A 4kb Memory Array for MRAM Development

by

Masood Qazi

Bachelor of Science in Electrical Science and Engineering
Bachelor of Science in Physics
Massachusetts Institute of Technology (2006)

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2007

©2007 Massachusetts Institute of Technology. All rights reserved.

Author: ________________________________
Department of Electrical Engineering and Computer Science
February 2, 2007

Certified by: ________________________________
John K. DeBrosse
Senior Technical Staff Member, IBM
Thesis Supervisor

Certified by: ________________________________
Anantha P. Chandrakasan
Professor of Electrical Engineering
Thesis Supervisor

Accepted by: ________________________________
Arthur C. Smith
Professor of Electrical Engineering
Chairman, Department Committee on Graduate Theses

A 4kb Memory Array for MRAM Development

by

Masood Qazi

Submitted to the Department of Electrical Engineering and Computer Science
on February 2, 2007, in Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

The circuits for a A 4kb array of Magnetic Tunnel Junctions (MTJs) have been designed and fabricated in a $0.18\mu m$ CMOS process with three levels of metal. Support circuitry for addressing, reading, writing, and test mode probing enables the characterization of the switching of a thin-film ferromagnetic layer in the MTJs. Specifically, novel mechanisms involving spin-transfer or thermal assistance can be studied and compared to current MRAM designs that switch the MTJ with current-induced magnetic fields. Using this array design, both high speed digital and quasi-static dI/dV experiments can be conducted to investigate the nature of the MTJ resistance hysteresis and process variation in addition to the switching behavior under both polarities of current.

Thesis Supervisor: John K. DeBrosse
Title: Senior Technical Staff Member, IBM

Thesis Supervisor: Anantha P. Chandrakasan
Title: Professor of Electrical Engineering

Acknowledgements

At the heart of the semiconductor industry is semiconductor memories, and at the heart of semiconductor memories is IBM. The kind of project I have had the fortune to undertake could only have come to fruition under the auspices of IBM memory development.

First, I would like to thank Andy Anderson for taking the risk to hire me in 2004. I am also grateful to my manager John Gabric for compelling me to meet his high expectations and return every year since my first assignment. I suspect that only later in my professional career will I fully appreciate how lucky I was to have him as my first "boss." I also appreciate the support from Bill Gallagher at IBM research for his role in conceiving a challenging, risky project for me and providing the guidance at critical junctures of this work that made it succeed as a thesis.

The MRAM processing technology involved in this project is based on the comprehensive body of knowledge and expertise cultivated by the MRAM team at IBM research. In particular, Solomon Assefa has played a central role in developing the process for fabricating the experimental magnetic tunnel junctions for which this 4kb array was intended. Furthermore, Jonathan Sun has been forthcoming in discussing his research on spin transfer effects in nanomagnets in addition to introducing me to the rich field of magnetism and magnetic materials. Janusz Nowak has also helped in characterization measurements on magnetic tunnel junctions that have guided my experiments.

For the circuit design, which is more immediate to the contributions of this thesis, I would like to thank Tom Maffitt for sharing his insight obtained over years of experience in DRAM and, more recently, MRAM design. I also appreciate Mark Jacunski's willingness and ability to teach me about memory circuit design, particularly for his methodical approach to integrated circuits and for taking time from his demanding responsibilities in embedded DRAM design. I would also like to thank Mark Lamorey for his extensive work on mask-related processing issues; he ultimately ensured that my design data got appropriately translated to physical masks for fabrication.

I am grateful to Mark Wood for not only his assistance in the layout of this project but for taking me through the elements of laying out a complex chip design with 10^6 to 10^9 transistors, drawing upon principles of hierarchy, robust wiring, techniques for tight pitch circuits, and device matching for analog circuits. His personality made the weeks of sitting with him in front of the layout software tools much more enjoyable than they should have been. This project also received significant contributions from Kim Maloney in laying out several circuit blocks of the 4kb Array.

For the formidable task of wafer-level test on a memory array with over 40 signals—several of which require timing control on a time scale of $10ns$—I cannot emphasize enough the vital role of John Parenteau and the memory tester which he helped develop over the course of twenty years. His test environment enabled me to exercise the array in several different ways, many of which were unanticipated during the design phase. In fact, the memory tester functioned as almost an extension of the integrated hardware on the wafer—in my incremental approach to extracting functionality from experimental, uncontrolled,

and less-understood magnetic memory elements. I would also like to thank Alan Yaeger for helping me with testing. His problem solving skills and fearless attitude in the face of new and unexpected challenges with electrical equipment helped me overcome severe obstacles to the data gathering phase of this project.

