(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4076724号
(P4076724)
(45)発行日　平成20年4月16日(2008.4.16)　(24)登録日　平成20年2月8日(2008.2.8)

(51)Int.Cl.　F I
H04L 12/42 (2006.01)　H04L 12/42　D

請求項の数 8　(全 27 頁)

(21)出願番号	特願2000-532850 (P2000-532850)
(86)(22)出願日	平成11年2月24日 (1999.2.24)
(65)公表番号	特表2002-504765 (P2002-504765A)
(43)公表日	平成14年2月12日 (2002.2.12)
(86)国際出願番号	PCT/US1999/003955
(87)国際公開番号	WO1999/043002
(87)国際公開日	平成11年8月26日 (1999.8.26)
審査請求日	平成17年3月15日 (2005.3.15)
(31)優先権主張番号	60/075,797
(32)優先日	平成10年2月24日 (1998.2.24)
(33)優先権主張国	米国(US)

(73)特許権者　500373758
シーゲイト　テクノロジー　エルエルシー
アメリカ合衆国，カリフォルニア，スコッツ　バレイ，ピー．オー．ボックス　６６３６０，　ディスク　ドライブ　９２０
(74)代理人　100066692
弁理士　浅村　皓
(74)代理人　100072040
弁理士　浅村　肇
(74)代理人　100107467
弁理士　員見　正文
(74)代理人　100094673
弁理士　林　鈺三

最終頁に続く

(54)【発明の名称】ダイナミック半二重によるループ・フェアネスの保持

(57)【特許請求の範囲】
【請求項１】
ループ・フェアネスを保持する通信チャネル・システムであって、
ファイバ・チャネル・アービトレーテッド・ループ・シリアル通信チャネルに取り付けられた第１のポートを有する第１のチャネル・ノードであって、前記第１のポートが、該ポートが取り付けられた通信チャネルの制御のためにアービトレーションする、第１のチャネル・ノードと、
前記第１のポートに作動的に結合されたフェアネス保持装置と、
を具備し、
前記第１のポートが前記通信チャネルのループの制御のためにアービトレーションし、制御が達成された後に、前記フェアネス保持装置が、所定量の使用が第１のポートと前記通信チャネルとの間で発生したか否かに少なくとも一部基づいて前記通信チャネルの制御を解放させ、前記所定量の使用が送信される第１の所定量のデータを有し、
送信されているデータの量を監視する第１のカウンタと、
前記第１のカウンタと作動的に結合し、前記第１のカウンタによってモニターされるデータの量が前記第１の所定量に達したか否かに少なくとも一部基づいて通信チャネルを解放するための制御を生ずる第１のコンパレータ回路と、
送信されるべく残っているデータの量を監視する第２のカウンタと、
前記第２のカウンタによって監視されるデータの量が第２の所定量のデータより少ないか否かに少なくとも一部基づいて、前記通信チャネルの制御を解放することを禁止する前

記第２のカウンタに作動的に結合する第２のコンパレータ回路と、
を備える前記通信チャネル・システム。
【請求項２】
前記第１のチャネル・ノードが、ダイナミック半二重をサポートし、
前記第１のカウンタおよび前記第１のコンパレータが、ダイナミック半二重コマンドの第１の部分による受信時に初期化される、
請求項１記載のシステム。
【請求項３】
前記第１の所定量のデータおよび前記第２の所定量のデータがプログラム可能な量である、請求項２記載のシステム。
【請求項４】
ファイバ・チャネル・アービトレーテッド・ループ・シリアル通信チャネルと、
前記第１のチャネル・ノードに作動的に結合された磁気ディスク記憶ドライブと、
第２のチャネル・ノードを有するコンピュータ・システムと、
をさらに具備し、
前記第２のチャネル・ノードが、前記ファイバ・チャネル・アービトレーテッド・ループ・シリアル通信チャネルを介して前記第１および第２のチャネル・ノード間でデータを転送するために、ファイバ・チャネル・ループ内の前記第１のチャネル・ノードに作動的に結合されている、
請求項１記載のシステム。
【請求項５】
（ａ）ファイバ・チャネル・アービトレーテッド・ループ・シリアル通信チャネルのループの制御のためにアービトレーションするステップと、
（ｂ）所定量の使用が第１のポートと前記通信チャネルとの間で発生したか否かに少なくとも一部基づいて前記通信チャネルの制御を解放させるステップであって、前記解放させるステップ（ｂ）がさらに、
（ｂ）（ｉ）第１の所定量のデータが送信されているか否かを判断するステップと、
（ｂ）（ｉｉ）前記判断するステップ（ｂ）（ｉ）に基づいてループの制御を解放するステップと、
（ｂ）（ｉｉｉ）第２の所定量のデータが転送されるべく残っているか否かを判断するステップと、
（ｂ）（ｉｖ）前記判断するステップ（ｂ）（ｉｉｉ）に基づいてループの制御を解放するステップと、
を含む、通信方法。
【請求項６】
前記判断するステップ（ｂ）（ｉｉｉ）が、
（ｂ）（ｉｉｉ）（Ａ）第１の値を得るために転送されたフレームの数を監視するステップと、
（ｂ）（ｉｉｉ）（Ｂ）前記第１の値を前記第１の所定量のデータと比較するステップと、
をさらに含む、請求項５記載の方法。
【請求項７】
（ｃ）ダイナミック半二重コマンドを受信するステップと、
（ｄ）該ダイナミック半二重コマンドの受信時に前記判断するステップ（ｂ）（ｉｉｉ）を初期化するステップと、
をさらに含む、請求項５記載の方法。
【請求項８】
前記初期化するステップ（ｄ）が、
（ｄ）（ｉ）前記第１の所定量のデータおよび前記第２の所定量のデータをプログラム可能にセットするステップをさらに含む、

請求項５記載の方法。
【発明の詳細な説明】
【０００１】
本出願は、１９９８年2月２４日出願の米国仮出願第６０／０７５，７９７号の３５Ｕ．Ｓ．Ｃ．１１９（ｅ）に基づく恩典を請求する。
（発明の分野）
本発明は、大容量記憶装置の分野に関する。特に、本発明は、データ伝送のダイナミック半二重モードによりループ・フェアネスを保持する装置および方法に対する改良ファイバ・チャネル・アービトレーテッド・ループ（ＦＣ－ＡＬ）装置および方法に関する。
【０００２】
（発明の背景）
あらゆるコンピュータ・システムの重要な一構成要素は、データを記憶する装置である。コンピュータ・システムは、データを記憶できる多くの異なる装置を有する。コンピュータ・システムにおける膨大な量のデータを記憶する一つの共通の場所は、ディスク・ドライブに関する。ディスク・ドライブの最も基本的な部品は、回転するディスクと、トランスデューサをディスク上の種々の位置に移動させるアクチュエータと、ディスクにデータを書き込んだりディスクからデータを読み出すために使用される回路とである。ディスク・ドライブは、首尾よくディスク面から読み出したりディスク面に書き込んだりできるようにデータを符号化する回路も備えている。マイクロプロセッサは、ディスク・ドライブの大部分の動作を制御するとともに、要求コンピュータにデータを返送したり、要求コンピュータからデータを取り込んでディスクに記憶する。
【０００３】
ディスク・ドライブとコンピュータ・システムの他の部分との間でデータを転送するためのインターフェースは、典型的には、スモール・コンピュータ・システム・インターフェース（ＳＣＳＩ）またはファイバ・チャネルのようなバスまたはチャネルである。このようなインターフェースのいくつかの機能は、異なる複数の製造者からの種々の装置が相互に交換できるように、かつ、全てが共通インターフェースに接続できるように、しばしば標準化される。このような標準は、典型的には、米国規格協会（ＡＮＳＩ）のような機構のいくつかの標準協会によって規定されている。
【０００４】
種々の記憶装置と種々のコンピュータとの間でデータを交換するための１つの標準インターフェースは、ファイバ・チャネルである。いくつかの実施の形態では、ファイバ・チャネル標準は、複数のアービトレーテッド・ループ（以下で更に説明する）を含む。いくつかの実施の形態では、ファイバ・チャネル標準は、データ転送を制御するＳＣＳＩ類似プロトコルをサポートしている。
【０００５】
ファイバ・チャネルは、スモール・コンピュータ・システム・インターフェース（ＳＣＳＩ）設計よりも顕著な利点を有する。ファイバ・チャネルは、従来のＳＣＳＩ設計の２～２０メガバイト／秒に比較して、現在の約１０６メガバイト／秒に達するかなり高い帯域幅を提供する。ファイバ・チャネルは、典型的なＳＣＳＩ環境における最大７個または１５個の装置に比較して、１２６個に現在達する装置（ホストを含む）を接続することができる点で、更に大きな接続機能を提供する。ファイバ・チャネルは、単一コネクタで取り付けることができ、かつ、スイッチを必要としない。同軸導体を使用するファイバ・チャネルは、ＳＣＳＩ環境における２５メートルに達する最大総延長に比較して、装置間が３０メートルに達する距離で動作し、また、全チャネルにわたり光ファイバを使用すると１０キロメートルに達する。
【０００６】
ＳＣＳＩ環境では、データ伝送での誤りはパリティを使用して検出されるのに対し、ファイバ・チャネルでは、誤りはランニング・ディスパリティおよびサイクリック・リダンダンシー・コード・チェック（ＣＲＣチェック）情報よって認識される。更なる情報は、本

発明者のウェストビーおよび本出願人のシーゲート・テクノロジー（株）による「マルチポート設計においてＣＲＣ発生器を使用したＣＲＣチェック」と題する米国特許第５，８０２，０８０号および「１６Ｂ／２０Ｂエンコーダ」と題する米国特許第５，６６３，７２４号に見い出すことができる。

【０００７】

ファイバ・チャネル・アービトレーテッド・ループ（ＦＣ－ＡＬ）は、バイト・オリエンテッドＤＣ平衡（０，４）ラン・レングス制限８Ｂ／１０Ｂのパーティションによるブロック転送コード方式を採用した工業標準システムである。ＦＣ－ＡＬは、１０６．２５ＭＨｚのクロック周波数で動作する。８Ｂ／１０Ｂエンコーダ／デコーダの一形式は、フラナジェックらによる「バイト・オリエンテッドＤＣ平衡（０，４）８Ｂ／１０Ｂパーティショント・ブロック転送コード」に対して１９８４年１２月４日に許可された米国特許第４，４８６，７３９号に説明されている。

【０００８】

ファイバ・チャネル・アービトレーテッド・ループ（ＦＣ－ＡＬ）は、それぞれ「ノード」と呼ばれる多数の装置を相互に接続可能にする。ノードは、ファイバ・チャネル「トポロジー」（この後で定義されている）に接続可能にしたインターフェースを有するコンピュータ・システムのあらゆる装置（コンピュータ，ワークステーション，プリンタ，ディスク・ドライブ，スキャナなど）である。各ノードは、他のノードに対するアクセスを得るためにＮＬポート（ノード－ループ・ポート）と呼ばれる少なくとも１つのポートを有する。２以上のポートを相互に接続する構成要素は、ひとまとめに、「トポロジー」または「ループ」と呼ばれる。各ノードは、提供されたトポロジーまたはループ内で他の全てのノードと通信する。

【０００９】

これらのポートはファイバ・チャネル・ノードにおける接続であり、データはファイバ・チャネルを介して他のノード（外側世界）のポートに転送できる。代表的なファイバ・チャネル・ドライブは、そのドライブのノード内にパッケージされた２つのポートを有する。各ポートは、情報をポートに搬送するものとポートから情報を搬送するものとの一対の「ファイバ」を含む。各「ファイバ」はシリアル・データ接続であり、また、一実施の形態では、各ファイバは、実際には、同軸ケーブル（例えば、ノードが互いに接近しているときに使用される同軸銅導体）であり、他の実施の形態では、ファイバは、（例えば、これらのノードが異なるキャビネット、特に、異なる建物におけるノードのように、適当な距離だけ分離されているときに）、そのパスのうちの少なくともいくつかに関する光ファイバとして実施される。各ポートに接続されたファイバ対（一つはポートへデータを搬送し、他はポートからデータを搬送する）は、「リンク」と呼ばれ、かつ、各トポロジーの一部である。これらのリンクは、ノード間で「フレーム」で「フレーム」にパッケージされた情報すなわち信号を搬送する。各リンクは、多数の形式のフレーム（例えば、初期化，データおよび制御フレーム）を取り扱う。

【００１０】

各ファイバは１方向のみにデータを搬送するので、複数のノードはループに沿って相互に接続され、そこでは、これらのノードは、転送すべきデータを有するときは、ループを制御するためのアービトレーションをしなければならない。「アービトレーション」は、複数のノードを協調させて、どのノードがループを制御するのかを決定する処理である。ファイバ・チャネル・アービトレーテッド・ループは、ハブまたはスイッチなしでループに多数のノードを取り付ける。これらのノード・ポートは、ポイント・ツー・ポイント・データ転送回路を確立するためにアービトレーション動作を使用する。ＦＣ－ＡＬは、各ポートが回路を確立するために少なくとも最小必要機能を含む分散トポロジーである。アービトレーション・ループ・トポロジーは、２および１２６ノード・ポート間で、任意の数のノードを接続するために使用される。

