(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5008845号
(P5008845)
(45)発行日　平成24年8月22日(2012.8.22)　(24)登録日　平成24年6月8日(2012.6.8)

(51)Int.Cl.　F I
G06F 3/06 (2006.01)　G06F 3/06 302A
G06F 3/06 540

請求項の数 3　(全 22 頁)

(21)出願番号　特願2005-252989(P2005-252989)
(22)出願日　平成17年9月1日(2005.9.1)
(65)公開番号　特開2007-66129(P2007-66129A)
(43)公開日　平成19年3月15日(2007.3.15)
審査請求日　平成20年1月9日(2008.1.9)

(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　110000062
特許業務法人第一国際特許事務所
(72)発明者　中村　崇仁
神奈川県川崎市麻生区王禅寺１０９９番地
株式会社　日立製作所　システム開発研究所内
(72)発明者　藤本　和久
神奈川県川崎市麻生区王禅寺１０９９番地
株式会社　日立製作所　システム開発研究所内

最終頁に続く

(54)【発明の名称】ストレージシステムとストレージ装置及びその制御方法

(57)【特許請求の範囲】
【請求項１】
ホストコンピュータからのデータを保存する１以上のメディアと、前記メディアを制御するメディア制御部、チャネルを介して前記ホストコンピュータに接続するチャネル制御部、及び前記ホストコンピュータからのデータを一時的に保存する揮発メモリであるキャッシュメモリ部を備えるストレージ装置を具備するストレージシステムにおいて、
前記メディアは、ＨＤＤと、フラッシュメモリとからなり、
前記ホストコンピュータからのリードの要求を受けた際に、格納先が前記ＨＤＤか前記フラッシュメモリかを判別し、格納先が前記ＨＤＤの場合には前記ＨＤＤから読み出したデータを前記キャッシュメモリ部に格納した後に前記ホストコンピュータに応答し、格納先が前記フラッシュメモリの場合には前記フラッシュメモリから読み出したデータを前記キャッシュメモリ部には格納せずに前記ホストコンピュータに応答し、
前記ストレージ装置は、前記フラッシュメモリに対する書き込み回数を平均化するように、デステージするデータを選択し、各々のフラッシュメモリに対するデステージした回数を記録し、デステージするデータを選択する際に、各々の前記フラッシュメモリに対するデステージした回数を比較し、デステージした回数が少ないものから優先的にデステージするデータを決定することを特徴とするストレージシステム。
【請求項２】
ホストコンピュータからのデータを保存する１以上のメディアと、前記メディアを制御するメディア制御部、チャネルを介して前記ホストコンピュータに接続するチャネル制御

部、及び前記ホストコンピュータからのデータを一時的に保存する揮発メモリであるキャッシュメモリ部を備えるストレージ装置を具備するストレージシステムにおいて、
前記メディアは、ＨＤＤと、フラッシュメモリとからなり、
前記ホストコンピュータからのリードの要求を受けた際に、格納先が前記ＨＤＤか前記フラッシュメモリかを判別し、格納先が前記ＨＤＤの場合には前記ＨＤＤから読み出したデータを前記キャッシュメモリ部に格納した後に前記ホストコンピュータに応答し、格納先が前記フラッシュメモリの場合には前記フラッシュメモリから読み出したデータを前記キャッシュメモリ部には格納せずに前記ホストコンピュータに応答し、
前記ストレージ装置は、前記フラッシュメモリに対する書き込み回数を平均化するように、デステージするデータを選択し、デステージするデータを選択する際に、各々の前記フラッシュメモリに対するデステージした回数およびアクセス回数を基に評価関数を算出し、評価関数が少ないものから優先的にデステージするデータを決定することを特徴とするストレージシステム。

【請求項３】
ホストコンピュータからのデータを保存するメディアを制御するメディア制御部、チャネルを介して前記ホストコンピュータに接続するチャネル制御部、及び前記ホストコンピュータからのデータを一時的に保存する揮発メモリであるキャッシュメモリ部を備えるストレージ装置を制御する方法において、
前記メディアは、ＨＤＤと、フラッシュメモリとからなり、
前記ホストコンピュータからのリードの要求を受けた際に、格納先が前記ＨＤＤか前記フラッシュメモリかを判別し、格納先が前記ＨＤＤの場合には前記ＨＤＤから読み出したデータを前記キャッシュメモリ部に格納した後に前記ホストコンピュータに応答し、格納先が前記フラッシュメモリの場合には前記フラッシュメモリから読み出したデータを前記キャッシュメモリ部には格納せずに前記ホストコンピュータに応答し、
前記フラッシュメモリに対する書き込み回数を平均化するように、デステージするデータを選択することを特徴とするストレージ装置の制御方法。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ストレージ装置に関し、また、その制御の方法に関するものである。

【背景技術】
【０００２】
近年、データセンタなどの情報ビジネスの現場において、ストレージシステムの総所有コスト（ＴＣＯ）の削減がますます重要になってきている。その一方で、データを長期的かつ確実に記録する要求も高まってきている。このことに関する例としては、金融機関および医療機関などの文書データは、消去せずに蓄積されることが法律によって義務付けられていることなどが挙げられる。こうした背景において、高信頼で大容量のストレージシステムが必要とされている。しかしながら、ハードディスクドライブ（以降では「ＨＤＤ」という）を用いた大規模なストレージシステムでは、一般に、ストレージ容量に比例して電力消費量が増大する。つまり、大容量のストレージシステムを所有することは、電気料金が含まれる総所有コストが上昇してしまうことを示している。こうした状況を鑑み、キャッシュ管理アルゴリズムにより、ＨＤＤの消費する電力を低減させる技術が提案されている（非特許文献１）。また、問題は、電気料金に限られたことではない。一般に、ストレージシステムの容量が大きくなるほど、設置するための床面積が増大する。このことも総所有コストの上昇につながってしまっている。
【０００３】
ところで、近年、不揮発メディアとして、フラッシュメモリが注目されている。フラッシュメモリは、一般にＨＤＤと比較し、数十倍以上低消費電力であり、高速に読み出しが可能である。また、ＨＤＤのように機械的な駆動部分が不要なため小型である。
【０００４】

しかしながら、フラッシュメモリには、情報を保持するセルの物理的な理由により、書き込み回数の制限がある。こうした制限に対し、上位装置に示すアドレスとセル位置との対応をもち、各セルに書き込まれる回数を均等化するように制御を行う、ウェアレベリングと呼ばれる技術などにより、フラッシュメモリの書き込み可能回数の向上が図られている。なお、以降では、情報を保持するための素子を単に「フラッシュメモリ」といい、上記のウェアレベリングや、上位装置に対するプロトコル処理などを行う機構を含めたものを「フラッシュメモリデバイス」という。こういった技術により、フラッシュメモリデバイスとしての書き込み回数制限に対する効率化が図られているものの、フラッシュメモリデバイスの書き込み回数制限は依然存在している。また、それに加えて、フラッシュメモリに書き込む際に、消去と呼ばれる操作が必要になる場合はＨＤＤ同等程度の速度となってしまうデメリットも存在している。

【０００５】

このようなフラッシュメモリを用いたストレージシステムを構成する技術としては、例えば特許文献１がある。特許文献１は、ＲＡＩＤ構成などにおいて頻繁にアクセスされるパリティデータをフラッシュメモリなどの半導体ディスクにおき、ストレージシステムの性能を向上させる技術を示している。しかし、ストレージシステムとしての書き込み回数制限を回避する手段は示されていない。また、ＨＤＤとフラッシュメモリを混在して１つのＲＡＩＤグループ、すなわち１つの仮想デバイスを構成するので、それぞれのメディアの特徴を考慮した仮想デバイスの制御はできていない。

【特許文献１】特開平６－３２４８１５号公報

【非特許文献１】ZHU, Q., DAVID, F., ZHOU, Y., DEVARAJ, C., AND CAO, P. ,”ReducingEnergy Consumption of Disk Storage Using Power-Aware Cache Management”. InProc. of the 10th Intl. Symp. on High Performance Computer Architecture(HPCA-10) (Feb. 2004).

【発明の開示】

【発明が解決しようとする課題】

【０００６】

以上のような背景から、低消費電力であり、設置面積が小さく、かつ、大容量を得ることのできる大規模まで構成可能なストレージシステムを提供することが課題となっている。

【０００７】

また、データを格納するメディアに応じた、高いシステム性能を提供することも課題となっている。

【０００８】

また、ストレージシステムとしての信頼性、可用性の向上も課題である。書き込み回数制限を有するメディアに対してはストレージシステムとして緩和を行わなくてはならない。

【課題を解決するための手段】

【０００９】

本発明は、ホストコンピュータからのデータを保存するメディアを制御するメディア制御部、チャネルを介して前記ホストコンピュータに接続するチャネル制御部、及び前記ホストコンピュータからのデータを一時的に保存する揮発メモリであるキャッシュメモリ部を備え、前記メディアは、ＨＤＤと、フラッシュメモリとからなり、前記ホストコンピュータからのリードの要求を受けた際に、格納先が前記ＨＤＤか前記フラッシュメモリかを判別し、格納先が前記ＨＤＤの場合には前記ＨＤＤから読み出したデータを前記キャッシュメモリ部に格納した後に前記ホストコンピュータに応答し、格納先が前記フラッシュメモリの場合には前記フラッシュメモリから読み出したデータを前記キャッシュメモリ部には格納せずに前記ホストコンピュータに応答し、前記フラッシュメモリに対する書き込み回数を平均化するように、デステージするデータを選択することを特徴とする。

【００１０】

すなわち、本発明は、ホストコンピュータからのデータを保存する１以上のメディアと、前記メディアを制御するメディア制御部、チャネルを介して前記ホストコンピュータに接続するチャネル制御部、及び前記ホストコンピュータからのデータを一時的に保存する揮発メモリであるキャッシュメモリ部を備えるストレージ装置を具備するストレージシステムにおいて、前記メディアは、ＨＤＤと、フラッシュメモリとからなり、前記ホストコンピュータからのリードの要求を受けた際に、格納先が前記ＨＤＤか前記フラッシュメモリかを判別し、格納先が前記ＨＤＤの場合には前記ＨＤＤから読み出したデータを前記キャッシュメモリ部に格納した後に前記ホストコンピュータに応答し、格納先が前記フラッシュメモリの場合には前記フラッシュメモリから読み出したデータを前記キャッシュメモリ部には格納せずに前記ホストコンピュータに応答し、前記ストレージ装置は、前記フラッシュメモリに対する書き込み回数を平均化するように、デステージするデータを選択し、各々のフラッシュメモリに対するデステージした回数を記録し、デステージするデータを選択する際に、各々の前記フラッシュメモリに対するデステージした回数を比較し、デステージした回数が少ないものから優先的にデステージするデータを決定することを特徴とするストレージシステムである。

