(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3781358号
(P3781358)
(45)発行日　平成18年5月31日(2006.5.31)　(24)登録日　平成18年3月17日(2006.3.17)

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

請求項の数 4　(全 9 頁)

(21)出願番号　特願2002-10958(P2002-10958)
(22)出願日　平成14年1月21日(2002.1.21)
(62)分割の表示　特願平8-2050の分割
原出願日　平成8年1月10日(1996.1.10)
(65)公開番号　特開2002-287908(P2002-287908A)
(43)公開日　平成14年10月4日(2002.10.4)
審査請求日　平成14年12月4日(2002.12.4)

(73)特許権者　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号
(74)代理人　100079108
弁理士　稲葉　良幸
(74)代理人　100093861
弁理士　大賀　眞司
(74)代理人　100075096
弁理士　作田　康夫
(72)発明者　小澤　匡二
神奈川県小田原市国府津２８８０番地　株式会社日立製作所　ストレージシステム事業部内

最終頁に続く

(54)【発明の名称】外部記憶制御装置及び外部記憶制御装置間データ転送方法

(57)【特許請求の範囲】

【請求項１】

複数の第一の外部記憶装置に接続され、該第一の外部記憶装置と上位装置とのデータ転送を制御する外部記憶制御装置であって、

複数の第二の外部記憶装置を有する他の外部記憶制御装置に接続され、前記上位装置からの書き込み命令により複数のライトデータが連続して格納されるメモリ手段と、

前記メモリ手段に格納された前記複数のライトデータを前記他の外部記憶制御装置に転送する手段と、

を備え、

前記第一の外部記憶装置と前記第二の外部記憶装置とに同一のデータが格納される場合に、前記上位装置から前記第一の外部記憶装置内のデータに対する前記書き込み命令を受け取ったとき、

前記書き込み命令で受け取った前記複数のライトデータを前記他の外部記憶制御装置へそのまま転送するよりも、前記書き込み命令で受け取った前記複数のライトデータを含む複数のライトレコードを、一又は複数のトラック分として前記他の外部記憶制御装置へまとめて転送する方が、転送処理時間が早い場合には、前記複数のライトレコードを前記他の外部記憶制御装置へまとめて転送し、転送処理時間が遅い場合には、前記書き込み命令で受け取った前記複数のライトデータをそのまま転送し、

前記他の外部記憶制御装置から前記書き込み命令に対するステータス報告を受けた後に、前記上位装置にステータス報告を送信する

ことを特徴とする外部記憶制御装置。
【請求項２】
請求項１記載の外部記憶制御装置において、
前記メモリ手段に格納されるライトデータは、前記外部記憶装置に対して書き込まれるものである
ことを特徴とする外部記憶制御装置。
【請求項３】
複数の第一の外部記憶装置に接続され、該第一の外部記憶装置と上位装置とのデータ転送を制御する外部記憶制御装置であって、
複数の第二の外部記憶装置を有する他の外部記憶制御装置に接続され、前記上位装置からの書き込み命令により複数のライトデータが連続して格納されるメモリ手段と、
前記メモリ手段に格納された前記複数のライトデータを前記他の外部記憶制御装置に転送する手段と、
を備え、
前記第一の外部記憶装置と前記第二の外部記憶装置とに同一のデータが格納される場合に、前記上位装置から前記第一の外部記憶装置内のデータに対する前記書き込み命令を受け取ったとき、
前記書き込み命令で受け取った前記複数のライトデータを前記他の外部記憶制御装置へそのまま転送するよりも、前記書き込み命令で受け取った前記複数のライトデータを含む複数のライトレコードを、一又は複数のトラック分として前記他の外部記憶制御装置へまとめて転送する方が、転送処理時間が早い場合には、前記複数のライトレコードを前記他の外部記憶制御装置へまとめて転送し、転送処理時間が遅い場合には、前記書き込み命令で受け取った前記複数のライトデータをそのまま転送し、
前記他の外部記憶制御装置から前記書き込み命令に対するステータス報告を受けた後に、前記上位装置にステータス報告を送信する
ことを特徴とする記憶システム。
【請求項４】
請求項３記載の記憶システムにおいて、
前記メモリ手段に格納されるライトデータは、前記外部記憶装置に対して書き込まれるものである
ことを特徴とする記憶システム。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、上位装置からの書き込み命令によりデータを記憶する外部記憶サブシステムに関わり、特に同一データを別の外部記憶サブシステムに転送する外部記憶サブシステムに関する。
【０００２】
【従来の技術】
上位装置で使用するデータを複数の外部記憶装置に保持する場合、同一データを保持する正副の外部記憶装置を異なる外部記憶制御装置配下に設けることがある。この場合、それらの外部記憶制御装置を接続し、正外部記憶装置を制御する外部記憶制御装置でライトコマンドを受領したときは、正側の外部記憶制御装置が副外部記憶装置を制御する外部記憶制御装置にこのライトコマンドを発行することで、正副外部記憶装置のデータの一致を図っている。
【０００３】
米国特許公報第５１５５８４５号には、正外部記憶装置を制御する外部記憶制御装置と副外部記憶装置を制御する外部記憶制御装置とを接続し、上位装置からのライトコマンドを受領した正側の外部記憶制御装置が副側の外部記憶制御装置に対してデータ転送を行い、正副外部記憶装置へのライト処理を並行に行うことが記載されている。

