行する(ステップ1203)。これにより、リードコマンドは、バス25、SW31、バス28を介してリード元のFMチップ32に送信される。次いで、DMA247が、

10

リード元のFMチップ32からリード対象のデータ要素を読み出して、そのデータ要素をバッファ244に格納する(ステップ1204)。

【0092】

次いで、FMバスプロトコル制御部246は、FMチップ32からコマンドに対するステータスを取得し、ステータスを制御レジスタ242に格納する。CPU21は、制御レジスタ242を参照して、リードが正常に終了していることをチェックし(ステップ1205)、正常終了している場合には、ECC回路241により、バッファ244に読み出したデータ要素に対して誤り訂正処理を行わせ、DRAM11に転送させる(ステップ1206)。これにより、DRAM11には、読み出し対象のデータ要素が格納される。なお、これ以降に、CPU21がDRAM11からリード対象のデータを読み出して、上位

20

の装置に送信することとなる。

【0093】

図12は、チップライト処理のフローチャートの一例である。

【0094】

チップライト処理とは、FMチップ32にデータ要素を書き込む処理である。ここで、チップライト処理の前においては、FMコントローラ20のCPU21が、ライト対象のFMチップ32を特定し、その特定したFMチップ32にデータ要素をライトするための制御用の設定を、FM I/F制御部24の制御レジスタ242に対して行う。また、ライト対象のデータ要素は、CPU21により、DRAM11に格納されている。

【0095】

CPU21は、DRAM11からライト対象のデータ要素を読み出し、そのデータ要素をECC回路241に渡す。ECC回路241は、ライト対象のデータ要素に対応するECCを生成し、ライト対象のデータ要素とECCとを含むデータ(ここで、この処理フローにおいては、ライトデータという)をバッファ244へ格納する(ステップ1301)。

30

【0096】

次いで、FM/SW制御部243が、ライト先のFMチップ32にバス25が接続されるように、SW31の切替信号を切替信号線26により送信する。これにより、SW31が接続を切り替えて、ライト先のFMチップ32がバス25に接続されるようにする(ステップ1302)。

40

【0097】

FM/SW制御部243が、ライト先のFMチップ32に繋がるCE信号線を介してCE信号を駆動して、ライト先のFMチップ32をアクティブにする(ステップ1303)。次いで、FM I/Fサブ制御部246は、バス25を介してプログラムコマンド(ライトコマンド)を発行する(ステップ1304)。これにより、プログラムコマンドは、バス25、SW31、バス28を介してライト先のFMチップ32に送信される。次いで、DMA247が、バッファ244からライトデータを読み出して、そのデータをFMチップ32に転送する(ステップ1305)。

【0098】

次いで、FMバスプロトコル制御部246は、FMチップ32からコマンドに対するス

50

データを取得し、ステータスを制御レジスタ242に格納する。CPU21は、制御レジスタ242を参照して、ライトが正常に終了していることをチェックし(ステップ1306)、正常終了している場合には、処理を終了する。

【0099】

図13は、チップ多重ライト処理のフローチャートの一例である。

【0100】

チップ多重ライト処理とは、複数のFMチップに複数のデータ要素を並行して書き込む処理である。ここで、チップ多重ライト処理の前においては、FMコントローラ20のCPU21が、ライト先の複数のFMチップ32を特定し、それらのFMチップ32にライトするための制御用の設定を、FM I/F制御部24の制御レジスタ242に対して行う。本実施形態では、同一のCE信号線27に接続されている複数のFMチップ32がライト先として特定される。また、ライト対象のデータ要素は、CPU21により、DRAM11に格納されている。

【0101】

CPU21は、FMチップ32(ここでは、例えば、SW#1に接続されているFMチップ#1-1とする)にライトするライト対象のデータ要素をDRAM11から読み出して、そのデータ要素をECC回路241に渡す。ECC回路241は、そのライト対象のデータ要素に対応するECCを生成し、ライト対象のデータ要素とECCとを含むデータ(この処理フローにおいては、ライトデータという)をバッファ#1へ格納する(ステップ1401)。次いで、FMチップ#1-1と同じCE信号線27-1に接続されているFMチップ32(例えば、SW#2に接続されているFMチップ#2-1)にライトするライト対象のデータ要素をDRAM11から読み出して、そのデータ要素をECC回路241に渡す。ECC回路241は、そのライト対象のデータ要素に対応するECCを生成し、そのライト対象のデータ要素とECCとを含むデータ(ライトデータ)をバッファ#2へ格納する(ステップ1402)。