【００１１】

いくつかの実施の形態では、各ノードは、冗長性をもたす二重ポート（それぞれ別個のポ

ートに接続される）を含み、その結果、一方のループが故障すると、他方のループがループ役割を遂行することができる。二重ポートはまた、２つのホスト（例えば、２つのホスト・コンピュータ）が１台のドライブを共有することを可能にする。

【００１２】

（フェアネス背景）

本発明の文脈における「フェアネス」は、ファイバ・チャネル・ループのような共有リソースが複数のユーザの各々に対して、例えば複数のディスク・ドライブ１００のそれぞれに対して、タイムリ・ベースで、すなわち、不都合な遅延なしに、かつ、各ディスク・ドライブ１００が転送するために必要とするデータ量に比例した時間量に対して、利用可能にされる動作モードである。「プレイング・アンフェア」は、１ユーザ、例えば１台のディスク・ドライブが、過度に頻繁にまたは過度に長期間にわたって、共有リソース、例えばファイバ・チャネル・ループの制御を行う動作モードであり、したがって、ループを使用する必要がある他のディスク・ドライブがそれを使用することを妨げる。

【００１３】

「フェアネス」は、各ポートが（優先度に基づいて）ループに対するアクセスを獲得可能にさせるとともに、（時間制限なしに）それが欲するだけの情報を送信可能にさせる。そのために、他の全てのポートがアクセスを使用してしまうまで、アービトレーション獲得を待機する必要がある。

【００１４】

ループ用のプロトコルは、各Ｌ＿ポート（ループ・ポートとも呼ばれ、Ｌ＿ポートは、ノーマル・ループ・ポート「ＮＬ＿ポート」、または、２つのループを相互接続するために一般に使用されるファブリック・ループ・ポート「ＦＬ＿ポート」である。）がループをアクセスするために連続的にアービトレーションをすることを可能にさせる。優先度は、アービトレーション・ループ物理アドレス（ＡＬ＿ＰＡ）に基づいて各参加Ｌ＿ポートに割り付けられる。他の優先度によるプロトコルのように、これは、低い優先度のＬ＿ポートがループに対するアクセスを獲得できないという状況に至る恐れがある。アクセス・フェアネス・アルゴリズムは、全てのＬ＿ポートがアービトレーションしかつループに対するアクセスに勝つ機会を与えられるアクセス・ウィンドウを設定する。全てのＬ＿ポートがループを１回アクセスする機会を有するときは、新しいアクセス・ウィンドウが開始される。Ｌ＿ポートは、再びアービトレーションすることができ、究極的に新しいアクセス・ウィンドウにおいてループに対するアクセスに勝つ。あらゆるＬ＿ポートが任意の１アクセス・ウィンドウにおいてループをアクセスすることを必要とすることはない。

【００１５】

アクセス・フェアネス・アルゴリズムを使用するＬ＿ポートがアービトレーションをしかつループに対するアクセスに勝ったときは、少なくとも１アイドルがＬ＿ポートによって送信されてしまうまで、Ｌ＿ポートは再びアービトレーションすべきでない。第１のＬ＿ポートがアービトレーションに勝ちアイドルを送信する間の時間は、アクセス・ウィンドウである。アクセス・ウィンドウの初期リセットを阻止するために、特殊なアービトレーション・プリミティブ信号（すなわち、ＡＲＢ（Ｆ０））を使用する。アクセス・フェアネス・アルゴリズムの詳細は、ファイバ・チャネルＦＣ－ＡＬ仕様（ＡＮＳＩ標準：ファイバ・チャネルＦＣ－ＡＬ　Ｘ３Ｔ１１／プロジェクト９６０Ｄ／Ｒｅｖ．４．５およびファイバ・チャネルＦＣ－ＡＬ２　Ｘ３Ｔ１１／プロジェクト１１３３Ｄ／Ｒｅｖ．６．３）に記載されている。

【００１６】

アクセス・フェアネス・アルゴリズムは、ＡＮＳＩ　Ｘ３．２３０，ＦＣ－ＰＨがクラス１接続に関する時間を制限しないように、Ｌ＿ポートがアービトレーションに勝てば、Ｌ＿ポートがループを制御する時間を制限することはしない。しかしながら、Ｅ＿Ｄ＿ＴＯＶよりも長くアクセスが拒否されると、アクセス・ウィンドウがリセットされて、Ｌ＿ポートがアービトレーションを開始することができる。

【００１７】

全てのＬ＿ポートがフェアネス・アルゴリズムを実施するけれども、ＦＬ＿ポートまたはＮＬ＿ポートは常にフェアネス・アルゴリズムを使用することが要求される。例えば、１つのＬ＿ポートが他のＬ＿ポートよりも多くのループ・アクセスを要求するときは、Ｌ＿ポートはアンフェアとなることを選択することができる。

【００１８】

全てのＬ＿ポート用のループに同等のアクセスを提供するために、各Ｌ＿ポートはアクセス・フェアネス・アルゴリズムを使用することが推奨される。Ｌ＿ポートがアクセス・フェアネス・アルゴリズムを使用しているときは、それは「フェア」Ｌ＿ポートと呼ばれる。

【００１９】

フェアＬ＿ポートがループに対してアービトレーションしかつループへのアクセスに勝ちかつ他のＬ＿ポートがアービトレーションしていることを検出しないときは、そのＬ＿ポートは、既存の回路を無期限に開き続けるかこの回路を閉じ、かつ、ループの所有権を保持して（すなわち、再アービトレーションなしに）ループ上の他のＬ＿ポートを開くことができる。

【００２０】

フェアＬ＿ポートがループに対するアクセスを有しかつ他のＬ＿ポートがアービトレーションしていることを検出すると、Ｌ＿ポートは最小可能時間でループを閉じることができる。Ｌ＿ポートは、ループを閉じて、異なるＬ＿ポートを開く前に次のアクセス・ウィンドウにおいて再びアービトレーションする。

【００２１】

いくつかのループの構成は、若干のＬ＿ポートが１回／アクセス・ウィンドウよりもループに対して多くのアクセスを有することを必要とすることがある。これらのＬ＿ポートの例は、限定的ではないが、サブシステム・コントローラまたはファイル・サーバを含む。

【００２２】

Ｌ＿ポートは、アクセス・フェアネス・アルゴリズムを使用しないように初期化することができる（または、一時的に選択することができる）。Ｌ＿ポートがフェアネス・アルゴリズムを使用していないときは、それは「アンフェア」Ｌ＿ポートと呼ばれる。アクセス・フェアネスに参加するか否かの判断は、実施する必要性に対して残される。

【００２３】

アンフェアＬ＿ポートがループに対してアービトレーションしかつループへのアクセスに勝ちかつ他のＬ＿ポートがアービトレーションしていることを検出しないときは、そのＬ＿ポートは、既存の回路を無期限に開き続けるかこの回路を閉じ、かつ、ループの所有権を保持して（すなわち、再アービトレーションなしに）ループ上の他のＬ＿ポートを開いてもよい。

【００２４】

アンフェアＬ＿ポートがループを制御しかつ他のＬ＿ポートがアービトレーションしていることを検出したときは、アンフェアＬ＿ポートは最小可能時間でループを閉じてもよい。アンフェアＬ＿ポートは、ループの所有権を保持するとともに（すなわち、再アービトレーションなしに）、ループ上の他のＬ＿ポートを開いてもよい。

【００２５】

参加しているＦＬ＿ポートは、常に、そのＡＬ＿ＰＡに基づきループ上の最優先Ｌ＿ポートである。ＦＬ＿ポートは、そのトラヒックの大部分はファブリックの残りをもつので、アクセス・フェアネス・アルゴリズムを使用することを免除される。

【００２６】

ＦＬ＿ポートがループを制御しかつ他のＮＬ＿ポートがアービトレーションしていることを検出すると、ＦＬ＿ポートは最小可能時間でループを閉じることができる。ＦＬ＿ポートは最高優先度を有しかつフェアネスから免除されているので、それは、常に、アービトレーションに勝つことになる。したがって、他のＮＬ＿ポートとの通信が必要である場合には、ＦＬ＿ポートは、ループに対するそのアクセスを保持するとともに（すなわち、再

アービトレーションなしに）、ループ上の他のＮＬ＿ポートを開いてもよい。
【００２７】
したがって、特にダイナミック半二重機能を含むファイバ・チャネル実施では、ループ・フェアネスを保持する構成に対する必要性が存在する。
【００２８】
（発明の概要）
ループ・フェアネスを保持する方法および装置を説明する。いくつかの実施の形態は、ダイナミック半二重特徴を含む。本発明の第１の態様は、１以上のポートを有する第１のチャネル・ノードを含むループ・フェアネスを保持する通信チャネル・システムを含み、各ポートは、ファイバ・チャネル・アービトレーテッド・ループ・シリアル通信チャネルをサポートしかつこれに取り付けられている。これらのポートの１つは、そのポートの取り付けられたチャネルを制御するためにアービトレーションし、チャネル・ループの制御において、アービトレーションに勝つと、フェアネスを保持している装置は、所定量の使用が第１のポートと通信チャネルとの間に発生したか否かに少なくとも一部基づいて通信チャネルの制御を解放させる。
【００２９】
本発明の他の態様は、ファイバ・チャネル・アービトレーテッド・ループ・セリアル通信チャネルのループを制御するためにアービトレーションすることと、第１のポートと通信チャネルとの間で所定量の使用が発生したか否かに少なくとも一部基づいて通信チャネルの制御を解放させることとを含む通信方法を提供する。
【００３０】
上記システムおよび方法のいくつかの実施の形態では、所定量の使用は、第１の所定量のデータの転送を含む。いくつかのそのような実施の形態では、第２の所定量よりも少ないデータが転送されるべく残っているときは、チャネルの制御の解放が禁止される。
【００３１】
（詳細な説明）
好ましい実施の形態の以下の詳細な発明では、その一部を形成するとともに、説明の手段として本発明を実施可能な特定の実施の形態を示す添付図面を参照する。他の実施の形態を利用することもでき、かつ、本発明の範囲から逸脱することなく構造的な変更を行うことができることを理解すべきである。
【００３２】
この出願で説明した本発明は、ハード・ディスク・ドライブ，ＺＩＰドライブ，フロッピ・ディスク・ドライブ，光ディスク・ドライブ，ＣＤＲＯＭ（「コンパクト・ディスク読出し専用メモリ」）ドライブおよび他の任意形式のドライブ，複数のドライブ・システム（例えば、「インエクスペンシィブ／インディペンデント・ディスク・ドライブのリダンダント・アレー」、すなわち、ＲＡＩＤ構成）、または、複数のドライブと他の複数のドライブまたは複数の情報処理システムとの間でデータを通信する他の複数の装置を含むあらゆる形式のディスク・ドライブに有用である。いくつかの実施の形態では、本発明は、（例えば、複数のファイバ・チャネル・ループを互いに接続するために使用される）ハブおよびスイッチ，ワークステーション，プリンタ，ファイバ・チャネル・アービトレーテッド・ループ上に接続されている他の装置または情報処理システムのような非ディスク装置用のノード・インターフェースにおいて有用である。
【００３３】
図１は、ファイバ・チャネル・ノード・インターフェースを有するディスク・ドライブ１００のブロック図を示す。更なる情報は、「ループ初期化および応答用の方法および専用フレーム・バッファ」と題する米国特許出願第０９／１９３，３８７号に見い出すことができる。
【００３４】
図１と図２とを参照すると、ファイバ・チャネル・ループ・インターフェース回路１２２０（ノード・インターフェース１２２０とも呼ばれる）は、ループ初期化および応答用の

