

メモリアイブ・ストレージタイプのストレージ・クラス・メモリで構成されたマルチ SSD の書き換え耐久性を考慮した性能評価

木下 怜佳¹ 松井 千尋² 安達 優¹ 竹内 健^{1,2}

1. 研究背景

近年、モノのインターネット (Internet of Things, IoT) などの普及により発生した膨大なデータを処理するため、データセンタではより高速、高信頼なストレージが求められている。そのため、データセンタで用いられるソリッド・ステート・ドライブ(SSD)の記憶媒体として、NAND 型フラッシュメモリよりも高速なストレージ・クラス・メモリ(SCM)が注目されている。SCM はアクセス性能の高いメモリアイブ SCM (M-SCM) と容量の大きいストレージタイプ SCM (S-SCM) とに分類できる。それぞれのメモリの代表例として、M-SCM には磁気抵抗メモリ(MRAM) [1]が、S-SCM には 3D XPoint [2]などが挙げられる。また、一般に SSD では書き換え耐久寿命を延ばすためにウェアレベリングが用いられている。しかし、ウェアレベリングは大量のデータコピーを必要とするため、SSD 性能が低下してしまう[3]。本論文では、M-SCM と S-SCM を組み合わせたマルチ SCM SSD[4]の性能向上のため、各 SCM からウェアレベリングを意図的に省いた場合の影響に関して検証する。また、書き換え耐久限界を考慮した上で SSD に最適なウェアレベリングに関して議論する。

2. マルチ SCM SSD におけるデータマネジメント

図 1 に M-SCM と S-SCM を用いたマルチ SCM SSD の構成を示す。M-SCM は S-SCM の不揮発性キャッシュとして用いられ、アクセスの頻度が高いデータを M-SCM に格納する。SCM の書き込み、読み出しの単位は 512 Byte で行われ、SSD 内のデータ移動は 16 KByte のページ単位で行われる。これは 512 Byte 単位でデータ移動を行うと、データ移動を管理するテーブルが大きくなり、DRAM が大量に必要になってしまうため、NAND 型フラッシュメモリの管理サイズを参考に 16 KByte 単位に設定した。

また、特定のアドレスへの書き換え集中を防ぐため、SSD には図 2 に示すウェアレベリングが用いられている。ウェアレベリングは、あるページ内のデータに上書きリクエストがあった場合、同ページ内で最も書き換え回数が多いセクタを探査し、書き換え回数が閾値を超えていた場合は別のページにデータをコピーする。コピーされたページは無効ページとなり、ウェアレベリングの閾値を更新する。閾値の更新間隔を 5 とし、定期的にウェアレベリングが行われるよう設定した [3]。

3. マルチ SCM SSD に最適なウェアレベリング

SSD エミュレータを用いてマルチ SCM SSD の性能を評価する。マルチ SCM SSD を構成する各 SCM の特性を表 1 に示す[1, 2, 5]。また、評価には図 3 に示す 4 つのワークロードを用いる[6, 7]。ワークロードは書き込みの多寡によって書き込みが支配的な (write-intensive) ワークロードと読み出しが支配的な (read-intensive) ワークロードとに分類される。さらに、平均アクセス頻度が 40 以上のものをホット、40 以下のものをコールドとする。これらを用いて、各 SCM にウェアレベリングを表 2 のように適用した際の性能を評価し、その結果を図 4 に示す。性能は input/output per second (IOPS) で評価し、各 Case における性能は両 SCM にウェアレベリングを適用した Case 1 の性能で正規化する。どちらの SCM でもウェアレベリングを行わない Case 4 が最も性能が高くなり、Case 1 の性能よりも最大で 28.2% 向上する。しかし、どちらの SCM にもウェアレベリング適用しないというのは SSD の寿命を考えると非現実的であるため、SSD 全体の寿命を考える必要がある。ここで、マルチ SCM SSD は 2 つの SCM を用いているが、そのうちどちらか一方でも書き換え寿命を迎えてしまえば SSD は使用できなくなる。理想的な状態は、両 SCM が同時に書き換え耐久限界を迎えることである。そこで、S-SCM が書き換え耐久限界(10^9 回)を迎えた時の各 Case における M-SCM の最大書き換え回数を評価し、その結果を図 5 に示した。図中の赤線は M-SCM における書き換え耐久限界(10^9 回)である。グラフがこの赤線上であれば、2 つの SCM が同時に寿命を迎えることを示している。図 5 より、読み出しがホットな prxy_1 では Case 4 が、prxy_1 を除いた 3 つのワークロードでは S-SCM のみウェアレベリングを適用した Case 3 が最も適していることが分かった。また、図 4 より Case 3 を選んだ際の SSD 性能は、最大で Case 1 の性能より 14.0% 性能が向上する。

4. まとめ

本論文では、2 種類の SCM を用いたマルチ SCM SSD の性能向上のため、各 SCM のウェアレベリングの有無による影響を調査した。その結果を表 3 にまとめる。表 3 より、どちらの SCM にもウェアレベリングを適用しない場合は最大で 28.2% 性能が向上するが、これは SSD の寿命を無視した結果である。しかし、S-SCM にのみウェアレベリングを

¹ 中央大学理工学部

² 中央大学研究開発機構

適用すると、マルチ SCM SSD 全体の書き換え耐久性を下げることなく、最大で 14.0%性能が改善されることが分かった。

謝辞

この成果の一部は、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] J. J. Kan *et al.*, in *IEEE Transactions on Electron Devices (TED)*, vol. 64, no. 9, pp. 3639-3646, Sep. 2017.
- [2] Intel Optane Technology, <https://www.intel.com/content/>

www.us/en/solid-state-drives/optane-ssd-dc-p4800x-brief.html.

