Designing a Secure DRAM+NVM Hybrid Memory Module

Wang Xu and Israel Koren
Department of Electrical and Computer Engineering, University of Massachusetts, Amherst
Amherst, MA, 01003, USA
wangx@umass.edu, koren@ecs.umass.edu

ABSTRACT

Non-Volatile Memory (NVM) such as PCM and STT-RAM has emerged as a potential alternative for main memory due to its high density and low power leakage. However, an NVM main memory system faces three challenges compared to Dynamic Random Access Memory (DRAM) – long latency, poor write endurance and data security. To address these three challenges, we propose a secure DRAM+NVM hybrid memory module. The hybrid module integrates a DRAM cache and a security unit (SU). DRAM cache can improve the performance of an NVM memory module and reduce the number of direct writes to the NVM. Our results show that a 256MB 2-way DRAM cache with a 1024B cache line performs well in an 8GB NVM main memory module. The SU is embedded in the onboard controller and includes an AES-GCM engine and an NVM vault. The AES-GCM engine implements encryption and authentication with low overhead. The NVM vault is used to store MAC tags and counter values for each cache line. According to our results, the proposed secure hybrid memory module improves the performance by 32% compared to an NVM-only memory module. and is only 6.8% slower than a DRAM-only memory module.

KEYWORDS

NVM, Hybrid DIMM, DRAM Cache, AES-GCM

1 INTRODUCTION

With the continuous growth of the software stack, a main memory system with high capacity and density is in demand. Unfortunately, the traditional main memory design that is based on Dynamic Random Access Memory (DRAM) technology now suffers easy loss of the stored charge in the 10nm and beyond technology. On the other hand, Non-Volatile Memories (NVM) such as PCM, STT-RAM and RRAM, have presented a clear road map for scaling down to nanoscale [1] . However, NVM has a low write endurance. An NVM cell could be written about 10^8 times before it fails, whereas a DRAM cell could be written about 10^{16} times. Second, an NVM read operation is 4X slower and an NVM write operation is 10X slower than DRAM [2]. Third, NVM's data security is a major concern [3]. In this paper, we propose a secure hybrid memory module depicted in Figure 1. In the proposed memory module, we not only build a DRAM cache to improve the performance and reduce the number of writes to the NVM, but also integrate a security unit (SU) between the DRAM cache and the NVM to provide confidentiality and integrity for the NVM data. To keep our memory module persistent, a backup power source is included in the memory module [4]. We have simulated different configurations for DRAM cache and conclude that a 2-way 256MB DRAM cache with a 1024B cache line is suitable for an 8GB hybrid memory module. We also implemented a modified AES-GCM in the SU. An NVM vault is also embedded in the SU to store MAC tags and counter values for each memory block (DRAM cache line). We suggest using PCM as the NVM main memory due to its high density and relatively low cost (only few dollars per GB). We also propose using STT-RAM as the NVM vault in the SU considering its shorter latency compared to other types of NVM [5] [6].

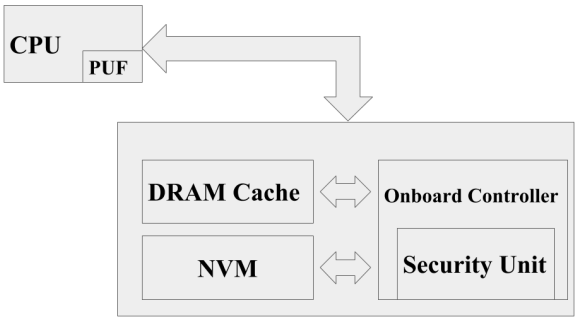

Figure 1: An Overview of Our Proposed Hybrid Memory Module

Table 1: System Configuration

CPU	
Frequency	2.67GHz
L1 Cache	I$: 4-way, 32KB, 4 cycles
	D$: 8-way, 32KB, 4 cycles
L2 Cache	8-way, 256KB, 8 cycles
L3 Cache	16-way, 8MB, 24 cycles
DRAM Cache	
Frequency	1600MHz
Channel	1
Row Buffer Size	2KB [12]
Policy	Open Page
tCL-tRCD-tRP	13.75ns-13.75nd-13.75ns [12]
NVM	
Total Size	8GB
Row Buffer Size	2KB [2]
Policy	Open Page
tCL-tRCD-tRP	13.75ns-55ns-150ns [2]
AES Latency-GHASH Latency	10 cycles - 69 cycles [13]

2 EXPERIMENTAL METHODOLOGY

To determine the configuration for DRAM cache, we developed our own DRAM cache simulator. We chose SPEC CPU2006 [7] , Rodinia [8] , NPB [9] and STREAM [10] as our benchmark suites. For performance evaluation. we integrated our own cache simulator into

Sniper Simulator [11] . The table above shows the system configuration for the performance evaluation. We chose some memory bound benchmarks from the following four benchmark suites.

