

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)公開番号

特開2022-171773

(P2022-171773A)

(43)公開日 令和4年11月11日(2022.11.11)

(51)国際特許分類

F I

G 0 6 F 12/00 (2006.01)

G 0 6 F 12/00 5 6 0 B

G 0 6 F 3/06 (2006.01)

G 0 6 F 12/00 5 9 7 U

G 0 6 F 3/06 3 0 1 M

審査請求 有 請求項の数 7 O L (全37頁)

(21)出願番号 特願2022-142772(P2022-142772)
 (22)出願日 令和4年9月8日(2022.9.8)
 (62)分割の表示 特願2019-13222(P2019-13222)の
 分割
 原出願日 平成31年1月29日(2019.1.29)

(71)出願人 318010018
 キオクシア株式会社
 東京都港区芝浦三丁目1番21号
 (74)代理人 110001737弁理士法人スズエ国際特許
 事務所
 (72)発明者 菅野 伸一
 東京都港区芝浦三丁目1番21号 キオ
 クシア株式会社内

(54)【発明の名称】 メモリシステムおよび制御方法

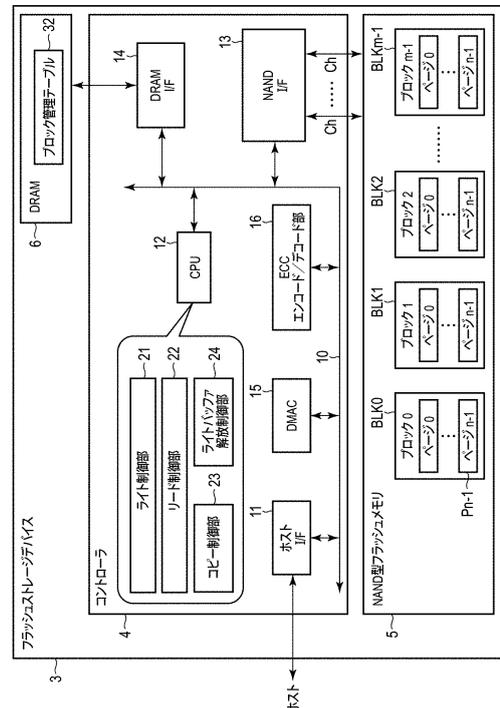
(57)【要約】

【課題】ホスト側の負担を軽減できるメモリシステムを実現する。

【解決手段】メモリシステムは、ホストのライトバッファからライトデータを取得し、書き込み動作を実行する。書き込み動作は、ライトデータを不揮発性メモリに複数回転送することを含む複数回の書き込みステップによって実行される。コントローラは、カウンタ値を、書き込み動作に必要な書き込みステップの回数に1を加えた値に設定し、1回の書き込みステップの終了の度にカウンタ値を1減らし、書き込み動作が終了してライトデータが読み出し可能となった場合、カウンタ値を1減らし、カウンタ値が零になった場合、ライトバッファ内の領域を解放するための解放要求をホストに送信する。

【選択図】図3

図3



【特許請求の範囲】

【請求項 1】

ホストに接続可能なメモリシステムであって、
複数のブロックを含む不揮発性メモリと、
前記不揮発性メモリを制御するコントローラと、を具備し、
前記コントローラは、
ライトデータが格納されている前記ホストのライトバッファ内の記憶位置を示す記憶位置情報を指定するライトコマンドを前記ホストから受信し、
前記記憶位置情報に基づいて前記ライトデータを前記ホストのライトバッファから取得し、

10

前記ライトデータを前記複数のブロックから選択された書き込み先ブロックの書き込み先位置に書き込む書き込み動作を実行し、

前記ライトデータの前記書き込み動作は、前記ライトデータを前記不揮発性メモリに複数回転送することを含む複数回の書き込みステップによって実行され、

前記コントローラは、
カウンタ値を、前記書き込み動作に必要な前記書き込みステップの回数に 1 を加えた値に設定し、

1 回の書き込みステップの終了の度に前記カウンタ値を 1 減らし、

前記書き込み動作が終了して前記ライトデータが読み出し可能となった場合、前記カウンタ値を 1 減らし、

20

前記カウンタ値が零になった場合、前記ライトデータが格納されている前記ライトバッファ内の領域を解放するための解放要求を前記ホストに送信する、
メモリシステム。

【請求項 2】

前記コントローラは、

前記書き込み動作の実行中に、前記ライトデータを読み出し対象データとして指定する第 1 のリードコマンドを受信した場合、前記カウンタ値を 1 増やし、

前記読み出し対象データを前記ライトバッファから読み出す動作と前記読み出した前記読み出し対象データを前記ホストに返す動作とを含む読み出し動作を実行し、

前記読み出し動作が終了した場合、前記カウンタ値を 1 減らす、

30

請求項 1 に記載のメモリシステム。

【請求項 3】

前記コントローラは、前記カウンタ値をライトコマンド毎に設定する、

請求項 1 または請求項 2 に記載のメモリシステム。

【請求項 4】

前記コントローラは、前記カウンタ値を前記不揮発性メモリの最小書き込み単位毎に設定する、

請求項 1 または請求項 2 に記載のメモリシステム。

【請求項 5】

前記コントローラは、前記ライトデータに対応する前記カウンタ値を前記最小書き込み単位毎に複数個設定し、前記ライトデータに対応する複数の前記カウンタ値の全てが零になった場合、前記解放要求を前記ホストに送信する、

40

請求項 4 に記載のメモリシステム。

【請求項 6】

前記書き込み動作がエラーとなった場合、前記カウンタ値を再度の書き込み動作に必要な書き込みステップの回数だけ増やす、

請求項 1 または請求項 2 に記載のメモリシステム。

【請求項 7】

前記コントローラは、前記ホストから受信されたリードコマンドによって指定される読み出し対象データが前記書き込み先ブロックから読み出し可能であり、且つ前記読み出し

50

対象データが前記ライトバッファに存在する場合、前記読み出し対象データを前記書き込み先ブロックから読み出す動作と前記読み出した前記読み出し対象データを前記ホストに返す動作とを含む読み出し動作を実行する、

請求項 1 または請求項 2 に記載のメモリシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は、不揮発性メモリを制御する技術に関する。

【背景技術】

【0002】

近年、不揮発性メモリを備えるメモリシステムが広く普及している。このようなメモリシステムの一つとして、NANDフラッシュ技術ベースのソリッドステートドライブ(SSD)が知られている。

データセンターのサーバにおいても、ストレージデバイスとしてSSDが使用されている。

【0003】

サーバのようなホスト計算機システムにおいて利用されるストレージデバイスにおいては、高いI/O性能が求められている。

このため、最近では、ホストがストレージデバイス内の不揮発性メモリを直接的に制御することを可能にする新たな技術が提案され始めている。

【先行技術文献】

【非特許文献】

【0004】

【非特許文献1】Yiying Zhang, 外, "De-indirection for flash-based SSDs with nameless writes." FAST. 2012, [online], [平成29年9月13日検索], インターネット URL: <https://www.usenix.org/system/files/conference/fast12/zhang.pdf>

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかし、ストレージデバイス内の不揮発性メモリを制御するためのホスト側の負担が増加されると、これによって十分なI/O性能が得られなくなる場合がある。このため、ホスト側の負担を軽減できるようにするための新たな技術の実現が要求される。

本発明が解決しようとする課題は、ホスト側の負担を軽減することができるメモリシステムおよび制御方法を提供することである。

【課題を解決するための手段】

【0006】

実施形態によれば、ホストに接続可能なメモリシステムは、複数のブロックを含む不揮発性メモリと、前記不揮発性メモリを制御するコントローラと、を具備する。前記コントローラは、ライトデータが格納されている前記ホストのライトバッファ内の記憶位置を示す記憶位置情報を指定するライトコマンドを前記ホストから受信する。前記コントローラは、前記記憶位置情報に基づいて前記ライトデータを前記ホストのライトバッファから取得する。前記コントローラは、前記ライトデータを前記複数のブロックから選択された書き込み先ブロックの書き込み先位置に書き込む書き込み動作を実行する。前記ライトデータの書き込み動作は、前記ライトデータを前記不揮発性メモリに複数回転送することを含む複数回の書き込みステップによって実行される。前記コントローラは、カウンタ値を、前記書き込み動作に必要な前記書き込みステップの回数に1を加えた値に設定し、1回の書き込みステップの終了の度に前記カウンタ値を1減らし、前記書き込み動作が終了して前記ライトデータが読み出し可能となった場合、前記カウンタ値を1減らし、前記カウンタ値が零になった場合、前記ライトデータが格納されている前記ライトバッファ内の領

10

20

30

40

50

域を解放するための解放要求を前記ホストに送信する。

【図面の簡単な説明】

【0007】

【図1】ホストと実施形態に係るメモリシステムとの関係を示すブロック図。

【図2】同実施形態のメモリシステムとホストとの間の役割分担を説明するための図。

【図3】同実施形態のメモリシステムの構成例を示すブロック図。

【図4】同実施形態のメモリシステムに設けられた、NANDインタフェースと複数のNAND型フラッシュメモリチップとの関係を示すブロック図。

【図5】複数の物理ブロックの集合によって構築される一つのブロック（スーパーブロック）の構成例を示す図。

【図6】同実施形態のメモリシステムに適用されるライトコマンドを説明するための図。

【図7】同実施形態のメモリシステムからホストに送られるアドレス記録要求を説明するための図。

【図8】同実施形態のメモリシステムに適用されるリードコマンドを説明するための図。

【図9】同実施形態のメモリシステムに適用されるQoSドメインクリエートコマンドを説明するための図。

【図10】同実施形態のメモリシステムからホストに送られるライトバッファ解放要求を説明するための図。

【図11】アドレス記録要求およびリードコマンドの各々に含まれる物理アドレスを規定するブロックアドレスおよびオフセットを説明するための図。

【図12】スーパーブロックが使用される場合におけるブロックアドレスとオフセットとの関係を説明するための図。

【図13】同実施形態のメモリシステムによって管理される複数のQoSドメインを説明するための図。

【図14】ホストと同実施形態のメモリシステムとによって実行されるデータ書き込み処理を説明するためのブロック図。

【図15】同実施形態のメモリシステムによって実行される、複数回の書き込みステップを含む書き込み動作を説明するための図。

【図16】ホストと同実施形態のメモリシステムとによって実行されるデータ読み出し処理を説明するためのブロック図。

【図17】リードコマンドによって指定されたデータの書き込み動作が終了していない場合に、このデータをホストのライトバッファから読み出すデータ読み出し処理を説明するためのブロック図。

【図18】同実施形態のメモリシステムによって管理される、複数のQoSドメインと共通フラッシュブロックプールとの関係を示すブロック図。

【図19】同実施形態のメモリシステムによって管理されるコマンド数カウンタを説明するためのブロック図。

【図20】同実施形態のメモリシステムによって実行されるコマンド数カウンタ更新処理の手順を示すフローチャート。

【図21】同実施形態のメモリシステムによって実行されるデータ読み出し動作の手順を示すフローチャート。

【図22】同実施形態のメモリシステムによって実行されるデータ書き込み動作とライトバッファ解放処理の手順を示すフローチャート。

【図23】リードコマンドによって指定されたデータがフラッシュメモリおよびホストのライトバッファの双方から読み出し可能である場合に実行されるデータ読み出し動作の手順を示すフローチャート。

【図24】リードコマンドによって指定されたデータがフラッシュメモリおよびホストのライトバッファの双方から読み出し可能である場合に実行されるデータ読み出し動作の別の手順を示すフローチャート。

【発明を実施するための形態】

10

20

30

40

50

【 0 0 0 8 】

以下、図面を参照して、実施形態を説明する。

まず、図 1 を参照して、本実施形態に係るメモリシステムとホストとの関係を説明する。

このメモリシステムは、不揮発性メモリにデータを書き込み、不揮発性メモリからデータを読み出すように構成された半導体ストレージデバイスである。このメモリシステムは、NANDフラッシュ技術ベースのフラッシュストレージデバイス 3 として実現されている。

【 0 0 0 9 】

ホスト（ホストデバイス）2 は、複数のフラッシュストレージデバイス 3 を制御するように構成されている。ホスト 2 は、複数のフラッシュストレージデバイス 3 によって構成されるフラッシュアレイをストレージとして使用するように構成された情報処理装置によって実現される。この情報処理装置はパーソナルコンピュータであってもよいし、サーバコンピュータであってもよい。

10

【 0 0 1 0 】

なお、フラッシュストレージデバイス 3 は、ストレージアレイ内に設けられる複数のストレージデバイスの一つとして利用されてもよい。ストレージアレイは、サーバコンピュータのような情報処理装置にケーブルまたはネットワークを介して接続されてもよい。ストレージアレイは、このストレージアレイ内の複数のストレージ（例えば複数のフラッシュストレージデバイス 3）を制御するコントローラを含む。フラッシュストレージデバイス 3 がストレージアレイに適用された場合には、このストレージアレイのコントローラが、フラッシュストレージデバイス 3 のホストとして機能してもよい。

20

【 0 0 1 1 】

以下では、サーバコンピュータのような情報処理装置がホスト 2 として機能する場合を例示して説明する。

ホスト（サーバ）2 と複数のフラッシュストレージデバイス 3 は、インタフェース 5 0 を介して相互接続される（内部相互接続）。この相互接続のためのインタフェース 5 0 としては、これに限定されないが、PCI Express（PCIe）（登録商標）、NVMe Express（NVMe）（登録商標）、Ethernet（登録商標）、NVMe over Fabrics（NVMeOF）等を使用し得る。

30

【 0 0 1 2 】

ホスト 2 として機能するサーバコンピュータの典型例としては、データセンター内のサーバコンピュータ（以下、サーバと称する）が挙げられる。

ホスト 2 がデータセンター内のサーバによって実現されるケースにおいては、このホスト（サーバ）2 は、ネットワーク 6 0 を介して複数のエンドユーザ端末（クライアント）6 1 に接続されてもよい。ホスト 2 は、これらエンドユーザ端末 6 1 に対して様々なサービスを提供することができる。

【 0 0 1 3 】

ホスト（サーバ）2 によって提供可能なサービスの例には、（1）システム稼働プラットフォームを各クライアント（各エンドユーザ端末 6 1）に提供するプラットフォーム・アズ・ア・サービス（PaaS）、（2）仮想サーバのようなインフラストラクチャを各クライアント（各エンドユーザ端末 6 1）に提供するインフラストラクチャ・アズ・ア・サービス（IaaS）、等がある。

40

【 0 0 1 4 】

複数の仮想マシンが、このホスト（サーバ）2 として機能する物理サーバ上で実行されてもよい。ホスト（サーバ）2 上で走るこれら仮想マシンの各々は、この仮想マシンに対応するクライアント（エンドユーザ端末 6 1）に各種サービスを提供するように構成された仮想サーバとして機能することができる。各仮想マシンにおいては、この仮想マシンに対応するエンドユーザ端末 6 1 によって使用される、オペレーティングシステムおよびユーザアプリケーションが実行される。

50

【 0 0 1 5 】

ホスト（サーバ）2 においては、フラッシュトランслーションレイヤ（ホスト F T L ）も実行される。このホスト F T L は、アクセス対象のデータを識別するためのデータ識別子（タグ）それぞれとフラッシュストレージデバイス 3 内の不揮発性メモリの物理アドレスそれぞれとの間のマッピングを管理するアドレス変換テーブルであるルックアップテーブル（L U T ）を含む。ホスト F T L は、この L U T を使用することによって、フラッシュストレージデバイス 3 内の不揮発性メモリ上のデータ配置を知ることができる。

