

(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2005-149082
(P2005-149082A)

(43) 公開日 平成17年6月9日(2005.6.9)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
G06F 3/06	G06F 3/06 302A	5B005
G06F 12/00	G06F 3/06 304F	5B014
G06F 12/08	G06F 12/00 514E	5B065
G06F 13/10	G06F 12/00 531D	5B082
	G06F 12/00 545A	
審査請求 未請求 請求項の数 15 O L (全 40 頁) 最終頁に続く		

(21) 出願番号 特願2003-385033 (P2003-385033)
(22) 出願日 平成15年11月14日 (2003.11.14)

(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 110000176
一色国際特許業務法人
(72) 発明者 金井 宏樹
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 加藤 将二
神奈川県横浜市戸塚区戸塚町393番地 日立ハイブリッドネットワーク株式会社内
(72) 発明者 矢内 裕介
神奈川県横浜市中区尾上町6丁目8番地 日立ソフトウェアエンジニアリング株式会社内

最終頁に続く

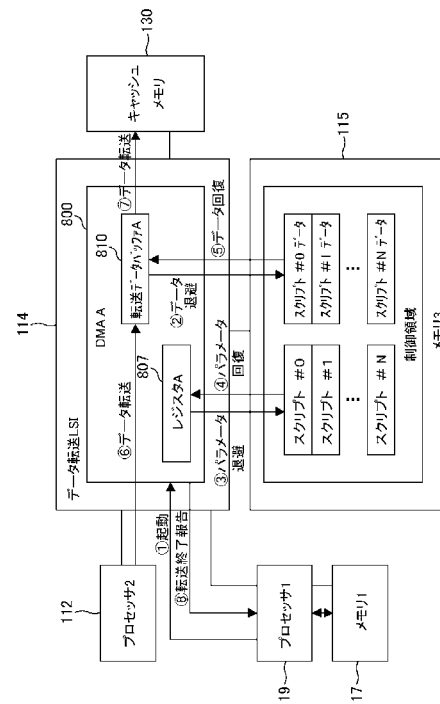
(54) 【発明の名称】 ストレージ制御装置、及びストレージ制御装置の制御方法

(57) 【要約】 (修正有)

【課題】 ストレージ装置に対するデータ入出力性能が、ストレージ装置間で行われる通信により制約されないストレージ制御装置、及びストレージ制御装置の制御方法を提供する。

【解決手段】 第1ストレージ制御装置と、第1ストレージ制御装置に接続される第2ストレージ制御装置とを備えるシステムにおける第1ストレージ制御装置であって、第1メモリと、第2メモリと、データ転送情報を第2メモリに書込む出力制御部と、データバッファ及びデータ転送レジスタを有し第2メモリからのデータ転送情報に基づいて第1メモリと第2ストレージ制御装置との間のデータ転送を制御するデータ転送制御部とを備え、データ転送制御部は第1データ転送情報に基づく第1データ転送を制御している途中で第2データ転送情報に基づく第2データ転送を制御する場合は、第1データ転送情報及びデータバッファに記憶されているデータを第2メモリに書出し第2メモリから第2データ転送情報を読み出して第2データ転送を制御する。

【選択図】 図2 1



【特許請求の範囲】

【請求項 1】

情報処理装置からデータ入出力要求を受信し、データを記憶する第 1 の記憶ボリュームに対してデータ入出力処理を行う第 1 のストレージ制御装置と、前記第 1 のストレージ制御装置と通信可能に接続され、データを記憶する第 2 の記憶ボリュームに対してデータ入出力処理を行う第 2 のストレージ制御装置とを備えて構成されるストレージシステムにおける、前記第 1 のストレージ制御装置であって、

前記第 1 のストレージ制御装置と前記第 2 のストレージ制御装置との間で送受信されるデータを記憶する第 1 のメモリと、

第 2 のメモリと、

前記第 1 のメモリにおけるデータの記憶位置と前記第 2 のストレージ制御装置におけるデータの記憶位置とを含むデータ転送情報を前記第 2 のメモリに書き込む入出力制御部と

データを記憶するデータバッファ、及び前記データ転送情報を記憶するデータ転送レジスタを有し、前記第 2 のメモリから前記データ転送レジスタに読み出した前記データ転送情報に基づいて、前記データバッファを介して、前記第 1 のメモリと前記第 2 のストレージ制御装置との間のデータ転送を制御するデータ転送制御部と、

を備え、

前記データ転送制御部は、

第 1 の前記データ転送情報に基づく第 1 の前記データ転送を制御している途中で、第 2 の前記データ転送情報に基づく第 2 の前記データ転送を制御する場合には、前記データ転送レジスタに記憶されている前記第 1 のデータ転送情報、及び前記データバッファに記憶されているデータを、前記第 2 のメモリに書き出し、

前記第 2 のメモリから前記データ転送レジスタに前記第 2 のデータ転送情報を読み出して、前記第 2 のデータ転送情報に基づいて前記第 2 のデータ転送を制御することを特徴とするストレージ制御装置。

【請求項 2】

前記データ転送制御部は、前記データバッファ及び前記データ転送レジスタをそれぞれ複数有し、

前記第 1 のデータ転送情報に基づく前記第 1 のデータ転送を制御している途中で、前記第 2 のデータ転送情報に基づく前記第 2 のデータ転送を制御する場合には、

前記第 1 のデータ転送情報を記憶している第 1 の前記データ転送レジスタ、及び前記第 1 のデータ転送により送受信されるデータを記憶している第 1 の前記データバッファから、前記第 2 のメモリへの、前記第 1 のデータ転送情報、及び前記第 1 のデータ転送により送受信されるデータの書き出しが完了する前に、前記第 2 のメモリから第 2 の前記データ転送レジスタに前記第 2 のデータ転送情報を読み出して、前記第 2 のデータ転送情報に基づいて、第 2 の前記データバッファを介して、前記第 2 のデータ転送を制御することを特徴とする請求項 1 に記載のストレージ制御装置。

【請求項 3】

前記データ転送は、前記第 1 のストレージ制御装置と前記第 2 のストレージ制御装置との間で送受信されるデータを少なくとも 1 つ以上に分割してなる各データブロックを単位として制御されること

を特徴とする請求項 1 に記載のストレージ制御装置。

【請求項 4】

前記第 1 のストレージ制御装置と、前記第 2 のストレージ制御装置との間は、少なくとも一つの以上のスイッチを介して通信可能に接続されていること

を特徴とする請求項 1 に記載のストレージ制御装置。

【請求項 5】

前記入出力制御部は、

前記情報処理装置から前記第 2 の記憶ボリュームに対する前記データ入出力要求を受信す

10

20

30

40

50

ると、前記データ転送情報を前記第2のメモリに書き込み、

前記データ転送制御部は、

前記第2のメモリから前記データ転送レジスタに前記データ転送情報を読み出し、

前記読み出した前記データ転送情報に基づいて、前記第1のメモリと前記第2のストレージ制御装置との間のデータ転送を制御すること

を特徴とする請求項1に記載のストレージ制御装置。

【請求項6】

前記入出力制御部は、

前記情報処理装置から前記第1の記憶ボリュームに対するデータ書き込み要求及び書き込みデータを受信すると、前記書き込みデータの複製を前記第2の記憶ボリュームにも書き込むべく、前記データ転送情報を前記第2のメモリに書き込み、

10

前記データ転送制御部は、

前記第2のメモリから前記データ転送レジスタに前記データ転送情報を読み出し、

前記読み出した前記データ転送情報に基づいて、前記第1のメモリから前記第2のストレージ制御装置への前記書き込みデータの送信を制御すること

を特徴とする請求項1に記載のストレージ制御装置。

【請求項7】

前記第2のメモリと、前記入出力制御部と、前記データ転送制御部とが形成された回路基板を備えるチャンネル制御部と、

前記第1のメモリが形成された回路基板を備えるキャッシュメモリ部と、

20

前記第1の記憶ボリュームに記憶されるデータの読み書きを行うディスク制御部とを備え、

前記第1のメモリには、

前記第1のストレージ制御装置と、前記情報処理装置又は前記第2のストレージ制御装置の少なくともいずれかとの間で送受信されるデータが記憶され、

前記入出力制御部は、

前記第1のメモリにおけるデータの記憶位置と、前記情報処理装置又は前記第2のストレージ制御装置におけるデータの記憶位置とを含むデータ転送情報を前記第2のメモリに書き込み、

前記データ転送制御部は、

30

前記第2のメモリから前記データ転送レジスタに読み出した前記データ転送情報に基づいて、前記データバッファを介して、前記第1のメモリと、前記情報処理装置又は前記第2のストレージ制御装置との間のデータ転送を制御すること

を特徴とする請求項1に記載のストレージ制御装置。

【請求項8】

情報処理装置からデータ入出力要求を受信し、データを記憶する第1の記憶ボリュームに対してデータ入出力処理を行う第1のストレージ制御装置と、前記第1のストレージ制御装置と少なくとも一つ以上のスイッチを介して通信可能に接続され、データを記憶する第2の記憶ボリュームに対してデータ入出力処理を行う第2のストレージ制御装置とを備えて構成されるストレージシステムにおける、前記第1のストレージ制御装置であって、

40

前記第1のストレージ制御装置と、前記情報処理装置又は前記第2のストレージ制御装置の少なくともいずれかとの間で送受信されるデータを記憶する第1のメモリが形成された回路基板を備えるキャッシュメモリ部と、

第2のメモリ、前記第1のメモリにおけるデータの記憶位置と前記情報処理装置又は前記第2のストレージ制御装置におけるデータの記憶位置とを含むデータ転送情報を前記第2のメモリに書き込む入出力制御部、及びデータを記憶する複数のデータバッファと前記データ転送情報を記憶する複数のデータ転送レジスタとを有し、前記第2のメモリから前記データ転送レジスタに読み出した前記データ転送情報に基づいて、前記データバッファを介して、前記第1のメモリと、前記情報処理装置又は前記第2のストレージ制御装置との間のデータ転送を制御するデータ転送制御部とが形成された回路基板を備えるチャンネル

50

制御部と、

前記第 1 の記憶ボリュームに記憶されるデータの読み書きを行うディスク制御部と、
を備え、

前記データ転送制御部は、

前記第 1 のデータ転送情報に基づく前記第 1 のデータ転送を制御している途中で、前記第 2 のデータ転送情報に基づく前記第 2 のデータ転送を制御する場合には、

前記第 1 のデータ転送情報を記憶している第 1 の前記データ転送レジスタ、及び前記第 1 のデータ転送により送受信されるデータを記憶している第 1 の前記データバッファから、前記第 1 のデータ転送情報、及び前記第 1 のデータ転送により送受信されるデータを前記第 2 のメモリへ書き出すと共に、前記書き出しが完了する前に、前記第 2 のメモリから
第 2 の前記データ転送レジスタに前記第 2 のデータ転送情報を読み出して、前記第 2 のデータ転送情報に基づいて、第 2 の前記データバッファを介して、前記第 2 のデータ転送を
制御すること

を特徴とするストレージ制御装置。

【請求項 9】

情報処理装置からデータ入出力要求を受信し、データを記憶する第 1 の記憶ボリュームに対してデータ入出力処理を行う第 1 のストレージ制御装置と、前記第 1 のストレージ制御装置と通信可能に接続され、データを記憶する第 2 の記憶ボリュームに対してデータ入出力処理を行う第 2 のストレージ制御装置とを備えて構成されるストレージシステムにおける、

前記第 1 のストレージ制御装置と前記第 2 のストレージ制御装置との間で送受信されるデータを記憶する第 1 のメモリと、

第 2 のメモリと、

前記第 1 のメモリにおけるデータの記憶位置と前記第 2 のストレージ制御装置におけるデータの記憶位置とを含むデータ転送情報を前記第 2 のメモリに書き込む入出力制御部と

データを記憶するデータバッファ、及び前記データ転送情報を記憶するデータ転送レジスタを有し、前記第 2 のメモリから前記データ転送レジスタに読み出した前記データ転送情報に基づいて、前記データバッファを介して、前記第 1 のメモリと前記第 2 のストレージ制御装置との間のデータ転送を制御するデータ転送制御部と、

を備える前記第 1 のストレージ制御装置の制御方法であって、

前記データ転送制御部は、

第 1 の前記データ転送情報に基づく第 1 の前記データ転送を制御している途中で、第 2 の前記データ転送情報に基づく第 2 の前記データ転送を制御する場合には、前記データ転送レジスタに記憶されている前記第 1 のデータ転送情報、及び前記データバッファに記憶されているデータを、前記第 2 のメモリに書き出し、

前記第 2 のメモリから前記データ転送レジスタに前記第 2 のデータ転送情報を読み出して、前記第 2 のデータ転送情報に基づいて前記第 2 のデータ転送を制御すること
を特徴とするストレージ制御装置の制御方法。

【請求項 10】

前記データ転送制御部は、前記データバッファ及び前記データ転送レジスタをそれぞれ複数有し、

前記第 1 のデータ転送情報に基づく前記第 1 のデータ転送を制御している途中で、前記第 2 のデータ転送情報に基づく前記第 2 のデータ転送を制御する場合には、

前記第 1 のデータ転送情報を記憶している第 1 の前記データ転送レジスタ、及び前記第 1 のデータ転送により送受信されるデータを記憶している第 1 の前記データバッファから、前記第 2 のメモリへの、前記第 1 のデータ転送情報、及び前記第 1 のデータ転送により送受信されるデータの書き出しが完了する前に、前記第 2 のメモリから第 2 の前記データ転送レジスタに前記第 2 のデータ転送情報を読み出して、前記第 2 のデータ転送情報に基づいて、第 2 の前記データバッファを介して、前記第 2 のデータ転送を制御すること

10

20

30

40

50

を特徴とする請求項 9 に記載のストレージ制御装置の制御方法。

【請求項 1 1】

前記データ転送は、前記第 1 のストレージ制御装置と前記第 2 のストレージ制御装置との間で送受信されるデータを少なくとも 1 つ以上に分割してなる各データブロックを単位として制御されること

を特徴とする請求項 9 に記載のストレージ制御装置の制御方法。

【請求項 1 2】

前記第 1 のストレージ制御装置と、前記第 2 のストレージ制御装置との間は、少なくとも一つの以上のスイッチを介して通信可能に接続されていること

を特徴とする請求項 9 に記載のストレージ制御装置の制御方法。

10

【請求項 1 3】

前記入出力制御部は、

前記情報処理装置から前記第 2 の記憶ボリュームに対する前記データ入出力要求を受信すると、前記データ転送情報を前記第 2 のメモリに書き込み、

前記データ転送制御部は、

前記第 2 のメモリから前記データ転送レジスタに前記データ転送情報を読み出し、

前記読み出した前記データ転送情報に基づいて、前記第 1 のメモリと前記第 2 のストレージ制御装置との間のデータ転送を制御すること

を特徴とする請求項 9 に記載のストレージ制御装置の制御方法。

【請求項 1 4】

20

前記入出力制御部は、

前記情報処理装置から前記第 1 の記憶ボリュームに対するデータ書き込み要求及び書き込みデータを受信すると、前記書き込みデータの複製を前記第 2 の記憶ボリュームにも書き込むべく、前記データ転送情報を前記第 2 のメモリに書き込み、

前記データ転送制御部は、

前記第 2 のメモリから前記データ転送レジスタに前記データ転送情報を読み出し、

前記読み出した前記データ転送情報に基づいて、前記第 1 のメモリから前記第 2 のストレージ制御装置への前記書き込みデータの送信を制御すること

を特徴とする請求項 9 に記載のストレージ制御装置の制御方法。

【請求項 1 5】

30

前記第 2 のメモリと、前記入出力制御部と、前記データ転送制御部とが形成された回路基板を備えるチャンネル制御部と、

前記第 1 のメモリが形成された回路基板を備えるキャッシュメモリ部と、

前記第 1 の記憶ボリュームに記憶されるデータの読み書きを行うディスク制御部と

を備え、

前記第 1 のメモリには、

前記第 1 のストレージ制御装置と、前記情報処理装置又は前記第 2 のストレージ制御装置の少なくともいずれかとの間で送受信されるデータが記憶され、

前記入出力制御部は、

前記第 1 のメモリにおけるデータの記憶位置と、前記情報処理装置又は前記第 2 のストレージ制御装置におけるデータの記憶位置とを含むデータ転送情報を前記第 2 のメモリに書き込み、

40

前記データ転送制御部は、

前記第 2 のメモリから前記データ転送レジスタに読み出した前記データ転送情報に基づいて、前記データバッファを介して、前記第 1 のメモリと、前記情報処理装置又は前記第 2 のストレージ制御装置との間のデータ転送を制御すること

を特徴とする請求項 9 に記載のストレージ制御装置の制御方法。

【発明の詳細な説明】

【技術分野】

【0001】

50

本発明は、ストレージ制御装置、及びストレージ制御装置の制御方法に関する。

【背景技術】

【0002】

近年の情報技術の進歩に伴い、ディザスタリカバリを実現するリモートコピーなど、複数のストレージ装置を通信可能に接続し、情報処理装置からデータ入出力要求を受信したストレージ装置が、他のストレージ装置に記憶されるデータをアクセスするような運用が行われるようになってきている。

【特許文献1】米国特許第5,742,792号明細書

【発明の開示】

【発明が解決しようとする課題】

10

【0003】

しかしながら、この場合には、他のストレージ装置に対するデータ入出力性能は、ストレージ装置間で行われる通信により制約される。

【0004】

本発明は上記課題を鑑みてなされたものであり、ストレージ制御装置、及びストレージ制御装置の制御方法を提供することを主たる目的とする。

【課題を解決するための手段】

【0005】

上記課題を解決するために、本発明は、情報処理装置からデータ入出力要求を受信し、データを記憶する第1の記憶ボリュームに対してデータ入出力処理を行う第1のストレージ制御装置と、前記第1のストレージ制御装置と通信可能に接続され、データを記憶する第2の記憶ボリュームに対してデータ入出力処理を行う第2のストレージ制御装置とを備えて構成されるストレージシステムにおける、前記第1のストレージ制御装置であって、

