

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3946873号  
(P3946873)

(45) 発行日 平成19年7月18日(2007.7.18)

(24) 登録日 平成19年4月20日(2007.4.20)

(51) Int. Cl.		F I		
<b>G06F 13/14</b>	<b>(2006.01)</b>	G06F 13/14	310H	
<b>G06F 3/06</b>	<b>(2006.01)</b>	G06F 3/06	540	
<b>G06F 13/12</b>	<b>(2006.01)</b>	G06F 13/12	330T	

請求項の数 14 (全 16 頁)

<p>(21) 出願番号 特願平10-189957                  (22) 出願日 平成10年6月19日(1998.6.19)                  (65) 公開番号 特開2000-10901(P2000-10901A)                  (43) 公開日 平成12年1月14日(2000.1.14)                  審査請求日 平成15年4月9日(2003.4.9)</p> <p>前置審査</p>	<p>(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</p> <p>(74) 代理人 100099298                  弁理士 伊藤 修</p> <p>(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</p> <p>(72) 発明者 藤林 昭                  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</p> <p>審査官 田内 幸治</p> <p style="text-align: right;">最終頁に続く</p>
--	---

(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接続された第1のインターフェースと、  
 複数のディスク装置と接続された第2のインターフェースと、  
 上記ディスク装置のデータを一時的に格納するキャッシュメモリと、  
 セレクタを有し、

上記セレクタは、上記第1インターフェースと接続されたポートと、かつ、上記第2の  
インターフェースと接続されたポートを含む複数の第1のポートと、上記キャッシュメモリと接続された複数の第2のポートを有し、

占有されている上記第2のポートへの接続要求と空いている上記第2のポートへの接続  
 要求を含む複数の接続要求を上記第1のポートが受けた場合に、空いている上記第2のポ  
 ートへの接続要求を優先的に処理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】

上記セレクタはさらに、  
 空いている上記第2のポートへの接続要求が上記第1のポートへ入力された場合に、上記  
 空いている第2のポートと、上記接続要求を受けた上記第1のポートを接続する手段を備  
 えることを特徴とする請求項1記載のディスクアレイ制御装置。

【請求項3】

上記セレクタはさらに、  
 複数の接続要求をキューイングするキューと、

10

20

上記キューイングされた複数の接続要求のそれぞれが、空いている上記第2のポートへの接続要求かどうかを、一番目にキューイングされた接続要求から順々に判定する手段を有することを特徴とする請求項1または2に記載のディスクアレイ制御装置。

【請求項4】

上記セレクトはさらに、

上記判定を受けた接続要求が、占有されている上記第2のポートへの接続要求だった場合に、該接続要求をキュー内の（空いているポートの数+1）番目にずらす手段を有することを特徴とする請求項3に記載のディスクアレイ制御装置。

【請求項5】

上記セレクトはさらに、

上記判定を受けた接続要求が、空いている上記第2のポートへの接続要求だった場合に、該空いている上記第2のポートと上記接続要求を受けた上記第1のポートを接続する手段を有することを特徴とする請求項3に記載のディスクアレイ制御装置。

【請求項6】

上記セレクトは、

上記第2のポートの数よりも多い数の上記第1のポートを備えたことを特徴とする請求項1乃至5のいずれかに記載のディスクアレイ制御装置。

【請求項7】

上記キャッシュメモリは2重化されていることを特徴とする請求項1乃至6のいずれかに記載のディスクアレイ制御装置。

【請求項8】

上記セレクトは、2重化されたセレクト回路を有することを特徴とする請求項1乃至7のいずれかに記載のディスクアレイ制御装置。

【請求項9】

ホストコンピュータと接続された第1のインターフェースと、複数のディスク装置と接続された第2のインターフェースと、上記ディスク装置に格納されている、又は格納されるデータを一時的に格納するキャッシュメモリと、セレクトを有し、

上記セレクトは、上記第1又は上記第2のインターフェースと上記キャッシュメモリとの間に接続され、

さらに、上記セレクトは、

上記第1のインターフェースに接続されたポートと、かつ、上記第2のインターフェースと接続されたポートを含む複数の第1のポートと、

上記キャッシュメモリに接続された複数の第2のポートを備え、

上記第2のポートのうちポートAへの第1の接続要求が上記第1のポートのいずれかから発行され、その後、上記ポートAへの第2の接続要求が上記第1のポートのいずれかから発行され、さらにその後、上記第2のポートのうち上記ポートAとは異なるポートBへの第3の接続要求が上記第1のポートのいずれかから発行された場合に、上記ポートAと上記第1の接続要求を発行した上記第1のポートを接続し、上記ポートBと上記第3の接続要求を発行した上記第1のポートを接続することを特徴とするディスクアレイ制御装置。

【請求項10】

ホストコンピュータと接続された第1のインターフェースと、複数のディスク装置と接続された第2のインターフェースと、上記ディスク装置に格納されている、又は格納されるデータを一時的に格納するメモリ部と、セレクト部を有し、

上記セレクト部は、上記第1又は上記第2のインターフェースと上記メモリ部との間に接続され、

さらに、上記セレクト部は、

10

20

30

40

50

上記第1のインターフェースに接続されたポートと、かつ、上記第2のインターフェースと接続されたポートを含む複数の入力ポートと、上記キャッシュメモリと接続された複数の出力ポートと、

上記入力ポートから上記出力ポートへの接続要求を、該接続要求の到着順に格納する手段と、

複数の接続要求を調停し、上記入力ポートからの接続要求に対して上記出力ポートを割り当てる調停手段を有し、

上記調停手段は、

上記到着順に格納されている接続要求のうちの1番目の接続要求が、空いている上記出力ポートへの接続要求である場合には、上記1番目の接続要求に上記出力ポートを割り当て

10

、  
上記到着順に格納されている接続要求のうちの1番目の接続要求が、占有されている上記出力ポートへの接続要求である場合には、2番目の接続要求を調べ、

さらに、上記2番目の接続要求が空いている上記出力ポートへの接続要求である場合には、上記2番目の接続要求に上記出力ポートを割り当て、

上記到着順に格納されている接続要求のうちの2番目の接続要求が、占有されている上記出力ポートへの接続要求である場合には、3番目の接続要求を調べ、

以後、最大で、空いている上記出力ポートの数と同じ回数だけ、上記接続要求への上記出力ポートの割り当てを繰り返すことを特徴とするディスクアレイ制御装置。

【請求項11】

20

上記メモリ部は、

物理的に独立しており、2重化された第1及び第2のキャッシュメモリ部を有し、

上記セクタ部は、

上記第1及び上記第2のキャッシュメモリ部に同時にアクセス可能な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。

