

SCM/NAND フラッシュハイブリッド SSD の ワークロード特性に応じた SCM のエラー救済手法

木下 怜佳¹ 松井 千尋¹ 山賀 祐典¹ 安達 優¹ 竹内 健¹

SCM/NAND フラッシュハイブリッド SSD において、ECC 及びペリファイ書き込みを適用した際の信頼性と SSD 性能の評価を行った。書き込みが支配的なワークロードでは、ペリファイ書き込みの性能への影響が大きいため ECC のみを適用するのが適切である。逆に、読み出しが支配的なワークロードでは ECC による影響が大きくなるため、ECC とペリファイ書き込みを組み合わせるべきである。

1. はじめに

近年、モノのインターネット (Internet of Things, IoT) やビッグデータの活用などにより、多種多様なデータを高速で処理することが求められており、データセンタでは hard-disk drive (HDD) よりも高速な solid-state drive (SSD) が用いられるようになった。また、更なる高速化、低電力化を求めて DRAM のように高速で、NAND フラッシュのように大容量な storage class memory (SCM) と NAND フラッシュを搭載したハイブリッド SSD が提案されている[1]。ただし SCM も NAND フラッシュも、他のメモリと同様に書き換え回数などによりエラーが発生するため、誤り訂正符号 (error-correcting code, ECC) やペリファイ書き込みなどの技術を用いてエラーを救済する必要がある。しかし、ECC やペリファイ書き込みは処理時間が長く、信頼性を高くするとその分 SSD 性能は悪化する。本論文では、ハイブリッド SSD に ECC やペリファイ書き込みを適用した際の信頼性と SSD 性能の評価を行った。

2. ハイブリッド SSD に ECC を適用した際の性能評価

本論文では、まず ECC による SCM/NAND フラッシュハイブリッド SSD の信頼性と性能のトレードオフについて検証した。図 1 (a) にハイブリッド SSD の構成を、図 1 (b) にデータマネージメントアルゴリズム、cold data eviction (CDE) を示す[1]。CDE では、アクセス頻度が高いデータ (ホットデータ) もしくはランダムデータ (データサイズが 8 kByte 以下) を SCM に、アクセス頻度が低いデータ (コールドデータ) かつーケンシャルデータ (データサイズが 8 kByte 以上) を NAND フラッシュに格納する。また、SCM の空き容量が 20% 以下となったとき、SCM 内のコールドかつーケンシャルなデータを NAND フラッシュに追い出す (evict) ことで、容量を確保する。データをメモリに書き込む際は ECC 符号化 (エンコード) の、データを読み出すときは ECC 復号 (デコード) の遅延が SSD 性能に影響する。Evict する際は SCM から読み出す際のデコード遅延と、NAND フラッシュに書き込む際のエンコード遅延両方の影響を受ける。

一般に SSD に用いられる ECC は、デコード遅延が比較的短い Bose-Chaudhuri-Hocquenghem (BCH) 符号である[2]。図 2 に BCH の構成、訂正可能 bit 数によるエンコード/デコード遅延と、許容可能な bit エラー率 (acceptable bit error rate, ABER) を示す。ユーザデータサイズが大きいほど同じ符号化率に対する ABER が高くなり、信頼性が向上する。NAND フラッシュは SCM に比べ信頼性が低いため、NAND フラッシュのユーザデータサイズは SCM より大きくした。本論文では、各メモリの最大符号化率は 9/10 とする。すなわち、SCM の最大訂正 bit 数は 35 bit、NAND フラッシュは 428 bit である。エンコードは訂正 bit 数によって遅延時間にはほぼ変化はないが、デコードは図

(a) に示すように各回路の処理時間を考える必要があり、訂正 bit 数が大きいほど遅延時間も大きくなる。[3]

SSD エミュレータを用いて、書き込みが支配的なワークロード (prxy_0) と読み出しが支配的なワークロード (prxy_1) を選んで性能を評価する[4]。図 3 に SCM と NAND フラッシュに BCH を適用した際の SSD 性能、input/output per second (IOPS) をそれぞれ示す。なお、SCM 容量は総 SSD 容量の 10% である。prxy_0 では、SCM と NAND フラッシュにそれぞれ 1 bit 訂正かけたときの性能を基準として、1 bit 訂正ごとに NAND フラッシュは 2.0×10^{-4} 、SCM は 5.6×10^{-3} だけ性能が低下する。また、prxy_1 では 1 bit 訂正ごとに NAND フラッシュは 4.4×10^{-4} 、SCM は 5.7×10^{-3} 性能が低下していることがわかる。prxy_0 も prxy_1 もホットランダムなワークロードであり、SCM にアクセスが集中するため SCM の 1 bit あたりの性能低下量は NAND フラッシュよりも大きくなる。

