

【公報種別】特許法第 17 条の 2 の規定による補正の掲載

【部門区分】第 6 部門第 3 区分

【発行日】平成23年2月3日(2011.2.3)

【公表番号】特表2010-512584(P2010-512584A)

【公表日】平成22年4月22日(2010.4.22)

【年通号数】公開・登録公報2010-016

【出願番号】特願2009-540484(P2009-540484)

【国際特許分類】

G 0 6 F 3/06 (2006.01)

G 0 6 F 12/00 (2006.01)

G 0 6 F 21/24 (2006.01)

【 F I 】

G 0 6 F 3/06 3 0 1 G

G 0 6 F 3/06 3 0 1 F

G 0 6 F 12/00 5 3 7 H

G 0 6 F 12/14 5 6 0 D

G 0 6 F 12/14 5 4 0 A

【手続補正書】

【提出日】平成22年12月6日(2010.12.6)

【手続補正 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

データ管理装置において、

空データトークン指令を生成するトークン指令生成モジュールであって、クライアントの要求用のストレージ要求に応じてストレージデバイスにデータセグメントトークンを記憶する要求を具え、前記データセグメントトークンが空データセグメントを含むデータセグメントに置換されるトークン指令生成モジュールと、

前記空データトークン指令を前記ストレージデバイスに転送するトークン指令転送モジュールと、

を具えることを特徴とする装置。

【請求項 2】

前記トークン指令生成モジュールが、前記ストレージデバイスの既存データを上書きする要求に応じて、空データトークン指令および安全消去コマンドの両方を生成し、前記既存データは、前記ストレージデバイスにおいて、前記空データトークン指令中のデータセグメント識別子と同一のデータセグメント識別子で識別され、前記安全消去コマンドは、前記ストレージデバイスが前記既存データを修正するように指令して、前記既存データはリカバリ不能となり、

前記トークン指令モジュールは、前記安全消去コマンドをさらに転送することを特徴とする請求項 1 に記載の装置。

【請求項 3】

前記既存データがリカバリ不能となるように、前記ストレージデバイスの前記既存データが上書きされたことの確認信号を受信する消去確認モジュールを更に具えることを特徴とする請求項 2 に記載の装置。

【請求項 4】

前記トークン指令生成モジュールが、前記ストレージデバイスの既存データを上書きする要求に回答して、空データトークン指令および暗号化キー消去コマンドの両方を生成し、前記既存データは、前記ストレージデバイスにおいて、前記空データトークン指令中の前記データセグメント識別子と同一のデータセグメント識別子で識別され、前記既存データは、前記ストレージデバイスに前記既存データを記憶するのに関連して受信した暗号化キーを用いて暗号化され、

前記暗号化キー消去コマンドは、前記既存データを記憶するために使用された前記暗号化キーを消去して、前記暗号化キーをリカバリ不能にすることを特徴とする請求項 1 に記載の装置。

【請求項 5】

前記空データトークン指令が、前記データセグメントトークンを具えることを特徴とする請求項 1 に記載の装置。

【請求項 6】

前記空データトークン指令は前記データセグメントトークンを含まず、前記データセグメントトークンを生成するための、前記ストレージデバイスに対するコマンドを具えることを特徴とする請求項 1 に記載の装置。

【請求項 7】

前記空データトークン指令は、一連の反復データからの反復データの少なくとも 1 つのインスタンスをさらに具えることを特徴とする請求項 1 に記載の装置。

【請求項 8】

前記空データトークン指令が、データセグメント識別子、データセグメント長、及びデータセグメントロケーションのうちの 1 以上を具えることを特徴とする請求項 1 に記載の装置。

【請求項 9】

前記ストレージデバイスから前記データセグメントを読み取るストレージ要求に応じて、前記ストレージデバイスで記憶される前記データセグメントトークンに対応する前記ストレージデバイスからメッセージを受信する読み取りトークン指令レシーバモジュールであって、前記メッセージが前記データセグメントを表す読み取りトークン指令レシーバモジュールと、

前記ストレージデバイスから受信した前記メッセージから構築された要求クライアントに回答信号を転送する、要求クライアント応答モジュールと、
を更に具えることを特徴とする請求項 1 に記載の装置。

【請求項 10】

前記メッセージに含まれる情報を使用して、前記データセグメントのデータを再構成するデータセグメント再生モジュールを更に具え、前記要求クライアントに送られる前記応答信号が前記データセグメントを具えることを特徴とする請求項 9 に記載の装置。

【請求項 11】

前記要求クライアントに送られる前記応答信号は、前記ストレージデバイスから受信した前記メッセージに含まれる情報を具えることを特徴とする請求項 9 に記載の方法。

【請求項 12】

前記メッセージが、前記データセグメントトークンを具えることを特徴とする請求項 9 に記載の装置。

【請求項 13】

前記空データトークン指令を有する前記ストレージ要求が、前記ストレージデバイスのストレージスペースをリザーブする要求を更に具え、要求されたりザーブしたストレージスペースは、前記空データトークン指令によって特定される長さと実質的に同等のストレージスペース量を具えることを特徴とする請求項 1 に記載の装置。

【請求項 14】

サーバで実行する前記要求クライアントを更に具え、前記要求クライアントは、アプリケーション、ファイルサーバおよびクライアントの少なくとも 1 つを具えることを特徴と

する請求項 1 に記載の装置。

【請求項 15】

データを管理するオペレーションを行うために実行可能なコンピュータ使用可能プログラムコードを有するコンピュータ読み取り媒体を具えたコンピュータプログラム製品において、前記コンピュータプログラム製品のオペレーションが、

空データトークン指令を有するストレージ要求を生成するステップであって、一連の回復データを含むデータセグメントに応じてデータセグメントトークンをストレージデバイスに記憶する要求を具え、前記データセグメントトークンが、前記データセグメントのデータが実質的にない前記データセグメント、前記データセグメントトークン、及び前記空データトークン指令に置換されるステップと、

前記ストレージデバイスに前記空データトークン指令を転送するステップと、
を具えることを特徴とするコンピュータプログラム製品。

【請求項 16】

ストレージ制御部によって管理される不揮発性ストレージメディアに記憶されるデータを管理するための方法であって、当該方法が、

不揮発性ストレージメディアを管理するように構成されるストレージ制御部に指令を送信するステップであって、前記不揮発性ストレージメディアにおける 1 又はそれ以上の以前に有効なブロックが、無効なブロックであることを前記指令が示し、前記以前に有効なブロックが既存データを記憶するステップと、

前記 1 又はそれ以上の以前に有効なブロックを、前記指令に応じた無効なブロックとして記憶するステップと、
を具えることを特徴とする方法。

【請求項 17】

請求項 16 に記載の方法において、前記ストレージ制御部が追記のみの書き込みプロセスを用いて前記不揮発性ストレージメディアにデータを記憶するよう構成され、かつ無効なブロックを記憶する不揮発性ストレージメディアを再利用するストレージスペースリカバリプロセスを用いて、前記不揮発性ストレージメディアにストレージスペースを回復させるように構成されることを特徴とする方法。

【請求項 18】

請求項 16 に記載の方法において、前記指令が 1 又はそれ以上の以前に有効なブロックに対するヒントを含み、前記ストレージ制御部が選択的に、前記ヒントに応じてストレージスペースリカバリプロセスを拡張することを特徴とする方法。

【請求項 19】

請求項 16 に記載の方法が、前記指示への適合を承認する受取り文字を受信するステップを更に具えること特徴とする方法。

【請求項 20】

請求項 16 に記載の方法において、前記不揮発性ストレージメディアが利用可能なストレージ容量を維持するのにストレージスペースリカバリを要求し、前記ストレージ制御部が無効なブロックとして記録したブロックを再利用することによってストレージスペースを回復させることを特徴とする方法。

【請求項 21】

請求項 16 に記載の方法において、前記無効なブロックを削除した既存データのブロック、及び消去した既存データのブロックのうちの 1 つを含むことを特徴とする方法。

【請求項 22】

ストレージ制御部によって管理される不揮発性ストレージメディアに記憶されるデータを管理するための方法であって、当該方法が、

不揮発性ストレージメディアを管理するように構成されるストレージ制御部に指令を送信するステップであって、前記不揮発性ストレージメディアの 1 又はそれ以上の以前に有効なデータパケットが、無効なデータパケットであることを前記指令が示し、前記以前に有効なデータパケットが既存データを記憶するステップと、

前記 1 又はそれ以上の以前に有効なデータパケットが、前記指令に応じた無効なデータパケットとして標識されるステップと、
を具えることを特徴とする方法。

【請求項 2 3】

請求項 2 2 に記載の方法において、前記ストレージ制御部が追記のみの書き込みプロセスを用いて前記不揮発性ストレージメディアにデータを記憶するよう構成され、かつ前記無効なデータパケットを記憶する不揮発性ストレージメディアを再利用するストレージスペースリカバリプロセスを用いて、前記不揮発性ストレージメディアにストレージスペースを回復させるように構成されることを特徴とする方法。

【請求項 2 4】

請求項 2 2 に記載の方法において、前記指令が 1 又はそれ以上の以前に有効なデータパケットに対するヒントを含み、前記ストレージ制御部が選択的に、前記ヒントに応じてストレージスペースリカバリプロセスを拡張することを特徴とする方法。

【請求項 2 5】

請求項 2 2 に記載の方法が、前記指示への適合を承認する受取り文字を受信するステップを更に具えること特徴とする方法。

【請求項 2 6】

請求項 2 2 に記載の方法において、前記不揮発性ストレージメディアが利用可能なストレージ容量を維持するのにストレージスペースリカバリを要求し、前記ストレージ制御部が無効なデータパケットとして標識されたデータパケットを再利用することによってストレージスペースを回復させることを特徴とする方法。

【請求項 2 7】

データを管理するための装置であって、当該装置が、
空データトークン指令を生成するトークン指令生成モジュールであって、クライアントの要求による関連データセグメントの削除に応じて、ストレージデバイスに空データセグメントを記憶する要求を具えるトークン指令生成モジュールと、
前記空データトークン指令を前記ストレージデバイスに転送するトークン指令転送モジュールと、
を具えることを特徴とする装置。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【発明の詳細な説明】

【発明の名称】空データトークン指令を有する要求デバイスからのデータを管理する装置、システムおよび方法

【技術分野】

【0001】

関連出願についての相互参照

この出願は、David Flynnらによって2006年12月6日に提出された米国仮出願番号第60/873,111号、発明の名称「エレメンタルブレードシステム」と、David Flynnらによって2007年9月22日に提出された米国仮出願番号60/974,470号、発明の名称「オブジェクト指向型ソリッドステージストレージ用の装置、システムおよび方法」と、の優先権を主張し、上記仮出願は、参照することで本明細書に組み込まれるものとする。

【背景技術】

【0002】

発明の分野

本発明は、データストレージデバイス中のデータ管理に関し、より詳しくは、空データ

セグメント指令を使用したストレージデバイス中のデータ管理に関する。

【0003】

関連技術の説明

通常、データがもはや有用でない場合、そのデータは消去してもよい。多くのファイルシステムにおいて、消去コマンドは、ファイルシステム中のディレクトリ項目を削除するが、データを含んでいるストレージデバイス中にはデータは残される。通常、データストレージデバイスは、この種の消去オペレーションには関連していない。データを消去する別の方法は、ゼロ、1、または、その他のヌルデータ文字を、データストレージデバイスに書き込み、実際に消去されたファイルを置換する。しかしながら、このことが非効率的であるのは、上書きされるデータを転送するとき、有効なバンド幅が使用されてしまうからである。

【0004】

いくつかのストレージデバイスにおいては、本明細書で記載されているソリッドステートストレージデバイス102のように、追記型及びガーベージコレクション型のストレージデバイスではないので、以前に記憶されていたデータをアップデートすることで、既存のデータは上書きされない。このようなデバイスに1のストリングまたはゼロのストリングで上書きする試みは、既存のデータを上書きする意図を満たすことなく、有用なスペースを占有するだけである。これらの追記型及びガーベージコレクション型のデバイスについては、ソリッドステートストレージデバイス102など、クライアント114は、通常、データを消去するためにデータを上書きする能力を有していない。

【0005】

反復文字のストリングまたは文字ストリングを受信したとき、受信データは、高度に圧縮可能であるが、通常、圧縮は、ストレージデバイスに転送する前にファイルシステムによってなされる。一般的なストレージデバイスは、圧縮データと未圧縮データとを区別することができない。さらに、このようなストレージデバイスは、消去されるファイルを読み取るコマンドを受信してから、ストレージデバイスは、要求デバイスにゼロ、1またはヌル文字のストリームを転送することができる。繰り返しとなるが、消去ファイルを表すデータを転送するためのバンド幅が必要となる。

【発明の概要】

【0006】

上述した説明から、ストレージデバイスが、空データセグメント、または、反復文字若しくは文字ストリングを有するデータ、を表すデータセグメントトークンを記憶することができるように、データを消去する指令を受信するストレージデバイスに関する装置、システムおよび方法が必要なことが明らかである。この装置、システムおよび方法は、既存のデータも消去することができ、最終的に使用されているストレージスペースは、小さなデータセグメントトークンを具える。開示される装置、システムおよび方法は、先行技術の欠陥のいくつか又は全てを解決する。

【0007】

本発明は、当分野の現状に合わせて開発され、具体的には、現在入手できるデータ管理システムによって完全に解決されていない当分野の問題および要求に対応している。従って、本発明は、上述した当分野の問題点の多くまたは全てを克服するデータ管理装置、システムおよび方法を提供するために開発された。

【0008】

データを管理する装置は、トークン指令生成モジュールおよびトークン指令転送モジュールを含む複数のモジュールを具える。トークン指令生成モジュールは、トークン指令を有するストレージ要求を生成する。このトークン指令は、ストレージデバイスにデータセグメントトークンを記憶する要求を具える。トークン指令は、データセグメントとして記憶される一連の反復同一文字または一連の反復同一文字ストリングを置換する。トークン指令は、少なくともデータセグメント識別子およびデータセグメント長を具える。データセグメントトークンおよびトークン指令は、データセグメントのデータが実質的にない。

トークン指令転送モジュールは、トークン指令をストレージデバイスに転送する。

【0009】

一実施例においては、トークン指令生成モジュールは、ストレージデバイスの既存データを上書きする要求にตอบสนองするトークン指令および安全消去コマンドの両方を生成する。既存データは、ストレージデバイスにおいて、トークン指令中のデータセグメント識別子と同一のデータセグメント識別子として識別される。安全消去コマンドは、ストレージデバイスが既存データを修正するように指令し、既存データはリカバリ不可能である。一実施例においては、トークン指令モジュールは、安全消去コマンドを転送する。さらなる実施例においては、本発明の装置は、消去確認モジュールを具え、この消去確認モジュールは、既存データがリカバリ不能なように、ストレージデバイスの既存データが文字で上書きされたという確認信号 (c o n f i r m a t i o n) を受信する。

【0010】

一実施例においては、トークン指令生成モジュールは、トークン指令と、ストレージデバイスの既存のデータを上書きする要求にตอบสนองする暗号化キー消去コマンドと、の両方を生成する。既存データは、ストレージデバイスでは、トークン指令中のデータセグメント識別子と同一のデータセグメント識別子として識別され、既存データは、ストレージデバイスの既存データの記憶と合わせて受信される暗号化キーを用いて暗号化される。一実施例においては、暗号化キー消去コマンドは、既存データを記憶するのに使用された暗号化キーを消去し、暗号化キーはリカバリ不能となる。

【0011】

一実施例においては、トークン指令は、データセグメントトークンを具える。別の実施例においては、トークン指令は、データセグメントトークンがなく、データセグメント指令を生成するストレージデバイス用コマンドを具える。別の実施例においては、トークン指令は、反復同一文字のインスタンスおよび/または反復同一文字ストリングのインスタンスの少なくとも一方を具える。

【0012】

一実施例においては、本発明の装置は、読み取りレシーバモジュール、読み取り要求モジュール、読み取りトークン指令レシーバモジュールおよび要求クライアント応答モジュールを具える。読み取りレシーバモジュールは、ストレージデバイスからデータセグメントを読み取るストレージ要求を受信する。ストレージ要求は、要求クライアントから受信される。読み取り要求転送モジュールは、ストレージデバイスにストレージ要求を転送する。読み取りトークン指令レシーバモジュールは、ストレージデバイスから要求されたデータセグメントトークンに対応するメッセージを受信し、このメッセージは、少なくともデータセグメント識別子およびデータセグメント長を具える。このメッセージは、データセグメントのデータが実質的でない。要求クライアント応答モジュールは、ストレージデバイスから受信されたメッセージから構築 (f o r m u l a t e) された要求クライアントにตอบสนอง信号を転送する。

【0013】

さらなる実施例においては、本発明の装置は、メッセージに含まれる情報を使用してデータセグメントのデータを再構成するデータセグメント再生成モジュールを具え、要求クライアントに送られる上記ตอบสนอง信号は、データセグメントを具える。関連した実施例においては、要求クライアントに送られるตอบสนอง信号は、ストレージデバイスから受信されたメッセージに含まれる情報を具える。別の実施例においては、このメッセージは、データセグメントトークンを具える。

【0014】

一実施例においては、一連の反復同一文字または文字ストリングは、データセグメントが空 (e m p t y) であることを示す。別の実施例においては、トークン指令を有するストレージ要求は、ストレージデバイスのストレージスペースをリザーブする要求を具える。要求されてリザーブされたストレージスペースは、データセグメント長と実質的に同様のストレージスペース量を具える。

【 0 0 1 5 】

別の装置は、上述したものと実質的に同様な、読み取りレシーバモジュール、読み取り要求転送モジュール、読み取りトークンレシーバモジュールおよび要求クライアント応答モジュールを具える。一実施例においては、本発明の装置は、メッセージ中に含まれる情報を使用して、データセグメントのデータを再構成するデータセグメント再生成モジュールを具え、要求クライアントに送られる応答信号は、データセグメントを具える。別の実施例においては、要求クライアントに送られる応答信号は、ストレージデバイスから受信されたメッセージに含まれる情報を具える。

【 0 0 1 6 】

本発明のシステムは、さらに、データを管理するために提供される。このシステムは、ストレージデバイスおよびクライアントと通信するサーバを具える。このサーバは、トークン指令生成モジュールおよびトークン指令転送モジュールを具える。このトークン指令生成モジュールは、トークン指令を有するストレージ要求を生成する。トークン指令は、データセグメントトークンをストレージデバイスに記憶する要求を具える。トークン指令は、データセグメントを置換する。トークン指令は、少なくともデータセグメント識別子およびデータセグメント長を具え、データセグメントトークンは、データセグメントのデータが実質的にない。トークン指令転送モジュールは、トークン指令をストレージ要求に転送する。

【 0 0 1 7 】

一実施例においては、本発明のシステムは、読み取りレシーバモジュール、読み取り要求転送モジュール、読み取りトークン指令レシーバモジュールおよび要求クライアント応答モジュールを具える。ストレージ要求はクライアントから受信される。読み取り要求転送モジュールはストレージ要求をストレージデバイスに転送する。読み取りトークン指令レシーバモジュールは、ストレージデバイスから、要求されたデータセグメントトークンに対応するメッセージを受信する。このメッセージは、少なくともデータセグメント識別子およびデータセグメント長を具える。このメッセージは実質的にデータセグメントデータがない。要求クライアント応答モジュールは、ストレージデバイスから受信したメッセージから構築されたクライアントに応答信号を転送する。一実施例においては、要求クライアントは、サーバ上で実行し、要求クライアントは、アプリケーション、ファイルサーバおよびクライアントの少なくとも1つを具える。

【 0 0 1 8 】

本発明の方法は、データを管理するために提供される。開示された実施例の方法は、記載された装置及びシステムのオペレーションに関して上述した機能を実行するのに必要なステップを具える。一実施例においては、この方法は、トークン指令を有するストレージ要求を生成するステップを具える。このトークン指令は、データセグメントトークンをストレージデバイスに記憶する要求を具える。トークン指令は、データセグメントを置換し、トークン指令は、少なくともデータセグメント識別子およびデータセグメント長を具える。データセグメントトークンは、データセグメントのデータが実質的にない。この方法は、トークン指令をストレージデバイスに転送するステップを具える。

【 0 0 1 9 】

データを管理する代替方法は、ストレージデバイスからデータセグメントを読み取るストレージ要求を受信するステップを具え、ストレージ要求は、要求クライアントから受信される。この方法は、ストレージデバイスにストレージ要求を転送するステップを具える。この方法は、ストレージデバイスから、要求されたデータセグメントトークンに対応するメッセージを受信するステップを具える。このメッセージは、少なくともデータセグメント識別子およびデータセグメント長を具える。このメッセージは、データセグメントのデータが実質的にない。この方法は、ストレージデバイスから受信されたメッセージから構築された要求クライアントに応答信号を転送するステップを具える。

【 0 0 2 0 】

本明細書全体に亘って、特徴、利点または同様の単語は、本発明によって実現される特

徴および利点の全てが、本発明のいずれかの１つの実施例に存在する可能性がある、または、存在することを意味するものではない。むしろ、特徴及び利点に関する用語は、一実施例に関連して記載される特定の特徴、利点または特性が、本発明の少なくとも１つの実施例に含まれることを意味すると理解されたい。従って、特徴、利点および同様の用語については、本明細書全体に亘って、必須ではないが、同一の実施例を意味する。

【００２１】

さらに、記載された本発明の特徴、利点および特性は、１又はそれ以上の実施例においてあらゆる適宜な方法で組み合わせることができる。当分野の当業者であれば、特定の実施例の１又はそれ以上の特定の特徴または利点なしで、本発明を実施可能であることを理解されたい。その他の例においては、更なる特徴および利点は、本発明の全ての実施例において示されていない所定の実施例を理解されたい。

【００２２】

本発明の特徴および利点は、以下の説明および添付の請求の範囲からより明らかになり、また、以下に記載する本発明の実施形態によって理解されるであろう。

【図面の簡単な説明】

【００２３】

本発明の利点が容易に理解されるように、簡単に上述した本発明のより具体的な説明は、添付の図面に例示した特定の実施例を参照することによってなされるであろう。これらの図面は、本発明の通常の実施例のみを示し、従って、本発明の範囲を限定するものではないと理解されたい。本発明は、添付の図面を利用して、具体的かつ詳細に記載および説明される。

【図１Ａ】図１Ａは、本発明によるソリッドステートストレージデバイス中のデータ管理システムの一実施例を示した概略ブロック図である。

【図１Ｂ】図１Ｂは、本発明によるストレージデバイス中のオブジェクト管理システムの一実施例を示した概略ブロック図である。

【図２Ａ】図２Ａは、本発明によるストレージデバイスのオブジェクト管理装置の一実施例を示した概略ブロック図である。

【図２Ｂ】図２Ｂは、本発明によるソリッドステートストレージデバイス中のソリッドステートストレージデバイス制御部の一実施例を示した概略ブロック図である。

【図３】図３は、本発明によるソリッドステートストレージデバイスのライトデータパイプラインおよびリードデータパイプラインを有するソリッドステートストレージ制御部を示す概略ブロック図である。

【図４Ａ】図４Ａは、本発明によるソリッドステートストレージ制御部のバンクインターリーブ制御部の一実施例を示した概略ブロック図である。

【図４Ｂ】図４Ｂは、本発明によるソリッドステートストレージ制御部のバンクインターリーブ制御部の代替実施例を示した概略ブロック図である。

【図５】図５は、本発明によるデータパイプラインを使用したソリッドステートストレージデバイスのデータ管理方法に関する一実施例を例示した概略フローチャートである。

【図６】図６は、本発明によるデータパイプラインを使用して、ソリッドステートストレージデバイスのデータ管理方法に関する別の実施例を示した概略フローチャート図である。

【図７】図７は、本発明によるバンクインターリーブを使用したソリッドステートストレージデバイスのデータ管理方法に関する一実施例を示した概略フローチャート図である。

【図８】図８は、本発明によるソリッドステートストレージデバイスの不要データコレクション装置に関する一実施例を示した概略ブロック図である。

【図９】図９は、本発明によるソリッドステートストレージデバイスの不要データコレクション装置の一実施例を示した概略フローチャート図である。

【図１０】図１０は、本発明によるトークン指令を生成する装置を有するシステムに関した一実施例を示した概略ブロック図である。

【図１１】図１１は、本発明によるトークン指令を生成および転送する方法の一実施例を

示した概略フローチャート図である。

【図 1 2】図 1 2 は、本発明によるデータセグメントトークンを読み取る方法に関する一実施例を示した概略ブロック図である。

【図 1 3】図 1 3 は、本発明によるデータセグメントトークンを管理する装置を有するシステムの一実施例を示した概略ブロック図である。

【図 1 4】図 1 4 は、本発明によるデータセグメントトークンを記憶する方法に関する一実施例を示した概略フローチャート図である。

【図 1 5】図 1 5 は、本発明によるデータセグメントトークンを読み取る方法に関する一実施例を示した概略フローチャート図である。

【発明を実施するための形態】

【0024】

本明細書に記載された多数の機能ユニットは、そのインプリメンテーション独立性をより具体的に強調するために、モジュールとしてラベルした。例えば、モジュールは、カスタム VLSI 回路またはゲートアレイ、論理チップなど標準規格の半導体、トランジスタまたは他のディスクリートコンポーネントを具えるハードウェア回路として実装可能である。さらに、モジュールは、フィールドプログラム可能ゲートアレイ、プログラム可能アレイロジック、プログラム可能ロジックデバイスなどプログラム可能ハードウェアデバイスに実装可能である。

【0025】

さらに、モジュールは、各種プロセッサによって実行されるソフトウェアに実装可能である。例えば、実行可能コードからなる特定モジュールは、オブジェクト、プロシージャまたはファンクションとして組織される物理的又は論理的ブロックの 1 又はそれ以上のコンピュータ指令を具える。にもかかわらず、実行可能な特定モジュールは、物理的に一カ所に配置される必要はないが、異なるロケーションに記憶される異なる指令を具え、論理的に同時に接続される場合、モジュールを具え、モジュール用の宣言された目的を達成する。

【0026】

実際に、実行可能コードからなるモジュールは、1 つの指令または多数の指令であり、複数の異なるコードセグメントに亘って、異なるプログラム中で、および、複数のメモリデバイスに亘って、分配され得る。同様に、オペレーションデータは、モジュール内で特定され、例示されており、適宜な形態で組み込まれ、あらゆる適宜な種類のデータ構造内で組織される。オペレーションデータは、1 つのデータセットとして集められ、異なるストレージデバイスを含む異なるロケーションに亘って配置され、少なくとも部分的に、システムまたはネットワークの単なる電気信号として存在し得る。モジュールまたはモジュールの一部がソフトウェアに実装される場合、ソフトウェア部は、1 又はそれ以上のコンピュータ可読メディアに記憶される。

【0027】

実施例に関連して記載される特定の特徴部、構造または特性を意味する、本明細書全体における「一実施例」、「実施例」または同様の用語については、本発明の少なくとも 1 つの実施例を含む。従って、本明細書全体における「一実施例においては」、「実施例においては」および同様の用語のフレーズの様相は、必須ではないが、同一実施例を意味する。

【0028】

信号担持媒体 (signal bearing medium) については、信号を生成可能なあらゆる形態、信号が生成可能なあらゆる形態、または、デジタル処理装置の機械可読指令からなるプログラムを実行可能なあらゆる形態をとることができる。信号担持媒体は、伝送線路、コンパクトディスク、デジタルビデオディスク、磁気テープ、ベルヌーイドライブ、磁気ディスク、パンチカード、フラッシュメモリ、集積回路または他のデジタル処理装置のメモリデバイスに組み込まれてもよい。

