

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7482994号
(P7482994)

(45)発行日 令和6年5月14日(2024.5.14)

(24)登録日 令和6年5月2日(2024.5.2)

(51)国際特許分類		F I	
G 0 6 F	13/12 (2006.01)	G 0 6 F	13/12 3 3 0 A
G 0 6 F	13/10 (2006.01)	G 0 6 F	13/10 3 4 0 A
G 0 6 F	13/14 (2006.01)	G 0 6 F	13/14 3 3 0 C
G 0 6 F	3/06 (2006.01)	G 0 6 F	3/06 3 0 1 Y
G 0 6 F	13/38 (2006.01)	G 0 6 F	13/38 3 2 0 A
請求項の数 8 (全29頁) 最終頁に続く			
(21)出願番号	特願2022-513694(P2022-513694)	(73)特許権者	390009531
(86)(22)出願日	令和2年8月13日(2020.8.13)		インターナショナル・ビジネス・マシー
(65)公表番号	特表2022-548208(P2022-548208		ンズ・コーポレーション
	A)		INTERNATIONAL BUSI
(43)公表日	令和4年11月17日(2022.11.17)		NESS MACHINES CORPO
(86)国際出願番号	PCT/IB2020/057624		RATION
(87)国際公開番号	WO2021/048658		アメリカ合衆国10504 ニューヨー
(87)国際公開日	令和3年3月18日(2021.3.18)		ク州 アーモンク ニュー オーチャード
審査請求日	令和5年1月20日(2023.1.20)		ロード
(31)優先権主張番号	16/567,320		New Orchard Road, A
(32)優先日	令和1年9月11日(2019.9.11)		rmonk, New York 105
(33)優先権主張国・地域又は機関	米国(US)		04, United States of
			America
		(74)代理人	100112690
			弁理士 太佐 種一
			最終頁に続く

(54)【発明の名称】 データ・ストレージ・システムにおけるリンク速度回復

(57)【特許請求の範囲】

【請求項1】

プロセッサが実行する方法であって、

順次リンク速度回復コマンドのメイン・ループの実行を第1の所定の最大回数、繰り返すことを含む、通信リンク上で目標リンク速度を実現すべくリンク速度回復を実行することを含み、リンク速度回復コマンドの各メイン・ループ実行が、各メイン・ループ実行内で順次リンク速度回復コマンドのサブループの実行を第2の所定の最大回数、繰り返すことを含み、順次リンク速度回復コマンドの各サブループ実行が、

順次リンク速度回復コマンドの第1のセットの順次における次のリンク速度回復コマンドを発行すること、

順次における次のリンク速度回復コマンドを発行した後、前記目標リンク速度が実現されているかどうかを判定すること、および

順次における次のリンク速度回復コマンドに応答して、前記目標リンク速度が実現されている場合、リンク速度回復の実行を終了させることを含み、

リンク速度回復コマンドの各メイン・ループ実行が、リンク速度回復コマンドの前記サブループの実行を前記第2の所定の最大回数、繰り返した後、

順次リンク速度回復コマンドの第2のセットの順次における次のリンク速度回復コマンドを発行すること、

順次における次のリンク速度回復コマンドを発行した後、前記目標リンク速度が実現されているかどうかを判定すること、および

順次における次のリンク速度回復コマンドに応答して、前記目標リンク速度が実現されている場合、リンク速度回復の実行を終了させることをさらに含む、前記プロセッサが実行する方法。

【請求項 2】

リンク速度回復コマンドの各メイン・ループ実行が、リンク速度回復コマンドの前記メイン・ループの実行を前記第 1 の所定の最大回数、繰り返した後、リンク速度回復コマンドに応答して、前記目標リンク速度が実現されていない場合、リンク速度回復の実行を終了させることをさらに含む、請求項 1 に記載の前記プロセッサが実行する方法。

【請求項 3】

リンク速度回復コマンドの各サブグループ実行が、

前記第 1 のセットのリンク速度回復コマンドに応答して、前記目標リンク速度が実現されていない場合、リンク速度回復コマンドの前記第 1 のセットの別の順次における次のリンク速度回復コマンドを発行すること、

前記第 1 のセットの別の順次における次のリンク速度回復コマンドを発行することの後、前記目標リンク速度が実現されているかどうかを判定すること、および

前記第 1 のセットの別の順次における次のリンク速度回復コマンドに応答して、前記目標リンク速度が実現されている場合、リンク速度回復の実行を終了させることをさらに含む、請求項 1 に記載の前記プロセッサが実行する方法。

【請求項 4】

リンク速度回復コマンドの各メイン・ループ実行が、

前記第 2 のセットのリンク速度回復コマンドに応答して、前記目標リンク速度が実現されていない場合、リンク速度回復コマンドの前記第 2 のセットの別の順次における次のリンク速度回復コマンドを発行すること、

前記第 2 のセットの別の順次における次のリンク速度回復コマンドを発行することの後、前記目標リンク速度が実現されているかどうかを判定すること、および

前記第 2 のセットの別の順次における次のリンク速度回復コマンドに応答して、前記目標リンク速度が実現されている場合、リンク速度回復の実行を終了させることをさらに含む、請求項 1 に記載の前記プロセッサが実行する方法。

【請求項 5】

順次リンク速度回復コマンドの前記サブグループの順次リンク速度回復コマンドの前記第 1 のセットが、

前記通信リンクに結合された上流ポートにおいてレジスタの構成ビットを設定すること、前記通信リンクに結合された下流ポートのレジスタの構成ビットを設定すべく帯域外信号を送信すること、前記上流ポートのディスエーブル・イネーブル・サイクルを開始すること、および前記下流ポートのディスエーブル・イネーブル・サイクルを開始すべく帯域外信号を送信することのうちの少なくとも 1 つを含む、請求項 1 に記載の前記プロセッサが実行する方法。

【請求項 6】

順次リンク速度回復コマンドの前記メイン・ループの順次リンク速度回復コマンドの前記第 2 のセットが、

順次リンク速度回復コマンドの前記サブグループを開始すること、前記通信リンクに結合された上流光トランシーバを有する上流ポートを含む通信パス・スイッチのリセットおよび再初期設定を開始すること、前記上流光トランシーバをリセットすること、ならびに前記通信リンクに結合された下流光トランシーバをリセットすべく帯域外信号を送信することのうちの少なくとも 1 つを含む、請求項 1 に記載の前記プロセッサが実行する方法。

【請求項 7】

請求項 1 から 6 のいずれか一項に記載の方法の各ステップを実行するためのプロセッサを備えるシステム。

【請求項 8】

請求項 1 から 6 のいずれか一項に記載の方法の各ステップをプロセッサに実行させるた

10

20

30

40

50

めのコンピュータ・プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、データ・ストレージ・システムにおけるリンク速度回復のためのコンピュータ・プログラム製品、システム、および方法に関する。

【背景技術】

【0002】

ストレージ・システムは、通常、ストレージ・コントローラと、ハードディスク・ドライブ、ソリッドステート・ドライブ、テープ・ドライブ、その他などの1つまたは複数のデータ・ストレージ・デバイスを含む。ストレージ・システムは、例えば、ボリュームなどのストレージ・サブユニットにデータを書き込むため、またはそのようなサブユニットからデータを読み取るための入出力命令もしくは入出力コマンドを発行するアプリケーションをホストするホストに、しばしば、接続される。

【0003】

特に企業レベルにおけるデータ・ストレージ・システムは、データ・ストレージ・システムのコンポーネントに障害が生じた際のデータ損失のリスクを低減すべく高いレベルの冗長性をもたらすように通常、設計される。このため、データの多数のコピーが、地理的に分散している可能性がある多数のシステム上に高い頻度で記憶される。ホストからデータ・ストレージ・システムに記憶されるべきデータは、ローカル・サイトにおける一次データ・ストレージ・デバイスに通常、向かわされ、その後、一次データ・ストレージ・デバイスから地理的に遠隔であることが可能な1つまたは複数の二次データ・ストレージ・デバイスにレプリケートされる。

【0004】

一部のコンピューティング環境において、ストレージ・エリア・ネットワークが、多数のホスト・システムと、多数のストレージ・デバイスを制御する多数のストレージ・コントロール・ユニット、例えば、ダイレクト・アクセス・ストレージ・デバイス(DASD)、冗長アレイ・オブ・インデペンデント・ディスク(RAID: Redundant Array of Independent Disks)、ジャスト・ア・バンチ・オブ・ディスク(JBOD: Just a Bunch of Disks)、その他との間の通信パスもしくは通信チャネルを提供する。ストレージ・エリア・ネットワークを経由する通信パスは、スイッチと、光ファイバもしくは他のタイプのケーブルで形成されてよい、または、例えば、無線であってよい通信リンクとを、通常、含む。

【0005】

ホストによって発行される入出力コマンドは、ストレージ・エリア・ネットワーク経由でストレージ・コントローラのポートに向かわせられる。ストレージ・コントローラの各ポートは、そのポートを、ストレージ・エリア・ネットワークの他のポートから区別すべくアドレスまたは他の識別情報を有する。ホストからのI/Oコマンドに応答して、ストレージ・コントローラは、ストレージ・エリア・ネットワークの通信パスによって1つまたは複数のストレージ・コントローラのポートに結合されたストレージ・デバイスからデータを読み取る、またはそのようなストレージ・デバイスにデータを書き込む。ストレージ・コントローラとストレージ・デバイスを結合する通信パスは、ホストとストレージ・コントローラを結合する通信パスと同様に、スイッチと、光ファイバもしくは他のタイプのケーブルで形成されてよい、または、例えば、無線であってよい通信リンクとを、通常、含む。

【0006】

ストレージ・エリア・ネットワークの通信パスのデータ搬送容量は、通常、様々な要因によって制限される。例えば、ストレージ・エリア・ネットワークのデータ搬送パスもしくはデータ搬送チャネルは、例えば、毎秒のバイト数またはビット数などの或る数のデータ単位で測定された指定されたデータ転送速度をサポートするように、しばしば、設計さ

10

20

30

40

50

れる、または構成される。したがって、特定のリンクのデータ搬送容量は、一般に、通信リンクのリンク速度によって制限される。

【0007】

例えば、ペリフェラル・コンポーネント・インタコネクト・エクスプレス（PCIe）リンクもしくはPCIeチャネルなどのシリアル・バス通信パスにおいて、通信リンクは、スイッチなどの上流デバイスとエンドポイント・デバイスなどの下流デバイスとの間でデータが移動することができる多数のレーンを有することがある。それ故、リンクのリンク速度は、リンクを「訓練すること」または「再訓練すること」として、しばしば、参照されるプロセスにおいてパスの上流コントローラと下流コントローラによってネゴシエートされるリンクに割り当てられたレーンのデータ転送レートおよびレーンの数の関数である。

10

【0008】

例えば、PCIeアドイン・カードの形態のスイッチが、ストレージ・コントローラの通信パス・ホスト（ルート・コンプレックス）に差し込まれたとき、ストレージ・デバイスにストレージ・コントローラを接続するリンクの上流コントローラと下流コントローラが、データ転送レートを含む様々なリンク・パラメータをネゴシエートすべく「訓練シーケンス」を交換する。このことは、比較的低いデータ転送レートを提供する単一のレーンを確立することからリンクを通常、開始するリンク訓練およびステータス状態マシン（LTSSM）の実行を介して実現される。しかし、リンク訓練プロセスによってレーンが追加されるにつれ、データ転送のために利用可能なリンク速度は、本明細書において「完全速度」もしくは「完全リンク速度」として参照される目標リンク速度が実現されるまで、増加する。

20

【0009】

訓練プロセスが終わると、データが、訓練プロセスによって実現された増加されたデータ転送レートで転送されてよい。しかし、リンク訓練プロセスは、リンクに関する完全速度もしくは目標速度を実現していないことがある。さらに、完全速度が、リンク訓練によって実現されている場合でさえ、データ伝送条件は変化して、リンク速度を、より低い、最適リンク速度未満もしくは目標リンク速度未満の速度に「下方訓練する」通信パスによるリンク・パラメータの再ネゴシエーションを生じさせることがある。また、リンクのレーンのうちの一部またはすべてのレーン上のデータ伝送が、様々な要因に起因して失われることもある。

30

【0010】

リンクが動作していない、または完全速度未満で動作している場合、通信パスは、完全速度を実現することを期待して、または完全速度を復元するようにリンクを再訓練すべくリンク・パラメータを再び再ネゴシエートしてよい。しかし、そのような訓練または再訓練の試みは、リンクを完全速度に復元することに高い頻度で失敗することを認識されたい。

【0011】

例えば、PCIe規格は、通常、ルート・コンプレックスである通信パス・ホストが、リンク上流コントローラまたはリンク下流コントローラのPCIエクスプレス・ケイパビリティ・リンク・ステータス・コントロール（PCI Express Capability Link Status Control）レジスタの構成ビットを設定することによってリンク再訓練を開始すべきことを規定する。しかし、現在、確立されているリンクのレーンが存在しない場合、ルート・コンプレックスは、下流コントローラの構成ビットを設定すべくリンク経由で下流コントローラと通信することができない。さらに、リンク・コントローラのPCIエクスプレス・ケイパビリティ・リンク・ステータス・コントロール・レジスタの構成ビットを設定することは、リンクを完全速度に復元するリンク再訓練を開始することに高い頻度で失敗する。

