(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5358736号
(P5358736)
(45)発行日　平成25年12月4日(2013.12.4)　(24)登録日　平成25年9月6日(2013.9.6)

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

請求項の数 13 (全 30 頁)

(21)出願番号　特願2012-516982(P2012-516982)
(86)(22)出願日　平成21年11月10日(2009.11.10)
(65)公表番号　特表2012-531656(P2012-531656A)
(43)公表日　平成24年12月10日(2012.12.10)
(86)国際出願番号　PCT/JP2009/005995
(87)国際公開番号　WO2011/058598
(87)国際公開日　平成23年5月19日(2011.5.19)
審査請求日　平成23年12月27日(2011.12.27)

(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　110000279
特許業務法人ウィルフォート国際特許事務所
(72)発明者　中村　崇仁
神奈川県横浜市戸塚区吉田町２９２番地
株式会社日立製作所　横浜研究所内
(72)発明者　新井　政弘
神奈川県横浜市戸塚区吉田町２９２番地
株式会社日立製作所　横浜研究所内

最終頁に続く

(54)【発明の名称】複数のコントローラを備えたストレージシステム

(57)【特許請求の範囲】
【請求項１】
第１のコントローラと、
前記第１のコントローラに第１パスを介して接続された第２のコントローラと
を備え、
ホスト装置が発行したＩ／Ｏ（Input/Output）コマンドを前記第１及び第２のコントローラのいずれかが受け付けた場合、そのＩ／Ｏコマンドを受けた第１又は第２のコントローラが、そのＩ／Ｏコマンドに従う処理であるＩ／Ｏ処理を行い、前記Ｉ／Ｏ処理において、そのＩ／Ｏコマンドに従うデータのＩ／Ｏを記憶デバイスに対して行い、
前記第１のコントローラが、
データ転送を制御する回路である第１の中継回路と、
前記第１の中継回路に第１の第２パスを介して接続された第１のプロセッサと
を有し、
前記第２のコントローラが、
データ転送を制御する回路であり、前記第１パスを介して前記第１の中継回路に接続された第２の中継回路と、
前記第２の中継回路に第２の第２パスを介して接続された第２のプロセッサと
を有し、
前記第１のプロセッサが、前記第１の中継回路を介することなく第１の第３パスを介して前記第２の中継回路に接続されており、前記第１のコントローラが前記Ｉ／Ｏ処理を行

っている場合、そのＩ／Ｏ処理において、前記第１の第３パスを介して前記第２の中継回路にアクセスし、
前記第２のプロセッサが、前記第２の中継回路を介することなく第２の第３パスを介して前記第１の中継回路に接続されており、前記第２のコントローラが前記Ｉ／Ｏ処理を行っている場合、そのＩ／Ｏ処理において、前記第２の第３パスを介して前記第１の中継回路にアクセスし、
前記第１の第３パスは、前記第１のプロセッサが前記第２の中継回路を直接利用するダイレクトパスであり、前記第２の第３パスは、前記第２のプロセッサが前記第１の中継回路を直接利用するダイレクトパスである、
記憶制御装置。
【請求項２】
請求項１記載の記憶制御装置であって、
前記第１の中継回路が、データの転送を行う回路である第１のデータ転送回路を有し、
前記第２の中継回路が、データの転送を行う回路である第２のデータ転送回路を有し、
前記第１のコントローラが、前記第１の中継回路に接続された第１のメモリと、プロセッサ毎に設けられておりデータ転送のパラメータである転送パラメータが蓄積される第１キューとを有し、
各第１キューは、前記第１のデータ転送回路についてのキューであり、
前記第２のコントローラが、前記第２の中継回路に接続された第２のメモリと、プロセッサ毎に設けられており転送パラメータが蓄積される第２キューとを有し、
各第２キューは、前記第２のデータ転送回路についてのキューであり、
前記転送パラメータは、前記第１及び第２プロセッサによって生成され、データの転送元の記憶領域のアドレスである転送元アドレスと、そのデータの転送先のアドレスである転送先アドレスとを含み、
前記第１のデータ転送回路が、複数の第１キューに蓄積されている転送パラメータの数をそれぞれ記憶する記憶領域である複数の第１インデックスと、前記複数の第１インデックスから一つの第１インデックスを選択する第１セレクタと、前記第１セレクタによって選択された第１インデックスに対応した第１キューから転送パラメータを取得しその転送パラメータが有する転送元アドレス及び転送先アドレスを設定する第１パラメータ取得回路と、前記設定された転送元アドレスが表す記憶領域内のデータを前記設定された転送先アドレスが表す記憶領域に転送する第１転送制御回路とを有し、
前記第２のデータ転送回路が、複数の第２キューに蓄積されている転送パラメータの数をそれぞれ記憶する記憶領域である複数の第２インデックスと、前記複数の第２インデックスから一つの第２インデックスを選択する第２セレクタと、前記第２セレクタによって選択された第２インデックスに対応した第２キューから転送パラメータを取得しその転送パラメータが有する転送元アドレス及び転送先アドレスを設定する第２パラメータ取得回路と、前記設定された転送元アドレスが表す記憶領域内のデータを前記設定された転送先アドレスが表す記憶領域に転送する第２転送制御回路とを有し、
前記第１及び第２プロセッサのうちのＩ／Ｏコマンドを受領したプロセッサである対象プロセッサが、下記の（Ａ）及び（Ｂ）：
（Ａ）データの転送先の記憶領域が、前記対象プロセッサを有するコントローラである対象コントローラに存在するか否か；
（Ｂ）前記対象プロセッサに対応した第１キューについての未完了の転送パラメータの数である第１の数と、前記対象プロセッサに対応した第２キューについての未完了の転送パラメータの数である第２の数、
に基づいて、前記第１及び第２のデータ転送回路のうちのいずれかのデータ転送回路を選択し、選択したデータ転送回路を有するコントローラ内の、前記対象プロセッサに対応したキューに、転送パラメータを格納する、
記憶制御装置。
【請求項３】

請求項２記載の記憶制御装置であって、
前記対象プロセッサが、下記（Ｘ）及び（Ｙ）の条件に適合するデータ転送回路を優先的に選択する、
（Ｘ）前記転送先記憶領域が前記対象コントローラ内にある場合、前記対象コントローラに存在する；
（Ｙ）前記第１の数と前記第２の数とのうちの少ない方に対応する、
記憶制御装置。
【請求項４】
請求項３記載の記憶制御装置であって、
前記第１のメモリが、第１のキャッシュメモリ領域を有し、
前記第２のメモリが、第２のキャッシュメモリ領域を有し、
前記第１のプロセッサが、受け付けたＩ／Ｏコマンドに従うデータの転送先アドレスとして前記第２のキャッシュメモリ領域の任意の位置のアドレスを含んだ転送パラメータを、前記第１の第３パスを介して、前記第１のプロセッサに対応した第２キューに格納し、
前記第２のプロセッサが、受け付けたＩ／Ｏコマンドに従うデータの転送先アドレスとして前記第１のキャッシュメモリ領域の任意の位置のアドレスを含んだ転送パラメータを、前記第２の第３パスを介して、前記第１のプロセッサに対応した第１キューに格納する、
記憶制御装置。
【請求項５】
請求項１記載の記憶制御装置であって、
プロセッサ毎に設けられておりデータ転送のパラメータである転送パラメータが蓄積される第１パラメータ領域と、
プロセッサ毎に設けられており転送パラメータが蓄積される第２パラメータ領域と
を有し、
前記第１の中継回路が、データの転送を行う回路である第１のデータ転送回路を有し、
前記第２の中継回路が、データの転送を行う回路である第２のデータ転送回路を有し、
前記第１パラメータ領域は、前記第１のデータ転送回路についての記憶領域であり、
前記第２パラメータ領域は、前記第２のデータ転送回路についての記憶領域であり、
前記第１のプロセッサが、前記第１のデータ転送回路を利用する場合、前記第１のプロセッサに対応した第１パラメータ領域に転送パラメータを格納し、前記第２のデータ転送回路を利用する場合、前記第１のプロセッサに対応した第２パラメータ領域に転送パラメータを格納し、
前記第２のプロセッサが、前記第２のデータ転送回路を利用する場合、前記第２のプロセッサに対応した第２パラメータ領域に転送パラメータを格納し、前記第１のデータ転送回路を利用する場合、前記第２のプロセッサに対応した第１パラメータ領域に転送パラメータを格納し、
前記第１のデータ転送回路が、いずれかの第１パラメータ領域から転送パラメータを取得し、その転送パラメータに従って、データ転送を実行し、
前記第２のデータ転送回路が、いずれかの第２パラメータ領域から転送パラメータを取得し、その転送パラメータに従って、データ転送を実行する、
記憶制御装置。
【請求項６】
請求項５記載の記憶制御装置であって、
前記第１及び第２プロセッサのうちのＩ／Ｏコマンドを受領したプロセッサである対象プロセッサが、データの転送先の記憶領域が、前記対象プロセッサを有するコントローラである対象コントローラに存在する場合、前記対象コントローラ内のデータ転送回路を選択し、選択したデータ転送回路を有するコントローラ内の、前記対象プロセッサに対応したパラメータ領域に、転送パラメータを格納する、
記憶制御装置。

【請求項７】
請求項５記載の記憶制御装置であって、
前記第１及び第２プロセッサのうちのＩ／Ｏコマンドを受領したプロセッサである対象プロセッサが、前記対象プロセッサに対応した第１パラメータ領域内の未処理の転送パラメータの数と、前記対象プロセッサに対応した第２パラメータ領域内の未処理の転送パラメータの数とのうち少ない方に対応したデータ転送回路を選択し、選択したデータ転送回路を有するコントローラ内の、前記対象プロセッサに対応したパラメータ領域に、転送パラメータを格納する、
記憶制御装置。
【請求項８】
請求項５記載の記憶制御装置であって、
前記第１のコントローラが、前記第１の中継装置に接続された第１のキャッシュメモリ領域を有し、
前記第２のコントローラが、前記第２の中継装置に接続された第２のキャッシュメモリ領域を有し、
前記第１のプロセッサが、前記第２のキャッシュメモリ領域内の任意のアドレスを転送先アドレスとした転送パラメータを、前記第１のプロセッサに対応した第２パラメータ領域に格納し、
前記第２のプロセッサが、前記第１のキャッシュメモリ領域内の任意のアドレスを転送先アドレスとした転送パラメータを、前記第２のプロセッサに対応した第１パラメータ領域に格納する、
記憶制御装置。
【請求項９】
請求項５記載の記憶制御装置であって、
前記第１のパラメータ領域は、Ｉ／Ｏコマンドがライトコマンドである場合の転送パラメータの格納先とされる第１のパラメータライト領域と、Ｉ／Ｏコマンドがリードコマンドである場合の転送パラメータの格納先とされる第１のパラメータリード領域とを有し、
前記第２のパラメータ領域は、Ｉ／Ｏコマンドがライトコマンドである場合の転送パラメータの格納先とされる第２のパラメータライト領域と、Ｉ／Ｏコマンドがリードコマンドである場合の転送パラメータの格納先とされる第２のパラメータリード領域とを有し、
前記第１のデータ転送回路は、前記第１のパラメータライト領域よりも前記第１のパラメータリード領域内の未処理の転送パラメータを優先的に取得し、
前記第２のデータ転送回路は、前記第２のパラメータライト領域よりも前記第２のパラメータリード領域内の未処理の転送パラメータを優先的に取得する、
記憶制御装置。
【請求項１０】
請求項５記載の記憶制御装置であって、
前記第１のデータ転送回路が、複数の第１パラメータ領域に蓄積されている未処理の転送パラメータの数をそれぞれ記憶する記憶領域である複数の第１インデックスと、前記複数の第１インデックスから一つの第１インデックスを選択する第１セレクタと、前記第１セレクタによって選択された第１インデックスに対応した第１パラメータ領域から転送パラメータを取得しその転送パラメータが有する転送元アドレス及び転送先アドレスを設定する第１パラメータ取得回路と、前記設定された転送元アドレスが表す記憶領域内のデータを前記設定された転送先アドレスが表す記憶領域に転送する第１転送制御回路とを有し、
前記第２のデータ転送回路が、複数の第２パラメータ領域に蓄積されている未処理の転送パラメータの数をそれぞれ記憶する記憶領域である複数の第２インデックスと、前記複数の第２インデックスから一つの第２インデックスを選択する第２セレクタと、前記第２セレクタによって選択された第２インデックスに対応した第２パラメータ領域から転送パラメータを取得しその転送パラメータが有する転送元アドレス及び転送先アドレスを設定

