(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6886486号
(P6886486)
(45)発行日　令和3年6月16日(2021.6.16)　(24)登録日　令和3年5月18日(2021.5.18)

(51)Int.Cl.　F I
G06F 3/06 (2006.01)　G06F 3/06 306Z
G06F 3/06 301Z

請求項の数 6　(全 41 頁)

(21)出願番号　特願2019-72298(P2019-72298)
(22)出願日　平成31年4月4日(2019.4.4)
(62)分割の表示　特願2017-511430(P2017-511430)の分割
原出願日　平成27年4月9日(2015.4.9)
(65)公開番号　特開2019-106224(P2019-106224A)
(43)公開日　令和1年6月27日(2019.6.27)
審査請求日　令和1年5月7日(2019.5.7)

(73)特許権者　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号
(74)代理人　110002365
特許業務法人サンネクスト国際特許事務所
(72)発明者　山本　貴大
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　藤本　和久
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　坏　弘明
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】ストレージシステム及びデータ制御方法

(57)【特許請求の範囲】
【請求項１】
データを記憶する複数の半導体記憶装置と、
仮想ボリュームに対して前記複数の半導体記憶装置に基づくプールから記憶領域を割り当てるプロセッサとを備え、
前記複数の半導体記憶装置には、書き込み上限回数とセル当たりの記憶容量が異なる複数の種類の半導体記憶装置が含まれ、セル当たりの記憶容量が大きい半導体記憶装置は、書き込み上限回数が少なく、記憶容量当たりのコストが小さく、
前記プロセッサは、前記仮想ボリュームの記憶領域ごとに書き込み頻度を記録し、
前記プロセッサは、前記仮想ボリュームの記憶領域の書き込み頻度に基づいて、前記記憶領域に割り当てられる半導体記憶装置の種類を変更し、
前記プロセッサは、前記半導体記憶装置の種類の変更では、書き込み頻度の高い記憶領域を書き込み上限回数が多い種類の半導体記憶装置に割り当て、書き込み頻度の低い記憶領域を書き込み上限回数が少ない種類の半導体記憶装置に割り当て、
前記プロセッサは、前記プール又は前記仮想ボリュームの容量と、前記プール又は前記仮想ボリュームへ想定される書き込み頻度と、前記半導体記憶装置の各種類について単位容量当たりの許容される書き込み頻度とに基づいて、前記記憶領域に割り当てるプールの基となる前記半導体記憶装置にかかる推奨される種類ごとの記憶容量にかかる情報を算出する、
ことを特徴とするストレージシステム。

【請求項２】
請求項１において、
前記記憶領域に割り当てるプールの基となる前記半導体記憶装置にかかる推奨される種類ごとの記憶容量にかかる情報は、前記プール又は前記仮想ボリュームの容量のうちの、当該種類の半導体記憶装置に基づく記憶領域の容量の比率であり当該種類の半導体記憶装置の使用する想定期間としての寿命が保証される推奨の比率である
ことを特徴とするストレージシステム。
【請求項３】
請求項１において、
前記記憶領域に割り当てるプールの基となる前記半導体記憶装置にかかる推奨される種類ごとの記憶容量は、前記半導体記憶装置への書き込み回数とその書き込み上限回数とに基づいて半導体記憶装置を交換する場合に、ストレージシステムの運用管理コストを小さくする前記半導体記憶装置の種類ごとの容量である
ことを特徴とするストレージシステム。
【請求項４】
請求項３において、
前記半導体記憶装置には、使用する想定期間が設定されており、
前記想定期間を使用するための半導体記憶装置が不足する場合に、前記半導体記憶装置の不足量を報知する
ことを特徴とするストレージシステム。
【請求項５】
請求項１において、
所定の閾値よりも前記書き込み頻度の高い記憶領域を、書き込み上限回数が少ない種類の半導体記憶装置から書き込み上限回数が多い種類の半導体記憶装置に割り当てを移動させ、
前記所定の閾値は、書き込み上限回数が少ない種類の半導体記憶装置への書き込み頻度に基づいて変更される
ことを特徴とするストレージシステム。
【請求項６】
データを記憶する複数の半導体記憶装置と、
仮想ボリュームに対して前記複数の半導体記憶装置に基づくプールから記憶領域を割り当てるプロセッサと、
を備えたストレージシステムにおける制御方法において、
前記複数の半導体記憶装置には、書き込み上限回数とセル当たりの記憶容量が異なる複数の種類の半導体記憶装置が含まれ、セル当たりの記憶容量が大きい半導体記憶装置は、書き込み上限回数が少なく、記憶容量当たりのコストが小さく、
前記プロセッサは、前記仮想ボリュームの記憶領域ごとに書き込み頻度を記録し、
前記プロセッサは、前記仮想ボリュームの記憶領域の書き込み頻度に基づいて、前記記憶領域に割り当てられる半導体記憶装置の種類を変更し、
前記プロセッサは、前記半導体記憶装置の種類の変更では、書き込み頻度の高い記憶領域を書き込み上限回数が多い種類の半導体記憶装置に割り当て、書き込み頻度の低い記憶領域を書き込み上限回数が少ない種類の半導体記憶装置に割り当て、
前記プロセッサは、前記プール又は前記仮想ボリュームの容量と、前記プール又は前記仮想ボリュームへ想定される書き込み頻度と、前記半導体記憶装置の各種類について単位容量当たりの許容される書き込み頻度とに基づいて、前記記憶領域に割り当てるプールの基となる前記半導体記憶装置にかかる推奨される種類ごとの記憶容量にかかる情報を算出する、
ことを特徴とするストレージシステム制御方法。
【発明の詳細な説明】
【技術分野】

【０００１】
本発明は、ストレージシステム及びデータ制御方法に関し、特に同一種別であって特性の異なる複数の記憶装置を利用して、記憶装置間でデータを自動的に再配置するストレージシステム及びデータ制御方法に適用して好適なものである。
【背景技術】
【０００２】
近年、ＨＤＤ（Hard Disk Drive）やＳＳＤ（Solid State Drive）の普及により、ストレージシステムに搭載される記憶装置の種類が多様化している。特にＳＳＤについては、データ格納方式としてＳＬＣ（Single Level Cell）、２ｂｉｔ ＭＬＣ（Multi Level Cell）又はＴＬＣ（Triple Level Cell）等があり、同一種別の記憶装置であっても、寿命（書き込み上限回数）及びビットコストが異なるという特性がある。
【０００３】
寿命は、一般にＳＬＣが最も長く、ＭＬＣ、ＴＬＣの順に短くなる。またビットコストは、ＴＬＣが最も低く、ＭＬＣ、ＳＬＣの順に高くなる。よってＳＳＤを搭載するストレージシステムにおいては、その特性を考慮してデータを適切に配置することが重要となる。
【０００４】
特許文献１及び２には、複数の記憶装置が搭載されたストレージシステムを管理又は制御する技術が開示されている。具体的に特許文献１には、複数の記憶装置をプールとして統合的に管理し、仮想的なボリュームをホスト装置に提供する技術が開示されている。
【０００５】
また特許文献２には、ＨＤＤとＳＳＤとを異なる階層の記憶装置として管理し、ホスト装置からのアクセス頻度に応じてデータを自動的に再配置し、高階層でのデータヒット率を向上させることにより、ホスト装置に対する応答性能を向上させる技術が開示されている。
【先行技術文献】
【特許文献】
【０００６】
【特許文献１】米国特許第７４４７８３２号明細書
【特許文献２】米国特許第８０４１６８２号明細書
【発明の概要】
【発明が解決しようとする課題】
【０００７】
しかし特許文献１及び２に記載の技術では、同一種別であって特性（寿命及びビットコスト）の異なる複数の記憶装置（例えばＳＳＤ）が混在したストレージシステムに対して書き込みを行った場合、特性が考慮されることなく、各記憶装置にデータが書き込まれる。
【０００８】
ＳＳＤの特性を考慮せずに書き込みを行うと、書き込み上限回数の少ないＳＳＤの寿命が書き込み上限回数の多いＳＳＤよりも先に尽きることになり、寿命の尽きたＳＳＤに対する交換回数が増え、ストレージシステムの運用管理コストが高くなるという課題がある。
【０００９】
本発明は以上の点を考慮してなされたもので、同一種別であって特性の異なる記憶装置の寿命を延ばし、運用管理コストを削減し得るストレージシステム及びデータ制御方法を提案する。
【課題を解決するための手段】
【００１０】
かかる課題を解決するために、本発明においては、第１の記憶装置と、第１の記憶装置よりも書き込み上限回数が少なく、かつ、単位面積当たりの記憶容量が多い第２の記憶装

置と、ホストに提供する仮想ボリュームに対して第１の記憶装置及び第２の記憶装置から記憶領域を割り当てるプロセッサとを備え、プロセッサは、仮想ボリュームに対して第２の記憶装置から割り当てられている記憶領域のうち、ホストからのライト頻度が予め定められたライト閾値よりも多い記憶領域に格納されているデータを第１の記憶装置の記憶領域に再配置する。

【００１１】

またかかる課題を解決するために、本発明においては、第１の記憶装置と、第１の記憶装置よりも書き込み上限回数が少なく、かつ、単位面積当たりの記憶容量が多い第２の記憶装置と、ホストに提供する仮想ボリュームに対して第１の記憶装置及び第２の記憶装置から記憶領域を割り当てるプロセッサとを備えたストレージシステムのデータ制御方法であって、プロセッサが、仮想ボリュームに対して第２の記憶装置から割り当てられている記憶領域のうち、ホストからのライト頻度が予め定められたライト閾値よりも多い記憶領域に格納されているデータを第１の記憶装置の記憶領域に再配置する第１のステップと、仮想ボリュームに対して第１の記憶装置から割り当てられている記憶領域のうち、ホストからのライト頻度がライト閾値以下である記憶領域に格納されているデータを第２の記憶装置の記憶領域に再配置する第２のステップとを備える。

【発明の効果】

【００１２】

本発明によれば、同一種別であって特性の異なる記憶装置の寿命を延ばし、運用管理コストを削減することができる。

【図面の簡単な説明】

【００１３】

【図１】第１の実施の形態における計算機システムの全体構成図である。
【図２】ストレージシステムの論理構成図である。
【図３】ＴＬＣ－ＭＬＣ間で行われるページ再配置処理の概念構成図である。
【図４】ＴＬＣ－ＴＬＣ間で行われるページ再配置処理の概念構成図である。
【図５】共有メモリの内部構成図である。
【図６】ローカルメモリの内部構成図である。
【図７】ページ毎モニタテーブルの論理構成図である。
【図８】パリティグループ毎モニタテーブルの論理構成図である。
【図９】パリティグループ毎再配置管理テーブルの論理構成図である。
【図１０】プール毎再配置管理テーブルの論理構成図である。
【図１１】ホストＩ／Ｏ処理のフローチャートである。
【図１２】デステージ処理のフローチャートである。
【図１３】寿命情報採取処理のフローチャートである。
【図１４】閾値決定処理のフローチャートである。
【図１５】再配置処理のフローチャートである。
【図１６】ＴＬＣ－ＭＬＣ間再配置処理のフローチャートである。
【図１７】ライトリバランス処理のフローチャートである。
【図１８】性能リバランス処理のフローチャートである。
【図１９】新規割り当て決定処理のフローチャートである。
【図２０】推奨容量算出方法の概念図である。
【図２１】パラメータ設定画面の画面構成の一例である。
【図２２】警告画面の画面構成の一例である。
【図２３】第２の実施の形態における計算機システムの全体構成図である。
【図２４】ストレージシステムの論理構成図である。
【図２５】ライトデモーション処理によるページ再配置処理の概念構成図である。
【図２６】パリティグループ毎再配置管理テーブルの論理構成図である。
【図２７】プール毎再配置管理テーブルの論理構成図である。
【図２８】再配置処理のフローチャートである。

【図２９】ティア間再配置処理のフローチャートである。
【図３０】ライトデモーション処理のフローチャートである。
【図３１】新規割り当て決定処理のフローチャートである。
【図３２】ライト追加可能量及びライト削減要求量の算出方法の概念図である。
【発明を実施するための形態】
【００１４】
以下、図面を参照しながら本発明の一実施の形態を説明する。なお以下に説明する実施の形態は、本発明を実現するための一例であって、本発明の技術的範囲を限定するものではない。また各図において共通の構成については、同一の参照番号を付してその説明を省略する。また各図において複数の同一部材を図示する場合は、「２０１Ａ」、「２０１Ｂ」のようにアルファベットを付して区別する一方で、総称して呼ぶ場合には「２０１」のようにアルファベットを省略する。さらに本発明の実施の形態は、汎用コンピュータ上で稼動するソフトウェアに実装してもよいし、専用ハードウェアに実装してもよい。またソフトウェアとハードウェアとの組み合わせに実装してもよい。以下の説明では、管理用の情報をテーブル形式で説明するが、管理用の情報は必ずしもテーブルによるデータ構造で表現されなくてもよく、「リスト」、「ＤＢ」、「キュー」などについて単に「情報」と呼ぶことがある。また「プログラム」を主語（動作主体）として本発明の実施の形態における各処理について説明する場合がある。プログラムはプロセッサによって実行されることで定められた処理をメモリ及び通信ポート（通信制御装置）を用いながら行うため、プロセッサを主語とした説明としてもよい。プログラムの一部又は全部は専用ハードウェアで実現してもよく、またモジュール化されていてもよい。各種プログラムはプログラム配布サーバや記憶メディアによって各計算機にインストールされてもよい。
【００１５】
（１）第１の実施の形態
第１の実施の形態は、記憶装置として半導体メモリ（ＳＳＤ）のみを搭載するストレージシステムにおいて、特性の異なる半導体メモリ間（例えばＴＬＣ－ＭＬＣ間）でデータを再配置することにより、半導体メモリの長寿命化を実現するとともに、ビットコストを低減しようとするものである。さらには同一種別の半導体メモリ間（例えばＴＬＣ－ＴＬＣ間）でデータを再配置することにより、負荷分散を実現して長寿命化を実現しようとするものである。
【００１６】
（１－１）計算機システムの全体構成
図１は、第１の実施の形態における計算機システム１の全体構成を示す。計算機システム１は、ホスト１０１、管理サーバ１０２、ネットワーク１０３及びストレージシステム１０４から構成される。ホスト１０１は、例えば一般的なサーバであり、ネットワーク１０３を介してストレージシステム１０４のポート１０５に接続される。ホスト１０１は、ストレージシステム１０４に対してリード要求又はライト要求を発行して、データの読み書きを実行する。
【００１７】
ネットワーク１０３は、例えばＳＡＮ（Storage Area Network）やイーサネット（登録商標）等の通信回線である。管理サーバ１０２は、ネットワーク１０３を介して、ストレージシステム１０４の保守Ｉ／Ｆ１０６又はポート１０５に接続される。ストレージ管理者は、管理サーバ１０２を操作して、ストレージシステム１０４を運用する上で必要な各種設定や管理を行う。
【００１８】
次にストレージシステム１０４の内部構成について説明する。ストレージシステム１０４は、ポート１０５、保守Ｉ／Ｆ１０６、プロセッサパッケージ１０７、キャッシュメモリ１０８、共有メモリ１０９、ドライブ１１１及びドライブ１１２等を備える。これらは内部ネットワーク１１０を介して互いに通信可能に接続される。
【００１９】

キャッシュメモリ１０８は、ストレージシステム１０４のＩ／Ｏ処理のスループットやレスポンスを向上させるために、データを一時的なキャッシュとして格納するための高速アクセス可能なメモリである。プロセッサパッケージ１０７は、ローカルメモリ１１３及びプロセッサ１１４を備えて構成される。なおここではプロセッサパッケージ１０７は、１０７Ａ及び１０７Ｂの２つを図示しているが、数はこれに限定されない。

【００２０】

プロセッサ１１４は、ホスト１０１からのリード要求及びライト要求を処理するために、ドライブ１１１及び１１２と、キャッシュメモリ１０８との間のデータの転送処理を実行する。共有メモリ１０９は、プロセッサ１１４がリード要求又はライト要求を処理し、またストレージシステム１０４の機能（ボリュームのコピー機能など）を実行する上で必要な制御用の情報を格納するメモリである。共有メモリ１０９は、複数のプロセッサパッケージ１０７（ここでは１０７Ａ及び１０７Ｂ）のプロセッサ１１４により共有される情報を格納する。

【００２１】

ローカルメモリ１１３は、プロセッサ１１４がリード要求又はライト要求を処理し、またストレージシステム１０４の機能を実行する上で必要な制御用の情報を格納するメモリである。ローカルメモリ１１３は、各プロセッサ１１４により占有される情報を格納する。ローカルメモリ１１３には、例えばプロセッサ１１４により実行するプログラムが格納される。

【００２２】

ドライブ１１１及び１１２は、複数の半導体メモリ（SSD：Solid State Drive）から構成される記憶装置である。ここでのドライブ１１１は、１つのセルに２ビットの情報を格納するＭＬＣ ＳＳＤ（Multi Level Cell SSD）であり、ドライブ１１２は、１つのセルに３ビットの情報を格納するＴＬＣ（Triple Level Cell SSD）である。

【００２３】

ＭＬＣ ＳＳＤと、ＴＬＣ ＳＳＤとの違いは、上記の通りセル当たりの情報量が異なることに加えて、寿命（書き込み上限回数）及びビットコストが異なる。寿命は、一般にセル当たりの情報量が多くなるほど短くなる。よってＭＬＣ ＳＳＤの方がＴＬＣ ＳＳＤよりも長い。またビットコストは、単位面積当たりの情報量が多いほど低く抑えられるため、ＭＬＣ ＳＳＤの方がＴＬＣ ＳＳＤよりも高い。

【００２４】

すなわちＭＬＣ ＳＳＤは、寿命が長くてビットコストが高く、ＴＬＣ ＳＳＤは、寿命が短くてビットコストが低いと言える。なおここではＭＬＣ ＳＳＤ及びＴＬＣ ＳＳＤを利用してドライブ１１１及び１１２を構成するとしているが、これに限らず、ＳＬＣ ＳＳＤ（Single Level Cell SSD）を利用するとしてもよい。