40

【0012】

リンク訓練を開始する知られている技術は、ルート・コンプレックスが、PCIe通信パスのハードウェア・デバイスをリセットする、またはディスエーブルにするーイネーブルにすることである。例えば、PCIeポートの光学トランシーバが、リセットされてよ

50

く、または1つまたは複数のP C I eポートを提供するP C I eカード全体が、リセットされてよい。別の実施例として、P C I eポートが、通信リンクを訓練すること、または再訓練することを開始すべくイネーブル・ディスエーブル・サイクルにおいて巡回させられてよい。この場合も、現在、確立されているリンクのレーンが存在しない場合、ルート・コンプレックスは、下流コンポーネントをリセットすべくリンク経由で下流コントローラと通信することができない。さらに、P C I eコンポーネントをリセットすること、または巡回させることは、リンクを完全速度に復元するリンク再訓練を開始することに高い頻度で失敗する。

【発明の概要】

【0013】

本説明によるデータ・ストレージ・システムにおけるリンク速度回復は、コンピュータ技術の重要な改良をもたらす。一態様において、通信リンク上で目標リンク速度を実現すべくリンク速度回復を実行することが、順次リンク速度回復コマンドのメイン・ループの実行を所定の最大回数、繰り返すことを含む。さらに、リンク速度回復コマンドの各メイン・ループ実行が、各メイン・ループ実行内の順次リンク速度回復コマンドのサブループの実行を別の所定の最大回数、繰り返すことを含む。

【0014】

一実施形態において、順次リンク速度回復コマンドのサブループに関する順次コマンド・セットの実行は、メイン・ループに関する順次コマンド・セットの実行と比べて、より高い頻度で成功することが可能である。それ故、各メイン・ループ内の順次リンク速度回復コマンドのサブループに関する順次コマンド・セットを繰り返し、したがって、順次リンク速度回復コマンドのメイン・ループに関する順次コマンド・セットの繰り返しよりも高い頻度で繰り返すことが、リンク訓練またはリンク再訓練が完全リンク速度の実現または回復をもたらす成功率を向上させることができる。

【0015】

別の態様において、各メイン・ループ内の順次リンク速度回復コマンドの各サブループ実行が、オプションとして、順次リンク速度回復コマンドのサブループ・セットの順次における次のリンク速度回復コマンドを発行すること、順次における次のリンク速度回復コマンドを発行した後、目標リンク速度が実現されているかどうかを判定すること、および順次における次のリンク速度回復コマンドに応答して目標リンク速度が実現された場合、リンク速度回復の実行を終了させることを含んでよい。さらに、リンク速度回復コマンドの各メイン・ループ実行は、オプションとして、リンク速度回復コマンドのサブループの実行を所定の最大回数、繰り返した後、メイン・ループに関する順次リンク速度回復コマンドのメイン・ループ・セットの順次における次のリンク速度回復コマンドを発行すること、順次における次のリンク速度回復コマンドを発行した後、目標リンク速度が実現されているかどうかを判定すること、および順次における次のリンク速度回復コマンドに応答して目標リンク速度が実現された場合、リンク速度回復の実行を終了させることを含んでよい。

【0016】

特定のリンク速度回復コマンドの実行は、目標リンク速度を実現することも、実現しないこともあることを理解されたい。順次における次のリンク速度回復コマンドを発行した後、目標リンク速度が実現されているかどうかを判定することによって、リンク速度回復は、リンク速度回復の効率を向上させるべく、目標リンク速度回復が実現されるとすぐに終了させられてよい。

【0017】

逆に、目標リンク速度回復が実現されていないと判定された場合、目標リンク速度回復が実現されるまでリンク速度回復を続けるべく、順次リンク速度回復コマンドのセットの別の順次における次のリンク速度回復コマンドが、即時に発行されてよい。したがって、一態様において、リンク速度回復コマンドの各サブループ実行は、オプションとして、サブループ・セットのリンク速度回復コマンドに応答して目標リンク速度が実現されてい

10

20

30

40

50

い場合、リンク速度回復コマンドのサブグループ・セットの別の順次における次のリンク速度回復コマンドを発行すること、サブグループ・セットの別の順次における次のリンク速度回復コマンドを発行した後、目標リンク速度が実現されているかどうかを判定すること、およびサブグループ・セットの別の順次における次のリンク速度回復コマンドに応答して目標リンク速度が実現された場合、リンク速度回復の実行を終了させることを含んでよい。さらに、リンク速度回復コマンドの各メイン・ループ実行は、オプションとして、メイン・ループ・セットのリンク速度回復コマンドに応答して目標リンク速度が実現されていない場合、リンク速度回復コマンドのメイン・ループ・セットの別の順次における次のリンク速度回復コマンドを発行すること、メイン・ループ・セットの別の順次における次のリンク速度回復コマンドを発行した後、目標リンク速度が実現されているかどうかを判定すること、およびメイン・ループ・セットの別の順次における次のリンク速度回復コマンドに応答して目標リンク速度が実現されている場合、リンク速度回復の実行を終了させることを含んでよい。

【0018】

別の態様において、リンク速度回復コマンドの各メイン・ループ実行は、オプションとして、リンク速度回復コマンドのメイン・ループの実行を所定の最大回数、繰り返した後、リンク速度回復コマンドに応答して目標リンク速度が実現されていない場合、リンク速度回復の実行を終了させることを含んでよい。コンピュータ・システム・リソースの割振りが、それらの制限内で完全リンク速度が実現されない場合にメイン・ループ実行の繰返しの回数を制限することによって向上させられてよいことを理解されたい。しかし、欠陥のある機器を交換して、本明細書において説明されるリンク速度回復を繰返すことなどの、他の技術が適用されてよい。

【0019】

さらに別の態様において、順次リンク速度回復コマンドのサブグループ・セットは、オプションとして、通信リンクに結合された上流ポートにおいてレジスタの構成ビットを設定すること、通信リンクに結合された下流ポートのレジスタの構成ビットを設定すべく帯域外信号を送信すること、上流ポートのディスエーブル・イネーブル・サイクルを開始すること、および下流ポートのディスエーブル・イネーブル・サイクルを開始すべく帯域外信号を送信することのうちの少なくとも1つを含んでよい。順次リンク速度回復コマンドのサブグループに関するこの順次コマンド・セットの実行は、メイン・ループに関する異なる順次コマンド・セットの実行と比べて、より高い頻度で成功することがあることを理解されたい。それ故、順次リンク速度回復コマンドのサブグループに関するこの順次コマンド・セットを、順次リンク速度回復コマンドのメイン・ループに関する順次コマンド・セットの繰返しと比べて、より頻繁に繰返すことが、リンク再訓練が完全リンク速度の回復をもたらす成功率を向上させることができる。

【0020】

さらに別の態様において、順次リンク速度回復コマンドのメイン・ループ・セットは、オプションとして、順次リンク速度回復コマンドのサブグループを開始すること、通信リンクに結合された上流光トランシーバを有する上流ポートを含む通信パス・スイッチのリセットおよび再初期設定を開始すること、上流光トランシーバをリセットすること、および通信リンクに結合された下流光トランシーバをリセットすべく帯域外信号を送信することのうちの少なくとも1つを含んでよい。順次リンク速度回復コマンドのメイン・ループに関するこの順次コマンド・セットの実行は、サブグループに関する順次コマンド・セットの実行が成功しない場合に成功することがあることを理解されたい。それ故、順次リンク速度回復コマンドのメイン・ループに関する順次コマンド・セットは、一部の実施形態において、順次リンク速度回復コマンドのサブグループに関する順次コマンド・セットの繰返しと比べて、より低い頻度で実行される可能性があるものの、順次リンク速度回復コマンドのメイン・ループに関する順次コマンド・セットは、リンク訓練またはリンク再訓練が完全リンク速度の実現または回復をもたらす成功率を向上させるべくリンク速度回復のための他の方法を提供するさらなる回復コマンドを追加する。

10

20

30

40

50

【図面の簡単な説明】

【0021】

【図1】本説明の一態様によるデータ・ストレージ・システムにおいてリンク速度回復を用いるコンピューティング環境の実施形態を示す図である。

【図2】本説明の一態様によるデータ・ストレージ・システムにおいてリンク速度回復を用いる、図1のコンピューティング環境のストレージ・コントローラおよびストレージ・ユニットを含むデータ・ストレージ・システムの実施例を示す図である。

【図3】本説明の一態様によるデータ・ストレージ・システムにおいてリンク速度回復を用いる、図1のコンピューティング環境のストレージ・コントローラを含む一次データ・ストレージ・システムおよび二次データ・ストレージ・システムの実施例を示す図である。

10

【図4】本説明の一態様によるデータ・ストレージ・システムにおいてリンク速度回復を用いる、図1のコンピューティング環境の通信パスの実施例を示す図である。

【図5】本説明の一態様によるデータ・ストレージ・システムにおいてリンク速度回復を用いる、図1のコンピューティング環境のホストの実施例を示す図である。

【図6】本説明の一態様によるデータ・ストレージ・システムにおいてリンク速度回復を用いる、図1のコンピューティング環境のコンポーネントの動作の実施例を示す図である。

【図7A】本説明の一態様によるデータ・ストレージ・システムにおいてリンク速度回復を用いる、図1のコンピューティング環境のコンポーネントの動作のループに関する順次コマンド・セットの実施例を示す図である。

【図7B】本説明の一態様によるデータ・ストレージ・システムにおいてリンク速度回復を用いる、図1のコンピューティング環境のコンポーネントの動作のループに関する順次コマンド・セットの実施例を示す図である。

20

【図8】本説明によるデータ・ストレージ・システムにおいてリンク速度回復を用いるコンピュータ実施形態を示す図である。

【発明を実施するための形態】

【0022】

本説明の一態様によるデータ・ストレージ・システムにおけるリンク速度回復は、コンピュータ技術の重要な改良をもたらす。一実施形態において、リンク速度回復動作は、完全リンク速度などの目標リンク速度を獲得すべく、または再獲得すべく通信パスの通信リンクを訓練することに向けられる。したがって、本明細書において使用される、通信リンクを「訓練すること」という術語は、後続の回に通信リンクを再訓練することを含む。例えば、本説明によるリンク速度回復は、通信リンクの1つまたは複数のレーンが失われた完全リンク速度の喪失の後に続いて通信リンクを再訓練することを容易化することができる。同様に、本説明によるリンク速度回復は、通信リンクが初回に初期設定されるときに完全リンク速度を実現すべく通信リンクを訓練することを容易化することができる。したがって、本明細書において使用される「リンク速度回復」という術語は、初回の、または後続の回の完全リンク速度などの目標リンク速度を実現すべく通信リンクを訓練すること、または再訓練することに適用される。

30

【0023】

一態様において、本説明のリンク速度回復は、メイン・ループにおいて、順次リンク速度回復コマンドのセットを実行することを含む。一実施形態において、メイン・ループは、必要に応じて繰り返され、そのため、メイン・ループの順次リンク速度回復コマンドのセットは、必要に応じて繰り返される。さらに、一実施形態において、メイン・ループは、やはり必要に応じて繰り返される順次リンク速度回復コマンドの別のセットのサブループを含む。このため、順次リンク速度回復コマンドの1つのセットの各メイン・ループの実行内で、順次リンク速度回復コマンドの別のセットのサブループが、必要に応じて繰り返される。

40

【0024】

本説明の一態様において、サブループの順次リンク速度回復コマンドのセットは、メイン・ループの順次リンク速度回復コマンドのセットが繰り返されるのと比べて、より頻繁

50

に繰り返される。したがって、リンク速度回復の有効性を最大化するように、メイン・ループ・セットのリンク速度回復コマンドと比較して異なるリンク速度回復コマンドが、サブグループ・セットに割り当てられてよい。例えば、一実施形態により各メイン・ループ実行内で順次リンク速度回復コマンドのサブグループ・セットの実行を繰り返すこと、および順次リンク速度回復コマンドのメイン・ループ・セットの実行を繰り返すことによって、完全リンク速度へのリンク速度訓練またはリンク速度再訓練の信頼性が、向上させられてよいものと考えられる。したがって、リンク訓練が完全リンク速度の回復をもたらす成功率もまた、向上させられることが可能である。特定の応用例に依存して、他の態様および他の利点の実現されてよい。