Finally, my mentor John DeBrosse has been involved in each step of this project, keeping me on a path—for over one and a half years—that ultimately resulted in viable integrated hardware. In working with him, I have experienced a form of teamwork beyond the mere partitioning of responsibilities; his feedback and ideas shaped my inchoate thoughts into a design for a 4kb memory array and exposed me to work in MRAM and DRAM beyond the scope of my project. Because of his experience in the multifaceted elements of memory development and ability to articulate his thought process, he has made work in MRAM circuit design challenging, exciting, and rewarding. In the course of my career, I hope to acquire such elements of technical leadership.

It is self-evident from the nature of the work described in this thesis how dependent it was upon these people. Remarkably, they made their contributions to my thesis in parallel to fulfilling their own work obligations. May this project ultimately reflect an additional capacity of theirs to advance memory technologies.

Masood Qazi
Cambridge, MA

Contents

1 Introduction **13**
1.1 The Memory Landscape . . . 14
1.2 Previous MRAM Work . . . 16
1.3 Problem Statement . . . 18
1.4 Contributions of this Work . . . 19

2 Magnetics Review **21**
2.1 The Magnetic Dipole . . . 21
2.2 Properties of Nanomagnets . . . 32
2.2.1 The Fields and Energy of a Nanomagnet . . . 33
2.3 Magnetization Dynamics . . . 40
2.4 The MTJ structure . . . 41
2.5 Spin Angular Momentum Transfer . . . 42

3 Design of the 4kb Array **47**
3.1 Overview . . . 47
3.2 The Cell . . . 49
3.3 Row Path . . . 54
3.4 Column Path . . . 56
3.4.1 Control Logic . . . 56
3.4.2 MBL/SBL grounding . . . 57
3.5 Magnet Wire . . . 59
3.6 Sense-amplifier . . . 62
3.6.1 Sizing of mirrors and source follower clamp device . . . 65
3.6.2 Design of transconductance amplifier . . . 67
3.6.3 Analysis of loop dynamics . . . 78
3.6.4 Transient operation . . . 81
3.7 Write Drivers . . . 82
3.8 Operation . . . 83
3.8.1 PULSE timing . . . 84
3.8.2 Standard Write . . . 84
3.8.3 Standard Read . . . 86

3.8.4 TMRBL . . . 87
3.8.5 TMSENSE . . . 88
3.9 Layout Floorplan . . . 88

4 Testing 89
4.1 Test Setup . . . 91
4.2 Experimental Results on Resistance Bitline . . . 94
4.2.1 Write Pulses . . . 94
4.2.2 Read Pulses . . . 97
4.2.3 Senseamp Reference Sweep . . . 106
4.3 Description of Initial Test Plan . . . 111
4.4 Results from Field Switching . . . 115

5 Conclusion 119
5.1 Summary of Contributions . . . 119
5.2 Future Work . . . 120

A Electromagnetics Reference 121
A.1 Maxwell's Equations . . . 121
A.2 Derivation of Spin-Transfer Switching Dynamics for a Mono-domain model . . . 122

Bibliography 129

List of Figures

1.1 The memory landscape . . . 16

2.1 The dipole field . . . 22
2.2 Calculation of magnetic field . . . 22
2.3 A current loop . . . 24
2.4 Classical precession . . . 27
2.5 Magnetized ellipsoid along the "easy" axis . . . 34
2.6 Magnetized ellipsoid along the "hard" axis . . . 34
2.7 The Stoner-Wolfarth Astroid for a monodomain magnet . . . 39
2.8 The MTJ stack . . . 41
2.9 Representation of spin torque due to current between two ferromagnets . . . 42

3.1 Top-level block diagram of the ADM . . . 48
3.2 Schematic cross-section of array . . . 49
3.3 Spin-transfer switching in MTJs . . . 50
3.4 Loadline of IV hyteresis in a bidirectional cell . . . 50
3.5 Vertical cross-section of the memory cell . . . 52
3.6 Cell layout . . . 53
3.7 Schematic of a "one out of eight" predecoder." . . . 55
3.8 Field produced by magnet wire . . . 60
3.9 Plot of field produced by magnet wire . . . 60
3.10 Circuits for one of three magnet wires . . . 61
3.11 Prior sense-amplifier topology. . . . 63
3.12 Sense-amplifier topology. . . . 65
3.13 Sizing of current mirror load . . . 67
3.14 Schematic of transconductance amplifier . . . 68
3.15 Full small-signal schematic of transconductance amplifier . . . 69
3.16 Simplified small-signal schematic of transconductance amplifier . . . 75
3.17 Feedback loop for M5 . . . 75
3.18 Small-signal schematic to calculate overal G_m . . . 76
3.19 Simulation of VTC of feedback amplifier . . . 77
3.20 Small signal circuit for stability analysis . . . 79
3.21 Bode plot of sense-amplifier loop transmission . . . 80

