(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2010-15195
(P2010-15195A)
(43)公開日　平成22年1月21日(2010.1.21)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 12/16	(2006.01)	G06F	12/16	320L	5B018
G06F 3/06	(2006.01)	G06F	12/16	320M	5B065
		G06F	3/06	540	
		G06F	3/06	305F	
		G06F	3/06	305C	

審査請求 未請求　請求項の数 7　OL　（全 17 頁）

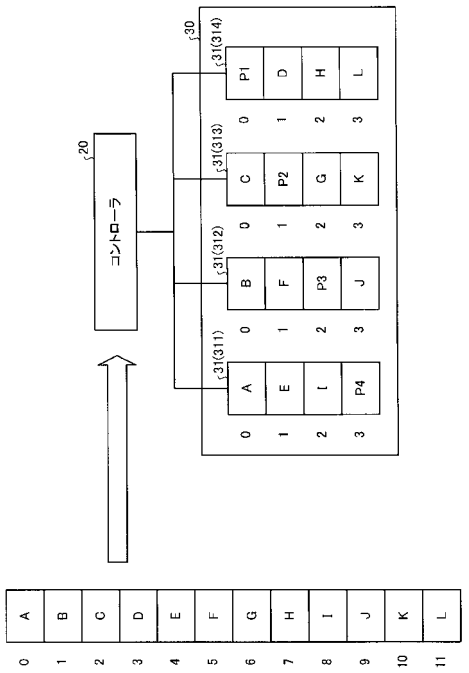
(21)出願番号　特願2008-171800 (P2008-171800)
(22)出願日　平成20年6月30日 (2008.6.30)

(71)出願人　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(74)代理人　100089118
弁理士　酒井　宏明
(72)発明者　福冨　和弘
東京都港区芝浦一丁目１番１号　株式会社東芝内
(72)発明者　佐藤　英昭
東京都港区芝浦一丁目１番１号　株式会社東芝内
(72)発明者　菅野　伸一
東京都港区芝浦一丁目１番１号　株式会社東芝内

最終頁に続く

(54)【発明の名称】記憶制御装置及び記憶制御方法

(57)【要約】
【課題】ＲＡＩＤ構成とされた複数の不揮発性半導体記憶装置を有効に活用することが可能な記憶制御装置及び記憶制御方法を提供する。
【解決手段】複数の不揮発性半導体記憶装置を用いて、当該不揮発性半導体記憶装置に記憶されるデータを復元可能なＲＡＩＤを構成し、外部から入力されるデータの読み込み要求に応じ、前記ＲＡＩＤを構成する不揮発性半導体記憶装置からデータを読み込み、この読み込み時にエラーが発生した場合には、当該読み込みエラーが発生したデータを復元し、前記読み込みエラーが発生した前記不揮発性半導体記憶装置上の領域に書き戻す。
【選択図】　図２



【特許請求の範囲】
【請求項１】
複数の不揮発性半導体記憶装置を接続可能なインタフェースと、
前記複数の不揮発性半導体記憶装置を用いて、記憶対象となるデータを、当該データを復元可能な復元情報とともに記憶するためのＲＡＩＤを構成する構成手段と、
外部から入力されるデータの読み込み要求に応じ、前記ＲＡＩＤを構成する不揮発性半導体記憶装置からデータを読み込む読込手段と、
前記読込手段による読み込み時にエラーが発生したデータを、前記復元情報に基づいて復元する復元手段と、
前記復元手段により復元されたデータを一時的に保存する保存手段と、
前記保存手段に保存されたデータを、当該データの読み込みエラーが発生した前記不揮発性半導体記憶装置上の領域に書き込む書込手段と、
を備えたことを特徴とする記憶制御装置。
【請求項２】
外部から入力されるデータの書き込み要求に応じ、前記ＲＡＩＤを構成する不揮発性半導体記憶装置にデータを書き込む書込手段を更に備え、
前記構成手段は、前記書込手段による書き込み時にエラーが発生した不揮発性半導体記憶装置を、前記ＲＡＩＤの構成から除外することを特徴とする請求項１に記載の記憶制御装置。
【請求項３】
前記構成手段は、前記書き込みエラーが発生した不揮発性半導体記憶装置以外の他の不揮発性半導体記憶装置により、前記ＲＡＩＤの構成を維持することが可能な場合に、前記書き込みエラーが発生した不揮発性半導体記憶装置を、前記ＲＡＩＤの構成から除外することを特徴とする請求項２に記載の記憶制御装置。
【請求項４】
前記不揮発性半導体記憶装置は、自己の記録媒体に記憶されたデータのリフレッシュを行うことを特徴とする請求項１～３の何れか一項に記載の記憶制御装置。
【請求項５】
前記不揮発性半導体記憶装置は、ＮＡＮＤ型フラッシュメモリを記録媒体として有することを特徴とする請求項１～４の何れか一項に記載の記憶制御装置。
【請求項６】
前記構成手段は、前記複数の不揮発性半導体記憶装置を、ＲＡＩＤ１、５、６又はこれらの組み合わせとした構成とすることを特徴とする請求項１～３の何れか一項に記載の記憶制御装置。
【請求項７】
複数の不揮発性半導体記憶装置を用いて、記憶対象となるデータを、当該データを復元可能な復元情報とともに記憶するためのＲＡＩＤを構成する記憶制御装置の記憶制御方法であって、
読込手段が、外部から入力されるデータの読み込み要求に応じ、前記ＲＡＩＤを構成する不揮発性半導体記憶装置からデータを読み込む読込工程と、
復元手段が、前記読込工程で読み込みエラーが発生したデータを、前記復元情報に基づいて復元する復元工程と、
保存手段が、前記復元工程で復元されたデータを一時的に保存する保存工程と、
書込手段が、前記保存工程に保存されたデータを、当該データの読み込みエラーが発生した前記不揮発性半導体記憶装置上の領域に書き込む書込工程と、
を含むことを特徴とする記憶制御方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ＲＡＩＤ構成とされた複数の不揮発性半導体記憶装置を制御する記憶制御装

