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

Perspectives on Blackcomb Simulation Tools

Jeffrey Vetter, ORNL

Robert Schreiber, HP Labs

Trevor Mudge, University of Michigan

Yuan Xie, Penn State University

Presented to

DOE X-Stack PI Meeting

Boston

28 May 2014

OAK RIDGE NATIONAL LABORATORY

MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

Georgia Tech College of Computing

Computational Science and Engineering

http://ft.ornl.gov ♦ vetter@computer.org

Blackcomb: Hardware-Software Co-design for Non-Volatile Memory in Exascale Systems (started 2010)

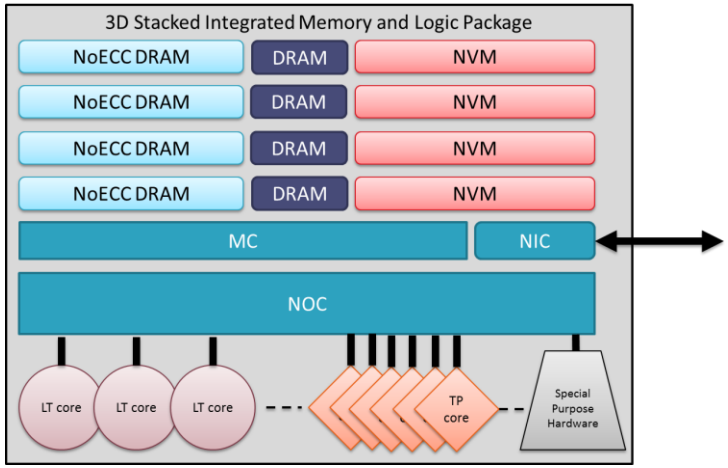
Jeffrey Vetter, ORNL
Robert Schreiber, HP Labs
Trevor Mudge, University of Michigan
Yuan Xie, Penn State University

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

FWP #**ERKJU59**

Objectives

- Rearchitect servers and clusters, using nonvolatile memory (NVM) to overcome resilience, energy, and performance walls in exascale computing:
  - Ultrafast checkpointing to nearby NVM
  - Reoptimize the memory hierarchy for exascale, using new memory technologies
  - Replace disk with fast, low-power NVM
  - Enhance resilience and energy efficiency
  - Provide added memory capacity



Established and Emerging Memory Technologies – A Comparison

	SRAM	DRAM	eDRAM	NAND Flash	PCRAM	STTRAM	ReRAM (1T1R)	ReRAM (Xpoint)
Data Retention	N	N	N	Y	Y	Y	Y	Y
Cell Size (F^2)	50-200	4-6	19-26	2-5	4-10	8-40	6-20	1- 4
Read Time (ns)	< 1	30	5	10^4	10-50	10	5-10	50
Write Time (ns)	< 1	50	5	10^5	100-300	5-20	5-10	10-100
Number of Rewrites	10^{16}	10^{16}	10^{16}	10^4-10^5	10^8-10^{12}	10^{15}	10^8-10^{12}	10^6-10^{10}
Read Power	Low	Low	Low	High	Low	Low	Low	Medium
Write Power	Low	Low	Low	High	High	Medium	Medium	Medium
Power (other than R/W)	Leakage	Refresh	Refresh	None	None	None	None	Sneak