3.22 Transient wave form of sense-amplifier 81
3.23 The outpout of the SA goes through two latches 82
3.24 Schematic of current driver 83
3.25 Timings for PULSE signal 84
3.26 Timing diagram for write cycle 85
3.27 Timing diagram for read cycle 86
3.28 ADM floor plan for major core circuits 88

4.1 MACE Block Diagram 93
4.2 Current write pulses at $V_{dd} = 1.8V$ 95
4.3 Voltage write pulses at $V_{dd} = 1.8V$ 95
4.4 Current write pulses at $V_{dd} = 3.6V$ 96
4.5 Voltage write pulses at $V_{dd} = 3.6V$ 96
4.6 $50mV$ read pulses with high resolution scope probe. 98
4.7 Close examination of a $50mV$ read pulses for resistance value in the middle of the expected operating range. 99
4.8 $50mV$ read pulses with high bandwidth scope probe. 100
4.9 Close examination of a $50mV$ read pulses for resistance value in the middle of the expected operating range (high bandwidth scope). 101
4.10 $100mV$ read pulses with high resolution scope probe. 102
4.11 Close examination of a $100mV$ read pulses for resistance value in the middle of the expected operating range. 103
4.12 $100mV$ read pulses with high bandwidth scope probe. 104
4.13 Close examination of a $100mV$ read pulses for resistance value in the middle of the expected operating range (high bandwidth scope). 105
4.14 The digital output correctly reads the resistance of the memory cell. 106
4.15 Read 0 failcount plot for a single cell. 108
4.16 Read 0 failcount plot for full array. 108
4.17 Successful extraction of resistance distribution on RBL. 110
4.18 Single cell failcount curves for both W1R0 and W0R0 test patterns. 114
4.19 Full array failcount curves for both W1R0 and W0R0 test patterns. 114
4.20 Cumulative distribution after W1 and W0 115
4.21 Histograms after W1 and W0 116
4.22 Two dimensional sweep of switching field 118

List of Tables

3.1 Small signal design considerations. 74
3.2 Functional description of normal operation 84
3.3 Timing values for a write cycle. 85
3.4 Timing values for a read cycle. 87

Chapter 1

Introduction

The three most significant semiconductor memories in today's integrated circuit market are DRAM, SRAM, and FLASH. Each type of memory has a distinct set of advantages in terms of speed, density, non-volatility, and power. SRAM offers the fastest speeds but compromises on density because of a six-transistor (or sometimes four-transistor) cell. DRAM offers higher density with a one-transistor cell and storage capacitor but operates at slower speeds than SRAM. FLASH offers the density of DRAM and non-volatility but has write cycles several orders of magnitude slower than the other two RAM memories. Across these three memories, power is also a consideration through transistor off current in SRAM, refresh requirements in DRAM, and large write voltages and currents in FLASH [1].

Electronic systems like personal computers, mainframes, and mobile phones benefit from the distinct advantages of each type of memory. Thus, a need exists to more effectively integrate the different types of memories into one unit. A non-volatile RAM memory could be a "game-changer" to the semiconductor memory industry by offering the advantages of multiple memories in one chip [2]. For portable systems, it would allow for increased space and energy efficiency. In general, it would simplify system architecture, reduce hardware cost, and enable instant-on functionality. One possible candidate for non-volatile RAM is magnetoresistive random access memory (MRAM), which is comprised of arrays of Magnetic Tunnel Junctions (MTJs) whose states are stored as high or low resistances, depending on

the parallel or anti-parallel alignment of two thin-film ferromagnetic layers. Some of the advantages of MRAM as a "universal memory" are: it can retain its state with zero power; it is radiation immune in space applications; it requires 400 times less write power than FLASH; it has unlimited write endurance; and, it has comparable densities and speeds to SRAM and DRAM. [3].

Conventional MRAM memories have manipulated ferromagnetic layers in MTJs through current-induced magnetic fields, posing problems for isolating bits and working within the operating range of CMOS technology. This project aims to make a first step towards the development of a new kind of MRAM memory, differing from its predecessors through novel switching mechanisms based on spin-transfer or thermal effects. The vehicle for this investigation will be a 4kb array development macro (ADM) designed as a functional memory unit that also allows detailed experimental modes to measure the switching and read characteristics of MTJs.

1.1 The Memory Landscape

Shown in Fig. 1.1 is a comparison of the cost-performance tradeoff made by several types of memories. On the horizontal axis is the *random* access time. [1] This value corresponds to the minimum time required between (1) a read or write operation at a given address in the memory and (2) a subsequent read or write operation at another, arbitrarily chosen, address location in the memory. For the vertical axis, the high-volume unit cost was divided by the memory size to give a cost per bit. One can also interpret this as a proxy for cell area, but the quotation in terms of $ permits comparison accross memory technologies that have different processing costs for the same die size.