する第２パラメータ取得回路と、前記設定された転送元アドレスが表す記憶領域内のデータを前記設定された転送先アドレスが表す記憶領域に転送する第２転送制御回路とを有する、
記憶制御装置。
【請求項１１】
請求項１記載の記憶制御装置であって、
障害の発生を監視する障害監視ユニットを備え、
前記障害監視ユニットは、前記第１のプロセッサの障害を検出した場合、前記第２のプロセッサに、前記第１のプロセッサの障害を通知する、
記憶制御装置。
【請求項１２】
請求項１１記載の記憶制御装置であって、
前記障害監視ユニットは、前記第１の中継装置が有する、
記憶制御装置。
【請求項１３】
請求項１記載の記憶制御装置であって、
第１のスイッチ装置と、第２のスイッチ装置とを有し、
前記第１パスは、前記第１及び第２の中継装置を含んだ複数の中継装置に接続された前記第１のスイッチ装置で実現され、
前記第１の第３パス及び前記第２の第３パスは、前記第１及び第２のプロセッサを含んだ複数のプロセッサに接続された前記第２のスイッチ装置で実現される、
記憶制御装置。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、複数のコントローラを備えたストレージシステムに関する。
【背景技術】
【０００２】
近年、低コストでありながら高性能かつ高機能のストレージシステムが市場から要求されている。ストレージシステムは、一般に、コントローラと物理記憶装置群とに大別することができるが、ストレージシステムを高性能かつ高機能にする方法としては、コントローラを高性能かつ高機能にする方法がある。具体的には例えば、プロセッサの周波数を上げる、コントローラ内でのデータ転送を行うＬＳＩ(Large Scale Integration)の性能を高くするなどの方法が考えられる。
【０００３】
しかし、コントローラを高性能かつ高機能にすると、通常、ストレージシステム自体のコストは上昇する。そのため、コントローラを高性能かつ高機能にすることに代えて、ストレージシステム内での処理を効率的にする必要があると考えられる。
【０００４】
ストレージシステム内での処理の効率化に関する技術が、例えば、特許文献１及び２に開示されている。特許文献１によれば、第１のコントローラと第２のコントローラとの間に、高レスポンスに適した第１種のパスと高スループットに適した第２種のパスとが設けられ、それら二種類のパスが使い分けられる。特許文献２によれば、第１種のリソースと第２種のリソースのうち一方の種類のリソースの負荷が高く他方の種類のリソースの負荷に余裕がある場合には、他方の種類のリソースの負荷がより高くなるような処理が実行される。
【先行技術文献】
【特許文献】
【０００５】

【特許文献１】特開２００１－４３０２６号公報
【特許文献２】特開２００８－１８６１０８号公報
【発明の概要】
【発明が解決しようとする課題】
【０００６】
第１及び第２のコントローラを有するストレージシステムでは、通常、第１のコントローラの電源と第２のコントローラの電源は異なっており、第１のコントローラに障害が生じた場合、第２のコントローラが、第１のコントローラに代わって稼働する。このため、第１のコントローラが稼働している間、第２のコントローラが有する資源が利用されないようになっており、第１のコントローラの負荷を低減することができない。
【０００７】
具体的には、例えば、第１のコントローラは、ホスト装置から受けたライト要求に付随するデータ（ライトデータ）を、ライトデータが失われないようにするために、第１のコントローラが有するキャッシュメモリ（第１のキャッシュメモリ）だけでなく、コントローラ間のパスを介して、第２のコントローラが有するキャッシュメモリ（第２のキャッシュメモリ）にも書き込むことができる。
【０００８】
しかし、それは、ライトデータのミラーリングにすぎず、第１のコントローラの負荷を低減することにはならない。
【０００９】
そこで、本発明の目的は、第１のコントローラの負荷を低減できるように第２のコントローラの資源を利用可能にすることである。
【課題を解決するための手段】
【００１０】
第１のコントローラと、前記第１のコントローラに第１パスを介して接続された第２のコントローラとを備える。第１のコントローラが、データ転送を制御する回路である第１の中継回路と、第１の中継回路に第１の第２パスを介して接続された第１のプロセッサとを有する。第２のコントローラが、データ転送を制御する回路であり、第１パスを介して第１の中継回路に接続された第２の中継回路と、第２の中継回路に第２の第２パスを介して接続された第２のプロセッサとを有する。第１のプロセッサが、第１の中継回路を介することなく第１の第３パスを介して第２の中継回路に接続されており、Ｉ／Ｏ処理において、第１の第３パスを介して前記第２の中継回路にアクセスする。第２のプロセッサが、第２の中継回路を介することなく第２の第３パスを介して第１の中継回路に接続されており、Ｉ／Ｏ処理において、第２の第３パスを介して第１の中継回路にアクセスする。
【００１１】
記憶制御装置は、物理記憶デバイスを有したストレージシステムであっても良いし、ストレージシステムに備えられる装置であっても良いし、ホスト装置とストレージシステムとの間の通信を中継する中継装置（例えばスイッチ装置）であっても良い。
【発明の効果】
【００１２】
第１のコントローラの負荷を低減できるように第２のコントローラの資源を利用することができる。
【図面の簡単な説明】
【００１３】
【図１】図１は、本発明の第１の実施形態にかかるストレージシステムを備えた計算機システムの構成の一例を示すブロック図である。
【図２】図２は、ストレージコントローラ２１のブロック図である。
【図３】図３は、ＳＰＵ２５ａの詳細ブロック図である。なお、他のＳＰＵ２５ｂも同様の構成を有する。
【図４】図４は、ＭＰ１１の詳細ブロック図である。

【図５】図５は、キャッシュメモリ２４ａ及び２５ｂの内部のデータ領域を示したブロック図である。
【図６】図６は、ストレージシステム２０がホスト装置１０からリード要求を受領した場合の処理（Ｉ／Ｏ処理のうちのリード処理）の流れを示すラダーチャートである。
【図７】図７は、ストレージコントローラ２１がホスト装置１０からライト要求を受領した場合の処理（Ｉ／Ｏ処理のうちのライト処理）の流れを示したラダーチャートである。
【図８】図８は、ＤＭＡＣ２５１の内部の構造、及び、ＤＭＡＣ２５１とパケットユニット２５２及びＤＭＡＣ制御エリア２４３１との関係を示すブロック図である。
【図９】図９は、転送パラメータ３０およびステータスメッセージ４０の詳細を示す。
【図１０】図１０は、ＤＭＡＣ選択処理のフローチャートである。
【図１１Ａ】図１１Ａは、キャッシュデータエリアが固定的に区分されている場合のリード時のキャシュ利用の一例を示す。
【図１１Ｂ】図１１Ｂは、キャッシュデータエリアが固定的に区分されている場合のライト時のキャシュ利用の一例を示す。
【図１２Ａ】図１２Ａは、本発明の第１の実施形態でのリード時のキャシュ利用の一例を示す。
【図１２Ｂ】図１２Ｂは、本発明の第１の実施形態でのライト時のキャシュ利用の一例を示す。
【図１３】図１３は、本発明の第２の実施形態にかかるストレージシステムのストレージコントローラの内部構成を示すブロック図である。
【図１４】図１４は、本発明の第３の実施形態におけるインデックスを示す。
【図１５】図１５は、本発明の第４の実施形態におけるインデックス及びパラメータキューの構成を示す。
【発明を実施するための形態】
【００１４】
以下、図面を参照して、本発明の幾つかの実施形態を説明する。
【実施例１】
【００１５】
図１は、本発明の第１の実施形態にかかるストレージシステムを備えた計算機システムの構成の一例を示すブロック図である。
【００１６】
計算機システムは、ストレージシステム２０と、ＳＡＮスイッチ３０と、複数のホスト装置１０とを有する。ストレージシステム２０が、ＳＡＮ（Ｓｔｏｒａｇｅ Ａｒｅａ Ｎｅｔｗｏｒｋ）を構成するＳＡＮスイッチ３０を介して、複数のホスト装置１０に接続される。例えば、ＳＡＮスイッチ３０は、ストレージシステム２０にホストチャネル４０を介して接続されるとともに、複数のホスト装置１０にそれぞれチャネル１１を介して接続される。
【００１７】
ストレージシステム２０は、ストレージコントローラ２１と、記憶媒体としての複数のＨＤＤ（Ｈａｒｄ Ｄｉｓｋ Ｄｒｉｖｅ）２３とを有する。ストレージコントローラ２１に、ＨＤＤチャネル２２（Ｈａｒｄ Ｄｉｓｋ Ｄｒｉｖｅ Ｃｈａｎｎｅｌ）を介して、複数のＨＤＤ２３が接続される。ＨＤＤチャネル２２は冗長構成にしてもよい。記憶媒体は、ＨＤＤ２３に代えて、フラッシュメモリやＤＲＡＭ（Ｄｙｎａｍｉｃ Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）などを用いたＳＳＤ（Ｓｏｌｉｄ Ｓｔａｔｅ Ｄｒｉｖｅ）、或いは、他の記憶媒体でもよい。
【００１８】
ストレージコントローラ２１は、ホスト装置１０からのライト要求に従い、ライト要求に付随するライト対象のデータをＨＤＤ２３に格納する。ストレージコントローラ２１は、また、ホスト装置１０からのリード要求に従い、リード対象のデータをＨＤＤ２３から読み出し、そのリード対象のデータをホスト装置１０へ送信する。

【００１９】
図２は、ストレージコントローラ２１のブロック図である。
【００２０】
本実施形態では、ストレージコントローラ２１は障害に備え冗長構成となっている。例えば、図２に示すように、ストレージコントローラ２１は、第１のコントローラ部２１ａ及び第２のコントローラ部２１ｂの二つのコントローラ部を備える。各コントローラ部２１ａ、２１ｂを点線で示す。図示の例では、第１のコントローラ部２１ａと第２のコントローラ部２１ｂは、同じ構成となっているが、異なった構成となっていてもよい。本実施形態では、後に説明する通り、稼働中において、第１のコントローラ部２１ａ及び第２のコントローラ部２１ｂのいずれも、自分のコントローラ部内の資源だけでなく相手のコントローラ部内の資源を活用することができるため、第１のコントローラ部２１ａ及び第２のコントローラ部２１ｂをともに稼働させることができる。つまり、ホスト装置が一方のコントローラのみに接続されている場合においても、接続された一方が稼動している間に他方が待機している必要は無く、両方が稼動することができる。
【００２１】
第１のコントローラ部２１ａには第１のＰＳ（Ｐｏｗｅｒ　Ｓｕｐｐｌｙ）５０ａから電力が供給され、第２のコントローラ部２１ｂには第２のＰＳ５０ｂから電力が供給される。
【００２２】
コントローラ部２１ａ（２１ｂ）は、ＳＰＵ（Ｓｔｏｒａｇｅ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）２５ａ（２５ｂ）と、ＭＰ（Ｍｉｃｒｏ　Ｐｒｏｃｅｓｓｏｒ）２６ａ（２６ｂ）と、ＨＰＣ（Ｈｏｓｔ　Ｃｈａｎｎｅｌ　Ｐｒｏｔｏｃｏｌ）２１２ａ（２１２ｂ）と、ＤＰＣ（Ｄｉｓｃ　Ｃｈａｎｎｅｌ　Ｐｒｏｔｏｃｏｌ　Ｔｉｐ）２１１ａ（２１１ｂ）と、キャッシュメモリ２４ａ（２４ｂ）と、を備える。
【００２３】
以下、第１のコントローラ部２１ａを例にとり、コントローラ部２１ａ及び２１ｂを詳細に説明する。なお、以下の説明では、適宜、第１のコントローラ部２１ａの要素の名称の前に「第１の」と付し、第２のコントローラ部２１ｂの要素の名称の前に「第２の」と付すことにする。
【００２４】
第１のＳＰＵ２５ａは、ハードウェア回路、例えばＬＳＩ(Large Scale Integration)である。第１のＳＰＵ２５ａは、第１のキャッシュメモリ２４ａに接続され、これを制御する。第１のＳＰＵ２５ａは、第１のＨＰＣ２１２ａを介してホストチャネル４０に接続される。第１のＨＰＣ２１２ａは、ホストチャネル４０の下位レベルのプロトコル処理を行って、第１のＳＰＵ２５ａに接続可能な伝送方式に変換する。第１のＳＰＵ２５ａは、また、第１のＤＰＣ２１１ａを介してＨＤＤチャネル２２に接続される。第１のＤＰＣ２１１ａは、ＨＤＤチャネル２２の下位レベルのプロトコル処理を行って、第１のＳＰＵ２５ａに接続可能な伝送方式に変換する。
【００２５】
第１のコントローラ部２１ａは第２のコントローラ部２１ｂに接続される。具体的には、
第１のＳＰＵ２５ａと第２のＳＰＵ２５ｂとが、２本のＳＰＵ間パスＣ１、Ｃ２により接続される。ＳＰＵ間パスの数は、２より多くても少なくてもよい。ＳＰＵ間パスＣ１、Ｃ２は、キャッシュメモリ２４ａ、２４ｂと後述するＤＭＡＣとの間のアクセスなどに用いられる。
【００２６】
本実施形態では、第１のＭＰ２６ａとして、ＭＰ１１及び１２があり、第２のＭＰ２６ｂとして、ＭＰ２１及び２２がある。
【００２７】
第１のＭＰ１１（１２）が、ＳＰＵ－ＭＰパスＮ１１（Ｎ１２）により第１のＳＰＵ２

５ａに接続されている。同様に、第２のＭＰ２１（２２）が、ＳＰＵ－ＭＰパスＮ２１（Ｎ２２）により第２のＳＰＵ２５ｂに接続されている。以下、ＭＰとそのＭＰを有するコントローラ部内のＳＰＵとを結ぶＳＰＵ－ＭＰパスを、「ノーマルパス」と呼ぶ。

