NVM-Charade: An Open-Sourced FPGA Based NVM Characterization Scheme

Gieseo Park[1], Mustafa Shihab[1], Lubaba Nahar[1], Wonil Choi[1],
David Donofrio[2], John Shalf[2], and Myoungsoo Jung[1]
[1]Department of Electrical Engineering, The University of Texas at Dallas
[2]Computer Architecture Laboratory, Lawrence Berkeley National Laboratory
{gieseo.park, mustafa.shihab, lubaba.nahar, wonil.choi}@utdallas.edu, {ddonofrio, jshalf}@lbl.gov, jung@utdallas.edu

Abstract

Accurate characterization of real device samples is essential for understanding the true potential of the emerging non-volatile memories (NVMs), and identifying their optimal placement in the memory hierarchy. Unfortunately, most existing publicly-available resources provide performance and reliability data acquired from analytic models or simulation based studies. In this work, we present NVM-Charade, an open-sourced, highly configurable FPGA based empirical data extraction scheme for various NVM technologies. At this juncture, our proposed scheme is a comprehensive evaluation setup with fine-grained user control that extracts reliable endurance, retention, latency, and power data for a real MRAM product.

1. Introduction

Over the past decade, semiconductor industry has demonstrated successful development of multiple promising yet realizable non-volatile memories (NVMs). While conventional memory technologies such as Static Random Access Memory (SRAM) and Dynamic Random Access Memory (DRAM) may struggle to keep up with the uphill challenges of better density, latency, and power, the research community is inspired by the potential of the emerging NVM technologies for successfully addressing those challenges.

Thanks to this, extensive research is being carried out to fit in the NVMs such as flash, phase change memory (PCM), magnetoresistive RAM (MRAM), and resistive RAM (RRAM) into the different layers of the existing memory hierarchy. Specifically, [6] proposed main memory designs that utilize MRAM and PCM for improving performance, while [3] explored designing hybrid main memory modules to combine the speed of DRAM with the persistence of NAND flash. Further, the NVMs are considered to be applied in modern systems as high density cache memories, in an attempt to secure more storage spaces with a lower power consumption. While the density of traditional SRAM based cache is limited by its six-transistor cell design, merging high-density NVMs with SRAM can provide a practical solution to this. Since PCM and MRAM can be optimized to achieve reasonably fast reads and writes, while maintaining their significantly smaller cell area, several studies have proposed to incorporate them in last level cache (LLC) design [5, 7]. Even though these prior approaches explored theoretical potential of the NVMs for a range of memory applications, for most part their results have not yet been verified for real devices.

Incorporating NVMs in the memory hierarchy is a non-trivial task that requires significant optimization of the memory system's configuration, which in turn depends on a wide spectrum of design parameters such as NVM cell design, memory management policy, pin configuration, and the fabrication process. Consequently, extensive empirical data, collected from real devices, is crucial for exploring the full design space of NVM based systems and identifying the required architectural modifications. Unfortunately, most existing publicly-available resources in NVM research provides data generated from overly simplified analytic models or from simulation based studies. Though NVM devices are now available from different manufacturers, lack of appropriate NVM controller and evaluation platform is the main challenge in extracting empirical data from those emerging NVM technologies.

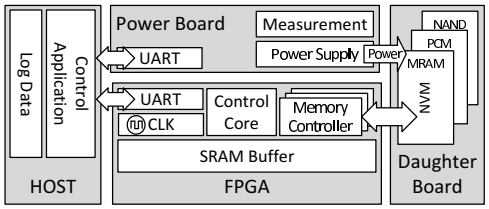


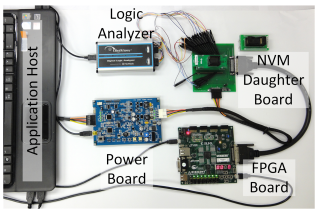
Figure 1: NVM-Charade architecture.

To address these limitations, in this work, we present **NVM Chara**cterization **d**ata **e**xtractor (**NVM-Charade**) – an open-sourced Field-Programmable Gate Array (FPGA) based NVM evaluation and characterization platform equipped with an efficient, highly configurable NVM controller. As shown in Figure 1, our design consists of a daughterboard that can accommodate different types of NVM products for evaluation, and is connected to our custom-made power measurement board and the FPGA, which in turn are controlled by the NVM control application we developed, that resides in the host. We utilize FPGA intellectual property (IP) core to implement our NVM controller because it provides the flexibility of using the same setup to evaluate a broad range of real NVM products at reasonable cost, requiring only appropriate re-configuration of the controller IP. Our proposed platform can evaluate NVM chips and extract empirical data regarding various device-level design parameters such as their retention time, endurance, latency and current consumption, under real working environment. The current **contributions** of this ongoing work are as follows:

– We describe the structural details of NVM-Charade platform, and the basic operation of the evaluation system.
– We collect empirical data for a real MRAM product, and also validate our characterization platform by comparing the extracted data with the manufacturer's specification.

2. Characterization Parameters

Efficacy of any NVM technology is defined by the potential applications in the memory hierarchy, based on its performance, energy and reliability parameters. Even though analytic models and simulation based studies can provide us some insights about a technology, we cannot solely depend on such models for two reasons. First, behavior of a mass-produced memory device with millions of cells and peripheral circuitry can significantly differ from a stand-alone single cell of the same technology. Second, real life operating environment (e.g.,

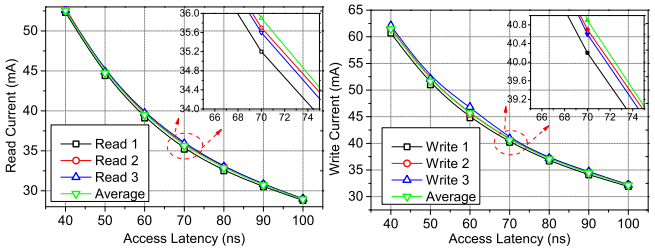


(a) Hardware setup.