【0029】

さらに、本発明の記載された特徴部、構造または特性は、1又はそれ以上の実施例の適宜な方法で組み合わせることができる。以下の説明において、多数の具体的詳細事項は、例えば、プログラミング、ソフトウェアモジュール、ユーザセクション、ネットワーク処理、データベースクエリ、データベース構造、ハードウェアモジュール、ハードウェア回路、ハードウェアチップなどに設けられることで、本発明の実施例を完全に理解されたい。しかしながら、当分野の当業者であれば、本発明が1又はそれ以上の具体的詳細事項なしで実施でき、あるいは、他の方法、コンポーネント、材料などを用いて実施できることを理解されたい。別の例においては、本発明の側面をあいまいになるのを避けるために、周知な構造、材料またはオペレーションを示していないか、または、詳細に記載していない。

【0030】

本明細書に含まれる概略的フローチャートは、論理フローチャート図として一般的に記載されている。このように、記載された順番およびラベルされたステップは、示される方法の一実施例を示す。他のステップ及び方法は、例示された方法の関数、論理または1又はそれ以上のステップに対する効果、またはその部分において同等であると考えられる。さらに、使用したフォーマットおよび記号は、方法の論理ステップを説明するために設けたものであり、この方法の範囲を限定するものではないと理解されたい。各種矢印および線がフローチャート図で使用されているが、対応する方法の範囲を限定するものではないことを理解されたい。実際に、いくつかの矢印または他のコネクタは、この方法の論理フローのみを示すために使用されている。例えば、矢印は、記載された方法の列挙されたステップ間における待機または不特定時間のモニタを示す。さらに、特定の方法が生じる順番は、示された対応するステップの順番に厳密に固執しなくてもよい。

【0031】

ソリッドステートストレージシステム

図1Aは、本発明によるソリッドステートストレージデバイスのデータ管理用のシステム100の一実施例を示した概略ブロック図である。システム100は、以下に示すように、ソリッドステートストレージデバイス102、ソリッドステートストレージ制御部104、ライトデータパイプライン106、リードデータパイプライン108、ソリッドステートストレージ110、コンピュータ112、クライアント114およびコンピュータネットワーク116を具える。

【0032】

システム100は、少なくとも1のソリッドステートストレージデバイス102を具える。別の実施例においては、システム100は、2又はそれ以上のソリッドステートストレージ102を具える。各ソリッドステートストレージデバイス102は、フラッシュメモリ、ナノランダムアクセスメモリ(「ナノRAMまたはNRAM」)、磁気抵抗RAM(「MRAM」)、ダイナミックRAM(「DRAM」)、フェーズチェンジRAM(「PRAM」)など、不揮発性ソリッドステートストレージ110を具える。このソリッドステートストレージデバイス102は、図2および図3で更に詳細に示す。ソリッドステートストレージデバイス102は、コンピュータネットワーク116を介して、クライアント114に接続したコンピュータ112内に示されている。一実施例においては、ソリッドステートストレージデバイス102は、コンピュータ112の内部にあり、周辺コンポーネントインターコネクタエクスプレス(「PCI-e」)バス、シリアルアドバンステクノロジーアタッチメント(「シリアルATA」)バスなど、システムバスを用いて接続されている。別の実施例においては、ソリッドステートストレージデバイス102は、コンピュータ112の外部にあり、ユニバーサルシリアルバス(「USB」)コネクション、米国電気電子技術者協会(「IEEE」)1394バス(「FireWire」)などを用いて接続される。その他の実施例においては、ソリッドステートストレージデバイス102は、InfinitibandまたはPCIエクスプレスアドバンススイッチング(「PCIe-AS」)など、外部の電気または光学バスエクステンション、または、バスネットワークングソリューション、を使用し、周辺コンポーネントインターコネクタ(「

PCI」) エクスプレスバスを用いたコンピュータ 112 に接続される。

【0033】

各種実施例において、ソリッドステートストレージ 102 は、デュアルインラインメモリモジュール(「DIMM」)、ドーターカードまたはマイクロモジュールの形態でもよい。別の実施例においては、ソリッドステートストレージデバイス 102 は、ラック取付ブレード内のエレメントである。別の実施例においては、ソリッドステートストレージデバイス 102 は、より高いレベルのアセンブリ(例えば、マザーボード、ラップトップ、画像処理装置)に直接統合したパッケージ内に含まれる。別の実施例においては、ソリッドステートストレージデバイス 102 を具える個々のコンポーネントは、中間パッケージなしで、より高いレベルのアセンブリに直接統合される。

【0034】

ソリッドステートストレージデバイス 102 は、1 又はそれ以上のソリッドステートストレージ制御部 104 を具え、各々は、ライトデータパイプライン 106 およびリードデータパイプラインを具え、各々は、ソリッドステートストレージ 110 を具え、これらは、図 2 および図 3 で以下に詳細に示す。

【0035】

システム 100 は、ソリッドステートストレージデバイス 102 に接続した 1 又はそれ以上のコンピュータ 112 を具える。コンピュータ 112 は、ホスト、サーバ、ストレージエリアネットワーク(「SAN」)のストレージ制御部、ワークステーション、パーソナルコンピュータ、ラップトップコンピュータ、手持ち型コンピュータ、スーパーコンピュータ、コンピュータクラスタ、ネットワークスイッチ、ルータまたはアプライアンス、データベースまたはストレージアプライアンス、データ取得システムまたはデータ獲得システム、診断システム、試験システム、ロボット、携帯型電子機器、ワイヤレスデバイスなどでもよい。別の実施例においては、コンピュータ 112 は、クライアントでもよく、ソリッドステートストレージデバイス 102 は、コンピュータ 112 から送られるデータ要求に自動で応答するように機能する。この実施例においては、コンピュータ 112 およびソリッドステートストレージデバイス 102 は、コンピュータネットワーク、システムバス、または、コンピュータ 112 と自立性ソリッドステートストレージデバイス 102 との間の接続に適切なその他の通信手段を用いて、接続可能である。

【0036】

一実施例においては、システム 100 は、1 又はそれ以上のコンピュータネットワーク 116 を介して、1 又はそれ以上のコンピュータ 112 に接続された 1 又はそれ以上のクライアント 114 を具える。クライアント 114 は、ホスト、サーバ、SANのストレージ制御部、ワークステーション、パーソナルコンピュータ、ラップトップコンピュータ、手持ち型コンピュータ、スーパーコンピュータ、コンピュータクラスタ、ネットワークスイッチ、ルータまたはアプライアンス、データベースまたはストレージアプライアンス、データ取得システムまたはデータ獲得システム、診断システム、試験システム、ロボット、携帯電子機器、ワイヤレスデバイスなどでもよい。コンピュータネットワーク 116 は、インターネット、ワイドエリアネットワーク(「WAN」)、メトロポリタンエリアネットワーク(「MAN」)、ローカルエリアネットワーク(「LAN」)、トークンリング、ワイヤレスネットワーク、ファイバチャネルネットワーク、SAN、ネットワーク接続ストレージ(「NAS」)、ESCON、または、ネットワークの組み合わせを具える。さらに、コンピュータネットワーク 116 は、Ethernet(登録商標)、トークンリング、Wi-Fi、WiMax など、ネットワーク技術の IEEE 802 ファミリのネットワークを含む。

【0037】

コンピュータネットワーク 116 は、サーバ、スイッチ、ルータ、ケーブル、ラジオ、および、コンピュータ 112 とクライアント 114 とのネットワーク構築を容易にするのに使用するその他の設備を具える。一実施例においては、システム 100 は、コンピュータネットワーク 116 を介して、同等に(peer to peer)通信する複数のコンピュー

タ 1 1 2 を具える。別の実施例においては、システム 1 0 0 は、コンピュータ 1 1 6 を介して同等に通信する複数のソリッドステートストレージデバイス 1 0 2 を具える。当分野の当業者であれば、1 又はそれ以上のコンピュータネットワーク 1 1 6 を具える他のコンピュータネットワーク 1 1 6 と、1 またはそれ以上のクライアント 1 1 4 間の単一または冗長接続を有する関連設備、または、1 又はそれ以上のソリッドステートストレージデバイス 1 0 2 を有する他のコンピュータ、または、1 又はそれ以上のコンピュータ 1 1 2 に接続された 1 又はそれ以上のソリッドステートストレージデバイス 1 0 2 と、を理解されたい。一実施例においては、システム 1 0 0 は、コンピュータ 1 1 2 なしで、コンピュータネットワーク 1 1 6 を介して、クライアント 1 1 4 に接続した 2 又はそれ以上のソリッドステートストレージデバイス 1 0 2 を具える。

【0038】

ストレージ制御部管理オブジェクト

図 1 B は、本発明によるストレージデバイスのオブジェクト管理用システム 1 0 1 の一実施例を示す概略ブロック図である。このシステム 1 0 1 は、ストレージ制御部 1 5 2 および 1 又はそれ以上のデータストレージデバイス 1 5 4 を含む 1 又はそれ以上のストレージデバイス 1 5 0 と、1 又はそれ以上の要求 (r e q u e s t i n g) デバイス 1 5 5 と、を有する。これらのストレージデバイス 1 5 0 は、互いにネットワークで接続され、1 又はそれ以上の要求デバイス 1 5 5 に結合される。要求デバイス 1 5 5 は、ストレージデバイス 1 5 0 a にオブジェクト要求を送る。オブジェクト要求は、オブジェクトを構築する要求、オブジェクトにデータを書き込む要求、オブジェクトからデータを読み取る要求、オブジェクトを削除する要求、オブジェクトをチェック (c h e c k p o i n t) する要求、オブジェクトをコピーする要求などである。当分野の当業者であれば、その他のオブジェクト要求も理解されたい。

【0039】

一実施例においては、ストレージ制御部 1 5 2 およびデータストレージデバイス 1 5 4 は別のデバイスである。別の実施例においては、ストレージ制御部 1 5 2 およびデータストレージデバイス 1 5 4 は、1 つのストレージデバイス 1 5 0 に統合される。別の実施例においては、データストレージデバイス 1 5 4 は、ソリッドステートストレージ 1 1 0 であり、ストレージ制御部 1 5 2 は、ソリッドステートストレージデバイス制御部 2 0 2 である。他の実施例においては、データストレージデバイス 1 5 4 は、ハードディスクドライブ、光学ドライブ、テープストレージなどでもよい。別の実施例においては、ストレージデバイス 1 5 0 は、異なる種類の 2 又はそれ以上のデータストレージデバイス 1 5 4 を具える。

【0040】

一実施例においては、データストレージデバイス 1 5 4 は、ソリッドステートストレージ 1 1 0 であり、ソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 のアレイとして構成させる。別の実施例においては、ソリッドステートストレージ 1 1 0 は 2 又はそれ以上のバンク 2 1 4 a - n において構成される。ソリッドステートストレージ 1 1 0 は、以下の図 2 B でより詳細に示す。

【0041】

これらのストレージデバイス 1 5 0 a - n は、互いにネットワークで接続され、分散したストレージデバイスとして機能する。要求デバイス 1 5 5 に接続されたストレージデバイス 1 5 0 a は、分散したストレージデバイスへのオブジェクト要求を制御する。一実施例においては、ストレージデバイス 1 5 0 および関連したストレージ制御部 1 5 2 は、オブジェクトを管理し、分散したオブジェクトファイルシステムとして要求デバイス 1 5 5 に存在する。この文脈において、並行オブジェクトファイルシステムは、分散したオブジェクトファイルシステム型の例である。別の実施例においては、ストレージデバイス 1 5 0 および関連するストレージ制御部 1 5 2 は、オブジェクトを管理し、分散したオブジェクトファイルサーバとして 1 又はそれ以上の要求デバイス 1 5 5 に存在する。この文脈において、並行オブジェクトファイルサーバは、分散したオブジェクトファイルサーバ型の

例である。これらの実施例およびその他の実施例において、要求デバイス 155 は、オブジェクトを排他的に管理するか、または、ストレージデバイス 150 と組み合わせたオブジェクトの管理に加わる；これは、通常、ストレージデバイス 150 の能力を制限せず、他のクライアント 114 のためのオブジェクトを完全に管理する。場合によっては、分散したストレージデバイスの各々、分散したオブジェクトファイルシステムおよび分散したオブジェクトファイルサーバは、1つのデバイスとして独立して機能することができる。ネットワーク接続されたストレージデバイス 150 a - n は、分散したストレージデバイス、分散したオブジェクトファイルシステム、分散したオブジェクトファイルサーバ、および、これらの組み合わせであって、1又はそれ以上の要求デバイス 155 に対して設定された1又はそれ以上の上記能力のイメージを有する組み合わせとして、機能できる。例えば、ストレージデバイス 150 は、第1の要求デバイス 155 a に対して分散したストレージデバイスとして機能するように設定でき、一方、要求デバイス 155 b に対して分散したストレージデバイスおよび分散したオブジェクトファイルシステムとして機能する。システム 101 が、1つのストレージデバイス 150 a を具える場合、ストレージデバイス 150 a のストレージ制御部 152 a は、オブジェクトを管理し、オブジェクトファイルシステムまたはオブジェクトファイルサーバとして要求デバイス 155 に存在してもよい。

【0042】

ストレージデバイス 150 が、分散したストレージデバイスとして互いにネットワークで接続される一実施例において、ストレージデバイス 150 は、1又はそれ以上の分散したストレージ制御部 152 によって管理された独立ドライブ冗長アレイ（「RAID」）として機能する。例えば、オブジェクトのデータセグメントを書き込む要求は、RAID レベルに応じて、パリティストライプを用いてデータストレージデバイス 154 a - n に亘ってストライプ化されたデータセグメントとなる。このような構成の利点の1つは、オブジェクト管理システムが、ストレージ制御部 152、データストレージデバイスまたはストレージデバイス 150 のその他のコンポーネントのいずれかのうち、1つのストレージデバイス 150 にエラーが生じたときに、利用可能なように継続できることである。

【0043】

冗長ネットワークを使用してストレージデバイス 150 と要求デバイス 155 を相互接続するとき、オブジェクト管理システムは、ネットワークの1つが機能している限り、ネットワークエラーの存在下で利用可能となるように続けることができる。1つのストレージデバイス 150 a を有するシステム 101 は、さらに、複数のデータストレージデバイス 154 a を具えることができ、ストレージデバイス 150 a のストレージ制御部 152 a は、RAID 制御部として機能し、ストレージデバイス 150 a のデータストレージデバイス 154 a に亘ってデータセグメントをストライプ化し、RAID レベルに応じて、パリティストライプを具えることができる。

【0044】

一実施例においては、1又はそれ以上のストレージデバイス 150 a - n は、ソリッドステートストレージデバイス制御部 202 およびソリッドステートストレージ 110 を有するソリッドステートストレージデバイス 102 であり、ソリッドステートストレージデバイス 102 は、DIMM 設定、ドーターカード、マイクロモジュールなどで構成され、コンピュータ 112 内にあってもよい。コンピュータ 112 は、サーバでもよく、または、互いにネットワーク接続されたソリッドステートストレージデバイス 102 を有し、分散した RAID 制御部として機能する同様のデバイスでもよい。有利には、ストレージデバイス 102 は、PCI-e、PCIe-AS、InfiniBand またはその他の高性能バス、スイッチ付きバス、ネットワーク接続バスまたはネットワークを用いて接続されてもよく、ソリッドステートストレージ 110 a - n に亘ってデータセグメントを自律的にストライプ化する1つ又は分散したソリッドステートストレージ制御部 202 を有する極めて小型で、高性能 RAID ストレージシステムを提供することができる。

【0045】

一実施例においては、ストレージデバイス 150 と通信するために要求デバイス 155 によって使用される同一ネットワークは、同等のストレージデバイス 150 b - n と通信し R A I D 機能を得るために、同等のストレージデバイス 150 a によって使用される。別の実施例においては、別のネットワークは、R A I D するために、ストレージデバイス 150 巻で使用される。別の実施例においては、要求デバイス 155 は、ストレージデバイス 150 に冗長要求を送ることによって R A I D プロセスに加わることができる。例えば、要求デバイス 155 は、第 1 のオブジェクトライト要求を第 1 のストレージデバイス 150 a に送り、同一データを有する第 2 のオブジェクトライト要求を第 2 のストレージデバイス 150 b に送り、単純なミラーリングを行う。

【0046】

ストレージデバイス 102 内でオブジェクト処理するための能力に関して、ストレージ制御部 152 は、1 つの R A I D レベルを用いて、1 つのデータセグメントまたはオブジェクトを記憶する能力を個別に具え、一方、別のデータセグメントまたはオブジェクトは、異なる R A I D レベルを用いて、または、R A I D ストライプ化せずに、記憶される。これらの複数の R A I D グループ分けは、ストレージデバイス 150 内の複数のパーティションに関連していてもよい。R A I D 0、R A I D 1、R A I D 5、R A I D 6 およびコンボジット R A I D タイプ 10、50、60 は、データストレージデバイス 154 a - n を具える各種 R A I D グループに亘って同時にサポート可能である。当分野の当業者であれば、同時にサポート可能なその他の R A I D タイプおよび設定も理解されたい。

【0047】

さらに、ストレージ制御部 152 は、R A I D 制御部として自律的に機能するので、R A I D 制御部は、累進的 R A I D を行うことができ、要求デバイス 155 に影響を与えず、R A I D レベルに変化を加えず、または、検出さえせずに、1 つの R A I D レベルから別の R A I D レベルへと、データストレージデバイス 154 に亘ってストライプ化したオブジェクトまたはオブジェクトの一部をトランスフォーム可能である。好適な実施例においては、1 つのレベルから別のレベルへと R A I D 設定を前進 (p r o g r e s s) させることは、オブジェクトまたはパケットベースでも自律的に行われ、ストレージデバイス 150 の 1 つまたはストレージ制御部 152 において機能する分散した R A I D 制御モジュールによって起動される。通常、R A I D の前進は、R A I D 1 などのより高性能でより低効率なストレージ設定から、R A I D 5 などのより低性能でより高いストレージ効率へと前進であり、トランスフォームは、アクセス頻度に基づいて動的に開始される。なお、R A I D 5 から R A I D 1 へ設定を前進させることも可能であることを理解されたい。R A I D 前進を開始するその他のプロセスは、ストレージシステム管理サーバ要求など、クライアントまたは外部エージェントから設定または要求できる。当分野の当業者であれば、自律的にオブジェクトを管理するストレージ制御部 152 を用いたストレージデバイス 102 の特徴および利点を理解されたい。

【0048】

ストレージ制御部管理オブジェクト用装置

図 2 A は、本発明によるストレージデバイス内のオブジェクト管理装置 200 の実施例を例示する概略的ブロック図である。この装置 200 は、オブジェクト要求レシーバモジュール 260、パーズィング (p a r s i n g) モジュール 262、コマンド実行モジュール 264、オブジェクトインデックスモジュール 266、オブジェクト要求キューイングモジュール 268、メッセージモジュール 270 を有するパケットタイザ (p a c k e t i z e r) 302、および、オブジェクトインデックス再構築モジュール 272 を含むストレージ制御部 152 を具え、これらを以下に示す。

【0049】

ストレージ制御部 152 は、図 1 B のシステム 101 に関連して記載されたストレージ制御部 152 と実質的に同様で、図 2 と関連して記載されたソリッドステートストレージデバイス制御部 202 であってもよい。この装置 200 は、1 又はそれ以上の要求デバイス 155 からオブジェクト要求を受信するオブジェクト要求レシーバモジュール 260 を

具える。例えば、記憶オブジェクトデータ要求に関して、ストレージ制御部 152 は、ストレージ制御部 152 に接続したデータストレージデバイス 154 内のデータパケットとしてデータセグメントを記憶する。オブジェクト要求は、一般的に、1 又はそれ以上のオブジェクトデータパケットに記憶された、または、記憶されるべきデータセグメントで、ストレージ制御部 152 によって管理されたオブジェクトに指令される。ストレージ制御部 152 が、ローカルまたはリモートダイレクトメモリアクセス（「DMA」、「RDMA」）転送を使用できる後期オブジェクト要求を介したデータで、後で充填されるオブジェクトを形成することを、オブジェクト要求は要求してもよい。

【0050】

一実施例においては、オブジェクト要求は、既に形成されたオブジェクトにオブジェクトの全てまたは一部を書き込む、書き込み要求である。一例においては、書き込み要求は、オブジェクトのデータセグメント用である。オブジェクトのその他のデータセグメントは、ストレージデバイス 150 またはその他のストレージデバイスに書き込まれてもよい。別の例においては、書き込み要求は、完全なオブジェクト用である。別の例においては、オブジェクト要求は、ストレージ制御部 152 によって管理されたデータセグメントからのデータを読み取ることである。さらなる別の実施例においては、オブジェクト要求は、データセグメントを削除する削除要求である。

【0051】

有利には、ストレージ制御部 152 は、新規オブジェクトを書き込むか、または、データを既存のオブジェクトに加える以上の書き込み要求をアクセプト可能である。例えば、オブジェクト要求レシーバモジュール 260 が受信する書き込み要求は、ストレージ制御部 152 によって記録されるデータの先のデータを加える要求、記憶したデータにデータを挿入する要求またはデータのセグメントを置換する要求を具えることができる。ストレージ制御部 152 によって維持されるオブジェクトインデックスは、その他のストレージ制御部で利用可能ではないが、サーバや他のコンピュータのフィルシステムのストレージ制御部の外側でのみ現在利用可能である、上記のような複雑な書き込みオペレーションに必要なフレキシビリティを提供する。

【0052】

装置 200 は、オブジェクト要求を 1 又はそれ以上のコマンドにパーズするパーズモジュール 262 を具える。一般的に、パーズモジュール 262 は、オブジェクト要求を 1 又はそれ以上のバッファにパーズする。例えば、オブジェクト要求の 1 又はそれ以上のコマンドは、コマンドバッファにパーズできる。通常、パーズモジュール 262 は、オブジェクト要求を調整し、オブジェクト要求の情報は、ストレージ制御部 152 によって理解され実行される。当分野の当業者であれば、1 又はそれ以上のコマンドにオブジェクト要求をパーズするパーズモジュール 262 のその他の機能も理解されたい。

【0053】

装置 200 は、オブジェクト要求からパーズされたコマンドを実行するコマンド実行モジュール 264 を具える。一実施例においては、コマンド実行モジュール 264 は、1 つのコマンドを実行する。別の実施例においては、コマンド実行モジュール 264 は、複数のコマンドを実行する。通常、コマンド実行モジュール 264 は、書き込みコマンドなど、オブジェクト要求からパーズされたコマンドを解釈し、次いで、サブコマンドを形成して、キューイングして（queue）、実行する。例えば、オブジェクトからパーズされた書き込みコマンドは、ストレージ制御部 152 に指令し、複数のデータセグメントを記憶する。さらに、オブジェクト要求は、暗号化、圧縮など必要な特性を具えていてもよい。コマンド実行モジュール 264 は、ストレージ制御部 152 に指令して、データセグメントを圧縮し、データセグメントを暗号化し、1 又はそれ以上のデータパケットおよび各データパケットの関連するヘッダを作成し、メディア暗号化キーを有するデータパケットを暗号化し、エラー修正コードを加え、そして、特定のロケーションにデータパケットを記憶することができる。特定のロケーションおよびその他のサブコマンドの

データパケットを記憶することは、その他のより低レベルのサブコマンドにブレークダウンされる。当分野の当業者であれば、コマンド実行モジュール 264 が、オブジェクト要求からパーシングされる 1 又はそれ以上のコマンドを実行可能であるその他の方法を理解されたい。

【0054】

装置 200 は、オブジェクトインデックスモジュール 266 を具え、このモジュール 266 は、オブジェクトを作成するか、または、オブジェクトのデータセグメントを記憶するストレージ制御部 152 に対応して、オブジェクトインデックスのオブジェクト項目を作成する。一般的に、ストレージ制御部 152 は、データセグメントと、記憶されたデータパケットが、データセグメントが記憶されるときに、割り当てられたロケーションと、からデータパケットを作成する。データセグメント内で受信したオブジェクトメタデータまたはオブジェクト要求の一部としてのオブジェクトメタデータは、同様に記憶できる。

【0055】

オブジェクトインデックスモジュール 266 は、データパケットが記憶され、データパケットの物理的アドレスが割り当てられるときに、オブジェクトインデックスにオブジェクト項目を作成する。オブジェクト項目は、オブジェクトのロジック識別子と、ストレージ制御部 152 が 1 又はそれ以上のデータパケットおよびあらゆるオブジェクトメタデータパケットを記憶したところに関連する 1 又はそれ以上の物理的アドレスと、の間におけるマッピングを具える。別の実施例においては、オブジェクトインデックスの項目は、オブジェクトのデータパケットが記憶される前に作成される。例えば、データパケットが早期に記憶される物理的アドレスをストレージ制御部 152 が決定する場合、オブジェクトインデックスモジュール 266 は、オブジェクトインデックスの項目を早期に作成できる。

【0056】

一般的には、オブジェクト要求またはオブジェクト要求群が、読み取り - 修正 - 書き込みオペレーション中、修正されたオブジェクトまたはデータセグメントになるとき、オブジェクトインデックスモジュール 266 は、修正したオブジェクトに関連するオブジェクトインデックス内の項目をアップデートする。一実施例においては、オブジェクトインデックスは、修正したオブジェクト用のオブジェクトインデックス中に、新規オブジェクトおよび新規項目を作成する。一般的に、オブジェクトの一部のみが修正され場合、オブジェクトは、修正されたデータパケットおよび未変化のいくつかのデータパケットを具える。この場合、新規項目は、もともと書き込まれていた未変化のデータパケットと、新規ロケーションに書き込まれた修正されたオブジェクトと、に対するマッピングを具える。

【0057】

別の実施例においては、オブジェクト要求レシーバモジュール 260 は、データブロックまたはその他のオブジェクトエレメントを消去するコマンドを具えたオブジェクト要求を受信し、ストレージ制御部 152 は、オブジェクトに対する基準、オブジェクトの関連性、および、消去したデータブロックのサイズなどの情報を含む消去パケットなど、少なくとも 1 つのパケットを記憶することができる。さらに、それは、さらに、消去したオブジェクトエレメントがゼロで充填されていることを示してもよい。従って、消去オブジェクト要求は、実際のメモリまたは消去されるストレージをエミュレートするのに使用可能であり、メモリまたはストレージは、実際に、メモリ / ストレージのセル中でゼロを用いて実際に記憶された適宜なメモリ / ストレージの一部を有している。

【0058】

有利には、データセグメントとオブジェクトのメタデータとの間のマッピングを示す項目を含んだオブジェクトインデックスを作成することで、ストレージ制御部 152 が、オブジェクトを自律的に処理して管理することが可能になる。この能力によって、ストレージデバイス 150 内にデータを記憶する大きなフレキシビリティをもたせることができる。オブジェクトのインデックス項目が作成されると、続いて、オブジェクトに関するオブジェクト要求を、ストレージ制御部 152 が効率的に応答可能である。

【 0 0 5 9 】

一実施例においては、ストレージ制御部 1 5 2 は、パーズングモジュール 2 6 2 によってパーズングする前に、オブジェクト要求レシーバモジュール 2 6 0 によって受信される 1 又はそれ以上のオブジェクト要求をキューイングするオブジェクト要求キューイングモジュール 2 6 8 を具える。オブジェクト要求キューイングモジュール 2 6 8 によって、オブジェクト要求が受信されるときと、オブジェクト要求が実行されるときとの間においてフレキシビリティをもたせることができる。

