

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4483168号  
(P4483168)

(45) 発行日 平成22年6月16日(2010.6.16)

(24) 登録日 平成22年4月2日(2010.4.2)

(51) Int.Cl. F I  
**G 0 6 F 3/06 (2006.01)**  
 G 0 6 F 3/06 3 0 1 A  
 G 0 6 F 3/06 5 4 0

請求項の数 13 (全 14 頁)

(21) 出願番号	特願2002-307790 (P2002-307790)	(73) 特許権者	000005108 株式会社日立製作所 東京都千代田区丸の内一丁目6番6号
(22) 出願日	平成14年10月23日(2002.10.23)	(74) 代理人	100100310 弁理士 井上 学
(65) 公開番号	特開2004-145497 (P2004-145497A)	(72) 発明者	森脇 紀彦 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所中央研究所内
(43) 公開日	平成16年5月20日(2004.5.20)	(72) 発明者	藤本 和久 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所中央研究所内
審査請求日	平成17年9月16日(2005.9.16)	審査官	田上 隆一

最終頁に続く

(54) 【発明の名称】 ディスクアレイ制御装置

(57) 【特許請求の範囲】

【請求項1】

ホストコンピュータとのインターフェースを有するチャンネルインターフェース部と、磁気ディスク装置とのインターフェースを有するディスクインターフェース部と、前記磁気ディスク装置に対しリード/ライトされるデータを一時的に格納するキャッシュメモリ部と、前記チャンネルインターフェース部及び前記ディスクインターフェース部と前記キャッシュメモリ部を接続する手段と、を有し、前記ホストコンピュータからのデータのリード/ライト要求に対し、前記チャンネルインターフェース部は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、前記ディスクインターフェース部は、前記磁気ディスク装置と前記キャッシュメモリ部との間のデータ転送を実行することにより、データのリード/ライトを行うディスクアレイ制御ユニットを、複数ユニット有するディスクアレイ制御装置であって、

前記複数のディスクアレイ制御ユニットを接続する手段は、スケラブルに増設可能な複数のスイッチ手段から構成され、前記スイッチ手段は、他のスイッチ手段との間に論理的な完全結合を形成し、また、スイッチの増設時に前記完全結合のバスポロジ再形成を行うための出力先切替用パステブルおよび切替手段を有し、

前記複数のスイッチ手段は、各宛先スイッチ手段に対する物理ルートが分離し冗長パスによる完全結合を形成するにあたり、複数の宛先スイッチ手段に対するパスを同一物理パス内にマージして設定することにより、必要最小限のポート数を有し、

前記スイッチ手段を4個結合する場合に、マトリクス状に配置された前記スイッチ手段

の各々は5個のスイッチ接続用ポートを有し、隣接するスイッチ手段とはそれぞれ2つの直接接続物理ポートを使用して接続することで冗長パスを形成し、また、対角に位置する前記スイッチ手段は、それぞれ1つの直接接続物理ポートを使用した接続、および、隣接スイッチを経由し、隣接スイッチ間の2つの直接物理ポートのうちの1つを共有使用する接続により、冗長パスを形成することを特徴とするディスクアレイ制御装置。

【請求項2】

ホストコンピュータとのインターフェースを有するチャンネルインターフェース部と、磁気ディスク装置とのインターフェースを有するディスクインターフェース部と、前記磁気ディスク装置に対しリード/ライトされるデータを一時的に格納するキャッシュメモリ部と、前記チャンネルインターフェース部及び前記ディスクインターフェース部と前記キャッシュメモリ部を接続する手段と、を有し、前記ホストコンピュータからのデータのリード/ライト要求に対し、前記チャンネルインターフェース部は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、前記ディスクインターフェース部は、前記磁気ディスク装置と前記キャッシュメモリ部との間のデータ転送を実行することにより、データのリード/ライトを行うディスクアレイ制御ユニットを、複数ユニット有するディスクアレイ制御装置であって、

前記複数のディスクアレイ制御ユニットを接続する手段は、スケラブルに増設可能な複数のスイッチ手段から構成され、前記スイッチ手段は、他のスイッチ手段との間に物理ルートが分離した冗長パスによる完全結合を形成し、また、スイッチの増設時に前記完全結合のバスポロジの再形成を行うための出力先切替パステーブルおよび切替手段を有し、

前記複数のスイッチ手段は、各宛先スイッチ手段に対する物理ルートが分離し冗長パスによる完全結合を形成するにあたり、複数の宛先スイッチ手段に対するパスを同一物理パス内にマージして設定することにより、必要最小限のポート数を有し、

前記スイッチ手段を4個結合する場合に、マトリクス状に配置された前記スイッチ手段の各々は5個のスイッチ接続用ポートを有し、隣接するスイッチ手段とはそれぞれ2つの直接接続物理ポートを使用して接続することで冗長パスを形成し、また、対角に位置する前記スイッチ手段は、それぞれ1つの直接接続物理ポートを使用した接続、および、隣接スイッチを経由し、隣接スイッチ間の2つの直接物理ポートのうちの1つを共有使用する接続により、冗長パスを形成することを特徴とするディスクアレイ制御装置。

【請求項3】

前記チャンネルインターフェース部及び前記ディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報及び前記磁気ディスク装置の管理情報を格納する共有メモリ部と、

前記チャンネルインターフェース部及び前記ディスクインターフェース部と前記共有メモリ部を接続する手段を有する請求項1または2に記載のディスクアレイ制御装置。

【請求項4】

前記チャンネルインターフェース部及び前記ディスクインターフェース部と前記キャッシュメモリ部を接続する手段と、前記チャンネルインターフェース部及び前記ディスクインターフェース部と前記共有メモリ部を接続する手段のそれぞれが、スケラブルに増設可能な複数のスイッチ手段から構成され、前記スイッチ手段は、他のスイッチ手段との間に物理ルートが分離した冗長パスによる完全結合を形成し、また、スイッチの増設時に前記完全結合のバスポロジの再形成を行うための出力先切替パステーブルおよび切替手段を有すること、を特徴とする請求項3記載のディスクアレイ制御装置。