【請求項12】

上記メモリ部は物理的に分離したキャッシュメモリ部と共有メモリ部を有し、上記キャッシュメモリ部は上記ディスク装置のデータを一時的に格納し、

上記共有メモリ部は上記キャッシュメモリ部及び上記ディスクアレイ制御装置に関する制御情報を格納し、

30

上記セクタ部は物理的に独立した第1及び第2のセクタを有し、

上記第1のセクタは上記キャッシュメモリ部に接続され、

上記第2のセクタは上記共有メモリ部に接続され、

さらに、上記ディスクアレイ制御装置は上記第1及び上記第2のインターフェースと上記キャッシュメモリ部または上記共有メモリ部との間を接続する、物理的に独立した複数のアクセスパスを有し、

上記第1及び上記第2のセクタのうち、少なくとも上記第1のセクタは上記調停手段を備えた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。

【請求項13】

上記キャッシュメモリ部は、物理的に独立しており、2重化された第1及び第2のキャッシュメモリ部を有し、

40

上記共有メモリ部は、物理的に独立しており、2重化された第1及び第2の共有メモリ部を有し、

上記第1のセクタは、上記第1及び上記第2のキャッシュメモリ部に同時にアクセス可能なことを特徴とする請求項12記載のディスクアレイ制御装置。

【請求項14】

上記第1及び上記第2のインターフェースが上記メモリ部にアクセスする際に、アドレスとコマンドを連続して送信し、上記メモリ部へのアクセスパスが確立した後に、データを送信することを特徴とする請求項10乃至13のいずれかに記載のディスクアレイ制御装置。

50

**【発明の詳細な説明】****【0001】****【産業上の利用分野】**

本発明は、データを分割して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。

**【0002】****【従来の技術】**

コンピュータの主記憶のI/O性能に比べて、2次記憶装置として用いられる磁気ディスク装置を使ったサブシステムのI/O性能は3~4桁程度小さく、従来からこの差を縮めること、すなわちサブシステムのI/O性能を向上する努力が各所でなされている。

10

サブシステムのI/O性能を向上させるための1つの方法として、複数の磁気ディスク装置でサブシステムを構成し、データを分割して複数の磁気ディスク装置に格納する手段、いわゆるディスクアレイと呼ばれるシステムが知られている。

**【0003】**

例えば、1つの従来技術(以下、従来技術1という)では、図2に示すようにホストコンピュータ101とディスク制御装置2間のデータ転送を実行する複数のチャンネルI/F部111と、磁気ディスク装置120とディスク制御装置2間のデータ転送を実行する複数のディスクI/F部112と、磁気ディスク装置120のデータを一時的に格納するキャッシュメモリ部115と、磁気ディスク120のデータ及びディスク制御装置2に関する制御情報を格納する共有メモリ部114を備え、キャッシュメモリ部115および共有メモリ部114は全チャンネルI/F部111及びディスクI/F部112からアクセス可能な構成となっている。

20

この従来技術1では、チャンネルI/F部111及びディスクI/F部112と共有メモリ部114、またチャンネルI/F部111及びディスクI/F部112とキャッシュメモリ部115間は1:1に接続されている。

**【0004】**

また、他の従来技術(以下、従来技術2という)では、図3に示すようにホストコンピュータ101とディスク制御装置3間のデータ転送を実行する複数のチャンネルI/F部111と、磁気ディスク装置120とディスク制御装置3間のデータ転送を実行する複数のディスクI/F部112と、磁気ディスク装置120のデータを一時的に格納するキャッシュメモリ部115と、磁気ディスク120のデータ及びディスク制御装置2に関する制御情報を格納する共有メモリ部114を備える。

30

そして、各チャンネルI/F部111及びディスクI/F部112と共有メモリ部114間は共有バス130で接続され、各チャンネルI/F部111及びディスクI/F部112とキャッシュメモリ部115間は共有バス130により接続されている。

**【0005】****【発明が解決しようとする課題】**

ディスクアレイシステムに対する高性能化要求に対して、これまでは、ディスクアレイ制御装置の大規模化や構成要素の高速化、例えばプロセッサ数やキャッシュ容量の増大、高性能プロセッサの適用、内部バス幅の拡大やバス転送能力の向上などで対応してきた。しかしながら、前記従来技術2では、内部バスの転送能力がシステムの大規模化および性能向上に追従するのが困難になりつつある。

40

**【0006】**

そこで内部バス性能を向上させて高いメモリアクセス性能を得るために、前記従来技術1のように、プロセッサとメモリ間を1:1に接続する方法が考えられる。

この方法によれば、メモリに接続したアクセスパス数に比例して内部バス性能が増加する。

しかしながら、搭載するプロセッサ数の増大に比例して、共有メモリおよびキャッシュメモリに接続するアクセスパス数も増加する。

そのため、内部バス性能を最大限に引き出すために、各プロセッサ-メモリ間のアクセス

50

を効率的に制御する必要がある。

【0007】

本発明の目的は上述の課題を解消し、プロセッサ - メモリ間のアクセスパスを効率的に使用し、メモリアクセススループット、特にキャッシュメモリへのアクセススループットの高いディスクアレイ制御装置を提供することにある。