【0025】

1つまたは複数のコンピュータのシステムが、本説明による書込み転送リソース管理を、動作の際、システムに実行させるソフトウェア、ファームウェア、ハードウェア、またはその組合せがシステムにインストールされているお陰で、本説明によるリンク速度回復のために構成されてよい。例えば、1つまたは複数のコンピュータ・プログラムが、例えば、ストレージ・コントローラ・プロセッサなどのデータ処理装置によって実行されたとき、装置にアクションを実行させる命令を含むお陰で、データ・ストレージ・システムにおいてリンク速度回復を実行するように構成されてよい。本明細書において使用される「圧縮」という用語は、もたらされるデータのストリングのサイズが、アルゴリズムの操作より前の元のデータのストリングと比べて縮小されるように、データのストリングを再エンコードすることによってデータを操作する任意の適切なアルゴリズムの圧縮プロセスを指す。

【0026】

本明細書において説明される操作は、手動で実行されることが示される場合を除き、ほとんど、またはまったくシステム・オペレータ介入なしに自動的に、または実質的に自動的に操作を実行するように構成された論理によって実行される。このため、本明細書において使用される「自動的」という用語は、完全に自動的であること、すなわち、動作が、グラフィカル・ユーザ選択インタフェースに対するユーザ入力などの人間の介入なしに1つまたは複数のハードウェア制御される、もしくはソフトウェア制御されるマシンによって実行されることを含む。本明細書において使用される「自動的」という用語は、大部分が自動的であること、すなわち、動作のほとんど（例えば、50%を超えるなどの）が、グラフィカル・ユーザ選択インタフェースに対するユーザ入力などの人間の介入なしに1つまたは複数のハードウェア制御される、もしくはソフトウェア制御されるマシンによって実行され、残りの動作（例えば、50%未満）が、手動で実行されること、すなわち、手動の動作が、動作の実行を導くべくグラフィカル・ユーザ選択インタフェースに対するユーザ入力などの人間の介入を伴って1つまたは複数のハードウェア制御される、もしくはソフトウェア制御されるマシンによって実行されることをさらに含む。

【0027】

本明細書において説明される機能要素の多くは、それらの要素の実装の独立性をより強調すべく「論理」とラベル付けされている。例えば、論理要素は、カスタムVLSI回路もしくはゲート・アレイ、論理チップなどの既製の半導体、トランジスタ、または他のディスクリートのコンポーネントを備えるハードウェア回路として実装されてよい。また、論理要素は、フィールド・プログラマブル・ゲート・アレイ、プログラマブル・アレイ論理、プログラマブル論理デバイス、またはそれに類するものなどのプログラマブル・ハードウェア・デバイスにおいて実装されてもよい。

【0028】

また、論理要素は、様々なタイプのプロセッサによって実行されるようにソフトウェアにおいて実装されてもよい。例えば、実行可能コードを含む論理要素は、例えば、オブジェクト、手続き、または関数として編成されてよいコンピュータ命令の1つまたは複数の物理ブロックまたは論理ブロックを備えてよい。それでも、識別される論理要素の実行可能ファイルは、物理的に一緒に配置される必要はなく、論理的に一緒に合わせられたとき

10

20

30

40

50

、論理要素を備え、論理要素に関する明示される目的を実現する、異なるロケーションに記憶された分離した命令を備えてよい。

【0029】

実際、論理要素のための実行可能コードは、単一の命令であってよく、または多数の命令であってよく、いくつかの異なるコード・セグメントにわたって、異なるプログラムの間、異なるプロセッサの間、およびいくつかのメモリ・デバイスにわたって分散されることさえある。同様に、動作データは、本明細書において論理要素内で識別され、例示されることがあり、任意の適切な形態で実体化されてよく、任意の適切なタイプのデータ構造内で編成されてよい。動作データは、単一のデータ・セットとして収集されてよく、または異なるストレージ・デバイスにわたることを含めて、異なるロケーションにわたって分散されてよい。

10

【0030】

図1～図5は、本説明によるデータ・ストレージ・システムにおいてリンク速度回復を用いるコンピューティング環境の実施形態を示す。この実施例において、複数のホスト1a、1b...1nが、ネットワーク6bを介してストレージ・コントローラ4a、4b...4nによってストレージ10a、10b...10nに記憶されたデータにアクセスすべくネットワーク6aを介して1つまたは複数のストレージ・コントローラ4a、4b...4nに入出力(I/O)要求をサブミットしてよい。各ストレージ・コントローラ、およびネットワーク6bを介してストレージ・コントローラによって制御されるストレージが、データ・ストレージ・システムを形成する。ホスト1a、1b...1nは、別々の物理デバイスであってよく、または、例えば、サーバのパーティションの割り当てられたリソースを使用して実装される仮想デバイスであってよい。同様の様態で、ストレージ・コントローラ4a、4b...4nは、別々の物理デバイスであってよく、または、例えば、1つまたは複数のサーバのパーティションの割り当てられたリソースを使用して実装される仮想デバイスであってよい。

20

【0031】

図2は、本説明によるリンク速度回復を用いるデータ・ストレージ・システム11の実施例をさらに詳細に示す。データ・ストレージ・システム11は、ストレージ・コントローラ4a、4b...4nと、ストレージ10a、10b...10nとを含む図1のデータ・ストレージ・システムと同様であり、それを代表する。図3は、一次データ・ストレージ・システムまたは二次データ・ストレージ・システム的一方または両方が、本説明の一態様によるリンク速度回復を用いる、一次データ・ストレージ・システム11aと、二次データ・ストレージ・システム11bとを有するストレージ・システムの実施例を示す。

30

【0032】

各データ・ストレージ・システム11(図2)、11a、11b(図3)は、1つまたは複数のストレージ・ドライブ10(図2)、10a(図3)、図10b(図3)によって表されるストレージにおけるボリューム12(図2)、ボリューム1、ボリューム2(図3)(例えば、LUN(論理ユニット)、論理デバイス、論理サブシステム、その他)におけるデータにアクセスするストレージ・コントローラもしくはストレージ・コントロール・ユニット4(図2)、4a(図3)、4bを含む。各ストレージ・コントローラ4、4a、4bは、単一のまたは多数のプロセッサ・コアを各々が有する、1つまたは複数のプロセッサもしくは中央処理装置によって提供されるプロセッサ・リソースを含むCPUコンプレックス14(図2)を含む。この実施形態において、プロセッサ・コアは、例えば、算術論理ユニット(ALU)、浮動小数点ユニット(FPU)、または様々なレベルのキャッシュ(L1キャッシュおよびL2キャッシュなどの)、あるいはその組合せなどの、命令を実行することに関与するCPUのコンポーネントを包含する。プロセッサ・コアは、本明細書で言及される論理要素に加えて、またはそれらの代わりに他の論理要素を有してよいことを理解されたい。

40

【0033】

50

各ストレージ・コントローラ 4（図 2）、4 a（図 3）、4 b は、ホストからの I/O データ要求に应答して、それぞれ、関連付けられたストレージ 10、10 a（図 3）、10 b にデータを書き込むこと、またはそれらからデータを読み取ることを含むストレージ動作を管理するためのストレージ・マネージャ 24 を含むメモリ 20（図 2）をさらに有する。ストレージ・マネージャ 24 によって管理されるストレージ動作は、例えば、データ・ストレージ・システム 11 a などの一次データ・ストレージ・システムの一次ボリューム 1（図 3）から、データ・ストレージ・システム 11 b などの二次データ・ストレージ・システムの二次ボリューム 2 に対するデータ・レプリケーション動作をさらに含む。ストレージ・マネージャ 24 は、一次データ・ストレージ・システム 11 a の一次ボリューム 1（図 3）のコピーを、二次データ・ストレージ・システム 11 b の二次ボリューム 2（図 3）として生成するように構成される。ボリュームのペア、ボリューム 1、ボリューム 2 は、一次ボリューム 1 に対する更新が二次ボリューム 2 にレプリケートされるようにコピー関係にある。

10

【0034】

各ストレージ・コントローラの CPU コンプレックス 14 は、各クラスタが、そのクラスタ独自の割り当てられたメモリ 20、ストレージ・マネージャ 24、キャッシュ、その他を有する、プロセッサの多数のクラスタを有してよい。各クラスタに割り当てられる処理リソースおよびメモリ・リソースは、特定の応用例に依存して、物理的、仮想、共有、移転可能、または専用であってよい。

【0035】

20

後段でさらに詳細に説明されるとおり、ストレージ・マネージャ 24 のマルチループ・リンク速度回復論理 32 が、一実施形態において、ストレージ・コントローラ 4 をストレージ 10 に結合するネットワーク 6 b（図 1）のリンク 38 などの通信リンクの訓練の成功を容易化する。その結果、ストレージ・コントローラ 4 をストレージ 10 に結合する通信パス 38 の一部であるリンク 34 が、完全リンク速度を確立すべく、または復元すべく確実に訓練されること、または再訓練されることが可能であり、それにより、データ・ストレージ・システム 11 のコンピュータ動作の効率を向上させることが可能である。特定の応用例に依存して、他の態様および他の利点が実現されてよい。

【0036】

図 2 および図 4 に示される例示される実施形態において、通信パス 38 は、ルート・コンプレックスで実装されてよい通信パス・ホスト 44 を含む PCI e 通信パスである。通信パス 38（図 2、図 4）は、実施例が、上流光トランシーバ 60 を有する上流入出力ポート 52（図 4）として示される、複数の上流 I/O ポートを有する通信パス・スイッチ 48 をさらに含む。一実施形態において、通信パス・スイッチ 48 は、通信パス・スイッチ 48 の集積回路デバイスをサポートする基板を有する取外し可能なアドイン・カードとして実装されてよい。通信パス・スイッチ 48 の PCI e カード上に配置されたコネクタが、通信パス・スイッチ 48 が、通信パス 38 の通信パス・ホスト 44 に取外し可能に結合されることを許す。

30

【0037】

通信リンク 34 は、上流ポート 52 の光トランシーバ 60 に光結合された 1 つまたは複数の光ファイバ・ケーブルを含む。リンク 34 は、例示される実施形態において光ファイバ・リンクとして説明されるものの、リンク 34 は、伝導性金属ケーブル、無線伝送、その他などの他のデータ伝送技術を用いてよいことを理解されたい。さらに、通信パス 38 は、PCI e シリアル・バス通信パスを提供するものとして示されるものの、本説明によるリンク速度回復を用いる通信パスは、他のタイプのシリアル・バス、パラレル・バス、その他などの他の通信プロトコルおよび通信技術を用いてよいことを理解されたい。

40

【0038】

通信パス 38 は、実施例が、光トランシーバ 72 を有する下流ポート 68 として示される、複数の下流ポートを有してよいエンドポイント・デバイス 64 をさらに含む。通信リンク 34 は、1 つの終端で、スイッチ 48 の上流ポート 52 の光トランシーバ 60 に結合

50

され、別の終端で、この実施形態においてストレージ・デバイス 10 のエンドポイント・デバイス 64 の下流ポート 68 の光トランシーバ 72 に結合される。

【0039】

通信パス 38 は、ストレージ・コントローラ 4 とストレージ 10 との間の通信パスを提供するものとして示されるものの、本説明によるリンク速度回復を用いる通信パスは、他のタイプのデバイス間の通信のために用いられてよいことを理解されたい。例えば、本説明によるリンク速度回復を用いる通信パスは、例えば、ホスト 1a などのホストとストレージ・コントローラ 4 との間の通信を提供してよい。特定の応用例に依存して、通信パスを利用する他のデバイスが、本説明によるリンク速度回復を用いることによって利益を得てもよい。

10

【0040】

例示される実施形態において、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 32 は、メモリ 20 に記憶され、CPU コンプレックス 14 によって実行されるソフトウェアとして示される。しかし、マルチループ・リンク速度回復論理 32 の論理機能は、特定の応用例に依存して、ハードウェア、ソフトウェア、ファームウェア、またはその 1 つまたは複数の任意の組合せとして実装されてよいことを理解されたい。例えば、マルチループ・リンク速度回復論理 32 の論理機能は、PCIe 通信パス 38 などの通信パスのためのドライバにおいて実装されてよく、通信パスのためのドライバに加えて、またはその代わりに、通信パス 38 自体のハードウェア、ソフトウェア、ファームウェア、またはその 1 つまたは複数の任意の組合せにおいて実装されてもよい。

20

【0041】

本説明によるリンク速度回復の別の態様において、通信パス 38 は、通信パス・ホスト 44 とエンドポイント・デバイス 64 との間の補助通信パスを提供する帯域外 (OOB) サブパス 76a、76b、およびエンドポイント帯域外コントローラ 80 を含む。一実施形態において、帯域外 (OOB) サブパス 76a、76b は、PCIe プロトコルの代わりに RS485 シリアル・プロトコルに準拠してよい。特定の応用例に依存して、他のプロトコルが、使用されてよい。後段でより詳細に説明されるとおり、帯域外サブパス 76a、76b、および帯域外コントローラ 80 は、通信リンク 34 とは独立に、ルート・コンプレックスなどの通信パス・ホスト 44 と、エンドポイント・デバイス 64 のコンポーネントとの間の通信を許す。例えば、通信リンク 34 が、通信の動作可能なレーンをまったく欠いている場合、通信は、それでも、後段でさらに詳細に説明されるとおり、通信エンドポイント・デバイス 64 においてリンク速度回復を開始すべく、帯域外サブパス 76a、76b、および帯域外コントローラ 80 によって提供されてよい。例示される実施形態において、通信パス・ホスト 44 は、帯域外サブパス 76a によって帯域外コントローラ 80 に結合され、コントローラ 80 は、通信パス・ホスト 44 と、エンドポイント・デバイス 64 のコンポーネントとの間の帯域外通信を許すべく帯域外サブパス 76b によってエンドポイント・デバイス 64 に結合される。