Not shown in this plot are considerations related to power consumption and maximum

[1] The data for Fig. 1.1 comes from the following chips: HYB18T1G160BF-5, IS42S32200C1-7TL, IS42S32800B-7TL (DRAM); CY7C1512V18, IS61LV6416-10TL, CY7C1041CV33-12ZXC (SRAM); CAT28F010H-90, NAND512W3A2BN6E, LHF00L13, SST29SF040-55-4C-NHE (FLASH); MR2A16ATS35C (MRAM). For hard drives, Maxtor Ultra 16 and Wester Digital Caviar SE 250GB hard drives were used. The datasheets, prices, and other specifications were accessed in Jan. 2007.

write/read bandwidths. Yet, an effort was made to select representative parts available for purchase from online electronics component sellers to give a reasonably fair comparison of random access capabilities.

Immediately one can see the ultimate in cost is the hard disk drive, and the ultimate in random access time is SRAM. DRAM offers a cheaper alternative to SRAM that is still fast enough to be sufficient in many applications. However, the low-performance of FLASH and hard disk drives will necessitate their accompaniment by DRAM or SRAM in electronic systems. This addition of FLASH or a hard disk drive brings two advantages; the cost of mass data storage can be significantly lowered and the data can be preserved during a power down and power up cycle (this second advantage is defined as non-volatility). To cope with the much slower random access time, techniques based on increasing the address locality of serially written and read data have been developed to maximize the bandwidth of these two memories. Finally, FLASH has asymmetrically faster read performance than write performance and has a smaller form factor than a hard disk drive. These features of FLASH combined with SRAM provides a viable alternative to the fifth memory in the landscape: MRAM.

The MRAM memory currently available from Freescale Semiconductor aptly describes MRAM's current status as costly, fast, and nonvolatile. Although the cell area of MRAM ($1.2-1.6\mu m$ for $180nm$ node) is between that of DRAM and SRAM, the magnetics processing and smaller market push its cost above both SRAM and DRAM. Without a compelling reason for simultaneous fast random access and non-volatility, this cost discrepancy makes SRAM+FLASH five to ten times cheaper than MRAM. Nevertheless, MRAM shows promise with better endurance than FLASH and less static power consumption than SRAM, especially with scaling to smaller technology nodes. As new applications and system designs emerge to leverage MRAM's unique combination of simultaneous nonvolatility and random access, the cost of magnetic processing decreases, and the acceptance of MRAM for mainstream use increases, MRAM will become more viable. For these reasons, MRAM is still worth pursuing at smaller semiconductor technology nodes.

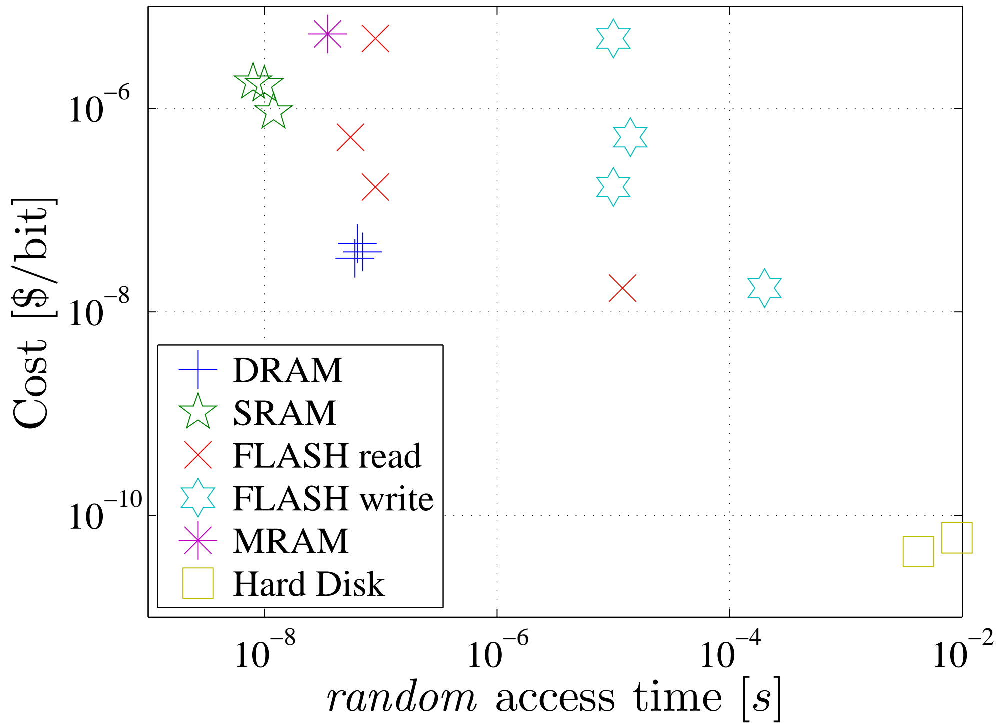


Figure 1.1: The memory landscape: a comparison of the cost-performance tradeoff made by several types of memories. Note that MRAM, FLASH, and hard disk drives are also nonvolatile.