【0102】

次いで、FM/SW制御部243が、ライト先の複数のFMチップ#1-1及び#2-1にバス25-1及び25-2が接続されるように、SW#1及び#2の切替信号を切替信号線26により送信する。これにより、SW#1及び#2が、接続を切り替えて、ライト先の複数のFMチップ#1-1及び#2-1が、バス25-1及び25-2に接続されるようにする(ステップ1403)。

【0103】

FM/SW制御部243が、ライト先のFMチップ#1-1及び#2-1に繋がるCE信号線27-1を介してCE信号を駆動して、そのCE信号線27-1に接続されている複数のFMチップ32をアクティブにする(ステップ1404)。

【0104】

次いで、FM I/Fサブ制御部#1は、バス25-1を介してプログラムコマンド(ライトコマンド)を発行する(ステップ1405)。これにより、プログラムコマンドは、バス25-1、SW#1、バス28を介して、ライト先のFMチップ#1-1に送信される。また、これと並行して、FM I/Fサブ制御部#2は、バス25-2を介してプログラムコマンド(ライトコマンド)を発行する(ステップ1406)。これにより、プログラムコマンドは、バス25-2、SW#2、バス28を介して、ライト先のFMチップ#2-1に送信される。

【0105】

次いで、DMA#1が、バッファ#1からライトデータを読み出し、そのデータをFMチップ#1-1に転送するとともに、ほぼ同時に(並行して)、DMA#2が、バッファ#2からライトデータを読み出し、そのデータをFMチップ#2-1に転送する(ステップ1307)。

【0106】

次いで、FMバスプロトコル制御部246は、FMチップ#1-1及び#2-1からコ

10

20

30

40

50

マンドに対するステータスを取得し、ステータスを制御レジスタ242に格納する。CPU21は、制御レジスタ242を参照して、ライトが正常に終了していることをチェックし(ステップ1308)、正常終了している場合には、処理を終了する。

【0107】

この多重ライト処理によると、複数のFMチップ32に対して、ほぼ同時に(並行して)、複数のデータ要素をライトすることができるので、ライト処理に要する時間を短縮することができる。

【0108】

なお、チップライト処理及びチップ多重ライト処理のいずれにおいても、消去回数を均等にすべくウェアレベリング処理を行うことができる。ウェアレベリング処理は、それらのライト処理とは非同期に行われても良い。

【0109】

ライト処理と非同期に行われるウェアレベリング処理によれば、例えば、任意のタイミングで、FMコントローラ20(例えばCPU21)が、消去回数が最も多いブロックを選択し、その選択したブロック内の有効ページから、消去回数が最も少ないブロックに有効データを移動する。移動元のブロックと移動先のブロックは、同一のFMチップ32にあっても良いし、異なるFMチップ32にあっても良い。後者の場合、異なるFMチップ32は、CE信号線27が共通であることが望ましい。

【0110】

ライト処理において行われるウェアレベリング処理によれば、FMコントローラ20は、ライト先のFMチップ32から消去回数の最も少ないブロックを書き込み先として選択し、そのブロックにデータ要素を書き込む。なお、この処理では、例えば、図13或いは図14において、FMコントローラ20は、CE信号線27を選択する段階で、最も消去回数合計が少ないCE信号線27を選択し、そのCE信号線27に接続されているFMチップ32から、そのFMチップ32において最も消去回数が少ないブロックが選択して良い。FMコントローラ20の記憶領域(例えばDRAM11)は、図19に例示する消去管理情報1901を記憶して良い。この情報1901は、CE信号線27毎及びブロック毎に、ブロックの消去回数を表す。CE信号線27の消去回数合計とは、そのCE信号線27を共通にする全てのFMチップ32の全てのブロックの消去回数の合計である。ブロックに対して消去処理を行った場合、FMコントローラ20は、そのブロックに対応する消去回数と、そのブロックを有するFMチップ32が接続されているCE信号線27に対応する消去回数合計とを更新して良い。この情報1901から、各ブロックの消去回数と、各CE信号線27についての消去回数合計とを特定することができる。

【0111】

次に、リクラメーション処理、すなわち、消去処理可能なブロックを生成する処理を説明する。

【0112】

図18は、リクラメーション処理のフローチャートの一例である。

【0113】

リクラメーション処理は、例えば、FMチップ32における利用可能な容量の枯渇をFMコントローラ20が検出したことを契機として、FMコントローラ20により実行される。容量の枯渇とは、空きブロックの数が所定割合(所定数)未満になったことを意味する。容量枯渇の検出は、任意の単位でよく、或るDIMM上の複数のFM毎でもよい。リクラメーション処理は、ライト処理において空きブロックが枯渇していることが検出されたときに開始されても良いし、ライト処理とは非同期に行われても良い。