専用送信フレーム・バッファ７３を含む。一実施の形態では、各ノード・インターフェース回路１２２０は、２つのポート１１６（ポートＡおよびポートＢを表す）を含む。（「ループ初期化」は、１以上の特殊非データ・フレームのシーケンスを送信し（かつ、これらのフレームに対する応答を監視し）てファイバ・チャネル・ループを初期化する。「応答」は、他のノードからの指令または質問に応答して送出された非データ・フレームである。）ファイバ・チャネル・アービトレーテッド・ループ通信チャネル１２５０（ループ１２５０またはファイバ・チャネル・ループ１２５０とも呼ばれる）は、ディスク記録ドライブ１００とコンピュータ（または、他の情報処理システム）１２０２または他の情報処理装置との間でデータを通信するために使用され得る。一実施の形態では、ファイバ・チャネル・ループ１２５０はシリアル通信チャネルであり、他の実施の形態では、ファイバ・チャネル・ループ１２５０を実施するために２以上の並列ライン（または、「複数のファイバ」）が使用される。このような専用送信フレーム・バッファ７３を備えると、他方のポートがデータ・フレームを送受信している間に、二重ポート・ノード１２２０の一方のポート１１６が初期化または応答フレームを送信することを可能にする。ポート１１６は、一方が入力データ用のライン１１７であり他方が出力データ用のライン１１８であるシリアル・ラインであり、両方のライン１１７，１１８は、通信チャネル・ループ１２５０に接続するとととともにその一部を形成している。また、２ポート・ノードの各ポート１１６に対して、専用受信バッファ（５３，５３’，５５）が設けられている。（プライム符号による参照番号を有する各ブロック（例えば、５３’）はプライムのない対応するブロック（例えば、５３）と同一機能を提供することに注意すべきである。）ファイバ・チャネル１２５０からフレームとともに受信するサイクリック・リダンダンシー・コード情報は、１以上のフレーム・バッファ（５３，５３’または５５）の１つに記憶されたのち、チェックされてフレーム・バッファ（５３，５３’また５５）にある間にデータの完全さを保証する。ループ１２５０の制御は、ループ１２５０の制御に対してアービトレーションするのに費やされる総時間量を減少させるために、プログラム可能な量のデータが送信に利用できる限り、保持される。

【００３５】

いくつかの実施の形態では、ディスク・ドライブ１００は、１以上のディスク・プラッタ１３４とディスク・プラッタ当り１以上の磁気読出し／書込みトランスデューサ１５０とアーム・アクチュエータ・アッセンブリ１２６とを有する磁気記憶ヘッド・ディスク・アッセンブリ（ＨＤＡ）１１４を含む。トランスデューサ（または、「ヘッド」）とＨＤＡインターフェース１１３との間の信号は、データをディスク・プラッタ１３４におよびそれから転送する。したがって、いくつかの実施の形態の「ディスク・ドライブ」（例えば、図１のディスク・ドライブ１２５６）は、ＨＤＡ１１４とＨＤＡインターフェース１１３（例えば、通常のＳＣＳＩドライブ）とを含み、１以上のそのような従来のディスク・ドライブ１２５６は、図１に示すように、ループまたはファイバ・チャネル・トポロジーに接続するために、外部ノード・インターフェース１２２０に接続される。他の実施の形態では、「ディスク・ドライブ」は、図２のディスク・ドライブ１００によって代表され、ディスク・ドライブ１００全体内のディスク・ドライブ１２５６と統合されるノード・インターフェース１２２０を含む。一実施の形態では、データは、続いて、オフ・チップ・バッファ１１１からおよびそれに転送される。本発明は、専用オン・チップ・バッファ１１９を備えており、これは、図示した実施の形態では、各ポート（すなわち、バッファ５３，５３’）用の受信非データ・フレーム・バッファ５３（代わって、「入力非データ・バッファ５３」と呼ばれる。）と、両ポート（他の実施の形態では、一度に１ポートによってのみ単一バッファが使用される。）により同時に使用可能とされる送信フレーム・バッファ７３と、ＣＲＣチェッカ５９６とともに共有データ・フレーム・バッファ５５とを含む。

【００３６】

一実施の形態では、ループ・ポート・トランシーバ・ブロック１１５（すなわち、１１５および１１５’）は、ポートＡおよびポートＢを介してこれに接続されているファイバ・

チャネル・ループ１２５０（図２を参照）に転送するデータを直並列変換および並直列変換する複数のポート・トランシーバを含む。いくつかの実施の形態では、トランシーバ１１５は、外部トランシーバとして実施される。他の実施の形態では、これらのトランシーバは、ブロック１１０にオン・チップで配置されている。いくつかの実施の形態では、右側（すなわち、図１のトランシーバ１１５または１１５’に対して右側）インターフェースは、１０ビット幅である並列入出力信号である。他の実施の形態では、それらは２０ビット幅である。ブロック１１０，１１１，１１２とポートＡトランシーバ１１５とポートＢトランシーバ１１５’とは、総合して、ファイバ・チャネル・ノード・インターフェース１２２０を形成している。いくつかの実施の形態では、トランシーバ１１５，１１５’は単一チップ１１０に集積される。他の実施の形態では、それらの直並列／並直列変換機能を含むトランシーバ１１５，１１５’は、チップ１１０から分離した回路上に実施される。

【００３７】

他の実施の形態では、トランシーバ１１５は、シリアル・ループ１２５０とチップ１１０との間の単なるインターフェースであり、１０ビット幅または２０ビット幅のデータに対する直並列変換／並直列変換がオン・チップで実行される。

【００３８】

図２は、コンピュータ・システム１２００の概要図である。都合のよいことに、本発明はコンピュータ・システム１２００に使用するのによく適している。コンピュータ・システム１２００はまた、電子システムまたは情報処理システムと呼ばれてもよく、中央処理装置（ＣＰＵ）とメモリとシステム・バスとを含む。コンピュータ・システム１２００は、中央処理装置１２０４とランダム・アクセス・メモリ（ＲＡＭ）１２３２と中央処理装置１２０４およびランダム・アクセス・メモリ１２３２を通信可能に接続するシステム・バス１２３０とを有するＣＰＵ情報処理システム１２０２を含む。ＣＰＵ情報処理システム１２０２は、ファイバ・チャネル・ノード・インターフェース１２２０を含む。１以上のディスク記憶情報処理システム１００～１００’のそれぞれは、１以上のディスク・ドライブ装置１２５６とファイバ・チャネル・ノード・インターフェース１２２０とを含む。

【００３９】

いくつかの実施の形態では、多数のディスク・ドライブ１２５６は、装置１００’がディスク・ドライブのＲＡＩＤアレーとなるように、単一ノード・インターフェース１２２０、例えばＲＡＩＤ（インエクスペンシィブ／インディペンデント・ディスク・ドライブのリダンダント・アレー）構成に接続されている。ＣＰＵ情報処理システム１２０２は、内部入出力バス１２１０と入出力バス１２１０に取り付けられたいくつかの周辺装置、例えば１２１２，１２１４および１２１６とを駆動する入出力インターフェース回路１２０９を含む。これらの周辺装置は、ハード・ディスク・ドライブ，磁気光学ドライブ，フロッピ・ディスク・ドライブ，モニタ，キーボードおよび他のそのような周辺装置を含むことができる。任意形式のディスク・ドライブまたは他の周辺装置は、ここで説明するファイバ・チャネル方法および装置（特に、例えばファイバ・チャネル・ノード・インターフェース１２２０における改良）を使用することができる。各装置では、任意の所与のループ１２５０に接続するためにＡポートまたはＢポートを使用することができる。

【００４０】

システム１２００の一実施の形態は、オプションとして、第２のＣＰＵ情報処理システム１２０２’（システム１２０２と同一または類似している）を含み、ＣＰＵ情報処理システム１２０２’は、中央処理装置１２０４’（中央処理装置１２０４と同一である）と、ランダム・アクセス・メモリ（ＲＡＭ）１２３２’（ＲＡＭ１２３２と同一である）と、中央処理装置１２０４’とＲＡＭ２３２’とを通信可能に結合するシステム・バス１２３０’とを有する。ＣＰＵ情報処理システム１２０２’は、それ自身のファイバ・チャネル・ノード・インターフェース１２２０’（ノード・インターフェース１２２０と同一である）を含むが、（ループ１２５０から分離され独立した）第２のファイバ・チャネル・ループ１２５０’を介して１以上のディスク・システム１００に接続されている（この説明

例では、それは、ディスク・システム１００’に単に接続されているが、他の実施の形態では、全ての装置またはディスク・システム１００～１００’に接続されている）。この構成は、２つのＣＰＵ情報処理システム１２０２，１２０２’が各ＣＰＵ情報処理システム１２０２用の個別的なファイバ・チャネル・ループを使用して１以上のディスク・システム１００を共有可能にさせる。更に他の実施の形態では、全ての装置１００～１００’と全てのＣＰＵ情報処理システム１２０２～１２０２’とが両方のループ１２５０，１２５０に接続されている。

【００４１】

本発明の一実施の形態を構築するのに使用されるファイバ・チャネル仕様は、下記のＡＮＳＩ標準を含む。

【００４２】

【表１】

ファイバ・チャネル	FC-PH	X3T11/プロジェクト 755D/Rev. 4.3 物理的及びシグナリング・インターフェース
ファイバ・チャネル	FC-AL	X3T11/プロジェクト 960D/Rev. 4.5 アービトレーテッド・ループ
ファイバ・チャネル	FC-AL2	X3T11/プロジェクト 1133D/Rev. 6.3 アービトレーテッド・ループ
ファイバ・チャネル	FCP	X3T10/ Rev. 012
SCSI用プロトコル		X3.2 69-199X

【００４３】

Ｉ．ダイナミック半二重によるループ・フェアネスの保持

本発明の一実施の形態は、ダイナミック半二重（ＤＨＤ）コマンドまたは命令によりループ・フェアネスを保持する（または、等価的に提供する）フェアネス保持装置１７５および方法と、ＤＨＤコマンドを使用したフェアネスを提供する方法とを提供する。

【００４４】

ファイバ・チャネルは本質的に全二重リンクである（ファイバ対において複数のフレームが同時に両方向に移動する）が、いくつかのループ・ポート（Ｌ＿ポート）の実施は１方向データ転送のみをサポートすることができる。あるポートが全二重動作を可能とするときでも、ループ・アービトレーションを損ない得る状況が存在する。

【００４５】

ダイナミック半二重（ＤＨＤ）は、ＯＰＥＮステートにあるポートによってループ上に送信されるループ・プリミティブであり、ＯＰＥＮＥＤステートにあるポートに対してそれ以上送信すべきフレームを有していないことを表示する。ＤＨＤは、複数のポートが確立された回路を下記により更に効率的に使用することを可能にする。

１．半二重データ転送のみを可能とするポートが再アービトレーションすることなく逆方向に複数のフレームを転送することを可能にする。

２．ＯＰＥＮステートにあるポートがそのデータ転送を完了したとしても、ＯＰＥＮＥＤステートにあるポートが全てのフレームを転送することを可能にする。

【００４６】

ＯＰＥＮステートにあるポートは、通常、最初のＣＬＲを送信してループを閉じる。全二重回路が存在するときは、ＯＰＥＮＥＤステートにあるポートは、ＣＬＳを受信して、バ

ッファ・ツー・バッファ・クレジット（ＢＢ＿Ｃｒｅｄｉｔ－Ｒ＿Ｒｄｙ）がもうないというまでフレームを送信し続けることができる。ＯＰＥＮＥＤポートがもはやフレームを送信できなくなると、それはＯＰＥＮポートにＣＬＳを返送しなければならない。

【００４７】

ＣＬＳを送信するよりもクローズ・イニシアティブを転送するのに有用と思われる少なくとも２つの場合が存在する。

１．いくつかの実施はノードで同時送受信転送を処理することができない。これらのノードは、しばしば、ＯＰＥＮポートに対してペンディングしているフレームを有するが、これらは半二重設計のため、存在する双方向回路の利点を利用できない。

２．全二重転送が可能であっても、ＯＰＥＮポートがＣＬＳを送信すれば、ＯＰＥＮＥＤポートは既存のクレジットに基づいてフレームを送信することのみが可能である。（例：ポートが新しいコマンドを受信するようにＯＰＥＮＥＤであり、２つのＲ＿Ｒｄｙを受信し、ＯＰＥＮポート用のデータを読み出している。ＯＰＥＮＥＤポートは、それが閉じなければならない前に、２つのデータ・フレームを送信することのみが可能である。）

【００４８】

この追加的な再アービトレーション・サイクルを避けるために、ＤＨＤプリミティブ信号が供給される。ＤＨＤは、ＣＬＳを送信する代わりにＯＰＥＮ　Ｌ＿ポートによって送信される。ＤＨＤを送信すると、ＯＰＥＮポートはＲ＿Ｒｄｙおよびリンク制御フレーム（ただし、データ・フレームなし）を送信し続けることができる。ＯＰＥＮＥＤポートは、それがＤＨＤを受信しクローズ・イニシアティブを保持していることを記憶しており、もはやＯＰＥＮポートに送信すべきフレームがないときは最初のＣＬＳを送信することが期待されている。