30

【0042】

一実施形態において、ストレージまたはストレージ・ドライブ 10 (図 2)、10a、10b...10n (図 1) は、通信パス・エンドポイント・デバイス 64 に加えて、ハードディスク・ドライブおよび磁気テープなどの 1 つまたは複数の順次アクセス・ストレージ・デバイスを備えてよく、または、例えば、ソリッド・ステート・ドライブ (SSD) などの非順次アクセス・ストレージ・デバイスを含んでよい。各ストレージ・ドライブ 10、10a、10b...10n は、順次アクセス・ストレージ・デバイスを備えても、非順次アクセス・ストレージ・デバイスを備えてもよく、またはジャスト・ア・バンチ・オブ・ディスクス (JBOD)、ダイレクト・アクセス・ストレージ・デバイス (DASD)、冗長アレイ・オブ・インデペンデント・ディスク (RAID) アレイ、仮想化デバイス、テープ・ストレージ、フラッシュ・メモリ、その他などのストレージ・デバイスのアレイを備えてよい。

40

【0043】

50

ストレージ・ドライブ 10、10 a、10 b... 10 n のストレージ単位は、ボリューム、トラック、エクステント、ブロック、ページ、セグメント、シリンダ、その他などのデータ・ストレージの部分単位でデータを記憶するように構成されてよい。本説明によるリンク速度回復は、ボリュームなどのストレージ部分単位に関連して説明されるものの、本説明によるリンク速度回復は、トラック、エクステント、ブロック、ページ、セグメント、シリンダ、その他などの他のストレージ部分単位に適用可能であることを理解されたい。

【0044】

システム・コンポーネント 1 a、1 b... 1 n、4、4 a、4 b、... 4 n、10、10 a、10 b... 10 n が、スイッチ、リンク、ならびにアダプタなどのエンドポイント・デバイス経由でこれらのコンポーネントの間で通信を可能にするネットワーク 6 a、6 b に接続される。このため、ネットワーク 6 a、6 b は、一実施形態において、ストレージ・エリア・ネットワーク (SAN)、ローカル・エリア・ネットワーク (LAN)、イントラネット、インターネット、ワイド・エリア・ネットワーク (WAN)、ピア・ツー・ピア・ネットワーク、無線ネットワーク、アービトレーテッド・ループ・ネットワーク、その他を備えてよいファブリックを含む。ストレージ・システムからホスト 1 a、1 b、... 1 n に至る通信パス、およびストレージ・コントローラ 4、4 a、4 b、... 4 n からストレージ 10、10 a、10 b... 10 n に至る通信パスは、例えば、ファイバ接続 (FICON) などの様々なアタッチメント・プロトコルに基づいてよい。ファブリックの他の通信パスは、ファイバ・チャネル・アービトレーテッド・ループ構成、シリアル・ループ・アーキテクチャ、あるいは PCI エクスプレス・インタフェースなどのペリフェラル・コンポーネント・インタコネクタ (PCI) インタフェースなどのバス・インタフェースを備えてよい。また、ファブリックの通信パスは、例えば、各ノードが個別のネットワーク (インターネット・プロトコル) アドレスを有するように、イーサネット (R) ・ネットワークの一部であってもよい。特定の応用例に依存して、モデム電話パス、無線ネットワーク、その他などの他のタイプの通信パスが、利用されてよい。

【0045】

通信パスに関連付けられた通信ソフトウェアが、通信プロトコルを制御し、存在する場合、通信プロトコルにより通信ハードウェアの動作を制御する命令および他のソフトウェアを含む。特定の応用例に依存して、他の通信パス・プロトコルが利用されてよいことを理解されたい。

【0046】

図 5 のホスト 1 a によって代表される典型的なホストは、CPU コンプレックス 202 と、ストレージ・コントローラ 4、4 a、4 b... 4 n を介して一次ストレージ 10 a (図 3) または二次ストレージ 10 b からデータを読み取ること、およびそれらにデータのアップデートを書き込むことを行うべく協働するオペレーティング・システム 206 およびアプリケーション 208 を有するメモリ 204 とを含む。適切なオペレーティング・システムの実施例が、z/OS オペレーティング・システムである。特定の応用例に依存して、他のタイプのオペレーティング・システムが用いられてよいことを理解されたい。

【0047】

本説明によるリンク速度回復は、マルチループ・リンク速度回復論理 32 (図 2) によって代表される論理を利用して、通信リンクを有する任意のコンピュータ・システムに適用されてよい。それ故、例えば、ホスト 1 a などの各ホストがやはり、リンク速度回復のためにマルチループ・リンク速度回復論理を用いてもよい。

【0048】

ホスト 1 a、1 b... 1 n、ストレージ・コントローラ 4、4 a、4 b、ストレージ・デバイス 10、10 a、10 b、通信パス 38、およびマルチループ・リンク速度回復論理 32 は、それぞれ、本説明によるリンク速度回復のために変形されている任意の計算デバイスを使用して実装されてよい。本明細書において説明されるとおりの変形に適した計算デバイスは、パーソナル・コンピュータ、ワークステーション、サーバ、メインフレ

10

20

30

40

50

ーム、ハンドヘルド・コンピュータ、パームトップ・コンピュータ、電話デバイス、ネットワーク機器、ブレード・コンピュータ、処理デバイス、その他などの、当技術分野において現在、知られているものを含む。ホスト1 a、1 b . . . 1 n、ストレージ・コントローラ4、4 a、4 b . . . 4 n、ストレージ・デバイス1 0、1 0 a、1 0 b . . . 1 0 n、通信パス3 8、およびマルチループ・リンク速度回復論理3 2は、ストレージ・エリア・ネットワーク、ワイド・エリア・ネットワーク、インターネット、イントラネットなどの任意の適切なネットワークにおける要素、あるいはクラウド・コンピューティング環境における要素であってよい。

【0049】

図6は、通信パス3 8（図4）に接続されたストレージ・コントローラ4のマルチループ・リンク速度回復論理3 2（図2）の動作の一実施形態を示す。この実施例において、マルチループ・リンク速度回復論理3 2（図2）の論理要素、および通信パス3 8（図4）は、図6、および本明細書における付随する説明に示されるとおりにリンク速度回復を実行するように構成される。

【0050】

一実施例において、リンク速度回復動作は、通信リンク3 4の1つまたは複数のレーンが失われた完全リンク速度の喪失の後に続いて通信リンク3 4上で、完全リンク速度などのリンク速度を再獲得すべく通信パス3 8の通信リンク3 4を再訓練することに向けられる。後段でさらに詳細に説明されるとおり、この実施形態のリンク速度回復は、順次リンク速度回復コマンドのメイン・ループの実行を所定の最大回数、繰り返すことを含む。図7 Aは、順次リンク速度回復コマンドのメイン・ループに関する順次コマンド・セットの実施例を示す。特定の応用例に依存して、図7 Aに示されるものに加えて、またはその代わりに、他の順次、他のタイプ、または他の量のリンク速度回復コマンドおよびリンク速度回復動作が、メイン・ループにおいて用いられてよいことを理解されたい。特定の応用例に依存して、リンク速度回復コマンドおよびリンク速度回復動作の異なるメイン・ループの数は、様々であってよいことをさらに理解されたい。この実施形態において、リンク速度回復コマンドの各メイン・ループ実行は、各メイン・ループ実行内で順次リンク速度回復コマンドのサブループの実行を第2の所定の最大回数、繰り返すことを含む。

【0051】

図7 Bは、順次リンク速度回復コマンドのサブループに関する順次コマンド・セットの実施例を示す。順次リンク速度回復コマンドを異なるセットに分離することが、リンク速度回復の信頼性を向上させることができるものと考えられる。例えば、一実施形態において、各メイン・ループ実行内で順次リンク速度回復コマンドのサブループの実行を繰り返すこと、および順次リンク速度回復コマンドのメイン・ループ実行を繰り返すことによって、完全リンク速度へのリンク速度回復の信頼性が向上させられることが可能であるものと考えられる。それ故、一実施形態において、順次リンク速度回復コマンドのサブループに関する順次コマンド・セットは、順次リンク速度回復コマンドのメイン・ループに関する順次コマンド・セットと比べて、より頻繁に繰り返される。そのような構成は、リンク再訓練が完全リンク速度の回復をもたらす成功率を向上させることができる。例えば、順次リンク速度回復コマンドのサブループに関する順次コマンド・セットの実行は、メイン・ループに関する順次コマンド・セットの実行の繰返しと比べて、より頻繁に成功することがある。このため、順次リンク速度回復コマンドのサブループに関する順次コマンド・セットを、順次リンク速度回復コマンドのメイン・ループに関する順次コマンド・セットの繰返しと比べて、より頻繁に繰り返すことが、リンク再訓練が完全リンク速度の回復をもたらす成功率を向上させることができる。特定の応用例に依存して、図7 Bに示されるものに加えて、またはその代わりに、他の順次、他のタイプ、または他の量のリンク速度回復コマンドおよびリンク速度回復動作が、サブループにおいて用いられてよいことを理解されたい。特定の応用例に依存して、リンク速度回復コマンドおよびリンク速度回復動作の異なるサブループの数は、様々であってよいことをさらに理解されたい。

【0052】

完全リンク速度の喪失に応答するストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2 (図 2) は、通信パス 3 8 の通信リンク 3 4 を訓練すべくリンク速度回復を開始する (図 6、ブロック 2 2 0)。順次リンク速度回復コマンドのメイン・ループが、図 6 のリンク速度回復において最初に開始され (図 6、ブロック 2 2 4)、メイン・ループ順次コマンド・セットの (図 7 A) の順次における次のリンク速度回復コマンドが、発行される (図 6、ブロック 2 2 6)。この実施例において、メイン・ループ・コマンドの順次 M L 1 ~ M L 4 の順次における次のメイン・ループ順次コマンドは、メイン・ループ順次コマンド・セットの第 1 のメイン・ループ・コマンド M L 1 である。図 7 A に示されるとおり、順次における次のメイン・ループ・コマンド M L 1 が、図 7 B に示される順次リンク速度回復コマンド S L 1 ~ S L 4 のサブグループを開始し (図 6、ブロック 2 2 8)、サブグループ順次コマンド・セット (図 7 B) の順次における次のリンク速度回復コマンドが、発行される (図 6、ブロック 2 3 2)。

【0053】

この実施例において、サブグループ・コマンドの順次 S L 1 ~ S L 4 の順次における次のサブグループ順次コマンドは、サブグループ順次コマンド・セットの第 1 のサブグループ・コマンド S L 1 である。図 7 B に示されるとおり、順次における次のサブグループ・コマンド S L 1 は、上流ポート 5 2 において通信リンク 3 4 (図 4) が再訓練 (または訓練) されるように命令する。この実施形態において、通信パス・スイッチ 4 8 は、通信パス・ホスト 4 4 からのコマンドに応答して上流ポート 5 2 を制御する上流リンク・コントローラ 2 3 4 を含む。したがって、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2 (図 2) が、上流ポート 5 2 に結合された通信リンク 3 4 を再訓練すること (または訓練すること) を開始すべく、この実施例ではルート・コンプレックスである通信パス・ホスト 4 4 に適切なコマンドを発行し、通信パス・ホスト 4 4 が、通信パス・スイッチ 4 8 およびスイッチ 4 8 の上流ポート 5 2 に適切なコマンドを発行する。

【0054】

この実施形態において、通信リンク 3 4 を再訓練すること (または訓練すること) は、上流ポート 5 2 のレジスタ 2 3 6 のビットを設定することによって開始されてよい。例えば、P C I e 規格は、ルート・コンプレックスが、P C I エクスプレス・ケイパビリティ・リンク・ステータス・コントロール・レジスタの構成ビットを設定することによってリンク再訓練を開始すべきことを規定する。それ故、一実施形態において、上流ポート 5 2 のレジスタ 2 3 6 は、P C I エクスプレス・ケイパビリティ・リンク・ステータス・コントロール・レジスタであってよい。通信リンク 3 4 の特定のプロトコルに依存して、レジスタ 2 3 6 は、他のフォーマットを有してよいことを理解されたい。

