

高性能と高信頼性を両立させた エンタープライズ向けSSD MK4001GRZB

MK4001GRZB Solid-State Drive for Enterprise Use Achieving High Performance and High Reliability

木内 英通

■ KIUCHI Hidemichi

ネットワーク上にあるデータセンターやサーバのサービスを利用するクラウドコンピューティングの進展に伴い、エンタープライズ向けのサーバやストレージシステムの更なる性能向上とTCO (Total Cost of Ownership)^(注1)改善が求められている。

東芝は、従来のストレージシステムで問題になっていたシステムメモリ (DRAM) とハードディスクドライブ (HDD) のアクセス性能差を吸収するために、NANDフラッシュメモリを搭載した記憶容量400 Gバイトの2.5型ソリッドステートドライブ (SSD) を開発した。高性能、及びエンタープライズ特有の要求である5年のサービスライフを考慮して、32 nmプロセスのSLC (Single Level Cell) 型NANDフラッシュメモリを採用し優れたランダムアクセス性能を達成するとともに、ウェア (疲弊) レベリング技術による長寿命化や、独自のエラー訂正機能などを付加することで堅牢 (けんろう) なデータ保護システムを実現した。

With the expansion of cloud computing in recent years, there is strong demand for enterprise servers and storage systems with higher performance and lower total cost of ownership (TCO).

In response to these market needs, Toshiba has developed the MK4001GRZB 2.5-inch solid-state drive (SSD) for enterprise use with a capacity of 400 GB and the world's top-class random access speed. The MK4001GRZB is equipped with NAND flash memory devices in order to compensate for existing performance gaps between onboard cache memories and hard disk drives (HDDs) in servers and storage systems. In particular, it incorporates state-of-the-art 32 nm-process single-level cell (SLC) NAND flash memories in consideration of the requirements for higher throughput and a five-year service life for enterprise storage devices. Furthermore, longer life and greater robustness of data integrity are provided by a wear leveling technology and our highly efficient proprietary error correction function.

1 まえがき

クラウドコンピューティングの普及が本格化するのに伴い、情報を蓄積し、制御、管理するためのエンタープライズ向けサーバやストレージシステムに対する性能要求が高まっている。

従来のストレージシステムでは、ホストシステムのDRAMに代表されるキャッシュメモリに対して、エンタープライズ向けハードディスクドライブ (eHDD) がストレージデバイスとして採用されてきたが、eHDDのアクセス性能はDRAMと比べ非常に低いため、並列処理させることで性能ギャップをカバーしてきた。

そこで東芝は、eHDDよりランダムアクセス性能で2桁以上高速なエンタープライズ向け2.5型SSD MK4001GRZBを開発することで、従来のシステムに対していっそうの性能向上とTCO改善を達成した。ここでは、MK4001GRZBの概要と、その性能を実現するために開発した技術について述べる。

(注1) コンピュータシステムの導入や、維持、管理などに掛かる費用の総計。



図1. MK4001GRZB — MK4001GRZBは、従来のeHDDとの互換性を考慮し、2.5型のSFF設計を踏襲した。

MK4001GRZB 400 GB SSD for enterprise use

2 MK4001GRZBの概要

MK4001GRZBの外観を図1に、主な仕様を表1に示す。表1では比較のため、2.5型eHDDの従来製品の仕様も並記した。

インタフェースには6 Gビット/sのSAS (Serial Attached Small Computer System Interface) を採用し、記憶容量は

表 1. MK4001GRZBの主な仕様

Main specifications of MK4001GRZB SSD (compared with conventional HDD)

項目	型名		
	MK4001GRZB	MBE2147RC	
種別	SSD	HDD	
形状	2.5型	2.5型	
記憶容量 (Gバイト)	400	147	
インタフェース	6 Gビット/s SAS	6 Gビット/s SAS	
性能	4 kバイトランダム読み出し (IOPS)	90,000	378
	4 kバイトランダム書き込み (IOPS)	17,000	312
	連続読み出し (Mバイト/s)	510	144
	連続書き込み (Mバイト/s)	230	144
信頼性	MTBF (10 ⁶ h)	2	1.6
	製品寿命	5	5
環境性	動作時温度範囲 (°C)	0 ~ 55	5 ~ 55
	動作時高度限界 (m)	-305 ~ 4,000	-305 ~ 3,048
	動作時耐衝撃 (G, 1 msのとき)	1,000	100
	非動作時耐衝撃 (G, 1 msのとき)	1,000	400
外形寸法 (mm)	15 (高さ) × 69.85 (幅) × 100.45 (奥行き)	15 (高さ) × 69.85 (幅) × 100.45 (奥行き)	
電気二重層コンデンサ*	実装	非実装	