【請求項5】

請求項1または2記載の複数の前記ディスクアレイ制御ユニットを収容する筐体と、複数の前記ディスクアレイ制御ユニットおよび前記複数のディスクアレイ制御ユニット間を接続するスイッチを収容する筐体を、ケーブルを用いて接続することを特徴とする前記ディスクアレイ制御ユニットの増設方法。

【請求項6】

10

20

30

40

50

請求項 1 または 2 記載の複数の前記ディスクアレイ制御ユニットを収容する筐体の 4 個を接続する場合、複数の前記ディスクアレイ制御ユニットを収容する筐体の 2 個と、複数の前記ディスクアレイ制御ユニットおよび前記複数のディスクアレイ制御ユニット間を接続するスイッチを収容する筐体の 2 個を、ケーブルを用いて接続することを特徴とする前記ディスクアレイ制御ユニットの増設方法。

【請求項 7】

請求項 1 または 2 記載の複数の前記ディスクアレイ制御ユニットおよび前記複数のディスクアレイ制御ユニット間を接続するスイッチを収容する筐体の 2 個のそれぞれは 2 個の前記スイッチを収容し、前記 2 個のスイッチのそれぞれは、他の 3 個のスイッチとケーブルを用いて接続されることを特徴とする前記ディスクアレイ制御ユニットの増設方法。

10

【請求項 8】

請求項 1 または 2 記載の複数の前記ディスクアレイ制御ユニットおよび前記複数のディスクアレイ制御ユニット間を接続するスイッチを収容する筐体の 2 個のそれぞれは 2 個の前記スイッチを収容し、前記 2 個のスイッチのそれぞれは、同一筐体内の前記スイッチとはバックプレーンを用いて接続し、同一筐体外の前記 2 個のスイッチとはケーブルを用いて接続されることを特徴とする前記ディスクアレイ制御ユニットの増設方法。

【請求項 9】

複数のディスクアレイ制御ユニット(DKC)と、該複数のディスクアレイ制御ユニットと接続される複数の接続ユニット(GSW)と、該接続ユニットを相互に接続するパスを有するディスクアレイ制御装置であって、

20

上記ディスクアレイ制御ユニットは、ホストコンピュータとのインターフェースを有する第 1 のインターフェース部と、第 1 の記憶装置とのインターフェースを有する第 2 のインターフェース部と、上記第 1 の記憶装置に対しリード/ライトされるデータを一時的に格納する第 2 の記憶装置と、上記第 1 のインターフェース部及び上記第 2 のインターフェース部と上記第 2 の記憶装置を接続する転送線路とを有し、

上記複数の接続ユニットのそれぞれは、

上記第 1 のインターフェースまたは第 2 のインターフェースと接続される複数のディスクポートと、上記接続ユニット相互を接続してデータを転送するためのパスと接続されるユニット間ポートと、上記データの転送宛先となる接続ユニットとデータを出力するべきユニット間ポートの対応を変化させる管理手段を有し、

30

前記複数の接続ユニットは、スケーラブルに増設可能であり、前記接続ユニットのそれぞれは、他の接続ユニットとの間に論理的な完全結合を形成し、各宛先接続ユニットに対する物理ルートが分離し冗長パスによる完全結合を形成するにあたり、複数の接続ユニットに対するパスを同一物理パス内にマージして設定することにより、必要最小限のポート数を有し、

前記接続ユニットを 4 個結合する場合に、マトリクス状に配置された前記接続ユニットの各々は 5 個のユニット間ポートを有し、前記ユニット間ポートのうち隣接する接続ユニットとはそれぞれ 2 つの直接接続物理ポートを使用して接続することで冗長パスを形成し、また、対角に位置する前記接続ユニットは、それぞれ 1 つの直接接続物理ポートを使用した接続、および、隣接の接続ユニットを経由し、接続ユニット間の 2 つの直接物理ポートのうちの 1 つを共有使用する接続により、冗長パスを形成することを特徴とするディスクアレイ制御装置。

40

【請求項 10】

上記管理手段は、データの宛先となる接続ユニットとデータを出力するべきユニット間ポートの対応を管理する管理テーブルを有することを特徴とする請求項 9 記載のディスクアレイ制御装置。

【請求項 11】

上記管理テーブルは、宛先となる接続ユニットを少なくとも 4 つ管理することを特徴とする請求項 10 記載のディスクアレイ制御装置。

50

## 【請求項 1 2】

上記管理テーブルは、少なくとも5つのユニット間ポートを有することを特徴とする請求項 1 0 記載のディスクアレイ装置。

## 【請求項 1 3】

上記管理テーブルは、相互に接続されている接続ユニットの数に応じて書き換えられることを特徴とする請求項 1 0 記載のディスクアレイ制御装置。

## 【発明の詳細な説明】

## 【0 0 0 1】

## 【発明の属する技術分野】

本発明は、データを複数の磁気ディスク装置に格納するディスクアレイ装置の制御装置に関する。

10

## 【0 0 0 2】

## 【従来技術】

## 【特許文献 1】

特開 2 0 0 1 - 2 5 6 0 0 3 号

銀行、証券、電話会社等の規模の大きな企業では、従来各所に分散していたコンピュータ及びストレージを、データセンターの中に集中化してコンピュータシステム及びストレージシステム構成することにより、コンピュータシステム及びストレージシステムの運用、保守、管理に要する費用を削減する傾向にある。

## 【0 0 0 3】

このような傾向の中で、大型/ハイエンドのディスクアレイ制御装置には、数百台以上のホストコンピュータへ接続するためのチャンネルインターフェースのサポート(コネクティビティ)、数百テラバイト以上の記憶容量のサポートが要求されている。

20

## 【0 0 0 4】

一方、近年のオープン市場の拡大、今後予想されるストレージ・エリア・ネットワーク(SAN)の普及により、大型/ハイエンドのディスクアレイ制御装置と同様の高機能・高信頼性を備えた小規模構成(小型筐体)のディスクアレイ制御装置への要求が高まっている。

