【公報種別】特許法第１７条の２の規定による補正の掲載
【部門区分】第６部門第３区分
【発行日】平成18年1月5日(2006.1.5)

【公表番号】特表2005-508551(P2005-508551A)
【公表日】平成17年3月31日(2005.3.31)
【年通号数】公開・登録公報2005-013
【出願番号】特願2003-542490(P2003-542490)
【国際特許分類】
Ｇ０６Ｆ　3/06　(2006.01)
Ｇ０６Ｆ　12/00　(2006.01)
Ｇ０６Ｆ　13/28　(2006.01)
【ＦＩ】
Ｇ０６Ｆ　3/06　３０４Ｂ
Ｇ０６Ｆ　3/06　５４０
Ｇ０６Ｆ　12/00　５１４Ｅ
Ｇ０６Ｆ　12/00　５３１Ｄ
Ｇ０６Ｆ　12/00　５４５Ａ
Ｇ０６Ｆ　13/28　３１０Ａ

【手続補正書】
【提出日】平成17年9月13日(2005.9.13)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
記憶アレイを含む記憶システムにおいてデータをミラー化するための方法であって、
第１のダイレクトメモリアクセスエンジンを含む第１の制御装置管理モジュールであって、前記記憶アレイに関する読み出し／書き込み動作を制御するための前記第１の制御装置管理モジュールを提供するステップ、
第２のダイレクトメモリアクセスエンジンを含む第２の制御装置管理モジュールであって、前記記憶アレイに関する読み出し／書き込み動作を制御し、前記第１のダイレクトメモリアクセスエンジンが前記第２の制御装置管理モジュールにデータをミラー化するのに用いられ、及び前記第２のダイレクトメモリアクセスエンジンが前記第１の制御装置管理モジュールにデータをミラー化するのに用いられる前記第２の制御装置管理モジュールを提供するステップ、及び
前記第１のダイレクトメモリアクセスエンジンを用いて、前記第１の制御装置管理モジュールから前記第２の制御装置管理モジュールへ第１のデータをミラー化するステップ、
を含む方法。
【請求項２】
前記第１の制御装置管理モジュールが第１のプロセッサを含み、前記ミラー化するステップが、前記第１のプロセッサを用いて前記ミラー化するステップが行われるかどうかを判定するステップを含む請求項１に記載の方法。
【請求項３】
前記第２の制御装置管理モジュールが第２のプロセッサを含み、前記ミラー化するステップが、前記第２のプロセッサから独立して前記第１のデータをミラー化するステップを含む請求項１に記載の方法。

【請求項４】
前記第２の制御装置管理モジュールが第２のプロセッサを含み、前記第２のプロセッサが、前記記憶アレイに関する前記読み出し／書き込み動作の前記制御するステップに用いられ、前記第２のプロセッサの割り込みを回避しながら前記ミラー化するステップが行われる請求項１に記載の方法。
【請求項５】
前記第２の制御装置管理モジュールが第２のプロセッサを含み、前記ミラー化するステップの間、前記第２のプロセッサが、第１の読み出し動作及び第１の書き込み動作のうち少なくとも１つを制御するのに用いられるように動作可能である請求項１に記載の方法。
【請求項６】
前記第２の制御装置管理モジュールが第２のプロセッサを含み、前記第２のダイレクトメモリアクセスエンジン及び前記第２のプロセッサを用いない間、前記ミラー化するステップが、前記第１のダイレクトメモリアクセスエンジンを用いて実行される請求項１に記載の方法。
【請求項７】
前記第２の制御装置管理モジュールが第２のプロセッサを含み、前記ミラー化するステップが、前記第２のプロセッサを用いることなく前記第２の制御装置管理モジュールの不揮発性メモリに前記第１のデータを記憶するステップを含む請求項１に記載の方法。
【請求項８】
前記第２の制御装置管理モジュールが不揮発性メモリを含み、前記ミラー化するステップが、