【 0 0 1 6 】

フラッシュストレージデバイス 3 は、N A N D 型フラッシュメモリのような不揮発性メモリを含む。不揮発性メモリは、各々が複数のページを含む複数のブロックを含む。これらブロックの各々はデータを消去する単位であり、複数のページの各々はデータの書き込みおよび読み出しの単位である。

【 0 0 1 7 】

フラッシュストレージデバイス 3 は、ローレベルアブストラクションを実行することができる。ローレベルアブストラクションは不揮発性メモリのアブストラクションのための機能である。ローレベルアブストラクションは、データ配置を補助する機能等を含む。データ配置を補助する機能の例は、ホスト 2 からのユーザデータが書き込まれるべき書き込み先ブロックを割り当てる機能、ユーザデータの書き込み先位置（書き込み先ブロック、この書き込み先ブロック内の位置）を決定する機能、ユーザデータが書き込まれたこの書き込み先位置（ブロックアドレス、オフセット）を上位階層（ホスト 2 ）に通知する機能、等を含んでいてもよい。

【 0 0 1 8 】

フラッシュストレージデバイス 3 は、ホスト 2 から受信される様々なコマンドを実行する。これらコマンドには、フラッシュストレージデバイス 3 内の不揮発性メモリにデータを書き込むためのライトコマンド、不揮発性メモリからデータを読み出すためのリードコマンドが含まれる。本実施形態においては、各リードコマンドは、読み出すべきデータが格納されている記憶位置を示す物理アドレスを指定する。この物理アドレスは、読み出し対象のブロックのブロックアドレスと、この読み出し対象のブロック内の読み出し対象の記憶位置を示すオフセット（ブロック内オフセット）とによって表される。

【 0 0 1 9 】

フラッシュストレージデバイス 3 は、読み出し対象のブロックのブロックアドレスとこの読み出し対象のブロック内の読み出し対象の記憶位置を示すオフセットとを指定するリードコマンドそれぞれをホスト 2 から受信し、受信された各リードコマンドに応じてデータ読み出し動作を実行する。

【 0 0 2 0 】

図 2 はフラッシュストレージデバイス 3 とホスト 2 との間の役割分担を示す。

ホスト（サーバ）2 においては、複数のエンドユーザにそれぞれ対応する複数の仮想マシン 4 0 1 が実行される。各仮想マシン 4 0 1 においては、対応するエンドユーザによって使用される、オペレーティングシステムおよびユーザアプリケーション 4 0 2 が実行される。

【 0 0 2 1 】

また、ホスト（サーバ）2 においては、複数のユーザアプリケーション 4 0 2 にそれぞれ対応する複数の I / O サービス 4 0 3 が実行される。これら I / O サービス 4 0 3 には、論理ブロックアドレス（L B A ）ベースのブロック I / O サービス、キー・バリューストアサービスなどが含まれてもよい。各 I / O サービス 4 0 3 は、タグそれぞれとフラッシュストレージデバイス 3 の物理アドレスそれぞれとの間のマッピングを管理するルックアップテーブル（L U T ）を含む。

【 0 0 2 2 】

ここで、タグとは、アクセス対象のデータを識別可能な識別子を意味する。このタグの典型的な例は、これに限定されないが、L B A のような論理アドレスである。あるいは、

10

20

30

40

50

キー・バリュー・ストアのキー、またはこのキーのハッシュ値、等がタグとして利用されてもよい。

【0023】

フラッシュストレージデバイス3の物理アドレスは、フラッシュストレージデバイス3に含まれる不揮発性メモリ内の記憶位置（物理記憶位置）を特定するためのアドレスである。

LBAベースのブロックI/Oサービスにおいては、論理アドレス（LBA）それぞれとフラッシュストレージデバイス3の物理アドレスそれぞれとの間のマッピングを管理するLUTが使用されてもよい。

【0024】

一方、キー・バリュー・ストアサービスにおいては、キー（またはキーのハッシュ値）それぞれとこれらキーに対応するデータが格納されているフラッシュストレージデバイス3の物理アドレスそれぞれとこれらキーに対応するデータそれぞれのデータ長との間のマッピングを管理するLUTが使用されてもよい。

【0025】

各エンドユーザは、使用すべきアドレッシング方法（LBA、キー・バリュー・ストアのキー、このキーのハッシュ値、等）を選択することができる。

ホスト（サーバ）2においては、複数の仮想マシン401にそれぞれ対応する複数のライトバッファ（WB）404が管理されていてもよい。あるユーザアプリケーション402からのライトデータは、このユーザアプリケーション402に対応する仮想マシン401用のライトバッファ（WB）404に一時的に格納されてもよい。

【0026】

ホスト（サーバ）2からフラッシュストレージデバイス3へのコマンドの送信およびフラッシュストレージデバイス3からホスト（サーバ）2へのコマンド完了のレスポンス等の返送は、ホスト（サーバ）2およびフラッシュストレージデバイス3の各々に存在するI/Oキュー500を介して実行される。

【0027】

フラッシュストレージデバイス3は、フラッシュストレージデバイス3内の不揮発性メモリを論理的に分割することによって得られる複数の領域それぞれをQoSドメイン601として管理する。これらQoSドメイン601の各々は、不揮発性メモリに含まれる複数のブロックの部分集合である。不揮発性メモリに含まれる複数のブロックの各々は一つのQoSドメイン601のみに属し、同じブロックが異なるQoSドメイン601に同時に属することはない。

【0028】

これらQoSドメイン601は、QoSドメインIDと称される識別子によってそれぞれ識別される。これらQoSドメインIDは、これら複数の領域（複数のQoSドメイン）をそれぞれアクセスするための複数の識別子として使用される。

本実施形態においては、各ライトコマンドは、書き込まれるべきデータ（ライトデータ）が格納されているホスト2のライトバッファ（WB）404内の記憶位置を示す記憶位置情報を指定する。また、各ライトコマンドは、データ（ライトデータ）が書き込まれるべき領域（QoSドメイン）の識別子（QoSドメインID）を指定することができる。あるQoSドメインIDを指定するライトコマンドをホスト2から受信した場合、フラッシュストレージデバイス3は、複数のQoSドメイン601によって共有される共通ブロック群602のうちの一つのブロックを選択し、この選択したブロックを、このQoSドメインIDを有するQoSドメイン用の書き込み先ブロックとして割り当てる。

【0029】

ここで、書き込み先ブロックとは、データが書き込まれるべきブロックを意味する。共通ブロック群602は、複数のQoSドメイン601によって共有されるフリーブロックの集合を含む共通フリーブロック群である。

フリーブロックとは、新たな書き込み先ブロックとして再使用（再割り当て）可能な状

10

20

30

40

50

態（フリー状態）のブロックを意味する。フリーブロックの典型的な例は、有効データを格納していないブロックである。有効データとは、L B Aのようなタグに関連付けられている最新のデータを意味する。つまり、ホスト2のL U Tから最新のデータとしてリンクされているデータは有効データである。また、無効データとは、L B Aのようなタグに関連付けられていないデータを意味する。ホスト2のL U Tからリンクされていないデータは無効データである。例えば、あるL B Aに対応する更新データがフラッシュストレージデバイス3に書き込まれた場合には、このL B Aに対応する以前のデータは無効データとなる。

【0030】

そして、フラッシュストレージデバイス3は、受信されたライトコマンドに関連付けられたデータをホスト2のライトバッファ（W B）404から取得し、このデータをQ o Sドメイン用に割り当てられた書き込み先ブロックに書き込む。なお、利用可能な書き込み先ブロックが既にこのQ o Sドメイン用に割り当てられているならば、共通ブロック群602のうちの一つのブロック（フリーブロック）をこのQ o Sドメイン用の書き込み先ブロックとして割り当てるという上述の動作を行う必要は無い。フラッシュストレージデバイス3は、受信されたライトコマンドに関連付けられたデータを、既に割り当てられているこの書き込み先ブロック内の次の利用可能ページに書き込む。

10

【0031】

受信されたライトコマンドに関連付けられたデータが書き込み先ブロックに書き込まれると、フラッシュストレージデバイス3は、このデータのタグと、このデータが書き込まれた不揮発性メモリ内の記憶位置を示す物理アドレス（ブロックアドレス、オフセット）とをホスト2にアドレス記録要求として送出する。ブロックアドレスは、この書き込み先ブロックを識別するための識別子である。通常、不揮発性メモリは複数の不揮発性メモリチップによって構成されるので、あるブロックのブロックアドレスは、不揮発性メモリチップのチップ番号と、このチップ内のブロック番号とによって表されてもよい。オフセットは、この書き込み先ブロック内の記憶位置を示す。

20

【0032】

このアドレス記録要求に応じて、ホスト2は、このデータのタグにこの物理アドレス（ブロックアドレス、オフセット）が関連付けられるようにL U Tを更新することができる。

30

各Q o Sドメインに含まれるブロックの各々について、ホスト2は、各ブロックに格納されているデータが有効データまたは無効データのいずれであるかを、L U Tを使用して管理することができる。また、ホスト2は、既にデータが書き込まれている一つのブロックを新たな書き込み先ブロックとして再使用可能な状態（フリーブロック）に遷移させるためのコマンド（リユースコマンド）をフラッシュストレージデバイス3に送信することができる。

【0033】

図3は、フラッシュストレージデバイス3の構成例を示す。

フラッシュストレージデバイス3は、コントローラ4および不揮発性メモリ（N A N D型フラッシュメモリ）5を備える。フラッシュストレージデバイス3は、ランダムアクセスメモリ、例えば、D R A M 6も備えていてもよい。

40

【0034】

N A N D型フラッシュメモリ5は、マトリクス状に配置された複数のメモリセルを含むメモリセルアレイを含む。N A N D型フラッシュメモリ5は、2次元構造のフラッシュメモリであってもよいし、3次元構造のフラッシュメモリであってもよい。

N A N D型フラッシュメモリ5のメモリセルアレイは、複数のブロックB L K 0 ~ B L K m - 1を含む。ブロックB L K 0 ~ B L K m - 1の各々は複数のページ（ここではページP 0 ~ P n - 1）を含む。ブロックB L K 0 ~ B L K m - 1は、消去単位として機能する。ブロックは、「消去ブロック」、「物理ブロック」、または「物理消去ブロック」と称されることもある。ページP 0 ~ P n - 1は、データ書き込み動作およびデータ読み込

50

み動作の単位である。

【 0 0 3 5 】

コントローラ 4 は、T o g g l e N A N D フラッシュインタフェース、オープン N A N D フラッシュインタフェース (O N F I) のような、N A N D インタフェース 1 3 を介して、不揮発性メモリである N A N D 型フラッシュメモリ 5 に電氣的に接続されている。コントローラ 4 は、N A N D 型フラッシュメモリ 5 を制御するように構成されたメモリコントローラとして動作する。このコントローラ 4 は、S y s t e m - o n - a - c h i p (S o c) のような回路によって実現されてもよい。

【 0 0 3 6 】

N A N D 型フラッシュメモリ 5 は、図 4 に示すように、複数の N A N D 型フラッシュメモリチップ (N A N D 型フラッシュメモリダイ) を含んでいてもよい。個々の N A N D 型フラッシュメモリチップは独立して動作可能である。このため、N A N D 型フラッシュメモリチップは、並列動作可能な単位として機能する。図 4 においては、N A N D インタフェース 1 3 に 1 6 個のチャンネル C h . 1 ~ C h . 1 6 が接続されており、1 6 個のチャンネル C h . 1 ~ C h . 1 6 の各々に 2 つの N A N D 型フラッシュメモリチップが接続されている場合が例示されている。この場合、チャンネル C h . 1 ~ C h . 1 6 に接続された 1 6 個の N A N D 型フラッシュメモリチップ # 1 ~ # 1 6 がバンク # 0 として編成されてもよく、またチャンネル C h . 1 ~ C h . 1 6 に接続された残りの 1 6 個の N A N D 型フラッシュメモリチップ # 1 7 ~ # 3 2 がバンク # 1 として編成されてもよい。バンクは、複数のメモリモジュールをバンクインタリーブによって並列動作させるための単位として機能する。図 4 の構成例においては、1 6 個のチャンネルと、2 つのバンクを使用したバンクインタリーブとによって、最大 3 2 個の N A N D 型フラッシュメモリチップを並列動作させることができる。

【 0 0 3 7 】

消去動作は、一つのブロック (物理ブロック) の単位で実行されてもよいし、並列動作可能な複数の物理ブロックの集合を含むブロックグループ (スーパーブロック) の単位で実行されてもよい。一つのブロックグループ (一つのスーパーブロック) は、これに限定されないが、N A N D 型フラッシュメモリチップ # 1 ~ # 3 2 から一つずつ選択される計 3 2 個の物理ブロックを含んでいてもよい。なお、N A N D 型フラッシュメモリチップ # 1 ~ # 3 2 の各々はマルチプレーン構成を有していてもよい。例えば、N A N D 型フラッシュメモリチップ # 1 ~ # 3 2 の各々が、2 つのプレーンを含むマルチプレーン構成を有する場合には、一つのスーパーブロックは、N A N D 型フラッシュメモリチップ # 1 ~ # 3 2 に対応する 6 4 個のプレーンから一つずつ選択される計 6 4 個の物理ブロックを含んでいてもよい。

【 0 0 3 8 】

図 5 には、3 2 個の物理ブロック (ここでは、N A N D 型フラッシュメモリチップ # 1 内の物理ブロック B L K 2、N A N D 型フラッシュメモリチップ # 2 内の物理ブロック B L K 3、N A N D 型フラッシュメモリチップ # 3 内の物理ブロック B L K 7、N A N D 型フラッシュメモリチップ # 4 内の物理ブロック B L K 4、N A N D 型フラッシュメモリチップ # 5 内の物理ブロック B L K 6、...、N A N D 型フラッシュメモリチップ # 3 2 内の物理ブロック B L K 3) を含む一つのスーパーブロック (S B) が例示されている。

【 0 0 3 9 】

書き込み先ブロックは一つの物理ブロックであってもよいし、一つのスーパーブロックであってもよい。なお、一つのスーパーブロックが一つの物理ブロックのみを含む構成が利用されてもよく、この場合には、一つのスーパーブロックは一つの物理ブロックと等価である。

【 0 0 4 0 】

次に、図 3 のコントローラ 4 の構成について説明する。

コントローラ 4 は、ホストインタフェース 1 1、C P U 1 2、N A N D インタフェース 1 3、および D R A M インタフェース 1 4、直接メモリアクセスコントローラ (D M A C

10

20

30

40

50

）15、ECCエンコード/デコード部16等を含む。これらホストインタフェース11、CPU12、NANDインタフェース13、DRAMインタフェース14、DMAC15、ECCエンコード/デコード部16は、バス10を介して相互接続される。

【0041】

このホストインタフェース11は、ホスト2との通信を実行するように構成されたホストインタフェース回路である。このホストインタフェース11は、例えば、PCIeコントローラ（NVMeコントローラ）であってもよい。あるいは、フラッシュストレージデバイス3がEthernet（登録商標）を介してホスト2に接続される構成においては、ホストインタフェース11は、NVMe over Fabrics（NVMeOF）コントローラであってもよい。

10

【0042】

ホストインタフェース11は、ホスト2から様々なコマンドを受信する。これらコマンドには、ライトコマンド、リードコマンド、コピーコマンド、イレースコマンド、リユースコマンド、QoSドメインクリエートコマンド、他の様々なコマンドが含まれる。