【 0 0 6 0 】

別の実施例においては、ストレージ制御部 1 5 2 は、1 又はそれ以上のデータセグメントから 1 又はそれ以上のデータパケットを形成するパケットタイザ (p a c k e t i z e r) 3 0 2 を具え、ここで、データパケットがデータストレージデバイス 1 5 4 のストレージに合うようにサイズ調整される。パケットタイザ 3 0 2 は、以下の図 3 でより詳細に示す。パケットタイザ 3 0 2 は、一実施例においては、各パケットのヘッダを作成するメッセージモジュール 2 7 0 を具える。このヘッダは、パケット識別子とパケット長を具える。パケット識別子は、パケットが形成されるためのオブジェクトにパケットに関連させる。

【 0 0 6 1 】

一実施例においては、各パケットは、自納式のパケット識別子を具え、ここで、パケット識別子は、オブジェクトと、パケット内に含まれるオブジェクトエレメントのオブジェクト内の関係性を識別するのに十分な情報を含んでいる。しかしながら、より効率的な好適な実施例は、コンテナにパケットを記憶することができる。

【 0 0 6 2 】

コンテナは、データを具え、このデータは、より効率的なパケットストレージを容易にし、オブジェクトとデータパケット、メタデータパケット、コンテナ内に記憶されるオブジェクトに関連したその他のパケットと、の関連性を構築するのを補助する。ストレージ制御部 1 5 2 は、通常、同様の方法で、オブジェクトおよびデータセグメントの一部として受信されるオブジェクトメタデータを処理することに留意されたい。一般的に、「パケット」は、データ、メタデータを具えるメタデータパケット、または、別のパケットタイプのパケットを意味する。オブジェクトは、1 又はそれ以上のコンテナに記憶可能で、コンテナは、一般的に、1 の固有オブジェクトのみに対するパケットを具える。オブジェクトは、複数のコンテナ間に配置可能である。一般的に、コンテナは、1 の論理消去ブロック (ストレージディビジョン) 内に記憶され、論理消去ブロック間でスプリットされることはない。

【 0 0 6 3 】

一例においては、コンテナは、2 又はそれ以上の論理 / パーチャルページ間でスプリットされてもよい。コンテナは、オブジェクト内のコンテナに関連するコンテナラベルによって識別される。コンテナは、多くのパケットに対してゼロを含み、一般的に、コンテナ内のパケットは、1 つのオブジェクトからなる。パケットは、オブジェクト属性エレメント、オブジェクトデータエレメント、オブジェクトインデックスエレメントなど、多数のオブジェクトエレメントタイプでもよい。2 以上のオブジェクトエレメントタイプを含むハイブリッドパケットを作成できる。各パケットは、同一エレメントタイプの多数のエレメントに対してゼロを含む。コンテナ内の各パケットは、一般的に、オブジェクトとの関連性を識別する固有の識別子を含む。

【 0 0 6 4 】

各パケットは、1 のコンテナに関連している。好適な実施例においては、コンテナは、消去ブロックに限定されず、各消去ブロックの開始点またはその付近で、コンテナパケットを見つけることができる。このことは、破損したパケットヘッダを有する消去ブロックに対するデータ損失を制限することを支援する。この実施例においては、オブジェクトインデックスが使用不能であり、消去ブロック内のパケットヘッダが破損している場合、パケットヘッダから消去ブロックエンドまでの内容は、連続するパケットのロケーションを決定する信頼性のある機構がない可能性が高いので、損失してしまう。別の実施例におい

ては、より信頼性のある方法は、ページバウンダリに限定されるコンテナを有することである。この実施例は、より多くのヘッダオーバーヘッドを必要とする。別の実施例においては、コンテナは、ページを横切りフローして、ブロックバウンダリを消去することができる。これは、必要とするヘッダオーバーヘッドは少なくなるが、パケットヘッダが破損している場合、データのより大きな部分が損失される。これらのいくつかの実施例に関して、いくつかのタイプの R A I D が、さらなるデータの完全性を保証するために使用されることが予測される。

【 0 0 6 5 】

一実施例においては、装置 2 0 0 は、データストレージデバイス 1 5 4 に記憶されるパケットヘッダからの情報を使用して、オブジェクトインデックスの項目を再構築するオブジェクトインデックス再構築モジュール 2 7 2 を具える。一実施例においては、オブジェクトインデックス再構築モジュール 2 7 2 は、各パケットが属するオブジェクトを判定するヘッダと、オブジェクトにおけるデータまたはメタデータがどこに属するかを判定するシークエンス情報とを読み込むことによって、オブジェクトインデックスの項目を再構築する。オブジェクトインデックス再構築モジュール 2 7 2 は、各パケットおよびタイムスタンプの物理アドレス情報、または、パケットの物理ロケーションとオブジェクト識別子とデータセグメントシークエンスとの間のマッピングを作成するシークエンス情報を使用する。タイムスタンプまたはシークエンス情報をオブジェクトインデックス再構築モジュール 2 7 2 が使用して、インデックスに対して行われる変化シークエンスをリプレイし、これによって、通常、最新状態を再構築 (r e e s t a b l i s h) する。

【 0 0 6 6 】

別の実施例においては、オブジェクトインデックス再構築モジュール 2 7 2 は、パケットの物理ロケーションと、オブジェクト識別子と、オブジェクトインデックス内に項目を再構築する各パケットのシークエンスナンバと、を識別する、コンテナパケット情報とともにパケットヘッダ情報を使用して、パケットを配置する。一実施例においては、消去ブロックは、タイムスタンプされるか、または、パケットが書き込まれときにシークエンスナンバが付与され、タイムスタンプまたは消去ブロックのシークエンス情報は、コンテナヘッダおよびパケットヘッダから集められた情報とともに使用され、オブジェクトインデックスを再構築する。別の実施例においては、消去ブロックがリカバリされるときに、タイムスタンプまたはシークエンス情報が書き込まれる。

【 0 0 6 7 】

オブジェクトインデックスが揮発性メモリに記憶されるとき、エラー、電力損失、または、ストレージ制御部 1 5 2 がオブジェクトインデックスを保存せずにシャットダウンする問題が、オブジェクトインデックスが再構築できない場合に、問題になるであろう。オブジェクトインデックス再構築モジュール 2 7 2 によって、オブジェクトインデックスが揮発性メモリに記憶され、これにより迅速なアクセスなど揮発性メモリの利点を得る。オブジェクトインデックス再構築モジュール 2 7 2 によって、ストレージデバイス 1 5 0 の外部にあるデバイスに依存せず自律的に、オブジェクトインデックスの迅速な再構築が可能となる。

【 0 0 6 8 】

一実施例においては、揮発性メモリのオブジェクトインデックスは、データストレージデバイス 1 5 4 に一時的に記憶される。特定の実施例においては、オブジェクトインデックスまたは「インデックスメタデータ」は、ソリッドステートストレージ 1 1 0 に一時的に記憶される。別の実施例においては、インデックスメタデータが、パケットを記憶するソリッドステートストレージ 1 1 0 a - 1 1 0 n - 1 から分離したソリッドステートストレージ 1 1 0 n に記憶される。インデックスメタデータは、データおよび要求デバイス 1 5 5 から伝達されるオブジェクトメタデータから独立的に管理され、ストレージ制御部 1 5 2 / ソリッドステートストレージデバイス制御部 2 0 2 によって管理される。オブジェクトからの他のデータおよびメタデータから分離したインデックスを管理して記憶することによって、オブジェクトメタデータを処理する必要なく、ストレージ制御部 1 5 2 / ソ

リッドステートストレージデバイス制御部 202 なしで、効率的データフローが可能となる。

【0069】

一実施例においては、オブジェクト要求レシーバモジュール 260 によって受信されるオブジェクト要求が書き込み要求を具える場合、ストレージ制御部 152 は、ローカルまたはリモートダイレクトメモリアクセス（「DMA」、「RDMA」）オペレーションとして、要求デバイス 155 のメモリからオブジェクトの 1 又はそれ以上のデータセグメントを受信される。好適な例において、ストレージ制御部 152 は、1 又はそれ以上の DMA または RDMA オペレーションにおいて要求デバイス 155 のメモリからデータを引き出す（pull）。別の例においては、要求デバイス 155 は、1 又はそれ以上の DMA または RDMA オペレーションにおいて、データセグメントをストレージ制御部 152 へプッシュする。別の実施例においては、オブジェクト要求は読み取り要求を具え、ストレージ制御部 152 は、1 又はそれ以上の DMA または RDMA オペレーションにおいて、オブジェクトの 1 又はそれ以上のデータセグメントを要求デバイス 155 のメモリに伝達する。好適な例においては、ストレージ制御部 152 は、1 又はそれ以上の DMA または RDMA オペレーションにおいて、データを要求デバイス 155 のメモリにプッシュする。別の実施例においては、要求デバイス 155 は、1 又はそれ以上の DMA または RDMA オペレーションにおいて、ストレージ制御部 152 からデータを引き出す。別の例においては、ストレージ制御部 152 は、1 又はそれ以上の DMA または RDMA オペレーションにおいて、要求デバイス 155 のメモリからオブジェクトコマンド要求セットを引き出す。別の実施例においては、要求デバイス 155 は、1 又はそれ以上の DMA または RDMA オペレーションにおいて、ストレージ制御部 152 にオブジェクトコマンド要求セットをプッシュする。

【0070】

一実施例においては、ストレージ制御部 152 は、ブロックストレージをエミュレートし、要求デバイス 155 とストレージ制御部 152 との間で通信されるオブジェクトは、1 又はそれ以上のデータブロックを具える。一実施例においては、要求デバイス 155 は、ドライバを具え、ストレージデバイス 150 はブロックストレージデバイスとして存在する。例えば、要求デバイス 155 は、要求デバイス 155 が記憶されるデータブロックを必要とする物理的アドレスとともに所定サイズのデータブロックを送ることができる。ストレージ制御部 152 は、データブロックを受信し、データブロックを伝達される物理的ブロックアドレスを使用するか、または、オブジェクト識別子としての物理的ブロックアドレスの変換（transformation）を使用する。次いで、ストレージ制御部 152 は、データブロックをパケット化してデータブロックを自在に記憶することで、オブジェクトまたはオブジェクトのデータセグメントとしてデータブロックを記憶する。次いで、オブジェクトインデックスモジュール 266 は、物理的なブロックベースのオブジェクト識別子と、ストレージ制御部 152 がデータブロックのデータを具えるデータパケットを記憶した実際の物理ロケーションとを使用して、オブジェクト内に項目を作成する。

【0071】

別の実施例においては、ストレージ制御部 152 は、ブロックオブジェクトをアクセプトすることでブロックストレージをエミュレートする。ブロックオブジェクトは、ブロック構造中に 1 又はそれ以上のデータブロックを具えることができる。一実施例においては、ストレージ制御部 152 は、その他のオブジェクトとしてブロックオブジェクトを処理する。別の実施例においては、オブジェクトは、完全なブロックデバイス、ブロックデバイスのパーティション、又は、トラック、セクタ、チャネルなどを含むブロックデバイスの他のロジック若しくは物理的サブエレメントを表す。特筆すべきことは、ブロックデバイス RAID 群を、プログレッシブ RAID などの異なる RAID 構造をサポートするオブジェクトに再マッピング（remap）する能力である。当分野の当業者であれば、従来または未来のブロックデバイスをオブジェクトへマッピングするその他のマッピングも

理解されたい。

【0072】

ソリッドステートストレージデバイス

図2Bは、本発明によるソリッドステートデバイス102内に書き込みデータパイプライン106および読み取りデータパイプライン108を具えるソリッドステートデバイス制御部202の一実施例201を示す概略的ブロック図である。ソリッドステートストレージデバイス制御部202は、各々がソリッドステートストレージ110を制御する多数のソリッドステートストレージ制御部0-N104a-nを具えることができる。描いた実施例においては、2つのソリッドステート制御部を示す：ソリッドステート制御部0104aおよびソリッドステートストレージ制御部N104nは、それぞれ、ソリッドステートストレージ110a-nを制御する。描いた実施例においては、ソリッドステートストレージ制御部0104aは、データチャネルを制御し、接続したソリッドステートストレージ110aはデータを記憶する。ソリッドステートストレージ制御部N104nは、記憶したデータと関連するインデックスメタデータチャネルを制御し、関連するソリッドステートストレージ110nはインデックスメタデータを記憶する。代替の実施例においては、ソリッドステートストレージデバイス制御部202は、1つのソリッドステートストレージ110aを有する1つのソリッドステート制御部104aを具える。別の実施例においては、複数のソリッドステートストレージ制御部104a-nおよび関連するソリッドステートストレージ110a-nがある。一実施例においては、1又はそれ以上のソリッドステート制御部104a-104n-1に関連するソリッドステートストレージ110a-110n-1に接続した制御部104a-104n-1は、データを制御し、ソリッドステートストレージ制御部104nに関連するソリッドステートストレージ110nに接続した少なくとも1つの制御部104nは、インデックスメタデータを制御する。

【0073】

一実施例においては、少なくとも1つのソリッドステート制御部104は、フィールドプログラマブルゲートアレイ（「FPGA」）であり、制御部機能は、FPGAにプログラムされる。特定の実施例においては、FPGAはXilinx（登録商標）FPGAである。別の実施例においては、ソリッドステートストレージ制御部104は、アプリケーション特異的集積回路（「ASIC」）またはカスタムロジックソリューションとして特別に設計されたコンポーネントを具える。各ソリッドステートストレージ制御部104は、通常、書き込みデータパイプライン106および読み取りデータパイプライン108を具えており、これらは、図3でさらに説明する。別の実施例においては、少なくとも1つのソリッドステートストレージ制御部104は、FPGA、ASICおよびカスタムロジックコンポーネントの組み合わせから構成される。

【0074】

ソリッドステートストレージ

ソリッドステートストレージ110は、バンク214内で構成され、2方向性ストレージ入力/出力（「I/O」）バス210を介して並行してアクセスできる、非揮発性ソリッドステートストレージエレメント216、218、220のアレイである。一実施例においては、ストレージI/Oバス210は、一回につき、一方向性通信が可能である。例えば、データがソリッドステートストレージ110に書き込まれるとき、データは、ソリッドステートストレージ110から読み取ることはできない。別の実施例においては、データは、両方で同時にフローすることができる。なお、データバスに関して本明細書で使用されるように、2方向性が意味するのは、データバスウェイは、一回につき一方向のみしかデータフローできないが、2方向性データバス上で一方向のデータフローが停止したときは、データは、2方向性データバス上の反対方向にデータをフローすることができることである。

【0075】

ソリッドステートストレージエレメント（例えば、SSS 0.0 216a）は、通

常、回路基板上のチップ（１又はそれ以上のチップのパッケージ）またはダイとして構成される。示されているように、例えば、以下の複数のエレメントが、チップパッケージ、チップパッケージスタックまたはいくつかのその他のパッケージエレメントと一緒にパッケージされている場合であっても、ソリッドステートストレージエレメント（例えば、２１６a）は、他のソリッドステートストレージエレメント（例えば、２１８a）から独立または半独立的に機能する。示されているように、ソリッドステートストレージ２１６、２１８、２２０のカラムは、バンク２１４として指定される。示されているように、ソリッドステートストレージ１１０内の $n \times m$ のソリッドステートストレージエレメント２１６、２１８、２２０のアレイにおいて、バンク当たり「 n 」バンク２１４a - n および「 m 」ソリッドステートストレージエレメント２１６a - m 、２１８a - m 、２２０a - m が存在していてもよい。一実施例においては、ソリッドステートストレージ１１０aは、８つのバンク２１４を有するバンク２１４あたり２０のソリッドステートストレージエレメント２１６、２１８、２２０を具え、ソリッドステートストレージ１１０nは、１つのバンク２１４を有するバンク２１４あたり、２つのソリッドステートストレージエレメント２１６、２１８を具える。一実施例においては、各ソリッドステートストレージエレメント２１６、２１８、２２０は、１つのレベルのセル（「SLC」）デバイスからなっている。別の実施例においては、各ソリッドステートストレージエレメント２１６、２１８、２２０は、複数レベルのセル（「MLC」）デバイスからなっている。

【００７６】

一実施例においては、共通ストレージＩ／Ｏバス２１０a列（例えば、２１６b、２１８b、２２０b）を共有する複数のバンクのソリッドステートストレージエレメントは、一緒にパッケージされる。一実施例においては、ソリッドステートストレージエレメント２１６、２１８、２２０は、チップ当たり１又はそれ以上のダイを有することができ、１又はそれ以上のチップは、垂直方向にスタックされ、各ダイは独立してアクセス可能である。別の実施例においては、ソリッドステートストレージエレメント（例えば、SSS 0.0 216a）は、ダイあたり１又はそれ以上のバーチャルダイと、チップ当たり１又はそれ以上のダイと、垂直方向にスタックされた１又はそれ以上のチップと、を有し、各バーチャルダイは独立してアクセス可能である。別の実施例においては、ソリッドステートストレージエレメントSSS 0.0 216aは、ダイあたり１又はそれ以上のバーチャルダイと、１又はそれ以上のいくつかまたは全部が垂直方向にスタックされたチップあたり１又はそれ以上のダイと、を有し、各バーチャルダイは、独立してアクセス可能である。

【００７７】

一実施例においては、２つのダイが、グループ当たり４つのスタックとともに垂直方向にスタックされ、８つのストレージエレメント（例えば、SSS 0.0 - SSS 0.8）216a - 220aを形成し、各々は、別のバンク214a - nにある。別の実施例においては、２０のストレージエレメント（例えば、SSS 0.0 - SSS 20.0）216は、バーチャルバンク214aを形成し、８のバーチャルバンクの各々は、２０のストレージエレメント（例えば、SSS 0.0 - SSS 20.8）216、218、220を有する。データは、特定のグループのストレージエレメント（SSS 0.0 - SSS 0.8）216a、218a、220aの全てのストレージエレメントに対して、ストレージＩ／Ｏバス210を介して、ソリッドステートストレージ１１０に送られる。ストレージ制御バス212aは、特定のバンク（例えば、バンク - 0 214a）を選択するために使用され、全てのバンク214に接続されたストレージＩ／Ｏバス210を介して受信するデータは、選択されたバンク214aに書き込まれる。

【００７８】

好適な実施例においては、ストレージＩ／Ｏバス210は、１又はそれ以上の独立Ｉ／Ｏバスからなり（210a. a - m、210n. a - mを具える「IIOba - m」）、各列内のソリッドステートストレージエレメントは、各ソリッドステートストレージエレメント216、218、220に並行してアクセスする独立Ｉ／Ｏバスの１つを共有する

ので、全てのバンク 2 1 4 は、同時にアクセスされる。例えば、ストレージ I / O バス 2 1 0 の 1 つのチャンネルは、各バンク 2 1 4 a - n の第 1 のソリッドステートストレージエレメント 2 1 6 a、2 1 8 a、2 2 0 a に同時にアクセスすることができる。ストレージ I / O バス 2 1 0 の第 2 のチャンネルは、各バンク 2 1 4 a - n の第 2 のソリッドステートストレージエレメント 2 1 6 b、2 1 8 b、2 2 0 b に同時にアクセスできる。ソリッドステートストレージ 2 1 6、2 1 8、2 2 0 の各列は、同時にアクセスされる。一実施例においては、ソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 が複数レベル（物理的にスタックされている）ある場合、ソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 の全ての物理的レベルは、同時にアクセスされる。本明細書で使用されるように、「同時」は、ほぼ同時のアクセスを含み、デバイスは、スイッチングノイズを回避するために、わずかに異なる間隔でアクセスされる。「同時」は、このような文脈で使用され、シーケンシャルまたは連続（*serial*）アクセスとは分け、コマンドおよび / またはデータは、交互に独立して送られる。

【0079】

一般的に、バンク 2 1 4 a - n は、ストレージ制御バス 2 1 2 を用いて独立して選択される。一実施例においては、バンク 2 1 4 は、使用可能チップまたは選択チップを用いて選択される。選択チップおよび使用可能チップの両方が使用可能な場合、ストレージ制御バス 2 1 2 は、複数レベルのソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 の 1 つのレベルを選択することができる。その他の実施例においては、ストレージ制御バス 2 1 2 によって他のコマンドを使用して、複数レベルのソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 の 1 つのレベルを独立して選択する。ソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 は、ストレージ I / O バス 2 1 0 およびストレージ制御バス 2 1 2 上で伝達される制御情報とアドレス情報との組み合わせを介して選択されてもよい。

【0080】

一実施例においては、各ソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 は、消去ブロックにパーティション化され、各消去ブロックは、ページにパーティション化される。一般的なページは、2 0 0 0 バイト（「2 k B」）である。一例においては、ソリッドステートストレージエレメント（例えば、SSS0.0）は、2 つのレジスタを具え、2 つのページをプログラムすることができ、2 つのレジスタソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 は、4 k B の容量を有する。さらに、2 0 のソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 のバンク 2 1 4 は、ストレージ I / O バス 2 1 0 のチャンネルを通る同一アドレスでアクセスされる 8 0 k B 容量のページを有することが可能である。

【0081】

8 0 k B のソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 のバンク 2 1 4 における上記グループのページは、バーチャルページと呼ばれる。同様にバンク 2 1 4 a の各ストレージエレメント 2 1 6 a - m の消去ブロックは、バーチャル消去ブロックを形成するようにグループ化されてもよい。好適な実施例においては、ソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 内の消去ブロックページは、消去コマンドをソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 内で受信したときに消去される。消去ブロック、ページ、プレーンまたはソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 内のその他のロジックおよび物理的ディビジョンのサイズおよび数が、技術の進歩に伴い、所定時間に亘って変化することが予想されるが、新規構成に一致する多数の実施例が、可能であり、本明細書の一般的説明と一致すると理解されたい。

【0082】

一般的に、パケットは、ソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 内の特定ロケーションに書き込まれるとき、このパケットは、特定バンクの特定エレメントの特定消去ブロックに特異的である特定ページ内のロケーションに書き込まれるよう

に意図されており、物理的アドレスは、ストレージ I/O バス 210 上で送られ、パケットがそれに続く。物理的アドレスは、パケットをページ内の指定ロケーションへと指令するための、ソリッドステートストレージエレメント 216、218、220 の情報を十分に保有している。ストレージエレメント（例えば、SSS 0.0 - SSS 0.N 216a、218a、220a）の列における全てのストレージエレメントは、ストレージ I/O バス 210 a . a 内の適宜なバスによって同時にアクセスされ、適宜なページに到達し、ストレージエレメント（SSS 0.0 - SSS 0.N 216a、218a、220a）の列において同様にアドレス化されたページに、データパケットを書き込むことを回避するので、データパケットが書き込まれるべき正規のページとともにソリッドステートストレージエレメント SSS 0.0 216a を具えるバンク 214a は、ストレージ制御バス 212 によって同時に選択される。

【0083】

同様に、ストレージ I/O バス 210 上を伝達される読み取りコマンドは、1つのバンク 214a と、このバンク 214a 内の適宜なページとを選択するために、ストレージ制御バス 212 上の同時コマンドを必要とする。好適な実施例においては、読み取りコマンドは、ページ全体を読み取り、さらに、バンク 214 において、並行して複数のソリッドステートストレージエレメント 216、218、220 があるので、バーチャルページ全体は、読み取りコマンドを用いて読み取られる。なお、読み取りコマンドは、バンクインターリーブに関して以下で説明するように、サブコマンドにブレイク可能である。さらに、バーチャルページは、書き込みオペレーションでアクセス可能である。

【0084】

消去ブロック消去コマンドは、特定の消去ブロックを消去する特定の消去ブロックアドレスを用いて、ストレージ I/O バス 210 を介して、消去ブロックを消去するように送信される。一般的に、消去ブロック消去コマンドは、バーチャル消去ブロックを消去するために、ストレージ I/O バス 210 の並行バスを介して送信可能であり、特定の消去ブロックアドレスを有する各々が、特定の消去ブロックを消去する。同時に、特定バンク（例えば、バンク - 0 214a）がストレージ制御バス 212 を介して選択され、全てのバンク（バンク 1 - N 214b - n）における同様にアドレス化されたブロックが消去されるのを防止する。さらに、その他のコマンドは、ストレージ I/O バス 210 とストレージ制御バス 212 との組み合わせを利用して特定ロケーションに送信可能である。当分野の当業者であれば、2方向性ストレージ I/O バス 210 およびストレージ制御バス 212 を用いて、特定のストレージロケーションを選択できるその他の方法を理解されたい。

【0085】

一実施例においては、パケットは、ソリッドステートストレージ 110 にシーケンシャルに書き込まれる。例えば、パケットは、ストレージエレメント 216 のバンク 214a のストレージ書き込みバッファへとストリームされ、バッファがフルの場合、パケットは、指定のバーチャルページにプログラムされる。次いで、パケットは、ストレージ書き込みバッファを補充（refill）し、フルの場合、パケットは、次のバーチャルページに書き込まれる。次のバーチャルページは、同一バンク 214a または他のバンク（例えば、214b）内にあってもよい。このプロセスは、一般的に、バーチャル消去ブロックが満たされるまで、バーチャルページごとに繰り返される。別の実施例においては、ストリーミングは、バーチャル消去ブロックバウンダリに亘って繰り返され、このプロセスは、バーチャル消去ブロックごとに繰り返される。

【0086】

読み取り、修正、書き込みオペレーションにおいて、オブジェクト関連データパケットは、配置され、読み取りオペレーションで読み取られる。修正されたオブジェクトのデータセグメントは、データセグメントが読み取られるロケーションに書き込まれない。その代わりに、修正されたデータセグメントは、データパケットに再度変換され、次いで、そのとき書き込まれているバーチャルページの次の利用可能ロケーションに書き込まれる。そ

れぞれのデータパケットのオブジェクトインデックス項目は、修正されたデータセグメントを含むパケットのポイントに修正される。修正されていない同一オブジェクトに関連したデータパケットのオブジェクトインデックスの1又はそれ以上の項目は、未修正データパケットのもともとのロケーションに対するポイントを具える。従って、例えば、オブジェクトの前バージョンを維持するために、もともとのオブジェクトを維持している場合、もともとのオブジェクトは、もともと書き込まれていた全てのデータパケットに対するオブジェクトインデックスのポイントを有する。この新規オブジェクトは、もともとのデータパケットに対するオブジェクトインデックスのポイントと、その時に書き込まれているバーチャルページにおける修正データパケットに対するポイントとを有する。

【0087】

コピーオペレーションにおいて、オブジェクトインデックスは、ソリッドステートストレージ110に記憶された多数のパケットにマップされたもともとのオブジェクトの項目を具える。コピーが行われるとき、新規オブジェクトが作成され、新規項目が、もともとのパケットに新規オブジェクトをマッピングするオブジェクトインデックスに作成される。新規オブジェクトは、ソリッドステートストレージ110に書き込まれ、そのロケーションは、オブジェクトインデックスの新規項目にマッピングされる。新規オブジェクトパケットは、コピーに伝達されていないもともとのオブジェクトで変更が生じた場合またはオブジェクトインデックスが損失または破損した場合に、参照されるもともとのオブジェクト内のパケットを識別するために使用できる。