20

前記第1のストレージ制御装置と前記第2のストレージ制御装置との間で送受信されるデータを記憶する第1のメモリと、第2のメモリと、前記第1のメモリにおけるデータの記憶位置と前記第2のストレージ制御装置におけるデータの記憶位置とを含むデータ転送情報を前記第2のメモリに書き込む入出力制御部と、データを記憶するデータバッファ、及び前記データ転送情報を記憶するデータ転送レジスタを有し、前記第2のメモリから前記データ転送レジスタに読み出した前記データ転送情報に基づいて、前記データバッファを介して、前記第1のメモリと前記第2のストレージ制御装置との間のデータ転送を制御するデータ転送制御部とを備え、前記データ転送制御部は、第1の前記データ転送情報に基づく第1の前記データ転送を制御している途中で、第2の前記データ転送情報に基づく第2の前記データ転送を制御する場合には、前記データ転送レジスタに記憶されている前記第1のデータ転送情報、及び前記データバッファに記憶されているデータを、前記第2のメモリに書き出し、前記第2のメモリから前記データ転送レジスタに前記第2のデータ転送情報を読み出して、前記第2のデータ転送情報に基づいて前記第2のデータ転送を制御することを特徴とするストレージ制御装置に関する。

30

【0006】

その他、本願が開示する課題、及びその解決方法は、発明を実施するための最良の形態の欄、及び図面により明らかにされる。

40

【発明の効果】

【0007】

ストレージ制御装置、及びストレージ制御装置の制御方法を提供することができる。

【発明を実施するための最良の形態】

【0008】

<<<第1の実施の形態>>>

===全体構成例===

まず、本実施の形態に係るストレージ制御装置A(第1のストレージ制御装置)100を含むストレージ装置A600と、ストレージ装置B600を備えて構成されるストレージシステムの全体構成を示すブロック図を図1に示す。なお、ストレージ装置A600の

50

構成とストレージ装置 B 6 0 0 の構成とは同様とすることができるため、特にストレージ装置 A 6 0 0 とストレージ装置 B 6 0 0 とを区別する必要の無い場合には、A や B という識別子をつけない。

【 0 0 0 9 】

ストレージ装置 6 0 0 は、ストレージ制御装置 1 0 0 とストレージ駆動装置 3 0 0 とを備える。ストレージ制御装置 A 1 0 0 は、情報処理装置 2 0 0 からデータ入出力要求を受信し、ストレージ駆動装置 A 3 0 0 が備えるデータを記憶する記憶ボリューム A (第 1 の記憶ボリューム) 3 1 0 に対してデータ入出力処理を行う。ストレージ制御装置 B (第 2 のストレージ制御装置) 1 0 0 は、ストレージ制御装置 A 1 0 0 と通信可能に接続され、ストレージ駆動装置 B 3 0 0 が備えるデータを記憶する記憶ボリューム B (第 2 の記憶ボリューム) 3 1 0 に対してデータ入出力処理を行う。

10

【 0 0 1 0 】

記憶ボリューム 3 1 0 は、例えばハードディスク装置などのディスクドライブにより提供される物理的な記憶領域である物理ボリュームと、物理ボリューム上に論理的に設定される記憶領域である論理ボリュームとを含む、データを記憶するための記憶リソースをいう。以下、記憶ボリューム 3 1 0 又は論理ボリュームを L U (Logical Unit) とも記す。

【 0 0 1 1 】

情報処理装置 2 0 0 は C P U やメモリを備えたコンピュータ等の情報機器である。情報処理装置 2 0 0 が備える C P U により各種プログラムが実行されることにより様々な機能が実現される。情報処理装置 2 0 0 は、例えばパーソナルコンピュータやワークステーションとすることもできるし、メインフレームコンピュータとすることもできる。

20

【 0 0 1 2 】

ストレージ制御装置 A 1 0 0 は、S A N (Storage Area Network) 5 0 0 を介して情報処理装置 2 0 0 と通信可能に接続されている。S A N 5 0 0 を介して行われる情報処理装置 2 0 0 とストレージ制御装置 A 1 0 0 との間の通信は、一般にファイバチャネルプロトコルに従って行われる。この場合 S A N 5 0 0 は、ファイバチャネルプロトコルに準拠した少なくとも一つ以上のスイッチ等の通信機器により構成される。情報処理装置 2 0 0 からは、ストレージ制御装置 A 1 0 0 に対して、ファイバチャネルプロトコルに従ってデータ入出力要求が送信される。ファイバチャネルプロトコルに従って通信が行われる場合には、送受信されるデータは所定のデータ長毎 (例えば 2 キロバイト毎) に一つ以上のデータブロックに分割される。通信は各データブロックを単位として制御されることになる。

30

【 0 0 1 3 】

もちろん、S A N 5 0 0 によらずに情報処理装置 2 0 0 とストレージ制御装置 A 1 0 0 とを接続するようにすることもできる。この場合、情報処理装置 2 0 0 とストレージ制御装置 A 1 0 0 との間の通信は、例えば F I C O N (Fibre Connection) (登録商標) や E S C O N (Enterprise System Connection) (登録商標) 等のメインフレーム系の通信プロトコルとすることもできるし、インターネット等で用いられている T C P / I P (Transmission Control Protocol / Internet Protocol) などの通信プロトコルとすることもできる。

【 0 0 1 4 】

またストレージ制御装置 A 1 0 0 は、S A N 5 0 0 を介して、ストレージ制御装置 B 1 0 0 と通信可能に接続されている。従って、ストレージ制御装置 A 1 0 0 とストレージ制御装置 B 1 0 0 との間の S A N 5 0 0 は、例えばファイバチャネルプロトコルに準拠した少なくとも一つ以上のスイッチ等の通信機器により構成される。ファイバチャネルプロトコルに従って通信が行われる場合には、送受信されるデータは所定のデータ長毎に一つ以上のデータブロックに分割される。通信は各データブロックを単位として制御されることになる。ストレージ制御装置 A 1 0 0 とストレージ制御装置 B 1 0 0 とを接続する S A N 5 0 0 は、ストレージ制御装置 A 1 0 0 と情報処理装置 2 0 0 とを接続する S A N 5 0 0 と通信可能に接続されているようにすることもできるし、通信可能に接続されていないようにすることもできる。もちろん、ストレージ制御装置 A 1 0 0 とストレージ制御装置 B

40

50

100との間は、SAN500によらずに接続されるようにすることもできる。この場合、ストレージ制御装置A100とストレージ制御装置B100との間の通信は、例えばFICON(登録商標)やESCON(登録商標)等のメインフレーム系の通信プロトコルとすることもできるし、インターネット等で用いられているTCP/IPなどの通信プロトコルとすることもできる。

【0015】

ストレージ装置B600は、ストレージ装置A600とは遠隔した場所に設置されるようにすることができる。この場合、ストレージ装置A600とストレージ装置B600との間でリモートコピーを行うようにすることにより、災害などでストレージ装置A600に記憶されるデータが消失あるいは読み書き不能となった場合であっても、ストレージ装置B600に記憶されているデータを用いて、情報処理装置200による情報処理サービスの提供を継続することを可能とするようにすることができる。リモートコピーとは、ストレージ装置A600に記憶されるデータの複製を、ストレージ装置B600に記憶すること、またはその技術をいう。この場合、ストレージ装置A600に記憶されるデータとストレージ装置B600に記憶されるデータとを一致させておくために、情報処理装置200によってストレージ装置A600にデータが書き込まれると、その書き込みデータの複製がストレージ装置A600からストレージ装置B600に送信される。そして、ストレージ装置B600が書き込みデータの複製をストレージ駆動装置B300に記憶する。ストレージ装置A600とストレージ装置B600との間で行われるリモートコピーについては後述する。

10

20

【0016】

また、ストレージ装置B600をストレージ装置A600と通信可能に接続することにより、ストレージコンソリデーションを実現することも可能である。ストレージコンソリデーションとは、複数のストレージ装置に分散して蓄積されている大量のデータを有効に活用するため、各ストレージ装置にばらばらに記憶されている大量のデータを統合管理する技術である。これにより、例えばストレージ装置A600を新たに情報処理システムに導入する場合に、それまで使用してきた古いストレージ装置B600をそのまま使用し、過去からの大量のデータの蓄積を有効に活用することができるようになる。また、異なるメーカーのストレージ装置600にそれぞれ記憶されているデータを統合管理することもできる。この場合、情報処理装置200は、ストレージ装置A600に対してデータ入出力要求を送信することにより、ストレージ装置A600に記憶されるデータのみならず、ストレージ装置B600に記憶されるデータに対してもアクセスすることが可能となる。ストレージコンソリデーションについても詳細は後述する。

30

【0017】

なお、図1に示すストレージシステムには、ストレージ装置A600とストレージ装置B600とが1台ずつ記載されているが、それぞれ複数台とする構成とすることもできる。

【0018】

=== 記憶ボリューム ===

ストレージ駆動装置300は多数のディスクドライブを備えている。これによりストレージ装置600は情報処理装置200に対して大容量の記憶領域を提供する。ストレージ駆動装置300は例えば複数のディスクドライブによりディスクアレイを構成することもできる。この場合、情報処理装置200に対して提供される記憶領域は、RAID (Redundant Arrays of Inexpensive Disks) により管理された複数のディスクドライブにより提供されるようにすることもできる。

40

【0019】

各記憶ボリューム310には、それぞれ固有の識別子(以下、LUN(Logical Unit Number)とも称する)が付与されている。詳細は後述するが、図14に示すように情報処理装置200がストレージ装置A600に送信するデータ入出力要求には、データ記憶領域を指定するためのLUNが記述されている。ストレージ装置A600は、情報処理装置2

50

00に対してLUN指定による記憶領域を提供する機能を備える。

【0020】

ストレージ制御装置100とストレージ駆動装置300との間は図1のように直接に接続される形態とすることもできるし、ネットワークを介して接続されるようにすることもできる。さらにストレージ駆動装置300はストレージ制御装置100と一体として構成されることもできる。

【0021】

===ストレージ制御装置===

ストレージ制御装置100はチャンネル制御部110、共有メモリ120、キャッシュメモリ(キャッシュメモリ部)130、ディスク制御部140、管理端末160、接続部150を備える。 10

チャンネル制御部110は情報処理装置200や他のストレージ装置600との間で通信を行うための通信インタフェースを備え、情報処理装置200や他のストレージ装置600との間でデータ入出力要求やデータを授受する機能を備える。

【0022】

各チャンネル制御部110は管理端末160と共に内部LAN151で接続されている。これにより各チャンネル制御部110に実行させるマイクロプログラム等を管理端末160から送信しインストールすることが可能である。チャンネル制御部110の構成については後述する。

【0023】

接続部150はチャンネル制御部110、共有メモリ120、キャッシュメモリ130、ディスク制御部140を相互に接続する。チャンネル制御部110、共有メモリ120、キャッシュメモリ130、ディスク制御部140間でのデータやコマンドの授受は接続部150を介することにより行われる。接続部150は例えばクロスバスイッチで構成される。 20

【0024】

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により共有される記憶メモリである。共有メモリ120は主に制御情報やコマンド等を記憶するために利用されるのに対し、キャッシュメモリ130は主にデータを記憶するために利用される。キャッシュメモリ130及び共有メモリ120はデータを記憶するためのメモリが形成された回路基板を備えて構成される。 30

【0025】

例えば、あるチャンネル制御部110が情報処理装置200から受信したデータ入出力要求がデータ書き込み要求であった場合には、当該チャンネル制御部110はデータ書き込み要求を共有メモリ120に書き込むと共に、情報処理装置200から受信した書き込み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にデータ書き込み要求が書き込まれたことを検出すると、当該データ書き込み要求に従ってキャッシュメモリ130から書き込みデータを読み出してストレージ駆動装置300に書き込む。

【0026】

またあるチャンネル制御部110が情報処理装置200から受信したデータ入出力要求がデータ読み出し要求であった場合には、読み出し対象となるデータがキャッシュメモリ130に存在するかどうかを調べる。ここでキャッシュメモリ130に存在すれば、チャンネル制御部110はそのデータを情報処理装置200に送信する。一方、読みだし対象となるデータがキャッシュメモリ130に存在しない場合には、当該チャンネル制御部110はデータ読み出し要求を共有メモリ120に書き込むと共に、共有メモリ120を監視する。データ読み出し要求が共有メモリ120に書き込まれたことを検出したディスク制御部140は、ストレージ駆動装置300から読みだし対象となるデータを読み出してこれをキャッシュメモリ130に書き込むと共に、その旨を共有メモリ120に書き込む。そして、チャンネル制御部110は読みだし対象となるデータがキャッシュメモリ130に書き 40 50

込まれたことを検出すると、そのデータを情報処理装置 200 に送信する。

【0027】

このようにチャンネル制御部 110 及びディスク制御部 140 の間では、キャッシュメモリ 130 を介してデータの授受が行われる。なお、キャッシュメモリ 130 と共有メモリ 120 とが、一体的に形成されているような構成とすることもできる。

【0028】

キャッシュメモリ 130 には、ストレージ制御装置 A100 と、情報処理装置 200 又はストレージ制御装置 B100 の少なくともいずれかとの間で送受信されるデータが記憶される。例えば、ストレージ装置 A600 が、情報処理装置 200 からストレージ装置 B600 に対するデータ書き込み要求と書き込みデータを受信した場合には、ストレージ装置 A600 のチャンネル制御部 110 は、当該データ書き込み要求と書き込みデータをキャッシュメモリ 130 に記憶し、ストレージ装置 B600 に送信する。そしてストレージ装置 B600 は、受信したデータ書き込み要求に従って書き込みデータをストレージ駆動装置 B300 に書き込む。また、ストレージ装置 A600 が、情報処理装置 200 からストレージ駆動装置 A300 に対するデータ書き込み要求と書き込みデータを受信した場合には、その書き込みデータをストレージ装置 B600 にリモートコピーする場合には、ストレージ装置 A600 のチャンネル制御部 110 は、キャッシュメモリ 130 に記憶された書き込みデータをストレージ装置 B600 に送信する。そしてストレージ装置 B600 は、受信した書き込みデータをストレージ駆動装置 B300 に書き込む。

10

【0029】

なお、チャンネル制御部 110 からディスク制御部 140 に対するデータの書き込みや読み出しの指示を共有メモリ 120 を介在させて間接的に行う構成の他、例えばチャンネル制御部 110 からディスク制御部 140 に対してデータの書き込みや読み出しの指示を共有メモリ 120 を介さずに直接に行う構成とすることもできる。

20

【0030】

また、チャンネル制御部 110 にディスク制御部 140 の機能を持たせることにより、チャンネル制御部 110 がストレージ駆動装置 300 に記憶されたデータの読み書きを行うようにすることもできる。

【0031】

ディスク制御部 140 はストレージ駆動装置 300 が備える記憶ボリューム 310 に記憶されるデータの読み書きを行う。例えば上述のように、チャンネル制御部 110 が情報処理装置 200 から受信したデータ書き込み要求に従ってストレージ駆動装置 300 へデータの書き込みを行う。

30

【0032】

各ディスク制御部 140 は管理端末 160 と共に内部 LAN151 で接続されており、相互に通信を行うことが可能である。これにより、各ディスク制御部 140 に実行させるマイクロプログラム等を管理端末 160 から送信しインストールすることが可能となっている。ディスク制御部 140 の構成については後述する。

【0033】

本実施例においては、共有メモリ 120 及びキャッシュメモリ 130 がチャンネル制御部 110 及びディスク制御部 140 に対して独立に設けられていることについて記載したが、本実施例はこの場合に限られるものでなく、共有メモリ 120 又はキャッシュメモリ 130 がチャンネル制御部 110 及びディスク制御部 140 の各々に分散されて設けられることも好ましい。この場合、接続部 150 は、分散された共有メモリ 120 又はキャッシュメモリ 130 を有するチャンネル制御部 110 及びディスク制御部 140 を相互に接続させることになる。

40

【0034】

=== 管理端末 ===

管理端末 160 はストレージ装置 600 を保守・管理するためのコンピュータである。管理端末 160 を操作することにより、例えばストレージ駆動装置 300 が備える物理デ

50

ディスク構成の設定や、LU310の設定、チャンネル制御部110やディスク制御部140において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できる。

管理端末160はストレージ制御装置100に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。また管理端末160は、ストレージ制御装置100及びストレージ駆動装置300の保守・管理を専用に行うコンピュータとすることもできるし、汎用のコンピュータに保守・管理機能を持たせたものとすることもできる。

【0035】

管理端末160の構成を示すブロック図を図2に示す。

管理端末160は、CPU161、メモリ162、ポート163、記録媒体読取装置164、入力装置165、出力装置166、記憶装置168を備える。 10

【0036】

CPU161は管理端末160の全体の制御を司るもので、メモリ162に格納されたプログラム162Aを実行することにより上記保守管理機能を実現する。メモリ162にはプログラム162Aが記憶されている。記録媒体読取装置164は、記録媒体167に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ162や記憶装置168に格納される。従って、例えば記録媒体167に記録されたプログラム162Aを、記録媒体読取装置164を用いて上記記録媒体167から読み取って、メモリ162や記憶装置168に格納することができる。また例えば同様に、チャンネル制御部110やディスク制御部140で実行させるための 20
マイクロプログラムが記録された記録媒体167からマイクロプログラムを記録媒体読取装置164を用いて読み取って、メモリ162や記憶装置168に格納し、チャンネル制御部110やディスク制御部140にインストールするようにすることもできる。

記録媒体167としてはフレキシブルディスクやCD-ROM、半導体メモリ等を用いることができる。記録媒体読取装置164は管理端末160に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。記憶装置168は、例えばハードディスク装置や半導体記憶装置等である。入力装置165はオペレータ等による管理 30
端末160へのデータ入力等のために用いられる。入力装置165としては例えばキーボードやマウス等が用いられる。出力装置166は情報を外部に出力するための装置である。出力装置166としては例えばディスプレイやプリンタ等が用いられる。ポート163 30
は内部LAN151に接続されており、これにより管理端末160はチャンネル制御部110やディスク制御部140等と通信を行うことができる。

【0037】

===チャンネル制御部===

チャンネル制御部110の構成を図3に示す。チャンネル制御部110は回路基板118を備えた一つのユニット化されたボードとして構成される。チャンネル制御部110は一枚もしくは複数枚の回路基板118を含んで構成される。回路基板118には、プロセッサ1(119)、プロセッサ2(112)、データ転送LSI(114)、メモリ1(117)、メモリ2(113)、メモリコントローラ1(111)、メモリコントローラ2(111)、メモリ3(115)及びコネクタ116が形成されている。 40