Blackcomb Use Cases

Device Level Cell and Array Modeling

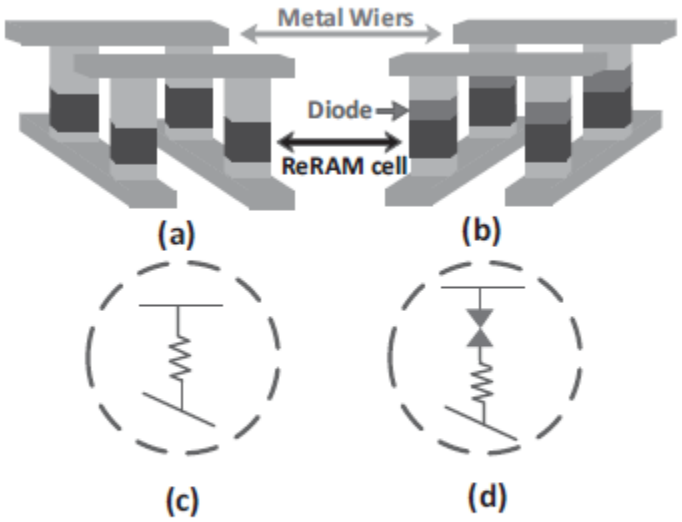


Fig. 1. Schematic view of 0T1R and 1D1R ReRAM structure: array structure of 0T1R(a) and 1D1R (b); circuit diagram of 0T1R (c) and 1D1R (d).

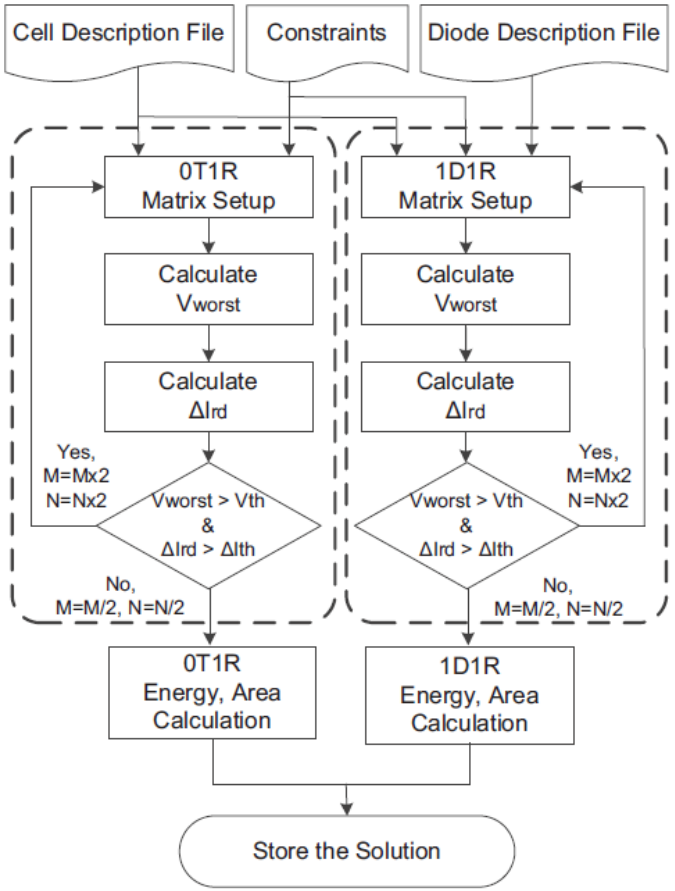
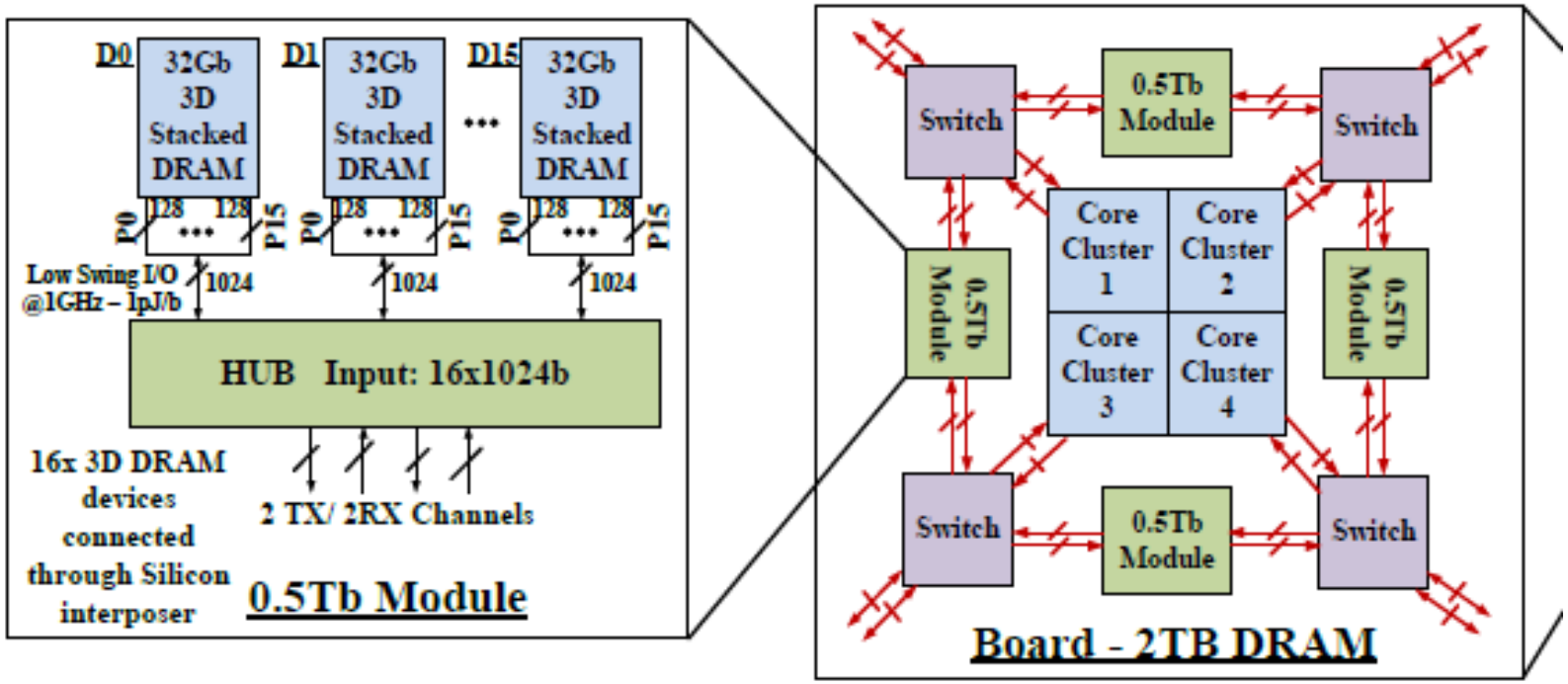