【００２８】

本実施形態では、一方のコントローラ部内のＭＰが他方のコントローラ部内のＳＰＵにおける資源を直接利用できるようにするための一つの工夫として、コントローラ部をまたぐＳＰＵ－ＭＰパスが用意される。以下、そのＳＰＵ－ＭＰパスを「ダイレクトパス」と呼ぶ。具体的には、第１のＳＰＵ２５ａに、第２のＭＰ２１（２２）が、ダイレクトパスＤ２１（２２）により接続される。同様に、第２のＳＰＵ２５ｂに、第１のＭＰ１１（１２）が、ダイレクトパスＤ１１（１２）により接続される。

【００２９】

すなわち、本実施形態によれば、第１のＳＰＵ２５ａには、第１のＭＰ１１（１２）がノーマルパスＮ１１（Ｎ１２）で接続されるとともに、第２のＭＰ２１（２２）がダイレクトパスＤ２１（Ｄ２２）で接続されることとなる。同様に、第２のＳＰＵ２５ｂには、第２のＭＰ２１（２２）がノーマルパスＮ２１（Ｎ２２）で接続されるとともに、第１のＭＰ１１（１２）がダイレクトパスＤ１１（Ｄ１２）で接続されることとなる。

【００３０】

この構成において、第１のＭＰ１１（１２）は、Ｉ／Ｏ処理（例えば、ライト要求に従う処理、及び、リード要求に従う処理）において、第１のＳＰＵ２５ａ内の資源（例えば後述のＤＭＡＣ）又は第１のキャッシュメモリ２４ａには、ノーマルパスＮ１１（Ｎ１２）を介してアクセスし、第２のＳＰＵ２５ｂ内の資源（例えば後述のＤＭＡＣ）又は第２のキャッシュメモリ２４ｂには、ダイレクトパスＤ１１（Ｄ１２）を介してアクセスする。また、第１のＭＰ１１（１２）は、ダイレクトパスＤ１１（Ｄ１２）の障害が検出された場合、第１のＳＰＵ２５ａ及びＳＰＵ間パスＣ１又はＣ２を介して、第２のＳＰＵ２５ｂ内の資源（例えば後述のＤＭＡＣ）又は第２のキャッシュメモリ２４ｂにアクセスすることができる。

【００３１】

一方、第２のＭＰ２１（２２）は、Ｉ／Ｏ処理において、第２のＳＰＵ２５ｂ内の資源（例えば後述のＤＭＡＣ）又は第２のキャッシュメモリ２４ｂに、ノーマルパスＮ２１（Ｎ２２）を介してアクセスし、第１のＳＰＵ２５ａ内の資源（例えば後述のＤＭＡＣ）又は第１のキャッシュメモリ２４ａに、ダイレクトパスＤ２１（Ｄ２２）を介してアクセスする。また、第２のＭＰ２１（２２）は、ダイレクトパスＤ２１（Ｄ２２）の障害が検出された場合、第２のＳＰＵ２５ｂ及びＳＰＵ間パスＣ１又はＣ２を介して、第１のＳＰＵ２５ａ内の資源（例えば後述のＤＭＡＣ）又は第１のキャッシュメモリ２４ａにアクセスすることができる。

【００３２】

図３は、ＳＰＵ２５ａの詳細ブロック図である。なお、他のＳＰＵ２５ｂも同様の構成を有する。

【００３３】

ＳＰＵ２５ａは、ＤＭＡＣ（Ｄｉｒｅｃｔ　Ｍｅｍｏｒｙ　Ａｃｃｅｓｓ　Ｃｏｎｔｒｏｌｌｅｒ）２５１と、パケットユニット２５２と、バッファ２５３と、障害管理ユニット２５７と、メモリコントローラ（ＭＣ）２５８と、複数のポートと、を備える。

【００３４】

各ポートは、プロトコルチップポート（Ｐ－Ｐ）２５４、ＳＰＵ間パスポート（Ｉ－Ｐ）２５５、及び、ＭＰポート（Ｍ－Ｐ）２５６のいずれかである。Ｉ－Ｐ２５５は、ＳＰＵ間パスＣ１又はＣ２が接続されたポートである。Ｍ－Ｐ２５６は、ノーマルパス又はダイレクトパスが接続されたポートである。Ｐ－Ｐ２５４は、プロトコルチップ（ＨＰＣ又はＤＰＣ）が接続されたポートである。ＭＣ２５８は、キャッシュメモリ２４ａ、２４ｂへのアクセスの制御を行う回路である。

【００３５】

パケットユニット２５２は、ＳＰＵ２５が受信したデータについて、宛先アドレスを識別し、適切なポート若しくはコンポーネントへのパケット転送を行う。ホスト装置１０若しくはＨＤＤ２３から、ＨＰＣ２１２若しくはＤＰＣ２１１がデータを受信した場合には、受信したデータを含んだパケットは、パケットユニット２５２を介してバッファ２５３に格納される（ただしデータ以外の要求などを含んだパケットはＭＰに転送される）。また、バッファ２５３に格納されたデータがホスト装置１０若しくはＨＤＤ２３へ転送される場合には、バッファ２５３に格納されたデータを含んだパケットは、パケットユニット２５２を介してＨＰＣ２１２若しくはＤＰＣ２１１に転送され、その後、ＨＰＣ２１２若しくはＤＰＣ２１１からホスト装置１０若しくはＨＤＤ２３へ転送される。

【００３６】

本実施形態においてバッファ２５３はＳＰＵ２５内に備えているが、構成によってはキャッシュメモリ２４の一部や、後述するＭＰ１１（１２，２１，２２）内のメモリ２６３の一部を利用して配しても良い。

【００３７】

ＤＭＡＣ２５１は、転送パラメータによるＭＰ２６ａ又は２６ｂからの指示に基づいて、バッファ２５３に格納されたデータの、キャッシュメモリ２４ａ、２４ｂへの転送を制御する。ＤＭＡＣ２５１は、また、転送パラメータによるＭＰ２６ａ、２６ｂからの指示に基づいて、キャッシュメモリ２４ａ、２４ｂに格納されたデータの、バッファ２５３へ転送を制御する。転送パラメータについては後述する。

【００３８】

障害管理ユニット２５７は、ＳＰＵ２５内の各資源を有効に使用するために、ＳＰＵ２５内の各資源の障害発生を監視するユニットである。障害管理ユニット２５７は、ＳＰＵ２５内の各資源（コンポーネント）に接続されている。ＳＰＵ２５内の或る資源で障害が発生した場合には、障害管理ユニット２５７は、障害の発生を検知し、各ＭＰ２６ａ、２６ｂに、障害が生じた資源を報告する。例えば、或るＭ－Ｐ２５６に障害が発生した場合、障害が発生したＭ－Ｐ２５６に接続しているＭＰ２６ａ又は２６ｂは、ノーマルパス又はダイレクトパス経由では、ＳＰＵ２５内の資源（例えばＤＭＡＣ２５１）にアクセスすることができなくなってしまう。このため、障害管理ユニット２５７は、障害が発生したＭ－Ｐ２５６に接続されているＭＰ２６ａ又は２６ｂが行う処理を、障害が発生していないＭ－Ｐ２５６に接続されているＭＰ２６ａ又は２６ｂに引き継がせる。

【００３９】

図４は、ＭＰ１１の詳細ブロック図である。なお、他のＭＰ１２、２１及び２２もそれぞれ同様の構成を有する。

【００４０】

ＭＰ１１は、ＣＰＵ２６１と、周辺ユニット２６２と、メモリ２６３と、を備える。

【００４１】

ＣＰＵ２６１は、メモリ２６３からプログラムを読み出し、そのプログラムを実行することで、プログラムに基づく処理を行う。

【００４２】

周辺ユニット２６２は、ＣＰＵ２６１とメモリ２６３に接続されているインターフェース回路である。周辺ユニット２６２には、ＳＰＵ２５ａに接続されるノーマルパスＮ１１と、ＳＰＵ２５ｂに接続されるダイレクトパスＤ１１とが接続される。周辺ユニット２６２は、ＣＰＵ２６１とメモリ２６３との間の通信や、ＣＰＵ２６１とノーマルパスＮ１１又はダイレクトパスＤ１１を介してのＳＰＵ２５ａ又は２５ｂとの通信を制御する。なおＣＰＵ２６１の種類によっては、メモリ２６３が周辺ユニット２６２を介さずにＣＰＵ２６１に直接接続される形態や、周辺ユニット２６２がＣＰＵ２６１と一体となっている形態であっても構わない。

【００４３】

メモリ２６３は、ＣＰＵ２６１（または、ＳＰＵ２５ａ又は２５ｂ）からアクセスされる。メモリ２６３は、プログラムエリア２６３と、ワークエリア２６３２と、ＤＭＡＣエ

リア２６３３と、メールボックス２６３４とを備える。
【００４４】
プログラムエリア２６３には、ＣＰＵ２６１が実行するプログラムが格納される。ワークエリア２６３２は、ＣＰＵ２６１がプログラムを実行するために確保された作業領域である。ＤＭＡＣエリア２６３３には、ＤＭＡＣ２５１からの転送ステータスが格納される。メールボックス２６３４は、他のＭＰ１２、２１又は２２との通信のために用いられる。
【００４５】
図５は、キャッシュメモリ２４ａ及び２４ｂの内部のデータ領域を示したブロック図である。
【００４６】
キャッシュメモリ２４ａ（２４ｂ）は、制御情報エリア２４２ａ（２４２ｂ）と、転送パラメータエリア２４３ａ（２４３ｂ）と、キャッシュデータエリア２４１ａ（２４１ｂ）と、を備える。
【００４７】
制御情報エリア２４２ａ（２４２ｂ）は、制御情報を記憶する。制御情報は、例えば、下記の（１）～（５）の情報：
（１）データを格納するスロット２４１１ａ（２４１１ｂ）が指定されるディレクトリ情報；
（２）スロット２４１１ａ（２４１１ｂ）の使用状況を示す情報；
（３）使用中のスロット２４１１ａ（２４１１ｂ）について、どのスロット２４１１ａ（２４１１ｂ）にどのデータが格納されているかを示すデータ格納情報；
（４）複数のＨＤＤ２３を仮想的に１つのボリュームとして提供するＲＡＩＤ制御の設定情報；
（５）バックアップ機能、スナップショット機能及びリモートコピー機能など、機能に関する情報、
を含む。
【００４８】
転送パラメータエリア２４３ａ（２４３ｂ）には、ＭＰ２６ａ又は２６ｂがセットした転送パラメータが格納される。転送パラメータエリア２４３に格納された転送パラメータは、ＤＭＡＣ２５１にフェッチされて実行される。転送パラメータについては後述する。
【００４９】
本実施形態では、転送パラメータエリア２４３ａ（２４３ｂ）はキャッシュメモリ２４ａ（２４ｂ）に設けられている。しかし、例えば、転送パラメータエリア２４３ａ及び／又は２４３ｂは、他の記憶資源、例えば、バッファ２５３及び／又はメモリ２６３（ＭＰ２６ａ（２６ｂ）内の記憶資源）に設けられてもよい。メモリ２６３が転送パラメータエリア２４３ａ（２４３ｂ）を有する場合、例えば、ＤＭＡＣエリア２６３３に転送パラメータエリア２４３ａ（２４３ｂ）が設けられる。
【００５０】
キャッシュデータエリア２４１ａ（２４１ｂ）は、複数のスロット２４１１ａ（２４１１ｂ）を有する。スロット２４１１ａ（２４１１ｂ）には、ホスト装置１０からのライトコマンドに従うライトデータ又はリードデータが一時的に格納される。図では、ライトデータは、アルファベット「Ｗ」と、数字との組合せであらわされている。リードデータは、アルファベット「Ｒ」と、数字との組合せであらわされている。
【００５１】
キャッシュデータエリア２４１ａ（２４１ｂ）には、複数のスロットが設けられており、ライトデータの書き込みエリアとリードデータの書き込みエリアといった複数のエリアが固定的に設けられていない。このため、ライトデータ（Ｗ）及びリードデータ（Ｒ）が、キャッシュデータエリア２４１ａ（２４１ｂ）の任意のスロット２４１１に格納される。特に、ライトデータ（Ｗ）は、キャッシュデータエリア２４１ａ及び２４１ｂの両方に

格納される。つまり、データのいわゆるダブルライト、言いかえれば、データのキャッシュミラーリングが行われる。

【００５２】

図９は、転送パラメータ３０およびステータスメッセージ４０の詳細を示す。

【００５３】

まず、転送パラメータ３０について以下に説明する。

【００５４】

転送パラメータ３０は、データの転送に関するパラメータである。転送パラメータ３０は、ＩＤフィールド３０１と、オペレーションフィールド３０２と、バッファアドレスフィールド３０３と、キャッシュメモリアドレスフィールド３０４と、サイズフィールド３０５とを備える。

【００５５】

ＩＤフィールド３０１は、ＭＰ２６ａ（２６ｂ）に付与されたＩＤが設定される領域である。

【００５６】

オペレーションフィールド３０２は、オペレーションの種類を表す情報が設定される領域である。オペレーションの種類には、例えば、キャッシュメモリ２４ａ（又は２４ｂ）からデータを転送する（読み出す）「リード」、単一のキャッシュメモリ２４ａ（又は２４ｂ）にデータを転送する（書き込む）「ライト」、及び、キャッシュメモリ２４ａ及び２４ｂの両方にデータを転送する（書き込む）「ダブルライト」がある。