【0114】

FMコントローラ20は、空きブロックが枯渇しているFMチップ(以下、図18において「空き枯渇チップ」と言う)32から、移動元のブロックを選択する(ステップ1701)。ここで、移動元のブロックは、空き枯渇チップ32(或いは、空き枯渇チップ32において末端ページまでデータが書き込まれているブロック(消去候補ブロック))の

10

20

30

40

50

うち、有効ページが最も少ないブロックであることが望ましい。なぜなら、移動させる有効データの総量が最も少なくして済み、以って、リクレーション処理にかかる時間及び負荷を抑えることが期待できるからである。なお、移動元ブロックは、空き枯渇チップ32以外のFMチップ32から選択されても良い。

【0115】

CPU21は、ステップ1701で選択した移動元ブロックを有するFMチップ32とデータ通信可能なFMIF制御部24（又は、バス25或いはSW31）と同一の制御部24（又は、バス25或いはSW31）に接続されている複数のFMチップ32に空きブロックが所定数以上あるか否かを判断する（ステップ1702）。ここで言う「所定数」は、全てのFMチップ32で一律であっても良いし異なっても良い。

10

【0116】

ステップ1702の判断の結果が肯定的であれば、CPU21は、同一の制御部24（又は、バス25或いはSW31）に接続されている複数のFMチップ32の空きブロックを移動先ブロックとして選択する（ステップ1703）。同一のバス25或いはSW31に接続されている複数のFMチップ32の空きブロックが優先的に移動先ブロックとして選択されて良い。もし、同一のバス25或いはSW31に接続されている複数のFMチップ32に所定数の空きブロックが無ければ、同一の制御部24における異なるバス25或いはSW31に接続されている複数のFMチップ32の空きブロックが移動先ブロックとして選択されても良い。その方が、ストライプ（アドレスの連続した2以上のデータ要素が同一CE信号線27の異なるFMチップ32（バス25が異なるFMチップ32）に配置されること）をより維持し易いと考えられる。例えば、図5において、リクレーション処理により、FM#1-1に格納されたデータ要素#0がFM#3-1に格納されると、データ要素#0とデータ要素#2が同じFMに存在することになる。この状態で、データ要素のリード/ライトが発生すると、バス25-3及びFM#3-1のビジー時間が重なるためリード/ライトに時間がかかる。リクレーションの範囲を制限することで、データ要素のストライプ状態が維持され、以降のリード/ライト処理においても効率の良いデータ転送ができる。

20

【0117】

ステップ1703の判断の結果が否定的であれば、CPU21は、異なる制御部24（又は、バス25或いはSW31）に接続されている複数のFMチップ32の空きブロックを移動先ブロックとして選択する（ステップ1704）。

30

【0118】

ステップ1703又は1704の後、すなわち、移動元ブロックと移動先ブロックとが決定した後、CPU21は、移動元ブロック内の有効データを移動先ブロックに移動する（ステップ1705）。すなわち、CPU21は、移動元ブロックから有効データを読み出してDRAM11に書き込み、その有効データをDRAM11から移動先ブロックに書き込む。この際、ECC回路241が誤り訂正処理を行う。なお、CPU21は、移動先ブロックを決定する前に移動元ブロックから有効データを読み出しDRAM11に書き込んで良い。また、DIMM30がECC回路を備える場合、DIMM30において誤り訂正処理が可能であるため、移動元ブロックのデータをDRAM11に格納することなく移動先ブロックにデータを転送することができる。同様に、FM32がECC回路を備える場合、FM32において誤り訂正処理が可能であるため、移動元ブロックのデータをDRAM11に格納することなく移動先ブロックにデータを転送することができる。このように、DIMM30またはFM32にECC回路がある場合は、FMIF制御部24がデータ転送を実行するため、CPU21の処理負荷を低減することができる。

40

【0119】

移動元ブロックから読み出された有効データが移動先ブロックに書き込まれると、移動元ブロック内のデータは全て無効データとなる。CPU21は、移動元ブロックに対して消去処理を行う（ステップ1706）。これにより、移動元ブロックが空きブロックとして管理され、再びライト先として選択され得る状態となる。なお、その消去処理の際、図

50

19に例示した情報1900が更新されて良い。すなわち、移動元ブロックに対応する消去回数と、移動元ブロックを有するFMチップ32が接続されているCE信号線27に対応する消去回数合計とが更新されて良い。

【0120】

以上が、本実施形態に係るリクラメーション処理である。

【0121】

なお、ステップ1703又は1704では、選択されたFMチップ32のうち消去回数が最も少ない空きブロックが移動先ブロックとして選択されることが望ましい。これにより、精度の高い消去回数の平準化が可能となる。また、複数のブロックを消去回数に応じて複数のグループに分けて、消去回数が少ないグループからブロックを選択することとしてもよい。この場合、ブロックを検索する時間が短縮される。