Fig. 4. First stage design flow. (V_{worst}, V_{th}: calculated value and constraint of worst case voltage. $\Delta I_{rd}, \Delta I_{th}$: calculated value and constraint of read noise margin. M, N: number of wordline and bitline.)

D. Niu, C. Xu *et al.*, *"Design of cross-point metal-oxide ReRAM emphasizing reliability and cost," in Proceedings of the International Conference on Computer-Aided Design. San Jose, California: IEEE Press, 2013, pp. 17-23.*

Tradeoffs in Exascale Memory Architectures

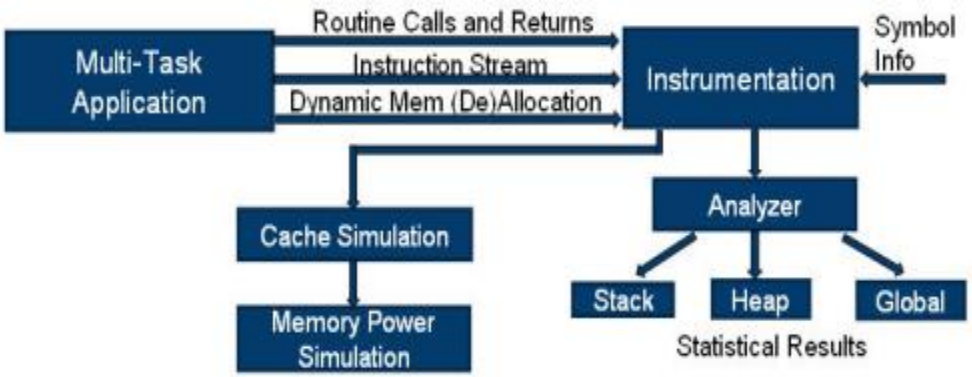


- Understanding the tradeoffs
  - ECC type, row buffers, DRAM physical page size, bitline length, etc

