

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5719013号
(P5719013)

(45) 発行日 平成27年5月13日 (2015. 5. 13)

(24) 登録日 平成27年3月27日 (2015. 3. 27)

(51) Int. Cl.	F I				
G06F 3/06	(2006.01)	G06F	3/06	301E	
G06F 3/08	(2006.01)	G06F	3/08	H	
G06F 13/10	(2006.01)	G06F	13/10	340A	
		G06F	3/06	304N	

請求項の数 13 (全 15 頁)

(21) 出願番号	特願2013-250303 (P2013-250303)	(73) 特許権者	503116280
(22) 出願日	平成25年12月3日 (2013. 12. 3)		エイチジーエスティーネザerlandビーブイ
(65) 公開番号	特開2014-110058 (P2014-110058A)		イ
(43) 公開日	平成26年6月12日 (2014. 6. 12)		オランダ国 1101 シーエム アムステルダム, ルナ アレナ, ヘリケルベルグ
審査請求日	平成25年12月3日 (2013. 12. 3)		ヴェヘ 238
(31) 優先権主張番号	13/705, 103	(74) 代理人	110000154
(32) 優先日	平成24年12月4日 (2012. 12. 4)		特許業務法人はるか国際特許事務所
(33) 優先権主張国	米国 (US)	(72) 発明者	ジョアオ アルカンタラ
			アメリカ合衆国 92604 カリフォルニア州 アーバイン レイクグラス 15
		(72) 発明者	ゾルタン スッポセヴ
			ドイツ連邦共和国 85716 ウンターシュライスハイム エッシェン シュトラーセ 74

最終頁に続く

(54) 【発明の名称】 フラッシュ読取コマンド完了の時間推定に基づくホスト読取コマンドリターンのリオーダーリング

(57) 【特許請求の範囲】

【請求項 1】

ソリッドステートドライブ (SSD) 装置がホストから受け取った読取コマンドに応答する前記ホストへのリターンを管理するための方法であって、

前記ホストからデータ読取コマンドを受け取るステップと、

それぞれの不揮発性メモリダイに送られた動作ごとに動作コストを生成するステップであって、各動作コストを周期的なイベントに基づいて経時でデクリメントするステップと、

前記読取コマンドに対応する前記それぞれの不揮発性メモリダイに送られた動作ごとの現在の動作コストに基づいて、前記読取コマンドの総コストを計算するステップと、

前記読取コマンドの前記総コストに基づいて、前記ホストから受け取った他のデータ読取コマンドに対して、要求されたデータをどんな順序で前記ホストに返すべきかを決定するステップと、

前記要求されたデータを、決定された前記順序で前記ホストに返すステップと、を含む方法。

【請求項 2】

動作コストを生成するステップが、(1) ターゲット不揮発性メモリダイがアイドル状態であるか否か、(2) ターゲット不揮発性メモリダイが消去動作でビジー状態であるか否か、(3) ターゲット不揮発性メモリダイがページプログラム動作でビジー状態であるか否か、および(4) ターゲット不揮発性メモリダイが読取動作でビジー状態であるか否

か、のうちの少なくとも1つに基づいて、動作コストを生成するステップを含む、
請求項1に記載の方法。

【請求項3】

動作コストを生成するステップが、不揮発性メモリダイに送られた対応する動作の完了までの時間を推定するステップを含む、

請求項1に記載の方法。

【請求項4】

動作コストを生成するステップが、不揮発性メモリダイ上の動作の物理的な位置に基づいて動作コストを生成するステップを含む、

請求項1に記載の方法。

10

【請求項5】

動作コストを生成するステップが、不揮発性メモリダイの動作履歴に基づいて動作コストを生成するステップを含む、

請求項1に記載の方法。

【請求項6】

前記要求されたデータを前記ホストに返すステップが、第1の読取コマンドに対応する第1の動作コストよりも、第2の読取コマンドに対応する第2の動作コストが少ないことに基づいて、前記第1の動作コストを有する第1の要求されたデータの前に、前記第2の動作コストを有する第2の要求されたデータを前記ホストに返すステップを含む、

請求項1に記載の方法。

20

【請求項7】

前記第1の読取コマンドを前記ホストから受け取った後、前記第2の読取コマンドを前記ホストから受け取った、

請求項6に記載の方法。

【請求項8】

前記動作コストが推定動作コストであり、

動作コストを生成するステップが、

前記推定動作コストと、不揮発性メモリダイに送られた対応する動作の完了までの時間との差を特定するステップと、

前記差に基づいて、前記推定動作コストを調整するステップと、

を含む、

請求項1に記載の方法。

30

【請求項9】

不揮発性メモリダイと、前記不揮発性メモリダイ上の動作を管理するためのコントローラと、を含むソリッドステートドライブ(SSD)であって、

前記コントローラがホストから受け取ったデータ読取コマンドの動作コストの要求を前記コントローラから受け取り、

それぞれの不揮発性メモリダイに送られた動作ごとの完了までの時間を示す動作コストを生成し、前記対応するカウンタに前記動作コストを記憶することを含み、各カウンタがそれぞれの不揮発性メモリダイに関連付けられたカウンタを管理し、

40

周期的なイベントを生成するタイマに基づいて各動作コストをデクリメントし、

前記コントローラが受け取った前記読取コマンドに対応するそれぞれの不揮発性メモリダイに送られた動作ごとの現在の動作コストを前記コントローラに返し、前記現在の動作コストが、前記読取コマンドに対応するそれぞれの不揮発性メモリダイに関連付けられた各カウンタの現在の値に基づく、

ように構成された動作モニタと、

前記それぞれの不揮発性メモリダイに送られた動作ごとの現在の動作コストに基づいて、前記読取コマンドの総コストを計算し、

前記読取コマンドの前記総コストに基づいて、前記ホストから受け取った他のデータ読取コマンドに対して、前記要求されたデータをどんな順序で前記ホストに返すべきかを決

50

定する、

ように構成された前記コントローラと、
を含む、SSD。

【請求項10】

1つまたは複数のチャンネルバス制御をさらに含み、

カウンタを管理する前記動作モニタが、前記1つまたは複数のチャンネルバス制御からの情報に基づいて動作ごとの動作コストを生成するステップであって、前記情報から、不揮発性メモリダイに送られた動作の期間が、前記不揮発性メモリダイに関連付けられたチャンネルバスを監視することにより決定されるステップを含む、