【0038】

プロセッサ1(119)、プロセッサ2(112)、データ転送LSI(114)、メモリコントローラ1(111)、メモリコントローラ2(111)、及びメモリ3(115)はそれぞれPCI(Peripheral Component Interconnect)バスで通信可能に接続されている。なお、必ずしも全てがPCIバスで接続されている必要はなく、一部又は全ての接続が他の規格に準拠する構成とすることもできる。このうち、メモリ1(117)、メモリ3(115)、メモリコントローラ1(111)により第2のメモリが形成される。またデータ転送LSI(114)、プロセッサ2(112)、メモリ2(113)、及びメモリコントローラ2(111)によりデータ転送制御部が形成される。またプロセッサ1(119)により入出力制御部が形成される。 50

【 0 0 3 9 】

プロセッサ 2 (1 1 2)、メモリ 2 (1 1 3) 及びメモリコントローラ 2 (1 1 1) は、情報処理装置 2 0 0 や他のストレージ装置 6 0 0 との間で通信を行うための通信インタフェース機能を提供する。例えば、他のストレージ装置 6 0 0 に送信するデータをファイバチャネルプロトコルに従ったデータ形式に変換して送信する。メモリコントローラ 2 (1 1 1) と接続されるコネクタ 1 1 6 は情報処理装置 2 0 0 や他のストレージ装置 6 0 0 と通信を行うためのコネクタである。プロセッサ 2 (1 1 2) は、メモリ 2 (1 1 3) に記憶された様々なプログラムを実行することにより、上記通信インタフェース機能を実現するための制御を行う。また、情報処理装置 2 0 0 や他のストレージ装置 6 0 0 との間で授受されるデータはメモリ 2 (1 1 3) に記憶される。つまりプロセッサ 2 (1 1 2) は、メモリ 2 (1 1 3) に記憶されるデータを情報処理装置 2 0 0 や他のストレージ装置 6 0 0 に送信し、また情報処理装置 2 0 0 や他のストレージ装置 6 0 0 から受信したデータをメモリ 2 (1 1 3) に記憶する。

【 0 0 4 0 】

プロセッサ 2 (1 1 2)、メモリ 2 (1 1 3)、メモリコントローラ 2 (1 1 1) 及びコネクタ 1 1 6 は、S A N 5 0 0 を構成するファイバチャネルスイッチのポートに接続されるポートとしても機能する。なお、ファイバチャネルの技術分野においては周知であるが、ポートには、S A N 5 0 0 上の機器を特定する識別子である W W N (World Wide Name) が付与されている。

【 0 0 4 1 】

プロセッサ 1 (1 1 9) は、メモリ 1 (1 1 7) に記憶される制御プログラム 7 0 0 を実行することにより、チャンネル制御部 1 1 0 全体の制御を司る。例えば情報処理装置 2 0 0 から送信されたデータ入出力要求を解析して、そのデータ入出力要求がストレージ駆動装置 A 3 0 0 に対するものなのか、ストレージ駆動装置 B 3 0 0 に対するものなのか、ストレージ駆動装置 A 3 0 0 に対するものである場合にはリモートコピーを行う必要があるのか、ないのか等を判断し、それに応じてデータ転送 L S I (1 1 4) にデータ転送の指示を行う。

【 0 0 4 2 】

データ転送 L S I (1 1 4) は、プロセッサ 1 (1 1 9) からの指示により、P C I バスのマスタとして、キャッシュメモリ 1 3 0 と、メモリ 2 (1 1 3) との間のデータ転送を制御する。具体的には、データ転送 L S I (1 1 4) は、プロセッサ 1 (1 1 9) からの指示により、プロセッサ 1 (1 1 9) がメモリ 1 (1 1 7) に書き込んだ、キャッシュメモリ 1 3 0 におけるデータの記憶位置と情報処理装置 2 0 0、ストレージ制御装置 B 1 0 0、又はストレージ装置 B 6 0 0 におけるデータの記憶位置とを含むデータ転送情報をメモリ 1 (1 1 7) から読み出して、データ転送情報に従って、キャッシュメモリ 1 3 0 とメモリ 2 (1 1 3) との間のデータ転送を行う。

【 0 0 4 3 】

キャッシュメモリ 1 3 0 とメモリ 2 (1 1 3) との間で行われるデータ転送の詳細については、後述する。

【 0 0 4 4 】

=== ディスク制御部 ===

次にディスク制御部 1 4 0 の構成を示す図を図 4 に示す。

ディスク制御部 1 4 0 は、インタフェース部 1 4 1、メモリ 1 4 3、C P U 1 4 2、N V R A M (nonvolatile random-access memory) 1 4 4、コネクタ 1 4 5 を備え、これらが一枚もしくは複数枚の回路基板に一体的なユニットとして形成されている。

【 0 0 4 5 】

インタフェース部 1 4 1 は、接続部 1 5 0 を介してチャンネル制御部 1 1 0 等との間で通信を行うための通信インタフェースや、ストレージ駆動装置 3 0 0 との間で通信を行うための通信インタフェースを備える。

C P U 1 4 2 は、ディスク制御部 1 4 0 全体の制御を司ると共に、チャンネル制御部 1 1

0 やストレージ駆動装置 300、管理端末 160 との間の通信を行う。CPU 142 によりメモリ 143 や NVRAM 144 に格納された各種プログラムが実行されることによりディスク制御部 140 の機能が実現される。ディスク制御部 140 により実現される機能としては、ストレージ駆動装置 300 に記憶されるデータの読み書きの制御や RAID の制御等である。

NVRAM 144 は CPU 142 の制御を司るプログラムを格納する不揮発性メモリである。NVRAM 144 に記憶されるプログラムの内容は、管理端末 160 を用いて書き込みや書き換えを行うことができる。

【0046】

=== 情報処理装置 ===

10

次に、情報処理装置 200 の構成を示すブロック図を図 5 に示す。

情報処理装置 200 は、CPU 201、メモリ 202、ポート 203、記録媒体読取装置 204、入力装置 205、出力装置 206 を備える。

【0047】

CPU 201 は情報処理装置 200 の全体の制御を司るもので、メモリ 202 に格納されたプログラムを実行することにより各種機能を実現する。例えば、CPU 201 がアプリケーションプログラム 202B を実行することにより各種情報処理サービスの提供が行われる。例えば銀行の自動預金預け払いサービスや、航空機の座席予約サービスを提供することができる。また CPU 201 が記憶デバイス管理プログラム 202A を実行することにより、記憶ポリューム 310 の管理が行われる。例えば、ストレージ駆動装置 A 300 が備える記憶ポリューム A 310 と、ストレージ駆動装置 B 300 が備える記憶ポリューム B 310 とを対応付け、上記記憶ポリューム A 310 にデータが書き込まれると、上記対応付けられた記憶ポリューム B 310 にそのデータの複製が書き込まれるようにするためのコマンドを、ストレージ制御装置 A 100 に送信することができる。また、記憶ポリューム 310 の LUN の設定なども行うようにすることができる。記録媒体読取装置 204 は、記録媒体 207 に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ 202 に格納される。従って、例えば記録媒体 207 に記録された記憶デバイス管理プログラム 202A やアプリケーションプログラム 202B を、記録媒体読取装置 204 を用いて上記記録媒体 207 から読み取って、メモリ 202 に格納するようにすることができる。記録媒体 207 としてはフレキシブルディスクや CD-ROM、半導体メモリ等を用いることができる。記録媒体読取装置 204 は情報処理装置 200 に内蔵されている形態とすることもできるし、外付されている形態とすることもできる。入力装置 205 はオペレータ等による情報処理装置 200 へのデータ入力等のために用いられる。入力装置 205 としては例えばキーボードやマウス等が用いられる。出力装置 206 は情報を外部に出力するための装置である。出力装置 206 としては例えばディスプレイやプリンタ等が用いられる。ポート 203 はストレージ装置 600 と通信を行うための装置である。ポート 203 には SAN 500 上の機器を特定するための識別子である WWN (World Wide Name) が付与されている。またポート 203 は、他の情報処理装置 200 との間で通信を行うために使用することもできる。この場合、例えば記憶デバイス管理プログラム 202A やアプリケーションプログラム 202B を、ポート 203 を介して他の情報処理装置 200 から受信して、メモリ 202 に格納することもできる。

20

30

40

【0048】

=== データ転送 ===

次に、本実施の形態に係るチャネル制御部 110 におけるキャッシュメモリ 130 と情報処理装置 200 又はストレージ制御装置 B 100 との間のデータ転送について説明する。

【0049】

メモリ 2 (113) は、図 8 に示すように、データ領域を有する。データ領域には、キャッシュメモリ 130 に転送されるデータ、あるいはキャッシュメモリ 130 から転送さ

50

れたデータが記憶される。データ領域に記憶されるこれらの各データの記憶位置や、各データの記憶領域のサイズ（エリア長）等は、図9に示すように、各データに対応してプロセッサ2用スクリプト（データ転送情報）としてメモリ1（117）のスクリプト領域に記憶される。プロセッサ2用スクリプトは、「スクリプト識別子」、「領域先頭アドレス」、「エリア長」、「連続Flag」の各欄を備える。「スクリプト識別子」は、プロセッサ2用スクリプト毎に付される識別番号を示す。「領域先頭アドレス」は、メモリ2（113）のデータ領域に記憶されるデータの記憶アドレスを示す。「エリア長」はデータの記憶領域のサイズを示す。「連続Flag」は、データ領域に記憶される複数のデータをひとまとまりとして転送したい場合に設定されるフラグである。メモリ1（117）に記憶されるこれらのプロセッサ2用スクリプトは、プロセッサ1（119）により設定され、メモリ1（117）のスクリプト領域に書き込まれる。 10

【0050】

一方、図10に示すように、キャッシュメモリ130のデータ領域には、メモリ2（113）に転送されるデータ、あるいはメモリ2（113）から転送されたデータが記憶される。データ領域に記憶されるこれらの各データの記憶位置や、各データの記憶領域のサイズ（エリア長）等は、図9に示すように、各データに対応してプロセッサ1用スクリプト（データ転送情報）としてメモリ1（117）のスクリプト領域に記憶される。プロセッサ1用スクリプトは、「スクリプト識別子」、「領域先頭アドレス」、「エリア長」、「連続Flag」の各欄を備える。これらの各欄の意味は、メモリ2（113）におけるプロセッサ2用スクリプトと同様である。メモリ1（117）に記憶されるこれらのプロセッサ1用スクリプトは、プロセッサ1（119）により設定され、メモリ1（117）のスクリプト領域に書き込まれる。 20

【0051】

次に、これらの各スクリプトに基づいてキャッシュメモリ130とメモリ2（113）との間のデータ転送を制御するデータ転送LSI（114）の構成について、図6及び図7を参照しながら説明する。

【0052】

本実施の形態に係るデータ転送LSI（114）は、4つのDMA800と、PCIインタフェース802とを備える。PCIインタフェース802は、PCIバスを介してメモリコントローラ111等と通信を行うための通信インタフェースである。なお、メモリコントローラ111やキャッシュメモリ等とは必ずしもPCIバスで接続されている必要はなく、一部又は全ての接続が他の規格に準拠する構成とすることもできる。その場合には、PCIインタフェース802に代わって、他の規格に準拠したインタフェースとすることもできる。 30

【0053】

DMA800は、DMA制御部A801、DMA制御部B801、転送データバッファA（データバッファ）810、転送データバッファB（データバッファ）810、レジスタA（データ転送レジスタ）807、レジスタB（データ転送レジスタ）807を備える。レジスタA807及びレジスタB807はそれぞれ、転送起動レジスタ803、転送レジスタ804、初期設定レジスタ805、スクリプトレジスタ806を備える。なお、DMA制御部A801とDMA制御部B801、転送データバッファA810と転送データバッファB810、及びレジスタA807とレジスタB807との構成はそれぞれ同様のため、特に区別する必要の無い場合には、AやBという識別子をつけない。 40

【0054】

DMA制御部801はデータ転送の制御を司る。例えばメモリ1（117）に記憶された上記スクリプト（データ転送情報）のレジスタA807又はレジスタB807への読み込みや、読み込んだスクリプトに基づいて転送データバッファA810又は転送データバッファB810を介して行うデータ転送の制御、データ転送終了ステータスの出力等を行う。DMA制御部801は、ハードウェアのみで構成されるようにすることもできるし、ハードウェアとソフトウェアとの組み合わせにより構成されるようにすることもできる。 50

またDMA制御部801はメモリ3(115)の制御を行う。従って、転送データバッファ810に記憶されるデータやレジスタ807に記憶されるスクリプトを、レジスタメモリ3(115)に書き出したり、メモリ3(115)から読み込んだりすることができる。

【0055】

転送起動レジスタ803は、「要求転送長」、「転送方向」、「プロセッサ2開始スクリプト番号」、「プロセッサ1開始スクリプト番号」の各レジスタを備える。

【0056】

「要求転送長」レジスタは、キャッシュメモリ130と情報処理装置200又はストレージ制御装置B100との間で転送されるデータの総データ長が記載される。要求転送長は、情報処理装置200から受信したデータ入出力要求に基づいて決定される。例えば情報処理装置200から8キロバイト(kB)のデータの書き込み要求を受信した場合には、「要求転送長」レジスタには8kBが記憶される。詳細は後述するが、要求転送長はプロセッサ1(119)がDMA800にデータの転送を開始させる際に、プロセッサ1(119)によりDMA800の転送起動レジスタ803に書き込まれる。

10

【0057】

「転送方向」レジスタは、情報処理装置200又はストレージ制御装置B100からキャッシュメモリ130への転送であるか、キャッシュメモリ130から情報処理装置200又はストレージ制御装置B100への転送であるかが記載される。転送方向は、プロセッサ1(119)が情報処理装置200から受信したデータ入出力要求に基づいて決定される。例えば、プロセッサ1(119)が情報処理装置200からストレージ駆動装置A300へのデータの書き込み要求を受信した場合には、転送方向はメモリコントローラ2(111)からキャッシュメモリ130となる。また、プロセッサ1(119)が情報処理装置200からストレージ駆動装置A300に記憶されたデータの読み出し要求を受信した場合には、転送方向はキャッシュメモリ130からメモリコントローラ2(111)となる。また、ストレージ制御装置A100からストレージ制御装置B100にデータを送信する場合には、転送方向はキャッシュメモリ130からメモリコントローラ2(111)となる。ストレージ制御装置B100からストレージ制御装置A100にデータを受信する場合には、転送方向はメモリコントローラ2(111)からキャッシュメモリ130となる。詳細は後述するが、転送方向は、要求転送長と同様に、プロセッサ1(119)がDMA800にデータの転送を開始させる際に、プロセッサ1(119)によりDMA800の転送起動レジスタ803に書き込まれる。

20

30

【0058】

「プロセッサ2開始スクリプト番号」レジスタには、メモリ1(117)に記憶されたプロセッサ2用スクリプトの識別子が記載される。DMA制御部801は、「プロセッサ2開始スクリプト番号」レジスタに記憶されたプロセッサ2用スクリプトをメモリ1(117)から読み出して、データの転送を開始する。プロセッサ2開始スクリプト番号は、プロセッサ1(119)がDMA800にデータの転送を開始させる際に、プロセッサ1(119)によりDMA800の転送起動レジスタ803に書き込まれる。あるいは、プロセッサ2用スクリプトに連続Flagが設定されている場合には、DMA制御部801により、プロセッサ2開始スクリプト番号が更新される。

40

【0059】

「プロセッサ1開始スクリプト番号」には、メモリ1(117)に記憶されたプロセッサ1用スクリプトの識別子が記載される。DMA制御部801は、「プロセッサ1開始スクリプト番号」レジスタに記憶されたプロセッサ1用スクリプトをメモリ1(117)から読み出して、データの転送を開始する。プロセッサ1開始スクリプト番号は、プロセッサ1(119)がDMA800にデータの転送を開始させる際に、プロセッサ1(119)によりDMA800の転送起動レジスタ803に書き込まれる。あるいは、プロセッサ1用スクリプトに連続Flagが設定されている場合には、DMA制御部801により、プロセッサ1開始スクリプト番号が更新される。

50

【 0 0 6 0 】

転送レジスタ 8 0 4 は、「転送単位」、「転送元アドレス」、「転送先アドレス」、「残転送長」の各レジスタを備える。

「転送単位」レジスタには 1 回のデータ転送におけるデータ転送長が記載される。1 回のデータ転送におけるデータ転送長は、DMA 制御部 8 0 1 により算出される。

「転送元アドレス」レジスタは、転送前のデータの記憶位置を示す。転送元アドレスは、キャッシュメモリ 1 3 0 におけるデータの記憶位置、情報処理装置 2 0 0 におけるデータの記憶位置、又はストレージ装置 B 6 0 0 におけるデータの記憶位置である。転送元アドレスは、転送方向と、プロセッサ 1 用スクリプトあるいはプロセッサ 2 用スクリプトとの記載に基づいて決定される。データの転送中は、転送の進捗と共に DMA 制御部 8 0 1 10 によって「転送元アドレス」レジスタの値が適宜インクリメント、あるいはデクリメントされる。

「転送先アドレス」レジスタは、転送後のデータの記憶位置を示す。転送先アドレスは、キャッシュメモリ 1 3 0 におけるデータの記憶位置、情報処理装置 2 0 0 におけるデータの記憶位置、又はストレージ装置 B 6 0 0 におけるデータの記憶位置である。転送先アドレスは、転送方向と、プロセッサ 1 用スクリプトあるいはプロセッサ 2 用スクリプトとの記載に基づいて決定される。データの転送中は、転送の進捗と共に DMA 制御部 8 0 1 によって「転送先アドレス」レジスタの値が適宜インクリメント、あるいはデクリメントされる。

「残転送長」レジスタは、要求転送長から、転送済みのデータ転送長を引いた値を示す 20 。残転送長は、データ転送の進捗と共に、DMA 制御部 8 0 1 によって適宜更新される。

【 0 0 6 1 】

初期設定レジスタ 8 0 5 は、「プロセッサ 2 用スクリプト領域先頭アドレス」、「プロセッサ 2 用スクリプト数」、「プロセッサ 2 用スクリプトサイズ」、「プロセッサ 1 用スクリプト領域先頭アドレス」、「プロセッサ 1 用スクリプト数」、「プロセッサ 1 用スクリプトサイズ」の各レジスタを備える。初期設定レジスタ 8 0 5 のこれらの値は、例えばチャンネル制御部 1 1 0 のパワーオンリセット時に、DMA 制御部 8 0 1 により、チャンネル制御部 1 1 0 が備える N V R A M から読み込まれるようにすることができる。