【００２５】

ドライブ１１１及び１１２について、以降は同一種別であって特性の異なる記憶装置であることを強調する場合にはドライブ１１１をＭＬＣ ＳＳＤと呼び、ドライブ１１２をＴＬＣ ＳＳＤと呼ぶ場合がある。また複数のＭＬＣ ＳＳＤから構成されるパリティグループをＭＬＣ ＰＧと呼び、複数のＴＬＣ ＳＳＤから構成されるパリティグループをＴＬＣ ＰＧと呼ぶ場合がある。

【００２６】

（１－２）ストレージシステムの論理構成

図２は、ストレージシステム１０４の論理構成を示す。プロセッサ１１４がホスト１０１に提供する仮想ボリューム２０２Ａ及び２０２Ｂは、ホスト１０１から認識される論理的な記憶領域であり、ホスト１０１からのリード要求又はライト要求の発行対象となる記憶領域である。

【００２７】

プール２０４は、１個以上のプールボリューム２０６Ａ～２０６Ｅにより構成される。プールボリューム２０６Ａ～２０６Ｅは、それぞれＭＬＣ ＳＳＤ又はＴＬＣ ＳＳＤの

何れかの記憶領域から構成される。一又は複数のプールボリューム２０６Ａ～２０６Ｅにより、パリティグループ（PG：Parity Group）２０５Ａ又は２０５Ｂが形成される。

【００２８】

なおここでのパリティグループ２０５Ａは、ＭＬＣ ＳＳＤのみで構成され、パリティグループ２０５ＢはＴＬＣ ＳＳＤのみで構成されているが、必ずしもこれに限らず、ＭＬＣ ＳＳＤ及びＴＬＣ ＳＳＤが混在して１つのパリティグループを構成するとしてもよい。プールボリューム２０６は、パリティグループ２０５の一部又は全部の領域が切り出されて使用される。

【００２９】

プロセッサ１１４は、例えばホスト１０１が仮想ボリューム２０２Ａに対してライト要求を発行した場合、この仮想ボリューム２０２Ａにおいてライト要求の対象となった仮想的な記憶領域に対し、未使用の実記憶領域をプール２０４から所定単位（ページ）で割り当てる。

【００３０】

なおページとは、データを書き込む際の最小単位の記憶領域である。ここでは仮想ボリューム２０２Ａ又は２０２Ｂに対して割り当てられた仮想ページを２０１Ａ～２０１Ｅとして図示しており、これらのページ２０１Ａ～２０１Ｅに対して割り当てるプールボリューム２０６Ａ～２０６Ｅにおける実ページを２０７Ａ～２０７Ｆとして図示している。

【００３１】

次回ホスト１０１から同じページ２０１Ａに対してリード要求又はライト要求が発行された場合、プロセッサ１１４は、既に割り当てられているプールボリューム２０６Ａの記憶領域２０７Ａに対してＩ／Ｏ処理を実行することにより、あたかもホスト１０１が仮想ボリューム２０２Ａに対してＩ／Ｏ処理を実行しているように処理することができる。

【００３２】

すなわち仮想ボリューム２０２Ａ又は２０２Ｂを用いて、使用する部分のみプールボリューム２０６Ａ～２０６Ｅの記憶領域（ページ）２０７Ａ～２０７Ｆを割り当てることにより、限られた記憶容量を効率的に使用することできる。

【００３３】

ここで、各仮想ボリューム２０２Ａ又は２０２Ｂを構成するページ２０１毎に、ホスト１０１からの単位時間当たりの書き込み回数（これをライト頻度と呼ぶ）は異なる。よって例えばライト頻度が高いページ２０１Ａは、特性として書き込み上限回数の多いＭＬＣ ＳＳＤで構成されたパリティグループ２０５Ａに配置することで、ＴＬＣ ＳＳＤのような書き込み上限回数の少ない記憶装置の寿命を延ばすことができる。

【００３４】

上記の再配置は、具体的にはページ２０７Ｃに格納されたデータを未使用のページ２０７Ｂにコピーし、仮想ボリューム２０２Ａのページ２０１Ｃとプールボリューム２０６Ｂのページ２０７Ｃとの対応付けを仮想ボリューム２０２Ａのページ２０１Ｃとプールボリューム２０６Ａのページ２０７Ｂとの対応付けに変更することにより実行する。

【００３５】

（１－３）ページ配置処理の概念構成

図３は、ＴＬＣ－ＭＬＣ間で行われるページ再配置処理の概念構成を示す。ここではＭＬＣ ＳＳＤで構成されたパリティグループ（以下、ＭＬＣ ＰＧと呼ぶ）２０５Ａと、ＴＬＣ ＳＳＤで構成されたパリティグループ（以下、ＴＬＣ ＰＧと呼ぶ）２０５Ｂとの間で、ページ３０１を再配置する。ページ３０１の再配置先は、ＭＬＣ ＰＧ２０５Ａ及びＴＬＣ ２０５Ｂ上のページ３０１毎に採取しているモニタ情報に基づいて決定される。

【００３６】

具体的には、ページ３０１毎にライト回数を一定期間採取し、周期の満了後、モニタ情報に基づいて算出されたＭＬＣ ＰＧ２０５Ａ及びＴＬＣ ＰＧ２０５Ｂ毎のライト頻度と、ＴＬＣ－ＭＬＣ間ライト閾値３０２とに基づいてページ３０１の再配置先を決定する

。
【００３７】
図３では、ＴＬＣ－ＭＬＣ間ライト閾値３０２が２０００以上のライト負荷のページ３０１をＭＬＣ　ＰＧ２０５Ａに再配置し（矢印３０３Ａ）、２０００未満のライト負荷のページ３０１をＴＬＣ　ＰＧに再配置する（矢印３０３Ｂ）。
【００３８】
このように寿命の異なるＳＳＤで構成されたＭＬＣ　ＰＧ２０５ＡとＴＬＣ　ＰＧ２０５Ｂとの間で、ライト頻度の高いページを書き込み上限回数の多いＳＳＤで構成されたＭＬＣ　ＰＧ２０５Ａに再配置し、ライト頻度の低いページを書き込み上限回数の少ないＳＳＤで構成されたＴＬＣ　ＰＧ２０５Ｂに再配置する。
【００３９】
この結果、書き込み上限回数の少ないＳＳＤで構成されたＴＬＣ　ＰＧ２０５Ｂにおけるライト回数を削減することができ、書き込み上限回数の少ないＳＳＤ（ここではＴＬＣ　ＳＳＤ）の寿命を延ばすことができる。
【００４０】
図４は、同一特性のドライブ間（ＴＬＣ－ＴＬＣ間又はＭＬＣ－ＭＬＣ間）で行われるページ再配置処理の概念構成を示す。同一特性のドライブ間で行われるページ再配置処理は、書き込み上限回数が同程度であるＳＳＤで構成されたパリティグループ間でのライト頻度を調整することを目的として行われる。
【００４１】
実際にはパリティグループ間のライト頻度を調整するライトリバランスと、ライトリバランスにより崩れたパリティグループ間のＩ／Ｏ頻度を調整する性能リバランスとの２種類のリバランス処理を行う。
【００４２】
ここではＴＬＣ　ＰＧ間で行うライトリバランス及び性能リバランスについて説明する。まずライトリバランスでは、寿命を保証するために削減が必要なライト頻度（これをライト削減要求量と呼ぶ）又は寿命を保証することができる範囲内で追加可能なライト頻度（これをライト追加可能量と呼ぶ）をパリティグループ毎に算出する。
【００４３】
このときライト削減要求量が正の値のパリティグループをライト削減ＰＧ２０５Ｂとし、ライト追加可能量が正の値のパリティグループをライト追加ＰＧ２０５Ｃとする。そしてライト削減ＰＧ２０５Ｂのライト高負荷なページ４０１Ａとライト追加ＰＧ２０５Ｃのライト低負荷なページ４０１Ｂとを互いに移動して（矢印４０３）、パリティグループ間のライト頻度を調整する。
【００４４】
ライトリバランスの際の移動対象となるページは、モニタ情報を元にした閾値に基づいて決定される。閾値には、ライト高負荷なページを決定する閾値（これをライト削減閾値呼ぶ）４０５と、ライト低負荷なページを決定する閾値（これをライト追加閾値と呼ぶ）４０６とがある。ライトリバランスの際、ライト削減ＰＧ２０５Ｂでは、ライト削減閾値４０５以上のライト頻度のページ４０１Ａを移動対象とし、ライト追加ＰＧ２０５Ｃでは、ライト追加閾値４０６以下のライト頻度のページ４０１Ｂを移動対象とする。そしてそれぞれのページを移動することでライト頻度を調整する。
【００４５】
移動するページ数は、１回のページ移動で移動するライト頻度を算出し、目標とするライト削減要求量を満たすために必要な移動ページ数（これを移動計画ページ数と呼ぶ）を予め算出することにより決定される。そしてこの算出された移動計画ページ数分だけページの移動（再配置）が行われる。
【００４６】
次に性能リバランスでは、ライトリバランスにより移動したライト頻度の分だけリード頻度をライトリバランスとは逆の方向に移動することで、各パリティグループのリード頻

度及びライト頻度を合算したＩ／Ｏ頻度をパリティグループ間で調整する。具体的には、ライト削減ＰＧ２０５Ｂのリード低負荷なページ４０２Ｂとライト追加ＰＧ２０５Ｃのリード高負荷なページ４０２Ａとを移動して（矢印４０４）、パリティグループ間のリード頻度を調整する。

【００４７】

性能リバランスの際の移動対象となるページは、モニタ情報を元にした閾値に基づいて決定される。閾値には、リード高負荷なページを決定する閾値（これをリード削減閾値と呼ぶ）４０７とリード低負荷なページを決定する閾値（これをリード追加閾値と呼ぶ）４０８がある。

【００４８】

性能リバランスの際、ライト削減ＰＧ２０５Ｂでは、リード追加閾値４０８以下のリード頻度のページ４０２Ｂを移動対象とし、ライト追加ＰＧ２０５Ｃでは、リード削減閾値４０７以上のリード頻度のページ４０２Ａを移動対象とする。そしてそれぞれのページを移動することでリード頻度を調整する。移動するページ数は、ライトリバランスと同様、移動計画ページ数を予め算出することにより決定される。そしてこの移動計画ページ数分だけページの移動（再配置）が行われる。

【００４９】

（１－４）メモリの内部構成

図５は、共有メモリ１０９の内部構成を示す。共有メモリ１０９には、ページ毎モニタテーブル５０１、パリティグループ毎モニタテーブル５０２、パリティグループ毎再配置管理テーブル５０３、プール毎再配置管理テーブル５０４、キャッシュ管理テーブル５０５、ダイナミックマッピングテーブル５０６及び論理物理アドレス変換テーブル５０７が格納される。

【００５０】

ページ毎モニタテーブル５０１は、Ｉ／Ｏ回数を含む各ページのモニタ情報を管理するテーブルであり、パリティグループ毎モニタテーブル５０２は、Ｉ／Ｏ回数を含む各パリティグループのモニタ情報を管理するテーブルである。またパリティグループ毎再配置管理テーブル５０３は、パリティグループ毎のページ再配置に関する制御情報を管理するテーブルである。

【００５１】

プール毎再配置管理テーブル５０４は、プール毎のページ再配置に関する制御情報を管理するテーブルであり、キャッシュ管理テーブル５０５は、キャッシュメモリ１０８にデータを格納する際にキャッシュメモリ１０８にあるデータのダーティ／クリーン状態管理を保持するテーブルである。

【００５２】

またプール毎再配置管理テーブル５０４は、仮想ボリューム２０２に対して書き込むデータをキャッシュする場合には、キャッシュメモリ１０８のアドレスと、対応する仮想ボリューム２０２のページ２０１を特定するアドレスとを対応づけて管理するテーブルである。

【００５３】

ダイナミックマッピングテーブル５０６は、仮想ボリューム２０２の各ページ２０１と、各ページ２０１に割り当てられているプールボリューム２０６のページ２０７と、各ページ２０１のモニタ情報との対応関係を管理するテーブルである。

【００５４】

論理物理アドレス変換テーブル５０７は、パリティグループと、プールボリュームと、プールボリュームのデータを格納する物理ディスクに対応するパリティグループのアドレスとの対応関係を管理するテーブルである。

【００５５】

図６は、ローカルメモリ１１３の内部構成を示す。ローカルメモリ１１３には、ホストＩ／Ｏ処理プログラム６０１、デステージ処理プログラム６０２、寿命情報採取処理プロ

グラム６０３、閾値決定処理プログラム６０４、再配置処理プログラム６０４Ａ及び新規割り当て決定処理プログラム６０５が格納される。これらの各種プログラムは、プロセッサ１１４により実行される。

【００５６】

ホストＩ／Ｏ処理プログラム６０１は、ホスト１０１からのＩ／Ｏ要求を受領した場合に仮想ボリューム２０２に対するリード要求又はライト要求を処理するプログラムである。デステージ処理プログラム６０２は、キャッシュメモリ１０８上の物理ディスクに未反映のデータを物理ディスクに格納するプログラムである。この処理は、ホスト１０１からのＩ／Ｏ要求に対する処理とは非同期に実行される。

【００５７】

寿命情報採取処理プログラム６０３は、ドライブ１１１及び１１２に対して所定周期でコマンドを発行して寿命情報を採取し、採取した情報を共有メモリ１０９に反映するプログラムである。閾値決定処理プログラム６０４は、所定周期で採取したモニタ情報とドライブ１１１及び１１２の寿命情報とに基づいて、ページ再配置のための閾値を算出するプログラムである。

【００５８】

再配置処理プログラム６０４Ａは、閾値決定処理プログラム６０４により呼び出されるプログラムであり、閾値決定処理プログラム６０４により決定された各種閾値に基づいて、ページを再配置するプログラムである。新規割り当て決定処理プログラム６０５は、ホストＩ／Ｏ処理プログラム６０１に同期して実行され、仮想ボリューム２０２における新規の仮想ページに対して実ページの割り当て先であるパリティグループ２０５を閾値に基づいて決定するプログラムである。

【００５９】

（１－５）テーブル構成

図７は、ページ毎モニタテーブル５０１の論理構成を示す。ページ毎モニタテーブル５０１は、ページ番号欄７０１、ライトＩ／Ｏカウンタ欄７０２、リードＩ／Ｏカウンタ欄７０３、合計Ｉ／Ｏカウンタ欄７０４及び新規ライトフラグ欄７０５から構成される。

【００６０】

ページ番号欄７０１には、ページ２０１を特定するページ番号が格納され、ライトＩ／Ｏカウンタ欄７０２には、一定周期のライト回数が格納される。またリードＩ／Ｏカウンタ欄７０３には、一定周期のリード回数が格納され、合計Ｉ／Ｏカウンタ欄７０４には、一定周期のリード回数及びライト回数の合計Ｉ／Ｏ回数が格納される。

【００６１】

周期は、上記した閾値決定処理プログラム６０４がモニタ情報を採取する周期と同じであり、閾値決定処理プログラム６０４は、この一定期間のモニタ情報を処理対象とする。また新規ライトフラグ欄７０５には、ページが新規割り当てページか否かを示すフラグが格納される。

【００６２】

図８は、パリティグループ毎モニタテーブル５０２の論理構成を示す。パリティグループ毎モニタテーブル５０２は、パリティグループ番号欄８０１、最大ライト頻度欄８０２、最小ライト頻度欄８０３、最大リード頻度欄８０４、最小リード頻度欄８０５、リード／ライト比率欄８０６、ライト追加可能量欄８０７、ライト削減要求量欄８０８、新規ライトＩ／Ｏカウンタ欄８０９、新規ライト比率欄８１０、平均Ｉ／Ｏ頻度欄８１１及び割り当てページ数欄８１２から構成される。

【００６３】

パリティグループ番号欄８０１には、パリティグループ２０５を特定するパリティグループ番号が格納され、最大ライト頻度欄８０２には、パリティグループ内のページの最大ライト頻度が格納される。また最小ライト頻度欄８０３には、パリティグループ内のページの最小ライト頻度が格納される。

【００６４】

最大リード頻度欄８０４には、パリティグループ内のページの最大リード頻度が格納され、最小リード頻度欄８０５には、パリティグループ内のページの最小リード頻度が格納される。またリード／ライト比率欄８０６には、パリティグループに対するリード回数とライト回数との比率が格納される。

【００６５】

ライト追加可能量欄８０７には、寿命を保証することができる範囲内でパリティグループに追加可能なライト頻度が格納され、ライト削減要求量欄８０８には、寿命を保証するためにパリティグループから削減が必要なライト頻度が格納される。ライト追加可能量欄８０７及びライト削減要求量欄８０８には、パリティグループのライト頻度及びＳＳＤの寿命情報に基づいて算出される値が格納され、何れかが正の値となる。

【００６６】

ライト追加可能量欄８０７及びライト削減要求量欄８０８に格納される値は、ＳＳＤから採取できる寿命情報がライト追加可能率及びライト削減要求率である場合、下記式１及び２を計算して算出することができる。

【００６７】

なおライト追加可能率とは、現状のライト頻度を１００％として、追加可能なライト頻度の割合であり、ライト削減要求率とは、現状のライト頻度を１００％として寿命を維持するために削減すべきライト頻度の割合である。

【００６８】

【数１】

ライト追加可能量［ＩＯＰＨ（Input Output Per Hour）］
＝Ｍｉｎ（パリティグループを構成する全ＳＳＤのライト追加可能率［％］）
×パリティグループのライト頻度［ＩＯＰＨ］
・・・・・・（１）

【００６９】

【数２】

ライト削減要求量［ＩＯＰＨ］
＝Ｍａｘ（パリティグループを構成する全ＳＳＤのライト削減要求率［％］）
×パリティグループのライト頻度［ＩＯＰＨ］
・・・・・・（２）

