

(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2009-266119

(P2009-266119A)

(43) 公開日 平成21年11月12日(2009.11.12)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 3/06 (2006.01)	G06F 3/06 302Z	5B005
G06F 12/08 (2006.01)	G06F 3/06 301M	5B065
	G06F 12/08 557	
	G06F 12/08 541Z	

審査請求 未請求 請求項の数 14 O L (全 26 頁)

(21) 出願番号 特願2008-117634 (P2008-117634)
 (22) 出願日 平成20年4月28日 (2008. 4. 28)

(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 100093861
 弁理士 大賀 真司
 (72) 発明者 平山 洋志
 神奈川県川崎市麻生区王禅寺1099番地
 株式会社日立製作所システム開発研究所
 内
 Fターム(参考) 5B005 JJ12 MM11
 5B065 BA01 CC08 CE26 CH15

(54) 【発明の名称】 ストレージ装置及びデータ転送方法

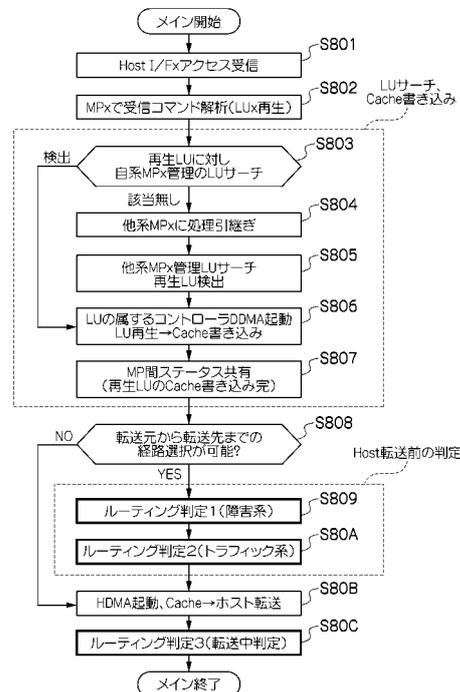
(57) 【要約】 (修正有)

【課題】 アクセス性能を向上させ得るストレージ装置を提供する。

【解決手段】 上位装置及び記憶デバイスと接続され、上位装置及び記憶デバイスとのデータの転送を制御するコントローラの複数と、コントローラ同士を接続するデータの転送路と、コントローラと接続され、コントローラを制御する複数のプロセッサとを備え、コントローラは、それぞれ同一又は異なるプロセッサに接続され、第1のプロセッサは、当該転送命令に基づいて、データの転送元となる第1のコントローラ及び当該データの転送先となる第2のコントローラを決定し、第1のコントローラ及び第2のコントローラ間の、転送路に属する経路を判定し、経路が複数あるときに、第2のプロセッサが当該第2のコントローラ以外に接続されているコントローラの数最大となる経路を特定経路として選択し、経路に沿って、第1のコントローラ及び第2のコントローラ間で前記データを転送する。

【選択図】 図14

図14



【特許請求の範囲】**【請求項 1】**

上位装置及び記憶デバイスと接続され、前記上位装置及び前記記憶デバイスとのデータの転送を制御するコントローラの複数と、

前記コントローラ同士を接続するデータの転送路と、

前記コントローラと接続され、前記コントローラを制御する複数のプロセッサとを備え、

前記複数のコントローラは、

それぞれ同一又は異なるプロセッサに接続され、

前記上位装置から転送命令を受信したコントローラに接続されている第 1 のプロセッサは、

当該転送命令に基づいて、データの転送元となる第 1 のコントローラ及び当該データの転送先となる第 2 のコントローラを決定し、

前記第 1 のコントローラ及び第 2 のコントローラ間の、前記転送路に属する経路を判定し、

前記経路が複数あるときに、前記第 2 のコントローラに接続されている第 2 のプロセッサが当該第 2 のコントローラ以外に接続されているコントローラの数最大となる経路を特定経路として選択し、

前記複数のプロセッサは、

前記特定経路に沿って、前記第 1 のコントローラ及び第 2 のコントローラ間で前記データを転送する

ことを特徴とするストレージ装置。

【請求項 2】

前記複数のコントローラは、

2 の倍数であり、かつ、4 つ以上の数のコントローラであり、

前記複数のプロセッサは、

2 の倍数であり、かつ、2 つ以上の数のプロセッサであると共に、前記コントローラの数よりも少ない数のプロセッサであり、1 つの前記プロセッサに対して、前記コントローラを前記プロセッサの数で割った数のコントローラが接続されている

ことを特徴とする請求項 1 に記載のストレージ装置。

【請求項 3】

前記コントローラは、

前記転送路を介するデータ転送において、転送タイムアウト又はデータ転送の際に付加されるパリティに基づいて、前記転送路の通信エラーの障害情報を検出し、

前記第 1 のプロセッサは、

前記通信エラーの障害情報が検出されたときに、前記通信エラーの障害情報が検出されていない別の前記転送路を含む経路を特定経路として選択する

ことを特徴とする請求項 1 に記載のストレージ装置。

【請求項 4】

前記コントローラは、

前記転送路を介するデータ転送の転送パケットに対して、無効なデータが含まれるパケット転送を検出することにより、前記接続路の帯域の浪費を検出し、

前記第 1 のプロセッサは、

前記転送路の帯域の浪費が最小の前記転送路を含む経路を特定経路として選択する

ことを特徴とする請求項 1 に記載のストレージ装置。

【請求項 5】

前記コントローラは、

転送先となる前記第 2 のコントローラが同一であり、かつ、複数の接続先から転送される転送パケットが時分割多重されているときに、前記転送路に出力される際の転送パケットの多重度を検出し、

10

20

30

40

50

前記第 1 のプロセッサは、

前記転送路に対する多重度が最小の前記転送路を含む経路を特定経路として選択することを特徴とする請求項 1 に記載のストレージ装置。

【請求項 6】

前記コントローラは、

転送先となる前記第 2 のコントローラが同一であり、かつ、複数の接続先から転送される転送パケットが時分割多重されているときに、前記転送路に出力される際の転送パケットの多重度を検出し、

前記第 1 のプロセッサは、

前記転送路の帯域の浪費に差異がないときに、前記転送路に対する多重度が最小の前記転送路を含む経路を特定経路として選択する

10

ことを特徴とする請求項 4 に記載のストレージ装置。

【請求項 7】

前記コントローラは、

前記転送路を介するデータ転送において、転送タイムアウト又はデータ転送の際に付加されるパリティに基づいて、前記転送路の通信エラーの障害情報を検出し、

前記第 1 のプロセッサは、

前記特定経路を選択する際、前記通信エラーの障害情報が検出されたときに、前記通信エラーの障害情報が検出されていない別の前記転送路を含む経路を特定経路として最優先に選択する

20

ことを特徴とする請求項 6 に記載のストレージ装置。

【請求項 8】

前記コントローラは、

前記特定経路に沿って、前記データの転送を開始した後に、前記転送路の通信エラーの障害情報を検出する

ことを特徴とする請求項 3 に記載のストレージ装置。

【請求項 9】

前記コントローラは、

前記通信エラーの障害情報が検出されなかったときに、前記転送路を介するデータ転送の転送パケットに対して、無効なデータが含まれるパケット転送を検出することにより、前記転送路の帯域の浪費を検出し、

30

前記第 1 のプロセッサは、

前記転送路の帯域の浪費が最小の前記転送路を含む経路を特定経路として改めて選択し

、
前記データの転送を開始した前記特定経路と、改めて選択した特定経路とが不一致のときに、前記データの転送を停止し、前記改めて選択した特定経路を選択し、

前記複数のプロセッサは、

前記改めて選択した特定経路に沿って、前記データの転送を再開する

ことを特徴とする請求項 8 に記載のストレージ装置。

【請求項 10】

40

前記コントローラは、

前記転送路の帯域の浪費に差異がないときに、転送先となる前記第 2 のコントローラが同一であり、かつ、複数の接続先から転送される転送パケットが時分割多重されているときに、前記転送路に出力される際の転送パケットの多重度を検出し、

前記第 1 のプロセッサは、

前記転送路の帯域の浪費に差異がないときに、前記転送路に対する多重度が最小の前記転送路を含む経路を特定経路として改めて選択し、

前記データの転送を開始した前記特定経路と、改めて選択した特定経路とが不一致のときに、前記データの転送を停止し、前記改めて選択した特定経路を選択し、

前記複数のプロセッサは、

50

前記改めて選択した特定経路を介して前記データの転送を再開することを特徴とする請求項 9 に記載のストレージ装置。

【請求項 11】

前記第 1 のプロセッサは、

前記特定経路を改めて選択する際、前記通信エラーの障害情報が検出されたときに、前記通信エラーの障害情報が検出されていない別の前記転送路を含む経路を特定経路として最優先に選択し、

前記通信エラーの障害情報が検出されなかったときに、前記コントローラにより検出された前記転送路の帯域の浪費が最小の前記転送路を含む経路を特定経路として改めて選択し、

前記転送路の帯域の浪費に差異がないときに、前記転送路に対する多重度が前記転送路を含む経路を特定経路として改めて選択する

ことを特徴とする請求項 10 に記載のストレージ装置。

【請求項 12】

上位装置及び記憶デバイスと接続され、前記上位装置及び前記記憶デバイスとのデータの転送を制御するコントローラの複数と、

前記コントローラ同士を接続するデータの転送路と、

前記コントローラと接続され、前記コントローラを制御する複数のプロセッサとを備え、

前記複数のコントローラは、

それぞれ同一又は異なるプロセッサに接続され、

前記上位装置から転送命令を受信したコントローラに接続されている第 1 のプロセッサは、

転送元となる第 1 のコントローラのパケットのヘッダー情報として、当該第 1 のコントローラ及び転送先となる第 2 のコントローラの識別情報と、前記第 1 のコントローラ及び第 2 のコントローラ間の前記転送路の識別情報と、前記データを転送する際に経由するコントローラの識別情報とを生成する

ことを特徴とするストレージ装置。

【請求項 13】

前記第 1 のプロセッサは、

前記データ転送命令に含まれるパケットのヘッダー情報を受信元において保持し、

前記第 1 のコントローラのパケットのヘッダー情報に従って、前記第 1 のコントローラのパケットを前記第 2 のコントローラにおいて受信すると、前記第 1 のコントローラのパケットのヘッダー情報を、前記受信元において保持した前記転送命令に含まれるパケットのヘッダー情報に変更した後、変更したパケットを該当する接続先に送信する

ことを特徴とする請求項 12 に記載のストレージ装置。

【請求項 14】

上位装置及び記憶デバイスと接続され、前記上位装置及び前記記憶デバイスとのデータの転送を制御するコントローラの複数と、

前記コントローラ同士を接続するデータの転送路と、

前記コントローラと接続され、前記コントローラを制御する複数のプロセッサとを備え、

前記複数のコントローラは、

それぞれ同一又は異なるプロセッサに接続され、

前記上位装置から転送命令を受信したコントローラに接続されている第 1 のプロセッサが、

当該転送命令に基づいて、データの転送元となる第 1 のコントローラ及び当該データの転送先となる第 2 のコントローラを決定し、

前記第 1 のコントローラ及び第 2 のコントローラ間の、前記転送路に属する経路を判定し、

10

20

30

40

50

前記経路が複数あるときに、前記第2のコントローラに接続されている第2のプロセッサが当該第2のコントローラ以外に接続されているコントローラの数最大となる経路を特定経路として選択し、

前記複数のプロセッサが、

前記特定経路に沿って、前記第1のコントローラ及び第2のコントローラ間で前記データを転送する

ことを特徴とするデータ転送方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ストレージ装置及びデータ転送方法に関し、特に、キャッシュメモリやホスト、HDD(Hard Disc Drive)ドライブとの各I/O(データ転送経路のことを示す。以下同じ。)を備えるコントローラにより構成されるストレージ装置及びデータ転送方法に関する。

【背景技術】

【0002】

従来、ストレージ装置は、例えば、ホストインターフェイス、キャッシュメモリ、HDDが接続されるドライブインターフェイスと、データ転送を制御するコントローラと、これらを接続するI/O、システム全体の制御を行うマイクロプロセッサを一つの単位(モジュール)として構成し、複数のモジュール間を、専用I/Oで接続することで構成される。各モジュールは、電源供給を独立させ、さらに、ホスト転送データを別モジュールのキャッシュメモリへ二重書きを行うなど、データ消失に対する信頼性向上を図っている(以上の内容は特許文献1を参照。)