10

【0122】

また、ステップ1701で、複数のブロックが移動元ブロックとして選択されても良いし、ステップ1703及び1704で、複数の空きブロックが移動先ブロックとして選択されても良い。この場合、複数の移動元ブロックは、CE信号線27を共通にする複数のFMチップ32から選択されることが望ましいが、異なるCE信号線27に接続されている複数のFMチップ32から選択されても良い。なぜなら、チップリード処理はチップライト処理と比べてチップビジーとなる時間が短いから、つまり、性能への影響が小さいからである。一方、複数の移動先ブロックは、CE信号線27を共通にする複数のFMチップ32から選択されることが望ましい。そして、有効データの移動(ステップ1705)の際には、CE信号線27を共通にする複数のFMチップ32に複数の有効データが並行して書き込まれることが望ましい。また、ステップ1706において消去処理を実行する際、同一CE27上に消去可能なブロックが存在する場合、それらも同時に(並行して)消去することができる。一般に消去処理は時間がかかるため、複数ブロックに対してまとめて消去処理を実行するのが効率的である。

20

【0123】

次に、本実施形態に係るリフレッシュ処理を説明する。

【0124】

リフレッシュ処理は、有効ページを有するブロックに対して定期的(例えば、そのブロックの前のリフレッシュ処理から30日経過した場合)に行われても良いし、リード時のECCエラーが所定bit数以上のブロックがある場合にそのブロックを対象に行われても良い。リフレッシュ処理は、FMコントローラ20により実行される。

30

【0125】

リフレッシュ処理では、FMコントローラ20は、リフレッシュ処理の対象のブロックを移動元のブロックとする。その後、(1)移動先のブロックの決定、(2)移動元ブロックから移動先ブロックへの有効データの移動、及び(3)移動元ブロックに対する消去処理が行われる。これら(1)~(3)については、上述したリクラメーション処理と同様である。

【0126】

以上、一実施形態を説明したが、本発明は、この実施形態に限定されるものでなく、その趣旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

40

【0127】

例えば、上記実施形態では、不揮発半導体記憶媒体の一例として、NAND型のフラッシュメモリが採用されているが、不揮発半導体記憶媒体は、これに限られない。例えば、記憶媒体は、相変換メモリでもよい。

【0128】

また、上記実施形態では、複数のFMチップ32を搭載するメモリモジュールは、DIMM30であるが、DIMM以外のメモリモジュールが採用されもよい。

【0129】

また、上記実施形態では、同一のDIMM30における複数のFMチップ32が同一の

50

C E 信号線 2 7 で接続されるが、異なる D I M M 3 0 の複数の F M チップ 3 2 が同一の C E 信号線 2 7 で接続されても良い。

【 0 1 3 0 】

また、上記実施形態では、フラッシュメモリデバイス 4 0 0 では、並行して複数の F M チップ 3 2 に複数のデータ要素を書き込むことができる。すなわち、時間当たり書き込めるデータの量が大きい。このため、R A I D コントローラ 3 0 1 は、フラッシュメモリデバイス 4 0 0 内のデータ量（又は、フラッシュメモリデバイス 4 0 0 間におけるデータ転送における転送単位のデータ量）を、フラッシュメモリデバイス 4 0 0 以外の種類の記憶デバイス（例えば、S S D デバイス 5 0 0、H D D デバイス（S A S）6 0 0、又は H D D デバイス（S A T A）7 0 0）が関わるデータ転送における転送単位のデータ量よりも大きくするように制御してもよい。

10

【 0 1 3 1 】

また、例えば、フラッシュメモリ P K G 1 0 の構成は、図 1 4 に示す構成でも良い。図 1 4 に示すフラッシュメモリ P K G は、C E 信号線 2 7 についても S W 3 3 で切り替えることができるようになっている。F M / S W 2 4 8 には、C E 信号線 2 7 が接続されている。C E 信号線 2 7 は、バス 2 5 とともに、S W # 1 に接続されている。C E 信号線 2 7 は、S W # 2 にも接続されている。S W 3 3（# 1、# 2）は、C E 信号線及びデータバスを含むバス 2 9 により複数の F M チップ 3 2 に接続されている。S W 3 3 は、切替信号線 2 6 により供給される切替信号に基づいて、複数のバス 2 9 のいずれか一つを選択的にバス 2 5 及び C E 信号線 2 7 に接続する。この構成によれば、F M I / F 制御部 2 4 において出力する C E 信号線 2 7 の本数を低減することができ、F M I / F 制御部 2 4 のチップサイズを小型化することができる。