請求項9に記載のSSD。

10

【請求項11】

前記不揮発性メモリダイ上の消去動作の完了時間に対応する値を管理するための1つまたは複数の消去コストレジスタと、

前記不揮発性メモリダイ上の読取動作の完了時間に対応する値を記憶するための1つまたは複数の読取コストレジスタと、

前記不揮発性メモリダイ上の書込動作の完了時間に対応する値を記憶するための1つまたは複数のプログラムコストレジスタと、

を含み、

動作ごとに動作コストを生成するステップが、前記消去コストレジスタ、前記読取コストレジスタおよび前記プログラムコストレジスタに関連する1つまたは複数のそれぞれの値に基づく、

請求項9に記載のSSD。

20

【請求項12】

カウンタを管理する前記動作モニタが、

カウンタからの動作コストと、対応する不揮発性メモリダイに送られた対応する動作の完了までの実際の時間との差を特定するステップと、

前記差に基づいて、前記カウンタの前記動作コストを調整するステップと、

を含む、

請求項9に記載のSSD。

30

【請求項13】

1つまたは複数の命令シーケンスを記憶するソリッドステートドライブコントローラであって、1つまたは複数のプロセッサによって実行される時、

ホストからデータ読取コマンドを受け取り、

それぞれの不揮発性メモリダイに送られた動作ごとに動作コストを生成し、各動作コストを周期的なイベントに基づいて経時でデクリメントし、

前記読取コマンドに対応する前記それぞれの不揮発性メモリダイに送られた動作ごとの現在の動作コストに基づいて、前記読取コマンドの総コストを計算し、

前記読取コマンドの前記総コストに基づいて、前記ホストから受け取った他のデータ読取コマンドに対して、要求されたデータをどんな順序で前記ホストに返すべきかを決定し

40

前記要求されたデータを、決定された前記順序で前記ホストに返す、
性能をもたらすコントローラ。

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は、ソリッドステートドライブ(SSD)からホストにデータを返す順序を決定するための推定機構に関する。

【背景技術】

【0002】

50

ソリッドステートドライブ（SSD）は、データを永続的に記憶するためにメモリユニットとして集積回路アセンブリを使用するデータ記憶装置である。SSD技術は、従来のブロック入力/出力（I/O）ハードディスクドライブ（HDD）と互換性がある電子インターフェースを使用する。SSDはいかなる移動機械部品も用いず、このことにより、可動読取書込ヘッドを使用し、回転している磁気ディスク上にデータを記憶させる従来のHDDと区別される。

【0003】

現在、ほとんどのSSDは、NAND型フラッシュメモリを使用する。NAND型フラッシュメモリは、電力がない状態でデータを保持し、（フラッシュメモリはEEPROMから開発されたため）電氣的に消去し、再プログラムすることができる不揮発性メモリの形態である。さらに、ハイブリッドドライブは、頻繁にアクセスされるデータの性能を改善するために典型的にはハードディスクドライブおよびSSDキャッシュを含む、SSDおよびHDDの両方の特徴を同じユニット内に併せ持つ。フラッシュメモリは、浮遊ゲートトランジスタから作られたメモリセルのアレイに情報を記憶する。

10

【0004】

エンタープライズSSDの実装は、例えばサーバにおいて、典型的にデスクトップならびにラップトップのHDDドライブおよびSSDドライブと異なり、より高い性能基準が保たれている。すなわち、エンタープライズSSDは、比較的高いI/O性能を提供することが典型的に要求される。したがって、最適な読取専用および読取/書込混合性能は、多くのSSD開発技術者の目標である。

20

【発明の概要】

【発明が解決しようとする課題】

【0005】

本発明の実施形態は、ソリッドステートドライブ（SSD）記憶装置で使用するのための、読取コマンドにตอบสนองするホストへのデータリターンに関する。最適な読取専用および読取/書込混合性能は、データをホストに返す順序に影響を受け、データをホストに返す順序は、ホストが読取コマンドを送った順序に必ずしも一致しない。

【課題を解決するための手段】

【0006】

一実施形態によれば、ソリッドステートドライブ（SSD）の動作モニタは、対応するフラッシュダイ上の読取動作を完了する推定時間を特徴付ける数値指標、すなわち動作コストを管理するのに使用するカウンタを管理するために構成される。動作モニタは、カウンタをデクリメントする周期的なイベントを生成するタイマを起動する。各カウンタ内に記憶される値は、対応するダイに送られたフラッシュ動作に対して生成され、一般に動作の履歴および物理的な位置に基づく。読取コマンドが特定のダイに送られる予定の場合は常に、コントローラファームウェアは例えば、その特定のフラッシュ読取動作の時間推定を検索することができる。この情報に基づいて、データをホストに返す最適な順序を決定することができる。この順序は、ホストへのデータ転送を予定しプログラムするために使用され、最適な順序は、他の読取コマンドによってブロックされる読取コマンドの数が最小であるシーケンスを表し、これはブロックコマンドが、ホストに返すことができる、それらの対応データを有していない場合である。

30

40

【0007】

SSD装置が受け取った読取コマンドにตอบสนองするホストへのリターンを管理するための方法の一実施形態によれば、ホストからデータコマンドを受け取り、それぞれの不揮発性メモリダイに送られた動作ごとに動作コストが生成される。各動作コストは、周期的なイベント（例えば、時間刻み）を生成するタイマに基づいてデクリメントされる。このように、対応するダイを読み取る時間が経時で減少すると、カウンタ内に示される動作コストが経時で減少する。ホストから受け取った読取コマンドの総コストは、読取コマンドに関連付けられた各それぞれの不揮発性メモリダイに送られた動作ごとの現在の動作コストに基づいて計算される。読取コマンドの総コストに基づいて、ホストから受け取った他の読