【 0 0 6 2 】

「プロセッサ 2 用スクリプト領域先頭アドレス」レジスタは、メモリ 1 (1 1 7) におけるスクリプト領域の先頭アドレスを示す。つまり、プロセッサ 2 用スクリプト 0 が記憶されるアドレスを示す。 30

「プロセッサ 2 用スクリプト数」レジスタは、プロセッサ 2 用スクリプトの数を示す。

「プロセッサ 2 用スクリプトサイズ」レジスタは、プロセッサ 2 用スクリプトのサイズを示す。このようにプロセッサ 2 用スクリプトのサイズを固定とすることにより、プロセッサ 2 用スクリプトの識別子を特定することにより、スクリプト領域の先頭アドレスを基準とした、目的のプロセッサ 2 用スクリプトの記憶アドレスを算出することができる。

「プロセッサ 1 用スクリプト領域先頭アドレス」レジスタ、「プロセッサ 1 用スクリプト数」レジスタ、及び「プロセッサ 1 用スクリプトサイズ」レジスタは、それぞれプロセ 40 ャッサ 2 用スクリプトと同様である。

【 0 0 6 3 】

スクリプトレジスタ 8 0 6 は、「プロセッサ 2 用実行スクリプト番号」、「プロセッサ 2 用有効エリア長」、「プロセッサ 1 用実行スクリプト番号」、「プロセッサ 1 用有効エリア長」の各レジスタを備える。

「プロセッサ 2 用実行スクリプト番号」レジスタは、現在実行中のプロセッサ 2 用スクリプトの識別子を示す。プロセッサ 2 用実行スクリプト番号は、DMA 制御部 8 0 1 により設定される。

「プロセッサ 2 用有効エリア長」レジスタは、プロセッサ 2 用スクリプトのエリア長欄に記載されたエリア長から、転送済みのデータサイズを引いた値である。データ転送の進 50

抄と共に、DMA制御部801によって適宜デクリメントされる。

「プロセッサ1用実行スクリプト番号」レジスタ、及び「プロセッサ1用有効エリア長」レジスタは、「プロセッサ2用実行スクリプト番号」レジスタ、及び「プロセッサ2用有効エリア長」レジスタと同様である。

【0064】

転送データバッファ810は、キャッシュメモリ130とメモリ2(113)との間でデータを転送する際にデータを記憶するバッファである。すなわちDMA制御部801は、キャッシュメモリ130とメモリ2(113)との間でデータを転送する際には、ストアアンドフォワード形式に、転送データバッファ810を介して行う。例えば、キャッシュメモリ130からメモリ2(113)へデータを転送する場合には、DMA制御部801は、キャッシュメモリ130から読み出したデータを転送データバッファ810に一旦書き込み、そして転送データバッファ810からデータを読み出してメモリ2(113)に書き込む。

10

【0065】

転送データバッファ810とメモリコントローラ2(111)との間の通信路は、転送データバッファA810と転送データバッファB810とで共用されている。また転送データバッファ810とキャッシュメモリ130との間の通信路は、転送データバッファA810と転送データバッファB810とでそれぞれ個別に設けられている。これは、転送データバッファ810とメモリコントローラ2(111)との間の通信は、同一回路基板118上における通信であるので、比較的高速にデータ転送が行えるのに対し、転送データバッファ810とキャッシュメモリ130との間は、異なる回路基板を跨る通信であるので、データ転送が比較的低速となるためである。このようにすることより、キャッシュメモリ130との間のデータ転送性能を向上させることが可能となり、DMA800内の2つの転送データバッファ810、DMA制御部801、及びレジスタ807を並列に実行させて2つのデータ転送を同時に行う場合にも、高いデータ転送性能を実現することが可能となる。また、転送データバッファ810とメモリコントローラ2(111)との間のデータ通信路が共有される結果、回路基板118上に占める通信路の面積を減少させることができる。これにより回路基板118上に実装される部品の配置自由度を増すことができるので、回路基板118のパターン設計の容易化を図ることも可能となる。さらに回路基板118、ひいてはストレージ制御装置100を小型化することも可能となる。

20

30

【0066】

===データ入出力の処理===

上述したように、情報処理装置200からデータ入出力要求が送信されると、プロセッサ1(119)はデータ入出力要求を解析して、そのデータ入出力要求がストレージ駆動装置A300に対するものなのか、ストレージ駆動装置B300に対するものなのか、ストレージ駆動装置A300に対するものである場合にはリモートコピーを行う必要があるのか、ないのか等を判断し、それに応じてデータ転送LSI(114)にデータ転送の指示を行う。その判断は、図11に示すように共有メモリ120に記憶されるペア管理テーブル710、及び構成情報管理テーブル720を参照することにより行われる。

【0067】

まず構成情報管理テーブル720を図13に示す。構成情報管理テーブル720は、ストレージ装置A600が、ストレージ装置B600の記憶ポリウム310を、ストレージ装置A600の記憶ポリウム310として情報処理装置200に提供する機能を実現するためのテーブルである。

40

【0068】

構成情報管理テーブル720には、ストレージ装置A600が備える記憶ポリウム310だけでなく、ストレージ装置B600が備える記憶ポリウム310についての情報も記述されている(なお、図13ではストレージ装置B600が備える記憶ポリウム310に関する情報のみが示してある)。

【0069】

50

図13において、PortIDの欄には、記憶ボリューム310に対応付けられているチャンネル制御部110のポートIDが記述される。WWNの欄には、ポートIDに対応するWWNが記述される。LUNの欄には、各記憶ボリューム310のLUNが記述される。Capacity(KB)の欄には、各記憶ボリューム310により提供される記憶容量が記述される。マッピングLUNの欄には、そのポートおよびLUNに対応づけられているストレージ装置B600が備える記憶ボリューム310のLUNが記述される。

【0070】

構成情報管理テーブル720の内容は、例えばストレージ装置A600が備える管理端末160などから登録されるようにすることができる。

【0071】

ストレージ装置A600は、構成情報管理テーブル720を用いることにより、ストレージ装置B600の記憶ボリューム310を、あたかもストレージ装置A600の記憶ボリューム310であるかのように情報処理装置200に提供する。つまり、情報処理装置200は、ストレージ装置B600の記憶ボリューム310に対するデータ入出力要求を、ストレージ装置A600に対して行うことができる。図15に示す例では、情報処理装置200は、ストレージ装置A600のK+1乃至Nで識別される記憶ボリューム310に対して、0乃至Kで識別される記憶ボリューム310と何ら変わることなく、データの読み書きを行うことができるが、実際には、K+1乃至Nで識別される記憶ボリューム310はストレージ装置B600が備える0乃至Mで識別される記憶ボリューム310である。従って、ストレージ装置A600は、情報処理装置200からK+1乃至Nで識別される記憶ボリューム310に対するデータ入出力要求を受信した場合には、ストレージ装置B600との間でデータ転送を行って、ストレージ装置B600の0乃至Mで識別される記憶ボリューム310にアクセスする。

【0072】

つぎに、情報処理装置200からストレージ装置A600に対して、ストレージ装置A600の記憶ボリューム310もしくはストレージ装置B600の記憶ボリューム310に対するデータ入出力要求があった場合の処理について説明する。

【0073】

情報処理装置200からストレージ装置A600に送信されるデータ入出力要求のデータフォーマットの一例を図14に示す。図14において、HostID欄は、そのデータ入出力要求を送信した情報処理装置200の識別子である。例えば情報処理装置200のポート203のWWNが記載される。PortID欄は、そのデータ入出力要求の対象となる記憶ボリューム310に対応付けられているチャンネル制御部110のポートIDである。LUN欄は、データ入出力要求の対象となる記憶ボリューム310のLUNである。アドレス欄は、データ入出力要求の対象となるデータの記憶アドレスが記載される。データ長欄は、データ入出力要求の対象となるデータのデータ長が記載される。

【0074】

情報処理装置200から送信されたデータ入出力要求が、データ書き込み要求であった場合を例に説明する。まず、ストレージ装置A600は、情報処理装置200からデータ書き込み要求を受信すると、プロセッサ1(119)は、このデータ書き込み要求とともに受信した書き込みデータを、キャッシュメモリ130に記憶する。

【0075】

つぎに、プロセッサ1(119)は、構成情報管理テーブル720を参照し、このデータ書き込み要求を実行できるかどうかを判断する。ここで実行できない場合とは、例えば、データ書き込み要求の処理対象となる記憶ボリューム310が存在しないか、存在した場合でも、書き込みデータのサイズが、記憶ボリューム310の記憶容量を超えている場合などである。データ入出力要求を実行できない場合には、情報処理装置200に対してその旨を記載したメッセージを送信し、当該データ入出力要求に対応する処理を終了する。

【0076】

一方、データ書き込み要求を実行できる場合には、プロセッサ1(119)は、ストレ

10

20

30

40

50

ージ装置 B 6 0 0 に対し、データ書き込み要求及び書き込みデータを送信する。送信は、まずプロセッサ 1 (1 1 9) がデータ転送のためのスクリプトをメモリ 1 (1 1 7) に書き込み、次にデータ転送 L S I (1 1 4) が、メモリ 1 (1 1 7) からレジスタ 8 0 7 にスクリプトを読み出し、読み出したスクリプトに基づいてメモリ 1 (1 1 7) とストレージ制御装置 B 1 0 0 との間のデータ転送を制御することにより行われる。そしてストレージ装置 B 6 0 0 は、このデータ書き込み要求を受信して、書き込みデータを記憶ボリューム 3 1 0 B に書き込む。データの読み出しの場合も同様である。

【 0 0 7 7 】

なお、ストレージ装置 A 6 0 0 からストレージ装置 B 6 0 0 に送信されるデータ書き込み要求は、ストレージ装置 B 6 0 0 が情報処理装置 2 0 0 から直接に受信する場合のデータ書き込み要求と同一のプロトコル (例えば、データフォーマットが同一) に従うものである。従って、ストレージ装置 B 6 0 0 が、ストレージ装置 A 6 0 0 に記憶ボリューム 3 1 0 を提供する装置として機能するように運用される場合には、ストレージ装置 B 6 0 0 の装置構成や仕様についてとくに変更や改造をする必要が無く、最小限の手間で、かつ、低コストでストレージ装置 B 6 0 0 の有効利用が可能となる。

【 0 0 7 8 】

次にペア管理テーブル 7 1 0 を図 1 2 に示す。

ペア管理テーブル 7 1 0 は、「ペア種類」欄、「複製方式」欄、「複製元装置」欄、「複製先装置」欄、「複製元ボリューム」欄、「複製先ボリューム」欄、「ペア状態」欄を備える。

【 0 0 7 9 】

ペアとは 2 つの記憶ボリューム 3 1 0 により形成される記憶ボリューム 3 1 0 の組み合わせを言う。また、ペアを形成する 2 つの記憶ボリューム 3 1 0 が同一のストレージ装置 6 0 0 にある場合を「ローカルペア」、異なるストレージ装置 6 0 0 にある場合を「リモートペア」と表現する。ペアを形成する記憶ボリューム 3 1 0 は、一方を主記憶ボリューム 3 1 0、他方を副記憶ボリューム 3 1 0 として管理される。一つの主記憶ボリューム 3 1 0 に対して複数の副記憶ボリューム 3 1 0 を組み合わせることも可能である。

【 0 0 8 0 】

情報処理装置 2 0 0 が、データの複製元となるストレージ装置 A 6 0 0 にリモートペアの形成を指示すると、ストレージ装置 A 6 0 0 は共有メモリ 1 2 0 のペア管理テーブル 7 1 0 を更新する。その後ストレージ装置 A 6 0 0 は、データの複製先となるストレージ装置 B 6 0 0 にリモートペアの形成を指示する。そしてストレージ装置 B 6 0 0 は、ストレージ装置 B 6 0 0 内の共有メモリ 1 2 0 のペア管理テーブル 7 1 0 を更新する。

【 0 0 8 1 】

ペア管理テーブル 7 1 0 の「ペア種類」欄は当該ペアがローカルペアであるかリモートペアであるかを示す。「複製方式」欄は、当該ペアがリモートペアである場合に、リモートコピーの方式が同期方式であるか非同期方式であるかを示す。「複製元装置」欄と「複製先装置」欄は、当該ペアがリモートペアである場合に、複製元のストレージ装置 6 0 0 と複製先のストレージ装置 6 0 0 を示す。「複製元ボリューム」欄は当該ペアの主記憶ボリューム 3 1 0 の L U N を、「複製先ボリューム」欄は当該ペアの副記憶ボリューム 3 1 0 の L U N を示す。

【 0 0 8 2 】

「ペア状態」欄は当該ペアの状態を示す。ペアの状態としては、「ペア中」、「スプリット中」、及び「リシンク中」がある。

「ペア中」の場合は、情報処理装置 2 0 0 から主記憶ボリューム 3 1 0 に書き込まれたデータの複製が副記憶ボリューム 3 1 0 にも反映される。このような主記憶ボリューム 3 1 0 と副記憶ボリューム 3 1 0 との対応付けにより、主記憶ボリューム 3 1 0 に記憶されている内容と副記憶ボリューム 3 1 0 に記憶されている内容との同一性を確保することができる。

「スプリット中」の場合は、情報処理装置 2 0 0 から主記憶ボリューム 3 1 0 にデータが書き込まれても副記憶ボリューム 3 1 0 には反映されない。

10

20

30

40

50

「リシンク中」は「スプリット中」から「ペア中」に移行する途中の状態である。すなわち、「スプリット中」に主記憶ポリューム310に対してなされたデータの更新を副論理ポリューム310に反映している状態である。反映が完了すると当該ペアの状態は「ペア中」になる。

【0083】

上記ペアの形成、ペアのスプリット、及びペアのリシンクは、記憶デバイス管理プログラム202Aが実行されている情報処理装置200に対して入力装置205からオペレータが指示入力を与えることにより行われるようにすることができる。オペレータにより与えられた指示入力はストレージ装置600のチャンネル制御部110に送信される。チャンネル制御部110は制御プログラム700を実行して、上記指示に従って、ペアの形成やペア状態の変更を行う。チャンネル制御部110は、形成したペアのペア状態に応じて、例えば「ペア中」主記憶ポリューム310に対してデータの書き込み要求を情報処理装置200から受信した場合には、書き込みデータの複製をストレージ装置B600に送信し、副記憶ポリューム310へ反映する。具体的には、まずストレージ装置A600は、情報処理装置200からデータ書き込み要求を受信すると、プロセッサ1(119)は、このデータ書き込み要求とともに受信した書き込みデータを、キャッシュメモリ130に記憶する。そしてプロセッサ1(119)は、ディスク制御部140に対して、記憶ポリュームA310へのデータの書き込みを指示すると共に、ペア管理テーブル710を参照して、データの複製が書き込まれる記憶ポリュームB310を特定する。そしてプロセッサ1(119)は、書き込みデータの複製を記憶ポリュームB310にも書き込むべく、データ転送のためのスクリプトをメモリ1(117)に書き込み、次にデータ転送LSI(114)が、メモリ1(117)からレジスタ807にスクリプトを読み出し、読み出したスクリプトに基づいてメモリ1(117)からストレージ制御装置B100への書き込みデータの送信を制御することにより行われる。そしてストレージ装置B600は、このデータ書き込み要求を受信して、書き込みデータを記憶ポリュームB310に書き込む。

10

20

【0084】

このように、本実施の形態に係るストレージシステムにおいては、ストレージ装置A600が情報処理装置200から受信したデータ入出力要求に応じて、ストレージ装置A600とストレージ装置B600との間でデータ転送が行われる。このデータ転送は、上述したようにプロセッサ1(119)を含む入出力制御部、及びデータ転送LSI(114)を含むデータ転送制御部により行われるが、通常は、ストレージ装置A600とストレージ装置B600との間で第1のデータ転送が行われている間は、第2のデータ転送は、第1のデータ転送が終了するまで実行されることはない。

30

【0085】

すなわち図16に示すように、情報処理装置200がストレージ装置B600に記憶されているデータを読み出す場合を例に説明すると、ストレージ装置A600が情報処理装置200から受信した(S1000)リードコマンド(データ読み出し要求)1をストレージ装置B600に送信(S1001)し、ストレージ装置B600が記憶ポリューム310からデータの読み出しを行って(S1002)、読みだされたデータをストレージ装置B600から受信して(S1003)、情報処理装置200に送信する(S1004)までは、リードコマンド2の処理は行われない。図16には、リードコマンド1の処理が終了した後に情報処理装置200からリードコマンド2の送信が行われる場合が示されているが、リードデータ(読み出しデータ)1の読み出し処理が終了する前にリードコマンド2が送信されたとしても、リードコマンド2の処理はリードデータ1の読み出しが終了してから行われる(S1005乃至S1009)。これは、データ転送LSI(114)が第1のデータ転送を行っている間に、第2のデータ転送を行うことができるように構成されていないからである。

40

【0086】

しかしながら本実施の形態に係るストレージ制御装置A100においては、図17及び図18に示すように、ストレージ装置B600からリードデータ1の受信が開始される前に情報処理装置200からリードコマンド2を受信した場合には、リードデータ1の受信

50

を開始する前に、リードコマンド2をストレージ装置B600に送信することができる。図17に示す例は、ストレージ装置B600からリードデータ1を受信する前にリードコマンド2をストレージ装置B600に送信し、ストレージ装置B600からリードデータ1の受信が終了した後にリードデータ2の受信が行われる場合の例である(S2000乃至S2009)。図18に示す例は、ストレージ装置B600からリードデータ1を受信する前にリードコマンド2をストレージ装置B600に送信し、ストレージ装置B600からリードデータ1の受信が終了する前に、リードデータ2の受信も行われる場合の例である(S3000乃至S3013)。

【0087】

なお、図16乃至図18において、一つのリードコマンドに対応して複数のリードデータが送信されるように記載されているが、これは、データ転送がデータブロック単位に制御されることを表したものである。