第１に、前記第１のデータを受信する前記不揮発性メモリの内容の部分を無効として記録し、第１のＤＭＡトランザクションにおいて前記不揮発性メモリに前記第１のデータを転送するステップ、及び
第２に、第２のＤＭＡトランザクションにおいて前記不揮発性メモリの内容の前記部分を有効として記録するステップと、を含む請求項１に記載の方法。
【請求項９】
前記第２の制御装置管理モジュールが不揮発性メモリを含み、前記ミラー化するステップが、
第１の文字列を前記不揮発性メモリの第１の記憶領域に記憶し、前記第１のデータを前記不揮発性メモリに転送し、前記不揮発性メモリの第２の記憶領域に前記第１の文字列を記憶するステップを含む請求項１に記載の方法。
【請求項１０】
前記ミラー化するステップが単一のＤＭＡトランザクションを用いて実行される請求項９に記載の方法。
【請求項１１】
前記第１の制御装置管理モジュールが、第１のプロセッサを含み、更に、前記第１のプロセッサを用い、かつ、パリティを判定するためにＸＯＲエンジンを用いて、前記記憶アレイに情報を記憶するステップを含む、請求項１に記載の方法。
【請求項１２】
記憶アレイを含む記憶システムにおいてデータをミラー化するための装置であって、
第１の制御装置管理モジュールは第１のプロセッサ及び第１のダイレクトメモリアクセスエンジンを含み、前記第１のプロセッサは記憶アレイに関する読み出し動作及び書き込み動作を制御するのに使用され、前記第１のダイレクトメモリアクセスエンジンは前記第１の制御装置管理モジュールによって受信されるデータを記憶するのに使用され、
第２の制御装置管理モジュールが第２のプロセッサ及び第２のダイレクトメモリアクセスエンジンを含み、前記第２のプロセッサは記憶アレイに関わる読み出し動作及び書き込み動作を制御するのに使用され、前記第２のダイレクトメモリアクセスエンジンは前記第２の制御装置管理モジュールによって受信されるデータを記憶するのに使用され、
第１のデータは前記第１の制御装置管理モジュールによってホストから受信され、前記

第１のデータは前記第２のプロセッサの割り込みを回避しながら、前記第１のダイレクトメモリアクセスエンジンを用いて、前記第１の制御装置管理モジュールから前記第２の制御装置管理モジュールへミラー化される装置。
【請求項１３】
前記第１の制御装置管理モジュールが不揮発性メモリを含み、前記第１のデータが前記不揮発性メモリに記憶される請求項１２に記載の装置。
【請求項１４】
前記第１のダイレクトメモリアクセスエンジンが前記第１のプロセッサから分離しているが通信し、前記第１のプロセッサは前記第１のダイレクトメモリアクセスエンジンを用いて前記第１のデータのミラー化を開始する請求項１２に記載の装置。
【請求項１５】
前記第１の制御装置管理モジュールがフィールドプログラマブルゲートアレイを有し、前記第１のダイレクトメモリアクセスエンジンが少なくともその部分と通信する請求項１２に記載の装置
【請求項１６】
前記装置が、更に、
第１の共有経路を有する第１のチャネルインタフェースモジュールを有し、前記第１のチャネルインタフェースモジュールは前記第１の制御装置管理モジュールと通信し、前記第１の共有経路が前記第１の制御装置管理モジュールと前記第２の制御装置管理モジュールとの間で前記第１のデータを転送するのに用いられる請求項１２に記載の装置。
【請求項１７】
前記装置が、更に、
前記第１のチャネルインタフェースモジュール及び前記第１の制御装置管理モジュールを相互接続する受動バックプレーンを有する請求項１６に記載の装置。
【請求項１８】
前記第２のプロセッサが、前記第１のデータが前記第２の制御装置管理モジュールにミラー化されている間、前記第２の制御装置管理モジュールに関連する動作を制御する請求項１２に記載の装置。
【請求項１９】
前記第１のデータが、前記第２のダイレクトメモリアクセスエンジンから独立して前記第２の制御装置管理モジュールにミラー化される請求項１２に記載の装置。