20

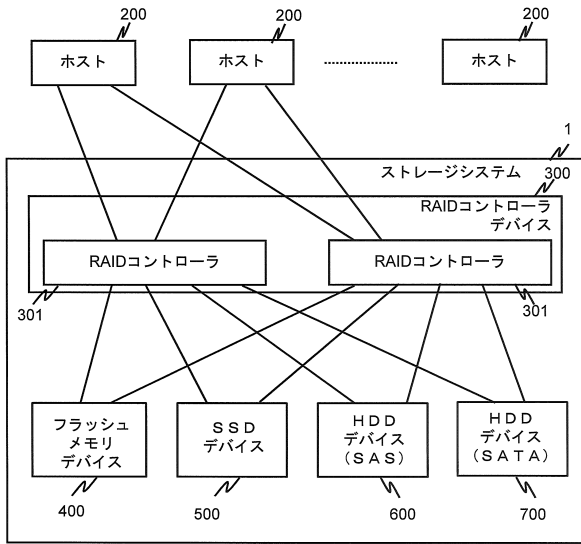
【 符号の説明 】

【 0 1 3 2 】

1 ... ストレージシステム、1 0 ... フラッシュメモリ P K G、4 0 0 ... フラッシュメモリデバイス

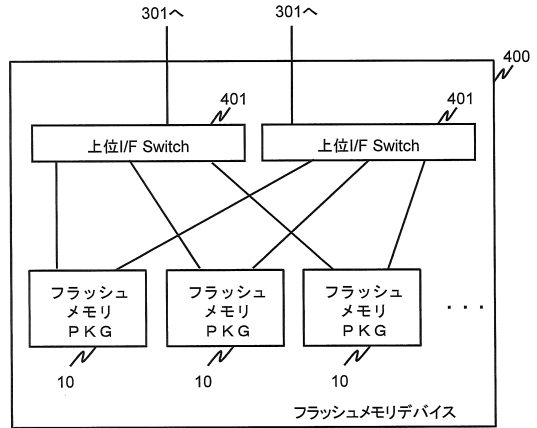
【 図 1 】

FIG. 1



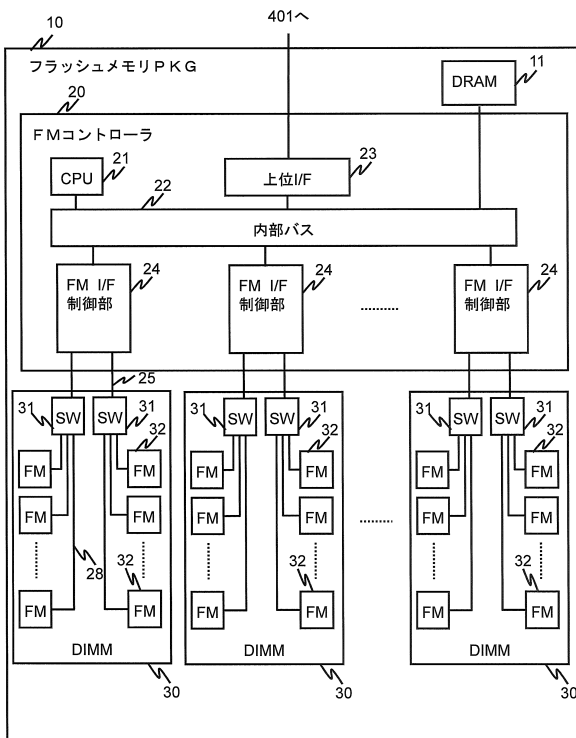
【 図 2 】

FIG. 2



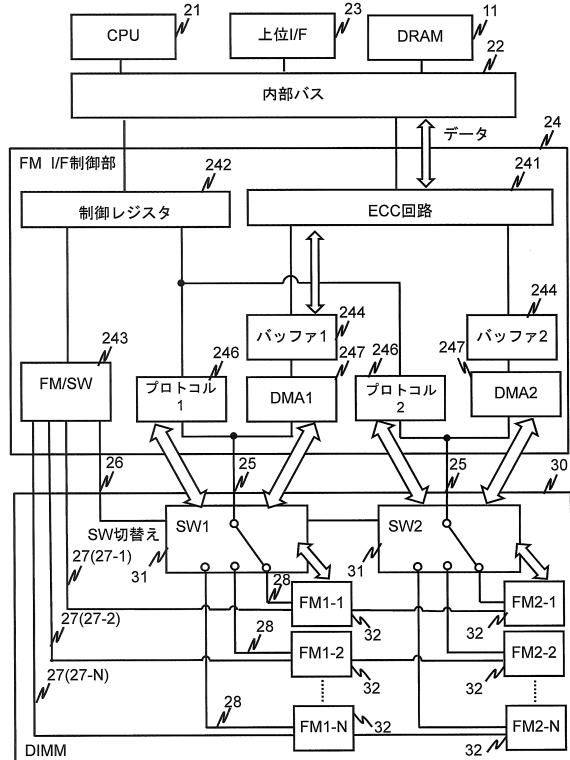
【 図 3 】

FIG. 3



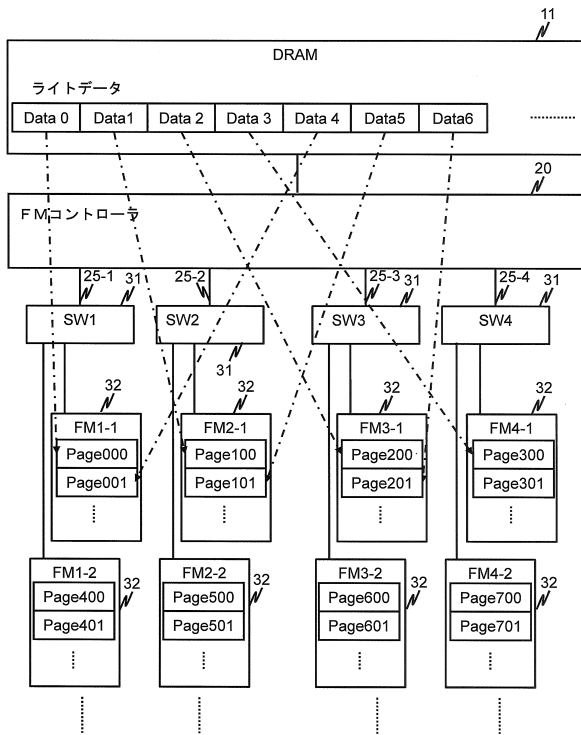
【 図 4 】

FIG. 4



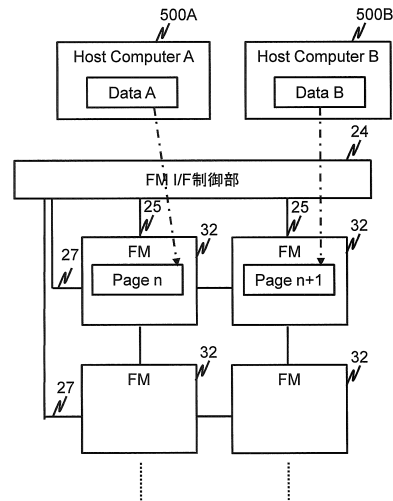
【 図 5 】

FIG. 5



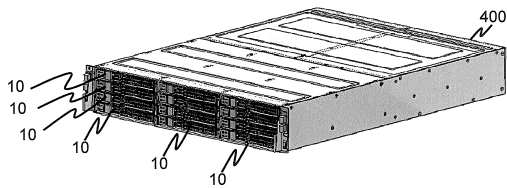