1.2 Previous MRAM Work

The switchable resistance of an MTJ structure based on the relative alignment of the magnetization of two ferromagnetic layers was first reported by Julliere in 1975 [4]. As one layer's magnetization varies from parallel to antiparallel alignment with the other, the density of electronic states at the energy level of conduction electrons changes for a given spin state, while it remains unchanged in the other layer. Thus, the read current, consisting of electrons traveling from one layer to the other, faces an impedance that depends on how well like spin states on the two sides of the MTJ match through an energy barrier [5]. Currently, MTJ technology has matured in terms of reliability in a CMOS manufacturing environment to the point where the change in resistance—70% to 200% of the low resistance value—is enough to provide a measurable signal for CMOS circuits [6], [7].

In fact, engineering one of the ferromagnetic layers to be fixed and the other to be switchable between parallel and antiparallel directions allows the design of nonvolatile MRAM memories. A selected MTJ can be switched by passing currents near it in order to manipulate its free layer magnetization through current-induced magnetic fields. The resistance can be sensed by setting a voltage across the MTJ and comparing the resulting current to a midpoint reference current [3]. Beyond device-level considerations of hysteresis and resistance values, two fundamental architectural issues must be addressed: isolation of cells and compatibility with the operating range of CMOS circuits. With this in mind, two main architectures have been proposed: (1) a cross point (XPT) architecture with MTJs directly connected between bitlines (BLs) and wordlines (WLs) at their points of perpendicular intersection and (2) an isolation cell transistor (1T1MTJ) architecture with an MTJ connected in series with a transistor at the intersection of a bitline and a read word line. Also, a second write wordline runs under the MTJ in the 1T1MTJ cell. So far, only 1T1MTJ arrays have been seriously pursued because of more robust electrical operation [8].

Promisingly, functional 1T1MTJ MRAM memories have achieved reasonable density (locally in terms of cell area, and globally in terms of array efficiency), speed, and power consumption with respect to their competitors (SRAM, DRAM, FLASH). A successful 16Mb chip has been reported by the IBM-Infineon MRAM Development Alliance that switches MTJs with current-induced magnetic fields. It was fabricated in a $0.18\mu m$ CMOS process and demonstrated read/write cycle times around $30ns$, high bit functionality, and non-volatility [9]. Furthermore, an arguably more robust toggle-mode MRAM has been demonstrated by a team at Freescale Semiconductor (originally developed under Motorola) which achieves improved write reliability with “toggle” MTJs that have two coupled free layers instead of only one free layer [10]. In fact, Freescale’s MR2A161A, a 4Mb MRAM chip with an SRAM-like 16 x 256k interface, is commercially available.

1.3 Problem Statement

Although 180nm node MRAM demonstrations show promise in achieving sufficient isolation of bits and compatibility with CMOS, scaling to smaller technology nodes amplifies these difficulties. In order to preserve the same thermal energy barrier in a smaller MTJ, a higher magnetic switching threshold must be engineered in order to compensate for the decrease in total magnetic moment. This magnetic constraint requires a larger current to switch. Firstly, this limits array size because of IR drops in wiring—whose resistance is also increasing with narrowing widths—ultimately reducing efficient usage of chip area. Secondly, it increases write power consumption beyond already tenuous WL and BL currents of $1mA - 10mA$. In addition, smaller spacing comparatively increases the disruptive effect of stray magnetic fields in "half-selected" (on active BL but not WL or vice versa) and other adjacent cells [8]. Although techniques such as cladding BL and WL wires with magnetically susceptible liners have the potential to mitigate these problems, methods beyond conventional field-switching MRAM could possibly achieve greater isolation and lower current [3].

In 1996, J. Slonczewski predicted the ability to switch parallel magnetic films by passing smaller currents directly through them, instead of passing larger currents adjacent to them for conventional field switching [11]. This so-called spin-transfer switching (STS) [2] is viable in smaller MTJs, as the spin of the conduction electrons passing through the MTJ structure can more strongly influence the macroscopic magnetization of the free layer. In 2004, STS phenomena has been reported in a spin-valve, a structure similar to an MTJ but with copper separating the magnetic layers instead of a tunneling oxide. A hysteresis with current switching was demonstrated, and sub-nanosecond speeds were observed [12]. Similar STS switching has also been reported in true MTJs with an oxide barrier between the ferromagnetic layers [13], [14].