置及び記憶制御方法に関する。
【背景技術】
【０００２】
従来、サーバ環境等で使用されるストレージシステムでは、耐障害性・冗長性を向上させるため、複数の磁気ディスク装置を用いることで、ＲＡＩＤ（Redundant Array of Independent/Inexpensive Disks）を構成することが行われている（例えば、非特許文献１参照）。例えば、ＲＡＩＤ５の構成では、３つ以上の磁気ディスク装置を使用し、データの復旧のためのパリティを当該データとともに各磁気ディスク装置に分散して記憶することで、データが破損した場合であっても、このパリティを用いることでデータを復旧することが可能である。このようなＲＡＩＤ構成を実現するコントローラでは、ある磁気ディスク装置からデータの読み込みが不可能になると、当該磁気ディスク装置が故障したと判断し、ＲＡＩＤ構成から除外することが一般に行われており、除外された磁気ディスク装置を新たな磁気ディスク装置に交換することで、従前のＲＡＩＤ構成を回復することが可能となっている。
【０００３】
一方、不揮発性の半導体記憶素子を記録媒体として用いたＳＳＤ（Solid State Drive）等の不揮発性半導体記憶装置が存在しており、磁気ディスク装置と同様に補助記憶装置（二次記憶装置）として使用することも行われている。この、不揮発性半導体記憶装置では、磁気ディスク装置のようにディスクを持たないため、データの読み書きが磁気ディスク装置に比べて高速であり、消費電力を抑えることができるため、サーバ環境での利用も期待されている。
【０００４】
記録媒体としてＮＡＮＤ型フラッシュメモリが採用された不揮発性半導体記憶装置では、データの読み込み時に生じる電荷変位や自然放電等により、記憶したデータが劣化し、正常に読み込みすることができなくなる可能性がある。そのため、このような不揮発性半導体記憶装置では、記憶されているデータを読み出して誤り訂正を行った後に再びＮＡＮＤ型フラッシュメモリに書き戻すリフレッシュ処理を所定時間毎に行うことで、データの損失を防ぐ機構が備えられている。なお、この場合、記憶素子（メモリセル）自体は故障しておらず、データが劣化しているだけであるため、正常なデータを再度書き込むことにより、当該記憶素子を再び使用することが可能となる。
【０００５】
【非特許文献１】D. Patterson, G. Gibson, and R. Katz. “A Case for Redundant Array of Inexpensive Disks (RAID)” ,Proceedings of the 1988 ACM SIGMOD, pp.109-116, June 1988.
【発明の開示】
【発明が解決しようとする課題】
【０００６】
ところで、劣化の程度によっては上記リフレッシュ処理により全てのデータを復元できるとは限らず、この場合、復元に失敗したデータの読み込みについては、読み込みエラーが発生することになる。そのため、不揮発性半導体記憶装置を用いて上記ＲＡＩＤを構成すると、従来技術と同様、読み込みエラーが発生した際にはその不揮発性半導体記憶装置が故障と判断されることになる。なお、上述したように読み込みエラーの発生したデータの格納領域に正常なデータを再度書き込むことで、故障と判断された不揮発性半導体記憶装置を復旧できる可能性があるが、上記従来技術は、磁気ディスク装置での使用が前提となるため、当該不揮発性半導体記憶装置を復旧することはできない。また、復旧可能な不揮発性半導体記憶装置であっても、ＲＡＩＤ構成から除外され、交換の対象とされてしまう可能性があるため、従来技術では、不揮発性半導体記憶装置を有効に活用できないという問題がある。
【０００７】
本発明は上記に鑑みてなされたものであって、ＲＡＩＤ構成とされた複数の不揮発性半

導体記憶装置を有効に活用することが可能な記憶制御装置及び記憶制御方法を提供することを目的とする。

【課題を解決するための手段】

【０００８】

上述した課題を解決し、目的を達成するために、本発明は、複数の不揮発性半導体記憶装置を接続可能なインタフェースと、前記複数の不揮発性半導体記憶装置を用いて、記憶対象となるデータを、当該データを復元可能な復元情報とともに記憶するためのＲＡＩＤを構成する構成手段と、外部から入力されるデータの読み込み要求に応じ、前記ＲＡＩＤを構成する不揮発性半導体記憶装置からデータを読み込む読込手段と、前記読込手段による読み込み時にエラーが発生したデータを、前記復元情報に基づいて復元する復元手段と、前記復元手段により復元されたデータを一時的に保存する保存手段と、前記保存手段に保存されたデータを、当該データの読み込みエラーが発生した前記不揮発性半導体記憶装置上の領域に書き込む書込手段と、を備える。

【０００９】

また、本発明は、複数の不揮発性半導体記憶装置を用いて、記憶対象となるデータを、当該データを復元可能な復元情報とともに記憶するためのＲＡＩＤを構成する記憶制御装置の記憶制御方法であって、読込手段が、外部から入力されるデータの読み込み要求に応じ、前記ＲＡＩＤを構成する不揮発性半導体記憶装置からデータを読み込む読込工程と、復元手段が、前記読込工程で読み込みエラーが発生したデータを、前記復元情報に基づいて復元する復元工程と、保存手段が、前記復元工程で復元されたデータを一時的に保存する保存工程と、書込手段が、前記保存工程に保存されたデータを、当該データの読み込みエラーが発生した前記不揮発性半導体記憶装置上の領域に書き込む書込工程と、を含む。

【発明の効果】

【００１０】

本発明によれば、ＲＡＩＤ構成とされた不揮発性半導体記憶装置に読み込みエラーが発生した場合であっても、この読み込みエラーが発生したデータを復元し、当該読み込みエラーが発生した不揮発性半導体記憶装置上の領域に書き戻すことで、当該不揮発性半導体記憶装置を正常な状態に復旧させることができるため、ＲＡＩＤ構成とされた複数の不揮発性半導体記憶装置を有効に活用することができる。

【発明を実施するための最良の形態】

【００１１】

以下、添付図面を参照して、本発明の最良な実施形態を詳細に説明する。なお、本発明は以下の記述に限定されるものではなく、本発明の要旨を逸脱しない範囲において適宜変更可能である。

【００１２】

［第１の実施形態］

図１は、第１の実施形態にかかるストレージシステム１００の概略構成を示したブロック図である。同図に示したように、ストレージシステム１００は、ホスト装置１０と、コントローラ２０と、複数の不揮発性半導体記憶装置３１から構成されるストレージ装置３０とを有している。

【００１３】

ホスト装置１０は、ＰＣ（Personal Computer）等であって、コントローラ２０に対してデータの書き込みや読み込みを要求する指示情報を出力する。以下、データの書き込みを要求する指示情報を「書き込み要求」と呼び、データの読み込みを要求する指示情報を「読み込み要求」と呼ぶ。なお、ホスト装置１０からコントローラ２０に出力される書き込み要求には、少なくとも書き込み対象のデータが含まれているものとし、また、読み込み要求には、読み込み先となるストレージ装置３０のアドレス情報（例えば、ＬＢＡ：Logical Block Addressing）が含まれているものとする。

【００１４】

コントローラ２０は、ストレージ装置３０を構成する複数の不揮発性半導体記憶装置３

１を、ＲＡＩＤ技術を用いて管理し、これら複数の不揮発性半導体記憶装置３１により論理的に構成される記憶領域に対し、データの書き込みや読み出しをホスト装置１０からの要求に応じて実行する。

【００１５】

具体的に、コントローラ２０は、複数の不揮発性半導体記憶装置３１をＲＡＩＤ１、５、６の何れか又はこれらの組み合わせとした構成とすることで、ストレージ装置３０の耐障害性・冗長性を実現している。以下、本実施形態ではストレージ装置３０をＲＡＩＤ５の構成とした態様について説明する。

【００１６】

ＲＡＩＤ５は、パリティと呼ばれる誤り訂正符号の記憶用に割り当てる記憶装置と、データの記憶用に割り当てる記憶装置とを、ストライプ毎に順次変更するものである。ＲＡＩＤ５を実装するディスクアレイ装置では、耐障害性の向上、大容量化、リード処理の高速化が実現できる。

【００１７】

図２は、ＲＡＩＤ５で構成されたストレージ装置３０の記憶領域を模式的に示した図である。なお、同図では４台の不揮発性半導体記憶装置３１（不揮発性半導体記憶装置３１１～３１４）によりストレージ装置３０を構成した例を示しており、当該ストレージ装置３０の記憶領域には１２個のデータＡ～Ｌが記憶されている。

【００１８】

ＲＡＩＤ５を構成する不揮発性半導体記憶装置３１の記憶領域は、コントローラ２０により、データの書き込みまたは読み込みの単位となる複数の論理ブロックに分割される。図２に示した例では、データＡ～Ｌの夫々やパリティＰ１～Ｐ４の夫々が格納された領域が、１つの論理ブロックを示している。

【００１９】

ここで、パリティＰ１～Ｐ４は、夫々同一のストライプグループ（０～３）に記憶された複数のデータから算出される復元情報であって、この復元情報に基づいて当該復元情報の生成元となったデータを復元することが可能となっている。例えば、パリティＰ１はストライプグループ０に記憶されたデータＡ、Ｂ、Ｃから生成されており、データＡ、Ｂ、Ｃのうち何れか一のデータにエラーが発生した場合であっても、残りのデータとパリティＰ１とからエラーの発生したデータを復元することが可能である。なお、データが格納される論理ブロック及びパリティが格納される論理ブロック（以下、パリティ領域という）は、所定のルールに基づき定められているものとするが、その配置位置は図２の例に限定されないものとする。