## 【0 0 0 5】

ディスクアレイ制御装置では、上記のように、小規模な構成から超大規模な構成まで、同一の高機能・高信頼なアーキテクチャで対応可能な、スケーラビリティのある構成の制御装置が必要となっており、そのためには、複数のディスクアレイ制御装置を用いて、これらを1つのディスクアレイ制御装置として運用できるディスクアレイ制御装置が必要となる。

30

## 【0 0 0 6】

図 2 は、従来技術の一例として特開 2 0 0 1 - 2 5 6 0 0 3 号で開示されているディスクアレイ制御装置の概要を示すものである。この従来技術では、ホストコンピュータ 5 0 とディスクアレイ制御装置 1 との間のデータ転送を実行する複数のチャンネル I F 部 1 1 と、磁気ディスク装置 5 とディスクアレイ制御装置 2 間のデータ転送を実行する複数のディスク I F 部 1 2 と、磁気ディスク装置 5 のデータを一時的に格納するキャッシュメモリ部 1 4 と、ディスクアレイ制御装置 1 に関する制御情報(例えば、チャンネル I F 部 1 1 及びディスク I F 部 1 2 とキャッシュメモリ部 1 4 との間のデータ転送制御に関する情報、磁気ディスク装置 5 に格納するデータの管理情報)を格納する共有メモリ部 1 3 とを備え、複数のディスクアレイ制御装置 1 間において、共有メモリ部 1 3 およびキャッシュメモリ部 1 4 は全てのチャンネル I F 部 1 1 及びディスク I F 部 1 2 からアクセス可能な構成となっている。この構成では、複数のディスクアレイ制御装置 1 の、チャンネル I F 部 1 1 及びディスク I F 部 1 2 と共有メモリ部 1 3 との間、及びチャンネル I F 部 1 1 及びディスク I F 部 1 2 とキャッシュメモリ部 1 4 との間は、ディスクアレイ制御ユニット間に跨るスイッチを用いた相互結合網 1 0、及び相互結合網 2 0 でそれぞれ接続されている。

40

## 【0 0 0 7】

図 3 は、従来技術のディスクアレイ制御装置 1 の内部をより詳細に示したものである。

50

各ディスクアレイ制御ユニット1は、共有メモリスイッチであるSM-SW110および、筐体間SMパス141を有しており、筐体間SMパス141を通して、相互結合網10に接続される。また、各ディスクアレイ制御ユニット1は、キャッシュメモリスイッチであるCM-SW111および、筐体間CMパス142を有しており、筐体間CMパス142を通して、相互結合網20に接続される。

【0008】

このように、前記従来技術のディスクアレイ制御装置では、複数のディスクアレイ制御ユニット間を、スイッチを用いた相互結合網で接続することにより、小規模な構成から超大規模な構成までスケラビリティのある構成の制御装置が可能となる。

【0009】

【発明が解決しようとする課題】

前記従来技術のディスクアレイ制御装置では、複数のディスクアレイ制御ユニット間を、スイッチを用いた相互結合網で接続することにより、小規模な構成から超大規模な構成までスケラビリティのある構成の制御装置が可能となるが、相互結合網の詳細な構成方法については、触れられていない。小規模な構成から超大規模な構成までスケラビリティを有するディスクアレイ制御装置を実現するには、スケラビリティのある相互結合網の実装方法が重要となる。つまり、接続されるディスクアレイ制御ユニットが少ない場合には、小規模の相互結合網を使用し、接続されるディスクアレイ制御ユニットが多い場合には、大規模の相互結合網を使用することが望ましい。また、小規模の相互結合網から大規模な相互結合網への規模の拡大および縮小は、装置の信頼性を維持したまま、サービスの

【0010】

そこで、本発明の目的は、スケラビリティのあるディスクアレイ制御装置を提供するための、スケラブルなディスクアレイ制御装置相互結合網を提供することにある。

【0011】

より具体的には、本発明の目的は、スケラビリティのあるディスクアレイ制御装置を提供する場合に、信頼性を維持したまま、サービス中断のない増設・減設が可能なディスクアレイ制御装置相互結合網を提供することにある。

また、本発明の目的は、スケラビリティのあるディスクアレイ制御装置を提供する場合に、ハード量が少なくかつコストスケラブルなディスクアレイ制御装置相互結合網の実装方式を提供することにある。

【0012】

【課題を解決するための手段】

上記目的は、チャンネルIF部と、ディスクIF部と、キャッシュメモリ部と、共有メモリ部と、チャンネルIF部及びディスクIF部とキャッシュメモリ部を接続する手段と、チャンネルIF部及びディスクIF部と共有メモリ部を接続する手段を有し、データのリード/ライトを行うディスクアレイ制御ユニットを、複数ユニット接続するための接続網を有するディスクアレイ制御装置であって、前記接続網がスケラブルに拡大可能な複数のスイッチ手段から構成され、前記スイッチ手段は、他のスイッチ手段との間に物理ルートが分離した冗長パスによる完全結合を形成し、さらに、前記スイッチは接続網構成の変更時に伴う冗長パス設定および切替を行うための出力先パステーブルを有する。また、前記スイッチ手段は必要な物理ポート数を削減する目的で、可能な限り物理リンクの共有を行って他のスイッチ手段との間にパスを設定する。

【0013】

また、スケラブル接続網を構成するスイッチ手段間の完全結合接続を行うためのケーブル長をできるだけ短くするために、複数の前記ディスクアレイ制御ユニットを収容する筐体と、複数の前記ディスクアレイ制御ユニットおよび前記複数のスイッチ手段を収容する筐体に分けて、前記複数の前記ディスクアレイ制御ユニットおよび前記複数のスイッチ手段を収容する筐体を隣接配置して接続する実装形態をとる。