【００５７】

バッファアドレスフィールド３０３は、バッファ２５３のアドレスを表す値（バッファアドレス値）が設定される領域である。

【００５８】

キャッシュアドレスフィールド３０４は、キャッシュメモリ２４ａ及び／又は２４ｂのアドレスを表す値（キャッシュアドレス値）が設定される領域である。キャッシュメモリ値は、例えば、キャッシュメモリ２４ａ及び／又は２４ｂの識別情報と、スロット２４１１の位置情報とを含んでよい。オペレーションの種類が「ダブルライト」のときには、当該アドレスフィールド３０４には、キャッシュメモリ２４ａ及び２４ｂの両方のキャッシュアドレス値が設定される。

【００５９】

サイズフィールド３０５は、転送対象のデータのデータサイズ（データ長）を表す情報（データサイズ情報）が設定されるフィールドである。

【００６０】

なお、ストレージコントローラ２１は、ライトの時に、バッファ２５３内の連続するライトデータを、複数のキャッシュメモリ２４に分割して格納するスキャッタ機能や、リードの時に、複数のキャッシュメモリ２４に分割されていたリードデータを、一つのバッファ内に連続して格納するギャザー機能を備えてもよい。これらスキャッタ機能若しくはギャザー機能を用いる場合には、転送パラメータ３０に、キャッシュアドレスフィールド３０４及びサイズフィールド３０５は複数組含まれることとなる。

【００６１】

次に、ステータスメッセージ４０について以下に説明する。

【００６２】

ステータスメッセージ４０は、転送パラメータ３０に対応するメッセージであり、転送パラメータ３０の実行の結果のステータスを表す。ステータスメッセージ４０は、ＩＤフィールド４０１と、ステータスフィールド４０２とを備える。

【００６３】

ＩＤフィールド４０１は、対応する転送パラメータ３０に設定されていたＩＤと同じＩＤが設定される領域である。従って、ＩＤフィールド４０１内のＩＤを用いて、そのＩＤを有するステータスメッセージ４０に対応する転送パラメータ３０を特定することができ

る。
【００６４】
ステータスフィールド４０２は、転送パラメータ３０の実行に関するステータスを表す情報が設定される領域である。ステータスは、例えば、「正常終了」、「転送中キャッシュメモリ障害検出」、「転送中パス障害検出」及び「転送パラメータフォーマット不正」がある。
【００６５】
ステータスメッセージ４０は、ＤＭＡＣ２５１が転送パラメータ３０の実行が完了したことを検知したときにＭＰ２６ａ、２６ｂに送信するメッセージである。ステータスメッセージ４０を受信することにより、ＭＰ２６ａ、２６ｂは転送状況を知ることができる。
【００６６】
すなわち、ＭＰ２６ａ、２６ｂが、受信していないステータスメッセージ４０に対応する転送パラメータ３０をカウントすることで、未完了の転送パラメータ３０の数を知ることができる。例えば、ＤＭＡＣ２５１の負荷が高ければ、結果として転送パラメータ３０の処理速度は遅くなる。具体的には、例えば、転送パラメータ３０の実行頻度よりも転送パラメータ３０が発行頻度の方が高ければ、未完了の転送パラメータ３０の数が増えていくことになる。つまり、未完了の転送パラメータ３０の数を知ることで、各ＭＰ２６ａ、２６ｂは、他のＭＰ２６ａ、２６ｂと通信せずともＤＭＡＣ２５１の負荷状況を知ることができる。
【００６７】
図８は、ＤＭＡＣ２５１の内部の構造、及び、ＤＭＡＣ２５１とパケットユニット２５２及びＤＭＡＣ制御エリア２４３１との関係を示すブロック図である。以下の説明では、図８に示すＤＭＡＣ２５１及びパケットユニット２５２は、ＳＰＵ２５ａ内の要素であるとする。
【００６８】
まず、ＤＭＡＣ制御エリア２４３１の内部構造について説明する。
【００６９】
ＤＭＡＣ制御エリア２４３１は、キャッシュメモリ２４ａにおける転送パラメータエリア２４３ａの内部に設けられた領域である。ＤＭＡＣ制御エリア２４３１は、複数のパラメータキュー２４３１１を備える。ＤＭＡＣ制御エリア２４３１内の複数のパラメータキュー２４３１１は、ストレージコントローラ２１内のＭＰに一対一で対応づけられる。つまり、各ＤＭＡＣ２５１について、ＭＰ毎にパラメータキュー２４３１１が設けられている。各パラメータキュー２４３１１には、例えば、対応するＭＰのＩＤ（対応ＭＰ－ＩＤ）が付与される。
【００７０】
次に、ＤＭＡＣ２５１の内部構造を説明する。
【００７１】
ＤＭＡＣ２５１は、インデックス２５１１と、セレクタ２５１２と、パラメータフェッチユニット２５１３と、アドレスレジスタ２５１４と、カウントレジスタ２５１５と、転送制御ユニット２５１６とを備える。
【００７２】
インデックス２５１１は、そのＤＭＡＣ２５１についてのパラメータキュー２４３１１と同様に、ストレージコントローラ２１内のＭＰに一対一で対応づけられる。つまり、ＤＭＡＣ２５１には、ＭＰ毎にインデックス２５１１が設けられており、各インデックス２５１１は各パラメータキュー２４３１１に対応している。各インデックス２５１１には、例えば、対応するＭＰのＩＤ(対応ＭＰ－ＩＤ)が付与される。すなわち、ＭＰ２６ａ、２６ｂが個々に有するＩＤは、各パラメータキュー２４３１１に付与された対応ＭＰ－ＩＤであるとともに、各インデックス２５１１に付与された対応ＭＰ－ＩＤでもある。
【００７３】
パラメータキュー２４３１１には、そのキュー２４３１１に対応したＭＰ２６ａ又は２

６ｂから転送パラメータ３０が格納される。パラメータキュー２４３１１からは、先に格納された転送パラメータ３０から先に、パラメータフェッチユニット２５１３によってフェッチされる。つまり、パラメータキュー２４３１１内の未処理の転送パラメータ３０は、格納された順に処理される。

【００７４】

インデックス２５１１には、そのインデックス２５１１に対応したパラメータキュー２４３１１についての未完了の転送パラメータ３０の数（つまりキュー２４３１１に蓄積されている転送パラメータ３０の数）が記憶されている。例えば、ＭＰ２６ａ又は２６ｂにより、そのＭＰに対応するパラメータキュー２４３１１に転送パラメータ３０が一つ格納された場合、そのＭＰ対応するインデックス２５１１が記憶する値に１がインクリメントされる。また、パラメータフェッチユニット２５１３により、パラメータキュー２４３１１から一つの転送パラメータ３０がフェッチされた場合、そのパラメータキュー２４３１１に対応するインデックス２５１１が記憶する値から１がデクリメントされる。

【００７５】

セレクタ２５１２は、インクリメントのあったインデックス２５１１（以下、対象インデックスという）を選択し、選択した対象インデックスの対応ＩＤをパラメータフェッチユニット２５１３に送る。セレクタ２５１２による対象インデックスの選択は、例えば、パラメータフェッチユニット２５１３が、使用可能な状況（「ＲＥＡＤＹ」という状況）であった場合に行われる。対象インデックスが複数あった場合、セレクタ２５１２が所定のルール（例えばラウンドロビン）で一つの対象インデックスを選択する。

【００７６】

パラメータフェッチユニット２５１３は、セレクタ２５１２から受け取った対応ＩＤが付与されている対象インデックス２５１１から転送パラメータ３０をフェッチする。パラメータフェッチユニット２５１３は、フェッチした転送パラメータ３０内のバッファアドレス値及びキャッシュアドレス値を、アドレスレジスタ２５１４にセットし、且つ、その転送パラメータ３０内のデータサイズ値をカウントレジスタ２５１５にセットする。また、転送制御ユニット２５１６が使用可能な状況であった場合には、パラメータフェッチユニット２５１３は、転送制御ユニット２５１６に対して、例えば、転送パラメータ３０のオペレーション種類及び上記受け取った対応ＩＤを伝えるとともに、そのオペレーション種類に応じた転送開始のトリガーをかける。トリガーとしては、リードのトリガーとライトのトリガーとがある。

【００７７】

転送制御ユニット２５１６は、パラメータフェッチユニット２５１３から転送開始のトリガーを受けた場合に、キャッシュメモリとバッファとの間の転送制御を行う。

【００７８】

例えば、リードのトリガーがかけられた場合には、転送制御ユニット２５１６は、レジスタ２５１４に設定されているキャッシュメモリ値が表すスロットに格納されているデータ（リードデータ）を、そのスロットから、レジスタ２５１４に設定されているバッファアドレス値が表すバッファ領域（バッファ２５３内の領域）に転送する。また、例えば、ライトのトリガーがかけられた場合には、転送制御ユニット２５１６は、レジスタ２５１４に設定されているバッファアドレス値が表すバッファ領域に格納されているデータ（ライトデータ）を、そのバッファ領域から、レジスタ２５１４に設定されているキャッシュメモリ値が表すスロットに転送する。このデータ転送が行われる都度に、転送制御ユニット２５１６によって、レジスタ２５１５に記憶されているデータサイズ値から、転送されたデータのサイズ分の値がデクリメントされる。つまり、このデータ転送は、レジスタ２５１５に記憶されているデータサイズ値分のデータが転送されるまで繰り返される。レジスタ２５１５に記憶されているデータサイズ値分のデータの転送が完了したときに（例えばレジスタ２５１５に記憶されているデータサイズ値がゼロになったときに）、転送制御ユニット２５１６が、処理した転送パラメータ３０に対応するステータスメッセージ４０を生成し、そのメッセージ４０を、ＭＰに送信する。具体的には、例えば、転送制御ユニ

ット２５１６が、パラメータフェッチユニット２５１３から受け取った対応ＩＤを含んだステータスメッセージ４０を作成し、そのメッセージ４０を、その対応ＩＤを有するＭＰに送信する。このデータ転送では、下記の三パターン：
（Ｐ１）転送元の記憶領域が、図示のＤＭＡＣ２５１を有するコントローラ部内に存在し、転送先の記憶領域も、図示のＤＭＡＣ２５１を有するコントローラ部内に存在し、それ故、ＳＰＵ間でのデータ転送が生じないパターン；
（Ｐ２）転送元の記憶領域が、図示のＤＭＡＣ２５１を有するコントローラ部内に存在し、転送先の記憶領域が、図示のＤＭＡＣ２５１を有するコントローラ部とは別のコントローラ部内に存在するパターン；
（Ｐ３）転送元の記憶領域が、図示のＤＭＡＣ２５１を有するコントローラ部とは別のコントローラ部内に存在し、転送先の記憶領域が、図示のＤＭＡＣ２５１を有するコントローラ部内に存在するパターン、
がある。転送元と転送先の一方の記憶領域は、キャッシュデータエリア２４１ａ又は２４１ｂ内のスロットであり、他方の記憶領域は、バッファ領域（又はＭＰ内のメモリなどの他の記憶資源）である。

【００７９】
以下、本実施形態で行われる処理の流れを説明する。

【００８０】
図６は、ストレージシステム２０がホスト装置１０からリード要求を受領した場合の処理（Ｉ／Ｏ処理のうちのリード処理）の流れを示すラダーチャートである。以下の説明では、ＨＰＣ２１２ａがリード要求を受領するとする。また、図６及び図７の説明では、どの要素がどちらのコントローラ部に存在するのかを区別し易くするため、第１のコントローラ部２１ａに存在する要素の参照番号の末尾に「ａ」を付し、第２のコントローラ部２１ｂに存在する要素の参照番号の末尾に「ｂ」を付すことにする。

【００８１】
ＨＰＣ２１２ａは、リード要求をホスト装置１０から受領する（ｓ１０００）。以下、リード要求に従うデータを「該当リードデータ」と言う。

【００８２】
ＨＰＣ２１２ａは、ホスト装置１０から受領したリード要求を、ＭＰ２６ａのメールボックス２６３４ａへ送付する（ｓ１００１）。

【００８３】
メールボックス２６３４ａ内のリード要求を確認したＣＰＵ２６１ａは、制御情報エリア２４２ａ及び２４２ｂ内の制御情報を参照し、キャシュヒットしたか否かを判定する。キャッシュヒットしたか否かとは、ホスト装置からのＩ／Ｏ要求で指定されている論理アドレス（例えば、ＬＵＮ（Logical Unit Number）及びＬＢＡ（Logical Block Address））に従う場所に格納されるデータがあるか否かである。キャッシュヒットした場合（ここでは、リード要求が指定する論理アドレスとスロット２４１１ａ又は２４１１ｂとの対応が管理されている場合）、Ｓ１００９に処理が移る。

【００８４】
一方、キャッシュヒットしなかった場合、ＣＰＵ２６１ａは、該当リードデータをバッファ２５３ａ（又はメモリ２６３ａ等の他の記憶資源）に転送するようＤＰＣ２１１ａに指示する（ｓ１００２）。その際、ＣＰＵ２６１ａは、例えば、リード要求が指定する論理アドレスを基に特定した物理アドレスを、ＤＰＣ２１１ａに通知する。

