The Non-Volatile Memory Technology Database (NVMDB)
UCSD-CSE Techreport CS2015-1011

Kosuke Suzuki
Fujitsu Laboratories Ltd.
kosuzuki@jp.fujitsu.com

Steven Swanson
UC San Diego
swanson@cs.ucsd.edu

Flash, phase-change, spin-torque, and resistive memories are rapidly transforming how system designers think about memory devices, memory hierarchies, processor architectures, storage systems, operating systems, and applications.

We have gathered a comprehensive survey of the (at last count) 340 non-volatile memory technology papers published between 2000 and 2014 in International Solid-State Circuits Conference (ISSCC), Symposia on VLSI Technology and Circuits (VLSI Technology, VLSI Circuits), and International Electron Devices Meeting (IEDM). The resulting data set provides a clear picture of how these memory technologies have evolved over time.

The links below provide access to a full bibliography for all 340 papers, an Excel spreadsheet summarizing the results, and a technical report describing our methodology and the contents of the spreadsheet.

The data is available online at http://nvmdb.ucsd.edu, and a paper describing the database and our methodology is included below as an appendix. It was presented at the 2015 International Memory Workshop.

If you use the NVMDB in your research please cite it using the this bibtex entry:

```
@TECHREPORT{NVMDB,
  AUTHOR =         {Kosuke Suzuki and Steven Swanson},
  TITLE =          {The Non-Volatile Memory Technology Database (NVMDB)},
  NUMBER =         {CS2015-1011},
  INSTITUTION =    {Department of Computer Science \& Engineering,
                    University of California, San Diego},
  NOTE =           {http://nvmdb.ucsd.edu},
  MONTH =          {May},
  YEAR  =          {2015},
  AUTHOR1_URL =    {http://nvmdb.ucsd.edu},
  AUTHOR1_EMAIL = {swanson@cs.ucsd.edu},
  URL   =          {http://nvmdb.ucsd.edu},
}
```

A Survey of Trends in Non-Volatile Memory Technologies: 2000–2014

Kosuke Suzuki
Fujitsu Laboratories Ltd.
kosuzuki@jp.fujitsu.com

Steven Swanson
UC San Diego
swanson@cs.ucsd.edu

***Abstract*—We present a survey of non-volatile memory technology papers published between 2000 and 2014 in leading journals and conference proceedings in the area of integrated circuit design and semiconductor devices. We present a summary of the data provided in these papers and use that data to model basic aspects of their performance at an architectural level. The full data set and complete bibliography will be published online.**

***Index terms*—STT-MRAM, ReRAM, PCM, PRAM, NAND flash, SCM, Trends survey**

I. Introduction

Flash, phase-change, spin-torque, and resistive memories are rapidly transforming how system designers think about memory devices, memory hierarchies, processor architectures, storage systems, operating systems, and applications. However, with the exception of flash memory, none of these memories are, as yet, commercially available. This makes fully appreciating their potential impact on these aspects of computer systems very difficult.

One source of potential guidance is the vast number of paper describing new cell designs and complete prototype memory devices. These published reports should make it possible to identify trends in memory technology evolution, predict what form a commercially viable device might take, and predict its performance and efficiency characteristics.

Collecting collating all this data is a logistical challenge. By our count, there have been over 300 such papers published since 2000, spanning a wide range of venues in several research communities. The papers present a large but messy data set: they present different sets of metrics and speak the different "languages" of their publication venues.

Finally, while these papers provide important technical information about prototype devices, they frequently omit higher-level metrics that would be of use to processor architects, storage engineers, operating system developers, and application programmers.

This paper presents a comprehensive survey of non-volatile memory technology papers published between 2000 and 2014 in International Solid-State Circuits Conference (ISSCC), Symposia on VLSI Technology and Circuits (VLSI Technology, VLSI Circuits), and International Electron Devices Meeting (IEDM). Besides, the survey includes the latest papers of ISSCC 2015. The resulting data set provides a clear picture of how these memory technologies have evolved over time.