【請求項２０】
前記第２の制御装置管理モジュールが不揮発性メモリを有し、前記第１のデータが前記第２のプロセッサから独立して前記不揮発性メモリに記憶される請求項１２に記載の装置。
【請求項２１】
前記第２の制御装置管理モジュールが不揮発性メモリを有し、前記第１のダイレクトメモリアクセスエンジンは、前記第１のデータを受信するべき前記不揮発性メモリの少なくとも部分に対して、前記部分が無効であるという指示を提供するのに用いられ、前記第１のデータが不揮発性メモリによって受信された後、前記第１のダイレクトメモリアクセスエンジンが前記部分を有効として記録するのに用いられる請求項１２に記載の装置。
【請求項２２】
前記第２の制御装置管理モジュールが少なくとも第１の記憶領域及び第２の記憶領域を有する不揮発性メモリを有し、
前記第１のダイレクトメモリアクセスエンジンは、前記第１のデータが前記不揮発性メモリによって受信される前に、前記第１の記憶領域での記憶のために第１の文字列を提供すること、及び、前記第１のデータが前記不揮発性メモリによって受信された後に、前記第２の記憶領域での記憶のために前記第１の文字列を提供することに用いられる請求項１２に記載の装置。
【請求項２３】

前記第１のダイレクトメモリアクセスエンジンが、単一のダイレクトメモリアクセストランザクションにおいて、前記第１の文字列を提供し、前記第１のデータを転送し、及び、前記第１の文字列を提供するよう、動作可能である請求項２１に記載の装置。
【手続補正２】
【補正対象書類名】明細書
【補正対象項目名】０００６
【補正方法】変更
【補正の内容】
【０００６】
制御装置３０、３４は、入出力モジュール－１　４２と入出力モジュール－２　４６の２つの入出力モジュールに接続されるファイバチャネルバス３８に接続される。各制御装置３０、３４は、ＣＰＵサブシステム５０、２倍のデータ転送速度の（ＤＤＲ：Double Data Rate）メモリ５４、制御論理５８、２つのホストポート６２ａ、６２ｂのデュアル(dual)ポートファイバチャネル接続、及び２つのディスクポート６６ａ、６６ｂのデュアルポートファイバチャネル接続を含む。ＣＰＵサブシステム５０は、データの分散化と、読み取り及び書き込みコマンドの開始及び実行とを含む、ディスクアレイへのデータの記憶に必要なタスクを実行する。ＤＤＲメモリ５４は、データ及びその他の情報のための不揮発性記憶領域である。制御論理５８は、ＣＰＵサブシステム５０と、ＤＤＲメモリ５４と、ホストポート６２ａ、６２ｂ及びディスクポート６６ａ、６６ｂとの仲介等のいくつかの機能を実行する。制御論理５８は、排他的論理和（ＸＯＲ：Exclusive OR）エンジン等のパリティ生成機能を含むその他の機能をも有していても良い。ホストポート６２ａ、６２ｂ及びディスクポート６６ａ、６６ｂは、ファイバチャネルバックプレーン３８との通信を提供する。入出力モジュール４２、４６は、ポートバイパス回路(port bypass circuit)としても良く知られている、各ホストポート１４、１８及び各ディスクポート２２、２６を各制御装置３０、３４に接続するために機能するリンク復元回路（ＬＲＣ：Link Resiliency Circuit）７０を含む。これによって、両制御装置３０、３４は、両ホストポート１４、１８及び両ディスクポート２２、２６にアクセス可能となる。
【手続補正３】
【補正対象書類名】明細書
【補正対象項目名】００１６
【補正方法】変更
【補正の内容】
【００１６】