【0043】

CPU12は、ホストインタフェース11、NANDインタフェース13、DRAMインタフェース14、DMAC15、ECCエンコード/デコード部16を制御するように構成されたプロセッサである。CPU12は、フラッシュストレージデバイス3の電源オンに 응답してNAND型フラッシュメモリ5または図示しないROMから制御プログラム（ファームウェア）をDRAM6にロードし、そしてこのファームウェアを実行することによって様々な処理を行う。なお、ファームウェアはコントローラ4内の図示しないSRAM上にロードされてもよい。このCPU12は、ホスト2からの様々なコマンドを実行するためのコマンド処理等を実行することができる。CPU12の動作は、CPU12によって実行される上述のファームウェアによって制御される。なお、コマンド処理の一部または全部は、コントローラ4内の専用ハードウェアによって実行してもよい。

20

【0044】

CPU12は、ライト制御部21、リード制御部22、コピー制御部23、およびライトバッファ解放制御部24として機能することができる。なお、これらライト制御部21、リード制御部22、コピー制御部23、およびライトバッファ解放制御部24の各々部または全部も、コントローラ4内の専用ハードウェアによって実現されてもよい。

30

【0045】

ライト制御部21は、ライトデータが格納されているホスト2のライトバッファ内の記憶位置を示す記憶位置情報を指定するライトコマンドをホスト2から受信する。ライト制御部21は、記憶位置情報に基づいてライトデータをホスト2のライトバッファから取得する。ライト制御部21は、このライトデータを、NAND型フラッシュメモリ5の複数のブロックから選択された書き込み先ブロックの書き込み先位置に書き込む書き込み動作を実行する。

【0046】

本実施形態では、ライトデータが書き込まれるべき記憶位置を示す物理アドレスを指定するタイプのライトコマンドが使用されてもよいし、QoSドメインIDとタグとを指定し物理アドレスを指定しないタイプのライトコマンドが使用されてもよい。

40

ホスト2が書き込み先ブロックおよびこの書き込み先ブロック内の書き込み先位置の双方を決定するという構成が利用されるケースにおいては、ライトデータが書き込まれるべき記憶位置を示す物理アドレスを指定するタイプのライトコマンドが使用される。

【0047】

一方、フラッシュストレージデバイス3が書き込み先ブロックおよびこの書き込み先ブロック内の書き込み先位置の双方を決定するという構成が利用されるケースにおいては、QoSドメインIDとタグとを指定し物理アドレスを指定しないタイプのライトコマンドが使用される。

【0048】

50

以下では、QoSドメインIDとタグとを指定し物理アドレスを指定しないタイプのライトコマンドが使用されるケースを想定する。

ライト制御部21は、ライトバッファ内の記憶位置情報とQoSドメインIDとタグとを指定するライトコマンドをホスト2から受信する。タグは、書き込むべきデータ(ユーザデータ)を識別可能な識別子であり、例えば、LBAであってもよいし、あるいはキー・バリュー・ストアのキーであってもよいし、このキーのハッシュ値であってもよい。ライトコマンドを受信した場合、ライト制御部21は、共通ブロック群602のうちの一つのブロックを選択し、選択したブロックを、ライトコマンドによって指定されたQoSドメインIDに対応するQoSドメイン用の書き込み先ブロックとして割り当てる。このQoSドメイン用の書き込み先ブロックとして割り当てられるブロックは、物理ブロックであつてもよいし、スーパーブロックであつてもよい。

10

【0049】

この場合、ライト制御部21は、共通ブロック群602から、最小プログラム/イレージサイクルカウントを有するブロック(物理ブロックまたはスーパーブロック)を選択してもよい。これにより、QoSドメイン間でウェアレベリングを実行することができる。利用可能な書き込み先ブロックが既にこのQoSドメイン用に割り当てられているならば、共通ブロック群602のうちの一つのブロックをこのQoSドメイン用の書き込み先ブロックとして割り当てる動作は実行されない。

【0050】

ライト制御部21は、ページ書き込み順序の制約、およびこの書き込み先ブロック内の不良ページ等を考慮して、データが書き込まれるべき、この書き込み先ブロック内の位置(書き込み先位置)を決定する。次いで、ライト制御部21は、記憶位置情報に基づいてホスト2のライトバッファからデータ(ライトデータ)を取得し、このライトデータを、この書き込み先ブロックの書き込み先位置に書き込む。この場合、ライト制御部21は、ホスト2からのライトデータのみならず、このライトデータとこのライトデータのタグの双方を書き込み先ブロックに書き込むことができる。

20

【0051】

あるQoSドメイン用の書き込み先ブロック全体がデータで満たされると、ライト制御部21は、共通ブロック群602のうちの一つのブロックをこのQoSドメイン用の新たな書き込み先ブロックとして割り当てる。例えば、あるライトコマンドに関連付けられたライトデータの一部の書き込みによって現在の書き込み先ブロック全体がデータで満たされると、ライト制御部21は、共通ブロック群602のうちの一つのブロックをこのQoSドメイン用の新たな書き込み先ブロックとして割り当て、このライトコマンドに関連付けられたライトデータの残り部分をこの新たな書き込み先ブロックに書き込む。

30

【0052】

ライトデータが書き込み先ブロックに書き込まれると、ライト制御部21は、このデータが書き込まれたNAND型フラッシュメモリ5内の記憶位置を示す物理アドレス(ブロックアドレス、オフセット)をホスト2にアドレス記録要求として送出する。

一つの物理ブロックが書き込み先ブロックとして割り当てられたケースにおいては、ブロックアドレスは、この物理ブロックを識別するためのブロックアドレスである。このブロックアドレスはチップ番号とこのチップ内のブロック番号とによって表されてもよい。

40

【0053】

一つのスーパーブロックが書き込み先ブロックとして割り当てられたケースにおいては、ブロックアドレスは、このスーパーブロックを識別するためのブロックアドレス(スーパーブロックアドレスとも云う)である。フラッシュストレージデバイス3内の全てのスーパーブロックに異なるスーパーブロックアドレスが付与されていてもよい。

【0054】

オフセットは、書き込み先ブロックの先頭から書き込み先位置までのオフセット、つまり書き込み先ブロックの先頭に対する書き込み先位置のオフセットを示す。書き込み先ブロックの先頭から書き込み先位置までのオフセットのサイズは、あるサイズを有する粒度

50

の倍数で表されてもよい。あるいは、オフセットは、一つのページのページアドレスと、このページ内のオフセットとによって表されてもよい。

【 0 0 5 5 】

リード制御部 2 2 は、ホスト 2 から受信されるリードコマンドに応じて、このリードコマンドによって読み出し対象データとして指定されたデータを N A N D 型フラッシュメモリ 5 から読み出し、そしてこの読み出したデータをホスト 2 に転送する。

ホスト 2 が書き込み先ブロックおよびこの書き込み先ブロック内の書き込み先位置の双方を決定するという構成が利用されるケース、またはフラッシュストレージデバイス 3 が書き込み先ブロックおよびこの書き込み先ブロック内の書き込み先位置の双方を決定するという構成が利用されるケースのいずれにおいても、物理アドレス（すなわち、ブロックアドレスおよびオフセット）を指定するリードコマンドを使用することができる。

【 0 0 5 6 】

物理アドレス（すなわち、ブロックアドレスおよびオフセット）を指定するリードコマンドをホスト 2 から受信した場合、リード制御部 2 2 は、これらブロックアドレスおよびオフセットに基づいて N A N D 型フラッシュメモリ 5 からデータをリードする。リード対象のブロックは、ブロックアドレスによって特定される。このブロック内のリード対象の記憶位置は、オフセットによって特定される。リードコマンドは、さらに、リード対象の Q o S ドメインの Q o S ドメイン I D を含んでいてもよい。

【 0 0 5 7 】

本実施形態では、あるライトデータの書き込み動作が終了してこのライトデータが N A N D 型フラッシュメモリ 5 から読み出し可能となる前に、このライトデータ（このライトデータの一部または全て）を読み出し対象データとして指定するリードコマンドがホスト 2 から受信された場合、リード制御部 2 2 は、このライトデータ（このライトデータの一部または全て）をホスト 2 のライトバッファから読み出す動作と読み出したライトデータ（このライトデータの一部または全て）をホスト 2 に返す動作とを含む読み出し動作を実行する。このように、リードコマンドによって読み出し対象データとして指定されたデータの書き込み動作が終了しておらず、このデータを N A N D 型フラッシュメモリ 5 から読み出す事ができない場合には、リード制御部 2 2 は、リードエラーをホスト 2 に返すのでは無く、ホスト 2 のライトバッファからこのデータを読み出してこのデータをホスト 2 に返す。

【 0 0 5 8 】

現代の N A N D 型フラッシュメモリの多くにおいては、プログラムディスタ urb を低減するための複雑な書き込み動作が実行される場合が多い。このため、現代の N A N D 型フラッシュメモリの多くにおいては、たとえブロック内のあるページにデータが書き込まれても、このページに書き込まれたデータは、その書き込み直後には読み出すことができず、このページに後続する 1 以上のページへのデータ書き込み後に、このページからのデータの読み出しが可能となるケースがある。

【 0 0 5 9 】

N A N D 型フラッシュメモリにおいては、セルの微細化が進んでおり、セルへのデータ書き込みによって隣接するセルの閾値レベルが変動するというプログラムディスタ urb が起こり得る。このため、N A N D 型フラッシュメモリメモリにおいては、プログラムディスタ urb による影響を考慮して、あるページ内の各セルへのデータの書き込みによって、その直前の一つ以上のページ内の各セルの閾値レベルを補正するといったプログラム動作が実行される場合がある。補正が完了する前のページからデータをリードすると、本来のデータとは異なる誤ったデータがリードされてしまう。補正を完了するタイミングは使用される N A N D 型フラッシュメモリのタイプによって異なる。

【 0 0 6 0 】

また、現代の N A N D 型フラッシュメモリにおいては、データ書き込み動作が、同じライトデータを N A N D 型フラッシュメモリに複数回転送することを含む複数回の書き込みステップによって実行されるケースもある。複数回の書き込みステップを含むデータ書き

10

20

30

40

50

込み動作の例の一つとしては、フォギー・ファイン書き込み動作が挙げられる。

【 0 0 6 1 】

複数回の書き込みステップを含むデータ書き込み動作は、少なくとも、フォギー書き込み動作のような第 1 段階の書き込み動作とファイン書き込み動作のような第 2 段階の書き込み動作とを含む。フォギー書き込み動作は各メモリセルの閾値分布を荒く設定する書き込み動作であり、ファイン書き込み動作は各メモリセルの閾値分布を調整する書き込み動作である。さらに、フォギー書き込み動作とファイン書き込み動作との間に、中間の書き込み動作が実行されてもよい。

【 0 0 6 2 】

また、現代の N A N D 型フラッシュメモリにおいては、データ書き込み動作は、隣接する幾つかのワード線を往復しながら実行される。例えば、2つのワード線を往復するフォギー・ファイン書き込み動作では、(1) 1 番目のワード線に接続されたメモリセル群に対するフォギー書き込み動作、(2) 2 番目のワード線に接続されたメモリセル群に対するフォギー書き込み動作、(3) 1 番目のワード線に接続されたメモリセル群に対するファイン書き込み動作、(4) 3 番目のワード線に接続されたメモリセル群に対するフォギー書き込み動作、(5) 2 番目のワード線に接続されたメモリセル群に対するファイン書き込み動作、...、というプログラム順序が適用される。

【 0 0 6 3 】

したがって、現代の N A N D 型フラッシュメモリにおいては、あるデータを書き込むための書き込み動作の開始から終了までに多くの時間が必要される。

本実施形態では、ライトデータの書き込み動作が終了（例えばフォギー書き込み動作とファイン書き込み動作の双方が終了）してこのライトデータが読み出し可能となる前に、このライトデータを読み出し対象データとして指定するリードコマンドがホスト 2 から受信された場合には、リード制御部 2 2 は、ホスト 2 のライトバッファからこの読み出し対象データを読み出し、この読み出したデータをホスト 2 に返す。したがって、ホスト 2 は、リードすべきデータが N A N D 型フラッシュメモリ 5 から読み出し可能であるか否かをチェックする処理およびライトバッファに向けたリード要求とフラッシュストレージデバイス 3 に向けたリードコマンドを選択的に発行する処理を行わずとも、フラッシュストレージデバイス 3 に向けたリードコマンドそれぞれを発行するというシンプルな処理を行うだけで、所望のデータをフラッシュストレージデバイス 3 から受け取ることができる。

【 0 0 6 4 】

コピー制御部 2 3 は、ホスト 2 から受信されるコピーコマンドに応じて、ある Q o S ドメイン用に割り当てられているあるブロック内に格納されているデータを、この Q o S ドメイン用のまたは別の Q o S ドメイン用のブロック（コピー先ブロック）にコピーする。コピー先ブロックは、不揮発性メモリ内に既に書き込まれているデータが書き込まれるべき（コピーされるべき）書き込み先ブロックである。このコピーコマンドは、コピー元 Q o S ドメインの Q o S ドメイン I D と、コピー元ブロックのブロックアドレスと、コピー先 Q o S ドメインの Q o S ドメイン I D とを指定してもよい。コピー元ブロックは一つの物理ブロックであってもよいし、一つのスーパーブロックであってもよい。

【 0 0 6 5 】

また、ホスト 2 は、所定の管理サイズ（例えば 4 K B ）単位でコピー元ブロック内の各データが有効データまたは無効データのいずれであるかを示す管理データ（マップデータ）をフラッシュストレージデバイス 3 に送信することができる。このマップデータはコピーコマンドに含まれていてもよいし、コピーコマンドとは別にホスト 2 からフラッシュストレージデバイス 3 に送信されてもよい。コピー制御部 2 3 は、ホスト 2 から受信されるマップデータおよびコピーコマンドに応じて、コピー元ブロック内に格納されている有効データそれぞれを、この Q o S ドメイン用のまたは別の Q o S ドメイン用のブロック（コピー先ブロック）にコピーすることができる。

【 0 0 6 6 】

コピー元ブロックからコピー先ブロックへのデータ（有効データ）それぞれのコピーが

10

20

30

40

50

完了すると、コピー制御部 23 は、コピーされたデータ毎にこのデータのタグとこのデータのコピー先物理アドレス（ブロックアドレス、オフセット）とをホスト 2 に通知する。

【0067】

ライトバッファ解放制御部 24 は、NAND 型フラッシュメモリ 5 から読み出し可能となっていないライトデータを読み出し対象データとして指定するリードコマンドが存在する場合には、このリードコマンドの実行が完了するまで、このライトデータが格納されているライトバッファ内の領域の解放を禁止する。そして、このリードコマンドの実行が完了した後に、ライトバッファ解放制御部 24 は、ライトデータが格納されているライトバッファ内の領域を解放するためのライトバッファ解放要求をホスト 2 に送信する。

【0068】

つまり、ライトバッファ解放制御部 24 は、あるライトデータの書き込み動作が終了してこのライトデータが読み出し可能となった場合、リードコマンドによって読み出し対象データとして指定されたこのライトデータをライトバッファから読み出してホスト 2 に返す読み出し動作が実行中であるか否かを判定する。この読み出し動作が実行中である場合、ライトバッファ解放制御部 24 は、この読み出し動作が終了するまでライトバッファ解放要求をホスト 2 に送信せず、この読み出し動作が終了した後、ライトバッファ解放要求をホスト 2 に送信する。よって、ライトバッファから読み出すべきことが決定されたデータまたはライトバッファから読み出し中のデータなどがライトバッファから消失されてしまうという、ライトバッファ解放要求とリードコマンドとの衝突（解放要求 / リードコマンド衝突）を回避することができる。