3. SCM にペリファイ書き込みと ECC を適用した際の性能評価

2章で示した通り、SCM はアクセスが集中するため ECC、主にデコードによる性能低下が大きくなる。そのため、ペリファイ書き込みで SCM のエラーを救済することも考慮に加えた。SCM の書き込みは、図 4 で示すように書き込みパルスと呼ばれる電圧を印加することで行われる。ペリファイ書き込みでは正しく書き込んでいるかを確認するため、書き込みを行った後一度読み出しを行う。このとき、書き込みが成功していなければ成功するまで、もしくは設定した回数上限まで書き込みと読み出し動作を繰り返す。図 4 にペリファイ書き込みによる BER の実測値を示す。ペリファイ書き込みの最大回数を 20 回に設定すると、BER は 49% 減少する[5]。図 1 (b) に示すように SCM にデータを書き込む際、ペリファイ書き込みを行いエラーを低減することで、SCM の読み出しの際に ECC 強度を弱くしても信頼性を維持することができる。

NAND フラッシュの ECC を符号化率 9/10 で固定し、2章と同じく prxy_0、prxy_1 のワークロードを用いて性能の評価を行った。図 5 に SCM の ECC とペリファイ書き込みによる SSD 性能低下の様子を示す。書き込みが支配的な prxy_0 ではペリファイ書き込みによる性能低下が非常に大きく、1 回あたり 1.4×10^{-2} 低下する。比べて、ECC は強度によってエンコード遅延がほぼ変わらず、また書き込みが支配的なワークロードで ECC デコード遅延の影響を受けないため、ECC による性能低下は 1 bit あたり 3.9×10^{-3} 程度である。一方、prxy_1 はペリファイ書き込みによる性能低下は 1 回あたり 7.8×10^{-3} 、ECC は 1 bit 訂正ごとに 4.2×10^{-3} 性能が低下する。prxy_1 は読み出しが支配的なので、prxy_0 に比べデコード遅延による SSD 性能低下が大きくなり、ペリファイ書き込みによる影響は小さくなる。

4. 結論

SCM/NAND フラッシュハイブリッド SSD において、ECC やベリファイ書き込みによる信頼性の向上と性能低下のトレードオフについて評価した。SCM へ ECC とベリファイ書き込みを適用した際の性能低下量を表 1 にまとめた。表より、書き込みが支配的なワークロードではベリファイ書き込みによる性能低下量が非常に大きく、ベリファイ書き込みより ECC によるエラー救済の方が適切であることがわかる。逆に、読み出しが支配的なワークロードではベリファイ書き込みによる影響が小さくなり、ECC による性能低下が大きくなるため、ECC とベリファイ書き込みを併用するべきである。

謝辞

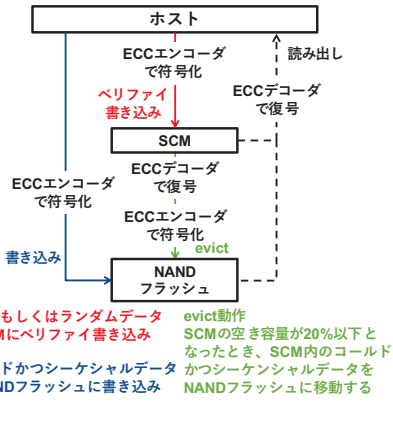
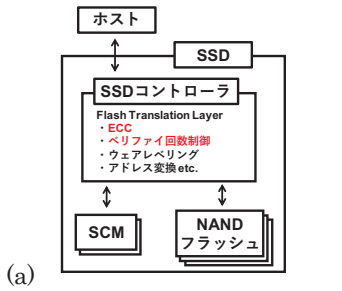
この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものです。

参考文献

- [1] C. Sun et al., "A High Performance and Energy-Efficient Cold Data Eviction Algorithm for 3D-TSV Hybrid ReRAM/MLC NAND SSD," IEEE Transactions on

Circuits and Systems I, vol. 61, no. 2, pp. 382-392, February 2014.