【０００４】
【発明が解決しようとする課題】
外部記憶装置のデータが、大型計算機システムで使用されているＣＫＤ（カウント、キー、データ）形式であると上位装置からみなされる場合、個々のレコード毎にそのデータ転送を指示するチャネルコマンドが連続して発行される。
【０００５】
ここで、この一回のコマンドチェインで複数レコードが連続して書き込まれる場合を考える。異なる外部記憶制御装置の配下に同一データを保持する正副外部記憶装置を有する外部記憶サブシステムにおいて、上位装置であるチャネル装置より正側の外部記憶制御装置であるディスク装置に対してＲ１、Ｒ２およびＲ３の連続した３個のレコードをライトするコマンドチェイン（ＤＥＦＩＮＥ　ＥＸＥＴＥＮＴ／ＬＯＤＡＴＥ　ＲＥＣＯＲＤ／ＷＲＩＴＥ（Ｒ１)／ＷＲＩＴＥ（Ｒ２）／ＷＲＩＴＥ（Ｒ３））が発行された場合、チャネル装置と正ディスク制御装置、および正ディスク制御装置とのデータのやり取りは図３の処理シーケンスの様に、書き込みレコードのデータ転送毎に正副ディスク制御装置間において数回のコマンドチェインが発生する。
【０００６】
また、このようなシステム構成をとってデータを二重化する場合、災害時にバックアップシステムへ切り替えるという観点から正副デ外部記憶制御装置間の距離は大きくなり、インタフェースには光ファイバが用いられる。このため、メタルケーブルではほぼ固定と考えられているコマンドチェイニング時間もインタフェースのケーブル長の影響が無視できない状態にある。さらに、上記のデータの二重化はバックアップ機能であるため副側外部記憶制御装置への書き込み処理による通常処理への影響は最小限に押さえる必要がある。
【０００７】
しかし、上記の様に１回のコマンドチェインにおいて複数回のライトが実行される場合、正外部記憶制御装置と副外部記憶制御装置との間でのコマンドチェインの回数が増えるため、副外部記憶制御装置への書き込み処理に要するコマンドチェイニング時間が無視できなくなり、バックアップ処理が通常処理のスループットを大幅に増大させることがある。
【０００８】
本発明の目的は、正副外部記憶制御装置間のコマンドチェイニング時間を考慮した副外部記憶制御装置への書き込み処理時間の最適化を行うことで、上述した状況下においても当初の目的を達成することのできる手段を提供することにある。
【０００９】
【課題を解決するための手段】
上記の目的を達成するため、本発明による外部記憶制御装置は、複数の第一の外部記憶装置に接続され、該第一の外部記憶装置と上位装置とのデータ転送を制御する外部記憶制御装置であって、複数の第二の外部記憶装置を有する他の外部記憶制御装置に接続され、前記上位装置からの書き込み命令により複数のライトデータが連続して格納されるメモリ手段と、前記メモリ手段に格納された前記複数のライトデータを前記他の外部記憶制御装置に転送する手段と、を備え、前記第一の外部記憶装置と前記第二の外部記憶装置とに同一のデータが格納される場合に、前記上位装置から前記第一の外部記憶装置内のデータに対する前記書き込み命令を受け取ったとき、前記書き込み命令で受け取った前記複数のライトデータを前記他の外部記憶制御装置へそのまま転送するよりも、前記書き込み命令で受け取った前記複数のライトデータを含む複数のライトレコードを、一又は複数のトラック分として前記他の外部記憶制御装置へまとめて転送する方が、転送処理時間が早い場合には、前記複数のライトレコードを前記他の外部記憶制御装置へまとめて転送し、転送処理時間が遅い場合には、前記書き込み命令で受け取った前記複数のライトデータをそのまま転送し、前記他の外部記憶制御装置から前記書き込み命令に対するステータス報告を受けた後に、前記上位装置にステータス報告を送信することを特徴とする。
正副外部記憶制御装置間のコマンドチェイニング時間を予測し、副外部記憶制御装置への書き込み処理を開始する以前に上記予測値および上位装置から正外部記憶制御装置に発