【0003】

一方、ストレージ装置は、複数のモジュールを接続することで、ホストやHDDドライブを拡張する構成とする場合がある。I/Oによるデータ転送は、転送元から転送先が同一でも経由するモジュール間I/Oの選択が可能な場合がある。従って、ストレージ装置では、各I/Oによる転送状況、障害状況に従い、最適な経路選択(ルーティング)が可能となる。経路選択方法の一例として、ノード間接続、つまりコントローラ間接続の利用状況に応じて重み付けを与え、転送元から転送先の間に含まれるノード間の重み付け総和から、最適経路を選択する方法が特許文献2に開示されている。

【特許文献1】特開2005-44010号公報

【特許文献2】特開平11-313069号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

ストレージ装置は、ホストとのI/OやHDDドライブとのI/Oの数を増すことで、拡張性(スケーラビリティ)に対応する。よって、ストレージ装置では、データ転送を制御するコントローラのホストとのI/O、HDDドライブとのI/Oの数を追加することで対応可能である。しかしながら、ストレージ装置におけるリード及びライトは、複数のI/Oにより同時に転送が行われる。従って、ストレージ装置では、I/O数の増加に比例して、コントローラ内で行われるI/Oの切り替え速度(スイッチ帯域)、キャッシュリード、ライト動作に対する帯域の向上が必要となる。特に、大規模なストレージ装置を構成する場合、必要なI/Oの数に対し、スイッチ帯域、キャッシュ帯域の確保が困難となる。つまり、ストレージ装置では、必要な帯域を備えたコントローラ等のデバイスが実現性困難となる。

【0005】

従って、ストレージ装置では、コントローラ間を接続するモジュール間I/Oの追加で、モジュールを4個、8個と接続し、拡張性に対応する。つまり、ストレージ装置では、マルチコントローラ構成で拡張性に対応する。各コントローラは、I/Oの数に必要なス

10

20

30

40

50

イチ帯域、キャッシュ帯域を備えており、インターフェイスに対する転送レートを確保しながら、ストレージ装置に必要な I / O の数の確保が可能である。

【 0 0 0 6 】

マルチコントローラ構成では、転送先及び転送元の I / O が同一のコントローラに含まれる場合に限定されず、別のコントローラの場合もある。この場合、データ転送は、コントローラ間に備えたモジュール間 I / O を経由して行われる。ストレージ装置では、複数のコントローラを経由してデータ転送が行われる場合に、複数の転送経路が存在する場所がある。一方で、コントローラ間のモジュール間 I / O は、転送障害や、同一 I / O に複数のコントローラからの転送が集中することによる転送レート低下が発生する。ストレージ装置では、そのような I / O を避け、転送に最適な経路選択をすることが望ましい。

10

【 0 0 0 7 】

経路選択の例として、全てのノード間の重み付けを一元管理しながら、重み付けの総和計算、経路選択を行う方法が特許文献 2 に開示されている。ここで、ノードは、ストレージ装置におけるコントローラ、コントローラ間の I / O に相当する。この方法をマルチコントローラ構成のストレージ装置に適用する場合には、次の課題がある。

【 0 0 0 8 】

この場合、インターフェイスに障害発生を考慮した重み付けが必要である。さらに、転送が発生するホストとの I / O や HDD ドライブとの I / O の数の変動により、各ノードの重みは、刻々と変化するため、転送前の経路選択に加え、転送中も経路判定を行うことが必要である。

20

【 0 0 0 9 】

また、実際のストレージ装置では、1つの制御手段（マイクロプロセッサ等）に対するコントローラ数を制限し、制御手段の処理能力不足による転送性能の低下を回避する。転送中に発生する障害管理は、一つのコントローラに集中させることが望ましいが、上記ストレージ装置では不可能である。従って、マルチコントローラ構成のストレージ装置における障害管理を考慮した転送経路選択が必要となる。

【 0 0 1 0 】

さらに、重み付け要素である障害状況と、転送状況、制御手段が管理するコントローラ数により、転送経路を決定する際には、各要素で重み付けを行い、最適な転送経路の選択を行う必要がある。

30

【 0 0 1 1 】

本発明は以上の点を考慮してなされたもので、アクセス性能を向上させ得るストレージ装置及びデータ転送方法を提案しようとするものである。

【課題を解決するための手段】

【 0 0 1 2 】

かかる課題を解決するために本発明においては、ストレージ装置であって、上位装置及び記憶デバイスと接続され、前記上位装置及び前記記憶デバイスとのデータの転送を制御するコントローラの複数と、前記コントローラ同士を接続するデータの転送路と、前記コントローラと接続され、前記コントローラを制御する複数のプロセッサとを備え、前記複数のコントローラは、それぞれ同一又は異なるプロセッサに接続され、前記上位装置から転送命令を受信したコントローラに接続されている第 1 のプロセッサは、当該転送命令に基づいて、データの転送元となる第 1 のコントローラ及び当該データの転送先となる第 2 のコントローラを決定し、前記第 1 のコントローラ及び第 2 のコントローラ間の、前記転送路に属する経路を判定し、前記経路が複数あるときに、前記第 2 のコントローラに接続されている第 2 のプロセッサが当該第 2 のコントローラ以外に接続されているコントローラの数最大となる経路を特定経路として選択し、前記複数のプロセッサは、前記特定経路に沿って、前記第 1 のコントローラ及び第 2 のコントローラ間で前記データを転送する。

40

【 0 0 1 3 】

また、本発明においては、ストレージ装置であって、上位装置及び記憶デバイスと接続

50

され、前記上位装置及び前記記憶デバイスとのデータの転送を制御するコントローラの複数と、前記コントローラ同士を接続するデータの転送路と、前記コントローラと接続され、前記コントローラを制御する複数のプロセッサとを備え、前記複数のコントローラは、それぞれ同一又は異なるプロセッサに接続され、前記上位装置から転送命令を受信したコントローラに接続されている第1のプロセッサは、転送元となる第1のコントローラのパケットのヘッダー情報として、当該第1のコントローラ及び転送先となる第2のコントローラの識別情報と、前記第1のコントローラ及び第2のコントローラ間の前記転送路の識別情報と、前記データを転送する際に経由するコントローラの識別情報とを生成する。

【0014】

さらに、本発明においては、データ転送方法であって、上位装置及び記憶デバイスと接続され、前記上位装置及び前記記憶デバイスとのデータの転送を制御するコントローラの複数と、前記コントローラ同士を接続するデータの転送路と、前記コントローラと接続され、前記コントローラを制御する複数のプロセッサとを備え、前記複数のコントローラは、それぞれ同一又は異なるプロセッサに接続され、前記上位装置から転送命令を受信したコントローラに接続されている第1のプロセッサが、当該転送命令に基づいて、データの転送元となる第1のコントローラ及び当該データの転送先となる第2のコントローラを決定し、前記第1のコントローラ及び第2のコントローラ間の、前記転送路に属する経路を判定し、前記経路が複数あるときに、前記第2のコントローラに接続されている第2のプロセッサが当該第2のコントローラ以外に接続されているコントローラの数最大となる経路を特定経路として選択し、前記複数のプロセッサが、前記特定経路に沿って、前記第1のコントローラ及び第2のコントローラ間で前記データを転送する。

【0015】

従って、マルチコントローラ構成のストレージ装置において、転送元から転送先までのデータ転送に複数の経路選択が可能な場合、その選択を適切に行うことができるため、所望のデータに効率的にアクセスさせることができる。

【発明の効果】

【0016】

本発明によれば、アクセス性能を向上させ得るストレージ装置及びデータ転送方法を実現することができる。

【発明を実施するための最良の形態】

【0017】

以下、本発明の一実施の形態を、図面を参照して詳細に説明する。なお、これにより本発明が限定されるものではない。

【0018】

(1) 実施例1

図1は、本発明のストレージ装置100について、第1の実施例を示す装置ブロック図である。図において、コントローラ1(1a~1d)(CTL0~CTL3)は、キャッシュメモリ4(4a~4d)(Cache0~Cache3)、ホストインターフェイス2(2a~2d)(Host I/F0~Host I/F3)、ドライブインターフェイス3(3a~3d)(Drv I/F0~Drv I/F3)、マイクロプロセッサ5(5a、5b)(MP0,MP1)との間のそれぞれのI/O(データ転送経路のことを示す。以下同じ。)を介したデータ転送を制御するコントローラである。コントローラ1同士は、I/O(デュアルI/O(Dual I/O)、クロスI/O(Cross I/O))により接続されている。

【0019】

図1においては、コントローラ1a及びコントローラ1bは、クロス0I/O(Cross 0 I/O)により接続されており、コントローラ1a及びコントローラ1cは、デュアル0I/O(Dual 0 I/O)により接続されており、コントローラ1b及びコントローラ1cは、クロス1I/O(Cross 1 I/O)により接続されており、コントローラ1b及びコントローラ1dは、デュアル1I/O(Dual 1 I/O)により接続されており、ループ状アーキテクチャを構成する。デュアルI/O及びクロスI/Oの利用方法の違いについては図4

及び図 5 にて説明する。

【 0 0 2 0 】

ホストインターフェイス 2 a ~ 2 d は、ホストとの通信プロトコルに従い、パケット単位での転送を制御するホストインターフェイスである。ドライブインターフェイス 3 a ~ 3 d は、ハードディスクドライブとの通信プロトコルに従い、セクタ単位でデータ転送を制御するドライブインターフェイスである。

【 0 0 2 1 】

さらに、本発明のストレージ装置 1 0 0 は、ホストインターフェイス 2 a ~ 2 d 及びドライブインターフェイス 3 a ~ 3 d からの転送データを一時的に蓄積するキャッシュメモリ 4 a ~ 4 d を備える。ストレージ装置 1 0 0 では、キャッシュメモリ 4 a ~ 4 d を設けることで、ドライブ書き込み時のホスト転送データの消失防止や、キャッシュヒットによるホストリード性能の向上が可能となる。ドライブインターフェイス 3 a ~ 3 d には、ハードディスクドライブが接続され、ホスト側から見たアクセス空間である論理ボリューム 8 a ~ 8 d (LU0 ~ LU3, LU: Logical Unit) が割り当てられる。

10

【 0 0 2 2 】

マイクロプロセッサ 5 a、5 b は、各コントローラコントローラ 1 a ~ 1 d、ホストインターフェイス 2 a ~ 2 d、ドライブインターフェイス 3 a ~ 3 d の制御手段である。マイクロプロセッサ 5 a、5 b は、メモリコントローラハブ 6 a、6 b (Memory Controller Hub :MCH) を介して、コントローラ 1 a ~ 1 d 及びプロセッサメモリ 7 a、7 b (MP Mem 0, MP Mem 1) と接続されている。例えば、メモリコントローラハブ 6 a は、コントローラ 1 a (CTL 0) 及びコントローラ 1 b (CTL 1) と接続されており、マイクロプロセッサ 5 a は、コントローラ 1 a、コントローラ 1 b、ホストインターフェイス 2 a、ホストインターフェイス 2 b、ドライブインターフェイス 3 a、ドライブインターフェイス 3 b を直接の制御対象とする。また、メモリコントローラハブ 6 a 及び 6 b は、MP I/O (MP I/O) によりコントローラ 1 a、1 b 及びコントローラ 1 c、1 d に接続されている。ストレージ装置 1 0 0 は、以上の要素より構成される。

20

【 0 0 2 3 】

次に、ストレージ装置 1 0 0 を構成するコントローラ 1 a ~ 1 d の内部構成を説明する。図 2 は、コントローラ 1 a ~ 1 d のブロック図である。図 2 では、コントローラ 1 に接続される各 I/O (MP I/O、クロス I/O、ホスト I/O (Host I/O)、ドライブ I/O (Drv I/O)、デュアル I/O) を介したデータ転送方法として、例えば、高速シリアルバス規格である PCI エクスプレス (PCI-Express) が採用される。

30

【 0 0 2 4 】

DDR I/F 1 0 は、キャッシュメモリ 4 とコントローラ 1 間のデータ転送を制御する。この場合、キャッシュメモリ 4 には、DDR (Double Data Rate Synchronous DRAM) に代表される高速メモリが用いられる。DDR I/F 1 0 及びキャッシュメモリ 4 は、キャッシュ I/O により接続されている。