【発明の効果】

【００１１】

本発明のストレージシステムでは、低消費電力であり、設置面積が小さく、かつ、大規模まで構成可能であり、また、データを格納するメディアに応じた、高いシステム性能を提供することもできる。また、各メディアに対する書き込み回数の低減を図るので、書き込み回数制限を有するメディアに対してもストレージシステムとして、信頼性、可用性の向上可能になるといった利点がある。

【発明を実施するための最良の形態】

【００１２】

本発明を実施するための最良の形態を説明する。

以下に本発明のストレージシステムとストレージ装置及びその制御方法の実施例について、図面に基づいて説明する。

【実施例１】

【００１３】

実施例１を説明する。図１は、本発明の第１の実施の形態のストレージシステムの構成のブロック図である。ストレージシステムは、ストレージ制御装置１およびハードディスクドライブ（ＨＤＤ）５０、フラッシュメモリデバイス（本図ではＦＭ制御部１６の内部にフラッシュメモリデバイスを備える例を示している）から構成されている。ストレージ制御装置１は、チャネル４を通じて、ＳＡＮスイッチ３などで構成されるＳＡＮ（Storage Area Network）を経て、ホストコンピュータ２に接続される。また、ディスク側チャネル６０を通じてデータを格納するメディアである複数のＨＤＤ５０と接続されている。ストレージ制御装置１は、複数のチャネル制御部１１と複数のキャッシュメモリ１３と制御メモリ部１７と複数のディスク制御部１４および複数のＦＭ制御部１６と、これらを、内部パス１５を介して接続する内部スイッチ１２から成る。チャネル制御部１１は、チャネル４を通じてホストコンピュータ２からの入出力要求を受け取り、この入出力要求の要求種類（例えば、リード要求、ライト要求）や対象アドレスなどを解釈し、図７以降で述べるような処理を行う。キャッシュメモリ部はＨＤＤやフラッシュメモリに格納されるべきデータやホストコンピュータ２に返すべきデータを一時的に格納する。制御メモリ部１７は、キャッシュメモリ部１３のディレクトリ情報や、ストレージシステムの構成情報を格納している。ディスク制御部１４は、チャネル制御部１１などの要求に基づき、ディスク側チャネル６０を通じてＨＤＤ５０の制御を行い、ホストコンピュータ２から要求されたデータを取り出しや格納を行う。この際、ディスク制御部１４がＨＤＤ５０に対してＲＡＩＤ制御を行いストレージシステムの信頼性、可用性および性能を向上させてもよい。ＦＭ制御部１６は、フラッシュメモリまたはフラッシュメモリデバイスの制御を行う。チャネル制御部１１などの要求に基づき、フラッシュメモリまたはフラッシュメモリデバイス

にホストコンピュータ２から要求されたデータを取り出しや格納を行う。この際、ＦＭ制御部１６がフラッシュメモリデバイスに対してＲＡＩＤ制御を行いストレージシステムの信頼性、可用性および性能を向上させてもよい。なお、本実施例では、ストレージシステムはＨＤＤ５０と接続されているが、ＨＤＤ５０ならびにディスク制御部１４を持たない構成でもよい。また、制御メモリ部１７に格納される情報は、物理的にキャッシュメモリ部１３と同一のメモリ上に配置してもよい。

【００１４】

図２はチャネル制御部１１の詳細の構成のブロック図である。チャネル制御部１１は、複数のプロセッサ１１１、メモリモジュール１１２、周辺処理部１１３、及び、複数のチャネルプロトコル処理部１１４と内部ネットワークインタフェース部１１７から成る。プロセッサ１１１はバス等で周辺処理部１１３に接続される。周辺処理部１１３は、メモリモジュール１１２に接続され、メモリモジュールの制御を行う。また、制御系バス１１５を介してチャネルプロトコル処理部１１４および内部ネットワークインタフェース部１１７にも接続される。周辺処理部１１３は、接続されるプロセッサ１１１及びチャネルプロトコル処理部１１４及び内部ネットワークインタフェース部１１７からのパケットを受け、パケットの示す転送先アドレスがメモリモジュール１１２上ならばその処理を行い、必要ならばデータを返す。また転送先アドレスがそれ以外ならば、適切なフォワーディングを行う。また、周辺処理部１１３は、その他のプロセッサ１１１がこの周辺処理部１１３に接続されるプロセッサ１１１と通信を行うためのメールボックス１１３１を持つ。プロセッサ１１１は、周辺処理部１１３を通してメモリモジュール１１２にアクセスし、メモリモジュール１１２に格納された制御プログラム１１２１に基づいて処理を行う。また、メモリモジュール１１２には、チャネルプロトコル処理部１１４がＤＭＡ（Ｄｉｒｅｃｔ Ｍｅｍｏｒｙ Ａｃｃｅｓｓ）を行うための転送リスト１１２３も格納されている。チャネルプロトコル処理部１１４は、チャネル４上のプロトコル制御を行い、ストレージシステム１内部で処理ができるようなプロトコル方式に変換する。チャネルプロトコル処理部１１４は、チャネル４を通じてホストコンピュータ２からの入出力要求を受けると、その入出力要求のホストコンピュータ番号やＬＵＮ（Logical Unit Number）やアクセス先アドレスなどをプロセッサ１１１に通知する。プロセッサ１１１は、その通知に基づき、ディレクトリ情報１３２３にアクセスし、入出力要求のデータを格納すべきアドレスまたは入出力要求のデータが存在する場合はメモリモジュール１１２上に転送リスト１１２３を作成し、それに基づきチャネルプロトコル処理部１１４に転送を行わせる。ホストコンピュータ２がリード要求したデータがキャッシュメモリ１３上にない場合で、ＨＤＤ５０に格納されているならば、ディスク制御部１４に、ＨＤＤ５０に格納されている要求データをキャッシュメモリ１３に格納する（この動作を「ステージング」という）ように指示を与えた後に転送リスト１１２３により転送させる。フラッシュメモリ上に格納されているならば、フラッシュメモリのアドレスを転送リストにセットする。転送リストは、キャッシュメモリ１３もしくはフラッシュメモリ上のアドレスのリストになっており、入出力要求がライトならば、データ転送系バス１１５を介して接続された内部ネットワークインタフェース部１１７を通じ、ホストコンピュータからのデータをリストに記載されたアドレスに書き込んでいく。またリードならば、同様にリストに記載されたアドレスからデータを読み込み、それをホストコンピュータに返す。これらの動作の詳細に関しては図７以降を用いて説明する。内部ネットワークインタフェース部１１７は、チャネル制御部１１内と他のストレージシステム１内部と内部パス１５を経て通信を行う際の、インタフェースとなる部位である。

【００１５】

なお、ディスク制御部１４もチャネル制御部１１とほぼ同じ構造を持つ。ただし、制御プログラム１１２１の内容が異なり、また、チャネルプロトコル処理部１１４はＨＤＤ５０との通信を行う（チャネル４とディスク側チャネル６０のプロトコルは異なってもよい。しかし、ディスク側チャネル６０上のプロトコル処理を行い、ストレージシステム１内部で処理が出来るように変換する意味ではチャネル制御部１１のチャネルプロトコル処理

部１１４と同様）。プロセッサ１１１はチャネル制御部１１からの要求や一定時間間隔で、キャッシュメモリ１３上のデータをハードディスクドライブ５０に書き込みを行う。また、キャッシュメモリ１３上にホストコンピュータから要求されたデータが存在していない場合は、チャネル制御部１１からの指示を受け、ＨＤＤ５０からデータを読み込み、キャッシュメモリ１３にそのデータを書き込む。これらの際には、プロセッサ１１１は、制御メモリ部１７に格納されたディレクトリ情報にアクセスし、ホストコンピュータ２の要求するデータを読み出すべきまたは格納すべきキャッシュメモリのアドレスを調査する。そして、要求されたデータがキャッシュメモリ１３上に存在しない場合でキャッシュメモリ１３に空き領域がないときは、要求されたデータを格納するため空き領域を作るべく、すでにあるデータをＨＤＤ５０に格納する（この動作をデステージと呼ぶ）。これらＨＤＤ５０の操作においては、ディスク制御部１４がディスク側チャネル６０を通じてＨＤＤ５０を制御する。この際、ＨＤＤ５０全体としての可用性および性能を向上させるため、ディスク制御部１４はＨＤＤ５０群に対してＲＡＩＤ制御を行う。

【００１６】

図３は、ＦＭ制御部１６の詳細の構成のブロック図であり、フラッシュメモリを一体化している。ＦＭ制御部１６は、内部ネットワークインタフェース部１６１、ＤＭＡコントローラ１６２、揮発メモリであるメモリモジュール１６４およびその制御を行うメモリコントローラ１６３、ならびにフラッシュメモリ１６６（図中ではＦＭ）とそれを制御するＦＭコントローラ１６５が備えられている。内部ネットワークインタフェース部１６１は、ＦＭ制御部１６内と他のストレージ制御装置１内部と内部パス１５を経て通信を行う際の、インタフェースとなる部位である。ＦＭ制御部１６内のＤＭＡコントローラ１６２は、ホストコンピュータ２からのライトの要求を処理する際キャッシュメモリに空き領域を作るなどの場合に、チャネル制御部１１のプロセッサ１１１がセットする転送リスト１６４１により、キャッシュメモリ部１３からフラッシュメモリ１６６へのデータの転送を行う。ＦＭコントローラ１６５は、内部ネットワークを通してなされたチャネル制御部１１からの読み込み要求や、ＤＭＡコントローラ１６２の書き込み要求により、フラッシュメモリ１６６を制御しデータのやり取りを行う。図３において、フラッシュメモリ１６６の実装形態としては、基板に直接配置することができる。この際は図４のコネクタやＦＭプロトコル処理部、図５のＦＭ側チャネルなどの部品が不要になるので、よりコンパクトにストレージシステムを実現することが可能になる。また、ＦＭコントローラ１６５にてそれぞれのフラッシュメモリ１６６に対するウェアレベリングなどを行ってもよい。