10

20

30

40

50

## 【 0 0 1 4 】

他の形態では、複数のディスクアレイ制御ユニット(DKC)と、複数のディスクアレイ制御ユニットと接続される接続ユニット(GSW)と、接続ユニットを相互に接続するパスを有するディスクアレイ制御装置であって、ディスクアレイ制御ユニットは、ホストコンピュータとのインターフェースを有する第1のインターフェース部と、第1の記憶装置とのインターフェースを有する第2のインターフェース部と、第1の記憶装置に対しリード/ライトされるデータを一時的に格納する第2の記憶装置と、第1のインターフェース部及び上記第2のインターフェース部と第2の記憶装置を接続する転送線路とを有する。ここで、接続ユニットは、第1のインターフェースまたは第2のインターフェースと接続される複数のディスクポートと、接続ユニット相互を接続してデータを転送するためのパスと接続されるユニット間ポートと、データの転送宛先(最終目的地)となる接続ユニットとデータを出力すべきユニット間ポートの対応を変化させる管理手段を有する。

一例として、管理手段は、データの宛先となる接続ユニットとデータを出力すべきユニット間ポートの対応を管理する管理テーブルを有する。典型的な例では、管理テーブルは、宛先となる接続ユニットを少なくとも4つ管理することが好ましい。また、管理テーブルは、少なくとも5つのユニット間ポートを有することが望ましい。

データの転送宛先となる接続ユニットとデータを出力すべきユニット間ポートの対応は、接続される接続ユニット数に応じて変化する。一例では、管理テーブルを、相互に接続されている接続ユニットの数に応じて書き換えることとすればよい。

## 【 0 0 1 5 】

その他、本願が開示する課題、及びその解決方法は、発明の実施形態の欄及び図面により明らかにされる。

## 【 0 0 1 6 】

## 【発明の実施の形態】

以下、本発明の実施例を図面を用いて説明する。

## 【 0 0 1 7 】

図1に示すように、ディスクアレイ制御装置100は複数のディスクアレイ制御ユニット1が規模に応じた結合スイッチ2により接続される構成を取る。ディスクアレイ制御ユニット1は、ホストコンピュータ50とのインターフェース部(チャンネルIF部)11と、磁気ディスク装置5とのインターフェース部(ディスクIF部)12と、共有メモリ部13と、キャッシュメモリ部14を有する。チャンネルIF部11及びディスクIF部12と共有メモリ部13の間は複数のディスクアレイ制御ユニット1-1を接続するための複数の結合スイッチ2より構成される結合網を介して接続され、チャンネルIF部11及びディスクIF部12とキャッシュメモリ部14の間も、複数のディスクアレイ制御ユニット1-1を接続するための複数の結合スイッチ2より構成される結合網を介して接続される。これら二つの結合網は独立した網として構成することも、単一の網として構成することも可能である。

## 【 0 0 1 8 】

図4に結合網を構成するための結合スイッチ2の接続構成の具体的な一例を示す。図4では、4個のディスクアレイ制御ユニット(以下、DKC: Disk Array Controllerと呼ぶ)1が1つの結合スイッチ2(以下、GSW: Global Connection Switchと呼ぶ)に収容され、4つのGSW2(本例ではGSW2-0~2-3)を互いに結合して、最大で16個までのDKCが結合可能であるようなディスクアレイ制御装置を例にとって説明する。GSW2-0にはDKC1-0~1-3が接続され、同様に、GSW2-1にはDKC1-4~1-7が、GSW2-2にはDKC1-8~1-11が、GSW2-3にはDKC1-12~1-15が接続される。また、GSW2-0~2-3はマトリクス状に接続されており、各GSW2は、他のGSW2と接続するために5つのポート(Port0~Port5)を使用する。これは、後述するように、ハード量削減のために、できるだけ少ないポート数で、かつ、各GSW2が他のGSW2との間に2本の物理的に分離した冗長パスを確保するのに最低限必要なポート

数である。隣接している G S W 間は、図 4 に示すように内周物理リンク ( 3 - 2 , 3 - 4 , 3 - 6 および 3 - 8 ) および外周物理リンク ( 3 - 1 , 3 - 3 , 3 - 5 および 3 - 7 ) で接続され、また、対角に位置する G S W 2 - 0 と G S W 2 - 2 , G S W 2 - 1 と G S W 2 - 3 間はそれぞれ対角物理リンク ( 3 - 9 , 3 - 1 0 ) を用いて接続される。

#### 【 0 0 1 9 】

図 5 を用いて具体的な隣接 G S W 2 間の接続を行うためのパス設定について説明する。G S W 2 - 0 から G S W 2 - 1 へ向かう接続については、パス 4 - 0 1 A が物理リンク 3 - 1 内に設定され、また、パス 4 - 0 1 B が物理リンク 3 - 2 内に設定される。G S W 2 - 1 から G S W 2 - 0 へ向かう接続については、パス 4 - 1 0 A が物理リンク 3 - 1 内に設定され、また、パス 4 - 1 0 B が物理リンク 3 - 2 内に設定される。このように物理的に分離した 2 本のパス ( 冗長パス ) を用意しておくことにより、片方のパスに障害が発生しても他方のパスを使うことで通信断を回避できる。パスの運用方法としては、( 1 ) 通常動作時には、片方のパスのみを使用し、通常動作パスに障害が発生した場合には他方のパスに切り替える方法、もしくは、( 2 ) 通常動作時に、双方のパスを負荷分散して使用し、パスに障害が発生した場合には障害パスを閉塞して、負荷分散を停止し、正常なパスのみを使用するように切り替えを行う方法などがある。他の全ての隣接 G S W 2 間についても、それぞれ 2 本のパスが設定されるが、設定方法については、前述した G S W 2 - 0 と G S W 2 - 1 の場合と同様であるため、説明を省略する。

図 6 を用いて、具体的な対角 G S W 2 間の接続を行うためのパス設定について説明する。G S W 2 - 0 から G S W 2 - 2 へ向かう接続については、パス 4 - 0 2 A が内周物理リンク 3 - 8 , 3 - 6 内に設定され、また、パス 4 - 0 2 B が対角物理リンク 3 - 9 内に設定される。G S W 2 - 2 から G S W 2 - 0 へ向かう接続については、パス 4 - 2 0 A が内周物理リンク 3 - 2 , 3 - 4 内に設定され、また、パス 4 - 2 0 B が対角物理リンク 3 - 9 内に設定される。G S W 2 - 1 から G S W 2 - 3 へ向かう接続については、パス 4 - 1 3 A が内周物理リンク 3 - 4 , 3 - 6 内に設定され、また、パス 4 - 1 3 B が対角物理リンク 3 - 1 0 内に設定される。G S W 2 - 3 から G S W 2 - 1 へ向かう接続については、パス 4 - 3 1 A が内周物理リンク 3 - 2 , 3 - 8 内に設定され、また、パス 4 - 3 1 B が対角物理リンク 3 - 1 0 内に設定される。このように、対角 G S W 2 間接続の場合には、冗長パスのうち 1 つのパスは 1 ホップでの接続となり、もう 1 つのパスは他の G S W 2 を経由した 2 ホップでの接続となる。