【 図 6 】

FIG. 6



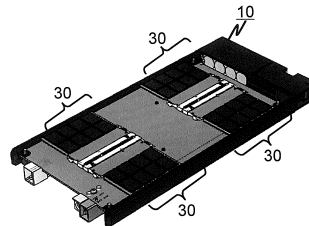
【 図 7 】

FIG. 7



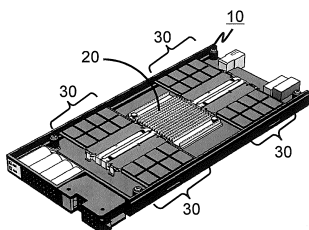
【 図 9 】

FIG. 9



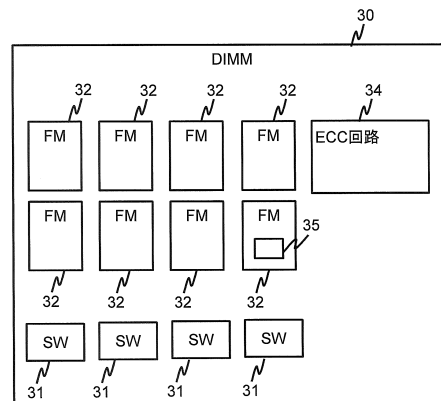
【 図 8 】

FIG. 8



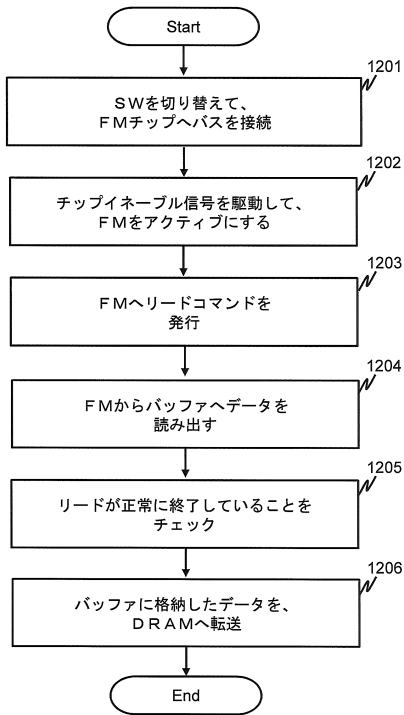
【 図 10 】

FIG. 10



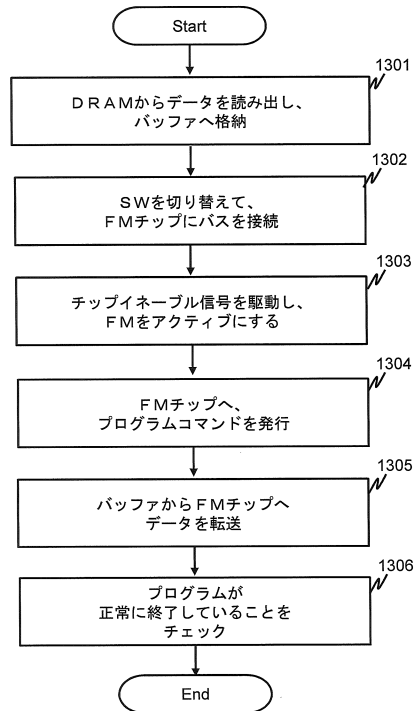
【 図 1 1 】

FIG. 11



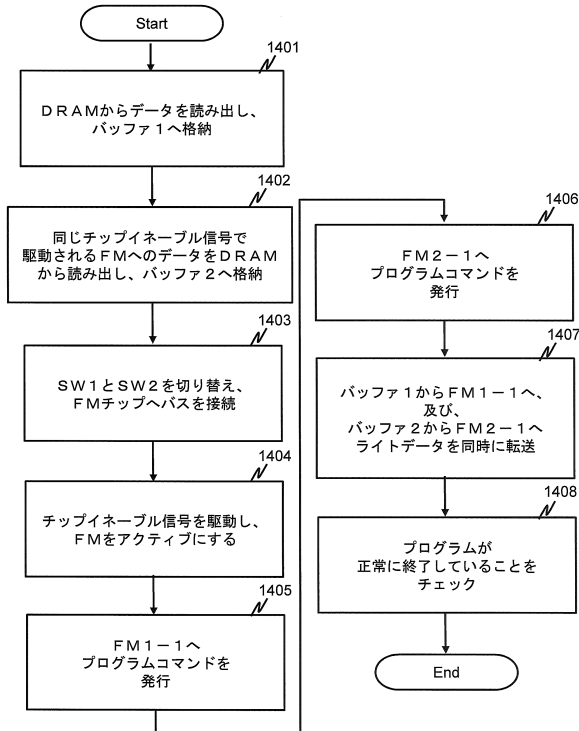
【 図 1 2 】

FIG. 12



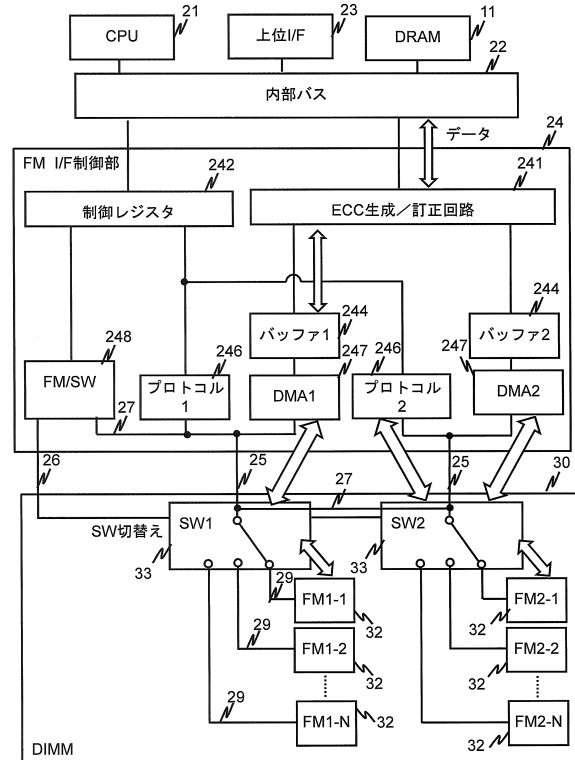
【 図 1 3 】

FIG. 13