【００１７】

図４は、ＦＭ制御部１６の別の詳細の構成のブロック図である。ここでは、記憶素子としてフラッシュメモリデバイス１６９を用いており、フラッシュメモリデバイスを別体としている。フラッシュメモリデバイス１６９は、コネクタ１６８を介してＦＭ制御部１６と接続されているため、脱着が可能となっている。そのため、フラッシュメモリデバイス１６９が故障した際には、交換することも可能となる（これを行うためには、チャネル制御部１１のプロセッサ１１１が、予めフラッシュメモリデバイス１６９間で冗長構成をとるように、転送リスト１６４１をセットしていればよい）。また、フラッシュメモリデバイス１６９自体をより容量の大きなものに交換することも可能となる。このフラッシュメモリデバイス１６９は、内部においてウェアレベリングなどの技術により信頼性、性能の向上が図られているものであり、外部とのやり取りは、専用プロトコルで行われる。そのため、ＦＭプロトコル処理部１６７にて、ストレージ制御装置１内部で処理できる形式に変換を行っている。

【００１８】

図５は、ＦＭ制御部１６の別の詳細の構成のブロック図である。ここでは、フラッシュメモリデバイス１６９をＦＭ側チャネル１６１０により接続している。この構成をとることにより、図４の特徴に加え、より多数のフラッシュメモリデバイス１６９を接続でき、大容量のストレージシステムを実現することができる。また、あるフラッシュメモリデバイス１６９の一部の領域は、後に説明する緊急デステージ領域１６９０としてもよい。

【００１９】
図６は、内部スイッチ部１２の詳細の構成のブロック図である。内部スイッチ部１２は、内部ネットワークインタフェース部１２１と複数のセレクタ１２２からなっている。セレクタ１２２は、ストレージ制御装置１の内部のチャネル制御部１１など各部から送られてきた要求の要求先を解析し、その要求先に接続される内部パス１５を制御する内部ネットワークインタフェース部１２１に要求を転送する。その際、各セレクタ間で要求転送先の内部ネットワークインタフェース部１２１競合を行う。この内部スイッチ部１２により、チャネル制御部１１からは、キャッシュメモリ部１３、制御メモリ部１７およびＦＭ制御部１６と直接のやり取りを行うことができる。ＦＭ制御部１６はチャネル制御部１１、キャッシュメモリ部１３および制御メモリ部１７とやり取りができる。また、ディスク制御部１４からは、キャッシュメモリ部１３と制御メモリ部１７に直接のやり取りができる。ＦＭ制御部１６とディスク制御部１４の接続に関する相違は、この内部スイッチ部１２はチャネル制御部ＦＭ制御部間接続１２３を備えていることにより、ＦＭ制御部１６はチャネル制御部１１と直接のやり取りが行える点である。

【００２０】
図７は、ホストコンピュータ２から、ＨＤＤ５０領域へのリードの要求が行われた場合の処理の流れを表す図である。ホストコンピュータ２からリード要求を、チャネル４を通じてチャネル制御部１１が受信する（ｓ７０１）。チャネル制御部１１のプロセッサ１１１は、受信した要求を解析し、ＬＵＮや対象論理ブロックアドレスを得る。ここでは、該当データがＨＤＤ５０に格納される領域のものであることを知る（ｓ７０２）。さらにチャネル制御部１１のプロセッサ１１１は、制御メモリ部１７に格納された、ライトキャッシュ領域ならびにリードキャッシュ領域のディレクトリ情報にアクセスし、キャッシュメモリ部１３に該当データの格納の有無を確認する（ｓ７０３、ｓ７０４。図中では１回のアクセスになっているが、実際には複数回のアクセスになる場合もある。以下も同様）。該当データがキャッシュメモリ部１３に既に存在していればｓ７１５以降の処理によりホストコンピュータ２に応答するが、ここでは、該当データがキャッシュメモリ部１３に存在しなかったとする。この場合、ディスク制御部に該当データをキャッシュメモリ部１３に転送させる（ステージング）が、キャッシュメモリ部１３に空き領域がない場合は、ステージングの前にデータを格納すべきキャッシュ領域を作らなければならない。どの領域に空き領域にすべきかを決定し、その操作をおこなうステップがｓ７０５、ｓ７０６である。領域が確保された後で、ステージングの要求を、制御メモリ部１７の通信領域１７３にメッセージを書き込むことを介してディスク制御部１４に通知する（ｓ７０７）。ディスク制御部１４は定期的もしくは一連の処理の終了毎に、制御メモリ部１７の通信領域１７３を読み込むことによりチャネル制御部１１からの要求があることを知る（ｓ７０８、ｓ７０９）。なお、こうした方式でチャネル制御部１１とディスク制御部１４が連携を行うのは、ＨＤＤ５０からデータを得られる時間が不定で、その他の処理時間と比較して長いためであり、この方式によりバックグラウンドで他の要求の処理などを行うことができる。ステージングの要求を認知したディスク制御部１４は、該当データが得られるようにディスク側チャネル６０を通じてＨＤＤ５０を制御する（ｓ７１０）。ＨＤＤ５０からのデータが得られると、ｓ７０５、ｓ７０６により確保した領域に、該当データを書き込む（ｓ７１１。ステージング）。また、ステージングが終了したことを、通信領域１７３を用いてチャネル制御部１１に通知する（ｓ７１２）。チャネル制御部１１は、ディスク制御部１４同様に通信領域１７３を読み込んでおり、ステージング終了のメッセージの存在を知る（ｓ７１３、ｓ７１４）。その後、チャネル制御部１１のプロセッサ１１１は、転送リスト１１２３をセットし（あらかじめステージング要求直後にセットしておいてもよい）、転送の指示を行う。チャネルプロトコル処理部１１４は、キャッシュメモリ部１３から該当データを読み込み、ホストコンピュータ２に転送する（ｓ７１５、ｓ７１６、ｓ７１７）。以上が、ＨＤＤ５０に格納されたデータのリードの処理の流れである。ＨＤＤ５０の速度が、不定であり、遅いため、キャッシュメモリ部１３を介したデータのやり取りが必要となる。

【００２１】
図８は、ホストコンピュータ２から、フラッシュメモリ領域へのリードの要求が行われた場合の処理の流れを表す図である。ホストコンピュータ２からリード要求を、チャネル４を通じてチャネル制御部１１が受信する（ｓ８０１）。チャネル制御部１１のプロセッサ１１１は、受信した要求を解析し、ＬＵＮや対象論理ブロックアドレスを得る。ここでは、該当データがフラッシュメモリに格納される領域のものであることを知る（ｓ８０２）。さらにチャネル制御部１１のプロセッサ１１１は、制御メモリ部１７に格納された、ライトキャッシュ領域のディレクトリ情報にアクセスし、キャッシュメモリ部１３のライトキャッシュ領域１３２に該当データの格納の有無を確認する（ｓ８０３、ｓ８０４）。図７の場合と違い、ライトキャッシュ領域１３２に対する調査のみでよい。ここでは、該当データがキャッシュメモリ部１３に存在しなかったとする。チャネル制御部１１のプロセッサ１１１は、転送リスト１１２３をセットし、チャネルプロトコル処理部１１４に転送の指示を行う。チャネルプロトコル処理部１１４は、内部スイッチ部１２のチャネル制御部ＦＭ制御部間接続１２３を介して、ＦＭ制御部１６にデータの要求を行う（ｓ８０５）。ＦＭ制御部１６は、フラッシュメモリ１６６もしくはフラッシュメモリデバイス１６９からデータを取り出し（ｓ８０６）、チャネル制御部１１にデータを返す（ｓ８０７）。こうして得たデータをチャネルプロトコル処理部はホストコンピュータ２に転送する（ｓ８０８）。以上が、フラッシュメモリ１６６もしくはフラッシュメモリデバイス１６９に格納されたデータのリード手順であった。フラッシュメモリの読み出し速度が一定で、高速のため（ｓ８０６）、キャッシュメモリ部１３を介さずにデータのやり取りが可能になる。また、直接転送を行うため、リードに関してはキャッシュ処理を行わず、リードキャッシュ領域１３１の調査（ｓ７０３、ｓ７０４）やキャッシュ領域の確保（ｓ７０５、ｓ７０６）の処理が不要になり、メディアからの読み出し高速化のみならず、付随する処理の高速化も可能になる。また、フラッシュメモリに対しては、リードキャッシュ領域１３１が不要なため、キャッシュメモリ部１３やそのリードキャッシュディレクトリ情報１７１１を格納する制御メモリ部１７の容量を小さくすることが可能になる効果もある。

【００２２】
続いて、フラッシュメモリ１６６またはフラッシュメモリデバイス１６９に好適なキャッシュメモリ部１３の制御方法を説明するが、まずその前に、ホストコンピュータ２からライト要求が行われた場合の処理の流れと、キャッシュメモリ部１３に関わる情報の説明を行う。

【００２３】
図１１Ａは、ホストコンピュータ２からライトの要求が行われた場合で、すでにキャッシュメモリ部１３上に該当アドレスのデータが存在する場合の処理の流れを表す図である。ライト要求については、対象がＨＤＤ５０領域、フラッシュメモリ領域いずれの場合もデータをキャッシュメモリ部１３に一時的に格納する。まず、ホストコンピュータ２からライト要求を、チャネル４を通じてチャネル制御部１１が受信する（ｓ１１０１）。チャネル制御部１１のプロセッサ１１１は、受信した要求を解析し、ＬＵＮや対象論理ブロックアドレスを得る（ｓ１１０２）。さらにチャネル制御部１１のプロセッサ１１１は、制御メモリ部１７に格納された、ライトキャッシュ領域ならびにリードキャッシュ領域のディレクトリ情報にアクセスし、キャッシュメモリ部１３に該当データの格納の有無を確認する（ｓ１１０３、ｓ１１０４。なお、リードキャッシュ領域１３１に存在した場合は、後に述べるような手順で該当スロットを無効化し、新たにライトキャッシュディレクトリ情報１７１２に加える）。ここでは、該当アドレスのデータがライトキャッシュ領域１３２に既に存在していたとする。この場合、同じキャッシュメモリのスロット（キャッシュの制御単位。後で詳述）にデータを格納する。チャネル制御部１１は、　ホストコンピュータ２にデータの送信要求を行い（ｓ１１０５）つつ、データを該当するスロットへ格納するように転送リスト１１２３をセットする。チャネル制御部１１のチャネルプロトコル処理部１１４は、応答したホストコンピュータ２からのデータを受信し（ｓ１１０６）、それを転送リスト１１２３に基づきキャッシュメモリ部１３に格納する（ｓ１１０７）。