本発明によれば、記憶アレイを有する記憶システムにおいてデータをミラー化するための方法及び装置が提供される。前記装置は、第１のプロセッサ及び第１のダイレクトメモリアクセスエンジンを含む第１の制御装置管理モジュールを含む。前記第１のプロセッサは、記憶アレイに関わる読み出し動作及び書き込み動作を制御するのに用いられる。前記第１のダイレクトメモリアクセスエンジンは、前記第１の制御装置管理モジュールによって受信されるデータを記憶するのに用いられる。前記装置はまた、第２のプロセッサ及び第２のダイレクトメモリアクセスエンジンを含む第２の制御装置管理モジュールを有する。前記第２のプロセッサは、記憶アレイに関する読み出し動作及び書き込み動作を制御するのに用いられる。前記第２のダイレクトメモリアクセスエンジンは、前記第２の制御装置管理モジュールから前記第１の制御装置管理モジュールにデータを転送するのに用いられることもできる。前記データは、前記第２のプロセッサの割り込みを回避しつつ、前記第１のダイレクトメモリアクセスエンジンを用いて前記第１の制御装置管理モジュールから前記第２の制御装置管理モジュールへミラー化される。前記第１のダイレクトメモリアクセスエンジンは前記第１のプロセッサから分離しているが通信し、前記第１のプロセッサは前記第１のダイレクトメモリアクセスエンジンを用いて前記データのミラー化を制御する。ある実施の形態において、前記第１の制御装置管理モジュールは、フィールドプログラマブルゲートアレイ(field programmable gate array)を有する。前記第１のダイレ

クトメモリアクセスエンジンは、フィールドプログラマブルゲートアレイの少なくとも一部と通信し、前記第１のダイレクトメモリアクセスエンジンは、フィールドプログラマブルゲートアレイの１部であることもできる。
【手続補正４】
【補正対象書類名】明細書
【補正対象項目名】００１７
【補正方法】変更
【補正の内容】
【００１７】
ある実施の形態において、前記装置は、第１の共有経路を有する第１のチャネルインタフェースモジュールを含む。前記第１のチャネルインタフェースモジュールは前記第１の制御装置管理モジュールと通信し、前記第１の共有経路は、前記第１の制御装置管理モジュールと前記第２の制御装置管理モジュールとの間でデータを転送するのに用いられる。受動バックプレーンは、前記第１のチャネルインタフェースモジュール及び前記第１の制御装置管理モジュールを相互接続する。前記第２のプロセッサは、前記データが前記第２の制御装置管理モジュールにミラー化されている間、前記第２の制御装置管理モジュールに関わる動作を制御する。前記データは、前記第２のダイレクトメモリアクセスエンジンから独立して前記第２の制御装置管理モジュールにミラー化される。前記第２の制御装置管理モジュール内には不揮発性メモリがあり、前記データは前記第２のプロセッサから独立して前記不揮発性メモリ内に記憶され得る。前記第１のダイレクトメモリアクセスエンジンは、前記データが記憶される前記不揮発性メモリの部分を無効として記録し、前記不揮発性メモリに前記データを転送する。前記データが記憶された前記不揮発性メモリの前記部分は、次に有効として記録される。
【手続補正５】
【補正対象書類名】明細書
【補正対象項目名】００２１
【補正方法】変更
【補正の内容】
【００２１】
ネットワーク記憶装置１００は、１つ以上のチャネルインタフェースモジュール（ＣＩＭ：Channel Interface Module）を有する。図２に示される実施の形態において、ＣＩＭ－１　１３６とＣＩＭ－２　１４０の、２つのＣＩＭがあるが、この数はネットワーク記憶装置１００が使用される構成及びアプリケーションに応じて変化しても良いことは理解されるであろう。各ＣＩＭ１３６、１４０は、２つのＣＩＭバスインタフェースポート１４４ａ、１４４ｂを有する。各ＣＩＭ１３６、１４０上では、１つのＣＩＭバスインタフェースポート１４４ａがＣＩＭバス接続１４６を介してＣＭＭ－Ａ１０４に接続される１つのバスと接続し、１つのＣＩＭバスインタフェースポート１４４ｂがＣＩＭバス接続１４６を介してＣＭＭ－Ｂ１０８に接続される１つのバスと接続する。