IOPS : Input Output per Second

MTBF : 平均故障間隔

*メモリバックアップ用

400 Gバイトで、32 nmプロセスのSLC型NANDフラッシュメモリを搭載した。同じシリーズには他に記憶容量が100 Gバイトと200 Gバイトの2タイプがある。16チャンネルの並列アクセス設計によって、優れたランダムアクセス性能を達成し、従来のサーバストレージシステムで使用されてきた15,000 rpmのeHDDに比べて2桁以上の性能向上が達成されている。また、eHDDのように磁気ヘッドや、磁気ディスク、スピンドルモータなどの機構がないため、耐衝撃性や騒音性能でも非常に優位である。

従来のサーバストレージシステムとの互換性を考慮し、筐体(きょうたい)は2.5型eHDDと同一のSFF (Small Form Factor) 設計とした。これによって、ユーザーはシステム設計を変えることなく、eHDDから置き換えるだけで性能を改善できる。

3 MK4001GRZBを支える基本技術

ここでは、MK4001GRZBを支える技術について述べる。

3.1 NAND技術

MK4001GRZBは、エンタープライズ向けストレージとして5年間の製品寿命が要求されており、今回は書換え耐性やデータ信頼性の点で有利な32 nmプロセスのSLC型NANDフラッシュメモリを搭載した(図2)。大容量を実現するため、

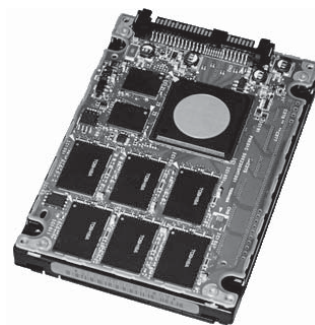


図2. NANDフラッシュメモリ実装基板 — MK4001GRZBは、32 nmプロセスのSLC型NANDフラッシュメモリを16個使用している。
Printed circuit board assembly (PCBA) equipped with NAND flash memories

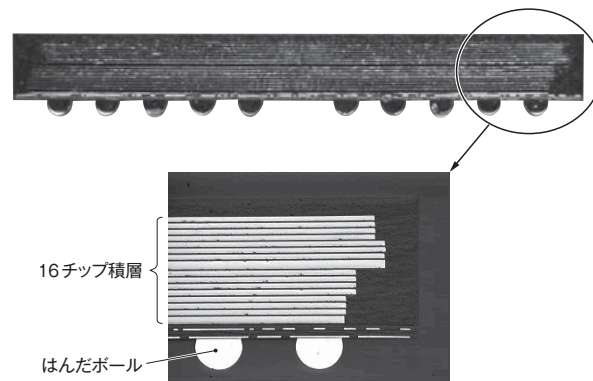


図3. NANDフラッシュメモリの断面構造 — MK4001GRZBは、400 Gバイトの大容量を達成するために、パッケージ内部で16段のスタック構造を採用している。
Cross-sectional structure of NAND flash memory

図3に示すように1パッケージに最大で16チップを積層している。このチップ多段積層技術によって、16個のパッケージで400 Gバイトの大容量が達成できる。

3.2 追記型アルゴリズム

MK4001GRZBは、高性能を達成するために追記型アルゴリズムを採用している。NANDフラッシュメモリに高速アクセスを行うために、実際にユーザーへ提供する論理容量以外に余裕容量と呼ばれる予備領域を追加した設計にしている。ホストシステムから新規にデータ書き込みの要求があった場合には、この予備領域へデータを書き込み、古いデータの書かれた領域でリード、モディファイ、又はライトを行わないことにより、高速なデータアクセスが可能になる。このため、HDDとは異なり、ユーザー論理アドレスと内部物理アドレスの関係は1対1ではなく、常にアドレス関係を変えながらデータを保存している。同時に、ガベージコレクションと呼ばれる内部データの最適な再置換アルゴリズムを採用することにより、メモリ割当てを高速に処理している。

3.3 ウェアレベリング技術

NANDフラッシュメモリはセルの構造上、プログラム／イレース (P/E：データ書込み／消去) を多数回行うと絶縁層の劣化が発生する。この絶縁層劣化が製品寿命に影響するため、これを防ぐことがSSD設計の重要な要件になる。

MK4001GRZBでは長寿命化設計として、静的ウェアレベリングと動的ウェアレベリングの2種類のウェアレベリング技術を用いて、業界最大クラスの製品寿命を達成している。

ウェアレベリングの基本的な考え方は、NANDフラッシュメモリの各セルに対するP/E回数を平準化して、最大限の製品寿命を引き出すことにある。静的ウェアレベリングでは装置内部の無効データのフリーブロック (次のデータ書込みが可能なブロック) への移行アルゴリズムを、動的ウェアレベリングでは無効データと有効データの再転換の制御アルゴリズムを行うことにより、NANDフラッシュメモリへのP/E回数を平準化している。

