

Blackcomb: Hardware-Software Co-design for Non-Volatile Memory in Exascale Systems

<http://ft.ornl.gov/trac/blackcomb>

Jeffrey Vetter, ORNL
Robert Schreiber, HP Labs

Trevor Mudge, University of Michigan
Yuan Xie, Penn State University

Background

Memory, not processing, is the crux of the exascale co-design problem. Exascale machines will push the limits of memory capacity, power, and performance. DRAM, the universal memory technology of today, may not scale to meet the needs of exascale. Disk storage, critical for checkpointing and for archiving computational inputs and results, may also fail to provide adequate performance, reliability, and power efficiency. We confront a memory/storage crisis.

The Blackcomb effort seeks to create and understand new memory technologies, develop their roles in exascale systems, adapt applications to them, and assess their relative merits. We focus on emerging nonvolatile memory (NVM) technologies, including spin-torque-transfer RAM (STT-RAM), phase-change RAM (PC-RAM), and memristor (resistive RAM, or R-RAM).

	SRAM	DRAM	NAND Flash	PC-RAM	STT-RAM	R-RAM (Xpoint)
Data Retention	N	N	Y	Y	Y	Y
Memory Cell Factor (F ²)	50-200	6	2-5	4-10	6-20	≤4
Read Time (ns)	< 1	30	10 ⁴	10-100	10-40	5-30
Write/Erase Time (ns)	< 1	50	10 ⁵	100-300	10-40	5-100
Number of Rewrites	10 ¹⁶	10 ¹⁶	10 ⁴ -10 ⁵	10 ⁸ -10 ¹⁵	10 ¹⁵	10 ⁶ -10 ¹²
Power Read/Write	Low	Low	High	Low	Low	Low
Power (other than R/W)	Leakage Current	Refresh Power	None	None	None	None

Figure 1 -- Characteristics of Memory Technologies

Objectives

- Understand and improve these new NVM technologies
- Propose new architectures that address the resilience, energy, and performance needs of Exascale applications
- Adopt the most promising NVM technologies
- Flatten the memory hierarchy
- Place low-power compute cores close to the data
- Replace mechanical disk-based storage with energy-efficient NVM
- Define programmer's APIs that expose nonvolatility and enable resilient applications that are easily written, energy efficient, and fast.

Approach

The project is structured around five tasks:

1. NVM Technology: identifies and characterizes the most promising NVM technologies. We will assess and improve wearout, error rate, durability, energy, latency.
2. Memory Architecture: explores the architecture space, considering how to assemble NVMs with a space of future processors, and also looks into the uses of NVM for resilience.
3. System Architecture: proposes a novel HPC system architecture. The idea is to explore the use of NVM as a single-level data store, co-located with ultra-low voltage processors and balanced network capability. This entails a design-space exploration of the various architecture options, as well as an analysis of the simplification and optimization of the software stack.
4. System and Runtime Software: identifies the most useful programming abstractions of the new NVM architectures. We will look into new programming paradigms that can help to fully take advantage memory nonvolatility. We will re-examine the Message Passing Interface (MPI) and Partitioned Global Address Space (PGAS) programming models, and respective I/O models, such as MPI-IO and the Hierarchical Data Format (HDF5) in light of the new memory and storage architectures.
5. Applications: identifies, characterizes, and transforms key DoE applications for NVM. The results of the characterization will be made available to the other work packages to provide a quantitative basis for research decisions. New programming and other software techniques will be ported to the selected applications and tested. We will seek to understand the sensitivity of studied applications to faults.