50

取コマンドに対して、要求されたデータをホストに返す順序が決定される。

【0008】

発明の概要の章で検討した実施形態は、本明細書で検討するすべての実施形態を示唆、説明、または教示することを意図していない。したがって本発明の実施形態は、この章で検討したものの追加の、またはそれとは異なる特徴を含んでもよい。

【0009】

本発明の実施形態は、添付図面の図において限定としてではなく一例として示し、図中の同じ参照符号は同様の要素を指す。

【図面の簡単な説明】

【0010】

【図1】本発明の実施形態が実施され得る例示的動作環境を示すブロック図である。

【図2】本発明の実施形態が実施され得る、メモリ動作モニタに注目した例示的動作環境の一部を示すブロック図である。

【図3】本発明の一実施形態による、メモリ動作モニタを示すブロック図である。

【図4】本発明の一実施形態による、ホストへのリターンを管理するための方法を示すフローチャートである。

【発明を実施するための形態】

【0011】

ソリッドステートドライブ（SSD）記憶装置で使用するための、ホストからの読取コマンドにตอบสนองするホストへのデータリターンを管理するための手法を記載する。以下の記述では、説明の目的で、本明細書に記載の本発明の実施形態を完全に理解するために多くの特定の詳細を述べる。しかしながら本明細書に記載の本発明の実施形態は、これらの特定の詳細なしに実施してもよいことが明らかであろう。他の事例では、本明細書に記載の本発明の実施形態を不必要に曖昧にしないように、よく知られている構造および装置をブロック図形式で示す。

【0012】

〔ソリッドステートドライブ構造〕

本発明の一実施形態は、ソリッドステートドライブ（SSD）記憶装置を管理するために使用してもよい。図1は、本発明の実施形態が実施され得る例示的動作環境を示すブロック図である。図1は、主要な通信インターフェース106を介してホストデバイス104と通信可能に接続されたSSD102を備えた一般的なSSD構造100を示す。本発明の一実施形態は、図1に示したような構造に限定されず、むしろ実施形態は、SSD以外の動作環境に適用可能であってもよく、図1に示した構造以外のSSD構造内で実施してもよい。例えば実施形態は、データを書き込み、かつ読み取るための不揮発性メモリ記憶構成要素に依存する他の環境において動作するように実施してもよい。

【0013】

ホストデバイス104（または単に「ホスト」）は、特に1つもしくは複数のメモリデバイスにデータI/O要求または呼び出しを行う任意の種類の実用ハードウェアまたは計算ソフトウェアを広く示す。例えば、ホスト104は、コンピュータ、タブレット、携帯電話、または一般に、メモリを含むもしくはメモリと相互作用する任意の種類の実用装置上で実行するオペレーティングシステムであってもよい。ホストデバイス104をSSD102に連結する主要インターフェース106は、例えばコンピュータの内部バスまたは通信ケーブルまたは無線通信リンクなどであってもよい。

【0014】

図1に示した例示的SSD102は、インターフェース110と、コントローラ112（例えば、その中にファームウェア論理を有するコントローラ）と、アドレッシング部114の機能ブロックと、データバッファキャッシュ116と、エラー訂正コード（ECC）118と、1つまたは複数の不揮発性メモリ構成要素120a、120b、120nとを含む。

【0015】

10

20

30

40

50

インターフェース 110 は、構成要素間、すなわち本内容において SSD 102 とホストデバイス 104 との間の相互作用の点であり、ハードウェアとソフトウェアの両方のレベルで適用可能である。これにより、入力/出力システムおよび関連するプロトコルを介して、構成要素同士が通信することができる。ハードウェアインターフェースは、機械的、電気的および論理的な信号によって、インターフェースおよび、信号の順序を付けるためのプロトコルにおいて典型的に表される。一般的かつ標準的なインターフェースのうちのいくつかの限定しない例には、SCSI (小型コンピュータシステムインターフェース)、SAS (シリアル接続 SCSI) および SATA (シリアル ATA) がある。

【0016】

SSD 102 はコントローラ 112 を含み、コントローラ 112 は不揮発性メモリ 120 a、120 b、120 n をホストデバイス 104 へなど、不揮発性メモリ構成要素 (例えば NAND フラッシュ) をホストへ橋架けする電子機器を組み込む。コントローラは典型的に、ファームウェアレベルコードを実行し、かつ SSD 性能において重要な因子である埋め込みプロセッサである。

【0017】

コントローラ 112 は、アドレッシング部 114 の機能ブロックを介して不揮発性メモリ 120 a、120 b、120 n とインターフェースする。アドレッシング部 114 の機能は、例えば、ホスト 104 から、SSD 102 上の、すなわち SSD 102 の不揮発性メモリ 120 a、120 b、120 n 上の、対応する物理的なブロックアドレスまでの論理ブロックアドレス (LBAs) 間のマッピングを管理するよう動作する。不揮発性メモリページおよびホストセクタは寸法が異なるため、SSD は、セクタへのホスト書込みデータまたはセクタからのホスト読取データを SSD が変換することができるデータ構造と、そのデータが実際に配置される物理的な不揮発性メモリページとを構築し維持しなければならない。このテーブル構造、すなわち「マッピング」は、DRAM または、コントローラ 112 およびアドレッシング部 114 にアクセス可能ないくつかの他のローカル揮発性メモリ構成要素などの SSD の揮発性メモリ 122 内に 1 セッション中、構築し維持されてもよい。あるいは、テーブル構造は、不揮発性メモリ 120 a、120 b、120 n などの SSD の不揮発性メモリ内に複数のセッションにわたって、より永続的に維持されてもよい。

