Innovating the SoC Design for Emerging Memory Technologies

Sungjoo Yoo

Embedded System Architecture Laboratory (ESAL)
Department of Electrical Engineering
Pohang University of Science and Technology (POSTECH)
Pohang, Korea 790-784
Tel: +82-54-279-2379
Fax: +82-54-279-8085
e-mail: sungjoo.yoo@gmail.com

Abstract –A new emerging memory technology, Phase-change RAM is gaining more and more attention as a complement or replacement of existing DRAM in the main memory subsystem. In order for PRAM to be applied to the main memory, its limitations of write endurance, long read/write latency, high write power consumption need to be overcome on the PRAM chip and the SoC utilizing the PRAM. In this paper, we explain recent works on innovating SoC designs to better utilize PRAM-based main memory.

I. Introduction

DRAM-based main memory subsystems are facing technology limits in sub-20nm technology [1]. The negative effects include the increase in refresh power (e.g., due to the increasing refresh rate mainly caused by process variation) and bit error rates (e.g., soft error rate).

Phase-change RAM (PRAM) is expected to be the most promising one of emerging memory technologies. Compared with DRAM, PRAM has the advantages of better scaling and non-volatility. Recently, large-size commercial PRAMs have been presented [2][3]. In addition, JEDEC released the LPDDR2-N interface specification [4].

In order for PRAM to compensate for or replace DRAM, however, its limitations need to be overcome. PRAM has longer read/write latency, higher write power consumption, and worse write endurance than DRAM. There have been presented several ideas on applying to PRAM the concepts of data encoding, error correction, and wear leveling. Those ideas can be implemented on the PRAM chip and/or the system-on-chip (SoC) connected to PRAM. Especially, functions needed to be implemented on the SoC requires innovations in existing SoC designs. In this paper, we explain those innovations as follows.

- New memory subsystem, especially, memory controller for hybrid DRAM/PRAM main memory
- PRAM-aware computation to resolve the PRAM limitations

II. Preliminary: Phase-Change RAM

The PRAM storage cell consists of GST material, heater and access transistor. PRAM stores information by changing the phase of GST material between amorphous (reset) and crystalline (set) states. In order to perform a reset operation, a high current flows through the heater, and generates heat (at about 600°C) to melt the GST material. Then, it is cooled down in a short period. For a set operation, a lower temperature (about 300°C) is generated and the GST material is cooled down in a longer period. It is the heating process that degrades write endurance.

PRAM chips, e.g., [3] typically have basic functions for write reduction, e.g., differential write [5] and invert coding [6]. There are also advanced function candidates which can be implemented on the PRAM chip. First, error correction capability can be enhanced by error correction pointer (ECP) [7] which stores, in the spare area, the information of error location and correct value, or SAFER [8] which groups data bits for multi-bit error correction based on SEC-DED capability. The PRAM lifetime can also be enhanced via on-chip wear leveling, e.g., start gap wear leveling [9], security refresh [10], etc.

III. Novel SoC Architectures for Phase-Change RAM

A. Memory Controller for DRAM/PRAM

The PRAM limitations can be practically mitigated by utilizing DRAM together with PRAM in a hybrid DRAM/PRAM main memory [11]. Both memories can be organized in a flat or hierarchical manner. In the flat structure, each memory has its own address subspace while, in the hierarchical structure, DRAM plays the role of last level cache and PRAM works as the large background memory. However, the hybrid DRAM/PRAM does not completely overcome the limitations. Thus, in addition to the basic functions explained in Section II, more advanced functions need to be realized in the main memory subsystem, especially, in the memory controller. We present four categories of advanced functions as follows.

The memory controller can give further reduction in bit updates by sophisticated data encoding. In [12], Tran et al. present a memory-mapped invert coding which enables heterogeneous granularities in invert coding thereby achieving significant bit updates with low PRAM area cost. The memory controller can keep track of write frequency for more efficient wear leveling called hot/cold swapping. Hot data are migrated to the location that cold data previously occupied thereby avoiding localized wear-outs which determine the PRAM lifetime [13][14].

Write performance can be improved by the memory controller adopting write cancellation, which preempts current write operation to serve a new read request [15]. A peak-power aware write scheduling can also be applied where the number of bit updates in the current write operation is estimated and then concurrent write operations are maximized based on the bit update information.

The hybrid DRAM/PRAM reduces the required capacity of DRAM [11]. However, a relatively large-size DRAM is still required due to write endurance and performance issues. The large-size DRAM incurs high power consumption due to refresh. In [16], a decay concept is applied to reduce DRAM refresh energy by evicting dead data from DRAM while considering total energy consumption of DRAM and PRAM.

B. PRAM-conscious Computation

PRAM-conscious computation has a large potential in improving the performance, lifetime and power efficiency of PRAM-based main memory. A key direction of PRAM-conscious computation will be to minimize bit updates while reducing the overhead of power, performance and area.