In addition the basic metrics provided by the papers, we also present data for several higher-level metrics (e.g., random access latency and bandwidth) based on a simple memory array model. This analysis should provide guidance to system architects as well as serve as the basis for more detailed studies of how these memory technologies will affect systems when they become commercially viable.

This short paper presents a summary of the data we collected. A complete version of the data set with a full bibliography and additional metrics is in preparation for posting online at http://nvmdb.ucsd.edu.

The remainder of this paper is organized as follows. Section II describes the methodology we used and describes each of the metrics we collected. Section III presents a sample of the data we have collected. Section IV concludes. Due to space constraints, the references section does not include citations for all the data in the figures.

II. Methodology

To collect our data, we surveyed all the papers in ISSCC, VLSI Technology, VLSI Circuits, and IEDM and identified the papers related to NAND flash, phase-change (PCM or PRAM), spin-torque MRAM (STT-MRAM), and resistive ram (ReRAM). This resulted in 340 papers. We also include data from the International Technology Roadmap for Semiconductors' (ITRS) [4] projections for 2013.

For each paper we recorded or computed two kinds of metrics: technology metrics and architectural metrics. Technology metrics are the basic facts about a devices, including its cell size, cell write time, cell read energy, etc. The architectural metrics are higher level measures of how the memory might perform in system. They include read and write bandwidth for a chip and the overall access latency. Table 1 summarizes the technology and architectural metrics we selected.

Most papers provide the majority of the technology metrics either directly or indirectly. For instance, some provide cell size directly (expressed in F^2) while others expressed it indirectly by providing the cell size in square microns the technology feature size.

Very few of the papers provide architectural metrics, and these metrics require us to make assumptions about parts of the system beyond the cell itself. For instance, to compute sequential read bandwidth or random write energy, we need to know the dimensions of the memory array. Likewise, to compute the random read latency for a technology, we need to know the RAS-CAS delay for the array.

III. Data

This section provides a sampling (for space reasons) of the technology and architectural metrics we collected. A complete set of graphs and the full bibliography will be available as a

Tech. Metric	Description
Process Technology	Minimum feature size in nm
Cell size	Size of a single cell in F^2
Read time	Cell read time in ns
Write time	Cell write time in ns
Chip capacity	Bits per die (if applicable)
Average cell size	Average cell size in μm^2, including peripheral logic
Endurance	Write cycle before the cell become unreliable
Bits/cell	Bits stored per cell
Average bit density	Average bit density in GB/cm^2, including peripheral logic

Arch. Metric	Description
Random read BW	Sustained bandwidth for 64 bit random reads
Random write BW	Sustained bandwidth for 64 bit random writes
Seq. read BW	Sustained bandwidth for sequential reads
Seq. write BW	Sustained bandwidth for sequential writes

Table 1: Technology and Architectural Metrics: Section III provides summary graphs for the metrics in bold.