【0018】

アドレッシング部 114 は、不揮発性メモリ 120 a、120 b、120 n に加えて、データバッファキャッシュ 116 およびエラー訂正コード (ECC) 118 と相互作用する。SSD 102 のデータバッファキャッシュ 116 は、ハードディスクドライブ内のキャッシュと類似した、キャッシュとしての DRAM を典型的に使用する。データバッファキャッシュ 116 は、キャッシュデータの今後の要求を速めるためのキャッシュとして役割を果たすと同時に、不揮発性メモリ構成要素間のデータ送信用のバッファまたは記憶領域として役割を果たす。データバッファキャッシュ 116 は典型的に、揮発性メモリと共に実装されるため、揮発性メモリに記憶されたデータはキャッシュに永久的には記憶されず、すなわちデータは永続的ではない。

【0019】

ECC 118 は、重複データまたはパリティデータをメッセージに追加するシステムであり、送信工程中、または記憶時に多数のエラーが生じた場合、受信機によって回復することができるようになっている。

【0020】

最後に、SSD 102 は 1 つまたは複数の不揮発性メモリ 120 a、120 b、120 n の構成要素を含む。限定しない例として、不揮発性メモリ構成要素 120 a、120 b、120 n は、フラッシュメモリ (例えば NAND もしくは NOR フラッシュ)、または、現在もしくは将来利用可能な他の種類のソリッドステートメモリとして実装してもよい。不揮発性メモリ 120 a、120 b、120 n の構成要素は、データが永続的に記憶される実際のメモリ電子構成要素である。SSD 102 の不揮発性メモリ 120 a、120

10

20

30

40

50

b、120nの構成要素は、ハードディスクドライブ(HDD)記憶装置内のハードディスクの類似物と考えることができる。

【0021】

ホストからの読取コマンドに応答するホストへのデータリターンを管理する本発明の実施形態を記載する。結果として、読取コマンドに関連するデータをホストに返すべき最適な順序を決定することができ、より高性能なSSD記憶装置が達成可能である。

【0022】

[メモリ動作モニタ]

図2は、本発明の一実施形態が実施され得る、メモリ動作モニタに注目した例示的動作環境の一部を示すブロック図である。例示的動作環境200は、CPU204に通信可能に連結されたメモリ動作モニタ202(以下「動作モニタ」と呼ぶ)を含む。動作モニタ202は、1つまたは複数のチャンネル206a、206b、206nにさらに連結され、各チャンネルは1つまたは複数の不揮発性メモリダイに連結される。例示的な目的として、図2は、ダイ208a0、ダイ208a1、ダイ208a2に連結され、それらと通信するチャンネル206aと、ダイ208b0、208b1、208b2に連結され、それらと通信するチャンネル206bと、ダイ208n0、208n1、208n2に連結され、それらと通信するチャンネル206nとを示す。図2に示したチャンネルおよび対応するメモリダイの数は、例を目的としており、本発明の実施形態は示した厳密な構造に限定されないことに留意されたい。

【0023】

実施形態によれば、動作モニタ202は図3を参照して説明するように動作する。一般に、動作モニタ202は、コントローラが実行可能な作業を行うのを支援するように構成されるハードウェアアクセラレータと見なすことができる。一実施形態によれば、動作モニタ202は、コントローラ112(図1)内に実装されるが、本発明の実施形態の実施はそれほど限定されない。したがって動作モニタ202は、本質的にSSD102(図1)内のいかなる場所にも、例えば1つまたは複数のメモリダイ208a0~208n2内にさえ実装することができるであろう。動作モニタ202は、ファームウェアまたはハードウェア内に実装されてもよい。図3を参照してより詳細に説明するように、動作モニタ202は推定器の機能として一般に機能する。

【0024】

動作環境200は、動作モニタ202の機能性に関連する指示、手順などを実行する専用または共用のCPU204を含んでもよい。CPU204は、メモリダイ208a0~208n2に関連する動作を行うための実行可能な論理を実行することができるSSD102のプロセッサである。同様に、チャンネル206a、206b、206nの各々は、それらそれぞれのメモリダイと同様に、互いにインターフェースする。チャンネルは、複数の実体または構成要素の間の通信「パイプライン」と考えることができる。したがって、SSD102のチャンネル206a、206b、206nは、動作モニタ202と各チャンネルのそれぞれのメモリダイとの間、およびCPU204と各チャンネルのそれぞれのメモリダイとの間の通信回線および動作回線として機能する。

【0025】

フラッシュメモリは、一種の不揮発性メモリである。フラッシュメモリは、一種の電氣的消去プログラム可能読取専用メモリ(EEPROM)である。メモリダイは、不揮発性メモリのハードウェアユニットである。フラッシュメモリダイは、個別の集積回路ダイであり、そのようなダイは少なくとも1つのフラッシュメモリ集積回路を含む。ダイは、データがフラッシュメモリ上に記憶される低レベル媒体である。

【0026】

図3は、本発明の一実施形態による、メモリ動作モニタを示すブロック図である。図3のメモリ動作モニタ202は、複数のダイカウンタを含み、ダイカウンタの各々は、1つまたは複数の対応するメモリダイに関連付けられている。例えば、ダイカウンタ302a0、302a1、302a2は、対応するダイ208a0、208a1、208a2(図

10

20

30

40

50

2)に関連付けられ、ダイカウンタ302b0、302b1、302b2は、対応するダイ208b0、208b1、208b2(図2)に関連付けられ、またダイカウンタ302n0、302n1、302n2は、対応するダイ208n0、208n1、208n2(図2)に関連付けられている。メモリダイの数と同様に、図3に示したダイカウンタの数は例を目的としており、本発明の実施形態は示した厳密な構造およびダイカウンタの数に限定されないことに留意されたい。

【0027】

ダイカウンタ302a0~302n2は、対応するダイ208a0~208n2(図2)上の読取動作に関連する完了時間を追跡するために使用されるデータ構造である。特定のダイ上の読取動作を完了する推定時間は、対応するダイカウンタに記憶されている「動作コスト」数値指標によって特徴付けられる。ダイカウンタ302a0~302n2に記憶される動作コスト値は、ダイ208a0~208n2に送られたすべてのフラッシュ動作に対して生成され、一実施形態によれば、ダイ履歴および動作の物理的な位置に基づく。