【００４９】

この記述における「フェアネス」は、ディスク・ドライブ１００のような複数のユーザのそれぞれに対して適時ベースにより、すなわち好ましくない遅延なしに、かつ、各ディスク・ドライブ１００が転送する必要があるデータ量に比例した時間量に対して、ファイバ・チャネル・ループ１２５０のような共有リソースを利用できるようにした動作モードである。「プレイング・アンフェア」は、１ユーザ、例えばＣＰＵ情報処理システム１２０２（イニシエータとして作動する）の１ポートが、全ての他のユーザが自分の順番が来るのを待機することなく、共有リソースの制御を獲得する動作モードである。プレイング・アンフェアは、例えばＣＰＵ情報処理システム１２０２が多数の低速度ディスク・ドライブ１００上で動作を開始可能にさせるので、しばしば、制限されたある種の（例えば、高速度装置が低速度装置より多くの順番が巡って来るのを可能にする）環境においては、好ましいものとなる。１ディスク・ドライブ１００がアンフェアを実行するのを可能にすると、その装置からのデータ転送がそうでないときよりも速く完了し得るが、他のディスク・ドライブ１００がサービスを受けられず、他の動作を実行するように開放させないため、また、ＣＰＵ情報処理システム１２０２が他の複数のディスク・ドライブ１００からデータを得るまでに待機するのが長くなり過ぎてしまうために、一般的には、システム・パフォーマンスを低下させる結果となる（図２を参照）。

【００５０】

本発明で目的とするファイバ・チャネル・アービトレーテッド・ループ設計では、ループ１２５０をアクセスするために、ループ・ポート１２２０はアービトレーションしなければならない。どのポート１２２０がループ１２５０の制御を獲得するのかを判断するために優先度システムが使用され、また、「フェアネス」方式は複数のポートが飢えていないことを保証するために使用される。目標装置のように、ディスク・ドライブ１００は、通常、フェアに行動し、低い優先度が与えられ、これが各ドライブ１００がループ１２５０に対して等しくアクセスをするのを保証する。イニシエータ（例えば、ＣＰＵ情報処理システム１２０２）は、ドライブ・キューをフルに保持するために、アンフェアに行動することができる（ドライブ・キューは、各ディスク・ドライブ１００がそれに向けられたコマンドを保持するペンディングおよびカレント・オペレーションのリストである）。「イ

ニシエータ」は、他の装置（「目標」と呼ばれる）によって実行されるべき入出力（Ｉ／Ｏ）処理を要求するファイバ・チャネル・ループ上の装置である。

【００５１】

「全二重」は、データがポート１２２０に受信されているのとちょうど同時にデータがそのポートから送信されることができる動作モードである。「半二重」は、データが交互的にのみであって同時ではなくポートから送信されたりそのポートに受信されることができる動作モードである。

【００５２】

本発明の一態様は、ファイバ・チャネル・アービトレーテッド・ループ・アーキテクチャー（ＦＣ－ＡＬ－２）に付加される「ダイナミック半二重」と呼ばれる特徴である。ポートが（下記章ＩＩ，ＩＩＩ，ＩＶおよびＶで説明されている設計のような）全二重動作を可能とするときであっても、ＤＨＤ特徴を設けることによってループ・アービトレーション・サイクルの数を減少させることができる状態がある。例えば、ＯＰＥＮポート（ＯＰＥＮ状態にあるポート１１６）がＣＬＳプリミティブ信号を送信するとき、ＯＰＥＮＥＤポートは既存のクレジットに基づいてフレームを送信できるだけである。（例えば、ノード・インターフェース１２２０のＡポート１１６はＯＰＥＮＥＤであって新しいコマンドを受信し、それは２つのＲ＿ＲＤＹを受信するとともにＯＰＥＮポート用のデータを読み込んでいた。ＯＰＥＮＥＤポートは、それが閉じられなければならない前に、２つのデータ・フレームを送信できるだけである。）ダイナミック半二重により、ＯＰＥＮポートは、ＣＬＳの代わりにプリミティブＤＨＤを送信するとともに、Ｒ＿ＲＤＹを送信し続ける。これは、データを読み出していたディスク・ドライブ１００が、ループに対してアービトレーションすることを必要とせずに新しいコマンドを受信するときに転送を完了することを可能にする。このＤＨＤ特徴は、アービトレーション・サイクルを減少させることができるが、ループ・フェアネスを損なう恐れがある。

【００５３】

例えば、本発明によるシステム１２００（図２を参照）では、情報処理システム１２０２は、（複数のデータ・フレームを返送する）ディスク・ドライブ１００に読出し動作コマンドを送出するとともに、（ディスク・ドライブ１００が、制御を放棄することなくかつループ１２５０の制御のために再アービトレーションする必要なく多数のデータ・フレームの転送を完了するためにループ１２５０上の制御を延長された期間維持することを可能にする）ＤＨＤプリミティブ信号を送出する。したがって、ＤＨＤプリミティブ信号を受信したディスク・ドライブ１００は、それが既にアービトレーションに勝ちかつ次のフェアネス・ウィンドウを待機していたとしても、再アービトレーション・ウィンドウを再び使用可能にされるからである。ループのフェアネス（全ての装置がループ・リソースに対して同等のアクセスを有する能力）は損なわれる。

【００５４】

最高の優先度アドレスを有しかつアンフェアを行う（すなわち、最高の優先度アドレスを使用して、それらが長時間その順番が来るのを待機していたとしても最低の優先度アドレスを有する他の装置の犠牲によりループの制御を行うことによって）イニシエータ情報処理システム１２０２（図２を参照）は、アービトレーションに勝って、ドライブ１００に新しいコマンド（例えば、ＤＨＤコマンド）を送出する。ドライブ１００は、イニシエータ１２０２用のデータを読み出していたので、ループ１２５０に対してアービトレーションを実行していた。ドライブ１００は、ＤＨＤコマンドを受信すると、そのデータ・フレームの全てと、動作を完了する応答とを送出することが可能とされる。新しいコマンドにより開放にならなかった他のドライブ（例えば、ドライブ１００’）は、アービトレーションで「フェア」に勝つまで待機する必要がある。一方、コマンドを完了したドライブ１００は、次のコマンドに対するデータを読み出すことが可能であった。前のコマンドは完了したので、イニシエータ１２０２は、そのドライブ１００に新しいコマンドを送出して、そのキューをフルに保持することができる。新しいコマンドは、ドライブ１００が他の読出し転送を完了することを可能にさせ得た。いくつかのドライブは、時間上で、ループ

１２５０に対するアクセスを「スターブド（ｓｔａｒｖｅｄ）」にするとともに、ＤＨＤを受信するのに使用するものよりも少ないコマンドで終了することができる。

【００５５】

本発明の改良されたフェアネス保持特徴は、１つのノード１１６がループ１２５０の制御を維持することを可能にさせられる時間を制限しており、したがって、他のノード１１６が合理的な時間内に順番が来ることを保証している（かくして、「スターベーション」（ｓｔａｒｖａｔｉｏｎ））を回避する。一実施の形態では、一つのノード１１６がループに対する制御を維持することが可能にさせられる時間は、任意の１つの動作で転送されるフレームの最大数を調節することによって調節される。一実施の形態では、ノード１１６がそのデータ転送の完了に十分近いときには、そうでないときに許容されるフレームの最大数を超えるものであっても、その転送を完了することが可能にさせられる。

【００５６】

本発明によれば、ダイナミック半二重を使用するときにループ・フェアネスを保持するために、数値によるフレーム限定特徴がノード・インターフェース１２２０に付加される。一実施の形態では、転送が完了に近くない限り、装置（例えば、ディスク・ドライブ１００）がＤＨＤコマンド（ＤＨＤ「プリミティブ」とも呼ばれる。）を受信した後にいくつの読出しデータ・フレーム（すなわち、送信すべきデータ・フレーム）が送信されることを許容されるかについて、その上限値が設定される。いくつかの実施の形態では、数値限定は、プログラム可能であり、（一実施の形態では）販売者固有ログイン・パラメータまたは（他の実施の形態では）モード・ページ設定に設定されるか、（他の実施の形態では）マイクロプロセッサを介してデフォルト値に設定される。これは、イニシエータ１２２０が、ドライブ１００が使用されているアプリケーション（すなわち、ＣＰＵ情報処理システム１２０２で実行されているソフトウェア）に対する数値限定をプログラム可能に調整することを可能にするか、そのドライブを合理的な値に設定することを可能にする。

【００５７】

一実施の形態では、本発明は、図１に示すようなフェアネス保持装置１７５を提供する。一実施の形態では、１６ビット・カウンタ６０（ここでは、ｄｈｄ＿ｃｎｔ＿ｏｕｔ（１５：０）とも呼ばれる。図１を参照）は、ＤＨＤが受信されるときにｉｎｔ＿ｚｅｒｏ信号１５９によってゼロにされるとともに、送信された各フレームに対して（インクリメンタ１６１を使用して）インクリメントされる。コンパレータ１６２は、ＤＨＤカウンタ１６０が限界レジスタ１６３（ｄｈｄ＿ｍａｘ＿ｆｒｍｓ（１５：０）とも呼ばれる。）に記憶された数値限定値に達した時をチェックして、データ転送を停止する（ループ１２５０を閉じる）時を判定する。コンパレータ１６２は、ｓｕｓｐｅｎｄ＿ｘｆｅｒ信号１６４を出力する。一実施の形態では、ｓｕｓｐｅｎｄ＿ｘｆｅｒ信号１６４は、データ・フレーム送信動作を停止させるとともにループを一時的に閉じるようにして、他の装置がループ１２５０を使用することを可能にする（すなわち、この実施の形態は、以下で説明する信号１６９および信号１７９を無視して、ｓｕｓｐｅｎｄ＿ｘｆｅｒ信号１７２をｓｕｓｐｅｎｄ＿ｘｆｅｒ信号１６４と同等にすることによって限界に達すると、転送を停止する）。他の実施の形態では、このフェアネス保持装置１７５用のｅｎａｂｌｅ＿ＤＨＤ＿ｓｕｓｐｅｎｄ信号１７０は、ＡＮＤゲート１７１への入力としても供給される（すなわち、この実施の形態は、以下で説明する信号１６９を無視して、イネーブル信号１７０が「オン」であれば限界に達して転送を停止し、イネーブル信号１７０が「オフ」であれば本発明のフェアネス保持特徴１７５をディセーブルする）。

【００５８】

他の実施の形態では、（ＤＨＤコマンドが受信されたときにのみというよりも）他のデータ転送に対してカウンタ１６０が活性化される。いくつかの実施の形態では、カウンタ１６０は、以上で説明した限界レジスタ１６３にロードされる値をロードすることによって初期化される減算カウンタにより置換され、また、この減算カウンタは、フレームが転送される毎に１つデクリメントされ、ゼロになると信号１６４をアクティブにし、したがって以上で説明したブロック１６０，１６１，１６２，１６３と等価な機能を提供する。

【００５９】
いくつかの実施の形態では、カウンタ１６０は、上述したように、転送されたフレームの数をモニタする。他の実施の形態では、カウンタ１６０は、転送されたデータ量のバイト若しくはワードの数また他の測定量をモニタする。
【００６０】
更に他の実施の形態では、カウンタ１６０は、転送中に経過した時間をモニタするとともに所定の時間が過ぎると信号１６４を活性化するタイマによって置換される。これは、以上で説明したものと同様の機能を提供するが、転送されたデータ量によるというよりも時間に基づくフェアネスを提供する。いくつかのそのような実施の形態では、所定量よりも少ないデータが転送されるべく残っているとき、および／または、ｅｎａｂｌｅ＿ＤＨＤ＿ｓｕｓｐｅｎｄ信号１７０がディセーブルされたとき、ＡＮＤゲート１７１はそのまま使用されてｓｕｓｐｅｎｄ＿ｘｆｅｒ信号を禁止する。
【００６１】
いくつかの実施の形態では、カウンタ１６５は、上述したように、転送されたフレームの数を監視する。他の実施の形態では、カウンタ１６５は、バイト若しくはワードの数または転送されたデータの量の他の測定値を監視する。
【００６２】
更に他の実施の形態では、カウンタ１６５は、転送中に経過した時間を監視するとともに所定の時間が過ぎると信号１６９を活性化するタイマで置換される。これは、上述したものと同様の機能を提供するが、転送されたデータの量というよりも時間に基づいたフェアネスを提供する。いくつかのそのような実施の形態では、所定量よりも少ないデータが転送されるべく残っているとき、および／または、ｅｎａｂｌｅ＿ＤＨＤ＿ｓｕｓｐｅｎｄ信号１７０がディセーブルされたときは、ＡＮＤゲート１７１がそのまま使用されてｓｕｓｐｅｎｄ＿ｘｆｅｒ信号を禁止する。
【００６３】
一実施の形態では、転送長カウンタ１６５ロジックは、フレームの代わりにワードの数を使用する。
【００６４】
上述の説明はダイナミック半二重特徴を含む実施の形態について特になされたが、本発明のフェアネス保持特徴１７５がＤＨＤ特徴から独立して提供される本発明の他の実施の形態が存在する。例えば、いくつかの実施の形態では、フェアネス保持特徴１７５が提供されるが、ＤＨＤ特徴は提供されない。他の実施の形態では、両者は提供されるが、これらの状況に対するフェアネスを強固にするために、ＤＨＤコマンドを受信したときおよび他の一定の状況においてフェアネス保持特徴が使用される。例えば、一実施の形態では、情報処理システム１２０２からドライブ１００への書込みデータの転送である。
【００６５】
一実施の形態では、転送長カウンタ１６５（ｘｆｒ＿ｃｎｔ＿ｏｕｔ（２６：０）とも呼ばれる。）が提供される。転送長カウンタ１６５は、ロジック１６６によってデクリメントされて、送信すべきワードの残りの数を示す。転送長カウンタ１６５およびディクリメンタ１６６とはデータ転送フレーム長カウンタ８２を形成する。この実施の形態は、転送長カウンタ１６５における残りのワードのカウントを最小長レジスタ１６８（ｄｈｄ＿ｃｍｐｌ＿ｌｅｎ（２６：０）とも呼ばれる。）に保持されたプログラマブル値と比較して、転送が完了に近いときには転送の停止を阻止する（すなわち、転送されるべく残っているワードが最小長レジスタ１６８に保持された値よりも小さいときには、その転送は「完了に近い」と定義される。）。一実施の形態では、ＡＮＤゲート１７１は、信号１６４と信号１６９との論理積を形成してｓｕｓｐｅｎｄ＿ｘｆｅｒ信号１７２を導き出し、続いてこれがデータ・フレーム送信動作を停止させて、ループを一時的に閉じて他の装置がループ１２５０を使用することができるようにする（すなわち、この一実施の形態は、以下で説明する信号１７０を無視し、残りのワードがレジスタ１６８における値よりも小さくなっていない限り、限界に達すれば、転送を停止する。）。他の実施の形態では、フェア