このような場合、フラッシュメモリに対するライト要求であったとしても、フラッシュメモリに書き込みをしなくともよく、書き込み回数の低減が図れる。このため、ライト要求の場合は、リード要求の場合と異なり、データの格納メディアがフラッシュメモリ１６６もしくはフラッシュメモリデバイス１６９の場合でも、キャッシュメモリ部１３を用いた処理を行う。

【００２４】

次に図１１Ｂを用いて、ホストコンピュータ２から、ライトの要求が行われた場合で、キャッシュメモリ部１３上に該当アドレスのデータが存在しないが既に空いているスロットがない場合の処理の流れを説明する。ｓ１１２４までは、図１１Ａのｓ１１０４と同様であるが、この図においては、該当アドレスのデータがライトキャッシュ領域１３２に存在しておらず、空きのスロットはない場合とする。そこで、まず、既にキャッシュメモリ部１３上に存在するデータを、格納されるべきＨＤＤ５０、フラッシュメモリに格納し（デステージング）、今回のホストコンピュータ２からのデータを格納するための領域を作る。まず、ＬＲＵ（Ｌｅａｓｔ　Ｒｅｃｅｎｔｌｙ　Ｕｓｅｄ）アルゴリズムなどで、デステージするデータを決定する（ｓ１１２５）。デステージするデータが決まったら、チャネル制御部１１のプロセッサ１１１は、ＦＭ制御部１６に、キャッシュメモリ部１３上のアドレスと格納先のフラッシュメモリ１６６もしくはフラッシュメモリデバイス１６９のアドレスの対応が記された転送リスト１６４１をセットする（ｓ１１２７）。なお、対象がＨＤＤ５０の場合は、図７同様に通信領域１７３を用いてディスク制御部１４にメッセージを送信することにより、デステージの要求が行われる。続いて、転送リスト１６４１に基づき、ＦＭ制御部１６のＤＭＡコントローラ１６２はキャッシュメモリ部１３から該当データを読み込み（ｓ１１２７、ｓ１１２８）、フラッシュメモリ１６６もしくはフラッシュメモリデバイス１６９に書き込む（ｓ１１２９）。一連の処理が終了するとＤＭＡコントローラ１６２はデステージが完了した旨をチャネル制御部１１のプロセッサ１１１に通知する（ｓ１１３０）。プロセッサ１１１は、デステージしたスロットに今回のホストコンピュータ２からのデータを格納するために、ディレクトリ情報を更新し（ｓ１１３１、ｓ１１３２）、ホストコンピュータ２にデータの送信要求を行い（ｓ１１３３）つつ、データを該当するスロットへ格納するように転送リスト１１２３をセットする。チャネル制御部１１のチャネルプロトコル処理部１１４は、応答したホストコンピュータ２からのデータを受信し（ｓ１１３４）、それを転送リスト１１２３に基づきキャッシュメモリ部１３に格納する（ｓ１１３５）。

【００２５】

次に、これまで述べてきたキャッシュメモリ部１３に関する情報を説明する。図９は、キャッシュメモリ部１３ならびに制御メモリ部１７に格納されるデータの詳細を示したブロック図である。キャッシュメモリ部１３は、リードキャッシュ領域１３１とライトキャッシュ領域１３２を持つ。リードキャッシュ領域１３１は、ホストコンピュータ２からリード要求されたデータを一時的に格納する。これにより、格納されているデータに対して再びホストコンピュータ２からリード要求を受け付けた場合に、ＨＤＤ５０から再び読み込むことなくリードキャッシュ領域１３１のデータを応答することによりストレージシステムとしての処理の高速化を図る。本実施例では、ＨＤＤ５０の読み出しまでの時間が不定であり、低速であるためリードキャッシュ領域１３１を設けているが、フラッシュメモリに格納されているデータに対してはリードキャッシュ領域１３１を用いない。ライトキャッシュ領域１３２は、ホストコンピュータ２からライト要求されたデータを一時的に格納し、格納されているデータと同じアドレスに対して再びライト要求を受け付けた場合には、ライトキャッシュ領域１３２上の該当データを上書きする。ＨＤＤ５０領域に対しては、高速なキャッシュメモリ部１３に一時的に書き込んでおくことで性能の向上も期待できる。また、キャッシュメモリ部１３上で上書きさせるため、フラッシュメモリに書き込む回数を低減することができ、ストレージシステムの信頼性、可用性を向上させることができる。なお、これらの管理している単位をここではスロットという。

【００２６】

次に制御メモリ部１７について説明する。制御メモリ部１７は、ディレクトリ情報１７１、構成情報１７２ならびに通信領域１７３を格納している。構成情報１７２は、ストレージシステムに関する情報である。例えば、チャネル制御部１１がいくつ存在していて、それぞれ接続されているチャネル４にどういったＬＵ（Ｌｏｇｉｃａｌ　Ｕｎｉｔ）を提供し、そのＬＵは、どのＨＤＤ５０もしくはフラッシュメモリデバイス１６６を用いてデータを格納するか、またそれはどういった形で仮想的に提供するか、といった情報である。通信領域１７３は、チャネル制御部１１とディスク制御部１４とが互いにメッセージを書き込みまたは読み込みを行い協調動作するための領域である。また、ディレクトリ情報１７１は、キャッシュメモリ部１３にどのようなデータが格納されているかを示す情報であり、どのＬＵＮや論理ブロックアドレス（ＬＢＡ、Ｌｏｇｉｃａｌ　Ｂｌｏｃｋ　Ａｄｄｒｅｓｓ）などで示されるホストコンピュータ２からのアドレスのデータがキャッシュメモリ部１３上のどこに格納されているかを示すリードキャッシュディレクトリ情報１７１１、ライトキャッシュディレクトリ情報１１７２、および、新しいデータをキャッシュメモリ部１３に格納するために、どの元からあるデータを無効化するかを決定するための基準となるリードおよびライトのアクセス順序リスト１７１３、１７１４を持つ。

【００２７】

なお、キャッシュメモリ部１３、制御メモリ部１７は、高速で書き込み回数制限のない揮発性のメモリで構成される。しかし、停電などに備え、一定時間は内容を保持できるバッテリバックアップの機能を備えている。しかし、そのバッテリの保つことの出来る時間を越えると考えられる更に大規模な停電の際には、図１１Ｂのｓ１１２５からｓ１１２９の処理を用いて、全ての一時的に格納されたキャッシュメモリ部１３上のデータをメディアに書き込み、バックアップバッテリが切れた際にもデータを失わないようにする。もしくは、ｓ１１２５からｓ１１２９を用いた処理はプロセッサ１１１の計算量が必要となってしまうのでこの方法を用いず、メディア上に予め用意された緊急デステージ領域に、キャッシュメモリ部１３および制御メモリ部１７のイメージをそのまま書き込む。これは、転送リスト１６４１をキャッシュメモリ部１３および制御メモリ部１７の全域を獲得するようにセットすることによって実行しても、ＤＭＡコントローラ１６２に、キャッシュメモリ部１３および制御メモリ部１７のイメージを取得する機能を設けることによって実行してもよい。特にフラッシュメモリは、低消費電力であるので、バックアップバッテリの限られた時間と電力で実行せねばならないこの処理のメディアとして適している。つまり、この緊急デステージ領域をフラッシュメモリ１６６もしくはフラッシュメモリデバイス１６９上に設けることにより大きな効果が得られる。例えば、図５ではあるフラッシュメモリデバイス１６９の一部の領域に緊急デステージ領域１６９０を割り当てている。

【００２８】

図１０Ａは、リードキャッシュディレクトリ情報１７１１、ライトキャッシュディレクトリ情報１１７２の詳細を示したブロック図である。ディレクトリ情報は、ＬＵＮ欄１００１、ＬＢＡ欄１００２、最終的にそのデータが格納されるメディアを示すメディア欄１００３および、それに続く、ホストコンピュータ２から指定されるアドレス（ＬＢＡ）とそのデータが格納されているキャッシュメモリ部１３上のアドレスの対応がリスト形式で格納されているホスト－キャッシュアドレス対応リスト１００４から成る。キャッシュメモリ部１３上のホストコンピュータ２の要求したデータがあるかどうかを調査する場合（図７のｓ７０３、ｓ７０４や図８のｓ８０３、ｓ８０４の操作に相当）、まず受領した要求のＬＵＮと一致するものをＬＵＮ欄１００１から見つける。次に対応するポインタが指し示すテーブルのＬＢＡ欄１００２と、受領した要求のＬＢＡの一致するものを調査する。ＬＢＡ欄１００２はＬＢＡの範囲が示されており、一致するＬＢＡ範囲のデータに関する情報が、その欄に対応するポインタが指す先の、ホスト－キャッシュアドレス対応リスト１００４に格納されている。これには、（範囲ではなく）ＬＢＡの値と、そのデータが格納されているキャッシュのアドレス（スロット番号）が記されている。このホスト－キャッシュアドレス対応リスト１００４の中に対応するＬＢＡがなければキャッシュメモリ部１３には該当データが格納されていないことになる。

【００２９】
図１０Ｂは、従来のアクセス順序リスト１７１３、１７１４の詳細を示したブロック図である。これは、どのスロットのデータが何番目に最近アクセスされたのかを示す情報である。先頭に格納されているものが最も最近にアクセスされたスロット、最後尾に格納されているものが最も昔にアクセスされたスロットである。リードキャッシュ領域１３１、ライトキャッシュ領域１３２のそれぞれに対応する情報がリスト構造で格納されている。例えば、ホストコンピュータ２からキャッシュメモリ部１３に格納されていないデータに対してのアクセス要求を受領した場合など、キャッシュメモリ部１３に開きスロットを作る場合、一般的なＬＲＵアルゴリズムを用いるとすると、キャッシュに新たな空きスロットを作る場合、最も昔にアクセスされたもの、すなわち、リストの最後尾に示されているスロットを無効化する。その後、新しいデータを該当スロットに格納し、該当スロットは最も最近にアクセスされたことになるので、アクセス順序リスト１７１３、１７１４の先頭に該当スロットをおく。このリスト操作の手順としては、まず、最後尾にあった該当スロットの情報をdelete操作により外し、先頭に該当スロットの情報をinsert操作により追加する。なお、この際、キャッシュディレクトリ情報１７１１または１７１２も更新する。該当スロットに元に格納されていたものを示す情報をホスト－キャッシュアドレス対応リスト１００４からdelete操作により外し、新たなＬＢＡに対応するものをこの該当スロット番号とともにinsert操作により追加する。また、ホストコンピュータ２からの要求を受けたが、キャッシュメモリ部１３に、すでに格納されていた場合は、上記の手順によりアクセス順序リスト１７１３、１７１４のみを更新し、キャッシュディレクトリ情報１７１１、１７１２は更新しない。