【0055】

ここにおいて、上流ポート 5 2 のレジスタ 2 3 6 のビットの単一の設定が、通信リンク 3 4 の再訓練の成功を実現しないことがあることを理解されたい。したがって、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2 (図 2) は、通信リンク再訓練を開始するレジスタ・ビットの設定の後に続いて、再訓練が完了することを許す適切な遅延の後、例えば、完全リンク速度などの目標リンク速度が、上流ポート 5 2 において開始された通信リンク再訓練 (または訓練) によって実現されているかどうかを判定する (図 6、ブロック 2 3 8)。一実施形態において、通信リンクのリンク速度は、通信リンク 3 4 のステータスをポーリングすることによって判定されてよい。通信リンク 3 4 に関して完全リンク速度が復元されている (または実現されている) 場合、リンク速度回復は、終了させられる (図 6、ブロック 2 4 0)。

【0056】

逆に、レジスタ 2 3 6 の適切な構成ビットの設定によって上流ポート 5 2 において開始されたリンク速度再訓練が、通信リンク 3 4 に関して目標とされる完全リンク速度を実現することに失敗したと判定された場合 (図 6、ブロック 2 3 8)、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2 (図 2) は、サブグループ順次コマンド・セット (図 7 B) のすべての順次リンク速度回復コマンドが完了されているかどうかを判

定する（ブロック 2 4 4）。

【0057】

この実施例において、サブグループ順次コマンド・セット（図 7 B）のすべての順次リンク速度回復コマンドが完了されているわけではない。したがって、サブグループ順次コマンド・セット（図 7 B）の順次における次のリンク速度回復コマンドが、発行される（図 6、ブロック 2 3 2）。この実施例において、サブグループ・コマンドの順次 S L 1 ~ S L 4 の順次における次のサブグループ順次コマンドは、サブグループ順次コマンド・セットの第 2 のサブグループ・コマンド S L 2 である。図 7 B に示されるとおり、順次における次のサブグループ・コマンド S L 2 は、下流ポート 6 8 において通信リンク 3 4（図 4）が再訓練される（または訓練される）ことを命令する。この実施形態において、通信パス・エンドポイント 6 4 は、通信パス・ホスト 4 4 からの命令に応答して下流ポート 6 8 を制御する下流リンク・コントローラ 2 4 6 を含む。

【0058】

本説明によるリンク速度回復の一態様において、帯域外サブパス 7 6 a が、通信パス・ホスト（ルート・コンプレックス） 4 4 と、エンドポイント帯域外コントローラ 8 0 との間の補助通信パスを提供し、コントローラ 8 0 は、補助帯域外通信パス 7 6 b によって通信パス・エンドポイント・デバイス 6 4 に結合される。帯域外サブパス 7 6 a、7 6 b、およびコントローラ 8 0 は、通信リンク 3 4 とは独立に、通信パス・ホスト 4 4 と、エンドポイント・デバイス 6 4 の下流リンク・コントローラ 2 4 6 との間の通信を許す。例えば、通信リンク 3 4 のための通信のレーンを確立する、または復元する前述される上流ポートにおける再訓練（または訓練）の失敗に起因して、例えば、通信リンク 3 4 が、通信のすべてのレーンを喪失した（またはレーンをまったく確立していない）場合、通信パス・ホスト 4 4 と、下流リンク・コントローラ 2 4 6 との間の通信は、それでも、帯域外サブパス 7 6 a、7 6 b、およびエンドポイント帯域外コントローラ 8 0 によって提供されてよい。したがって、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2（図 2）は、下流ポート 6 8 に結合された通信リンク 3 4 を再訓練することを開始すべく、この実施例においてはルート・コンプレックスである通信パス・ホスト 4 4 に適切なコマンドを発行し、これに応答して、通信パス・ホスト 4 4 が、エンドポイント帯域外コントローラ 8 0 および帯域外サブパス 7 6 a、7 6 b を介して通信パス・エンドポイント 6 4 およびエンドポイント 6 4 の下流ポート 6 8 に適切なコマンドを発行する。

【0059】

この場合も、この実施形態において、通信リンク 3 4 を再訓練すること、または訓練することは、下流ポート 6 8 のレジスタ 2 4 8 のビットを設定することによって開始されてよい。前述されるとおり、P C I e 規格は、ルート・コンプレックスが、P C I エクスプレス・ケイパビリティ・リンク・ステータス・コントロール・レジスタの構成ビットを設定することによってリンク再訓練を開始すべきことを規定する。それ故、一実施形態において、下流ポート 6 8 のレジスタ 2 4 8 は、P C I エクスプレス・ケイパビリティ・リンク・ステータス・コントロール・レジスタであってよい。通信リンク 3 4 の特定のプロトコルに依存して、レジスタ 2 4 8 は、他のフォーマットを有してよいことを理解されたい。

【0060】

ここにおいて、下流ポート 6 8 のレジスタ 2 4 8 のビットの単一の設定が、通信リンク 3 4 の再訓練または訓練の成功を実現しないことがあることを理解されたい。したがって、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2（図 2）は、下流ポートにおいて通信リンク再訓練を開始するレジスタ・ビットの設定の後に続いて、再訓練が完了することを許す適切な遅延の後、例えば、完全リンク速度などの所望されるリンク速度が、下流ポート 6 8 において開始された通信リンク再訓練によって実現されているかどうかを判定する（図 6、ブロック 2 3 8）。一実施形態において、通信リンクのリンク速度は、通信リンク 3 4 のステータスをポーリングすることによって判定されてよい。通信リンク 3 4 に関して完全リンク速度が復元されている場合、リンク速度回復は、終了させられる（図 6、ブロック 2 4 0）。

【0061】

逆に、レジスタ248の適切な構成ビットの設定によって下流ポート68において開始されたリンク速度再訓練が、通信リンク34に関して目標とされる完全リンク速度を実現することに失敗したと判定された場合（図6、ブロック238）、ストレージ・コントローラ4のマルチループ・リンク速度回復論理32（図2）は、サブグループ順次コマンド・セット（図7B）のすべての順次リンク速度回復コマンドが完了されているかどうかを判定する（ブロック244）。

【0062】

この実施例において、サブグループ順次コマンド・セット（図7B）のすべての順次リンク速度回復コマンドが完了されているわけではない。したがって、サブグループ順次コマンド・セット（図7B）の順次における次のリンク速度回復コマンドが、発行される（図6、ブロック232）。この実施例において、サブグループ・コマンドの順次SL1～SL4の順次における次のサブグループ順次コマンドは、サブグループ順次コマンド・セットの第3のサブグループ・コマンドSL3である。図7Bに示されるとおり、順次における次のサブグループ・コマンドSL3は、上流ポート52においてディスエーブル・イネーブル・サイクルを開始することによって通信リンク34（図4）が上流ポート52において再訓練されること、または訓練されることを命令する。したがって、ストレージ・コントローラ4のマルチループ・リンク速度回復論理32（図2）は、上流ポート52に結合された通信リンク34を再訓練すること、または訓練することを開始すべく、この実施例においてはルート・コンプレックスである通信パス・ホスト44に適切なコマンドを発行し、これに
20 応答して、通信パス・ホスト44が、上流ポート52においてディスエーブル・イネーブル・サイクルを開始すべく通信パス・スイッチ48およびスイッチ48の上流ポート52に適切なコマンドを発行する。この実施形態において、通信リンク34を訓練すること、または再訓練することは、上流ポートをディスエーブルにし、適切な持続時間、待機し、次に、ディスエーブル・イネーブル・サイクルにおいて上流ポートをイネーブルにすることによって開始されてよい。特定の応用例に依存して、例えば、100マイクロ秒などの、50マイクロ秒から1ミリ秒の範囲内の待機が、適切であることがある。

【0063】

上流ポート52の単一のディスエーブル・イネーブル・サイクリングが、通信リンク34の訓練または再訓練の成功を実現しないことがあることを理解されたい。したがって、
30 ストレージ・コントローラ4のマルチループ・リンク速度回復論理32（図2）は、ディスエーブル・イネーブル・サイクル、およびディスエーブル・イネーブル・サイクルによって開始された再訓練が完了することを許す適切な遅延の後に続いて、例えば、完全リンク速度などの目標とされるリンク速度が、上流ポート52において開始された通信リンク再訓練によって復元されているかどうかを判定する（図6、ブロック238）。通信リンク34に関して完全リンク速度が復元されている場合、リンク速度回復は、終了させられる（図6、ブロック240）。

【0064】

逆に、上流ポート52のディスエーブル・イネーブル・サイクリングによって上流ポート52において開始されたリンク速度再訓練が、通信リンク34に関して完全リンク速度
40 を実現することに失敗したと判定された場合（図6、ブロック238）、ストレージ・コントローラ4のマルチループ・リンク速度回復論理32（図2）は、サブグループ順次コマンド・セット（図7B）のすべての順次リンク速度回復コマンドが完了されているかどうかを再び判定する（ブロック244）。

【0065】

この実施例において、サブグループ順次コマンド・セット（図7B）のすべての順次リンク速度回復コマンドが完了されているわけではない。したがって、サブグループ順次コマンド・セット（図7B）の順次における次のリンク速度回復コマンドが、発行される（図6、ブロック232）。この実施例において、サブグループ・コマンドの順次SL1～SL4の順次における次のサブグループ順次コマンドは、サブグループ順次コマンド・セットの第4
50

のサブグループ・コマンド S L 4 である。図 7 B に示されるとおり、順次における次のサブグループ・コマンド S L 4 は、下流ポート 6 8 においてディスエーブル・イネーブル・サイクルを開始することによって下流ポート 6 8 において通信リンク 3 4（図 4）が再訓練される、または訓練されることを命令する。したがって、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2（図 2）は、下流ポート 6 8 に結合された通信リンク 3 4 を再訓練することを開始すべく、この実施例においてはルート・コンプレックスである通信パス・ホスト 4 4 に適切なコマンドを発行し、これに回答して、通信パス・ホスト 4 4 が、下流ポート 6 8 においてディスエーブル・イネーブル・サイクルを開始すべく、帯域外サブパス 7 6 a、エンドポイント帯域外コントローラ 8 0、および帯域外サブパス 7 6 b を介して通信パス・エンドポイント・デバイス 6 4 およびその下流ポート 6 8 に適切なコマンドを発行する。

【0066】

この実施形態において、通信リンク 3 4 を訓練すること、または再訓練することは、下流ポートをディスエーブルにし、適切な持続時間、待機し、次に、ディスエーブル・イネーブル・サイクルにおいて下流ポートをイネーブルにすることによって開始されてよい。特定の応用例に依存して、例えば、100 マイクロ秒などの、50 マイクロ秒から 1 ミリ秒の範囲内の待機が、適切であることがある。下流ポート 6 8 の単一のディスエーブル・イネーブル・サイクリングが、通信リンク 3 4 の訓練または再訓練の成功を実現しないことがあることを理解されたい。したがって、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2（図 2）は、ディスエーブル・イネーブル・サイクルを開始した後、かつディスエーブル・イネーブル・サイクルによって開始された訓練または再訓練が完了することを許す適切な遅延の後、ディスエーブル・イネーブル・サイクルの後に続いて、例えば、完全リンク速度などの目標とされるリンク速度が、下流ポート 6 8 において開始された通信リンク訓練または再訓練によって実現されているかどうかを判定する（図 6、ブロック 238）。通信リンク 3 4 に関して完全リンク速度が復元されている場合、リンク速度回復は、終了させられる（図 6、ブロック 240）。

【0067】

逆に、下流ポート 6 8 のディスエーブル・イネーブル・サイクリングによって下流ポート 6 8 において開始されたリンク速度再訓練が、通信リンク 3 4 に関して完全リンク速度を実現することに失敗したと判定された場合（図 6、ブロック 238）、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2（図 2）は、サブグループ順次コマンド・セット（図 7 B）のすべての順次リンク速度回復コマンドが完了されているかどうかを再び判定する（ブロック 244）。

【0068】

この実施例において、サブグループ順次コマンド・セット（図 7 B）のすべての順次リンク速度回復コマンド S L 1 ~ S L 4 が、前段で説明されるとおり完了されている。したがって、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2（図 2）は、サブグループ実行最大限度に達しているかどうかを判定する（図 6、ブロック 250）。一実施形態において、順次リンク速度回復コマンド S L 1 ~ S L 4 のサブグループの実行は、例えば、5 回であってよい変数「S」によって表される最大限まで各メイン・ループ内で繰り返される。特定の応用例に依存して、5 ~ 10 の範囲内などの他の最大限度が選択されてよいことを理解されたい。メイン・ループ内の順次リンク速度回復コマンド S L 1 ~ S L 4 のサブグループ実行の繰返しは、最大限 S に達していないと判定された場合（図 6、ブロック 250）、順次リンク速度回復コマンド S L 1 ~ S L 4 のサブグループ実行は、最大限 S に達するまで、前述されるとおりブロック 228 ~ 244 において繰り返される。順次リンク速度回復コマンド S L 1 ~ S L 4 のサブグループ実行の繰返しは、メイン・ループ内で最大限 S に達したと判定されると（図 6、ブロック 250）、順次リンク速度回復コマンド M L 1 ~ M L 4 のメイン・ループ実行が、続けられる（ブロック 252 A、252 B）。