10

【0028】

メモリ動作モニタ202は、ダイカウンタ302a0~302n2に関連付けられたタイマ304を含む。タイマ304は、時間刻みなどのダイカウンタ302a0~302n2をすべてデクリメントする周期的なイベントを生成する。限定しない例として、タイマ304は1 μ sの時間刻みを生成してもよい。タイマ304は一旦初期化されると、いずれの読取コマンドにかかわらず時を刻む。タイマ304は、例えば、設定可能であることが多い固定周波数で、デクリメントするデジタルカウンタとして実装されてもよい。より精巧なタイマでは、比較論理を使用して、タイマ値を、ソフトウェアまたは他の論理によって設定された特定値と比較してもよい。この比較論理は、タイマ値が既定値と一致すると、何らかの作用を引き起こす。この種のタイマは、例えばパルスを生成するために使用することができる。一実施形態によれば、パルスを生成するために比較論理を使用するデジタルカウンタは、ダイカウンタ302a0~302n2内の値をデクリメントするために使用される。しかしながら、使用されるタイマの種類は実施態様によって異なってもよい。

20

【0029】

このように、動作モニタ202は、動作コストを備えたダイカウンタ302a0~302n2をロードし、タイマ304が時間を刻むごとにこれらの動作コスト値をデクリメントすることによって、ソリッドステートドライブを含むすべてのフラッシュダイ上のフラッシュ読取動作の完了時間を推定かつ管理する。

30

【0030】

メモリ動作モニタ202は、1つまたは複数のプログラムコストレジスタ314a、314b、314c、314nを含む。プログラムコストレジスタ314a~314nは、書込動作としても知られているページプログラム動作を完了するのにかかる時間を考慮するのに使用される。図3に示したプログラムコストレジスタの数は、例を目的としており、本発明の実施形態は示した厳密な構造およびプログラムコストレジスタの数に限定されない。コストレジスタの数は、設計上の選択の問題であってもよく、例えば不揮発性メモリ120a~120n(図1)およびSSD102(図1)構造に基づいてもよい。しかしながら、一実施形態によれば、プログラムコストに対して4つの取り得る値を保持するために4つのコストレジスタが使用され、そのようなプログラムコスト値は、不揮発性メモリ120a~120n内のページ位置によるばらつきをプログラムすることを考慮する。一般的なプログラムコストレジスタ314a~314nの値は不揮発性メモリ120a~120nから予め決定することができるが、これらの値は要求されたデータが記憶される不揮発性メモリ上の関連位置に基づいて、および不揮発性メモリの種類などに基づいて変化する。

40

【0031】

動作モニタ202は、消去コストレジスタ310および読取コストレジスタ312を含

50

む。消去コストレジスタ310は、消去動作を完了するのにかかる時間を示す値を記憶するために使用され、読取コストレジスタ312は、読取動作を完了するのにかかる時間を示す値を記憶するために使用される。単純にするため、単一の消去コストレジスタ310および単一の読取コストレジスタ312を図3に示した。本明細書で記載したより広い教示を比較的単純に実施するため、図示したように本発明の実施形態を実施することができるであろう。すなわち、単純な実施では、単純なレジスタを利用して、様々なコスト情報を記憶してもよい。

【0032】

しかしながら、読取、消去およびプログラムのコストは、SSDの寿命にわたって、およびダイ208a0~208n2内の様々なブロックにわたって変化する。したがって、実施形態の実施は、コストが推定される粒度レベルの点で異なってもよい。よって、実施形態の比較的より複雑な実施は、メモリ308を使用して、ダイ、ブロックまたはページ単位で、およびドライブの寿命にわたるこれらのコスト変化として、読取、消去、プログラムのコストの多くのばらつきを記憶してもよい。コスト数値指標の多くのばらつきを記憶する、より大きいメモリ308は、より正確にコスト推定を行いやすいため、コスト粒度対推定精度は実施レベルで管理することができるトレードオフである。

【0033】

プログラム(すなわち書込動作)コスト、読取コスト、および消去コストが考慮される1つの理由は、不揮発性メモリダイ208a0~208n2上の読取動作の完了時間、ひいてはデータをホスト104(図1)に返す最適な順序が、例えば、ホスト書込動作およびガーベジコレクション動作(すなわちランダムミックスシナリオ)による読取動作、ならびに、同じダイ(すなわちランダム読取シナリオ)をターゲットとする他のホストコマンドによる読取動作と同時に実行する書込、消去、読取の動作により影響を受けることである。したがって、データをホスト104(図1)に返すべきコマンドの最適な順序は、少なくとも部分的に、背景動作と同様にランダムミックスシナリオおよびランダム読取シナリオにより、ホストがSSD102(図1)にコマンドを送った順序に必ずしも一致しない。

【0034】

動作モニタ202の制御論理306は、ダイカウンタ302a0~302n2を取り込んで管理し、かつ、プログラムコストレジスタ314a~314n、消去コストレジスタ310および読取コストレジスタ312などの様々なレジスタ、ならびに動作モニタ202の代替の実施形態で実施されてもよい任意の他のレジスタを構成するように実行する論理である。制御論理306はまた、複数のダイ208a0~208n2(図2)にわたる、または複数の不揮発性メモリ120a~120n(図1)にわたる様々な動作コスト(例えば、消去コスト、読取コスト)を平均するなど、現在または実際の動作完了時間に基づいてレジスタ動作コストを調整するなどの、データをホスト104に返す最適な順序の決定を支援して、他の作業および計算を行ってもよい。制御論理306の範囲は、限定されるものではなく、他の機能を行うためのそのような論理の展開が具体的に考慮される。

【0035】

動作モニタ202は、1つまたは複数のチャンネルバス制御316a、316b、316nを含み、各々はそれぞれのチャンネル206a、206b、206n(図2)に対応する。一実施形態によれば、チャンネルバス制御316a~316nは、例えば、その上でチャンネル206a~206nがダイ208a0~208n2への動作を促進するチャンネルバス(図2)をスヌープすることにより、ダイ208a0~208n2(図2)上の動作の開始および終了を検出するように構成される。そのような検出された動作活動は、ダイカウンタ302a0~302n2の動作コスト値を取り込んで管理するために制御論理306によって使用することができる。