【００２０】

ストレージ装置３０は、ＮＡＮＤ型フラッシュメモリ等の不揮発性半導体素子を記録媒体とする複数の不揮発性半導体記憶装置３１を有し、コントローラ２０によるＲＡＩＤ管理の下、データを記憶するストレージとして機能する。なお、ストレージ装置３０を構成する不揮発性半導体記憶装置３１の個数は、コントローラ２０が使用するＲＡＩＤの規約に応じた個数（例えば、ＲＡＩＤ１であれば２個以上、ＲＡＩＤ５ならば３個以上）であれば特に問わないものとする。

【００２１】

＜コントローラ２０の構成＞

次に、図３を参照して、コントローラ２０の構成について詳細に説明する。図３は、コントローラ２０の詳細構成を示したブロック図である。同図に示したように、コントローラ２０は、ホスト側Ｉ／Ｆ部２１と、コマンド処理部２２と、ストレージ側Ｉ／Ｆ部２３とを備えている。

【００２２】

ホスト側Ｉ／Ｆ部２１は、ホスト装置１０と接続するためのインタフェース装置であって、ホスト装置１０とコントローラ２０（コマンド処理部２２）との間で行われるデータの授受を制御する。

【００２３】
コマンド処理部２２は、復元情報生成部２２１、復元処理部２２２、キャッシュ管理部２２３を有し、ホスト側Ｉ／Ｆ部２１を介して入力されるホスト装置１０からの要求に応じて、ストレージ装置３０に対しデータの書き込みや読み込みをストレージ側Ｉ／Ｆ部２３を介して行う。
【００２４】
なお、コマンド処理部２２は、ＡＳＩＣやＣＰＵ等の処理装置、コントローラ２０の動作を制御する所定のプログラムが格納されたＲＯＭや当該処理装置のワーク領域となるＲＡＭ等の記憶装置を備え（何れも図示せず）、これら処理装置と記憶装置に格納されたプログラムとの協働により、復元情報生成部２２１、復元処理部２２２及びキャッシュ管理部２２３の各機能部を実現する。
【００２５】
ここで、復元情報生成部２２１は、書き込み対象となるデータのパリティを生成する機能部である。なお、復元情報生成部２２１は、書き込み先となる領域に既存のデータが存在する場合、当該既存のデータと、このデータに係るパリティと、書き込み対象のデータとから、新たなパリティを生成する。
【００２６】
復元処理部２２２は、読み込みエラーが発生したデータについて、当該データと同一のストライプに記憶された他のデータ及びパリティを用いて、読み込みエラーが発生したデータの復元を行う機能部である。
【００２７】
また、キャッシュ管理部２２３は、不揮発性半導体記憶装置３１へ書き込むデータ及び不揮発性半導体記憶装置３１から読み込むデータを一時的に保存して管理するとともに、読み込みエラー発生時に復元処理部２２２により復元されたデータを一時的に保存する機能部である。
【００２８】
コマンド処理部２２は、上記した各機能部（復元情報生成部２２１、復元処理部２２２、キャッシュ管理部２２３）との協働により、ストレージ装置３０に対するデータの書き込み又は読み込みを制御する。
【００２９】
具体的に、コマンド処理部２２は、ホスト装置１０からデータの書き込み要求を受け付けると、このデータの書き込み先が、どの不揮発性半導体記憶装置３１のどの領域（論理ブロック）に対応するのかを特定する。また、復元情報生成部２２１は、書き込み対象のデータに基づいてパリティを生成する。次いで、コマンド処理部２２は、特定した領域へのデータの書き込みと、当該領域に応じたパリティ用の領域へのパリティの書き込みとを、書き込み先となる不揮発性半導体記憶装置３１に要求することで、書き込み対象のデータと当該データについてのパリティとをストレージ装置３０に書き込む。
【００３０】
また、書き込み先となる領域に既存のデータが存在する場合には、既存のデータを新たなデータに更新することになる。この場合、コマンド処理部２２は、書き込み先として特定した領域に格納されている既存のデータと、当該データに係るパリティとの読み込みを、該当する不揮発性半導体記憶装置３１に要求することで、既存のデータ及びパリティを読み込む。このとき、復元情報生成部２２１は、読み込まれた既存のデータ及びパリティと、書き込み対象のデータとから新たなパリティを生成し、この生成された新たなパリティと書き込み対象のデータとを、キャッシュ管理部に保存し、不揮発性半導体記憶装置３１の該当する領域に書き込むことで、データの更新を行う。
【００３１】
また、コマンド処理部２２は、ホスト装置１０からデータの読み込み要求を受け付けると、この読み込み先が、どの不揮発性半導体記憶装置３１のどの領域（論理ブロック）に対応するのかを特定する。そして、コマンド処理部２２は、特定した領域からのデータの

読み込みを、読み込み先となる不揮発性半導体記憶装置３１に要求することで、読み込み対象のデータをストレージ装置３０から読み込み、ホスト装置１０に出力する。

【００３２】

なお、不揮発性半導体記憶装置３１の記録媒体がＮＡＮＤ型フラッシュメモリの場合、データの読み込みの際に発生する電荷変位や自然放電等の理由により、記憶セル上のデータが破損する可能性がある。一般に、ＮＡＮＤ型フラッシュメモリを記録媒体とする記憶装置には、破損したデータの誤りを検出し訂正する機構が設けられているが、必ずしも全ての誤りを訂正できるとは限らず、この場合、データを読み込み時にエラーが発生することになる。そのため、本実施形態では、このようなデータの破損について、復元処理部２２２が読み込みエラーの発生したデータを当該データのパリティに基づいて復元し、キャッシュ管理部２２３へ一時的に保存した後、該当部分への書き込み要求があった場合に、キャッシュ上のデータを元にして書き込むデータを生成し該当する領域に書き込むことで不揮発性半導体記憶装置３１の復旧を行う。これにより、読み込みエラーの発生した不揮発性半導体記憶装置３１を、正常な状態に復旧させることができる。

【００３３】

ストレージ側Ｉ／Ｆ部２３は、不揮発性半導体記憶装置３１と接続するためのインタフェース装置であって、コントローラ２０（コマンド処理部２２）と不揮発性半導体記憶装置３１の間で行われるデータの授受を制御する。なお、ストレージ側Ｉ／Ｆ部２３は、不揮発性半導体記憶装置３１毎に設けられているものとするが、これに限らず、一のストレージ側Ｉ／Ｆ部２３と複数の不揮発性半導体記憶装置３１とが接続される態様としてもよい。

【００３４】

＜コントローラ２０の動作＞

次に、コントローラ２０の動作について説明する。まず、図４を参照して、ストレージ装置３０にデータを書き込む際の動作について説明する。図４は、コントローラ２０により実行される書き込み処理の手順を示したフローチャートである。なお、本処理の前提として、ストレージ装置３０はＲＡＩＤ５で構成されているものとし、データの書き込み及び読み込みはストレージ装置３０のストライプ単位で行われるものとする。

【００３５】

まず、コマンド処理部２２は、ホスト側Ｉ／Ｆ部２１を介しホスト装置１０からデータの書き込み要求を受け付けると（ステップＳ１１）、書き込み先となる領域がどの不揮発性半導体記憶装置３１のどの領域に対応するのかを特定する（ステップＳ１２）。なお、書き込み先となる領域は一であってもよいし、複数であってもよい。