【0069】

したがって、メイン・ループ順次コマンド・セット(図7A)の順次における次のリンク速度回復コマンドが、発行される(図6、ブロック256)。この実施例において、メイン・ループ・コマンドの順次ML1~ML4の順次における次のメイン・ループ順次コマンドは、メイン・ループ順次コマンド・セットの第2のメイン・ループ・コマンドML2である。図7Aに示されるとおり、順次における次のメイン・ループ・コマンドML2は、通信パス・スイッチ48がリセットされること、および、次に、スイッチ48の構成が再初期設定されることを命令する。

【0070】

したがって、ストレージ・コントローラ4のマルチループ・リンク速度回復論理32(図2)は、上流ポート52に結合された通信リンク34を再訓練すること、または訓練することを開始すべく、この実施例においてはルート・コンプレックスである通信パス・ホスト44に適切なコマンドを発行し、これに回答して、通信パス・ホスト44が、リセットすべく、次に、通信パス・スイッチ48の構成を再初期設定すべく通信パス・スイッチ48に適切なコマンドを発行する。

【0071】

一実施形態において、通信パス・スイッチ48のリセットは、通信パス・スイッチ48の構成レジスタのすべての構成ビットをデフォルト値にリセットする基本リセットである。したがって、基本リセットの後に続いて、リセットされたビットは、適宜、再初期設定される。様々なレベルのレジスタ・ビットがリセットされ、次に、再初期設定されるハード・リセットまたはソフト・リセットなどの他のタイプのリセットが実行されてよいことを理解されたい。

【0072】

通信パス・スイッチをリセットし、次に、再初期設定することに応答して、通信リンク34の訓練が自動的に開始されることが知られる。しかし、本説明によるリンク速度回復の一態様において、通信パス・スイッチ48をリセットし、次に、再初期設定することは、通信リンク34の訓練または再訓練の成功を実現しないことがあることを認識されたい。したがって、ストレージ・コントローラ4のマルチループ・リンク速度回復論理32(図2)は、通信パス・スイッチ48をリセットし、次に、再初期設定することの後に続いて、適切な遅延の後、例えば、完全リンク速度などの目標リンク速度が、上流ポート52において開始された通信リンク再訓練または訓練によって実現されているかどうかを判定する(図6、ブロック260)。一実施形態において、通信リンクのリンク速度は、通信リンク34のステータスをポーリングすることによって判定されてよい。通信リンク34に関して完全リンク速度が復元されている場合、リンク速度回復は、終了させられる(図6、ブロック240)。

【0073】

逆に、通信パス・スイッチ48をリセットし、次に、再初期設定することによって上流ポート52において開始されたリンク速度訓練またはリンク速度再訓練が、通信リンク34に関して完全リンク速度を実現することに失敗したと判定された場合(図6、ブロック260)、ストレージ・コントローラ4のマルチループ・リンク速度回復論理32(図2)は、メイン・ループ順次コマンド・セット(図7A)のすべての順次リンク速度回復コマンドが完了されているかどうかを判定する(ブロック264)。

【0074】

この実施例において、メイン・ループ順次コマンド・セット(図7A)のすべての順次リンク速度回復コマンドが完了されているわけではない。したがって、メイン・ループ順次コマンド・セット(図7A)の順次における次のリンク速度回復コマンドが、発行される(図6、ブロック256)。この実施例において、メイン・ループ・コマンドの順次ML1~ML4の順次における次のメイン・ループ順次コマンドは、メイン・ループ順次コマンド・セットの第3のメイン・ループ・コマンドML3である。図7Aに示されるとおり、順次における次のメイン・ループ・コマンドML3は、上流ポート52の光トランシーバ60がリセットされること、およびリブートされることを命令する。

【0075】

したがって、ストレージ・コントローラ4のマルチループ・リンク速度回復論理32（図2）は、上流ポート52に結合された通信リンク34を再訓練することを開始すべく、この実施例においてはルート・コンプレックスである通信パス・ホスト44に適切なコマンドを発行し、これに应答して、通信パス・ホスト44が、上流ポート52の光トランシーバ60をリセットすべく通信パス・スイッチ48に適切なコマンドを発行する。

【0076】

上流ポート52の光トランシーバ60をリセットすることに応答して、通信リンク34の訓練が自動的に開始されることが知られる。しかし、本説明によるリンク速度回復の一態様において、上流ポート52の光トランシーバ60をリセットすることは、通信リンク34の訓練または再訓練の成功を実現しないことがあることを認識されたい。したがって、ストレージ・コントローラ4のマルチループ・リンク速度回復論理32（図2）は、上流ポート52の光トランシーバ60をリセットすること、および光トランシーバ60がリブートして、訓練が完了するのに十分な時間、待機することの後に続いて、例えば、完全リンク速度などの所望されるリンク速度が、上流ポート52の光トランシーバ60をリセットすることによって上流ポート52において開始された通信リンク再訓練によって実現されているかどうかを判定する（図6、ブロック260）。一実施形態において、特定の応用例に依存して、約5秒の待機が適切であることがある。前述されるとおり、通信リンクのリンク速度は、通信リンク34のステータスをポーリングすることによって判定されてよい。通信リンク34に関して完全リンク速度が復元されている場合、リンク速度回復は、終了させられる（図6、ブロック240）。

【0077】

逆に、上流ポート52の光トランシーバ60をリセットすることによって開始されたリンク速度訓練またはリンク速度再訓練が、通信リンク34に関して完全リンク速度を実現することに失敗したと判定された場合（図6、ブロック260）、ストレージ・コントローラ4のマルチループ・リンク速度回復論理32（図2）は、メイン・ループ順次コマンド・セット（図7A）のすべての順次リンク速度回復コマンドが完了されているかどうかを判定する（ブロック264）。

【0078】

この実施例において、メイン・ループ順次コマンド・セット（図7A）のすべての順次リンク速度回復コマンドが完了されているわけではない。したがって、メイン・ループ順次コマンド・セット（図7A）の順次における次のリンク速度回復コマンドが、発行される（図6、ブロック256）。この実施例において、メイン・ループ・コマンドの順次ML1～ML4の順次における次のメイン・ループ順次コマンドは、メイン・ループ順次コマンド・セットの第4のメイン・ループ・コマンドML4である。図7Aに示されるとおり、順次における次のメイン・ループ・コマンドML4は、下流ポート68の下流光トランシーバ72がリセットされることを命令する。

【0079】

したがって、ストレージ・コントローラ4のマルチループ・リンク速度回復論理32（図2）は、下流ポート68に結合された通信リンク34を訓練すること、または再訓練することを開始すべく、この実施例においてはルート・コンプレックスである通信パス・ホスト44に適切なコマンドを発行し、これに应答して、通信パス・ホスト44が、下流ポート68において光トランシーバ72のリセットおよびリブートを開始すべく、帯域外サブパス76a、エンドポイント帯域外コントローラ80、および帯域外サブパス76bを介して下流ポート68の下流光トランシーバ72に適切なコマンドを発行する。

【0080】

下流ポート68の光トランシーバ72をリセットすることに応答して、通信リンク34の訓練が自動的に開始されることが知られる。しかし、本説明によるリンク速度回復の一態様において、下流ポート68の光トランシーバ72の単一のリセットが、通信リンク34の訓練または再訓練の成功を実現しないことがあることを認識されたい。したがって、

ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2 (図 2) は、下流ポート 6 8 の光トランシーバ 7 2 をリセットすること、および光トランシーバ 7 2 がリブートして、訓練が完了するのに十分な時間、待機することの後に続いて、例えば、完全リンク速度などの所望されるリンク速度が、下流ポート 6 8 の光トランシーバ 7 2 をリセットすることによって下流ポート 6 8 において開始された通信リンク訓練または通信リンク再訓練によって実現されているかどうかを判定する (図 6、ブロック 2 6 0)。一実施形態において、特定の応用例に依存して、約 5 秒の待機が適切であることがある。通信リンク 3 4 に関して完全リンク速度が復元されている場合、リンク速度回復は、終了させられる (図 6、ブロック 2 4 0)。

【0081】

逆に、下流ポート 6 8 の光トランシーバ 7 2 をリセットすることによって開始されたリンク速度訓練またはリンク速度再訓練が、通信リンク 3 4 に関して完全リンク速度を実現することに失敗したと判定された場合 (図 6、ブロック 2 6 0)、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2 (図 2) は、メイン・ループ順次コマンド・セット (図 7 A) のすべての順次リンク速度回復コマンドが完了されているかどうかを判定する (ブロック 2 6 4)。この実施例において、メイン・ループ順次コマンド・セット (図 7 A) のすべての順次リンク速度回復コマンド M L 1 ~ M L 4 は、前述されたとおり完了されている。したがって、ストレージ・コントローラ 4 のマルチループ・リンク速度回復論理 3 2 (図 2) は、メイン・ループ実行最大限度が達せられているかどうかを判定する (図 6、ブロック 2 6 8)。一実施形態において、順次リンク速度回復コマンド M L 1 ~ M L 4 のメイン・ループの実行は、例えば、3 回であってよい変数「M」によって代表される最大限度まで繰り返される。しかし、特定の応用例に依存して、例えば、3 ~ 5 の範囲内などの他の最大限度が選択されてよいことを理解されたい。順次リンク速度回復コマンド M L 1 ~ M L 4 のメイン・ループ実行の繰返しは、最大限度 M に達していないと判定された場合 (図 6、ブロック 2 6 8)、順次リンク速度回復コマンド M L 1 ~ M L 4 のメイン・ループ実行は、最大限度 M が達せられるまで、前述されたとおり、ブロック 2 2 4 ~ 2 6 4 において繰り返される。

【0082】

完全リンク速度に通信リンク 3 4 を訓練すること、または再訓練することは、メイン・ループ繰返しの最大限度 M が達せられる前に頻繁に実現されるものと考えられる。例えば、一実施形態により各メイン・ループ実行内で順次リンク速度回復コマンドのサブループの実行を繰り返すこと、および順次リンク速度回復コマンドのメイン・ループの実行を繰り返すことによって、完全リンク速度へのリンク速度回復の信頼性が向上させられてよいものと考えられる。しかし、順次リンク速度回復コマンド M L 1 ~ M L 4 のメイン・ループ実行の繰返しは、完全リンク速度に通信リンク 3 4 を再訓練することを実現することなしに最大限度 M に達していると判定された場合 (図 6、ブロック 2 6 8)、一実施形態において、順次リンク速度回復コマンド M L 1 ~ M L 4 のメイン・ループ実行は、終了させられてよい (ブロック 2 4 0)。さらに、通信パスの障害のある、または欠陥のあるコンポーネントが、交換されてよく、図 6 のリンク速度回復プロセスが、繰り返されてよい。

【0083】

図の計算コンポーネントはそれぞれ、図 8 に示されるコンピュータ・システム 1 0 0 2 などの 1 つまたは複数のコンピュータ・システムにおいて実装されてよい。コンピュータ・システム／サーバ 1 0 0 2 は、コンピュータ・システムによって実行されている、プログラム・モジュールなどのコンピュータ・システム実行可能命令の一般的な脈絡において説明されてよい。一般に、プログラム・モジュールは、特定のタスクを実行する、または特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、コンポーネント、論理、データ構造などを含んでよい。コンピュータ・システム／サーバ 1 0 0 2 は、タスクが、通信ネットワークを介して結び付けられた遠隔処理デバイスによって実行される、分散型クラウド・コンピューティング環境において実施されてよい。分散型クラウド・コンピューティング環境において、プログラム・モジュールは、メモリ・ストレージ・デバ

10

20

30

40

50

イスを含むローカル・コンピュータ・システム記憶媒体と遠隔コンピュータ・システム記憶媒体の両方に配置されてよい。

【0084】

図8に示されるとおり、コンピュータ・システム／サーバ1002は、汎用コンピューティング・デバイスの形態で示される。コンピュータ・システム／サーバ1002のコンポーネントは、1つまたは複数のプロセッサまたは処理装置1004、システム・メモリ1006、ならびにシステム・メモリ1006を含む様々なシステム構成要素をプロセッサ1004に結合するバス1008を含んでよいが、これらには限定されない。バス1008は、様々なバス・アーキテクチャのいずれかを使用する、メモリ・バスもしくはメモリ・コントローラ、周辺バス、アクセラレーテッド・グラフィックス・ポート、およびプロセッサ・バスもしくはローカル・バスを含む、いくつかのタイプのバス構造のいずれかの1つまたは複数を表す。例として、限定としてではなく、そのようなアーキテクチャは、インダストリ・スタンダード・アーキテクチャ（ISA）バス、マイクロ・チャンネル・アーキテクチャ（MCA）バス、エンハンスドISA（EISA）バス、ビデオ・エレクトロニクス・スタンダーズ・アソシエーション（VESA）ローカル・バス、およびペリフェラル・コンポーネント・インターコネクト（PCI）バスを含む。