行されるコマンドから副外部記憶制御装置への書き込み処理時間を予測することができる。また、上記予測処理時間と副外部記憶制御装置に対して専用コマンドでライトレコード分のデータ、あるいはライトレコードを含む一又は複数のトラック分のデータを一回で転送した場合のデータ処理時間を比較し、副外部記憶制御装置に対して最適なコマンドチェインを選択し、該データをまとめて書き込む命令を有する。

本発明による記憶システムは、複数の第一の外部記憶装置に接続され、該第一の外部記憶装置と上位装置とのデータ転送を制御する外部記憶制御装置であって、複数の第二の外部記憶装置を有する他の外部記憶制御装置に接続され、前記上位装置からの書き込み命令により複数のライトデータが連続して格納されるメモリ手段と、前記メモリ手段に格納された前記複数のライトデータを前記他の外部記憶制御装置に転送する手段と、を備え、前記第一の外部記憶装置と前記第二の外部記憶装置とに同一のデータが格納される場合に、前記上位装置から前記第一の外部記憶装置内のデータに対する前記書き込み命令を受け取ったとき、前記書き込み命令で受け取った前記複数のライトデータを前記他の外部記憶制御装置へそのまま転送するよりも、前記書き込み命令で受け取った前記複数のライトデータを含む複数のライトレコードを、一又は複数のトラック分として前記他の外部記憶制御装置へまとめて転送する方が、転送処理時間が早い場合には、前記複数のライトレコードを前記他の外部記憶制御装置へまとめて転送し、転送処理時間が遅い場合には、前記書き込み命令で受け取った前記複数のライトデータをそのまま転送し、前記他の外部記憶制御装置から前記書き込み命令に対するステータス報告を受けた後に、前記上位装置にステータス報告を送信することを特徴とする。

正副外部記憶制御装置間のコマンドチェイニング時間を予測し、副外部記憶制御装置への書き込み処理を開始する以前に上記予測値および上位装置から正外部記憶制御装置に発行されるコマンドから副外部記憶制御装置への書き込み処理時間を予測することができる。また、上記予測処理時間と副外部記憶制御装置に対して専用コマンドでライトレコード分のデータ、あるいはライトレコードを含む一又は複数のトラック分のデータを一回で転送した場合のデータ処理時間を比較し、副外部記憶制御装置に対して最適なコマンドチェインを選択し、該データをまとめて書き込む命令を有する。

【００１０】

上記のコマンドチェイニング時間は正副外部記憶制御装置間で発行される特定のコマンドから次のコマンドへのコマンドチェイニング時間を計測する手段か、正副外部記憶制御装置間のインタフェースケーブル長をあらかじめ外部より設定し、このインタフェースケーブル長からコマンドチェイニング時間を算出する手段によって実現される。