【0069】

NAND インタフェース 13 は、CPU 12 の制御の下、NAND 型フラッシュメモリ 5 を制御するように構成されたメモリ制御回路である。

DRAM インタフェース 14 は、CPU 12 の制御の下、DRAM 6 を制御するように構成された DRAM 制御回路である。DRAM 6 の記憶領域は、ブロック管理テーブル 32 を格納するために使用される。また、DRAM 6 の記憶領域は、ライトデータ等を一時的に格納するための内部バッファとしても利用され得る。なお、内部バッファがフラッシュストレージデバイス 3 内に存在しない、または内部バッファの容量がほぼゼロに近い、というバッファレス構成がフラッシュストレージデバイス 3 に適用されてもよい。

【0070】

ブロック管理テーブル 32 は、NAND 型フラッシュメモリ 5 内の複数のブロックそれぞれに対応する複数の管理テーブルを含む。各管理テーブルは、この管理テーブルに対応するブロックを管理するための管理情報（メタデータ）を格納するために使用される。メタデータは、これに限定されないが、書き換え回数（プログラム / イレーズサイクルの数）、オープン / クローズ状態、書き込み先ページ範囲、等を含んでもよい。オープン / クローズ状態は、このブロックがオープン状態またはクローズ状態のいずれであることを示す。オープン状態は、このブロックが書き込み先ブロックとして使用中の状態を示す。書き込み先ブロックには、ホスト 2 からのデータが書き込まれるホスト書き込み用の書き込み先ブロックと、不揮発性メモリ内に既に書き込まれているデータのコピー用の書き込み先ブロックとがある。データコピー用の書き込み先ブロックはコピー先ブロックと称される。

【0071】

クローズ状態は、ブロックがデータで満たされてアクティブブロックとして管理されている状態を示す。書き込み先ページ範囲は、どのページからどのページまでが現在書き込み（プログラム）動作中であることを示す。

DMA C 15 は、CPU 12 の制御の下、ホスト 2 内のメモリ（ライトバッファ）と DRAM 6（内部バッファ）との間のデータ転送を実行する。ホスト 2 のライトバッファから内部バッファにライトデータを転送すべき場合には、CPU 12 は、ライトバッファ上の位置を示す転送元アドレス、データサイズ、内部バッファ上の位置を示す転送先アドレスを DMA C 15 に対して指定する。

10

20

30

40

50

【 0 0 7 2 】

E C C エンコード / デコード部 1 6 は、N A N D 型フラッシュメモリ 5 にデータをライトすべき時、書き込むべきライトデータをエンコード (E C C エンコード) することによってこのデータにエラー訂正コード (E C C) を付加する。N A N D 型フラッシュメモリ 5 からデータがリードされた時、E C C エンコード / デコード部 1 6 は、リードされたデータに付加されている E C C を使用して、このデータのエラー訂正を行う (E C C デコード) 。

【 0 0 7 3 】

図 6 は、フラッシュストレージデバイス 3 に適用されるライトコマンドを示す。

ライトコマンドは、フラッシュストレージデバイス 3 にデータの書き込みを要求するコマンドである。このライトコマンドは、コマンド I D、Q o S ドメイン I D、タグ、長さ、ライトバッファアドレス、等を含んでもよい。 10

【 0 0 7 4 】

コマンド I D はこのコマンド (このライトコマンド) の固有の識別子である。Q o S ドメイン I D は、データが書き込まれるべき Q o S ドメインを一意に識別可能な識別子である。あるエンドユーザに対応するアプリケーションからのライト要求に応じてホスト 2 から送信されるライトコマンドは、このエンドユーザに対応する Q o S ドメインを指定する Q o S ドメイン I D を含む。

【 0 0 7 5 】

タグは、書き込まれるべきライトデータを識別するための識別子である。このタグは、上述したように、L B A のような論理アドレスであってもよいし、キー・バリュー・ストアのキーであってもよいし、このキーのハッシュ値であってもよい。 20

長さは、書き込まれるべきライトデータの長さを示す。この長さ (データ長) は、上述の粒度 (G r a i n としても参照される) の数によって指定されてもよいし、L B A の数によって指定されてもよいし、あるいはバイトによって指定されてもよい。

【 0 0 7 6 】

ライトバッファアドレスは、書き込まれるべきライトデータが格納されているホストメモリ (ホスト 2 のライトバッファ) 内の記憶位置を示す記憶位置情報である。ライトバッファアドレスは、データポインタとしても参照される。

コントローラ 4 は、ホスト 2 から受信される Q o S ドメインクリエートコマンドに基づいて、Q o S ドメインクリエートコマンドによって指定された容量を有する Q o S ドメインを作成することができる。この場合、コントローラ 4 は、図 2 の共通ブロック群 6 0 2 のうちの、指定された容量に対応する個数のブロックを、この Q o S ドメイン用に確保 (予約) してもよい。ホスト 2 が Q o S ドメインクリエートコマンドを送信する度に、新たな Q o S ドメインが作成される。 30

【 0 0 7 7 】

コントローラ 4 は、共通ブロック群 6 0 2 に加え、各 Q o S ドメインに対応するアクティブブロックリスト (アクティブブロックプール) も管理することができる。

各ブロックの状態は、有効データを格納しているアクティブブロック (クローズ状態のブロック) と、有効データを格納しておらず書き込み先ブロックとして再使用可能なフリーブロックとに大別される。ある Q o S ドメインにおいては、アクティブブロックである各ブロックは、この Q o S ドメインに対応するアクティブブロックリストによって管理される。一方、各 Q o S ドメインの各フリーブロックは、共通ブロック群 6 0 2 によって管理される。 40

【 0 0 7 8 】

ある Q o S ドメイン I D を指定するライトコマンドをホスト 2 から受信した時、コントローラ 4 は、共通ブロック群 6 0 2 内の一つのブロック (フリーブロック) を選択し、選択されたブロックを、この Q o S ドメインに書き込み先ブロックとして割り当てる。さらに、コントローラ 4 は、この書き込み先ブロック内の位置 (書き込み先位置) を決定する。書き込み先ブロック内の書き込み先位置は、ページ書き込み順序の制約および不良ペー 50

ジ等を考慮して決定される。そして、コントローラ 4 は、ホスト 2 からのデータを、書き込み先ブロック内の書き込み先位置に書き込む。

【0079】

この書き込み先ブロック全体がユーザデータで満たされたならば、コントローラ 4 は、この書き込み先ブロックをこの QoS ドメインのアクティブブロックリスト（アクティブブロックプール）に移動する。そして、コントローラ 4 は、共通ブロック群 602 内の一つのブロック（フリーブロック）を再び選択し、選択されたブロックをこの QoS ドメインに新たな書き込み先ブロックとして割り当てる。

【0080】

ある QoS ドメインに割り当て可能なブロックの数の上限は、この QoS ドメインの容量に対応するブロックの個数までに制限される。共通ブロック群 602 から一つのブロックがこの QoS ドメインに割り当てられると、コントローラ 4 は、この QoS ドメインに割り当て可能なブロックの現在の個数を 1 だけ減らす。この QoS ドメインの一つのブロックを指定する上述のリユースコマンドまたはイレーズコマンドによってこのブロックが共通ブロック群 602 に返却されると、コントローラ 4 は、この QoS ドメインに割り当て可能なブロックの現在の個数を 1 だけ増やす。

10

【0081】

図 7 は、フラッシュストレージデバイス 3 からホスト 2 に送られるアドレス記録要求を示す。

アドレス記録要求は、ライトコマンドに関連付けられたライトデータが書き込まれた NAND 型フラッシュメモリ 5 内の記憶位置（物理記憶位置）を示す物理アドレスをホスト 2 に通知するために使用される。このアドレス記録要求は、QoS ドメイン ID、タグ、物理アドレス、長さを含んでもよい。

20

【0082】

QoS ドメイン ID は、ライトデータが書き込まれた QoS ドメインを識別する識別子である。タグは、図 6 のライトコマンドに含まれていたタグである。物理アドレスは、ライトコマンドに関連付けられたライトデータが書き込まれた物理記憶位置を示す。物理アドレスは、ブロックアドレスとオフセットとによって表される。長さは、書き込まれたライトデータの長さを示す。

【0083】

フラッシュストレージデバイス 3 から受信されるアドレス記録要求に基づいて、ホスト 2 は、ライトデータが書き込まれた記憶位置を示す物理アドレス（ブロックアドレス、オフセット）がこのライトデータのタグ（例えば LBA）に関連付けられるように、この QoS ドメインに対応する LUT を更新することができる。また、ホスト 2 は、各 QoS ドメインに対応する LUT の内容に基づいて、各 QoS ドメインで使用されているブロック毎に有効 / 無効管理情報を管理することができる。あるブロックに対応する有効 / 無効管理情報は、このブロックに格納されている複数のデータそれぞれが有効データであるか無効データであるかを示す。

30

【0084】

図 8 は、フラッシュストレージデバイス 3 に適用されるリードコマンドを示す。

40

リードコマンドは、フラッシュストレージデバイス 3 にデータの読み出しを要求するコマンドである。このリードコマンドは、コマンド ID、QoS ドメイン ID、物理アドレス、長さ、リードバッファアドレスを含んでもよい。

【0085】

コマンド ID はこのコマンド（このリードコマンド）の固有の識別子である。QoS ドメイン ID は、読み出されるべきデータが格納されている QoS ドメインを識別する識別子である。物理アドレスは、読み出されるべきデータが格納されている物理記憶位置を示す。物理アドレスは、ブロックアドレスとオフセットとによって表される。このブロックアドレスは、読み出し対象のブロックのブロックアドレスである。オフセットは、読み出し対象のブロック内の読み出し対象の記憶位置を示す。長さは、読み出されるべきデータ

50

の長さを示す。リードバッファアドレスは、読み出されたデータが転送されるべきホストメモリ（ホスト2のリードバッファ）内の位置を示す。

【0086】

図9は、QoSドメインクリエートコマンドを示す。

QoSドメインクリエートコマンドは、QoSドメインを作成するためのコマンドである。このQoSドメインクリエートコマンドは、コマンドID、QoSドメインID、容量を含んでもよい。

コマンドIDはこのコマンド（このQoSドメインクリエートコマンド）の固有の識別子である。QoSドメインIDは、作成すべきQoSドメインの識別子である。容量は、作成すべきQoSドメイン用に確保すべき容量を示す。コントローラ4は、この容量に対応する個数のブロックを共通ブロック群602から確保し、このQoSドメインIDを指定するライトコマンドに応じて、この確保されたブロックのうちの一つのブロックをこのQoSドメイン用の書き込み先ブロックとして割り当てる。

【0087】

図10は、フラッシュストレージデバイス3からホスト2に送られるライトバッファ解放要求を示す。

ライトバッファ解放要求は、NAND型フラッシュメモリ5への書き込みが終了することによってNAND型フラッシュメモリ5から読み出し可能になったデータが格納されているライトバッファ内の領域を解放するために使用される。ライトバッファ解放要求は、ライトコマンドIDを含んでもよい。ライトコマンドIDは、解放可能なライトデータに対応するライトコマンドの識別子である。例えば、あるライトコマンドに関連付けられたライトデータの書き込みが終了した場合、フラッシュストレージデバイス3のコントローラ4は、このライトコマンドの識別子（コマンドID）を含むライトバッファ解放要求をホスト2に送信してもよい。あるいは、ライトバッファ解放要求は、ライトバッファ内の解放可能領域を示す記憶位置情報を含んでもよい。

【0088】

図11は、物理アドレスを規定するブロックアドレスおよびオフセットを示す。

ブロックアドレスはある一つのブロックBLKを指定する。各ブロックBLKは、図11に示されているように、複数のページ（ここでは、ページ0～ページn）を含む。

ページサイズ（各ページのユーザデータ格納領域）が16Kバイトであり、粒度（Granularity）が4KBのサイズであるケースにおいては、このブロックBLKは、 $4 \times (n + 1)$ 個の領域に論理的に分割される。

【0089】

オフセット+0はページ0の最初の4KB領域を示し、オフセット+1はページ0の2番目の4KB領域を示し、オフセット+2はページ0の3番目の4KB領域を示し、オフセット+3はページ0の4番目の4KB領域を示す。

オフセット+4はページ1の最初の4KB領域を示し、オフセット+5はページ1の2番目の4KB領域を示し、オフセット+6はページ1の3番目の4KB領域を示し、オフセット+7はページ1の4番目の4KB領域を示す。

【0090】

図12は、スーパーブロックが使用されるケースにおけるブロックアドレスとオフセットとの関係を示す。

ここでは、図示を簡単化するために、ある一つのスーパーブロックSB#1が4つのブロックBLK#11、BLK#21、BLK#31、BLK#41から構成されている場合が想定されている。スーパーブロックSB#1のブロックアドレス（スーパーブロックアドレス）はSB#1である。4つのブロックBLK#11、BLK#21、BLK#31、BLK#41は、異なる4個のNAND型フラッシュメモリチップからそれぞれ選択されたブロックであってもよい。

【0091】

コントローラ4は、ブロックBLK#11のページ0、ブロックBLK#21のページ

10

20

30

40

50

0、ブロック B L K # 3 1 のページ 0、ブロック B L K # 4 1 のページ 0、ブロック B L K # 1 1 のページ 1、ブロック B L K # 2 1 のページ 1、ブロック B L K # 3 1 のページ 1、ブロック B L K # 4 1 のページ 1、... という順序でデータを書き込む。

【 0 0 9 2 】

オフセット + 0 はブロック B L K # 1 1 のページ 0 の最初の 4 K B 領域を示し、オフセット + 1 はブロック B L K # 1 1 のページ 0 の 2 番目の 4 K B 領域を示し、オフセット + 2 はブロック B L K # 1 1 のページ 0 の 3 番目の 4 K B 領域を示し、オフセット + 3 はブロック B L K # 1 1 のページ 0 の 4 番目の 4 K B 領域を示す。

【 0 0 9 3 】

オフセット + 4 はブロック B L K # 2 1 のページ 0 の最初の 4 K B 領域を示し、オフセット + 5 はブロック B L K # 2 1 のページ 0 の 2 番目の 4 K B 領域を示し、オフセット + 6 はブロック B L K # 2 1 のページ 0 の 3 番目の 4 K B 領域を示し、オフセット + 7 はブロック B L K # 2 1 のページ 0 の 4 番目の 4 K B 領域を示す。

【 0 0 9 4 】

同様に、オフセット + 1 2 はブロック B L K # 4 1 のページ 0 の最初の 4 K B 領域を示し、オフセット + 1 3 はブロック B L K # 4 1 のページ 0 の 2 番目の 4 K B 領域を示し、オフセット + 1 4 はブロック B L K # 4 1 のページ 0 の 3 番目の 4 K B 領域を示し、オフセット + 1 5 はブロック B L K # 4 1 のページ 0 の 4 番目の 4 K B 領域を示す。

【 0 0 9 5 】

オフセット + 1 6 はブロック B L K # 1 1 のページ 1 の最初の 4 K B 領域を示し、オフセット + 1 7 はブロック B L K # 1 1 のページ 1 の 2 番目の 4 K B 領域を示し、オフセット + 1 8 はブロック B L K # 1 1 のページ 1 の 3 番目の 4 K B 領域を示し、オフセット + 1 9 はブロック B L K # 1 1 のページ 1 の 4 番目の 4 K B 領域を示す。