【0085】

コンピュータ・システム／サーバ1002は、通常、様々なコンピュータ・システム可読媒体を含む。そのような媒体は、コンピュータ・システム／サーバ1002によってアクセス可能である任意の利用可能な媒体であってよく、そのような媒体は、揮発性媒体と不揮発性媒体、取外し可能な媒体と取外し可能でない媒体の両方を含む。

【0086】

システム・メモリ1006は、ランダム・アクセス・メモリ（RAM）1010またはキャッシュ・メモリ1012、あるいはその両方などの揮発性メモリの形態でコンピュータ・システム可読媒体を含むことが可能である。コンピュータ・システム／サーバ1002は、他の取外し可能な／取外し可能でない、揮発性／不揮発性のコンピュータ・システム記憶媒体をさらに含んでよい。単に例として、ストレージ・システム1013が、取外し可能でない、不揮発性の磁気媒体（図示されず、通常、「ハード・ドライブ」と呼ばれる）から読み取ること、およびそのような磁気媒体に書き込むことを行うために備えられることが可能である。図示されないものの、取外し可能な、不揮発性の磁気ディスク（例えば、「フロッピー・ディスク」）から読み取ること、およびそのような磁気ディスクに書き込むことを行うための磁気ディスク・ドライブ、ならびにCD-ROM、DVD-ROM、または他の光媒体などの取外し可能な、不揮発性の光ディスクから読み取ること、またはそのような光ディスクに書き込むことを行うための光ディスク・ドライブが、備えられることが可能である。そのような事例において、各媒体は、1つまたは複数のデータ媒体インタフェースによってバス1008に接続されることが可能である。後段でさらに示され、説明されるとおり、メモリ1006は、本発明の実施形態の機能を実行するように構成されたプログラム・モジュールのセット（例えば、少なくとも1つ）を有する少なくとも1つのプログラム製品を含んでよい。

【0087】

例として、限定としてではなく、プログラム・モジュール1016のセット（少なくとも1つ）を有するプログラム／ユーティリティ1014、ならびにオペレーティング・システム、1つまたは複数のアプリケーション・プログラム、他のプログラム・モジュール、およびプログラム・データが、メモリ1006に記憶されてよい。オペレーティング・システム、1つまたは複数のアプリケーション・プログラム、他のプログラム・モジュール、およびプログラム・データのそれぞれ、あるいはその何らかの組合せは、ネットワーク環境の実装形態を含んでよい。コンピュータ・システム1002のコンポーネントは、本明細書において説明される本発明の実施形態の機能または方法、あるいはその組合せを一般に実行するプログラム・モジュール1016として実装されてよい。図1のシステムは、1つまたは複数のコンピュータ・システム1002において実装されてよく、こ

ここで、システムが複数のコンピュータ・システム 1002 において実装される場合、それらのコンピュータ・システムは、ネットワークを介して通信してよい。

【0088】

また、コンピュータ・システム／サーバ 1002 は、キーボード、ポインティング・デバイス、ディスプレイ 1020、その他などの 1 つまたは複数の外部デバイス 1018、ユーザがコンピュータ・システム／サーバ 1002 と対話することを可能にする 1 つまたは複数のデバイス、またはコンピュータ・システム／サーバ 1002 が他の 1 つまたは複数のコンピューティング・デバイスと通信することを可能にする任意のデバイス（例えば、ネットワーク・カード、モデム、その他）と通信してもよい。そのような通信は、入出力（I/O）インタフェース 1022 経由で行われることが可能である。さらに、コンピュータ・システム／サーバ 1002 は、ネットワーク・アダプタ 1024 を介してローカル・エリア・ネットワーク（LAN）、汎用ワイド・エリア・ネットワーク（WAN）、またはパブリック・ネットワーク（例えば、インターネット）、あるいはその組合せなどの 1 つまたは複数のネットワークと通信することができる。図示されるとおり、ネットワーク・アダプタ 1024 が、バス 1008 を介してコンピュータ・システム／サーバ 1002 の他のコンポーネントと通信する。図示されないものの、他のハードウェア・コンポーネントまたはソフトウェア・コンポーネント、あるいはその組合せが、コンピュータ・システム／サーバ 1002 と連携して使用されることも可能であることを理解されたい。例は、マイクロコード、デバイス・ドライバ、冗長な処理装置、外部ディスク・ドライブ・アレイ、RAID システム、テープ・ドライブ、およびデータ・アーカイブ・ストレージ・システム、その他を含むが、これらには限定されない。

【0089】

i、j、および n などの本明細書において使用される参照文字は、同一の値を表すことも、異なる値を表すこともある、要素の実例の可変の数を表すべく使用され、説明される異なる実例において異なる要素または同一の要素に関して使用される場合、同一の値を表すことも、異なる値を表すこともある。

【0090】

本発明は、システム、方法、またはコンピュータ・プログラム製品、あるいはその組合せであることが可能である。コンピュータ・プログラム製品は、プロセッサに本発明の態様による動作を実行させるためのコンピュータ可読プログラム命令をその上に有するコンピュータ可読記憶媒体（または複数の媒体）を含んでよい。

【0091】

コンピュータ可読記憶媒体は、命令実行デバイスによって使用されるように命令を保持すること、および記憶することができる有形のデバイスであることが可能である。コンピュータ可読記憶媒体は、例えば、電子ストレージ・デバイス、磁気ストレージ・デバイス、光ストレージ・デバイス、電磁ストレージ・デバイス、半導体ストレージ・デバイス、または以上の任意の適切な組合せであってよいが、これらには限定されない。コンピュータ可読記憶媒体のより具体的な例の非網羅的なリストは、以下、すなわち、ポータブル・コンピュータ・ディスク、ハードディスク、ランダム・アクセス・メモリ（RAM）、読取り専用メモリ（ROM）、消去可能なプログラマブル読取り専用メモリ（EPROM もしくはフラッシュ・メモリ）、スタティック・ランダム・アクセス・メモリ（SRAM）、ポータブル・コンパクト・ディスク読取り専用メモリ（CD-ROM）、デジタル・バーサタイル・ディスク（DVD）、メモリ・スティック、フロッピー・ディスク、命令が記録されているパンチカードもしくは溝の中の隆起構造などの機械的に符号化されたデバイス、および以上の任意の適切な組合せを含む。本明細書において使用されるコンピュータ可読記憶媒体は、電波もしくは他の自由に伝播する電磁波、導波路もしくは他の伝達媒体を介して伝播する電磁波（例えば、光ファイバ・ケーブルを通過する光パルス）、または配線を介して伝送される電気信号などの一過性の信号そのものであると解釈されるべきではない。

【0092】

本明細書において説明されるコンピュータ可読プログラム命令は、コンピュータ可読記憶媒体からそれぞれのコンピューティング／処理デバイスに、またはネットワーク、例えば、インターネット、ローカル・エリア・ネットワーク、ワイド・エリア・ネットワーク、または無線ネットワーク、あるいはその組合せを介して外部コンピュータもしくは外部ストレージ・デバイスにダウンロードされることが可能である。ネットワークは、銅伝送ケーブル、伝送光ファイバ、無線伝送、ルータ、ファイアウォール、スイッチ、ゲートウェイ・コンピュータ、またはエッジ・サーバ、あるいはその組合せを備えてよい。各コンピューティング／処理デバイスにおけるネットワーク・アダプタ・カードまたはネットワーク・インタフェースが、ネットワークからコンピュータ可読プログラム命令を受信し、それぞれのコンピューティング／処理デバイス内のコンピュータ可読記憶媒体に記憶されるようにコンピュータ可読プログラム命令を転送する。

10

【0093】

本発明の動作を実行するためのコンピュータ可読プログラム命令は、アセンブラ命令、命令セット・アーキテクチャ（ISA）命令、機械命令、機械依存命令、マイクロコード、ファームウェア命令、状態設定データ、またはJava（R）、Smalltalk（R）、C++、もしくはそれに類するものなどのオブジェクト指向プログラミング言語、および「C」プログラミング言語もしくはそれに類似したプログラミング言語などの従来の手続き型プログラミング言語を含め、1つまたは複数のプログラミング言語の任意の組合せで書かれたソース・コードもしくはオブジェクト・コードであってよい。コンピュータ可読プログラム命令は、全体がユーザのコンピュータ上で実行されても、一部がユーザのコンピュータ上で実行されても、スタンドアロンのソフトウェア・パッケージとして実行されても、一部がユーザのコンピュータ上で、かつ一部が遠隔コンピュータ上で実行されても、全体が遠隔コンピュータもしくは遠隔サーバ上で実行されてもよい。全体が遠隔コンピュータもしくは遠隔サーバ上で実行されるシナリオにおいて、遠隔コンピュータは、ローカル・エリア・ネットワーク（LAN）もしくはワイド・エリア・ネットワーク（WAN）を含む任意のタイプのネットワークを介してユーザのコンピュータに接続されてよく、または接続は、外部コンピュータに対して行われてもよい（例えば、インターネット・サービス・プロバイダを使用してインターネットを介して）。一部の実施形態において、例えば、プログラマブル論理回路、フィールド・プログラマブル・ゲート・アレイ（FPGA）、またはプログラマブル論理アレイ（PLA）を含む電子回路が、本発明の態様を実行するために、電子回路をカスタマイズするようにコンピュータ可読プログラム命令の状態情報を利用することによってコンピュータ可読プログラム命令を実行してよい。

20

30

【0094】

本発明の態様は、本発明の実施形態による方法、装置（システム）、およびコンピュータ・プログラム製品のフローチャートまたはブロック図あるいはその両方を参照して本明細書において説明される。フローチャートまたはブロック図あるいはその両方の各ブロック、ならびにフローチャートまたはブロック図あるいはその両方におけるブロックの組合せは、コンピュータ可読プログラム命令によって実施されることが可能であることが理解されよう。

【0095】

これらのコンピュータ可読プログラム命令は、そのコンピュータまたは他のプログラマブル・データ処理装置のプロセッサを介して実行されるそれらの命令が、フローチャートまたはブロック図あるいはその両方の1つまたは複数のブロックに指定される機能／動作を実施する手段を作り出すべく、汎用コンピュータ、専用コンピュータ、または他のプログラマブル・データ処理装置のプロセッサに提供されてマシンを作り出すものであってよい。また、これらのコンピュータ可読プログラム命令は、命令が記憶されているコンピュータ可読記憶媒体が、フローチャートまたはブロック図あるいはその両方の1つまたは複数のブロックに指定される機能／動作の態様を実施する命令を含む製造品を備えるべく、特定の状態で機能するようにコンピュータ、プログラマブル・データ処理装置、または他のデバイス、あるいはその組合せを導くことができるコンピュータ可読記憶媒体に記憶さ

40

50

れてもよい。

【0096】

また、コンピュータ可読プログラム命令は、コンピュータ、他のプログラマブル装置、または他のデバイス上で実行される命令が、フローチャートまたはブロック図あるいはその両方の1つまたは複数のブロックに指定される機能／動作を実施するように、コンピュータによって実施されるプロセスを作り出すべく、コンピュータ、他のプログラマブル・データ処理装置、または他のデバイスにロードされ、コンピュータ、他のプログラマブル装置、または他のデバイスの上で一連の動作ステップを実行させるものであってもよい。

【0097】

図におけるフローチャートおよびブロック図は、本発明の様々な実施形態によるシステム、方法、およびコンピュータ・プログラム製品の可能な実装形態のアーキテクチャ、機能、および動作を例示する。これに関して、フローチャートまたはブロック図における各ブロックは、指定された論理機能を実施するための1つまたは複数の実行可能命令を備える、命令のモジュール、セグメント、または部分を表すことが可能である。一部の代替の実装形態において、ブロックに記載される機能は、図に記載される順序を外れて生じてよい。例えば、連続して示される2つのブロックが、実際には、実質的に同時に実行されてよく、またはそれらのブロックが、ときとして、関与する機能に依存して、逆の順序で実行されてよい。また、ブロック図またはフローチャートあるいはその両方の各ブロック、ならびにブロック図またはフローチャートあるいはその両方におけるブロックの組合せは、指定された機能もしくは動作を実行する、または専用ハードウェア命令とコンピュータ命令の組合せを実行する専用ハードウェア・ベースのシステムによって実施されることが可能であることに留意されたい。

【0098】

「或る実施形態」、「実施形態」、「いくつかの実施形態」、「その実施形態」、「それらの実施形態」、「1つまたは複数の実施形態」、「一部の実施形態」、および「一実施形態」という用語は、そうではないことが明記されない限り、「本発明の1つまたは複数の（ただし、すべてではない）実施形態」を意味する。

【0099】

「含む」、「備える」、「有する」という用語、およびその変形は、そうではないことが明記されない限り、「含むが、限定されない」ことを意味する。

【0100】

アイテムの列挙されるリストは、そうではないことが明記されない限り、アイテムのいずれもが、またはすべてが相互排他的であることを暗示するものではない。

【0101】

「或る」および「その」という用語は、そうではないことが明記されない限り、「1つまたは複数の」を意味する。

【0102】

互いに通信するデバイスは、そうではないことが明記されない限り、互いに継続的に通信している必要はない。さらに、互いに通信するデバイスは、直接に通信しても、1つまたは複数の仲介を介して通信してもよい。