So far, experiments on MTJ structures have been mostly done with isolated conductive paths to external probes in the development of STS MRAM. The first functional MRAM

[2] Spin-transfer switching is also referred to as spin angular momentum (SMT) tranfer and spin torque tranfer (STT).

array with support circuits for addressing, reading, and spin-transfer writing MTJs has been reported in December 2005 by a team at Sony [15]. Their investigation is not as aggressive as this project in terms of write currents, and they leave unanswered to what extent their array can operate beyond a probabilistic switching regime. Write error rates that meet industry standard specifications have yet to be demonstrated in an STS MRAM array.

Another approach to mitigate the write current requirement of field switched MRAM has been proposed by [16] as thermally assisted switching in which the MTJ's hysteresis thresholds—in magnetic field—become smaller with increasing temperature. This thermally assisted switching (TAS) has been demonstrated by [17] with FET isolated MTJs in a homogeneous external field; a shrinking hysteresis was measured as a heating current through the device was increased. To date, no arrays with thermally switched MRAM memory cells, and locally generated high speed write fields have been reported.

1.4 Contributions of this Work

A 4kb memory array with a one-transistor one-MTJ cell that supports bidirectional currents through the memory element has been developed. Full functionality of the fabricated array circuitry has been demonstrated on a dummy bitline of resistor cells, and the array has also been used on experimental MTJs to explore spin-transfer switching along with other magnetic and electrical properties.

This application of the 4kb array has led to a methodology for testing future iterations of MTJ hardware based on extracting resistance distributions before and after application of write pulses, and varying write conditions while reading at a fixed, optimum read reference. These experiments will allow one to seek answers for the following questions:

- What are the fastest reliable write cycles possible? What is the switching time as a function of write current, especially in the super-threshold deterministic switching regime?

- What types of resistance values, and resistance changes between the two states are achievable in scaled MTJs?
- What is the quantitative variance of the above measurable quantities? How big is the design window for a Spin-MRAM product demonstrator?
- Can STS switching work with very low error rates similar to the soft error rates (SER) of DRAM and SRAM? What is the effect of read current intensity on the disturbance of the MTJs?
- How well do current theoretical models describe the spin transfer switching?
- What circuit techniques will be needed to make Spin-MRAM work?

In the following chapters, magnetism related to MRAM will be reviewed (chapter 2); the design of the 4kb array will be outlined (chapter 3); and initial test results on integrated hardware will be presented (chapter 4).

Chapter 2

Magnetics Review

Operationally, MRAM is very simple to describe, but an explanation from basic physical principles requires a greater degree of technical sophistication. This chapter aims to outline key results from electromagnetism and specific magnetics theories that the MRAM circuit designer needs. This understanding of MTJ operation will allow the reader to appreciate the design considerations and the implications of experimental results for the 4kb array.

2.1 The Magnetic Dipole

The magnetic dipole [1] is the basic unit of magnetic interaction. The magnetic field produced by a magnetic dipole $\vec{m} = m\hat{z}$ is given by: [18, p. 409]

$$\vec{H}_{dip} = \frac{m}{r^3}\left(2\cos\theta\hat{r} + \sin\theta\hat{\theta}\right) \tag{2.1}$$

This field, along with the coordinate system used herein is depicted in Fig. 2.1.

In fact, the magnetic field of an arbitrary distribution of static currents, as shown in

[1] The discussion of the magnetic dipole in this section, including the chosen examples, is a compendium of results from textbooks by Purcell [18], Griffiths [19], Jackson [20], and Sakurai [21]. Further explanation can be found in the textbooks, and page numbers have been provided. The units used in this chapter are CGS; the use of SI units will be explicitly highlited.