【0008】

【課題を解決するための手段】

上記目的を達成するため、本発明は、

ホストコンピュータとの1つ以上のインターフェース部と、複数の磁気ディスク装置との1つ以上のインターフェース部と、磁気ディスク装置のデータ及びディスクアレイ制御装置に関する制御情報を格納する物理的に独立した1つ以上の共有メモリ部を有し、前記ホストコンピュータとのインターフェース部、または前記複数の磁気ディスク装置とのインターフェース部からは、セレクトを介して前記共有メモリ部にアクセス可能であり、前記ホストコンピュータとのインターフェース部、または前記複数の磁気ディスク装置とのインターフェース部と前記セレクト間と、前記セレクトと前記共有メモリ部間はアクセスパスにより接続されているディスクアレイ制御装置であって、

前記セレクトは、前記ホストコンピュータとのインターフェース部、または前記複数の磁気ディスク装置とのインターフェース部からの複数の入力ポートと、

前記共有メモリ部への複数の出力ポートを相互に接続する手段と、

前記複数の入力ポートから出力ポートへの接続要求を前記接続要求が到着した順に格納する手段と、複数の前記接続要求間の調停を行い、各出力ポートに前記入力ポートからの接続要求を割り当てる調停手段を有しており、

前記調停手段は、前記到着順に格納された接続要求の中の先頭の要求が、現在空いている出力ポートへの要求であれば、該要求へ出力ポートを割り当て、前記到着順に格納された接続要求の中の先頭の要求が、現在使用中の出力ポートへの要求であれば、2番目の要求を調べ、2番目の接続要求が、現在空いている出力ポートへの要求であれば、該要求へ出力ポートを割り当て、2番目の接続要求が、現在使用中の出力ポートへの要求であれば、3番目の要求を調べ、それ以降、多くとも現在空いている出力ポートの数に等しい回数だけ、上記出力ポートへの接続要求の調停（割り当て）を繰り返すようにしている。

【0009】

また、前記共有メモリ部は各々物理的に独立した前記共有メモリ部間で二重化されており、前記セレクトから前記二重化された共有メモリ部両方への同時アクセスが発生するようにしている。

【0010】

また、前記共有メモリ部が、磁気ディスク装置のデータを一時的に格納するキャッシュメモリ部と、前記キャッシュメモリ部及び前記ディスクアレイ制御装置に関する制御情報を格納する共有メモリ部に物理的に分割されており、

前記キャッシュメモリ部に繋がるセレクトと、前記共有メモリ部に繋がるセレクトが物理的に独立しており、

前記ホストコンピュータとのインターフェース部、及び前記複数の磁気ディスク装置とのインターフェース部から、前記キャッシュメモリ部、または前記共有メモリ部へのアクセスパスが物理的に独立しており、

少なくとも前記キャッシュメモリ部に繋がるセレクトに前記調停手段を備えるようにしている。

【0011】

また、前記共有メモリ部は各々物理的に独立した前記共有メモリ部間で二重化されており、前記キャッシュメモリ部は各々物理的に独立した前記キャッシュメモリ部間で二重化されており、少なくとも前記キャッシュメモリ部に繋がるセレクトから前記二重化されたキャッシュメモリ部両方への同時アクセスが発生し、少なくとも前記キャッシュメモリ部に繋がるセレクトに前記調停手段を備えるようにしている。

## 【 0 0 1 2 】

また、前記ホストコンピュータとのインターフェース部、または前記複数の磁気ディスク装置とのインターフェース部から、前記共有メモリ部、または前記キャッシュメモリ部へのアクセスの際に、最初にアドレスとコマンドを連続して送出し、前記共有メモリ部、または前記キャッシュメモリ部までのアクセスパスが確立した後に、データを送出するようにしている。

## 【 0 0 1 3 】

## 【 発明の実施形態 】

本発明の実施例について、以下詳細に説明する。

## 《 実施例 1 》

図 1 に、本発明の一実施例を示す。

ディスクアレイ制御装置 1 は、チャンネル I / F 部 1 1 1、ディスク I / F 部 1 1 2、セクタ部 1 1 3、共有メモリ部 1 1 4 と、アクセスパス 0 1 3 5、アクセスパス 1 1 3 6 から構成される。

## 【 0 0 1 4 】

図 1 3 に示すように、チャンネル I / F 部 1 1 1 は、ホストコンピュータとの 1 つの I / F (ホスト I / F) 5 1 と、1 つのマイクロプロセッサ 5 0 と、1 つの共有メモリアクセス回路 5 2 と、共有メモリ部 1 1 4 への 1 つのアクセスパス I / F から構成される。

データ書き込み時は、ホスト I / F 5 1 は、ホストコンピュータ 1 0 1 から送られてきたデータをパケットに分割して共有メモリアクセス回路 5 2 へ送出する。共有メモリアクセス回路 5 2 は、ホスト I / F 5 1 から送られてくる複数のパケットを 1 つのアクセスパスを使用して共有メモリ部 1 1 4 へ送出する。

データ読み出し時は、共有メモリアクセス回路 5 2 は、共有メモリ部 1 1 4 から送られてきた複数のパケットをホスト I / F 5 1 へ送出する。ホスト I / F 5 1 は、共有メモリアクセス回路 5 2 から送られてきた複数のパケットを 1 つのデータにまとめ、ホストコンピュータ 1 0 1 へ送出する。

マイクロプロセッサ 5 0 は、ホスト I / F 5 1 及び共有メモリアクセス回路 5 2 でのデータの送受信を制御する。

ディスク I / F 部 1 1 2 は、複数の磁気ディスク装置 1 2 0 との 1 つの I / F (ドライブ I / F)、1 つのマイクロプロセッサ、共有メモリ部 1 1 4 への 1 つのアクセス回路と、共有メモリ部 1 1 4 への 1 つのアクセスパス I / F から構成されている。図 1 3 に示すホスト I / F 5 1 をドライブ I / F で置き換えた構成となる。データの書き込みおよび読み出し時には、チャンネル I / F 部 1 1 1 の説明で述べた処理と少なくとも同様の処理が行われる。