【００８５】
ＤＰＣ２１１ａは、ＨＤＤ２３（例えば通知された物理アドレス）にアクセスして該当リードデータを受け取る（ｓ１００３）。

【００８６】
ｓ１００４で、ＤＰＣ２１１ａは、受け取った該当リードデータをバッファ２５３ａ（又はメモリ２６３ａ等の他の記憶資源）に格納する。そして、ＣＰＵ２６１ａは、キャッシュメモリ２４ａ及び２４ｂにおける複数の確保可能スロットから任意のスロット２４１

１ａ又は２４１１ｂを確保する。確保可能スロットとは、例えば、フリー又はクリーンのスロットとして管理されているスロットである。フリーのスロットとは、空きのスロットである。クリーンのスロットとは、ＨＤＤ２３に格納済みのデータを記憶しているスロットである。図６の説明では、キャッシュメモリ２４ｂからスロット２４１１ｂが確保されたとする。

【００８７】

ＤＰＣ２１１ａは、ＣＰＵ２６１ａに対して、該当リードデータのバッファ２５３ａへの格納が終了したことを知らせる終了通知を送信する（ｓ１００５）。

【００８８】

ｓ１００６で、以下の処理（ｓ１００６－１）～（ｓ１００６－４）が行われる。

【００８９】

（ｓ１００６－１）終了通知を受けたＣＰＵ２６１ａが、ＤＭＡＣ選択処理を行う。すなわち、ＣＰＵ２６１ａは、ＳＰＵ２５ａ及び２５ｂに存在する複数のＤＭＡＣ２５１ａ及び２５１ｂから一つのＤＭＡＣ２５１ａ又は２５１ｂを選択する。ＤＭＡＣ選択処理では、後述するように、データの転送先の記憶領域を有するコントローラ部内のＤＭＡＣが、データの転送先の記憶領域を有しないコントローラ部内のＤＭＡＣよりも優先的に選択される。そのため、ここでは、ＤＭＡＣ２５１ｂが選択されたとする。

【００９０】

（ｓ１００６－２）ＣＰＵ２６１ａは、下記（Ｒ０１）～（Ｒ０５）を含んだ転送パラメータ３０：

（Ｒ０１）ＩＤ；

（Ｒ０２）オペレーション種類「ライト」（キャッシュメモリにデータを書くため）、

（Ｒ０３）確保したスロットのアドレス値、

（Ｒ０４）該当リードデータを格納しているバッファ領域のアドレス値；

（Ｒ０５）該当リードデータのデータサイズ値、

を作成する。

【００９１】

（ｓ１００６－３）ＣＰＵ２６１ａは、その転送パラメータ３０を、ｓ６０００で選択したＤＭＡＣ２５１ｂについての複数のパラメータキュー２４３１１ｂのうちの、ＣＰＵ２６１ａを有するＭＰ２６ａに対応したパラメータキュー２４３１１ｂにセットする。

【００９２】

（ｓ１００６－４）ＣＰＵ２６１ａは、上記選択したＤＭＡＣ２５１ｂを起動する。

【００９３】

起動したＤＭＡＣ２５１ｂは、ＤＭＡＣ制御エリア２４３１ｂ内のパラメータキュー２４３１１ｂから転送パラメータ３０をフェッチし、その転送パラメータ３０に基づき、転送元のバッファ２５３ａから転送先のスロット２４１１ｂに該当リードデータを転送する（ｓ１００７）。

【００９４】

該当リードデータがスロット２４１１ｂに格納された後、ＤＭＡＣ２５１ｂは、転送が終了した旨のステータスメッセージ４０を、ＭＰ２６ａ内のＤＭＡＣエリア２６３３に、例えばダイレクトパスＤ１１又はＤ１２を介して送信する（ｓ１００８）。

【００９５】

ステップＳ１００９は、ステップＳ１００８でＤＭＡＣエリア２６３３ａに格納されたステータスメッセージ４０をＣＰＵ２６１ａが確認したとき、若しくは、ステップＳ１００２においてキャッシュヒットしたときに、実行される。ｓ１００９で、以下の処理（ｓ１００９－１）～（ｓ１００９－４）が行われる。

【００９６】

（ｓ１００９－１）ＣＰＵ２６１ａが、ＤＭＡＣ選択処理を行う。ここでは、例えば、ＤＭＡＣ２５１ａが選択されたとする。

【００９７】

（ｓ１００９－２）ＣＰＵ２６１ａは、下記（Ｒ１１）～（Ｒ１５）を含んだ転送パラメータ３０：
（Ｒ１１）ＩＤ；
（Ｒ１２）オペレーション種類「リード」（キャッシュメモリからデータを読むため）、
（Ｒ１３）該当リードデータを格納しているスロット２４１１ｂのアドレス値、
（Ｒ１４）バッファ領域のアドレス値；
（Ｒ１５）該当リードデータのデータサイズ値、
を作成する。

【００９８】

（ｓ１００９－３）ＣＰＵ２６１ａは、その転送パラメータ３０を、ＤＭＡＣ２５１ａについての複数のパラメータキュー２４３１１ａのうちの、ＣＰＵ２６１ａを有するＭＰ２６ａに対応したパラメータキュー２４３１１ａにセットする。

【００９９】

（ｓ１００９－４）ＣＰＵ２６１ａは、上記選択されたＤＭＡＣ２５１ａを起動する。

【０１００】

起動したＤＭＡＣ２５１ａは、ＤＭＡＣ制御エリア２４３１ａ内のパラメータキュー２４３１１ａから転送パラメータ３０をフェッチし、その転送パラメータ３０に基づき、スロット２４１１ｂからバッファ領域へ該当リードデータを転送する（ｓ１０１０）。

【０１０１】

転送終了後、ＤＭＡＣ２５１ａは、転送が終了した旨のステータスメッセージ４０をＭＰ２６ａのＤＭＡＣエリア２６３３ａに送信する（ｓ１０１１）。

【０１０２】

ステップｓ１０１１でＤＭＡＣエリア２６３３ａに格納されたステータスメッセージ４０を確認したＣＰＵ２６１ａは、バッファ２５３に格納されている該当リードデータをホスト装置１０に転送するようＨＰＣ２１２に指示する（ｓ１０１２）。

【０１０３】

指示を受けたＨＰＣ２１２は、バッファ２５３より該当リードデータを読み出す（ｓ１０１３）。

【０１０４】

ＨＰＣ２１２は、バッファ２５３から読みだした該当リードデータをホスト装置１０へ転送する（ｓ１０１４）。

【０１０５】

全ての該当リードデータの転送が正常に終了した場合は、ＨＰＣ２１２は、その旨（正常終了）をホスト装置１０に応答する（ｓ１０１５）。

【０１０６】

さらに、ＨＰＣ２１２は、ホスト装置１０に応答したことをＭＰ２６ａに通知する（ｓ１０１６）。これにより、ＭＰ２６ａは、リード要求の処理が終了したことを認識する。

【０１０７】

図７は、ストレージコントローラ２１がホスト装置１０からライト要求を受領した場合の処理（Ｉ／Ｏ処理のうちのライト処理）の流れを示したラダーチャートである。以下の説明では、ＨＰＣ２１２ａがライト要求を受領するとする。

【０１０８】

ＨＰＣ２１２ａは、ライト要求をホスト装置１０から受領する（ｓ２０００）。以下、ライト要求に従うデータを「該当ライトデータ」と言う。

【０１０９】

ＨＰＣ２１２ａは、ホスト装置１０から受領したライト要求を、ＭＰ２６ａのメールボックス２６３４ａへ送付する（ｓ２００１）。

【０１１０】

メールボックス２６３４ａ内のライト要求を確認したＣＰＵ２６１ａは、制御情報エリア２４２ａ及び２４２ｂの制御情報を参照し、キャッシュヒットしたか否かを判定する。

【０１１１】
ここで、該当ライトデータは、キャッシュメモリ２４ａ及び２４ｂの両方に格納されるべきデータ、つまり、ダブルライトされるべきデータである。
【０１１２】
そのため、キャッシュヒット判定では、ライト要求が指定する論理アドレスと対応するスロット２４１１ａ及び２４１１ｂがあるか否かが判定される。キャッシュヒットした場合、それら対応するスロット２４１１ａ及び２４１１ｂが確保される。キャッシュヒットしなかった場合、ＣＰＵ２６１ａは、キャッシュデータエリア２４１ａ及び２４１ｂからスロット２４１１ａ及び２４１１ｂを確保する。
【０１１３】
スロット２４１１ａ及び２４１１ｂを確保できた場合は、ＣＰＵ２６１ａは、該当ライトデータの受け入れ準備ができた旨をホスト装置１０に伝えるようＨＰＣ２１２ａに指示する（ｓ２００２）。
【０１１４】
ＣＰＵ２６１ａからの指示を受けたＨＰＣ２１２ａは、該当ライトデータの受け入れ準備完了した旨のメッセージをホスト装置１０に送信する（ｓ２００３）。
【０１１５】
ホスト装置１０は、該当ライトデータをＨＰＣ２１２ａへ送信する（ｓ２００４）。
【０１１６】
該当ライトデータを受領したＨＰＣ２１２ａは、その該当ライトデータをバッファ２５３ａに格納する（ｓ２００５）。
【０１１７】
該当ライトデータをバッファ２５３ａに格納した後、ＨＰＣ２１２ａは、ＣＰＵ２６１ａに対して、該当ライトデータのバッファ２５３ａへの格納が終了したことを知らせる終了通知を送信する（ｓ２００６）。
【０１１８】
ｓ２００７で、以下の処理（ｓ２００７－１）～（ｓ２００７－４）が行われる。
【０１１９】
（ｓ２００７－１）終了通知を受けたＣＰＵ２６１ａが、ＤＭＡＣ選択処理を行う。ここでは、ＤＭＡＣ２５１ａが選択されたとする。
【０１２０】
（ｓ２００７－２）ＣＰＵ２６１ａは、下記（Ｗ０１）～（Ｗ０５）を含んだ転送パラメータ３０：
（Ｗ０１）ＩＤ；
（Ｗ０２）オペレーション種類「ダブルライト」（ライト要求に従うデータのキャッシュメモリへのデータ転送であるため）、
（Ｗ０３）確保したスロット２４１１ａ及び２４１１ｂのアドレス値、
（Ｗ０４）該当ライトデータを格納しているバッファ領域のアドレス値；
（Ｗ０５）該当ライトデータのデータサイズ値、
を作成する。
【０１２１】
（ｓ２００７－３）ＣＰＵ２６１ａは、その転送パラメータ３０を、ｓ６０００で選択したＤＭＡＣ２５１ａについての複数のパラメータキュー２４３１１ａのうちの、ＣＰＵ２６１ａを有するＭＰ２６ａに対応したパラメータキュー２４３１１ａにセットする。
【０１２２】
（ｓ２００７－４）ＣＰＵ２６１ａは、上記選択したＤＭＡＣ２５１ａを起動する。
【０１２３】
起動したＤＭＡＣ２５１ａは、ＤＭＡＣ制御エリア２４３１ａ内のパラメータキュー２４３１１ａから転送パラメータ３０をフェッチし、その転送パラメータ３０に基づき、バッファ２５３からスロット２４１１ａ及び２４１１ｂに該当ライトデータを転送する（ｓ

２００８）。
【０１２４】
両データ転送が終了した後、ＤＭＡＣ２５１ａは、転送が終了した旨のステータスメッセージ４０をＭＰ２６ａのＤＭＡＣエリア２６３３ａに送信する（ｓ２００９）。
【０１２５】
ステップＳ２００９でＤＭＡＣエリア２６３３ａに格納されたステータスメッセージ４０を確認したＣＰＵ２６１ａが、ＨＰＣ２１２ａにライト終了を通知する（ｓ２０１０）。
【０１２６】
ライト終了の通知を受けたＨＰＣ２１２ａはが、ライト終了をホスト装置１０に応答する（ｓ２０１１）。
【０１２７】
その後、ｓ２０１２で、以下の処理（ｓ２０１２－１）～（ｓ２０１２－４）が行われる。
【０１２８】
（ｓ２０１２－１）ＣＰＵ２６１ａは、該当ライトデータが格納されているスロット２４１１ａ及び２４１１ｂのうちのいずれかを選択する。ここではスロット２４１１ａが選択されたとする。
【０１２９】
（ｓ２０１２－２）ＣＰＵ２６１ａは、ＤＭＡＣ選択処理を行う。ここでは、例えば、ＤＭＡＣ２５１ａが選択されたとする。
【０１３０】
（ｓ２０１２－３）ＣＰＵ２６１ａは、下記（Ｗ１１）～（Ｗ１５）を含んだ転送パラメータ３０：
（Ｗ１１）ＩＤ；
（Ｗ１２）オペレーション種類「リード」（キャッシュメモリからデータを読むため）、
（Ｗ１３）該当ライトデータを格納しているスロット２４１１ａのアドレス値、
（Ｗ１４）バッファ領域のアドレス値；
（Ｗ１５）該当ライトデータのデータサイズ値、
を作成する。
【０１３１】
（ｓ２０１２－４）ＣＰＵ２６１ａは、その転送パラメータ３０を、ＤＭＡＣ２５１ａについての複数のパラメータキュー２４３１１ａのうちの、ＣＰＵ２６１ａを有するＭＰ２６ａに対応したパラメータキュー２４３１１ａにセットする。
【０１３２】
（ｓ２０１２－５）ＣＰＵ２６１ａは、上記選択されたＤＭＡＣ２５１ａを起動する。
【０１３３】
起動したＤＭＡＣ２５１ａは、ＤＭＡＣ制御エリア２４３１ａ内のパラメータキュー２４３１１ａから転送パラメータ３０をフェッチし、その転送パラメータ３０に基づき、スロット２４１１ａからバッファ領域へ該当ライトデータを転送する（ｓ２０１３）。
【０１３４】
転送終了後、ＤＭＡＣ２５１ａは、転送が終了した旨のステータスメッセージ４０を、ＭＰ２６ａのＤＭＡＣエリア２６３３ａへ送信する（ｓ２０１４）。
【０１３５】
ｓ２０１４でＤＭＡＣエリア２６３３ａに格納されたステータスメッセージ４０を確認したＣＰＵ２６１ａは、バッファ領域に格納された該当ライトデータをＨＤＤ２３へ転送するようＤＰＣ２１１ａに指示する（ｓ２０１５）。
【０１３６】
ＣＰＵ２６１ａからの指示を受けたＤＰＣ２１１ａは、バッファ領域より該当ライトデータを読み出す（ｓ２０１６）。