- SPEC CPU2006: gobmk, sjeng, gcc, wrf, zeusmp, GemsFDTD, lbm, mcf, soplex
- Rodinia: backprop, nw, hotspot, bfs, cfd
- NPB: is, cg, ua, mg, lu
- STREAM: STREAM

In the experiments, we run SPEC in single thread and others in 4 threads.

3 DRAM CACHE

We simulated different levels of associativity, cache line sizes and total cache sizes using our DRAM cache simulator. According to Figure 2(a), we conclude that increasing the level of associativity does not benefit the performance of a DRAM cache. We choose 2-way considering the DRAM row organization and security concern. Figure 2(b) shows that when cache line size reaches 1024B, the hit rate stays the same. A 1024B cache line size indeed provides the best system performance as shown in Figure 2(c). We select 256MB as our DRAM cache total size according to the result shown in Figure 2(d).

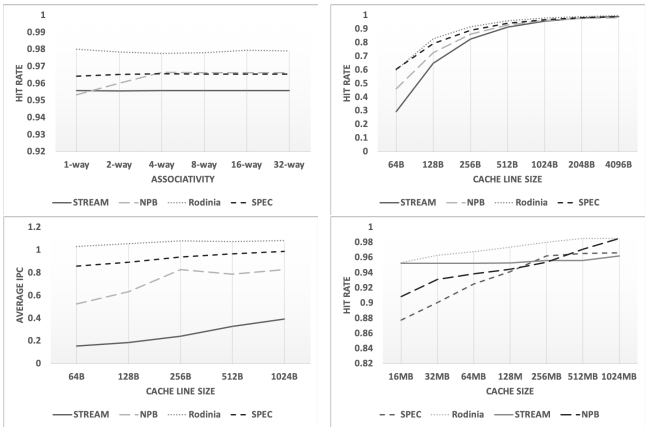

Figure 2: (a): Average Hit Rate for Different Levels of Associativity for Cache Size of 256MB and Cache Line Size of 1024B; (b): Average Hit Rate for Different Cache Line Size for Cache Size of 256MB and Associativity of 2; (c): Average IPC for Different Cache Line Size for Cache Size of 256MB and Associativity of 2; (d) Average Hit Rate for Different Total Cache Size for Cache Line Size of 1024B and Associativity of 2

4 SECURITY UNIT

Figure 3 shows the block diagram of the Security Unit (SU). The SU is located in the onboard controller between the DRAM cache and NVM main memory. It includes a security engine to implement AES-GCM [13] and an NVM vault to store all the counter values and MAC tags. The AES-GCM combines encryption and authentication into one single algorithm including AES-CTR mode and GHASH [13] . Because the SU operations could be easily parallelized with the NVM read and write operations, the SU almost has no performance overhead.

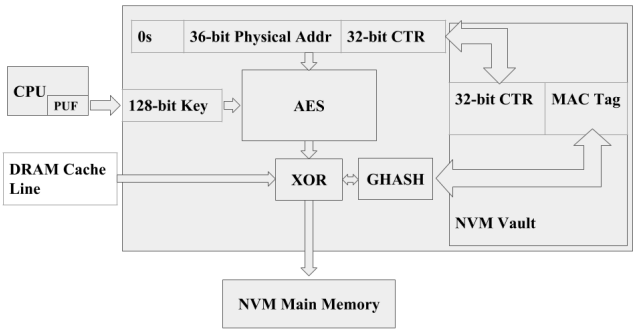

Figure 3: A Block Diagram of Security Unit

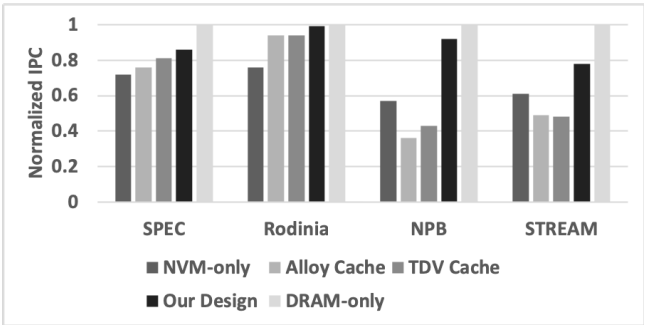

Figure 4: Normalized IPC Comparison with DRAM-only, Alloy, TDV and NVM-only for Different Benchmark Suites

5 PUTTING IT ALL TOGETHER

We compared our design with DRAM-only traditional memory system, NVM-only memory system, Alloy memory system [14] and TDV memory system [15] . Shown as Figure 5, on average, our design is faster than Alloy and TDV by 16.9% and 13.5 %, respectively. Our secure hybrid memory module also improves the performance by 32.0% compared to a NVM-only memory module and is only 6.8% slower compared to a DRAM-only memory module.