ここで、上記に示した個数は一実施例に過ぎず、上記に限定するものではない。

## 【 0 0 1 5 】

共有メモリ部 1 1 4 は、磁気ディスク装置 1 2 0 へ記録するデータや、そのデータの管理情報及びシステム情報などの管理情報を格納する。

セクタ部 1 1 3 には、2 つのチャンネル I / F 部 1 1 1、2 つのディスク I / F 部 1 1 2 からそれぞれ 1 本ずつ、計 4 本のアクセスパス 0 1 3 5 が接続されている。

また、セクタ部 1 1 3 には、2 つの共有メモリ部 1 1 4 へのアクセスパス 1 1 3 6 が 1 本ずつ、計 2 本接続されている。

これら 1 つのセクタ部 1 1 3 とそれに繋がる 2 つのチャンネル I / F 部 1 1 1 及び 2 つのディスク I / F 部 1 1 2 で 1 つのグループを形成し、セクタグループ 1 5 0 と呼ぶ。

本実施例では、ディスクアレイ制御装置 1 が 2 つのセクタグループ 1 5 0 を有する。チャンネル I / F 部及びディスク I / F 部とセクタ部間のアクセスパスと、セクタ部と共有メモリ部間のアクセスパスの間には上記のようなパス数の関係があるため、セクタ部 1 1 3 ではチャンネル I / F 部 1 1 1 及びディスク I / F 部 1 1 2 からの 4 本のアクセスパス 0 : 1 3 5 からの要求の内、共有メモリ部 1 1 4 へのアクセスパス 1 : 1 3 6 の数に相当する 2 個だけを選択して実行する機能を持つ。

10

20

30

40

50

ここで、上記個数は一実施例に過ぎず、個数を上記に限定するものではない。

#### 【0016】

1つのセクタ部113から共有メモリ部114へ接続されるアクセスパスの数を、チャンネルI/F部111及びディスクI/F部112から1つのセクタ部113に接続されるアクセスパスの数より少なくし、チャンネルI/F部111とディスクI/F部112の合計数よりもセクタ部113の数が少なくなるように上記個数を設定すると、共有メモリ部114それぞれへ接続されるアクセスパス数を削減することができる。

共有メモリ部のLSIピンネック及びパッケージのコネクタネックという問題が生じた場合、上記のようにすることで、LSIのピンネック及びパッケージのコネクタネックを解消することができる。

10

#### 【0017】

次に、セクタ部113内部の構成について述べる。

図4は、セクタ部113内の構成を示している。

セクタ部113は、チャンネルI/F部111、またはディスクI/F部112とのI/Fポート210と、共有メモリ部114とのI/Fポート211と、両者間を互いに接続するセクタ206と、I/Fポート210及び211で入出力を行う際のデータのエラーチェック部201と、チャンネルI/F部111、またはディスクI/F部112から送出されたアドレス、コマンド、及びデータをバッファリングするバッファ202と、チャンネルI/F部111、またはディスクI/F部112から送出されたアドレス及びコマンドを解析するアドレス・コマンド(addr、cmd)解析部203と、解析結果を共有メモリ部114とのI/Fポート211への接続要求として、要求の到着順に管理するキュー管理部204と、キュー管理部内に登録された接続要求に基づいてアービトレーションを行い、共有メモリとのI/Fポート211への接続権を決定するアービトレーション部205を有する。

20

#### 【0018】

共有メモリ部のLSIのピンネック及びパッケージのコネクタネックが生じた場合には、前述のように、チャンネルI/F部111、またはディスクI/F部112とのI/Fポート210の数よりも、共有メモリ部114とのI/Fポート211の数を少なくすることで、それらのネックを解消できる。

本実施例では、チャンネルI/F部111、またはディスクI/F部112とのI/Fポート210の数を4つ、共有メモリ部114とのI/Fポート211の数を2つとした。

30

#### 【0019】

図12は、アドレス・コマンド(addr、cmd)解析部203、キュー管理部204、及びアービトレーション部205の詳細構成について示している。

アドレス・コマンド(addr、cmd)解析部203は、チャンネルI/F部111、またはディスクI/F部112とのI/Fポート210の数に相当する4つのバッファ220を有し、バッファ内に各I/Fポート210からのアドレス(addr)及びコマンド(cmd)を格納する。

アドレスは長さが4バイトで最初の1バイトに出力ポート番号(port No.)が示されている。

40

コマンドは長さが4バイトで最初の1バイトにアクセスの種類(読み出し:RD、書き込み:WR、2重読み出し:2R、2重書き込み:2W)が示されている。

ここで、共有メモリ部114が2重化されている場合、2重読み出し及び2重書き込みを行う場合がある。

このような2重アクセス時には、2つのポートを同時に使用するため、両方のポートの使用権を獲得する必要がある。

#### 【0020】

port No.抽出部221では、アドレスから要求ポート番号を取り出す。

本実施例では、port 0に“00”、port 1に“11”を割り当てた。

cmd種類抽出部222では、コマンドからアクセスの種類を取り出す。

50

本実施例では、RDに“00”、WRに“01”、2Rに“10”、2Wに“11”を割り当てた。

使用ポート決定部223では、アクセスの種類が2重アクセスでない場合port No.をそのまま出力し、2重アクセスの場合それを示す“01”を出力する。

キュー管理部204では、アドレス・コマンド(ad r、c m d)解析部203から出力されたport No.を到着順に管理テーブル224に登録する。

#### 【0021】

アービトレーション部205では、管理テーブル224の先頭から要求port No.を1つ取り出し、バッファ227に格納する。

そして、バッファ226に格納された使用中のport No.とバッファ227内の要求port No.を比較器228で比較する。 10

port No.が異なる場合、その番号をセクタ切り替え信号SEL0、SEL1としてセクタ206へ出力し、キュー管理部204内の順番入替え部225へキューの順番を1つ進めるように指示する。

port No.が等しい場合、順番入替え部225へキューの順番を入れ替えるように指示する。

順番の入れ替え方法については、図6の調停のフローの説明で述べる。

ここで、アドレス、コマンドの長さ、アドレスまたはコマンド内のport No.またはc m d種類がそれぞれ示されている場所、port No.またはc m d種類へのビットの割り当て方は一実施例に過ぎず、上記に限定するものではない。 20