【0088】

有益なことに、パケットをシーケンシャルに書き込むことは、ソリッドステートストレージ110のより均等な利用を容易にし、ソリッドストレージデバイス制御部202がソリッドステートストレージ110内のストレージホットスポットおよび各種バーチャルページの利用レベルをモニタすることを可能にする。さらに、パケットをシーケンシャルに書き込むことは、下記で詳細に示すように、強力で有効な不要データ(garbage)コレクションシステムを容易にする。当分野の当業者であれば、データパケットのシーケンシャルストレージのその他の利点を理解されたい。

【0089】

ソリッドステートストレージデバイス制御部

各種実施例においては、ソリッドステートストレージデバイス制御部202は、データバス204、ローカルバス206、バッファ制御部208、バッファ0-N 222a-n、マスタ制御部224、ダイレクトメモリアクセス(「DMA」)制御部226、メモリ制御部228、ダイナミックメモリアレイ230、静的ランダムメモリアレイ232、管理制御部234、管理バス236、システムバス240へのブリッジ238、および、種々雑多なロジック242も具え、以下にこれを示す。他の実施例においては、システムバス240は、1又はそれ以上のネットワークインターフェースカード(「NICs」)244に結合し、そのうちのいくつかは、リモートDMA(「RDMA」)制御部246、1又はそれ以上の中央処理装置(CPU)248、1又はそれ以上の外部メモリ制御部250および関連する外部メモリアレイ252、1又はそれ以上のストレージ制御部254、ピア(peer)制御部256およびアプリケーション特異的处理装置258を具え、これらは以下で説明する。システムバス240に接続されたコンポーネント244-258は、コンピュータ112に配置されてもよく、その他のデバイスであってもよい。

【0090】

一般的に、ソリッドステートストレージ制御部104は、ストレージI/Oバス210を介してソリッドステートストレージ110とデータを通信する。ソリッドステートストレージが、バンク214内に配列され、各々のバンク214が、同時にアクセスされる複数のストレージエレメント216、218、220を具える一般的な実施例においては、ストレージI/Oバス210は、バスのアレイであり、ストレージエレメント216、218、220の各列の1つは、バンク214に亘って存在している。本明細書で使用されるように、用語「ストレージI/Oバス」は、1つのストレージI/Oバス210、また

は、バス 204 に独立したデータアレイを意味する。好適な実施例においては、ストレージエレメントの列（例えば、216a、218a、220a）にアクセスする各ストレージ I/O バス 210 は、ストレージエレメントの列 216a、218a、220a にアクセスされるストレージディビジョン（例えば、消去ブロック）のロジック - 物理マッピングを具えることができる。このマッピングによって、ストレージディビジョンの物理アドレスにマッピングされる論理アドレスが、第 1 のストレージディビジョンが機能していない場合、部分的に機能していない場合、アクセス不能な場合、または、その他何らかの問題がある場合に、異なるストレージディビジョンに再マップ可能である。再マップは、図 4A および 4B の再マップモジュール 430 に関連してさらに説明する。

【0091】

データは、システムバス 240、ブリッジ 238、ローカルバス 206、バッファ 222 を介して、また、最終的にデータバス 204 を介して、要求デバイス 155 からソリッドステートストレージ制御部 104 と通信可能である。データバス 204 は、一般的に、バッファ制御部 208 を用いて制御される 1 又はそれ以上のバッファ 222a - n に接続される。バッファ制御部 208 は、通常、ローカルバス 206 からバッファ 222 へ、データバス 204 からパイプラインインプットバッファ 306 およびアウトプットバッファ 330 を介して、データの転送を制御する。一般的に、バッファ制御部 208 は、どのように、要求デバイス 155 からきたデータがバッファ 222 に一時的に記憶可能で、次いで、データバス 204 上を転送されるかを制御でき、あるいは、その逆も同様に制御でき、異なるクロックドメインを考慮し、データ不一致 (collision) を防ぐことができる。バッファ制御部 208 は、通常、マスタ制御部 224 とともに作動し、データフローに調和させる。データがくるときは、データは、システムバス 240 の上に到達し、ブリッジ 238 を介してローカルバス 206 に転送される。

【0092】

一般的に、データは、マスタ制御部 224 およびバッファ制御部 208 によって指令されるように、ローカルバス 206 から 1 又はそれ以上のデータバッファ 222 へと転送される。次いで、データは、バッファ 222 からデータバス 204 へと、ソリッドステート制御部 104 を介して、NAND フラッシュまたはその他のストレージメディアなどソリッドステートストレージ 110 へとフローする。好適な実施例においては、データと、このデータとともにくる関連したアウトオブバンドメタデータ（「オブジェクトメタデータ」）は、1 又はそれ以上のソリッドステートストレージ制御部 104a - 104n - 1 および関連するソリッドステートストレージ 110a - 110n - 1 を具える 1 又はそれ以上のデータチャンネルを用いて通信され、一方、少なくとも 1 つのチャンネル（ソリッドステートストレージ制御部 104n、ソリッドステートストレージ 110n）は、インデックス情報や、ソリッドステートストレージデバイス 102 に内部生成されるその他のメタデータなど、インバンドメタデータ専用である。

【0093】

通常、ローカルバス 206 は、2 方向性バスであるか、または、ソリッドステートストレージデバイス制御部 202 の内部のデバイス間、および、ソリッドステートストレージデバイス 102 の内部デバイスと、システムバス 240 に接続したデバイス 244 - 258 との間において、データおよびコマンドの通信を可能にするバスのセットである。ブリッジ 238 は、ローカルバス 206 とシステムバス 240 との間の通信を容易にする。当分野の当業者であれば、リング構造またはスイッチ式スター設定、ならびに、バス 240、206、204、210 およびブリッジ 238 の機能など、その他の実施例を理解されたい。

【0094】

通常、システムバス 240 は、コンピュータ 112 のバスであり、または、ソリッドステートストレージデバイス 102 がインストールまたは接続されているその他のデバイスである。一実施例においては、システムバス 240 は、PCI - e バス、シリアルアドバンステクノロジーアタッチメント（「シリアル ATA」）バス、パラレル ATA などでも

よい。別の実施例においては、システムバス 240 は、スモールコンピュータシステムインターフェース（「SCSI」）、FireWire、ファイバーチャネル、USB、PCIe - AS など、外部バスである。ソリッドステートストレージデバイス 102 は、デバイスの内側に適用するように、または、外側に接続されたデバイスとして、パッケージされてもよい。

【0095】

ソリッドステートストレージデバイス制御部 202 は、ソリッドステートストレージ 102 内のより高レベルの機能を制御するマスタ制御部 224 を具える。マスタ制御部 224 は、各種実施例において、オブジェクト要求およびその他の要求を解釈することによってデータフローを制御し、インデックスの作成を指令し、データに関連したオブジェクト識別子を、関連するデータの物理ロケーションにマップして、DMA 要求と調和する本明細書に記載される機能の多くは、マスタ制御部 224 によって完全に、部分的に制御される。

【0096】

一実施例においては、マスタ制御部 224 は、内蔵型制御部を使用する。別の実施例においては、マスタ制御部 224 は、動的メモリアレイ 230（動的ランダムアクセスメモリ「DRAM」）、静的メモリアレイ 232（静的ランダムアクセスメモリ「SRAM」）など、ローカルメモリを使用する。一実施例においては、ローカルメモリは、マスタ制御部 224 を用いて制御される。別の実施例においては、マスタ制御部 224 は、メモリ制御部 228 を介してローカルメモリにアクセスする。別の実施例においては、マスタ制御部 224 は、Linux サーバを運用し、World Wide Web、ハイパーテキストマークアップ言語（「HTML」）など、各種共通サーバインターフェースをサポートすることができる。別の実施例においては、マスタ制御部 224 は、ナノプロセッサを使用する。マスタ制御部 224 は、プログラマブル若しくは標準ロジック、または、上記に挙げた制御部タイプの組み合わせ、を使用して構成されていてもよい。当分野の当業者であれば、マスタ制御部 224 の多数の実施例を理解されたい。

【0097】

一実施例においては、ストレージ制御部 152 / ソリッドステートストレージデバイス制御部 202 は、複数のデータストレージデバイス / ソリッドステートストレージ 110 a - n を管理し、マスタ制御部 224 は、ソリッドステートストレージ制御部 104 a - n など、内部制御部中のワークロードを分割する。例えば、マスタ制御部 224 は、データストレージデバイス（例えば、ソリッドステートストレージ 110 a - n）に書き込まれるオブジェクトを分割することができ、オブジェクトの一部は、接続したデータストレージデバイスの各々に記憶される。この特徴は、オブジェクトのより迅速なストレージおよびアクセスを可能にする性能向上である。一実施例においては、マスタ制御部 224 は、FPGA を使用して実装される。別の実施例においては、マスタ制御部 224 内のファームウェアは、管理バス 236、NIC 244 に接続したネットワークを介したシステムバス 240、または、システムバス 240 に接続したその他のデバイスを介してアップデート可能である。

【0098】

一実施例においては、オブジェクトを管理するマスタ制御部 224 は、ブロックストレージをエミュレートし、コンピュータ 112、または、ストレージデバイス / ソリッドステートストレージ 102 に接続したその他のデバイスは、ブロックストレージデバイスとしてストレージデバイス / ソリッドステートストレージデバイス 102 を観察（view）し、ストレージデバイス / ソリッドステートストレージデバイス 102 内の特定の物理アドレスにデータを送る。次いで、マスタ制御部 224 は、ブロックを分割し、それをオブジェクトとしてデータブロックに記憶する。さらに、マスタ制御部 224 は、ブロックと、このブロックとともに送られる物理アドレスを、マスタ制御部 224 によって決定される実際のロケーションにマップする。このマッピングは、オブジェクトインデックスに記憶される。通常、ブロックエミュレーションに関し、ブロックデバイスアプリケーション

ンプログラムインターフェース(「API」)は、コンピュータ112、クライアント114、または、ブロックストレージデバイスとしてストレージデバイス/ソリッドステートストレージデバイス102を利用することを望むその他のデバイス内のドライバに設けられる。

【0099】

別の実施例においては、マスタ制御部224は、NIC制御部244および内蔵型RDMA制御部246を調整し、データおよびコマンドセットをジャストインタイムRDMA転送する。NIC制御部244は、非透明ポートの背後に隠れ、カスタムドライバを使用することができる。さらに、クライアント114のドライバは、標準スタックAPIを使用し、I/Oメモリドライバを介してコンピュータネットワーク116にアクセスでき、NIC244とともに機能する。

【0100】

一実施例においては、マスタ制御部224も、独立ドライブ冗長アレイ(「RAID」)制御部である。データストレージデバイス/ソリッドステートストレージデバイス102が、1又はそれ以上の他のデータストレージデバイス/ソリッドステートストレージデバイス102とネットワーク接続される場合、マスタ制御部224は、1段(tier)RAID、複数段RAID、プログレッシブRAID用のRAID制御部でもよい。さらに、マスタ制御部224によって、いくつかのオブジェクトはRAIDアレイに記憶でき、他のオブジェクトはRAID以外に記憶できる。別の実施例においては、マスタ制御部224は、分散型RAID制御部エレメントでもよい。別の実施例においては、マスタ制御部224は、多数のRAID、分散型RAID、本明細書に記載されるその他の機能を実現することができる。

【0101】

一実施例においては、マスタ制御部224は、単一または冗長ネットワークマネージャ(例えば、スイッチ)を調整して、ルーティングを構築し、バンド幅使用(bandwidth utilization)、フェイルオーバー(failover)のバランスをとる。別の実施例においては、マスタ制御部224は、統合型アプリケーション特異的ロジック(ローカルバス206を介して)と、関連するドライバソフトウェアを調整する。別の実施例においては、マスタ制御部224は、接続されたアプリケーション特異的ロジック258またはロジック(外部システムバス240)および関連するドライバソフトウェアを調整する。別の実施例においては、マスタ制御部224は、リモートアプリケーション特異的ロジック(コンピュータネットワーク116を介して)および関連するドライバソフトウェアを調整する。別の実施例においては、マスタ制御部224は、ローカルバス206または外部バス接続ハードディスクドライブ(「HDD」)ストレージ制御部を調整する。

【0102】

一実施例においては、マスタ制御部224は、1又はそれ以上のストレージ制御部254と通信し、ストレージデバイス/ソリッドステートストレージデバイス102は、SCSIバス、インターネットSCSIバス、インターネットSCSI(「iSCSI」)、ファイバチャネルを介して接続されたストレージデバイスとして存在してもよい。その間、ストレージデバイス/ソリッドステートストレージデバイス102は、オブジェクトを自律的に管理することができ、オブジェクトファイルシステムまたは分散型オブジェクトファイルシステムとして存在してもよい。マスタ制御部224は、ピア制御部256および/またはアプリケーション特異的プロセッサ258によってアクセスすることができる。

【0103】

別の実施例においては、マスタ制御部224は、自律的に統合された管理制御部を調整し、FPGAコードおよび/または制御部ソフトウェアを定期的に確認し、実行中(リセット)FPGAコードを確認し、および/または、電源が入っている間(リセット)制御部ソフトウェアを確認し、外部リセット要求をサポートし、ウォッチドッグタイムアウト

によるリセット要求をサポートし、電圧、電流、電力、温度ならびにその他の環境測定値および中断閾値設定をサポートする。別の実施例においては、マスタ制御部 224 は、不要データコレクションを管理し、再利用のための消去ブロックをフリーにする。別の実施例においては、マスタ制御部 224 は、摩耗レベルを管理する。別の実施例においては、マスタ制御部 224 によって、データストレージデバイス/ソリッドステートストレージデバイス 102 が、複数のバーチャルデバイスにパーティション化され、パーティションベースメディアの暗号化が可能となる。さらなる別の実施例においては、マスタ制御部 224 は、有利な複数ビット ECC 修正部を有するソリッドステートストレージ制御部 104 をサポートする。当分野の当業者であれば、ストレージ制御部 152 において、または、より具体的には、ソリッドステートストレージデバイス 102 において、マスタ制御部 224 のその他の特性および機能を理解されたい。

【0104】

一実施例においては、ソリッドステートストレージデバイス制御部 202 は、動的ランダムメモリアレイ 230 および/または静的ランダムメモリアレイ 232 を制御するメモリ制御部 228 を具える。上述したように、メモリ制御部 228 は、独立しているか、または、マスタ制御部 224 と統合されている。通常、メモリ制御部 228 は、DRAM (動的ランダムメモリアレイ 230) および SRAM (静的ランダムメモリアレイ 232) など、いくつかのタイプの揮発メモリを制御する。他の例においては、メモリ制御部 228 は、電氣的商況可能プログラム可能読み取り専用メモリ (「EEPROM」) などのその他のメモリタイプを制御できる。他の実施例においては、メモリ制御部 228 は、2 又はそれ以上のタイプを制御し、メモリ制御部 228 は、1 以上の制御部を具えてもよい。一般的に、メモリ制御部 228 は、できるだけ多くの SRAM 232 が実行可能であるように、DRAM 230 によって制御し、SRAM 232 を補助する。

【0105】

一実施例においては、オブジェクトインデックスは、メモリ 230、232 に記憶され、次いで、ソリッドステートストレージ 110n のチャネルまたはその他の非揮発性メモリに定期的にオフロードされる。当分野の当業者であれば、動的メモリアレイ 230、メモリ制御部 228 および静的メモリアレイ 232 のその他の使用および設定を理解されたい。

【0106】

一実施例においては、ソリッドステートストレージデバイス制御部 202 は DMA 制御部 226 を具え、この DMA 制御部 226 は、ストレージデバイス/ソリッドステートストレージデバイス 102 および 1 又はそれ以上の外部メモリ制御部 250、関連する外部メモリアレイ 252 および CPU 248 間の DMA オペレーションを制御する。なお、外部メモリ制御部 250 および関連する外部メモリアレイ 252 は、ストレージデバイス/ソリッドステートストレージデバイス 102 の外部にあるので、外部と呼ぶ。さらに、DMA 制御部 226 も、NIC 244 を介した要求デバイスおよび関連する RDMA 制御部 246 を用いて RDMA オペレーションを制御することができる。DMA および RDMA は、以下でより詳細に説明する。

【0107】

一実施例においては、ソリッドステートストレージデバイス制御部 202 は、管理バス 236 に接続された管理制御部 234 を具える。通常、管理制御部 234 は、環境メトリクスおよびストレージデバイス/ソリッドステートストレージデバイス 102 のステータスを管理する。管理制御部 234 は、管理バス 236 を介してデバイス温度、ファンスピード、電力供給設定などをモニタすることができる。管理制御部 234 は、FPGA コードおよび制御部ソフトウェアのストレージに関して、消去可能プログラム可能読み取り専用メモリ (「EEPROM」) の読み取りおよびプログラムをサポートすることができる。通常、管理バス 236 は、ストレージデバイス/ソリッドステートストレージデバイス 102 内に各種コンポーネントを接続させる。管理制御部 234 は、ローカルバス 206 を介して、アラート、中断などと通信でき、システムバス 240 または他のバスに対して

別のコネクションを具えることができる。一実施例においては、管理バス236は、インターインテグレイテッド回路(Inter-Integrated Circuit:「I2C」)である。当分野の当業者であれば、管理バス236によってストレージデバイス/ソリッドステートストレージデバイス102のコンポーネントに接続された、管理制御部234のその他の関連する機能および利用を理解される。

【0108】

一実施例においては、ソリッドステートストレージデバイス制御部202は、特定のアプリケーションにカスタマイズ可能な種々雑多なロジック242を具える。通常、ソリッドステートデバイス制御部202またはマスタ制御部224が、FPGAまたはその他の設定可能制御部を用いて設定される場合、特定のアプリケーション、カスタマ要求、ストレージ要求に基づいて、カスタムロジックが含まれる。

【0109】

データパイプライン

図3は、本発明によるソリッドステートストレージデバイス102における書き込みデータパイプライン106および読み取りデータパイプライン108を有する、ソリッドステートストレージ制御部104の一実施例300を示した概略ブロック図である。実施例300は、データバス204、ローカルバス206およびバッファ制御208を具え、これらは、図2のソリッドステートストレージデバイス制御部202に関連して記載されたものと実質的に同等である。書き込みデータパイプライン106は、パケットタイザ302およびエラー修正コード(「ECC」)ジェネレータ304を具える。他の実施例においては、書き込みデータパイプライン106は、インプットバッファ306、書き込み同期バッファ308、書き込みプログラムモジュール310、圧縮モジュール312、暗号化モジュール314、不要データコレクタバイパス316(読み取りデータパイプライン108内の一部を有する)、メディア暗号化モジュール318、および、書き込みバッファ320を具える。読み取りデータパイプライン108は、読み取り同期バッファ328、ECC修正モジュール322、デパケットタイザ(depacketizer)324、アライメントモジュール326およびアウトプットバッファ330を具える。他の実施例においては、読み取りデータパイプライン108は、メディア解読モジュール332、不要データコレクタバイパス316の一部、解読モジュール334、展開(decompression)モジュール336および読み取りプログラムモジュール338を具えることができる。ソリッドステートストレージ制御部104は、さらに、制御およびステータスレジスタ340、制御キュー342、バンクインターリーブ制御部344、同期バッファ346、ストレージバス制御部348および多重チャネル(「MUX」)350を具えることができる。ソリッドステート制御部104のコンポーネントおよび関連する書き込みデータパイプライン106および読み取りデータパイプライン108は、以下で説明する。他の実施例においては、同期ソリッドステートストレージ110を使用することができ、同期バッファ308、328を省くことができる。

【0110】

書き込みデータパイプライン

書き込みデータパイプライン106は、直接または間接的に別の書き込みデータパイプライン106を介して、ソリッドステートストレージに書き込まれるデータ又はメタデータセグメントを受信する、パケットタイザ302を具え、ソリッドステートストレージ110用のサイズの1又はそれ以上のパケットを作成する。データまたはメタデータセグメントは、通常、オブジェクトの一部であるが、オブジェクト全体を具えることもできる。別の実施例においては、データセグメントは、データブロックの一部であるが、データブロック全体を具えることもできる。一般的に、オブジェクトは、コンピュータ112、クライアント114またはその他のコンピュータ若しくはデバイスから受信し、ソリッドステートストレージデバイス102またはコンピュータ112にストリーミングされるデータセグメントでソリッドステートストレージデバイス102に転送される。さらに、データセグメントは、データパーセルなど、別の名称として周知であるが、本明細書で規定され

るように、オブジェクトまたはデータブロックの全てまたは一部を具える。

【0111】

各オブジェクトは、1又はそれ以上のパケットとして記憶される。各オブジェクトは、1又はそれ以上のコンテナパケットを有することができる。各パケットはヘッダを含む。ヘッダは、ヘッダタイプフィールドを具えることができる。タイプフィールドは、データ、オブジェクト属性、メタデータ、データセグメントデリミタ（マルチパケット）、オブジェクトストラクチャ、オブジェクトリンケージなどを具えることができる。ヘッダは、パケットに含まれるデータのバイト数など、パケットサイズに関する情報を具えることができる。パケットの長さは、パケットタイプによって構築可能である。ヘッダは、オブジェクトパケットの関係性を構築する情報を具えることができる。一例では、オブジェクト内のデータセグメントの位置を識別するために、データパケットのオフセットを使用することができる。当分野の当業者であれば、パケットサイズ302によってデータに加えられヘッダに含まれるその他の情報、および、データパケットに加えられ他の情報を理解されたい。

【0112】

各パケットは、データまたはデータセグメントからのヘッダおよび可能であればデータを具えることができる。各パケットのヘッダは、パケットが属するオブジェクトにパケットに関連させる、関連情報を具える。例えば、ヘッダは、データパケットを形成していたデータセグメント、オブジェクトまたはデータブロックを示すオブジェクト識別子およびオフセットを具えることができる。さらに、ヘッダは、パケットを記憶するための、ストレージバス制御部によって使用される論理アドレスを具えることができる。また、ヘッダは、パケットに含まれるバイト数など、パケット寸法に関する情報を具えることができる。さらに、ヘッダは、データセグメントまたはオブジェクトを再構築するとき、データセグメントがオブジェクト内のその他のパケットに対して属しているところを識別するシーケンス番号を具えることができる。ヘッダは、ヘッダタイプフィールドを具えることができる。タイプフィールドは、データ、オブジェクト属性、メタデータ、データセグメントデリミタ（マルチパケット）、オブジェクトストラクチャ、オブジェクトリンケージなどを具えることができる。当分野当業者であれば、パケットサイズ302によってデータ又はメタデータに加えられヘッダに含まれてもよいその他の情報、および、パケットに加えることができる情報を理解されたい。

【0113】

書き込みデータパイプライン106は、パケットサイズ302から受信する1又はそれ以上のパケットに関する1又はそれ以上のエラー修正コード（「ECC」）を生成するECCジェネレータ304を具える。ECCジェネレータ304は、通常、エラー修正アルゴリズムを使用し、パケットとともに記憶されるECCを生成する。パケットとともに記憶されたECCは、通常、トランスミッションおよびストレージを介してデータに入ったエラーを検出し修正するために使用される。一実施例においては、パケットは、長さNの未コード化ブロックとしてECCジェネレータ304内にストリーミングされる。長さSのシンδροームが計算され、追加され、長さN+Sの符号化ブロックとして出力される。N値およびS値は、特別な性能、効率および強力（robustness）メトリクスを得るために選択されるアルゴリズムの特性に基づいている。好適な実施例においては、ECCブロックとパケットとの間において、固定された関係性はなく、パケットは、1以上のECCブロックを具えることができ、ECCブロックは、1以上のパケットを具えることができ、第1のパケットは、ECCブロック内のいずれかで終了（end）してもよく、第2のパケットは、同一のECCブロック内において第1のパケットのエンドの後から開始することができる。好適な実施例においては、ECCアルゴリズムは、動的に修正されない。好適な実施例においては、データパケットとともに記憶されたECCは、3以上のビットにおけるエラーを修正するのに十分に強力である。

【0114】

有利には、シングルビット以上の修正またはダブルビット修正でさえ可能な強力なECC

C アルゴリズムを用いることで、ソリッドステートストレージ 110 の寿命を延長させることができる。例えば、ソリッドステートストレージ 110 内のストレージ媒体としてフラッシュメモリが使用される場合、フラッシュメモリは、消去サイクルにつきエラーなしで、約 100,000 回書き込み可能である。この利用制限は、強力な ECC アルゴリズムを使用して拡張することができる。ECC ジェネレータ 304 を有し、ソリッドステートストレージデバイス 102 に搭載された ECC 修正モジュール 322 を対応させることで、ソリッドステートストレージデバイス 102 は、エラーを内部で修正することができ、シングルビット修正など、弱い ECC アルゴリズムを使用した場合よりも長い寿命を持つことができる。なお、その他の実施例においては、ECC ジェネレータ 304 は、より弱いアルゴリズムを使用することができ、シングルビットまたはダブルビットエラーを修正できる。別の実施例においては、ソリッドステートストレージデバイス 110 は、容量を増大するためにマルチレベルセル（「MLC」）フラッシュなど信頼性が低いストレージを具備してもよく、ストレージは、より強力な ECC アルゴリズムなしで、十分な信頼性がなくてもよい。

【0115】

一実施例においては、書き込みパイプライン 106 は、入力バッファ 306 を具備し、この入力バッファ 306 は、ソリッドステートストレージ 110 に書き込まれるデータセグメントを受信し、パケットサイズ 302（またはより複雑な書き込みデータパイプライン 106 用のその他のステージ）など、書き込みデータパイプライン 106 の次のステージが準備され、次のデータセグメントを処理されるまでに、入ってくるデータセグメントを記憶する。通常、入力バッファ 306 によって、受信したレートデータセグメントと、適宜なサイズのデータバッファを用いて書き込みデータパイプライン 106 によって処理されるデータセグメントとの間の不一致を許容することができる。さらに、入力バッファ 306 によって、データバス 204 のオペレーション効率を改善するために、データバス 204 は、書き込みデータパイプライン 106 によって維持できるよりも大きなレートで、データを書き込みデータパイプライン 106 に転送することができる。通常、書き込みデータパイプライン 106 は、入力バッファ 306 を具備していないときは、バッファリング機能は、書き込みパイプライン 106 の外側ではあるがソリッドステートストレージ 102 内、ネットワークインターフェースカード（「NIC」）内などコンピュータ 112 内、または、例えば、リモートダイレクトメモリアクセス（「RDAM」）を使用するときの別のデバイスなど、あらゆる場所で実行される。

【0116】

別の実施例においては、書き込みデータパイプライン 106 は、ソリッドステートストレージ 110 にパケットを書き込む前に、ECC ジェネレータ 304 から受信したパケットをバッファリングする書き込み同期バッファ 308 を具備することができる。書き込み同期バッファ 308 は、ローカルクロックドメインとソリッドステートストレージクロックドメインとの間のバウンダリに配置され、クロックドメイン差を考慮するためにバッファリングする。他の実施例においては、同期型ソリッドステートストレージ 110 を使用することができ、同期用バッファ 308、328 を省くことができる。

【0117】

一実施例においては、書き込みデータパイプライン 106 は、さらに、メディア暗号化モジュール 318 を具備することができ、このモジュール 318 は、パケットサイズ 302 から直接または間接的に、1 又はそれ以上のパケットを受信して、パケットを ECC ジェネレータ 304 にパケットを送る前に、ソリッドステートストレージデバイス 102 に固有の暗号化キーを用いて、1 又はそれ以上のパケットを暗号化する。通常、ヘッダを含む完全なパケットは暗号化される。別の実施例においては、ヘッダは、暗号化されない。本明細書においては、暗号化キーは、ソリッドステートストレージ 110 を統合する、暗号化保護を必要とする具体例の外部で管理されるシークレット暗号化キーを意味することを理解されたい。メディア暗号化モジュール 318 および関連するメディア暗号化モジュール 332 は、ソリッドステートストレージ 110 に記憶されるデータの安全レベルを提供す