【００３６】

続いて、復元情報生成部２２１は、ステップＳ１２で特定した領域に既存のデータが記憶されているか否かを判定する（ステップＳ１３）。ここで、既存のデータが存在すると判定した場合（ステップＳ１３；Ｙｅｓ）、復元情報生成部２２１は、ステップＳ１２で特定された領域について、データとパリティとの読み込みをストレージ側Ｉ／Ｆ部２３を介してストレージ装置３０に要求することで、既存のデータと当該データに係るパリティとをストレージ装置３０から読み込む（ステップＳ１４）。

【００３７】

次いで、復元情報生成部２２１は、ステップＳ１４で読み込んだ既存のデータ及びパリティと、書き込み対象のデータとから新たなパリティを生成し（ステップＳ１５）、ステップＳ１６の処理に移行する。

【００３８】

なお、ステップＳ１４の読み込みの際、読み込みエラーが発生した場合には、読み込みエラーの発生した既存のデータと同一のストライプに記憶された他のデータをストレージ装置３０から読み込み、当該他のデータと書き込み対象のデータとから、新たなパリティの生成を行うものとする。

【００３９】

一方、ステップＳ１３において、既存のデータが存在しないと判定した場合（ステップＳ１３；Ｎｏ）、復元情報生成部２２１は、書き込み対象のデータからパリティを生成し（ステップＳ１５）、ステップＳ１６の処理に移行する。

【００４０】

続くステップＳ１６では、コマンド処理部２２が、書き込み対象のデータをステップＳ１２で特定した領域に書き込むとともに、当該データの書き込み領域に応じたパリティ領域にステップＳ１５で生成したパリティを書き込む（ステップＳ１６）。ここで、コマンド処理部２２は、データ及び／又はパリティの書き込み時に、書き込みエラーが発生したか否かを判定し、正常に書き込みが行われたと判定した場合には（ステップＳ１７；Ｎｏ）、ステップＳ２２の処理に直ちに移行する。

【００４１】

また、ステップＳ１７において、書き込みエラーを検出した場合（ステップＳ１７；Ｙｅｓ）、コマンド処理部２２は、書き込み先となる不揮発性半導体記憶装置３１に障害が発生したと判断し、当該不揮発性半導体記憶装置３１を除いた残りの不揮発性半導体記憶装置３１でＲＡＩＤ５の構成を維持する縮退動作が可能か否かを判定する（ステップＳ１８）。ここで、縮退動作が不可能と判定した場合（ステップＳ１８；Ｎｏ）、コマンド処理部２２は、書き込みが失敗したことを示す応答をホスト装置１０に出力し（ステップＳ１９）、本処理を終了する。

【００４２】

また、ステップＳ１８において、縮退動作が可能と判定した場合、コマンド処理部２２は、障害が発生した不揮発性半導体記憶装置３１をＲＡＩＤ５の構成から除外し（ステップＳ２０）、縮退動作としたストレージ装置３０に書き込み対象のデータと、ステップＳ１５で生成したパリティとを書き込むと（ステップＳ２１）、ステップＳ２２の処理に移行する。

【００４３】

続くステップＳ２２において、コマンド処理部２２は、ステップＳ１２で特定した全ての領域にデータを書き込んだか否かを判定し、未処理の領域が存在すると判定した場合には（ステップＳ２２；Ｎｏ）、ステップＳ１３の処理に再び戻り、他のストライプに含まれた領域を処理対象とする。また、ステップＳ２２で特定した全ての領域にデータを書き込んだと判定した場合（ステップＳ２２；Ｙｅｓ）、コマンド処理部２２は、書き込みが終了したことを示す応答をホスト装置１０に出力し（ステップＳ２３）、本処理を終了する。

【００４４】

次に、図５を参照して、ストレージ装置３０からデータを読み込む際の動作について説明する。図５は、コントローラ２０により実行される読み込み処理の手順を示したフローチャートである。なお、本処理の前提として、ストレージ装置３０はＲＡＩＤ５で構成されているものとし、データの書き込み及び読み込みはストレージ装置３０のストライプ単位で行われるものとする。

【００４５】

まず、コマンド処理部２２は、ホスト側Ｉ／Ｆ部２１を介しホスト装置１０からデータの読み込み要求を受け付けると（ステップＳ３１）、読み込み先となる領域がどの不揮発性半導体記憶装置３１のどの領域に対応するのかを特定する（ステップＳ３２）。なお、読み込み先となる領域は一であってもよいし、複数であってもよい。

【００４６】

続いて、コマンド処理部２２は、ステップＳ１２で特定した領域に対応する各不揮発性半導体記憶装置３１に対し、当該領域の読み込みを要求することで、ストレージ装置３０から読み込み対象のデータを読み込む（ステップＳ３３）。このとき、コマンド処理部２２は、ステップＳ３３の読み込みの際に読み込みエラーが発生したか否かを判定する（ステップＳ３４）。ここで、コマンド処理部２２が、正常に読み込みできたと判定した場合（ステップＳ３４；Ｎｏ）、キャッシュ管理部２２３へ読み込んだデータを保存し（ステ

ップＳ３８）、ステップＳ３９の処理に移行する。
【００４７】
一方、ステップＳ３４において、コマンド処理部２２が、読み込みエラーが発生したと判定した場合（ステップＳ３４；Ｙｅｓ）、復元処理部２２２は、読み込みエラーが発生したデータと同一のストライプグループに記憶された未読み込みのデータと、パリティとをストレージ装置３０から読み込み、当該データ及びパリティと、ステップＳ３３で読み込んだデータとから読み込みエラーが発生したデータを復元する（ステップＳ３５）。
【００４８】
続いて、復元処理部２２２は、ステップＳ３５の処理でデータを復元できたか否かを判定する。ここで、パリティが読み込めない等の理由によりデータを復元することができないと判定した場合（ステップＳ３６；Ｎｏ）、復元処理部２２２は、読み込みが失敗したことを示す応答をホスト装置１０に出力し（ステップＳ３７）、本処理を終了する。また、ステップＳ３６において、復元処理部２２２が、データを復元できたと判定した場合（ステップＳ３６；Ｙｅｓ）、この復元したデータをキャッシュ管理部２２３に保存し（ステップＳ３８）、ステップＳ３９の処理に移行する。
【００４９】
続くステップＳ３９において、コマンド処理部２２は、ステップＳ３２で特定した全ての領域からデータを読み込んだか否かを判定する（ステップＳ３９）。ここで、未処理の領域が存在すると判定した場合（ステップＳ３９；Ｎｏ）、ステップＳ３３の処理に再び戻り、他の領域を処理対象とする。また、ステップＳ３２で特定した全ての領域からデータを読み込んだと判定した場合（ステップＳ３９；Ｙｅｓ）、コマンド処理部２２は、キャッシュ管理部２２３を参照し各領域から読み込んだデータをホスト装置１０に出力し（ステップＳ４０）、読み込み終了を示す応答をホスト装置１０に出力し（ステップＳ４１）、さらに以下に記載するようにデータの書き込み時にキャッシュ管理部２２３のデータを参照して不揮発性半導体記憶装置３１へ書き込み（ステップＳ４２）、本処理を終了する。
【００５０】
上記ステップＳ３４で読み込みエラーが発生した場合、復元したデータはキャッシュ管理部２２３にのみ保存されることになる。このため、不揮発性半導体記憶装置３１とキャッシュ管理部２２３とのデータの一貫性が保たれていない状態となるが、例えば、当該データの書き込み時にキャッシュ管理部２２３のデータを参照して不揮発性半導体記憶装置３１へ書き込むことで一貫性が保持できる。これにより、不揮発性半導体記憶装置３１に書き込む場合には書き込み時間がかかり、読み込み処理が完了せず、データの要求元（ホスト装置）への出力が遅くなることを防止することができる。
【００５１】
以上のように、第１の実施形態によれば、読み込み時にエラーが発生した場合、この読み込みエラーが発生したデータを復元してキャッシュ管理部２２３へ保存し、以降、書き込み要求があったときに当該読み込みエラーが発生した不揮発性半導体記憶装置３１上の領域に書き込むことで、当該不揮発性半導体記憶装置３１を正常な状態に復旧させることができるため、ＲＡＩＤ構成とされた複数の不揮発性半導体記憶装置３１を有効に活用することができる。
【００５２】
［第２の実施形態］
第１の実施形態では、ストレージ装置３０をＲＡＩＤ５の構成としたが、データの復元を可能とするストレージシステムであれば特に問わず、上述したようにＲＡＩＤ１やＲＡＩＤ６の構成を採用してもよい。以下、第２の実施形態として、ストレージ装置３０をＲＡＩＤ１の構成とした場合について説明する。なお、第１の実施形態と同様の構成要素については、同一の符号を付与し説明を省略する。
【００５３】
＜コントローラ４０の構成＞