【0088】

=== データ転送の流れ ===

次に、本実施の形態に係るデータ転送の処理の流れについて図25乃至図29のフローチャートに従って説明する。

【0089】

まずストレージ制御装置A100が、情報処理装置200からデータ入出力要求を受信すると、プロセッサ2(112)はメモリ2(113)にデータ領域を確保する(S4000)。データ領域を確保するとは、情報処理装置200から送信されたデータ入出力要求が書きこみ要求の場合は、書き込みデータをメモリ2(113)に記憶するということである。また情報処理装置200から送信されたデータ入出力要求が読み出し要求の場合は、読み出しデータを記憶するための記憶領域をメモリ2(113)に確保するということである。ここで確保されるデータ領域は一つとは限らない。メモリ2(113)におけるデータの記憶可能領域の分布状況等によっては、一つのデータ入出力要求に係るデータを複数に分割してメモリ2(113)に記憶せざるを得ない場合もあるからである。

【0090】

続いてプロセッサ2(112)は、情報処理装置200から受信したデータ入出力要求とメモリ2(113)に確保したデータ領域に関する情報とをプロセッサ1(119)に送信する(S4001)。

そうするとプロセッサ1(119)はプロセッサ2(112)から送信されたデータ入出力要求を解析する(S4002)。これによりプロセッサ1(119)は要求転送長と転送方向とを認識する。

続いてプロセッサ1(119)は、要求転送長で指定されたサイズのデータ領域をキャッシュメモリ130に確保する(S4003)。ここでもキャッシュメモリ130における記憶可能メモリ空間の分布状況等によっては、複数のデータ領域が確保される場合がある。

【0091】

そしてプロセッサ1(119)は、キャッシュメモリ130に確保した各データ領域に対応して、プロセッサ1用スクリプトを作成する。同様に、プロセッサ2(112)から送信された、データ入出力要求とメモリ2(113)に確保されたデータ領域に関する情報とから、メモリ2(113)に確保された各データ領域に対応してプロセッサ2用スクリプトを作成する(S4004)。複数のデータ領域が確保された場合には、各プロセッサ1用スクリプト及び各プロセッサ2用スクリプトをひとまとまりに管理するために、連続Flagがセットされる。

これによりキャッシュメモリ130におけるデータの記憶位置を示す情報を含むプロセッサ1用スクリプトと、情報処理装置200又はストレージ装置B600におけるデータの記憶位置を示す情報を含むプロセッサ2用スクリプトがメモリ1(117)に書き込まれる。

【0092】

その後プロセッサ1(119)は、データ転送LSI(114)が備えるDMA800

10

20

30

40

50

の内のレジスタ A 8 0 7 又はレジスタ B 8 0 7 に、要求転送長、転送方向、プロセッサ 2 用スクリプトの識別子、及びプロセッサ 1 用スクリプトの識別子を含む転送起動情報を送信し、データ転送 L S I (1 1 4) が備える D M A 8 0 0 の転送起動レジスタ 8 0 3 にこれらのデータを書き込む。これによりプロセッサ 1 (1 1 9) はデータ転送 L S I (1 1 4) を起動する (S 4 0 0 5) 。

そうすると、D M A 8 0 0 は D M A 転送処理を開始する (4 0 0 6) 。 D M A 転送処理については、図 2 6 に示すフローチャートに従って説明する。

【 0 0 9 3 】

まず D M A 制御部 8 0 1 が、転送起動情報に記載されたプロセッサ 1 用スクリプトの識別子とプロセッサ 2 用スクリプトの識別子とにより、プロセッサ 1 用スクリプトとプロセッサ 2 用スクリプトとをそれぞれ取得する (S 5 0 0 0) 。スクリプトの取得は図 2 7 に示すフローチャートに従って行われる。

【 0 0 9 4 】

まず D M A 制御部 8 0 1 は、転送起動情報に基づいて、プロセッサ 1 (1 1 9) から送信された転送起動情報に記載されたプロセッサ 1 用スクリプトの識別子と、初期設定レジスタ 8 0 5 に記憶されたプロセッサ 1 用スクリプトサイズとから、メモリ 1 (1 1 7) におけるプロセッサ 1 用スクリプトの記憶アドレスを算出する (S 6 0 0 0) 。プロセッサ 1 用スクリプトの記憶アドレスが算出できたら、続いてメモリ 1 (1 1 7) のその記憶アドレスからプロセッサ 1 用スクリプトを読み出す (S 6 0 0 1) 。ここで、D M A 制御部 8 0 1 は、プロセッサ 1 用スクリプトに記載された領域先頭アドレスを、転送方向に応じて転送レジスタ 8 0 4 の転送先アドレス欄又は転送元アドレス欄に記載する。そしてスクリプトレジスタ 8 0 6 のプロセッサ 1 用実行スクリプト番号欄、プロセッサ 1 用有効エリア長欄にそれぞれ、プロセッサ 1 用スクリプトに記載されたスクリプト識別子、エリア長欄に記載されたエリア長を書き込む (S 6 0 0 2) 。

【 0 0 9 5 】

同様に、D M A 制御部 8 0 1 は、転送起動情報に基づいて、プロセッサ 1 (1 1 9) から送信された転送起動情報に記載されたプロセッサ 2 用スクリプトの識別子と、初期設定レジスタ 8 0 5 に記憶されたプロセッサ 2 用スクリプトサイズとから、メモリ 1 (1 1 7) におけるプロセッサ 2 用スクリプトの記憶アドレスを算出する (S 6 0 0 0) 。プロセッサ 2 用スクリプトの記憶アドレスが算出できたら、続いてメモリ 1 (1 1 7) のその記憶アドレスからプロセッサ 2 用スクリプトを読み出す (S 6 0 0 1) 。

ここで、D M A 制御部 8 0 1 は、プロセッサ 2 用スクリプトに記載された領域先頭アドレスを、転送方向に応じて転送レジスタ 8 0 4 の転送元アドレス欄又は転送先アドレス欄に記載する。そしてスクリプトレジスタ 8 0 6 のプロセッサ 2 用実行スクリプト番号欄、プロセッサ 2 用有効エリア長欄にそれぞれ、プロセッサ 2 用スクリプトに記載されたスクリプト識別子、エリア長欄に記載されたエリア長を書き込む (S 6 0 0 2) 。

【 0 0 9 6 】

続いて D M A 制御部 8 0 1 は転送単位を決定する (S 5 0 0 1) 。転送単位の決定は図 2 8 に示すフローチャートに従って行われる。

まず D M A 制御部 8 0 1 は、転送レジスタ 8 0 4 の残転送長欄に残転送長を書き込む (S 7 0 0 0) 。転送開始時は、要求転送長を書き込む。

そしてプロセッサ 1 の有効エリア長とプロセッサ 2 の有効エリア長と、残転送長のうちの最小のものを転送単位とする (S 7 0 0 1 乃至 S 7 0 0 7) 。

そして D M A 制御部 8 0 1 は、転送レジスタ 8 0 4 に記憶された転送元アドレスに記憶されたデータを、転送先アドレスに、転送単位のデータサイズだけ転送する (S 5 0 0 2) 。転送は、転送データバッファ 8 1 0 を介して行われる。

転送実行中は、転送レジスタ 8 0 4 の転送元アドレス、転送先アドレス、及び残転送長が順次更新されると共に、スクリプトレジスタ 8 0 6 のプロセッサ 2 用有効エリア長及びプロセッサ 1 用有効エリア長が順次転送済みデータのサイズの分ずつ減少される。

スクリプトレジスタ 8 0 6 のプロセッサ 2 用有効エリア長、プロセッサ 1 用有効エリア

長、又は残転送長のいずれかの値が0になったら、DMA制御部801は転送終了判定処理を行う(S5003)。

【0097】

転送終了判定処理は図29に示すフローチャートに従って行われる。

まずDMA制御部801は、転送レジスタ804の残転送長を確認する。残転送長が0の場合は要求転送長の全データの転送が完了しているので、“No”に進み、処理を終了する(S8000)。

【0098】

一方、残転送長が0でない場合は、プロセッサ1用有効エリア長あるいはプロセッサ2用有効エリア長の少なくともいずれかが一方が0である。この場合、有効エリア長が0となった方のスクリプトには、連続Flagで連結された、次に実行すべきスクリプトが存在する。S8001乃至S8004の処理により、連続Flagで連結された、次に実行すべきスクリプトが存在するスクリプトは、プロセッサ1用スクリプトであるのか、あるいはプロセッサ2用スクリプトであるのかが決定される。

【0099】

まずS8001において、スクリプトレジスタ806のプロセッサ1用有効エリア長が0であるか否かを確認する。0であれば“Yes”に進み、S8002においてプロセッサ1用開始スクリプト番号を更新する。0でなければ“No”に進む。

続いてプロセッサ2用有効エリア長が0であるか否かを確認する(S8003)。0であれば“Yes”に進み、S8004においてプロセッサ2用開始スクリプト番号を更新する。0でなければ“No”に進む。

続いてDMA制御部801はS5000を再び実行する。この際DMA制御部801は、S8002又はS8004において更新された転送起動レジスタ803のプロセッサ1用開始スクリプト番号又はプロセッサ2用開始スクリプトで指定されるスクリプトを読み出す。そしてDMA制御部801は新たに読み出したスクリプトに従って、DMA転送を行う。

最終的に、S5003における転送終了判定処理において転送レジスタ804の残転送長が0になったら、DMA制御部801は転送処理を終了する。そしてDMA制御部801は、転送処理の結果が記載された終了ステータスをメモリ1(117)に書き込み(S4007)、プロセッサ1(119)に終了通知を行う(S4008)。この終了通知は、例えば割り込み信号の送信により行うことができる。

【0100】

プロセッサ1(119)は、メモリ1(117)から終了ステータスを読み出して、その内容に応じた処理を実行する(S4009)。そしてプロセッサ1(119)はプロセッサ2(112)に終了通知を送信する(S4010)。

そうするとプロセッサ2(112)は、データ入出力要求がデータの読み出し要求であった場合には、メモリ2(113)のデータ領域に転送された読み出しデータを読み出して、情報処理装置200に送信することができる。またデータ入出力要求がデータの書き込み要求であった場合には、書き込み終了報告を情報処理装置200に送信する。

【0101】

ここで、上述したように、本実施の形態に係るDMA800は、メモリ3(115)の制御を行うことができる。従って、レジスタ807に記憶されるデータ転送情報、及び転送データバッファ810に記憶されているデータを、メモリ3(115)へ書き出したり、メモリ3(115)から読み込んだりすることができる。その様子を図19及び図20に示す。メモリ3(115)へのデータの書き出しや読み込みの制御はDMA制御部801により行われる。

【0102】

このため、本実施の形態に係るストレージ制御装置100では、第1のデータ転送情報に基づく第1のデータ転送を制御している途中でも、第2のデータ転送情報に基づく第2のデータ転送を制御することができる。つまり、レジスタ807に記憶されている第1のデータ転送情報、及び転送データバッファ810に記憶されているデータを、メモリ3(

10

20

30

40

50

115)に書き出して、メモリ3(115)からレジスタ807に第2のデータ転送情報を読み出すことにより、第1のデータ転送を制御している途中でも、第2のデータ転送を制御することができるのである。図20に示すように、メモリ3(115)に書き出されるデータ転送情報やデータは一つに限られず、複数とすることができる。

【0103】

なお、図20には、メモリ3(115)に書き出されるデータ転送情報のうちの少なくとも一部が記載されている。すなわち、“転送元アドレス”、“転送先アドレス”、“転送長”、“転送コマンド”、“制御情報”、“チェックコード”である。“転送元アドレス”、及び“転送先アドレス”は、転送レジスタ804に記憶されている、それぞれ転送元アドレス、転送先アドレスである。また“転送長”は、転送レジスタ804に記憶されている残転送長とすることができるし、転送起動レジスタ803に記憶される要求転送長から残転送長を引いた値とすることもできる。また“転送コマンド”は、転送起動レジスタ803に記憶されている転送方向とすることもできるし、情報処理装置200からストレージ制御装置100に送信されたデータ入出力要求とすることもできる。後者の場合には、DMA800にはデータ入出力要求を記憶するレジスタが設けられ、そのレジスタに記憶されるデータ入出力要求がメモリ3(115)に書き出される。“制御情報”は、DMA制御部801がデータ転送の制御の際に記憶しておく各種ステータス情報とすることができる。また“チェックコード”は、転送されるデータに基づいて計算されるチェックコードの途中結果である。チェックコードは例えばチェックサムや、CRC(Cyclic Redundancy Check)とすることができる。

【0104】

このように、レジスタ807に記憶されるデータ転送情報、及び転送データバッファ810に記憶されているデータを、メモリ3(115)へ書き出すことにより、第1のデータ転送情報に基づく第1のデータ転送を制御している途中に、第2のデータ転送情報に基づく第2のデータ転送を制御することができる。第1のデータ転送情報に基づく第1のデータ転送を制御している途中に、第2のデータ転送情報に基づく第2のデータ転送を制御する場合の処理の流れを図21及び図22を用いて説明する。

【0105】

図21及び図22に記載されている丸付き数字は処理の順序を示す。すなわちまずプロセッサ1(119)は、DMA800にデータ転送の開始を指示する(DMA800を起動)。そうすると、DMA制御部801は、レジスタ807に記憶されている第1のデータ転送情報と転送データバッファ810に記憶されているデータ転送中のデータをメモリ3(115)に書き出す(退避させる)。書き出しが完了したら、DMA制御部801は、メモリ3(115)から、第2のデータ転送情報と第2のデータ転送情報に対応するデータを、それぞれレジスタ807と転送データバッファ810に読み込む。そしてDMA制御部801は、転送データバッファ810を介して第2のデータ転送を行う。第2のデータ転送が終了したら、プロセッサ1(119)に終了報告を送信する。

【0106】

このようにすることにより、本実施の形態に係るストレージ制御装置100では、第1のデータ転送が終了するまで待たなくても、第2のデータ転送を行うことができる。例えばストレージ装置A600からストレージ装置B600の記憶ボリューム310に記憶されたデータの読み出しを複数回行う場合に、ストレージ制御装置A100からストレージ制御装置B100に送信された第1のデータ読み出し要求に応じて読み出されるデータがストレージ制御装置B100からストレージ制御装置A100に送信されてくる前に、あるいは、ストレージ制御装置B100からストレージ制御装置A100へデータが送信されてきた後でも全てのデータブロックの送信が終了する前に、第2のデータ読み出し要求をストレージ制御装置B100に送信してしまうことが可能となる。この場合、第2のデータ読み出し要求を送信する際には、データ転送LSI(114)のレジスタ807には第1のデータ読み出し要求に対応する第1のデータ転送情報が記憶されているが、第1のデータ転送情報をメモリ3(115)に書き出すことにより、第2のデータ読み出し要求に

10

20

30

40

50

対応する第2のデータ転送情報をレジスタ807に読み出して第2のデータ転送を行うことが可能となる。そして、第1のデータ読み出し要求に応じてストレージ装置B600の記憶ポリューム310から読み出されたデータを受信する際には、レジスタ807に記憶されている第2のデータ転送情報をメモリ3(115)に書き出して、メモリ3(115)に書き出されていた第1のデータ転送情報を再びレジスタ807に読み出して、データを受信すればよい。

【0107】

このようにすれば、例えばストレージ制御装置A100とストレージ制御装置B100との間でデータ転送を行う場合の転送待ち時間に、他のデータ転送を行うことができる。また、例えばストレージ制御装置B100のキャッシュメモリ130に読み出しデータが記憶されていない場合(キャッシュミスヒットの場合)であっても、記憶ポリュームB310からデータを読みだしている間に、他のデータ転送を行うことも可能となる。これによりデータ転送のスループットを向上させることができるので、データ入出力性能を向上させることができる。

【0108】

上述したように、メモリ3(115)には複数のデータ転送情報やデータを書き出すことができるので、もちろん、第2のデータ読み出し要求に応じて読み出されるデータがストレージ制御装置A100に送信されてくるまでの間に、第3のデータ読み出し要求をストレージ制御装置B100に送信し、さらに多重度を上げたデータ入出力を行うことも可能である。

【0109】

このように本実施の形態においては、通信可能に接続されたストレージ装置B600に記憶されたデータをアクセスする場合においても、通信によるデータ転送遅れ時間を隠蔽し、性能を低下させずに、データの入出力を行うことが可能となる。

【0110】

また、ストレージ装置A600とストレージ装置B600との間の通信がファイバチャネルプロトコルに従って行われる場合など、データ転送がデータブロックを単位として行われる場合には、第1のデータ転送によるデータブロックや第2のデータ転送によるデータブロックが混在してデータ転送が行われることになる。複数のストレージ装置A600やストレージ装置B600を備えたストレージシステムにおいては、さらに多くのデータブロックが混在することになる。本実施の形態に係るストレージ制御装置においては、かかる場合においても、第1のデータ転送のデータブロックの送受信中に第2のデータ転送のデータブロックの送受信を行う場合には、第1のデータ転送に係るデータ転送情報やデータブロックをメモリ3(115)に書き出すことにより、第2のデータ転送を行うことができる。そして第2のデータ転送の終了後に第1のデータ転送に係るデータ転送情報やデータブロックをメモリ3(115)から読みだして、第1のデータ転送を再開することもできる。

【0111】

また図7に示したように、本実施の形態に係るDMA800には、DMA制御部801、転送データバッファ810、及びレジスタ807がそれぞれ2つずつ設けられている。そのため、以下に述べるように、本実施の形態に係るストレージ制御装置100においては、より好ましい態様のデータ転送を行うことが可能である。

【0112】

すなわち、図23に示すように、第1のデータ転送のデータブロックの送受信中に第2のデータ転送のデータブロックの送受信を行う場合には、一方のDMA800のレジスタ807に記憶されている第1のデータ転送情報と転送データバッファ810に記憶されているデータ転送中のデータブロックのメモリ3(115)への書き出しが完了する前に、第2のデータ転送情報と第2のデータ転送情報に対応するデータブロックを、他方のDMA800のレジスタ807と転送データバッファ810に読み込むようにする。このようにすることにより、レジスタ807に記憶されている第1のデータ転送情報と転送データバ

10

20

30

40

50

ッファ 810 に記憶されているデータ転送中のデータブロックのメモリ 3 (115) への書き出しが完了するまで待たなくても、第 2 のデータ転送情報と第 2 のデータ転送情報に対応するデータブロックを、他方の DMA 800 のレジスタ 807 と転送データバッファ 810 に読み込み、第 2 のデータ転送をより早く開始することが可能となる。このため、ストレージ装置 A 600 とストレージ装置 B 600 との間のデータ入出力処理をさらにより高速に行うことが可能となるのである。