【００１１】

一方、副外部記憶制御装置への書き込み処理時間は上記情報および正外部記憶制御装置に発行されるコマンド情報から算出可能である。即ち、データ入出力に使用されるコマンドチェインは、データ転送を開始するコマンドの前に、特定かつ少なくとも２個のチャネルコマンドを含む。例えば、外部記憶制御装置の場合には、ＤＥＦＩＮＥ ＥＸＴＥＮＴコマンドとＬＯＣＡＴＥ ＲＥＣＯＲＤコマンドがデータ転送前に発行され、ＬＯＣＡＴＥ ＲＥＣＯＲＤコマンドにより、処理レコード数とデータ長が与えられるため、副外部記憶制御装置へのデータ転送量を算出することが可能である。

【００１２】

また、正副外部記憶制御装置間のコマンドチェイニング時間はインタフェースケーブル長と外部記憶制御装置の処理能力には依存するが、コマンドチェイニングの前後のコマンドにはほとんど依存しない。従って２個のコマンドチェイニング時間を計測することによってデータ転送中のコマンドチェイニング時間を予測することができる。尚、バックアップシステムの設置においては正副外部記憶制御装置間のインタフェースケーブル長、および当該外部記憶制御装置のサポート機能は既知のものであり、システム設置時に該情報を正側の外部記憶制御装置に設定しておくことも可能であるので設定されたインタフェースケーブル長からコマンドチェイニング時間を予測することもできる。

【００１３】

そして、コマンドチェイニング時間が予測可能であれば上記データ転送量からチャネルからのコマンドチェインを副側の外部記憶制御装置に発行した場合の処理時間が予測可能となるため、ライトレコード分のデータ、あるいはライトレコードを含む一又は複数のトラック分のデータをまとめて書き込む命令により１回で転送した場合の処理時間とを比較して、最適なコマンドチェインを副側の外部記憶制御装置に対して発行することができる。

【００１４】

また、本発明による外部記憶制御装置は、書き込み命令に対して、外部記憶制御装置内データバッファから書き込み命令を受領する外部記憶制御装置へのデータを転送するために、対象データが別の外部記憶制御装置のデータバッファ内に存在することを確認する手段、および上記のデータを転送する手段を備える。さらに、データ転送において制御バイト、チェックバイトを含んだデータバッファ内の物理データをまとめて転送することでさらに転送処理時間の短縮を図ることができる。

【００１５】

加えて、上位装置からの読み出し命令に対して上位装置からの読み出し対象データが読み出し命令を受領した外部記憶制御装置以外の外部記憶装置の外部記憶制御装置バッファ内に存在していた場合、このデータを上記の転送方式により読み出し命令受領した外部記憶制御装置に転送することにより外部記憶装置をアクセスすることなく読み出し処理時間の短縮を図ることができる。

【００１６】

【発明の実施の形態】

以下、本発明の一実施例を示す。図１は本発明の一実施例におけるデータ処理システムの構成図である。このデータ処理システムは、主記憶装置１、チャネル装置２、およびバックアップシステムである主記憶装置６、チャネル装置７とを含む。さらに、バッファメモリ４を有する正ディスク制御装置３と正ディスク装置５およびバッファメモリ９を内蔵した副ディスク制御装置８と副ディスク装置１０を含む。チャネル装置２と正ディスク制御装置３はインタフェースケーブル１１によって、正ディスク制御装置３と副ディスク制御装置８とはインタフェースケーブル１２によって接続されている。

【００１７】

正ディスク制御装置３は図２に示す様に、バッファメモリ４の他に、チャネルコマンド解析部３１、副ディスク制御装置発行コマンド判断部３２、副ディスク制御装置管理情報テーブル３３、タイマ３４、副ディスク制御装置コマンド発行部３５、ケーブル長・副ディスク制御装置機能判断部３６とから構成される。３１～３３、３５～３６はディスク制御装置３が内蔵するマイクロプロセッサが実行するマイクロプログラムによって実現される機能ブロックである。タイマ３４は一定周期でカウントアップされるハードウェアのカウンタである。また、副ディスク制御装置８も同様の構成を有している。