る。例えば、データが、メディア暗号化モジュール 3 1 8 で暗号化されるとき、ソリッドステートストレージ 1 1 0 が、異なるソリッドステートストレージ制御部 1 0 4、ソリッドステートストレージデバイス 1 0 2 またはコンピュータ 1 1 2 に接続される場合、ソリッドステートストレージ 1 1 0 の内容は、通常、大きな労力を要せずに、ソリッドステートストレージ 1 1 0 にデータを書き込む間に使用される同一の暗号化キーを使用せずに読み取ることができない。

【0 1 1 8】

通常の実施例においては、ソリッドステートストレージデバイス 1 0 2 は、非揮発性ストレージの暗号化キーを記憶できず、暗号化キーへのいかなる外部からのアクセスもできない。暗号化キーは、初期渦中、ソリッドステートストレージ制御部 1 0 4 に設けられる。ソリッドステートストレージデバイス 1 0 2 は、暗号化キーとともに使用される、非シークレット暗号化ナンス (n o n c e) を使用および記憶することができない。異なるナンスは、パケット毎に記憶できる。データセグメントは、暗号化アルゴリズムによる保護を改善するために、独自のナンスを有する複数のパケット間にスプリットできる。暗号化キーは、クライアント 1 1 4、コンピュータ 1 1 2、キーマネージャ、または、ソリッドステートストレージ制御部 1 0 4 によって使用される暗号化キーを管理するその他のデバイスから受信することができる。別の実施例においては、ソリッドステートストレージ 1 1 0、は、2 又はそれ以上のパーティションを有し、ソリッドステートストレージ制御部 1 0 4 は、まるで 2 又はそれ以上のソリッドステートストレージ制御部 1 0 4 かのように動作し、各々は、ソリッドステートストレージ 1 1 0 内の 1 のパーティションで機能する。この実施例においては、独自のメディア暗号化キーは、各パーティションで使用できる。

【0 1 1 9】

別の実施例においては、書き込みデータパイプライン 1 0 6 は、さらに、パケットタイザ 3 0 2 にデータセグメントを送る前に、直接または間接的に、入力バッファ 3 0 6 から受信するデータまたはデータセグメントを暗号化する暗号化モジュール 3 1 4 を具えることができ、データセグメントは、データセグメントとともに受信した暗号化キーを用いて暗号化される。暗号化モジュール 3 1 4 は、メディア暗号化モジュール 3 1 8 とは異なり、このメディア暗号化モジュールにおいては、データを暗号化する暗号化モジュール 3 1 4 が使用する暗号化キーは、ソリッドステートストレージデバイス 1 0 2 内に記憶される全てのデータに共通でなくてもよいが、以下で記載するように、受信するデータセグメントとともに受信されるオブジェクトベースに基づいて変化してもよい。例えば、暗号化モジュール 3 1 4 によって暗号化されるべきデータセグメント用の暗号化キーは、データセグメントとともに受信され、データセグメントが属するオブジェクトを書き込むコマンドの一部として受信される。ソリッドステートストレージデバイス 1 0 2 は、暗号化キーとともに使用される各オブジェクトパケットにおいて、非シークレット暗号法 (c r y p t o g r a p h i c) ナンスを使用して記憶することができる。異なるナンスは、パケット毎に記憶できる。データセグメントは、暗号化アルゴリズムによる保護を改善する目的のために、独自のナンスを有する複数のパケット間にスプリットできる。一実施例においては、メディア暗号化モジュール 3 1 8 によって使用されるナンスは、暗号化モジュール 3 1 4 によって使用されるものと同一である。

【0 1 2 0】

暗号化キーは、クライアント 1 1 4、コンピュータ 1 1 2、キーマネージャ、または、データセグメントを暗号化するのに使用される暗号化キーを保持したその他デバイスから受信することができる。一実施例においては、ソリッドステートストレージデバイス 1 0 2、コンピュータ 1 1 2、クライアント 1 1 4、または、プライベートキーおよびパブリックキーを安全に転送して保護する産業上の標準的な方法を実行する能力を有するその他の外部エージェント、のうちの 1 つから、暗号化キーは、ソリッドステートストレージ制御部 1 0 4 に転送される。

【0 1 2 1】

一実施例においては、暗号化モジュール 3 1 4 は、第 1 のパケットを暗号化し、第 1 の暗号化キーはこのパケットとともに受信され、また、第 2 のパケットを暗号化し、第 2 の暗号化キーは第 2 のパケットとともに受信される。別の実施例においては、暗号化モジュール 3 1 4 は、第 1 のパケットを暗号化し、第 1 の暗号化キーは、このパケットとともに受信され、第 2 のデータパケットを暗号化せず次のステージにパスする。有利には、ソリッドステートストレージデバイス 1 0 2 の書き込みデータパイプライン 1 0 6 に含まれる暗号化モジュール 3 1 4 によって、オブジェクト毎またはセグメント毎のデータ暗号化が、1 のファイルシステムまたはその他のファイルシステムなしで、対応するオブジェクトまたはデータセグメントを記憶するのに使用される暗号化キーのトラックを、保持することが可能となる。各要求デバイス 1 5 5 または対応するキーマネージャは、要求デバイス 1 5 5 によって送られるオブジェクトまたはデータセグメントのみを暗号化するのに使用される暗号化キーを独立して管理することができる。

【0 1 2 2】

別の実施例においては、書き込みデータパイプライン 1 0 6 は、パケットタイザ 3 0 2 にデータセグメントを送る前に、メタデータセグメントに関するデータを圧縮する圧縮モジュール 3 1 2 を具える。圧縮モジュール 3 1 2 は、通常、セグメントのストレージサイズを低減するのに当分野で周知な圧縮ルーチンを用いて、データまたはメタデータセグメントを圧縮する。例えば、データセグメントは、5 1 2 のゼロからなるストリングを具える場合、圧縮モジュール 3 1 2 は、5 1 2 のゼロを、5 1 2 のゼロを示すコード又はトークンで置換でき、コードは、5 1 2 のゼロがとるスペースよりも大幅に小型である。

【0 1 2 3】

一実施例においては、圧縮モジュール 3 1 2 は、第 1 の圧縮ルーチンを用いて第 1 のセグメントを圧縮し、圧縮せずに第 2 のセグメントをパスする。別の実施例においては、圧縮モジュール 3 1 2 は、第 1 の圧縮ルーチンを用いて第 1 のセグメントを圧縮し、第 2 の圧縮ルーチンを用いて第 2 のセグメントを圧縮する。ソリッドステートストレージデバイス 1 0 2 内にフレキシビリティを有することに利点があり、クライアント 1 1 4、または、ソリッドステートストレージデバイス 1 0 2 にデータを書き込むその他のデバイスは圧縮ルーチンを各々特定でき、あるいは、一方が圧縮ルーチンを特定することができ、他方がいずれの圧縮ルーチンを特定しなくてもよい。圧縮ルーチンの選択は、オブジェクトごと (per object) のタイプまたはオブジェクトクラスベースに基づいたデフォルトの設定に従って選択されてもよい。例えば、特定のオブジェクトの第 1 のオブジェクトは、デフォルトの圧縮ルーチン設定を作動しないようにでき、同一オブジェクトクラスおよびオブジェクトタイプの第 2 のオブジェクトは、デフォルトの圧縮ルーチンを使用することができ、同一オブジェクトクラスおよびオブジェクトタイプの第 3 のオブジェクトは、圧縮を利用しなくてもよい。

【0 1 2 4】

一実施例においては、書き込みデータパイプライン 1 0 6 は、不要データコレクションシステムのデータバイパスの一部として、読み取りデータパイプライン 1 0 8 からデータセグメントを受信する不要データコレクションバイパス 3 1 6 を具える。不要コレクションシステムは、通常、パケットがデリション用に標識されるか、修正されるので、もはや有効でないパケットを標識し、修正されたデータが、異なるロケーションに記憶される。いくつかのポイントでは、不要データコレクションシステムは、ストレージの特定のセクションがリカバリ可能かを判定する。この判定は、入手可能ストレージ容量の不足、閾値に達した無効データの割合、有効なデータの統合、閾値に達したストレージセクションのエラー検出率、または、データ分散に基づいた性能の向上によってもよい。いくつかの因子は、不要データコレクションアルゴリズムによって考慮され、ストレージセクションをリカバリするときを判定する。

【0 1 2 5】

ストレージセクションがリカバリのために標識されると、通常、そのセクション中の有効なパケットは再配置されなければならない。不要データコレクタバイパス 3 1 6 によ

て、パケットは、読み取りデータパイプライン 108 に読み取られ、次いで、ソリッドステートストレージ制御部 104 からルートアウトされずに、書き込みデータパイプライン 106 に直接転送される。好ましい実施例においては、不要データコレクタバイパス 316 は、ソリッドステートストレージデバイス 102 内で機能する自律的不要データコレクタシステムの一部である。これによって、ソリッドステートストレージデバイス 102 は、データを管理することが可能となり、データは、ソリッドステートストレージ 110 に亘って系統的に分散され、性能、データ信頼性が改善し、ソリッドステートストレージ 110 のいずれか 1 のロケーションまたはエリアの過剰使用または使用不足を回避し、ソリッドステートストレージ 110 の寿命を延長させる。

【0126】

不要データコレクタバイパス 316 は、セグメントを書き込みデータパイプライン 106 に挿入し、その他のセグメントは、クライアント 114 またはその他のデバイスによって書き込まれるように調整する。描かれた実施例においては、不要データコレクタバイパス 316 は、読み取りデータパイプライン 108 において、書き込みデータパイプライン 106 内のパケットサイズ 302 (packetizer) の前およびデパケットサイズ 324 (depacketizer) の後にあるが、読み取りおよび書き込みデータパイプライン 106、108 内のどこかに配置されていてもよい。不要データコレクタバイパス 316 は、ソリッドステートストレージ 110 内のストレージ効率を改善するために、バーチャルページの残りの部分を埋めるように、書き込みパイプライン 106 のフラッシュ中に使用されてもよく、これにより、不要データコレクションの頻度を低減することができる。

【0127】

一実施例においては、書き込みデータパイプライン 106 は、有効な書き込みオペレーション用のデータをバッファする書き込みバッファ 320 を具える。通常、書き込みバッファ 320 は、ソリッドステートストレージ 110 の少なくとも 1 つのバーチャルページを埋めるのに十分なパケット容量を具える。これによって、書き込みオペレーションは、中断せずに、ソリッドステートストレージ 110 にページデータの全体を送ることができる。書き込みデータパイプライン 106 の書き込みバッファ 320 と、読み取りデータパイプライン 108 内のバッファを、同一容量、または、ソリッドステートストレージ 110 内のストレージ書き込みバッファよりも大きい容量にサイジングすることで、書き込みおよび読み取りデータは、複数コマンドの代わりに、1 つの書き込みコマンドが、ソリッドステートストレージ 110 にバーチャルページデータの全体を送るようにクラフト化されるので、より効率的になる。

【0128】

書き込みバッファ 320 が埋められる一方、ソリッドステートストレージ 110 は、その他の読み取りオペレーションに使用されてもよい。このことが有利であるのは、書き込みバッファがより小さいか、または、書き込みバッファのないその他のデバイスが、データがストレージ書き込みバッファに書き込まれ、ストレージ書き込みバッファ内にフローするデータがストールするとき、ソリッドステートストレージをタイアップできるからである。読み取りオペレーションは、ストレージ書き込みバッファ全体がうまり、プログラム化されるまで、ブロックされる。書き込みバッファのないシステムまたは書き込みバッファが小さいシステムに対する別のアプローチは、読み取り可能にするために、フルではないストレージ書き込みバッファをフラッシュすることである。ページを埋めるのに複数の書き込み/プログラムサイクルも必要となるので、これは非効率的である。

【0129】

バーチャルページよりも大きなサイズの書き込みバッファ 320 を有する描かれた実施例に関して、多数のサブコマンドを具える 1 つの書き込みコマンドは、次いで、1 のプログラムコマンドが続き、各ソリッドステートストレージエレメント 216、218、220 におけるストレージ書き込みバッファから、各ソリッドステートストレージエレメント 216、218、220 内の指定されたページに、ページデータを転送する。この技術は

、バッファが埋まっているとき、データ信頼性および耐久性を低下させ、読み取りコマンドおよびその他のコマンドのための指定バンクを解放させることで周知である部分的なページプログラミングを省くことができる利点がある。

【0130】

一実施例においては、書き込みバッファ320は、ピンポンバッファであり、このバッファの片面が埋まり、次いで、適宜な時間での転送が指定され、ピンポンバッファのもう一方の面が埋まる。別の実施例においては、書き込みバッファ320は、バーチャルページデータセグメントよりも大きい容量を有する先入れ先出し(「FIFO」)レジスタを具える。当分野の当業者であれば、バーチャルページデータがデータをソリッドステートストレージ110に書き込む前に記憶させることができるその他の書き込みバッファ320設定を理解されたい。

【0131】

別の実施例においては、書き込みバッファ320は、バーチャルページよりも小さなサイズであり、ページ情報よりも少ない情報が、ソリッドステートストレージ110内のストレージ書き込みバッファに書き込まれる。実施例においては、書き込みデータパイプライン106内のストールが読み取りオペレーションを妨害するのを防ぐために、データは、1つのロケーションから不要データコレクションプロセスの一部としての別のロケーションへと移動される必要がある、不要データコレクションシステム用いてデータをキューイングする。書き込みデータパイプライン106においてデータストールが生じた場合、データは、不要データコレクタバイパス316を介して、書き込みバッファ320へと供給可能であり、次いで、ソリッドステートストレージ110内のストレージ書き込みバッファに供給され、データをプログラミングする前に、バーチャルページのページを埋める。この方法では、書き込みデータパイプライン106内のデータストールは、ソリッドステートストレージデバイス102から読み取りをストールしない。

【0132】

別の実施例においては、書き込みデータパイプライン106は、書き込みデータパイプライン106内の1又はそれ以上のユーザ定義可能機能を有する書き込みプログラムモジュール310を具える。書き込みプログラムモジュール310によって、ユーザは書き込みデータパイプライン106をカスタマイズすることができる。ユーザは、特定のデータ要求またはアプリケーションに基づいて書き込みデータパイプライン106をカスタマイズすることができる。ソリッドステートストレージ制御部104はFPGAである場合、ユーザは、カスタムコマンドおよび機能を有する書き込みデータパイプライン106を比較的容易にプログラムすることができる。さらに、ユーザは、書き込みプログラムモジュール310を使用し、ASICを用いたカスタム機能を具えるが、しかしながら、ASICをカスタマイズすることは、FPGAを用いるよりも難しい。書き込みプログラムモジュール310は、バッファおよびバイパス機構を具えることができ、これらによって、第1のデータセグメントが書き込みプログラムモジュール310で実行され、第2のデータセグメントは、書き込みデータパイプライン106を介して繰り返すことができる。別の実施例においては、書き込みプログラムモジュール310は、ソフトウェアを介してプログラム可能なプロセッサコアを具えることができる。

【0133】

書き込みプログラムモジュール310は、入力バッファ306と圧縮モジュール312との間に示されているが、書き込みプログラムモジュール310は、書き込みデータパイプライン106内のあらゆる位置に存在でき、各種ステージ302-320中に分散可能である。さらに、独立してプログラムおよびオペレーション可能な各種ステート302-320中に分散される、複数の書き込みプログラムモジュール310があってもよい。さらに、ステージ302-320の順番は変更可能である。当分野の当業者であれば、特定のユーザ要求に基づいてステージ302-320の順番を作動可能に変更したものを理解されたい。

【0134】

読み取りデータパイプライン

読み取りデータパイプライン 108 は、データエラーが ECC ブロックにある場合に、要求されたパケットが、要求されたパケットの各 ECC ブロックで記憶される ECC を用いることによって、ソリッドステートストレージ 110 から受信されることを判定する ECC 修正モジュール 322 を具える。次いで、ECC コレクションモジュール 322 は、エラーが存在している場合、要求されたパケットにおけるエラーを修正し、これらのエラーは、ECC を用いて修正可能である。例えば、ECC が、6 ビットのエラーを検出できるが、3 ビットのエラーしか修正できない場合、ECC 修正モジュール 322 は、エラー中の 3 ビットまで、要求されたパケットの ECC ブロックを修正する。ECC 修正モジュール 322 は、エラー中のビットを、正しい 1 またはゼロのステートに変更することで、エラー中のビットを修正し、要求されたデータパケットは、ソリッドステートストレージ 110 に書き込まれたときのものと同一であり、ECC は、パケット用に生成された。

【0135】

ECC が修正できるよりも多く、要求されたパケットがエラー中のビットを含んでいると、ECC 修正モジュール 322 が判定した場合、ECC 修正モジュール 322 は、要求されたパケットの破損した ECC ブロック中のエラーを修正できず、中断信号を送る。一実施例においては、ECC 修正モジュール 322 は、要求されたパケットがエラー状態にあることを示すメッセージを有する中断信号を送る。メッセージは、ECC 修正モジュール 322 がエラーを修正できず、エラーを修正するための ECC 修正モジュール 322 の能力がないことが示される。別の実施例においては、ECC 修正モジュール 322 は、要求されたパケットの破損 ECC ブロックを、中断信号および / またはメッセージとともに送る。

【0136】

好適な実施例においては、ECC 修正モジュール 322 によって修正不可能な要求されたパケットの、破損した ECC ブロックまたは破損した ECC ブロックの一部は、マスタ制御部 224 によって読み取られ、修正され、ECC 修正モジュール 322 に返され、読み取りデータパイプライン 108 によってさらに処理される。一実施例においては、要求されたパケットの破損した ECC ブロックの一部ブロックまたはその一部は、データを要求しているデバイスに送られる。要求デバイス 155 は、ECC ブロックを修正することができ、バックアップまたはミラーコピーなど、別のコピーを使用してデータを置換することができ、次いで、要求されたデータパケットの置換データを使用するか、それを読み取りデータパイプライン 108 に返すことができる。要求デバイス 155 は、エラー中の要求されたパケットのヘッダ情報を使用し、破損した要求パケットを置換するか、または、パケットが属するオブジェクトを置換するかを必要とするデータを識別する。別の好適な実施例においては、ソリッドステートストレージ制御部 104 は、いくつかのタイプの RAID を使用してデータを記憶し、破損したデータをリカバリすることができる。別の実施例においては、ECC コレクションモジュール 322 は、中断信号および / またはメッセージを送り、受信するデバイスは、要求されたデータパケットに関連した読み取りオペレーションができない。当分野の当業者であれば、要求されたパケットの 1 又はそれ以上の ECC ブロックが破損していて、ECC 修正モジュール 322 がエラーを修正できないことを判定した ECC 修正モジュール 322 の結果として、とるべきその他の選択肢やアクションを理解されたい。

【0137】

読み取りデータパイプライン 108 は、デパケットタイザ 324 を具え、このデパケットタイザ 324 は、直接または間接的に、ECC 修正モジュール 322 から要求パケットの ECC ブロックを受信し、チェックして、1 又はそれ以上のパケットヘッダを除去する。一実施例においては、ヘッダは、読み取りデータパイプライン 108 に送られたパケットが、要求されたパケットであることを確証するために使用可能なハッシュコードを具える。さらに、デパケットタイザ 324 は、パケットタイザ 302 によって加えられた要求パケットからヘッダを除去することができる。デパケットタイザ 324 は、所定のパケット上で機能

しないように指令できるが、修正されずに先に送られる。例としては、再構築プロセスの途中で要求されるコンテナラベルであってもよく、ヘッダ情報は、オブジェクトインデックス再構築モジュール 272 によって要求される。さらなる例としては、ソリッドステートストレージデバイス 102 内で使用されるように設計された各種タイプのパケットを転送することを含む。別の実施例においては、デパケットタイザ 324 オペレーションは、パケットタイプに応じたものであってもよい。

【0138】

読み取りデータパイプライン 108 は、デパケットタイザ 324 からデータを受信し、不要なデータを除去するアライメントモジュール 326 を具える。一実施例においては、ソリッドステートストレージ 110 に送られる読み取りコマンドは、パケットデータをリカバリさせる。データを要求しているデバイスは、リカバリしたパケットの全てのデータを必要としなくてもよく、アライメントモジュール 326 は、不要なデータを除去する。リカバリしたページ内の全てのデータが要求されているデータである場合、アライメントモジュール 326 は、いずれのデータも除去しない。

【0139】

アライメントモジュール 326 は、次のステージにデータセグメントを送る前に、データセグメントを要求しているデバイスと互換性のある形態で、オブジェクトのデータセグメントとしてのデータをリフォーマットする。通常、データは、読み取りデータパイプライン 108 によって処理されるとき、データセグメントまたはパケットのサイズは、各種ステージで変更される。アライメントモジュール 326 は、受信したデータを使用して、データを、要求デバイス 155 に送り、応答を形成するのに加えられるのに適したデータセグメントにフォーマットする。例えば、第 1 のデータパケットの一部は、第 2 のデータパケットの一部のデータと組み合わせることができる。データセグメントが、要求デバイス 155 によって要求されるデータよりも長い場合、アライメントモジュール 326 は、不要データを捨てることができる。

【0140】

一実施例においては、読み取りデータパイプライン 108 は、読み取りデータパイプライン 108 によって処理される前に、ソリッドステートストレージ 110 から読み取られる 1 又はそれ以上の要求パケットをバッファリングする読み取り同期バッファ 328 を具える。読み取り同期バッファ 328 は、ソリッドステートストレージクロックドメインと、ローカルバスクロックドメインとの間のバウンダリにあり、クロックドメインの差を考慮するバッファリングを行う。

【0141】

別の実施例においては、読み取りデータパイプライン 108 は、出力バッファ 330 を具え、この出力バッファ 330 は、アライメントモジュール 326 からの要求パケットを受信し、要求デバイス 155 に転送される前にパケットを記憶する。出力バッファ 330 は、データセグメントが、読み取りデータパイプライン 108 のステージから受信されるときと、データセグメントが、ソリッドステートストレージ制御部 104 のその他部分または要求デバイス 155 に転送されるときと、の間の差を考慮している。さらに、出力バッファ 330 によって、データバス 204 は、データバス 204 のオペレーション効率を改善するために、読み取りデータパイプライン 108 によって維持可能なレートよりも高いレートで、読み取りデータパイプライン 108 からのデータを受信できる。

【0142】

一実施例においては、読み取りデータパイプライン 108 は、メディア解読モジュール 332 を具え、このメディア解読モジュール 332 は、ECC 修正モジュール 322 からの 1 又はそれ以上の暗号化された要求パケットを受信し、1 又はそれ以上の要求されたパケットをデパケットタイザ 324 に送る前に、ソリッドステートストレージデバイス 102 に固有の暗号化キーを用いて 1 又はそれ以上の要求されたパケットを解読する。通常、メディア解読モジュール 332 によってデータを解読するのに使用される暗号化キーは、メディア解読モジュール 318 によって使用される暗号化キーと同一である。別の実施例に

においては、ソリッドステートストレージ 110 は、2 又はそれ以上のパーティションを具備していてもよく、ソリッドステートストレージ制御部 104 は、2 又はそれ以上のソリッドステートストレージ制御部 104 であるかのように動作し、各々は、ソリッドステートストレージ 110 内の 1 つのパーティションで機能する。この実施例においては、固有のメディア暗号化キーは、各パーティションで使用されてもよい。

【0143】

別の実施例においては、読み取りデータパイプライン 108 は、解読モジュール 334 を具備し、この解読モジュールは、データセグメントを出力バッファ 330 に送る前に、デパケットライザ 324 によってフォーマットされるデータセグメントを解読する。データセグメントは、読み取り要求と共に受信された暗号化キーを用いて解読され、読み取り要求は、読み取り同期バッファ 328 が受信する要求パケットのリカバリを開始させる。解読モジュール 334 は、第 1 のパケットの読み取り要求とともに受信される暗号化キーを有する第 1 のパケットを解読でき、次いで、異なる暗号化キーを有する第 2 のパケットを解読することができ、または、解読せずに読み取りデータパイプライン 108 の次のステージに第 2 のパケットを送ることができる。通常、解読モジュール 334 は、メディア解読モジュール 332 が要求されたパケットを解読するのに使用するものとは異なる暗号化キーを使用してデータセグメントを解読する。パケットが非シークレット暗号化法ナンスで記憶されているときは、ナンスを、暗号化キーとともに使用して、データパケットを解読する。暗号化キーは、クライアント 114、コンピュータ 112、キーマネージャ、または、ソリッドステートストレージ制御部 104 によって使用される暗号化キーを管理するその他のデバイスから、受信できる。

【0144】

別の実施例においては、読み取りデータパイプライン 108 は、デパケットライザ 324 によってフォーマットされたデータセグメントを展開する展開モジュール 336 を具備する。好適な実施例においては、展開モジュール 336 は、1 又はそれ以上のパケットヘッダおよびコンテナラベルに記憶された圧縮情報を使用し、圧縮モジュール 312 によってデータを圧縮するのに使用されたルーチンと相補的なルーチンを選択する。別の実施例においては、展開モジュール 336 によって使用される展開ルーチンは、展開しようとするデータセグメントを要求しているデバイスによって、指令される。別の実施例においては、展開モジュール 336 は、オブジェクト毎 (per object) のタイプまたはオブジェクトクラスベースのデフォルトの設定によって、展開ルーチンを選択する。第 1 のオブジェクトの第 1 のパケットは、デフォルトの展開ルーチンを無効にすることができ、同一オブジェクトクラスおよびオブジェクトタイプの第 2 のオブジェクトの第 2 のパケットは、デフォルトの展開ルーチンを使用することができ、同一オブジェクトクラスおよびオブジェクトタイプの第 3 のオブジェクトの第 3 のパケットは、展開を利用しなくてもよい。

【0145】

別の実施例においては、読み取りデータパイプライン 108 は、読み取りデータパイプライン 108 内の 1 又はそれ以上のユーザ定義可能機能を含む読み取りプログラムモジュール 338 を具備する。読み取りプログラムモジュール 338 は、書き込みプログラムモジュール 310 と同様の特性を有し、ユーザは、読み取りデータパイプライン 108 にカスタム機能を加えることができる。読み取りプログラムモジュール 338 は、図 3 に示されるように配置可能であり、または、読み取りデータパイプライン 108 内で別のポジションに配置可能であり、あるいは、読み取りデータパイプライン 108 内の複数のロケーションの複数の部分を具備してもよい。さらに、独立して機能する読み取りデータパイプライン 108 内の複数のロケーション内に、複数の読み取りプログラムモジュール 338 があってもよい。当分野の当業者であれば、読み取りデータパイプライン 108 内に、読み取りプログラムモジュール 338 のその他の形態を理解されたい。書き込みデータパイプライン 106 に関して、読み取りデータパイプライン 108 のステージは、再配置可能であり、当分野の当業者であれば、読み取りデータパイプライン 108 内のその他の順番を理