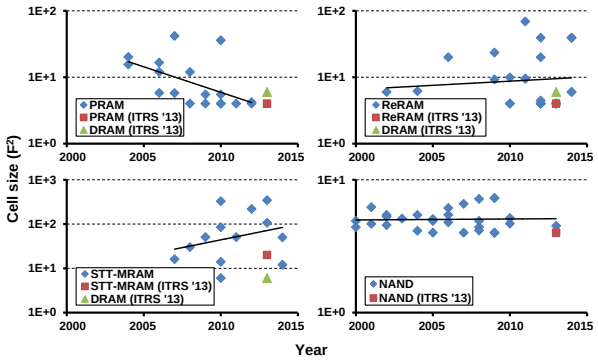

Figure 1: Cell size

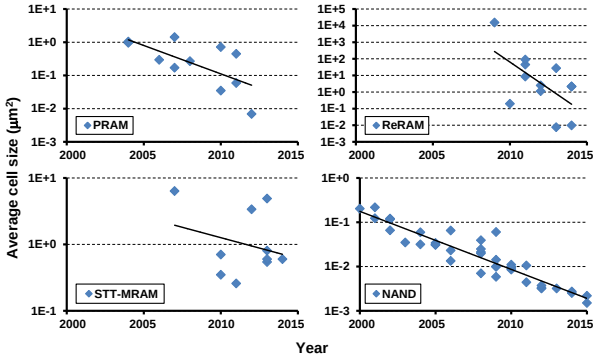

Figure 2: Average cell size

tech report and all of the data will be available online.

Cell size Cell size is a phisycal size of a single cell. The effective area required to store a bit may be smaller due multi-level cells and/or 3D stacking. Figure 1 shows cell size of each NVM. PRAM has already shrunk to 4 F^2.

ReRAM devices use several techniques to switch cell resistance (e.g., bipolar or unipolar, filament or non-filament) and use a range of materials (e.g., metal oxide or Electro-chemical Metalization Bridge) [4]. These difference affect cell geometry, so there is no clear trend.

STT-MRAM cell designs also differ. This is because some papers use two magnetic tunnel junctions (MTJ) to make read time shorter [7]. Cell size of 1-MTJ cells can be as small as 6 F^2 [8]. Using two MTJs increases the size to over 100 F^2.

Cell size of NAND has been around 4 F^2 since 2000.

Average cell size Average cell size is similar to cell size, but it includes the peripheral cirucuits. The smallest average cell sizes for PRAM and ReRAM are 0.0069 μm^2 [1] and 0.0076 μm^2 [10] respectively – $\sim$5$\times$ larger than the smallest NAND cell (0.0015 μm^2 [3]). STT-MRAM cells are much larger: 0.26 μm^2 [5].

Average bit density Average bit density is capacity (Gbit) devided by die size (cm^2). Figure 3 shows average bit density of each NVM. PRAM density is doubling roughly every 1.66 years and has reached 13.5 Gbit/cm^2 [1].

NAND density has been doubling every 1.91 years since 2000. Currently the highest density device achieves 185.8 Gbit/cm^2 [3].

ReRAM devices are denser than PRAM. SanDisk and Toshiba have reported a 24.5 Gbit/cm^2 [10] device. STT-MRAM's best reported density to date is 0.36 Gbit/cm^2 [5]. For comparison, modern DRAMs achieve 9.1 Gbit/cm^2 [9].

Read & write time Read and write times (Figures 5 and 6) reflect the time required to read from a single cell, but they do not include the latency of the peripheral circuitry (e.g., the row and column decoders). For devices with different set and reset times, we report the large of the two. For NAND, we report the program time.

Read time increases with capacity for all technologies except STT-MRAM (e.g., [1, 2]). STT-MRAM read time has been decreasing due to the use of dual MTJ cells.

Write times for PRAM and ReRAM are rising, and although ReRAM is denser, its write time is worse. For example, write time of 32 Gbit ReRAM and 16 Gbit ReRAM are 23 μs and 10 μs respectively [2, 10]. On the other hand, write time of 8 Gbit PRAM is 150 ns [1]. STT-MRAM has good write time but capacity is less than 100 Mbit.

Random read/write bandwidth To compute random read/write bandwidth we use the cell read and write times and assume a DRAM-like DIMM-based architecture (Figure 4) which has 64 I/O pins for data. The read/write timing diagrams per pin are shown in Figure 9 and 10. We assumed read time and write time to set t_{RCD} and t_{RP_WRITE} in these figures. We also assumed burst length (BL) of four [6]. As a result, block size of each access is 32 bytes. The other parameters (t_{RP_READ}, t_{RTP}, etc) were also taken from a DRAM datasheet [6].