- [3] T. Onagi *et al.*, *Japanese Journal of Applied Physics (JJAP)*, vol. 54, no. 4S, pp. 04DE04-1-04DE04-7, Mar. 2015.
- [4] R. Kinoshita *et al.*, in *IEEE Non Volatile Memory Technology Workshop (NVMTS)*, Oct. 2018.
- [5] C. Matsui *et al.*, *Japanese Journal of Applied Physics (JJAP)*, vol. 54, no. 4S, pp. 04CE02-1-04CE02-9, Feb. 2017.
- [6] MSR Cambridge Traces, <http://iotta.snia.org/traces/388>.
- [7] S. Okamoto *et al.*, in *IEEE International Memory Workshop (IMW)*, May 2015, pp. 157-160.

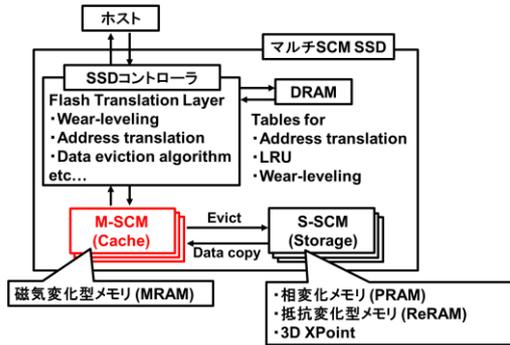


図1 マルチ SCM SSD の構成 [4]

表1 不揮発性メモリ特性 [1, 2, 5]

	M-SCM	S-SCM
ホストからのアクセス単位	セクタ (512 Byte)	
SSD内のデータ移動単位	ページ (16 KByte)	
I/O	1066 MHz	
V _{DD} (core, I/O)	1.8 V / 1.2 V	
読み出しレイテンシ	100 ns	1 us
書き込みレイテンシ	100 ns	1 us
書き換え耐久性	10 ⁹	10 ⁵
SSD 容量	ユーザーデータサイズ * 1.25 (Over provisioning : 25%)	
各容量	SSD容量の5%	SSD容量の95%

表2 ウェアレベリング適用条件

	M-SCM	S-SCM
Case 1	ウェアレベリング有り	ウェアレベリング有り
Case 2	ウェアレベリング有り	ウェアレベリング無し
Case 3	ウェアレベリング無し	ウェアレベリング有り
Case 4	ウェアレベリング無し	ウェアレベリング無し

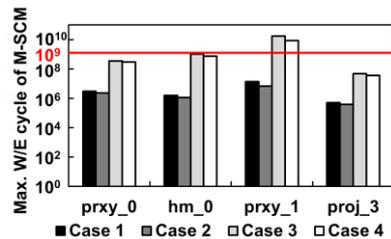


図5 S-SCM が書き換え耐久限界(10⁵回)まで書かれた場合の M-SCM の最大書き換え回数

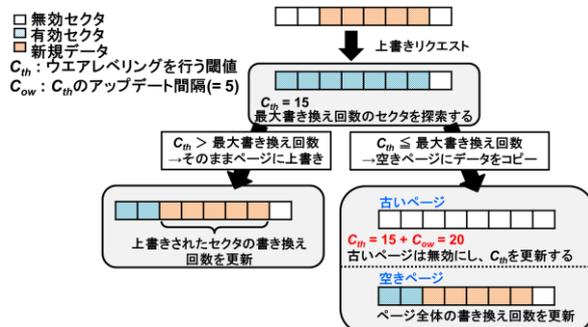


図2 ウェアレベリング動作の例 [3]

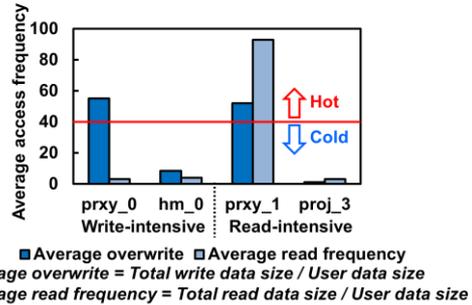


図3 ワークロード特性 [6, 7]

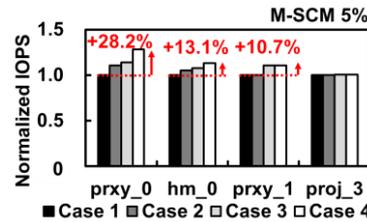


図4 ウェアレベリングによる性能への影響

表3 最適なウェアレベリング適用条件

ワークロード	IOPSが最も高いCase (IOPS改善率)	SSD書き換え耐久性を考慮したCase
prxy_0 (Write-hot)	Case 4 (+28.2%)	Case 3 (+14.0%)
hm_0 (Write-cold)	Case 4 (+13.1%)	Case 3 (+7.6%)
prxy_1 (Read-hot)	Case 4 (+10.7%)	Case 4 (+10.7%)
proj_3 (Read-cold)	Case 4 (+0.7%)	Case 3 (+0.6%)