【 0 0 2 5 】

PCI I/F 1 1 a ~ 1 1 e (PCI I/F) は、物理レイヤー、リンクレイヤー、トランザクションレイヤーの各レイヤーにおける通信プロトコルに従い、データ転送を制御する。トランザクションレイヤーのデータ転送は、TLP (Transaction Layer Packet) 形式のパケット単位で行われる。

40

【 0 0 2 6 】

メモリコントローラ 1 2 (MEM CTL) は、キャッシュメモリに対するリード、ライト、RMW (Read Modify Write)、DRAM に対するリフレッシュ動作など DRAM に対する各種コマンド発行を制御する。また、メモリコントローラ 1 2 は、コントローラ 1 a 内部の高速スイッチ (SW) 1 5 と接続されている。

【 0 0 2 7 】

各ブリッジ回路 1 3 a ~ 1 3 e (BRG MP, BRG Cross, BRG D, BRG Dual, BRG D) は、コントローラ 1 a 内部の高速スイッチ 1 5 と接続され、転送された TLP に対する受信バ

50

ッファ、送信バッファを含むブリッジ回路である。ブリッジ回路 13 a ~ 13 e は、バッファを備えることで、TLP を構成するヘッダ情報の保持、書き換えが可能であり、コントローラ 1 内、コントローラ 1 間の TLP 転送に適したパケット変換機能も備える。これについては、図 8 ~ 図 13 にて説明する。

【 0 0 2 8 】

DMA 回路 14 a、14 b (DMA, H DMA) は、ブリッジ回路 13 e (BRG H)、ブリッジ回路 13 c (BRG D) と接続されており、転送元の各 I/O を介したリクエスタとして、転送動作を制御する。DMA 回路 20 (Multi DMA) は、マイクロプロセッサ 5、キャッシュメモリ 4 との各 I/O を介した転送を制御する。

【 0 0 2 9 】

高速スイッチ 15 は、異なる I/O を介して、同一の I/O への TLP 転送要求 (REQ) が発生した場合、アービタ 16 に従い、各 I/O を介した TLP 転送を時分割多重して転送する。

【 0 0 3 0 】

アービタ 16 は、同時発生 of TLP 転送要求 (REQ) に対し、決められた優先順位に従った I/O 選択命令を生成する。また、アービタ 16 は、同時発生 of TLP 転送要求 (REQ) の数、或いは、一定時間内に TLP 転送要求 (REQ) を生成した転送 I/O 数とある一定時間辺りに発生した TLP 転送要求 (REQ) 回数から転送先の I/O 上の TLP 多重度を検出する。

【 0 0 3 1 】

トラフィックモニタ 17 (Traffic Mon.) は、ある一定時間間隔で、転送先の I/O 上の TLP 転送の占有時間を計測することで、トラフィック量を算出する。トラフィックモニタ 17 は、例えば、単位時間当たりで、転送元の I/O を介して、出力の I/O に接続される時間を計測する。計測は、アービタ 16 で生成される選択信号を利用したり、高速スイッチ 15 に、転送先の I/O 上を流れる TLP 転送の期間に同期して、転送状態を示すステータス検出機能を持たせることで実現する。

【 0 0 3 2 】

ダミー転送モニタ 21 (Dmy Xfer Mon.) は、コントローラ 1 内部でダミーデータを伴う転送を検出する機能である。ダミーデータを伴う転送は、受信した TLP データ量が、受信バッファ量を満たさず、一定時間それ以降の受信がない場合に、ダミーデータを付加して受信バッファ量を強制的に満たし、転送 (吐き出し) を行う処理である。

【 0 0 3 3 】

障害モニタ 18 (障害 Mon.) は、特に、コントローラ間 I/O (デュアル I/O 及びクロス I/O) の障害検出機能を備える。I/O 障害は、例えば、ノン・ポストド (Non-Posted) で TLP 転送を行う際に応答 (コンプリーションステータス) タイムアウトや、PCI エクスプレスで転送の際に付加されるパリティエラーを受信側で検出した場合などが挙げられる。

【 0 0 3 4 】

MSI・MPI/F レジスタ 19 は、各モニタ 17、18、21 で検出した多重度、トラフィック量、ダミーデータ転送の検出、障害検出を反映し、MPI/O を介してマイクロプロセッサ 5 a、5 b や、プロセッサメモリ 7 a、7 b に転送する。

【 0 0 3 5 】

ここで、上述した多重度、トラフィック量、ダミーデータ転送検出の具体例について説明する。図 3 は、各転送元 I/O (MP I/O、Host I/O、Drv I/O) 介した、転送先 I/O (Dual I/O) への TLP 転送、多重化の方法である。

【 0 0 3 6 】

転送元 I/O (MP I/O、Host I/O、Drv I/O) のブリッジ回路 13 a、13 c、13 e (BRG MP, BRG H, BRG D) は、受信 TLP を格納するバッファと、TLP 受信間隔を計測するタイマーから構成される。バッファ容量は、例えば、ストレージ装置 100 を構成するハードディスクドライブのアクセス単位 (セクタ) が 512 バイトに対し、2 倍の容量

10

20

30

40

50

(1024バイト(Byte)、1kバイト)を備える。バッファは、高速スイッチ15で他I/Oと多重する際のオーバーヘッドを最小とするためであり、あるI/Oを介した転送中に、別のI/OのTLP転送を継続するために設けられる。

【0037】

転送元I/Oに対するブリッジ回路13a、13c、13eは、コントローラ1内レジスタの数バイト単位アクセスから、数Kバイト単位のデータ転送のTLP転送に対応した構成である。ブリッジ回路13a、13c、13eは、TLPの受信でタイマーを起動し、次のTLP受信までの時間間隔を測定する。ブリッジ回路13a、13c、13eは、タイムアウト検出なら、受信TLPにダミーデータを付加することで1kバイトのバッファ容量を強制的に満たし、高速スイッチ15へ転送する。ブリッジ回路13a、13c、13eは、同時に、アービタ16へTLP転送要求(REQ)を出力する。アービタ16は、TLP転送要求(REQ)発生の順に、アクナリッジ(Acknowledge: ACK)を各ブリッジ回路13a、13c、13eへ返す。アービタ16は、TLP転送要求(REQ)が同時発生した場合には、優先順位に従い、ACKを返す。ブリッジ回路13a、13c、13eは、ACKを受信後、バッファ内のデータを転送する。アービタ16は、ACKを返した順に高速スイッチ15に転送されたTLPを時分割多重し、出力I/O上に伝送させる。

10

【0038】

アービタ16は、一方で、同時発生のTLP転送要求(REQ)の数、或いは、ある一定時間辺りに発生したTLP転送要求(REQ)の回数から、転送先のI/O上のTLP多重度を検出する。トラフィックモニタ17は、ある一定時間間隔で、出力I/O(Dual I/O)上のTLP占有時間を計測することで、トラフィック量を算出する。ダミー転送モニタ21は、コントローラ1内部でダミーデータを伴う転送を検出する機能であり、ブリッジ回路13a、13c、13eを構成するタイマーのタイムアウト検出に相当する。ダミー転送モニタ21は、一定時間単位でタイムアウト検出回数を計測し、閾値との比較を行う。ダミー転送モニタ21は、閾値との比較で、ダミーデータ転送の頻度を計測することでI/O帯域の浪費を検出可能となる。

20

【0039】

次に、I/O帯域の浪費について、図4を用いて説明する。図4では、例えば、MPI/Oにデータ量小のTLPが受信され、ブリッジ回路13a、13c、13eを構成するバッファ量を満たさないまま、ダミーデータと共に転送される場合を示している。ダミーデータは、転送先では無意味なデータであるが、出力I/O上にはデータとして存在するため、I/O帯域の浪費が発生する。ストレージ装置100では、ダミーデータ転送の発生頻度が高い場合、I/O帯域の浪費が増加し、他I/Oの転送レート低下を引き起こすこととなる。

30

【0040】

図4では、一例としてMPI/Oを用いて説明したが、ホストI/O、ドライブI/Oに接続されるホストインターフェイス2、ドライブインターフェイス3との間のレジスタアクセス等が発生する場合にも、ダミーデータを伴う転送が発生する。従って、図3では、全てのブリッジ回路13a、13c、13eにタイマーとダミーデータ付加機能を備える。

40

【0041】

次に、各コントローラ1で検出の多重度、トラフィック量、ダミーデータ検出、障害検出など各ステータスの共有方法を説明する。この場合、マイクロプロセッサ5a(MP0)は、直接制御するコントローラ1a、1b(CTL0,CTL1)、ホストインターフェイス2a、2b(Host I/F 0,1)、ドライブインターフェイス3a、3b(Drv I/F 0,1)に対するステータスや、制御レジスタへのアクセスを、メモリコントローラハブ6a(MCH0)を介したプロセッサメモリ7a(MP Mem 0)との間のTLP転送や、マイクロプロセッサのレジスタアクセスによって行う(自系(CTL0/1)ステータス)。マイクロプロセッサ5b(MP1)は、マイクロプロセッサ5aと同様に、コントローラ1c、1d(CTL2,CTL3)、ホストインターフェイス2c、2d(Host I/F 2,3)、ドライブインターフェイス

50

ス 3 c、3 d (Drv I/F 2,3) からのステータスを取得し、プロセッサメモリ 7 b に更新する (自系 (CTL2/3) ステータス)。

【 0 0 4 2 】

ここで、1つのマイクロプロセッサ 5 から見て、直接制御可能なコントローラ 1、ホストインターフェイス 2、ドライブインターフェイス 3 を「自系」と定義し、デュアル I/O を挟み、反対側にあるコントローラ等を「他系」と定義する。例えば、マイクロプロセッサ 5 a においては、コントローラ 1 a、1 b は「自系」、コントローラ 1 c、1 d は「他系」となり、マイクロプロセッサ 5 b においては、コントローラ 1 c、1 d は「自系」、コントローラ 1 a、1 b は「他系」となる。

【 0 0 4 3 】

一方、各マイクロプロセッサ 5 は、「他系」のコントローラ 1 や I/O のステータスの取得が必要な場合がある。例えば、マイクロプロセッサ 5 a が、コントローラ 1 c あるいは 1 d のステータスや、I/O を介してホストインターフェイス 2 c、2 d、ドライブインターフェイス 3 c、3 d へアクセスする場合はこれに相当する。

【 0 0 4 4 】

この場合、ストレージ装置 1 0 0 では、マイクロプロセッサ 5 a が、デュアル I/O を介して、「他系」のプロセッサメモリ 7 b から、「自系」のプロセッサメモリ 7 a にデータ転送を行うことで、ステータスを共有する。また、ストレージ装置 1 0 0 では、マイクロプロセッサ 5 a が、直接、「他系」のコントローラ 1 c、1 d 内のレジスタ、I/O に対してアクセスすることも考えられる。ここで、ストレージ装置 1 0 0 では、プロセッサメモリ 7 a、7 b 間の通信でステータスを共有する場合、転送経路が 2 種類存在する。そこで、ストレージ装置 1 0 0 では、例えば、ステータス等をプロセッサメモリ 7 b からプロセッサメモリ 7 a へ転送する場合には、マイクロプロセッサ 5 a が、「他系」のコントローラ 1 c、1 d のデュアル I/O に対する障害、多重度、トラフィック量、ダミーデータ転送検出から、転送に最適なデュアル I/O を選択する。

【 0 0 4 5 】

なお、MSI・MPI/F レジスタ 1 9 及びマイクロプロセッサ 5 間の通信は、MPI/O 経由の通信に限定されず、例えば、ハードワイヤーで直接マイクロプロセッサ 5 に接続される割り込み信号や、専用のレジスタ I/F バスを設ける場合も考えられる。ストレージ装置 1 0 0 では、割り込み通知に、MSI (Message Signaled Interrupt) (図示せず) を用いる場合、MPI/O を経由し、TLP としてプロセッサメモリ 7 へ転送する。

【 0 0 4 6 】

次に、図 1 のストレージ装置 1 0 0 で行われるホスト、HDD ドライブ (LU 8 a) 間のデータ転送について説明する。図 6 は、ホストへのデータ転送 (LU リード) の一例として、LU 8 a (LU 0) を再生し、キャッシュメモリ 4 a (Cache 0) に書き込み、キャッシュメモリ 4 a から、各ホストインターフェイス 2 a ~ 2 d (Host I/F 0 - 3) に転送する場合のデータ転送経路を示している。LU 8 a のキャッシュメモリ 4 a への書き込みは、LU 8 a が属するコントローラ 1 a (CTL 0) を制御するマイクロプロセッサ 5 a (MP 0) が、コントローラ 1 a の DMA 回路 1 4 b (DMA) を起動することで行われる。