【0113】

その様子を示したのが図 24 である。DMA 制御部 801、転送データバッファ 810、及びレジスタ 807 が 1 つずつしか設けられていない構成の DMA 800 によりデータ転送を行う場合に第 1 のデータ転送から第 2 のデータ転送に切り替える場合の様子を示した図 22 と、DMA 制御部 801、転送データバッファ 810、及びレジスタ 807 が 2 つずつ設けられている構成の本実施の形態に係る DMA 800 を用いて第 1 のデータ転送から第 2 のデータ転送に切り替える場合の様子を示した図 24 とを比較すると分かるように、DMA 制御部 801、転送データバッファ 810、及びレジスタ 807 がそれぞれ 2 つずつ設けられている構成の本実施の形態に係るデータ転送 LSI (114) によれば、DMA 転送の切り替えオーバーヘッドを隠蔽することができるのである。そのためストレージ装置 A 600 とストレージ装置 B 600 との間のデータ入出力処理をより高速に行うことが可能となるのである。

【0114】

<<< 第 2 の実施の形態 >>>

次に、第 2 の実施の形態について説明する。第 2 の実施の形態においては、チャンネル制御部 110 における通信インタフェース機能はプロトコル LSI (Large-Scale Integrated circuit) 112 により実現される。

【0115】

第 2 の実施の形態に係るチャンネル制御部 110 の構成を図 30 に示す。

チャンネル制御部 110 は回路基板 118 を備えた一つのユニット化されたボードとして構成される。チャンネル制御部 110 は一枚もしくは複数枚の回路基板 118 を含んで構成される。回路基板 118 には、プロセッサ 1 (119)、プロトコル LSI (112)、データ転送 LSI (114)、メモリ 1 (117)、メモリコントローラ 1 (111)、メモリ 3 (115) 及びコネクタ 116 が形成されている。

【0116】

プロセッサ 1 (119)、プロトコル LSI (112)、データ転送 LSI (114)、メモリコントローラ 1 (111)、メモリ 1 (117)、及びメモリ 3 (115) はそれぞれ PCI (Peripheral Component Interconnect) バスで通信可能に接続されている。なお、必ずしも全てが PCI バスで接続されている必要はなく、一部又は全ての接続が他の規格に準拠する構成とすることもできる。このうち、メモリ 1 (117)、メモリ 3 (115)、メモリコントローラ 1 (111) により第 2 のメモリが形成される。またデータ転送 LSI (114)、及びプロトコル LSI (112) によりデータ転送制御部が形成される。またプロセッサ 1 (119) により入出力制御部が形成される。

【0117】

プロトコル LSI (112) は、情報処理装置 200 や他のストレージ装置 600 との間で通信を行うための通信インタフェース機能を提供する。例えば、他のストレージ装置 600 に送信されるデータをファイバチャンネルプロトコルに従ったデータ形式に変換する。プロトコル LSI (112) と接続されるコネクタ 116 は情報処理装置 200 や他のストレージ装置 600 と通信を行うためのコネクタである。プロトコル LSI (112) は、上記通信インタフェース機能を実現するための制御を行う。また、情報処理装置 200 や他のストレージ装置 600 との間でのデータの授受は、プロトコル LSI (112) が備えるバッファメモリを介して行われる。

【0118】

プロトコル LSI (112)、及びコネクタ 116 は、SAN 500 を構成するファイ

10

20

30

40

50

バチャンネルスイッチのポートに接続されるポートとしても機能する。なお、ファイバチャンネルの技術分野においては周知であるが、ポートには、SAN 500上の機器を特定する識別子であるWWNが付与されている。

【0119】

プロセッサ1(119)は、メモリ1(117)に記憶される制御プログラム700を実行することにより、チャンネル制御部110全体の制御を司る。例えば情報処理装置200から送信されたデータ入出力要求を解析して、そのデータ入出力要求がストレージ駆動装置A300に対するものなのか、ストレージ駆動装置B300に対するものなのか、ストレージ駆動装置A300に対するものである場合にはリモートコピーを行う必要があるのか、ないのか等を判断し、それに応じてデータ転送LSI(114)にデータ転送の指示を行う。

10

【0120】

データ転送LSI(114)は、プロセッサ1(119)からの指示により、PCIバスのターゲットとして、PCIバスからの動作要求により、キャッシュメモリ130と情報処理装置200又はストレージ装置B600との間のデータ転送を制御する。具体的には、データ転送LSI(114)は、プロセッサ1(119)からの指示により、プロセッサ1(119)がメモリ1(117)に書き込んだ、キャッシュメモリ130におけるデータの記憶位置と情報処理装置200、ストレージ制御装置B100、又はストレージ装置B600におけるデータの記憶位置とを含むデータ転送情報をメモリ1(117)から読み出して、PCIバスからのアクセスによりデータ転送を行う。

20

【0121】

=== データ転送 ===

次に、第2の実施の形態に係るチャンネル制御部110におけるキャッシュメモリ130と情報処理装置200又はストレージ制御装置B100との間のデータ転送について説明する。

【0122】

まず、プロトコルLSI(112)からみたPCIバス空間を図31に示す。プロトコルLSI(112)のPCIバス空間には、データ空間と制御空間とが形成されている。制御空間は、物理的には図9に示したメモリ1(117)のスク립ト領域に形成されている。

30

PCIバスのデータ空間には、キャッシュメモリ130に転送されるデータ、あるいはキャッシュメモリ130から転送されたデータが記憶される。PCIバスの制御空間には、データ空間に記憶されるこれらの各データの記憶位置や、各データの記憶領域のサイズ(エリア長)等が、各データに対応してプロセッサ2用スク립ト(データ転送情報)として記憶される。プロセッサ2用スク립トは、第1の実施の形態におけるプロセッサ2用スク립トと同様である。これらのプロセッサ2用スク립トは、プロセッサ1(119)により設定され、プロトコルLSI(112)のPCIバス空間の制御空間、すなわちメモリ1(117)のスク립ト領域に書き込まれる。

【0123】

第2の実施の形態におけるキャッシュメモリ130のデータ領域には、第1の実施の形態と同様に、図10に示すように、PCIバス空間のデータ空間に転送されるデータ、あるいはPCIバス空間のデータ空間から転送されたデータが記憶される。キャッシュメモリ130のデータ領域に記憶されるこれらの各データの記憶位置や、各データの記憶領域のサイズ(エリア長)等は、図9に示すように、各データに対応してプロセッサ1用スク립ト(データ転送情報)としてメモリ1(117)のスク립ト領域に記憶される。プロセッサ1用スク립トは、第1の実施の形態におけるプロセッサ1用スク립トと同様である。メモリ1(117)に記憶されるこれらのプロセッサ1用スク립トは、プロセッサ1(119)により設定され、メモリ1(117)のスク립ト領域に書き込まれる。

40

【0124】

50

=== データ転送の流れ ===

次に、本実施の形態に係るデータ転送の処理の流れについて、ストレージ装置 A 6 0 0 が情報処理装置 2 0 0 から、ストレージ装置 B 6 0 0 に記憶されるデータの読み出し要求を受信した場合の処理を例に、図 3 2 のフローチャートに従って説明する。

【 0 1 2 5 】

まずストレージ制御装置 A 1 0 0 が、情報処理装置 2 0 0 からデータ読み出し要求を受信すると、プロセッサ 1 (1 1 9) はメモリ 1 (1 1 7) とキャッシュメモリ 1 3 0 とにデータ領域を確保する (S9000)。ここでメモリ 1 (1 1 7) にデータ領域を確保するとは、プロセッサ 1 用スクリプト及びプロセッサ 2 用スクリプトを記憶するための記憶領域をメモリ 1 (1 1 7) に確保するということである。また、キャッシュメモリ 1 3 0 にデータ領域を確保するとは、読み出しデータを記憶するための記憶領域をキャッシュメモリ 1 3 0 に確保するということである。ここで確保されるデータ領域は複数の場合もある。

【 0 1 2 6 】

続いてプロセッサ 1 (1 1 9) は、プロセッサ 1 用スクリプトを作成する。プロセッサ 1 用スクリプトには、キャッシュメモリ 1 3 0 のアドレスがデータの送信先アドレスとして記憶されている。そして作成したプロセッサ 1 用スクリプトをメモリ 1 (1 1 7) に記憶する (S9001)。次にプロセッサ 1 (1 1 9) は、プロセッサ 2 用スクリプトを作成して、作成したプロセッサ 2 用スクリプトをメモリ 1 (1 1 7) に記憶する (S9002)。プロセッサ 2 用スクリプトには、P C I アドレスがデータの送信元アドレスとして記憶されている。

【 0 1 2 7 】

そしてプロセッサ 1 (1 1 9) は、プロトコル L S I (1 1 2) に対してプロセッサ 2 用スクリプトの番号を送信することによりデータ転送要求を行う (S9003)。

そうすると、プロトコル L S I (1 1 2) はメモリ 1 (1 1 7) からプロセッサ 2 用スクリプトを読みだして (S9004)、ストレージ装置 B 6 0 0 との間で S A N 5 0 0 を介して通信を行う (S9005)。この通信によりプロトコル L S I (1 1 2) はストレージ装置 B 6 0 0 に対してデータ読み出し要求を送信する。そしてストレージ装置 B 6 0 0 から読み出しデータが送信されてきたら、プロトコル L S I (1 1 2) は、P C I アドレスを送信先として読み出しデータを送信する (S9006)。

【 0 1 2 8 】

一方、プロセッサ 1 (1 1 9) は、データ転送 L S I (1 1 4) に対してプロセッサ 1 用スクリプトの番号を送信する。そうするとデータ転送 L S I (1 1 4) は、プロセッサ 1 用スクリプトの番号に基づいて、メモリ 1 (1 1 7) からプロセッサ 1 用スクリプトを読みだす (S9007)。そしてデータ転送 L S I (1 1 4) は、ターゲットとして、P C I バスの動作に応じて、プロセッサ 1 用スクリプトに従ってキャッシュメモリ 1 3 0 への D M A 転送を開始する (S9008)。D M A 転送処理については、図 2 6 乃至図 2 9 に示すフローチャートに示される処理と同様である。

【 0 1 2 9 】

D M A 転送処理が終了したら、D M A 制御部 8 0 1 は、転送処理の結果が記載された終了ステータスをメモリ 1 (1 1 7) に書き込み (S9009)、プロセッサ 1 (1 1 9) に終了通知を行う (S9010)。この終了通知は、例えば割り込み信号の送信により行うことができる。

そしてプロセッサ 1 (1 1 9) は、メモリ 1 (1 1 7) から終了ステータスを読み出して、その内容に応じた処理を実行する (S9011)。

【 0 1 3 0 】

ここで、第 2 の形態に係る D M A 8 0 0 も、第 1 の実施の形態に係る D M A 8 0 0 と同様に、メモリ 3 (1 1 5) の制御を行うことができる。従ってレジスタ 8 0 7 に記憶されるデータ転送情報、及び転送データバッファ 8 1 0 に記憶されているデータを、メモリ 3 (1 1 5) へ書き出したり、メモリ 3 (1 1 5) から読み込んだりすることができる。

このため、本実施の形態に係るストレージ制御装置 1 0 0 でも、第 1 のデータ転送情報

10

20

30

40

50

に基づく第1のデータ転送を制御している途中でも、第2のデータ転送情報に基づく第2のデータ転送を制御することができる。

【0131】

例えば、S9005の処理、すなわちプロトコルLSI(112)によるストレージ装置B(600)へのアクセスは比較的長時間を要するので、この間に複数のデータ読み出し要求をストレージ装置B600へ送信するようにすることができる。また各データ読み出し要求に応じてストレージ装置B600から送信されてくる複数の読み出しデータを受信するようにすることができる。このとき一つのデータ読み出し要求に対して複数の分割されたデータ(データブロック)を受信するようにすることができる。

【0132】

またストレージ装置A600からストレージ装置B600の記憶ポリウム310に記憶されたデータの読み出しを複数回行う場合には、ストレージ制御装置A100からストレージ制御装置B100に送信された第1のデータ読み出し要求に応じて読み出されるデータがストレージ制御装置B100からストレージ制御装置A100に送信されてくる前に、あるいは、ストレージ制御装置B100からストレージ制御装置A100へデータが送信されてきた後でも全てのデータブロックの送信が終了する前に、第2のデータ読み出し要求をストレージ制御装置B100に送信してしまうことが可能となる。この場合、第2のデータ読み出し要求を送信する際には、データ転送LSI(114)のレジスタ807には第1のデータ読み出し要求に対応する第1のデータ転送情報が記憶されているが、第1のデータ転送情報をメモリ3(115)に書き出すことにより、第2のデータ読み出し要求に対応する第2のデータ転送情報をレジスタ807に読み出して第2のデータ転送を行うことが可能となる。そして、第1のデータ読み出し要求に応じてストレージ装置B600の記憶ポリウム310から読み出されたデータを受信する際には、レジスタ807に記憶されている第2のデータ転送情報をメモリ3(115)に書き出して、メモリ3(115)に書き出されていた第1のデータ転送情報を再びレジスタ807に読み出して、データを受信すればよい。

【0133】

このようにすれば、例えばストレージ制御装置A100とストレージ制御装置B100との間でデータ転送を行う場合の転送待ち時間に、他のデータ転送を行うことができる。また、例えばストレージ制御装置B100のキャッシュメモリ130に読み出しデータが記憶されていない場合(キャッシュミスヒットの場合)であっても、記憶ポリウムB310からデータを読みだしている間に、他のデータ転送を行うことも可能となる。これによりデータ転送のスループットを向上させることができるので、データ入出力性能を向上させることができる。

【0134】

また上述したように、メモリ3(115)には複数のデータ転送情報やデータを書き出すことができるので、もちろん、第2のデータ読み出し要求に応じて読み出されるデータがストレージ制御装置A100に送信されてくるまでの間に、第3のデータ読み出し要求をストレージ制御装置B100に送信し、さらに多重度を上げたデータ入出力を行うことも可能である。

【0135】

このように本実施の形態においては、通信可能に接続されたストレージ装置B600に記憶されたデータをアクセスする場合においても、通信によるデータ転送遅れ時間を隠蔽し、性能を低下させずに、データの入出力を行うことが可能となる。

【0136】

また、ストレージ装置A600とストレージ装置B600との間の通信がファイバチャネルプロトコルに従って行われる場合など、データ転送がデータブロックを単位として行われる場合には、第1のデータ転送によるデータブロックや第2のデータ転送によるデータブロックが混在してデータ転送が行われることになる。複数のストレージ装置A600やストレージ装置B600を備えたストレージシステムにおいては、さらに多くのデータ

10

20

30

40

50

ブロックが混在することになる。本実施の形態に係るストレージ制御装置においては、かかる場合においても、第1のデータ転送のデータブロックの送受信中に第2のデータ転送のデータブロックの送受信を行う場合には、第1のデータ転送に係るデータ転送情報やデータブロックをメモリ3(115)に書き出すことにより、第2のデータ転送を行うことができる。そして第2のデータ転送の終了後に第1のデータ転送に係るデータ転送情報やデータブロックをメモリ3(115)から読みだして、第1のデータ転送を再開することもできる。

【0137】

また、第2の実施の形態に係るDMA800にも、DMA制御部801、転送データバッファ810、及びレジスタ807がそれぞれ2つずつ設けられている。そのため、本実施の形態に係るストレージ制御装置100においても、第1のデータ転送のデータブロックの送受信中に第2のデータ転送のデータブロックの送受信を行う場合には、一方のDMA800のレジスタ807に記憶されている第1のデータ転送情報と転送データバッファ810に記憶されているデータ転送中のデータブロックのメモリ3(115)への書き出しが完了する前に、第2のデータ転送情報と第2のデータ転送情報に対応するデータブロックを、他方のDMA800のレジスタ807と転送データバッファ810に読み込むようにすることができる。このようにすることにより、レジスタ807に記憶されている第1のデータ転送情報と転送データバッファ810に記憶されているデータ転送中のデータブロックのメモリ3(115)への書き出しが完了するまで待たなくても、第2のデータ転送情報と第2のデータ転送情報に対応するデータブロックを、他方のDMA800のレジスタ807と転送データバッファ810に読み込み、第2のデータ転送をより早く開始することが可能となる。このため、ストレージ装置A600とストレージ装置B600との間のデータ入出力処理をさらにより高速に行うことが可能となるのである。

【0138】

以上発明を実施するための最良の形態について説明したが、上記実施の形態は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