【００３０】
以上に示した、図１０Ｂのアクセス順序リストをフラッシュメモリ１６６およびフラッシュメモリデバイス１６９に用いると次のような不都合が起こる。アクセス順序によりデステージするスロットを決定しているので、あるメディアにばかりデステージすなわち書き込みが起こってしまうことがある。例えば、スロット数が３のキャッシュメモリが備わっている場合に、ＦＭ１に対するアドレスＡ、Ｂに対するデータと、ＦＭ２に対するアドレスＣ、Ｄに対するデータを、Ａ、Ｂ、Ｃ、Ａ、Ｂ、Ｄのパターンの繰り返しでライト要求された場合は、ＦＭ２に集中してデステージが起こってしまう。こういった状況では、メディアの書き込み制限回数に早々に到達してしまい、ストレージシステムとして十分な信頼性、可用性を提供できない恐れがある。なお、本実施例では、ＬＲＵアルゴリズムを用いて説明しているが、アクセス頻度を用いるなどのＬＲＵ以外のアルゴリズムを用いた場合も同様のことが起こりうる。

【００３１】
以下では、こうした不都合を回避するキャッシュ制御方法を説明する。図１０Ｃは、好適なキャッシュ制御を行うためのアクセス順序リスト１７１３、１７１４の詳細を示したブロック図である。図１０Ｂと異なり、書き込み回数制限があるフラッシュメモリに関しては、各メディア毎（各フラッシュメモリデバイス１６９毎）にリスト１０６１が設けられている。ＨＤＤ５０に関しては、各々のＨＤＤ５０毎に設けても、またはＨＤＤ５０全体として１つのリスト１０６２としてもよい。さらに、各リストにデステージ回数欄１０６３、アクセス回数欄１０６４、容量欄１０６５を持つ表を備える。デステージ回数欄１０６３は、そのメディアにデステージを行った総回数を示す。アクセス回数欄１０６４は、そのメディアに領域に対して起こったリード、ライトも含めたアクセス回数を示す。これはキャッシュメモリ部１３上にデータが存在した場合も回数に含めている。また容量欄１０６５は、そのメディアの容量を示している。同じ頻度でフラッシュメモリデバイス１６９にデステージしたとしても、容量が大きいほど、フラッシュメモリ内部の１つのセルに書き込まれる頻度は少なくなることを考慮する場合に用いる。さらに、各メディアにリストを持つ。リストの内容は図１０Ｂの場合とほぼ同様で、ＬＲＵアルゴリズムを用いるとすると、どのスロットが何番目に最近アクセスされたのかを示す情報が格納されている。ただし、この情報は、このメディアに格納されるべきデータに関するもののみである。

なお、ＨＤＤ５０に関するリストにはデステージ制限欄１０６６がある。この欄が「なし」の場合は、フラッシュメモリとＨＤＤ５０で異なるキャッシュ制御を行い、次に説明する図１２Ａのｓ１２０１でメディア種類をＨＤＤと認識する。この場合、ライトキャッシュ領域１３２は、予めＨＤＤ５０領域とフラッシュメモリ領域に分けられている。こうした分割をなくキャッシュメモリをより柔軟に用いたい場合は、この欄を「あり」とし、フラッシュメモリと同等のキャッシュ制御を行う。この場合は、メディア種類がＨＤＤであっても、図１２Ａのｓ１２０１でフラッシュメモリとして処理が選択される。また、デステージ回数欄１０６３は、一定時間間隔でゼロにリセットしてもよい。リセットを行う際は全メディアに関してのデステージ回数欄１０６３をゼロにする。

【００３２】

次に、図１０Ｃのアクセス順序リスト１７１３、１７１４を用いた、デステージするスロットを決定する処理の説明を、図１２Ａを用いて行う。これは、チャネル制御部１１のプロセッサ１１１により実行されるもので、デステージを行わねばならないことが判明した後の処理である図１１Ｂのｓ１１２５以降の処理に相当する。まず、ホストコンピュータ２から受領したライト要求が、ＨＤＤに対するものかフラッシュメモリに対するものか、メディア種類を判別する（ｓ１２０１）。これは、ディレクトリ情報のメディア欄１００３を用いて知ることができる。メディア種類がＨＤＤならば、ＨＤＤのアクセス順序リスト１０６２を用いて最も昔にアクセスされたスロットを探し、そのスロットをデステージさせるメッセージを送り、ｓ１２０７に処理を移す（ｓ１２０２）。メディア種類がフラッシュメモリ、もしくはＨＤＤであってもデステージ制限欄１０６６が「あり」の場合は、各メディアのデステージ回数欄１０６３を比較して、デステージ回数が最小のメディアを探す（ｓ１２０３）。そのメディアのリストの要素数を調べることで、使用しているスロットがあるかを調査する（ｓ１２０４）。ないならば、次にデステージ回数の少ないメディアを探して再びｓ１２０４に処理を移す（ｓ１２０５）。あるならば、そのアクセス順序リスト１０６１を用いて最も昔にアクセスされたスロットを探し、そのスロットをデステージするように指示する（ｓ１２０６）。次に該当メディアのデステージ回数欄１０６３に１を足す（ｓ１２０７）。さらに、該当スロットをディレクトリ情報およびアクセス順序リストから外し（ｓ１２０８）、さらにステージするデータに関してディレクトリ情報およびアクセス順序リストの最新に追加する（ｓ１２０９）。その後、該当スロットに今回受領したデータを書き込むことを行う（ｓ１２１０）。なお、ｓ１２０３でデステージ回数の最小のメディアを探すが、容量欄１０６５の説明で述べたとおりセル自体の書き込み回数を考慮するならば、その評価関数として、デステージ回数を容量で除算したものを用いてもよい。

【００３３】

また、デステージ回数のみでなく、アクセス回数も考慮した場合の同処理を、図１２Ｂを用いて説明する。まず、各メディアのデステージ回数欄１０６３とアクセス回数欄１０６４について調査し、α、βを定数とし、αとデステージ回数の積とβをアクセス回数で除したものの和を評価関数とし、これが最少となるメディアを探す（ｓ１２４１）。そのメディアのリストの要素数を調べることで、使用しているスロットがあるかを調査する（ｓ１２４２）。ないならば、次に少ないメディアを探して再びｓ１２４２に処理を移す（ｓ１２４３）。あるならば、そのアクセス順序リスト１０６１、１０６２を用いて最も昔にアクセスされたスロットを探し、そのスロットをデステージするように指示する（ｓ１２４４）。以降の処理は、図１２Ａのｓ１２０７と同様である。この処理を用いた場合は、フラッシュメモリ領域とＨＤＤ領域に分けておく必要もなく、またメディアがＨＤＤの場合のライト要求後に同じデータに対するリード要求が多発するなどの、キャッシュメモリを用いた性能向上が可能なパターンなどにも、柔軟に対応することができる。なお、メディアをインデックス順で選択することも可能である。

【実施例２】

【００３４】

実施例２を説明する。図１３に本発明の第２の実施の形態のストレージシステムのブロ

ック図を示す。この形態では、ＦＭ制御部がキャッシュメモリ部１３や制御メモリ部１５と同等の位置に実装されている。一般にキャッシュメモリ部１３や制御メモリ部１５は、チャネル制御部１１やディスク制御部１４と異なり、チャネル４やディスク側チャネル６０が接続できる位置に実装しなくてもよい。例えば、チャネル４接続する作業が容易なように装置の前面に配置するといったことや、ディスク側チャネル６０の伝送路的制限からＨＤＤ５０の付近に配置せねばならないといった実装上の制限はない。そのため、この位置に配置することにより、さらにコンパクトに実装できる。この場合は、図３あるいは図４のＦＭ制御部１６の形式が好適である。

【００３５】

さらに、図１４にＦＭ制御部１６を高機能にした高機能ＦＭ制御部１６０のブロック図を示す。さらに大規模な構成をとる場合は、高機能ＦＭ制御部１６０の使用がより好ましい。ここでは、ＦＭ制御部１６もチャネル制御部１１同様に複数のプロセッサ１１１を備え、制御プログラム１１２１に処理方法を設定することにより高度な信頼性、可用性を得るための制御や管理性向上に向けた制御が行えるようになっている。

【００３６】

以上実施例で説明したが、本発明の他の実施形態１は、前記ストレージ装置は、前記第１のメディアを制御する第１のメディア制御部から前記チャネル制御部に直接に転送するパスを備えるストレージシステムである。

【００３７】

本発明の他の実施形態２は、前記ストレージ装置は、前記キャッシュメモリ部から、前記第１のメディアを制御する第１のメディア制御部に直接に転送するパスを備えるストレージシステムである。

【００３８】

本発明の他の実施形態３は、前記メディアは、書き込み回数に制限を有する第１のメディアと、書き込み回数に制限のない第２のメディアとからなるストレージシステムである。

【００３９】

本発明の他の実施形態４は、前記第２のメディアは、前記第１のメディアよりも読み込み速度が遅く、消費電力が大きいが、第１のメディアよりも書き込み可能回数がはるかに多いメディアであるストレージシステムである。

【００４０】

本発明の他の実施形態５は、前記ストレージ装置は、前記第２のメディアを制御する第２のメディア制御部から、該第２のメディアに格納されたデータを一時的に格納するキャッシュメモリ部に転送するパスを備えるストレージシステムである。

【００４１】

本発明の他の実施形態６は、前記チャネル制御部は、前記ホストコンピュータからのリード要求を受領し、該リード要求の対象データが前記キャッシュメモリ部に格納されていないが前記第１のメディアに格納されている場合、該第１のメディアを制御する第１のメディア制御部に対し、該当データを直接に転送するストレージシステムである。