【 0 0 9 6 】

オフセット + 2 0 はブロック B L K # 2 1 のページ 1 の最初の 4 K B 領域を示し、オフセット + 2 1 はブロック B L K # 2 1 のページ 1 の 2 番目の 4 K B 領域を示し、オフセット + 2 2 はブロック B L K # 2 1 のページ 1 の 3 番目の 4 K B 領域を示し、オフセット + 2 3 はブロック B L K # 2 1 のページ 1 の 4 番目の 4 K B 領域を示す。

【 0 0 9 7 】

同様に、オフセット + 2 8 はブロック B L K # 4 1 のページ 1 の最初の 4 K B 領域を示し、オフセット + 2 9 はブロック B L K # 4 1 のページ 1 の 2 番目の 4 K B 領域を示し、オフセット + 3 0 はブロック B L K # 4 1 のページ 1 の 3 番目の 4 K B 領域を示し、オフセット + 3 1 はブロック B L K # 4 1 のページ 1 の 4 番目の 4 K B 領域を示す。

【 0 0 9 8 】

例えば、ある L B A (L B A x) を指定するライトコマンドに対応する 4 K バイトデータをオフセット + 8 に対応する位置に書き込む場合には、コントローラ 4 は、タグ (= L B A x)、ブロック番号 (= S B # 1)、オフセット (= + 8)、長さ (= 1) をこのライトコマンドに対するアドレス記録要求としてホスト 2 に返してもよい。

【 0 0 9 9 】

図 1 3 は、フラッシュストレージデバイス 3 によって管理される複数の Q o S ドメインを示す。

図 1 3 では、Q o S ドメイン # 0、Q o S ドメイン # 1、Q o S ドメイン # 2、...、Q o S ドメイン # n - 1 が既に作成されている場合が明示されている。図 1 3 では、これら Q o S ドメインは四角によって表されている。ある Q o S ドメインを表す四角の縦幅は、この Q o S ドメインの容量を表している。

【 0 1 0 0 】

ユーザアプリケーション # 0 は、Q o S ドメイン # 0 の Q o S ドメイン I D # 0 を含むリード/ライトコマンドを使用して、Q o S ドメイン # 0 をリード/ライトアクセスすることができる。同様に、ユーザアプリケーション # n - 1 は、Q o S ドメイン # n - 1 の Q o S ドメイン I D # n - 1 を含むリード/ライトコマンドを使用して、Q o S ドメイン

n - 1 をリード / ライトアクセスすることができる。

【 0 1 0 1 】

図 1 4 は、ホスト 2 とフラッシュストレージデバイス 3 とによって実行されるデータ書き込み処理を示す。

(1) ホスト 2 においては、ホスト F T L 7 0 1 が実行される。このホスト F T L 7 0 1 は、L B A のようなタグとフラッシュストレージデバイス 3 の物理アドレスそれぞれとの間のマッピングを L U T を使用して管理する。あるユーザアプリケーションからのライト要求に応じて、ホスト F T L 7 0 1 は、このユーザアプリケーションに対応する Q o S ドメインの Q o S ドメイン I D を指定するライトコマンドをフラッシュストレージデバイス 3 に送信する。ホスト F T L 7 0 1 は、Q o S ドメイン I D のみを指定すればよく、データが書き込まれるべきブロックおよびデータが書き込まれるべきページを指定する必要は無い。このため、ホスト F T L 7 0 1 は、フラッシュストレージデバイス 3 内の不良ブロック、不良ページなどを管理する必要がなくなる。

10

【 0 1 0 2 】

(2) フラッシュストレージデバイス 3 のコントローラ 4 は、複数の Q o S ドメイン 6 0 1 によって共有される共通ブロック群 (フリーブロック群) 6 0 2 のうちの一つのブロックを選択する。コントローラ 4 は、この選択したブロックを、受信されたライトコマンドによって指定された Q o S ドメイン I D を有する Q o S ドメイン用の書き込み先ブロックとして割り当てる。共通ブロック群 (フリーブロック群) 6 0 2 においては、不良ブロック以外のフリーブロック群のみが管理されていてもよい。例えば、最小プログラム / イレージサイクル数を有するブロックが共通ブロック群 6 0 2 から選択されてもよい。そして、コントローラ 4 は、この書き込み先ブロックをデータが書き込まれるべきブロックとして決定し、さらに、データが書き込まれるべき、この書き込み先ブロック内の書き込み先位置を決定する。なお、上述したように、既にこの Q o S ドメイン用の書き込み先ブロックが割り当てられているならば、コントローラ 4 は、この既に割り当てられている書き込み先ブロックをデータが書き込まれるべきブロックとして決定すればよく、共通ブロック群 6 0 2 から一つのブロックをこの Q o S ドメイン用の書き込み先ブロックとして割り当てる動作を実行する必要は無い。

20

【 0 1 0 3 】

(3) コントローラ 4 は、決定された書き込み先ブロックを示すブロックアドレスと決定した書き込み先位置を示すオフセットとを含む物理アドレス記録要求をホスト 2 に送信して、このライトコマンドに関連付けられたライトデータが書き込まれる物理アドレスをホスト 2 に通知する。

30

【 0 1 0 4 】

(4) コントローラ 4 は、ライトコマンドに含まれるライトバッファアドレスに基づいて、ホスト 2 のライトバッファ 7 0 2 からライトデータを取得する。ライトバッファ 7 0 2 からのライトデータの取得は D M A 転送を使用して実行されてもよい。そして、コントローラ 4 は、ライトデータを書き込み先ブロック内の書き込み先位置に書き込むデータ書き込み動作を実行する。

【 0 1 0 5 】

(5) ライトデータの書き込みが終了してこのライトデータが N A N D 型フラッシュメモリ 5 から読み出し可能になった後、コントローラ 4 は、このライトデータが格納されているライトバッファ 7 0 2 内の領域を解放するための解放要求をホスト 2 に送出する。この解放要求は、上述したように、書き込みが完了したライトデータに対応するライトコマンドの識別子を含んでもよい。あるいは、この解放要求は、このライトデータの格納位置を示すライトバッファアドレス、このライトデータの長さを含んでもよい。ホスト 2 は、この解放要求に応じて、このライトデータが格納されているライトバッファ 7 0 2 内の領域を解放する。このライトバッファ 7 0 2 内の解放された領域は他のライトデータの格納等のために利用可能となる。

40

【 0 1 0 6 】

50

図15は、同じライトデータをNAND型フラッシュメモリ5に複数回転送することを含む複数回の書き込みステップによって実行されるデータ書き込み動作を示す。

ここでは、4つのワード線を往復する場合のフォギー・ファイン書き込み動作を例示する。また、ここでは、NAND型フラッシュメモリ5が、メモリセル当たり4ビットのデータを格納するQLC-フラッシュである場合を想定する。NAND型フラッシュメモリ5内の一つの特定の書き込み先ブロック(ここでは、書き込み先ブロックBLK#1)に対するフォギー・ファイン書き込み動作は以下のように実行される。

【0107】

(1)まず、4ページ(P0~P3)分のライトデータがページ単位でNAND型フラッシュメモリ5に転送され、この書き込み先ブロックBLK#1内のワード線WL0に接続された複数のメモリセルに、これら4ページ(P0~P3)分のライトデータを書き込むためのフォギー書き込み動作が実行される。

10

【0108】

(2)次いで、次の4ページ(P4~P7)分のライトデータがこのNAND型フラッシュメモリ5にページ単位で転送され、この書き込み先ブロックBLK#1内のワード線WL1に接続された複数のメモリセルに、これら4ページ(P4~P7)分のライトデータを書き込むためのフォギー書き込み動作が実行される。

【0109】

(3)次いで、次の4ページ(P8~P11)分のライトデータがこのNAND型フラッシュメモリ5にページ単位で転送され、この書き込み先ブロックBLK#1内のワード線WL2に接続された複数のメモリセルに、これら4ページ(P8~P11)分のライトデータを書き込むためのフォギー書き込み動作が実行される。

20

【0110】

(4)次いで、次の4ページ(P12~P15)分のライトデータがこのNAND型フラッシュメモリ5にページ単位で転送され、この書き込み先ブロックBLK#1内のワード線WL3に接続された複数のメモリセルに、これら4ページ(P12~P15)分のライトデータを書き込むためのフォギー書き込み動作が実行される。

【0111】

(5)ワード線WL3に接続された複数のメモリセルに対するフォギー書き込み動作が終了すると、書き込み対象のワード線はワード線WL0に戻り、ワード線WL0に接続された複数のメモリセルに対するファイン書き込み動作の実行が可能となる。そして、ワード線WL0に対するフォギー書き込み動作で使用された4ページ(P0~P3)分のライトデータと同じ4ページ(P0~P3)分のライトデータがページ単位でNAND型フラッシュメモリ5に再び転送され、この書き込み先ブロックBLK#1内のワード線WL0に接続された複数のメモリセルに、これら4ページ(P0~P3)分のライトデータを書き込むためのファイン書き込み動作が実行される。これにより、ページP0~P3に対するフォギー・ファイン書き込み動作が終了する。この結果、これら4ページ(P0~P3)分のデータの読み出しが可能となる。

30

【0112】

(6)次いで、次の4ページ(P16~P19)分のライトデータがこのNAND型フラッシュメモリ5にページ単位で転送され、この書き込み先ブロックBLK#1内のワード線WL4に接続された複数のメモリセルに、これら4ページ(P16~P19)分のライトデータを書き込むためのフォギー書き込み動作が実行される。

40

【0113】

(7)ワード線WL4に接続された複数のメモリセルに対するフォギー書き込み動作が終了すると、書き込み対象のワード線はワード線WL1に戻り、ワード線WL1に接続された複数のメモリセルに対するファイン書き込み動作の実行が可能となる。そして、ワード線WL1に対するフォギー書き込み動作で使用された4ページ(P4~P7)分のライトデータと同じ4ページ(P4~P7)分のライトデータがページ単位でNAND型フラッシュメモリ5に再び転送され、この書き込み先ブロックBLK#1内のワード線WL1

50

に接続された複数のメモリセルに、これら 4 ページ (P 4 ~ P 7) 分のライトデータを書き込むためのファイン書き込み動作が実行される。これにより、ページ P 4 ~ P 7 に対するフォギー・ファイン書き込み動作が終了する。この結果、4 ページ (P 4 ~ P 7) 分のデータの読み出しが可能となる。

【 0 1 1 4 】

(8) 次いで、次の 4 ページ (P 2 0 ~ P 2 3) 分のライトデータがこの NAND 型フラッシュメモリ 5 にページ単位で転送され、この書き込み先ブロック B L K # 1 内のワード線 W L 5 に接続された複数のメモリセルに、これら 4 ページ (P 2 0 ~ P 2 3) 分のライトデータを書き込むためのフォギー書き込み動作が実行される。

【 0 1 1 5 】

(9) ワード線 W L 5 に接続された複数のメモリセルに対するフォギー書き込み動作が終了すると、書き込み対象のワード線はワード線 W L 2 に戻り、ワード線 W L 2 に接続された複数のメモリセルに対するファイン書き込み動作の実行が可能となる。そして、ワード線 W L 2 に対するフォギー書き込み動作で使用された 4 ページ (P 8 ~ P 1 1) 分のライトデータと同じ 4 ページ (P 8 ~ P 1 1) 分のライトデータがページ単位で NAND 型フラッシュメモリ 5 に再び転送され、この書き込み先ブロック B L K # 1 内のワード線 W L 2 に接続された複数のメモリセルに、これら 4 ページ (P 8 ~ P 1 1) 分のライトデータを書き込むためのファイン書き込み動作が実行される。これにより、ページ P 8 ~ P 1 1 に対するフォギー・ファイン書き込み動作が終了する。この結果、4 ページ (P 8 ~ P 1 1) 分のデータの読み出しが可能となる。

【 0 1 1 6 】

図 1 6 は、ホスト 2 とフラッシュストレージデバイス 3 とによって実行されるデータ読み出し処理を示す。

(1) ホスト F T L 7 0 1 があるユーザアプリケーションからのリード要求を受信した場合、ホスト F T L 7 0 1 は、ホスト F T L 7 0 1 内の L U T を参照して、このリード要求によって指定される L B A のようなタグに対応する物理アドレスを取得する。ホスト F T L 7 0 1 は、このユーザアプリケーションに対応する Q o S ドメインの Q o S ドメイン I D と、取得した物理アドレス (ブロックアドレスとオフセット) と、長さ、リードバッファアドレスとを指定するリードコマンドを、フラッシュストレージデバイス 3 に送信する。

【 0 1 1 7 】

(2) コントローラ 4 がこのリードコマンドを受信した場合、コントローラ 4 は、このリードコマンドによって指定されるブロックアドレスとオフセットに基づいて、NAND 型フラッシュメモリ 5 からデータを読み出す。そして、コントローラ 4 は、この読み出したデータを、D M A 転送を使用して、リードバッファアドレスによって指定されるホスト 2 のリードバッファ 7 0 3 内の位置に転送する。

【 0 1 1 8 】

図 1 7 は、リードコマンドによって指定されたデータの書き込み動作が終了していない場合に、このデータをホスト 2 のライトバッファから読み出すデータ読み出し処理を示す。

(1) ホスト F T L 7 0 1 があるユーザアプリケーションからのリード要求を受信した場合、ホスト F T L 7 0 1 は、ホスト F T L 7 0 1 内の L U T を参照して、このリード要求によって指定される L B A のようなタグに対応する物理アドレスを取得する。ホスト F T L 7 0 1 は、このユーザアプリケーションに対応する Q o S ドメインの Q o S ドメイン I D と、取得した物理アドレス (ブロックアドレスとオフセット) と、長さ、リードバッファアドレスとを指定するリードコマンドを、フラッシュストレージデバイス 3 に送信する。

【 0 1 1 9 】

(2) コントローラ 4 がこのリードコマンドを受信した場合、コントローラ 4 は、このリードコマンドによって指定されるブロックアドレスとオフセットと長さに基づいて、

10

20

30

40

50

リードコマンドによって指定された読み出し対象データの書き込み動作が終了しているか否かを判定する。指定されるブロックアドレスを有するブロックがクローズ状態のブロックであるならば、リードコマンドによって指定された読み出し対象データの書き込み動作が終了していると判定される。

【0120】

一方、指定されるブロックアドレスを有するブロックがオープン状態のブロックであるならば、このブロックの現在の書き込み先ページ範囲に基づいて、読み出し対象データの書き込み動作が終了しているか否かが判定される。読み出し対象データが格納されているこのブロック内の位置が現在の書き込み先ページ範囲よりも先行するページに属しているならば、読み出し対象データの書き込み動作が終了していると判定される。読み出し対象データが格納されているこのブロック内の位置が現在の書き込み先ページ範囲内のいずれかのページに属しているならば、読み出し対象データの書き込み動作が終了していないと判定される。

10

【0121】

読み出し対象データの書き込み動作が終了しておらずこの読み出し対象データをNAND型フラッシュメモリ5から読み出すことができない場合、コントローラ4は、DMA転送を使用して、読み出し対象データをホスト2のライトバッファから読み出す。各オープン状態のブロック(書き込み先ブロック)に書き込まれるライトデータ部それぞれに対応するライトバッファ内の位置は、書き込み先ブロック毎に用意されたライトバッファアドレスリストを使用して管理されてもよい。ある書き込み先ブロックに対応するライトバッファアドレスリストにおいては、この書き込み先ブロックに書き込むべきことが決定されたライトデータ毎に、このライトデータが格納されているライトバッファアドレスとこのライトデータに対応する書き込み先位置を示すオフセットとが記録されてもよい。コントローラ4は、リードコマンド内のブロックアドレスを有する書き込み先ブロックに対応するライトバッファアドレスリストを参照することによって、読み出し対象データが格納されているライトバッファアドレスを特定してもよい。