ネス保持特徴１７５に対するｅｎａｂｌｅ＿ＤＨＤ＿ｓｕｐｅｎｄ信号１７０は、ＡＮＤゲート１７１への入力としても供給される。転送を停止させる論理式（ゲート１７１の出力）は、以下のようになる。

【００６６】

【数１】

suspend_xfer signal 172 = (enable_DHD_suspend = '1'

AND (dhd_cnt_out(15:0) = dhd_max_frms(15:0))

AND (xfr_cnt_out(26:0) ≥ dhd_cmpl_len(26:0)))

【００６７】

したがって、ループ・アービトレーション・フェアネスを保持するために、ＤＨＤプリミティブが受信され、かつ、読出しデータが「開」ポート（このポートは「開放」である。）に利用可能であれば、このポートは、ｄｈｄ＿ｍａｘ＿ｆｒｍカウントに達するまでデータを送信したのち、ループを閉じる。転送が終りに近い（すなわち、ｄｈｄ＿ｃｍｐ＿ｌｎにおける値よりも小さい）ときは、このポートは、直ちに閉じるよりも転送を完了させる。転送はそのように行われるので、再びアービトレーションを行う必要はない。ｄｈｄ＿ｍａｘ＿ｆｒｍおよびｄｈｄ＿ｃｍｐ＿ｌｎにおける値は、モード・ページ初期化によるかログ・イン値によりセットされて、この特徴をもっと柔軟性のあるものにする。

【００６８】

データ・フレームが送信されるべきであるとき、ＤＨＤプリミティブが受信されると、カウント６０がリセットされ（すなわち、ゼロの値にセットされ）、次いで、カウンタ１６０は、送信された各データ・フレームに対して“１”だけインクリメントされる。カウンタ１６０が（転送が終り近くになっていない限り）限界レジスタ１６３に保持されている最大許容値に達すると、ループ１２５０に対する次の時間アービトレーションに勝つまで、それ以上のデータ・フレームは送信されない。

【００６９】

「ｓｕｓｐｅｎｄ＿ｘｆｅｒ」出力信号１７２は、ポートＡ，Ｂ開放制御回路４２への入力としてループ制御回路４０によって使用されて（図５および以下の説明を参照）、ＤＨＤが受信されると、ループ上に閉を送信する。そうでなければＤＨＤモードは非常に長い転送を許容してループのフェアネスを損なうので、ｓｕｓｐｅｎｄ＿ｘｆｅｒ信号１７２は、ループ・フェアネスを回復すなわち保持する。

【００７０】

一実施の形態では、ｓｕｓｐｅｎｄ＿ｘｆｅｒ信号１７２はまた、入力ポート開制御ステート・マシン４２／４２’として供給されて、ポート開制御ステート・マシン４２／４２’にループを「閉」にさせ（すなわち、他の装置のポートがそのループに対するアービトレーションを可能にするためにループ１２５０の制御を解放するようにさせ）、これによって、ポートがループ１２５０の制御をアンフェアに維持することを防止する。いくつかの実施の形態では、所定量のデータ（例えば、ワード数）が転送されると、ｓｕｓｐｅｎｄ＿ｘｆｅｒ信号１７２が活性化される。他の実施の形態では、所定の時間が経過すると、ｓｕｓｐｅｎｄ＿ｘｆｅｒ信号１７２が活性化される。これらの形式の実施の形態のうちのいくつかでは、転送されるべき所定量よりも少ないデータ（例えば、フレーム数）が残っているときは、ｓｕｓｐｅｎｄ＿ｘｆｅｒ信号１７２の活性化が禁止される。上記形式の実施の形態のうちのいくつかでは、ｅｎａｂｌ＿ＤＨＤ＿ｓｕｓｐｅｎｄ信号１７０がディセーブル（不活性）されると、ｓｕｓｐｅｎｄ＿ｘｆｅｒ信号１７２の活性化が禁止される。

【００７１】

ＩＩ．ループ初期化および応答用の専用フレーム・バッファ

本発明の一実施の形態に関して、フレーム・バッファは、第３世代特定用途向け集積回路（ＡＳＩＣ）に付加されて、両方のポートが同時に活性であることを可能にさせる。非データ・フレーム（図１の「受信非データ・フレーム・バッファ」５３，５３’とも呼ばれる。）を受信する２つのバッファは、ノードの両方のポートで同時に複数のコマンドおよび複数のＦＣＰフレーム（ｆｉｂｒｅ－ｃｈａｎｎｅｌ－ｐｒｏｔｏｃｏｌフレーム）を受信できるように（および、全二重動作ができるように、すなわち、あるポートの一方のファイバを介して受信すると同時に、同じポートの他方のファイバを介して送信するように）設けられている。これは、ディスク・ドライブ１００が転送の中断または終了まで待機するよりも同じポートを介したおよび／または他のポートを介したデータ転送中に一方のポートを介して新しいコマンド（または、他の非データ・フレーム）を受信できるようにさせる。従来のアプローチにおけるよりも早くコマンドを有することによって、本発明は、データ転送が進行している間にコマンドを分類して最適化することを可能にさせ、これによって、システム１２００の性能を改善する。

【００７２】

図３は、ファイバ・チャネル・ノード・インターフェース・チップ１１０のブロック図である。本発明におけるファイバ・チャネル・ノード・インターフェース・ロジック１１０は、アービトレーテッド・ループ・ロジックおよびフレーミング・ロジックを含むファイバ・チャネル・プロトコルに対して責任を持つ。一実施の形態は、ファイバ・チャネル・プロトコル（ＦＣＰ）標準によって定義されたＳＣＳＩ上位プロトコルのみを使用してクラス－３ＳＣＳＩ実施（上述したＦＣ　ＡＬ仕様）に対して最適化される。ファイバ・チャネル・ノード・インターフェース・ロジック１１０は、種々のバッファ帯域幅をサポートするとともに、二重ポートおよび全二重動作を支援するように４つのオン・チップ・フレーム・バッファ（５３，５３’５５および７３）を含む。ファイバ・チャネル・ノード・インターフェース・ロジック１１０はまた、マイクロプロセッサ１１２に対してインターフェースを行い、マイクロプロセッサ１１２がファイバ・チャネル・ノード・インターフェース・ロジック１１０を構築するとともにファイバ・チャネル・ノード・インターフェース・ロジック１１０の現在状態についてのステータス情報を読み出せるようにする。

【００７３】

ファイバ・チャネル・ノード・インターフェース・ロジック１１０は、２つのループ・ポート回路２０（一方はポートＡ用で、他方はポートＢ用であり、各ポートはループ通信をサポートするためにデータ入力インターフェースおよびデータ出力インターフェースを有する。）と、ループ制御回路４０（フレーム送信回路４０とも呼ばれる。）と、受信パス・ロジック５０と、転送制御回路６０と、単一フレーム送信回路７０と、送信パス・マルチプレクサ（ｍｕｘ）７９と、データ・フレーム送信パス・ロジック８０と、マイクロプロセッサ・インターフェース９０とを有する。これらのブロックは、受信フレーム処理，送信データ・フレーム発生，単一送信フレーム発生，転送制御およびプロセッサ・インターフェース処理のような機能をサポートする。

【００７４】

マイクロプロセッサ・インターフェース回路９０は、マイクロプロセッサ１１２がファイバ・チャネル・ノード・インターフェース・ロジック１１０のレジスタおよびカウンタにアクセスできるようにする。（「マイクロプロセッサ」を説明するとき、そのような用語は任意の適当なプログラマブル・ロジック装置を含む。）インターフェース・レジスタは、ファイバ・チャネル・インターフェースの応答よりも前に外部マイクロプロセッサ１１２によって初期化される。出力転送はこのインターフェースを介して初期化され、また、受信された転送のステータスはこのインターフェースを介して入手可能である。

【００７５】

図３用の入力信号は、ファイバ・チャネル１６からのデータ入力をポートＡ用のループ・ポート回路２０に搬送するＡ＿ＩＮ３０２１と、ファイバ・チャネル１６からのデータ入力をポートＢ用のループ・ポート回路２０に搬送するＢ＿ＩＮ３０２２とを含む。ＤＡＴＡ　ＦＲＯＭ　ＯＦＦ‐ＣＨＩＰ　ＢＵＦＦＥＲ３０５１は、オフ・チップ・バッファ１

１１からのデータを受信パス５０に搬送する。ＴＯ　ＯＦＦ‐ＣＨＩＰ　ＢＵＦＦＥＲ３０５２は、受信パス５０からオフ・チップ・バッファ１１１にデータを搬送する。ＢＵＦＦＥＲ　ＳＴＡＴＵＳ３０６１は、転送制御６０にステータスを提供する。ＭＰＵインターフェース９０へのＭＰＵ　ＡＤＲＥＳＳ３０９１およびＭＰＵ　ＤＡＴＡ３０９５はマイクロプロセッサ１１２からのアドレスおよびデータをそれぞれ供給する。ＭＰＵインターフェース９０へのＲＥＡＤ＿ＥＮＡＢＬＥ３０９２およびＷＲＩＴＥ＿ＥＮＡＢＬＥ３０９３はマイクロプロセッサ１１２からのイネーブル信号を供給する。信号ＭＰＵ３０７６は、マイクロプロセッサ１１２が送信フレームバッファ７３をアクセスすることを可能にさせる。Ａ＿ＯＵＴ３０２３は、ポートＡに対してループ・ポート回路２０からファイバ・チャネル１６にデータを搬送し、また、Ｂ＿ＯＵＴ３０２４は、ポートＢに対してループ・ポート回路２０からファイバ・チャネル１６にデータを搬送する。

【００７６】

図４は、ファイバ・チャネル・ループ・ポート回路２０のブロック図である。本発明の一実施の形態のファイバ・チャネル設計は、周辺装置の直接アタッチメント用の二重ポート・ファイバ・チャネル・インターフェースをサポートするために２つの同じループ・ポート回路２０を含む。一実施の形態では、ファイバ・チャネル・ループ・ポート回路２０は、受信レジスタ２１と、８Ｂ／１０Ｂデコーダ・ロジック２２と、ワード同期ステート・マシン２３と、受信クロック喪失検出器２４と、同期喪失タイマ２５と、アービトレーテッド・ループ・ロジック２６と、８Ｂ／１０Ｂエンコーダ２７とを含む。