Impact

- Better energy scalability: NV memories have zero standby power



U.S. DEPARTMENT OF
ENERGY

Office of
Science

- Increase system reliability: MRAM/PCRAM are resilient to soft errors
- Boost performance: NVM will be much faster than magnetic disk
- Improve programmability and application fault tolerance with enhanced programming models

Challenges

- Understand and mitigate limitations of NVMs as a general purpose memory: higher write overheads and lower endurance than SRAM or DRAM
- Need novel analytical/simulation hybrid model to understand tradeoffs between energy efficiency, resilience, and performance
- Evaluate productivity of proposed programming models that exploit NVM to improve fault-tolerance of distributed applications

Research Products & Artifacts

- Identify and characterize a few key DOE applications, making results available to the other areas of the study to provide a quantitative basis for research decisions
- Broadly explore opportunities for NVM memory in high performance computing, such as using active NVM for IO staging
- Explore novel architectures that include NVM in the memory and storage hierarchies; investigate the bandwidth requirements for 3D-stacked and side-stacked NVM; develop peripheral circuitry design in NVM devices that can achieve higher memory density
- Investigate NVM-based programming to improve resilience and productivity, including efficient checkpointing
- Apply these new technologies to exascale proxy applications and test them on realistic datasets

Recent Publications

1. D. Li, J.S. Vetter, G. Marin, C. McCurdy, C. Ciria, Z. Liu, and W. Yu, "Identifying Opportunities for Byte-Addressable Non-Volatile Memory in Extreme-Scale Scientific Applications," in *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. Shanghai: IEEE, 2012.
2. D. Li, J.S. Vetter, and W. Yu, "Classifying Soft Error Vulnerabilities in Extreme-Scale Scientific Applications Using a Binary Instrumentation Tool," in *SC12: ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis*. Salt Lake City, 2012 (to appear).
3. D-H. Yoon, N. Muralimanohar, J. Chang, P. Ranganathan, N. P. Jouppi and M. Erez. FREE-p: Protecting non-volatile memory against both hard and soft errors. *Intl. Symp. on High Perf. Comp. Architecture (HPCA)*, 2011.
4. C. Xu and X. Dong and N. Jouppi and Y. Xie. Design implications of memristor-based RRAM cross-point structures. *Design, Automation and Test in Europe (DATE)*, 2011.
5. D. Yoon, T. Gonzalez, P. Ranganathan, and R. Schreiber. Co-design of efficient advanced memory hierarchies. *ACM International Conference on Computing Frontiers*, May 2012.
6. D. Yoon, R. Schreiber, J. Chang, N. Muralimanohar, P. Ranganathan, and P. Faraboschi. VerMem: Versioned memory using multilevel-cell NVRAM. *Non-Volatile Memories Workshop (NVMW)*, poster, 2012.
7. C. Yang, Y. Emre, C. Chakrabarti, and T. Mudge. Flexible product code-based ECC schemes for MLC NAND FLASH memories. *IEEE Workshop on Signal Processing Systems, (SiPS 2011)*, October 2011.
8. D. Yoon, J. Chang, N. Muralimanohar, and P. Ranganathan. BOOM: Enabling mobile memory based low-power server DIMMs, *39th International Symposium on Computer Architecture (ISCA-39)*, June 2012
9. P. Tandon, J. Chang, R. Dreslinski, P. Ranganathan, T. Mudge, T.F. Wenisch. PicoServer revisited: On the profitability of eliminating intermediate cache levels. *2012 Workshop Duplicating, Deconstructing, and Debunking (WDDD)*, June 2012, pp. 1-10.
10. J. Chang, K. Lim, T. Mudge, P. Ranganathan, D. Roberts, M. Shah. A limits study of benefits from nanostore-based future data-centric system architectures. *ACM International Conference on Computing Frontiers (CF' 12)*, May 2012, pp. 33-42.
11. X. Dong, C. Xu, Y. Xie, N. Jouppi. NVSim: A circuit-level performance, energy, and area model for emerging nonvolatile memory. *IEEE Trans. on CAD of Integrated Circuits and Systems* 31(7): 994-1007 (2012)
12. D. Niu, C. Xu, N. Muralimanohar, N. Jouppi, Y. Xie. Design trade-offs for high density cross-point resistive memory. *ISLPED 2012*, pp. 209-214
13. A. Jog, A. K. Mishra, C. Xu, Y. Xie, V. Narayanan, R. Iyer, C. Das. Cache revive: Architecting volatile STT-RAM caches for enhanced performance in CMPs. *DAC 2012*, pp. 243-252