The best random read bandwidth of PRAM, ReRAM, STT-

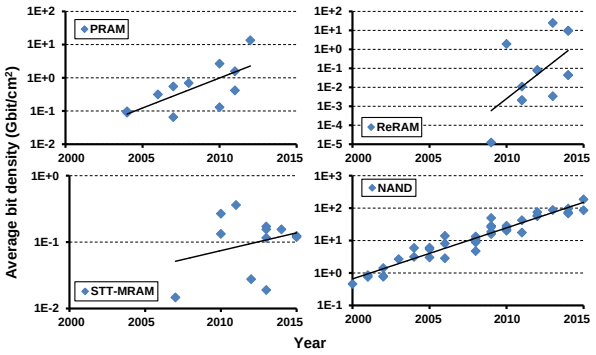

Figure 3: Average bit density

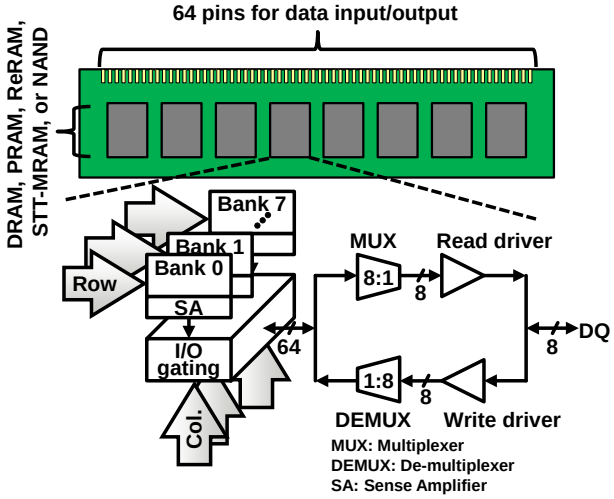

Figure 4: Memory architecture for read/write bandwidth

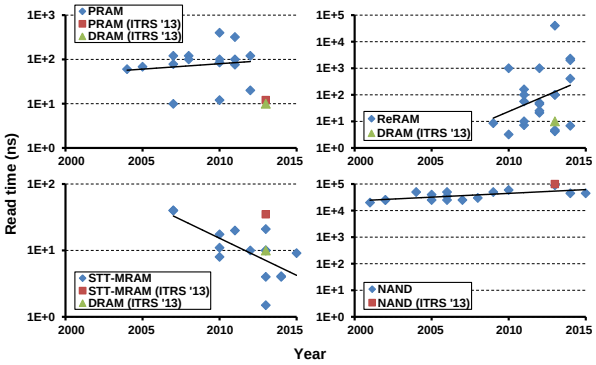

Figure 5: Read time

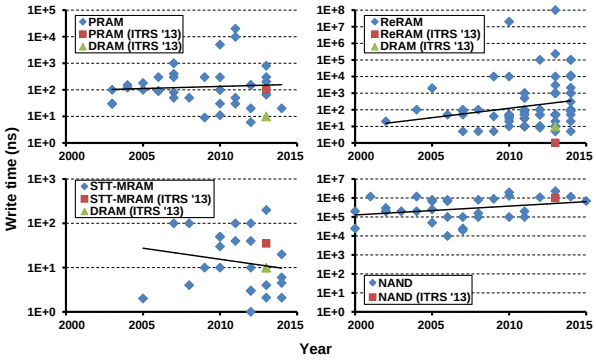

Figure 6: Write time

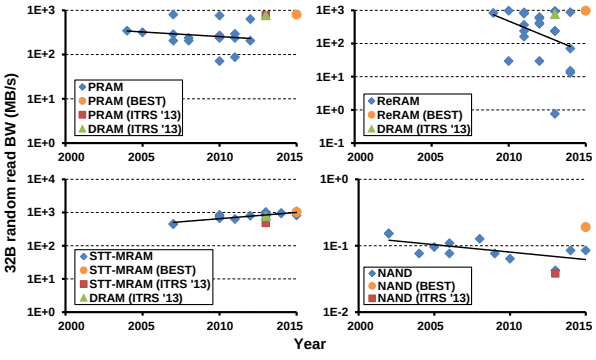

Figure 7: 32B random read BW

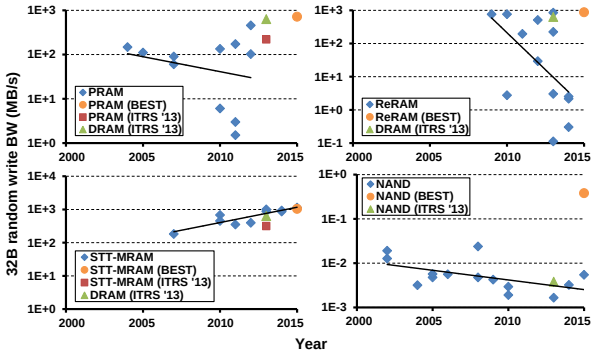

Figure 8: 32B random write BW

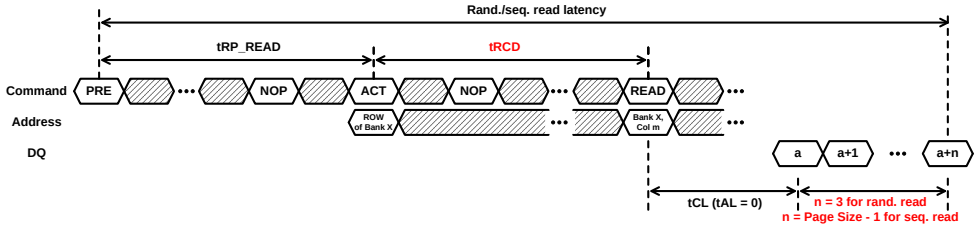

Figure 9: Definition of seq./rand. read latency

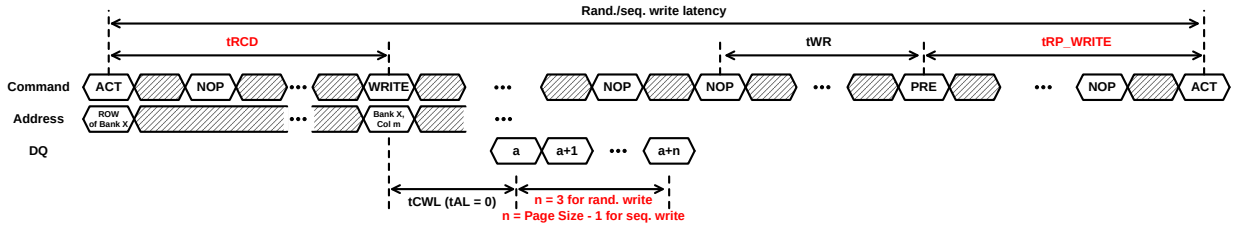

Figure 10: Definition of seq./rand. write latency

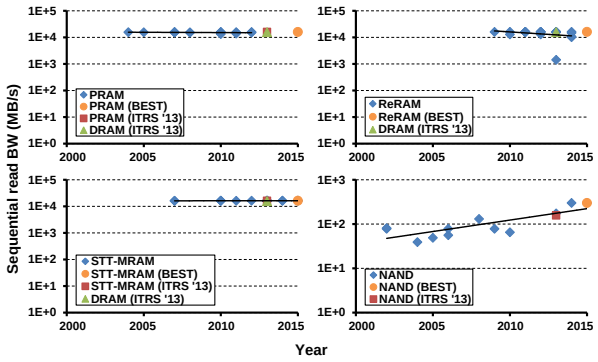

Figure 11: Sequential read BW

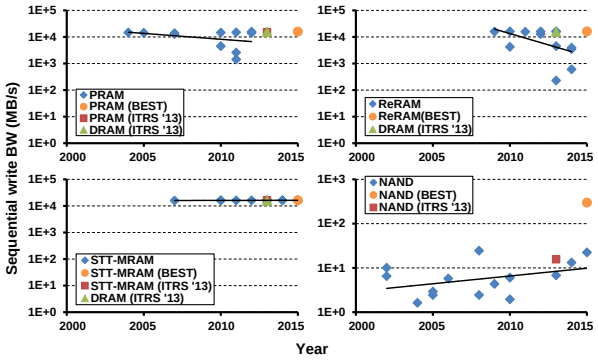

Figure 12: Sequential write BW

MRAM are 802, 974, 1030 MB/s respectively (Figure 7). On the other hand, the best random write bandwidth of them are 716, 874, 1044 MB/s (Figure 8). The read and write bandwidths for the best of these memories may exceed DRAM's (775 MB/s for read, 632 MB/s for write).

