

(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2007-310861  
(P2007-310861A)

(43) 公開日 平成19年11月29日(2007.11.29)

(51) Int. Cl. F I テーマコード (参考)  
**G06F 3/06 (2006.01)** G06F 3/06 301J 5B065  
 G06F 3/06 304F

審査請求 未請求 請求項の数 16 O L (全 27 頁)

<p>(21) 出願番号 特願2006-240031 (P2006-240031)                  (22) 出願日 平成18年9月5日 (2006.9.5)                  (31) 優先権主張番号 特願2006-114933 (P2006-114933)                  (32) 優先日 平成18年4月18日 (2006.4.18)                  (33) 優先権主張国 日本国 (JP)</p>	<p>(71) 出願人 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 100079108                  弁理士 稲葉 良幸                  (74) 代理人 100093861                  弁理士 大賀 真司                  (72) 発明者 角井 健太郎                  神奈川県小田原市中里322番2号 株式会社日立製作所SANソリューション事業部内                  (72) 発明者 阿知和 恭介                  神奈川県小田原市中里322番2号 株式会社日立製作所SANソリューション事業部内</p>
--	---

最終頁に続く

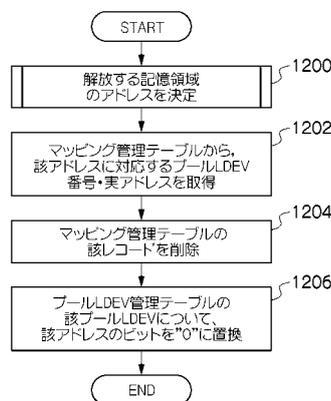
(54) 【発明の名称】 ストレージシステム及びその制御方法

(57) 【要約】

【解決課題】 AOU技術を利用したストレージシステムにおいて、記憶領域の利用効率を改善する。

【解決手段】 ストレージシステムのコントローラは、上位装置による仮想ボリュームへのアクセスに応じて、上位装置がアクセスした仮想ボリュームの領域に実ボリュームの記憶領域を割り当て、この割り当てを維持する必要性が低下したことを検出し、その検出結果に基づいて仮想ボリュームの領域に対する、前記実ボリュームの記憶領域の割り当てを解放する。

【選択図】 図12



## 【特許請求の範囲】

## 【請求項 1】

記憶デバイスと、通信経路を介して接続される上位装置からの要求に応じて、前記記憶デバイスにデータを入出力する制御を行うコントローラと、を備えるストレージシステムにおいて、

前記上位装置がアクセスする仮想ボリュームと、

前記記憶デバイスの記憶領域を有する実ボリュームと、を有し、

前記実ボリュームが前記仮想ボリュームに割り当てられることによって、前記仮想ボリュームを介して前記上位装置が前記実ボリュームの記憶領域にアクセスするように構成され、

10

前記コントローラは、前記上位装置による前記仮想ボリュームへのアクセスに応じて、前記上位装置がアクセスした仮想ボリュームの領域に前記実ボリュームの記憶領域を割り当て、この割り当てを維持する必要が低下したことを検出し、その検出結果に基づいて前記仮想ボリュームの領域に対する、前記実ボリュームの記憶領域の割り当てを解放する、ストレージシステム。

## 【請求項 2】

第 1 の仮想ボリュームと第 2 の仮想ボリュームと、を備え、当該第 1 の仮想ボリュームと第 2 の仮想ボリュームと、がコピーペアの関係を有してなり、前記コントローラは、このコピー関係が解消されたことを検出して、コピー先となる仮想ボリュームへの前記記憶領域の割り当てを解放する、請求項 1 記載のストレージシステム。

20

## 【請求項 3】

第 1 の仮想ボリュームと第 2 の仮想ボリュームと、を備え、当該第 1 の仮想ボリュームと第 2 の仮想ボリュームとがコピーペアの関係を有してなり、前記コントローラは、両ボリューム間のコピーが終了し、そしてコピー先ボリュームのデータが外部媒体にバックアップされた後、コピー先ボリュームへの前記記憶領域の割り当てを解放する、請求項 1 記載のストレージシステム。

## 【請求項 4】

前記上位装置からのアクセスの状況を判定するための制御情報を記憶する制御メモリを備え、前記コントローラは、前記制御情報の内容にしたがって前記解放を実行する、請求項 1 記載のストレージシステム。

30

## 【請求項 5】

前記制御情報は、前記上位装置からの実アクセスがあった記憶領域か、前記上位装置からの将来アクセスが予測される記憶領域かを示す情報である、請求項 4 記載のストレージシステム。

## 【請求項 6】

前記コントローラは、前記実ボリュームの記憶領域について前記制御情報を定期的に走査し、この結果に基づいて前記解放を実行する請求項 4 記載のストレージシステム。

## 【請求項 7】

前記コントローラは、前記上位装置のアプリケーションプログラムの実行結果に基づいて、前記解放を実行する請求項 1 記載のストレージシステム。

40

## 【請求項 8】

前記コントローラは、前記上位装置のアプリケーションプログラムによるデータの削除を契機として、このデータに対応する前記実ボリュームの記憶領域について前記解放を実行する、請求項 1 記載のストレージシステム。

## 【請求項 9】

前記コントローラは、前記上位装置から前記仮想ボリュームへのアクセスコマンドを監視し、このアクセスコマンドに基づいて、前記解放を実行する請求項 1 記載のストレージシステム。

## 【請求項 10】

記憶デバイスと、コントローラと、を有し、通信経路を介して接続される上位装置から

50

の要求に応じて、前記記憶デバイスにデータを入出力する制御を行うストレージ装置において、

前記上位装置がアクセスする仮想ボリュームと、

前記記憶デバイスの記憶領域を有する実ボリュームと、を有し、

前記実ボリュームが前記仮想ボリュームに割り当てられることによって、前記仮想ボリュームを介して前記上位装置が前記実ボリュームの記憶領域にアクセスするように構成され、さらに、

前記コントローラは、前記上位装置による前記仮想ボリュームへのアクセスに応じて、前記上位装置がアクセスした仮想ボリュームの領域に前記実ボリュームの記憶領域を割り当て、そして前記上位装置が前記仮想ボリュームに対して続いてアクセスする領域を予測し、この予測された領域に前記実ボリュームの記憶領域を割り当てる、ストレージシステム。

10

#### 【請求項 1 1】

前記予測されたアドレスに対する前記上位装置のアクセス状況に基づいて、前記コントローラは前記予測された領域への前記実ボリュームの記憶領域の割り当てを解放する、請求項 1 0 記載のストレージシステム。

#### 【請求項 1 2】

前記仮想ボリュームの領域について、前記上位装置がアクセスした領域か前記予測された領域かを区別する制御情報を備え、前記コントローラはこの制御情報に基づいて前記解放を実行する、請求項 1 0 記載のストレージシステム。

20

#### 【請求項 1 3】

前記コントローラは前記制御情報を定期的に参照し、その参照結果に基づいて前記解放を実行する、請求項 1 2 記載のストレージシステム。

#### 【請求項 1 4】

前記コントローラは、前記実ボリュームの記憶領域のうち前記仮想ボリュームに対して割り当てられている割合を算出し、この割合がある値以上の場合に、前記解放を実行する、請求項 1 記載のストレージシステム。

#### 【請求項 1 5】

前記コントローラは前記解放された、前記実ボリュームの記憶領域のデータを消去する、請求項 1 記載のストレージシステム。

30

#### 【請求項 1 6】

記憶デバイスと、通信経路を介して接続される上位装置からの要求に応じて、前記記憶デバイスにデータを入出力する制御を行うコントローラと、を備えるストレージシステムの制御において、

前記ストレージシステムは、前記上位装置がアクセスする仮想ボリュームと、前記記憶デバイスの記憶領域を有する実ボリュームと、を有し、前記実ボリュームが前記仮想ボリュームに割り当てられることによって、前記仮想ボリュームを介して前記上位装置が前記実ボリュームの記憶領域にアクセスするように構成され、

前記コントローラは、前記上位装置による前記仮想ボリュームへのアクセスに応じて、前記上位装置がアクセスした仮想ボリュームの領域に前記実ボリュームの記憶領域を割り当てるステップと、この割り当てを維持する必要が低下した、前記仮想ボリュームの領域に対して、前記実ボリュームの記憶領域の割り当てを解放するステップと、を実行するように構成された、ストレージシステムの制御方法。

40

#### 【発明の詳細な説明】

#### 【技術分野】

#### 【0001】

本発明は、ハードディスクドライブ（以下、HDD、と称する）等の記憶デバイスに対するデータの記憶制御を行うストレージシステムに関するものである。

#### 【背景技術】

#### 【0002】

50

ストレージシステムと、このストレージシステムにネットワークなどの通信経路を介して接続するサーバやホストコンピュータなどの上位装置とを備えた、コンピュータシステムで扱われるデータ量の増加に伴い、ストレージシステムが備える記憶領域が大容量化されてきている。ストレージ装置は、上位装置がアクセス可能なボリュームを論理的に定義し、この論理ボリュームの物理的記憶領域に上位装置がアクセスすることによって、記憶デバイスに対するデータの入出力が可能となる。

#### 【0003】

最近では、上位装置が取り扱うデータ量が大幅に増えていることから、ボリュームの記憶容量であるボリュームサイズを大幅に増やさなくてはならなくなっている。最初から大きなサイズの記憶領域を上位装置に割り当てておけば、上位装置にとって記憶容量の不足は起こらず、途中で上位装置に割り当てられている記憶領域のサイズを拡大する必要もない。しかしながら、上位装置である計算機がそれほど大量のデータを使わない場合には、計算機に割り当てられた記憶領域に未使用の領域が発生して記憶領域に無駄が発生する。そこで、上位計算機のプロセッサがストレージ装置の仮想ボリュームに書き込みを行ったときに初めて、書き込みが行われた仮想ボリューム上の領域に実記憶領域を割り当てることが提案されている（特開2005-11316号公報）。さらに、米国特許第6,823,442号公報には、上位装置がアクセスする仮想ボリュームをストレージシステムに設け、この仮想ボリュームに物理的記憶領域を割り当てることが記載されている。

10

#### 【0004】

なお、本発明に関連する従来例として、特開2005-128771号公報には、データを記憶する物理ストレージを有しアクセス要求を受けてその物理ストレージをアクセスするデータアクセスサーバと、そのデータアクセスサーバにアクセス要求を送りデータアクセスサーバをデータ保存のために利用してデータファイリングを行なうデータファイルサーバとを備えたデータファイルシステム等に関し、論理ブロックが不使用になったときに物理ブロックを有効に解放することが記載されている。また、特開2004-127301号公報には、仮想記憶空間と、物理記憶空間と、仮想記憶空間の複数のアドレスを物理記憶空間の各アドレスに関連付けるように構成されたマッピングシステムと、順序に従ってマッピングシステムの除去部分を指定する優先順位付け規定に従って、マッピングシステムの一部を自動的に除去するように構成されたコントローラと、を備える記憶システムが記載されている。