【００４２】

本発明の他の実施形態７は、前記ストレージ装置は、前記第１のメディアに対する書き込み回数を平均化するように、デステージするデータを選択するストレージシステムである。

【００４３】

本発明の他の実施形態８は、前記ストレージ装置は、各々の第１のメディアに対するデステージした回数を記録するストレージシステムである。

【００４４】

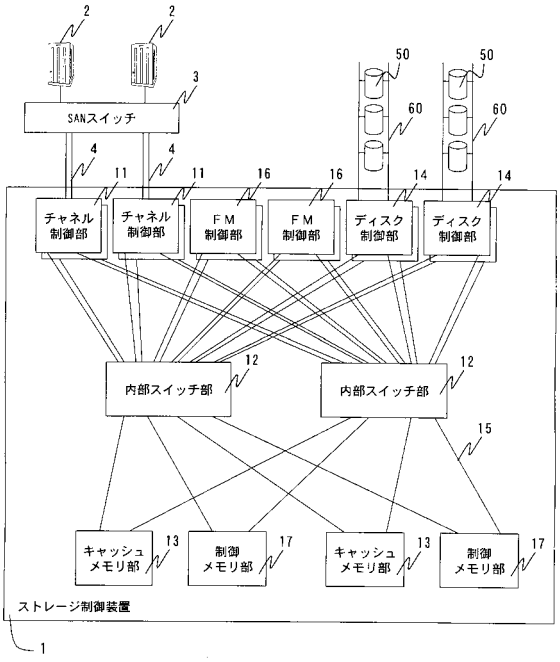
本発明の他の実施形態９は、前記ストレージ装置は、デステージするデータを選択する際に、各々の前記第１のメディアに対するデステージした回数を比較し、デステージした回数が少ないものから優先的にデステージするデータを決定するストレージシステムであ

る。
【００４５】
本発明の他の実施形態１０は、前記ストレージ装置は、各々の第１のメディアに格納されるべきデータに対するアクセスの時間または頻度の情報を、各々の第１のメディアに関連して記録するストレージシステムである。
【００４６】
本発明の他の実施形態１１は、前記ストレージ装置は、デステージするデータを選択する際に、各々の前記第１のメディアに対するデステージした回数およびアクセス回数を基に評価関数を算出し、評価関数が少ないものから優先的にデステージするデータを決定するストレージシステムである。
【００４７】
本発明の他の実施形態１２は、前記ストレージ装置は、キャッシュメモリ部をバックアップするバッテリを備えるストレージシステムである。
【００４８】
本発明の他の実施形態１３は、前記第１のメディア制御部は、前記チャネル制御部の機能の一部を行う高機能メディア制御部であるストレージシステムである。
【００４９】
本発明の他の実施形態１４は、ホストコンピュータからのデータを保存するメディアを制御するメディア制御部、チャネルを介して前記ホストコンピュータに接続するチャネル制御部、及び前記ホストコンピュータからのデータを一時的に保存する揮発メモリであるキャッシュメモリ部を備えるストレージ制御装置において、前記メディアのうちの書き込み回数に制限を有する第１のメディアを制御する第１のメディア制御部から前記チャネル制御部に直接に転送するパスを備えるストレージ装置である。
【００５０】
本発明の他の実施形態１５は、前記キャッシュメモリ部から、前記第１のメディアを制御する第１のメディア制御部に直接に転送するパスを備えるストレージ装置である。
【００５１】
本発明の他の実施形態１６は、前記第１のメディアよりも読み込み速度が遅く、消費電力が大きいが、第１のメディアよりも書き込み可能回数がはるかに多い第２のメディアを制御する第２のメディア制御部から、該第２のメディアに格納されたデータを一時的に格納するキャッシュメモリ部に転送するパスを備えるストレージ装置である。
【００５２】
本発明の他の実施形態１７は、前記チャネル制御部は、前記ホストコンピュータからのリード要求を受領し、該リード要求の対象データが前記キャッシュメモリ部に格納されていないが前記第１のメディアに格納されている場合、該第１のメディアを制御する第１のメディア制御部に対し、該当データを直接に転送するよう指示するストレージ装置である。
【００５３】
本発明の他の実施形態１８は、ホストコンピュータからのデータを保存するメディアを制御するメディア制御部、チャネルを介して前記ホストコンピュータに接続するチャネル制御部、及び前記ホストコンピュータからのデータを一時的に保存する揮発メモリであるキャッシュメモリ部を備えるストレージ制御装置を制御する方法において、前記チャネル制御部が前記ホストコンピュータからのリード要求を受領し、該要求の対象データが前記キャッシュメモリ部に格納されていないが書き込み回数に制限を有する第１のメディアに格納されている場合、該第１のメディアを制御する第１のメディア制御部に対し、該当データを直接転送するよう指示するストレージ装置の制御方法である。
【００５４】
本発明の他の実施形態１９は、前記第１のメディアに対する書き込み回数を平均化するように、デステージするデータを選択するストレージ装置の制御方法である。
【図面の簡単な説明】

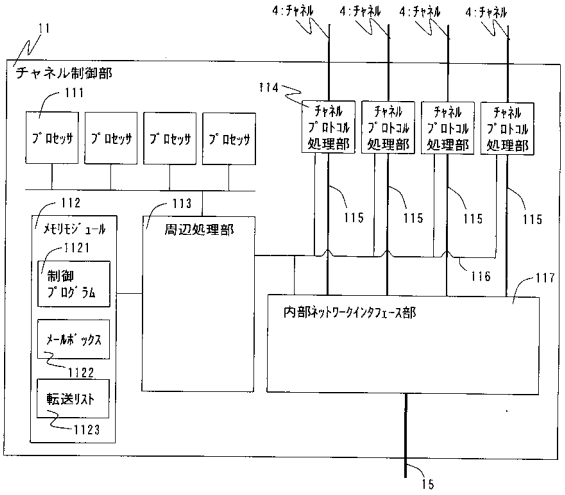
【００５５】
【図１】実施例１のストレージシステムの構成のブロック図。
【図２】チャネル制御部１１の詳細の構成のブロック図。
【図３】ＦＭ制御部１６の詳細の構成のブロック図。
【図４】ＦＭ制御部１６の別の詳細の構成のブロック図。
【図５】ＦＭ制御部１６の別の詳細の構成のブロック図。
【図６】内部スイッチ部１２の詳細の構成のブロック図。
【図７】ホストコンピュータ２から、ＨＤＤ５０領域へのリードの要求が来た場合の処理の流れを表す図。
【図８】ホストコンピュータ２から、フラッシュメモリ領域へのリードの要求が来た場合の処理の流れを表す図。
【図９】キャッシュメモリ部１３ならびに制御メモリ部１７に格納されるデータの詳細を示したブロック図。
【図１０Ａ】リードキャッシュディレクトリ情報１７１１、ライトキャッシュディレクトリ情報１１７２の詳細を示したブロック図。
【図１０Ｂ】従来のアクセス順序リスト１７１３、１７１４の詳細を示したブロック図。
【図１０Ｃ】好適なキャッシュ制御を行うためのアクセス順序リスト１７１３、１７１４の詳細を示したブロック図。
【図１１Ａ】ホストコンピュータ２から、ライトの要求が来た場合で、すでにキャッシュメモリ部１３上に該当アドレスのデータが存在する場合の処理の流れを表す図。
【図１１Ｂ】ホストコンピュータ２から、ライトの要求が来た場合で、キャッシュメモリ部１３上に該当アドレスのデータが存在しないが既に空いているスロットがない場合の処理の流れを表す図。
【図１２Ａ】デステージするスロットを決定する処理を表す図。
【図１２Ｂ】ステージ回数のみでなく、アクセス回数も考慮した場合のスロット決定の処理を表す図。
【図１３】実施例２のストレージシステムのブロック図。
【図１４】高機能ＦＭ制御部１６０のブロック図。
【符号の説明】
【００５６】
１　　ストレージ制御装置
２　　ホストコンピュータ
３　　ＳＡＮスイッチ
４　　チャネル
１１　チャネル制御部
１２　内部スイッチ
１３　キャッシュメモリ部
１４　ディスク制御部
１５　内部パス
１６　ＦＭ制御部
１７　制御メモリ部
５０　ハードディスクドライブ
６０　ディスク側チャネル
１１１　プロセッサ
１１２、１６４　メモリモジュール
１１３　周辺処理部
１１４　チャネルプロトコル処理部
１１５　データ転送系バス
１１６　制御系バス
１１７、１２１、１６１　内部ネットワークインタフェース部

１２２　セレクタ
１２３　チャネル制御部ＦＭ制御部間接続
１３１　リードキャッシュ領域
１３２　ライトキャッシュ領域
１６０　高機能ＦＭ制御部
１６２　ＤＭＡコントローラ
１６３　メモリコントローラ
１６５　ＦＭコントローラ
１６６　フラッシュメモリ
１６７　ＦＭプロトコル処理部
１６８　コネクタ
１６９　フラッシュメモリデバイス
１７２　構成情報
１７３　通信領域
１００１　ＬＵＮ欄
１００２　ＬＢＡ欄
１００３　メディア欄
１００４　ホストーキャッシュアドレス対応リスト
１０６１、１０６２、１７１３、１７１４　アクセス順序リスト
１０６３　デステージ回数欄
１０６４　アクセス回数欄
１０６５　容量欄
１０６６　デステージ制限欄
１１２１　制御プログラム
１１２２　メールボックス
１１２３、１６４１　転送リスト
１６１０　ＦＭ側チャネル
１６９０　緊急デステージ領域
１７１１　リードキャッシュディレクトリ情報
１７１２　ライトキャッシュディレクトリ情報