図 7 は別の例であり、対角 G S W 2 間の接続は、図 7 のようなパス設定方法も可能である。図 7 は図 6 とパストポロジは等しく、2 ホップパスの周回方向を全て逆に設定したものと等しい。このような、パスの設定は、G S W 2 内に構成される出力先テーブル ( 後述 ) に、出力を希望する宛先 G S W 2 とそれに対応する出力ポートの関係を設定することで行われる。

#### 【 0 0 2 0 】

以上は G S W 2 が 4 個搭載される最大構成時についてのパス設定について説明を行ったが、次に、小規模構成から G S W 2 の数をスケラブルに増やしていく場合のパス設定についての説明を行う。増設もしくは減設を行う際には、それまでに行われているサービスが中断しないことが好ましい。また、信頼性の条件も変えないことが好ましい。

#### 【 0 0 2 1 】

D K C 1 の数が 4 個以下の時には G S W 2 - 0 のみ用いればよいが、D K C 1 の個数を 5 個以上 9 個未満に増やす場合には、G S W 2 - 0 に G S W 2 - 1 を追加接続する。図 8 に G S W 2 - 0 に G S W 2 - 1 を追加接続した場合の構成を示す。この場合、内周リンク 3 - 1 および外周リンク 3 - 2 内に、それぞれ、双方向のパスを設定することで、互いの宛先に対する冗長パスの設定が可能である。つまり、外周リンク 3 - 1 にパス 4 - 0 1 A と 4 - 1 0 A を設定し、内周物理リンク 3 - 2 にパス 4 - 0 1 B と 4 - 1 0 B を設定する。次に、D K C 1 の個数を 9 個以上 1 3 個未満に増やす場合には、搭載されている G S W 2 - 0 , G S W 2 - 1 に対して、G S W 2 - 3 をさらに追加接続する。この場合、隣接の G S W である G S W 2 - 1 と G S W 2 - 2 の接続については、内周物理リンク 3 - 3 および

外周物理リンク 3 - 4 内に、それぞれ、双方向のパスを設定することで、互いの宛先に対する冗長パスの設定が可能である。

図 9 に対角の G S W である、G S W 2 - 0 と G S W 2 - 2 の接続についての、パス設定を示す。つまり、G S W 2 - 0 から G S W 2 - 2 へ向かう方向については、外周物理リンク 3 - 1、3 - 3 にパス 4 - 0 2 C を、また、対角物理リンク 3 - 9 にパス 4 - 0 2 B を設定する。また、G S W 2 - 2 から G S W 2 - 0 へ向かう方向については、内周物理リンク 3 - 2、3 - 4 にパス 4 - 2 0 A を、また、対角物理リンク 3 - 9 にパス 4 - 2 0 B を設定することで、冗長パスの設定が可能となる。なお、パス 4 - 0 2 C およびパス 4 - 2 0 B は、逆方向のパスとして設定する方法も考えられる。どちらの設定方法をとっても、全ての接続先に対して冗長パスが確保されるため、一点故障に耐えられるトポロジとなる。

10

#### 【 0 0 2 2 】

次に、D K C 1 の個数を 1 3 個以上に増やす場合には、搭載されている G S W 2 - 0、G S W 2 - 1 および G S W 2 - 2 に対して、G S W 2 - 3 をさらに追加接続する。この場合には、最大構成となり、図 5 および図 6 で説明したパスの設定を行えば良い。ここで、対角 G S W 接続のみに着目した場合、図 9 の構成から図 6 の構成への移行を考慮すると、パス 4 - 0 2 C のパス張替えを行う必要があることがわかる。つまり、G S W 2 - 3 を追加する場合には、G S W 2 - 3 に接続するためのパス設定だけではなく、図 9 のパス 4 - 0 2 C ( G S W 2 - 0 から G S W 2 - 1 を経由して G S W 2 - 2 に向かうパス ) を図 7 に示すパス 4 - 0 2 A ( G S W 2 - 0 から G S W 2 - 2 を経由して G S W 2 - 2 に向かうパス ) に切り替えを行う。この切り替えを実現するために、G S W 2 が保有する出力先テーブルの設定を変更可能な構成としておく。

20

ここで、図 1 9 を用いて G S W 2 の内部構成例について示す。G S W 2 は、D K C 1 もしくは他の G S W 2 からのデータが入力される G S W 入力ポート 2 3、D K C 1 もしくは他の G S W 2 に対してデータを出力する G S W 出力ポート 2 4、G S W 入力ポート 2 3 より入力されたデータを一時蓄えてエラーチェックなどを行う入力バッファ 2 1、G S W 入力ポート 2 3 と G S W 出力ポート 2 4 の接続切替をおこなうためのセレクタ 2 2、および、出力先テーブル 2 0 より構成される。出力先テーブル 2 0 は、全ての入力データより参照することができる。

出力先テーブル 2 0 には、出力を希望する G S W 2 およびこれに対応する出力先ポート番号とその設定が有効であるか無効であるかを示す V A L I D フラグにより構成される。