まず、第２の実施形態に係るコントローラ４０について説明する。コントローラ４０は、ストレージ装置３０を構成する二つの不揮発性半導体記憶装置３１を、ＲＡＩＤ１の技術を用いて管理し、これら複数の不揮発性半導体記憶装置３１により論理的に構成された記憶領域に対し、データの書き込みや読み出しをホスト装置１０からのアクセス要求に応じて行う。

【００５４】

図６は、ＲＡＩＤ１で構成されたストレージ装置３０の記憶領域を模式的に示した図である。ＲＡＩＤ１は、ミラーリングとも呼ばれ、少なくとも２台以上の不揮発性半導体記憶装置３１に同一のデータを、同一のストライプ領域に同時に書き込みすることで、耐障害性・冗長性を確保している。なお、同図では２台の不揮発性半導体記憶装置３１（不揮発性半導体記憶装置３１１、３１２）によりストレージ装置３０を構成した例を示しており、当該ストレージ装置３０の記憶領域に１２個のデータＡ～Ｌが記憶されている。

【００５５】

このＲＡＩＤ１の構成の場合、一方の不揮発性半導体記憶装置３１についてデータの読み込みエラーが発生した場合、他方の不揮発性半導体記憶装置３１から同一のデータを読み込むことで、システム自体は問題無く稼動し続けることができる。

【００５６】

図７は、第２の実施形態に係るコントローラ４０の詳細構成を示したブロック図である。同図に示したように、コントローラ４０は、ホスト側Ｉ／Ｆ部２１と、コマンド処理部４１と、ストレージ側Ｉ／Ｆ部２３とを備えている。

【００５７】

ここで、コマンド処理部４１は、復元処理部４１１、キャッシュ管理部４１２を有し、ホスト側Ｉ／Ｆ部２１を介して入力されるホスト装置１０のアクセス要求に応じて、ストレージ側Ｉ／Ｆ部２３を介して接続されるストレージ装置３０に対し、データの書き込みや読み込みを行う。

【００５８】

なお、コマンド処理部４１は、ＡＳＩＣやＣＰＵ等の処理装置、コントローラ２０の動作を制御する所定のプログラムが格納されたＲＯＭや当該処理装置のワーク領域となるＲＡＭ等の記憶装置を備え（何れも図示せず）、これら処理装置と記憶装置に格納されたプログラムとの協働により、復元処理部４１１及びキャッシュ管理部４１２の各機能部を実現する。

【００５９】

具体的に、コマンド処理部４１は、ホスト装置１０からデータの書き込み要求を受け付けると、このデータの書き込み先となる領域（論理ブロック）を各不揮発性半導体記憶装置３１から夫々特定する。そして、コマンド処理部２２は、書き込み対象のデータを不揮発性半導体記憶装置３１の個数に応じた数だけ複製すると、各不揮発性半導体記憶装置３１の特定した領域に夫々書き込む。

【００６０】

また、コマンド処理部２２は、ホスト装置１０からデータの読み込み要求を受け付けると、アクセス対象となる一の不揮発性半導体記憶装置３１から、この読み込み先に対応する領域（論理ブロック）を特定する。そして、コマンド処理部２２は、アクセス対象となる不揮発性半導体記憶装置３１の特定した領域からデータを読み込み、ホスト装置１０に出力する。ここで、アクセス対象となる不揮発性半導体記憶装置３１は、予め定められているものとしてもよいし、負荷分散などの理由により他の不揮発性半導体記憶装置３１と動的に切り替わるものとしてもよい。以下、アクセス対象の不揮発性半導体記憶装置３１を「主記憶装置」と呼び、他の不揮発性半導体記憶装置３１を「待機記憶装置」と呼ぶ。

【００６１】

なお、不揮発性半導体記憶装置３１の記録媒体がＮＡＮＤ型フラッシュメモリの場合、上記したように読み込みの際に発生する電荷変位や自然放電等の理由により、記憶セル上のデータが破損する可能性がある。そのため、コマンド処理部４１では、主記憶装置に発

生したデータの破損について、復元処理部４１１が当該データと同一のデータを待機記憶装置から読み込み、キャッシュ管理部４１２へ一時的に保存する。そして、コマンド処理部４１は、読み込みエラーの発生した部分への書き込み要求があった場合に、キャッシュ上のデータを元にして書き込むデータを生成し、該当する領域に書き込むことでデータの復元を行う。つまり、待機記憶装置に記憶されるデータは、主記憶装置に記憶されたデータを復元するための復元情報として機能する。これにより、読み込みエラーの発生した主記憶装置を、正常な状態に復旧することができる。

【００６２】

＜コントローラ４０の動作＞

次に、コントローラ４０の動作について説明する。まず、図８を参照して、ストレージ装置３０にデータを書き込む際の動作について説明する。

【００６３】

図８は、コントローラ４０により実行される書き込み処理の手順を示したフローチャートである。なお、本処理の前提として、ストレージ装置３０はＲＡＩＤ１で構成されているものとし、データの書き込みは論理ブロック単位で行われるものとする。

【００６４】

まず、コマンド処理部４１は、ホスト側Ｉ／Ｆ部２１を介しホスト装置１０からデータの書き込み要求を受け付けると（ステップＳ５１）、ストレージ装置３０を構成する各不揮発性半導体記憶装置３１について書き込み先となる領域を特定する（ステップＳ５２）。

【００６５】

続いて、コマンド処理部４１は、書き込み対象のデータを不揮発性半導体記憶装置３１の個数分複製すると、ステップＳ５２で特定した各不揮発性半導体記憶装置３１の領域に、書き込み対象のデータを書き込む（ステップＳ５３）。ここで、コマンド処理部２２は、データの書き込みの際に書き込みエラーが発生したか否かを判定し、正常に書き込みが行われたと判定した場合には（ステップＳ５４；Ｎｏ）、ステップＳ５８の処理に直ちに移行する。

【００６６】

また、ステップＳ５４において、書き込みエラーを検出した場合（ステップＳ５４；Ｙｅｓ）、コマンド処理部４１は、書き込み先となった不揮発性半導体記憶装置３１に障害が発生したと判定し、当該不揮発性半導体記憶装置３１を除いた残りの不揮発性半導体記憶装置３１でシステムを維持する縮退動作が可能か否かを判定する（ステップＳ５５）。ここで、縮退動作が不可能と判定した場合（ステップＳ５５；Ｎｏ）、コマンド処理部２２は、書き込みが失敗したことを示す応答をホスト装置１０に出力し（ステップＳ５６）、本処理を終了する。

【００６７】

また、ステップＳ５５において、縮退動作が可能と判定した場合、コマンド処理部４１は、障害が発生した不揮発性半導体記憶装置３１をＲＡＩＤ１の構成から除外し（ステップＳ５７）、ステップＳ５８の処理に移行する。