20

30

【特許文献1】特開2005-11316号公報

【特許文献2】米国特許第6,823,442号公報

【特許文献3】特開2005-128771号公報

【特許文献4】特開2004-127301号公報

【発明の開示】

【発明が解決しようとする課題】

#### 【0005】

本願出願人はストレージシステムにおいて記憶資源を有効に活用するためにAllocation on Use(AOU)という技術を開発している。このAOU技術は、ストレージシステムが、それ自身は記憶領域を持たない仮想ボリュームを上位装置に提供し、この仮想ボリュームにはプールと呼ばれる記憶領域の集合を対応させている。上位装置が仮想ボリュームにアクセスすると、ストレージシステムは上位装置によってアクセスされた仮想ボリュームの領域にプールに含まれる記憶領域を割り当てる。この割り当ては、上位装置が仮想ボリュームにアクセスすることによって実行される。一つのプールには複数の仮想ボリュームが対応可能であるために、ストレージシステムはプールの記憶領域を有効活用することができる。ストレージシステムはあらかじめ大きな、見かけの容量を持つ仮想ボリュームを用意しておき、ストレージシステムは、プールの利用状況に応じて、プールに記憶容量を追加する。

40

#### 【0006】

AOU技術は、プールの記憶領域が仮想ボリュームの領域に割り当てられる構成を備え

50

ているため、上位装置が仮想ボリュームの領域にアクセスするにしたがって仮想ボリュームに割り当てられる記憶領域は増加する一方である。記憶領域に対する上位装置のアクセスが今後実質的に無いと判断される場合に、仮想ボリュームに対する記憶領域の割り当てが固定されたままであると、この記憶領域の利用効率が低下する。また、記憶領域の仮想ボリュームの領域への割り当てが、上位装置が仮想ボリュームにアクセスする際に行われるために、上位装置が仮想ボリュームを介してプールの記憶領域にアクセスする際のアクセススピードに改善の余地がある。

**【 0 0 0 7 】**

そこで、本発明は、A O U技術を利用したストレージシステムにおいて、記憶領域の利用効率を改善することにある。本発明の他の目的は、A O U技術を利用したストレージシステムにおいて、上位装置が仮想ボリュームを介して記憶領域にアクセスする際のアクセススピードを改善することにある。さらに本発明の他の目的は、A O U技術を利用したストレージシステムにおいて、上位装置が仮想ボリュームを介して記憶領域にアクセスする際のアクセススピードを改善するとともに、記憶領域の利用効率を改善することにある。

10

**【 課題を解決するための手段 】****【 0 0 0 8 】**

本発明は、仮想ボリュームに一旦記憶領域を割り当てた後、何らかの事情を契機として、この割当を解放することにより、記憶資源の有効利用が可能なストレージシステムを提供することを特徴とするものである。この契機とはさまざまなものが考えられるが、例えば、ボリュームペアのペア状態の解除、正ボリュームのスナップショットを記録する副ボリュームを外部媒体にバックアップした後のタイミング、上位装置からストレージシステムに発行されたデータ削除のコマンドの処理のタイミング、そして、上位装置に対して予め割り当てられているものの、上位装置からのアクセスが無い記憶領域について、定期的に解放処理を実行するタイミング、がある。

20

**【 0 0 0 9 】**

本発明は、記憶デバイスと、通信経路を介して接続される上位装置からの要求に応じて、前記記憶デバイスにデータを入出力する制御を行うコントローラと、を備えるストレージシステムにおいて、前記上位装置がアクセスする仮想ボリュームと、前記記憶デバイスの記憶領域を有する実ボリュームと、を有し、前記実ボリュームが前記仮想ボリュームに割り当てられることによって、前記仮想ボリュームを介して前記上位装置が前記実ボリュームの記憶領域にアクセスするように構成され、前記コントローラは、前記上位装置による前記仮想ボリュームへのアクセスに応じて、前記上位装置がアクセスした仮想ボリュームの領域に前記実ボリュームの記憶領域を割り当て、この割り当てを維持する必要が低下したことを検出し、その検出結果に基づいて前記仮想ボリュームの領域に対する、前記実ボリュームの記憶領域の割り当てを解放する、ことを特徴とするものである。

30

**【 0 0 1 0 】**

さらに、本発明は、記憶デバイスと、コントローラと、を有し、通信経路を介して接続される上位装置からの要求に応じて、前記記憶デバイスにデータを入出力する制御を行うストレージ装置において、前記上位装置がアクセスする仮想ボリュームと、前記記憶デバイスの記憶領域を有する実ボリュームと、を有し、前記実ボリュームが前記仮想ボリュームに割り当てられることによって、前記仮想ボリュームを介して前記上位装置が前記実ボリュームの記憶領域にアクセスするように構成され、さらに、前記コントローラは、前記上位装置による前記仮想ボリュームへのアクセスに応じて、前記上位装置がアクセスした仮想ボリュームの領域に前記実ボリュームの記憶領域を割り当て、そして前記上位装置が前記仮想ボリュームに対して続いてアクセスする領域を予測し、この予測された領域に前記実ボリュームの記憶領域を割り当てる、ことを特徴とするものである。

40

**【 発明の効果 】****【 0 0 1 1 】**

以上説明したように、本発明によれば、仮想ボリュームに一旦記憶領域を割り当てた後、何らかの事情を契機として、この割当を解放することにより、記憶資源の有効利用が可

50

能なストレージシステムを提供することができる。更に本発明によれば、本発明の他の目的は、A O U技術を利用したストレージシステムにおいて、上位装置が仮想ボリュームを介して記憶領域にアクセスする際のアクセススピードを改善することができる。さらにA O U技術を利用したストレージシステムにおいて、上位装置が仮想ボリュームを介して記憶領域にアクセスする際のアクセススピードを改善するとともに、記憶領域の利用効率を改善することができる。

【発明を実施するための最良の形態】

【0012】

以下、本発明の実施の形態を図面に基づいて詳細に説明する。なお、以後説明される図面においては、同一部には同一符号を付し、その繰り返しの説明は省略又は簡略化される。

10

【0013】

図1は、本発明が適用されるストレージシステム(ストレージ装置ということもある。)600を含んだ、記憶制御システムを示すハードウェアブロック図である。ストレージシステム600は複数の記憶デバイス300と、情報処理装置200からの入出力要求にตอบสนองして記憶デバイス300への入出力制御を行う記憶デバイス制御装置(コントローラ)100と、を備えて構成されている。

【0014】

情報処理装置200は上位装置に相当し、CPUやメモリ等を備えたサーバ(ホスト)や、ストレージ装置の管理用コンピュータからなり、例えば、ワークステーション、メインフレームコンピュータ、パーソナルコンピュータ等で構成される。情報処理装置200は複数台のコンピュータをネットワーク接続して構成することもできる。情報処理装置200はオペレーティングシステム上で動作するアプリケーションプログラムを実装している。アプリケーションプログラムとしては、例えば、銀行の自動預金預け払いシステムや航空機の座席予約システム等がある。また、サーバとしては、アプリケーションを実行する、更新系サーバや、更新系サーバのバックエンドでバックアップなどを実行するバックアップサーバが存在する。

20

【0015】

情報処理装置1乃至3(200)はLAN(Local Area Network)400を介してストレージ装置600と接続されている。LAN400は、例えば、イーサネット(登録商標)やFDDI等の通信網であり、情報処理装置1乃至3(200)とストレージシステム600との間の通信はTCP/IPプロトコルスイートによって行われる。情報処理装置1乃至3(200)からはストレージシステム600に対してファイル名指定によるデータアクセス要求(ファイル単位でのデータ入出力要求。以下、ファイルアクセス要求と称する。)が後述のチャンネル制御部CHN1乃至CHN4(110)に送信される。

30

【0016】

LAN400にはバックアップデバイス910が接続されている。バックアップデバイス910は、例えば、MO、CD-R、DVD-RAM等のディスク系デバイス、DATテープ、カセットテープ、オープンテープ、カートリッジテープ等のテープ系デバイスである。バックアップデバイス910はLAN400を介して記憶デバイス制御装置100との間で通信を行うことにより記憶デバイス300に記憶されているデータのバックアップデータを記憶する。また、バックアップデバイス910は情報処理装置1(200)と接続し、情報処理装置1(200)を介して記憶デバイス300に記憶されているデータのバックアップを取得する。

40

【0017】

記憶デバイス制御装置100はチャンネル制御部CHN1乃至4(110)を備える。記憶デバイス制御装置100は、チャンネル制御部CHN1乃至4(110)とLAN400とを介して、情報処理装置1乃至3(200)と、バックアップデバイス910と、記憶デバイス300との間のライトアクセス又はリードアクセスを媒介する。チャンネル制御部CHN1乃至4(110)は情報処理装置1乃至3(200)からのファイルアクセス要

50

求を個々に受け付ける。すなわち、チャンネル制御部CHN1乃至4(110)には個々にLAN400上のネットワークアドレス(例えば、IPアドレス)が割り当てられていて、それぞれが個別にNASとして振る舞い、個々のNASがあたかも独立したNASが存在するようにNASとしてのサービスを情報処理装置1乃至3(200)に提供することができる。このように1台のストレージシステム600に個別にNASとしてのサービスを提供するチャンネル制御部CHN1乃至4(110)を備えるように構成したことで、従来、独立したコンピュータで個々に運用されていたNASサーバが一台のストレージシステム600に集約される。そして、これによりストレージシステム600の統括的な管理が可能となり、各種設定・制御や障害管理、バージョン管理といった保守業務の効率化が図られる。

10

**【0018】**

情報処理装置3乃至4(200)はSAN500を介して記憶デバイス制御装置100と接続されている。SAN500は、記憶デバイス300が提供する記憶領域におけるデータの管理単位であるブロックを単位として情報処理装置3乃至4(200)との間でデータの授受を行うためのネットワークである。SAN500を介して行われる情報処理装置3乃至4(200)と記憶デバイス制御装置100との間の通信は一般にSCSIプロトコルに従って行われる。情報処理装置3乃至4(200)からは、ストレージシステム600に対して、ファイバチャンネルプロトコルに従ってブロック単位のデータアクセス要求(以下、ブロックアクセス要求と称する。)が送信される。

**【0019】**

SAN500にはSAN対応のバックアップデバイス900が接続されている。SAN対応バックアップデバイス900は、SAN500を介して記憶デバイス制御装置100との間で通信を行うことにより、記憶デバイス300に記憶されているデータのバックアップデータを記憶する。

20

**【0020】**

記憶デバイス制御装置100はチャンネル制御装置CHN1乃至4(110)に加えて、更にチャンネル制御部CHF1乃至2,CHA1乃至2(110)を備える。記憶デバイス制御装置100はチャンネル制御部CHF1乃至2(110)とSAN500とを介して、情報処理装置3乃至4(200)及びSAN対応バックアップデバイス900との間の通信を行う。チャンネル制御部は、上位装置からのアクセスコマンドを処理する。

30

**【0021】**

情報処理装置5(200)は、更に、LAN400やSAN500等のネットワークを介さずに記憶デバイス制御装置100と接続されている。この情報処理装置5(200)の例は、例えばメインフレームコンピュータである。情報処理装置5(200)と記憶デバイス制御装置100との間の通信は、例えばFICON(Fiber Connection)(登録商標)やESCON(Enterprise System Connection)(登録商標)、ACONARC(Advanced Connection Architecture)(登録商標)、FIBARC(Fiber Connection Architecture)(登録商標)などの通信プロトコルに従って行われる。情報処理装置5(200)からはストレージシステム600に対して、これらの通信プロトコルに従ってブロックアクセス要求が送信される。記憶デバイス制御装置100は、チャンネル制御部CHA1乃至2(110)を通じて情報処理装置5(200)との間で通信を行う。