【 0 0 4 7 】

キャッシュメモリ 4 a への書き込み後、転送先のホストインターフェイス 2 が属するコントローラ 1 内の DMA 回路 1 4 a (H DMA) を、各コントローラ 1 を制御するマイクロプロセッサ 5 が起動し、LU 8 a のデータを転送する。図 6 では、各ホストインターフェイス 2 への転送経路 (xfer0, xfer1, xfer2, xfer3) を図示している。図 6 中、転送元 (キャッシュメモリ 4 a) と、転送先 (Host I/F 3) が属するコントローラ 1 d が互いに隣接しない転送 (xfer3) では、複数の転送経路が発生する。ここで、デュアル I/O、クロス I/O で接続されないコントローラ 1 間の転送経路 (xfer3) を「対角経路」と呼ぶことにする。図 6 において、コントローラ 1 c (CTL 2) 経由の対角経路 xfer3 の場合、転送先コントローラ 1 d (CTL 3) の DMA 回路 1 4 a (H DMA) を起動し、コントローラ 1 c、1 d を同一のマイクロプロセッサ 5 b (MP 1) で転送管理する。

10

20

30

40

50

【0048】

一方で、コントローラ1a(LU8aのデータが書き込まれたキャッシュメモリ4a)とコントローラ1cとはデュアル0I/O(Dual 0I/O)で接続され、そのI/O障害はコントローラ1c側で受信データの転送タイムアウトにより管理される。従って、転送時の障害把握をマイクロプロセッサ5bで一元管理可能となる。なお、LU8aのデータのキャッシュメモリ4aへの書込みは、図6の一例に限定されず、例えば、転送先のホストインターフェイス3の属するコントローラ1へ転送し、転送先のホストインターフェイス3の属するキャッシュメモリ4に書き込むことも考えられる。この場合も、LU8aからキャッシュメモリ4dまでの伝送経路は、コントローラ1b又はコントローラ1cを経由する経路が存在する。本発明の内容は、この場合にも適用される。

10

【0049】

図7は、ホストインターフェイス2aから各LU8へのデータ転送(LU書込み)経路を示している。図7において、ホストインターフェイス2aの受信データは、それぞれが属するコントローラ1aのキャッシュメモリ4aと、デュアル0I/Oを介して対称位置にあるコントローラ1cのキャッシュメモリ4cに二重書きされる。ストレージ装置100では、二重書きによりホストからの転送データの消失を回避することができる。マイクロプロセッサ5aは、二重書き終了後、目的のLU8aが存在するドライブインターフェイス3aへ転送し、LU8aを更新する。マイクロプロセッサ5aは、転送先のコントローラ1aのDMA回路14b(DMA)で転送を制御する。なお、ホストインターフェイス2aから受信されたデータのキャッシュメモリ4への書込みは、図7に限定されず、例えば、キャッシュメモリ4bやキャッシュメモリ4dに二重書きする場合も考えられる。

20

【0050】

以上、図5~図7により、コントローラ1間を接続するデュアルI/O、クロスI/O上を通過するI/Oアクセスについて説明する。デュアル0I/O、デュアル1I/Oには、「他系」のプロセッサメモリ7の共有や、「他系」のコントローラ1、ホストインターフェイス2、ドライブインターフェイス3のレジスタへのアクセスに伴う通信(図5)と、ホストからの転送データ二重書き処理(図7)、ホストへの転送時に発生するデータ転送(図6)が多重される。特に、マイクロプロセッサ5や、ホストインターフェイス2、ドライブインターフェイス3のレジスタアクセスは、図4で説明したダミーデータ転送を伴うため、デュアルI/Oは、それらアクセスの発生によって、帯域の浪費が発生しやすいことになる。一方で、クロス0I/O、クロス1I/Oは、コントローラ1間のデータ通信のみに利用されるため、上記問題が発生しない。

30

【0051】

次に、図6及び図7で説明したホスト及びLU8間のデータ転送について、コントローラ1、マイクロプロセッサ5で行われる具体的な転送方法を説明する。

【0052】

図8~図12は、図1、図6及び図7のストレージ装置100を構成するコントローラ1内、コントローラ1間のデータ転送のみに適用されるTLPフォーマットを示す図である。TLPフォーマットは、PCIエクスプレスのTLP形式に準拠し、さらに、コントローラ1内、コントローラ1間のTLP転送に最適化している。

40

【0053】

図8は、転送元からのライトコマンド31を示し、ヘッダー部とライトデータから構成される。図9は、ライトコマンド31に対するステータス応答を必要とするノン・ポストド(Non-Posted)転送の応答パケット(ライト応答32)であり、ヘッダーとエラーステータスから構成される。図10は、リードコマンド33を示し、ヘッダーのみで構成される。図11は、リードコマンドに対するリード応答34(コンプリーション)であり、ヘッダーと転送元(コンプリータ)からの転送データより構成される。

【0054】

図12は、各コマンドのヘッダー構成ビットに対し、コントローラ1内、コントローラ1間のTLP転送に最適化した部分を示す。図12において、TLPフォーマット(31

50

～ 34)には、データの転送元を示すコンプリータIDとデータの転送先を示すリクエストIDに含める情報として、コンプリータ、或いは、リクエストが属するコントローラを識別するLSI番号や、リクエスト、或いは、リクエストが接続されるI/Oを識別するFunc.(Function)番号が与えられる。さらに、TLPフォーマット(31～34)には、図6における転送経路(xfer3)の対角経路選択に対応する目的で、リクエスト、コンプリータがそれぞれ属するコントローラ1に挟まれた経路コントローラを識別するルートID1、ルートID2が与えられる。なお、コンプリータID、リクエストID、ルートID1及びルートID2は、マイクロプロセッサ5により付与される。

【0055】

ルートID1には、コントローラ1a～1dのコントローラ番号(0～3)が与えられる。しかしながら、図6の転送経路(xfer0-2)や、図7の転送経路(xfer0-3)の各経路識別のため、0～3以外のコード値は、無効、つまり、経由するコントローラ1がないことを示す。ルートID2は、例えば、経由するコントローラ1におけるデータの転送元、つまり、コントローラ1への入力I/O(デュアルI/O又はクロスI/O)を規定する。ルートID2では、逆に、転送先、つまりコントローラ1からの出力I/Oの規定も考えられる。タグ(Tag)番号は、例えば、あるリード命令に対し、そのTLPに含まれるタグ番号を、全てのリード応答34に引継ぎ、リクエスト要求に対する応答パケットの識別に用いる。

【0056】

なお、図8～図12は、全てPCIエクスプレスのTLPフォーマットに準拠しており、その他ヘッダーを構成する各ビットについての説明は省略する。

【0057】

また、コントローラと接続されるI/O(Host I/O, Drv I/O, MP I/O)間のTLP転送には、以上説明したTLPヘッダーは適用されない。当該TLP転送には、PCIエクスプレスのTLPフォーマットに完全に準拠したフォーマットが適用され、転送互換性を維持する。TLP互換については、図13で説明する。

【0058】

図13は、図8～図12で説明したTLP形式で行われるコントローラ1内、コントローラ1間のデータ転送の説明図であり、一例として、図6における転送経路(xfer3)(コントローラ2経由の転送)に適用した場合を示している。つまり、図13は、ホストインターフェイス2c(Host I/F3)にリード命令が受信され、キャッシュメモリ4a(Cache0)にLU8a(LU0)のデータを書込み、コントローラ2b(CTL2)経由でキャッシュデータのホストインターフェイス2cへの転送を行う場合である。

【0059】

図13において、ホストインターフェイス2cよりリード命令を含むTLPが受信されると、コントローラ1dを管理するマイクロプロセッサ5b(MP1)は、受信TLPから転送命令を解析し、再生するLU8を検索する(LU検索は、図14を用いて後述する)。一方、コントローラ1dのブリッジ回路13e(BRGH)においてヘッダーを一時退避する。ヘッダー退避は、例えば、ブリッジ回路13eを構成するバッファからヘッダー部分をプロセッサメモリ7bへ転送することで行われる。

【0060】

マイクロプロセッサ5bは、LUサーチ後、DMA回路14b(DMA)の起動で再生LU8aのデータをキャッシュメモリ4aに格納する。一方、マイクロプロセッサ5bは、退避したリード命令に対するヘッダー情報を、リード命令を受信したI/Oと、コントローラ1cの番号(3)に対応したリクエストIDに変換する。さらに、マイクロプロセッサ5bは、タグ番号を含む引き継がれるヘッダー情報も、プロセッサメモリ7bを介してコントローラ1aのメモリコントローラ12に転送する。

【0061】

つまり、マイクロプロセッサ5bは、リード命令に対するコンプリーションステータス(CplSTS)に、転送先を示すコンプリータIDのLSI番号としてコントローラ1cの

10

20

30

40

50

番号(3)を与え、転送先I/Oとしてホストインターフェイス2を示すDMA回路14a(HDMA)(0x13)を与える。また、マイクロプロセッサ5bは、リクエストIDのLSI番号としてコントローラ1aの番号(0)を与え、転送元I/Oとしてキャッシュを示すメモリコントローラ12(0x05)を与える。マイクロプロセッサ5bは、タグ番号を継承する。さらに、マイクロプロセッサ5bは、ルーティング結果に基づき、経由コントローラ番号、I/O(Route ID 1-2)を格納する。図13の場合、コントローラ1c経由であるため、コントローラ1cにおける入力は、デュアルI/Oであり、ルートID1にコントローラ1cの番号(2)、ルートID2にデュアルI/Oの番号(0x01)を格納する。ルーティングについては、図14以降で説明する。

【0062】

マイクロプロセッサ5bは、コントローラ1aでのコンプリーションステータス転送準備が完了すると、転送先のコントローラ1cのDMA回路14a(HDMA)を起動し、TLP単位でキャッシュデータ転送を制御する。コントローラ1aのメモリコントローラ12(MEM CTL)では、ルートID1、ルートID2により、高速スイッチ15に対し、転送先I/Oへのデータ転送を要求する。コントローラ1aのメモリコントローラ12では、ルートID2が「0x01」であることから、転送先I/OとしてデュアルI/Oへの転送を要求し、コントローラ1a(CTL0)からコントローラ1c(CTL2)へのTLP転送を行う。

【0063】

コントローラ1cにおいて、デュアル0I/O経由でTLPを受信したブリッジ回路13d(BRG Dual)は、ルートID1が「0x02」であることを検出し、自己のコントローラ番号(2)との一致を検出して、経由コントローラであることを識別する。ブリッジ回路13dは、その認識結果と、ルートID2が「0x01」であることと、入力が「デュアルI/O」であることから、コントローラ1間のI/Oであるクロス1I/Oを転送先I/Oとして認識する。さらに、ブリッジ回路13dは、高速スイッチ15に対し、転送先I/O(クロス1I/O)へのデータ転送を要求し、コントローラ1cのクロス1I/Oに接続されるコントローラ1d(CTL3)に転送する。

【0064】

コントローラ1dにおいて、クロス1I/O経由でTLPを受信したブリッジ回路13b(BRG Cross)は、コンプリータIDから、コントローラ識別情報である「0x03」を検出する。さらに、ブリッジ回路13bは、自己のコントローラ番号との一致を検出し、コンプリーションステータスの受信コントローラであることを識別する。また、ブリッジ回路13bは、転送先I/OがDMA回路14a(HDMA)(0x13)であることから、ホストI/Oを転送先として認識する。そして、ブリッジ回路13bは、高速スイッチ15に対し、ホストI/Oへのデータ転送を要求、ホストインターフェイス2d(Host I/F3)へTLP転送する。一方、ブリッジ回路13e(BRG H)は、リードコマンド受信時に保持しておいたタグ番号と転送TLPに対するタグ番号との比較を行い、一致したものに対し、受信時のヘッダーに付け替え、外部I/Oの応答用コンプリーションステータスに変換する。

【0065】

以上より、コントローラ1間の転送は、図8~図12のTLPフォーマットを用い、外部I/O間の転送は、TLPヘッダー付け替えでPCIエクスプレスに完全に準拠したTLPを用い、データ転送が行われる。