【００１８】

この正ディスク制御装置３において、タイマ３４、副ディスク制御装置コマンド発行部３５、およびケーブル長・副ディスク制御装置機能判断部３６により、正副ディスク制御装置間のコマンドチェイニング時間の計測をおこなう。そして、正副ディスク装置のペア確立時に発行されるコマンド処理時にタイマ３４にて計測されたコマンドチェイニング時間Ｔｓｇが正ディスク制御装置内の管理テーブル３３に格納される。

【００１９】

また、正ディスク制御装置は副ディスク制御装置のサポート機能レベルを正副ディスク装置のペア確立時に取得し、同様に正ディスク制御装置内の管理テーブル３３に格納する。この情報は後の副ディスク制御装置に対するライトコマンド発行時、副ディスク制御装置に対してライトレコード分のデータ、あるいはライトレコードを含む一又は複数のトラック分のデータをまとめて書き込む命令を受領可能か否かの判断に用いる。

【００２０】

サポート機能レベル情報により副ディスク制御装置発行コマンド判断部３２が専用コマン

ドを受領不可能と判断した場合は、チャネル装置２よりライトコマンドを受領したときに、副ディスク制御装置コマンド発行部３５より無条件にチャネル装置からのコマンドチェインが副ディスク制御装置に対して発行される。そのため、この機能をサポートしていない従来機種との接続も可能となる。

【００２１】

チャネル装置２より正ディスク制御装置３にライトコマンドを発行したとき、チャネルコマンド解析部３１ではこのライトコマンドが正副ペアを形成するディスク装置に対するものであるかの判断を行う。

【００２２】

そして、正副ペア形成ディスク装置に対するものであった場合、副ディスク制御装置管理情報テーブル３３からの情報、およびチャネルコマンド解析部３１にて解析されたＬＯＣＡＴＥコマンドの転送レコード数Ｎ、平均レコード長Ｌから、チャネル装置から、受け取ったコマンドチェインをそのまま副ディスク制御装置に発行した場合と、ライトレコードを含む１トラック分をまとめて転送した場合の比較グラフを作成する（図６参照）。ここで、直線Ｉは既知となった正副ディスク制御装置間のコマンドチェイニング時間Ｔｓｇ、上記解析値、および正副ディスク制御装置３、８間のデータ転送速度Ｖから以下の式で表すことができる。

【００２３】

Ｔｓ＝ＮＬ／Ｖ＋（Ｎ－１）Ｔｓｇ

直線ＩＩは該レコード長Ｌの１トラック分のレコード数Ｎａ分をまとめて転送した場合の転送時間Ｔｓ’（一定値）を表したものである。

【００２４】

Ｔｓ’＝ＮａＬ／Ｖ＋Ｔｓｇ

副ディスク制御装置発行コマンド判断部３２ではこの比較グラフと該コマンドチェインにて転送されるレコード数Ｎからチャネル装置から受け取ったコマンドチェインをそのまま副ディスク制御装置に発行するか、１トラック分をまとめて転送するかを判断する。

【００２５】

図４、５、６に示す８レコード／トラックのトラックに対して６レコードライトするケースを想定した場合、該グラフは直線Ｉ、ＩＩの接点を含むレコード数（４レコード）を境に１トラックのデータをまとめて転送した方がコマンドチェイン回数が減ることにより処理時間が少なくてすむことが予測できる。

【００２６】

副ディスク制御装置コマンド発行部３５では上記副ディスク制御装置発行コマンド判断部３２の指示に従って副ディスク制御装置８に対して最適化されたコマンドチェインを発行する。