40

**【0022】**

SAN500には、他のストレージシステム610が接続されている。ストレージシステム610は、記憶デバイス制御装置100にストレージシステムが有する記憶資源を提供するという機能を、情報処理装置及びストレージ装置600に提供する。ストレージ装置の情報処理装置に対する見かけ上の記憶領域が、ストレージ装置610によって拡張されたことになる。なお、ストレージシステム610はSAN500以外にもATMなどの通信回線によりストレージシステム600に接続していることもある。また、ストレージシステム610をストレージシステム600に直接接続することもできる。

**【0023】**

50

このように、ストレージシステム600内にチャンネル制御部CHN1乃至4(110)、チャンネル制御部CHF1乃至2(110)、チャンネル制御部CHA1乃至2(110)を混在させて装着させることにより、異種ネットワークに接続できるストレージシステムを実現できる。即ち、このストレージシステム600はチャンネル制御部CHN1乃至4(110)を用いてLAN400に接続し、かつチャンネル制御部CHF1乃至2(110)を用いてSAN500に接続するという、SAN-NAS統合ストレージシステムである。

#### 【0024】

接続部150は、各々のチャンネル制御部110、共有メモリ120、キャッシュメモリ130、及び各々のディスク制御部140を相互に接続する。これらのチャンネル制御部110、共有メモリ120、キャッシュメモリ130、及びディスク制御部140相互間のコマンド又はデータの送受信は接続部150を介して行われる。接続部150は例えば高速スイッチングによりデータ伝送を行う超高速クロスバススイッチ等の高速バスで構成される。これにより、チャンネル制御部110相互間の通信パフォーマンスが大幅に向上するとともに、高速なファイル共有機能や高速フェイルオーバー等が可能になる。

10

#### 【0025】

共有メモリ120、及びキャッシュメモリ130はチャンネル制御部110、及びディスク制御部140が共用するメモリ装置である。共有メモリ120は主として制御情報やコマンド等を記憶するために利用され、キャッシュメモリ130は主としてデータを記憶するために利用される。例えば、あるチャンネル制御部110が情報処理装置200から受け取ったデータ入出力コマンドが書き込みコマンドであった場合に、当該チャンネル制御部110はその書き込みコマンドを共有メモリ120に書き込むとともに、情報処理装置200から受け取ったライトデータをキャッシュメモリ130に書き込む。一方、ディスク制御部140は共有メモリ120を監視しており、共有メモリ120に書き込みコマンドが書き込まれたと判定すると、当該書き込みコマンドに従ってキャッシュメモリ130からライトデータを読み出して、これを記憶デバイス300に書き込む。

20

#### 【0026】

一方、あるチャンネル制御部110が情報処理装置200から受け取ったデータ入出力コマンドが読み出しコマンドであった場合に、当該チャンネル制御部110はその読み出しコマンドを共有メモリ120に書き込むとともに、読み出し対象となるデータがキャッシュメモリ130に存在しているか否かをチェックする。ここで、読み出し対象となるデータがキャッシュメモリ130に存在している場合には、チャンネル制御部110はそのデータをキャッシュメモリ130から読み取って情報処理装置200に送信する。読み出し対象となるデータがキャッシュメモリ130に存在していない場合には、共有メモリ120に読み出しコマンドが書き込まれたことを検出したディスク制御部140は読み出し対象となるデータを記憶デバイス300から読み出してこれをキャッシュメモリ130に書き込むとともに、その旨を共有メモリ120に書き込む。チャンネル制御部110は共有メモリ120を監視することにより、読み出し対象となるデータがキャッシュメモリ130に書き込まれたことを検出すると、そのデータをキャッシュメモリ130から読み出して情報処理装置200に送信する。

30

40

#### 【0027】

ディスク制御部140はチャンネル制御部110から送信された論理アドレス指定による記憶デバイス300へのデータアクセス要求を、物理アドレス指定によるデータアクセス要求に変換し、チャンネル制御部110から出力されるI/O要求に応答して記憶デバイス300へのデータの書き込み又は読み取りを行う。記憶デバイス300がRAID構成されている場合には、ディスク制御部140はRAID構成に従ったデータのアクセスを行う。すなわち、ディスク制御部は、記憶デバイスであるHDDの制御、RAIDグループの制御を行う。複数のHDDの記憶領域によってRAIDグループが構成されている。

#### 【0028】

記憶デバイス300は単一又は複数のディスクドライブ(物理ボリューム)を備えてお

50

り、情報処理装置 200 からアクセス可能な記憶領域を提供する。記憶デバイス 300 が提供する記憶領域には単一又は複数の物理ボリュームの記憶空間を合わせた論理ボリュームが設定されている。記憶デバイス 300 に設定される論理ボリュームには、情報処理装置 200 からアクセス可能なユーザ論理ボリュームや、チャンネル制御部 110 の制御のために使用されるシステム論理ボリュームがある。システム論理ボリュームにはチャンネル制御部 110 で実行されるオペレーティングシステムも格納されている。また、記憶デバイス 300 が、上位装置に提供する論理ボリュームには、各々のチャンネル制御部 110 がアクセス可能な論理ボリュームが割り当てられている。もとより、複数のチャンネル制御部 110 が同一の論理ボリュームを共有することもできる。

#### 【0029】

尚、記憶デバイス 300 として、例えば、ハードディスク装置等を用いることができる。記憶デバイス 300 のストレージ構成として、例えば、複数の記憶デバイス 300 により RAID 方式のディスクアレイを構成してもよい。また、記憶デバイス 300 と記憶デバイス制御装置 100 は直接接続してもよく、ネットワークを介して接続してもよい。更に、記憶デバイス 300 は記憶デバイス制御装置 100 と一体に構成してもよい。

#### 【0030】

管理端末 160 はストレージシステム 600 を保守・管理するためのコンピュータ装置であり、内部 LAN 151 を通じて各々のチャンネル制御部 110 とディスク制御部 140 に接続している。オペレータは管理端末 160 を操作することで記憶デバイス 300 のディスクドライブの設定、論理ボリュームの設定、チャンネル制御部 110 やディスク制御部 140 で実行されるマイクロプログラムのインストール等を行うことが可能である。この種の制御は、管理端末に基づくもののほか、上位装置（ホスト）上で動作するプログラムがネットワークを経由するものに基づいても良い。

#### 【0031】

図 2 は図 1 に示す記憶制御システムの機能ブロック図である。チャンネル制御部 110 は、マイクロプロセッサ CT1 とローカルメモリ LM1 を備え、チャンネルコマンド制御プログラムはローカルメモリ LM1 に記憶されている。マイクロプロセッサ CT1 はローカルメモリ LM1 を参照してチャンネルコマンド制御プログラムを実行する。チャンネルコマンド制御プログラムは、ホストに LU を提供する。チャンネルコマンド制御プログラムはホストから LU に対して送られるアクセスコマンドを処理し、LDEV へのアクセスへと変換する。チャンネルコマンド制御プログラムはホストからのアクセスに依らず、LDEV に対してアクセスすることもある。LDEV とは、RAID グループの一部を論理的に切り出した論理ボリュームである。但し、仮想 LDEV はホストのアクセスを受けるが、実記憶領域を備えていない。ホストは LDEV ではなく、LU にアクセスする。LU とは、ホストがアクセスする記憶領域の単位である。一部の LU は仮想 LDEV に割り当てられている。以後仮想 LDEV に割り当てられている LU を実ボリュームに割り当てられている LU と区別するために、便宜上仮想 LU という。

#### 【0032】

ディスク制御部 140 は、マイクロプロセッサ CT2 とローカルメモリ LM2 とを備えている。ローカルメモリ LM2 には、RAID 制御プログラムと HDD 制御プログラムが記憶されている。マイクロプロセッサ CT2 はローカルメモリ LM2 を参照して、RAID 制御プログラムと HDD 制御プログラムを実行する。RAID 制御プログラムは、複数の HDD から RAID グループを構成し、上位のチャンネルコマンド制御プログラムに LDEV を提供する。HDD 制御プログラムは、上位の RAID 制御プログラムの要求に応じて HDD に対してデータの読み書きを実行する。ホスト 200A は LU10 を介して LDEV12A にアクセスする。ホスト 200B に対する記憶領域は AOU 技術によって構成されている。ホスト 200B は仮想 LU14 を介して仮想 LDEV16 にアクセスする。仮想 LDEV16 にはプール 18 が割り当てられており、このプールには LDEV12B, 12C が割り当てられている。仮想 LU と仮想 LDEV とが仮想ボリュームに相当する。プールとは、仮想 LDEV に割り当てられる LDEV の集合をいう。なお、チャンネル I

10

20

30

40

50

／F，I/Oパスとは、ホストがストレージサブシステムにアクセスするインターフェースであり、ファイバチャネル、iSCSIなどをいう。

【0033】

図3は仮想ボリュームとプールとの関係を示すブロック図である。ホストは仮想ボリューム16に対してアクセスする。アクセスされた仮想ボリュームの領域はプール(物理記憶装置)18にマッピングされる。このマッピングは仮想ボリュームに対するホストからのアクセスに応じて動的に生成され、以後、ストレージシステムがこのマッピング情報を利用する。仮想ボリュームの領域のうち使用されていない領域の分は物理記憶装置を消費しないので、あらかじめ十分大きな仮想ボリュームを用意しておき、プール18の利用状況を見て々にプール18Aを仮想ボリュームの割り当てに追加するという運用が可能になる。

10

【0034】

チャンネル制御部110がチャンネルコマンド制御プログラムにしたがって、AOUの動作を実行するに際して、管理テーブルが利用される。各管理テーブルについて説明する。図4はLU管理テーブルであり、LUの一覧を保持する。LUはチャンネル番号とLUNの組で特定される。仮想ビットが「OFF」の場合は、LUはRAID制御プログラムが提供するLDEVにマッピングされる。仮想ビットが「ON」の場合、LUは仮想LDEVにマッピングされていることを示している。

【0035】

図5は仮想LDEV管理テーブルであり、仮想LDEVの一覧を保持する。仮想LDEVは仮想LDEV番号で特定される。仮想LDEV番号はRAID制御プログラムが提供するLDEVのLDEV番号とは別の管理体系を持つ。プール番号は仮想LDEVとのマッピング対象となるプールを特定するためのものである。

20

【0036】

図6は、RAID制御プログラムが、チャンネル制御部及び上位装置に対して提供するLDEVの一覧を示す。LDEVはLDEV番号で特定される。用途レジスタが"L"である場合は、LDEVがLUに対してマッピングされていることを示している。"P"である場合は、LDEVがプールを構成すること、換言すれば、LDEVがプールに属していることを示している。"N"である場合は、LDEVが未使用であることを示している。

【0037】

図7はプール管理テーブルであり、プールの一覧を示す。プールはプール番号で特定される。LDEV番号リストには、そのプールを構成するLDEV(プールLDEV)がリストアップされている。図8はマッピング管理テーブルであり、仮想LDEVの記憶領域と、それに対応するプールLDEVの記憶領域のマッピングを保持する。このテーブルは仮想LDEV毎に存在する。図に示す例は、仮想LDEV番号が0001の場合である。このテーブルには、仮想LDEVの領域であるアドレスと、それに対応するプールLDEVのあるアドレス(実アドレス)が記録される。

30

【0038】

ここでいうアドレスとは、仮想LDEVおよびプールLDEVそれぞれをある一定のサイズで区切って通番を振ったものであり、仮想LDEVとそれが使用するプールLDEV群との間に対応がとれるようなものであれば良い。このアドレスの体系は、ホストがLUにアクセスする際に用いるアドレッシング体系とは共通でも良いしあるいは異なっても良い。