また、共有メモリ部114が2重化されていない場合、2重アクセスは生じないため、c m d種類抽出部222及び使用port決定部223は必要なくなり、port No.抽出部221の出力をキュー管理部204へ直接入力すればよい。

#### 【0022】

次に、セクタ部113での処理の手順について述べる。

図5は、チャンネルI/F部111、またはディスクI/F部112とのI/Fポート210のうちの1つにおける処理フローを示している。

まず、ステップ301で、チャンネルI/F部111、またはディスクI/F部112内の共有メモリアクセス回路からアクセス要求(REQ ON)が来るまで待機する。

アクセス要求が来たら、ステップ302でアドレス(ad r)及びコマンド(c m d)を解析する。 30

ステップ303でアドレス(ad r)、コマンド(c m d)にエラーがないかどうかチェックし、エラーがあった場合ステップ315でエラー処理を行い、ステップ301のアクセス要求待機状態に戻る。

エラーがなかった場合、ステップ304で、共有メモリ部114とのI/Fポート211への接続要求として、キューに登録する。

そして、そのキューの内容に基づいてアービトレーションを行う。

ステップ305で、要求した共有メモリ部114とのI/Fポート211が獲得できるまで待機する。

獲得できたら、ステップ306でセクタ206を切り替えて要求を出したI/Fポート210と獲得したI/Fポート211を接続する。 40

#### 【0023】

次に、ステップ307で共有メモリ(SM)部114へアクセス要求(REQ ON)を出し、アドレス(ad r)とコマンド(c m d)を送出する。

ステップ308で、共有メモリ部114からアクセス承認(ACK ON)が返ってくるまで待機する。

アクセス承認(ACK ON)が返ってきたら、ステップ309でチャンネルI/F部111、またはディスクI/F部112内の共有メモリアクセス回路へアクセス承認(ACK ON)を返す。

ステップ310で、データの書き込み時は共有メモリアクセス回路から送られてきたデー 50



タを共有メモリ部 1 1 4 へ送信する。

また、データの読み出し時は共有メモリ部 1 1 4 から送られてきたデータを共有メモリアクセス回路へ送信する。

【 0 0 2 4 】

その際、ステップ 3 1 1 でエラーのチェックを行う。

エラーを発見した場合、ステップ 3 1 5 でエラー処理を行い、ステップ 3 0 1 のアクセス要求待機状態へ戻る。

エラーが無い場合、ステップ 3 1 2 でステータス ( S t a t u s ) が届くのをチェックし、ステータス ( S t a t u s ) が届くまでデータを送信する。

ステータス ( S t a t u s ) が届いたら、ステップ 3 1 3 で共有メモリ部へ、アクセス承認 ( A C K O N ) を取り下げるよう指示し、ステップ 3 0 1 のアクセス要求待機状態に戻る。

10

【 0 0 2 5 】

次に、ステップ 3 0 4 でのアービトレーション ( 調停 ) の方法について述べる。

図 6 は、調停のフローを示している。

ステップ 4 0 1 で、空いている出力ポートがあるかどうか調べ、空きポートができるまで待つ。

ステップ 4 0 1 で空きポートがあれば、ステップ 4 0 2 でキュー管理部 2 0 4 に到着順に格納された接続要求の中の先頭の要求を調べる。

ステップ 4 0 3 で現在空いている出力ポートへの要求であれば、ステップ 4 0 4 でその要求へ出力ポートを割り当てる。

20

ステップ 4 0 3 で、キュー管理部 2 0 4 に到着順に格納された接続要求の中の先頭の要求が、現在使用中の出力ポートへの要求であれば、ステップ 4 0 6 でキュー先頭の要求を ( 空きポート数 + 1 ) 番目にし、ステップ 4 0 1 へ戻る。

ステップ 4 0 4 で出力ポートを割り当てたら、ステップ 4 0 5 でキューの順番を 1 つ進め、ステップ 4 0 1 へ戻る。

上記の制御を行うことにより、共有メモリ部側の I / F ポート 2 1 1 を効率よく割り当てることが可能となり、高いスループットのデータ転送を実現できる。

【 0 0 2 6 】

また、図 9 に示すように、共有メモリ部 1 1 4 を各々物理的に独立した共有メモリ部 1 1 4 間で二重化して二重化領域 ( 1 6 0 ) を形成する。すなわち、2 個の共有メモリ部 1 1 4 を 2 重化した場合には各共有メモリ部に同一のデータが書き込まれる。また、各共有メモリ部全体を二重化したり、各共有メモリ部の一部分を二重化したりすることができる。そして、セレクト部 1 1 3 から二重化された共有メモリ部 1 1 4 両方への同時アクセス ( 2 重アクセス ) が発生するディスクアレイ制御装置 4 では、図 6 のステップ 4 0 2、4 0 3 においてキュー先頭の要求が 2 重アクセスかどうかを調べ、2 重アクセスの場合、要求される 2 つのポートが空いていればポートを割り当て、そうでなければステップ 4 0 6 へ進むという処理を行う。

30

これによって、共有メモリ部 1 1 4 に格納したデータの信頼性を向上させることが可能となる。

40

また、磁気ディスク装置 1 2 0 に記録するデータの転送時、共有メモリ部 1 1 4 との I / F ポート 2 1 1 を効率よく割り当てることが可能となり、高いスループットのデータ転送を実現できる。

【 0 0 2 7 】

《 実施例 2 》

図 1 に示すディスクアレイ制御装置の構成を、図 1 0 に示すように、共有メモリ部 1 1 4 を、磁気ディスク装置 1 2 0 に記録するデータを一時的に格納するキャッシュメモリ部 1 1 5 と、キャッシュメモリ部 1 1 5 及びディスクアレイ制御装置 5 に関する制御情報を格納する共有メモリ部 1 1 4 に物理的に分割し、キャッシュメモリ部 1 1 5 に繋がるセレクト部 ( C M セレクト部 ) 1 2 3 と、共有メモリ部に繋がるセレクト部 ( S M セレクト部 )