【００７０】

新規ライトＩ／Ｏカウンタ欄８０９には、パリティグループに対する新規ライト回数が格納され、新規ライト比率欄８１０には、パリティグループにおける書き込み処理のうち、更新ライトと新規ライトとの比率が格納される。また平均Ｉ／Ｏ頻度欄８１１には、パリティグループにおける各ページの平均Ｉ／Ｏ頻度が格納され、割り当てページ数欄８１２には、パリティグループに割り当てられているページ数が格納される。

【００７１】

図９は、パリティグループ毎再配置管理テーブル５０３の論理構成を示す。パリティグループ毎再配置管理テーブル５０３は、パリティグループ番号欄９０１、メディアタイプ欄９０２、移動元ＰＧ種別欄９０３、移動先ＰＧ欄９０４、ライト削減閾値欄９０５、ライト追加閾値欄９０６、リード削減閾値欄９０７、リード追加閾値欄９０８、移動計画ページ数欄９０９Ａ及び９０９Ｂ、移動実績ページ数欄９１０Ａ及び９１０Ｂ並びに新規ライト可能量欄９１１から構成される。

【００７２】

パリティグループ番号欄９０１には、パリティグループを特定するパリティグループ番号が格納され、メディアタイプ欄９０２には、パリティグループを構成するＳＳＤの特性の情報が格納される。また移動元ＰＧ種別欄９０３には、ライト削減ＰＧ又はライト追加ＰＧの何れかを示す情報が格納される。移動元ＰＧ種別は、パリティグループ毎モニタテ

ーブル５０２のライト追加可能量欄８０７又はライト削減要求量欄８０８に格納される情報に基づいて決定される。

【００７３】

具体的には、ライト追加可能量欄８０７に正の値が格納されている場合、このパリティグループの移動元ＰＧ種別欄９０３には、ライト追加ＰＧを示す情報が格納される。またライト削減要求量欄８０８に正の値が格納されている場合、このパリティグループの移動元ＰＧ種別９０３には、ライト削減ＰＧを示す情報が格納される。

【００７４】

移動先ＰＧ欄９０４には、ページ再配置実行時の移動先のパリティグループ番号が格納される。移動先のパリティグループ番号は、再配置進捗度が最低のパリティグループを移動先として決定してもよい。再配置進捗度は、移動計画ページ数欄９０９Ａ及び９０９Ｂと、移動実績ページ数欄９１０Ａ及び９１０Ｂとを用いて、下記式３を計算して算出することができる。

【００７５】

【数３】

再配置進捗度＝移動実績ページ数／移動計画ページ数・・・・・・（３）

【００７６】

ライト削減閾値欄９０５には、ライト削減ＰＧでのライトリバランス時の移動対象のページを決定するための閾値が格納され、ライト追加閾値欄９０６には、ライト追加ＰＧでのライトリバランス時の移動対象のページを決定するための閾値が格納される。

【００７７】

リード削減閾値欄９０７には、ライト追加ＰＧでの性能リバランス時の移動対象ページを決定するための閾値が格納され、リード追加閾値欄９０８には、ライト削減ＰＧでの性能リバランス時の移動対象ページを決定するための閾値が格納される。ライト削減閾値、ライト追加閾値、リード削減閾値及びリード追加閾値は、下記式４を計算して算出することができる。

【００７８】

【数４】

ライト削減閾値［ＩＯＰＨ］＝最大ライト頻度［ＩＯＰＨ］×０．７
ライト追加閾値［ＩＯＰＨ］＝最小ライト頻度［ＩＯＰＨ］×１．３
リード削減閾値［ＩＯＰＨ］＝最大リード頻度［ＩＯＰＨ］×０．７
リード追加閾値［ＩＯＰＨ］＝最小リード頻度［ＩＯＰＨ］×１．３
・・・・・・（４）

【００７９】

移動計画ページ数（ライトリバランス）欄９０９Ａには、同一特性のＳＳＤにより構成されるパリティグループ間でライト頻度を調整するために必要なページの移動数が格納され、移動計画ページ数（性能リバランス）欄９０９Ｂには、同一特性のＳＳＤにより構成されるパリティグループ間でリード頻度を含むＩ／Ｏ頻度を調整するために必要なページの移動数が格納される。ライトリバランス用の移動計画ページ数は、下記式５を計算して算出することができる。

【００８０】

【数５】
ライトリバランス用の移動計画ページ数［ページ］
＝目標移動量［ＩＯＰＨ］÷１回のページ移動で変化するＩ／Ｏ頻度［ＩＯＰＨ／ページ］
目標移動量［ＩＯＰＨ］
＝Ｍｉｎ（ライト追加可能量［ＩＯＰＨ］、ライト削減要求量［ＩＯＰＨ］）
１回のページ移動で変化するＩ／Ｏ頻度［ＩＯＰＨ／ページ］
＝（ライト追加ＰＧの移動対象ページの平均Ｉ／Ｏ頻度［ＩＯＰＨ］
－ライト削減ＰＧの移動対象ページの平均Ｉ／Ｏ頻度［ＩＯＰＨ］）÷２
・・・・・・（５）

【００８１】
また性能リバランス用の移動計画ページ数は、下記式６を計算して算出することができる。

【００８２】
【数６】
性能リバランス用の移動計画ページ数［ページ］
＝目標移動量［ＩＯＰＨ］÷１回のページ移動で変化するＩ／Ｏ頻度［ＩＯＰＨ／ページ］
目標移動量［ＩＯＰＨ］
＝（（ライトリバランスで移動したライト頻度［ＩＯＰＨ］×２）＋
（ライト追加ＰＧの平均Ｉ／Ｏ頻度［ＩＯＰＨ］
－ライト削減ＰＧの平均Ｉ／Ｏ頻度［ＩＯＰＨ］））÷２
１回のページ移動で変化するＩ／Ｏ頻度［ＩＯＰＨ］
＝（ライト追加ＰＧの移動対象ページの平均Ｉ／Ｏ頻度［ＩＯＰＨ］
－ライト削減ＰＧの移動対象ページの平均Ｉ／Ｏ頻度［ＩＯＰＨ］）÷２
・・・・・・（６）

【００８３】
移動実績ページ数（ライトリバランス）欄９１０Ａには、ライトリバランスのために他のパリティグループに移動したページ数が格納され、移動実績ページ数（性能リバランス）欄９１０Ｂには、性能リバランスのために他のパリティグループに移動したページ数が格納される。新規ライト可能量欄９１２には、パリティグループで処理することのできる新規ライト回数が格納される。新規ライト可能量は、下記式７を計算して算出することができる。

【００８４】
【数７】
新規ライト可能量［Ｉ／Ｏ］
＝パリティグループの移動後のライト頻度［ＩＯＰＨ］
×新規割り当て比率［％］×再配置周期［Ｈ］
・・・・・・（７）

【００８５】
図１０は、プール毎再配置管理テーブル５０４の論理構成を示す。プール毎再配置管理テーブル５０４は、プール番号欄１００１、寿命制御再配置周期欄１００２、ＴＬＣ－ＭＬＣ間ライト閾値欄１００３、新規ライト閾値欄１００４、ワークロードタイプ欄１００５、同種ドライブ間新規割り当てポリシ欄１００６、新規割り当てバッファ欄１００７及び再配置バッファ欄１００８から構成される。

【００８６】
プール番号欄１００１には、プールを特定するプール番号が格納され、寿命制御再配置周期欄１００２には、プールにおいて寿命制御によるページの再配置を実行する周期が格納される。またＴＬＣ－ＭＬＣ間ライト閾値欄１００３には、ＴＬＣ　ＰＧとＭＬＣ　ＰＧとの間でページを再配置する際にどちらの種類のパリティグループにページを配置するのかを決定するための閾値が格納される。ＴＬＣ－ＭＬＣ間ライト閾値は、下記式８を計

算して算出することができる。
【００８７】
【数８】
ＴＬＣ－ＭＬＣ間ライト閾値［ＩＯＰＨ］
＝Ａｖｅ（全ＴＬＣ　ＰＧの限界ライト閾値［ＩＯＰＨ］）
ＴＬＣ　ＰＧの限界ライト閾値［ＩＯＰＨ］
＝ＩＦ（当該パリティグループの移動元ＰＧ種別がライト追加ＰＧ？）
当該パリティグループのライト頻度［ＩＯＰＨ］＋ライト追加可能量［ＩＯＰＨ］
　　ＥＬＳＥ　ＩＦ（当該パリティグループの移動元ＰＧ種別がライト削減ＰＧ？）
当該ＰＧのライト頻度［ＩＯＰＨ］－ライト削減要求量［ＩＯＰＨ］
・・・・・・（８）

【００８８】
新規ライト閾値欄１００４には、プールにおいて新規ページに対する割り当て処理を行う際、ＴＬＣ　ＰＧとＭＬＣ　ＰＧのどちらから新規ページに対する実ページの割り当てを行うかを決定するための閾値が格納される。新規ライト閾値は、ページ移動時のペナルティ（＝１ページあたりのデータサイズ）としてストレージシステム１０４が指定してもよい。
【００８９】
ワークロードタイプ欄１００５には、プールに関連付けられている仮想ボリュームに対してホスト１０１から発行されるＩ／Ｏ要求の特性の情報が格納される。ワークロードタイプには、例えばＷｒｉｔｅ　ｉｎｔｅｎｓｉｖｅ、Ｒｅａｄ　ｉｎｔｅｎｓｉｖｅ、Ｕｎｋｎｏｗｎ等がある。
【００９０】
Ｗｒｉｔｅ　ｉｎｔｅｎｓｉｖｅは、ホスト１０１からのＩ／Ｏ要求のうち、ライト比率が高いことを意味し、Ｒｅａｄ　ｉｎｔｅｎｓｉｖｅは、ホスト１０１からのＩ／Ｏ要求のうち、リード比率が高いことを意味し、Ｕｎｋｎｏｗｎは、ホスト１０１からのＩ／Ｏ要求のうち、リード／ライト比率が不明であることを意味する。ワークロードタイプは、ユーザが指定してもよいし、ストレージシステム１０４が自動で決定してもよい。
【００９１】
同種ドライブ間新規割り当てポリシ欄１００６には、新規割り当て処理でＴＬＣ　ＰＧ内又はＭＬＣ　ＰＧ内でどのパリティグループから新規ページに対する実ページの割り当てを行うかを決定するための情報が格納される。新規割り当てポリシには、例えばパリティグループ間で順番に割り当てるラウンドロビン、容量の多いパリティグループから割り当てる容量優先又はライト追加可能量の多いパリティグループから割り当てる寿命優先がある。新規割り当てポリシは、ユーザが指定してもよいし、ストレージシステム１０４が自動で決定してもよい。
【００９２】
新規割り当てバッファ欄１００７には、プールで新規割り当て処理のために使用するバッファのプール容量に対する比率が格納される。再配置バッファ欄１００８には、プールでページの再配置処理のために使用するバッファのプール容量に対する比率が格納される。なおこのプール毎再配置管理テーブル５０４内には、再配置の制御情報のデフォルト値を格納するエントリが１つ存在し、ユーザからの指定がない制御情報にはデフォルト値が格納される。
【００９３】
（１－６）フローチャート
図１１は、ホストＩ／Ｏ処理のフローチャートを示す。このホストＩ／Ｏ処理は、ストレージシステム１０４がホスト１０１からのＩ／Ｏ要求を受領したことを契機として、プロセッサ１１４とホストＩ／Ｏ処理プログラム６０１との協働により実行される。説明の便宜上、処理主体をホストＩ／Ｏ処理プログラム６０１として説明する。

【００９４】
まずホストＩ／Ｏ処理プログラム６０１は、ホスト１０１からのＩ／Ｏ処理要求を受領すると、受領したＩ／Ｏ処理要求が仮想ボリューム２０２に対してデータを書き込むライト要求であるか否かを判断する（Ｓ１１０１）。
【００９５】
ライト要求である場合（Ｓ１１０１：Ｙ）、ホストＩ／Ｏ処理プログラム６０１は、ダイナミックマッピングテーブル５０６を参照して、ライト対象の仮想ページに対して実ページが割り当て済みであるか否かを判断する。そして未割り当ての場合には、未使用の実ページを割り当てる新規割り当て決定処理を実行する（Ｓ１１０９）。新規割り当て決定処理の詳細については後述する（図１９）。
【００９６】
新規割り当て決定処理を実行した後、次いでホストＩ／Ｏ処理プログラム６０１は、仮想ボリューム２０２上のライト対象のアドレスに対応した領域がキャッシュメモリ１０８上に確保されているか否かを確認し、確保されていない場合にはキャッシュメモリ１０８上の領域を確保する（Ｓ１１１０）。
【００９７】
次いでホストＩ／Ｏ処理プログラム６０１は、ホスト１０１に対してライト用のデータを送信するように通知する。ホストＩ／Ｏ処理プログラム６０１は、ホスト１０１からデータが送信されると、このデータをキャッシュメモリ１０８上の確保した領域に書き込む（Ｓ１１１１）。そしてホストＩ／Ｏ処理プログラム６０１は、まだドライブ１１１又は１１２に書き込みが完了していない領域であることを示すダーティフラグをキャッシュ管理テーブル５０５においてＯＮに設定する。
【００９８】
ここで、ダーティフラグは、キャッシュメモリ１０８にのみデータが格納されており、ドライブ１１１又は１１２にはまだ格納されていない状態を示す情報である。ダーティフラグは、キャッシュメモリ１０８の領域を管理するキャッシュ管理テーブル５０５においてＯＮ又はＯＦに設定される。キャッシュ管理テーブル５０５は、後述するデステージ処理（図１２）において参照される。
【００９９】
そしてダーティフラグがＯＮである場合にはキャッシュメモリ１０８上の領域に格納されているデータは、ドライブ１１１又は１１２に書き込まれる。ドライブ１１１又は１１２に書き込まれた後は、ダーティフラグはＯＦＦに設定され、リード処理に対応してドライブ１１１又は１１２から読み込んだデータをキャッシュメモリ１０８に格納した場合を含めて、キャッシュ管理テーブル５０５にはクリーンフラグがＯＮに設定される。
【０１００】
このようにキャッシュ管理テーブル５０５は、キャッシュメモリ１０８上のアドレスと対応する仮想ボリューム２０２のアドレスと、キャッシュメモリ１０８上のデータの状態とを少なくとも管理する。なおキャッシュメモリ１０８のアドレスに対応する仮想ボリューム２０２のアドレスは、仮想ボリューム２０２のデータを置くためにキャッシュメモリ１０８上の領域を確保した場合にのみ、有効な値である仮想ボリューム２０２のアドレスが格納される。
【０１０１】
以上の処理を行った後、ホストＩ／Ｏ処理プログラム６０１は、ホスト１０１にＩ／Ｏ処理（ライト処理）が完了したことを通知して（Ｓ１１１２）、本処理を終了する。
【０１０２】
ステップＳ１１０１に戻り、ホストＩ／Ｏ処理プログラム６０１は、ホスト１０１から受領したＩ／Ｏ要求が仮想ボリューム２０２からデータを読み込むリード要求である場合（Ｓ１１０１：Ｎ）、キャッシュ管理テーブル５０５を参照して、リード要求に対応する仮想ボリューム２０２上のアドレスに対応したデータがキャッシュメモリ１０８上に格納されているか否かを判断する（Ｓ１１０２）。

【０１０３】
リード要求に対応する仮想ボリューム２０２上のアドレスに対応したデータがキャッシュメモリ１０８上に格納されている場合をキャッシュヒットと呼ぶ。キャッシュヒットである場合（Ｓ１１０２：Ｙ）、ホストＩ／Ｏ処理プログラム６０１は、キャッシュメモリ１０８上のデータをホスト１０１に転送するとともに（Ｓ１１０８）、Ｉ／Ｏ処理（リード処理）が完了したことをホスト１０１に通知して、本処理を終了する。
【０１０４】
これに対し、キャッシュヒットしなかった場合（Ｓ１１０２：Ｎ）、ホストＩ／Ｏ処理プログラム６０１は、リード対象の仮想ボリューム２０２のアドレスに対応したデータを格納するための領域をキャッシュメモリ１０８上に確保する（Ｓ１１０３）。次いでホストＩ／Ｏ処理プログラム６０１は、ダイナミックマッピングテーブル５０６を参照して、リード対象の仮想ボリューム２０２のアドレスにプール２０４から実ページが割り当てられているか否かを確認する。
【０１０５】
仮想ボリューム２０２に実ページが割り当てられていない場合、ホストＩ／Ｏ処理プログラム６０１は、ダイナミックマッピングテーブル５０６を参照して、デフォルト値を格納するページを用いてデフォルト値の格納ページのドライブ１１１又は１１２のアドレスを算出し、ドライブ１１１又は１１２からデフォルト値をキャッシュメモリ１０８の領域に転送する（Ｓ１１０５）。
【０１０６】
ここで、デフォルト値の場合は、ダイナミックマッピングテーブル５０６の仮想ボリューム２０２及び論理アドレスに対応する、プールボリューム番号及び論理アドレスには、デフォルト値格納ページのあるプールボリューム番号及び論理アドレスが設定されている。デフォルト値格納ページは、プール２０４に１つ以上あればよい。容量効率を考えればプール２０４にデフォルト値ページは１又は２つである。
【０１０７】
デフォルト値格納ページのアドレスと対応付けられている仮想ボリューム２０２の論理アドレスは、ホスト１０１から新規にデータの書き込みがあった際に、ホスト１０１のデータ書き込み用のページで未だどの仮想ボリューム２０２のアドレスにも対応付けられていない未使用のページと対応付け直される。
【０１０８】
これに対し、仮想ボリューム２０２に実ページが割り当てられている場合、ホストＩ／Ｏ処理プログラム６０１は、ダイナミックマッピングテーブル５０６を参照して、プールボリューム番号及び論理アドレスを取得し、更に論理物理アドレス変換テーブル５０７を参照して、物理ドライブ番号及び物理開始アドレスを算出することで、リード対象の仮想ボリューム２０２のアドレスに対応したデータが格納されているドライブ１１１又は１１２のアドレスを算出する（Ｓ１１０４）。
【０１０９】
次にホストＩ／Ｏ処理プログラム６０１は、算出したアドレスからデータをキャッシュメモリ１０８上の領域に転送する（Ｓ１１０５）。そしてリード時にページ毎モニタテーブル５０１を参照して、リードＩ／Ｏカウンタの数値をカウントアップする（Ｓ１１０６）。
【０１１０】
そしてホストＩ／Ｏ処理プログラム６０１は、ドライブ１１１又は１１２からキャッシュメモリ１０８上に格納したデータをホスト１０１に転送するとともに（Ｓ１１０７）、Ｉ／Ｏ処理（リード処理）が完了したことをホスト１０１に通知して、本処理を終了する。
【０１１１】
以上の処理により、ホストＩ／Ｏ処理を行うとともに、必要なモニタ情報を採取することができる。