40

【0039】

図9は図8の制御テーブルで示された、仮想LDEVのアドレスとプールに属するLDEVの実記憶領域との対応関係をグラフ化したものである。アクセスビットが"1"である場合、仮想LDEVの領域とプールLDEVの記憶領域とのマッピングが、上位装置が仮想LDEVへアクセスコマンドを発行したことに基づくもの、すなわち仮想LDEVの仮想記憶領域とプールLDEVの実記憶領域間の対応付けが、ホストが仮想LUにアクセスしたことに基づくものであることを示している。アクセスビットが"0"である場合は、そ

50

のマッピングがそれ以外の契機によるものであることを示している。詳しくは後述するが、例えば、ホストからの将来アクセスを予測して、割当を行った場合である。

#### 【0040】

図10は、プールLDEVの記憶領域が仮想LDEVの領域にマッピングされている状態に係わるプールLDEV管理テーブルである。マッピングの状態は、各々のプールLDEVについて、プールLDEVのアドレス毎に設定される制御ビットの配列（制御情報）として保持されている。プールLDEV管理テーブルには該ビット配列へのポインタが記録されている。ビット配列の先頭をアドレス0とし、以降アドレス順にビットが設定されている。ビットが"0"ならばこのビットに対応する記憶領域は仮想LDEVに対してマッピングされていないことを示し、"1"ならばマッピング済みであることを示している。図10に基づく、プールLDEV番号0003のLDEVでは、アドレス0,1,2の領域が仮想LDEVにマッピングされている。プールLDEV番号0004のLDEVの全領域は仮想LDEVにマッピングされていない。プールLDEV番号0005のアドレス0の領域は、仮想LDEVにマッピングされている。

10

#### 【0041】

以上説明した制御テーブルは、チャンネル制御部のチャンネルコマンド制御プログラムがアクセス可能なメモリ空間に保持されている。制御テーブルは、図1の管理端末によってその内容を変更することができる。このメモリ空間としては、図1の共有メモリ120、又は、チャンネル制御部のローカルメモリLM1（図2）がある。

#### 【0042】

図11は、ホストからLUにアクセスがあった場合のチャンネルコマンド制御プログラムによるAOUの動作を示すフローチャートである。チャンネル制御部はチャンネルコマンド制御プログラムに基づいて、図11のAOU処理を実行する。チャンネル制御部は、LUに対する、ホストからのアクセスコマンドを受領すると（1100）、アクセス対象LUの仮想ビットがONであるか否かをチェックする（1102）。仮想ビットがOFFの場合は、ステップ1104に移行して、ホストのLUに対するアクセスは、実ボリューム（LDEV）にマッピングされる（図4のLDEV番号0001,0002）。

20

#### 【0043】

ステップ1102において、仮想ビットがONである場合には、ホストからのアクセス先であるLUのアドレスをマッピング管理用の仮想LDEVのアドレスに変換する（1106）。次いで、チャンネル制御部のコントローラCT1は、このアドレスが図8のマッピング管理テーブルに未登録か否かを判定する（1108）。このアドレスが図8のマッピング管理テーブルに未登録の場合は、チャンネル制御部はプール管理テーブルに基づいて、仮想LDEVが対応するプールに属するLDEV番号リストを取得する（1112）。チャンネル制御部はプールLDEVのリストのX番目（X:=1・・・）のLDEVを取得する（1114）。チャンネル制御部はX番目のLDEVが存在するか否かを判定する（1116）。

30

#### 【0044】

この判定が肯定されると、チャンネル制御部は、プールLDEV管理テーブル（図10）からLDEVのビット配列を取得する（1118）。チャンネル制御部は、ビット配列に"0"のビットがあるか否かを判定（1120）し、これが肯定された場合には、このビットに"1"を設定する（1122）。この判定が否定された場合には、プールLDEVリストの次の番号（X++）のLDEVを取得する。ステップ1124では、マッピング管理テーブル（図8）に、ビット"1"にした、対応するアドレスを実アドレスとして登録する。次いで、マッピング管理テーブルのアクセスビットを"1"に設定する（1126）。次いで、チャンネル制御部はホストからの仮想ボリューム（0001）へのアクセスを、プールLDEV番号・実アドレスへのアクセスにマッピングする（1128）。

40

#### 【0045】

チャンネルアダプタはステップ1108を否定判定すると、ステップ1110において、図8のマッピング管理テーブルのアクセスビットが"0"か否かを判定する。これが肯定

50

判定されると、アクセスビットが"1"に設定され(1130)、チャンネル制御部はホストからの仮想ボリューム(0001)へのアクセスを、プールLDEV番号・実アドレスへのアクセスにマッピングする(1132)。ステップ1116において、X番目のLDEVが存在しない場合には、チャンネル制御部は管理端末に対してプールLDEV不足エラーを通知する。図11の処理によれば、チャンネル制御部はホストから仮想ボリュームにアクセスがあると、アクセスがあった領域に実記憶領域のアドレスをマッピングする。

【0046】

図12は、プールLDEVの記憶領域を仮想ボリュームの記憶領域へマッピングするためのレコードを、マッピング管理テーブルから削除することによりプールLDEVの記憶領域を仮想ボリュームへのマッピングから解放して、他の仮想ボリュームに対する記憶領域に開放するなどとして有効活用するための、チャンネル制御部がチャンネルコマンド制御プログラムに基づいて実行するフローチャートである。先ず、チャンネル制御部は、仮想ボリュームの記憶領域のうち、この解放の対象となるべきアドレスを決定する(1200)。

10

【0047】

チャンネル制御部は、仮想ボリュームのアドレスが決定された後、マッピング管理テーブル(図8)から、仮想ボリュームのアドレスに対応するプールLDEV番号・実アドレスを取得する(1202)。次いで、マッピング管理テーブルについて、仮想ボリュームのアドレスとLDEV番号・実アドレスとのマッピングに関するレコードを削除する(1204)。次いで、チャンネルアダプタは、プールLDEV管理テーブル(図10)の該プールLDEVについて、仮想LDEVのアドレスに対応する、プールLDEVのアドレスについて、このアドレスに対するビットを"0"に置換する(1206)。これにより、プールLDEVの実記憶領域がマッピングから解放される。

20

【0048】

次に、解放の契機の第1の例について説明する。第1の例に係わる契機は、ボリュームペアの解除を契機とする解放である。図13はボリュームペアを説明するブロック図である。図13において、LU1とLU2との間に、バックアップなどの目的でボリュームペアが形成されている。LU1が副ボリュームであり、LU2が正ボリュームである。ホストのアクセスはLU2に対して行われる。チャンネル制御部110は、定期的に正ボリュームLU2(コピー元)のコピーを実行してボリューム内容のスナップショットを副ボリュームLU1に対して作成する。副ボリュームLU1は仮想LDEV1に対応付けられており、この仮想LDEVはプール内のプールLDEVにマッピングされている。副ボリュームに対応するプールLDEVのデータは、テープデバイスなど外部媒体にバックアップされる。正ボリュームと副ボリュームとのペア関係が解除された場合、コピー先ボリューム(副ボリューム)LU1の内容は不要である。この副ボリュームが仮想LDEVにマッピングされている場合、ペア関係の解除を契機として、仮想LDEVとプールLDEVの記憶領域とのマッピングを解消してプールLDEVの記憶領域をマッピング前の仮想ボリュームから解放する。

30

【0049】

図14は、チャンネル制御部110におけるボリュームペア制御プログラムとチャンネルコマンド制御プログラムとの関係を示すブロック図である。ボリュームペア制御プログラムは、チャンネルコマンド制御プログラムに対してコマンドを発行し、LU間のコピーを制御する。チャンネル制御部は、ペアボリュームの設定、解除、ステータスの変更などの制御を管理端末からの指示で行い、又は、上位装置がチャンネルI/Fを経由して発行するコマンドに依る指示によって行うこともできる。チャンネル制御部はボリュームペア制御プログラムからのペアボリューム解除の通知を受けてプール領域(実ボリューム)の記憶領域を解放する。

40

【0050】

図15は、ボリュームペア管理テーブルである。ペア関係にある正副ボリュームが、それぞれチャンネル番号とLUNの組で保持される。ボリュームペア制御プログラムはペア関係が設定されると、そのペアに対応するレコードを生成し、ペア関係が解除されると該レ

50

コードを削除する。ステータスが"P"である場合、ボリュームペアは「ペア状態」にある。同様に"S"である場合は「サスペンド状態」、"C"である場合は「コピー状態」である。ペア状態(P)は、正ボリュームで内容が更新されると、それが即座に副ボリュームにコピーされ、正副ボリュームの内容が一致していることを示す。サスペンド状態(S)は、この更新コピー処理が停止している状態を示す。コピー状態(C)は、サスペンド状態からペア状態に移行する間、正副ボリュームの内容の差分をコピー中であることを示す。ボリュームの内容の一致状態はボリュームペア毎にビット配列に保持され、ボリュームペア管理テーブルにはビット配列へのポインタが保持される。ビット配列の先頭をアドレス0とし、以降アドレス順にビットが"1"ならばアドレスの記憶領域は一致、"0"ならば不一致である。このテーブルはチャンネル制御部のプログラムがアクセス可能なメモリ空間に保持される。

10

#### 【0051】

図16はペアボリュームのペア解除に応じた解放の動作を示すフローチャートである。チャンネル制御部は、ボリュームペア制御プログラムとチャンネルコマンド制御プログラムに基づいて、図16の処理を実行する。チャンネル制御部は上位装置からペアボリュームのペアを解除する解除指示を受領する(1600)。チャンネル制御部は、副ボリュームのLUが仮想ボリュームLDEVにマッピングされているかをチェックする(1602)。この判定が肯定されると、ボリュームペア制御プログラムは、仮想LDEVの全領域のアドレスについて解放をチャンネルコマンド制御プログラムについて指示する(1604)。

#### 【0052】

次いで、ボリュームペア管理テーブル(図15)からボリュームペアのレコードを削除する(1606)。チャンネルコマンド制御プログラムは、ボリュームペア制御プログラムから解放指示を受け取ると、指示された仮想LDEVのアドレスについてマッピングを解放する。したがって、この記憶領域に関するマッピングを解放することにより、記憶領域の利用効率が向上する。

20

#### 【0053】

次に解放処理の他の実施形態について説明する。図17は、AOU技術において、ストレージシステムがホストから仮想ボリュームにアクセスしていない仮想ボリュームの領域に対して、事前にプールLDEVの記憶領域をマッピングするという「投機的割当」について説明したフローチャートである。チャンネル制御部がホストから受領したアクセスコマンドに応じて仮想LDEVとプールLDEVのマッピングを生成する場合、該アクセスコマンドのアクセス先アドレスに続いて他のアドレスにもアクセスがあるという実情とに基づいて、このアクセスが予測される記憶領域、例えば、ホストからアクセスがあった領域の直後のアドレスに対して(ホストからのアクセスが予測したアドレスにあるか否かに拘わらず)マッピングを実行する。これは一般に、ホストからのアクセスはLUの特定の領域に集中するという予測を基にしている。但し、ホストから仮想ボリュームに対するアクセスのパターンによっては次にアクセスされる領域の予測内容は異なることがあり得る。ここでは、投機的割当を説明するに際して、次にホストからアクセスがあるであろうアドレスの予測の一例として、ホストによって特定されたアドレスに続く幾分かのアドレス範囲、特に「直後」を例示したが、この他の予測、例えばある種のファイルシステムではLUを「シリンダグループ」という単位に分割して利用するので、あるアドレスにアクセスがあった場合はそのアドレスが属するシリンダグループに対して集中的に(不連続なアドレスであっても)割当を行なう、という「予測」の仕方も考えられる。これら予測した領域に上位装置であるホスト計算機からのアクセスが無い場合に、予測した領域に対する仮想ボリュームへの割当を解放することで記憶領域を有効利用できる。