Mudge, et al. "Optimizing DRAM Architectures for Energy-Efficient, Resilient Exascale Memories," *SC13, 2013*

Identifying Opportunities for Byte-Addressable Non-Volatile Memory in Extreme-Scale Scientific Applications

- **Problem**
  - Do specific memory workload characteristics of scientific apps map well onto NVRAMs' features?
  - Can NVRAM be used as a solution for future Exascale systems?

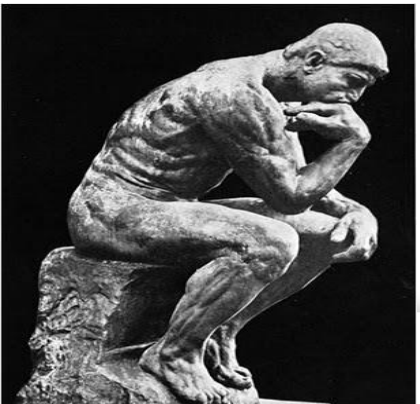


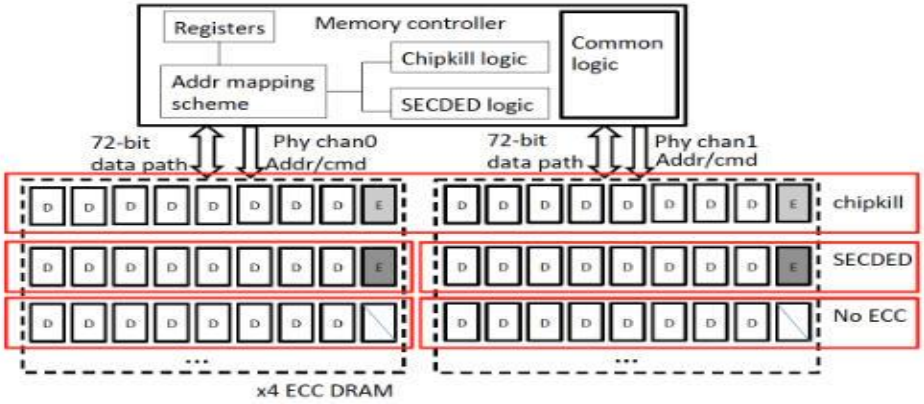
- **Solution**
  - Develop a binary instrumentation tool to investigate memory access patterns related to NVRAM
  - Study realistic DOE applications (Nek5000, S3D, CAM and GTC) at fine granularity
- **Impact**
  - Identify large amount of commonly existing data structures that can be placed in NVRAM to save energy
  - Identify many NVRAM-friendly memory access patterns in DOE applications
  - Received attention from both vendor and apps teams

D. Li, J.S. Vetter, G. Marin, C. McCurdy, C. Cira, Z. Liu, and W. Yu, "Identifying Opportunities for Byte-Addressable Non-Volatile Memory in Extreme-Scale Scientific Applications," in *IEEE International Parallel & Distributed Processing Symposium (IPDPS). Shanghai: IEEEE, 2012*

Rethinking Algorithm-Based Fault Tolerance

- Algorithm-based fault tolerance (ABFT) has many attractive characteristics
  - Can reduce or even eliminate the expensive periodic checkpoint/rollback
  - Can bring negligible performance loss when deployed in large scale
  - No modifications from architecture and system software
- However
  - ABFT is completely opaque to any underlying hardware resilience mechanisms
  - These hardware resilience mechanisms are also unaware of ABFT
  - Some data structures are over-protected by ABFT and hardware