3.4 堅牢なデータ保護機構

ストレージ製品では、ミッションクリティカルなデータが取り扱われるため、堅牢なデータ保護機能が差異化技術になる。

このためMK4001GRZBでは、NANDフラッシュメモリのエラー特性に適したエラー訂正機能によって堅牢なデータ保護機能を実現している。また、内部にセルの疲弊をモニタする機能や余裕容量の状態モニタ機能などを搭載することによって、データの状態を自己診断するS.M.A.R.T. (Self-Monitoring, Analysis and Reporting Technology) 機能を用いて故障を事前検知し、システムにアラームを送ることでデータ保護を向上させている。

4 その他の技術

4.1 機構設計

MK4001GRZBの機構設計の概要を図4に示す。

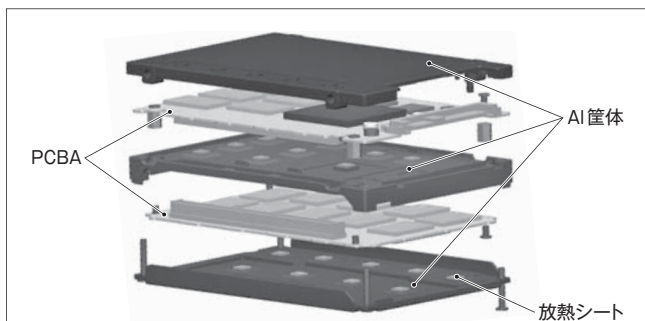


図4. MK4001GRZBの機構設計 — MK4001GRZBは、3枚のAI筐体によって2枚のPCBAを挟む構造を採用することで、NANDフラッシュメモリから発生する熱を効率よく放熱させるとともに、衝撃印加時のPCBA内部のねじれ力を分散している。

Mechanical design of MK4001GRZB

16個のNANDフラッシュメモリを実装するため印刷回路基板アセンブリ (PCBA) には二層構造を採用し、NANDフラッシュメモリへの並列アクセスによって発生する熱を効率よく放熱させるためアルミニウム (Al) 筐体で挟み込む構造を採用した。また、この構造にすることで、衝撃印加時のPCBA内部のねじれ力を分散する設計にした。

4.2 実装信頼性

MK4001GRZBは、HDDと異なりPCBAに実装している半導体素子数が多いため、半導体素子のシリコン (Si) の熱膨張係数と基板材料の熱膨張係数の差により、熱歪 (ひずみ) が発生する。この歪応力を抑制するための実装技術が重要であり、当社ははんだボール及び筐体をモデル化して熱発生時の大規模な非線形構造解析シミュレーションを実施し、最適な実装設計を行った。

装置全体をモデル化し、様々な条件で熱によって発生する応力を分析し、それぞれを最小化させるために素子の実装位置及び筐体とのねじ止め位置を決定した。また、装置動作時に受ける熱サイクルをモデル化し、そのサイクルによって受ける