30

40

#### 【0054】

図17において、チャンネル制御部がホストからのアクセスコマンドを受領すると(1700)、アクセス対象のLUに仮想ボリュームが割り当てられているか、即ち、仮想ビットが"ON"か否かを判定する(1702)。これが否定判定されると、チャンネル制御部はホストからのアクセスが実ボリュームに対してあったと判定し、ホストからのアクセスを

50

マッピング管理テーブルに基づいて対応する番号のLDEVへのアクセスにマッピングする(1708)。ステップ1702の判定において肯定判定されると、ホストからのLUに対するアクセス先アドレスを、マッピング管理用の、仮想ボリュームのアドレス(a1)に変換する(1704)。さらに、ホストからのLUに対するアクセス先アドレスの直後のアドレスを仮想ボリュームのアドレス(a2)に変換する(1706)。

**【0055】**

次いで、チャンネル制御部は、アドレスA(a1又はa2)はマッピング管理テーブル(図8)に未登録か否かをチェックする(1709)。これが否定された場合、即ち、アドレスa1が登録されている場合、マッピング管理テーブルに基づいて、チャンネル制御部はアドレスa1を対応するLDEV番号によって特定されるLDEVおよび実アドレスへマッピングする(1712)。ステップ1709において、肯定判定された場合は、プール管理テーブル(図7)からプールLDEV番号リストを取得する(1710)。チャンネル制御部はプールLDEVリストのX番目(1番目)のLDEVを取得する(1714)。チャンネル制御部は、X番目のLDEVが存在するか否かをチェックする(1716)。これが肯定判定された場合、チャンネル制御部は、プールLDEV管理テーブルから該LDEVのビット配列を取得する(1718)。次いで、チャンネル制御部はビット配列に"0"のビット配列があるか否かをチェックする(1724)。

10

**【0056】**

この判定が否定された場合、ステップ1714に移行して、チャンネル制御部は(X+1)番目のLDEVを取得する。この判定が肯定された場合には、"0"のビットを"1"に置換する(1726)。次いで、チャンネル制御部は、マッピング管理テーブル(図8)に該ビットに対応するアドレスを実アドレスとして登録する(1728)。ステップ1730において、仮想ボリュームのアドレス(A)がa1であるか否かが判定され、これが肯定された場合は、チャンネル制御部はマッピング管理テーブル(図8)のアクセスビットを"1"に設定する(1734)。これが否定判定された場合、チャンネル制御部はアクセスビット"0"に設定する(1734)。アクセスビット"0"は、既述のとおりマッピングがホストから仮想ボリュームへのアクセスに依らないこと、すなわち"投機的割当"であることを示している。

20

**【0057】**

次いで、チャンネルアダプタは、仮想ボリュームに対するホストのアクセスをプールLDEV番号・実アドレスへのアクセスにマッピングする。ステップ1738において、ホストのアクセス先アドレスがa1である場合(肯定判定)の、アドレスがa2に置き換えられた割当処理が繰り返される。すなわち、ステップ1709に処理がリターンして既述の各処理が繰り返される。ステップ1716において、X番目のLDEVが存在しない場合には、ステップ1720に処理が移行する。ここでは、ホストのアクセス先アドレスがa1であるか否かが判定され、アクセス先アドレスがa1の場合、チャンネル制御部はプールLDEV不足エラーを判定する(1722)。一方、アクセス先がa2の場合は、ホストからのアクセスが現実にあるわけではないのでこのLDEV不足エラー処理を行わず処理を終了する。

30

**【0058】**

以上説明した処理によれば、ホストから仮想ボリュームにアクセスがあることに依り、ホストのアクセス先アドレスにプールに属する記憶領域が割り当てられると共に、ホストがアクセスするであろう、仮想ボリュームのアクセス先をチャンネル制御部が予測して、この予測アドレスに記憶領域をあらかじめマッピングしておく。したがって、ホストが予測したアドレスに実際にアクセスした場合には、マッピングの処理を経ることなく、マッピングされた記憶領域に書き込みコマンドを直ちに実行することができる。

40

**【0059】**

図18のフローチャートに示すように、アクセスビットが"0"の記憶領域にホストがアクセスした場合、アクセスビットは"1"に変更される。チャンネル制御部は、ホストからのアクセス先アドレスをマッピング管理用のアドレスに変更する(1800)。チャンネル制

50

御部はマッピング管理テーブルから L D E V 番号・実アドレスを取得する ( 1 8 0 2 )。チャンネル制御部は図 8 に示す制御テーブルのアクセスビットが " 0 " であるか否かを判定して ( 1 8 0 4 )、これが肯定判定された場合は、アクセスビットを " 1 " に設定する ( 1 8 0 6 )。

#### 【 0 0 6 0 】

図 1 9 は、アクセスビットが " 0 " の記憶領域について、マッピングを解放するフローチャートである。アクセスビットが " 0 " のまま残っている、L D E V の記憶領域はホストからのアクセスが当面期待し得ない余剰領域のまま放置されることになる。これでは記憶資源を有効活用されない。そこで、この余剰領域の整理を目的として、余剰領域についてのマッピングを解放することとした。ステップ 1 9 0 0 において、チャンネル制御部は、マッピング管理テーブル ( 図 8 ) の 1 番目のレコードを取得し、該レコードについてのアクセスビットが " 0 " であるか否かをチェックする ( 1 9 0 2 )。これが肯定判定された場合は、該レコードのプール L D E V 番号と実アドレスを取得した上で ( 1 9 0 3 )、このレコードを削除する ( 1 9 0 4 )。次いで、プール L D E V 管理テーブルの該プール L D E V について、該アドレスのビットを " 0 " にする ( 1 9 0 5 )。

10

#### 【 0 0 6 1 】

以上の処理は、最終番目のレコードまで処理が継続される ( 1 9 0 6 )。マッピング管理テーブルのレコードのうち、アクセスビットが " 0 " であるものは実際にはホストからアクセスされていない記憶領域である。このような余剰領域について、定期的にマッピング管理テーブルの全レコードを走査しアクセスビットに基づいてマッピング情報の削除を行う。アクセスビットは、ストレージシステムが解放処理を行う場合における、解放する、或いは解放しないを判定するための制御情報である。

20

#### 【 0 0 6 2 】

次に、マッピングを解放する第 3 の例について説明する。図 2 0 はこの例を説明するブロック図である。この例は、ストレージシステムがペアボリュームからバックアップ取得後にマッピングされた記憶領域を解放することを示している。L U 1 と L U 2 とがペアボリュームになっている。コピー元 L U ( 正ボリューム ) がマッピングされるのは通常の L D E V でも、仮想 L D E V でもよい。コピー元 L U 3 は実記憶領域を備える L D E V に直接マッピングされている。コピー元 L U 2 は仮想 L D E V 2 にマッピングされている。ストレージシステムは、定期的にコピー ( 再同期 ) を行い、ボリューム内容のスナップショットをコピー先 ( 副ボリューム ) L U 1 に作成、そこからテープデバイスにデータバックアップを取得する。

30

#### 【 0 0 6 3 】

テープデバイス 2 1 0 はホスト 2 0 0 A に接続されていてもよいし、テープデバイス 2 1 2 がストレージに接続されて、ホストからの EXTENDED COPY 命令によりストレージシステムがテープデバイス 2 1 2 へのバックアップを制御しても良い。この再同期、バックアップ取得という一連の処理が終了すると、副ボリューム L U 1 の記憶内容は不要になる。この副ボリュームが仮想 L D E V にマッピングされているので、それに割り当てられたプール L D E V の記憶領域のマッピングを解消して記憶領域を解放する。上記の再同期、バックアップ取得、領域解放という一連の処理は、ホスト上で動作するバックアップアプリケーション A P によって行われる。そのため、アプリケーションからストレージサブシステムに対して、領域解放の指示を I / O パスやネットワークを通じて送出する。

40

#### 【 0 0 6 4 】

図 2 1 に示すように、バックアップアプリケーションは、チャンネルコマンド制御プログラムとボリュームペア制御プログラムにコマンドを発行し、L U のスナップショット取得とテープデバイス 2 1 2 へのコピーを行う。ボリュームペア制御プログラムに対するコマンド発行はネットワークを経由して制御端末から行うこともできるし、チャンネル I / F を経由して行うこともできる。テープデバイスの制御はバックアップアプリケーション自身が行うこともできるし、チャンネルコマンド制御プログラムが制御するようコマンドで指示することもできる。

50

## 【 0 0 6 5 】

図 2 2 は、ペアボリュームからテープデバイスにバックアップを取得した後、記憶領域を解放する動作を示すフローチャートである。ホストのバックアップアプリケーションは、ペアボリューム間のコピーをボリュームペア制御プログラムに対して指示する(2200, 2214)。ボリュームペア制御プログラムはボリュームペアのステータスをチェックする(2216)。ステップ 2216 が肯定判定(ステータスが"P")の場合、は何もせずバックアップアプリケーションの指示を待つ。ステップ 2216 において否定判定の場合は、チャンネルコマンド制御プログラムにペアボリューム間のコピーを指示する(2218)。このときステータスは"C"に変更される。ボリュームペア制御プログラムは、チャンネルコマンド制御プログラムに依るコピー完了を受けてステータスを"P"にする。

10

## 【 0 0 6 6 】

バックアップアプリケーションは、ペアステータスが"P"になるまでウエイトし(2202)、次いで、ペアボリュームのコピー停止を指示する(2204)。ボリュームペア制御プログラムは、ペアボリュームのコピー停止を受けてステータスを"S"に変更する(2226)。バックアップアプリケーションは副ボリュームからテープへのコピーを指示する(2206)。この指示を受けてチャンネルコマンド制御プログラムは副ボリュームからテープへコピーを実行する(2230)。バックアップアプリケーションはコピー完了までウエイトし(2208)、副ボリュームがアクセスするLUは仮想LDEVかを判定する。これが肯定判定された場合は、バックアップアプリケーションは仮想ボリュームLDEVの全領域のアドレスについてマッピングを解放することをチャンネルコマンド制御プログラムに指示する。

20

## 【 0 0 6 7 】

チャンネルコマンド制御プログラムは、コピー完了と解放指示を受けて、仮想LDEVの全領域のアドレスについて、このアドレスにマッピングされた、プール領域のLDEVの記憶領域を仮想ボリュームへのマッピングから解放する(2232)。以上の処理によって、ボリュームペアの関係にあるボリューム間でコピーが完了する過程で、副ボリュームに対応する仮想ボリュームの記憶領域が解放される。