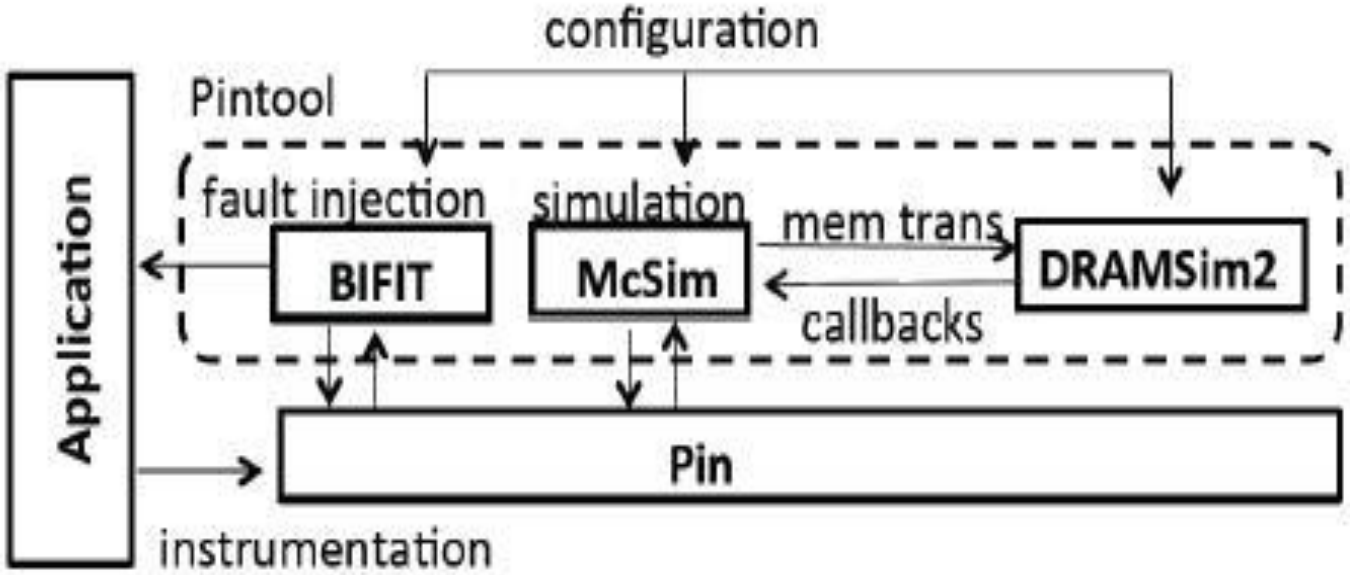




D. Li, C. Zizhong, W. Panruo, and S. Vetter Jeffrey, "Rethinking Algorithm-Based Fault Tolerance with a Cooperative Software-Hardware Approach," Proc. International Conference for High Performance Computing, Networking, Storage and Analysis (SC13), 2013,

Evaluation

- We use four ABFT (FT-DGEMM, FT-Cholesky, FT-CG and FT-HPL)



- We save up to 25% for system energy (and up to 40% for dynamic memory energy) with up to 18% performance improvement

Simulation Tools

Prediction Techniques Ranked

	Speed	Ease	Flexibility	Accuracy	Scalability
Ad-hoc Analytical Models	1	3	2	4	1
Structured Analytical Models	1	2	1	4	1
Aspen	1	1	1	4	1
Simulation – Functional	3	2	2	3	3
Simulation – Cycle Accurate	4	2	2	2	4
Hardware Emulation (FPGA)	3	3	3	2	3
Similar hardware measurement	2	1	4	2	2
Node Prototype	2	1	4	1	4
Prototype at Scale	2	1	4	1	2
Final System	-	-	-	-	-

Architectural Simulators

- Sniper, McSim, Zsim, Marss, Gem5 etc.
  - These simulators model processor components, e.g. cache, memory, core
  - They provide stats on cache/memory access, their overlap, hit rate, etc.
  - We have used McSim and Sniper
- NVRAM cache parameters are derived from device level simulators like NVSim and imported into Sniper, McSim, etc