【００６８】

続くステップＳ５８において、コマンド処理部４１は、ステップＳ５２で特定した全ての領域にデータを書き込んだか否かを判定する（ステップＳ５８）。ここで、未処理の領域が存在すると判定した場合には（ステップＳ５８；Ｎｏ）、ステップＳ５３の処理に再び戻り、他の領域を処理対象とする。また、ステップＳ５２で特定した全ての領域にデータを書き込んだと判定した場合（ステップＳ５８；Ｙｅｓ）、コマンド処理部４１は、書き込みが終了したことを示す応答をホスト装置１０に出力し（ステップＳ５９）、本処理を終了する。

【００６９】

次に、図９を参照して、ストレージ装置３０からデータを読み込む際の動作について説明する。図９は、コントローラ４０により実行される読み込み処理の手順を示したフロー

チャートである。なお、本処理の前提として、ストレージ装置３０はＲＡＩＤ１で構成されているものとし、データの書き込みは論理ブロック単位で行われるものとする。

【００７０】

まず、コマンド処理部４１は、ホスト側Ｉ／Ｆ部２１を介しホスト装置１０からデータの読み込み要求を受け付けると（ステップＳ６１）、読み込み先となる領域が主記憶装置のどの領域に対応するのかを特定する（ステップＳ６２）。

【００７１】

続いて、コマンド処理部４１は、主記憶装置に対し、ステップＳ６２で特定した領域のデータの読み込みを要求することで、読み込み対象となったデータをストレージ装置３０から読み込む（ステップＳ６３）。このとき、コマンド処理部４１は、ステップＳ６３の読み込みの際に読み込みエラーが発生したか否かを判定する（ステップＳ６４）。ここで、コマンド処理部４１が、正常に読み込みできたと判定した場合（ステップＳ６４；Ｎｏ）、キャッシュ管理部４１２へ読み込んだデータを保存し（ステップＳ６８）、ステップＳ６９の処理に移行する。

【００７２】

一方、ステップＳ６４において、コマンド処理部４１が、読み込みエラーが発生したと判定した場合（ステップＳ６４；Ｙｅｓ）、復元処理部４１１は、読み込みエラーの発生したデータと同一のデータを待機記憶装置から読み込む（ステップＳ６５）。続いて、復元処理部４１１は、待機記憶装置からデータを読み込めたか否かを判定する（ステップＳ６６）。ここで、何れの待機記憶装置からもデータを読み込めないと判定した場合（ステップＳ６６；Ｎｏ）、復元処理部４１１は、読み込みが失敗したことを示す応答をホスト装置１０に出力し（ステップＳ６７）、本処理を終了する。

【００７３】

また、ステップＳ６６において、復元処理部４１１は、待機記憶装置からデータを読み込めたと判定した場合（ステップＳ６６；Ｙｅｓ）、ステップＳ６５で読み込まれたデータをキャッシュ管理部４１２へ保存し（ステップＳ６８）、ステップＳ６９の処理に移行する。

【００７４】

続くステップＳ６９において、コマンド処理部４１は、ステップＳ６２で特定した全ての領域からデータを読み込んだか否かを判定する（ステップＳ６９）。ここで、未処理の領域が存在すると判定した場合には（ステップＳ６９；Ｎｏ）、ステップＳ６３の処理に再び戻り、次の領域を処理対象とする。また、ステップＳ６２で特定した全ての領域からデータを読み込んだと判定した場合（ステップＳ６９；Ｙｅｓ）、コマンド処理部２２は、各領域から読み込んだデータをホスト装置１０に出力すると（ステップＳ７０）、読み込み終了を示す応答をホスト装置１０に出力し（ステップＳ７１）、さらに以下に記載するようにデータの書き込み時にキャッシュ管理部２２３のデータを参照して不揮発性半導体記憶装置３１へ書き込み（ステップＳ７２）、本処理を終了する。

【００７５】

なお、上記ステップＳ６４で読み込みエラーが発生した場合、復元したデータはキャッシュ管理部４１２にのみ保存されることになる。このため、不揮発性半導体記憶装置３１とキャッシュ管理部４１２とのデータの一貫性が保たれていない状態となるが、例えば、当該データの書き込み時にキャッシュ管理部４１２のデータを参照して不揮発性半導体記憶装置３１へ書き込むことで一貫性が保持できる。これにより、不揮発性半導体記憶装置３１に書き込む場合には書き込み時間がかかり、読み込み処理が完了せず、データの要求元（ホスト装置）への出力が遅くなることを防止することができる。

【００７６】

以上のように、第２の実施形態によれば、読み込み時にエラーが発生した場合、この読み込みエラーが発生したデータを復元してキャッシュ管理部４１２へ保存し、以降、書き込み要求があったときに当該読み込みエラーが発生した不揮発性半導体記憶装置３１上の領域に書き込むことで、当該不揮発性半導体記憶装置３１を正常な状態に復旧させること

ができるため、ＲＡＩＤ構成とされた複数の不揮発性半導体記憶装置３１を有効に活用することができる。

【００７７】

以上、発明の実施の形態について説明したが、本発明はこれに限定されるものではなく、本発明の主旨を逸脱しない範囲での種々の変更、置換、追加などが可能である。

【図面の簡単な説明】

【００７８】

【図１】ストレージシステムの構成を示した図である。

【図２】ストレージ装置をＲＡＩＤ５とした場合の記憶領域を模式的に示した図である。

【図３】第１の実施形態に係るコントローラの構成を示した図である。

【図４】第１の実施形態に係る書き込み処理の手順を示したフローチャートである。

【図５】第１の実施形態にかかる読み込み処理の手順を示したフローチャートである。

【図６】第２の実施形態に係るコントローラの構成を示した図である。

【図７】ストレージ装置をＲＡＩＤ１とした場合の記憶領域を模式的に示した図である。

【図８】第２の実施形態に係る書き込み処理の手順を示したフローチャートである。

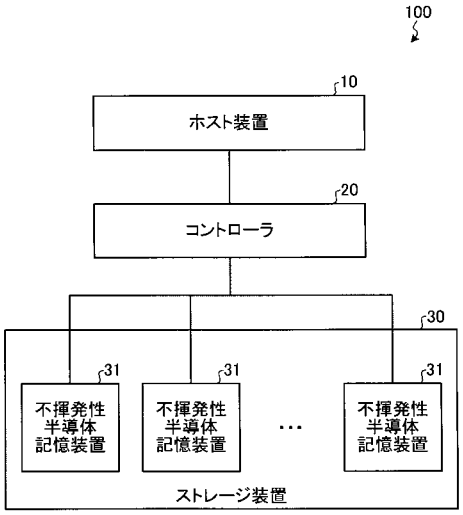
【図９】第２の実施形態にかかる読み込み処理の手順を示したフローチャートである。

【符号の説明】

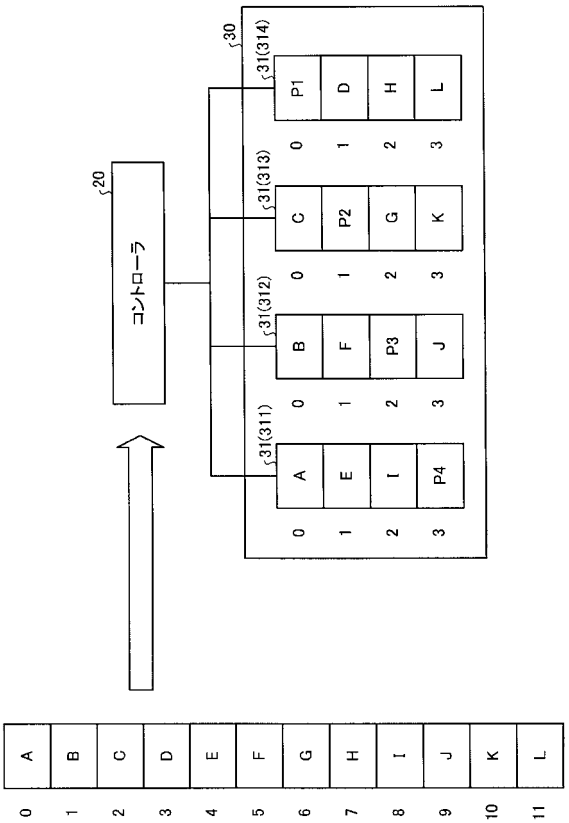
【００７９】

１００　ストレージシステム
１０　ホスト装置
２０　コントローラ
２１　ホスト側Ｉ／Ｆ部
２２　コマンド処理部
２２１　復元情報生成部
２２２　復元処理部
２２３　キャッシュ管理部
２３　ストレージ側Ｉ／Ｆ部
３０　ストレージ装置
３１　不揮発性半導体記憶装置
４０　コントローラ
４１　コマンド処理部
４１１　復元処理部
４１２　キャッシュ管理部