【０１１２】
図１２は、デステージ処理のフローチャートを示す。このデステージ処理は、ホストＩ／Ｏ処理（図１１）とは非同期で、プロセッサ１１４とデステージ処理プログラム６０２との協働により適宜実行される。説明の便宜上、処理主体をデステージ処理プログラム６０２として説明する。
【０１１３】
まずデステージ処理プログラム６０２は、キャッシュ管理テーブル５０５を参照して、ダーティフラグのＯＮ又はＯＦＦを確認し、ドライブ１１１又は１１２に書き込みがなされていない未反映データがキャッシュメモリ１０８上にあるか否かを判断する（Ｓ１２０１）。
【０１１４】
デステージ処理プログラム６０２は、未反映データがキャッシュメモリ１０８上にある場合、キャッシュ管理テーブル５０５から仮想ボリューム番号及び論理アドレスを取得する。そしてこの仮想ボリューム番号及び論理アドレスを元にダイナミックマッピングテーブル５０６を参照して、プールボリューム番号及び論理アドレスを取得する。
【０１１５】
このときプールボリューム番号及び論理アドレスがデフォルト値格納ページのアドレスであった場合、デステージ処理プログラム６０２は、新規データを書き込むためにダイナミックマッピングテーブル５０６から新規空きページを割り当てる。そしてこの割り当てページのプールボリューム番号及び論理アドレスをダイナミックマッピングテーブル５０６の対応する仮想ボリューム番号及び論理アドレスに対応付けて格納する。
【０１１６】
既にページが割り当てられている場合は、デフォルト値のプールボリューム番号及び論理アドレスとは異なるプールボリューム番号及び論理アドレスの値が仮想ボリュームの論理アドレスに対応付けて格納されている。デステージ処理プログラム６０２は、プールボリューム番号及び論理アドレスを取得した後、論理物理アドレス変換テーブル５０７を参照して、ドライブ１１１又は１１２のアドレスを算出する（Ｓ１２０２）。
【０１１７】
次いでデステージ処理プログラム６０２は、算出したドライブ１１１又は１１２のアドレスに対して、キャッシュメモリ１０８上の未反映データを書き込む（Ｓ１２０３）。そしてダイナミックマッピングテーブル５０６のページ番号に対応するページ毎モニタテーブル５０１を参照して、ライトＩ／Ｏカウンタの数値をカウントアップする（Ｓ１２０４）。
【０１１８】
次いでデステージ処理プログラム６０２は、ページ毎モニタテーブル５０１の新規ライトフラグ欄７０５を参照して、デステージ処理対象のページが新規割り当てページであるか否かを判断する（Ｓ１２０５）。新規割り当てページである場合（Ｓ１２０５：Ｙ）、デステージ処理プログラム６０２は、このページの格納先のパリティグループに対応するパリティグループ毎モニタテーブル５０２を参照して、新規ライトＩ／Ｏカウンタの数値をカウントアップする（Ｓ１２０６）。
【０１１９】
これに対し、デステージ処理プログラム６０２は、テージ処理対象のページが新規割り当てページでない場合（Ｓ１２０５：Ｎ）、ステップＳ１２０１に移行する。ステップＳ１２０１においてデステージ処理プログラム６０２は、キャッシュメモリ１０８上に未反映データがさらにあるか否かを判断する。そしてさらなる未反映データがない場合には（Ｓ１２０１：Ｎ）、本処理を終了する。
【０１２０】
以上の処理により、キャッシュメモリ１０８上のデータを非同期にドライブ１１１又は１１２に格納するとともに、必要なモニタ情報を採取することができる。
【０１２１】

図１３は、寿命情報採取処理のフローチャートである。この寿命情報採取処理は、プロセッサ１１４と寿命情報採取処理プログラム６０３との協働により一定周期で実行される。説明の便宜上、処理主体を寿命情報採取処理プログラム６０３として説明する。

【０１２２】

まず寿命情報採取処理プログラム６０３は、ドライブ１１１又は１１２に対して寿命採取のコマンドを発行する（Ｓ１３０１）。次いで寿命情報採取処理プログラム６０３は、寿命情報としてライト追加可能率又はライト削減要求率を受信する（Ｓ１３０２）。そして受信した寿命情報をローカルメモリ１１３に格納して（Ｓ１３０３）、本処理を終了する。

【０１２３】

以上の処理により、ドライブ１１１又は１１２から寿命情報を採取することができる。

【０１２４】

図１４は、閾値決定処理のフローチャートである。この閾値決定処理は、プロセッサ１１４と閾値決定処理プログラム６０４との協働により一定周期で実行される。一定周期の情報は、プール毎再配置管理テーブル５０４の寿命制御再配置周期欄１００２に格納される。説明の便宜上、処理主体を閾値決定処理プログラム６０４として説明する。

【０１２５】

まず閾値決定処理プログラム６０４は、全てのパリティグループについて、ページ毎のモニタ情報の集計が完了したか否かを判断する（Ｓ１４０１）。すなわちパリティグループ毎モニタテーブル５０２の各欄に情報が格納されているか否かを判断する。

【０１２６】

集計が完了していない場合（Ｓ１４０１：Ｎ）、閾値決定処理プログラム６０４は、寿命情報採取処理プログラム６０３を呼び出してドライブ１１１又は１１２から寿命情報を採取し（Ｓ１４０２）、ページ毎のモニタ情報をパリティグループ毎に集計する（Ｓ１４０３）。集計が完了している場合（Ｓ１４０１：Ｙ）、閾値決定処理プログラム６０４は、ページ再配置のための各種閾値を算出する（Ｓ１４０４）。

【０１２７】

ここで算出される各種閾値は、パリティグループ毎モニタテーブル５０２、パリティグループ毎再配置管理テーブル５０３及びプール毎再配置管理テーブル５０４の各欄に格納される。そして閾値決定処理プログラム６０４は、再配置処理プログラム６０４Ａを呼び出して再配置処理を実行した後（Ｓ１４０５）、本処理を終了する。

【０１２８】

以上の処理により、パリティグループ毎にモニタ情報を集計し、集計したモニタ情報に基づいて、各種閾値を算出することができる。そして算出した閾値を用いて再配置処理を実行することができる。

【０１２９】

図１５は、再配置処理のフローチャートである。この再配置処理は、プロセッサ１１４と閾値決定処理プログラム６０４により呼び出される再配置処理プログラム６０４Ａとの協働により実行される。説明の便宜上、処理主体を再配置処理プログラム６０４Ａとして説明する。

【０１３０】

まず再配置処理プログラム６０４Ａは、閾値決定処理プログラム６０４により算出されたＴＬＣ－ＭＬＣ間ライト閾値に基づいて、ＴＬＣ　ＰＧとＭＬＣ　ＰＧとの間でページを再配置するＴＬＣ－ＭＬＣ間再配置処理を実行する（Ｓ１５０１）。

【０１３１】

このＴＬＣ－ＭＬＣ間再配置処理により、ＴＬＣ　ＳＳＤ（ドライブ１１２）の寿命を延ばすことができる。またビットコストを削減することができる。

【０１３２】

次に再配置処理プログラム６０４Ａは、パリティグループ毎再配置管理テーブル５０３を参照し、ライトリバランス用の移動計画ページ数に基づいて、同一特性のＳＳＤで構成

されたパリティグループ間（ＴＬＣ　ＰＧ間又はＭＬＣ　ＰＧ間）でページを再配置するライトリバランス処理を実行する（Ｓ１５０２）。

【０１３３】

このライトリバランス処理により、同一特性のＳＳＤで構成されたパリティグループ間でライト負荷を分散して、寿命を平準化することができる。

【０１３４】

次に再配置処理プログラム６０４Ａは、パリティグループ毎再配置管理テーブル５０３を参照して、ライト削減ＰＧのライトリバランス用の移動計画ページ数が０以下であるか否かを判断する（Ｓ１５０３）。ライト削減ＰＧのライトリバランス用の移動計画ページ数が０以下である場合（Ｓ１５０３：Ｙ）、再配置処理プログラム６０４Ａは、ライト頻度の調整が完了していると判断して、次に性能リバランス処理を実行する（Ｓ１５０４）。

【０１３５】

これに対し、ライト削減ＰＧのライトリバランス用の移動計画ページ数が０よりも大きい場合（Ｓ１５０３：Ｎ）、再配置処理プログラム６０４Ａは、現在のパリティグループの構成では、ライト頻度を調整できず、寿命を維持することができないため、警告画面を表示して、寿命を保証するために追加すべきＴＬＣ　ＳＳＤ又はＭＬＣ　ＳＳＤの容量をユーザに通知する（Ｓ１５０５）。

【０１３６】

以上の処理により、パリティグループ間でのライト頻度及びリード頻度を調整することができる。またライト頻度及びリード頻度を調整することができない場合には寿命を維持するために必要なＴＬＣ　ＳＳＤ又はＭＬＣ　ＳＳＤの容量をユーザに通知することができる。

【０１３７】

図１６は、ＴＬＣ－ＭＬＣ間再配置処理のフローチャートである。このＴＬＣ－ＭＬＣ間再配置処理は、プロセッサ１１４と再配置処理プログラム６０４Ａとの協働により実行される。説明の便宜上、処理主体を再配置処理プログラム６０４Ａとして説明する。

【０１３８】

まず再配置処理プログラム６０４Ａは、閾値決定処理プログラム６０４により算出されたＴＬＣ－ＭＬＣ間ライト閾値に基づいて、全てのパリティグループ内の各ページをＴＬＣ　ＰＧ又はＭＬＣ　ＰＧに再配置する。このとき再配置について未判定のページがあるか否かを判断する（Ｓ１６０１）。

【０１３９】

未判定のページがある場合（Ｓ１６０１：Ｙ）、再配置処理プログラム６０４Ａは、パリティグループ毎再配置管理テーブル５０３を参照して、この未判定の対象ページが所属するパリティグループのメディアタイプがＴＬＣであるか否かを判断する（Ｓ１６０２）。メディアタイプがＴＬＣである場合（Ｓ１６０２：Ｙ）、再配置処理プログラム６０４Ａは、対象ページのライト頻度がＴＬＣ－ＭＬＣ間ライト閾値以上であるか否かを判断する（Ｓ１６０３）。

【０１４０】

対象ページのライト頻度がＴＬＣ－ＭＬＣ間ライト閾値以上である場合（Ｓ１６０３：Ｙ）、再配置処理プログラム６０４Ａは、このページをＴＬＣ　ＰＧからＭＬＣ　ＰＧに移動する（Ｓ１６０４）。これに対し、対象ページのライト頻度がＴＬＣ－ＭＬＣ間ライト閾値未満である場合（Ｓ１６０３：Ｎ）、再配置処理プログラム６０４Ａは、何もせずにステップＳ１６０１に移行する。

【０１４１】

ステップＳ１６０２に戻り、メディアタイプがＭＬＣである場合（Ｓ１６０２：Ｎ）、対象ページのライト頻度がＴＬＣ－ＭＬＣ間ライト閾値未満であるか否かを判断する（Ｓ１６０５）。対象ページのライト頻度がＴＬＣ－ＭＬＣ間ライト閾値未満である場合（Ｓ１６０５：Ｙ）、再配置処理プログラム６０４Ａは、このページをＭＬＣ　ＰＧからＴＬ

Ｃ　ＰＧに移動する（Ｓ１６０６）。
【０１４２】
これに対し、対象ページのライト頻度がＴＬＣ－ＭＬＣ間ライト閾値以上である場合（Ｓ１６０５：Ｎ）、再配置処理プログラム６０４Ａは、何もせずにステップＳ１６０１に移行する。再配置処理プログラム６０４Ａは、全てのパリティグループ内の各ページについて判定を終えると、本処理を終了する。以上の処理により、ＴＬＣ－ＭＬＣ間でページを再配置して、ＴＬＣ　ＳＳＤ（ドライブ１１２）の寿命を延ばすとともに、ビットコストを削減することができる。
【０１４３】
図１７は、ライトリバランス処理のフローチャートである。このライトリバランス処理は、プロセッサ１１４と再配置処理プログラム６０４Ａとの協働により実行される。説明の便宜上、処理主体を再配置処理プログラム６０４Ａとして説明する。
【０１４４】
まず再配置処理プログラム６０４Ａは、閾値決定処理プログラム６０４により算出されたライトリバランス用移動計画ページ数、ライト削減閾値及びライト追加閾値に基づいて、同一特性のＳＳＤで構成されたパリティグループ間（ＴＬＣ　ＰＧ間又はＭＬＣ　ＰＧ間）でページを再配置する。このとき再配置について未判定のページがあるか否かを判断する（Ｓ１７０１）。
【０１４５】
未判定のページがない場合には（Ｓ１７０１：Ｎ）、再配置処理プログラム６０４Ａは、本処理を終了する。これに対し、未判定のページがある場合（Ｓ１７０１：Ｙ）、再配置処理プログラム６０４Ａは、パリティグループ毎再配置管理テーブル５０３を参照して、この未判定の対象ページが所属するパリティグループの移動元ＰＧ種別がライト削減ＰＧであるか否かを判断する（Ｓ１７０２）。
【０１４６】
対象ページが所属するパリティグループの移動元ＰＧ種別がライト削減ＰＧである場合（Ｓ１７０２：Ｙ）、再配置処理プログラム６０４Ａは、ページ毎モニタテーブル５０１のライトＩ／Ｏカウンタを参照して、この対象ページのライト頻度を取得する。そして取得したライト頻度がライト削減閾値以上であるか否かを判断する（Ｓ１７０３）。
【０１４７】
対象ページのライト頻度がライト削減閾値未満である場合（Ｓ１７０３：Ｎ）、再配置処理プログラム６０４Ａは、ステップＳ１７０１に移行する。これに対し、対象ページのライト頻度がライト削減閾値以上である場合（Ｓ１７０３：Ｙ）、再配置処理プログラム６０４Ａは、この対象ページの移動先のパリティグループを決定する。
【０１４８】
移動先のパリティグループを決定する際、再配置処理プログラム６０４Ａはパリティグループ毎再配置管理テーブル５０３を参照して、ライト追加ＰＧのうち、ライトリバランス用の移動実績ページ数がライトリバランス用の移動計画ページ数未満であるパリティグループが存在するか否かを判断する（Ｓ１７０４）。
【０１４９】
ライト追加ＰＧのうち、ライトリバランス用の移動実績ページ数がライトリバランス用の移動計画ページ数未満であるパリティグループが存在する場合（Ｓ１７０４：Ｙ）、再配置処理プログラム６０４Ａは、このパリティグループにライト負荷の高い対象ページを移動しても寿命を維持することができると判断して、このパリティグループを移動先ＰＧとしてパリティグループ毎再配置管理テーブル５０３に登録し、対象ページをこのパリティグループに移動する（Ｓ１７０５）。
【０１５０】
これに対し、ライト追加ＰＧのうち、ライトリバランス用の移動実績ページ数がライトリバランス用の移動計画ページ数未満であるパリティグループが存在しない場合（Ｓ１７０４：Ｎ）、再配置処理プログラム６０４Ａは、この対象ページについては判定を終え、

ステップＳ１７０１に移行する。

【０１５１】

ステップＳ１７０２に戻り、対象ページが所属するパリティグループの移動元ＰＧ種別がライト削減ＰＧでない場合（Ｓ１７０２：Ｎ）、すなわち対象ページが所属するパリティグループの移動元ＰＧ種別がライト追加ＰＧである場合、再配置処理プログラム６０４Ａは、ページ毎モニタテーブル５０１のライトＩ／Ｏカウンタ欄７０２を参照して、この対象ページのライト頻度を取得する。そして取得したライト頻度がライト追加閾値未満であるか否かを判断する（Ｓ１７０６）。

【０１５２】

対象ページのライト頻度がライト追加閾値以上である場合には（Ｓ１７０６：Ｎ）、再配置処理プログラム６０４Ａは、ステップＳ１７０１に移行する。これに対し、対象ページのライト頻度がライト追加閾値未満である場合（Ｓ１７０６：Ｙ）、再配置処理プログラム６０４Ａは、この対象ページの移動先のパリティグループを決定する。

【０１５３】

移動先のパリティグループを決定する際の処理は、上記のステップＳ１７０４及びＳ１７０５と同様であるため説明を省略する。以上の処理により、同一特性のＳＳＤで構成された異なるパリティグループ間でページを再配置してライト負荷を分散し、寿命を平準化することができる。

【０１５４】

図１８は、性能リバランス処理のフローチャートである。この性能リバランス処理は、プロセッサ１１４と再配置処理プログラム６０４Ａとの協働により実行される。説明の便宜上、処理主体を再配置処理プログラム６０４Ａとして説明する。

【０１５５】

なお性能リバランス処理は、ライトリバランス処理（図１７）で調整したライト頻度と同程度のリード頻度のページをライトリバランス処理で移動させた方向とは逆方向に移動する点で、ライトリバランス処理と異なり、他の基本的な処理内容は同様である。

【０１５６】

まず再配置処理プログラム６０４Ａは、閾値決定処理プログラム６０４により算出された性能リバランス用の移動計画ページ数、ライト削減閾値及びライト追加閾値に基づいて、同一特性のＳＳＤで構成されたパリティグループ間（ＴＬＣ　ＰＧ間又はＭＬＣ　ＰＧ間）でページを再配置する。このとき再配置について未判定のページがあるか否かを判断する（Ｓ１８０１）。