20

【0122】

(3)コントローラ4は、ライトバッファから読み出したデータを、DMA転送を使用して、リードバッファアドレスによって指定されるホスト2のリードバッファ703内の位置に転送する。

30

図18は、フラッシュストレージデバイス3のコントローラ4によって管理される、複数のQoSドメインと共通フラッシュブロックプールとの関係を示す。

【0123】

図18においては、図示の簡単化のために、QoSドメイン#0とQoSドメイン#n-1の2つのQoSドメインのみが例示されている。共通フラッシュブロックプール801は、複数のQoSドメインによって共有される上述の共通ブロック群602を管理するためのリストである。

【0124】

QoSドメイン#0のフラッシュブロックプール#0は、QoSドメイン#0に属するアクティブブロックそれぞれを管理するための上述のアクティブブロックリストである。

40

コントローラ4がQoSドメイン#0のQoSドメインID#0を指定するライトコマンドを受信した場合、コントローラ4は、QoSドメイン#0用の書き込み先ブロック(オープンブロック)が割り当て済みであるか否かを判定する。

【0125】

書き込み先ブロック(オープンブロック)が割り当て済みでないならば、コントローラ4は、共通フラッシュブロックプール801の内のフリーブロックをQoSドメイン#0用の書き込み先ブロック(オープンブロック)として割り当てる。そして、コントローラ4は、この書き込み先ブロック内の書き込み先位置を決定し、このライトコマンドに関連付けられたライトデータを、この書き込み先ブロック内のこの書き込み先位置に書き込む。

50

【 0 1 2 6 】

一方、書き込み先ブロック（オープンブロック）が既に割り当て済みならば、コントローラ 4 は、この書き込み先ブロック内の書き込み先位置を決定し、このライトコマンドに関連付けられたライトデータを、この書き込み先ブロック内のこの書き込み先位置に書き込む。

【 0 1 2 7 】

この書き込み先ブロック全体がホストからのデータで満たされると、コントローラ 4 は、この書き込み先ブロックをフラッシュブロックプール # 0 によって管理するとともに（クローズ）、共通フラッシュブロックプール 8 0 1 の内のフリーブロックを Q o S ドメイン # 0 用の新たな書き込み先ブロック（オープンブロック）として割り当てる。

10

【 0 1 2 8 】

コントローラ 4 が Q o S ドメイン ID # 0 を示すコピー元 Q o S ドメインと、コピー元ブロックのブロックアドレスと、Q o S ドメイン ID # 0 を示すコピー先 Q o S ドメインとを指定するコピーコマンドを受信した場合、コントローラ 4 は、Q o S ドメイン # 0 用のコピー先ブロック（オープンブロック）が割り当て済みであるか否かを判定する。

【 0 1 2 9 】

コピー先ブロック（オープンブロック）が割り当て済みでないならば、コントローラ 4 は、共通フラッシュブロックプール 8 0 1 の内のフリーブロックを Q o S ドメイン # 0 用のコピー先ブロック（オープンブロック）として割り当てる。そして、コントローラ 4 は、コピー元ブロックのブロックアドレスに基づいて、フラッシュブロックプール # 0 内のブロックをコピー元ブロックとして選択し、この選択したコピー元ブロック内のデータをコピー先ブロックにコピーする。上述のマッピングデータがホスト 2 から受信されている場合には、このマッピングデータに基づき、コントローラ 4 は、選択したコピー元ブロック内の有効データそれぞれをコピー先ブロックにコピーする。

20

【 0 1 3 0 】

一方、コピー先ブロックが既に割り当て済みならば、コントローラ 4 は、選択したコピー元ブロック内のデータをこのコピー先ブロックにコピーする。

コピー先ブロック全体がホストからのデータで満たされると、コントローラ 4 は、このコピー先ブロックをフラッシュブロックプール # 0 によって管理するとともに（クローズ）、共通フラッシュブロックプール 8 0 1 の内のフリーブロックを Q o S ドメイン # 0 用の新たなコピー先ブロック（オープンブロック）として割り当てる。

30

【 0 1 3 1 】

更新データの書き込みまたはデータコピー動作によって Q o S ドメイン # 0 に属する一つのブロック内のデータ全てが無効データになった場合、ホスト 2 は、このブロックのブロックアドレスを指定するリユースコマンドをフラッシュストレージデバイス 3 に送信する。

【 0 1 3 2 】

コントローラ 4 がこのリユースコマンドをホスト 2 から受信した場合、コントローラ 4 は、このリユースコマンドによって指定されるフラッシュブロックプール # 0 内のブロックを共通フラッシュブロックプール 8 0 1 に返却して、このブロックを任意の Q o S ドメイン用の新たな書き込み先ブロック（またはコピー先ブロック）として再使用可能な状態にする。

40

【 0 1 3 3 】

なお、更新データの書き込みまたはデータコピー動作によって Q o S ドメイン # 0 に属する一つのブロック内のデータ全てが無効データになった場合、ホスト 2 は、このブロックのブロックアドレスを指定するイレースコマンドをフラッシュストレージデバイス 3 に送信してもよい。コントローラ 4 がこのイレースコマンドをホスト 2 から受信した場合、コントローラ 4 は、このイレースコマンドによって指定されるフラッシュブロックプール # 0 内のブロックに対する消去動作を実行し、且つこのブロックを共通フラッシュブロックプール 8 0 1 に返却して、このブロックを任意の Q o S ドメイン用の新たな書き込み先

50

ブロック（またはコピー先ブロック）として再使用可能な状態にする。

【0134】

QoSドメイン#n-1においても、コントローラ4は、書き込み先ブロック/コピー先ブロックの割り当て、データで満たされた書き込み先ブロック/コピー先ブロックをフラッシュブロックプール#n-1によって管理する処理、リユースコマンド/イレースコマンドによって指定されるブロックを共通フラッシュブロックプール801に返却する処理等を実行する。

【0135】

図19は、コントローラ4によって管理される、コマンド数カウンタを含むブロック管理テーブル32を示す。

例えば、ブロックBLK#1に対応するブロック管理テーブル32においては、ブロックBLK#1の状態を管理するためのメタデータとして、コマンド数カウンタ（以下、イン・フライトコマンド数カウンタと称する）が追加されている。

【0136】

イン・フライトコマンド数カウンタはライトコマンド毎に用意されていてもよい。この場合、あるライトコマンドIDに対応するイン・フライトコマンド数カウンタ値は、このライトコマンドに関連付けられたライトデータが格納されているライトバッファ内の領域が解放可能か否かを判定するために使用される。コントローラ4は、各書き込み先ブロックについてイン・フライトコマンド数カウンタを管理する。

【0137】

あるライトコマンドが受信され、このライトコマンドに関連付けられたライトデータを書き込み先ブロックBLK#1に書き込む場合を想定する。この場合、書き込み先ブロックBLK#1に対応するブロック管理テーブル32においては、このライトコマンドのライトコマンドIDに対応するイン・フライトコマンド数カウンタが管理される。

【0138】

コントローラ4は、このイン・フライトコマンド数カウンタ値を、データ書き込み動作に必要な書き込みステップの回数に1を加えた値に初期設定する。このライトコマンドに関連付けられたライトデータを、2回の書き込みステップを含むフォギー・ファイン書き込み動作によって書き込み先ブロックBLK#1に書き込むケースにおいては、データ書き込み動作に必要な書き込みステップの回数は2である。なお、書き込みステップの回数に1を加えるのは、書き込みエラーによって再書き込み動作が必要とされる場合があるからである。

【0139】

1回の書き込みステップの終了の度に、コントローラ4は、このイン・フライトコマンド数カウンタ値を1減少する。

データ書き込み動作が終了してライトデータが読み出し可能となった場合、コントローラ4は、このイン・フライトコマンド数カウンタ値を1減少する。

【0140】

データ書き込み動作の実行中に、このライトデータ（このライトデータの一部または全て）を読み出し対象データとして指定するリードコマンドがホスト2から受信された場合、コントローラ4は、このイン・フライトコマンド数カウンタ値を1増やす。

リードコマンドによって指定された読み出し対象データをライトバッファから読み出してホスト2に返す読み出し動作が終了した場合、コントローラ4は、このイン・フライトコマンド数カウンタ値を1減少する。

【0141】

書き込みエラーによって再書き込み動作が必要となった場合は、コントローラ4は、このイン・フライトコマンド数カウンタ値を、再書き込み動作に必要な書き込みステップの回数だけ増やす。

イン・フライトコマンド数カウンタ値が零になった場合、コントローラ4は、このライトデータが格納されているライトバッファ内の領域を解放するためのライトバッファ解放

10

20

30

40

50

要求をホスト 2 に送信する。このライトバッファ解放要求はこのライトコマンドのコマンド ID を含む。

【 0 1 4 2 】

イン・フライトコマンド数カウンタをライトコマンド毎に用意する代わりに、4 K B のような最小書き込み単位毎に用意してもよい。イン・フライトコマンド数カウンタが 4 K B のような最小書き込み単位毎に用意されている場合には、あるライトコマンドに対応するライトデータ全体に対応する複数のイン・フライトコマンド数カウンタ値の全てが零になった時点で、このライトデータが格納されているライトバッファ内の領域を解放するためのライトバッファ解放要求をホスト 2 に送信してもよい。

【 0 1 4 3 】

なお、このように、ライトコマンド単位でライトバッファ解放要求をホスト 2 に送信する代わりに、4 K B のような最小書き込み単位でライトバッファ解放要求をホスト 2 に送信してもよい。

図 2 0 のフローチャートは、コントローラ 4 によって実行されるイン・フライトコマンド数カウンタ更新処理の手順を示す。

【 0 1 4 4 】

コントローラ 4 は、まず、イン・フライトコマンド数カウンタ値を、データ書き込み動作に必要な書き込みステップの回数（つまりデータ転送回数）に設定する（ステップ S 1 1）。そして、コントローラ 4 は、エラーのための再書き込み処理のために、イン・フライトコマンド数カウンタ値を 1 増やす（ステップ S 1 2）。そして、コントローラ 4 は、受信されたライトコマンドに関連付けられたライトデータを書き込み先ブロックに書き込みための書き込み動作を開始する（ステップ S 1 3）。

【 0 1 4 5 】

このライトデータのための一回の書き込みステップが終了する度（ステップ S 1 4 の Y E S）、コントローラ 4 は、イン・フライトコマンド数カウンタ値を 1 減らし（ステップ S 1 5）。そして、コントローラ 4 は、データ書き込み動作が終了してこの書き込み先ブロックからのライトデータの読み出しが可能になったか否かを判定する（ステップ S 1 6）。

【 0 1 4 6 】

データ書き込み動作が終了してこの書き込み先ブロックからのライトデータの読み出しが可能になった場合（ステップ S 1 6 の Y E S）、コントローラ 4 は、イン・フライトコマンド数カウンタ値を 1 減らす（ステップ S 1 7）。

データ書き込み動作の実行中に、このライトデータを読み出し対象データとして指定するリードコマンドが受信されたならば（ステップ S 1 8 の Y E S）、コントローラ 4 は、イン・フライトコマンド数カウンタ値を 1 増やす（ステップ S 1 9）。そして、このリードコマンドの実行が完了すると（ステップ S 2 0 の Y E S）、コントローラ 4 は、イン・フライトコマンド数カウンタ値を 1 減らす（ステップ S 2 1）。

【 0 1 4 7 】

もし書き込みエラーが発生したならば（ステップ S 2 2 の Y E S）、コントローラ 4 は、このイン・フライトコマンド数カウンタ値を、再書き込み動作に必要な書き込みステップの回数だけ増やす（ステップ S 2 3）。

図 2 1 のフローチャートは、データ読み出し動作の手順を示す。

【 0 1 4 8 】

ホスト 2 からリードコマンドを受信すると（ステップ S 2 1 の Y E S）、コントローラ 4 は、このリードコマンドによって読み出し対象データとして指定されたデータが N A N D 型フラッシュメモリ 5 から読み出し可能であるか否かを判定する（ステップ S 2 2）。

【 0 1 4 9 】

ステップ S 2 2 では、コントローラ 4 は、リードコマンドによって指定されるブロックアドレスを有するブロックがクローズ状態またはオープン状態のいずれであるかを判定する。このブロックがクローズ状態であるならば、読み出し対象データの書き込み動作は既

10

20

30

40

50

に終了している。したがって、コントローラ 4 は、読み出し対象データが NAND 型フラッシュメモリ 5 から読み出し可能であると判定する。

【 0 1 5 0 】

一方、このブロックがオープン状態であるならば、コントローラ 4 は、読み出し対象データが格納されているこのブロック内の位置が、このブロックの現在の書き込み先ページ範囲よりも先行するページに属しているか否かを判定する。読み出し対象データが格納されているこのブロック内の位置が、このブロックの現在の書き込み先ページ範囲よりも先行するページに属しているならば、読み出し対象データの書き込み動作は既に終了している。したがって、コントローラ 4 は、読み出し対象データが NAND 型フラッシュメモリ 5 から読み出し可能であると判定する。

10

【 0 1 5 1 】

読み出し対象データが格納されているこのブロック内の位置が、このブロックの現在の書き込み先ページ範囲よりも先行するページに属しておらず、現在の書き込み先ページ範囲内のいずれかのページに属しているならば、読み出し対象データの書き込み動作は終了していない。したがって、コントローラ 4 は、読み出し対象データが NAND 型フラッシュメモリ 5 から読み出し可能ではないと判定する。

【 0 1 5 2 】

読み出し対象データが NAND 型フラッシュメモリ 5 から読み出し可能である場合には（ステップ S 2 1 の YES）、コントローラ 4 は、読み出し対象データを NAND 型フラッシュメモリ 5 から読み出し、読み出したデータをホスト 2 に返す（ステップ S 2 3）。

20

【 0 1 5 3 】

読み出し対象データが NAND 型フラッシュメモリ 5 から読み出し可能でない場合には（ステップ S 2 1 の NO）、つまりライトデータの書き込み動作が終了してこのライトデータが読み出し可能となる前にこのライトデータ（このライトデータの一部又は全て）を読み出し対象データとして指定するリードコマンドが受信された場合には、コントローラ 4 は、このライトデータが格納されているライトバッファ内の領域の解放を禁止する（ステップ S 2 4）。この場合、たとえこのライトデータの書き込み動作が終了してこのライトデータが読み出し可能となっても、コントローラ 4 は、このライトデータが格納されているライトバッファ内の領域を解放するためのライトバッファ解放要求をホスト 2 に送信しない。ステップ S 2 4 では、このライトデータのライトコマンドに対応するこのイン・フライトコマンド数カウンタ値を 1 増やすことによって、このライトデータが格納されているライトバッファ内の領域の開放を禁止してもよい。

30

【 0 1 5 4 】

コントローラ 4 は、この読み出し対象データをホスト 2 のライトバッファから読み出し、読み出した読み出し対象データをホスト 2 に返す（ステップ S 2 5）。このステップ S 2 5 のデータ読み出し動作が終了すると、つまりこのリードコマンドの実行が完了すると、コントローラ 4 は、このライトデータが格納されているライトバッファ内の領域の開放禁止状態を解除する（ステップ S 2 6）。ステップ S 2 6 では、このライトデータのライトコマンドに対応するこのイン・フライトコマンド数カウンタ値を 1 減らすことによって、このライトデータが格納されているライトバッファ内の領域の開放禁止状態を解除して