Sequential read/write bandwidth We assumed the same memory architecture to compute sequential bandwidths, and we assume that the device reads (or writes) an entire row of the memory array and streams it out over the pins. Papers describing NAND devices provided a page size, but we had to estimate the page size for the other technologies. The main limiter on page size is the need to limit write power. Based on the write energy values given the papers, we use page sizes of 512 B, 1 kB, and 2 kB, respectively, for PRAM, ReRAM, and STT-MRAM. Figures 11 and 12 summarize the data. Sequential access is competitive with DRAM.

IV. Conclusion

We have surveyed NVM technologies papers over the last 14 years and presented a range of device-level and architecture-level metrics. The data illuminate several trends in the evolution of these technologies and will allow system designers to make better estimates of how these memories will affect the design and performance of future systems.

V. References

[1] Y. Choi, I. Song, M.-H. Park, H. Chung, S. Chang, et al. A 20nm 1.8v 8gb pram with 40mb/s program bandwidth. In *ISSCC*, pages 46–48, Feb 2012.

[2] R. Fackenthal, M. Kitagawa, W. Otsuka, K. Prall, D. Mills, et al. 19.7 a 16gb reram with 200mb/s write and 1gb/s read in 27nm technology. In *ISSCC*, pages 338–339, Feb 2014.