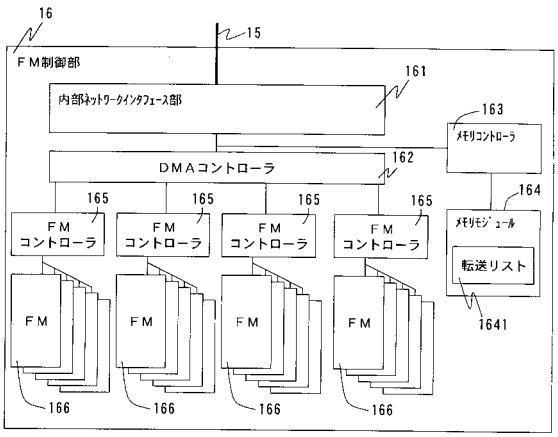
【図１】


2
2
3
SANスイッチ
4
4
50
50
60
60
11
11
16
16
14
14
チャネル制御部
チャネル制御部
ＦＭ制御部
ＦＭ制御部
ディスク制御部
ディスク制御部
12
12
内部スイッチ部
内部スイッチ部
15
13
17
13
17
キャッシュメモリ部
制御メモリ部
キャッシュメモリ部
制御メモリ部
ストレージ制御装置
1

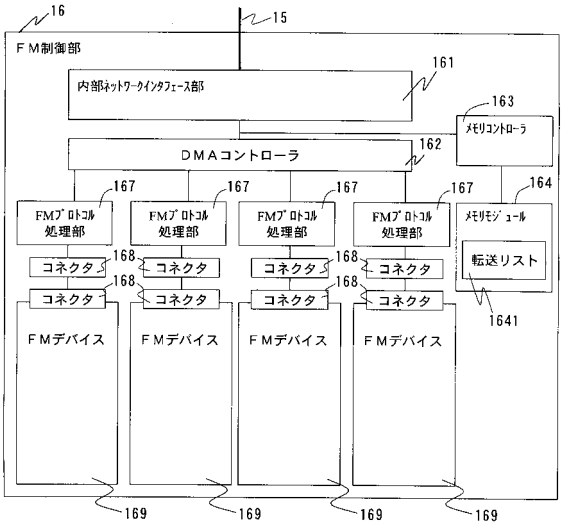
【図２】


4:チャネル
4:チャネル
4:チャネル
4:チャネル
11
チャネル制御部
111
プロセッサ
プロセッサ
プロセッサ
プロセッサ
114
チャネルプロトコル処理部
チャネルプロトコル処理部
チャネルプロトコル処理部
チャネルプロトコル処理部
115
115
115
115
112
メモリモジュール
1121
制御プログラム
メールボックス
1122
転送リスト
1123
113
周辺処理部
116
117
内部ネットワークインタフェース部
15

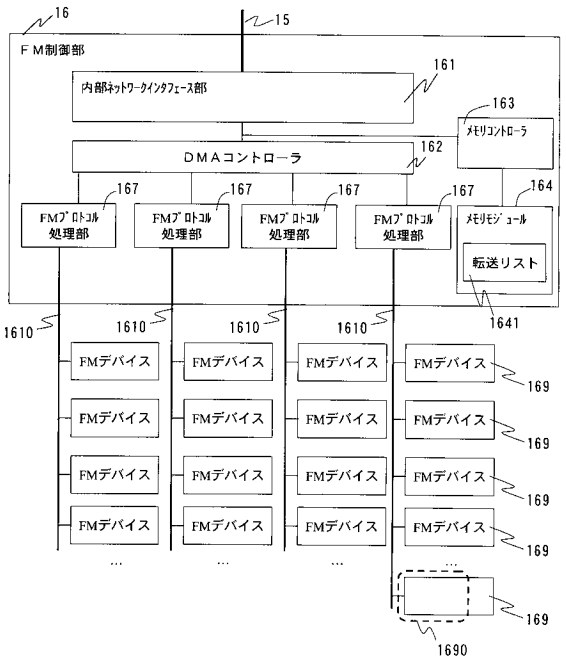
【図３】


16
15
ＦＭ制御部
161
内部ネットワークインタフェース部
163
メモリコントローラ
DMAコントローラ
162
165
165
165
165
164
ＦＭコントローラ
ＦＭコントローラ
ＦＭコントローラ
ＦＭコントローラ
メモリモジュール
転送リスト
ＦＭ
ＦＭ
ＦＭ
ＦＭ
1641
166
166
166
166

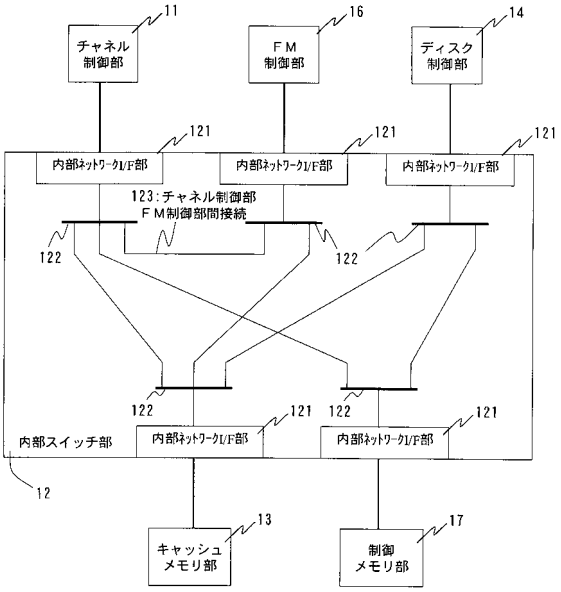
【図４】


16
15
ＦＭ制御部
161
内部ネットワークインタフェース部
163
メモリコントローラ
DMAコントローラ
162
167
167
167
167
164
FMプロトコル処理部
FMプロトコル処理部
FMプロトコル処理部
FMプロトコル処理部
メモリモジュール
168
168
168
168
コネクタ
コネクタ
コネクタ
コネクタ
転送リスト
168
168
168
168
コネクタ
コネクタ
コネクタ
コネクタ
ＦＭデバイス
ＦＭデバイス
ＦＭデバイス
ＦＭデバイス
1641
169
169
169
169

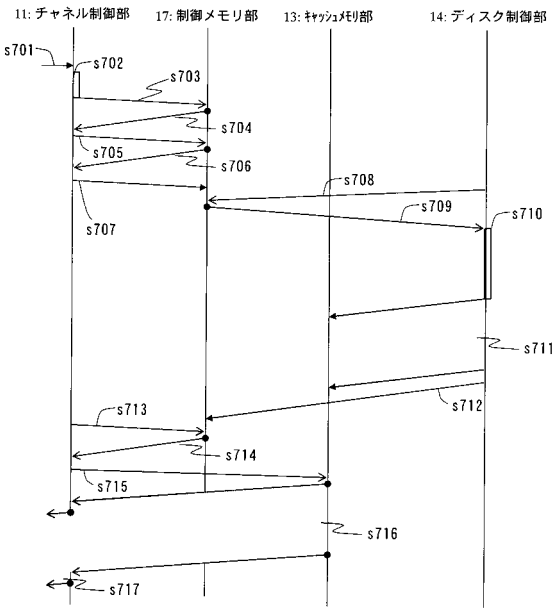
【図５】


16
15
ＦＭ制御部
161
内部ネットワークインタフェース部
163
メモリコントローラ
162
ＤＭＡコントローラ
167
FMプロトコル処理部
164
メモリモジュール
転送リスト
1641
1610
FMデバイス
169
1690

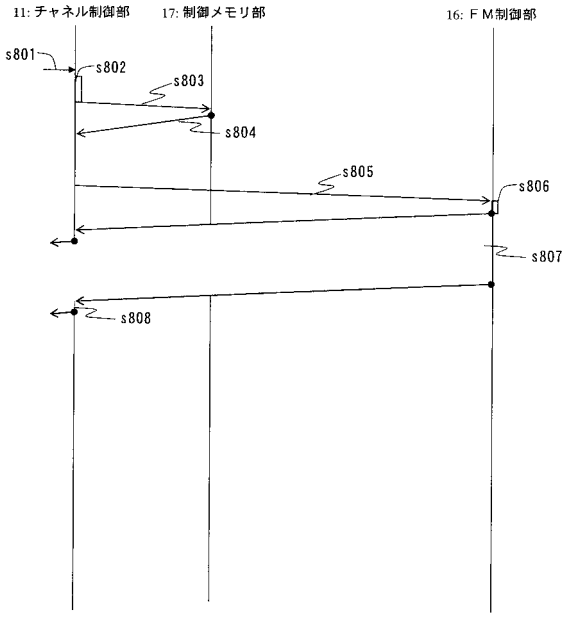
【図６】


11
チャネル制御部
16
ＦＭ制御部
14
ディスク制御部
121
内部ネットワークI/F部
123：チャネル制御部ＦＭ制御部間接続
122
内部スイッチ部
12
13
キャッシュメモリ部
17
制御メモリ部

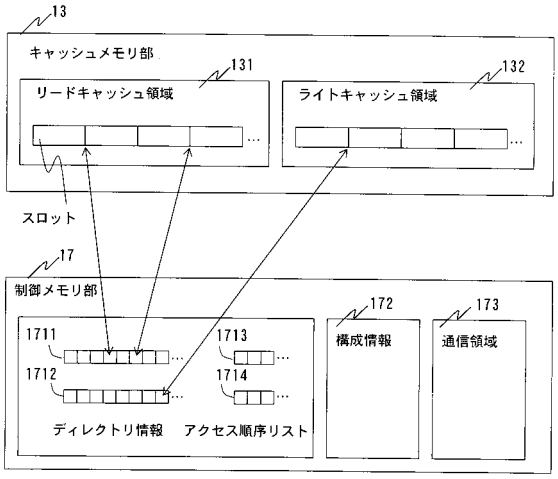
【図７】


11: チャネル制御部
17: 制御メモリ部
13: ｷｬｯｼｭﾒﾓﾘ部
14: ディスク制御部
s701
s702
s703
s704
s705
s706
s707
s708
s709
s710
s711
s712
s713
s714
s715
s716
s717