解されたい。

【0146】

ソリッドステートストレージ制御部104は、制御およびステータスレジスタ340と、関連する制御キュー342を具える。制御およびステータスレジスタ340および制御キュー342は、書き込みデータパイプライン106および読み取りデータパイプライン108内で処理されるデータに関連した、制御およびシークエンスコマンドならびにサブコマンドを用意にする。例えば、パケットサイズ302のデータセグメントは、ECCジェネレータ304に関連した制御キュー342において、1又はそれ以上の関連する制御コマンドまたは指令を具えることができる。データセグメントが、パケット化されるとき、指令またはコマンドのいくらかは、パケットサイズ302内で実行されてもよい。その他のコマンドまたは指令は、データセグメントから作成された新規に形成されたデータパケットが、次のステージに送られるとき、制御およびステータスレジスタ340を介して次の制御キュー342に送られる。

【0147】

コマンドまたは指令は、制御キュー342に同時にロードされてもよく、パケットは、書き込みデータパイプライン106に送られ、各パイプラインステージは、各々のパケットがそのステージで実行されるとき、適宜なコマンドまたは指令を引き出す。同様に、コマンドまたは指令は、制御キュー342に同時にロード可能であり、パケットは、読み取りデータパイプライン108から要求されており、各パイプラインステージは、各々のパケットがそのステージで実行されるときに、適宜なコマンドまたは指令を引き出す。当分野の当業者であれば、制御およびステータスレジスタ340、ならびに、制御キュー342のその他の特性および機能を理解されたい。

【0148】

ソリッドステートストレージ制御部104またはソリッドステートストレージデバイス102も、バンクインターリーブ制御部344、同期バッファ346、ストレージバス制御部348およびマルチプレクサ(「MUX」)350を具えることができ、これらは、図4Aおよび図4Bに関連して記載される。

【0149】

バンクインターリーブ

図4Aは、本発明によるソリッドステートストレージ制御部104内のバンクインターリーブ制御部344の一実施例400を示した概略ブロック図である。バンクインターリーブ制御部344は、制御およびステータスレジスタ340に接続され、MUX350を介してストレージI/Oバス210およびストレージ制御バス212に接続され、ストレージ制御部348および同期バッファ346に接続され、これらは以下で説明する。バンクインターリーブ制御部344は、ソリッドステートストレージ110内のバンク214に関して、読み取りエージェント402、書き込みエージェント404、消去エージェント406、管理エージェント408、読み取りキュー410a-n、書き込みキュー412a-n、消去キュー414a-n、管理キュー416a-nを具え、バンク制御部418a-n、バスアービター420、ステータスMUX422を具え、これらは以下で説明する。ストレージバス制御部348は、リマッピングモジュール430を有するマッピングモジュール424、ステータス取得モジュール426、および、NANDバス制御部428を具え、これらは以下で説明する。

【0150】

バンクインターリーブ制御部344は、バンクインターリーブ制御部344において、1又はそれ以上のコマンドを2又はそれ以上のキューに指令し、ソリッドステートストレージ110のバンク214中で、上記キューに記憶されたコマンドを調整し、その結果、第1のタイプのコマンドは1つのバンク214aで実行され、第2のタイプのコマンドは第2のバンク214bで実行される。1又はそれ以上のコマンドは、コマンドタイプによってキューに分けられる。ソリッドステートストレージ110の各バンク214は、バンクインターリーブ制御部344内の関連するキューセットを有し、各々のキューセットは

、各コマンドタイプ用のキューを具える。

【0151】

バンクインターリーブ制御部344は、ソリッドステートストレージ110のバンク214中で、キューに記憶されたコマンドの実行を調整する。例えば、第1のタイプのコマンドは、1つのバンク214aで実行され、第2位のタイプのコマンドは、第2のバンク214bで実行される。通常、コマンドタイプおよびキュータイプは、読み取りコマンドおよび書き込みコマンド、ならびに、キュー410、412を具えるが、さらに、ストレージメディア特異的なその他のコマンドおよびキューを具えることができる。例えば、図4Aに描かれる実施例において、消去キューおよび管理キュー414、416が含まれ、フラッシュメモリ、NRAM、MRAM、DRAM、PRAMなどに適している。

【0152】

その他のタイプのソリッドステートストレージ110に関し、その他のタイプのコマンドおよび対応するキューも、本発明の範囲から逸脱することなく含むことができる。FPGAソリッドステートストレージ制御部104のフレキシブルな性質は、ストレージメディアにおけるフレキシビリティを付与する。フラッシュメモリが、別のソリッドステートストレージタイプに変更された場合、バンクインターリーブ制御部344、ストレージバス制御部348、および、MUX350は、データパイプライン106、108、および、その他のソリッドステートストレージ制御部104機能に重大な影響を与えずに、メディアタイプを収容できるように変更可能である。

【0153】

図4Aに描かれた実施例においては、バンクインターリーブ制御部344は、各バンク214に関して、ソリッドステートストレージ110からのデータを読み取るための読み取りキュー410、ソリッドステートストレージ110に対する書き込みコマンド用の書き込みキュー412、ソリッドステートストレージ中の消去ブロックを消去するための消去キュー414、管理コマンド用の管理キュー416を具える。さらに、バンクインターリーブ制御部344は、関連する読み取り、書き込み、消去および管理エージェント402、404、406、408を具える。別の実施例においては、制御およびステータスレジスタ340ならびに制御キュー342、または、同様のコンポーネントは、バンクインターリーブ制御部344なしで、ソリッドステートストレージ110のバンク214に送られるデータ用コマンドをキューイングする。

【0154】

エージェント402、404、406、408は、一実施例においては、特定のバンク214a用に設計された適宜なタイプのコマンドを、バンク214aの正しいキューに指令する。例えば、読み取りエージェント402は、バンク-1214b用の読み取りコマンドを受信し、読み取りコマンドをバンク-1読み取りキュー410bに指令できる。書き込みエージェント404は、ソリッドステートストレージ110のバンク-0214aにおけるロケーションにデータを書き込むための書き込みコマンドを受信し、次いで、下記コマンドをバンク-0書き込みキュー412aに送ることができる。同様に、消去エージェント406は、バンク-1214bの消去ブロックを消去する消去コマンドを受信し、次いで、消去コマンドをバンク-1消去キュー414bに送ることができる。管理エージェント408は、通常、バンク-0214aなど、バンク214の設定レジスタを読み取るためのリセットコマンドまたは要求など、管理コマンド、ステータス要求など、を受信する。管理エージェント408は、管理コマンドをバンク-0管理キュー416aに送る。

【0155】

さらに、エージェント402、404、406、408は、通常、キュー410、412、414、416のステータスをモニタし、ステータス、または、キュー410、412、414、416がフル、ほぼフル、機能しないときの中断信号またはメッセージを送る。一実施例においては、エージェント402、404、406、408は、コマンドを受信し、関連するサブコマンドを作成する。一実施例においては、エージェント402、

404、406、408は、制御及びステータスレジスタ340を介したコマンドを受信し、キュー410、412、414、416に送られる関連サブコマンドを作成する。当分野の当業者であれば、エージェント402、404、406、408のその他の機能も理解されたい。

【0156】

キュー410、412、414、416は、通常、コマンドを受信し、ソリッドステートストレージバンク214に送られる必要があるまでコマンドを記憶する。通常の実施例においては、キュー410、412、414、416は、先入れ先出し（「FIFO」）レジスタまたはFIFOとして機能する同様のコンポーネントである。別の実施例においては、キュー410、412、414、416は、データにマッチする順番、重要度の順番またはその他の基準でコマンドを記憶する。

【0157】

バンク制御部418は、通常、キュー410、412、414、416からのコマンドを受信し、適宜なサブコマンドを作成する。例えば、バンク-0書き込みキュー412aは、ページデータパケットをバンク-0 214aに書き込むコマンドを受信することができる。バンク-0制御部418aは、適宜な時間で書き込みコマンドを受信し、1又はそれ以上の書き込みコマンドを作成することができ、書き込みバッファ320に記憶される各データパケットは、バンク-0 214aに書き込まれる。例えば、バンク-0制御部418aは、バンク0 214aおよびソリッドステートストレージアレイ216を確認するコマンドを作成し、1又はそれ以上のデータパケットを書き込む適宜なロケーションを選択肢、ソリッドステートストレージメモリアレイ216内の入力バッファをクリアし、1又はそれ以上のデータパケットを入力バッファに転送し、入力バッファを選択したロケーションにプログラムし、データが正しくプログラムされたことを検証し、プログラムに破損がある場合、1又はそれ以上のマスタ制御部224の中断を行い、同一の物理ロケーションへの書き込みを再度行い、異なる物理ロケーションへの書き込みを行う。さらに、例示の書き込みコマンドとともに、ストレージバス制御部348は、1又はそれ以上のコマンドを、ストレージI/Oバス210a用の第1の物理アドレスにマップされたコマンドの論理アドレス、および、ストレージI/Oバス210b用の第2の物理アドレスにマップされたコマンドの論理アドレスで、ストレージI/Oバス210a-nの各々に複製した。

【0158】

通常、バスアービター420は、バンク制御部418中から選択され、バンク制御部418内の出力キューからサブコマンドを引き出し、バンク214の性能を最適化するように、これらをシークエンス状態でストレージバス制御部348に送る。別の実施例においては、バスアービター420は、高レベル中断信号に応答し、通常の実施例を修正することができる。別の実施例においては、マスタ制御部224は、制御およびステータスレジスタ340を介してバスアービター420を制御することができる。当分野の当業者であれば、バスアービター420が、バンク制御部418からソリッドステートストレージ110へのコマンドシークエンスを制御しインターリーブすることができるその他の手段を理解されたい。

【0159】

バスアービター420は、通常、適宜なコマンドと、コマンドタイプが必要とされるときは、バンク制御部418の関連データと、の選択を調整し、コマンド及びデータをストレージバス制御部348に送る。一般的に、バスアービター420は、さらに、適宜なバンク214を選択するのに、ストレージ制御バス212にコマンドを送る。フラッシュメモリ、または、非同期式2方向性シリアルストレージI/Oバス210を有するその他のソリッドステートストレージ110の場合、1つのコマンド（制御情報）またはデータセットのみが、同時に転送可能である。例えば、書き込みコマンドまたはデータが、ストレージI/Oバス210のソリッドステートストレージ110に転送され、読み取りコマンド、データが読み取られるときは、消去コマンド、管理コマンドまたはその他のステータ

スコマンドは、ストレージ I / O バス 2 1 0 に転送できない。

【 0 1 6 0 】

例えば、バンク - 0 の書き込みオペレーション中、バスアービター 4 2 0 は、バンク - 0 制御部 4 1 8 a を選択し、この制御部 4 1 8 a は、ストレージバス制御部 3 4 8 に次のシーケンスを実行させるキューのトップに書き込みコマンドまたは一連の書き込みサブコマンドを有することができる。バスアービター 4 2 0 は、書き込みコマンドをストレージバス制御部 3 4 8 に送り、ストレージ制御バス 2 1 2 を介してバンク - 0 2 1 4 a を選択することによって、書き込みコマンドをセットアップし、バンク - 0 2 1 4 a に関連したソリッドステートストレージエレメント 1 1 0 の入力バッファをクリアするコマンドを送り、バンク - 0 2 1 4 a に関連したソリッドステートストレージエレメント 2 1 6、2 1 8、2 2 0 のステータスを確認するコマンドを送る。次いで、ストレージバス制御部 3 4 8 は、ストレージ I / O バス 2 1 0 に書き込みサブコマンドを転送し、論理消去ブロックアドレスからマッピングされるときに、個々の物理的消去ソリッドステートストレージエレメント 2 1 6 a - m 用の論理消去ブロックのアドレスを含む、物理アドレスを具える。次いで、ストレージバス制御部 3 4 8 は、M U X 3 5 0 を介して、書き込み同期バッファ 3 0 8 をストレージ I / O バス 2 1 0 へと書き込みバッファ 3 2 0 をマルチプレックス (M U X) し、書き込みデータを適宜なページにストリーミングする。ページがフルのとき、次いで、ストレージバス制御部 3 4 8 は、バンク - 0 2 1 4 a に関連するソリッドステートストレージエレメント 2 1 6 a - m に、入力バッファを、ソリッドステートストレージエレメント 2 1 6 a - m 内のメモリセルにプログラムさせる。最終的に、ストレージバス制御部 3 4 8 は、ステータスを確認して、ページが正しくプログラムされていることを確かめる。

【 0 1 6 1 】

読み取りオペレーションは、上述の書き込み例と同様である。読み取りデータパイプライン 1 0 8 中、通常、バスアービター 4 2 0 またはバンクインターリーブ制御部 3 4 4 のその他のコンポーネントは、データと、関連するステータス情報を受信し、データを読み取りデータパイプライン 1 0 8 に送り、ステータス情報を制御およびステータスレジスタ 3 4 0 に送る。通常、バスアービター - 4 2 0 からストレージバス制御部 3 4 8 に送られる読み取りデータコマンドは、M U X 3 5 0 に、ストレージ I / O バス 2 1 0 の読み取りデータを読み取りデータパイプライン 1 0 8 にゲート制御し、ステータス M U X 4 2 2 を介してステータス情報を適宜な制御及びステータスレジスタ 3 4 0 に送る。

【 0 1 6 2 】

バスアービター - 4 2 0 は、各種コマンドタイプとデータアクセスモードを調整し、適宜なコマンドタイプまたは関連するデータのみが、所定時間に、バスに存在する。バスアービター - 4 2 0 が、書き込みコマンドを選択しており、書き込みサブコマンドと関連するデータとが、ソリッドステートストレージ 1 1 0 に書き込まれているとき、バスアービター - 4 2 0 は、ストレージ I / O バス 2 1 0 のその他コマンドタイプを許可しない。有利には、バスアービター 4 2 0 は、バンク 2 1 4 ステータスに関して受信したステータス情報とともに、予測したコマンド実行時間など、タイミング情報を使用し、バスのアイドル時間を最小化または省く目的として、バスの各種コマンドの実行を調整する。

【 0 1 6 3 】

バスアービター 4 2 0 を介してマスタ制御部 2 2 4 は、通常、ステータス情報とともに、キュー 4 1 0、4 1 2、4 1 4、4 1 6 に記憶されたコマンドの予測完了時間を使用し、コマンドに関連したサブコマンドが 1 つのバンク 2 1 4 a で実行されているとき、他のコマンドの他のサブコマンドが、他のバンク 2 1 4 b - n で実行される。1 つのコマンドが、2 1 4 a で完全に実行されるとき、バスアービター 4 2 0 は、別のコマンドをバンク 2 1 4 a に指令する。バスアービター - 4 2 0 は、キュー 4 1 0、4 1 2、4 1 4、4 1 6 に記憶されたコマンドと、キュー 4 1 0、4 1 2、4 1 4、4 1 6 に記憶されていない他のコマンドを調整することができる。

【 0 1 6 4 】

例えば、消去コマンドは、ソリッドステートストレージ 110 内の消去ブロック群を消去するように送信可能である。消去コマンドは、書き込みまたは読み取りコマンドよりも 10 乃至 1000 倍以上実行され、プログラムコマンドよりも 10 乃至 100 倍以上実行される。Nバンク 214 に関して、バンクインターリーブ制御部 344 は、消去コマンドを N コマンドにスプリットすることができ、各々は、バンク 214 a のバーチャル消去ブロックを消去する。バンク - 0 214 a が消去コマンドを実行している間、バスアービタ - 420 は、その他のバンク 214 b - n で実行されるその他のコマンドを選択することができる。さらに、バスアービタ - 420 は、ストレージバス制御部 348、マスタ制御部 224 など、他のコンポーネントとともに機能することができ、バス間のコマンド実行を調整する。バスアービタ - 420、バンク制御部 418、キュー 410、412、414、416 およびバンクインターリーブ制御部 344 のエージェント 402、404、406、408 を使用してコマンド実行を調整することで、バンクインターリーブ機能なしで、その他のソリッドステートストレージシステムを介して、大幅に性能を向上させることができる。

【0165】

一実施例においては、ソリッドステートストレージ制御部 104 は、ソリッドステートストレージ 110 のストレージエレメント 216、218、220 の全てをサーブする 1 つのバンクインターリーブ制御部 344 を具える。別の実施例においては、ソリッドステート制御部 104 は、ストレージエレメント 216 a - m、218 a - m、220 a - m の各列に対するバンクインターリーブ制御部 344 を具える。例えば、1 つのバンクインターリーブ制御部 344 は、ストレージエレメント SSS0 . 0 - SSS0 . N216 a、218 a、220 a の一列をサーブし、第 2 のバンクインターリーブ制御部 344 は、ストレージエレメント SSS1 . 0 - SSS1 . N216 b、218 b、220 b などの第 2 の列をサーブする。

【0166】

図 4 B は、本発明によるソリッドステートストレージ制御部 104 のバンクインターリーブ制御部 344 の代替実施例 401 を示した概略ブロック図である。図 4 B に示された実施例に描かれたコンポーネント 210、212、340、346、348、350、402 - 430 は、図 4 A に関連して描かれたバンクインターリーブ装置 400 と実質的に同じであるが、ただし、各バンク 214 が、1 つのキュー 432 a - n を具えて、バンク（例えば、バンク 0 214 a）用の読み取りコマンド、書き込みコマンド、消去コマンド、管理コマンドが、バンク 214 a の 1 つのキュー 432 a に対して指令される点を除く。一実施例においては、キュー 432 は、FIFO である。別の実施例においては、キュー 432 は、キュー 432 が記憶されている順番とは別の順番で、キュー 432 から引き出されたコマンドを有することができる。別の代替実施例においては（図示せず）、読み取りエージェント 402、書き込みエージェント 404、消去エージェント 406 および管理エージェント 408 は、適宜なキュー 432 a - n に対して、1 つのエージェント割当コマンドに組み込むことができる。

【0167】

別の代替実施例においては（図示せず）、コマンドは、1 つのキューに記憶され、これらのコマンドは、記憶された方法以外の順番でキューから引き出し可能であり、バンクインターリーブ制御部 344 は、1 つのバンク 214 a でコマンドを実行でき、他のコマンドは、残りのバンク 214 b - n で実行される。当分野の当業者であれば、他のコマンドが他のバンク 214 b - n で実行されている間に、1 つのバンク 214 a でコマンドを実行できる、他のキュー設定およびタイプを容易に理解されたい。

【0168】

ストレージ特異的コンポーネント

ソリッドストレージ制御部 104 は、同期バッファ 346 を具え、この同期バッファ 346 は、ソリッドステートストレージ 110 から送られて受信したコマンドおよびステータスメッセージをバッファリングする。同期バッファ 346 は、ソリッドステートストレ

ージクロックドメインとローカルバスクロックドメインとの間のバウンダリに配置され、クロックドメインの差を考慮するバッファリングを提供する。同期バッファ 346、書き込み同期バッファ 308 および読み取り同期バッファ 328 は、独立しており、同時に機能して、データ、コマンド、ステータスメッセージなどをバッファリングすることができる。好適な実施例においては、同期バッファ 346 は、クロックドメインをクロスする信号が最小のところに配置される。当分野の当業者であれば、クロックドメイン間の同期は、設計事項のいくつかを最適にするために、ソリッドステートストレージデバイス 102 内の他のロケーションに自由に移動可能であることを理解されたい。

【0169】

ソリッドステートストレージ制御部 104 は、ストレージバス制御部 348 を具え、このストレージバス制御部 348 は、ソリッドステートストレージ 110 に送られ、読み取られるデータに関するコマンドと、ソリッドステートストレージ 110 のタイプに基づいてソリッドステートストレージ 110 から受信したステータスメッセージとを解釈し翻訳する。例えば、ストレージバス制御部 348 は、性能特性の異なるストレージ、製造者の異なるストレージなど、異なるタイプのストレージに対する異なるタイミング要求を具えることができる。さらに、ストレージバス制御部 348 は、ストレージ制御バス 212 に制御コマンドを送る。

【0170】

好適な実施例においては、ソリッドステートストレージ制御部 104 は、MUX 350 を具え、MUX 350 は、マルチプレクサ 350a - n のアレイを具え、各マルチプレクサは、ソリッドステートストレージアレイ 110 の列の専用となる。例えば、マルチプレクサ 350a は、ソリッドステートストレージエレメント 216a、218a、220a に関連している。MUX 350 は、書き込みデータパイプライン 106 からのデータと、ストレージ I/O バス 210 を介してストレージバス制御部 348 からソリッドステートストレージ 110 へのコマンドをルーティングし、ストレージ I/O バス 210 を介してソリッドステートストレージ 110 から読み取りデータパイプライン 108 へのデータおよびステータスメッセージをルーティングし、ストレージバス制御部 348、同期バッファ 346 およびバンクインターリーブ制御部 344 を介して制御およびステータスレジスタ 340 にルーティングする。

【0171】

好適な実施例においては、ソリッドステートストレージ制御部 104 は、ソリッドステートストレージエレメントの各列（例えば、SSS0.1 216a、SSS0.2 218a、SSS0.N 220a）用の MUX 350 を具える。MUX 350 は、書き込みデータパイプライン 106 からのデータと、ストレージ I/O バス 210 を介してソリッドステートストレージ 110 へ送られるコマンドをコンバインし、コマンドから読み取りデータパイプライン 108 によって処理されるようにデータを分割する。書き込みバッファ 320 に記憶されたパケットは、ソリッドステートストレージエレメントの各列（SSSx.0 から SSSx.N 216、218、220）用の書き込み同期バッファ 308 から、ソリッドステートストレージエレメントの各列（SSSx.0 から SSSx.N 216、218、220）用の MUX 350 へと、書き込みバッファ 320 からのバスに指令される。コマンドおよび読み取りデータは、ストレージ I/O バス 210 から MUX 350 によって受信される。さらに、MUX 350 は、ステータスメッセージをストレージバス制御部 348 に指令する。

【0172】

ストレージバス制御部 348 は、マッピングモジュール 424 を具える。マッピングモジュール 424 は、消去ブロックの 1 又はそれ以上の物理アドレスに消去ブロックの論理アドレスをマッピングする。例えば、ブロック 214a あたり 20 のストレージエレメント（例えば、SSS0.0 から SSS M.0 216）のアレイを有するソリッドステートストレージ 110 は、ストレージエレメントあたり 1 の物理アドレスで、消去ブロックの 20 の物理アドレスにマッピングされる特定の消去ブロック用論理アドレスを具える

ことができる。ストレージエレメントが同時にアクセスされるので、ストレージエレメント 216 a、218 a、220 a の列の各ストレージエレメントにおける同一ポジションの消去ブロックは、物理アドレスを共有するであろう。列中の全ての消去ブロックではなく（例えば、ストレージエレメント S S S 0 . 0 , 0 . 1 , . . . 0 . N 216 a、218 a、220 a において）、1 の消去ブロックを選択するために、1 つのバンク（この場合、バンク - 0 214 a）が選択される。

【0173】

このような消去ブロックに関する論理 - 物理マッピングは、1 つの消去ブロックが損傷またはアクセス不能になった場合、マッピングが別の消去ブロックにマッピングされるように変更可能であるので、有益である。これによって、1 つのエレメントの消去ブロックが失敗したとき、バーチャル消去ブロック全体を損失するリスクを緩和することができる。再マッピングモジュール 430 は、消去ブロックの論理アドレスのマッピングを、バーチャル消去ブロックの 1 又はそれ以上の物理アドレス（ストレージエレメントのアレイに亘って分散させる）に変更する。例えば、バーチャル消去ブロック 1 は、ストレージエレメント S S S 0 . 0 216 a の消去ブロック 1、ストレージエレメント S S S 1 . 0 216 b の消去ブロック 1、. . .、ストレージエレメント M . 0 216 m の消去ブロック 1 にマッピング可能であり、バーチャル消去ブロック 2 は、ストレージエレメント S S S 0 . 1 218 a の消去ブロック 2、ストレージエレメント S S S 1 . 1 218 b の消去ブロック 2、. . .、ストレージエレメント M . 1 218 m の消去ブロック 2 などにマッピング可能である。

【0174】

ストレージエレメント S S S 0 . 0 216 a の消去ブロック 1 が損傷し、摩耗などによるエラーが生じるか、何らかの理由で利用できない場合、再マッピングモジュール 430 は、バーチャル消去ブロック 1 の消去ブロックをポインティングした論理アドレス用に、論理 - 物理マッピングを変更することができる。ストレージエレメント S S S 0 . 0 216 a のスペア消去ブロック（消去ブロック 221 と呼ぶ）が利用可能であり、その時点でマッピングされていない場合、再マッピングモジュール 430 は、ストレージエレメント S S S 0 . 0 216 a の消去ブロック 221 をポインティングするために、バーチャル消去ブロック 1 の消去ブロックのマッピングを変更でき、ストレージエレメント S S S 1 . 0 216 b の消去ブロック 1、ストレージエレメント S S S 2 . 0（図示せず）の消去ブロック 1、. . .、ストレージエレメント M . 0 216 m の消去ブロック 1 のポインティングを継続する。マッピングモジュール 424 または再マッピングモジュール 430 は、上述した順番（ストレージエレメントの消去ブロック 1 に対してバーチャル消去ブロック 1、ストレージエレメントの消去ブロック 2 に対してバーチャル消去ブロック 2）で消去ブロックをマッピングでき、あるいは、その他の基準に基づいた別の順番でストレージエレメント 216、218、220 の消去ブロックをマッピングできる。

【0175】

一実施例においては、消去ブロックは、アクセス時間によってグループ分け可能である。データを特定消去ブロックのページにプログラム（書き込み）するなど、コマンドを実行する時間を意味する、アクセス時間によるグループ分けは、コマンド完了を平均化（level）することができ、バーチャル消去ブロックの消去ブロックに亘って実行されるコマンドは、最低の消去ブロックによって制限されることはない。他の実施例においては、消去ブロックは、摩耗レベル、健全（health）などによってグループ分け可能である。当分野の当業者であれば、消去ブロックをマッピングまたは再マッピングするときに考慮すべきその他のファクタを理解されたい。

【0176】

一実施例においては、ストレージバス制御部 348 は、ステータス獲得モジュール 426 を具え、このステータス獲得モジュール 426 は、ソリッドステートストレージ 110 からのステータスメッセージを受信し、ステータスメッセージをステータス MUX 422 に送る。別の実施例においては、ソリッドステートストレージ 110 は、フラッシュメモ

リであり、ストレージバス制御部 348 は、NAND バス制御部 428 を具える。NAND バス制御部 428 は、読み取りデータパイプライン 106 および書き込みデータパイプライン 108 からソリッドステートストレージ 110 内の正しいロケーションにコマンドを送り、フラッシュメモリの特性に基づいてコマンド実行のタイミングを調整する。ソリッドステートストレージ 110 が別のタイプのソリッドステートストレージである場合、NAND バス制御部 428 は、ストレージタイプ特異的なバス制御部によって置換できる。当分野の当業者であれば、NAND バス制御部 428 のその他の機能も理解されたい。

【0177】

フローチャート

図 5 は、本発明によるデータパイプラインを用いてソリッドステートストレージデバイス 102 におけるデータ管理方法 500 の一実施例を示した概略フローチャート図である。この方法 500 は、スタートして (502)、入力バッファ 306 が、ソリッドステートストレージ 110 に書き込まれる 1 又はそれ以上のデータセグメントを受信する (504)。1 又はそれ以上のデータセグメントは、通常、オブジェクトの少なくとも一部を具えるが、オブジェクト全体であってもよい。パケットタイザ 302 は、オブジェクトと関連した 1 又はそれ以上のオブジェクト特異的なパケットを作成できる。パケットタイザ 302 は、各パケットにヘッダを加え、通常、各パケットは、パケット長およびオブジェクト内のパケットシーケンス番号を具える。パケットタイザ 302 は、入力バッファ 306 内に記憶される 1 又はそれ以上のデータまたはメタデータセグメントを受信し (504)、ソリッドステートストレージ 110 用にサイズを調整された 1 又はそれ以上のパケットを作成することによって 1 又はそれ以上のデータまたはメタデータセグメントをパケット化し (506)、各パケットは、1 又はそれ以上のセグメントから 1 つのヘッダおよびデータを具えている。