## 【 0 0 6 8 】

次に、記憶領域の解放に関する他の実施形態について図 2 3 を利用して説明する。この実施形態は、ホスト 2 0 0 上で動作するアプリケーションが削除した記憶領域に関するマッピングを解放するというものである。ここでは、アプリケーションソフトウェアの一例としてDBMS(データベースマネジメントシステム)を取り上げる。DBMSはLUの記憶領域に対して、表領域、索引、ログといったデータの読み書きを行う。これらのデータはいずれも、あるサイズの領域(セグメント)を占有し、その占有する領域内に論理的な構造が形成されている。このセグメントは、さらに細かい領域(データブロック)に分割されている。DBMSはその動作中、LU内のセグメントの論理的構造に基づいて、データを記録・更新・削除する。このため、DBMSはセグメントS1のあるデータブロックS2が未使用であるか、使用中であるか、という情報を管理する表を維持している。この表をフリーリストS3と呼ぶ。これらデータブロックのあるLUが仮想LDEVに対応する場合、ホスト(DBMS)からアクセスがあるとマッピングが生成される。この例ではデータブロックのサイズの単位と、マッピングされる領域のサイズの単位が等しいが、この二つのサイズの関係はこれに限定されるものではない。

30

40

## 【 0 0 6 9 】

図 2 4 は、DBMSとチャンネルコマンド制御プログラムとの関係を示すブロック図である。DBMSは、チャンネルコマンド制御プログラムにコマンドを発行しLUにアクセスする。ホストは、記憶領域のマッピングを削除、解放するようチャンネル制御部に指示する。この指示のコマンド発行はネットワークを経由してSVPから行うこともできるし、チャンネルI/Fを経由して行うこともできる。図 2 5 は、フリーリストを示すテーブルである。フリーリストはセグメント毎に存在する。データブロック番号で特定される各データブロックについて、使用ビットが"1"である場合は使用中、"0"である場合は未使用を示

50

す。

【0070】

図26は、DBMSによるデータの記録を実行するフローチャートである。図26において、DBMSはフリーリストを走査し(2600)、使用ビットが"0"のデータブロックがあるか否かをチェックする(2602)。この判定が肯定された場合には、チャンネル制御部はフリーリストのこのデータブロックについて使用ビットを"1"に設定する(2604)。次いで、当該データブロックにデータを書き込む(2606)。ステップ2604において、否定判定されるとデータブロック不足エラー処理を行う(2608)。

【0071】

図27は、DBMSによるデータの削除を実行するフローチャートである。DBMSはフリーリストにおいて削除対象データブロックの使用ビットを"0"に設定する(2700)。削除対象データブロックのアドレスを指定し領域解放コマンドを発行する。

10

【0072】

図28は、チャンネル制御部がフリーリストを参照して、記憶領域の解放を実行するためのフローチャートである。チャンネル制御部は、フリーリストから使用ビットが"0"であるデータブロックを抽出する(2800)。そして、該データブロックのアドレスをマッピング管理用の、仮想ボリュームのアドレスに変換する(2802)。次いで、このアドレスが図8のマッピング管理テーブルに登録されているか否かをチェックする(2804)。これが肯定判定された場合には、このアドレスに対応したプールLDEVの記憶領域を解放する(2806)。チャンネルコマンド制御プログラムは、図28の処理を定期的に

20

【0073】

図29は、図28の変形例に係わるもので、ホストからチャンネル制御部に対するI/Oを監視するプログラムによって、記憶領域を解放するための、プログラムの動作に係わるブロック図である。ホストのI/O監視プログラムは、DBMSから発行されるアクセスコマンドを監視し、フリーリストのあるレコードについて使用ビットを"0"に設定するアクセスを認識した場合、チャンネルコマンド制御プログラムに対してそのレコードが示すアドレスのマッピングを削除、解放するよう指示する。この指示のコマンド発行はネットワークを経由してSVPから行うこともできるし、チャンネルI/Fを経由して行うこともできる。図30に示すように、I/O監視プログラムをチャンネル制御部に実装しても良い。

30

【0074】

図31は、図29の動作を示すフローチャートである。I/O監視プログラムはホストからのアクセスコマンドを検知して(3100)、このアクセスコマンドがフリーリストへの書き込みか否かをチェックする(3102)。このチェックが肯定された場合は、あるデータブロックの使用ビットを"0"に設定するようなコマンドであるかどうかをチェックする(3104)。この判定が肯定判定されると、この使用ビットに対応するデータブロックのアドレスをマッピング管理用アドレスに変換する(3106)。次いで、アドレスがマッピング管理用テーブルに登録されているか否かをチェックして、これが肯定判定されるとこのアドレスに対応する記憶領域を解放する。

【0075】

以上説明した記憶領域の解放処理において、解放処理を直ちに実行するのではなく、マッピング管理テーブルの解放すべき実アドレスに対してアクセスビットを"0"にまず設定し、然るべき期間経過の後図19の解放処理を実行するようにしても良い。これにより解放処理を遅延させることができる。また、図32のフローチャートのようにしても良い。即ち、チャンネル制御部は、プールLDEV管理テーブルから全プールLDEVのビット配列を取得する(3000)。次いで、全ビットのうち"0"となっているビットの割合を算出して(3002)、ある値よりこれが多いか否かを判定する(3004)。この判定が否定された場合は、記憶領域が不足しつつあるとして、チャンネル制御部はマッピング管理テーブルを走査して、図19の処理に基づいて解放処理を行う(3006)。

40

【0076】

50

図33は、解放する記憶領域のデータを消去する処理を示すフローチャートである。チャンネル制御部は、解放する記憶領域のアドレスを決定する(3300)。次いで、マッピング管理テーブルから、該アドレスに対応するプールLDEV番号・実アドレスを取得する(3302)。次いでマッピング管理テーブルの該レコードを削除する(3304)。チャンネル制御部はプールLDEV管理テーブルの該プールLDEVについて該アドレスのビットを"0"に置換する(3306)。次いで、該プールLDEVの該アドレスの記憶領域に特定データパターンを書き込む(3308)。

【0077】

特定のデータパターンの書き込みの例としては、全ビットを"0"にする、ランダムなビット列にする、またこれらを複数回繰り返すことがある。このデータ消去によって、該記憶領域に再びマッピングが生成される際に、もともと記録されていたデータが無関係な他のホストやアプリケーションによって読み出されるといったセキュリティ上のリスクを低減することができる。

10

【0078】

なお、プールLDEVとして、他のストレージシステムの外部記憶資源を利用することもできる。この場合、第1のストレージシステムの仮想ボリュームに第2のストレージシステムの記憶資源が割り当てられる。上位装置は第1の仮想ボリュームにアクセスすることによって、第2のストレージシステムの記憶資源にアクセスすることができる。

【0079】

図34は、ストレージシステム600に他のストレージシステム600Aがストレージシステム600の外部から接続されている状態を示している。なお、図34を図2と対比して参照されたい。図34において、501はネットワークであり、502はI/Oパスであり、503AはチャンネルI/F(エクスターナルモード)、503B, 503CはチャンネルI/F(ターゲットモード)、301はハードディスクユニット、601は外部ストレージサブシステムの内の外部LUを示す。

20

【0080】

チャンネルI/F(ターゲットモード)とは、ホストがストレージサブシステムにアクセスするインタフェースであり、チャンネルI/F(エクスターナルモード)とは、ストレージサブシステムが外部ストレージサブシステムにアクセスするインタフェースであり、外部LUとは、外部ストレージサブシステム内に定義されたボリューム(LU)の一群をいう。

30

【0081】

図35は、ホストが、ストレージシステム600の仮想ボリュームを介して外部LUをストレージシステム600のボリュームとして認識している状態を示すブロック図である。12Dは外部ストレージシステム600A内に定義されて設けられた外部LDEVである。外部LDEV12DはLU10に割り当てられている。12Eは外部LDEVであり、ストレージシステム600のLDEV12B(12C)と同様にプール18を構成している。ストレージシステム600の仮想LDEV16はプール18をマッピングに使用している。

【0082】

外部LDEVとは、チャンネルコマンド制御プログラムが、エクスターナルモードのチャンネルI/Fを経由して認識した外部ストレージサブシステムのLUを、仮想的にLDEVと同等の記憶領域としたものである。通常の(RAID制御プログラムによって提供される)LDEVと同様に、LUにマップして上位ホストからのアクセスを受け付けることもできるし、プールの一員として仮想LDEVとのマッピングに使用することもできる。プールを構成するLDEVは、通常のLDEVのみであっても外部LDEVのみであってもよいし、これらが混在してもよい。外部LDEVは、LDEVの一覧を示すテーブル(図6に示す)の上では通常のLDEVと同等に扱われる。