【0066】

次に、図6で説明したデータ転送制御について、図14のフローチャートを用いて説明する。図14では、コマンド受信、再生LUの検索から、データ転送経路選択の要否判定、経路選択処理、データ転送までの制御を示す。図14のフローチャートは、マイクロプロセッサ5がコントローラ1等を制御することで成し遂げられる。まず、マイクロプロセッサ5は、ホストからの転送命令をホストインターフェイス2により受信する(S801)。続いて、転送命令を受信したホストインターフェイス2を制御するマイクロプロセッ

10

20

30

40

50

サ5は、転送命令の受信をホストインターフェイス2、或いは、コントローラ1からの割り込み(MSI)を介して検知する。さらに、マイクロプロセッサ5は、受信した転送命令をプロセッサメモリ7に転送し、命令を解析する(S802)。

【0067】

次に、S803からS807の範囲で行われるLU検索、キャッシュ書込み処理を説明する。続いて、マイクロプロセッサ5は、S802の処理後、LU再生命令を検出し、「自系」のコントローラ1で管理するLU8を検索する(S803)。そして、マイクロプロセッサ5は、再生目的のLU8が検出された場合(S803:検出)、S806にジャンプし、再生LU8が属するコントローラ1のDMA回路14b(DMA)を起動し、キャッシュメモリ4へ再生LU8のデータを書き込む。

10

【0068】

これに対して、マイクロプロセッサ5は、S803で該当LU8がない場合(S803:該当なし)、「他系」のマイクロプロセッサ5へ処理を引き継ぐ(S804)。具体的には、「自系」のプロセッサメモリ7内の受信コマンドを、「他系」のプロセッサメモリ7に転送する。「他系」のマイクロプロセッサ5は、自己の管理するLU8を検索し、再生LUを検出し(S805)、再生LU8が属するコントローラ1のDMA回路14b(DMA)を起動し、キャッシュメモリ4へ再生LU8のデータを書き込む(S806)。

【0069】

図6の例では、転送命令を受信したホストインターフェイス2d(Host I/F3)を制御するマイクロプロセッサ5b(MP1)は、管理するLU8c、8d(LU2, LU3)の範囲内でLU8a(LU0)を検索する。続いて、マイクロプロセッサ5bは、結果該当LU8aがないため、マイクロプロセッサ5a(MP0)に処理を引き継ぎ、マイクロプロセッサ5aが管理するLU8a、8b(LU0, LU1)の範囲内でLU8aの検索を行う。そして、マイクロプロセッサ5aは、検索の結果、LU8aを検出し、キャッシュメモリ4(Cache0)への書き込みを制御する。

20

【0070】

最後に、マイクロプロセッサ5は、S807で再生LU8のキャッシュメモリ4への書き込み完了を、ステータスとしてマイクロプロセッサ5間で共有する。ステータスの共有方法については、図5で説明の方法が適用される。

【0071】

なお、マイクロプロセッサ5別に管理されるLU情報は、プロセッサメモリ7やキャッシュメモリ4の一部を用いて、共有する場合も考えられる。従って、LU検索処理(S803~S805)は、単独のマイクロプロセッサ5で行うこともできる。

30

【0072】

続いて、マイクロプロセッサ5は、再生LU8のキャッシュメモリ4への書込み後、キャッシュメモリ4から転送先までの経路選択可否を判定する(S808)。図6の転送経路(xfer0, xfer1, xfer2)は、経路上に障害が発生しない限り、最適経路は一意に決まる。しかしながら、転送経路(xfer3)は、経路長、つまり、通過するコントローラ1の数が同一であり、転送経路が複数存在することから、各コントローラ1の動作状態の変化に応じて、最適な転送経路の選択が可能である。

40

【0073】

続いて、マイクロプロセッサ5は、S808の経路選択可否判定の際において、再生LU8のデータが書き込まれたキャッシュメモリ4、転送命令を受信したホストインターフェイス2それぞれが属するコントローラ情報を、プロセッサメモリ7で共有している。そして、マイクロプロセッサ5は、例えば、キャッシュメモリ4a(Cache0)と、ホストインターフェイス2d(Host I/F3)について、コントローラ1同士が隣接せず、対角位置であることから、転送経路が「xfer3」と判定し、転送経路の選択処理を行う(S808:YES)。

【0074】

続いて、マイクロプロセッサ5は、S809、S80Aのルーティング判定1、2を順

50

に行い、コントローラ障害や、コントローラ間 I / O の多重度、トラフィック量、ダミーデータ転送検出から、最適な転送経路を判定する (S 8 0 9、S 8 0 A)。マイクロプロセッサ 5 は、転送を行う際には、図 8 ~ 図 1 3 で説明したコンプリーションステータス (T L P フォーマット) の生成方法、D M A 起動による転送方法に従い、キャッシュメモリ 4 上の再生 L U 8 のデータを転送する (S 8 0 B)。

【 0 0 7 5 】

これに対して、マイクロプロセッサ 5 は、S 8 0 8 で経路選択が不要の場合 (S 8 0 8 : N O) (図 6 の転送経路の「xfer 0, xfer 1, xfer2」)、S 8 0 B の処理に移行する (S 8 0 B)。なお、マイクロプロセッサ 5 は、S 8 0 B 実行前に S 8 0 9 を実行し、経路上の I / O に障害がある場合に対応する場合も考えられる。

10

【 0 0 7 6 】

最後に、マイクロプロセッサ 5 は、コンプリーションステータス転送中にも、S 8 0 9 及び S 8 0 A 相当の処理である S 8 0 C のルーティング判定 3 を行い、変化するコントローラ 1 の状態に応じて最適な経路選択を行う (S 8 0 C)。

【 0 0 7 7 】

具体例として、図 6 の転送経路が「xfer 3」を図 1 4 のフローチャートに適用した場合を説明する。図 6、図 1 4 において、マイクロプロセッサ 5 b (M P 1) は、ホストインターフェイス 2 d (H o s t I / F 3) により転送命令を受信する (S 8 0 1)。続いて、マイクロプロセッサ 5 b は、転送命令を、ホストインターフェイス 2 d を制御するプロセッサメモリ 7 b (M P M e m 1) に転送する。

20

【 0 0 7 8 】

続いて、マイクロプロセッサ 5 b は、命令内容の解析を行い、再生 L U 8 a を検出する (S 8 0 2)。この場合、マイクロプロセッサ 5 b は、当該マイクロプロセッサ 5 b の管理 L U が L U 8 c、8 d (L U 2, L U 3) であり、該当なしと判断する (S 8 0 3 : 該当なし)。続いて、マイクロプロセッサ 5 b は、「他系」のマイクロプロセッサ 5 a (M P 0) へ処理を引き渡す (S 8 0 4)。

【 0 0 7 9 】

続いて、マイクロプロセッサ 5 a は、当該マイクロプロセッサ 5 a の管理 L U が L U 8 a、8 b (L U 0, L U 1) であり、再生 L U 8 a を検出する (S 8 0 5)。続いて、マイクロプロセッサ 5 a は、コントローラ 1 a (C T L 0) の D M A 回路 1 4 b (D D M A) を起動し、再生 L U 8 a のデータをキャッシュメモリ 4 a (C a c h e 0) へ書き込む (S 8 0 6)。続いて、マイクロプロセッサ 5 a は、キャッシュメモリ 4 a の再生 L U 8 a のデータの書込み終了ステータスを、プロセッサメモリ 7 a、7 b (M P M e m 0, M P M e m 1) で共有する (S 8 0 7)。

30

【 0 0 8 0 】

続いて、マイクロプロセッサ 5 a は、キャッシュメモリ 4 a と転送命令を受信したホストインターフェイス 2 d を対角位置と判断し (S 8 0 8 : Y E S)、S 8 0 9 の移行の処理を行い、最適な転送経路である「xfer 3」の選択を行う (ルーティング判定 1 (S 8 0 9)、ルーティング判定 2 (S 8 0 A))。続いて、マイクロプロセッサ 5 a は、キャッシュメモリ 4 a から再生 L U 8 a のデータの転送を行う (S 8 0 B)。続いて、マイクロ

40

【 0 0 8 1 】

次に、S 8 0 9 のルーティング判定 1 について説明する。図 1 5 のフローチャートは、コントローラ 1 の障害状況に応じた経路判定方法を示している。特に、経路選択に関係の深い各コントローラ間 I / O (デュアル I / O 及びクロス I / O) の障害に対する判定方法である。マイクロプロセッサ 5 は、S 9 0 1 にて、各コントローラ間 I / O (デュアル I / O 及びクロス I / O) の障害状況を取得し、各プロセッサメモリへの書込みで全てのコントローラ間 I / O の障害状況を共有する。障害検出方法の一例として、ノンポストド (N o n - P o s t e d) 転送される応答パケットの受信でタイムアウトが発生した場合や、スト

50

レージ装置起動の際に行われる I / O の自己診断で、コントローラ 1 間の通信不良が検出された場合などが挙げられる。

【 0 0 8 2 】

続いて、マイクロプロセッサ 5 は、S 9 0 2 において、共有している障害情報から「自系」のコントローラ 1 におけるコントローラ間 I / O 障害の有無を確認する。例えば、「自系」のマイクロプロセッサ 5 b は、図 6 の「xfer 3」の場合、コントローラ 1 c、1 d のコントローラ間 I / O 障害を確認する（デュアル 0 I / O、クロス 1 I / O）。続いて、「自系」のマイクロプロセッサ 5 b は、コントローラ間 I / O 障害が無い場合（S 9 0 2 : N O）、コントローラ 1 c 経路を選択する（S 9 0 3）。

【 0 0 8 3 】

これにより、ストレージ装置 1 0 0 では、「xfer 3」の選択経路に対する障害管理が、転送先であるホストインターフェイス 2 d を制御するマイクロプロセッサ 5 b で一元管理可能となる。つまり、ストレージ装置 1 0 0 では、転送元直後のコントローラ間 I / O から、転送先のコントローラ間 I / O まで唯一のマイクロプロセッサ 5 b で管理されるため、障害発生を速やかに把握することができる。障害は、プロセッサメモリ 7 で共有することも考えられるが、デュアル I / O で障害発生の場合には、共有不可能である。従って、S 9 0 2 から S 9 0 3 の優位性が理解される。

【 0 0 8 4 】

続いて、マイクロプロセッサ 5 は、S 9 0 2 でコントローラ間 I / O 障害発生の場合（S 9 0 2 : Y E S）、S 9 0 4 において、「他系」のコントローラ 1 のコントローラ間 I / O 障害を確認する（S 9 0 4）。具体的に、S 9 0 2 では、「自系」のマイクロプロセッサ 5 b が、デュアル 0 I / O、クロス 1 I / O を判定し、コントローラ 1 c 経路の可否を判定する（S 9 0 2）。S 9 0 4 では、「他系」のマイクロプロセッサ 5 a が、デュアル 1 I / O、クロス 0 I / O を判定し、コントローラ 1 b 経路の可否を判定する（S 9 0 4）。

【 0 0 8 5 】

そして、マイクロプロセッサ 5 は、S 9 0 4 で障害検出した場合（S 9 0 4 : Y E S）、経路確保が不可能と判断する（S 9 0 5）。最後に、マイクロプロセッサ 5 は、S 9 0 1 からのリトライを判断する（S 9 0 6）。そして、マイクロプロセッサ 5 は、リトライ回数を越えた場合や、リトライ処理を行わない場合（S 9 0 6 : N O）、エラー終了とする。これに対して、マイクロプロセッサ 5 は、S 9 0 4 でコントローラ間 I / O 障害が無いと判断した場合（S 9 0 4 : N O）、コントローラ 1 b 経路を選択する（S 9 0 7）。

【 0 0 8 6 】

次に、S 8 0 A のルーティング判定 2 について説明する。図 1 6 のフローチャートは、図 2、図 3 で説明したコントローラ間 I / O の多重度、トラフィック量、ダミーデータ検出に基づいた転送経路の選択方法を示している。