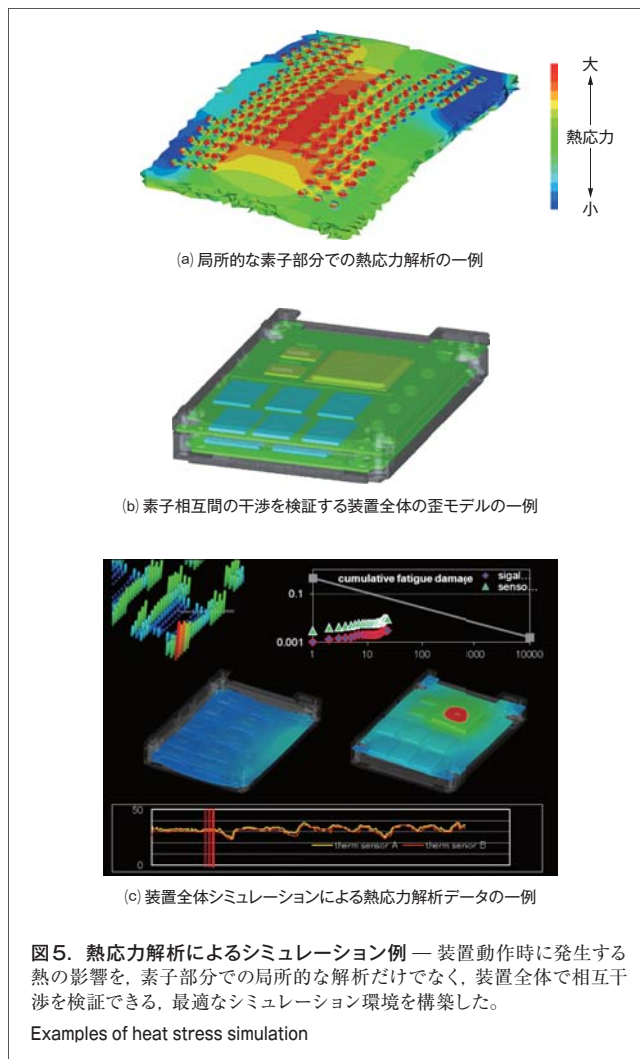


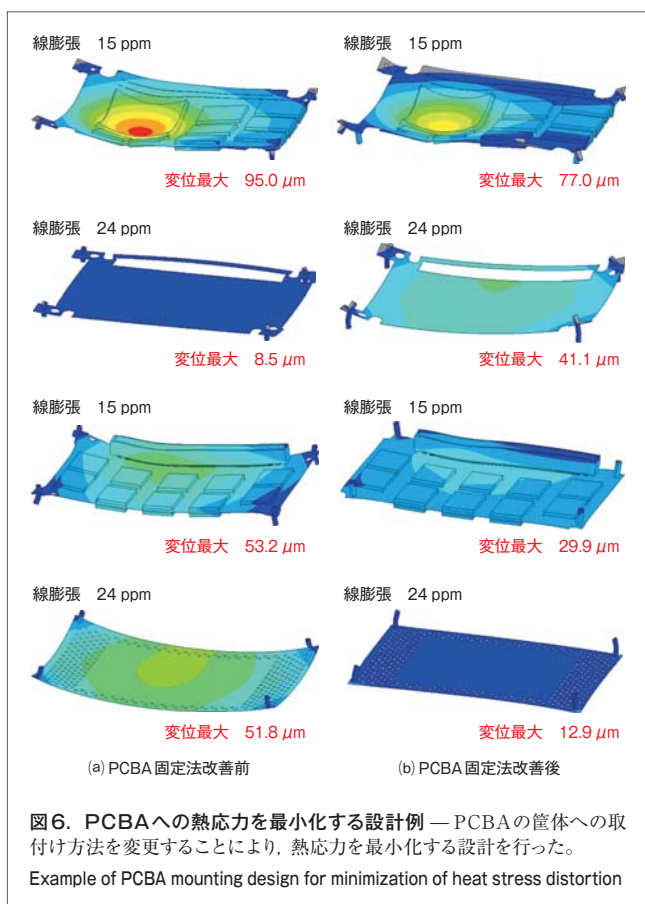
図5. 熱応力解析によるシミュレーション例 — 装置動作時に発生する熱の影響を、素子部分での局所的な解析だけでなく、装置全体で相互干渉を検証できる、最適なシミュレーション環境を構築した。

Examples of heat stress simulation

はんだ付け部への応力予測モデルを作成し、はんだ付けの信頼性を検証した。シミュレーション実施時のデータの一例を図5に示す。

一般的な解析では、図5(a)に示す局所的な素子部分での熱応力解析を行う。半導体素子などの実装密度が低いPCBAを用いる装置では、このような局所的な解析でも問題の有無を判断できるが、今回のSSDのような実装密度が高いPCBAでは、素子相互間の干渉が問題になる。そのため、図5(b)に示すような装置全体の歪モデルを作成し、装置全体としてシミュレーションを行うことにより、PCBAの構造設計を最適化した。

また、このシミュレーションモデルを用いて、PCBAの筐体への固定方法など構造設計を変更しながら、装置動作時に発生する熱によるPCBAへの応力を最小化する設計も行った。図6に実施した解析の一例を示す。



4.3 耐衝撃設計技術

MK4001GRZBの開発では、キー技術の一つである耐衝撃性能を改善するため、衝撃印加時の筐体及びPCBAへの歪応力を解析して応力集中を最小にする筐体構造設計を行った。この解析によって、衝撃印加時のAl筐体へのねじれの発生と内部のPCBAへの伝搬解析を行い、局所的な応力集中がない、2枚のPCBAの接合方法及び、筐体とのねじ止め位置を最適化した。

5 あとがき

当社が保有する様々な技術を活用することにより、優れたランダムアクセス性能を持つエンタープライズ向けSSDの製品化を実現できた。これにより、サーバやストレージシステムの大幅なTCO改善に貢献できる。

今後のエンタープライズ向けSSDの市場では、更なるコスト低減を達成するために、MLC (Multi Level Cell) 型NANDフラッシュメモリの採用が必須になる。今後も、当社の総合力を生かしたエンタープライズ向け製品の開発を継続し、社会に貢献していく。



木内 英通 KIUCHI Hidemichi

セミコンダクター&ストレージ社 ストレージプロダクツ事業部 エンタープライズSSD設計部グループ長。エンタープライズSSDの設計に従事。
Storage Products Div.