【００２７】

【発明の効果】

本発明によれば、正側のディスク制御装置と副側のディスク制御装置間のデータ転送処理において、チャネル装置から受け取ったコマンドチェインをそのまま副側のディスク制御装置に発行する場合と、ライトレコード分のデータ、あるいはライトレコードを含む一又は複数のトラック分のデータをまとめて書き込む命令により１回で転送する場合とを比較し、処理時間が最小となるように副側のディスク制御装置に発行するコマンドチェインを選択することにより、バックアップ処理である副側のディスク装置への書き込み処理時間を最小限に押さえることが可能である。

【図面の簡単な説明】

【図１】本発明の実施例におけるデータ処理システムの構成図である。

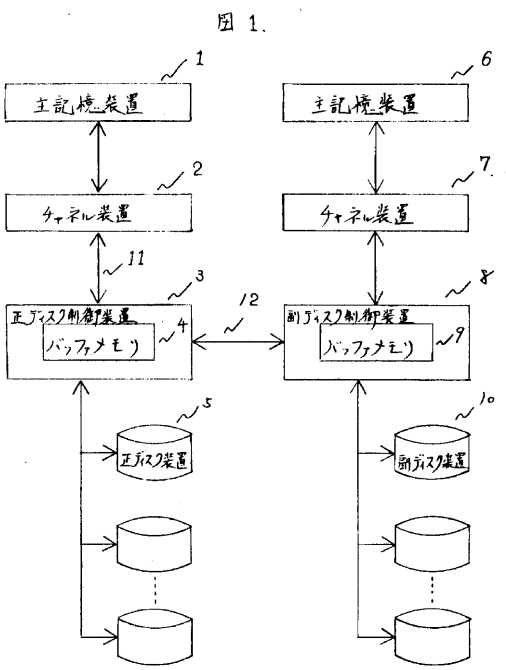
【図２】本発明の実施例におけるディスク制御装置の内部構成図である。

【図３】上位装置であるチャネル装置、正ディスク制御装置、および副ディスク制御装置間のデータのやり取りを示す図である。

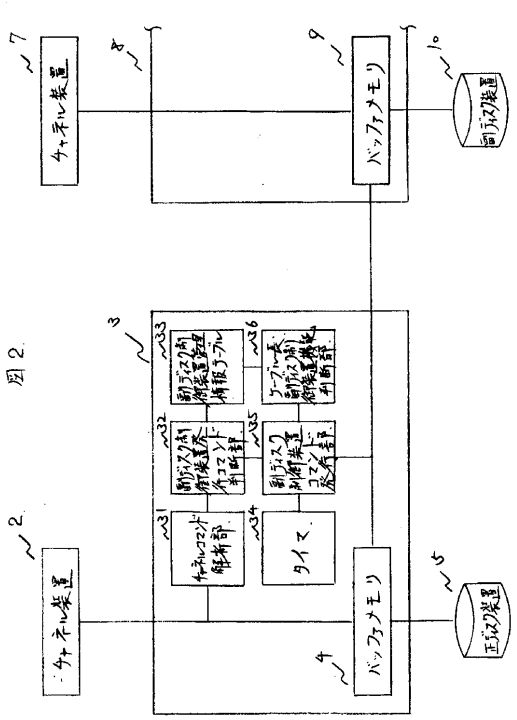
【図４】本発明の実施例において、トラック単位で正副ディスク制御装置間のデータ転送

を行った場合を示す図である。
【図５】本発明の実施例におけるディスク制御装置内データバッファ格納されている８レコード／１トラックのデータを示す図である。
【図６】本発明の実施例における転送レコード数と処理時間の関係を表した図である。
【符号の説明】
３　　正ディスク制御装置
８　　副ディスク制御装置
１２　正副ディスク制御装置間インタフェースケーブル
３２　副ディスク制御装置発行コマンド判断部
３３　副ディスク制御装置管理情報テーブル
３４　タイマ
３５　副ディスク制御情報コマンド発行部
３６　ケーブル長・副ディスク制御装置機能判断部