30

図 1 0 を用いて出力先テーブル 2 0 の設定変更について説明する。図 9 に示す G S W 搭載状態に対応する G S W 2 - 0 の出力先テーブル 2 0 の構成を図 1 0 の左側に示す ( 2 0 - 1 )。隣接 G S W 2 - 1 との接続には図 4 で示す P o r t 0 ( リンク 3 - 1 ) と P o r t 1 ( リンク 3 - 2 ) を使用する。対角 G S W 2 - 2 との接続には図 4 で示す P o r t 0 ( リンク 3 - 1 ) と P o r t 2 ( リンク 3 - 9 ) を使用する。これらの V A L I D フラグは有効 ( “ 1 ” ) に設定されている。出力先テーブル 2 0 には、リルート用に G S W 2 - 3 を経由して G S W 2 - 2 に到達するパス設定が格納されているが、この時点では、隣接 G S W 2 - 3 は搭載されていないため、G S W 2 - 3 への隣接接続に対応する V A L I D フラグは全て無効 ( “ 0 ” ) に設定されている。次に、G S W 2 - 3 を追加搭載した場合 ( 図 5、図 6 に対応 ) の G S W 2 - 0 の出力先テーブルの構成 2 0 - 1 を図 1 0 の右側に示す ( 2 0 - 2 )。この時点では、新たに搭載された隣接 G S W 2 - 3 の接続に対応する P o r t 3 ( リンク 3 - 8 ) と P o r t 4 ( リンク 3 - 7 ) を有効にし ( V A L I D フラグ = “ 1 ” )、また、対角 G S W 2 - 2 との接続に使用している P o r t 0 ( リンク 3 - 1 ) を P o r t 3 ( リンク 3 - 8 ) に切り替えるように V A L I D フラグの設定を行う。

40

図 1 1 に、図 1 0 に示した出力先テーブル 2 0 の切替に対応する G S W 増設およびパス設定変更の様子を示す ( 対角に位置する G S W 2 - 0 および G S W 2 - 2 間に設定されるパスのみ示す )。より詳細な切替手順としては、サービス断を防止するため、P o r t 3 ( リンク 3 - 8 ) の設定を有効にした後、一定時間経過後に、それまで使用していた P o r t 0 ( リンク 3 - 1 ) の設定を無効にする。このように、G S W 2 の増設時に出力先テーブルの変更を行うことで、サービス断のない増設が可能となる。なお、G S W 2 の減設に

50



については、増設と逆の手順をおこなうことにより、サービス断を防止できる。出力先テーブルの設定および切替は、ディスクアレイ制御装置に接続される（もしくはその一部である）管理ターミナル（図示されていない）より行うことができる。もしくは、接続されているGSW2を認識して、自律的に各GSW2が各自の出力先テーブル20の設定を行っても良い。

#### 【0023】

次に、本発明の別の実施例として、GSW2の増設に適した実装構成について説明する。図12に示す例では、4個のDKC1を1つのDKC筐体30に実装する。具体的には、DKC1-0~1-3はDKC筐体30-Aに実装され、DKC1-4~1-7はDKC筐体30-Bに実装され、DKC1-8~1-11はDKC筐体30-Cに実装され、DKC1-12~1-15はDKC筐体30-Dに実装される。4つのDKC筐体30はGSW筐体50に実装されたGSW2-0~2-3とケーブル80-1~80-4を使用して接続される。GSW筐体50内のGSW2間のポートの接続は、筐体のバックプレーンもしくはケーブル接続によって行われる。現実的には、GSW2をGSW筐体50に搭載する方式は、配置上の制約を受けることとなるため好ましくない。その理由は、それぞれのDKC30筐体には、複数のディスクユニット筐体（以下、DKU: Disk Unit）60が接続されるため、複数のDKC筐体30を用いる場合には図15に示すような実装配置となるからである。ここで、DKC筐体30に接続されるDKU60は、各DKC1より制御される多数の磁気ディスクが搭載された筐体である。このような配置をとる場合には、DKC筐体30とGSW筐体50を接続は、GSW筐体50をいかなる場所に置いた場合にも、2つのDKC筐体30の配置距離以上の長さのケーブル接続が必要となる部分が出てきてしまう。そこで、GSW筐体50の配置制限を無くし、また、DKC筐体間接続に必要なとされるケーブル長をより短くする方法として、図13に示すように、GSW2を不均等にDKC筐体内に実装する例を示す。

図13においては、DKC筐体30AとDKC筐体30Dは図12と同様とし、DKC筐体30Bには、GSW用スロット31にGSW2-0とGSW2-1を実装し、また、DKC筐体30Cには、GSW2-2とGSW2-3を実装する。DKC筐体30BとDKC筐体30Cを隣接して配置し、別の筐体に搭載されるGSW2間をケーブル80-5~80-10で接続し、DKC筐体30Bの外側にDKC筐体30Aを配置して、DKC筐体30Cの外側にDKC筐体30Dを配置することで、全てのケーブル長は2つのDKC筐体の配置距離より短くすることが可能となる。

図14は全てのDKC筐体をGSW2搭載可能な筐体として構成した場合の例である。この場合、DKC筐体30AおよびDKC筐体30DのGSW2スロット31は空きスロットとなるが、図13の構成と比較して、DKC筐体30の開発が1種類で済むという利点がある。

図16に搭載DKC1の増加に対して、必要となるGSW1およびDKC筐体30の個数の関係を示す。DKC筐体30の個数を2個に増やした時点で、GSW2-0とGSW2-1をDKC筐体30Bに実装する。また、DKC筐体30の個数を3個に増やした時点で、GSW2-2をDKC筐体30Cに実装する。さらに、DKC筐体30の個数を4個に増やした時点で、GSW2-3をDKC筐体30Cに実装する。このように、搭載するDKC1およびDKC筐体に比例して、GSW2を搭載すればよいため、コストスケラブルなシステムを提供することが可能となる。

#### 【0024】

図17を使って次に、DKC筐体30Bおよび30Dに搭載されるGSW2間接続の実装構成例を説明する。DKC筐体30Bに搭載されるGSW2-0およびGSW2-1からは、DKC筐体30Cに面する端面からケーブルを出して、DKC筐体30Cに搭載されるGSW2-2およびGSW2-3に接続される。また、GSW2-0からは、DKC筐体30Aに面する端面からケーブルを出して、DKC筐体30A内のDKC1に接続され、GSW2-2からは、DKC筐体30Dに面するボードの端面からケーブルを出して、DKC筐体30D内のDKC1に接続される。図17の例では、同一DKC筐体内のGS