【 0 0 8 7 】

マイクロプロセッサ 5 は、S 1 0 1 において、ルーティング判定 1（S 8 0 9）の結果から、各コントローラ間 I / O 障害の有無を確認する。そして、マイクロプロセッサ 5 は、障害 I / O が有る場合（S 1 0 1 : Y E S）、経路選択は不可能であり、ルーティング判定 1 の結果を、最終判定結果とする（S 1 0 2）。また、マイクロプロセッサ 5 は、各コントローラ間 I / O 障害が無く、ルーティング判定 1（S 8 0 9）の結果を優先する場合にも適用され（S 1 0 1 : Y E S）、ルーティング判定 1 の結果を、最終判定結果とする（S 1 0 2）。つまり、ストレージ装置 1 0 0 では、ルーティング判定 1（S 8 0 9）の S 9 0 3 による経路判定が行われた場合に相当するため、このような場合、前述のように転送経路の障害管理に有利である。

【 0 0 8 8 】

これに対して、マイクロプロセッサ 5 は、S 1 0 1 から S 1 0 3 に遷移の場合（S 1 0 1 : N O）、各コントローラ間 I / O（デュアル I / O 及びクロス I / O）に対する多重度、トラフィック量、ダミーデータ検出の各ステータス情報を取得する（S 1 0 3）。各

10

20

30

40

50

情報は、プロセッサメモリ7への共有や、コントローラ1内のMSI・MPI/レジスタ19から直接取得する。続いて、マイクロプロセッサ5は、S104からS106において、取得したステータス情報から経路判定を行う。

【0089】

マイクロプロセッサ5は、最初の判定条件として、デュアルI/Oにおけるダミーデータ転送から判定する(S104)。さらに、マイクロプロセッサ5は、S104でダミーデータ転送検出がある場合(S104: YES)、S108において、ダミーデータ転送の発生I/Oを判定し(デュアル0I/O又はデュアル1I/O)、ダミーデータ転送の無いデュアルI/Oに対する経路を選択する(S108)。マイクロプロセッサ5は、図6の「xfer 3」適用の場合、例えば、デュアル0I/Oにダミーデータ転送を検出の場合、デュアル1I/O経由、つまりコントローラ1b経由を選択する。

10

【0090】

なお、マイクロプロセッサ5は、デュアル0I/O及びデュアル1I/Oが共にダミーデータ転送を検出の場合、ステータス情報の更新、取得を繰り返し、検出頻度の低いデュアルI/Oの経路を選択する。もしくは、マイクロプロセッサ5は、S105以降の判定に委ねるようにすることも考えられる。

【0091】

これに対して、マイクロプロセッサ5は、S104でダミーデータ転送検出がない場合(S104: NO)、S105において、デュアル0I/O及びデュアル1I/Oのケット多重度を比較する(S105)。そして、マイクロプロセッサ5は、多重度に差が発生の場合(S105: YES)、S108に処理を移し、多重度が小さいデュアルI/Oを選択する(S108)。これにより、マイクロプロセッサ5は、高速スイッチ15で多重化の際に発生するオーバーヘッドの転送への影響に配慮した経路選択を行うことができる。

20

【0092】

これに対して、マイクロプロセッサ5は、S105において多重度に差が無い場合(S105: NO)、S106において、デュアル0I/O及びデュアル1I/Oのトラフィック量から判断する(S106)。そして、マイクロプロセッサ5は、トラフィック量に差が発生の場合(S106: YES)、S108において、トラフィック量が小さいデュアルI/Oを選択する(S108)。これにより、マイクロプロセッサ5は、トラフィック量、つまり、デュアルI/Oに対する転送帯域に余裕がある転送経路の選択を行うことができる。

30

【0093】

以上より、マイクロプロセッサ5は、デュアルI/Oによる経路判定について、S104(ダミーデータ転送の有無)に最大の重み付けを与え、次に、S105(多重度の差)、S106(トラフィック量の差)の順に重み付ける。S106によりもS105に重み付けを与える(多重度優先)理由は、ケット多重化により発生する転送オーバーヘッド増加が、転送効率の低下、つまり、デュアルI/Oの帯域を浪費するためである。

【0094】

これに対して、マイクロプロセッサ5は、S106においてトラフィック量の差が無い場合(S106: NO)、S107において、クロスI/Oにおける同様のステータス情報(多重度、トラフィック量等)から、クロス0I/O又はクロス1I/O経路を判断する(S107)。この場合、マイクロプロセッサ5は、S107における判定の優先順位について、S105からS106と同様、「(多重度) > (トラフィック量)」の重み付けを与える。

40

【0095】

また、クロスI/Oは、デュアルI/Oのようなプロセッサアクセス、キャッシュ二重書きも無いため、帯域の浪費は発生しない。従って、マイクロプロセッサ5は、経路選択を行う際には「(Dual I/O) > (Cross I/O)」とし、デュアルI/Oに重み付けをおいた判定を行う。

50

【 0 0 9 6 】

以上より、マイクロプロセッサ 5 は、上記において説明したルーティング判定 1 (図 1 5) 及びルーティング判定 2 (図 1 6) を順番に行うことで、図 6 の「xfer 3」の経路選択を最適に行うことが可能となる。

【 0 0 9 7 】

次に、S 8 0 C のルーティング判定 3 についてについて説明する。図 1 7 のフローチャートは、図 1 4 の S 8 0 B の起動後で、ステータス情報の変化が発生した場合の制御方法を示している。マイクロプロセッサ 5 は、S 8 0 B 起動後、まず、各コントローラ 1 に対するステータス情報からデュアル I / O 及びクロス I / O の障害有無を判定する (S 1 1 1) 。マイクロプロセッサ 5 は、障害発生の場合 (S 1 1 1 : Y E S) 、図 1 5 のルーティング判定 1 を行う (S 1 1 2) 。続いて、マイクロプロセッサ 5 は、S 1 1 3 にて障害が発生したデュアル I / O 及びクロス I / O を回避したルーティングが可能な場合 (S 1 1 3 : N O) 、S 1 1 5 に進む一方、当該ルーティングが不可能な場合 (S 1 1 3 : Y E S) 、エラー終了する。

10

【 0 0 9 8 】

これに対して、マイクロプロセッサ 5 は、S 1 1 1 で障害が無い場合 (S 1 1 1 : N O) 、図 1 6 のルーティング判定 2 を行う (S 1 1 4) 。続いて、マイクロプロセッサ 5 は、S 1 1 5 において、S 1 1 4 或いは S 1 1 2 ~ S 1 1 3 の判定結果から、前回のルーティング結果と比較する (S 1 1 5) 。

【 0 0 9 9 】

そして、マイクロプロセッサ 5 は、差異が無い場合 (S 1 1 5 : N O) 、S 1 1 1 に遷移し、以降の処理を継続する (S 1 1 1 ~ S 1 1 5) 。これに対して、マイクロプロセッサ 5 は、差異発生の場合 (S 1 1 5 : Y E S) 、転送を一時停止するか否かを判定し (S 1 1 6) 、一時停止の場合 (S 1 1 6 : Y E S) 、S 1 1 7 にて、DMA 回路 1 4 を一時停止し、新しい転送経路の選択を、コンプリーションステータスにおけるルート I D 1 、ルート I D 2 を付け替えることで行い、DMA 回路 1 4 を再起動する (S 1 1 7) 。続いて、マイクロプロセッサ 5 は、S 1 1 6 で転送一時停止をしない場合 (S 1 1 6 : N O) 、判定 3 を終了し、受信命令に対する転送終了まで、経路を維持したまま転送する。

20

【 0 1 0 0 】

以上より、マイクロプロセッサ 5 は、転送途中においても、コントローラ 1 の動作状況の変化に応じて経路を最適化することが可能となる。

30

【 0 1 0 1 】

(2) 実施例 2

図 1 8 は、本発明のストレージ装置 1 0 0 の別の構成例を示している。図 1 8 の本実施例におけるストレージ装置 1 0 0 では、図 1 のストレージ装置 1 0 0 の構成におけるコントローラ 1 a ~ 1 d の代わりに、メモリコントローラハブ (Memory Controller Hub : MCH) 6 a ~ 6 d に、ホストインターフェイス 2 a ~ 2 d 、ドライブインターフェイス 3 a ~ 3 d を接続する。また、キャッシュメモリ 4 a ~ 4 d は、マイクロプロセッサ 5 a ~ 5 d に接続され、プロセッサメモリ 7 a 、7 b (MP Mem) は、キャッシュメモリ 4 a ~ 4 d に統合される。なお、プロセッサメモリ 7 は、統合されることに限らず、マイクロプロセッサ 5 a ~ 5 d 上の別 I / O に接続されることも考えられる。

40

【 0 1 0 2 】

さらに、本実施例におけるストレージ装置 1 0 0 では、メモリコントローラハブ 6 間の T L P 転送を制御する目的で、MCH ブリッジ 2 2 a ~ 2 2 d を備える。メモリコントローラハブ 6 間の接続は、図 1 と同様、メモリコントローラハブ 6 a 及びメモリコントローラハブ 6 b をクロス 0 I / O で接続し、メモリコントローラハブ 6 a 及びメモリコントローラハブ 6 c をデュアル 0 I / O で接続し、メモリコントローラハブ 6 b 及びメモリコントローラハブ 6 d をデュアル 1 I / O で接続し、メモリコントローラハブ 6 c 及びメモリコントローラハブ 6 d をクロス 1 I / O で接続する。なお、本実施例におけるストレージ装置 1 0 0 では、その他の構成は図 1 と同様であり、同一の参照数字をつけ説明を省略す

50

る。

【0103】

図18の構成に対するデータ転送の一例として、LU8aからホストインターフェイス2dへのデータ転送を示している。これは、図6の「xfer3」と同様の転送であり、転送経路が2種類存在することを示している。従って、実施例1と同様、最適な転送経路の選択が必要となる。

【0104】

この場合、マイクロプロセッサ5は、最適経路選択の際には、第1の実施例と同様、最初に、各マイクロプロセッサ5の制御対象であるメモリコントローラハブ6の間のI/O障害を判定する。マイクロプロセッサ5は、メモリコントローラハブ6とMCHブリッジ22より全てのメモリコントローラハブ間I/Oに対する障害の有無を検出する。マイクロプロセッサ5は、取得したメモリコントローラハブ間I/Oの障害情報より、障害の無い経路選択を行う。

10

【0105】

一方、マイクロプロセッサ5は、実施例1の図3で説明したダミーデータ転送、I/O多重度、トラフィック量それぞれの検出を、例えば、メモリコントローラハブ6内に同様の回路を設けることで、行うことができる。トラフィック量の検出については、MCHブリッジ22に含めることも可能である。マイクロプロセッサ5は、メモリコントローラハブ6、MCHブリッジ22から得られる検出情報を用い、図16で説明した方法に従い、経路選択を行う。

20

【0106】

メモリコントローラハブ6間を流れるTLPについては、図8～図12のヘッダー構成を用いることで、選択した転送経路に従ったパケット転送が可能となる。具体的に、メモリコントローラハブ6間を流れるTLPについては、第1の実施例と同様、コントローラ番号を、転送元、転送先、経由するメモリコントローラハブ6の識別情報、転送に利用するI/Oの識別情報に置き換えることで対応可能である。

【0107】

図13で説明の packets ヘッダの付け替えについては、同様の要領でメモリコントローラハブ6にて行われる。

【0108】

図8～図12のヘッダー構成を備えるTLP転送は、メモリコントローラハブ6間のI/Oに限定して行う。マイクロプロセッサ5、ホストインターフェイス2、ドライブインターフェイス3への転送時は、メモリコントローラハブ6間を転送したTLPが到着後、各I/Oにおける転送命令受信時に保持したヘッダーに付け替えを行うことで、TLP互換性を維持する。

30

【0109】

このようにして、ストレージ装置100では、マイクロプロセッサ5が、ホストから受信した転送命令を解析することにより、データの転送経路を判定し、データの転送経路が複数存在するときに、転送先の転送経路に接続されているコントローラ1を制御するマイクロプロセッサ5について、当該マイクロプロセッサ5に接続されているコントローラ1の数が最大となるデータの転送経路を選択して、当該転送経路を介してデータを転送する。

40

【0110】

従って、コントローラ1を複数有するマルチコントローラ構成のストレージ装置において、転送元から転送先までのデータ転送に複数の経路選択が可能の場合、その選択を適切に行うことができるため、所望のデータに効率的にアクセスさせることができる。