【００７７】

一実施の形態では、各ループ・ポート回路２０は、１０ビット・データ・インターフェースを使用して外部トランシーバ１１５（図１を参照）にインターフェースする。そのような実施の形態では、トランシーバ１１５は、並列インターフェース（例えば、１０ビット幅または２０ビット幅インターフェース）におよびそれからシリアル・データを並直列変換しかつ直並列変換する。他の実施の形態では、これらのトランシーバ１１５はチップ１１０に集積される。（ファイバ・チャネルから入力される）並列データは、各トランシーバ１１５の受信機部からの受信クロックを使用して捕捉されて、並列８Ｂ／１０Ｂデコーダを使用してデコードする前に２０ビット幅フォーマットに変換される。次いで、（特別な順序集合（ｏｒｄｅｒｅｄ　ｓｅｔ）を表すために使用された）１６ビット・データ＋２ｋキャラクタは、アービトレーテッド・ループ・ロジック２６に配置される前にワードの有効性についてチェックされる。アービトレーテッド・ループ・ロジック２６の出力は、送信機クロックに再同期されて、受信フレーミング・ロジックに渡されてもよいし、エンコーダ２７を介してループ１２５０上に再送信されてもよい。一実施の形態では、エンコーダ２７は、各動作中に１つの８ビット・キャラクタを１つの１０ビット・キャラクタに変換する。他の実施の形態では、２つ以上の８ビット・キャラクタが、対応した数の１０ビット・キャラクタに各動作において変換される。（「１６Ｂ／２０Ｂエンコーダ」と題する米国特許第５，６６３，７２４号を参照。）アービトレーテッド・ループ・ロジック２６は、ループ・ステート・マシンと順序集合デコーダと弾性挿入および削除機能とを含む。ループ・ポート回路２０は、ファイバ・チャネル・アービトレーテッド・ループＡＮＳＩ標準（すなわち、上述したＦＣ－ＡＬおよび／またはＦＣ－ＡＬ２）に定義されているアービトレーション・ループ・プロトコルを実施する。

【００７８】

一実施の形態では、ファイバ・チャネル・データは、シリアルに送信されて、トランシーバ１１５によって１０ビット並列データに変換される。受信レジスタ２１は、トランシーバ１１５の受信機部によって発生されたクロックを使用してトランシーバ１１５からの１０ビット・データ（Ａ＿ＩＮ３０２１またはＢ＿ＩＮ３０２２）を捕捉する。データは、８Ｂ／１０Ｂデコーダ２２を通過する前に２０ビット幅（すなわち、１０ビット・キャラクタ幅）に直ちに変換される。一実施の形態では、デコーダ２２は「８Ｂ／１０Ｂデコーダ」と呼ばれているが、各動作中に１つの１０ビット・キャラクタを１つの８ビット・キャラクタに変換する。他の実施の形態では、２つ以上の１０ビット・キャラクタは、対応

する数の８ビット・キャラクタに各動作において変換される。
【００７９】
８Ｂ／１０Ｂデコーダ２２は、受信レジスタ２１によって捕捉された符号化データを入力する。２つの１０ビット・キャラクタは、並列にデコードされて、２つの８ビット・キャラクタを出力する。入力キャラクタのランニング・ディスパリティがチェックされ、また、エラー・ステータスがワード同期ステート・マシン２３とアービトレーテッド・ループ・ロジック２６とに渡される。負のランニング・ディスパリティは、ランニング・ディスパリティ・エラーに従って次の順序集合上に強制設定される。符号化規則の違反もチェックされ、コード違反ステータスがワード同期ステート・マシン２３に渡される。
【００８０】
受信クロック喪失検出器２４は、トランシーバ１１５からの受信クロックが停止した時を検出する。受信クロック喪失が検出されると、ワード同期ステート・マシン２３がリセットされて、データはアービトレーテッド・ループ・ロジック２６のＦＩＦＯに行くのが阻止される（ＦＩＦＯは、ファースト・イン・ファースト・アウト・メモリであり、典型的には、異なる速度を有するバスまたは処理間をインターフェースするのに使用される）。カレント・フィル・ワード（ＣＦＷ、以下で更に説明する）は、ワード同期が再獲得されるまで送信される。
【００８１】
ワード同期ステート・マシン２３は、ワード同期用の入力ストリームを監視する。３つの有効な順序集合が正しいバイト／制御キャラクタ・アライメントで検出されたときに、ワード同期が検出され、かつ、干渉のない無効キャラクタが検出される。「ワード同期喪失」がＦＣ－ＰＨ（すなわち、ＦＣ－ＰＨ物理およびシグナリング・インターフェースＸ３Ｔ１１／プロジェクト７５５Ｄ／Ｒｅｖ．４．３）標準によって定義されている。ワード同期が達成されると、データがアービトレーテッド・ループ・ロジック２６のＦＩＦＯに入力される。
【００８２】
同期喪失タイマ２５は、（３つの有効な順序集合を検出するのに１フレーム時間かかるかもしれないので）ワード同期喪失条件が１フレーム時間以上の間存在した時を判断するために使用される。このタイマが時間切れとなったときは、マイクロプロセッサ１１２は、ＬＯＳＳ－ＯＦ－ＳＹＮＣ割込み信号４０２５で割り込みされる結果、処理を開始することができる。
【００８３】
アービトレーテッド・ループ・ロジック２６は、ループ弾性ＦＩＦＯと、ループＦＩＦＯ制御ロジックと、順序集合デコード・ロジックと、ループ・ステート・マシン・ロジックと、カレント・フィル・ワード選択ロジックと、ループ出力マルチプレクサ・ロジックと、種々の機能とを含む。ループ弾性ＦＩＦＯは、（受信クロックによってクロッキングされた）入力データを送信クロックと再同期するために必要なバッファリングを提供する。ループＦＩＦＯ制御ロジックは、アービトレーテッド・ループ・ロジック２６のステートを監視して、挿入または削除動作を必要であるか否かを判断する。順序集合は、順序集合認識ロジックによってデコードされる。これらの順序集合は、フレーム・デリミタおよびアービトレーテッド・ループ順序集合を含むＦＣ－ＰＨ定義順序集合（すなわち、ＦＣ－ＰＨ物理およびシグナリング・インターフェースＸ３Ｔ１１／プロジェクト７５５Ｄ／Ｒｅｖ．４．３）を含む。カレント・フィル・ワード選択ロジックは、ループ・ステータスおよびデコードされた順序集合を監視してカレント・フィル・ワード（ＣＦＷ）を決定する。アービトレーテッド・ループがイネーブルされると、ハードウェア・ステート・マシンは、順序集合デコーダを使用して、ＦＣ－ＡＬ標準（すなわち、ファイバ・チャネルＦＣ－ＡＬ１アービトレーテッド・ループ標準Ｘ３Ｔ１１／プロジェクト９６０Ｄ／Ｒｅｖ．４．５またはファイバ・チャネルＦＣ－ＡＬ２アービトレーテッド・ループ標準Ｘ３Ｔ１１／プロジェクト１１３３Ｄ／Ｒｅｖ．６．３）に説明されている複数のループ機能を実行する。入力ＬＯＯＰ　Ａ　ＴＲＡＮＳＭＩＴ　ＣＯＮＴＲＯＬ　ＯＵＴＰＵＴ６４２

５およびＬＯＯＰ　Ｂ　ＴＲＡＮＳＭＩＴ　ＣＯＮＴＲＯＬ　ＯＵＴＰＵＴ６４２７は、図５においてロジックからアービトレーテッド・ループ・ロジック２６への入力を供給する。出力ＬＯＯＰ　Ａ　ＳＴＡＴＥＳ　ＡＮＤ　ＣＯＮＴＲＯＬ６４２２，ＬＯＯＰ　Ｂ　ＳＴＡＴＥＳ　ＡＮＤ　ＣＯＮＴＲＯＬ６４３２は、各ループの出力を制御して、ループ制御ロジックにステータスを提供し、続いて、これがループ制御４０に要求を発生する（図５を参照）。出力ＬＯＯＰ　Ａ　ＤＡＴＡ４０２６，ＬＯＯＰ　Ｂ　ＤＡＴＡ４０２７は、データを各ローカル・ポートに供給する。

【００８４】

一実施の形態では、８Ｂ／１０Ｂエンコーダ・ロジック２７は、アービトレーテッド・ループ・ロジック２６から１６ビット・データおよび２ｋキャラクタ（下位ｋが常に０である。）を受け取る。一実施の形態では、入力は、２つの１０ビット・キャラクタに符号化され、これらは分離されてトランシーバ１１５（図１参照）に１度に１つ出力され、トランシーバ１１５はデータをシリアル・ストリームに変換する。他の実施の形態では、両方とも１０ビット・キャラクタ（すなわち、２０ビット）が並列にトランシーバ１１５に送出され、トランシーバ１１５はデータをシリアル・ストリームに変換する。送信マルチプレクサ７９（図３を参照）はまた、エンド・オブ・フレーム（ＥＯＦ）・デリミタが転送されている時を表すためにステータスを提供して、エンコーダ２７が現在ランニング・ディスパリティに基づいて正しい形式（すなわち、フレイバ（ｆｌａｖｏｒ））のＥＯＦを選択できるようにする。また、（Ｏｐｅｎｄステートにある）ポートが送信しているとき、または、アービトレーテッド・ループ・ロジック２６がプリミティブを送信しているときは、各非ＥＦＯプリミティブの開始時にランニング・ディスパリティが強制的に負にされる。出力信号Ａ＿ＯＵＴ３０２３およびＢ＿ＯＵＴ３０２４は各トランシーバ１１５，１１５’にデータを送信する。

【００８５】

図５は、ループ制御回路４０（フレーム送信（ＸＭＩＴ）回路４０とも呼ばれる。）のブロック図である。ループ制御回路４０（図３および図５を参照）は、複数フレームすなわちＲ＿ＲＤＹを送信し始めるために（ポートＡおよびポートＢのアービトレーテッド・ループ・ロジック２６における）適当なアービトレーテッド・ループ・ステート・マシンに対する要求を発生するとともに送信フレーミング・ステート・マシン７２，８１に対する要求を発生する制御ロジックを含む。

【００８６】

送信データ・シーケンサ・ロジック４１は、転送がマイクロプロセッサ１１２によって要求されるときに活性化されるロジックを含む。送信データ・シーケンサ・ロジック４１は、入力信号ＴＲＡＮＳＭＩＴ　ＳＴＡＴＵＳ　ＩＮＰＵＴＳ６４１１を使用して転送を監視するとともに、転送の各段階に関して「イネーブル」（すなわち、イネーブル信号ＴＲＡＮＳＭＩＴ　ＣＯＮＴＲＯＬ　ＯＵＴＰＵＴＳ６４１３）を発生する。これは、転送レディーおよびＦＣＰ応答がマイクロプロセッサ１１２の介入なしに発生されることを可能にさせる。

【００８７】

ループ・ポートＡ／Ｂ開放制御ステート・マシン４２（ポートＡ），４２’（ポートＢ）は、ポートが他のＬ＿ポートによって開放される場合、または、ループ１２５０がフレームを送信するために開放される場合を取り扱う。このロジックは、アービトレーションする要求とループ１２５０を閉じる要求とＲ＿ＲＤＹおよび種々のフレームを送信する要求とを発生し、半二重または全二重動動作に対して構成することができる。

【００８８】

次の条件は、アービトレーションする要求を開始するために満足される必要がある。

－マイクロプロセッサ１１２から送信ポート・イネーブル付きフレームを送信する要求

－送信ポートが監視ステートにある

－転送長カウントがゼロでない

－転送を中断するマイクロプロセッサ１１２からの要求がない

－（非データ転送、または、満足されたデータしきい値でまだ送信されていない転送レディー付きデータ書込み転送、または、満足されたデータしきい値および満足されたデータ・フレーム・バッファしきい値でのデータ読出し転送）。

【００８９】

ポートが半二重モード用に構成されているときは、Ｒ＿ＲＤＹはＯｐｅｎｅｄステートにおいてのみ送信することができる。ポートが全二重モード用に構成されているときは、Ｏｐｅｎｅｄステートまたは Ｏｐｅｎステートにおいて Ｒ＿ＲＤＹを送信することができる。Ｒ＿ＲＤＹを送信させる条件は、“利用可能なＢｕｆｅｒ－ｔｏ－Ｂｕｆｆｅｒ Ｃｒｅｄｉｔ（ＢＢ＿Ｃｒｅｄｉｔ）および最大ＢＢ＿Ｃｒｅｄｉｔよりも小さい未解決Ｒ＿ＲＤＹ”を含む。（Ｂｕｆｅｒ－ｔｏ－Ｂｕｆｆｅｒ制御ロジック６０３は、接続されたポートにバッファ・クレジットを発行してフレームを送出できるようにする。このクレジットは、Ｒ＿ＲＤＹを送出することによって発行される。）

【００９０】

ポートが半二重モード用に構成されているときは、Ｏｐｅｎステートにおいてのみフレームを送信することができる。ポートが全二重モード用に構成されているときは、Ｏｐｅｎステートにおいて、または、ポートがフレーム受信者によって全二重モードで開放されればＯｐｅｎｅｄステートにおいて、フレームを送信することができる。