【0036】

動作モニタ202のメモリ308は、様々な目的での動作モニタ202によって使用可能なメモリであり、それらのうちのいくつかは、所望のコスト数値指標の粒度、および動

10

20

30

40

50

作モニタ 202 内で具体化されてもよい自動調整特徴の文脈において、より詳細に本明細書で述べる。メモリ 308 の種類および形態は限定されず、関連する設計制約、目標などに基づいて実施態様によって異なる。

【0037】

[ホストへのリターンを管理するための方法]

図 4 は、本発明の一実施形態による、ホストへのリターンを管理するための方法を示すフローチャートである。図 4 に示した工程は、SSD 102 (図 1) などの SSD 内で実施してもよい。さらに具体的には、図 4 に示した工程は、一実施形態による SSD 102 のコントローラ 112 (図 1) 内で実施してもよい。さらに一実施形態によれば、図 4 の工程はメモリ動作モニタ 202 (図 2) 内で実施する。プロセス論理は、SSD 102 内のアナログもしくはデジタルハードウェア回路として、または SSD 102 内の CPU 204 (図 2) もしくは MPU などのプロセッサによって実行されるファームウェア命令として実施してもよい。

10

【0038】

図 4 のブロック 402 では、データ読取コマンドをホストから受け取る。例えば、データ読取コマンドは、SSD 102 (図 1) のインターフェース 110 (図 1) を介してコントローラ 112 (図 1) においてホストデバイス 104 (図 1) から受け取る。待ち行列型コマンドシステムでは、ホストは多数の、すなわち SSD 102 が同時かつ瞬時に処理することができる量を超える読取および書込コマンドを SSD 102 などの記憶装置に送ることができる。顕著には、ホストがコマンドを送る順序は、必ずしも、最良の SSD 性能のための、ホストに回答し、またデータを返す最良の順序ではない。これは、例えばターゲットダイがアイドル状態であるか否か、および、ターゲットダイが消去動作、ページプログラム動作または読取動作でビジー状態であるか否かに基づいて、フラッシュメモリダイからデータを読み取る待ち時間が異なることが一因である。

20

【0039】

コントローラ 112 (図 1) は、不揮発性メモリ 120a ~ 120n (図 1) 内のどこにデータが記憶されているかの「知識」を有する。このように、コントローラ 112 は、要求を受け取る順序で要求に回答してデータをメモリから取り出してもよいが、例えば、不揮発性メモリが現在何をどのように動作しているか、メモリの前の動作は何であったか、関連データはメモリ内のどこに位置しているかなどによって、動作はかかる時間量が異なる。著しくは、本明細書で説明した実施形態は、性能目的でデータをホストに返す最適な順序を決定する。対照的に、次いで最適でないリターン順序が用いられる場合、システムはデータをホストに返すことができる前に、他のメモリ動作が完了するのをより長くおよび/またはより頻繁に待つ必要がある場合が多い。

30

【0040】

ブロック 404 では、動作コストが、それぞれの不揮発性メモリダイに送られる動作ごとに生成される。例えば、消去コストレジスタ 310 (図 3)、読取コストレジスタ 312 (図 3) およびプログラムコストレジスタ 314a ~ 314n (図 3) に基づいて、ならびに、ターゲットメモリダイが現在何を処理しているのか、または何を待ち行列に入れているのかに基づいて、メモリ動作モニタ 202 (図 2) の制御論理 306 (図 3) は、ターゲットダイ (例えば、図 2 の 1 つまたは複数のダイ 208a0 ~ 208n2) に対する動作コストを、対応するダイカウンタ 302a0 ~ 302n2 (図 3) 内に、構築かつ記憶する。動作コストは、必ずしもではないが、ホストからの読取要求に回答して生成されてもよいことに留意されたい。

40

【0041】

一実施形態によれば、カウンタに記憶される動作コスト値は、特定の読取コマンドについて、ダイに送られたすべての不揮発性メモリ動作に対して自動的に生成される。先述したように、制御論理 306 (図 3) は、チャンネルバス制御 316a ~ 316n (図 3) によってそれぞれのチャンネルバスから得られた情報、例えば動作の開始および終了時間を使用し、読取コマンドに回答して様々なダイカウンタ 302a0 ~ 302n2 (図 3) に

50

対して現在の値を計算することができる。あるいは、カウンタに記憶される動作コスト値は、すべての不揮発性メモリダイに対して、自動的に生成され、常時管理されるため、ホストからの読取コマンドに回答して要求される時、利用可能である。このように、制御論理 306 は、チャンネルバス制御 316a ~ 316n によってそれぞれのチャンネルバスから得た情報を使用し、経時で様々なダイカウンタ 302a0 ~ 302n2 の初期化およびフラッシングを引き起こすことができる。

【0042】

ブロック 406 では、各動作コストが、周期的なイベントを生成するタイマに基づいて経時でデクリメントされる。例えば、対応するダイカウンタ 302a0 ~ 302n2 (図 3) 内のターゲットダイ (例えば、図 2 の 1 つまたは複数のダイ 208a0 ~ 208n2) に対して生成された動作コストは、タイマ 304 (図 3) の時間刻みに基づいて、経時でデクリメントされる。1 つまたは複数のダイが消去動作、読取動作、または書込動作で現在ビジー状態であると仮定すると、次いで、対応するダイ内の動作コストは、それぞれの現在の動作の完了までの時間を反映する。次いで、時間が経過し、動作が完了すると、現在の動作の動作コストを現在の動作完了に一致する 0 にデクリメントする目的で、ダイカウンタ内の動作コストは、デクリメントするタイマ 304 の時間刻みに基づいて、相応に減少している。