【図面の簡単な説明】

【0139】

【図1】本実施の形態に係るストレージシステムの全体構成を示すブロック図である。

【図2】本実施の形態に係る管理端末の構成を示すブロック図である。

【図3】本実施の形態に係るチャンネル制御部の構成を示すブロック図である。

【図4】本実施の形態に係るディスク制御部の構成を示すブロック図である。

【図5】本実施の形態に係る情報処理装置の構成を示すブロック図である。

【図6】本実施の形態に係るデータ転送LSIの構成を示すブロック図である。

【図7】本実施の形態に係るDMAの構成を示すブロック図である。

【図8】本実施の形態に係るメモリ2を示す図である。

【図9】本実施の形態に係るメモリ1を示す図である。

【図10】本実施の形態に係るキャッシュメモリを示す図である。

【図11】本実施の形態に係る共有メモリを示す図である。

【図12】本実施の形態に係るペア管理テーブルを示す図である。

【図13】本実施の形態に係る構成情報管理テーブルを示す図である。

【図14】本実施の形態に係るデータ入出力要求を示す図である。

【図15】本実施の形態に係る記憶ボリュームの管理を示すブロック図である。

【図16】他の実施形態に係るリードコマンドの処理を示すフローチャートである。

【図17】本実施の形態に係るリードコマンドの処理を示すフローチャートである。

【図18】本実施の形態に係るリードコマンドの処理を示すフローチャートである。

【図19】本実施の形態に係るデータ転送の様子を示すブロック図である。

【図20】本実施の形態に係るメモリ3を示す図である。

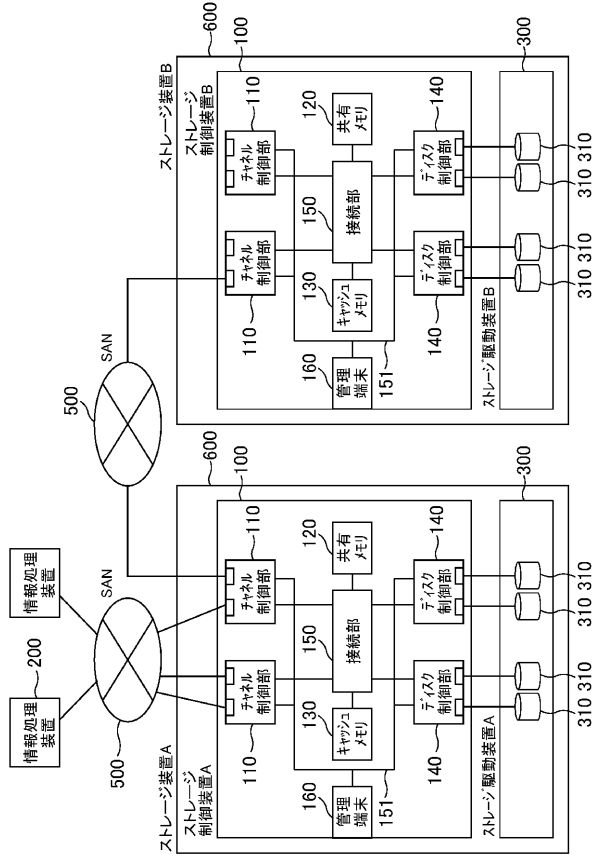
- 【図 2 1】本実施の形態に係るデータ転送の概要を示すブロック図である。
 【図 2 2】本実施の形態に係るデータ転送の概要を示すフローチャートである。
 【図 2 3】本実施の形態に係るデータ転送の概要を示すブロック図である。
 【図 2 4】本実施の形態に係るデータ転送の概要を示すフローチャートである。
 【図 2 5】本実施の形態に係るデータ転送の処理の流れを示すフローチャートである。
 【図 2 6】本実施の形態に係るデータ転送の処理の流れを示すフローチャートである。
 【図 2 7】本実施の形態に係るデータ転送の処理の流れを示すフローチャートである。
 【図 2 8】本実施の形態に係るデータ転送の処理の流れを示すフローチャートである。
 【図 2 9】本実施の形態に係るデータ転送の処理の流れを示すフローチャートである。
 【図 3 0】第 2 の実施の形態に係るチャンネル制御部の構成を示すブロック図である。 10
 【図 3 1】第 2 の実施の形態に係るプロトコル L S I のメモリ空間を示す図である。
 【図 3 2】第 2 の実施の形態に係るデータ転送の処理の流れを示すフローチャートである

。【符号の説明】

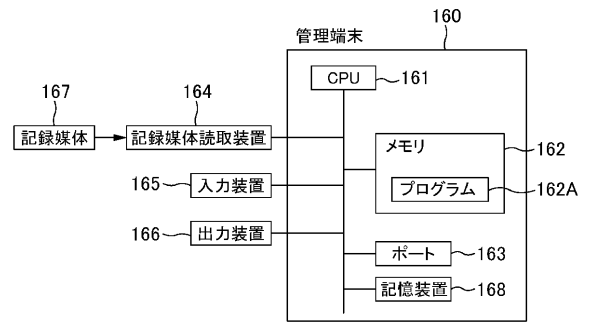
【 0 1 4 0 】

1 0 0	ストレージ制御装置	1 1 0	チャンネル制御部	
1 1 2	プロセッサ 2	1 1 3	メモリ 2	
1 1 4	データ転送 L S I	1 1 5	メモリ 3	
1 1 7	メモリ 1	1 1 8	回路基板	
1 1 9	プロセッサ 1	1 2 0	共有メモリ	20
1 3 0	キャッシュメモリ	1 4 0	ディスク制御部	
1 5 0	接続部	1 6 0	管理端末	
2 0 0	情報処理装置	3 0 0	ストレージ駆動装置	
3 1 0	記憶ポリューム	5 0 0	S A N	
6 0 0	ストレージ装置			
7 1 0	ペア管理テーブル	7 2 0	構成情報管理テーブル	
8 0 0	D M A	8 0 1	D M A 制御部	
8 0 2	P C I インタフェース	8 0 3	転送起動レジスタ	
8 0 4	転送レジスタ	8 0 5	初期設定レジスタ	
8 0 6	スクリプトレジスタ	8 0 7	レジスタ	30
8 1 0	転送データバッファ			

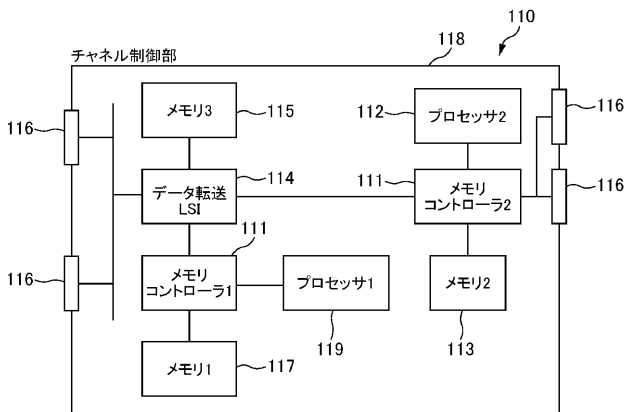
【 図 1 】



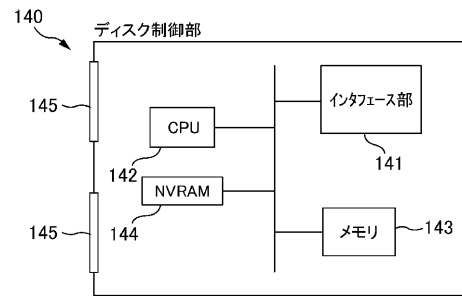
【 図 2 】



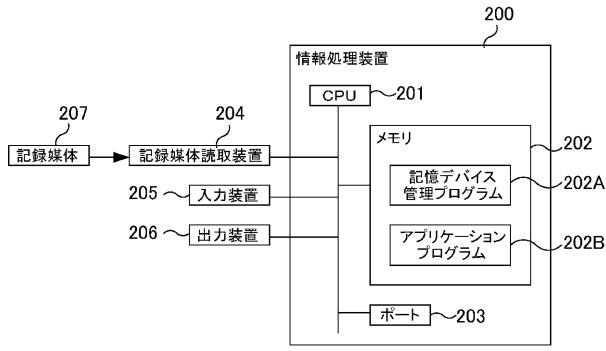
【 図 3 】



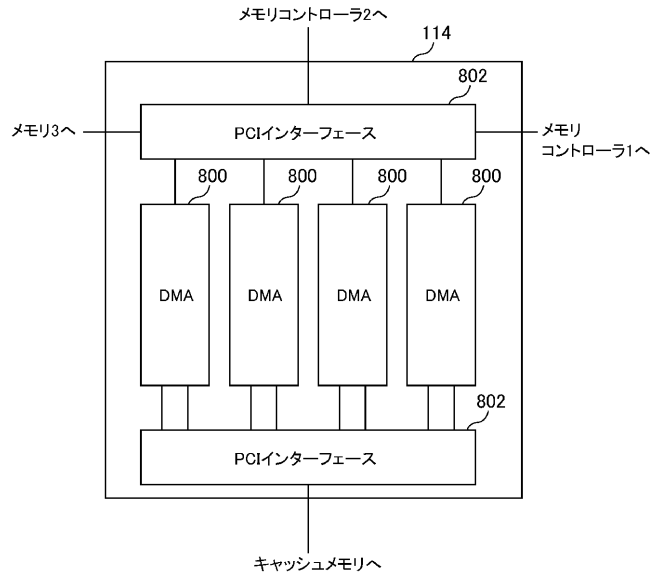
【 図 4 】



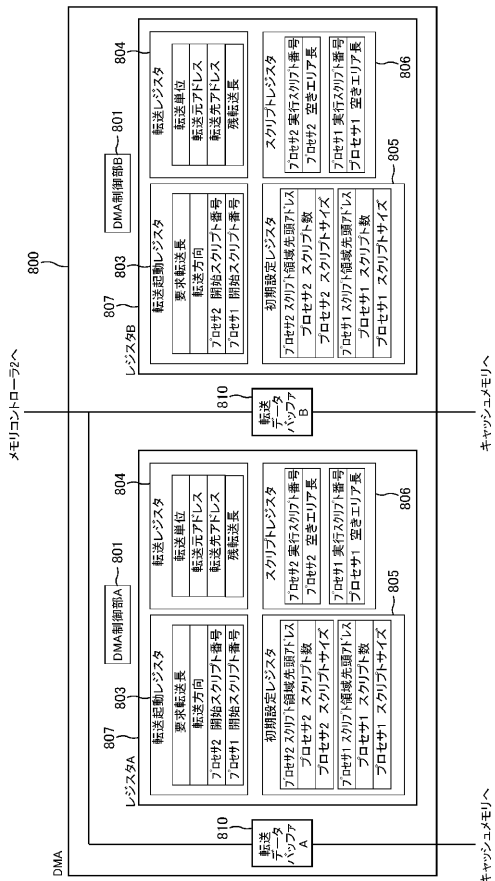
【 図 5 】



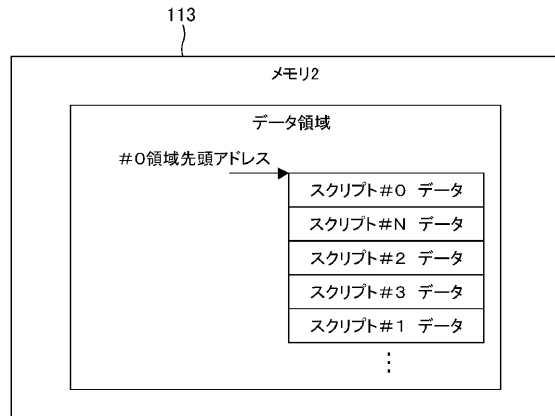
【 図 6 】



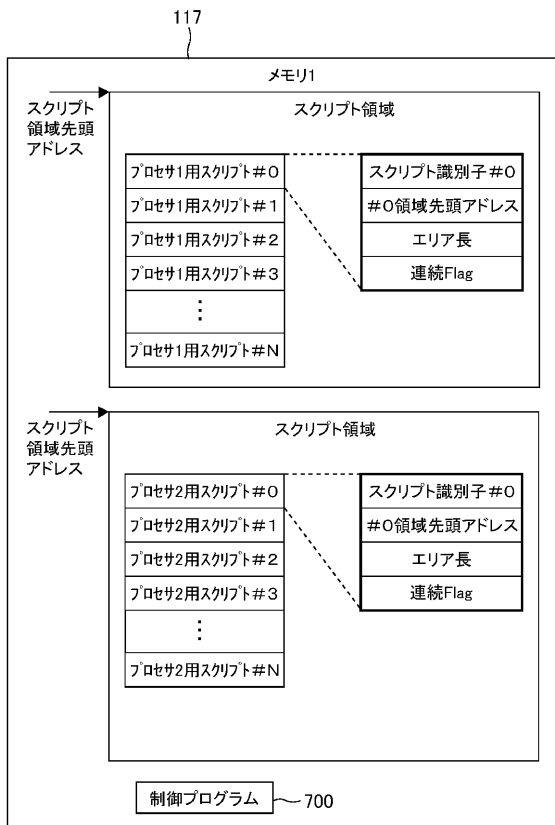
【 図 7 】



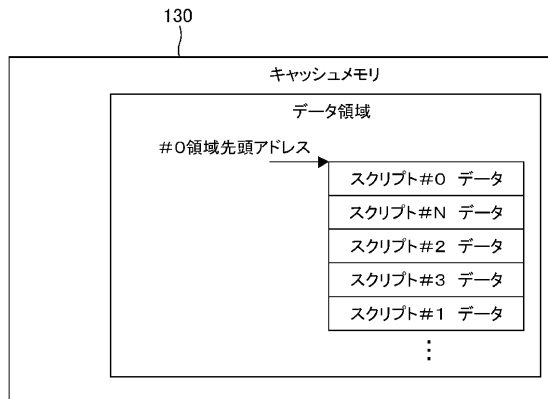
【 図 8 】



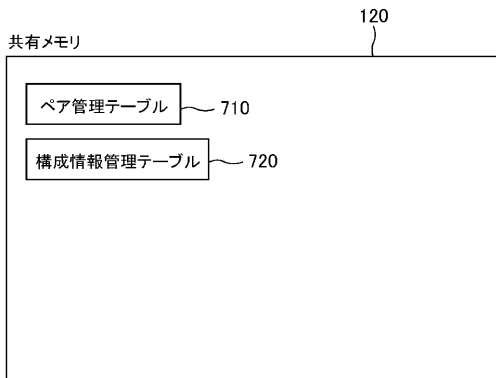
【 図 9 】



【 図 10 】



【 図 11 】



【 図 13 】

720

Port ID	WWN	LUN	Capacity (KB)	マッピングLUN
010001	XXZZYYXX00002141	3	3,072	01:03
010006	ad00bfffz00a1ffd	5	5,120	3B:20
⋮	⋮	⋮	⋮	⋮
032F31	XXZZYYXX00002142	2	45,897	02:01

【 図 12 】

710

ピア管理テーブル

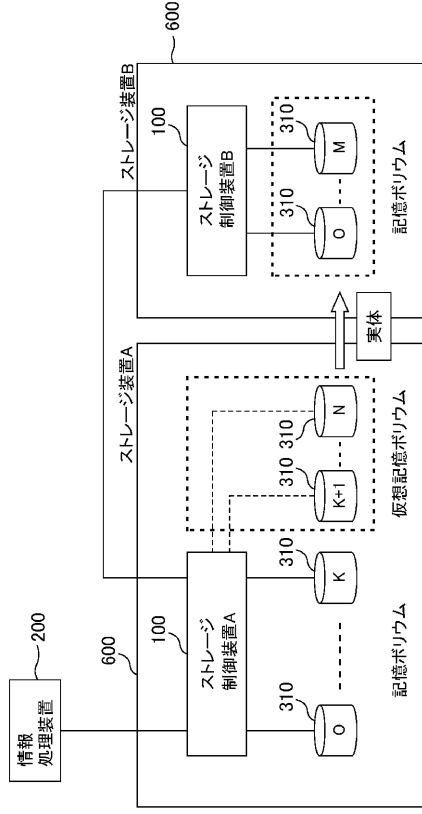
ペア種類	複製方式	複製元装置	複製先装置	複製元ボリューム	複製先ボリューム	ペア状態
ローカル	—	—	—	0	2	ペア中
ローカル	—	—	—	1	3	ペア中
リモート	非同期	A	B	0	0	スプリット中
⋮	⋮	⋮	⋮	⋮	⋮	⋮

