High Performance and Energy-efficient Hybrid ReRAM/MLC NAND Flash SSD with Intelligent Data Management Algorithm

CHAO SUN†‡, HIROKI FUJII‡, KOUSUKE MIYAJI†, KOH JOHGUCHI†,
KAZUHIDE HIGUCHI‡ and KEN TAKEUCHI†

1. Introduction

A 3D TSV-integrated hybrid solid-state drive (SSD) using ReRAM and NAND flash memories is a promising solution to achieve high performance and low power consumption. This paper proposes detailed specifications for the ReRAM and architecture for the hybrid SSD. Requirements for the ReRAM in the proposed 3D TSV-integrated hybrid ReRAM/MLC NAND SSD are clarified for the first time. Suitable SSD data management algorithms are also proposed. Finally, the proposed hybrid SSD performance, energy and endurance are evaluated.

2. ReRAM I/F and Overwrite Specifications

The block diagram of the proposed SSD is shown in Fig. 1. For ReRAM, multiple write verification is necessary to achieve more than 10^6 P/E cycles [1]. In addition, the limited Program/Erase (P/E) cycles of ReRAM require logical-physical address translation and wear leveling tables in the SSD controller. Therefore, NAND I/F which allows a variable access time is proposed for ReRAM rather than DRAM I/F. Next, the overwrite policy is discussed. Fig. 2 illustrates the NAND flash and ReRAM overwrite policy. For NAND-based SSD, frequent random overwrites create many invalid pages. As a result, serious performance degradation is induced due to long garbage collection latency [3]. On the other hand, the proposed ReRAM access unit is a sector and thus partial overwrite is possible. The fragmentation problems can be solved by the hybrid SSD architecture and suitable data management algorithms.

3. Proposed SSD Data Management Algorithms

Three data management algorithms are proposed for the 3D hybrid SSD. The key idea is to store fragmented data (less than page size) to ReRAM and sequential data to MLC NAND. Fig. 3(a) introduces the concept of the used sector flag table (USFT) located in the SSD controller. By using USFT, page utilization ratio R is calculated. First, anti-fragmentation (AF) algorithm is proposed (Fig. 3(b)). When a write request is received, USFT of the target LPA is updated. If R is larger than the threshold R_{TH}, the data are written to MLC NAND. Otherwise, they are written to ReRAM. Therefore, only fragmented pages are written in ReRAM. As data accumulate in ReRAM, page data become not fragmented. When R becomes higher than R_{TH}, the data are evicted to MLC NAND. Second, most-recently-used (MRU) algorithm is proposed to catch the hot data. LPAs of the write request from the host are stored in a MRU table in a FIFO order. If the write LPA is found in the MRU table, the data go to ReRAM. Finally, reconsider-as-a-fragmentation (RAAF) algorithm is proposed to suppress overwrites to the MLC NAND after the data eviction from ReRAM to MLC NAND. Once data are evicted from ReRAM to MLC NAND, R is permanently greater than R_{TH}. The data stay in MLC NAND. In this scheme, the target LPA USFT is reset to 0 when the data are written to the MLC NAND. Small size overwrites to MLC NAND pages are recognized as fragmented data again. Thus, AF works again to store hot small data in ReRAM.

4. Evaluation and results

The proposed hybrid SSD was evaluated by the developed TLM (transaction level modeling) -based emulator. The profile of input write data obtained from a financial server [4] is shown in Fig. 4. Compared with MLC NAND-based SSD, the hybrid SSD shows 11-times higher performance and 93% lower write energy with 3D TSV interconnects (Figs. 5(a) and 5(b)). Furthermore, the slope of the average MLC NAND P/E cycles decreases by 6.9-times in Fig. 5(c), corresponding to a reduction in the replacement cost of a SSD storage system. The slope of the ReRAM P/E cycles is limited to 28-times that of the MLC NAND. Assuming MLC NAND endurance is 3×10^3, the required P/E cycles for ReRAM is less than 10^5. Fig. 6 shows the SSD valid page map indicating that data fragmentation is well suppressed by the proposed algorithms. Moreover, Fig. 7 shows that both ReRAM write and read latency should be less than 3μs to maintain high performance and low power operation.

5. Summary

Over 10-times high-speed, energy-efficient 3D TSV-integrated hybrid ReRAM/MLC NAND SSD with intelligent data management algorithms has been proposed. Furthermore, ReRAM specifications are also proposed. NAND-like ReRAM I/F, sector-access overwrite policy are necessary. Both ReRAM write and read latency should be less than 3μs and the required endurance for ReRAM is 10^5.

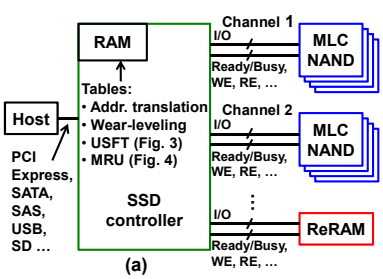
6. Acknowledgement

This work is partly supported by NEDO.

References

[1] K. Higuchi *et al.*, "50nm HfO_2 ReRAM with 50-Times Endurance Enhancement by Set/Reset Turnback Pulse & Verify Scheme," *SSDM*, pp. 1011-1012, Sep. 2011.
[2] http://www.elpida.com/ja/ news/2012/01-24r.html.
[3] K. Takeuchi *et al.*, "A 56-nm CMOS 99-mm^2 8-Gb Multi-Level NAND Flash Memory With 10-MB/s Program Throughput," *JSSC*, pp. 219-232, Jan. 2007.
[4] http://traces.cs.umass.edu/index.php/Storage/Storage.