【0043】

NAND フラッシュなどの特定の種類の不揮発性メモリについては、メモリから読み取ったり、またはメモリに書き込んだりすることができるデータの最大量は 1 ページである。例えば、1 ページは 8 Kb のデータであり得る。したがって、読取または書込コマンドは、複数のダイにわたって広げられることが多い。読取動作について記述するホストデータコマンドを受け取ると、コントローラ 112 (図 1) のファームウェアは、データコマンドをダイ読取コマンドに分けることができ、ダイ読取コマンドの数はコマンドの長さに依存する。例えば、8 Kb ページで、ホストからの 64 Kb の読取コマンドは、ダイからの 8 つの読取り、すなわち 64 Kb の読取コマンドに対して各 8 Kb の 8 つの読取りを必要とするであろう。

【0044】

ブロック 408 では、ホストから受け取った読取コマンドの総コストが、それぞれの不揮発性メモリダイに送られた動作ごとの現在の動作コストに基づいて計算される。例えば、ダイからデータの各セグメントを検索するのに必要な全ての読取動作の完了までの最大時間が計算される。64 Kb の読取コマンド例を続け、ホストからの読取コマンドを満たすために読み取る必要がある 8 つのダイのそれぞれに対する動作コストが、このセットのダイ読取動作の完了までの最大時間を決定するために比較され、読取コマンドの総動作コストを計算する。説明したように、各読取動作の完了までの時間は、各ダイが現在何を行っているかを考慮しており、各ダイがその動作寿命の間に何を行ったかを考慮してもよい。

【0045】

例えば一実施形態によれば、動作モニタ 202 (図 3) は記述したようなダイカウンタ 302a0 ~ 302n2 (図 3) 内に動作コストを構築かつ維持する役割を果たし、一方、コントローラ 112 (図 1) のファームウェアは、ブロック 408 でのように、総動作コストを計算する役割を果たす。しかしながら、機能性の分布は実施態様によって異なり、記述した分布に限定されない。例えば、代替の一実施形態では、総動作コストは、動作モニタ 202 によって計算され、コントローラ 112 のファームウェアに返されてもよい。

【0046】

ブロック 410 では、ブロック 408 で計算された読取コマンドの総コストに基づいて、ホストから受け取った他の読取コマンドに対して、要求されたデータをどんな順序でホストに返すべきかが決定される。説明したように、ホストがコマンドを送る順序は、必ずしも最良の SSD 性能のための、応答し、データをホストに返す最良の順序ではない。図

10

20

30

40

50

4の方法は、様々な読取コマンドをホストから受け取る順序を第一に考慮せずに、データをホストに返す「最良の」または「最適な」順序を推定する。そのため、動作コストが最小である、すなわち完了までの時間が最短であるダイ動作は、動作コストおよび完了時間がより大きいダイ動作の前に、ホストに返すことができる。その結果、より短い動作のリターンをブロックする、より長い動作を回避するようにリターンは待ち行列に入れられる。

【0047】

例えば一実施形態によれば、コントローラ112(図1)のファームウェアは、ブロック410のように、リターン順序を決定する役割を果たす。しかしながら、機能性の分布は実施態様によって異なり、記述した分布に限定されない。例えば、代替の一実施形態では、リターン順序は動作モニタ202によって計算され、コントローラ112のファームウェアに返されてもよい。

10

【0048】

動作コスト推定工程がどのように進むかを要約すると、一実施形態によれば、読取コマンドをホスト104(図1)から受け取ると、コントローラ112(図1)は、読取コマンドによってターゲットとされた(図2のダイ208a0~208n2の)各不揮発性メモリダイの動作コストについて、動作モニタ202(図2および図3)に問い合わせる。コントローラ112は、各ターゲットダイの動作についての現在の状況を考慮して、各ターゲットダイの読取りを完了する時間量について動作モニタ202に効果的に問い合わせる。次いでコントローラ112は、すべてのターゲットダイに対する動作コストに基づいて、読取コマンドの総コストを計算する。読取コマンドの総コストから、コントローラは、ホストからの他の読取コマンドに対して読取コマンドに回答するデータを返す高性能な順序を決定することができる。

20

【0049】

[推定器機構の自動調整]

フラッシュメモリの特性および作業能力は経時で、そして使用で変わり(例えばプログラム時間、消去時間などを)、また動作コスト推定が部分的にフラッシュの特性に基づくため、初期動作コスト値は、フラッシュが不揮発性メモリに使用される一実施形態において、ドライブの寿命の間、調整されてもよい。そのため、消去コストレジスタ310、読取コストレジスタ312、およびプログラムコストレジスタ314a~314n(図3)内の値は、使用で生じる完了時間のばらつきに基づいて、すなわちフラッシュの動作履歴に基づいて経時で調整されてもよい。

30

【0050】

一実施形態によれば、動作モニタ202(図2)は、フラッシュの物理的性質による経時での動作コスト変化を特定するために、様々な動作に対して各ダイが使用される回数を追跡する。この目的で、動作モニタ202は、チャンネルバス制御316a~316n(図3)から得てもよい、フラッシュからの実際のデータを使用し、フラッシュの劣化に伴い、動作コストレジスタ値を調整してもよい。そのような情報は、各フラッシュメモリ構成要素に関連した正確なダイ依存の劣化を与え、そして、動作モニタ202のメモリ308(図3)に記憶することができるであろう。そのような情報の正確さは実施態様によって異なってもよく、このように「ダイごと」または「ブロックごと」で収集し、記憶することができるであろう。したがって、メモリ308は、フラッシュ劣化データに実施される精度レベルに基づいて、相応に大きさを形成することができる。

40

【0051】

上述の明細書では、本発明の実施形態について、実施態様によって異なり得る多数の特定の詳細を参照し説明した。したがって本発明のもの、および本出願人により発明であるように意図されたものの唯一の指標は、特許請求範囲が由来する特定の形式の、いかなる以下の補正も含む本出願に由来する一組の特許請求範囲である。このような特許請求範囲に含まれる用語の、本明細書において明示的に記載したいかなる定義も、特許請求の範囲に使用したような用語の意味を規定するものとする。したがって請求項内で明示的に記載