40

【図面の簡単な説明】

【0083】

50

【図1】本発明が適用されるストレージシステムを含んだ、記憶制御システムを示すハードウェアブロック図である。

【図2】図1に示す記憶制御システムの機能ブロック図である。

【図3】仮想ボリュームとプールとの関係を示すブロック図である。

【図4】LU管理テーブルである。

【図5】仮想LDEV管理テーブルである。

【図6】RAID制御プログラムが、チャンネル制御部及び上位装置に対して提供するLDEVの一覧を示すテーブルである。

【図7】プール管理テーブルである。

【図8】マッピング管理テーブルであり

10

【図9】図8の制御テーブルで示された、仮想LDEVのアドレスとプールに属するLDEVの実記憶領域との対応関係を示すブロック図である。

【図10】プールLDEVの記憶領域が仮想LDEVの領域にマッピングされている状態を示すプールLDEV管理テーブルである。

【図11】ストレージシステムのAOU処理動作を示すフローチャートである。

【図12】チャンネル制御部がチャンネルコマンド制御プログラムに基づいて実行するフローチャートである。

【図13】ボリュームペアを説明するブロック図である。

【図14】チャンネル制御部におけるボリュームペア制御プログラムとチャンネルコマンド制御プログラムとの関係を示すブロック図である。

20

【図15】ボリュームペア管理テーブルである。

【図16】ペアボリュームのペア解除に応じた解放の動作を示すフローチャートである。

【図17】"投機的割当"について説明したフローチャートである。

【図18】アクセスビットに対する処理動作を説明するフローチャートである。

【図19】アクセスビットが"0"の記憶領域について、マッピングを解放するフローチャートである。

【図20】コピーペアのボリュームを用いたバックアップ動作を示す、機能ブロック図である。

【図21】バックアップアプリケーションとチャンネル制御部との動作関係を示す機能ブロック図である。

30

【図22】バックアップの動作を示すフローチャートである。

【図23】データベースマネジメントシステム(DBMS)に関するアプリケーションプログラムの動作を示す、ストレージシステムの機能ブロック図である。

【図24】DBMSとチャンネルコマンド制御プログラムとの関係を示すブロック図である。

【図25】フリーリストを示すテーブルである。

【図26】DBMSによるデータの記録を実行するフローチャートである。

【図27】DBMSによるデータの削除を実行するフローチャートである。

【図28】チャンネル制御部がフリーリストを参照して、記憶領域の解放を実行するためのフローチャートである。

40

【図29】ホストからチャンネル制御部に対するI/Oを監視するプログラムによって、記憶領域を解放するための、プログラムの動作に係わるブロック図である。

【図30】図29の変形例である。

【図31】図29に係わるシステムの動作を示すフローチャートである。

【図32】プールLDEV管理テーブルのビット配列に基づいて解放を実行するためのフローチャートである。

【図33】解放される記憶領域のデータを消去するためのフローチャートである。

【図34】ストレージシステムに他のストレージシステムが外部から接続しているシステムを示すブロック図である。

【図35】図34のシステムにおける、ストレージシステム間のボリュームの対応関係を

50

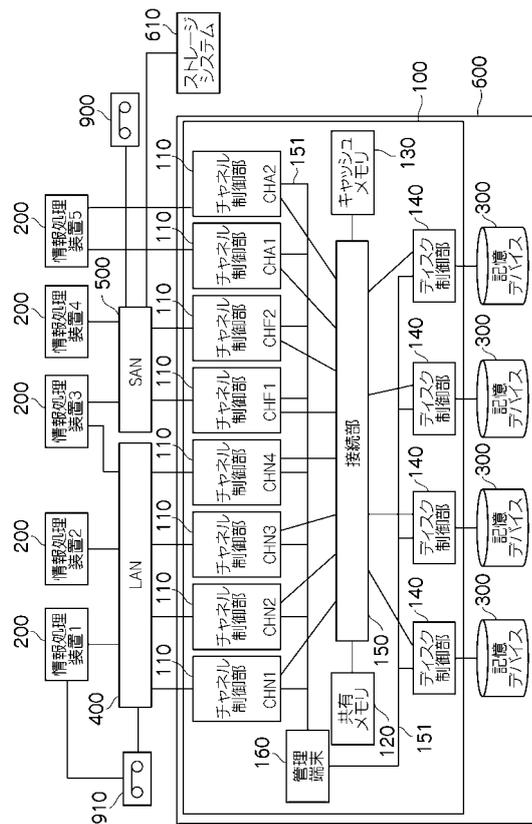
示すブロック図である。

【符号の説明】

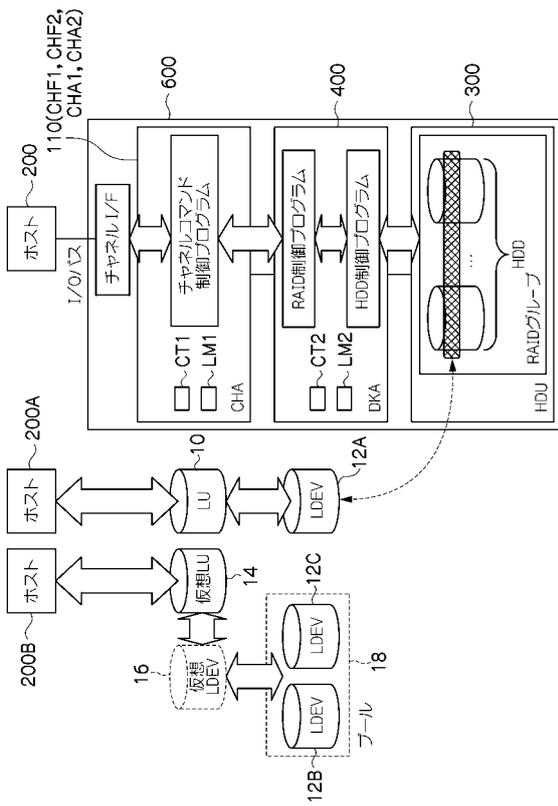
【0084】

100 コントローラ、110 チャネル制御部、140 ディスク制御部、200 上位装置、600 ストレージシステム

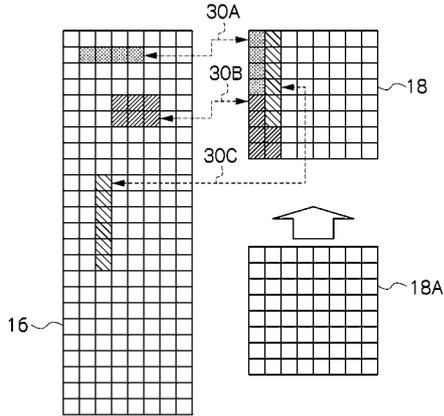
【図1】



【図2】



【 図 3 】



【 図 4 】

通番	チャンネル番号	LUN	仮想ビット	LDEV番号
0	1A-0	0	OFF	0001
1	1A-0	1	ON	0001#
2	1B-0	0	OFF	0002

【 図 5 】

通番	仮想LDEV番号	プール番号
0	0001#	01
1	0002#	02

【 図 6 】

通番	LDEV番号	用途シジスタ
0	0001	L
1	0002	L
2	0003	P
3	0004	P
4	0005	P
5	0006	N

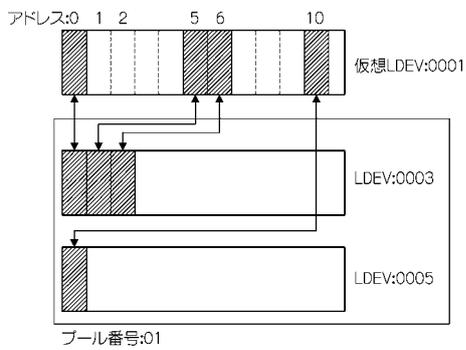
【 図 7 】

通番	プール番号	LDEV番号リスト
0	01	0003,0005
1	02	0004

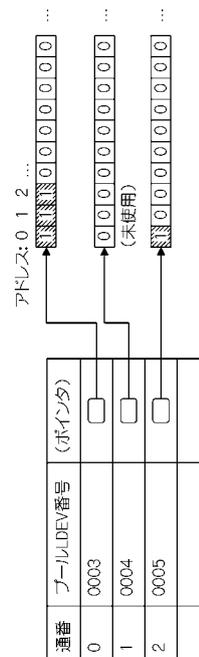
【 図 8 】

仮想LDEV番号:0001#				
通番	アドレス	プールLDEV番号	実アドレス	アクセスビット
0	0	0003	0	1
1	5	0003	1	1
2	6	0003	2	0
3	10	0005	0	1