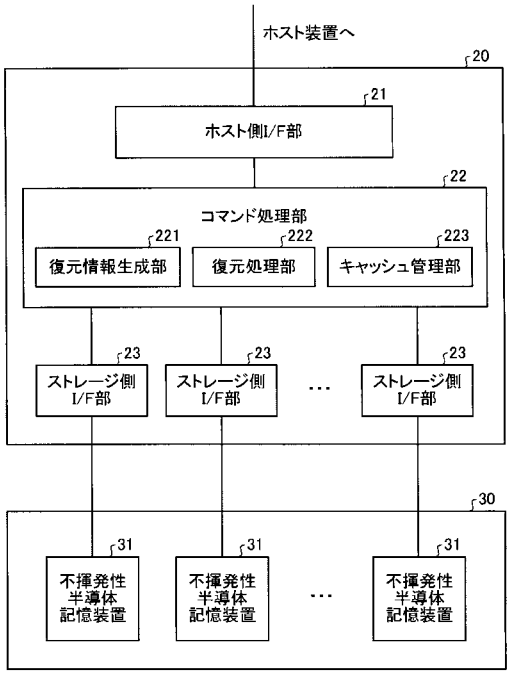
【図１】


100
10
ホスト装置
20
コントローラ
30
31
不揮発性半導体記憶装置
31
不揮発性半導体記憶装置
31
不揮発性半導体記憶装置
ストレージ装置

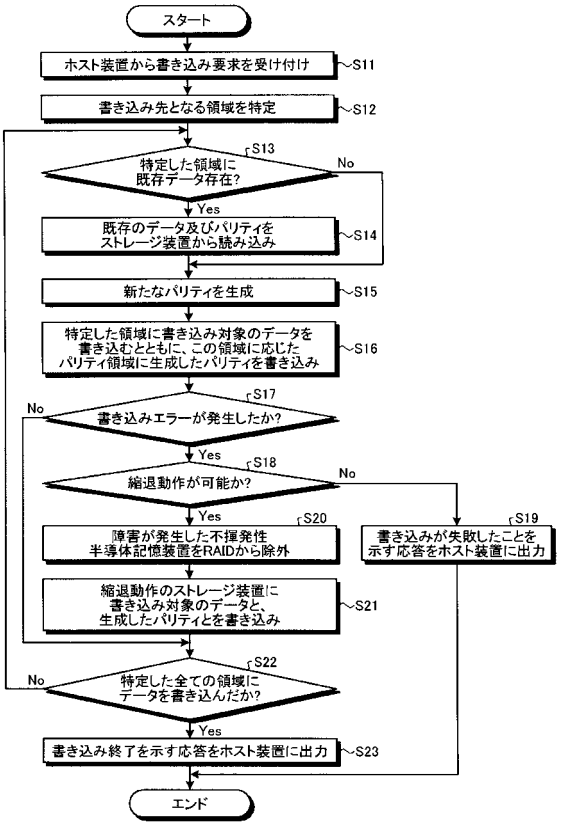
【図２】


30
31(314)
P1 D H L
31(313)
C P2 G K
31(312)
B F P3 J
31(311)
A E I P4
0 1 2 3
20
コントローラ
A B C D E F G H I J K L
0 1 2 3 4 5 6 7 8 9 10 11

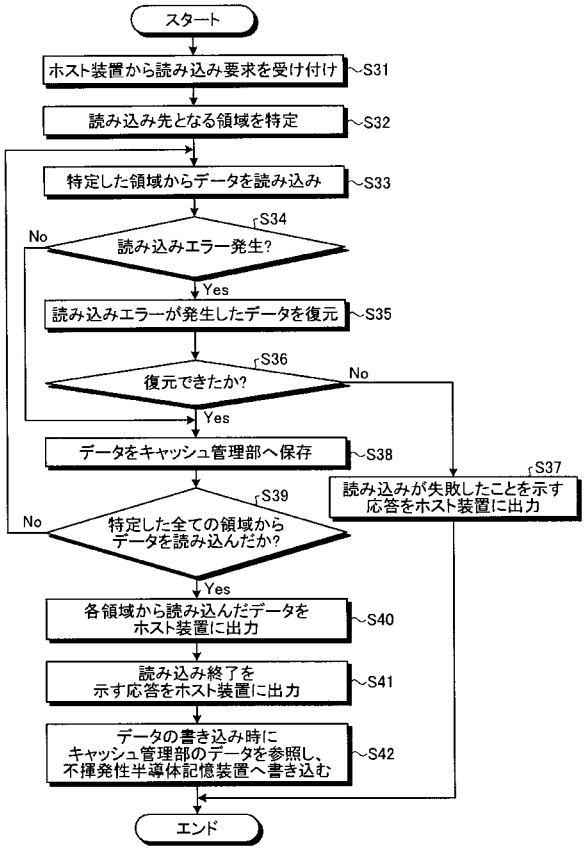
【図３】


ホスト装置へ
20
21
ホスト側I/F部
22
コマンド処理部
221
復元情報生成部
222
復元処理部
223
キャッシュ管理部
23
ストレージ側I/F部
23
ストレージ側I/F部
23
ストレージ側I/F部
30
31
不揮発性半導体記憶装置
31
不揮発性半導体記憶装置
31
不揮発性半導体記憶装置

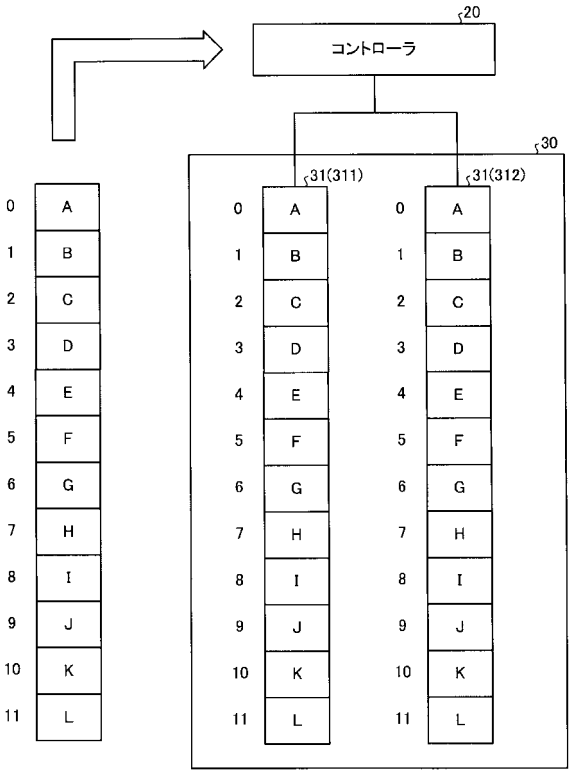
【図４】


スタート
ホスト装置から書き込み要求を受け付け
S11
書き込み先となる領域を特定
S12
S13
特定した領域に既存データ存在?
No
Yes
既存のデータ及びパリティをストレージ装置から読み込み
S14
新たなパリティを生成
S15
特定した領域に書き込み対象のデータを書き込むとともに、この領域に応じたパリティ領域に生成したパリティを書き込み
S16
S17
書き込みエラーが発生したか?
No
Yes
S18
縮退動作が可能か?
No
Yes
S20
障害が発生した不揮発性半導体記憶装置をRAIDから除外
S19
書き込みが失敗したことを示す応答をホスト装置に出力
縮退動作のストレージ装置に書き込み対象のデータと、生成したパリティとを書き込み
S21
S22
特定した全ての領域にデータを書き込んだか?
No
Yes
書き込み終了を示す応答をホスト装置に出力
S23
エンド