50

1 1 3を物理的に独立させた構成にする。

そして、チャンネルI/F部1 1 1、及びディスクI/F部1 1 2から、キャッシュメモリ部1 1 5、または共有メモリ部1 1 4へのアクセスパス0 1 3 5、アクセスパス1 1 3 6を物理的に独立させ、少なくともキャッシュメモリ部1 1 5に繋がるセクタ部(CMセクタ部)1 2 3において、実施例1で述べた調停を行う。これは、共有メモリ部に格納されるのは、キャッシュメモリ部1 1 5及びディスクアレイ制御装置5に関する制御情報であり、制御情報のデータ量は少ないので、ポートが使用中となる時間は小さく、すぐにポートは使用可能となるため、調停をしなくても格別の差し障りはないからである。

【0028】

また、図11に示すように、共有メモリ部1 1 4、及びキャッシュメモリ部1 1 5を各々物理的に独立した共有メモリ部1 1 4、及びキャッシュメモリ部1 1 5間でそれぞれ二重化して二重化領域(160)を形成し、少なくともキャッシュメモリに繋がるセクタ部(CMセクタ部)1 2 3から二重化されたキャッシュメモリ部1 1 5両方への同時アクセス(2重アクセス)が発生するディスクアレイ制御装置6では、図6のステップ402、403においてキュー先頭の要求が2重アクセスかどうかを調べ、2重アクセスの場合、要求される2つのポートが空いていればポートを割り当て、そうでなければステップ406へ進むという処理を、キャッシュメモリ部に繋がるセクタ部(CMセクタ部)1 2 3で行う。

これによって、共有メモリ部1 1 4に格納したデータの信頼性を向上させることが可能となる。

また、磁気ディスク装置120に記録するデータの転送時、キャッシュメモリ部1 1 5とのI/Fポート2 1 1を効率よく割り当てることが可能となり、高いスループットのデータ転送を実現できる。

【0029】

《実施例3》

図7は、チャンネルI/F部1 1 1またはディスクI/F部1 1 2内の共有メモリ(SM)アクセス回路から共有メモリ部1 1 4へ、またはチャンネルI/F部1 1 1またはディスクI/F部1 1 2内のキャッシュメモリ(CM)アクセス回路からキャッシュメモリ部1 1 5へデータの書き込みを行うときの処理の流れを示している。

データの書き込み時は、ステップ501でSMまたはCMアクセス回路からセクタ部1 1 3または1 2 3へアクセス要求(REQ)を出し、続けてステップ502、503でアドレス(ADR)、コマンド(CMD)を送出する。

ステップ504、505でセクタ部1 1 3または1 2 3において、アービトレーションを行い、セクタを切り替えて共有メモリ部1 1 4またはキャッシュメモリ部1 1 5へのポートを割り当てる。

ステップ506で、セクタ部1 1 3または1 2 3から共有メモリ部またはキャッシュメモリ部へアクセス要求(REQ)を出し、続けてステップ507、508でアドレス(ADR)、コマンド(CMD)を送出する。

【0030】

ステップ509で、共有メモリ部1 1 4またはキャッシュメモリ部1 1 5において、アクセスするメモリモジュールの選択を行い、選択後、ステップ510でセクタ部1 1 3または1 2 3を経由してSMまたはCMアクセス回路へアクセス承認(ACK ON)を返す。

SMまたはCMアクセス回路はACK ONを受けたら、ステップ511でデータを送出する。

共有メモリ部1 1 4またはキャッシュメモリ部1 1 5はデータを全て受け取ったら、ステップ512で後処理を行い、ステップ513でセクタ部1 1 3または1 2 3を経由してSMまたはCMアクセス回路へステータス(STATUS)を返す。

セクタ部1 1 3または1 2 3はSTATUSを受けたら、ステップ514で共有メモリ部1 1 4またはキャッシュメモリ部1 1 5へ、アクセス承認を取り下げるように指示を出

10

20

30

40

50

す(A C K O F F)。

S MまたはC Mアクセス回路はS T A T U Sを受けたら、ステップ5 1 5でセクタ部1 1 3または1 2 3に、アクセス承認を取り下げるように指示を出す(A C K O F F)。

【0 0 3 1】

図8は、チャンネルI / F部1 1 1またはディスクI / F部1 1 2内の共有メモリアクセス回路へ共有メモリ部1 1 4から、またはチャンネルI / F部1 1 1またはディスクI / F部1 1 2内のキャッシュメモリアクセス回路へキャッシュメモリ部1 1 5からデータの読み出しを行うときの処理の流れを示している。

データの読み出し時の処理ステップ6 0 1 ~ 6 1 0までは、データの書き込み時の処理ステップ5 0 1 ~ 5 1 0までと同じである。

10

【0 0 3 2】

その後、共有メモリ部1 1 4またはキャッシュメモリ部1 1 5において、ステップ6 1 1で読み出しの前処理を行う。

ステップ6 1 2で、セクタ部1 1 3または1 2 3を経由してS MまたはC Mアクセス回路へデータを送出する。

データの送出自ら終了したら、共有メモリ部1 1 4またはキャッシュメモリ部1 1 5において、ステップ6 1 3で後処理を行い、ステップ6 1 4でセクタ部1 1 3または1 2 3を経由してS MまたはC Mアクセス回路へS T A T U Sを返す。

セクタ部1 1 3または1 2 3はS T A T U Sを受けたら、ステップ6 1 5で共有メモリ部1 1 4またはキャッシュメモリ部1 1 5へ、アクセス承認を取り下げないように指示を出す(A C K O F F)。

20

S MまたはC Mアクセス回路はS T A T U Sを受けたら、ステップ6 1 6でセクタ部1 1 3または1 2 3へ、アクセス承認を取り下げないように指示を出す(A C K O F F)。

【0 0 3 3】

上記のように、チャンネルI / F部1 1 1、またはディスクI / F部1 1 2から、共有メモリ部1 1 4またはキャッシュメモリ部1 1 5へのアクセスの際に、最初にアドレスとコマンドを連続して送出し、共有メモリ部1 1 4、またはキャッシュメモリ部1 1 5までのアクセスパスが確立した(ステップ5 1 0または6 1 0)後データを送出することによって、セクタ部1 1 3または1 2 3において、転送データをバッファリングする必要がなくなり、セクタ部1 1 3または1 2 3での制御が簡単化され、メモリへのアクセススループットの向上が可能となる。