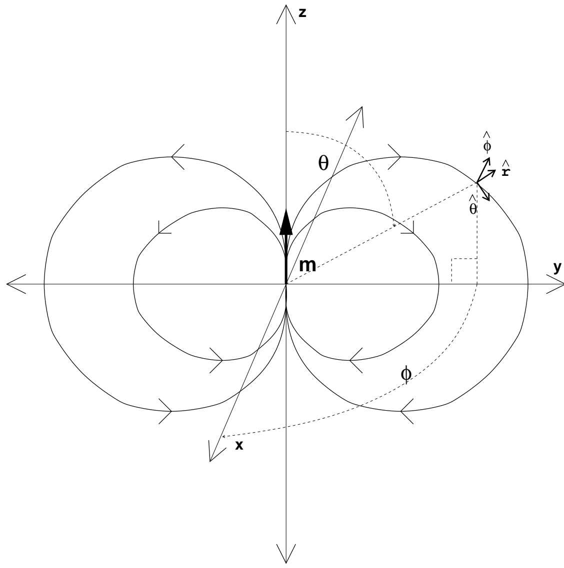


Figure 2.1: The field produced by an ideal dipole at the origin

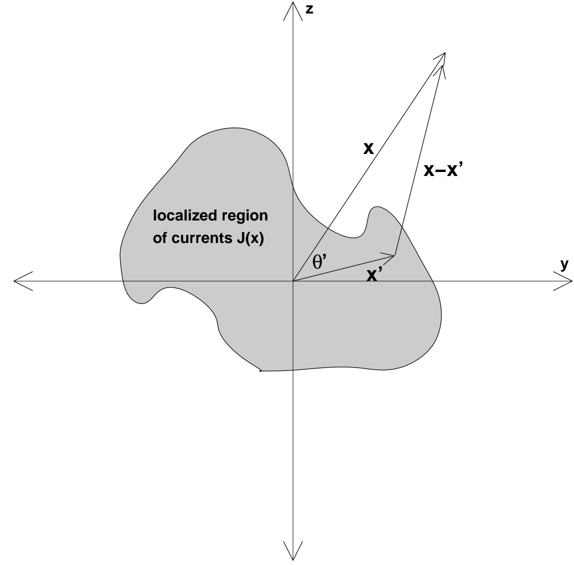


Figure 2.2: The setup for the calculation of an aribitrary distribution of static currents

Fig. 2.2, can be obtained by evaluating the vector potential $\vec{A}(\vec{x})$: [19, p. 234]

$$\vec{A}(\vec{x}) = \frac{1}{c}\int \frac{\vec{J}(\vec{x'})}{|\vec{x}-\vec{x'}|} d^3\vec{x'} \tag{2.2}$$

and translating to field with [2]

$$\vec{B} = \vec{\nabla} \times \vec{A} \tag{2.3}$$

At this point, it is useful to examine the expansion of the $1/|\vec{x}-\vec{x'}|$ term in the denominator of Eq. 2.2

$$\frac{1}{|\vec{x}-\vec{x'}|} = \frac{1}{|\vec{x}|}\sum_{n=0}^{\infty}\left(\frac{|\vec{x'}|}{|\vec{x}|}\right)^n P_n(\cos\theta')$$

where $P_n(x)$ signifies the legendre polynomial series. This expression leads to a multipole

[2] $\vec{x}$ signifies the cartesian position vector: $\vec{x} = x\hat{x} + y\hat{y} + z\hat{z}$. Furthermore, the unit position vector will be given as $\hat{r} = \vec{x}/|\vec{x}|$ and sometimes r will be used in place of $|\vec{x}|$.

expansion of $\vec{A}(\vec{x})$: [19, p. 234]

$$\vec{A} = \frac{1}{c|\vec{x}|}\int \left[\vec{J}(\vec{x'}) + \frac{1}{|\vec{x}|}\vec{J}(\vec{x'})|\vec{x'}|\cos\theta' + \frac{1}{|\vec{x}|^2}\vec{J}(\vec{x'})|\vec{x'}|^2\left(\frac{3}{2}\cos^2\theta' - \frac{1}{2}\right) + \ldots\right] d^3\vec{x'} \quad (2.4)$$

The first term based on $\int \vec{J}(\vec{x'})d^3\vec{x'}$ must be zero because there is no net growth or decrease in charge by construction of the example as a *localized* distribution of currents. Namely, the average current in the x, y, and z directions must be zero. One can show that this first term in Eq. 2.4 is merely a vector whose components are directly proportional to the average current along the corresponding axes. For example, assuming that the region is bounded by x-z planes located at $y = a$ and $y = b$:

$$\begin{aligned}\int J_y(\vec{x'})d^3\vec{x'} &= \int_a^b dy' \iint dx'dz' J_y(\vec{x'}) \\ &= \int_a^b dy' I_y(y') \\ &= (b-a)\cdot < I_y >\end{aligned}$$

which must be zero since $< I_y >= 0$ is an equivalent statement of the fact that the current distribution is localized in y. With the condition that there are no sources and sinks of charge in the distribution, one can make an even stronger statement that $I_y(y)$ is identically zero.

Therefore, the $1/|\vec{x}|^2$ term will dominate the expression for $\vec{A}$ at sufficiently far enough distances. Although the mathematical development of Eq. 2.2 and the interpretation of $\vec{J}$ showed this to be true, the fundamental reason comes from two of Maxwell's equations. $\vec{\nabla}\cdot\vec{B} = 0$ allows $\vec{B}$ to be expressed in the form of Eq. 2.3, and $\vec{\nabla}\times\vec{B} = \frac{4\pi}{c}\vec{J}$ allows a solution for $\vec{A}(\vec{x})$ in the form of Eq. 2.2. [3]

Now the dipole moment vector $\vec{m}$ can be redefined in terms of the prefactor of the $1/|\vec{x}|^2$

[3]Eq. 2.2 is obtained by choosing $\vec{\nabla}\cdot A = 0$ and then applying Poisson inversion. [22, p. 596] It is not the only possible solution.