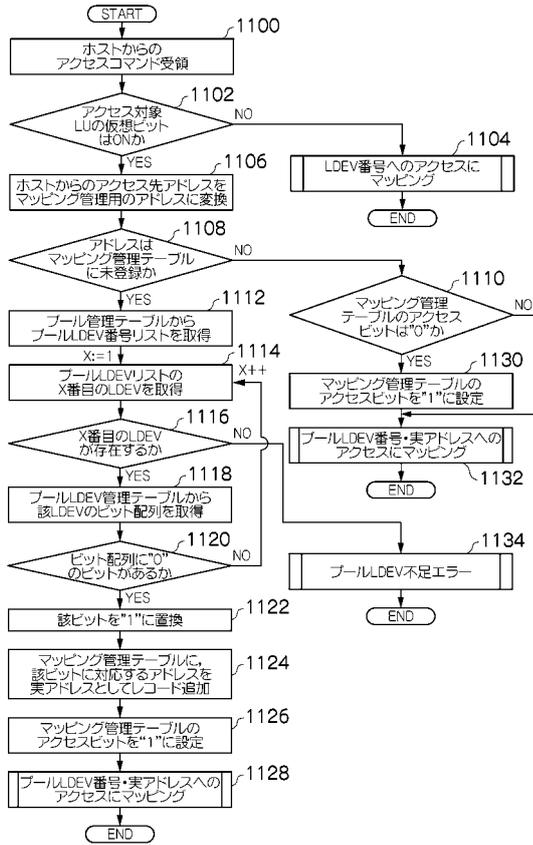
【 図 9 】



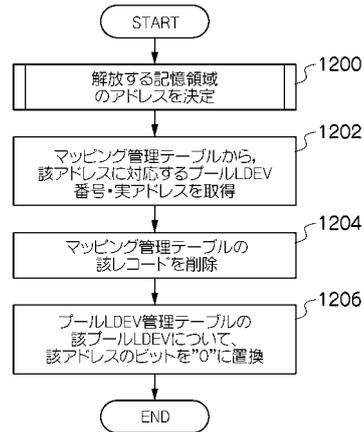
【 図 10 】



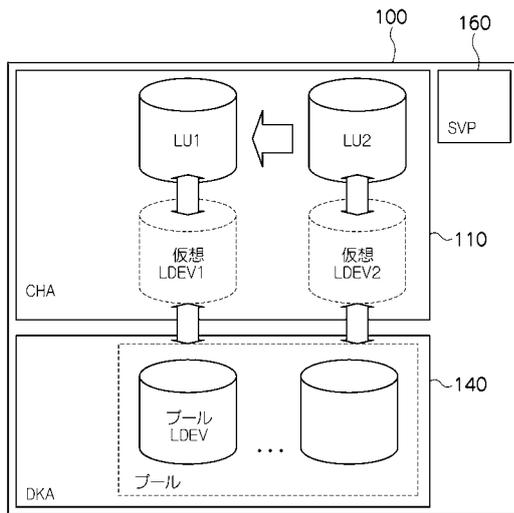
【 図 1 1 】



【 図 1 2 】



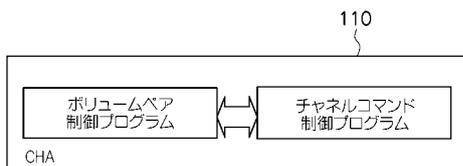
【 図 1 3 】



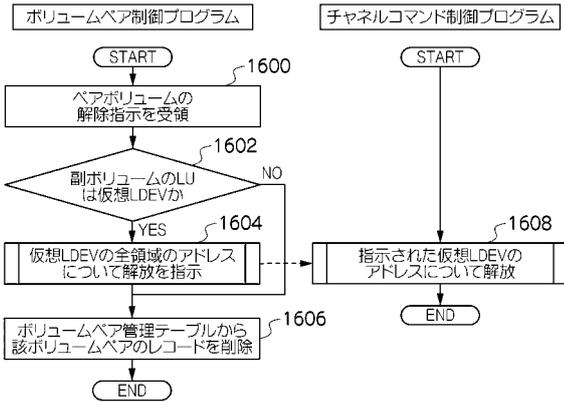
【 図 1 5 】



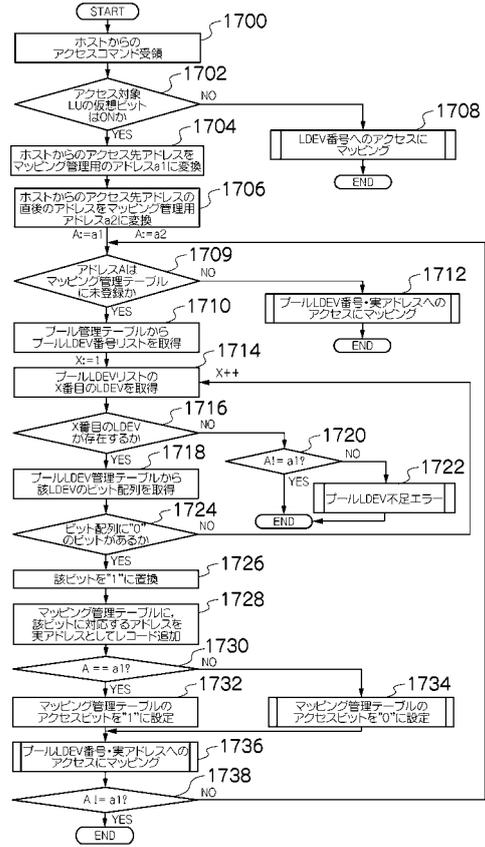
【 図 1 4 】



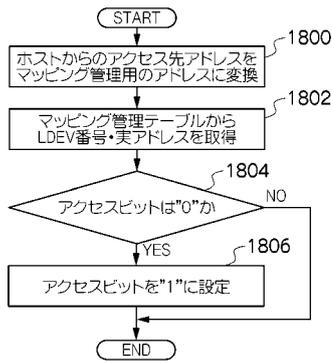
【図16】



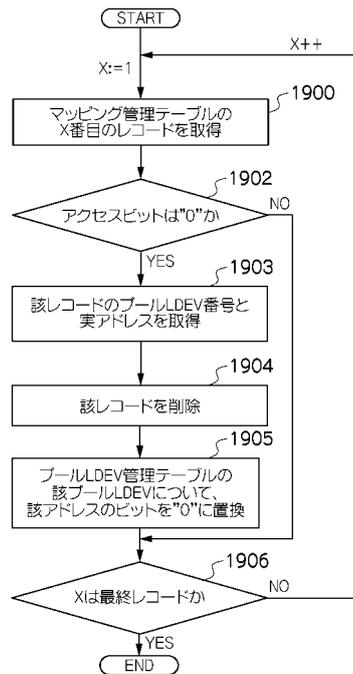
【図17】



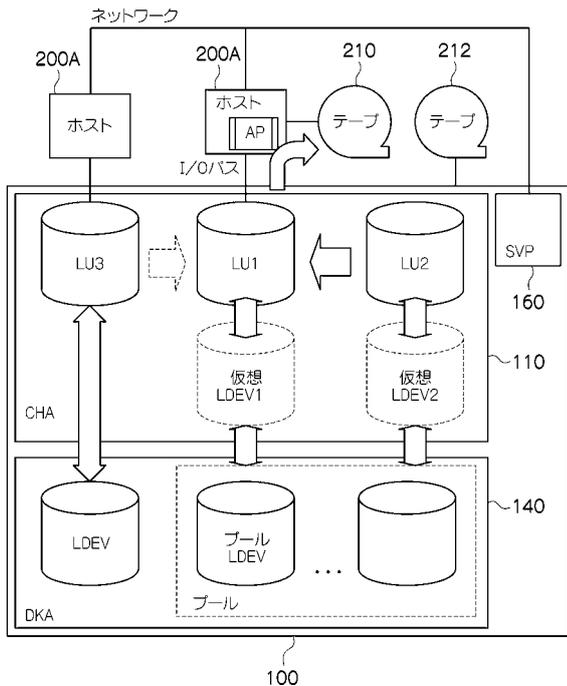
【図18】



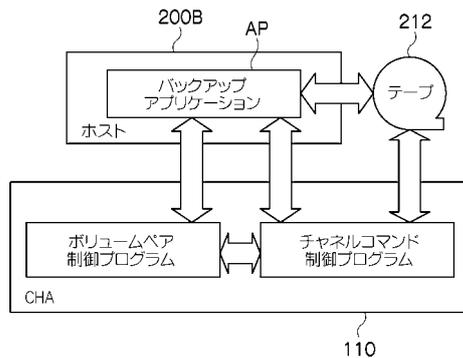
【図19】



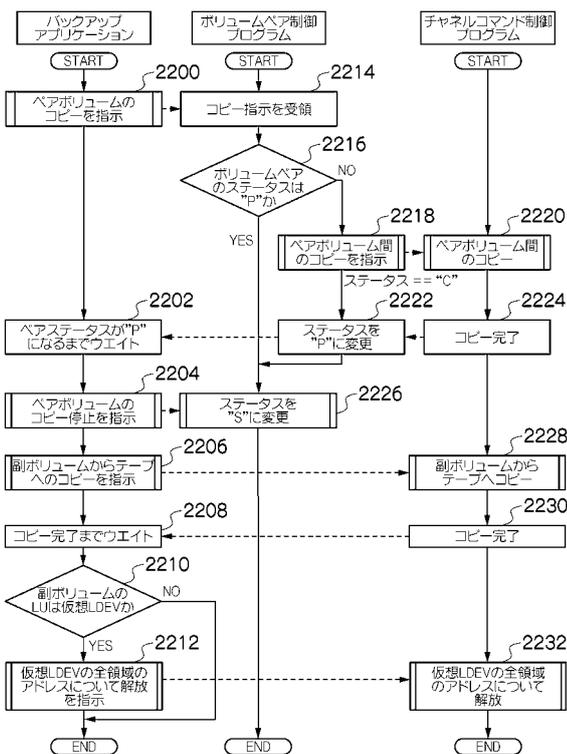
【図20】



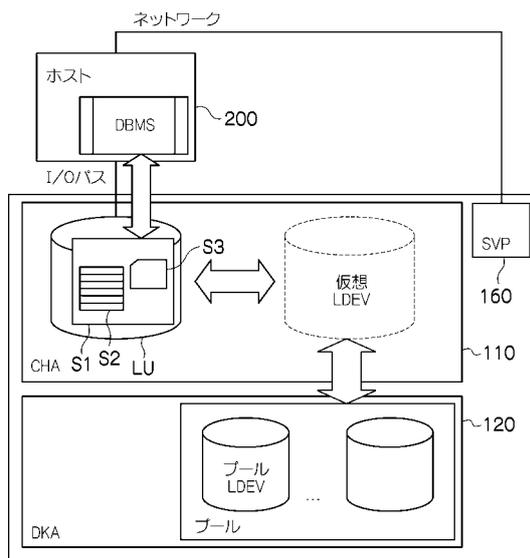
【図21】



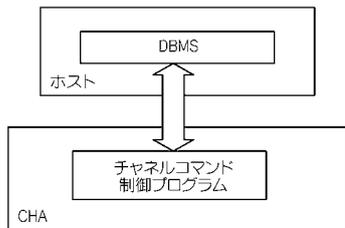
【図22】



【図23】



【 図 2 4 】

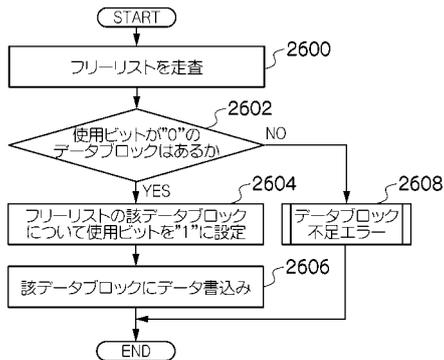


【 図 2 5 】

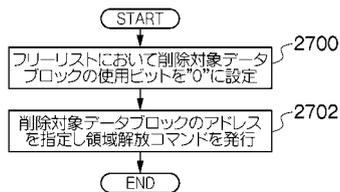
セグメント番号:01

通番	データブロック番号	アドレス	使用ビット
0	01	0	1
1	02	1	0

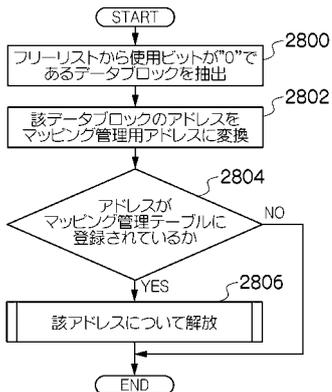
【 図 2 6 】



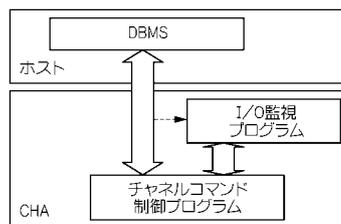
【 図 2 7 】



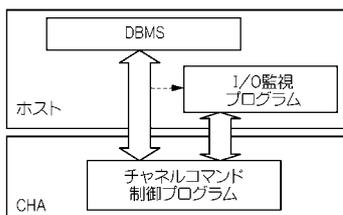
【 図 2 8 】



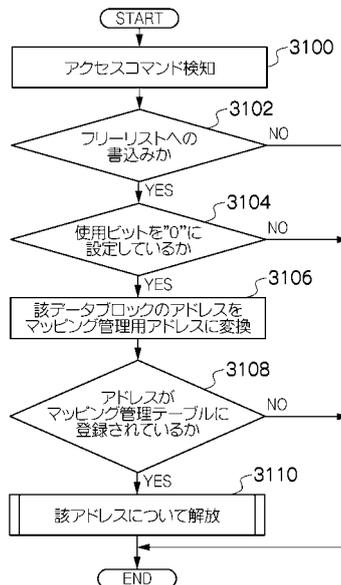
【 図 3 0 】



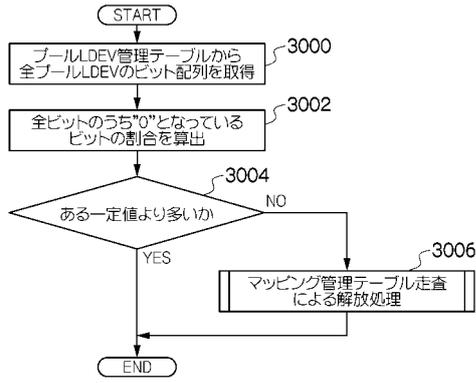
【 図 2 9 】



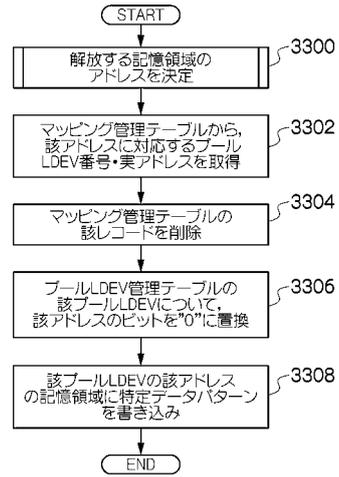
【 図 3 1 】



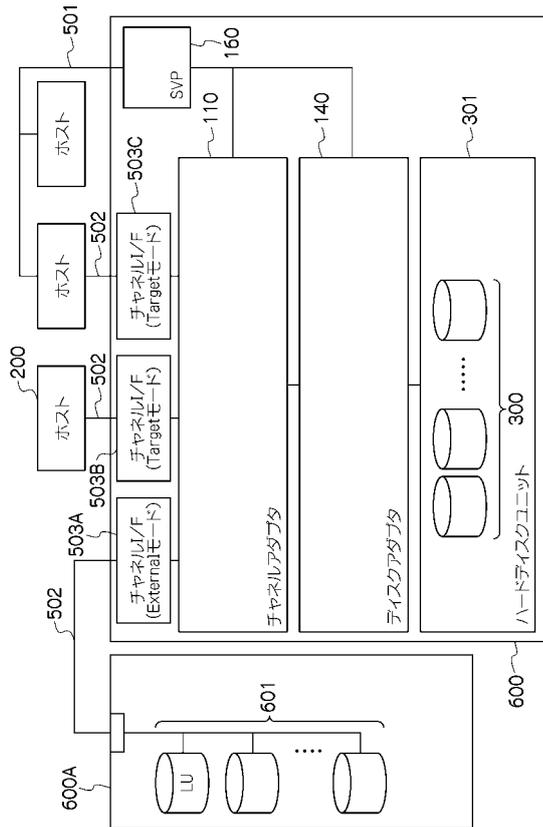
【 図 3 2 】



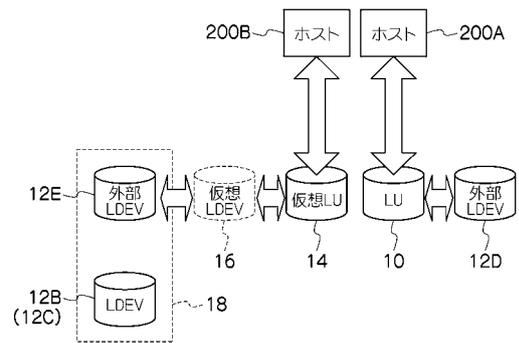
【 図 3 3 】



【 図 3 4 】



【 図 3 5 】



フロントページの続き

Fターム(参考) 5B065 BA01 CA12 CA30 CC03 CE22 CE24 CH20 EA12 EA33