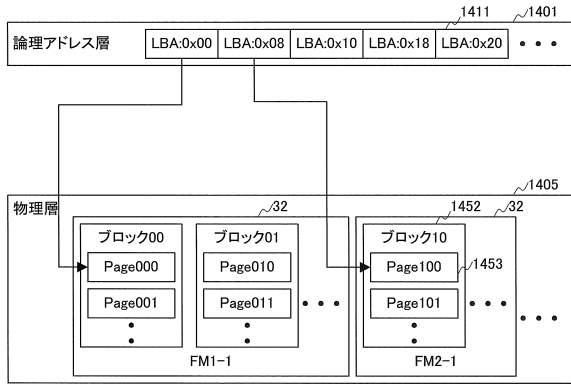
【 図 1 4 】

FIG. 14



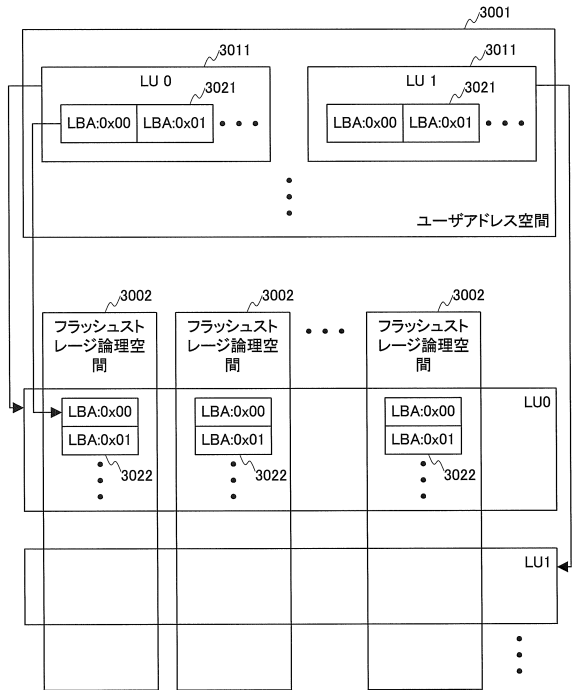
【図15】

FIG. 15



【図16】

FIG. 16



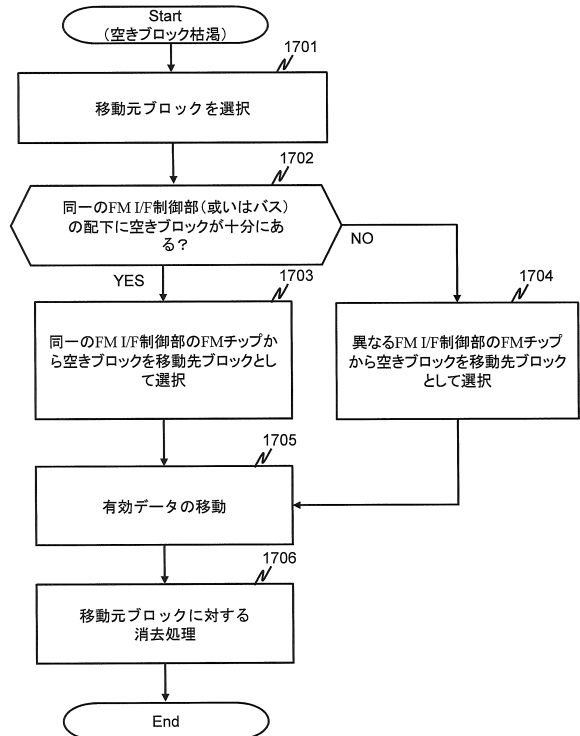
【図17】

FIG. 17

ブロック番号	ブロック内ページ番号	属性	アドレス
00	000	有効	0x00
	001	空き	なし
	002	空き	なし
	003	空き	なし
⋮			
10	100	有効	0x08
	101	空き	なし
	102	空き	なし
	103	空き	なし
⋮			

【図18】

FIG. 18



【図19】

FIG. 19

1900
↙

CE信号線 番号	FMチップ 番号	ブロック 番号	ブロック 消去回数	合計
0	0	0	30	500
		1	40	
		⋮	⋮	
	1	⋮		
	⋮	⋮		
1	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

フロントページの続き

(51)Int.Cl. F I
G 0 6 F 13/14 (2006.01) G 0 6 F 13/14 3 1 0 F

(72)発明者 小川 純司
日本国東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内

(72)発明者 小関 英通
日本国東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内

審査官 桜井 茂行

(56)参考文献 特開2010-049586(JP,A)
特表2008-523528(JP,A)
特開2005-258874(JP,A)

(58)調査した分野(Int.Cl., DB名)
G 0 6 F 1 3 / 1 6 - 1 3 / 1 8
G 0 6 F 1 2 / 0 0 - 1 2 / 0 6
G 0 6 F 3 / 0 6
G 0 6 F 1 2 / 1 6
G 0 6 F 1 3 / 1 4