【産業上の利用可能性】

【0111】

本発明は、キャッシュメモリやホスト、HDDドライブとの各I/Oを備えるコントローラにより構成されるストレージ装置に広く適用することができる。

50

【図面の簡単な説明】

【0112】

【図1】本発明が適用されるストレージ装置の構成図である。

【図2】ストレージ装置を構成するコントローラの内部構成図である。

【図3】I/Oポートに対する多重度、トラフィック量、ダミー転送検出の説明図である。

。

【図4】I/Oポートに対する多重度、トラフィック量、ダミー転送検出の説明図である。

。

【図5】ストレージ装置におけるマイクロプロセッサ管理のコントローラ割当とプロセッサメモリによるステータス共有方法を示す説明図である。

10

【図6】ストレージ装置におけるデータ転送経路の説明図である。

【図7】ストレージ装置におけるデータ転送経路の説明図である。

【図8】コントロール内、コントロール間のデータ転送に用いられるパケットフォーマットの説明図である。

【図9】コントロール内、コントロール間のデータ転送に用いられるパケットフォーマットの説明図である。

【図10】コントロール内、コントロール間のデータ転送に用いられるパケットフォーマットの説明図である。

【図11】コントロール内、コントロール間のデータ転送に用いられるパケットフォーマットの説明図である。

20

【図12】コントロール内、コントロール間のデータ転送に用いられるパケットフォーマットの説明図である。

【図13】コントロール内、コントロール間のパケットデータ転送例に対する説明図である。

【図14】本発明のデータ転送方法のメインルーチンを説明するフローチャートである。

【図15】コントローラ障害発生時に適用されるルーティング方法を説明するフローチャートである。

【図16】コントローラ間I/Oのパケット多重度、トラフィック量、ダミー転送検出の各状況からルーティング判定の方法を説明するフローチャートである。

【図17】データ転送中に行われるルーティング判定の方法を説明するフローチャートである。

30

【図18】本発明が適用されるストレージ装置の第2の構成図である。

【符号の説明】

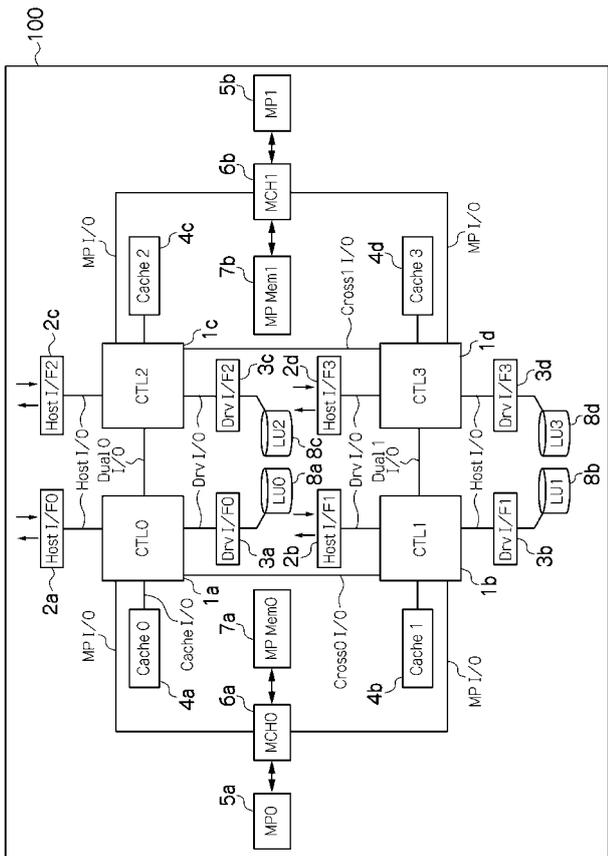
【0113】

1 コントローラ、 2 ホストインターフェイス、 3 ドライブインターフェイス、 4 キャッシュ、 5 マイクロプロセッサ、 6 メモリコントローラハブ、 7 プロセッサメモリ、 8 LU、 10 DDRメモリアンターフェイス回路 (DDR I/F)、 11 PCIインターフェイス回路 (PCI I/F)、 12 メモリコントローラ、 13 I/Oブリッジ回路、 14 DMA回路、 15 高速スイッチ、 16 アービタ、 17 トラフィックモニタ、 18 障害モニタ、 19 MSI・MPI/Fレジスタ、 20 ダミーマルチDMA回路、 21 ダミー転送検出回路、 22 MCHブリッジ、 100 ストレージ装置