A quick, partial overview of architectural simulators - Mar 2014

	CPU-GPU Hybrid	Detailed Processor Modeling	Detailed Main Memory Modeling	Uses Pin?	Coherence	x86 support	Full-System	Parallelization
McSimA+	No	?	Yes	Yes	?	x86-64	No	?
Multi2Sim	Yes	Yes (OoO)	Yes	No	?	x86	No	No
MacSim	Yes	?	Yes (DRAMsim)	Yes	?	x86	No	?
GEMS	No	Yes (OoO)	No	Simics	Yes	Yes (x86, sparc …)	Yes	No
Gem5	No	Yes (OoO, inorder …)	Yes (DRAMsim)	No	Yes (GEMS)	x86	Yes	No
Marss	No	Yes	Yes (DRAMsim)	Qemu	Yes	x86-64	Yes	Yes
Sniper	No	No (suitable only for uncore)	No	Yes	?	x86-64	No	Yes
GPGPUsim	No	?	Yes	No	No	N/A	N/A	N/A
Gem5-GPU X86	Yes	Yes	Yes (DRAMsim)	No	Yes(GEMS)	X86-64	Yes	No
Gem5-GPU Alpha	Yes	Yes	Yes (DRAMsim)	No	Yes(GEMS)	Alpha	Yes	No
ESESC	Yes	Yes (OoO, inorder)	?	Qemu	Yes	No (Arm, MIPS)	No	No
Graphite	No	No	No	Yes	?	Yes	No	Yes
Zsim	No	Yes	Yes	Yes	Yes	Yes	No	Yes

Architectural Memory Simulators

- DRAMsim2
  - a cycle accurate DRAM system simulator at the architectural level
- NVmain
  - a cycle accurate main memory simulator to simulate NVM at the architectural level
- Often integrated with processor-simulators, such as McSim, Zsim, Gem5, Marss etc.

- FPGA Emulation for Memory Behaviors
  - Using FPGA connected to simulator running on host to investigated complex, high throughput memory behaviors

Device-level Simulators

- CACTI (and its variants)
  - An integrated cache and memory access time, cycle time, area, leakage, and dynamic power model
  - Primarily for SRAM, eDRAM and DRAM (2D and 3D)
- NVSim
  - Models the area, timing, dynamic energy and leakage power of NVM technologies
  - For STT-RAM, PCM, ReRAM (NVRAMs), NAND flash
  - Extension of PCRAMsim

Q & A

More info: vetter@computer.org

Contributors and Recent Sponsors

- Future Technologies Group: http://ft.ornl.gov
  - Publications: https://ft.ornl.gov/publications

- Department of Energy Office of Science
  - Vancouver Project: https://ft.ornl.gov/trac/vancouver
  - Blackcomb Project: https://ft.ornl.gov/trac/blackcomb
  - ExMatEx Codesign Center: http://codesign.lanl.gov
  - Cesar Codesign Center: http://cesar.mcs.anl.gov/
  - SciDAC: SUPER, SDAV http://science.energy.gov/ascr/research/scidac/scidac-institutes/
  - CS Efforts: http://science.energy.gov/ascr/research/computer-science/
- DOE 'Application' offices
- National Science Foundation Keeneland Project: http://keeneland.gatech.edu
- NVIDIA CUDA Center of Excellence at Georgia Tech
- Other sponsors
  - ORNL LDRD, NIH, AFRL, DoD
  - DARPA (HPCS, UHPC, AACE)

References

- http://nvsim.org/
- http://wiki.nvmain.org/
- http://quid.hpl.hp.com:9081/cacti/
- http://www.hpl.hp.com/research/cacti/
- http://wiki.umd.edu/DRAMSim2/
- http://scale.snu.ac.kr/mcsim.en.html
- http://snipersim.org/

Integration

- SST can be used with Gem5, which can be used with NVMain/DRAMsim2 (for memory) and CACTI/NVSim (for cache)