図２に示される実施の形態において、ＣＩＭ－１　１３６が第１のデータバス１２０及び第３のデータバス１２８に接続し、ＣＩＭ－２　１４０が第２のデータバス１２４及び第４のデータバス１３２に接続する。各ＣＩＭ１３６、１４０は、ホストコンピュータ（図示せず）に接続されるホストチャネル１５２に接続する２つのホストポート１４８を有する。また、各ＣＩＭ１３６、１４０は、１つ以上の記憶装置（図示せず）に接続されるディスクチャネル１５８に接続する２つのディスクポート１５６をも有する。記憶装置は、ＲＡＩＤアレイ等の記憶アレイであっても良い。下記に詳述するように、別の実施の形態において、ＣＩＭは、要求されるアプリケーション及びチャネルインタフェースに応じて、複数のホストポート或いは複数のディスクポートを含んでも良い。
【手続補正６】
【補正対象書類名】明細書
【補正対象項目名】００２６

【補正方法】変更
【補正の内容】
【００２６】
メモリインタフェース１７６は、メモリ１６４とインタフェースＦＰＧＡ１６８間のインタフェースとしての役割を果たす。ＸＯＲエンジン１８０は、書き込まれるデータに関するパリティ情報を得るために、記憶されるデータに関するＸＯＲ動作を実行するように働く。ＸＯＲエンジン１８０は、ディスクアレイ内の故障ドライブからデータを復元するためにパリティ情報の使用が必要とされる状況においても使用される。ＸＯＲエンジン１８０は、ＰＣＩインタフェース１７２を通じてＣＰＵサブシステム１６０に接続する。複数のデータＦＩＦＯ１９２はメモリインタフェース１７６及びブリッジコア１８４に接続し、順次バックプレーンインタフェース１１２に接続する。複数のデータＦＩＦＯは、読み取り及び書き込み動作を処理するためにＣＭＭ１０４によって使用される待ち行列として働く。下記に詳述されるように、ＤＭＡエンジン１８８は、ＣＭＭが冗長性を提供するように動作している時、別のＣＭＭにＤＭＡデータを提供するように働く。ある実施の形態において、ＤＭＡエンジン１８８はまた、ＸＯＲ動作を実行するために、ＸＯＲエンジン１８０と関連して用いられ、メモリ１６４内の２つの領域からデータを読み出し、ＸＯＲエンジン１８０にデータを提供し、メモリ１６４内の第３の領域にＸＯＲエンジンの出力を書き込む。
【手続補正７】
【補正対象書類名】明細書
【補正対象項目名】００２８
【補正方法】変更
【補正の内容】
【００２８】
チャネルインタフェース２０４は、交換ＰＣＩＸ経路２０８、２１２をホストポート１４８及びディスクポート１５６に接続する。チャネルインタフェース２０４は、交換ＰＣＩＸ経路２０８、２１２上で送信されるデータを監視し、データがホストポート１４８或いはディスクポート１５６のどちらに経路付けされるかを判定するように動作可能である。この監視及び経路付けは、データを適切なディスク又はホスト位置に通過させ、ミラー化データをホスト又はディスクポート１４８、１５６にまで通過させない。チャネルインタフェース２０４は、アプリケーションにとって最適なチャネル媒体上での通信を可能にする。例えば、ホストチャネル１５２及びディスクチャネル１５８がファイバチャネルを使用する場合、チャネルインタフェース２０４は交換ＰＣＩＸ経路２０８、２１２とファイバチャネルとの間のインタフェースの役割を果たすであろう。同様に、ホストチャネル１５２及びディスクチャネル１５８がＳＣＳＩチャネルを使用する場合、チャネルインタフェース２０４は交換ＰＣＩＸ経路２０８、２１２とＳＣＳＩチャネルとの間のインタフェースの役割を果たすであろう。ホストチャネル１５２及びディスクチャネル１５８の両方が同じチャネル媒体を使用する場合、ＣＩＭ１３６は、ホストポート１４８及びディスクポート１５６を使用してホストチャネル１５２及びディスクチャネル１５８の両方との通信のため用いられることができる。
【手続補正８】
【補正対象書類名】明細書
【補正対象項目名】００３０
【補正方法】変更
【補正の内容】
【００３０】