† Faculty of Science and Engineering, Chuo University
‡ Graduate School of Engineering, University of Tokyo



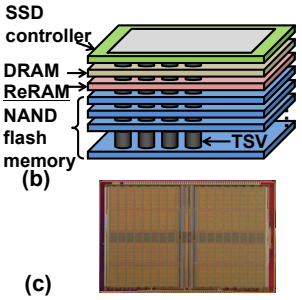


Fig. 1 (a) Block diagram of the proposed 3D TSV-integrated hybrid ReRAM/MLC NAND SSD. Proposed ReRAM uses NAND-like I/F. (b) Physical image of the proposed SSD. (c) 64Mb 50nm ReRAM test chip photograph [2].

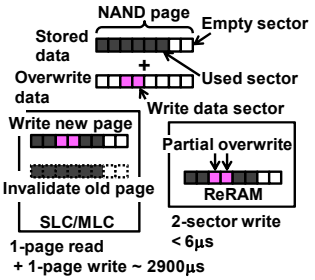
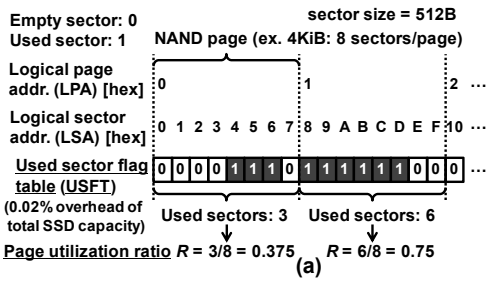


Fig. 2 Partial write or overwrite policy of the SLC/MLC NAND and ReRAM.



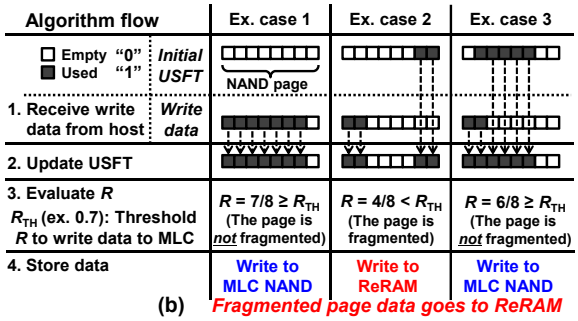


Fig. 3 (a) Used sector flag table (USFT) and page utilization ratio *R* introduction. (b) Proposed anti-fragmentation (AF) algorithm.

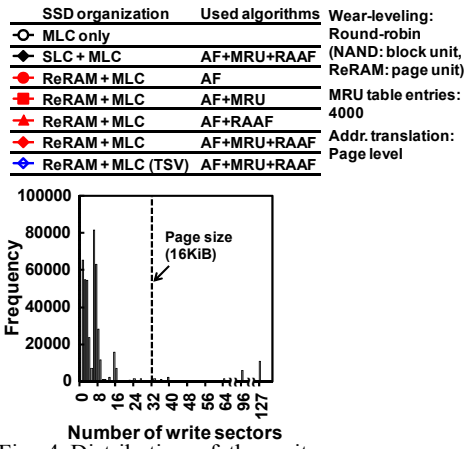


Fig. 4 Distribution of the write data size for SSD evaluation [4].

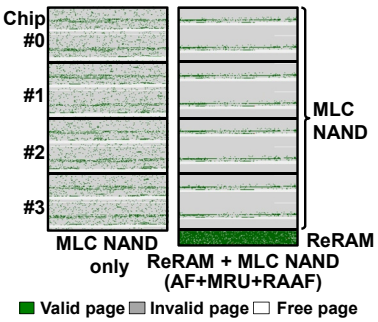
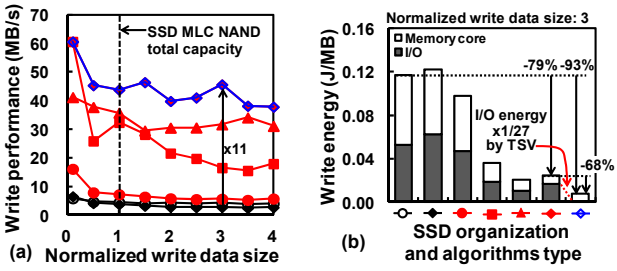


Fig. 6 SSD valid page location comparison.



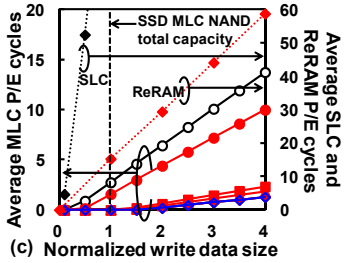


Fig. 5(a) Write performance, (b) write energy and (c) average P/E cycles of the evaluated SSDs. The horizontal axis for (a) and (c) is the data size written to the SSD normalized by the SSD MLC NAND total capacity.

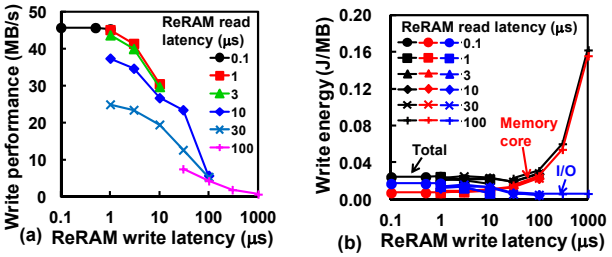


Fig. 7 SSD write performance and energy dependence on ReRAM read/write latencies.