【図１】

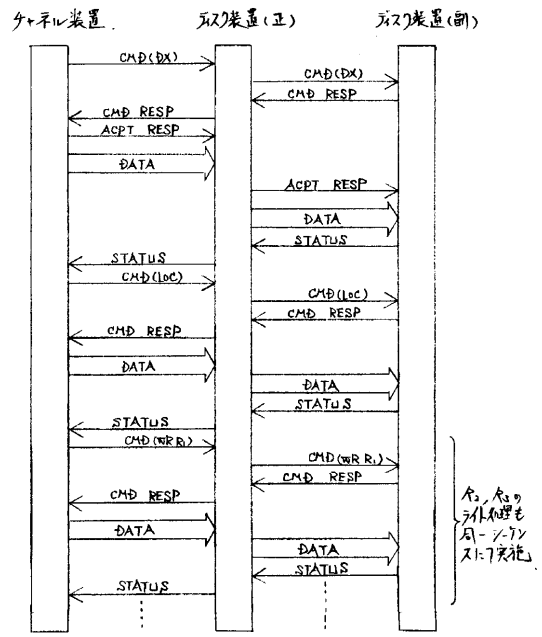


【図２】



【図３】

図３



【図４】

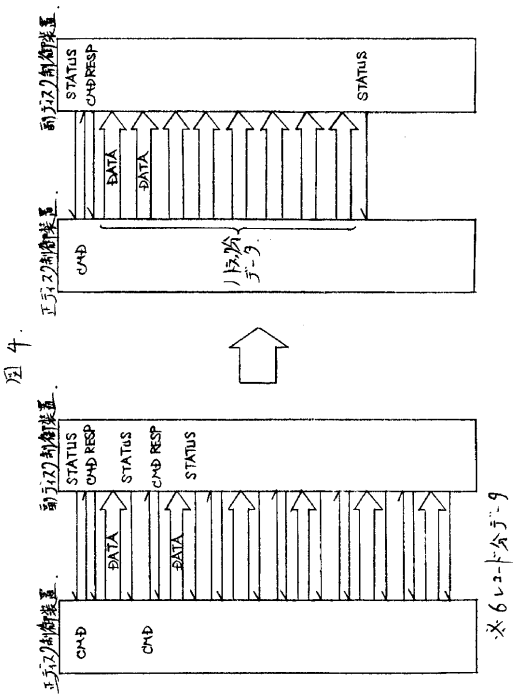
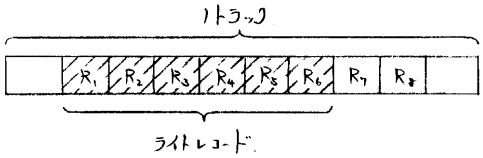


図４

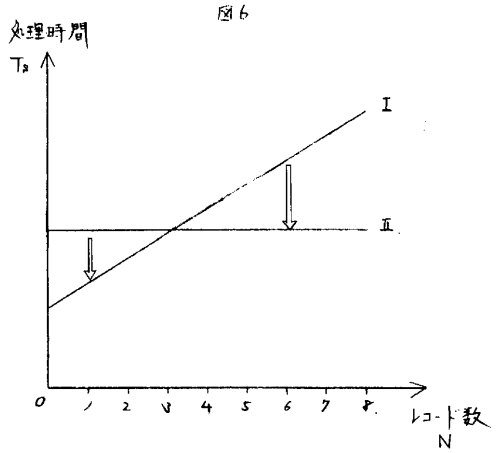
【図５】

図５



【図６】

図６



フロントページの続き

(72)発明者 佐野 一英
神奈川県小田原市国府津２８８０番地 株式会社日立製作所 ストレージシステム事業部内
(72)発明者 小出 雄
神奈川県小田原市国府津２８８０番地 株式会社日立製作所 ストレージシステム事業部内
(72)発明者 中村 勝憲
神奈川県小田原市国府津２８８０番地 株式会社日立製作所 ストレージシステム事業部内

審査官 木村 貴俊

(56)参考文献 特開平０７－２０１１３２（ＪＰ，Ａ）
特開平０６－０６７８１１（ＪＰ，Ａ）
特開平０７－２７１５２１（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F 3/06- 3/08、
12/00-12/16、
13/10-13/14