- [2] Y. Lee et al., "6.4Gb/s Multi-Threaded BCH Encoder and Decoder for Multi-Channel SSD Controllers," IEEE Solid-State Circuits Conference Digest of Technical Papers, pp. 426-428, February 2012.
- [3] H. Takishita et al., "ReRAM-based SSD Performance Considering Verify-program Cycles and ECC Capability," Non-Volatile Memories Workshop, March 2017.
- [4] MSR Cambridge Traces, <http://iota.snia.org/traces/388>.
- [5] S. Ning et al., "50 nm Al_xO_y resistive random access memory array program bit error reduction and high temperature operation," Japanese journal of Applied Physics, vol. 53, No. 4S, pp. 04ED09-1-04ED09-7, February 2014.
- [6] C. Matsui et al., "Error-Correction & Set/Reset Verify Strategy of Storage Class Memory (SCM) for SCM/NAND Flash Hybrid and All-SCM Storage," Solid State Devices and Materials, pp. 783-784 September 2017.



ホットもしくはランダムデータ → SCMにベリファイ書き込み
 コールドかつシークルデータ → NANDフラッシュに書き込み
 evict動作
 SCMの空き容量が20%以下となったとき、SCM内のコールドかつシークルデータを NANDフラッシュに移動する

図 1 (a) SSD 構成と (b) CDE アルゴリズム[1]

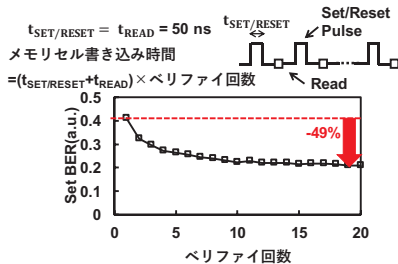


図 4 ベリファイ書き込み[5]

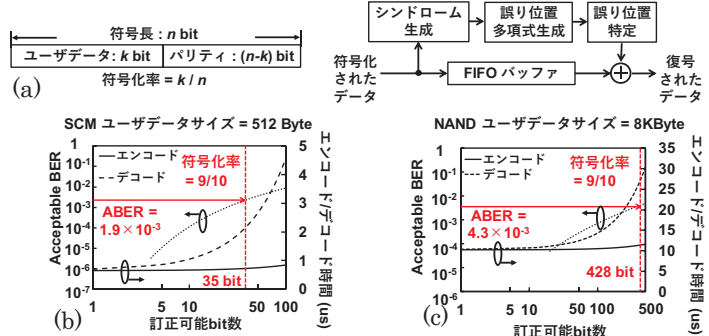


図 2 (a) BCH とデコーダ構成及び訂正可能 bit 数における (b) SCM, (c) NAND フラッシュの遅延, ABER[2, 3]

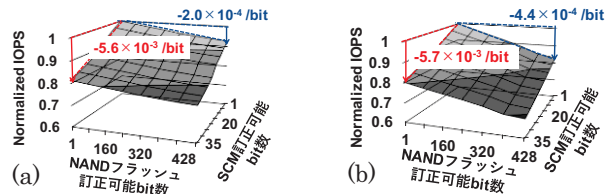


図 3 ハイブリッド SSD に ECC を適用した際の (a) prxy_0, (b) prxy_1 の性能

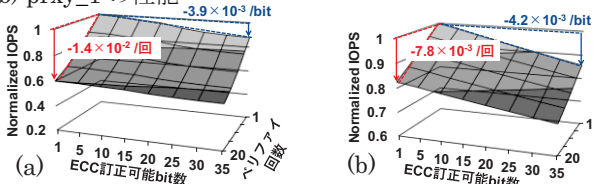


図 5 SCM に ECC とベリファイ書き込みを適用した際の (a) prxy_0, (b) prxy_1 の性能[6]

表 1 SCM のエラー救済手法

SCM のエラー救済手法	単位あたりの SSD 性能低下量 (IOPS)	
	書き込みが支配的なワークロード prxy_0	読み出しが支配的なワークロード prxy_1
ECC	$-3.9 \times 10^{-3}/\text{bit}$	$-4.2 \times 10^{-3}/\text{bit}$
ベリファイ	$-1.4 \times 10^{-2}/\text{回}$	$-7.8 \times 10^{-3}/\text{回}$
最適なエラー救済手法	ECCのみ	ECC+ベリファイ書き込み