40

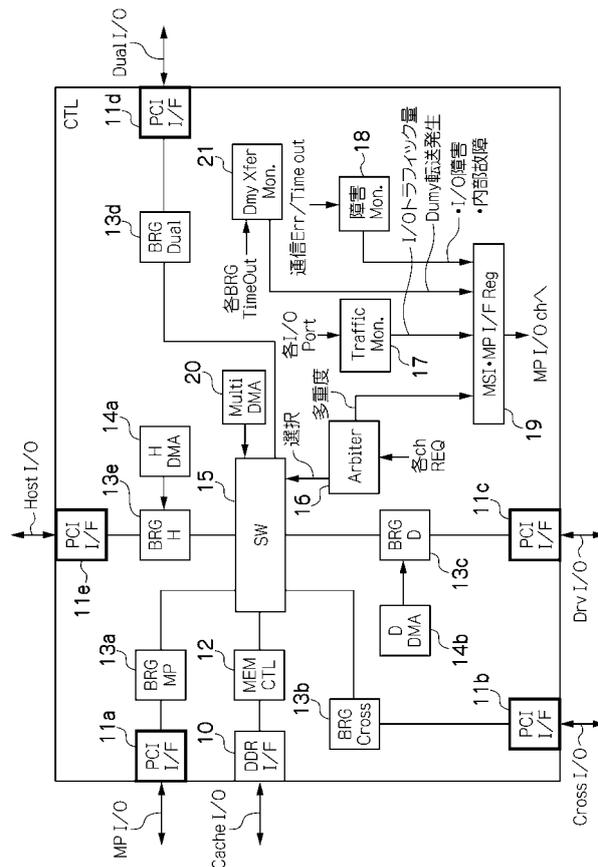
【図1】

図1



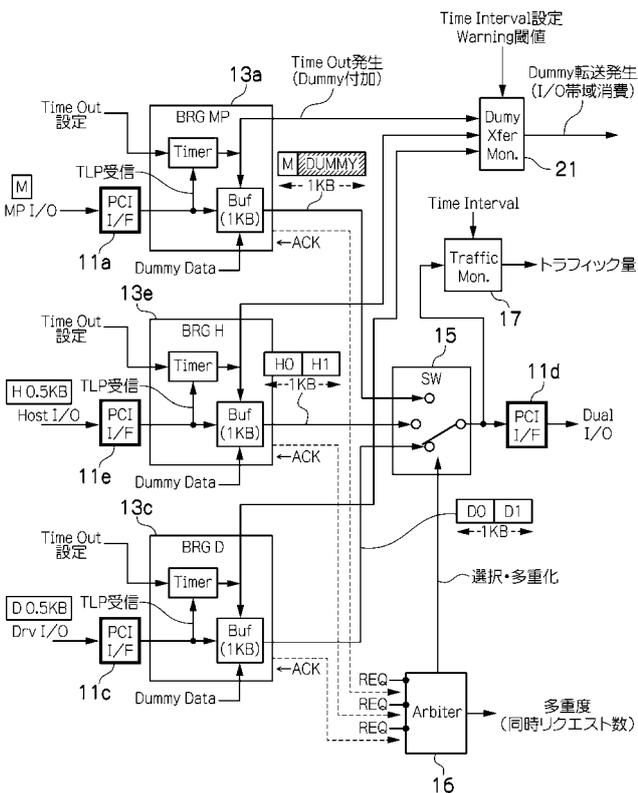
【図2】

図2



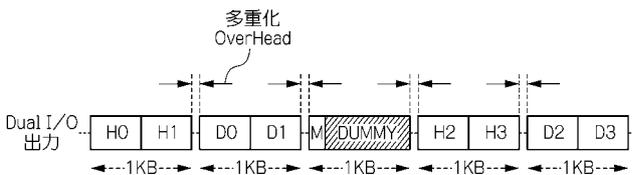
【図3】

図3



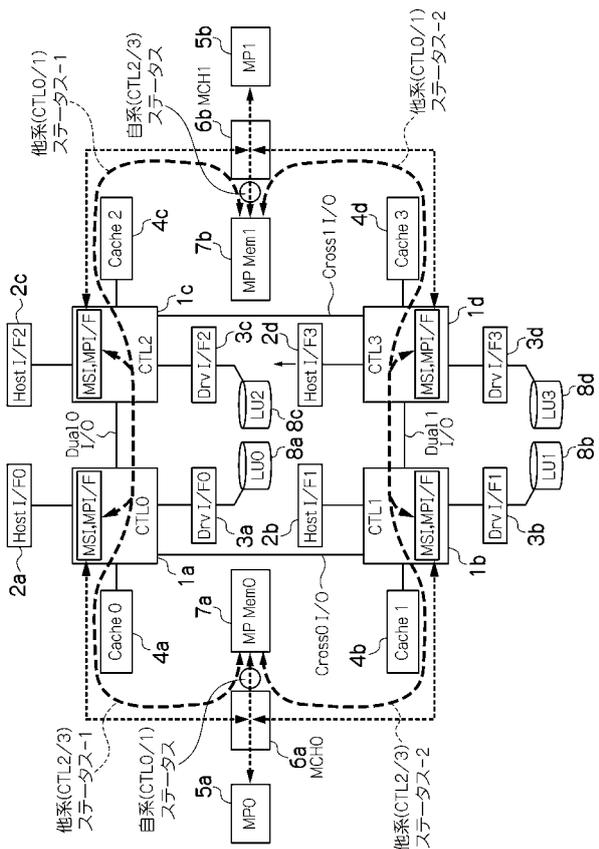
【図4】

図4



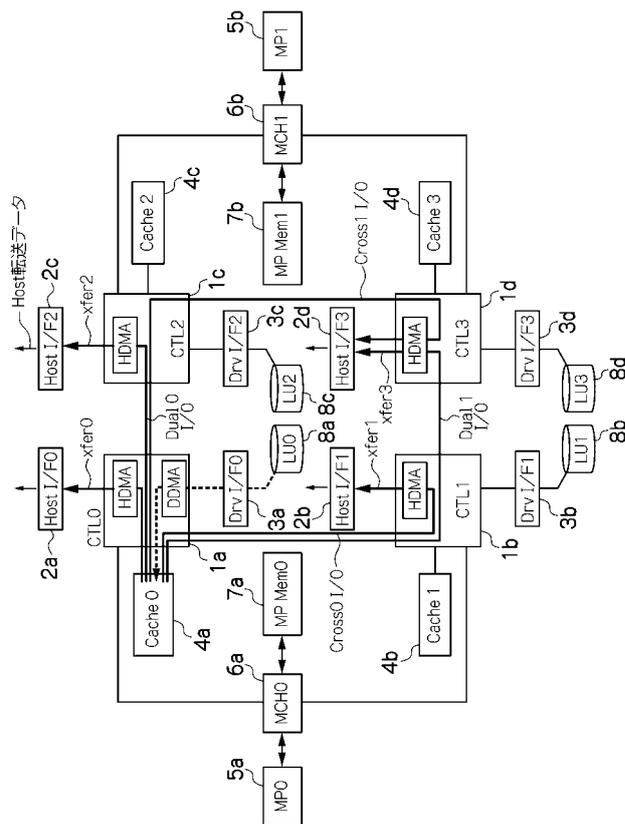
【図5】

図5



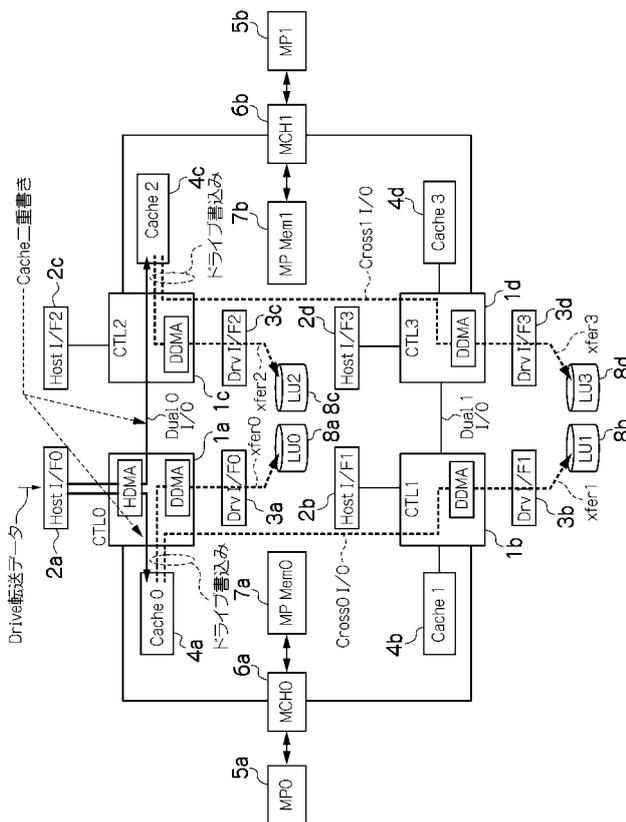
【図6】

図6



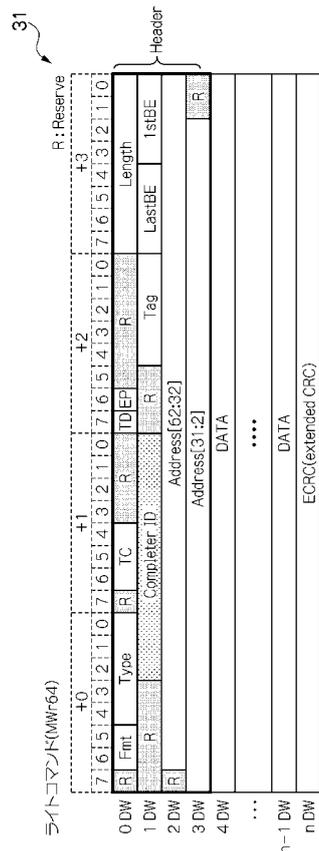
【図7】

図7

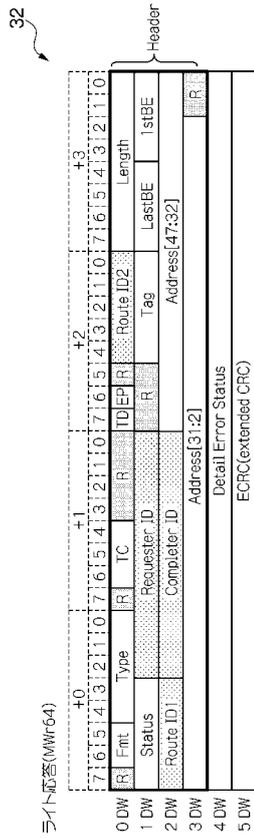


【図8】

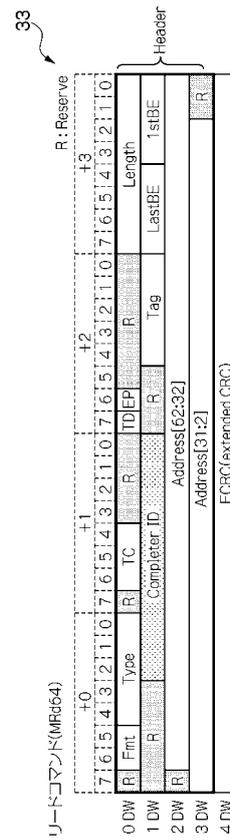
図8



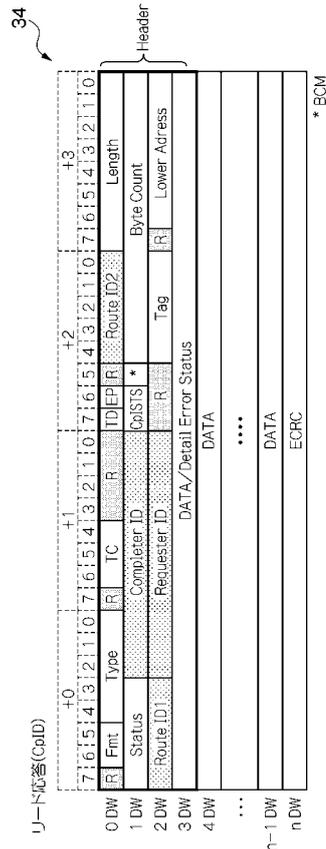
【 図 9 】
図9



【 図 10 】
図10



【 図 11 】
図11



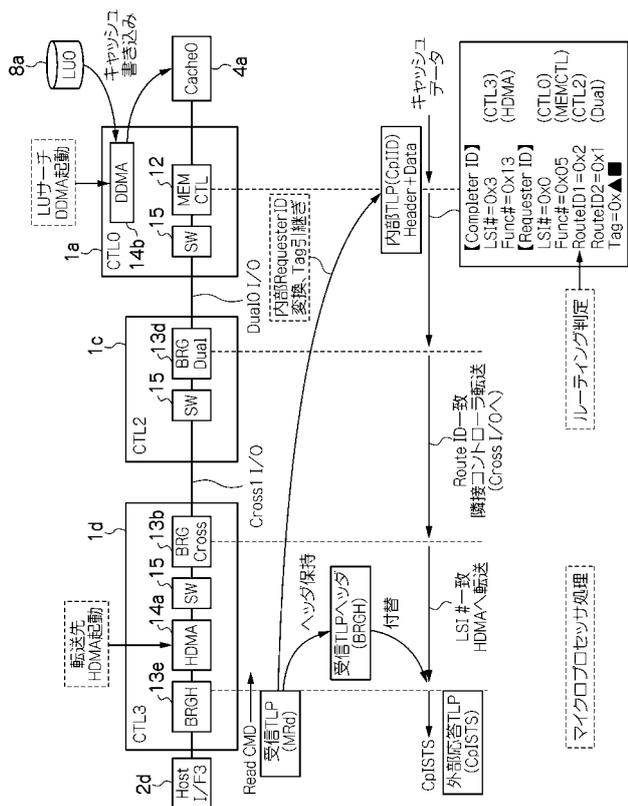
【 図 12 】
図12

各構成ビットの説明(抜粋)

Completer ID	Requester ID	LSI #	Func. #	コントローラ識別番号(0:CTL0, 1:CTL1, 2:CTL3, 3:CTL3)	転送元識別
Route ID1	Route ID2	R-LSI #	R-Func #	0x00: BRGMIP	0x10: BRGMIP
				0x01: BRG DUAL	0x11: BRG DUAL
				0x02: BRG CROSS	0x12: BRG CROSS
				0x03: HDMA(BRGH)	0x13: HDMA(BRGH)
				0x04: DDMA(BRGD)	0x14: DDMA(BRGD)
				0x05: MEMCTL	0x15: MEMCTL
				經由コントローラ識別番号	
				(0:CTL0, 1:CTL1, 2:CTL3, 3:CTL3, other:經由なし)	
				經由コントローラにおける転送元/0	
				0x01: BRG DUAL	
				0x02: BRG CROSS	

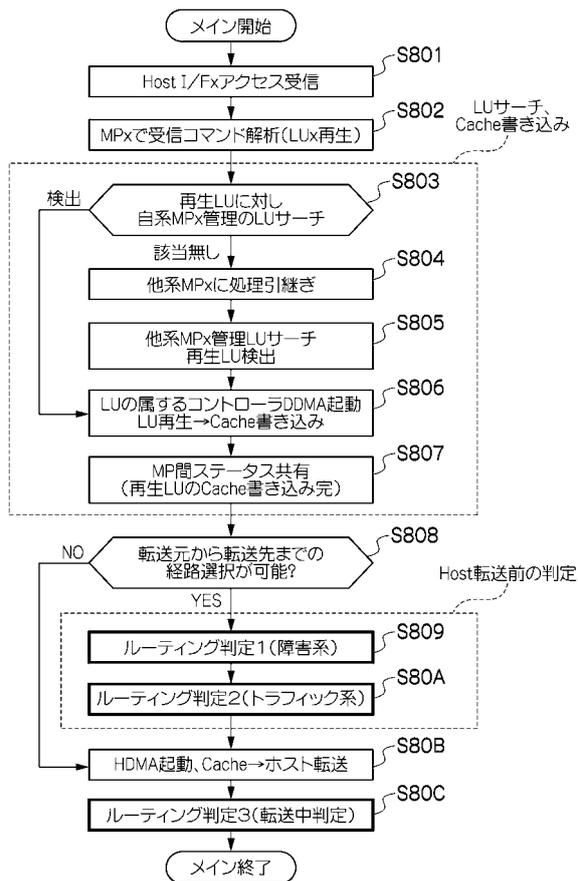
【図13】

図13



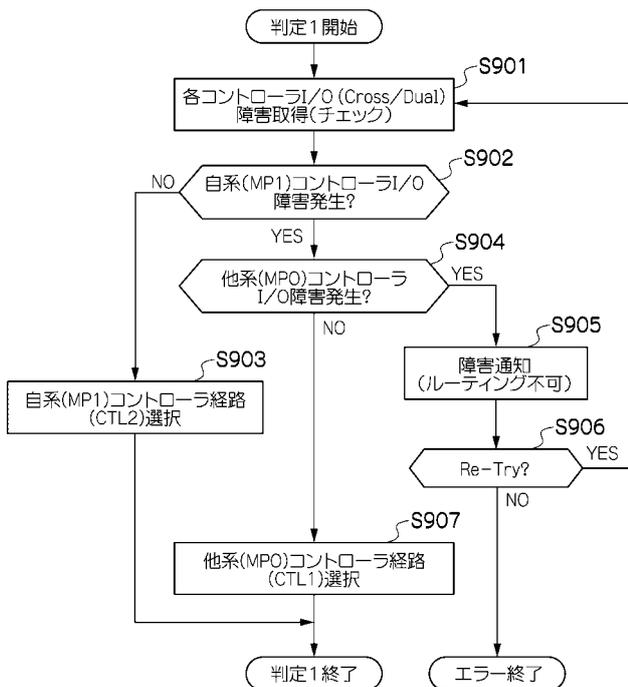
【図14】

図14



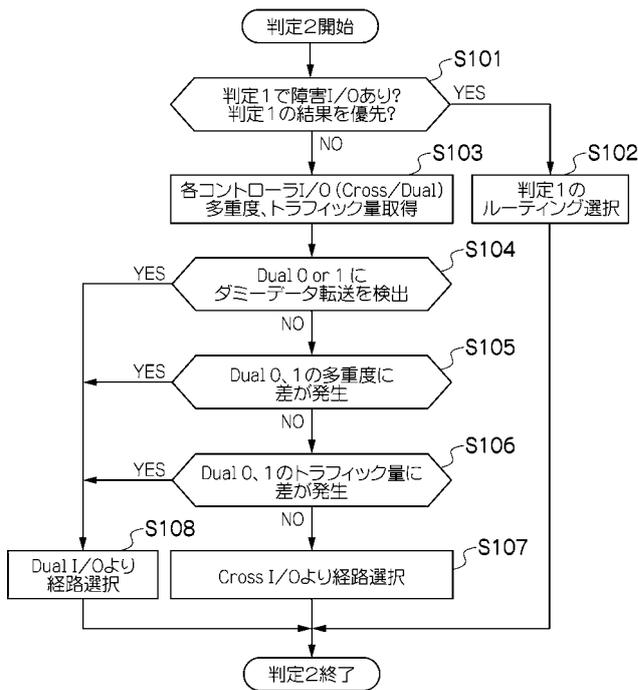
【図15】

図15



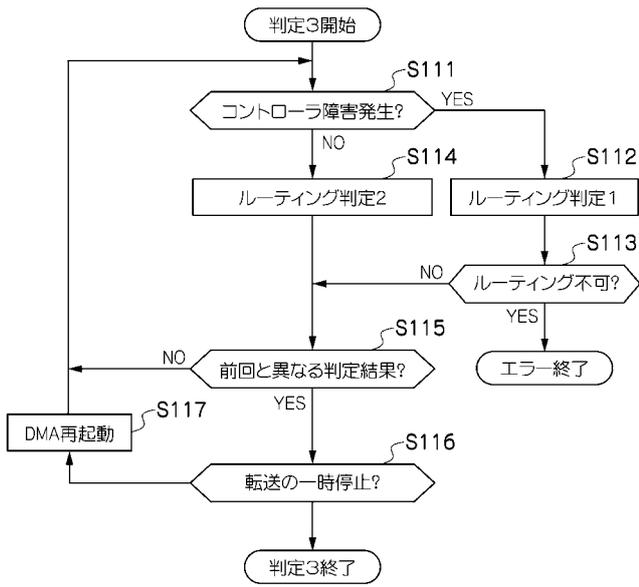
【図16】

図16



【図17】

図17



【図18】

図18