【 図 14 】

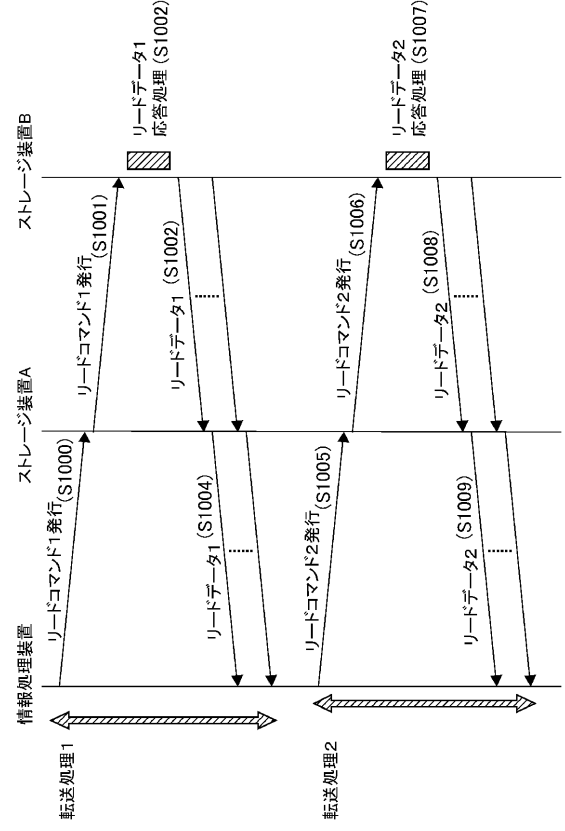
730

ホストID	ポートID	LUN	アドレス	データ長
-------	-------	-----	------	------

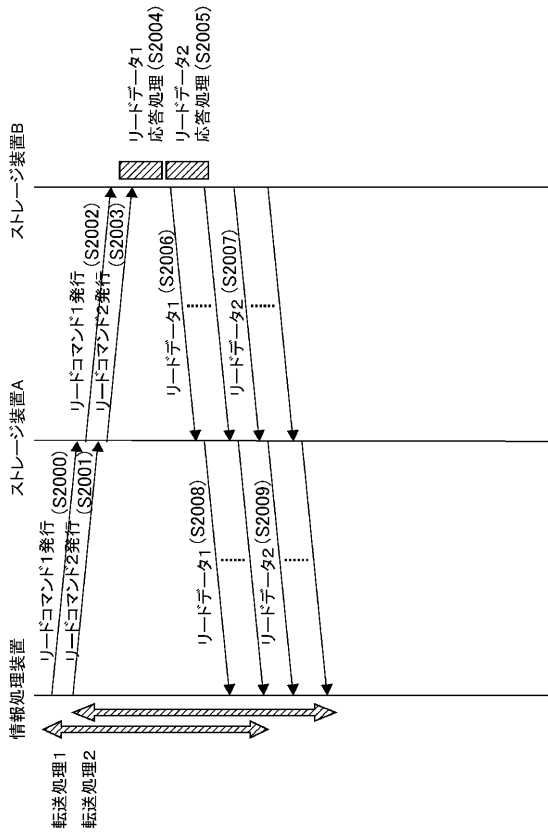
【 図 1 5 】



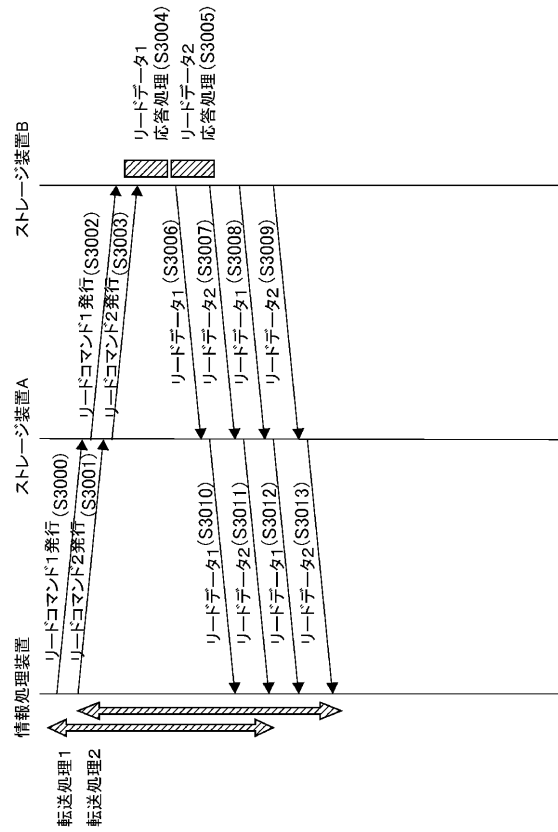
【 図 1 6 】



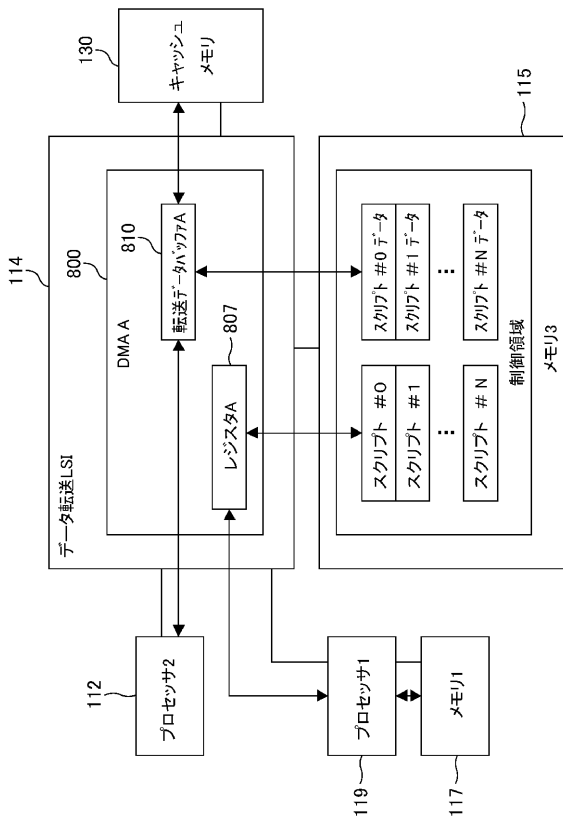
【 図 1 7 】



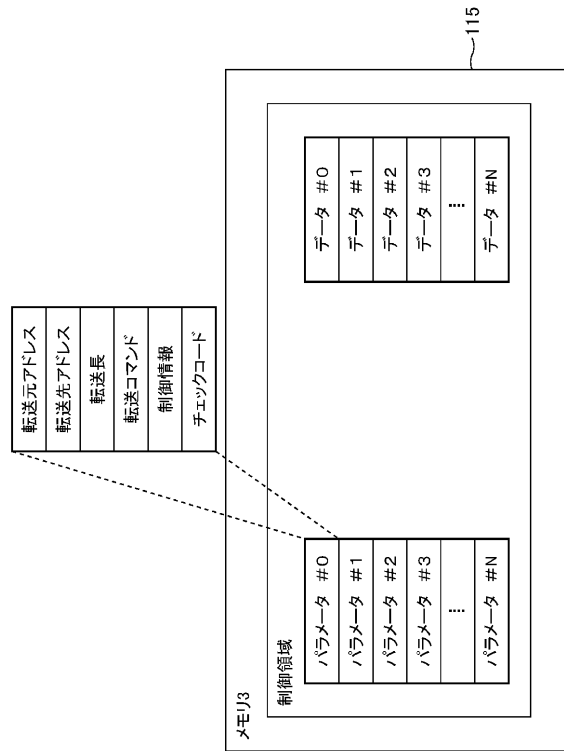
【 図 1 8 】



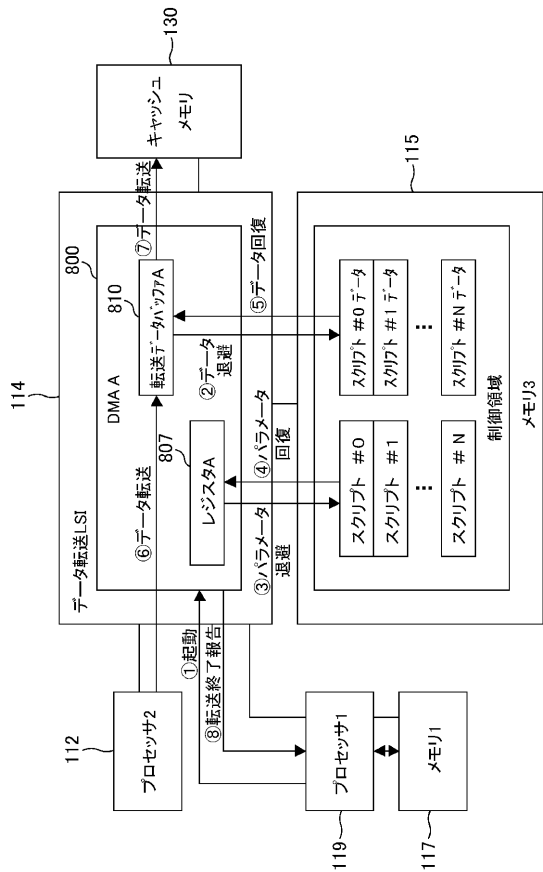
【図 19】



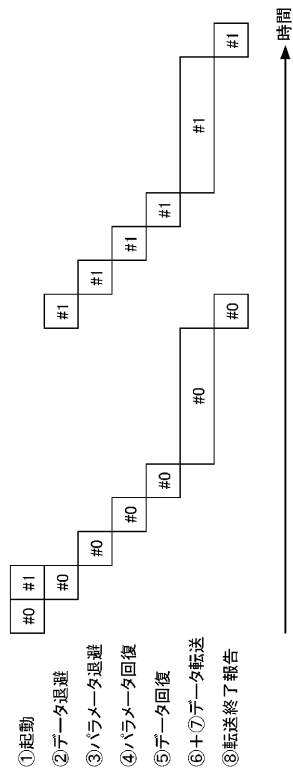
【図 20】



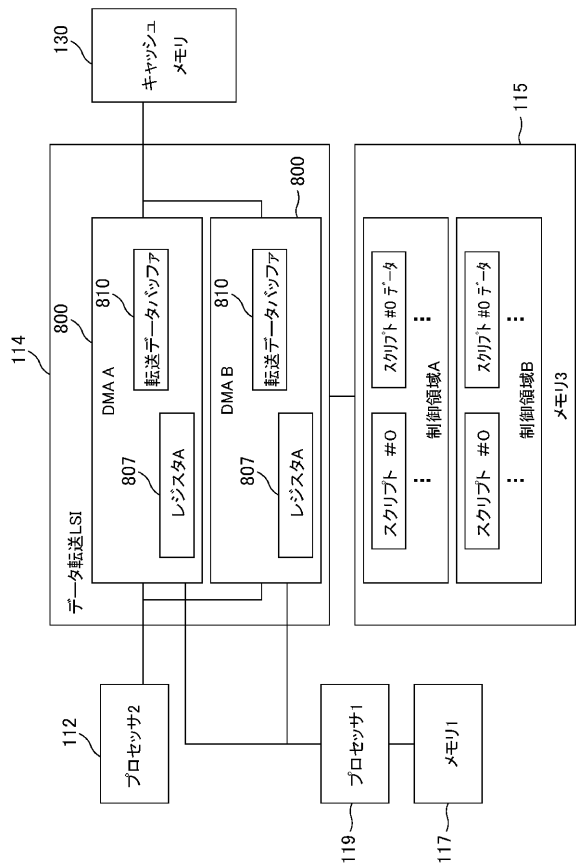
【図 21】



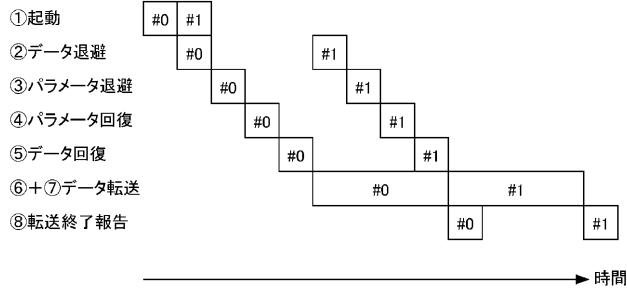
【図 22】



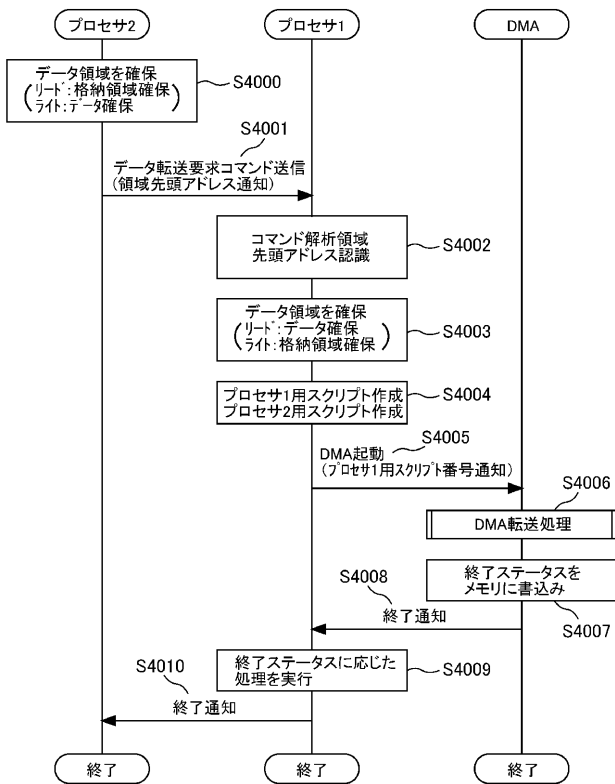
【 図 2 3 】



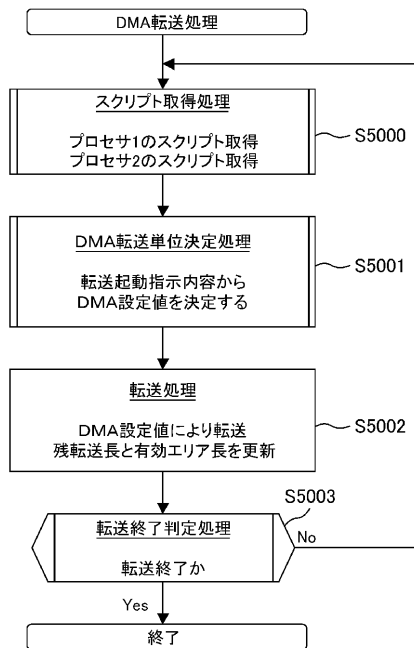
【 図 2 4 】



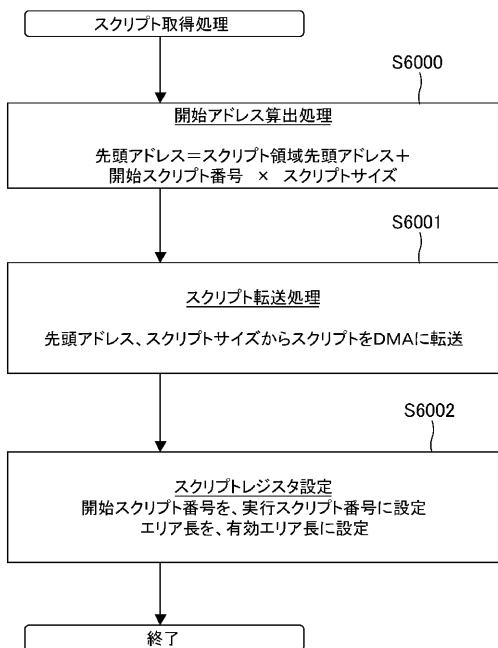
【 図 2 5 】



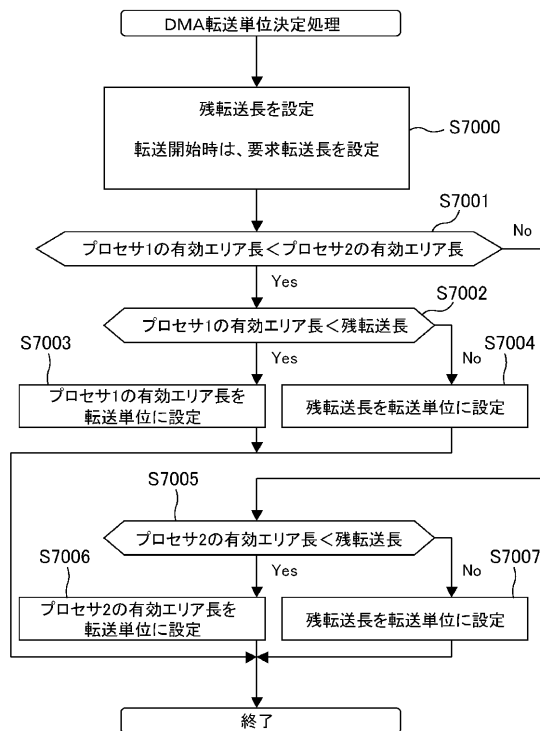
【 図 2 6 】



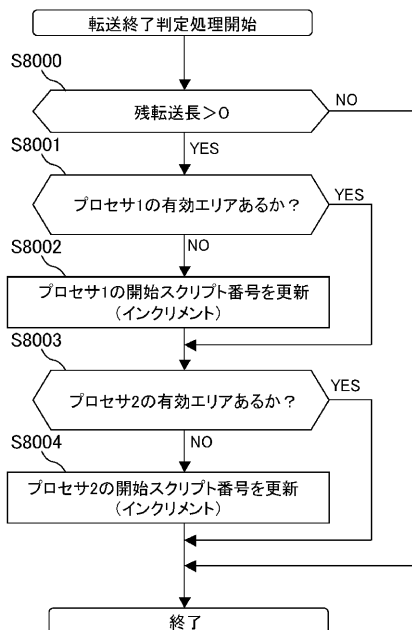
【 図 2 7 】



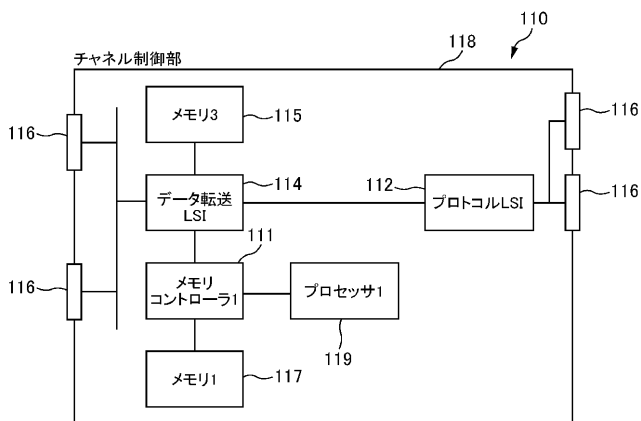
【 図 2 8 】



【 図 2 9 】

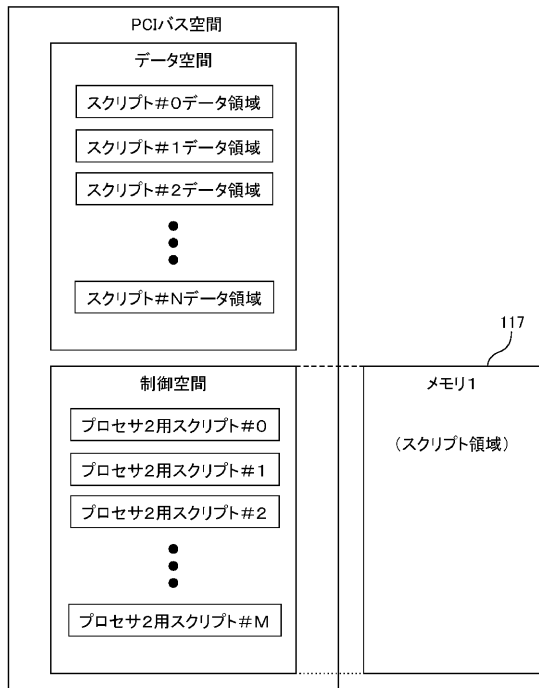


【 図 3 0 】

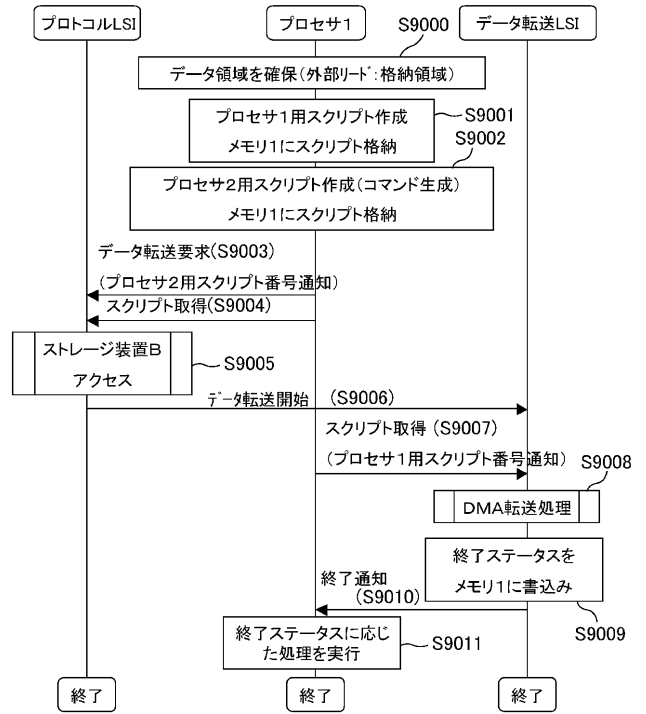


【 図 3 1 】

プロトコルLSIメモリ空間(プロトコルLSIのPCIバス空間)



【 図 3 2 】



フロントページの続き

(51) Int.Cl. ⁷	F I	テーマコード(参考)
	G 0 6 F 12/08	5 0 1 E
	G 0 6 F 12/08	5 4 1 Z
	G 0 6 F 12/08	5 5 7
	G 0 6 F 13/10	3 4 0 A

F ターム(参考) 5B005 JJ01 KK02 MM11 NN12 WW13
5B014 EB04
5B065 BA01 CA11 CA13 CH01
5B082 DE04 HA08