【０１３７】
ＤＰＣ２１１ａは、該当ライトデータをＨＤＤ２３に格納する（ｓ２０１７）。
【０１３８】
全てのデータ転送が正常に終了した場合は、ＤＰＣ２１１ａはその旨をＣＰＵ２６１ａに終了通知を送付する（ｓ２０１８）。
【０１３９】
終了通知を受けたＣＰＵ２６１ａは、該当ライトデータが格納されていたスロット２４１１ａを開放する。解放されたスロット２４１１ａは、クリーンのスロットとして管理される。
【０１４０】
図１０は、ＤＭＡＣ選択処理のフローチャートである。なお、以下の説明では、ＭＰ２６ａ及び２６ｂを「ＭＰ２６」と総称する。
【０１４１】
ｓ３００１で、ＭＰ２６が、ＤＭＡＣ２５１毎のＤＭＡＣスコアをすべてリセットする。
【０１４２】
ｓ３００２で、ＭＰ２６は、転送先（又は転送元）のバッファ２５３の位置に応じてＤＭＡＣスコアの変更を行う。具体的には、例えば、ＭＰ２６ａは、転送先（又は転送元）のバッファ２５３を含んだＳＰＵ２５内のＤＭＡＣ２５１のＤＭＡＣスコアを変更せず、転送先のバッファ２５３を含んでいないＳＰＵ２５内のＤＭＡＣ２５１のＤＭＡＣスコアに第１の所定値を加算する。
【０１４３】
ｓ３００３で、ＭＰ２６は、転送先（又は転送元）のキャッシュメモリの位置に応じてＤＭＡＣスコアの変更を行う。具体的には、例えば、ＭＰ２６は、転送先（又は転送元）のキャッシュメモリが接続されているＳＰＵ２５内のＤＭＡＣ２５１のＤＭＡＣスコアを変更せず、転送先のキャッシュメモリが接続されていないＳＰＵ２５内のＤＭＡＣ２５１のＤＭＡＣスコアに第２の所定値を加算する。第２の所定値は第１の所定値と同じでも異なっていても良い。
【０１４４】
次に、ＭＰ２６は、各パラメータキュー２４３１１における未完了の転送パラメータ３０の数をチェックする（ｓ３００４）。ここでは、ＭＰ２６は、各ＤＭＡＣについて、そのＤＭＡＣについての未完了の転送パラメータの数と第３の所定値（０より大きい値）との乗数を、そのＤＭＡＣのＤＭＡＣスコアに加算する。
【０１４５】
最後に、ＭＰ２６は、すべてのＤＭＡＣスコアを比較し、ＤＭＡＣ２５１を選択する（ｓ３００５）。具体的には、例えば、ＤＭＡＣスコアが最少のＤＭＡＣ２５１が選択される。ＤＭＡＣスコアが最少のＤＭＡＣ２５１は、他のＤＭＡＣ２５１よりも転送先記憶領域に近く、及び／又は、他のＤＭＡＣ２５１よりも負荷が低い。そのため、ＤＭＡＣスコアが最少のＤＭＡＣ２５１が、ＤＭＡＣ選択処理を行っている時点において、データ転送を行うのに最適なＤＭＡＣ２５１であると考えられる。
【０１４６】
さらに、ＤＭＡＣ２５１に障害が発生した場合に、障害が発生していない資源に処理を引き継がせることなどにより、上記第１～第３の所定値の少なくとも一つがいずれかのＭＰ２６によってより小さい値にされても良い。すなわち、ＳＰＵ２５内の正常なＤＭＡＣ２５１の数に不均衡が発生しても、スコアの重み付けを変えることで、引き続きＳＰＵ２５内の資源を有効に使うことが可能となる。
【０１４７】
本実施形態では、コントローラ部２１ａ及び２１ｂの一方が他方のコントローラ部内のハードウェア資源を利用している。これにより、ストレージコントローラ２１の負荷を低減することができる。

【０１４８】
また、本実施形態では、第１のＭＰ１１（１２）が、ダイレクトパスＤ１１（１２）を介して第２のＳＰＵ２５ｂに接続され、第２のＭＰ２１（２２）が、ダイレクトパスＤ２１（２２）を介して接続される。このため、コントローラ部２１ａ及び２１ｂの一方が他方のコントローラ部内のハードウェア資源を利用する際に、ＳＰＵ間パスＣ１及びＣ２の経由を不要にすることができる。
【０１４９】
また、本実施形態では、一つのＤＭＡＣが複数のＭＰから使用されることにより生じ得る下記課題を、ソフトウェア制御ではなくハードウェア制御によって解決される。
【０１５０】
例えば、或るＭＰ（ｘ）が転送パラメータをパラメータキューにセットした後、インデックスが更新される前に、他のＭＰ（ｙ）が転送パラメータを上書きしてしまうことがあり得る。この場合は、ＭＰｘのセットした転送パラメータは実行されない。その上、転送パラメータが実行されなかったことをＭＰ（ｘ）は検知できない。このため、バッファやキャッシュメモリに新たなデータが反映されないまま、ホスト装置対して処理の完了を通知してしまうことになる。つまり、データ消失や誤ったデータの返送といった致命的な不具合がおこってしまう可能性がある。
【０１５１】
この課題を、ソフトウェアでの制御によって解決することも考え得る。ここで、ソフトウェアの制御によって解決するとは、例えば、ＭＰ間であらかじめ調停を行って権利を取得したＭＰが転送パラメータをセットするなどである。しかしながら、ソフトウェアの制御によって上記不具合を解決した場合には、頻繁なアクセスが必要なＤＭＡＣなどの資源に関しては、相対的にオーバーヘッドが大きく、資源の有効活用ができなくなってしまうという更なる課題が生じるおそれがある。
【０１５２】
本実施形態では、パラメータキュー２４３１１及びインデックス２５１１がＭＰ２６毎に設けられていることで、ハードウエアの制御によって、上記のようなデータ消失や誤ったデータの返送といった不具合を防止することができる。
【０１５３】
本実施形態にあっては、各ＳＰＵ２５ａ、２５ｂ内に障害管理ユニット２５７がある。このため、障害管理ユニット２５７が、ＳＰＵ２５ａ、２５ｂ内で障害が発生した資源に対して、対策を施すことが可能となる。例えば、障害管理ユニット２５７が、障害が発生したＭＰポート２５６に接続されたＭＰ２６ａ、２６ｂが行う処理を、障害が発生していないＭＰポート２５６に接続されたＭＰ２６ａ、２６ｂに引き継がせることが可能となる。これにより、ＳＰＵ２５中の資源を有効に使うことができる。
【０１５４】
本実施形態にあっては、キャッシュメモリ２４ａ、２４ｂのキャッシュデータエリア２４１は、下記に示すように、ライトデータの書き込みエリアとリードデータの書き込みエリアを予め固定的に用意しておく必要はない。
【０１５５】
図１１Ａ及び図１１Ｂによれば、キャッシュデータエリア（Ａ）に、リードキャッシュエリア（Ａ）と、ライトキャッシュエリア（Ａ）と、ミラーエリア（Ａ）とが予め用意されている。同様に、キャッシュデータエリア（Ｂ）に、リードキャッシュエリア（Ｂ）と、ライトキャッシュエリア（Ｂ）と、ミラーエリア（Ｂ）とが予め用意されている。ライトキャッシュエリア（Ａ）とミラーエリア（Ｂ）はペアになっており、同様に、ライトキャッシュエリア（Ｂ）とミラーエリア（Ａ）はペアになっている。
【０１５６】
リードキャッシュエリアは、リードデータが格納される専用エリアである。図１１Ａに示すように、リードキャッシュエリア（Ａ）には、リードキャッシュエリア（Ａ）と同じコントローラ部内にあるＭＰ（Ａ）によって取得されたリードデータ（Ｒ１）～（Ｒ１０

）のみが格納され、ライトデータや、リードキャッシュエリア（Ａ）と別のコントローラ部内にあるＭＰ（Ｂ）によって取得されたリードデータは格納されない。

【０１５７】

ライトキャッシュエリアは、ライトデータが格納される専用エリアであり、それとペアになったミラーエリアは、ライトデータのミラーが格納される専用エリアである。図１１Ｂに示すように、ライトキャッシュエリア（Ａ）には、ライトキャッシュエリア（Ａ）と同じコントローラ部内にあるＭＰ（Ａ）によって取得されたライトデータ（Ｗ１）～（Ｗ５）のみが格納され、リードデータや、他のＭＰ（Ｂ）が受けたライトデータや、そのライトデータのミラーデータは格納されない。また、図１１Ｂに示すように、ミラーエリア（Ｂ）には、ＭＰ（Ａ）によって取得されたライトデータのミラー（Ｗ１）～（Ｗ５）のみが格納され、リードデータや、ＭＰ（Ａ）及び（Ｂ）が受けたライトデータ（オリジナルデータ）は格納されない。

【０１５８】

しかし、本実施形態では、図１１Ａ及び図１１Ｂを参照して説明したリードキャッシュエリア、ライトキャッシュエリア及びミラーエリアのいずれも、キャッシュデータエリアに用意されていない。本実施形態では、データの種類に限らず、キャッシュデータエリアの任意のスロットに格納することができる。そのため、図１２Ａに示すように、例えば、第１のＭＰ２４ａが取得したリードデータ（Ｒ１）～（Ｒ４０）の一部を第１のキャッシュメモリ２４ａに格納し残りを第２のキャッシュメモリ２４ｂに格納することができる。また、図１２Ｂに示すように、キャッシュメモリ２４ａ及び２４ｂにおけるそれぞれのキャッシュデータエリアを、第１のＭＰ２４ａが取得したライトデータ（Ｗ１）～（Ｗ２０）で満杯にすることもできる。

【実施例２】

【０１５９】

図１３は、本発明の第２の実施形態にかかるストレージシステムのストレージコントローラの内部構成を示すブロック図である。

【０１６０】

本実施形態では、ノーマルパス及びダイレクトパスがＭＰスイッチ（ＭＰ　Ｓｗｉｔｃｈ）２７０で実現されている。ＭＰスイッチ２７０は、複数のＭＰ２６と複数のＳＰＵ２５が接続されたスイッチ装置である。各ＭＰ２６は、ＭＰスイッチ２７０を介して所望のＳＰＵ２５にアクセスすることができる。

【０１６１】

また、本実施形態では、ＳＰＵ間パスがＳＰＵスイッチ（ＳＰＵ　Ｓｗｉｔｃｈ）２８０で実現されている。複数のＳＰＵ２５のうちの一つと複数のＳＰＵ２５のうちの別の一つとの間の通信は、ＳＰＵスイッチ２８０を介して行われる。

【０１６２】

ＳＰＵ２５の数、ＭＰ２６の数、キャッシュメモリ２４の数を、スケーラブルにする場合には、本実施形態のような構成をとることができる。この場合、ＭＰスイッチ２７０とＳＰＵスイッチ２８０は論理的な構成としてもよく、物理的には同一のスイッチで構成しても良い。例えばPCI-Expressのvirtual channel技術が用いられてもよい。またＭＰスイッチ２７０およびＳＰＵスイッチ２８０は、物理的には複数のスイッチをカスケード接続した形態でも良い。

【０１６３】

また、さらにキャッシュメモリのスケーラビリティを向上させるため、キャッシュメモリを独立させてもよい。この場合、例えば、キャッシュメモリ機能のみを持ったＳＰＵがＳＰＵスイッチに接続されてもよい。