図４に示される実施の形態において、バスインタフェースポート１４４及びＣＩＭバス接続１４６を通じて、第１の交換ＰＣＩＸ経路２０８が第１のデータバス１２０と通信し、第２の交換ＰＣＩＸ経路２１２が第３のデータバス１２８と通信する。下記に詳述するように、ＰＣＩＸブリッジ２００は、一方のＣＭＭが他方のＣＭＭと通信するための通信

経路として使用されても良い。ネットワーク制御装置上に存在する残りのＣＩＭに対しても、同様の構成が使用されると理解されるであろう。例えば、図２に示される実施の形態において、ＣＩＭ－２　１４０は第２のデータバス１２４及び第４のデータバス１３２に接続され、従って、ＣＩＭ－２　１４０は、第２のデータバス１２４及び第４のデータバス１３２とそれぞれ通信する交換ＰＣＩＸ経路２０８、２１２を有するであろう。同様に、２つを超えるＣＩＭが存在する場合、アプリケーションによる要求に応じて受動バックプレーン１１６上の適切なバスと通信するよう構成されるであろう。

【手続補正９】
【補正対象書類名】明細書
【補正対象項目名】００３３
【補正方法】変更
【補正の内容】
【００３３】
当業者によって理解されるように、冗長制御装置は、記憶サブシステムに取り付けられた２つの制御装置間のデータのミラー化を必要とする。これは、制御装置がホストコンピュータからデータを受信し、データをキャッシュし、データが書き込まれたというメッセージをホストコンピュータに送信する、書き戻しキャッシュ(write back cache)の使用による。従って、データが実際に制御装置に格納され、ディスクアレイ内のドライブに書き込まれるためにそこで待機している時、ホストコンピュータは、データが書き込まれたと判定する。故障の際に、このデータが失われないよう確保するのを助けるため、冗長制御装置はこのデータを他の制御装置にミラー化し、従って、他の制御装置にデータの別のコピーができる。このことは、キャッシュの一貫性として既知である。ある実施の形態において、ＣＭＭ１０４、１０８は、ネットワーク記憶装置１００ａにキャッシュの一貫性を提供するためにデータをミラー化する。これは、ＣＭＭ－Ａ１０４とＣＭＭ－Ｂ１０８との間にＤＭＡ経路を設けることによって実現できる。これは、図３に関して上述したように、インタフェースＦＰＧＡ１６８内のＤＭＡエンジン１８８を提供することによって達成されることが出来、共有経路は、図４に関して上述したように、ＰＣＩＸブリッジ２００を利用する。各ＣＭＭ１０４、１０８は、他のＣＭＭにデータを送信するためにこのＤＭＡ経路を用いる。ＤＭＡ経路を利用することによって、２つのＣＭＭ１０４、１０８は、ホストチャネル１５２又はディスクチャネル１５８を使用する必要なくデータをミラー化でき、従って、ディスクチャネル１５８又はホストチャネル１５２のチャネル帯域幅はデータをミラー化しているＣＭＭ１０４、１０８によって消費されない。

【手続補正１０】
【補正対象書類名】明細書
【補正対象項目名】００３６
【補正方法】変更
【補正の内容】
【００３６】
他の実施の形態において、図１３に示された２つのＤＭＡトランザクションは、単一の順序付けられたＤＭＡトランザクションに組み合わせられる。この実施の形態において、各ＣＭＭ１０４、１０８内のＤＤＲメモリ１６４は、ユーザデータに関連するメタデータを記憶するための２つの記憶領域を有する。ＤＭＡ転送を開始する場合、第１のユニークな文字列が第１の記憶領域に記憶され、その後に転送されるデータが続く。ＤＭＡ転送の最後に、第１のユニークな文字列は第２の記憶領域に記憶される。ＣＭＭ１０４、１０８が故障から回復されなければならない際には、第１の及び第２の記憶領域内に記憶された文字列が比較される。比較はＣＭＭの故障が起こった際にのみ実行され、従って、通常の動作の間、パフォーマンスを低下させない。文字列が一致した場合、ミラー化されたデータが有効であるということを示す。