30

【0 0 3 4】

【発明の効果】

本発明によれば、ホストコンピュータとのインターフェース部、または複数の磁気ディスク装置とのインタフェース部と共有メモリ部の間にあるセクタ部において、ホストコンピュータとのインターフェース部、または複数の磁気ディスク装置とのインタフェース部から共有メモリ部へのアクセス要求を、共有メモリ部へのアクセスパスへ効率よく振り分けることが可能となる。

それによって、ディスクアレイ制御装置のデータ転送のスループットを向上することができる。

40

【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である。

【図2】従来のディスクアレイ制御装置の構成を示す図である。

【図3】従来のディスクアレイ制御装置の構成を示す図である。

【図4】本発明によるディスクアレイ制御装置内のセクタ部の構成を示す図である。

【図5】セクタ部での動作フローを示す図である。

【図6】セクタ部内のアービトレーション部での動作フローを示す図である。

【図7】共有メモリ部またはキャッシュメモリ部へデータの書き込みを行うときのシーケンスを示す図である。

【図8】共有メモリ部またはキャッシュメモリ部からデータの読み出しを行うときのシー

50

ケースを示す図である。

【図9】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である。

【図10】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である。

【図11】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である。

【図12】本発明によるディスクアレイ制御装置内のセクタ部の部分の詳細構成を示す図である。

【図13】チャンネルI/F部の構成を示す図である。

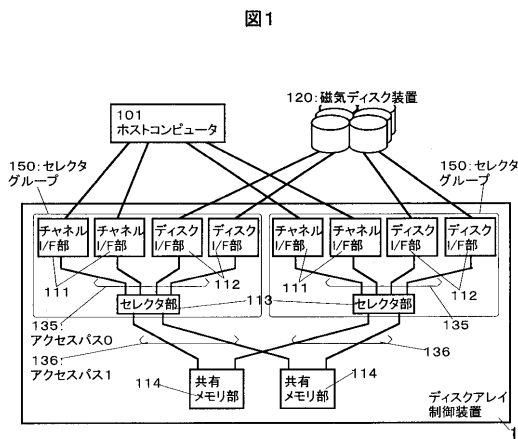
【符号の説明】

- 1、4、5、6 ディスクアレイ制御装置
- 50 マイクロプロセッサ
- 51 ホストI/F
- 52 共有メモリアクセス回路
- 101 ホストコンピュータ
- 111 チャンネルI/F部
- 112 ディスクI/F部
- 113 セクタ部(SMセクタ部)
- 114 共有メモリ部
- 115 キャッシュメモリ部
- 120 磁気ディスク装置
- 123 CMセクタ部
- 135 アクセスパス0
- 136 アクセスパス1
- 150 セクタグループ