【０１５７】

未判定のページがない場合には（Ｓ１８０１：Ｎ）、再配置処理プログラム６０４Ａは、本処理を終了する。これに対し、未判定のページがある場合（Ｓ１８０１：Ｙ）、再配置処理プログラム６０４Ａは、パリティグループ毎再配置管理テーブル５０３を参照して、この未判定の対象ページが所属するパリティグループの移動元ＰＧ種別がライト削減ＰＧであるか否かを判断する（Ｓ１８０２）。

【０１５８】

対象ページが所属するパリティグループの移動元ＰＧ種別がライト削減ＰＧである場合（Ｓ１８０２：Ｙ）、再配置処理プログラム６０４Ａは、ページ毎モニタテーブル５０１のリードＩ／Ｏカウンタを参照して、この対象ページのリード頻度を取得する。そして取得したリード頻度がリード追加閾値未満であるか否かを判断する（Ｓ１８０３）。

【０１５９】

対象ページのリード頻度がリード追加閾値以上である場合には（Ｓ１８０３：Ｎ）、再配置処理プログラム６０４Ａは、ステップＳ１８０１に移行する。これに対し、対象ページのリード頻度がリード追加閾値未満である場合（Ｓ１８０３：Ｙ）、再配置処理プログラム６０４Ａは、この対象ページの移動先のパリティグループを決定する。

【０１６０】

移動先のパリティグループを決定する際、再配置処理プログラム６０４Ａはパリティグ

ループ毎再配置管理テーブル５０３を参照して、ライト追加ＰＧのうち、性能リバランス用の移動実績ページ数が性能リバランス用の移動計画ページ数未満であるパリティグループが存在するか否かを判断する（Ｓ１８０４）。

【０１６１】

ライト追加ＰＧのうち、性能リバランス用の移動実績ページ数が性能リバランス用の移動計画ページ数未満であるパリティグループが存在する場合（Ｓ１８０４：Ｙ）、再配置処理プログラム６０４Ａは、このパリティグループにリード負荷の低い対象ページを移動させても高負荷にならないと判断して、このパリティグループを移動先ＰＧとしてパリティグループ毎再配置管理テーブル５０３に登録し、対象ページをこのパリティグループに移動する（Ｓ１８０５）。

【０１６２】

これに対し、ライト追加ＰＧのうち、性能リバランス用の移動実績ページ数が性能リバランス用の移動計画ページ数未満であるパリティグループが存在しない場合（Ｓ１８０４：Ｎ）、再配置処理プログラム６０４Ａは、この対象ページについては判定を終え、ステップＳ１８０１に移行する。

【０１６３】

ステップＳ１８０２に戻り、対象ページが所属するパリティグループの移動元ＰＧ種別がライト削減ＰＧでない場合（Ｓ１８０２：Ｎ）、すなわち対象ページが所属するパリティグループの移動元ＰＧ種別がライト追加ＰＧである場合、再配置処理プログラム６０４Ａは、ページ毎モニタテーブル５０１のリードＩ／Ｏカウンタ欄７０３を参照して、この対象ページのリード頻度を取得する。そして取得したリード頻度がリード削減閾値以上であるか否かを判断する（Ｓ１８０６）。

【０１６４】

対象ページのリード頻度がリード削減閾値未満である場合には（Ｓ１８０６：Ｎ）、再配置処理プログラム６０４Ａは、ステップＳ１８０１に移行する。これに対し、対象ページのリード頻度がリード削減閾値以上である場合（Ｓ１８０６：Ｙ）、再配置処理プログラム６０４Ａは、この対象ページの移動先のパリティグループを決定する。

【０１６５】

移動先のパリティグループを決定する際の処理は、上記のステップＳ１８０４及びＳ１８０５と同様であるため説明を省略する。以上の処理により、同一特性のＳＳＤで構成された異なるパリティグループ間でページを再配置してリード負荷を分散し、Ｉ／Ｏ頻度を平準化することができる。

【０１６６】

図１９は、新規割り当て決定処理のフローチャートを示す。この新規割り当て決定処理は、ホスト１０１から新規の仮想ページに対するライト要求を受領したことを契機として、プロセッサ１１４と新規割り当て決定処理プログラム６０５との協働により実行される。説明の便宜上、処理主体を新規割り当て決定処理プログラム６０５として説明する。

【０１６７】

まず新規割り当て決定処理プログラム６０５は、プール毎再配置管理テーブル５０４を参照して、新規ページの割り当て対象の仮想ボリューム２０２に記憶領域を提供するプール２０４のワークロードタイプが「Ｕｎｋｎｏｗｎ」であるか否かを判断する（Ｓ１９０１）。

【０１６８】

プール２０４のワークロードタイプが「Ｕｎｋｎｏｗｎ」ではない場合（Ｓ１９０１：Ｎ）、新規割り当て決定処理プログラム６０５は、更に「Ｗｒｉｔｅ　ｉｎｔｅｎｓｉｖｅ」であるか否かを判断する（Ｓ１９０６）。ワークロードタイプが「Ｗｒｉｔｅ　ｉｎｔｅｎｓｉｖｅ」である場合（１９０６：Ｙ）、新規割り当て決定処理プログラム６０５は、新規ページに対するライト頻度は多いと予測して、書き込み上限回数が比較的多いＭＬＣ　ＰＧを新規ページに対する実ページの割り当て先のパリティグループに設定する（Ｓ１９０７）。

【０１６９】
これに対し、ワークロードタイプが「Ｒｅａｄ　ｉｎｔｅｎｓｉｖｅ」である場合（Ｓ１９０６：Ｎ）、新規割り当て決定処理プログラム６０５は、新規ページに対するライト頻度は少ないものと予測して、書き込み上限回数が少ないＴＬＣ　ＰＧを新規ページに対する実ページの割り当て先のパリティグループに設定する（Ｓ１９０４）。
【０１７０】
ステップＳ１９０１に戻り、ワークロードタイプが「Ｕｎｋｎｏｗｎ」である場合（Ｓ１９０１：Ｙ）、新規割り当て決定処理プログラム６０５は、ホスト１０１からのＩ／Ｏ特性が分からないため、新規ページに対する将来のライト頻度を予測して、割り当て先を決定する。まずは新規ページの予測ライト頻度を算出する（Ｓ１９０２）。
【０１７１】
例えば新規割り当て決定処理プログラム６０５は、ホスト１０１からの１ページ当たりの平均ライト頻度をモニタ情報として採取して見積もることにより、予測ライト頻度を算出する。
【０１７２】
次いで新規割り当て決定処理プログラム６０５は、プール毎再配置管理テーブル５０４を参照して、予測ライト頻度が新規ライト閾値未満であるか否かを判断する（Ｓ１９０３）。予測ライト頻度が新規ライト閾値未満の場合（Ｓ１９０３：Ｙ）、新規割り当て決定処理プログラム６０５は、書き込み上限回数の少ないＴＬＣ　ＰＧを割り当て先のパリティグループに設定する（Ｓ１９０４）。
【０１７３】
これに対し、予測ライト頻度が新規ライト閾値以上の場合（Ｓ１９０３：Ｎ）、新規割り当て決定処理プログラム６０５は、書き込み上限回数の多いＭＬＣ　ＰＧを割り当て先のパリティグループに設定する（Ｓ１９０７）。
【０１７４】
次いで新規割り当て決定処理プログラム６０５は、割り当て先に設定した特性のパリティグループについて、新規ページの割り当てが可能か否かを判断するため、パリティグループ毎モニタテーブル５０２及びパリティグループ毎再配置管理テーブル５０３を参照して、新規ライト可能量が新規ライトＩ／Ｏカウンタよりも大きいか否かを割り当て先に設定した特性のパリティグループごとに判断する（Ｓ１９０５）。
【０１７５】
新規ライト可能量が新規ライトＩ／Ｏカウンタよりも大きいパリティグループが存在する場合（Ｓ１９０５：Ｙ）、新規割り当て決定処理プログラム６０５は、このパリティグループから新規ページを割り当てるために、ページ毎モニタテーブル５０１を参照して、このパリティグループにおける何れかのページについて新規ライトフラグを設定し（Ｓ１９０９）、本処理を終了する。
【０１７６】
これに対し、新規ライト可能量が新規ライトＩ／Ｏカウンタよりも大きいパリティグループが存在しない場合（Ｓ１９０５：Ｎ）、新規割り当て決定処理プログラム６０５は、寿命を維持するにあたり新規に許容可能なライト頻度を超えているため、ユーザに推奨容量を通知する（Ｓ１９０８）。
【０１７７】
そして新規割り当て決定処理プログラム６０５は、判断対象のパリティグループから新規ページを割り当てるために、何れかのパリティグループにおける何れかのページについて新規ライトフラグを設定し（Ｓ１９０９）、本処理を終了する。
【０１７８】
以上の処理により、ホスト１０１から新規ページにデータを書き込むライト要求を受領した場合、新規ページに対するライト頻度を考慮して、寿命特性に応じたＳＳＤにより構成されたパリティグループから新規ページに対して実ページを割り当てることができる。
【０１７９】

（１－７）推奨容量算出方法の概念構成

図２０は、ホスト１０１からのライト頻度に対して寿命を保証するにあたり推奨される書き込み上限回数の異なる記憶装置の容量比率の算出方法の考え方を示す。図２０は、プール２０４内の各ページ２０７（又は仮想ボリューム２０２内の各ページ２０１）のライト頻度の分布を表す。グラフ２００７は、左からライト頻度が多い順番に全ページ２０７を並べたときの各ページ２０７のライト頻度を示す。縦軸はライト頻度であり、横軸はページ数である。

【０１８０】

ＴＬＣ－ＭＬＣ間ライト閾値２００５は、ＴＬＣ　ＰＧとＭＬＣ　ＰＧのどちらにページを配置するのかを決める閾値であり、ＴＬＣ－ＭＬＣ間ライト閾値２００５とグラフ２００７との交点がＴＬＣとＭＬＣの推奨される容量比率（２００６）となる。ＴＬＣ－ＭＬＣ間ライト閾値２００５は、ユーザが指定してもよいし、ストレージシステム１０４が算出してもよい。

【０１８１】

ストレージシステム１０４が算出する場合、推奨容量比率は、顧客要件のライト頻度（＝Ｗｈｏｓｔ）、プール２０４（又は仮想ボリューム２０２）容量（＝Ｃ）及び各ＳＳＤの寿命を保証するにあたり許容できるライト頻度の相関グラフ２００１、２００２を利用し、下記式９を計算して算出することができる。

【０１８２】

ここでグラフ２００１は、寿命を保証するにあたり単位容量当たりに許容できるライト頻度（＝Ｗｔｌｃ）を傾きとしたＴＬＣ　ＳＳＤの容量と許容できるライト頻度との相関を表す。またグラフ２００２は、寿命を保証するにあたり単位容量当たりに許容できるライト頻度（＝Ｗｍｌｃ）を傾きとしたＭＬＣ　ＳＳＤの容量と許容できるライト頻度との相関を表す。

【０１８３】

【数９】

ＴＬＣ　ＳＳＤ推奨容量

＝（（（Ｗｍｌｃ×Ｃ）－Ｗｈｏｓｔ）÷（Ｗｍｌｃ－Ｗｔｌｃ））÷Ｃ）

ＭＬＣ　ＳＳＤ推奨容量

＝（（Ｃ－ＴＬＣ推奨容量）÷Ｃ）

・・・・・・（９）

【０１８４】

以上により、顧客要件ライト頻度とプール容量とを満たすＴＬＣ　ＳＳＤとＭＬＣ　ＳＳＤとの容量比率を算出することができる。

【０１８５】

（１－８）画面構成

図２１は、プール毎にパラメータを設定する際の画面構成の一例を示す。プール単位のＧＵＩ画面２１０１は、設定対象のプール２０４を特定できるプール番号を表示する領域２１０２と、寿命制御再配置のＯＮ／ＯＦＦを設定する領域２１０３と、寿命制御再配置をＯＮにした場合の詳細設定のＯＮ／ＯＦＦを設定する領域２１０４と、詳細設定の内容を設定する領域２１０５とから構成される。本画面で設定した情報は、プール毎再配置管理テーブル５０４の各欄に格納される。

【０１８６】

寿命制御再配置領域２１０３の設定がＯＦＦの場合、閾値決定処理プログラム６０４が閾値決定処理を実行することはなく、よって再配置処理プログラム６０４Ａがページの再配置を実行することはないが、寿命制御の精度の低下を防ぐため、モニタ情報は寿命制御再配置のＯＮ／ＯＦＦにかかわらず採取される。

【０１８７】

寿命制御再配置領域２１０３の設定がＯＮの場合、上記説明してきた通りページの再配

置が行われる。この場合、詳細設定領域２１０４の項目設定領域２１０５が入力可能に表示される。詳細設定領域２１０４の設定がＯＦＦの場合、項目設定領域２１０５の各種パラメータには、デフォルト値又はストレージシステム１０４内で自動的に算出された値が設定される。

【０１８８】

再配置周期領域２１０６には、寿命制御のための再配置を実行する周期が設定される。この周期はユーザが指定することができる。例えばユーザが「７ｄａｙｓ」と指定すると、７日周期でページの再配置が実行される。

【０１８９】

ワークロードタイプ領域２１０７には、ホスト１０１からのＩ／Ｏ特性が設定される。このＩ／Ｏ特性はユーザが指定することができる。ホスト１０１からのＩ／Ｏ特性を予め把握している場合は、ユーザがＩ／Ｏ特性を指定することにより、新規割り当て先のＳＳＤの特性を明示的に指定することができる。

【０１９０】

具体的には「Ｗｒｉｔｅ　ｉｎｔｅｎｓｉｖｅ」が指定された場合、ホスト１０１のＩ／Ｏ特性は、ライト高負荷であるため、書き込み上限回数の比較的多いＭＬＣ　ＰＧから新規ページに対して実ページの割り当てが行われる。

【０１９１】

また「Ｒｅａｄ　ｉｎｔｅｎｓｉｖｅ」が指定された場合、ホスト１０１のＩ／Ｏ特性は、ライト低負荷であるため、書き込み上限回数の少ないＴＬＣ　ＰＧから新規ページに対して実ページの割り当てが行われる。

【０１９２】

またホスト１０１のＩ／Ｏ特性をユーザが把握していない場合、ユーザは「Ｕｎｋｎｏｗｎ」を指定する。この場合、ストレージシステム１０４が新規ページに対して実ページの割り当て先のパリティグループを自動的に決定することになる。

【０１９３】

同種ドライブ間新規割り当てポリシ領域２１０８には、新規ページに対する実ページの割り当て時にＴＬＣ　ＰＧ又はＭＬＣ　ＰＧのうち、何れの特性のパリティグループからページを割り当てるかを決定するポリシが設定される。

【０１９４】

例えばユーザにより「ラウンドロビン」が指定された場合、各パリティグループから均等にページが割り当てられる。また「容量優先」が指定された場合、容量が少なパリティグループから優先してページが割り当てられる。また「寿命優先」が指定された場合、寿命が長いパリティグループから優先してページが割り当てられる。

【０１９５】

バッファサイズ領域２１０９及び２１１０には、特性がＭＬＣ　ＰＧであるパリティグループの容量に対するバッファの割合が設定される。新規割り当て用領域２１０９には、特性がＭＬＣ　ＰＧであるパリティグループから新規ページを割り当てる際に使用するバッファが設定される。新規割り当て時にＭＬＣ　ＰＧから割り当てられるはずのライト高負荷なページがＭＬＣ　ＰＧの残容量が足りないことによってＴＬＣ　ＰＧから割り当てられることを防ぐ効果がある。

【０１９６】

新規割り当てバッファは、再配置周期毎に確保しなおされる。このため、新規割り当てバッファは、周期内に予想されるホスト１０１から新規ページに対して書き込まれるデータ量に基づいて最適なサイズを見積もることができる。

【０１９７】

再配置バッファ用領域２１１０には、ページの再配置時に使用されるバッファが設定される。再配置バッファにより、再配置時に単位時間当たりに移動できるデータサイズを調整する。このため、再配置バッファを多くとることにより、再配置時のスループットを増やす効果がある。

【０１９８】
図２２は、ストレージシステム１０４がユーザに通知する警告画面の画面構成の一例を示す。警告画面は、ホスト１０１からのライト頻度が多く、現在の構成では目標期間よりもＳＳＤの寿命が短くなること又は寿命を保証するにあたり過剰にＳＳＤを搭載していることをユーザに通知することができる。
【０１９９】
画面を表示する契機は、周期的に実行されるページ再配置処理の完了後にストレージシステム１０４が自動で表示するとしてもよいし、ユーザがストレージシステム１０４に対する操作により、任意のタイミングで表示させてもよい。後者の場合、任意の操作が行われたタイミングでのモニタ情報に基づいて推奨容量が算出される。
【０２００】
プール単位のＧＵＩ画面２２０１は、設定対象のプール２０４を特定できるプール番号を表示する領域２２０２と、警告の内容を通知する領域２２０３から構成される。警告の内容を通知する領域２２０３は、現在のドライブ構成から追加が必要又は削減が可能なＴＬＣ又はＭＬＣの容量を通知する領域２２０４及び現在のホスト１０１からのＩ／Ｏ要求の情報から推奨されるＭＬＣとＴＬＣの容量を通知する領域２２０５から構成される。なお通知する容量の情報は、ＴＬＣとＭＬＣの容量の比率で表してもよい。
【０２０１】
（１－９）第１の実施の形態による効果
以上のように第１の実施の形態におけるストレージシステム１０４によれば、ＴＬＣ－ＭＬＣ間でページを再配置することにより、寿命の異なるＳＳＤから構成されたパリティグループ間でライト頻度を調整することができる。またライトリバランス処理を行うことにより、同一特性のＳＳＤで構成されたパリティグループ間でライト頻度を調整することができる。よって寿命劣化の激しいＳＳＤに対するライト頻度のページを寿命劣化の緩やかなＳＳＤに移動して、ＳＳＤの保守交換回数を削減することができる。またストレージシステム１０４のコストを削減することができる。
【０２０２】
（２）第２の実施の形態
第２の実施の形態は、記憶装置として半導体メモリ（ＳＳＤ）だけでなく、ハードディスクドライブ（ＨＤＤ）を搭載するストレージシステムにおいて、記憶装置（ＳＳＤ及びＨＤＤ）を性能に応じた階層に分類し、ホストからのアクセス頻度に応じて適切な階層の記憶装置にデータを配置する階層制御を行う点で、第１の実施の形態と異なる。
【０２０３】
（２－１）計算機システムの全体構成
図２３は、第２の実施の形態における計算機システム１Ａの全体構成を示す。計算機システム１Ａは、ＳＡＳ（Serial Attached SCSI）規格のＨＤＤ（ドライブ２３０１）を搭載している点及びＳＳＤ（ドライブ１１１及び１１２）をティア１に設定し、ＳＡＳ規格のＨＤＤ（ドライブ２３０１）をティア２に設定するとともに、各ページのＩ／Ｏ頻度に応じてティア間でデータを再配置する点で、第１の実施の形態と異なる。
【０２０４】
例えばティア１からはＩ／Ｏ頻度が１００［ＩＯＰＳ］のページを割り当てるように設定されており、ティア２からはＩ／Ｏ頻度が１０［ＩＯＰＳ］のページを割り当てるように設定されているとする。一方でＩ／Ｏ頻度が５０［ＩＯＰＳ］のページがティア２から割り当てられており、Ｉ／Ｏ頻度が２０［ＩＯＰＳ］のページがティア１から割り当てられているとする。
【０２０５】
この場合、ストレージシステム１０４Ａは全体として１０（ティア２の上限ＩＯＰＨ）＋２０＝３０［ＩＯＰＳ］の性能しか発揮することができない。そこでＩ／Ｏ頻度が５０［ＩＯＰＨ］のページをティア２からティア１に移動（再配置）すると、ストレージシステム１０４Ａは全体として５０＋２０＝７０［ＩＯＰＳ］の性能を発揮することができる