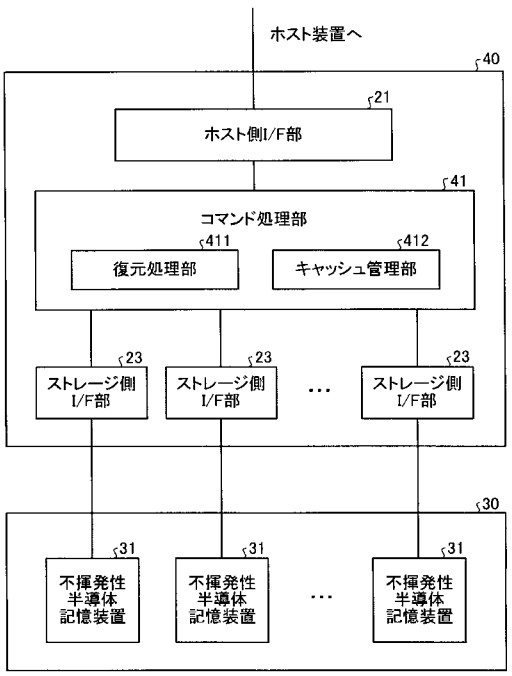
【図５】


スタート
ホスト装置から読み込み要求を受け付け
S31
読み込み先となる領域を特定
S32
特定した領域からデータを読み込み
S33
S34
読み込みエラー発生?
No
Yes
読み込みエラーが発生したデータを復元
S35
S36
復元できたか?
No
Yes
データをキャッシュ管理部へ保存
S38
S37
読み込みが失敗したことを示す応答をホスト装置に出力
S39
特定した全ての領域からデータを読み込んだか?
No
Yes
各領域から読み込んだデータをホスト装置に出力
S40
読み込み終了を示す応答をホスト装置に出力
S41
データの書き込み時にキャッシュ管理部のデータを参照し、不揮発性半導体記憶装置へ書き込む
S42
エンド

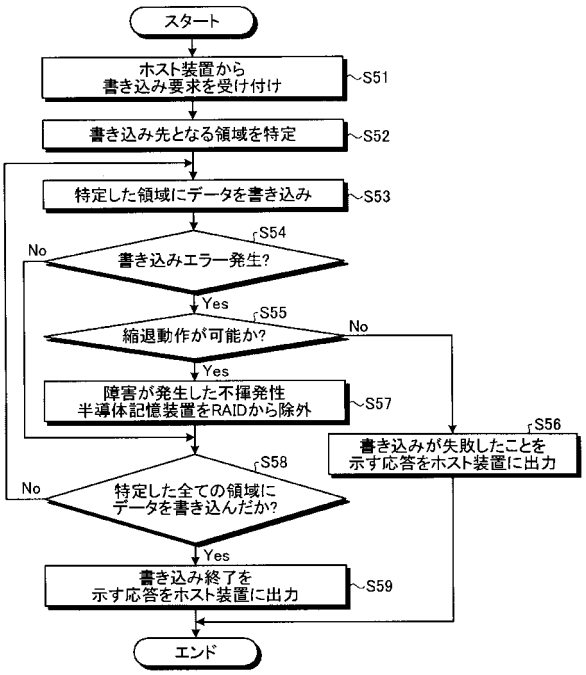
【図６】


20
コントローラ
30
31(311)
31(312)
0 A
1 B
2 C
3 D
4 E
5 F
6 G
7 H
8 I
9 J
10 K
11 L

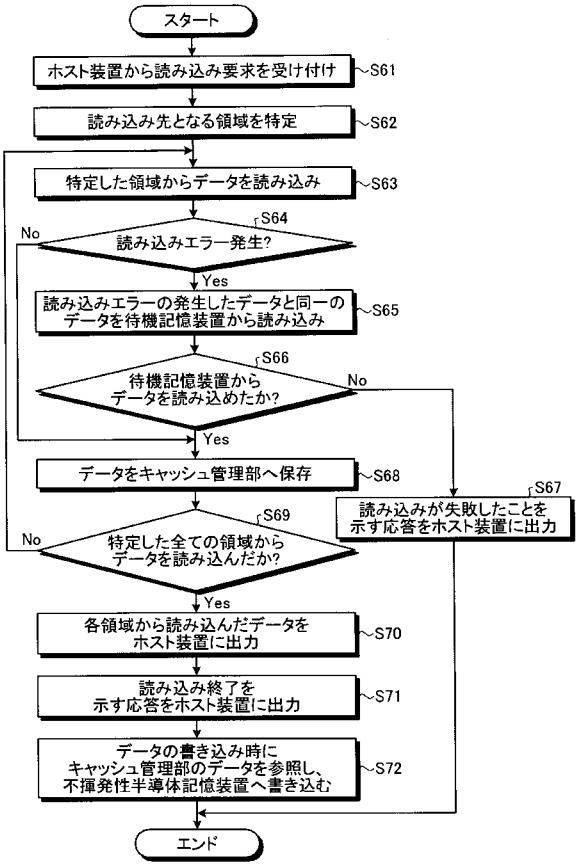
【図７】


ホスト装置へ
40
21
ホスト側I/F部
41
コマンド処理部
411
復元処理部
412
キャッシュ管理部
23
ストレージ側I/F部
...
30
31
不揮発性半導体記憶装置

【図８】


スタート
ホスト装置から書き込み要求を受け付け
S51
書き込み先となる領域を特定
S52
特定した領域にデータを書き込み
S53
S54
書き込みエラー発生?
No
Yes
S55
縮退動作が可能か?
No
Yes
障害が発生した不揮発性半導体記憶装置をRAIDから除外
S57
S56
書き込みが失敗したことを示す応答をホスト装置に出力
S58
特定した全ての領域にデータを書き込んだか?
No
Yes
書き込み終了を示す応答をホスト装置に出力
S59
エンド

【図９】


スタート
ホスト装置から読み込み要求を受け付け
S61
読み込み先となる領域を特定
S62
特定した領域からデータを読み込み
S63
S64
読み込みエラー発生?
No
Yes
読み込みエラーの発生したデータと同一のデータを待機記憶装置から読み込み
S65
S66
待機記憶装置からデータを読み込めたか?
No
Yes
データをキャッシュ管理部へ保存
S68
S67
読み込みが失敗したことを示す応答をホスト装置に出力
S69
特定した全ての領域からデータを読み込んだか?
No
Yes
各領域から読み込んだデータをホスト装置に出力
S70
読み込み終了を示す応答をホスト装置に出力
S71
データの書き込み時にキャッシュ管理部のデータを参照し、不揮発性半導体記憶装置へ書き込む
S72
エンド

フロントページの続き

(72)発明者　浅野　滋博
東京都港区芝浦一丁目１番１号　株式会社東芝内

Ｆターム(参考) 5B018 GA04 GA06 HA35 KA22 MA22 NA06 QA03
5B065 BA05 CA30 EA03 EA24