【実施例３】

【０１６４】

図１４は、本発明の第３の実施形態におけるインデックスを示す。

【０１６５】

インデックス２５１１が、コンシューマインデックス２５１１１と、プロデューサインデックス２５１１２とで構成されている。

【０１６６】

コンシューマインデックス２５１１１は、パラメータキューに蓄積されている一以上の転送パラメータのうちのどの転送パラメータまでフェッチしたかを表す値を記憶する。

【０１６７】

プロデューサインデックス２５１１２は、パラメータキューに蓄積されている最後尾の転送パラメータを表す値である。

【０１６８】

このようなインデックス２５１１を用いて、パラメータキューのどこから転送パラメータを読み出すべきかを特定することができる。

【実施例４】

【０１６９】

図１５は、本発明の第４の実施形態におけるインデックス及びパラメータキューの構成を示す。

【０１７０】

ＭＰ毎に、リード用パラメータキュー２６３３１Ｒとライト用パラメータキュー２６３３１Ｗとがある。また、ＭＰ毎に、リード用パラメータキュー２６３３１Ｒに対応したリード用インデックス２５１１Ｒと、ライト用パラメータキュー２６３３１Ｗに対応したライト用インデックス２５１１Ｗとがある。

【０１７１】

ＭＰは、リード要求での処理において生成した転送パラメータを、リード用パラメータキュー２６３３１Ｒに格納する。リード用インデックス２５１１Ｒは、リード用パラメータキュー２６３３１Ｒでの未処理の転送パラメータの数を表す値を記憶する。

【０１７２】

ＭＰは、ライト要求での処理において生成した転送パラメータを、ライト用パラメータキュー２６３３１Ｗに格納する。ライト用インデックス２５１１Ｗは、ライト用パラメータキュー２６３３１Ｗでの未処理の転送パラメータの数を表す値を記憶する。

【０１７３】

例えば、ライト要求の処理では、該当ライトデータがキャッシュメモリに格納されれば、ホスト装置に対してライトの完了とすることができるが、リード要求の処理では、該当リードデータがキャッシュメモリに格納されるだけでなくホスト装置に転送されてリードの完了となる。

【０１７４】

そこで、本実施形態では、ＤＭＡＣは、ライト用パラメータキュー２６３３１Ｗに格納されている転送パラメータよりも、リード用パラメータキュー２６３３１Ｒに格納されている転送パラメータを優先的にフェッチする。これにより、リード性能を高めることができる。

【０１７５】

以上、本発明の幾つかの実施形態を説明したが、本発明は、これらの実施形態に限定されるものでなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【０１７６】

上述した実施形態に従う記憶制御装置を別の観点から抽象的に表現すれば、例えば、下記の通りになる。

＜表現１＞

ホスト装置から発行されたＩ／Ｏコマンドに従うデータの転送を制御する回路である中継回路と、

前記中継回路に接続された複数のプロセッサと

プロセッサ毎に設けられておりデータ転送のパラメータである転送パラメータが蓄積さ

れるパラメータ領域と
を有し、
前記中継回路が、データの転送を行う回路であるデータ転送回路を有し、
前記パラメータ領域は、前記データ転送回路についての記憶領域であり、
各プロセッサが、そのプロセッサに対応したパラメータ領域に転送パラメータを格納し、
前記データ転送回路が、いずれかのパラメータ領域から転送パラメータを取得し、その転送パラメータに従って、データ転送を実行する、
記憶制御装置。
＜表現２＞
表現１記載の記憶制御装置であって、
前記データ転送回路が複数あり、
各データ転送回路について、プロセッサ毎のパラメータ領域があり、
前記複数のプロセッサのうちのＩ／Ｏコマンドを受領したプロセッサである対象プロセッサが、前記対象プロセッサに対応した複数のパラメータ領域のうち未完了の転送パラメータの数が最も少ないパラメータ領域に対応したデータ転送回路を選択し、選択したデータ転送回路についての、前記対象プロセッサに対応したパラメータ領域に、転送パラメータを格納する、
記憶制御装置。
＜表現３＞
表現１又は２記載の記憶制御装置であって、
前記複数のプロセッサにそれぞれ対応した複数のキャッシュメモリ領域を有し、
前記複数のプロセッサのうちの或るプロセッサが、前記複数のプロセッサのうちの別のプロセッサに対応したキャッシュメモリ領域内の任意のアドレスを転送先アドレスとした転送パラメータを、前記或るプロセッサに対応したパラメータ領域に格納する、
記憶制御装置。
＜表現４＞
表現１乃至３のうちのいずれかに記載の記憶制御装置であって、
各パラメータ領域は、Ｉ／Ｏコマンドがライトコマンドである場合の転送パラメータの格納先とされるパラメータライト領域と、Ｉ／Ｏコマンドがリードコマンドである場合の転送パラメータの格納先とされるパラメータリード領域とを有し、
前記データ転送回路は、前記パラメータライト領域よりも前記パラメータリード領域内の未完了の転送パラメータを優先的に取得する、
記憶制御装置。
＜表現５＞
表現１乃至４のうちのいずれかに記載の記憶制御装置であって、
前記データ転送回路が、複数のパラメータ領域に蓄積されている未完了の転送パラメータの数をそれぞれ記憶する記憶領域である複数のインデックスと、前記複数のインデックスから一つのインデックスを選択するセレクタと、前記セレクタによって選択されたインデックスに対応したパラメータ領域から転送パラメータを取得しその転送パラメータが有する転送元アドレス及び転送先アドレスを設定するパラメータ取得回路と、前記設定された転送元アドレスが表す記憶領域内のデータを前記設定された転送先アドレスが表す記憶領域に転送する転送制御回路とを有する、
記憶制御装置。
【符号の説明】
【０１７７】
２０：ストレージシステム