【0178】

通常、第 1 のパケットは、パケットが作成されるオブジェクトを識別するオブジェクト識別子を具える。第 2 のパケットは、ソリッドステートストレージデバイス 102 によって使用される情報を有するヘッダを具え、第 2 のパケットは、第 1 のパケット中で識別されたオブジェクトに関連することができ、オフセット情報は、オブジェクト内の第 2 のパケットおよびデータを配置する。このソリッドステートストレージデバイス制御部 202 は、バンク 214 と、パケットがストリーミングされる物理領域を管理する。

【0179】

ECC ジェネレータ 304 は、パケットタイザ 302 からのパケットを受信し、データパケット用の ECC を生成する (508)。通常、パケットと ECC ブロックとの間には固定した関連性はない。ECC ブロックは、1 又はそれ以上のパケットを具えることができる。パケットは、1 又はそれ以上の ECC ブロックを具えることができる。パケットは、ECC ブロック内のいずれの場所でもスタートおよび終了することができる。パケットは、第 1 の ECC ブロック内のいずれの場所からスタートすることができ、連続する ECC ブロックのいずれの場所で終了することができる。

【0180】

書き込み同期バッファ 308 は、ECC ブロックをソリッドステートストレージ 110 に書き込む前に、対応する ECC ブロックに分散するように、パケットをバッファリングし (510)、次いで、ソリッドステートストレージ制御部 104 は、クロックドメイン差を考慮した適宜の時間でデータを書き込むことができ (512)、方法 500 は終了する (514)。書き込み同期バッファ 308 は、ローカルクロックドメインとソリッドステートストレージ 110 クロックドメインとの間のバウンダリに配置される。なお、この方法 500 は、便宜上、1 又はそれ以上のデータセグメントを受信するステップと、1 又はデータパケットを書き込むステップと、を表しているが、通常、データセグメントのストリームおよびグループが受信される。通常、ソリッドステートストレージ 110 の完全バーチャルページを具える多数の ECC ブロックが、ソリッドステートストレージ 110 に書き込まれる。通常、パケットタイザ 302 は、1 サイズのデータセグメントを受信し、

別のサイズの packets を生成する。これは、データ若しくはメタデータセグメント、または、データ若しくはメタデータセグメントの一部が、セグメントのデータ全てが packets にキャプチャーされるように、データ packets を形成するためにコンバインされる必要がある。

【0181】

図6は、本発明によるデータパイプラインを使用してソリッドステートストレージデバイス102のデータを管理する方法600の別の実施例を示した概略フローチャート図である。この方法600はスタートし(602)、入力バッファ306が、ソリッドステートストレージ110に書き込まれる1又はそれ以上のデータまたはメタデータセグメントを受信する(604)。パケットサイズ302は、ヘッダを各 packet に加え、各 packet は、オブジェクト内の packet 長を具える。パケットサイズ302は、入力バッファ306に記憶された1又はそれ以上のセグメントを受信し(604)、ソリッドステートストレージ110用にサイズを調整された1又はそれ以上の packet を作成することによって、1又はそれ以上のセグメントを packet 化し(606)、各 packet は、ヘッダと、1又はそれ以上のセグメントからのデータとを具える。

【0182】

ECCジェネレータ304は、パケットサイズ302からの packet を受信し、packet 用の1又はそれ以上のECCブロックを生成する(608)。書き込み同期バッファ308は、ECCブロックをソリッドステートストレージ110に書き込む前に、対応するECCブロック内に分散されるように packet をバッファリングし(610)、次いで、ソリッドステートストレージ制御部104は、クロックドメイン差を考慮して適宜な時間でデータを書き込む(612)。データが、ソリッドステートストレージ110から要求されるとき、1又はそれ以上のデータ packet を具えるECCは、読み取り同期バッファ328に読み取られ、バッファリングされる(614)。packet のECCブロックは、ストレージI/Oバス210を介して受信される。ストレージI/Oバス210は2方向性であるので、データが読み取られるとき、書き込みオペレーション、コマンドオペレーションが中止される。

【0183】

ECCコレクションモジュール322は、読み取り同期バッファ328で保持される要求された packet のECCブロックを受信し、必要に応じて各ECCブロック内のエラーを修正する(616)。ECC修正モジュール322は、ECCブロックに1又はそれ以上のエラーが存在し、そのエラーがECCシンドロームを使用して修正可能であるか判定し、ECC修正モジュール322が、ECCブロックのエラーを修正する(616)。ECC修正モジュール322が、検出されたエラーがECCを用いて修正できないと判定した場合は、ECC修正モジュール322は、中断信号を送る。

【0184】

デパケットサイズ324は、ECCコレクションモジュール322がエラーを修正した後、要求された packet を受信し(618)、各 packet の packet ヘッダをチェックおよび除去することによって packet を脱 packet 化する(618)。アライメントモジュール326は、脱 packet 化後、 packet を受信し、不要なデータを除去し、セグメントまたはオブジェクトを要求するデバイスと適合するフォームにおいてオブジェクトのデータまたはメタデータセグメントとしてデータ packet をリフォーマットする(620)。出力バッファ330は、脱 packet 化後要求される packet を受信し、要求デバイス155に転送する前に packet をバッファリングし(622)、この方法600は終了する(624)。

【0185】

図7は、本発明によるバンクインターリーブを使用して、ソリッドステートストレージデバイス102のデータを管理する方法700に関する一実施例を示した概略フローチャート図である。この方法700は、スタートし(702)、バンクインターリーブ制御部344は、1又はそれ以上のコマンドを2又はそれ以上のキュー410、412、414

、416に指令する(704)。通常、エージェント402、404、406、408は、コマンドタイプによってコマンドをキュー410、412、414、416に指令する(604)。キュー410、412、414、416の各セットは、各コマンドタイプのキューを具える。バンクインターリーブ制御部344は、バンク214中で、キュー410、412、414、416に記憶されたコマンドの実行を調整し(706)、第1のタイプのコマンドは、1つのバンク214aで実行され、第2のタイプのコマンドは、第2のバンク214bで実行され、この方法700は終了する(708)。

【0186】

ストレージスペースリカバリ

図8は、本発明によるソリッドステートストレージデバイス102の不要データコレクション装置800に関する一実施例を示した概略ブロック図である。この装置800は、シークエンス式ストレージモジュール802、ストレージディビジョン選択モジュール804、データリカバリモジュール806、および、ストレージディビジョンリカバリモジュール808を具え、これらは以下で説明される。他の実施例においては、この装置800は、不要データマーキングモジュール812および消去モジュール810を具える。

【0187】

装置800は、ストレージディビジョン内におけるページにデータバケットをシークエンシャルに書き込むシークエンス式ストレージモジュール802を具える。バケットは、バケットが新規バケットまたは修正されたバケットであるかにかかわらずシークエンシャルに記憶される。一実施例においては、シークエンス式ストレージモジュール802は、ストレージディビジョンのページ内の第1のロケーションにバケットを書き込み、次いで、ページの次のロケーションにバケットを書き込み、それから、次、次と、ページが埋まるまで続けられる。次いで、シークエンス式ストレージモジュール802は、ストレージディビジョン内の次のページを埋めるようにスタートする。これは、ストレージディビジョンが埋まるまで繰り返される。

【0188】

好適な実施例においては、シークエンス式ストレージモジュール802は、バンク(バンク-0 214a)のストレージエレメント(例えば、SSS0.0からSSSM.0 216)のストレージ書き込みバッファにバケットを書き込むことを開始する。ストレージ書き込みバッファがフルの場合、ソリッドステートストレージ制御部104は、ストレージ書き込みバッファのデータを、バンク214aのストレージエレメント216内の指定されたページにプログラムすることができる。次いで、別のバンク(例えば、バンク-1 214b)が選択され、シークエンス式ストレージモジュール802は、バンク214bのストレージエレメント218のストレージ書き込みバッファにバケットを書き込むことを開始し、第1のバンク-0 214aは、指定されたページをプログラミングする。このバンク214bのストレージ書き込みバッファがフルの場合、ストレージ書き込みバッファの内容は、各ストレージエレメント218の別の指定ページにプログラムされる。1つのバンク214aをページにプログラムする間に、別のバンク214bのストレージ書き込みバッファが埋められる可能性があるので、このプロセスは有効である。

【0189】

ストレージディビジョンは、ソリッドステートストレージデバイス102のソリッドステートストレージ110の一部を具える。通常、ストレージディビジョンは消去ブロックである。フラッシュメモリに関して、消去ブロックの消去オペレーションは、各セルをチャージングすることによって、1を消去ブロックの全ビットに書き込む。これは、ロケーションが全て1として開始するプログラムオペレーションと比較して、冗長的なプロセスであり、データが書き込まれるとき、いくつかのビットは、ゼロとして書き込まれたセルをディスチャージすることによってゼロに変更される。なお、ソリッドステートストレージ110がフラッシュメモリでないか、あるいは、消去サイクルが、読み取りまたはプログラミングなど、その他のオペレーションとして同様の時間を費やすフラッシュメモリを有する場合は、ストレージディビジョンは、消去されるのに必要ではない。

【0190】

本明細書で使用されるように、ストレージディビジョンは、消去ブロックに対する領域と同等であるが、消去されても消去されなくてもよい。本明細書で消去ブロックが使用される場合、消去ブロックは、ストレージエレメント（例えば、SSS0.0 216a）内の指定されたサイズの特定期間を指し、通常、所定量のページを具える。「消去ブロック」は、フラッシュメモリと関連して使用されるとき、通常、書き込まれる前に消去されるストレージディビジョンである。「消去ブロック」は、「ソリッドステートストレージ」とともに使用される場合、消去されても消去されなくてもよい。本明細書で使用されるように、消去ブロックは、1つの消去ブロック、または、ストレージエレメント（例えば、SSS0.0からSSSM.0 216a-n）の各列の消去ブロックを有する消去ブロック群、を具えることができ、消去ブロックは、バーチャル消去ブロックも意味する。バーチャル消去ブロックと関連した論理コンストラクトを指す場合、消去ブロックは、論理消去ブロック（「LEB」）も意味する。

【0191】

通常、パケットは、処理の順番によって、シークエンシャルに記憶される。一実施例においては、書き込みデータパイプライン106が使用される場合、シークエンス式ストレージモジュール802は、パケットが書き込みデータパイプライン106からくる順番でパケットを記憶する。有効データは、下記で説明するように、リカバリオペレーション中ストレージディビジョンからリカバリされるので、上記順番は、別のストレージディビジョンから読み取られる有効データのパケットと混合した、要求デバイス155からきたデータセグメントの結果でもよい。リカバリされた有効データパケットを書き込みデータパイプライン106に再ルーティングすることは、図3のソリッドステートストレージ制御部104に関連して上述したように、不要データコレクタバイパス316を具えてもよい。

【0192】

装置800は、リカバリ用ストレージディビジョンを選択するストレージディビジョン選択モジュール804を具える。リカバリ用ストレージディビジョンを選択することで、データを書き込むシークエンス式ストレージモジュール802によってストレージディビジョンを再利用することができ、従って、リカバリされたストレージディビジョンをストレージプールに加えるか、または、ストレージディビジョンが使用不能、信頼性がない、リフレッシュすべき、あるいは、ストレージプールから一時的または永続的にストレージディビジョンをとると判定した後、ストレージディビジョンから有効データをリカバリする。別の実施例においては、ストレージディビジョン選択モジュール804は、不要データの多いストレージディビジョンまたは消去ブロックを識別することによって、リカバリ用ストレージディビジョンを選択する。

【0193】

別の実施例においては、ストレージディビジョン選択モジュール804は、摩耗の少ないストレージディビジョンまたは消去ブロックを識別することによってリカバリ用ストレージディビジョンを選択する。例えば、摩耗の少ないストレージディビジョンまたは消去ブロックを識別することは、不要データの少ない、消去サイクルの数が少ない、ビットエラー率の低い、または、プログラムカウンタの低い（回数の低い、バッファ中のページデータは、ストレージディビジョンのページに書き込まれ；プログラムカウンタは、デバイスが製造されたときから、ストレージディビジョンが最後に消去されたときから、その他の任意のイベントから、および、これらの組み合わせから、測定されてもよい）ストレージディビジョンを識別することを含む。このストレージディビジョン選択モジュール804は、摩耗量の少ないストレージディビジョンを判定するために、上記パラメータまたは他のパラメータの組み合わせを使用することができる。摩耗量の低いストレージディビジョンを判定することによって、リカバリ用ストレージディビジョンの選択は、使用されているストレージディビジョンを見つけるのに望ましく、摩耗レベルなどに合わせてリカバリ可能である。

【 0 1 9 4 】

別の実施例においては、ストレージディビジョン選択モジュール 8 0 4 は、摩耗量の多いストレージディビジョンまたは消去ブロックを識別することによってリカバリ用ストレージディビジョンを選択する。例えば、摩耗量の多いストレージディビジョンまたは消去ブロックを識別することは、消去サイクル数の多いストレージディビジョン、ビットエラー率の高いストレージディビジョン、リカバリ不可能 E C C ブロックを有するストレージディビジョン、または、プログラムカウントの高いストレージディビジョンを識別することを含む。さらに、ストレージディビジョン選択モジュール 8 0 4 は、摩耗量の多いストレージディビジョンを判定する上記パラメータまたは他のパラメータの組み合わせを利用することができる。摩耗量の多いストレージディビジョンを判定することによってリカバリ用ストレージディビジョンを選択することは、過剰使用されているストレージディビジョンを見つけるのに望ましく、消去サイクルなどを使用してストレージディビジョンをリフレッシュすることによってリカバリ可能であり、または、利用不能としてサービスからストレージディビジョンを引き下げる (r e t i r e) ことができる。

【 0 1 9 5 】

装置 8 0 0 は、データリカバリモジュール 8 0 6 を具え、このデータリカバリモジュール 8 0 6 は、リカバリ用に選択されたストレージディビジョンから有効データバケットを読み取り、他のデータバケットを有する有効データバケットをキューイングし、シークエンس式ストレージモジュール 8 0 2 によってシークエンシャルに書き込み、シークエンス式ストレージモジュール 8 0 2 によって書き込まれた有効データの新規物理アドレスを有するインデックスをアップデートする。通常、このインデックスは、データオブジェクトから由来するバケットがソリッドステートストレージ 1 1 0 に記憶される物理アドレスに、オブジェクトのデータオブジェクト識別子をマッピングするオブジェクトインデックスである。

【 0 1 9 6 】

一実施例においては、装置 8 0 0 は、ストレージディビジョンリカバリモジュール 8 0 8 を具え、このストレージディビジョンリカバリモジュール 8 0 8 は、使用または再使用のためのストレージディビジョンを準備し、データリカバリモジュール 8 0 6 がストレージディビジョンから有効データをコピーするのが完了した後、データバケットをシークエンシャルに書き込むために、利用可能なストレージディビジョンをシークエンス式ストレージモジュール 8 0 2 に標識する。別の実施例においては、装置 8 0 0 は、データを記憶するために利用不可能なとき、リカバリ用に選択されたストレージディビジョンを標識するストレージディビジョンリカバリモジュール 8 0 8 を具える。通常、これは、ストレージディビジョンまたは消去ブロックが、信頼性のあるデータストレージ用に使用される状況ではないと、摩耗量の多いストレージディビジョンまたは消去ブロックを識別するストレージディビジョン選択モジュール 8 0 4 によるものである。

【 0 1 9 7 】

一実施例においては、装置 8 0 0 は、ソリッドストレージデバイス 1 0 2 のソリッドストレージデバイス制御部 2 0 2 中にある。別の実施例においては、装置 8 0 0 は、ソリッドステートストレージデバイス制御部 2 0 2 を制御する。別の実施例においては、装置 8 0 0 の一部が、ソリッドステートストレージデバイス制御部 2 0 2 中にある。別の実施例においては、データリカバリモジュール 8 0 6 によってアップデートされたオブジェクトインデックスは、ソリッドステートストレージデバイス制御部 2 0 2 中に配置される。

【 0 1 9 8 】

一実施例においては、ストレージディビジョンは、消去ブロックであり、装置 8 0 0 は、データリカバリモジュール 8 0 6 が、選択された消去ブロックから有効データバケットをコピーした後で、ストレージディビジョンリカバリモジュール 8 0 8 が、利用可能として消去ブロックを標識する前に、リカバリ用に選択された消去ブロックを消去する消去モジュール 8 1 0 を具える。読み取りオペレーションまたは書き込みオペレーションよりも大幅に長い時間を要する消去オペレーションを有するフラッシュメモリおよびその他のソ

リッドストレートストレージに関して、新規データを書き込むことを可能にする前にデータブロックを消去することは、効率的なオペレーションとして望ましい。ソリッドステートストレージ 110 がバンク 214 中に配置される場合、消去モジュール 810 による消去オペレーションは、1つのバンクで実行され、その他のバンクは、読み取りオペレーション、書き込みオペレーションまたはその他のオペレーションを実行する。

【0199】

一実施例においては、装置 800 は、不要データ標識モジュール 812 を具え、このモジュール 812 は、データパケットがもはや有効ではないことを示すオペレーションに応じて、ストレージディビジョン中のデータパケットが無効であることを識別する。例えば、データパケットが削除される場合、不要データ標識モジュール 812 は、データパケットが無効であることを識別できる。読み取り・修正・書き込みオペレーションは、無効であるとして識別されるデータパケットに関する別の方法である。一実施例においては、不要データ標識モジュール 812 は、インデックスをアップデートすることによって、無効であるデータパケットを識別することができる。別の実施例においては、不要データ標識モジュール 812 は、無効なデータパケットが削除されたことを示す別のデータパケットを記憶することによって、データパケットが無効であることを識別できる。このことが有利であるのは、ソリッドステートストレージ 110 中に、データパケットが削除されたという情報を記憶することによって、オブジェクトインデックス再構築モジュール 272 または同様のモジュールが、無効データパケットが削除されたことを示す項目を有するオブジェクトインデックスを再構築することができることである。

【0200】

一実施例においては、装置 800 は、性能全体を改善するために、フラッシュコマンドに続いて、バーチャルページデータの残りの部分を埋めるのに使用でき、フラッシュコマンドは、書き込みパイプライン 106 が空になり、全てのパケットが、非揮発性ソリッドステートストレージ 110 内に書き込まれるまで、書き込みパイプライン 106 にデータが流れるのを中止する。これは、必要とされる不要データコレクションの量、ストレージディビジョンを消去するのにかかる時間、バーチャルページをプログラムするのに必要な時間を低減することができる利点を有する。例えば、ソリッドステートストレージ 110 のバーチャルページ内に書き込むために 1つの小さなパケットのみが準備されるときに、フラッシュコマンドが受信される場合がある。このようにほぼ空のバーチャルページをプログラムすることで、無駄なスペースをすぐにリカバリする必要性が生じ、ストレージディビジョン内の有効データを、収集された不必要な不要データにして、消去される、リカバリされる、および、シークエン스式ストレージモジュール 802 によって書き込む利用可能スペースのプールにリターンされる、ストレージディビジョンにする。

【0201】

不要データパケットを実際に消去するよりも、データパケットを無効として標識する方が有効であるのは、上述したように、フラッシュメモリおよび他の同様のストレージに関して、消去オペレーションが大幅に時間をとるからである。装置 800 に記載されるように、不要データコレクションシステムが、ソリッドステートストレージ 110 内で自律的に機能できることで、読み取りオペレーション、書き込みオペレーションおよびその他のより短時間のオペレーションから消去オペレーションを分ける方法を提供することができる。ソリッドステートストレージデバイス 102 は、他の多くのソリッドステートストレージシステムまたはストレージデバイスよりも迅速に機能することができる。

【0202】

図 9 は、本発明によるストレージリカバリのための方法 900 に関する一実施例を示す概略フローチャート図である。方法 900 はスタートし (902)、シークエン式ストレージモジュール 802 は、ストレージディビジョン中にデータパケットをシークエンシャルに書き込む (904)。ストレージディビジョンは、ソリッドステートストレージデバイス 102 中のソリッドステートストレージ 110 の一部である。通常、ストレージディビジョンは消去ブロックである。データパケットは、オブジェクト由来であり、データ

パケットは、処理の順番によってシークエンシャルに記憶される。

【0203】

ストレージディビジョン選択モジュール804は、リカバリ用ストレージディビジョンを選択し(906)、データリカバリモジュール806は、リカバリ用に選択されたストレージディビジョンから有効データパケットを読み取る(908)。通常、有効データパケットは、消去または削除が標識されていないデータパケットであるか、標識されている他の不要データであり、有効または「優良」データとみなされる。データリカバリモジュール806は、シークエンス式ストレージモジュール802によってシークエンシャルに書き込まれるようにスケジューリングされた他のデータパケットを有する有効データパケットをキューイングする(910)。データリカバリモジュール806は、シークエンス式ストレージモジュール802によって書き込まれる有効データの新規物理アドレスを有するインデックスをアップデートする(912)。このインデックスは、オブジェクト識別子に対するデータパケットの物理アドレスのマッピングを具える。データパケットは、ソリッドステートストレージ110中に記憶される際に記憶されるものであり、オブジェクト識別子は、データパケットに対応する。

【0204】

データリカバリモジュール806は、ストレージディビジョンから有効データをコピーし終えてから、ストレージディビジョンリカバリモジュール808は、データパケットをシークエンシャルに書き込むシークエンス式ストレージモジュール802が利用可能であるとして、リカバリ用に選択されたストレージディビジョンを標識し(914)、方法900は終了する(916)。

【0205】

空データセグメント指令

通常、データがもはや有用できないとき、そのデータは消去可能である。多くのファイルシステムでは、消去コマンドは、ファイルシステム中のディレクトリ項目を削除し、データを含むストレージデバイスの所定位置にあるデータをそのままに残しておく。通常、データストレージデバイスは、この種の消去オペレーションに含まれない。データを消去する別の方法は、ゼロ、1またはその他のヌルデータ文字をデータストレージデバイスに書き込み、実際に消去されたファイルに置換することである。しかしながら、これが非効率なのは、上書きするデータを転送している間、有効なバンド幅が使用されるからである。さらに、ストレージデバイス中のスペースは、不要データを上書きするのに使用されるデータによって占有される。

【0206】

ここで記載されるソリッドステートストレージデバイス102のようないくつかのストレージデバイスにおいては、追記型及びガーベジコレクション型のストレージデバイスではないので、既に記憶されたデータをアップデートすることで既存のデータを上書きすることはない。上記デバイス上で、1のストリングまたはゼロのストリングを有するデータを上書きする試みは、既存データを上書きする所望の意図を完了せずに、有用なスペースを埋める。ソリッドステートストレージデバイス102など、追記型及びガーベジコレクション型のデバイスについては、クライアント114は、通常、データを消去するためにデータを上書きする能力を有しない。

【0207】

反復文字のストリングまたは文字ストリングを受信するときに、受信したデータは、高度に圧縮可能であるが、通常、圧縮は、ストレージデバイスに転送する前に、ファイルシステムによってなされる。通常のストレージデバイスは、圧縮データと未圧縮データとの区別をつけることはできない。さらに、ストレージデバイスは、消去されたファイルを読み取るコマンドを受信することができ、ストレージデバイスは、ゼロ、1またはヌル文字のストリームを要求デバイスに転送することができる。再度となるが、バンド幅は、消去されたファイルを示すデータを転送するのに必要である。

【0208】

上述した説明から、ストレージデバイスが、空データセグメントまたは反復文字若しくは文字ストリングを有するデータを表すデータセグメントトークンを記憶することができるように、データを消去すべきという指令を受信するストレージデバイスのための装置、システムおよび方法に関する必要性が存在することは明らかである。装置、システムおよび方法は、さらに、既存データを消去することができ、得られる使用されたストレージスペースは、スモールデータセグメントトークンをもつ。当分野の欠点のいくつかまたは全部を克服する装置、システムおよび方法が示される。

【0209】

図10は、本発明によるトークン指令を生成する装置を有するシステム1000の一実施例を示した概略ブロック図である。この装置は、トークン指令生成モジュール1002、トークン指令転送モジュール1004、読み取りレシーバモジュール1006、読み取り要求転送モジュール1008、読み取りトークン指令レシーバモジュール1010、要求クライアント応答モジュール1012およびセグメント再生成モジュール1014を具え、これらは以下で説明する。一実施例においては、この装置は、ストレージ制御部152およびデータストレージデバイス154を具えるストレージデバイス150に接続したサーバ112内にあり、これらは、上述したものと実質的に同等である。

【0210】

一実施例においては、この装置は、トークン指令を有するストレージ要求を生成するトークン指令生成モジュール1002を具える。トークン指令は、ストレージデバイス150上のデータセグメントトークンを記憶する要求を具える。トークン指令は、ストレージデバイス150に送られて、データセグメントトークンが所定位置に送られていなかった場合にデータセグメントとして記憶される、一連の反復同一文字、または、一連の反復同一文字ストリングに置換されるように意図される。一実施例においては、一連の反復同一文字は、データセグメントが空であることを示唆する。例えば、一連の反復同一文字は、ゼロまたは1であってもよく、ゼロまたは一で埋められたデータはセグメントは、空として解釈される。

【0211】

トークン指令は、少なくともデータセグメント識別子およびデータセグメント長を具える。データセグメント識別子は、通常、オブジェクトID、ファイル名であり、あるいは、ストレージデバイスに反復同一文字または文字ストリングを記憶するためにシークするときの、ファイルシステム、アプリケーション、サーバ112などに対するその他の周知な他の識別子である。データセグメント長は、一連の反復同一文字または文字ストリングによって求められるストレージスペースである。データセグメントトークンおよびトークン指令は、通常、一連の反復同一文字など、データセグメントのデータを具えていない。

【0212】

しかしながら、トークン指令は、反復同一文字または文字ストリングの少なくとも1つのインスタンスなど、データセグメントトークンを形成する他の関連情報を具える。さらに、トークン指令は、データセグメントロケーション、ファイルシステムからのアドレス、データセグメントに対応するデータストレージデバイスのロケーションなど、メタデータを具えることができる。当分野の当業者であれば、トークン指令とともに含むことができるその他の情報を認識されたい。一実施例においては、トークン指令生成モジュール1002は、トークン指令とともにデータセグメントトークンを生成する。

【0213】

一実施例においては、トークン指令生成モジュール1002は、ストレージデバイス150にある既存データを上書きする要求に回答して、トークン指令および安全消去コマンドの双方を生成する。既存データは、トークン指令におけるデータセグメント識別子と同一のデータセグメント識別子を有するストレージデバイス150で識別されたデータを具える。通常、データを上書きする要求は、無効または不要データとしてデータを単に標識するか、データまたは他の典型的な削除オペレーションに対するポイントを削除するのに十分ではないが、データは、リカバリ不可能なように上書きされるように要求されるとこ