【００９１】

フレームを送信する要求は、次の条件が全て満足されたときに発生される。

－データ・フレーム・バッファ５５が、利用可能なデータを有する。

－Ｂｕｆｅｒ－ｔｏ－Ｂｕｆｆｅｒ Ｃｒｅｄｉｔが利用可能である（受信されたＲ＿ＲＤＹ）。

－（ブロック６０９において）非データ転送またはデータ読出し転送および転送長カウンタがゼロでない。

【００９２】

ループ１２５０を閉にする（ポート１１６によって解放されるべき通信チャネルの制御）条件は、次のものを含む。

－Ｏｐｅｎｅｄステートに入ったときにＢｕｆｅｒ－ｔｏ－Ｂｕｆｆｅｒ Ｃｒｅｄｉｔが利用可能でない。

－フェアネスのために、すなわち、（図１の制限レジスタ１６３で指定された）転送可能とされるフレームの数についての所定の制限がＤＨＤカウンタ１６０によって到達し、かつ、（オプションとして）（最小長レジスタ１６３によって指定された）最小長よりも多いフレームの数が転送されるべく残っており、かつ、（オプションとして）ｅｎａｂｌｅ＿ＤＨＤ＿ｓｕｓｐｅｎｄ信号１７０が活性である。

－未解決Ｒ＿ＲＤＹがなく、かつ、Ｏｐｅｎｅｄステートのときにそれ以上のＢＢ＿Ｃｒｅｄｉｔが利用できない。

－ポートがＯｐｅｎｅｄステートにあるときにプロセッサ・ビジー要求が活性である。

－転送が完了した。

－データ読出し転送動作およびデータが利用可能でない。

－ＣＬＳプリミティブが受信され、かつ、それ以上のＢＢ＿Ｃｒｅｄｉｔが利用可能でない。

－マイクロプロセッサ中断要求が待機中であり、かつ、ロジックがフレーム間にある。

【００９３】

図５において、ループ・ポートＡ／Ｂ開放割込み制御ステート・マシン４６（ポートＡ），４６’（ポートＢ）は、ループ１２５０がＯｐｅｎ割込みステートにあるときの場合を取り扱う。このロジック４６，４６’は、フレームを送信する要求を発生する。各ポート（それぞれ４６，４６’）に対して１つのステート・マシンがある。これらのステート・マシンは、マイクロプロセッサ１１２がフレームを要求するときに、フレームを送信する要求を発生し、かつ、ＥＯＦの送信を監視する。送信が完了すると、送信完了がマイクロプロセッサ１１２に対して発生される。

【００９４】
ブロック４０への入力は、ＰＯＲＴ　ＢＢ＿ＣＲＥＤＩＴ　ＡＶＡＩＬＡＢＬＥ　ＴＯ　ＴＲＡＮＳＭＩＴ　Ｒ＿ＲＤＹ６０１７と、ＰＯＲＴ　ＣＲＥＤＩＴ　ＡＶＡＩＬＡＢＬＥ　ＴＯ　ＴＲＡＮＳＭＩＴ　Ａ　ＦＲＡＭＥ６０２０と、ＬＯＯＰ　Ａ　ＳＴＡＴＥＳ　ＡＮＤ　ＣＯＮＴＲＯＬ６４２２と、ＬＯＯＰ　Ｂ　ＳＴＡＴＥＳ　ＡＮＤ　ＣＯＮＴＲＯＬ６４３２（図４を参照）と、ＤＡＴＡ　ＡＶＡＩＬＡＢＬＥ６０１９とを含む。ブロック４０からの出力は、ＴＲＡＮＳＭＩＴ　ＣＯＮＴＲＯＬ　ＯＵＴＰＵＴＳ６４１３と、ＬＯＯＰ　Ａ　ＴＲＡＮＳＭＩＴ　ＣＯＮＴＲＯＬ　ＯＵＴＰＵＴＳ６４２５と、ＬＯＯＰ　Ｂ　ＴＲＡＮＳＭＩＴ　ＣＯＮＴＲＯＬ　ＯＵＴＰＵＴＳ６４２７とを含む。

【００９５】
ＩＩＩ．フレーム受信用の専用フレーム・バッファ
二重ポート・ファイバ・チャネル・アービトレーテッド・ループ設計１２００では、オン・チップ・フレーム・バッファ１１９のバッファを使用して、入フレームおよび出フレームを管理することができる。受信されたフレームおよび送信されたフレームは、通常、より低い転送速度で大きなオフ・チップ領域（例えば、オフ・チップ・バッファ１１１）に記憶される。オフ・チップ・バッファ１１１が単一ポートに対して全転送速度が可能なときであっても、二重ポート設計に対しては、必要とされる帯域幅は遙かに大きく、追加コストが加わる。ＦＣ－ＡＬ　ＡＳＩＣ１１０（図１を参照）におけるオン・チップ・フレーム・バッファ１１９は、パフォーマンス，シリコン面積およびコスト間のバランスを考慮して種々の方法で構成される。この仕様は、各ポートに同時に非データ形式フレームを受信するとともに、大きな専用データ・フレーム・バッファ５５（これもオン・チップ・フレーム・バッファ１１９全体の構成要素）を提供するように、専用フレーム・バッファ５３，５３’（オン・チップ・フレーム・バッファ１１９全体の構成要素）の使用を詳述している。本発明による二重ポート設計では、複数のフレームを同時に両方のポート１１６で受信することができる。これらのフレームは、通常、受信された後に大きなオフ・チップ・メモリ１１１に移動されて記憶される。更なる情報は、「ループ初期化および応答用の方法および専用フレーム・バッファ」と題する米国特許出願第０９／１９３，３８７号に見い出すことができる。

【００９６】
ＩＶ．オン・チップ・メモリでデータ保全用にファイバ・チャネルＣＲＣを使用すること
本発明の一態様によれば、ファイバ・チャネル・フレームを一時的に記憶するフレーム・バッファは、最大ファイバ・チャネル・インターフェース・データ転送速度でフレームを受信することを可能とする。その後、より低くより管理し易い速度でオフ・チップ・メモリにフレームを転送することができる。パリティ，ＣＲＣまたは他の冗長機能のような種々のメカニズムは、オプションとして、フレーム・バッファにデータを記憶している間にデータを保護するために使用される。

【００９７】
一実施の形態では、受信したファイバ・チャネル・サイクリック・リダンダンシー・コード（ＣＲＣ）をデータとともにフレーム・バッファ通過させることによってデータ保全チェックが強化され（すなわち、ＣＲＣは、フレームとともにフレーム・バッファに記憶されて、後の時点でフレームとともに読み出される）、ＲＡＭをより幅広にする付加的なパリティ・ビットを省略することができる。

【００９８】
Ｖ．アービトレーテッド・ループ・オーバヘッドを軽減する方法および装置
ファイバ・チャネル・アービトレーテッド・ループ設計１２００では、ループ・ポート１１６のノード・インターフェース１２２０は、ループ１２５０にアクセスするためにアービトレーションする必要がある。どのポートがループ１２５０の制御を獲得するのかを決定するために優先度システムが使用され、また、ポートが足りなくないことを保証するために「フェアネス」方式が使用される。目標装置として、ドライブ１００は、通常、ＣＰＵ情報処理システム１２０２よりも低い優先度が与えられ、その結果、ドライブ１００は

、複数のより高い優先度の装置がそれらのアクセスを完了するまで、アービトレーションに勝つのを待たねばならないかもしれない。ループ・ポート１１６のノード・インターフェース１２２０がループ１２５０の制御を獲得すると、それは、不必要なアービトレーション・サイクルを避けるために、ループ１２５０を閉じる前に、可能な限り多くのフレームを送出する。しかし、データがもはや得られないときは、ループ・ポート１１６のノード・インターフェース１２２０は、ループ１２５０を閉じて、他のポートがループ１２５０にアクセスすることを許容する。これは、ある他のコントローラ・アーキテクチャーに使用される方法である。本発明は、ポートのデータ利用可能性に基づいて、ループ１２５０を閉じるのか否かの判断に関するルールを変更することによってループ・パフォーマンスを高め、したがって総合ループ・オーバーヘッドを軽減する機構を提供する。

【００９９】

ある他のコントローラ・アーキテクチャーでは、エンド・オブ・フレーム・デリミタが送信されているとき、ポートは、他のフレームが利用可能か否かを判断する。データがもはや得られないときは（例えば、全フレームが送信用に利用できないときは）、ループ１２５０が閉じられる。その直ぐ後にデータが再び利用可能となるので、ポートは、後で再びアービトレーションし、転送を継続する前にアービトレーションに勝つ必要がある。転送の最終フレームが利用可能となるときにこれが発生すれば、転送の完了は遅延され、次のコマンドが可能となる前に遅延をもたらすかもしれない。本発明は、データがポートに対して直ぐに利用可能となるのであれば、あるポートによってループ１２５０を開放のままにすることができるコントローラ・アーキテクチャー設計用の機構を提供する。一実施の形態では、次の条件の両方が満足されるときは、更なるデータがポートに利用可能となるのが十分に予測される際は（ポートがそのループの制御を保持するのを正当化するために）、ループ１２５０を開放したままとする。

－少なくともＸフレームが利用可能なオフ・チップであり、かつ

－データの少なくともＹワードがデータ・フレーム・バッファ５５で利用可能である。

一実施の形態では、所定の量のデータが利用可能（オン・チップでは少なくとも１／２フレームが利用可能で、オフ・チップでは少なくとも１フレームが利用可能）であれば、ループ１２５０は開に保持されるが、フレームの転送は、１つのフレーム全部がオン・チップで利用可能になるまで、開始しないであろう。

【０１００】

本発明の１つの目的は、データがポート１１６に対して直ぐに利用可能であるときは、ループ１２５を開に保持して、付加アービトレーション・サイクルを回避することである。延長された期間（例えば、ヘッド切換えを実行するのに必要な時間）の間待機となるのであれば、データが利用可能となるまで待機してループ１２５０を開のままにしてはならない。なぜならば、これは、ループ１２５０上で他のポートが転送を実行するのを阻止するからである。

【０１０１】

（結論）

以上、１つ以上のポート（１１６）を有する第１のチャネル・ノード（１２２０）を含むループ・フェアネスを保持する改良された通信チャネル・システム（１２００）について説明した。各ポート（１１６）は、ファイバ・チャネル・アービトレーテッド・ループ・シリアル・通信チャネル（１２５０）をサポートし、かつ、これに取り付けられる。これらのポートの１つは、そのポートが取り付けられたチャネル（１２５０）の制御のためにアービトレーションするであろう。ここで、チャネル・ループ（１２５０）の制御は、アービトレーションに勝つと、第１のポートと通信チャネル（１２５０）との間で所定量の使用が発生したか否かに少なくとも一部基づいてフェアネス保持装置（１７５）は通信チャネルの制御を解放させる。