【図１】

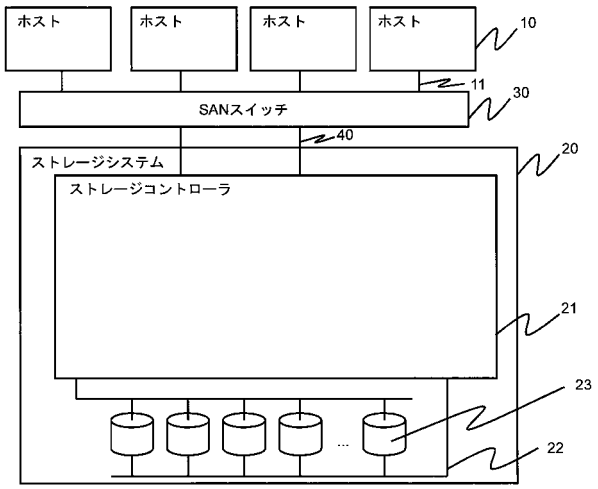

Fig.1

【図２】

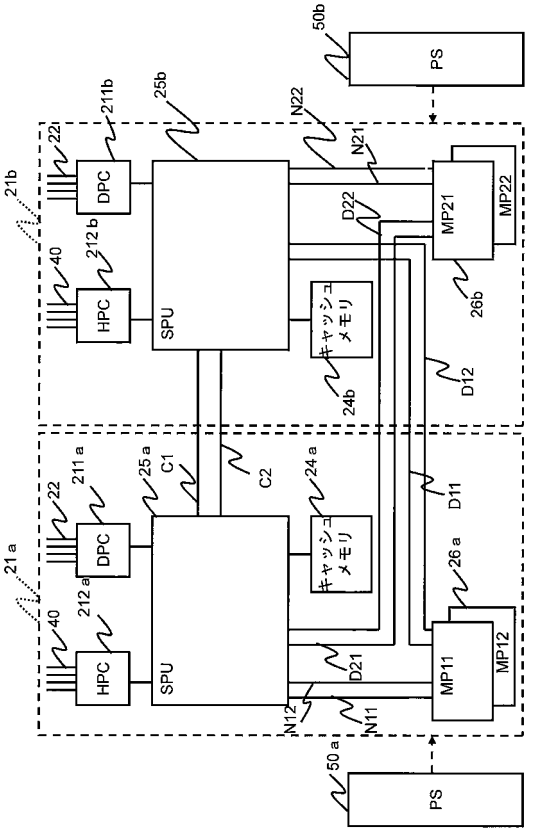

Fig.2

【図３】

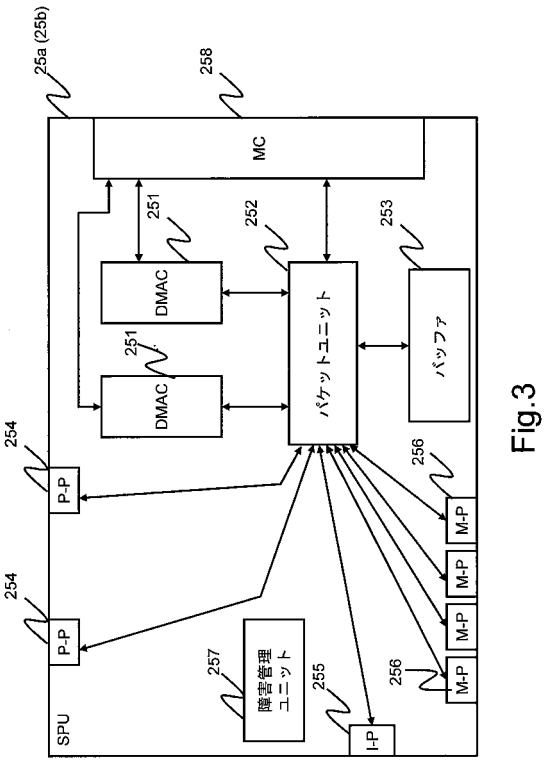

Fig.3

【図４】

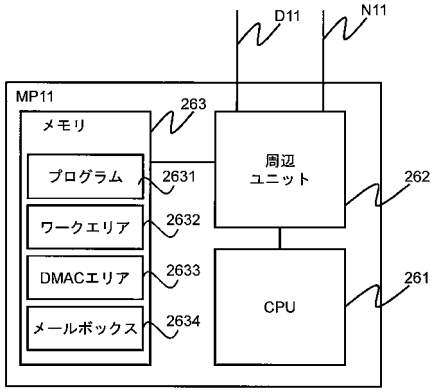

Fig.4

【図５】

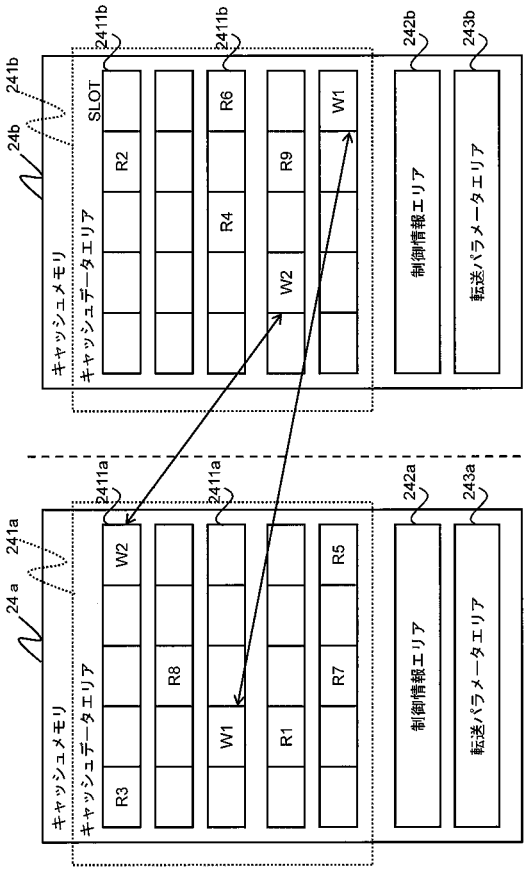

Fig.5

【図６】

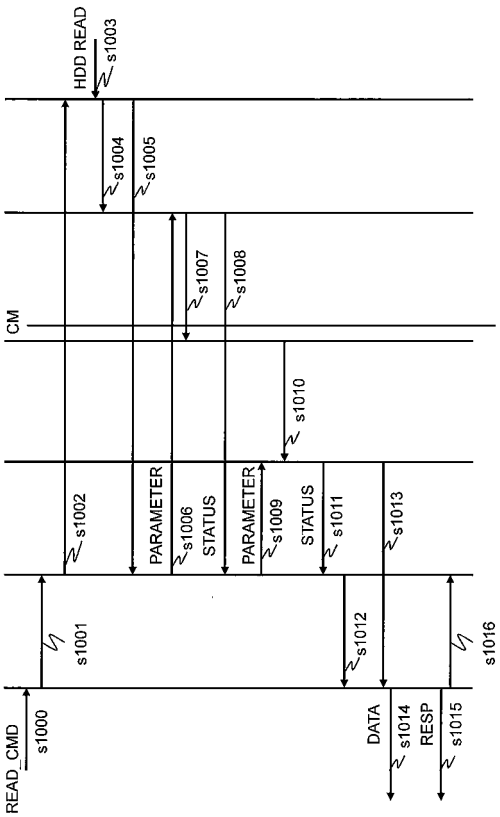

Fig.6

【図７】

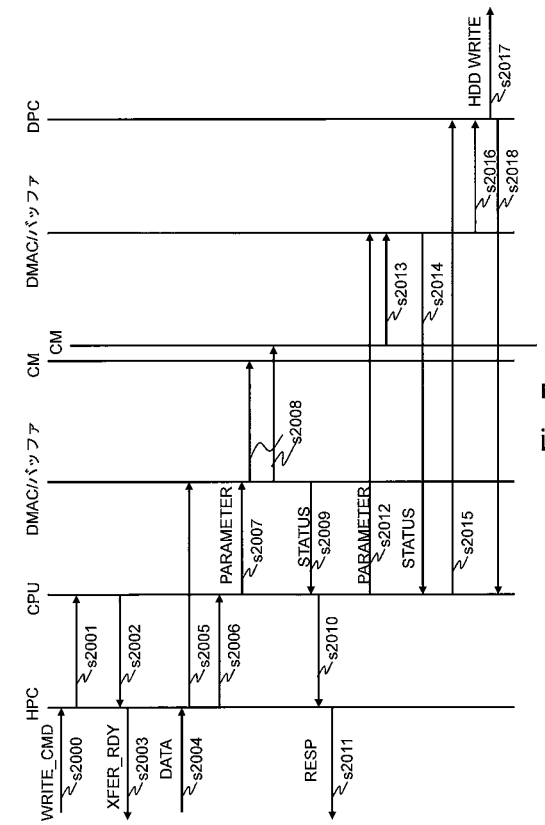

Fig.7

【図８】

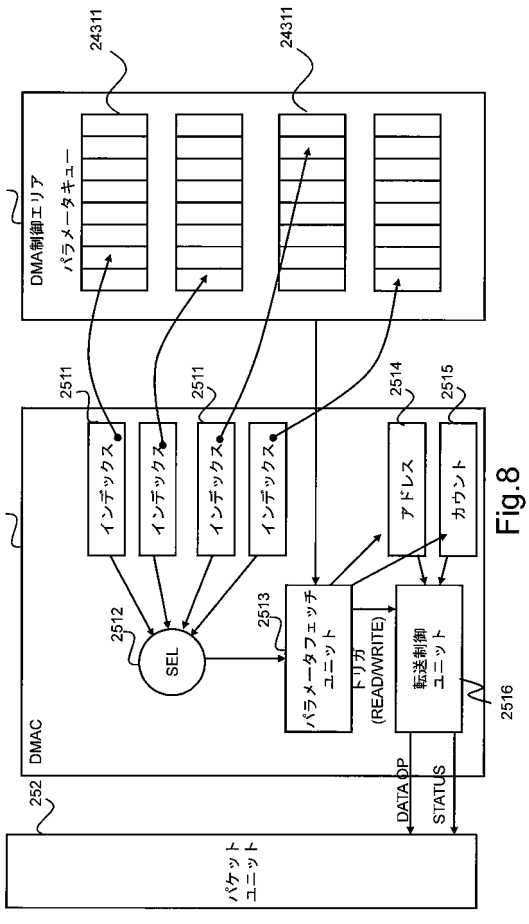

Fig.8

【図９】

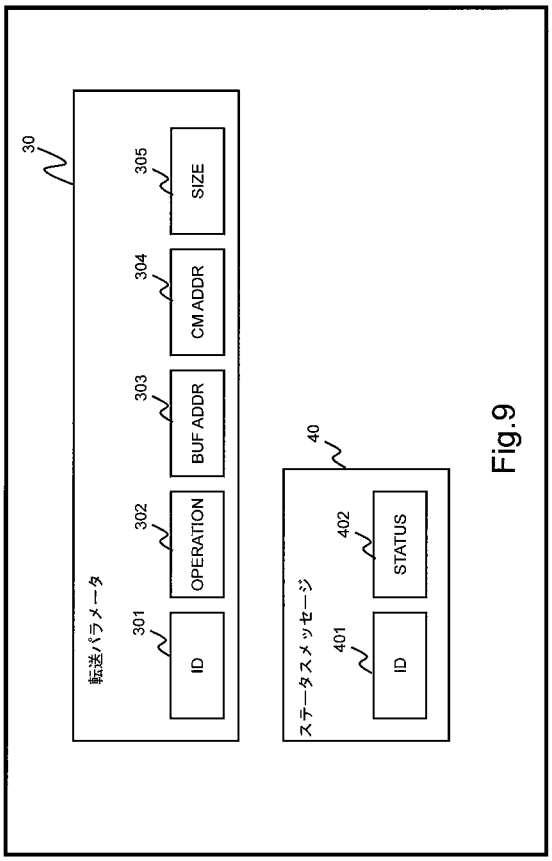

Fig.9

【図１０】

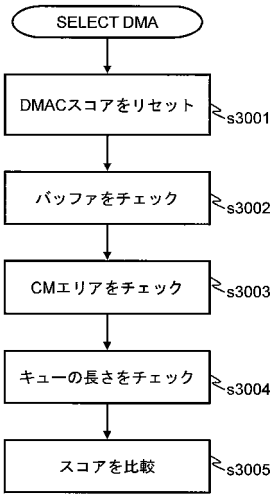

Fig.10

【図１１Ａ】

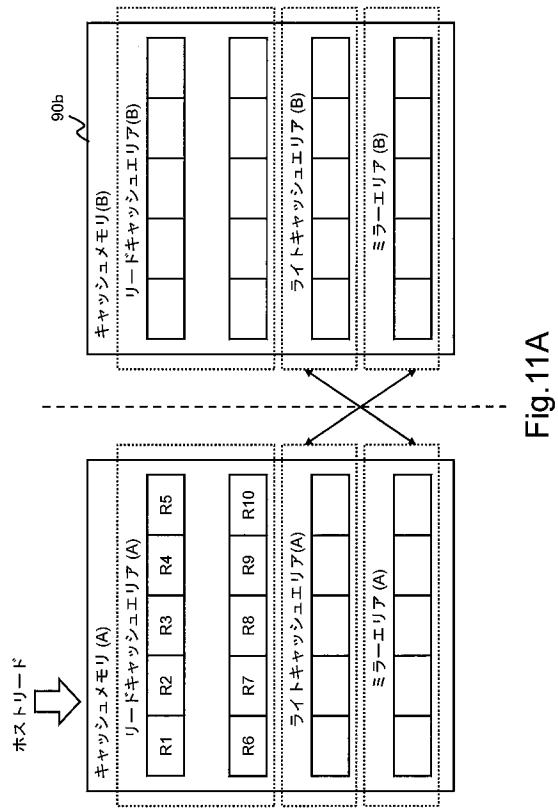

Fig.11A

【図１１Ｂ】

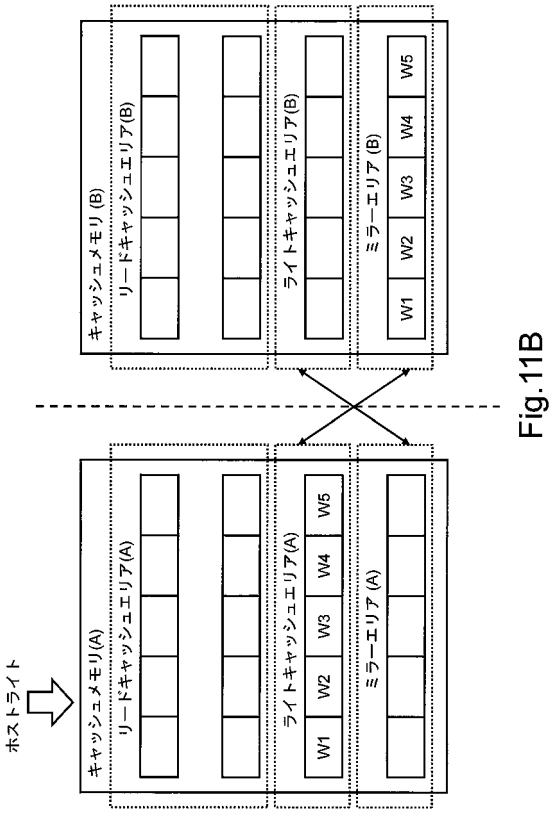

Fig.11B

【図１２Ａ】

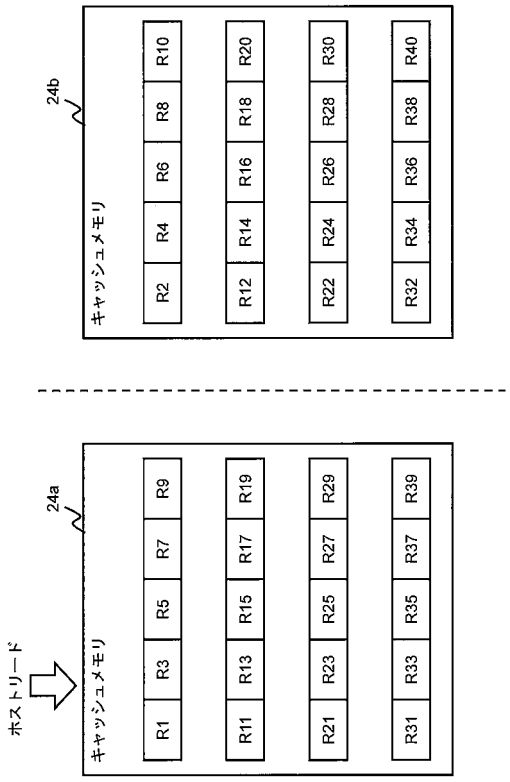

Fig.12A

【図１２Ｂ】

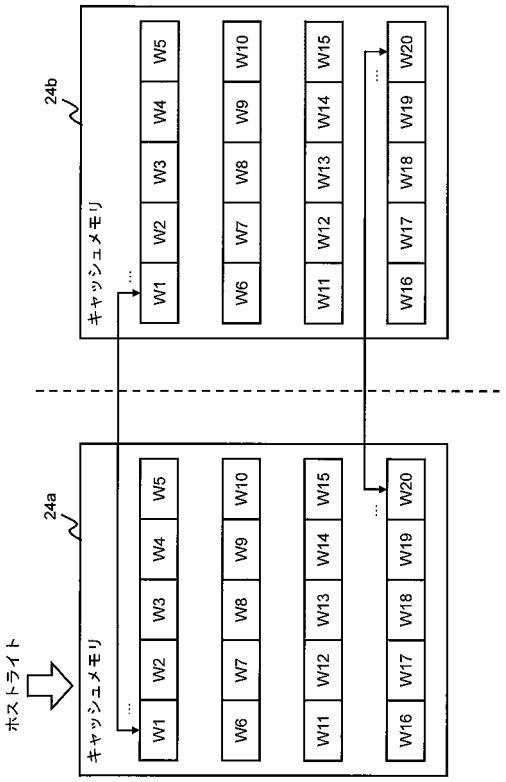

Fig.12B

【図１３】

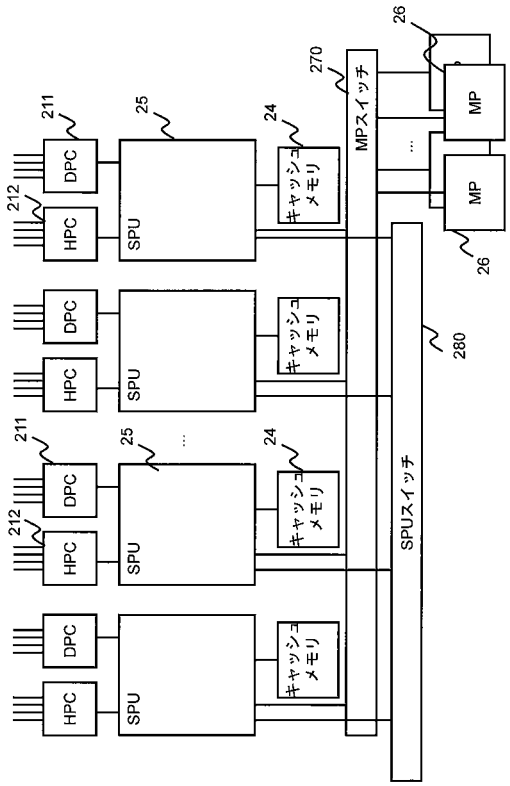

Fig.13

【図１４】

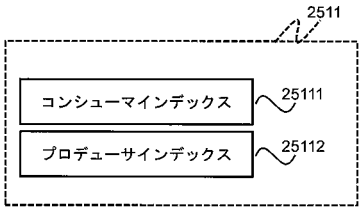

Fig.14

【図１５】

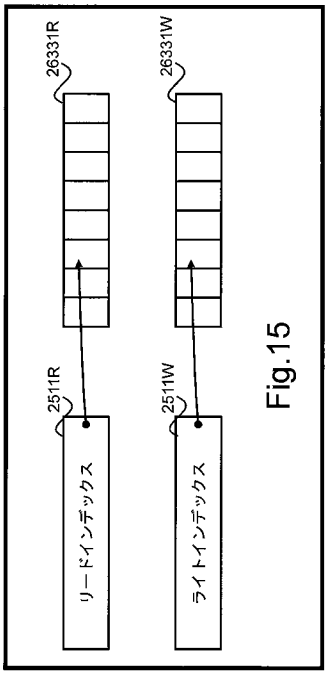

Fig.15

フロントページの続き

(72)発明者　福田　秀明
神奈川県小田原市中里３２２番２号　株式会社日立製作所　ＲＡＩＤシステム事業部内
(72)発明者　箕輪　信幸
神奈川県小田原市中里３２２番２号　株式会社日立製作所　ＲＡＩＤシステム事業部内

審査官　木村　貴俊

(56)参考文献　特開平０９－１２８３０５（ＪＰ，Ａ）
特開平０８－３３５１４４（ＪＰ，Ａ）
特開２００８－１３４７７６（ＪＰ，Ａ）
特開２００１－３１８９０４（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　３／０６－　３／０８
Ｇ０６Ｆ　１２／００－１２／１６
Ｇ０６Ｆ　１３／００