50

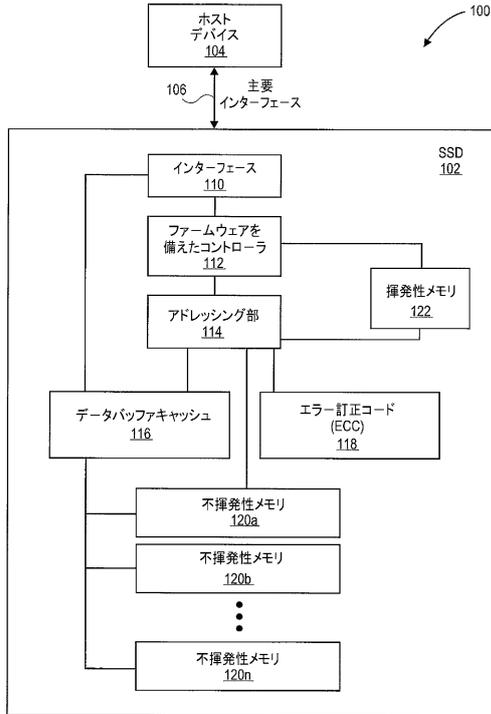
しない限界、素子、特性、特徴、利点、または属性により、このような請求項の範囲が決して制限されるべきではない。したがって本明細書と添付図面は限定的ではなく例示的であると解釈すべきである。

【符号の説明】

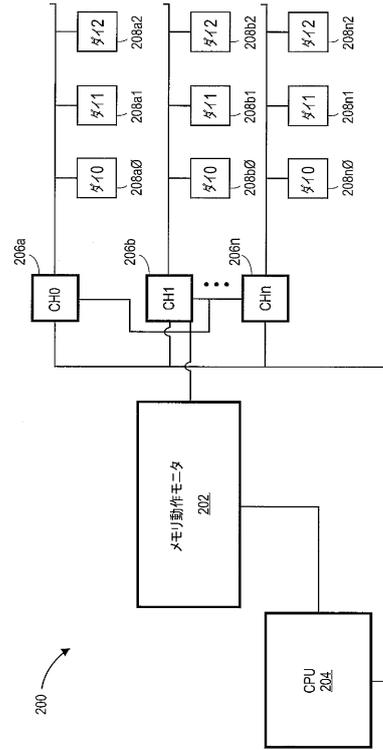
【0052】

100	ソリッドステートドライブ (SSD) 構造	
102	SSD	
104	ホストデバイス	
106	主要インターフェース	
110	インターフェース	10
112	コントローラ	
114	アドレッシング部	
116	データバッファキャッシュ	
118	エラー訂正コード (ECC)	
120 a、120 b、120 n	不揮発性メモリ構成要素	
122	揮発性メモリ	
200	動作環境	
202	メモリ動作モニタ	
204	CPU	
206 a、206 b、206 n	チャンネル	20
208 a0、208 a1、208 a2、208 b0、208 b1、208 b2、208 n0、208 n1、208 n2	メモリダイ	
302 a0、302 a1、302 a2、302 b0、302 b1、302 b2、302 n0、302 n1、302 n2	ダイカウンタ	
304	タイマ	
306	制御論理	
308	メモリ	
310	消去コストレジスタ	
312	読取コストレジスタ	
314 a、314 b、314 c、314 n	プログラムコストレジスタ	30
316 a、316 b、316 n	チャンネルバス制御機構	
402、404、406、408、410、412	ブロック	

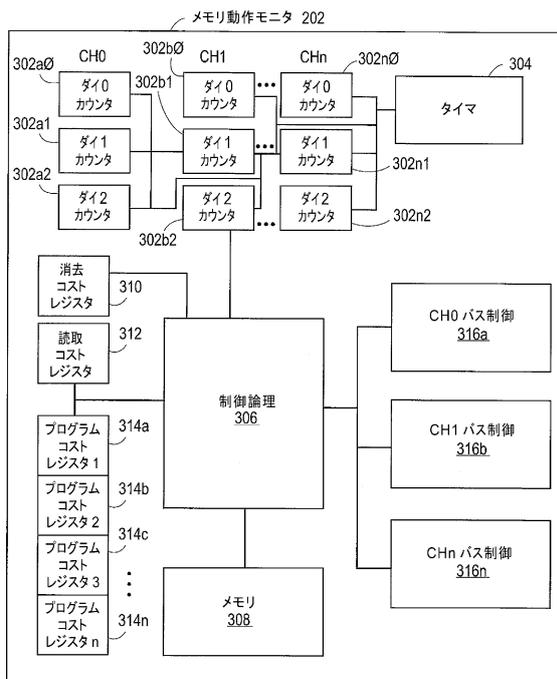
【図1】



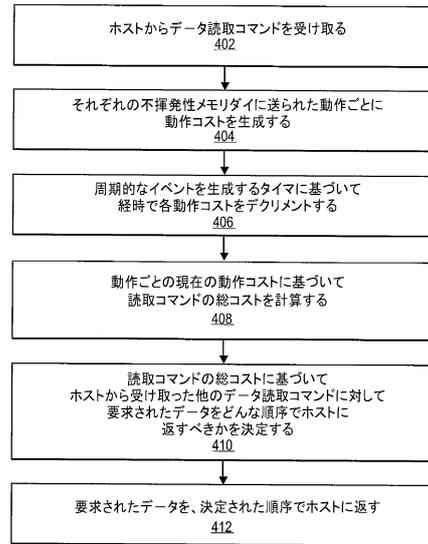
【図2】



【図3】



【図4】



フロントページの続き

審査官 小林 秀和

(56)参考文献 国際公開第2012/037293(WO, A1)
特開2012-014433(JP, A)
国際公開第2011/000082(WO, A1)

(58)調査した分野(Int.Cl., DB名)

G 0 6 F	3 / 0 6
G 0 6 F	3 / 0 8
G 0 6 F	1 2 / 0 0
G 0 6 F	1 3 / 1 0
G 0 6 F	1 3 / 1 6