40

【 0 1 5 5 】

図 2 2 のフローチャートは、データ書き込み動作とライトバッファ解放処理の手順を示す。

ライトデータが格納されているライトバッファ内の記憶位置を示す記憶位置情報を指定するライトコマンドをホスト 2 から受信すると（ステップ S 3 1）、コントローラ 4 は、この記憶位置情報に基づいて、このライトデータをライトバッファから取得する（ステップ S 3 2）。コントローラ 4 は、NAND 型フラッシュメモリ 5 に含まれる複数のブロックから選択された書き込み先ブロック内の書き込み先位置に書き込む書き込み動作を実行する（ステップ S 3 3）。

50

【 0 1 5 6 】

この書き込み動作が終了してこのライトデータが読み出し可能になった場合（ステップ S 3 4 の Y E S ）、コントローラ 4 は、このライトデータ（このライトデータの一部分又は全て）を読み出し対象データとして指定するリードコマンドが既に受信され、且つこの読み出し対象データをライトバッファから読み出す動作とこの読み出した読み出し対象データをホスト 2 に返す動作とを含む読み出し動作が実行中であるか否かを判定する（ステップ S 3 5 ）。

【 0 1 5 7 】

この読み出し動作が実行中である場合（ステップ S 3 5 の Y E S ）、コントローラ 4 は、このライトデータが格納されているライトバッファ内の領域を解放を禁止する。つまり、コントローラ 4 は、このライトデータが格納されているライトバッファ内の領域を解放するためのライトバッファ解放要求をホスト 2 に送信せず、ライトバッファ解放要求の発行を、このリードコマンドが完了するまで、つまりこの読み出し動作が終了するまで待機する（ステップ S 3 6 、 S 3 7 ）。

【 0 1 5 8 】

そして、この読み出し動作が終了した場合（ステップ S 3 7 の Y E S ）、つまりこのリードコマンドの実行が完了した場合、コントローラ 4 は、このライトデータが格納されているライトバッファ内の領域を解放するためのライトバッファ解放要求をホスト 2 に送信する（ステップ S 3 8 ）。

【 0 1 5 9 】

図 2 3 のフローチャートは、リードコマンドによって指定されたデータがフラッシュメモリおよびホストのライトバッファの双方から読み出し可能である場合に実行されるデータ読み出し動作の手順を示す。

ホスト 2 からリードコマンドを受信すると（ステップ S 4 1 の Y E S ）、コントローラ 4 は、このリードコマンドによって指定された読み出し対象データが書き込み先ブロックから読み出し可能であり且つこの読み出し対象データがライトバッファに存在しているか否かを判定する（ステップ S 4 2 ）。

【 0 1 6 0 】

例えば、ライトデータを書き込み先ブロックに書き込む書き込み動作が終了してからこのライトデータに対応するライトバッファ解放要求の送信までの期間中に、このライトデータ（このライトデータの一部分または全て）を読み出し対象データとして指定するリードコマンドが受信された場合には、コントローラ 4 は、この読み出し対象データが書き込み先ブロックから読み出し可能であり且つこの読み出し対象データがライトバッファに存在していると判定してもよい。

【 0 1 6 1 】

この読み出し対象データが書き込み先ブロックから読み出し可能であり且つこの読み出し対象データがライトバッファに存在している場合（ステップ S 4 2 の Y E S ）、コントローラ 4 は、読み出し対象データをライトバッファから読み出す動作と読み出した読み出し対象データをホスト 2 に返す動作とを含む読み出し動作を実行する（ステップ S 4 3 ）。

【 0 1 6 2 】

通常、ホスト 2 とフラッシュストレージデバイス 3 との間のデータ転送は高速に実行することができる。したがって、読み出し対象データをライトバッファから読み出すことにより、リードレイテンシを低減することが可能となる。

このリードコマンドの実行が管理用するまで、つまりこの読み出し動作が終了するまで、コントローラ 3 は、このライトデータが格納されているライトバッファ内の領域の解放を禁止する（ステップ S 4 4 ）。すなわち、読み出し動作の実行中は、コントローラ 3 は、このライトデータが格納されているライトバッファ内の領域を解放するためのライトバッファ解放要求をホスト 2 に送信しない。これにより、このライトデータの一部分または全てである読み出し対象データの読み出し動作が終了するまで、この読み出し対象データ

10

20

30

40

50

をライトバッファに維持することができる。

【0163】

図24のフローチャートは、リードコマンドによって指定されたデータがフラッシュメモリおよびホストのライトバッファの双方から読み出し可能である場合に実行されるデータ読み出し動作の別の手順を示す。

上述の図23のフローチャートで説明したデータ読み出し動作は、リードレイテンシの低減を優先するために、読み出し対象データをライトバッファから読み出した。図24のフローチャートで説明するデータ読み出し動作では、読み出し可能なデータが格納されているライトバッファ内の領域をできるだけ早期に解放することを優先するために、読み出し対象データは、ライトバッファからではなく、以下説明するように、書き込み先ブロックから読み出される。

【0164】

ホスト2からリードコマンドを受信すると(ステップS51のYES)、コントローラ4は、このリードコマンドによって指定された読み出し対象データが書き込み先ブロックから読み出し可能であり且つこの読み出し対象データがライトバッファに存在しているかを判定する(ステップS52)。

【0165】

例えば、ライトデータを書き込み先ブロックに書き込む書き込み動作が終了してからこのライトデータに対応するライトバッファ解放要求の送信までの期間中に、このライトデータ(このライトデータの一部または全て)を読み出し対象データとして指定するリードコマンドが受信された場合には、コントローラ4は、この読み出し対象データが書き込み先ブロックから読み出し可能であり且つこの読み出し対象データがライトバッファに存在していると判定してもよい。

【0166】

この読み出し対象データが書き込み先ブロックから読み出し可能であり且つこの読み出し対象データがライトバッファに存在している場合(ステップS52のYES)、コントローラ4は、読み出し対象データを書き込み先ブロックから読み出す動作と読み出した読み出し対象データをホスト2に返す動作とを含む読み出し動作を実行する(ステップS53)。この場合、コントローラ4は、この読み出し対象データとして指定されたライトデータをライトバッファ内に維持する必要は無いので、この読み出し動作の開始の前に、またはこの読み出し動作の実行中に、このライトデータが格納されているライトバッファ内の領域を解放するための解放要求をホスト2に送信することができる。

【0167】

本実施形態のフラッシュストレージデバイス3においては、上述のライトコマンドのみならず、以下のタイプのライトコマンドを適用することもできる。

<ブロックアドレスを指定するライトコマンド>

このライトコマンドは、コマンドID、QoSドメインID、物理アドレス、タグ、長さ、ライトバッファアドレス、等を含んでもよい。

【0168】

コマンドIDはこのコマンド(このライトコマンド)の固有の識別子である。QoSドメインIDは、データが書き込まれるべきQoSドメインを一意に識別可能な識別子である。あるエンドユーザに対応するアプリケーションからのライト要求に応じてホスト2から送信されるライトコマンドは、このエンドユーザに対応するQoSドメインを指定するQoSドメインIDを含む。

【0169】

物理アドレスは、データが書き込まれるべきブロックのブロックアドレスを指定し、このブロック内の位置(例えば、ページアドレスおよびページ内のオフセット)は指定しない。

タグは、書き込まれるべきライトデータを識別するための識別子である。このタグは、上述したように、LBAのような論理アドレスであってもよいし、キー・バリュー・スト

10

20

30

40

50

アのキーであってもよいし、このキーのハッシュ値であってもよい。

【0170】

長さは、書き込まれるべきライトデータの長さを示す。この長さ（データ長）は、上述の粒度（Grain）の数によって指定されてもよいし、LBAの数によって指定されてもよいし、あるいはバイトによって指定されてもよい。

ライトバッファアドレスは、書き込まれるべきライトデータが格納されているホストメモリ（ホスト2のライトバッファ）内の位置を示す。ライトバッファアドレスは、データポインタとしても参照される。

【0171】

ホスト2からライトコマンドを受信した時、コントローラ4は、ライトコマンドによって指定されたブロックアドレスを有するブロック内の書き込み先位置を決定する。この書き込み先位置は、ページ書き込み順序の制約およびパッドページ等を考慮して決定される。そして、コントローラ4は、ホスト2からのデータを、ライトコマンドによって指定されたブロックアドレスを有するこのブロック内のこの決定された書き込み先位置に書き込む。

10

【0172】

このタイプのライトコマンドを使用するケースにおいては、ホスト2は、ブロック割り当てコマンドをフラッシュストレージデバイス3に送付することによって、特定のQoSドメイン用のブロックの割り当てをフラッシュストレージデバイス3に要求してもよい。ブロック割り当てコマンドはQoSドメインIDを指定する。コントローラ4がブロック割り当てコマンドを受信した場合、コントローラ4は、共通ブロック群602内の一つのブロック（フリーブロック）を選択し、選択されたブロックを、このQoSドメインに書き込み先ブロックとして割り当てる。そして、コントローラ4は、この割り当てられたブロックのブロックアドレスをホスト2に通知する。ホスト2は、この割り当てられたブロックのブロックアドレスを指定するライトコマンドをフラッシュストレージデバイス3に送信することができる。

20

【0173】

<ブロックアドレスおよびオフセットを指定するライトコマンド>

このライトコマンドは、コマンドID、QoSドメインID、物理アドレス、タグ、長さ、ライトバッファアドレス、等を含んでもよい。

30

コマンドIDはこのコマンド（このライトコマンド）の固有の識別子である。QoSドメインIDは、データが書き込まれるべきQoSドメインを一意に識別可能な識別子である。あるエンドユーザに対応するアプリケーションからのライト要求に応じてホスト2から送信されるライトコマンドは、このエンドユーザに対応するQoSドメインを指定するQoSドメインIDを含む。

【0174】

物理アドレスは、データが書き込まれるべきブロックのブロックアドレスと、このブロック内書き込み先の位置を示すオフセット（例えば、ページアドレスおよびページ内のオフセット）とを指定する。

タグは、書き込まれるべきライトデータを識別するための識別子である。このタグは、上述したように、LBAのような論理アドレスであってもよいし、キー・バリュー・ストアのキーであってもよいし、このキーのハッシュ値であってもよい。

40

【0175】

長さは、書き込まれるべきライトデータの長さを示す。この長さ（データ長）は、上述の粒度（Grain）の数によって指定されてもよいし、LBAの数によって指定されてもよいし、あるいはバイトによって指定されてもよい。

ライトバッファアドレスは、書き込まれるべきライトデータが格納されているホストメモリ（ホスト2のライトバッファ）内の位置を示す。

【0176】

以上説明したように、本実施形態によれば、リードコマンドによって読み出し対象デー

50

タとして指定されたデータの書き込み動作が終了しておらず、このデータがNAND型フラッシュメモリ5から読み出す事ができない場合には、ホスト2のライトバッファからこのデータが読み出されてホスト2に返される。したがって、ホスト2は、リードすべきデータがNAND型フラッシュメモリ5から読み出し可能であるか否かをチェックする処理およびライトバッファに向けたリード要求とフラッシュストレージデバイス3に向けたリードコマンドを選択的に発行する処理を行わずとも、フラッシュストレージデバイス3に向けたリードコマンドそれぞれを発行するというシンプルな処理を行うだけで、所望のデータをフラッシュストレージデバイス3から受け取ることができる。

【0177】

さらに、このリードコマンドの実行が完了するまで、このライトデータが格納されているライトバッファ内の領域の解放は禁止される。したがって、ライトバッファから読み出すべきことが決定されたデータ、またはライトバッファから読み出し中のデータなどがライトバッファから消失されてしまうという、ライトバッファ解放要求とリードコマンドとの衝突（解放要求/リードコマンド衝突）を回避することができる。そして、このリードコマンドの実行が完了した後に、このライトデータが格納されているライトバッファ内の領域を解放するための解放要求がホスト2に送信される。したがって、ライトバッファの解放制御のためのホスト側の負担も軽減することができる。

【0178】

なお、本実施形態では、ライトデータの書き込み動作が終了してこのライトデータがNAND型フラッシュメモリ5から読み出し可能となる前に、このライトデータを読み出し対象データとして指定するリードコマンドがホスト2から受信された場合、コントローラ4は、ライトバッファからこの読み出し対象データを読み出した。しかし、別の実施形態では、ライトデータの書き込み動作が終了してこのライトデータがNAND型フラッシュメモリ5から読み出し可能となる前に、このライトデータを読み出し対象データとして指定するリードコマンドがホスト2から受信された場合、コントローラ4は、ライトバッファからこの読み出し対象データをリードコマンドに対する応答として返すように要求する転送要求をホスト2に送信してもよい。この転送要求は、読み出し対象データが格納されているライトバッファ内の位置を示す記憶位置情報を含んでもよい。

【0179】

この転送要求の受信に応じて、ホスト2のホストFTL701は、ライトバッファから読み出し対象データを読み出し、読み出した読み出し対象データをリードバッファに転送する。そして、ホストFTL701は、転送要求に対するレスポンスをフラッシュストレージデバイス3に返す。コントローラ4は、このレスポンスの受信をこのリードコマンドの実行の完了として扱ってもよい。この場合、このレスポンスが受信されるまで、コントローラ4は、このライトデータが格納されているライトバッファ内の領域の解放を禁止する。このレスポンスの受信の後、コントローラ4は、ライトデータが格納されているライトバッファ内の領域を解放するための解放要求をホスト2に送信する。この場合、コントローラ4は、ライトデータの書き込み動作が終了してライトデータがNAND型フラッシュメモリ5から読み出し可能となり、且つこのレスポンスが受信された後、ライトデータが格納されているライトバッファ内の領域を解放するための解放要求をホスト2に送信してもよい。

【0180】

また、本実施形態では、不揮発性メモリとしてNAND型フラッシュメモリを例示した。しかし、本実施形態の機能は、例えば、MRAM (Magnetoresistive Random Access Memory)、PRAM (Phase change Random Access Memory)、ReRAM (Resistive Random Access Memory)、又は、FeRAM (Ferroelectric Random Access Memory)のような他の様々な不揮発性メモリにも適用できる。

【0181】

10

20

30

40

50

本発明のいくつかの実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。

【符号の説明】

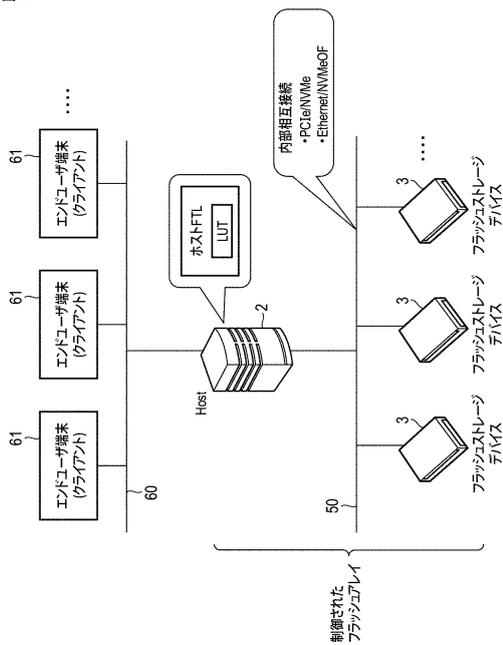
【0182】

2 ... ホスト、3 ... フラッシュストレージデバイス、4 ... コントローラ、5 ... NAND型フラッシュメモリ、21 ... ライト制御部、22 ... リード制御部、23 ... コピー制御部、24 ... ライトバッファ解放制御部。