ようになる。

【０２０６】

（２－２）ストレージシステムの論理構成

図２４は、ストレージシステム１０４Ａの論理構成を示す。ストレージシステム１０４Ａは、プールボリューム２０６Ａ～２０６ＣがＭＬＣ　ＳＳＤ又はＴＬＣ　ＳＳＤから構成され、プールボリューム２０６Ｄ及び２０６ＥがＳＡＳ　ＨＤＤから構成されており、各プールボリューム２０６Ａ～２０６Ｅがティア１又は２に分類されている点で、第１の実施の形態と異なる。

【０２０７】

そして第２の実施の形態のストレージシステム１０４Ａは、第１の実施の形態において説明してきたページの再配置に加えて、ＳＳＤから構成されるティア１においてホスト１０１からのライト負荷が高くなり、ＳＳＤの寿命を保証することができなくなった場合、ライト負荷の高い例えばページ２０７Ｄをティア２に移動するライトデモーション処理を実行する。これによりＳＳＤの寿命劣化を防ぐことができる。

【０２０８】

（２－３）ページ配置処理の概念構成

図２５は、ライトデモーション処理によるページ再配置処理の概念構成を示す。ここではＭＬＣ　ＰＧ２０５Ａ又はＴＬＣ　ＰＧ２０５Ｂから構成されるティア１と、ＳＡＳ　ＨＤＤで構成されたパリティグループ（以下、ＳＡＳ　ＰＧと呼ぶ）２０５Ｃから構成されるティア２との間で、ページ３０１を再配置する。ライトデモーション処理は、閾値決定処理時にティア１内でのページ再配置だけではＳＳＤの寿命を維持することができないと判断された場合に実行される。

【０２０９】

具体的には、閾値決定処理においてライトリバランスによる移動計画ページ数を算出した後、ライト削減ＰＧのライト削減要求量を満たすだけのページを移動することができなかった場合、移動することができなかった分のライト頻度に基づいて、ライトデモーション処理による移動計画ページ数を算出する。そしてライトリバランス実行時に、ライトデモーション処理による移動計画ページ数分だけＭＬＣ　ＰＧ２０５Ａ又はＴＬＣ　ＰＧ２０５ＢからＳＡＳ　ＰＧ２０５Ｃにライトデモーション閾値２５０１以上のライト頻度のページを移動する。

【０２１０】

（２－４）テーブル構成

図２６は、第２の実施の形態におけるパリティグループ毎再配置管理テーブル５０３Ａの論理構成を示す。パリティグループ毎再配置管理テーブル５０３Ａは、パリティグループ毎再配置管理テーブル５０３（図９）の各欄に加えて、ライトデモーション閾値欄２６０１、ライトデモーション用移動計画ページ数欄２６０２、ライトデモーション用移動実績ページ数欄２６０３及びティアレベル欄２６０４から構成される。

【０２１１】

ライトデモーション閾値欄２６０１には、ライトデモーション処理対象のページを決定するための閾値が格納され、ライトデモーション用移動計画ページ数欄２６０２には、ライトデモーション処理により移動するページ数が格納される。またライトデモーション用移動実績ページ数欄２６０３には、ライトデモーション処理により移動したページ数が格納される。ティアレベル欄２６０４には、パリティグループが所属する階層順序（例えばティア１、ティア２、ティア３）が格納される。なおここでは階層順序の値が小さいティアほど高性能のドライブで構成されている。

【０２１２】

図２７は、第２の実施の形態におけるプール毎再配置管理テーブル５０４Ａの論理構成を示す。プール毎再配置管理テーブル５０４Ａは、プール毎再配置管理テーブル５０４（図１０）の各欄に加えて、ティア間Ｉ／Ｏ閾値欄２７０１から構成される。ティア間Ｉ／Ｏ閾値欄２７０１には、ページをどの階層に配置するのかを決定するための閾値が格納さ

れる。
【０２１３】
（２－５）フローチャート
図２８は、第２の実施の形態における再配置処理のフローチャートを示す。この再配置処理は、プロセッサ１１４と閾値決定処理プログラム６０４により呼び出される再配置処理プログラム６０４Ａとの協働により実行される。説明の便宜上、処理主体を再配置処理プログラム６０４Ａとして説明する。
【０２１４】
なお前提として閾値決定処理プログラム６０４は、閾値決定処理（図１４）においてティア間Ｉ／Ｏ閾値及びライトデモーション用移動計画ページ数を算出し、各種閾値をパリティグループ毎再配置管理テーブル５０３Ａ及びプール毎再配置管理テーブル５０４Ａにそれぞれ格納しているものとする。
【０２１５】
まず再配置処理プログラム６０４Ａは、ティア間再配置処理を実行する（Ｓ２８０１）。ここでは再配置処理プログラム６０４Ａは、プール毎再配置管理テーブル５０４Ａのティア間Ｉ／Ｏ閾値欄２７０１を参照して、対象ページを配置するティアを決定する。
【０２１６】
その後再配置処理プログラム６０４Ａは、第１の実施の形態における再配置処理（図１５）と同様にＴＬＣ－ＭＬＣ間再配置処理を実行し（Ｓ１５０１）、次いでライトリバランス処理を実行する（Ｓ１５０２）。ライトリバランス処理を実行した後、再配置処理プログラム６０４Ａは、パリティグループ毎再配置管理テーブル５０３Ａを参照して、ライトデモーション用移動計画ページ数が０であるか否かを判断する（Ｓ２８０２）。
【０２１７】
ライトデモーション用移動計画ページ数が０でない場合（Ｓ２８０２：Ｎ）、再配置処理プログラム６０４Ａは、ライトリバランス処理を実行しただけではライト頻度を調整しきれなかったため、ライトデモーション処理を実行する（Ｓ２８０３）。その後再配置処理プログラム６０４Ａは、性能リバランスを実行して（Ｓ１５０４）、本処理を終了する。
【０２１８】
これに対し、ライトデモーション用移動計画ページ数が０である場合（Ｓ２８０２：Ｙ）、再配置処理プログラム６０４Ａは、ライトデモーション処理を実行する必要はないため、性能リバランスを実行して（Ｓ１５０４）、本処理を終了する。以上の処理により、ホスト１０１からのライト頻度が過多であるためティア１を構成するＳＳＤの寿命を維持することができない場合、ティア２にライト負荷が高いページを移動して、ティア１内のＳＳＤの寿命を維持することができる。
【０２１９】
図２９は、ティア間再配置処理のフローチャートを示す。このティア間再配置処理は、プロセッサ１１４と再配置処理プログラム６０４Ａとの協働により実行される。説明の便宜上、処理主体を再配置処理プログラム６０４Ａとして説明する。
【０２２０】
まず再配置処理プログラム６０４Ａは、閾値決定処理プログラム６０４により算出されたティア間Ｉ／Ｏ閾値に基づいて、全てのパリティグループ内の各ページをティア１又は２に再配置する。このとき再配置について未判定のページがあるか否かを判断する（Ｓ２９０１）。
【０２２１】
未判定のページがある場合（Ｓ２９０１：Ｙ）、再配置処理プログラム６０４Ａは、パリティグループ毎再配置管理テーブル５０３Ａを参照して、この未判定の対象ページが所属するティアがティア２であるか否かを判断する（Ｓ２９０２）。
【０２２２】
対象ページが所属するティアがティア２である場合（Ｓ２９０２：Ｙ）、再配置処理プ

ログラム６０４Ａは、プール毎再配置管理テーブル５０４Ａを参照して、対象ページのＩ／Ｏ頻度がティア間Ｉ／Ｏ閾値以上であるか否かを判断する（Ｓ２９０３）。

【０２２３】

対象ページのＩ／Ｏ頻度がティア間Ｉ／Ｏ閾値以上である場合（Ｓ２９０３：Ｙ）、再配置処理プログラム６０４Ａは、対象ページをティア２からティア１に移動する（Ｓ２９０４）。これに対し、対象ページのＩ／Ｏ頻度がティア間Ｉ／Ｏ頻度未満である場合（Ｓ２９０３：Ｎ）、再配置処理プログラム６０４Ａは、何もせずにステップＳ２９０１に移行する。

【０２２４】

ステップＳ２９０２に戻り、対象ページが所属するティアがティア１の場合（Ｓ２９０２：Ｎ）、再配置処理プログラム６０４Ａは、プール毎再配置管理テーブル５０４Ａを参照して、対象ページのＩ／Ｏ頻度がティア間Ｉ／Ｏ閾値未満であるか否かを判断する（Ｓ２９０５）。

【０２２５】

対象ページのＩ／Ｏ頻度がティア間Ｉ／Ｏ閾値未満である場合（Ｓ２９０５：Ｙ）、再配置処理プログラム６０４Ａは、対象ページをティア１からティア２に移動する（Ｓ２９０６）。これに対し、対象ページのＩ／Ｏ頻度がティア間Ｉ／Ｏ頻度以上である場合（Ｓ２９０５：Ｎ）、再配置処理プログラム６０４Ａは、何もせずにステップＳ２９０１に移行する。

【０２２６】

再配置処理プログラム６０４Ａは、全てのパリティグループ内の各ページについて判定を終えると、本処理を終了する。以上の処理により、ホスト１０１からのＩ／Ｏ頻度及び各ティアの性能に応じて、各ティアにページを再配置することができる。

【０２２７】

図３０は、ライトデモーション処理のフローチャートを示す。このライトデモーション処理は、プロセッサ１１４と再配置処理プログラム６０４Ａとの協働により実行される。説明の便宜上、処理主体を再配置処理プログラム６０４Ａとして説明する。

【０２２８】

まず再配置処理プログラム６０４Ａは、閾値決定処理プログラム６０４により算出されたライトデモーション用移動計画ページ数、ライト削減閾値及びライトデモーション閾値に基づいて、異なるティア間でページを再配置する。このとき再配置について未判定のページがあるか否かを判断する（Ｓ３００１）。

【０２２９】

未判定のページがない場合（Ｓ３００１：Ｎ）、再配置処理プログラム６０４Ａは本処理を終了する。これに対し未判定のページがある場合（Ｓ３００１：Ｙ）、再配置処理プログラム６０４Ａは、パリティグループ毎再配置管理テーブル５０３Ａを参照して、この未判定の対象ページが所属するパリティグループの移動元ＰＧ種別がライト削減ＰＧであるか否かを判断する（Ｓ３００２）。

【０２３０】

対象ページが所属するパリティグループの移動元ＰＧ種別がライト削減ＰＧでない場合（Ｓ３００２：Ｎ）、すなわちライト追加ＰＧである場合、再配置処理プログラム６０４Ａは対象ページに対してライトデモーション処理を実行する必要はないと判断して、ステップＳ３００１に移行する。

【０２３１】

これに対し、対象ページが所属するパリティグループの移動元ＰＧ種別がライト削減ＰＧである場合（Ｓ３００２：Ｙ）、再配置処理プログラム６０４Ａは、ページ毎モニタテーブル５０１のライトＩ／Ｏカウンタ欄７０２を参照して、この対象ページのライト頻度を取得する。

【０２３２】

そして再配置処理プログラム６０４Ａは、パリティグループ毎再配置管理テーブル５０

３Ａのライト削減閾値欄９０５及びライトデモーション閾値欄２６０１を参照して、ライト削減閾値及びライトデモーション閾値を取得する。

【０２３３】

そして取得した対象ページのライト頻度がライトデモーション閾値以上であり、かつ、ライト削減閾値未満であるか否かを判断する（Ｓ３００３）。対象ページのライト頻度がライトデモーション閾値以上であり、かつ、ライト削減閾値未満でない場合（Ｓ３００３：Ｎ）、再配置処理プログラム６０４Ａはステップ３００１に移行する。

【０２３４】

これに対し、対象ページのライト頻度がライトデモーション閾値以上であり、かつ、ライト削減閾値未満である場合（Ｓ３００３：Ｙ）、再配置処理プログラム６０４Ａは、対象ページが所属するパリティグループのライトデモーション用移動実績ページ数がライトデモーション用移動計画ページ数未満であるか否かを判断する（Ｓ３００４）。

【０２３５】

対象ページが所属するパリティグループのライトデモーション用移動実績ページ数がライトデモーション用移動計画ページ数未満でない場合（Ｓ３００４：Ｎ）、再配置処理プログラム６０４Ａは、ステップＳ３００１に移行する。

【０２３６】

これに対し、対象ページが所属するパリティグループのライトデモーション用移動実績ページ数がライトデモーション用移動計画ページ数未満である場合（Ｓ３００４：Ｙ）、再配置処理プログラム６０４Ａは、対象ページをティア１からティア２に移動する（Ｓ３００５）。再配置処理プログラム６０４Ａは、未判定の全てのページについて判定を終えると、本処理を終了する。

【０２３７】

図３１は、新規割り当て決定処理のフローチャートを示す。この新規割り当て決定処理は、ホスト１０１から新規ページに対するライト要求を受領したことを契機として、プロセッサ１１４と新規割り当て決定処理プログラム６０５との協働により実行される。説明の便宜上、処理主体を新規割り当て決定処理プログラム６０５として説明する。

【０２３８】

ステップＳ１９０１～Ｓ１９０５までは、第１の実施の形態における新規割り当て決定処理（図１９）と同様であるため、ここでの説明は省略する。ステップＳ１９０５において、新規割り当て決定処理プログラム６０５は、割り当て先に設定した特性のティア１のパリティグループについて、新規ライト可能量が新規ライトＩ／Ｏカウンタよりも大きいか否かを判断する（Ｓ１９０５）。

【０２３９】

新規ライト可能量が新規ライトＩ／Ｏカウンタよりも大きいパリティグループが存在しない場合（Ｓ１９０５：Ｎ）、再配置処理プログラム６０４Ａは、ティア１からこれ以上ページを割り当てると、ＳＳＤの寿命を維持することができなくなると判断して、新規ページに対する実ページの割り当て先をティア２に設定する（Ｓ３１０１）。

【０２４０】

ステップＳ１９０９は、第１の実施の形態における新規割り当て決定処理（図１９）と同様であるため説明は省略する。以上の処理により、寿命維持のために必要な分のライト頻度をティア１から削減することができる。なお新規割り当て時にホスト１０１からのライト要求がシーケンシャルライトである場合、新規割り当て決定処理プログラムを実行せずに、ティア２からページを割り当てるようにしてもよい。

【０２４１】

（２－６）第２の実施の形態による効果

以上のように第２の実施の形態におけるストレージシステム１０４Ａによれば、ホスト１０１からティア１に対するライト頻度をティア２に移動することにより、一定期間当たりに許容できるライト頻度を超えたＳＳＤの寿命劣化を緩やかにすることができる。よってＳＳＤの保守交換回数を削減でき、ストレージシステム１０４Ａのコストを削減するこ

とができる。
【０２４２】
（３）第３の実施の形態
第３の実施の形態では、第１の実施の形態において説明したライト追加可能量及びライト削減要求量（図８）を算出する手法について説明する。第１の実施の形態においては、ＳＳＤからライト追加可能率及びライト削減要求率を採取し、これらを用いて上記式１及び２を計算することによりライト追加可能量及びライト削減要求量を算出するとしたが、ライト追加可能率及びライト削減要求率をＳＳＤから直接採取することができない場合、以下の手法を用いてライト追加可能量及びライト削減要求量を算出することができる。
【０２４３】
図３２は、ＳＳＤの寿命情報として磨耗指標（Wear out Indicator）を取得し、磨耗指標からライト追加可能量及びライト削減要求量を算出するための考え方を示す。グラフの縦軸は、ＳＳＤの磨耗指標（寿命率）３２０１であり、横軸は、ＳＳＤを使用しはじめてからの経過年数（経過時間率）３２０２である。
【０２４４】
寿命率３２０１は、ＳＳＤの消去回数から算出される磨耗指標を意味しており、一般的にＳ．Ｍ．Ａ．Ｒ．Ｔ．情報として取得できることが知られている。磨耗指標は、値が１００（＝L_e）に達したとき、ＳＳＤの寿命を意味し、保守交換が必要となる。経過時間率は、目標寿命期間（例えば、３年や５年）を１００％（＝T_e）とした経過時間の割合を意味する。
【０２４５】
直線３２１５は、区間Δｔ１（＝T_1-T_0）の寿命率の変化ΔＬ１（＝L_1-L_0）を示しており、直線の傾きは、区間Δｔ１のライト頻度（＝W_0）を表している。直線３２１５のライト頻度が継続すると、目標寿命期間を達成する前にＳＳＤの寿命が尽きる。
【０２４６】
このため、直線３２１６で示すライト頻度（＝W_1）までライト頻度を削減し、区間Δｔ２（＝T_e-T_1）の間の寿命率の増加量がΔＬ（＝L_e-L_1）となるように寿命劣化の速度を調整する。また一般的にＳＳＤに対するライト頻度は、ＷＡ（Write Amplification）により、ホスト１０１がＳＳＤに対して発行したライト要求の回数よりもＳＳＤ内部のフラッシュチップに対するライト回数の方が大きくなることが知られている。ＳＳＤのＷＡ値は、（フラッシュチップに対するライト回数÷ホスト１０１からのライト要求の回数）により算出される。
【０２４７】
以上の数値を用いてストレージシステム１０４は、寿命を維持するために必要なライト削減要求量を算出する。ライト削減要求量は、下記式１０を計算して算出することができる。
【０２４８】
【数１０】

$$\text{ライト削減要求量［ＩＯＰＨ］} = W_0 \times WA \times \left(1 - \left(\left(\Delta t1 \div \Delta L1\right) \times \left(\Delta L2 \div \Delta t2\right)\right)\right) \quad \cdots\cdots (10)$$