Charac. Parameter	Spec. Data	Empirical Data
Read cycle $[ns]$	35	40
Write cycle $[ns]$	35	40
Stand by I_{CC} $[mA]$	9	9
Read I_{CC} $[mA]$	55	53
Write I_{CC} $[mA]$	105	62
Endurance [cycles]	∞	$>10^7$
Retention [years]	>20	>20

(b) Data validation.

Figure 2: NVM-Charade: hardware setup (a), and data validation with evaluated MRAM's empirical data (b).



(a) Read performance. (b) Write performance.

Figure 3: Extracted empirical data for evaluated MRAM chip: read data (a), and write data (b).

temperature, humidity, etc.) can influence the performance of NVM devices. Therefore, collecting empirical data from a real product under practical working conditions is necessary for reliable understanding of the NVM technology. Our proposed NVM-Charade can test and evaluate a broad range of NVM technologies (at this point we have evaluated a MRAM device) for the following characterization parameters:

Latency. Our NVM-Charade reports varying read and write latency for real NVM products, where the accuracy depends on the FPGA specification. A high clock frequency corresponds to a short cycle time, which can offer more accurate latency values by understanding diverse device-level conditions.

Current (Power) consumption. Our custom designed current measurement board can examine current values every 10ms with an 1mA precision. Utilizing this *power board*, NVM-Charade provides current consumed during reads, writes, and standby mode, which in turn can be used for measuring dynamic and static power consumption values.

Reliability. NVM-Charade reports reliability in terms of endurance and retention. The controller writes data at the same memory address repeatedly for 10^7 times to verify endurance. On the other hand, the memory is tested at an specified elevated temperature to emulate and verify 20 year data retention.

3. Implementation and Result

Setup and Operation. We prototyped NVM-Charade based on a *Digilent Nexys3* [2] development board with a *Xilinx Spartan-6* [4] FPGA core. The FPGA has a maximum clock frequency of 500MHz, which in turn can provide up to 2*ns* precision in timing measurements. The FPGA controller maintains the required memory transaction protocols for the target NVM chip, which is mounted on our custom daughterboard (Figure 2a). As the first case for our evaluation platform, we evaluated a MRAM product from *Everspin – Toggle MR2A16A* [1]. A generic computer is used as the host system running the control application, exploiting its two USB-UART converters, one for processing FPGA command and data validity checking, and the other for collecting power consumption data from the power board.

For keeping this initial prototype simple, we assigned most of the basic functions to the FPGA, while processing the complex features on the host-side application. The FPGA sends signal to the NVM through a very-high-density cable interconnect (VHDCI), and communicate with the host application via a universal asynchronous receiver/transmitter (UART). To overcome the slow throughput of UART and accelerate the evaluation process, we incorporate a pattern generation feature on the FPGA. This allows us to save time by not needing to send entire data, and retrieve the latency and reliability verification result only. The power measurement board supplies V_{CC} from the host to the underlying NVM daughterboard, measures current consumption every 10ms with 1mA precision, and sends the measured data to the host via UART.

Evaluation Result and Design Validation. Since MRAM operates on fixed latency based on the access cycle provided by the user, we set both read and write cycle time as predefined parameters. We then calculate read and write current against varying access cycles in the $40 \sim 100ns$ range, which is presented in Figure 3a and 3b. Endurance is tested by performing writing (switching bits) to one specific address and checking data validity after every 1K writes – this is repeated for more than 10^7 times. Finally, to verify data retention, we designed an accelerated stress test using MRAM's retention model – after writing data, device is stored at the maximum allowable temperature (150°C as per Everspin specification), and periodically checked for error. From the retention model, we can determine that retaining data for 60 minutes at 150°C is equivalent to 20 years retention time at 25°C.

Table 2b summarizes the empirical data extracted with our proposed NVM-charade platform, as well as the corresponding values specified by Everspin. By comparing the values one can validate that our NVM-Charade can accurately extract empirical data by evaluating the MRAM device.

4. Future Work and Conclusion

In this work, we proposed NVM-Charade – an open-sourced FPGA based empirical data extraction platform for emerging NVM technologies. We presented the core design of NVM-Charade platform and carried out a sample evaluation of a real MRAM product. We validated the accuracy of data from NVM-Charade by comparing it with manufacturer data. We plan to extend this work and develop controller IPs for other NVM technologies such as PCM and MLC/TLC NAND flash.

Acknowledgement

This research is supported by the ASCR Office in the DOE Office of Science under contract number DE-AC02-05CH1123.

References

[1] "MR2A16A," http://goo.gl/SdvjyV, Everspin Technologies.
[2] "Nexys3 Board - Reference Manual," https://www.digilentinc.com/Data/Products/NEXYS3/Nexys3_rm.pdf, Digilent Inc.
[3] "NVDIMM," http://goo.gl/XgPVQ8, Micron.
[4] "Spartan 6 FPGA Family," http://www.xilinx.com/products/silicon-devices/fpga/spartan-6.html, Xilinx.
[5] P. Mangalagiri *et al.*, "A low-power phase change memory based hybrid cache architecture," in *GLSVLSI*, 2008.
[6] M. Qureshi *et al.*, "Scalable high performance main memory system using phase-change memory technology," *Comp. Arch. News*, 2009.
[7] C. W. Smullen *et al.*, "Relaxing non-volatility for fast and energy-efficient stt-ram caches," in *HPCA*, 2011.