るに送られる。例えば、データを上書きする要求は、データが機密情報と考えられ、安全上の理由から破壊されなければならないところで要求される。

【0214】

安全消去コマンドは、ストレージデバイス150が既存データを上書きするように指令し、既存データはリカバリ不可能となる。次いで、ストレージデバイス150は、既存データを上書き、リカバリ、消去するとともに、データセグメントトークンを作成する。結果として、既存データは、リカバリ不可能であり、データセグメントトークンは、ストレージデバイス150に記憶され、ストレージスペースは、通常、既存データよりも大幅に少ない。

【0215】

さらなる実施例においては、装置は、消去確認モジュール1016を具え、この消去確認モジュール1016は、ストレージデバイス150の既存データが、文字とともに上書きされたことの確認信号を受信する。この確認信号は、要求デバイスまたはクライアント114に送られ、既存データがリカバリ不能である条件におかれたことを確認するために使用可能である。他の実施例においては、安全消去コマンドは、ストレージデバイス150が、特定の文字または文字ストリングを有する既存データを上書きするように指令でき、実行コマンドは、複数回実行することができる。当分野の当業者であれば、既存データがリカバリ不能であることを確認するために、1又はそれ以上の安全消去コマンドを設定する他の方法を理解されたい。

【0216】

データは、暗号化可能であり、次いで、ストレージデバイス150に記憶され、データを記憶することと関連してストレージデバイス150が受信する暗号化キーを使用して暗号化がなされる。既存データが、記憶される前に受信された符号化キーを用いて暗号化される場合、別の実施例においては、トークン指令生成モジュール1002は、既存データを上書きするために要求を受信するのに応じて、トークン指令を生成することともに暗号化消去コマンドを生成する。暗号化消去コマンドは、既存データを記憶するのに使用される暗号化キーを消去し、暗号化キーはリカバリ不能になる。

【0217】

一実施例においては、暗号化キーを消去することは、要求デバイスから暗号化キーを消去することを具える。別の実施例においては、暗号化キーの消去は、サーバ、キー保管場所(vault)、キーが記憶されるその他のロケーションを具える。暗号化キーを消去することは、暗号化キーは、いずれの方法でもリカバリすることができないように、他のデータまたは一連の文字を用いて暗号化キーを置換することを含むことができる。通常、暗号化ルーチンが、既存データを解読する試みを邪魔するのに十分強固である既存データを暗号化するのに使用された場合、暗号化キーを消去することは、ストレージデバイス150の既存データをリカバリ不能にする。既存データを上書きする要求は、データを安全上の理由から上書きする安全消去指令、データを消去するためのデータを上書きする要求、または、反復同一文字または文字ストリングで既存データを置換するようにシークする要求などでもよい。一実施例においては、安全消去指令は、デバイスに、暗号化キーを安全に消去させ、既存データを安全に消去させることを両方行う。一実施例においては、暗号化キーの消去によって、不要データコレクションプロセスがストレージスペースリカバリプロセスの一部としてデータを消去するまで、ストレージデバイス150のデータの安全消去を延期することが可能となる。当分野の当業者であれば、暗号化キーを消去するその他の方法、および、既存データを上書きする要求を受信するその他の方法を理解されたい。

【0218】

一実施例においては、トークン指令は、データセグメントトークンを具え、トークン指令転送モジュール1004は、トークン指令とともにデータセグメントトークンを転送する。一実施例においては、トークン指令は、データセグメントトークンを具えずに、データセグメントを生成するためのストレージデバイス150用のコマンドを具える。この実

施例においては、トークン指令転送モジュール 1004 は、データセグメントトークンを生成するコマンドを有するトークン指令を転送し、データセグメントトークンを転送しない。

【0219】

この装置は、ストレージデバイス 150 にトークン指令を転送するトークン指令転送モジュール 1004 を具える。通常、トークン指令転送モジュール 1004 は、ストレージ要求の一部としてトークン指令を転送する。ストレージ要求は、オブジェクト要求フォーム、データ要求フォームまたは当分野の当業者に周知の他のフォームであってもよい。トークン指令生成モジュール 1002 は、安全消去指令を生成し、トークン指令転送モジュール 1004 は、安全消去指令をストレージデバイス 150 に転送する。トークン指令生成モジュール 1002 は、消去暗号化キーコマンドを生成し、必要がある場合には、コマンドを実行する別のデバイスに消去暗号化キーコマンドが転送される。

【0220】

一実施例においては、トークン指令転送モジュール 1004 は、データセグメントトークンなしでトークン指令を転送する。この実施例においては、トークン指令は、データセグメントトークンを作成するための、ストレージデバイス 150 に関する指令及び情報を具える。別の実施例においては、トークン指令転送モジュール 1004 は、データセグメントトークンを具えるトークン指令を転送する。この実施例においては、ストレージデバイス 150 は、トークン指令を受信したデータセグメントが、データセグメントを表し、適宜なアクションをとり、データセグメントトークンを記憶し、データセグメントトークンは、通常のデータとしてデータセグメントトークンを単に記憶するのではなく、データセグメントを表す。

【0221】

特定の実施例においては、装置は、ストレージデバイス 150 と、ストレージデバイス 150 にストレージ要求を転送する読み取り要求転送モジュール 1008 と、ストレージデバイス 150 からセグメントトークンを読み取るストレージ要求を受信する読み取りレシーバモジュール 1006 を具える。通常、ストレージ要求は、外部クライアント 114、サーバ 112 で運転しているアプリケーションまたはファイルサーバなど、サーバ 112 内部のクライアント 114 などの要求クライアント 114 から受信される。当分野の当業者であれば、読み取りレシーバモジュール 1006 がストレージ要求を受信できる要求クライアント 114 として機能するその他のデバイスおよびソフトウェアを理解されたい。

【0222】

ストレージ要求は、トークン指令転送モジュール 1004 によってストレージデバイス 150 に転送されるトークン指令に記憶されるように要求されたデータセグメントトークンに対応する、データセグメントを読み取るための要求 (request) を具える。一実施例においては、要求クライアント 114 は、データセグメントがデータセグメントトークンの形態で記憶されたことを認識しない。別の実施例においては、要求デバイスは、データセグメントが、データセグメントトークンとして記憶されたことを認識するが、データセグメントトークンに記憶された情報を認識しない。

【0223】

特定の実施例においては、装置は、ストレージデバイス 150 から要求されたデータセグメントトークンに対応するメッセージを受信する読み取りトークン指令レシーバモジュール 1010 を具え、このメッセージは、少なくともデータセグメント識別子およびデータセグメント長を具える。このメッセージは、通常、データセグメントのデータを具えない。さらに、このメッセージは、データセグメントロケーションまたは反復同一文字若しくは文字ストリングなど、データセグメントトークンに記憶されたその他の情報を具えることができる。特定の実施例においては、この装置は、ストレージデバイス 150 から受信したメッセージから構築 (formulate) された要求クライアント 113 に、応答信号を転送する要求クライアント応答モジュール 1012 を具える。

【0224】

一実施例においては、読み取りトークン指令レシーバモジュール1010も、メッセージにおいて、既存データが文字で上書きされたことの確認信号を受信し、既存データはリカバリ不能となり、既存データは、ストレージデバイス150に以前に記憶されており、メッセージにおいて、受信されたデータセグメントトークンから同一データセグメント識別子として参照される。この確認信号は、データセグメントを読み取るためにストレージ要求から独立してストレージデバイス150から受信されてもよい。

【0225】

別の実施例においては、要求クライアント114は、データセグメントを要求する場合、この装置は、メッセージ中に含まれる情報を使用し、データセグメントのデータを再構築するデータセグメント再生成モジュール1014を具える。この場合、要求クライアントに送られる応答信号は、再構成されたデータセグメントを含む。別の実施例においては、要求クライアントに送られる応答信号は、ストレージデバイス150から受信したメッセージ維持に含まれる情報を具える。要求クライアント114は、次いで、データセグメントを再構築し、いくつか他の方法において、この情報を使用することができる。別の実施例においては、メッセージは、データセグメントトークンを具える。データセグメントトークンは、要求クライアント114に送る前にデータセグメントを再構築するために、データセグメント再生成モジュール1014によって使用可能であり、あるいは、要求クライアント応答モジュール1012が、データセグメントトークンを単純に送る(f o r w a r d)することができる。

【0226】

一実施例においては、トークン指令を有するストレージ要求は、さらに、ストレージデバイス150にストレージスペースをリザーブする要求を具え、要求されたりリザーブしたストレージスペースは、データセグメント長とほぼ同一のストレージスペース量である。別の実施例においては、ストレージスペースをリザーブする要求は、データセグメント長とは異なるストレージスペース量である。例えば、ストレージデバイス150がソリッドステートストレージデバイス102である場合、ソリッドステートストレージデバイス102は、ハードドライブまたはその他の低価格長期ストレージに接続でき、ソリッドステートストレージ110は、長期ストレージ用キャッシュとして設定される。ストレージをリザーブする要求は、ソリッドステートストレージデバイス102に、ソリッドステートストレージデバイス102にデータを書き込む準備をする際、長期ストレージに対するキャッシュの一部としてフラッシュさせる。当分野の当業者であれば、所望されるストレージスペースをリザーブする要求であるその他の条件を理解されたい。

【0227】

一実施例においては、装置は、読み取りレシーバモジュール1006、読み取り要求転送モジュール1008、読み取りトークン指令レシーバモジュール1010および要求クライアント応答モジュール1012を具えてもよく、これらは、上述したものと実質的に同等なものである。一実施例においては、モジュール1006 - 1012は、トークン指令生成モジュール1002またはトークン指令転送モジュール1004を具える装置から独立していてもよい。一実施例においては、この装置は、データセグメント再生成モジュール1014を具え、これは、上述したデータセグメント再生成モジュール1014と実質的に同じである。

【0228】

図11は、本発明によるトークン指令を生成し転送する方法1100に関する一実施例を示した概略フローチャート図である。この方法1100は、スタートし(1102)、トークン指令生成モジュール1002は、トークン指令を有するストレージ要求を生成し(1104)、トークン指令は、ストレージデバイス150にデータセグメントトークンを記憶する要求を具える。トークン指令転送モジュール1004は、ストレージデバイス150にトークン指令を転送し(1106)、方法1100は終了する(1108)。一実施例においては、ストレージ要求は、データセグメントトークンを記憶するトークン指

令を具え、ストレージ要求は、データセグメントのデータが実質的にない。別の実施例においては、ストレージ要求は、データセグメントからデータを具える。好適な実施例においては、ソフトウェアアプリケーションは、トークン指令を使用してストレージ要求を作成し、データセグメントの作成を回避する。別の実施例においては、ソフトウェアアプリケーションは、トークン指令の作成を要求する。

【0229】

図12は、本発明によるデータセグメントトークンを読み取る方法1200に関する一実施例を示した概略フローチャート図である。この方法1200は、スタートし(1202)、読み取りレシーバモジュール1006は、要求クライアント114から、ストレージデバイス150からのデータセグメントを読み取るストレージ要求を受信する(1204)。読み取り要求転送モジュール1008は、ストレージデバイス150にストレージ要求を転送する(1206)。

【0230】

読み取りトークン指令レシーバモジュール1010は、ストレージデバイス150から、要求されたデータセグメントトークンに対するメッセージを受信し(1208)、このメッセージは、少なくともデータセグメント識別子およびデータセグメント長を具える。このメッセージは、データセグメントのデータが実質的にない。要求クライアント応答モジュール1012は、ストレージデバイス150から受信したメッセージから構築された要求クライアントに応答信号を転送し(1210)、この方法1200は、終了する(1212)。

【0231】

図13は、本発明によるデータセグメントトークンを管理する装置を有するシステム1300に関する一実施例を示した概略ブロック図である。このシステム1300は、書き込み要求レシーバモジュール1302およびデータセグメントトークンストレージモジュール1304を有する装置を具え、各種実施例においては、トークン生成モジュール1306、読み取り要求レシーバモジュール1308、読み取りデータセグメントトークンモジュール1310、転送データセグメントトークンモジュール1314および転送データセグメントモジュール1316を有する読み取り要求応答モジュール1312、再構築データセグメントモジュール1318、安全確認モジュール1322を有する安全消去モジュール1320、ストレージスペースリザーブモジュール1324、を具え、これらは、以下で説明する。システム1300は、ストレージ制御部152およびデータストレージデバイス154を有するストレージデバイス10を具え、これらは、上述したものと実質的に同じである。システム1300は、要求デバイス1326を具え、これは以下に記載されているように、ストレージデバイス150と接続している。

【0232】

描かれた実施例においては、モジュール1302 - 1324は、ストレージデバイス150またはストレージ制御部152に含まれる。別の実施例においては、1又はそれ以上のモジュール1302 - 1324の少なくとも一部分は、ストレージデバイス150の外側に配置される。更なる実施例においては、要求デバイス1326は、ドライバ、ソフトウェアまたは1又はそれ以上のモジュール1302 - 1324のその他の機能の形態において、モジュール1302 - 1324の一部を具える。例えば、トークン生成モジュール1306および再構成データセグメントモジュール1318は、要求デバイス1326中に示されている。当分野の当業者であれば、モジュール1302 - 1324の機能を分配し、実装するその他の方法を理解されたい。

【0233】

この装置は、要求デバイス1326からのストレージ要求を受信する書き込み要求レシーバモジュール1302を具え、ストレージ要求は、ストレージデバイス150内にデータセグメントを記憶する要求を具える。データセグメントは、一連の反復同一文字または文字ストリングを具える。通常、一連の反復同一文字は、データセグメントが空であることを意味する。これは、一連の反復同一文字が1又はゼロである場合に正しい。この装置

は、ストレージデバイス 150 通にデータセグメントトークンを記憶するデータセグメントトークンストレージモジュール 1304 を具える。データセグメントトークンは、少なくともデータセグメント識別子およびデータセグメント長を具える。データセグメントトークンは、データセグメントからの実際のデータが実質的にない。

【0234】

データセグメントトークンは、多数のフォームで記憶可能である。一実施例においては、インデックス中の項目を具えており、このインデックスは、ストレージデバイス 150 に記憶された情報およびデータに対応する。例えば、このインデックスは、図 2 に描かれた装置 200 に関連して上述したように、オブジェクトインデックスであってもよい。さらに、このインデックスは、ファイルシステムインデックス、ブロックストレージインデックス、または、当分野の当業者に周知なその他のインデックスであってもよい。別の実施例においては、データセグメントトークンは、ストレージデバイス 150 に記憶されるメタデータを具え、あるいは、ストレージデバイス 150 に記憶されるメタデータのフォームである。別の実施例においては、セグメントトークンが、ストレージデバイス 150 上のメタデータとして記憶され、インデックス中の項目を具える。当分野の当業者であれば、データセグメントトークンをストレージする他の方法を理解されたい。

【0235】

一実施例においては、ストレージ要求は、データセグメントトークンを記憶するトークン指令を具え、このストレージ要求は、データセグメントのデータが実質的にない。トークン指令は、データセグメントトークンを生成するデータセグメントトークンまたはコマンドを具えることができる。トークン指令は、データセグメントトークンを具えていない場合、データトークンストレージモジュール 1304 は、トークン中の情報からデータセグメントトークンを生成する。トークン指令がデータセグメントトークンを具える場合は、データセグメントトークンストレージモジュール 1304 が、トークン指令中のデータセグメント識別子によって識別されるデータセグメントを表すデータ構造としてデータセグメントを認識し、適宜にデータセグメントトークンを記憶する。

【0236】

通常、データセグメントトークンストレージモジュール 1304 が、データセグメントトークンを認識する場合、データセグメントトークンは、ストレージデバイス 150 に記憶される他のデータとは何らかの方法で区別される。例えば、要求デバイス 1326 は、データを圧縮し、圧縮オブジェクト、ファイルまたはセグメントを送るが、ストレージデバイス 150 は、他のストレージ要求の方法によって受信された他のデータから圧縮データセグメントを区別できない。

【0237】

データセグメントトークンストレージモジュール 1304 は、受信されたデータセグメントがデータセグメントトークンであることを認識する場合、データセグメントトークンストレージモジュール 1304 は、読み取るとき、データセグメントトークンが、データセグメントトークンではなくデータセグメントとして存在するような方法で、データセグメントトークンを記憶することができる。当分野の当業者であれば、受信したデータセグメントトークンが、データセグメントではなくデータセグメントトークンであると認識した後に、データセグメントトークンストレージモジュール 1304 がデータセグメントトークンを記憶することができるような他の方法を理解されたい。

【0238】

別の実施例においては、ストレージ要求は、データセグメントからデータを具える。この実施例においては、この装置は、データセグメントからデータセグメントトークンを生成するトークン生成モジュール 1306 を具え、データセグメントトークンは、データセグメントを記憶するストレージ要求に応答して作成される。更なる実施例においては、トークン生成モジュール 1306 は、可能なドライバフォームにおいて、要求デバイス 1326 中にある。

【0239】

一実施例においては、この装置は、既存データがリカバリ不能であるように、文字を用いて既存データを上書きする安全消去モジュール 1 3 2 0 を具え、既存データは、ストレージ要求で識別されたデータセグメントと同一のデータセグメント識別子として識別された、ストレージデバイス 1 5 0 中に既に記憶されたデータセグメントのデータを具える。この実施例においては、データセグメントトークンは、データセグメント識別子およびデータセグメント長とともに記憶され、データセグメントトークンに記憶された同一のデータセグメント識別子によって識別される既存データは、既存データを上書きすることで消去される。通常、既存文字は、ゼロ、1、または、その他の文字ストリングによって上書きされ、データは破壊され、リカバリ不能となる。

【0 2 4 0】

さらなる実施例においては、安全消去モジュール 1 3 2 0 は、既存データが上書きされることを示すメッセージを転送する消去確認モジュール 1 3 2 2 を具える。通常、メッセージは、要求デバイス 1 3 2 6 に送られる。消去確認メッセージは、安全消去モジュール 1 3 2 0 が既存データを上書きした後に転送される。このメッセージは、ストレージ要求として同一トランザクションにおいて、あるいは、異なるトランザクションにおいて、転送されてもよい。

【0 2 4 1】

別の実施例においては、安全消去モジュール 1 3 2 0 は、ストレージスペースリカバリオペレーション中に、既存データを上書きする。例えば、ストレージデバイス 1 5 0 がソリッドステートストレージデバイス 1 0 2 である場合、上述したように、ストレージスペースリカバリオペレーションは、図 8 に描かされた装置 8 0 0 に関して記載された不要データコレクションに関連してもよい。なお、既存データを上書きする要求を含むストレージスペースリカバリオペレーションは、通常、迅速化されることが可能であり、既存データが記憶されるストレージロケーションが、確認メッセージが消去確認モジュール 1 3 2 2 によって送られる前に、必ずリカバリされる。一実施例においては、既存データは、安全消去が要求されていることを示すように、標識され、あるいは、識別される。通常、確認メッセージは、消去を標識した既存データが上書きされてリカバリ不能になるまで、送られない。別の実施例においては、安全消去モジュール 1 3 2 0 は、後期ストレージスペースリカバリが無効であることを既存データに標識するのみである。別の実施例においては、安全消去は、既存データが無効であることを示すインデックスをアップデートし、後期ストレージスペースリカバリ中にデータが上書きされるまで、このデータへのアクセスを阻止する。

【0 2 4 2】

一実施例においては、安全消去モジュール 1 3 2 0 は、データセグメントが記憶される毎に、既存データを上書きする。別の実施例においては、既存データを上書きする要求を具え、安全消去モジュール 1 3 2 0 は、既存データを上書きする要求に応答して、既存データを上書きする。別の実施例においては、安全消去モジュール 1 3 2 0 は、既存データが消去されたという確認信号に関するメタ情報を記憶し、次の読み取りが消去を指示することができる。

【0 2 4 3】

安全消去が受信された場合の他の実施例において、既存データが削除される。一実施例においては、データを削除することは、インデックス項目、アドレスなどを削除することを含める。好適な実施例においては、データセグメントトークンが記憶されるとき、対応する既存データは、無効あるいはストレージリカバリの準備ができていることを標識される。このデータは、ストレージリカバリオペレーションまたは不要データコレクションオペレーションなどにおいて後でリカバリ可能である。

【0 2 4 4】

特定の実施例においては、この装置は、データセグメントを読み取るためのストレージ要求を受信する読み取り要求レシーバモジュール 1 3 0 8、ストレージ要求によって要求されるデータセグメントに対応するデータセグメントトークンを読み取る読み取りデータ

セグメントトークンモジュール 1310、および、要求デバイス 1326 に応答信号を転送する読み取り要求応答モジュール 1312 を具える。この応答信号は、要求されたデータセグメントに対応するデータセグメントトークンを使用して生成される。

【0245】

一実施例において、データセグメントを読み取るストレージ要求は、ストレージ要求に関連しており、ストレージ要求が成功したことを確認するのに役立つ。別の実施例においては、データセグメントを読み取る要求は、ストレージ要求から独立しており、ストレージ要求を作成した要求デバイス 1326 または別の要求デバイス 1326 によって開始されてもよい。

【0246】

一実施例においては、要求デバイス 1326 が、実際のデータの位置にあるデータセグメントトークンからの情報を受信する場合、読み取り要求応答モジュール 1312 は、要求デバイス 1326 へのメッセージに応答して転送する転送データセグメントトークンモジュール 1314 を具える。このメッセージは、少なくともデータセグメント識別子およびデータセグメント長を具えるが、さらに、データセグメントロケーション、反復同一文字又は文字ストリングの少なくとも 1 つのインスタンス、あるいは、その他の関連情報を具えてもよい。通常、メッセージは、データセグメントトークンに含まれるもの以外で、データセグメントの実際のデータを具えていない。

【0247】

別の実施例においては、要求デバイス 1326 は、データセグメントを受信するとされる場合、この装置は、データセグメントトークンを使用して、データセグメントのデータを再構成する再構成データセグメントモジュール 1318 を具える。読み取り要求応答モジュール 1312 は、さらに、要求デバイス 1326 に再構成要求データセグメントを転送する転送データセグメントモジュール 1316 を具える。別の実施例において、再構成データセグメントモジュール 1318 は、可能であれば、ドライバのフォームにおいて、要求デバイス 1326 に存在し、転送データセグメントトークンモジュール 1314 は、データセグメントトークン情報を有するメッセージを要求デバイス 1326 に送る。再構成データセグメントモジュール 1318 は、要求デバイス 1326 で、メッセージから要求されたデータセグメントを再構成する。

【0248】

一実施例においては、システム 1300 は、読み取り要求レシーバモジュール 1308、読み取りデータセグメントトークンモジュール 1310、読み取り要求応答モジュール 1312 を含む別の装置を具え、これらは、上述したものと実質的に同一である。この装置は、書き込み要求レシーバモジュール 1302 およびデータセグメントトークンストレージモジュール 1304 を具える装置から独立していてもよい。一実施例においては、読み取り要求応答モジュール 1312 は、転送データセグメントトークンモジュール 1314 および / または転送データセグメントモジュール 1316 を具え、この装置は、再構成データセグメントモジュール 1318 を具えてもよく、モジュール 1314、1316、1318 が上述したものと実質的に同じである。

【0249】

図 14 は、本発明によるデータセグメントトークンを記憶する方法 1400 に関する一実施例を示す概略フローチャート図である。この方法 1400 は、スタートし (1402)、書き込み要求レシーバモジュール 1302 が、要求デバイス 1326 からのストレージ要求を受信し (1404)、上記ストレージ要求は、ストレージデバイス 150 にデータセグメントを記憶する要求を具える。データセグメントは、一連の反復同一文字または文字ストリングを具える。データセグメントトークンストレージモジュール 1304 は、ストレージデバイス 150 にデータセグメントトークンを記憶し (1406)、方法 1400 は終了する (1408)。データセグメントトークンは、少なくともデータセグメント識別子およびデータセグメント長を具え、データセグメントトークンは、ほとんどの部分に関してデータセグメントからのデータを具えていない。

【 0 2 5 0 】

図 1 5 は、本発明によるデータセグメントトークンを読み取る方法 1 5 0 0 に関する一実施例を示した概略フローチャート図である。この方法 1 5 0 0 は、スタートし (1 5 0 2)、読み取り要求レシーバモジュール 1 3 0 8 が、ストレージデバイス 1 5 0 からデータセグメントを読み取るストレージ要求を受信する (1 5 0 4)。データセグメントは、データセグメントトークンによってストレージデバイス 1 5 0 に表現され、このデータセグメントは、一連の反復同一文字又は文字ストリングを具える。このデータセグメントトークンは、少なくともデータセグメント識別子およびデータセグメント長を具え、データセグメントトークンは、データセグメントからデータを具えていない。読み取りデータセグメントトークンモジュール 1 3 1 0 は、ストレージ要求において要求されたデータセグメントに対応するデータセグメントを読み取り (1 5 0 6)、読み取り要求応答モジュール 1 3 1 2 は、要求 ストレージデバイス 1 5 0 に応答信号を転送し (1 5 0 8)、この方法 1 5 0 0 は終了する (1 5 1 0)。この応答信号は、要求されたデータセグメントに応答してデータセグメントトークンを使用して生成される。

【 0 2 5 1 】

本発明は、本発明の意図または本質的特徴から逸脱することなく、その他の特定の形態で実施され得る。開示された実施例は、例示のみを目的とするものであって、限定を意図するものではない。本発明の範囲は、従って、上述した説明ではなく、添付の請求の範囲によって規定される。請求の範囲内にあり、これと同等の変更は、本発明の範囲に含まれることとなる。

【 手続補正 3 】

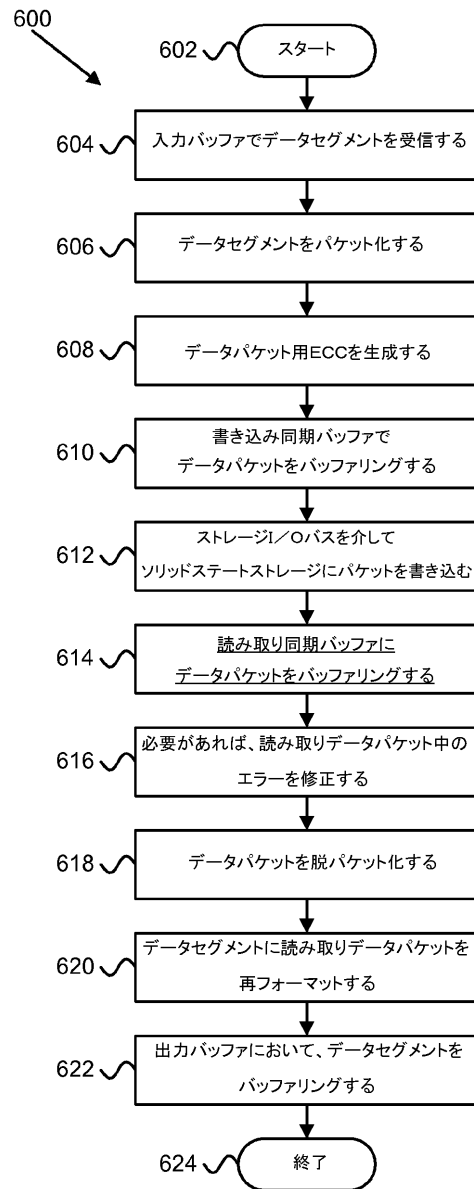
【 補正対象書類名 】 図面

【 補正対象項目名 】 図 6

【 補正方法 】 変更

【 補正の内容 】

【図 6】



【手続補正 4】

【補正対象書類名】図面

【補正対象項目名】図 1 5

【補正方法】変更

【補正の内容】

【図 15】