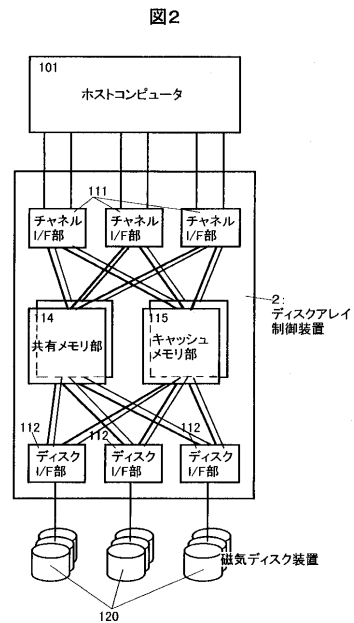
10

20

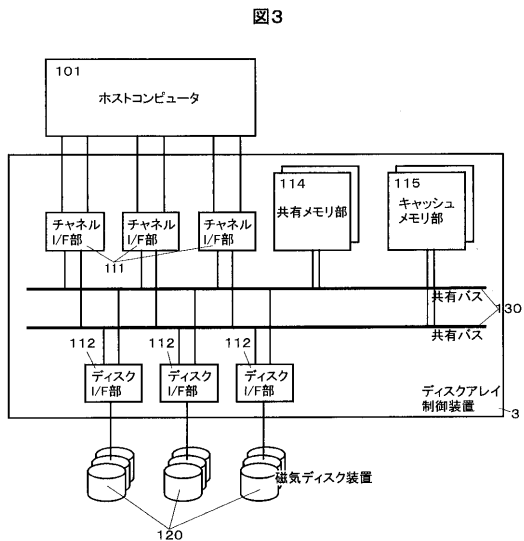
【図1】



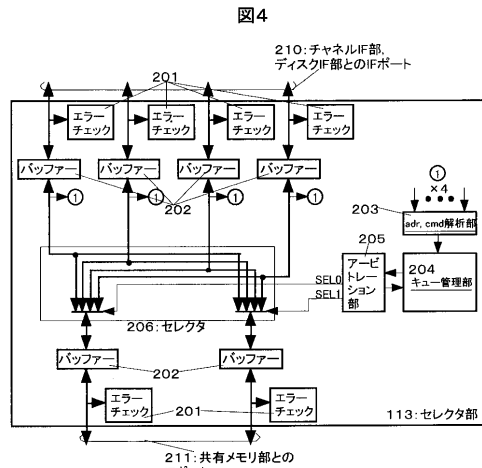
【図2】



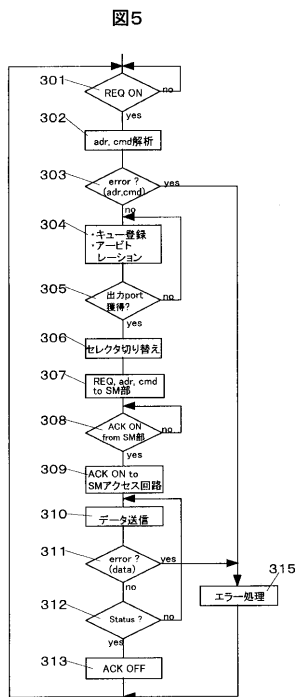
【 図 3 】



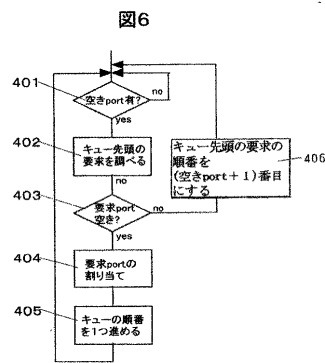
【 図 4 】



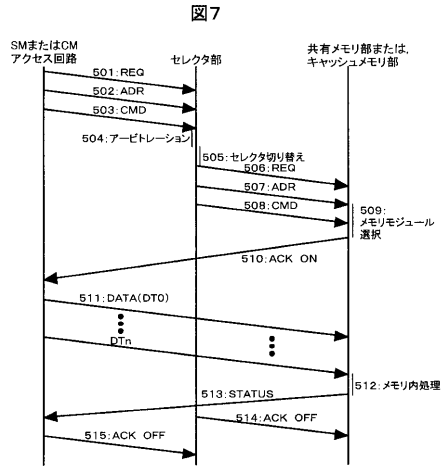
【 図 5 】



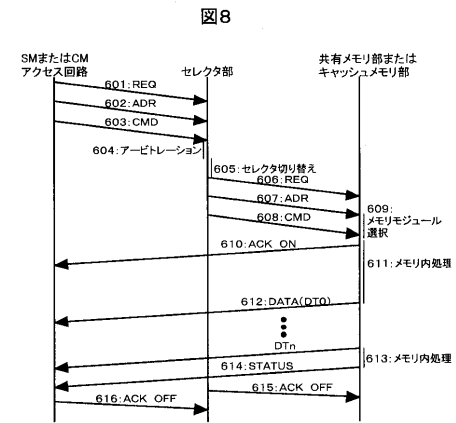
【 図 6 】



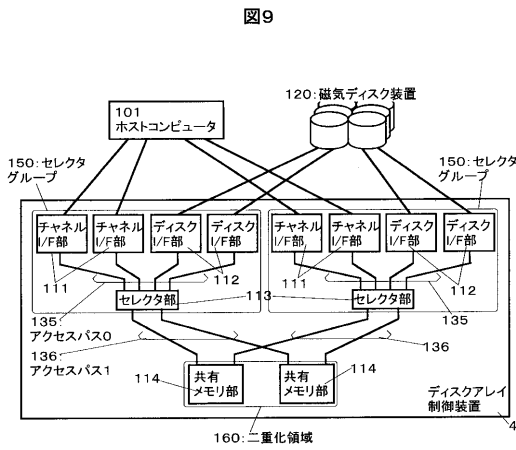
【 図 7 】



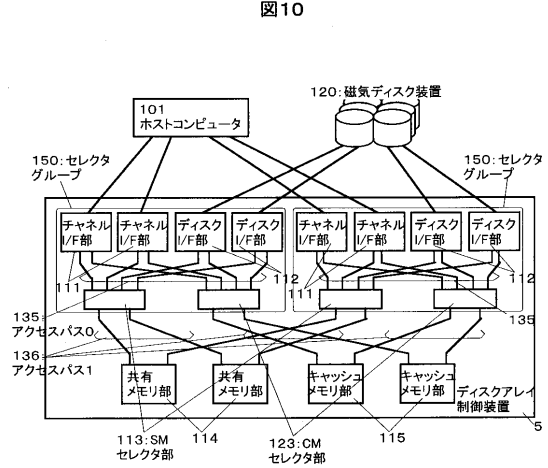
【 図 8 】



【 図 9 】

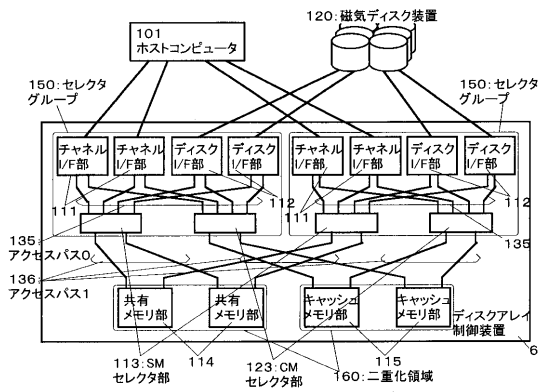


【 図 10 】



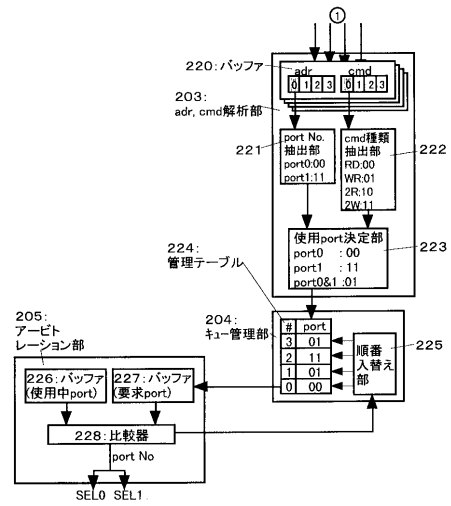
【 図 1 1 】

図 11



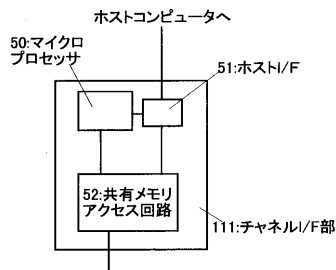
【 図 1 2 】

図 12



【 図 1 3 】

図 13



---

フロントページの続き

- (56)参考文献 特開平05 - 143242 (JP, A)  
特開平06 - 012359 (JP, A)  
特開平08 - 180007 (JP, A)  
特開平07 - 020994 (JP, A)  
特開平09 - 198308 (JP, A)  
特開平8 - 328760 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 3/06-3/08  
G06F 13/10-13/14  
G06F 12/00  
G06F 15/16-15/177