[3] J.-W. Im, W.-P. Jeong, D.-H. Kim, S.-W. Nam, D.-K. Shim, et al. 7.2 a 128gb 3b/cell v-nand flash memory with 1gb/s i/o rate. In *ISSCC*, pages 130–131, Feb 2015.

[4] ITRS 2013 Emerging Research Devices (ERD), 2013.

[5] J. Kim, T. Kim, W. Hao, H. Rao, K. Lee, et al. A 45nm 1mb embedded stt-mram with design techniques to minimize read-disturbance. In *VLSI Circuits*, pages 296–297, June 2011.

[6] Micron 4Gb: x4, x8, x16 DDR3L SDRAM Description, 2011.

[7] H. Noguchi, K. Ikegami, N. Shimomura, T. Tetsufumi, J. Ito, et al. Highly reliable and low-power nonvolatile cache memory with advanced perpendicular stt-mram for high-performance cpu. In *VLSI Circuits*, pages 1–2, June 2014.

[8] S. Oh, J. Jeong, W. Lim, W. Kim, Y. Kim, et al. On-axis scheme and novel mtj structure for sub-30nm gb density stt-mram. In *Electron Devices Meeting (IEDM), 2010 IEEE International*, pages 12.6.1–12.6.4, Dec 2010.

[9] T.-Y. Oh, H. Chung, Y.-C. Cho, J.-W. Ryu, K. Lee, et al. 25.1 a 3.2gb/s/pin 8gb 1.0v lpddr4 sdram with integrated ecc engine for sub-1v dram core operation. In *ISSCC*, pages 430–431, Feb 2014.

[10] T. yi Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. Lee, et al. A 130.7mm2 2-layer 32gb reram memory device in 24nm technology. In *ISSCC*, pages 210–211, Feb 2013.