【図面】

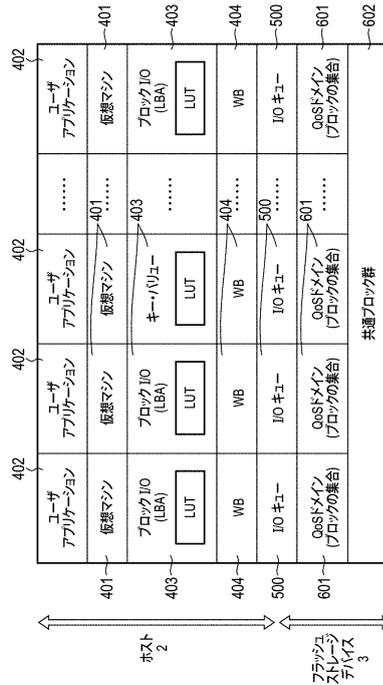
【図1】

図1

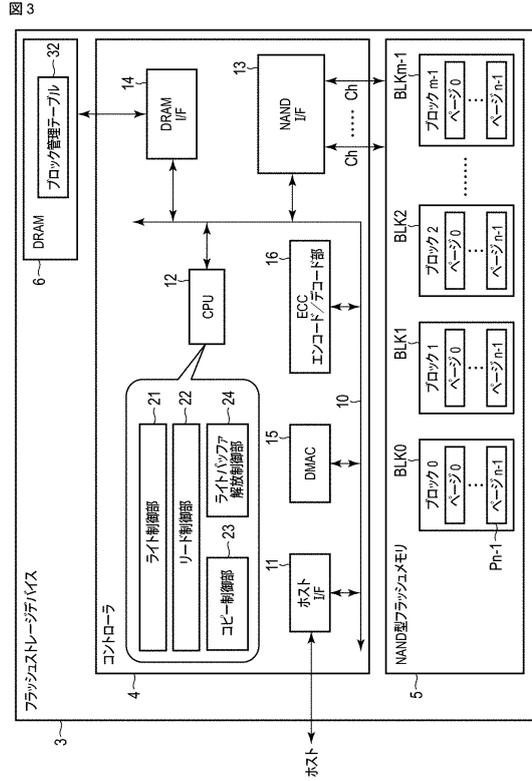


【図2】

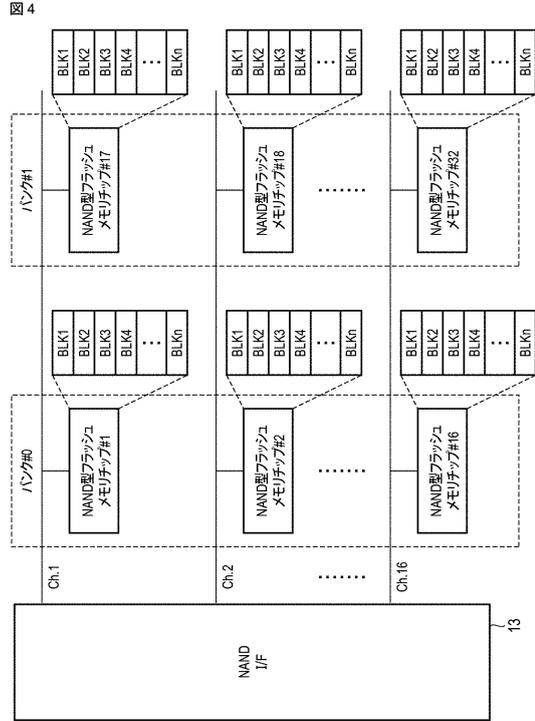
図2



【 図 3 】



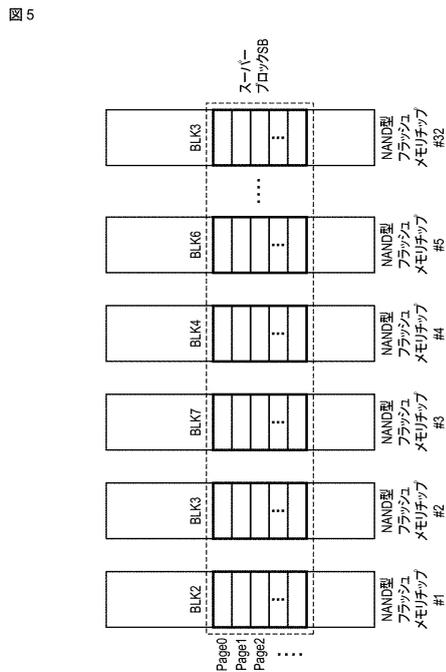
【 図 4 】



10

20

【 図 5 】



【 図 6 】

図6
ライトコマンド

	説明
コマンド ID	ライトコマンドのコマンド ID
QoSドメイン ID	データが書き込まれるべきQoSドメインの識別子
タグ	書き込まれるべきデータの識別子 (例 LBA, キーバリューストアのキー、このキーのハッシュ値など)
長さ	書き込まれるべきデータの長さ
ライトバッファアドレス	書き込まれるべきデータが格納されているライトバッファ内の位置

30

40

50

【 図 7 】

図 7

アドレス記録要求

	説明
QoSドメイン ID	データが書き込まれるべきQoSドメインの識別子
タグ	書き込まれるべきデータの識別子 (例 LBA、キー・バリューストアのキー、このキーのハッシュ値など)
物理アドレス	物理アドレスはデータが書き込まれた物理記憶位置を示す。 物理アドレスはブロックアドレスおよびオフセットによって表される。
長さ	書き込まれたデータの長さ

【 図 8 】

図 8

リードコマンド

	説明
コマンド ID	リードコマンドのコマンド ID
QoSドメイン ID	読み出されるべきデータが格納されているQoSドメインの識別子
物理アドレス	物理アドレスは読み出されるべきデータが格納されている 物理記憶位置を示す。 物理アドレスはブロックアドレスおよびオフセットによって表される。
長さ	読み出されるべきデータの長さ
リードバッファアドレス	データが転送されるべきリードバッファ内の位置

10

【 図 9 】

図 9

QoSドメインクリエートコマンド

	説明
コマンド ID	QoSドメインクリエートコマンドのコマンド ID
QoSドメイン ID	作成すべきQoSドメインの識別子
容量	作成すべきQoSドメインの容量

【 図 10 】

図 10

ライトバッファ解放要求

	説明
ライトコマンド ID	ライトバッファ内の解放可能なライトデータに対応する ライトコマンド識別子 (ライトバッファ解放要求は、ライトバッファ内の解放可能領域を 示す記憶位置情報を含んでもよい)

20

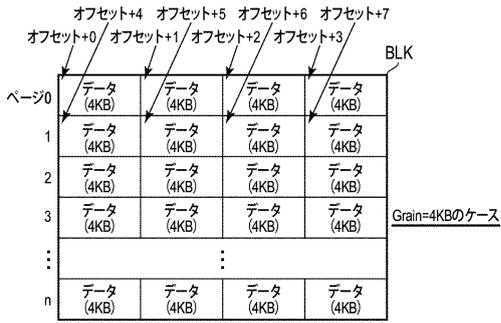
30

40

50

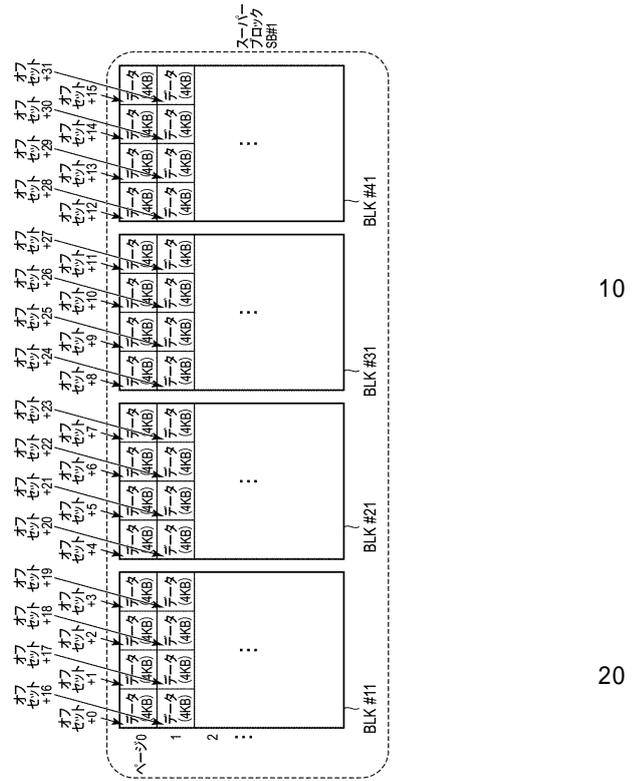
【 図 1 1 】

図 11



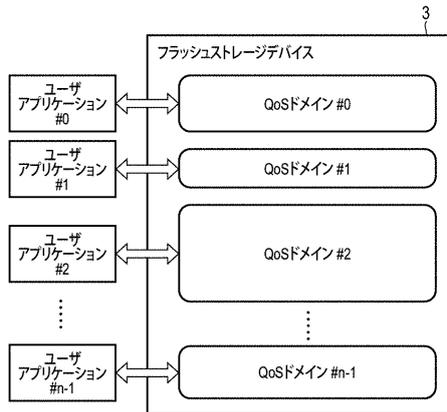
【 図 1 2 】

図 12



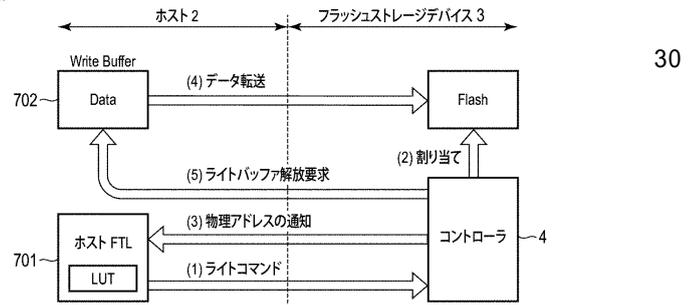
【 図 1 3 】

図 13



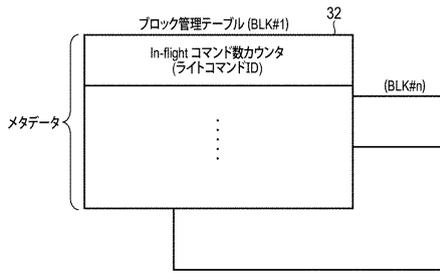
【 図 1 4 】

図 14



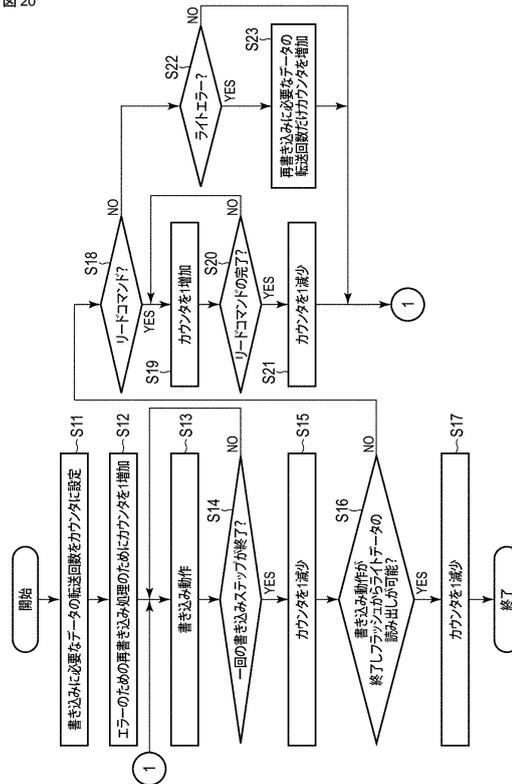
【 図 19 】

図 19



【 図 20 】

図 20

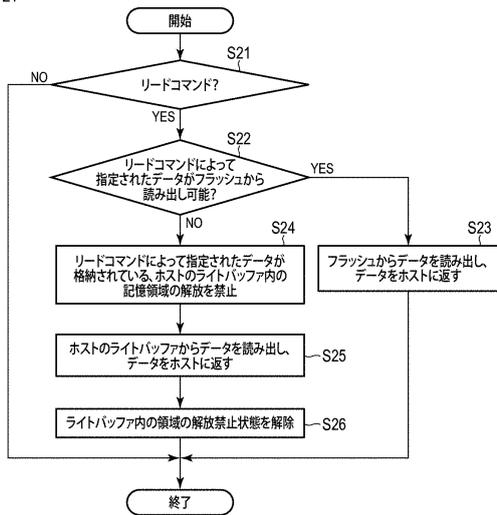


10

20

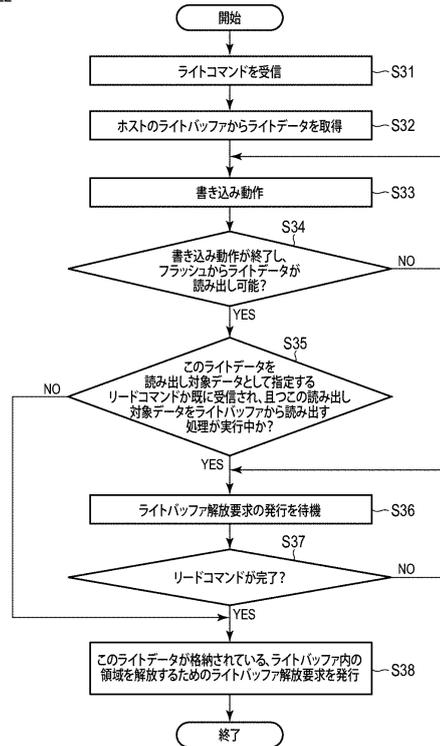
【 図 21 】

図 21



【 図 22 】

図 22



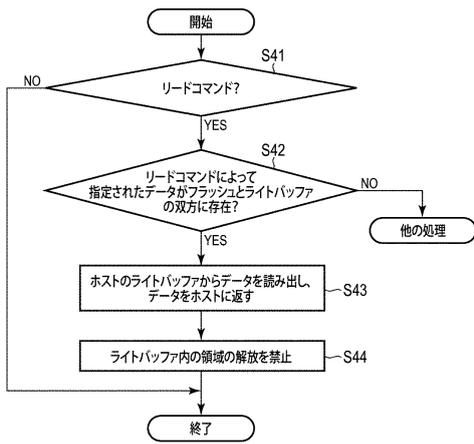
30

40

50

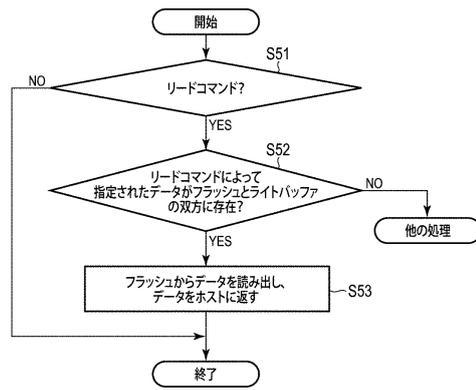
【 図 2 3 】

図 23



【 図 2 4 】

図 24



10

20

30

40

50