【０２４９】
またライト頻度が目標寿命期間を達成するにあたり余裕がある場合、ストレージシステム１０４は、ライト追加可能量を算出する。ライト追加可能量は、下記式１１を計算して算出することができる。
【０２５０】

【数１１】

ライト追加可能量［ＩＯＰＨ］
＝W_0×ＷＡ×（（（Δｔ１÷ΔＬ１）×（ΔＬ２÷Δｔ２））－１）
・・・・・・（１１）

【０２５１】

以上により、一般的に取得可能なＳＳＤの寿命情報からＳＳＤ　ＰＧに対するライト追加可能量及びライト削減要求量（図８）を算出することができる。そして算出したライト追加可能量及びライト削減要求量に基づいてページを再配置することにより、ストレージシステム１０４のコストを削減することができる。

【０２５２】

なお上記説明してきた実施の形態においては、「書き込み」上限回数の少ないＳＳＤのライト回数を削減するようにページを再配置することで、書き込み上限回数の少ないＳＳＤの寿命を延ばす構成について説明してきたが、「書き換え」上限回数の少ないＳＳＤについても上記構成を採用することにより、同様に寿命を延ばすことができる。ＳＳＤにおける書き換えとは、複数ページから構成されるブロックを一旦消去し、その後ブロック内の各ページにデータを書き込む一連の処理をいう。よって書き換えを１回行うと、消去と書き込みとを両方行うことになる。すなわち書き換え上限回数の少ないＳＳＤと、書き込み上限回数の少ないＳＳＤとを同様に取り扱うことで、書き換え上限回数の少ないＳＳＤの寿命を延ばすことができる。

【０２５３】

また上記説明してきた実施の形態においては、書き込み上限回数の少ないＳＳＤの「ライト頻度」（ライト回数）を削減するようにページを再配置することで、書き込み上限回数の少ないＳＳＤの寿命を延ばす構成について説明してきたが、ライトする「データ量」を削減するようにページを再配置することでも、同様に書き込み上限回数の少ないＳＳＤの寿命を延ばすことができる。ライトするデータ量が大きい場合には複数ページにデータをライトする必要があり、よってライト回数も増加するためである。

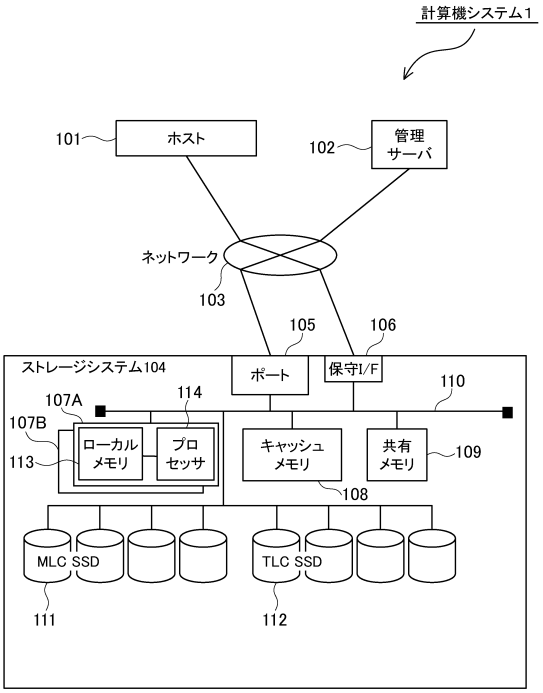
【符号の説明】

【０２５４】

１、１Ａ	計算機システム
１０４、１０４Ａ	ストレージシステム
１１４	プロセッサ
１１１	ＭＬＣ　ＳＳＤ
１１２	ＴＬＣ　ＳＳＤ
２３０１	ＳＡＳ　ＨＤＤ
２０２	仮想ボリューム
２０１、３０１、４０１、２０７	ページ
２０６	プールボリューム
２０５	パリティグループ
３０２	ＴＬＣ－ＭＬＣ間ライト閾値

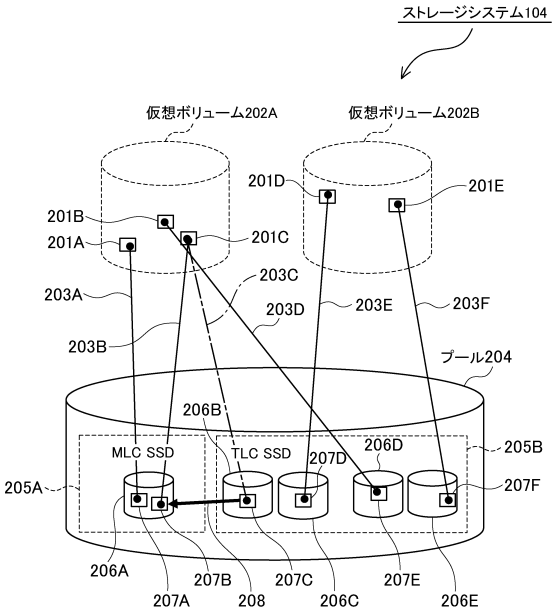
【図１】

図 1



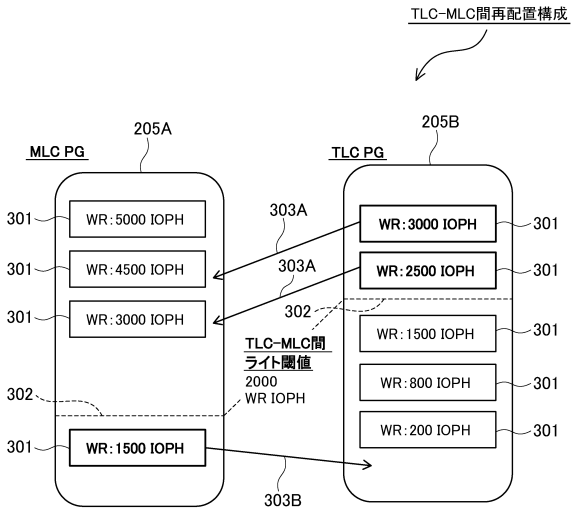
【図２】

図 2



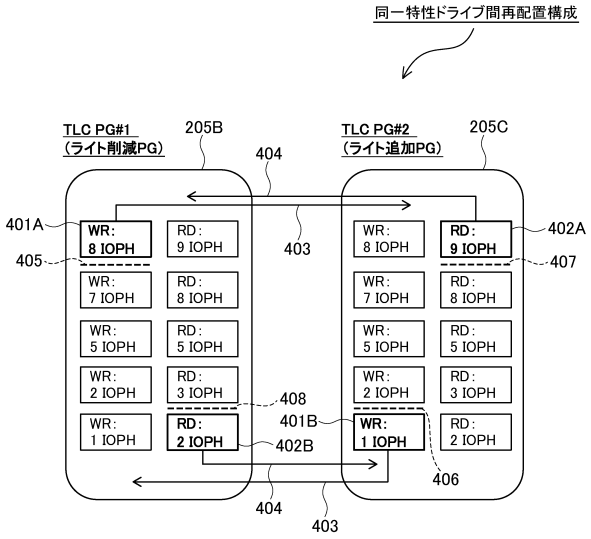
【図３】

図 3



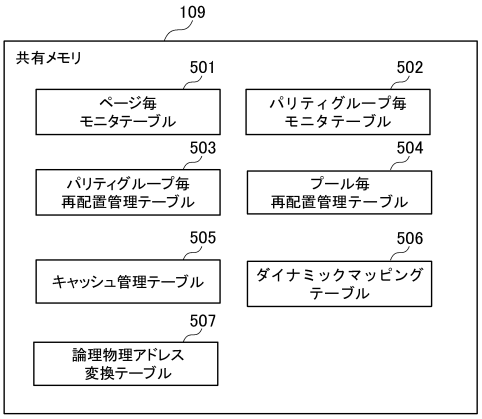
【図４】

図 4



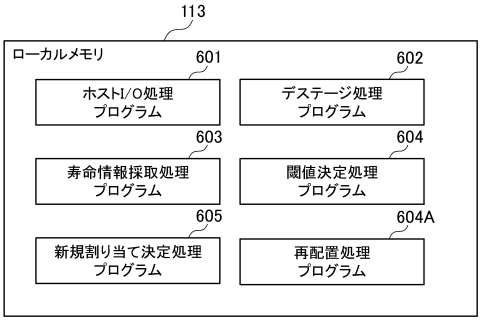
【図５】

図 5



【図６】

図 6



【図７】

図 7

ページ毎モニタテーブル501

701	ページ番号	0	1	2	3
702	ライト I/O カウンタ	2000	12100	156	56
703	リード I/O カウンタ	5010	6520	4600	300
704	合計 I/O カウンタ	7010	18620	4756	356
705	新規ライトフラグ	0	1	1	0

【図８】

図 8

パリティグループ毎モニタテーブル502

801	パリティグループ番号	0	1	2
802	最大ライト頻度	22187	1962	4703
803	最小ライト頻度	5120	453	1085
804	最大リード頻度	4118	8018	17862
805	最小リード頻度	950	1850	4122
806	リード/ライト比率	16/84	80/20	79/21
807	ライト追加可能量	0	1790	2170
808	ライト削減要求量	2447	0	0
809	新規ライト I/O カウンタ	896	202	69
810	新規ライト比率	16%	40%	6%
811	平均 I/O頻度	8800	12000	9000
812	割り当てページ数	458	128	386

【図９】

図 9

パリティグループ毎再配置管理テーブル503

901	パリティグループ番号	0	1	2
902	メディアタイプ	TLC	TLC	MLC
903	移動元PG種別	ライト削減PG	ライト追加PG	ライト追加PG
904	移動先PG	1	0	3
905	ライト削減閾値	15531	Invalid	Invalid
906	ライト追加閾値	Invalid	589	1411
907	リード削減閾値	Invalid	5613	12503
908	リード追加閾値	1236	Invalid	Invalid
909A	移動計画ページ数（ライトリバランス）	156	70	1486
910A	移動実績ページ数（ライトリバランス）	154	56	562
909B	移動計画ページ数（性能リバランス）	84	134	896
910B	移動実績ページ数（性能リバランス）	0	0	232
911	新規ライト可能量	846	1026	2394

【図１０】

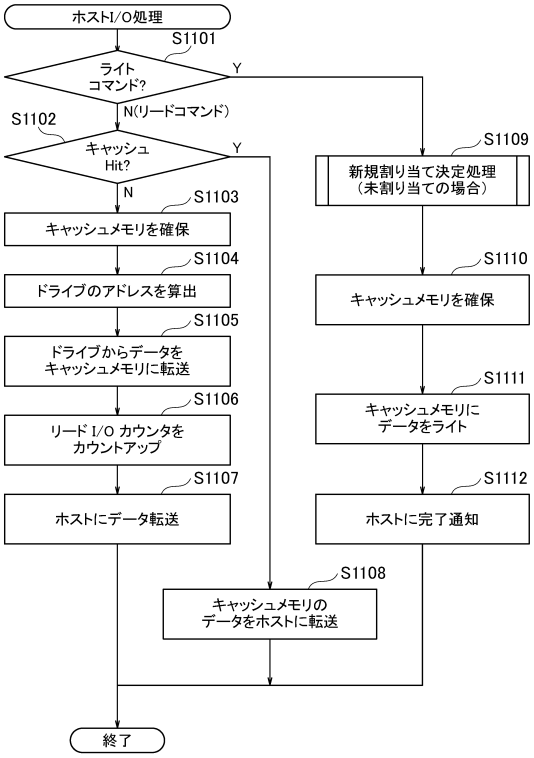
図 10

プール毎再配置管理テーブル504

1001	プール番号	0	1	2
1002	寿命制御再配置周期 [hour]	168	24	72
1003	TLC-MLC間ライト閾値	7040	9541	4530
1004	新規ライト閾値	7550	8205	4260
1005	ワークロードタイプ	Write intensive	Unknown	Read intensive
1006	同種ドライブ間新規割り当てポリシ	寿命優先	容量優先	ラウンドロビン
1007	新規割り当てバッファ	3	3	5
1008	再配置バッファ	2	2	3