10

20

30

40

50

W間接続，つまりG S W 2 - 0とG S W 2 - 1，および，G S W 2 - 2とG S W 2 - 3は，ケーブルを用いて接続しているが，ケーブルではなく，バックプレーンで接続する構成としてもよい。また，同一D K C筐体内のG S Wと筐体内D K Cへの接続は，ケーブルを用いて接続しているが，ケーブルではなく，バックプレーンで接続する構成としてもよい。このような実装構成をとることで，D K C筐体30間の接続に必要なケーブル長を最短とすることができる。

図18はD K C筐体内のG S W間接続およびG S Wと筐体内D K Cの接続の別の実装例を示す図である。同一D K C筐体内のG S W間接続およびG S Wと筐体内D K Cへの接続は，図18に示すようにミドルプレーンを用いて接続する構成も考えられる。G S W 2のボードはハーフサイズのボードを使用して，スロットの両面から実装する実装形態とすることができる。この場合，図17に実装されるG S W 2ボードと比較して，G S W 2のボードコスト削減およびスロット数の削減が可能になり，よりコンパクトなD K C筐体30が実現可能となる。

【0025】

【発明の効果】

本発明によれば、信頼性が変わらず、かつ、サービス中断のない増設・減設が可能なスケラビリティのあるディスクアレイ制御装置を、少ないハード量で提供することが可能となる。

【図面の簡単な説明】

【図1】本発明によるディスクアレイ制御装置の構成を示す図。

【図2】従来のディスクアレイ制御装置の構成を示す図。

【図3】従来のディスクアレイ制御装置の他の構成を示す図。

【図4】従来のディスクアレイ制御装置の他の構成を示す図。

【図5】本発明によるディスクアレイ制御装置の他の構成を示す図。

【図6】本発明によるディスクアレイ制御装置の他の構成を示す図。

【図7】図1に示すディスクアレイ制御ユニット内の詳細構成を示す図。

【図8】図7に示すディスクアレイ制御ユニットを複数台接続する構成を示す図。

【図9】図5に示すディスクアレイ制御ユニット内の詳細構成を示す図。

【図10】図9に示すディスクアレイ制御ユニットを複数台接続する構成を示す図。

【図11】図6に示すディスクアレイ制御ユニット内の詳細構成を示す図。

【図12】図11に示すディスクアレイ制御ユニットを複数台接続する構成を示す図。

【図13】図7に示すディスクアレイ制御ユニットを複数台接続する他の構成を示す図。

【図14】本発明によるディスクアレイ制御ユニットの筐体への搭載例を示す図。

【図15】本発明によるディスクアレイ制御ユニットを搭載した筐体を複数台接続する構成を示す図。

【図16】本発明によるディスクアレイ制御ユニット筐体およびスイッチボードのスケラビリティを示す図。

【図17】本発明によるディスクアレイ制御ユニット筐体へのスイッチボード実装例を示す図。

【図18】本発明によるディスクアレイ制御ユニット筐体へのスイッチボード実装例を示す図。

【図19】図19はG S W 2の内部構成例を示す図。

【符号の説明】

1：ディスクアレイ制御装置，2：結合スイッチ，1-1...ディスクアレイ制御ユニット、5...磁気ディスク装置，11...チャンネルI F部、12...ディスクI F部、13...共有メモリ部、14...キャッシュメモリ部、50...ホストコンピュータ。