6 CONCLUSION

In this paper, we propose a secure DRAM+NVM hybrid memory module to address three challenges of NVM, namely, long latency, poor write endurance and data security. Our proposal consists of a DRAM cache along with NVM in the same memory module to improve the performance and reduce the number of direct writes to the NVM. According to our experiments, a 2-way 256MB DRAM cache with a 1024 cache line size is appropriate for an 8GB NVM memory module. Additionally, we integrate a security unit in the onboard controller between DRAM cache and NVM to provide confidentiality and authentication with low overhead. According to our results, our proposed memory module improves the performance by 32.0% compared to an NVM only memory module and is only 6.8% slower than a DRAM only memory module.

REFERENCES

[1] S. Yu and P. Chen, Emerging Memory Technologies: Resent Trends and Prospects, IEEE Solid-State Circuits Magazine, vol. 8, no. 2, pp. 43-46, 1994

[2] Benjamin C. Lee et al., Architecting Phase Change Memory as a Scalable DRAM Alternative, Proceedings of the 36th Symposium on Computer Architecture, pp. 2-13, 2009
[3] Sparsh Mittal and Ahmed Izzat Alsabi, A Survey of Techniques for Improving Security of Non-Volatile Memories, Journal of Hardware and System Security, vol. 2, issue 2, pp. 179-200, 2018
[4] Micron Technology, Inc, Micron's NVDIMMs: Persistent Memory Performance, https://www.micron.com/~/media/documents/products/product-flyer/nvdimm_flyer.pdf, 2016
[5] M. Chang, P. Rosenfeld, S. Lu and B. Jacob, Technology Comparison for Large Last-Level Caches: Low-Leakage SRAM, Low Write-Energy STT-RAM, and Refresh-Optimized eDRAM, Proceedings of IEEE 9th International Symposium on High Performance Computer Architecture, pp. 143-154, 2013
[6] NVM Technologies, http://research.cs.wisc.edu/sonar/tutorial/01-technology.pdf
[7] SPEC CPU Benchmark Suite, https://www.spec.org/cpu
[8] Shuai Che, Micheal Boyer, Jayuan Meng, David Tarjan, Jeremy W. Sheaffer, Sang-Ha Lee and Kevin Skadron, Rodinia: A Benchmark Suite for Heterogeneous Computing, Proceedings of the 2009 IEEE International Symposium on Workload Characterization, pp. 44-54, 2009
[9] NAS Parallel Benchmark Suite, https://www.nas.nasa.gov
[10] John D.McCalpin, Memory Bandwidth and Machine Balance in Current High Performance Computers, IEEE Computer Architecture Newsletter, vol. 2, pp. 19-25, 1995
[11] Trevor E. Carlson, Wim Heirman, Stijin Eyerman, Iberhim Hur and Lieven Eechhout, "An Evaluation of High Level Mechanistic Core Models", ACM Transactions on Architecture and Code Optimization, vol. 11, issue 3, pp. 1-25, 2014
[12] Micron DDR4 SDRAM, https://www.micron.com
[13] Karim M.Abdellatif, Roselyne Chotin-Avot and Habib Mehrez, AES-GCM and AEGIS: Efficient and High Speed Hardware Implementations, Journal of Signal Processing Systems, vol. 88, issue 1, pp. 1-12, 2017
[14] M. K. Qureshi and G. H. Loh, Fundamental Latency Trade-off in Architecting DRAM Caches: Outperforming Impractical SRAM-Tags with a Simple and Practical Design, Proceedings of 45th Annual IEEE/ACM International Symposium on Microarchitecture, pp. 235-246, 2012
[15] T. Lu, Y. Liu, H. Pan and M. Chen, TDV Cache: Organizing Off-Chip DRAM Cache of NVMM from a Fusion Perspective, Proceedings of IEEE International Conference on Computer Design, pp. 65-72, 2017