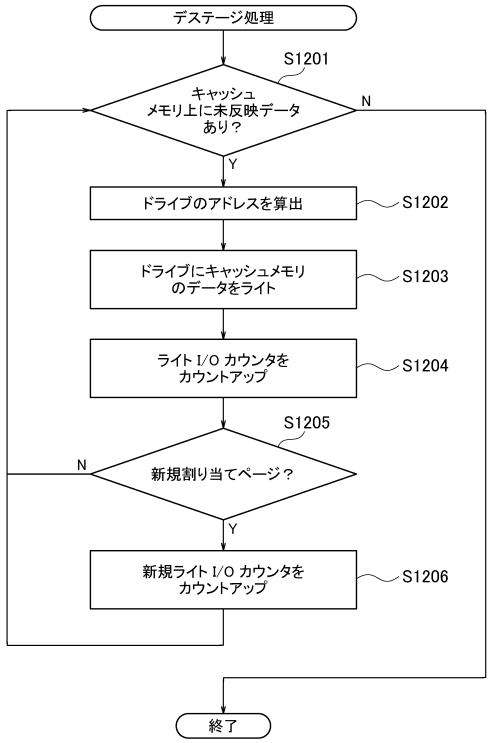
【図１１】

図 11



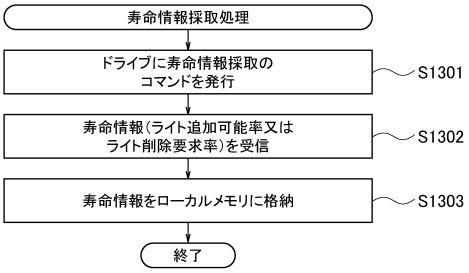
【図１２】

図 12



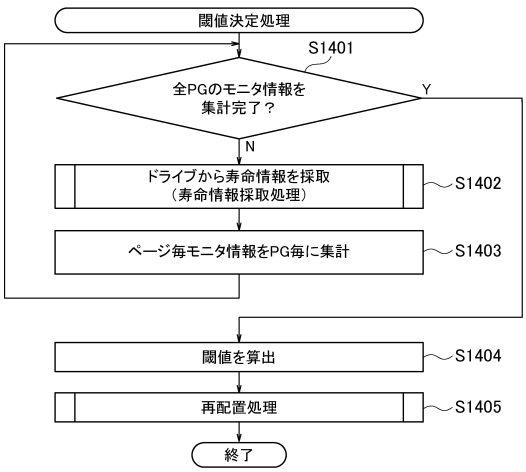
【図１３】

図 13



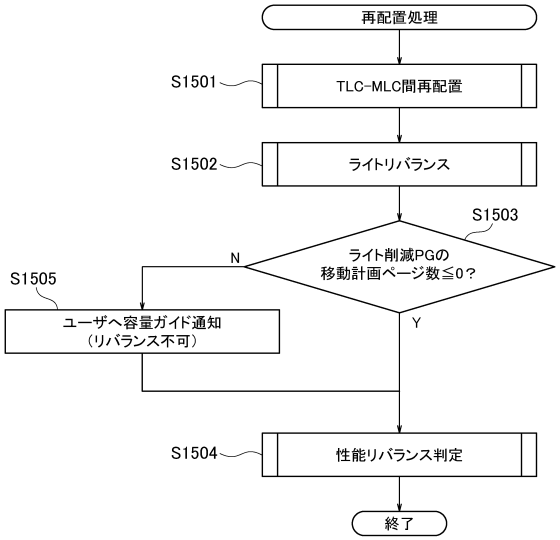
【図１４】

図 14



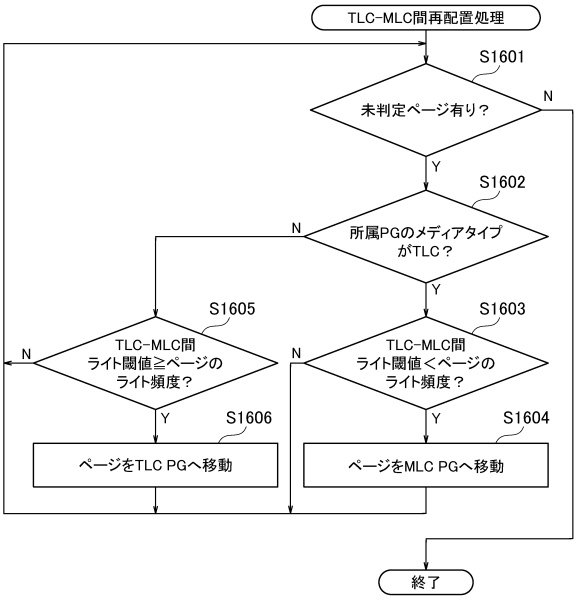
【図１５】

図 15



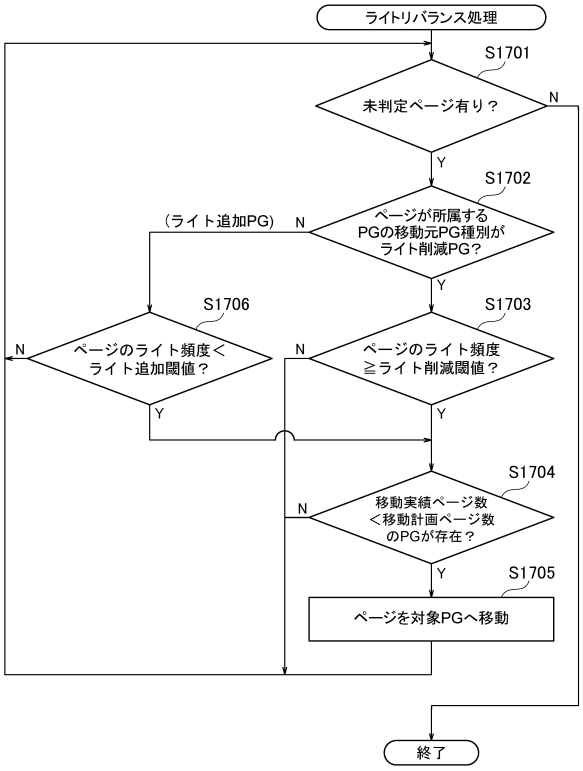
【図１６】

図 16



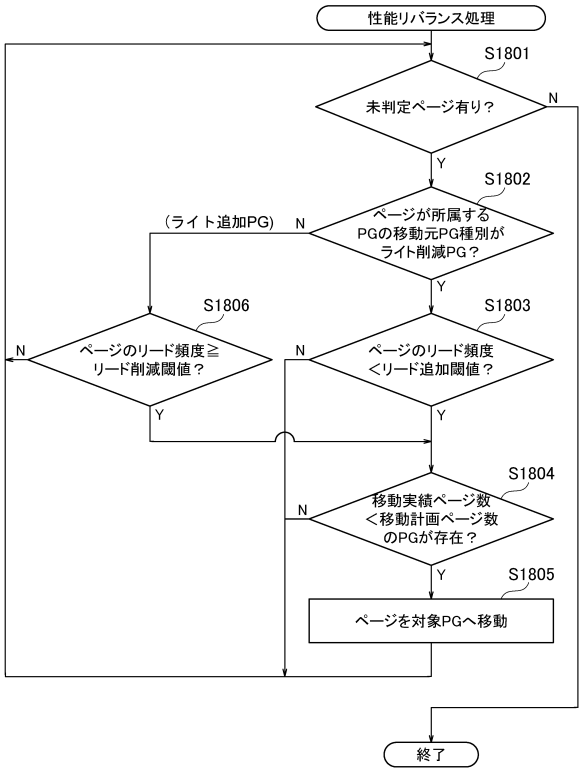
【図１７】

図 17



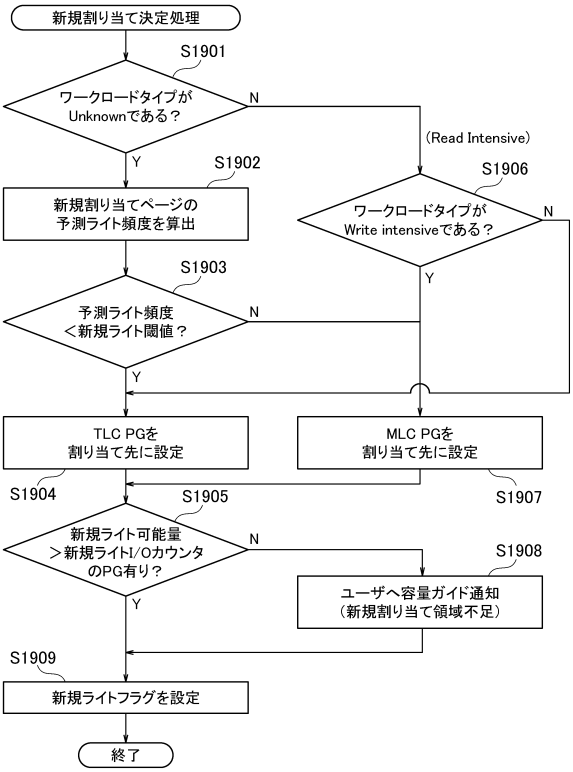
【図１８】

図 18



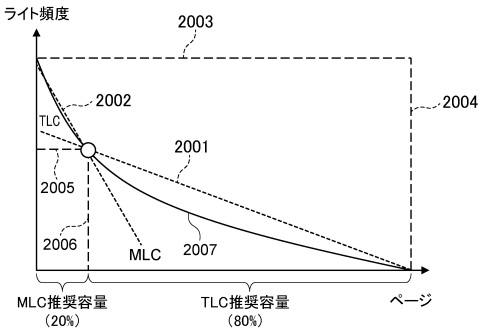
【図１９】

図 19



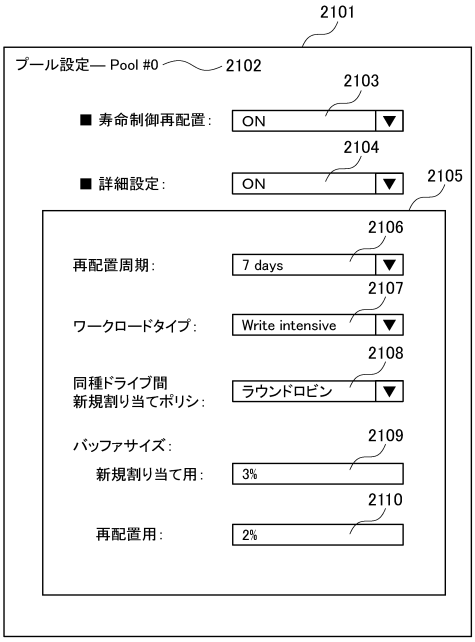
【図２０】

図 20



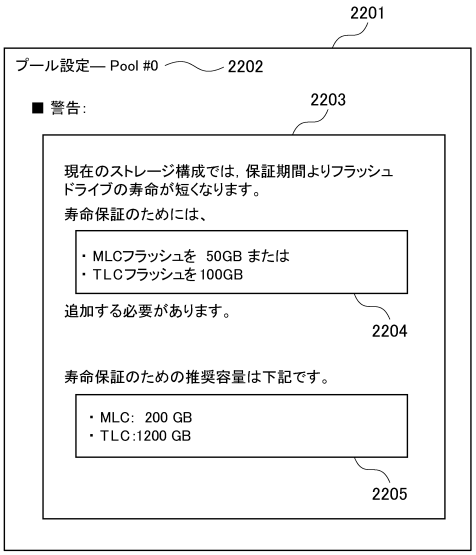
【図２１】

図 21



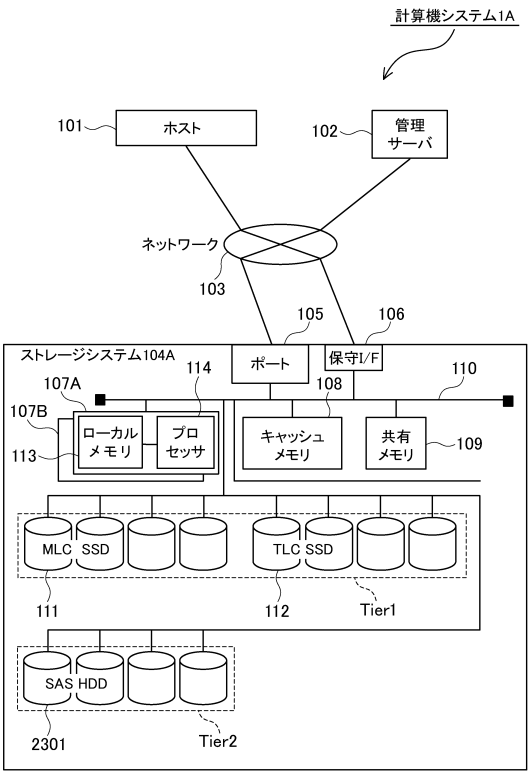
【図２２】

図 22



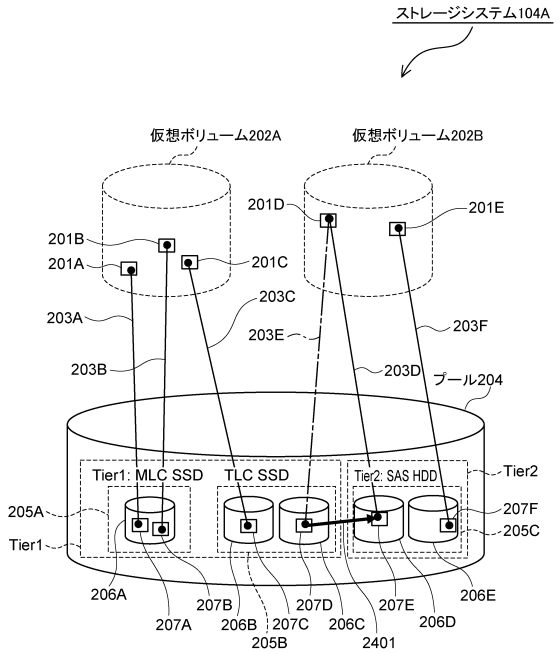
【図２３】

図 23



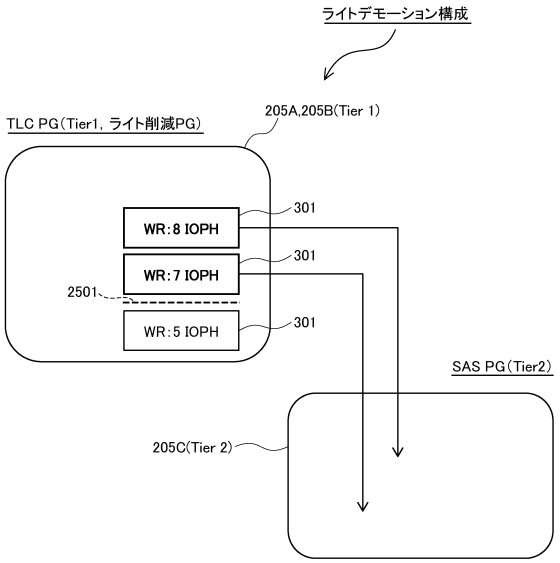
【図２４】

図 24



【図２５】

図 25



【図２６】

図 26

パリティグループ毎再配置管理テーブル503A

901	パリティグループ番号	0	1	2
902	メディアタイプ	TLC	SAS	MLC
903	PGタイプ	ライト削減PG	ライト追加PG	ライト追加PG
904	移動先PG	1	–	3
905	ライト削減閾値	15531	–	–
906	ライト追加閾値	–		
907	リード削減閾値	–		
908	リード追加閾値	1236	–	–
909	移動計画ページ数（ライトリバランス）			
910	移動実績ページ数（ライトリバランス）			
911	移動計画ページ数（性能リバランス）			
912	移動実績ページ数（性能リバランス）			
913	新規ライト可能量			
2601	ライトデモーション閾値			
2602	移動計画ページ数（ライトデモーション）			
2603	移動実績ページ数（ライトデモーション）			
2604	Tier Level	1	2	1

【図２７】

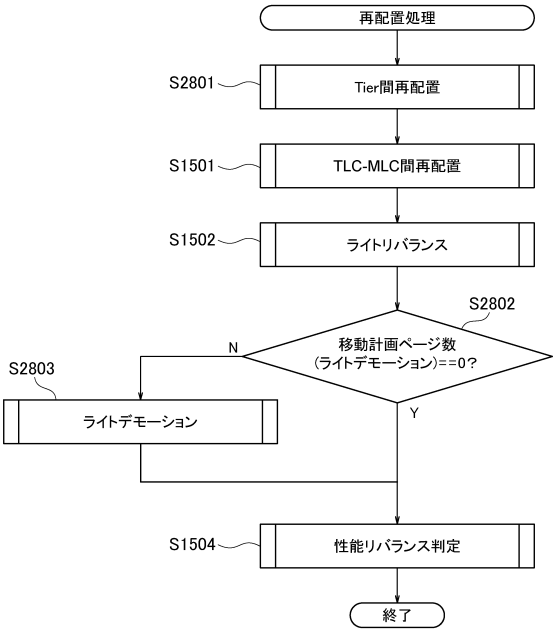
図 27

プール毎再配置管理テーブル504A

1001	プール番号	0	1	2
1002	寿命制御再配置周期[hour]	168	24	72
1003	TLC-MLC間ライト閾値	7040	9541	4530
1004	新規ライト閾値	7550	8205	4260
1005	ワークロードタイプ	Write intensive	Unknown	Read intensive
1006	同種ドライブ間新規割り当てポリシ	寿命優先	容量優先	ラウンドロビン
1007	新規割り当てバッファ	3	3	5
1008	再配置バッファ	2	2	3
2701	Tier間I/O閾値	5680	7520	2602

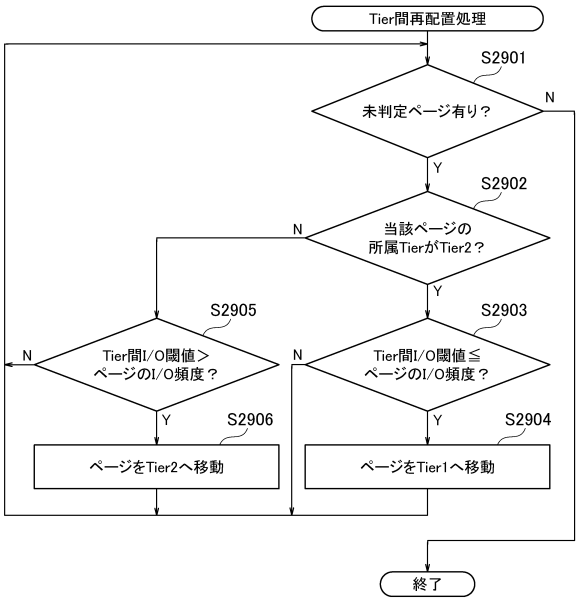
【図２８】

図 28



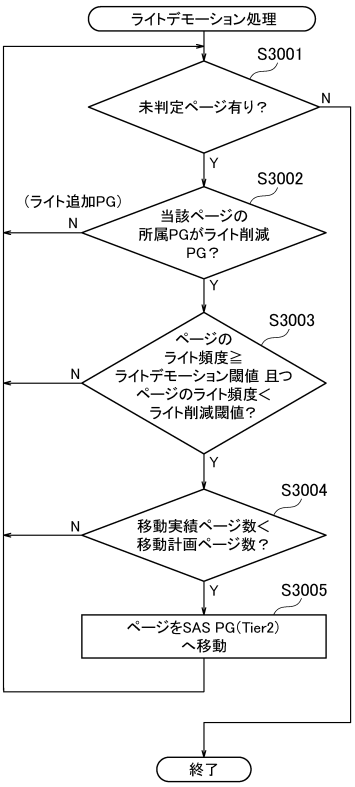
【図２９】

図 29



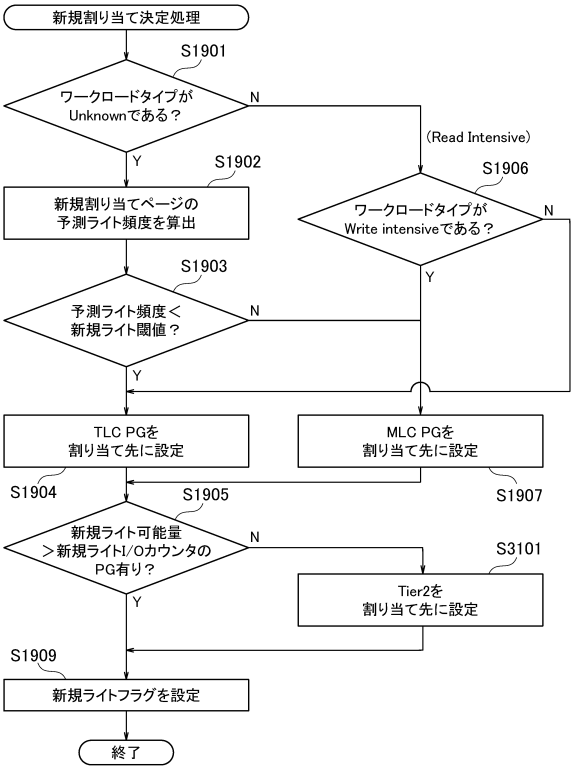
【図３０】

図 30



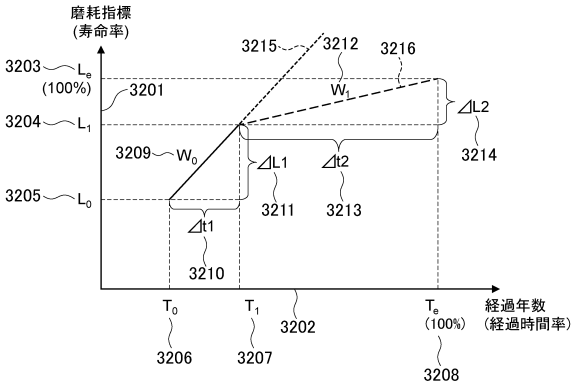
【図３１】

図 31



【図３２】

図 32


磨耗指標
(寿命率)
3203
L_e
(100%)
3201
3204
L_1
3209
W_0
3205
L_0
3215
3212
3216
W_1
⊿L2
3214
⊿L1
⊿t2
3211
3213
⊿t1
3210
T_0
3206
T_1
3207
3202
T_e
(100%)
3208
経過年数
(経過時間率)

フロントページの続き

審査官　田名網　忠雄

(56)参考文献　特表２０１５－５０５０７８（ＪＰ，Ａ）
国際公開第２０１２／１６４７１４（WO，Ａ１）
特開２０１０－１０８２４６（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　　３／０６－３／０８
Ｇ０６Ｆ　１２／００－１２／０６
Ｇ０６Ｆ　１６／００－１６／９５８