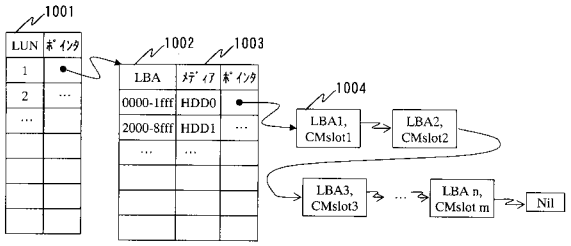
【図８】


11: チャネル制御部
17: 制御メモリ部
16: ＦＭ制御部
s801
s802
s803
s804
s805
s806
s807
s808

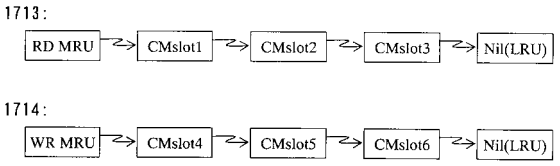
【図９】



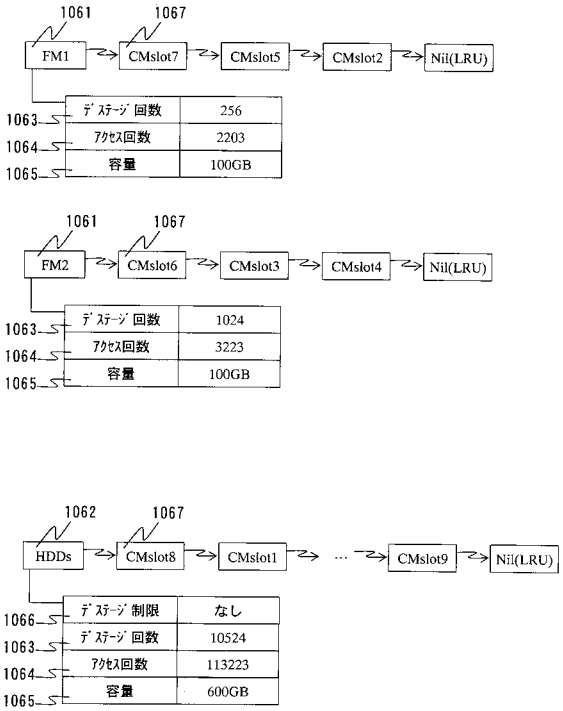
【図１０Ａ】



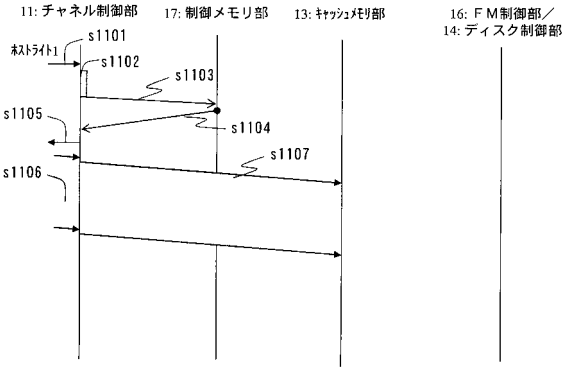
【図１０Ｂ】



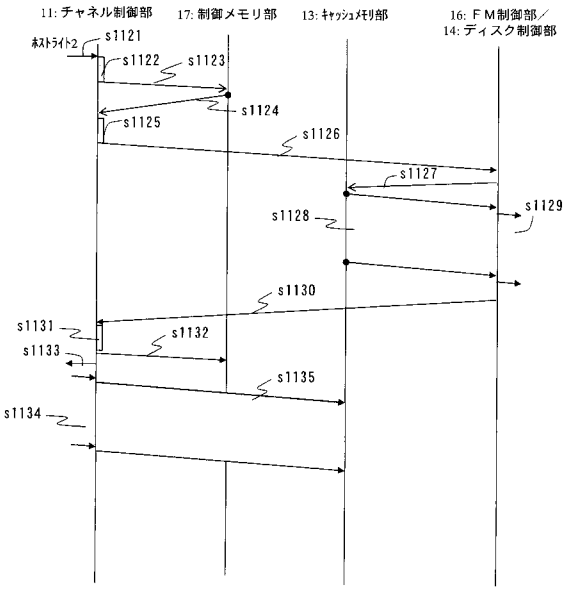
【図１０Ｃ】



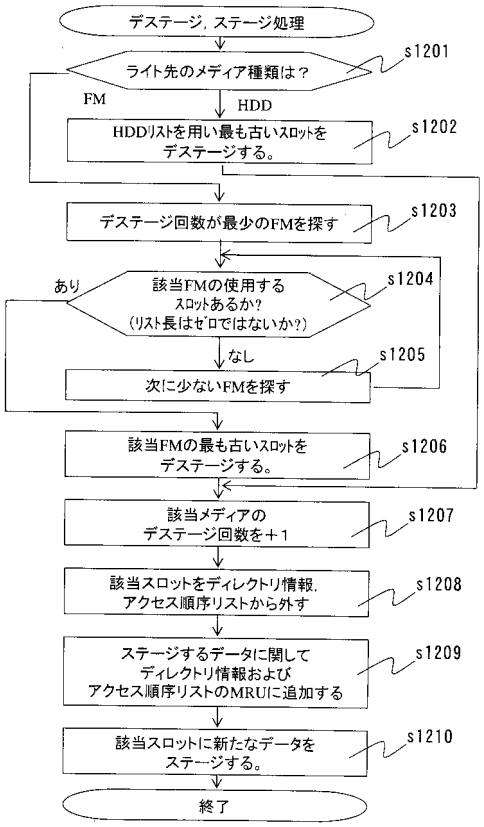
【図１１Ａ】



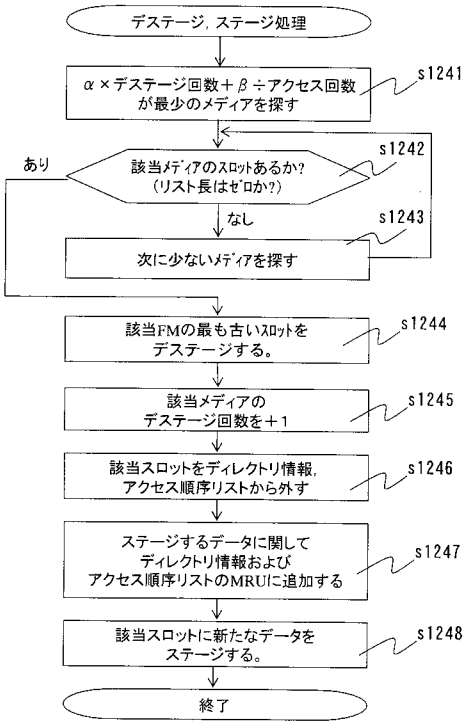
【図１１Ｂ】



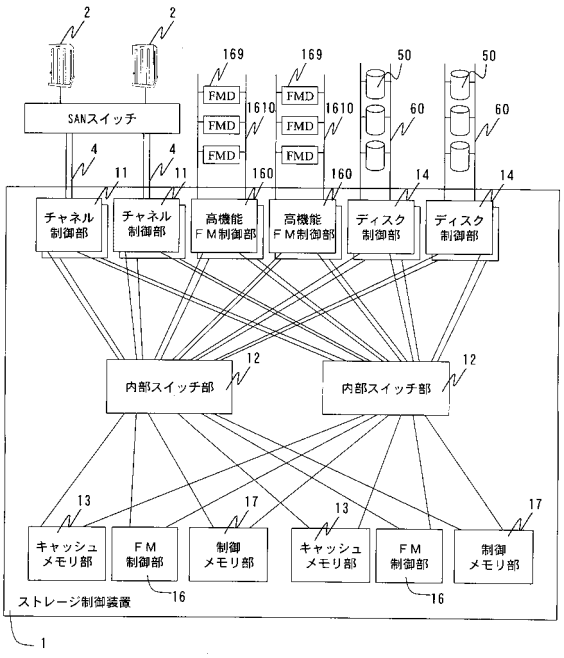
【図１２Ａ】



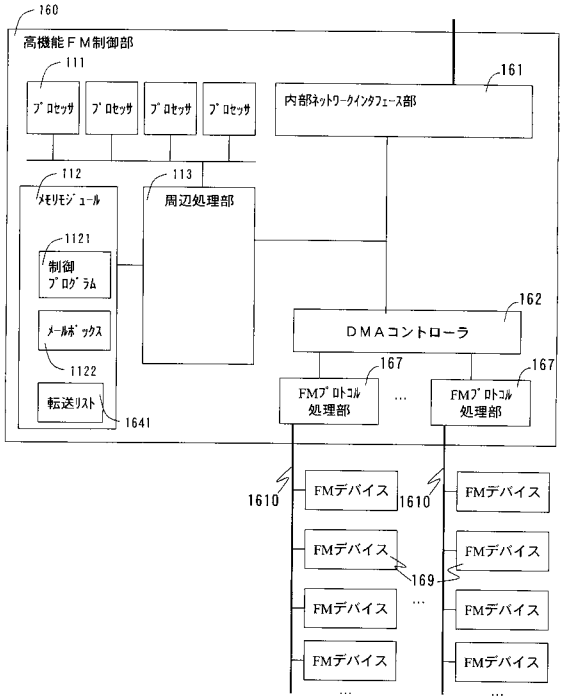
【図１２Ｂ】



【図１３】



【図１４】


160
高機能ＦＭ制御部
111
ﾌﾟﾛｾｯｻ
ﾌﾟﾛｾｯｻ
ﾌﾟﾛｾｯｻ
ﾌﾟﾛｾｯｻ
161
内部ﾈｯﾄﾜｰｸｲﾝﾀﾌｪｰｽ部
112
ﾒﾓﾘﾓｼﾞｭｰﾙ
113
周辺処理部
1121
制御
ﾌﾟﾛｸﾞﾗﾑ
ﾒｰﾙﾎﾞｯｸｽ
1122
転送ﾘｽﾄ
1641
162
ＤＭＡコントローラ
167
FMﾌﾟﾛﾄｺﾙ
処理部
167
FMﾌﾟﾛﾄｺﾙ
処理部
1610
FMデバイス
FMデバイス
FMデバイス
FMデバイス
1610
FMデバイス
FMデバイス
FMデバイス
FMデバイス
169

フロントページの続き

(72)発明者　藤林　昭
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社　日立製作所　システム開発研究所内

審査官　菅原　浩二

(56)参考文献　特開２００４－０２１８１１（ＪＰ，Ａ）
特開２０００－２７６４０２（ＪＰ，Ａ）
特開２００５－２２２５５０（ＪＰ，Ａ）
特開２００３－２２８５１３（ＪＰ，Ａ）
特開２００３－２７１４４４（ＪＰ，Ａ）
特開２００４－０７８９６３（ＪＰ，Ａ）
特開平０６－０７５８３６（ＪＰ，Ａ）
特開２００４－１０２５３９（ＪＰ，Ａ）
特開２００５－１１５６０３（ＪＰ，Ａ）
特開２００５－１１５８５７（ＪＰ，Ａ）
特表２００７－５２５７５３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　　３／０６