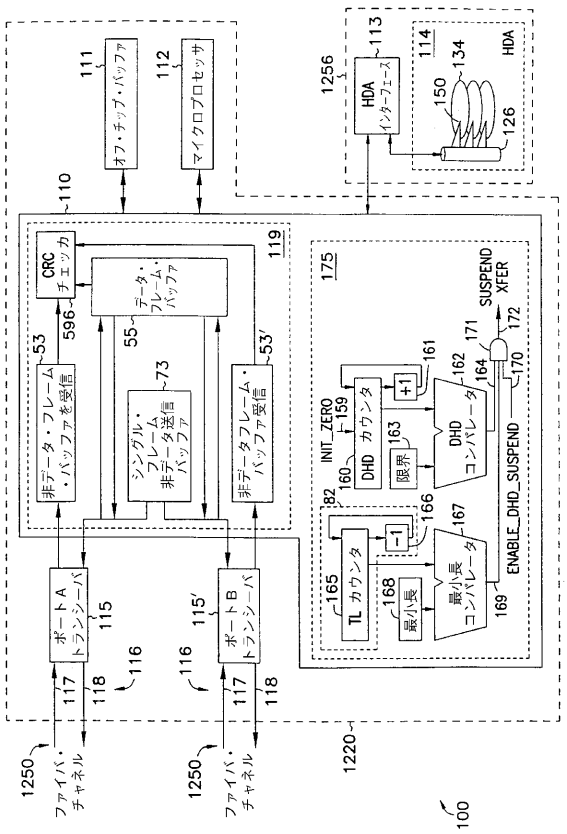
【０１０２】

一実施の形態では、所定量の使用は、通信チャネルの制御を所定の時間保持することを含む。

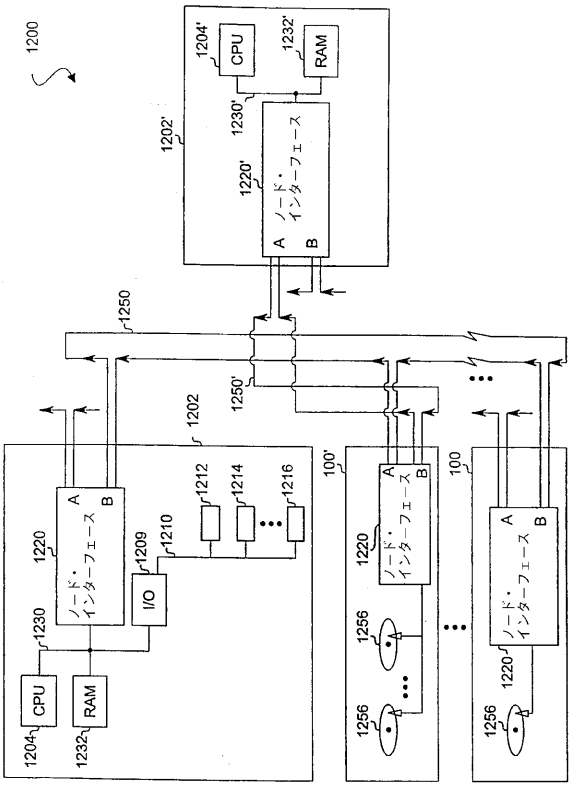
【０１０３】
他の実施の形態では、所定量の使用は、第１の所定量のデータを転送することを含む。そのような一実施の形態では、システム（１２００）は、転送されたデータの量を監視する第１のカウンタ（１６２）をさらに含む。第１のカウンタ（１６０）に作動的に結合された第１のコンパレータ回路（１６２）は、第１のカウンタ（１６０）によって監視されたデータの量が第１の所定量のデータに達したか否かに少なくとも一部基づいて通信チャネルの制御を解放させる。一実施の形態では、第１の所定量のデータはレジスタ（１６３）に保持される。
【０１０４】
いくつかの実施の形態では、システム（１２００）は、転送されるべく残っているデータの量を監視する第２のカウンタ（１６５）をさらに含む。第２のコンパレータ回路（１６７）は、第２のカウンタ（１６５）に作動的に結合され、第２のカウンタ（１６５）によって監視されたデータの量が第２の所定量のデータよりも小さいか否かに少なくとも一部基づいて通信チャネルの制御の解放を禁止する。一実施の形態では、第２の所定量のデータはレジスタ（１６８）に保持される。
【０１０５】
いくつかの実施の形態では、システム（１２００）はダイナミック半二重をサポートし、ここで、第１のカウンタ（１６０）および第１のコンパレータ（１６２）はダイナミック半二重コマンドの第１の部分による受信時に初期化される。
【０１０６】
いくつかの実施の形態では、データの第１の所定量およびデータの第２の所定量は、プログラム可能な量である。
【０１０７】
システム（１２００）のいくつかの実施の形態は、第１のチャネル・ノード（１２２０）に作動的に結合された磁気ディスク記憶装置（１１４）をさらに含む。コンピュータ・システム（１２０２）は第２のチャネル・ノード（１２２０）を含み、ここで、第２のチャネル・ノード（１２２０）は、ファイバ・チャネル・アービトレーテッド・ループ・シリアル通信チャネルを介して第１および第２のチャネル・ノード間でデータを転送するために、ファイバ・チャネル・ループ（１２２０）で第１のチャネル・ノード（１２２０）に作動的に結合されている。
【０１０８】
本発明の他の態様は、（ａ）ファイバ・チャネル・アービトレーテッド・ループ・シリアル通信チャネルのループの制御のためにアービトレーションすることと、（ｂ）第１のポートと通信チャネルとの間で所定量の使用が発生したか否かに少なくとも一部基づいて通信チャネルの制御を解放させることとを含む。
【０１０９】
方法の一実施の形態では、解放するステップ（ｂ）は、（ｂ）（ｉ）通信チャネルの制御が所定の時間維持されたか否かを判断するステップと、（ｂ）（ｉｉ）判断するステップ（ｂ）（ｉ）に基づいてループの制御を解放するステップとをさらに含む。
【０１１０】
方法の他の実施の形態では、解放するステップ（ｂ）は、（ｂ）（ｉｉｉ）第１の所定量のデータが転送されたか否かを判断するステップと、（ｂ）（ｉｖ）判断するステップ（ｂ）（ｉｉｉ）に基づいてループの制御を解放するステップとをさらに含む。そのような一実施の形態では、判断するステップ（ｂ）（ｉｉｉ）は、（ｂ）（ｉｉｉ）（Ａ）第１の値を得るために転送されたフレームの数を監視するステップと、（ｂ）（ｉｉｉ）（Ｂ）第１の値を第１の所定量のデータと比較するステップとをさらに含む。
【０１１１】
方法の他の実施の形態では、解放するステップ（ｂ）は、（ｂ）（ｖ）第２の所定量のデータが転送されたか否かを判断するステップと、（ｂ）（ｖｉ）判断するステップ（ｂ）（ｖ）に基づいてループの制御の解放を禁止するステップとをさらに含む。

【０１１２】
上述した方法のいくつかの実施の形態は、（ｃ）ダイナミック半二重コマンドを受信するステップと、（ｄ）判断するステップ（ｂ）（ｉｉｉ）をダイナミック半二重コマンドの受信時に初期化するステップとをさらに含む。これらの実施の形態のいくつかでは、初期化するステップ（ｄ）は、（ｄ）（ｉ）第１の所定量のデータおよび第２の所定量のデータをプログラム可能にセットするステップをさらに含む。
【０１１３】
本発明の別の他の態様は、ループ・フェアネスを保持するファイバ・チャネル・ノード・コントローラ・システムを提供する。このシステムは、ファイバ・チャネル・アービトレーテッド・ループ・シリアル通信チャネル（１２５０）と、ファイバ・チャネル・アービトレーテッド・ループ・シリアル通信チャネル（１２５０）に取り付けられた第１のポート（１１６）を有する第１のチャネル・ノード（１２２０）であって、第１のポートが、そのポートが取り付けられた通信チャネルの制御のためにアービトレーションする、第１のチャネル・ノード（１２２０）と、所定量の使用が第１のポートと通信チャネルとの間で発生したか否かに少なくとも一部基づいて通信チャネルの制御を解放させる、ここで説明したようなフェアネス保持手段とを含む。
【０１１４】
上述の説明は例示を意図しかつ限定的でないことを理解すべきである。本発明の種々の実施の形態の多数の特徴および効果が種々の実施の形態の構造および機能の詳細とともに以上の説明で明らかにされたが、詳細に対する多くの他の実施の形態および変更は、以上の説明を再検討すると当該技術分野に習熟する者に明らかである。したがって、本発明の範囲は、付記する請求の範囲をこのような請求の範囲が主張する等価物の完全な範囲とともに参照して、判断すべきである。
【図面の簡単な説明】
【図１】　ファイバ・チャネル・ノード・インターフェースを有するディスク・ドライブ１００のブロック図である。
【図２】　本発明を組み入れている情報処理システム１２００のブロック図である。
【図３】　ファイバ・チャネル・ノード・インターフェース・チップ１１０のブロック図である。
【図４】　ファイバ・チャネル・ループ・ポート回路２０のブロック図である。
【図５】　ファイバ・チャネル・ループ制御回路４０のブロック図である。

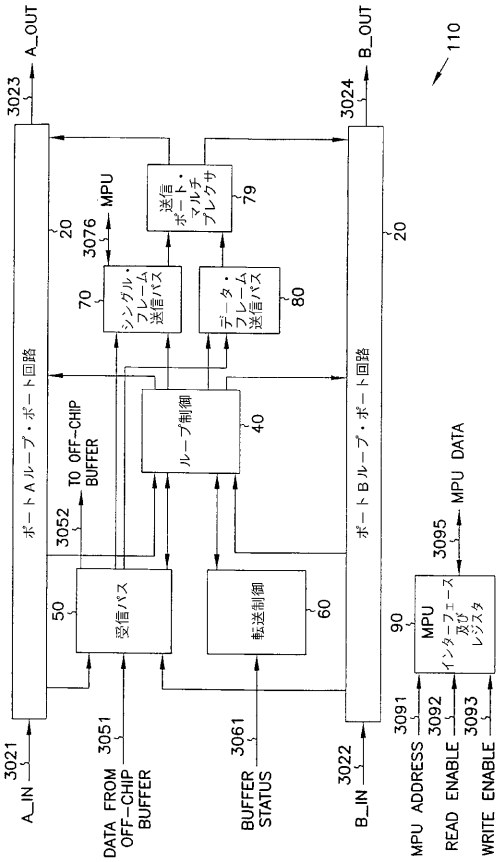
【図１】


オフ・チップ・バッファ
マイクロプロセッサ
HDA
インターフェース
CRC チェッカ
データ・フレーム・バッファ
非データ・フレーム・バッファを受信
シングル・フレーム非データ送信バッファ
非データフレーム・バッファ受信
ポートA トランシーバ
ポートB トランシーバ
INIT_ZERO
DHD カウンタ
限界
DHD コンパレータ
TL カウンタ
最小長
最小長 コンパレータ
ENABLE_DHD_SUSPEND
SUSPEND XFER
ファイバ・チャネル
100
110
111
112
113
114
115
116
117
118
119
126
134
150
159
160
161
162
163
164
165
166
167
168
169
170
171
172
175
596
1220
1250
1256

【図２】


CPU
RAM
ノード・インターフェース
I/O
1200
1202
1202'
1204
1204'
1209
1210
1212
1214
1216
1220
1220'
1230
1230'
1232
1232'
1250
1250'
1256
100
100'

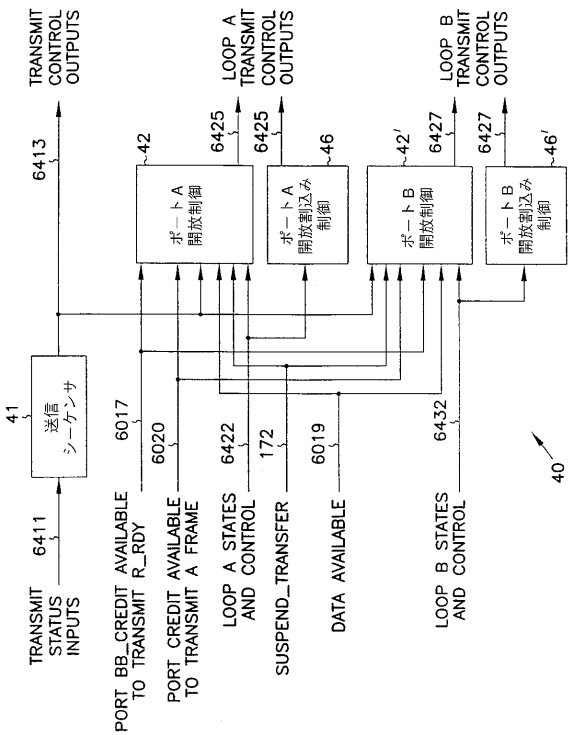
【図３】


ポートAループ・ポート回路
ポートBループ・ポート回路
受信バス
転送制御
ループ制御
送信 ポート・マルチプレクサ
シングル・フレーム送信バス
データ・フレーム送信バス
MPU インターフェース及びレジスタ
A_IN
A_OUT
B_IN
B_OUT
DATA FROM OFF-CHIP BUFFER
TO OFF-CHIP BUFFER
BUFFER STATUS
MPU ADDRESS
READ ENABLE
WRITE ENABLE
MPU DATA
MPU
110
3021
3022
3023
3024
3051
3052
3061
3076
3091
3092
3093
3095

【図４】

受信クロック喪失検出器
同期喪失タイマ
ワード同期ステート・マシン
アービトレーテッド・ループ・ロジック
8B/10B デコーダ
8B/10B エンコーダ
受信レジスタ
LOSS OF SYNC
LOOP A/B DATA
LOOP A/B DATA STATES AND CONTROL
A_OUT/B_OUT
A_IN/B_IN
LOOP A/B TRANSMIT CONTROL OUTPUTS
20
21
22
23
24
25
26
27
3021/3022
3023/3024
4025
4026/4027
6422/6432
6425/6427

【図５】


TRANSMIT CONTROL OUTPUTS
6413
LOOP A TRANSMIT CONTROL OUTPUTS
6425
6425
LOOP B TRANSMIT CONTROL OUTPUTS
6427
6427
42
ポートA開放制御
46
ポートA開放割込み制御
42'
ポートB開放制御
46'
ポートB開放割込み制御
41
送信シーケンサ
40
TRANSMIT STATUS INPUTS
6411
PORT BB_CREDIT AVAILABLE TO TRANSMIT R_RDY
6017
PORT CREDIT AVAILABLE TO TRANSMIT A FRAME
6020
LOOP A STATES AND CONTROL
6422
SUSPEND_TRANSFER
172
DATA AVAILABLE
6019
LOOP B STATES AND CONTROL
6432

フロントページの続き

(72)発明者　ミラー、マイケル、エイチ
アメリカ合衆国　ミネソタ、イーデン　プレイリー、　パーク　ビュー　レーン　６８５０
(72)発明者　ウエストビイ、ジュディ、リン
アメリカ合衆国　ミネソタ、ブルーミントン、ダブリュ、　ワンハンドレッド　アンド　イレブンス　ストリート　７９０６

審査官　矢頭　尚之

(58)調査した分野(Int.Cl.，ＤＢ名)
H04L 12/42