【0103】

互いに通信するいくつかのコンポーネントを有する実施形態の説明は、すべてのそのようなコンポーネントが要求されることを暗示するわけではない。それどころか、様々なオプションのコンポーネントが、本発明の多種多様な可能な実施形態を例示すべく説明される。

【0104】

単一のデバイスまたは品目について本明細書において説明されるとき、単一のデバイス／品目の代わりに複数のデバイス／品目（それらが協働するか否かにかかわらず）が使用されてよいことが直ちに明白であろう。同様に、複数のデバイスまたは品目について本明細書において説明される場合（それらが協働するか否かにかかわらず）が、その複数のデ

バイスまたは品目の代わりに単一のデバイス／品目が使用されてよいこと、あるいは示される数のデバイスまたはプログラムの代わりに異なる数のデバイス／品目が使用されてよいことが直ちに明白であろう。デバイスの機能または特徴、あるいはその組合せは、代替として、そのような機能／特徴を有するものとして明示的に説明されない他の1つまたは複数のデバイスによって実体化されてよい。それ故、本発明の他の実施形態は、そのデバイス自体を含む必要はない。

【0105】

本発明の様々な実施形態の以上の説明は、例示および説明の目的で提示されてきた。以上の説明は、網羅的であることも、開示される形態そのものに本発明を限定することも意図していない。前段の教示に鑑みて、多くの変形形態および変更形態が可能である。本発明の範囲は、この詳細な説明によってではなく、本明細書に添付された特許請求の範囲によって限定されることが意図される。以上の詳細、実施例、およびデータは、本発明の構成の製造および使用の完全な説明を提供する。本発明の多くの実施形態が、本発明の思想および範囲を逸脱することなく作成されることが可能であるので、本発明は、本明細書に添付された特許請求の範囲に存する。

10

20

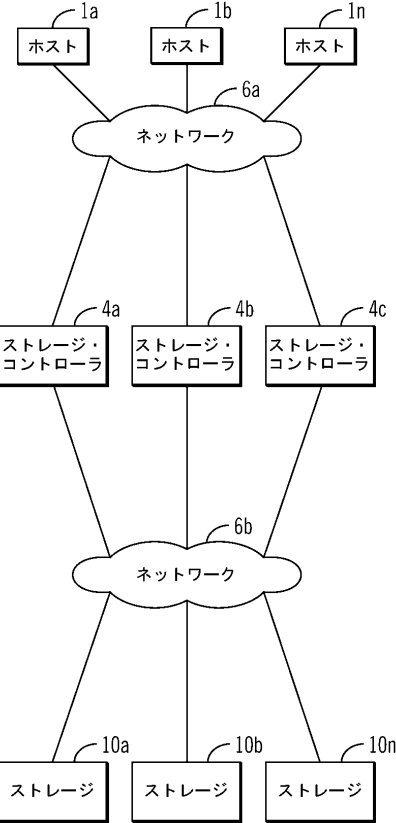
30

40

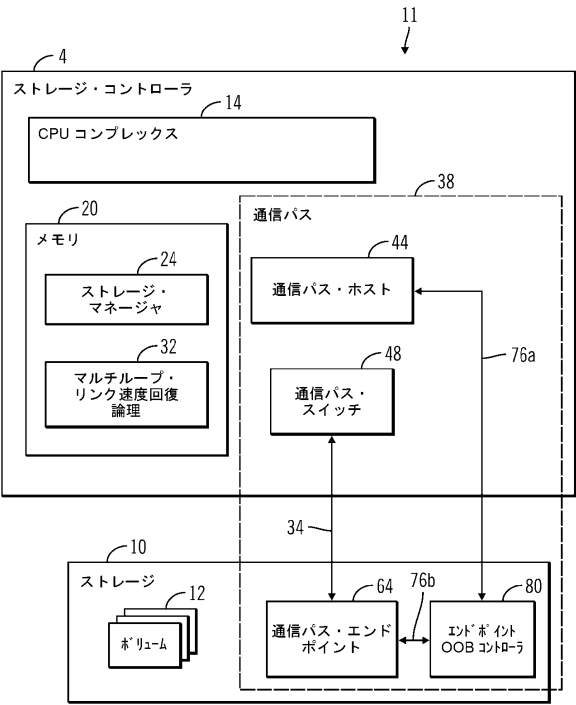
50

【図面】

【図 1】



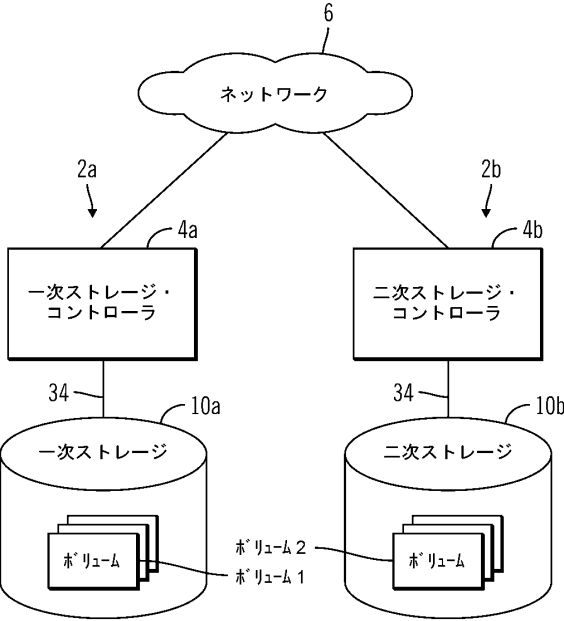
【図 2】



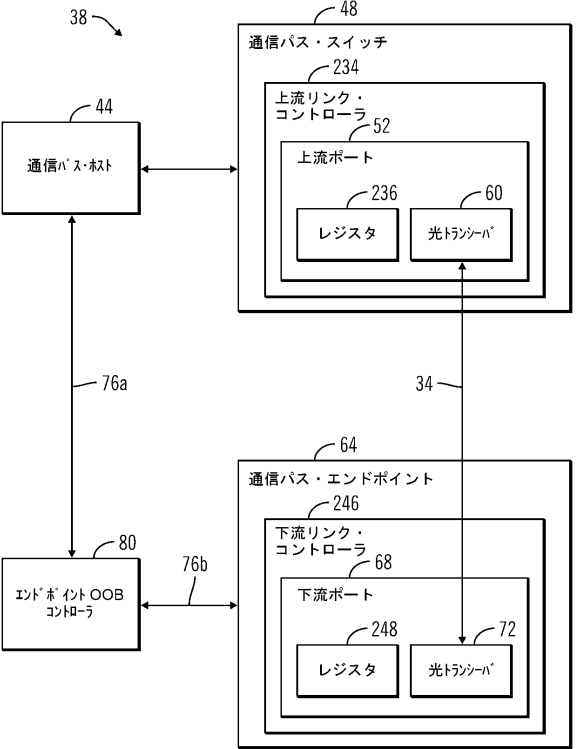
10

20

【図 3】



【図 4】

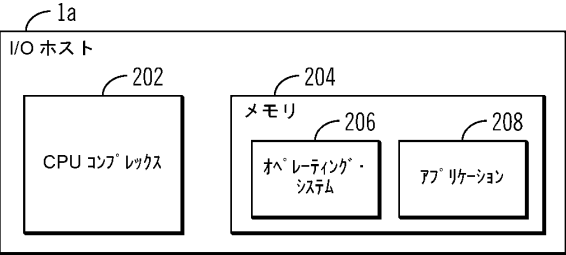


30

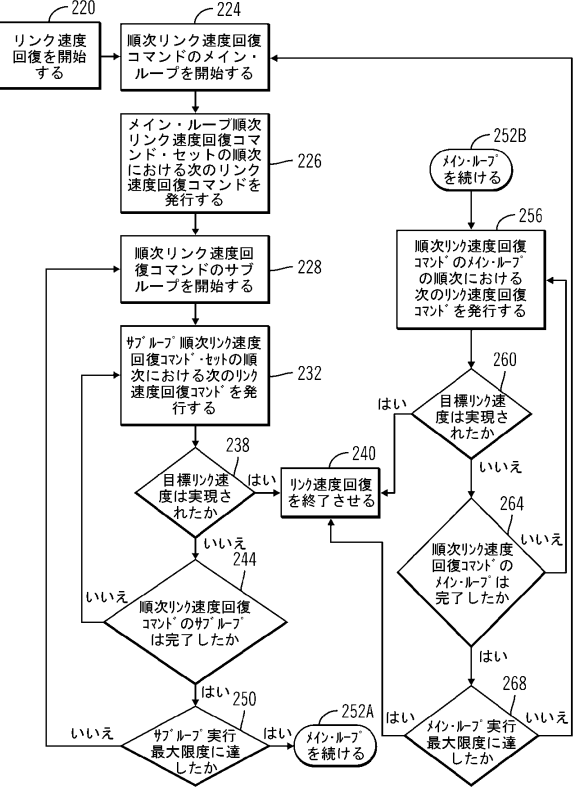
40

50

【図 5】



【図 6】



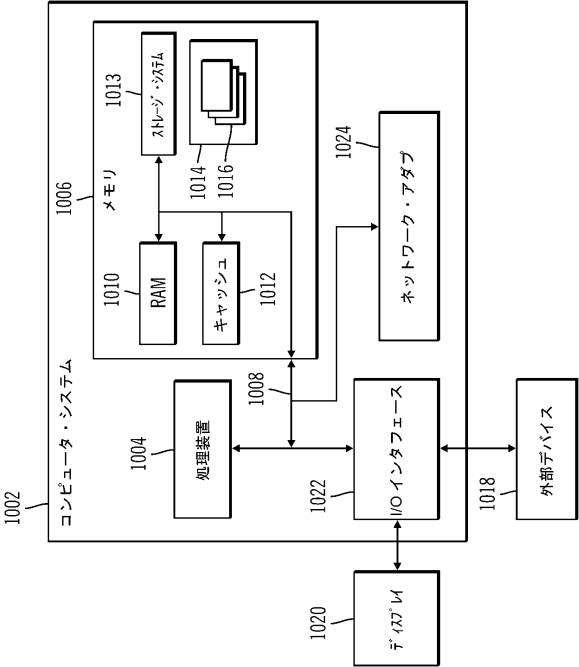
【図 7 A】

メイン・ループ順次コマンド・セット	
メイン・ループ・コマンド 順次番号	コマンド
ML1	サブ・ループ・コマンドを開始する
ML2	通信バス・スイッチをリセット・再初期設定する
ML3	上流光トランシーバをリセットする
ML4	下流ポートにおいて下流トランシーバをリセットすべく OOB 信号を送信する

【図 7 B】

サブ・ループ順次コマンド・セット	
サブ・ループ・コマンド・順 次番号	コマンド
SL1	上流ポートにおいてリンクを再訓練する
SL2	下流ポートにおいてリンクを再訓練すべく OOB 信号を送信する
SL3	上流ポートをディスエーブル・イネーブル・サイクルする
SL4	下流ポートをディスエーブル・イネーブル・サイクルすべく OOB 信号を送信する

【図 8】



10

20

30

40

50

フロントページの続き

(51)国際特許分類

G 0 6 F 13/42 (2006.01)

F I
G 0 6 F 13/38 3 5 0
G 0 6 F 13/42 3 1 0

(72)発明者 バーク、シェーマス
 アメリカ合衆国8 5 7 4 4-0 0 0 2 アリゾナ州ツーソン サウス・リタ・ロード9 0 0 0

(72)発明者 カーソン、マシュー
 アメリカ合衆国9 5 1 2 0-6 0 9 9 カリフォルニア州サンノゼ アルマデン・リサーチ・センタ
 ー ハリー・ロード6 5 0

(72)発明者 バッチェラー、ゲイリー、ウィリアム
 アメリカ合衆国8 5 7 4 4-0 0 0 2 アリゾナ州ツーソン サウス・リタ・ロード9 0 0 0

(72)発明者 マックック、ハイディ、リーン
 アメリカ合衆国8 5 7 4 4-0 0 0 2 アリゾナ州ツーソン サウス・リタ・ロード9 0 0 0

審査官 田中 啓介

(56)参考文献 米国特許出願公開第2 0 1 3 / 0 0 5 1 4 8 3 (U S , A 1)
 特開2 0 1 1 - 2 4 8 8 1 4 (J P , A)
 特開2 0 1 5 - 1 2 1 8 3 7 (J P , A)

(58)調査した分野 (Int.Cl., D B名)
 G 0 6 F 3 / 0 6 - 3 / 0 8
 G 0 6 F 1 1 / 0 7、1 1 / 1 4
 G 0 6 F 1 1 / 2 8 - 1 1 / 3 6
 G 0 6 F 1 3 / 1 0 - 1 3 / 1 4
 G 0 6 F 1 3 / 3 8 - 1 3 / 4 2