Recently, an example of PRAM-conscious video processing function was presented in [17]. We present two methods of inter-block differential data encoding and inter-frame multiple experts for write reduction in PRAM. The inter-block differential data encoding exploits the fact that video data values have high spatial locality. For a group of four pixels, given a baseline pixel value, a differential encoding is applied among the four pixels and between the given baseline pixel value and the max or min pixel value of the four pixels. In the inter-frame multiple experts, multiple coding methods are evaluated on a video frame basis and the most successful coding method in recent frames is selected and applied for the current frame data.

IV. Prospects

PRAM has the capability of multi-level cell (MLC) memory. It is expected that SLC PRAM can be utilized for main memory while MLC PRAM for storage or large-scale main memory. To apply MLC PRAM, there are two physical limitations: R drift and thermal issues.

R drift is the phenomenon that the resistance value continues to increase for some period after the program operation (set and reset in SLC PRAM, and writing 00, 01, 10, and 10 in MLC PRAM) [18]. A read operation to the PRAM cell under R drift can leave a un-determined state after the read operation. Thus, R drift latency needs to be respected, especially, for MLC PRAM where voltage margin is much narrower than in SLC PRAM.

The GST material in PRAM cell can re-crystallized to give low resistance value even long after the program operation. It is typically caused by high ambient temperature. Both R drift and re-crystallization need to be addressed for MLC PRAM in a cost-effective manner.

V. Summary

In this paper, we introduced a new emerging memory technology, PRAM, to compensate for or replace DRAM in the main memory. In order to overcome the PRAM limitations, both memory and SoC architectures need to be innovated. On the SoC side, the memory controller needs to be equipped with advanced methods for write reduction, wear leveling, performance improvement and power reduction in PRAM. In addition, PRAM-aware computation, which has a large potential in resolving the PRAM issues, needs to be investigated.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0007909 and 2010-0015336) and IC Design Education Center (IDEC).

References

[1] International Technology Roadmap for Semiconductors (ITRS), available at www.itrs.net.

[2] C. Villa, et al., "A 45nm 1Gb 1.8V Phase-Change Memory," Proc. ISSCC, 2010.

[3] H. Chung, et al., "A 58nm 1.8V 1Gb PRAM with 6.4MB/s Program BW," Proc. ISSCC, 2011.

[4] JEDEC Standard, Low Power Double Data Rate 2 (LPDDR2), JESD209-2E, April 2011.

[5] B. D. Yang, et al., "A Low Power Phase-Change Random Access Memory Using a Data-Comparison Write Scheme," Proc. ISCAS, 2007.

[6] S. Cho and H. Lee, "Flit-N-Write: A Simple Deterministic Technique to Improve PRAM Write Performance, Energy and Endurance," Proc. MICRO, 2009.

[7] S. Schechter, et al., "Use ECP, not ECC, for Hard Failures in Resistive Memories," Proc. ISCA, 2010.

[8] N. Seong, et al., "SAFER: Stuck-At-Fault Error Recovery for Memories," Proc. MICRO, 2010.

[9] M. K. Qureshi, et al., "Enhancing Lifetime and Security of PCM-Based Main Memory with Start-Gap Wear Leveling," Proc. MICRO, 2009.

[10] N. H. Seong, et al., "Security Refresh: Prevent Malicious Wear-out and Increase Durability for Phase-Change Memory with Dynamically Randomized Address Mapping," Proc. ISCA, 2010.

[11] M. K. Qureshi, V. Srinivasan, and J. A. Rivers, "Scalable High Performance Main Memory System Using Phase-Change Memory Technology," Proc. ISCA, 2009.

[12] A. Tran, S. Yoo, S. Lee, and C. Park, "Memory-Mapped Invert Coding for PRAM Main Memory," EMT (emerging memory technology) workshop co-located with ISCA, June 2010.

[13] J. Dong, et al., "Wear Rate Leveling: Lifetime Enhancement of PRAM with Endurance Variation," Proc. DAC, 2011.

[14] J. Yun, S. Yoo, and S. Lee, "Bloom Filter-based Dynamic Wear Leveling for Phase Change RAM," Proc. DATE, 2012.

[15] M. K. Qureshi, et al., "Improving Read Performance of Phase Change Memories via Write Cancellation and Write Pausing," Proc. HPCA, 2010.

[16] H. Park, S. Yoo, and S. Lee, "Power Management of Hybrid DRAM/PRAM-based Main Memory," Proc. DAC, 2011.

[17] S. Kwon, S. Yoo, and S. Lee, "Optimizing Video Application Design for Phase-Change RAM-based Main Memory," to appear in IEEE Trans. on VLSI.

[18] D. Ielmini, A. L. Lacaita, and D. Mantegazza, "Recovery and Drift Dynamics of Resistance and Threshold Voltages in Phase-Change Memories," IEEE Trans. on Electron Devices, vol. 54, no. 2, Feb. 2007.