10

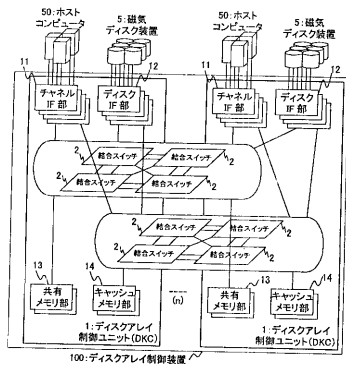
20

30

40

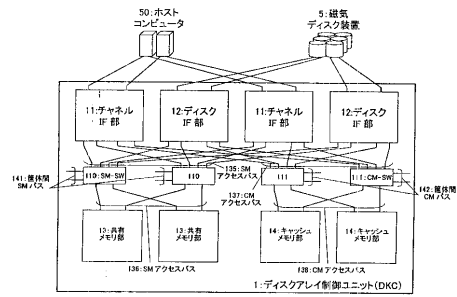
【図1】

図1



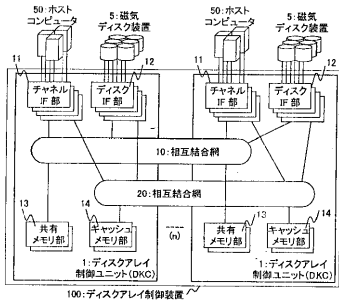
【図3】

図3



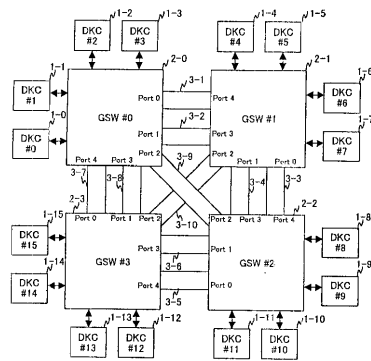
【図2】

図2



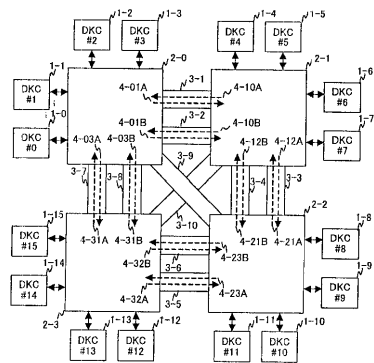
【図4】

図4



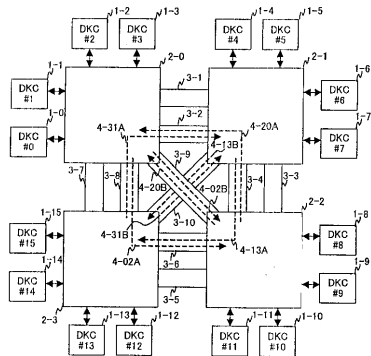
【図5】

図5



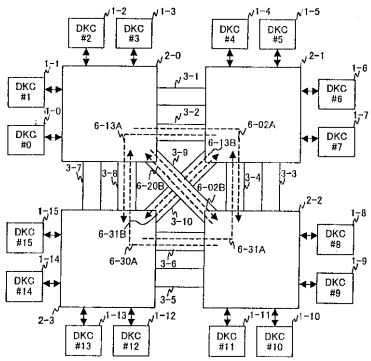
【図6】

図6

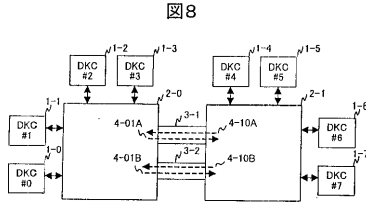


【図7】

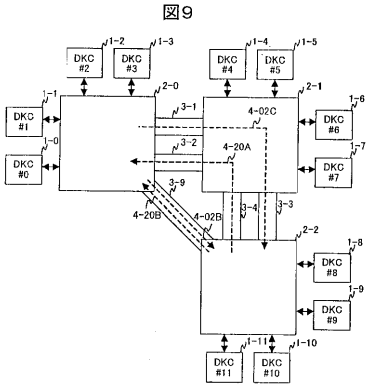
図7



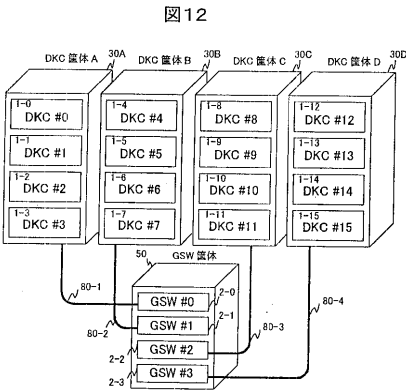
【図8】



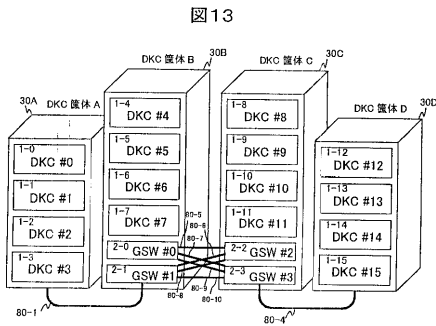
【図9】



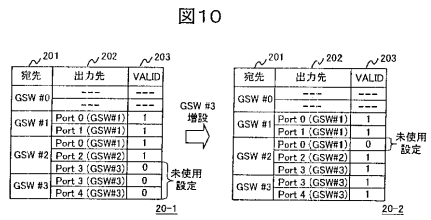
【図12】



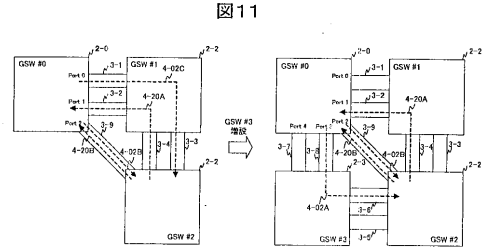
【図13】



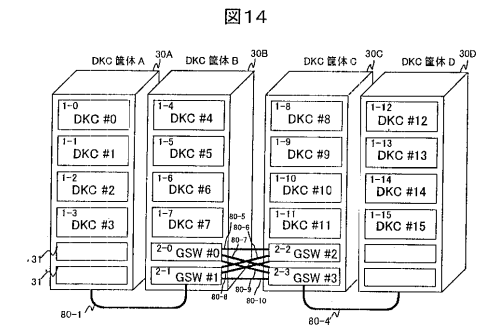
【図10】



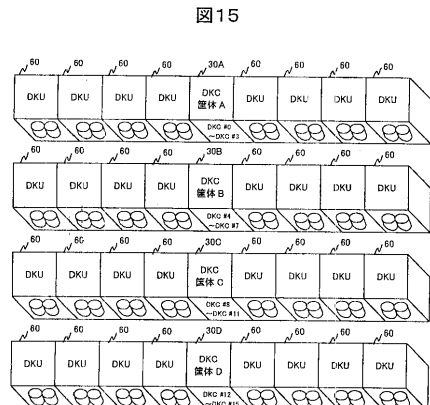
【図11】



【図14】

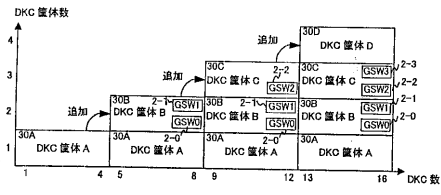


【図15】



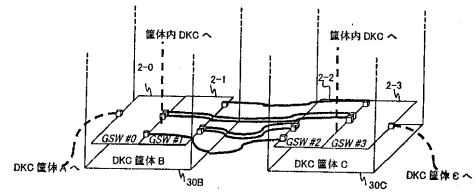
【図16】

図16



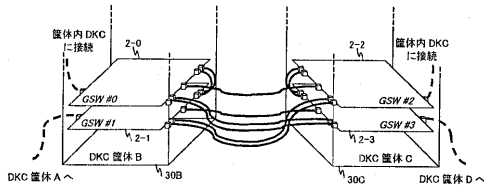
【図18】

図18



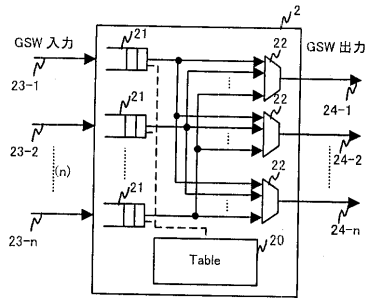
【図17】

図17



【図19】

図19



---

フロントページの続き

(56)参考文献 特開平 1 1 - 1 7 5 2 6 0 ( J P , A )  
特開 2 0 0 1 - 2 5 6 0 0 3 ( J P , A )  
特開 2 0 0 1 - 0 2 7 9 7 2 ( J P , A )

(58)調査した分野(Int.Cl. , D B 名)  
G06F 3/06