term in Eq. 2.4:

$$\vec{m} \times \hat{x} = \frac{1}{c} \int \vec{J}(\vec{x'})|\vec{x'}| \cos\theta' d^3\vec{x'} \tag{2.5}$$

restates the vector potential of a dipole moment as:

$$\vec{A} = \frac{\vec{m} \times \hat{r}}{|\vec{x}|^2} \tag{2.6}$$

Taking the curl of this equation recovers $\vec{B}$ as given in Eq. 2.1. Note that $\vec{H}$ is defined as:

$$\vec{H} = \vec{B} - 4\pi\vec{M} \tag{2.7}$$

and is equivalent to $\vec{B}$ outside of the presence of magnetic media, which is represented by nonzero $\vec{M}$, and will be further discussed later. In examples of practical interest, it is sometimes easier to solve Maxwell's equations in terms of $\vec{H}$.

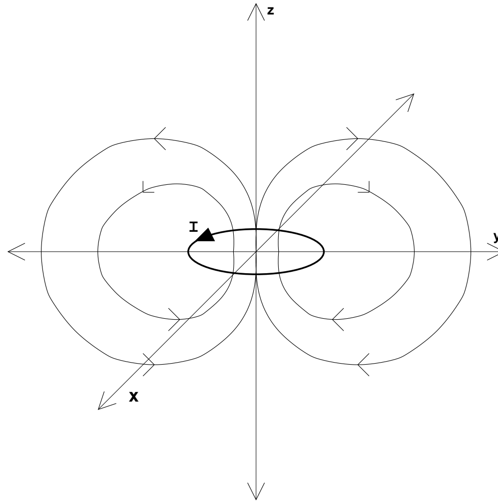


Figure 2.3: A prototypical current loop useful for evaluating the properties of an ideal dipole.

A useful example for working with dipoles is a current loop as shown in Fig. 2.3. Evaluating its dipole moment via the right hand side of Eq. 2.5

$$\frac{1}{c} \int \vec{J}(\vec{x'})|\vec{x'}| \cos\theta' d^3\vec{x'} = \frac{1}{c} \int I|\vec{x'}| \cos\theta' d\vec{l} \qquad [19, \text{p.}236]$$

and associating this with the left hand side of Eq. 2.5 (in addition to applying vector identities as in [20, p. 185]) gives:

$$\vec{m} = \frac{I}{c}\int d\vec{a} = \frac{I}{c}\,\vec{a} = \frac{I}{c}\,(\text{area of loop})\;\hat{z} \tag{2.8}$$

This is the dipole moment of a current loop. At far distances relative to the size of the current loop, the field will approach that of Eq 2.1. Thus, an ideal dipole will behave like this current loop in the limit of arbitrarily large current, vanishingly small area, and constant $I|\vec{a}|$.

This concrete example of a dipole allows one to apply the lorentz force law on the moving charges in the loop:

$$\vec{F} = q\frac{\vec{v}}{c}\times\vec{B} \tag{2.9}$$

to derive the torque on a dipole like the one in Fig. 2.1 from a uniform external field $\vec{H} = H\hat{z}$:

$$\vec{\Gamma} = \vec{m}\times\vec{H} \tag{2.10}$$

The work done by a magnetic field on a dipole in moving from one orientation at (θ_1, ϕ_1) to another oerientation with (θ_2, ϕ_2) is:

$$\begin{aligned} W &= \int_{\theta_1}^{\theta_2}\Gamma d\theta \\ &= \int_{\theta_1}^{\theta_2}|\vec{m}||\vec{H}|\sin\theta d\theta \\ &= -|\vec{m}||\vec{H}|\left(\cos\theta_2 - \cos\theta_1\right) \end{aligned} \tag{2.11}$$

This expression is independent of the path in θ-ϕ space because the cross product results in zero torque on the azimuthal component of rotation. Hence, this conservative torque

contributes an energy term dependent on the dipole's deviation from the field:

$$U = -\vec{m} \cdot \vec{H} \tag{2.12}$$

This equation allows a direct derivation of the force on a dipole, which is non-zero only in the presence of a non-uniform magnetic field:

$$\begin{aligned} \vec{F} &= -\vec{\nabla} \cdot U \\ &= m_x \vec{\nabla} H_x + m_y \vec{\nabla} H_y + m_z \vec{\nabla} H_z \end{aligned} \tag{2.13}$$

When a classical, massive body in free space with a magnetic dipole moment experiences a torque from a suddenly applied, uniform external field as described by eq. 2.10, the body, if free to move, will rigidly rotate towards allignment with the applied field, and in the presence of damping will settle into alignment with the field. This direct rotation is simply described by classical mechanics:

$$\begin{aligned} \vec{\Gamma} &= \frac{d\vec{L}}{dt} \\ \Gamma &= \mathcal{I} \frac{d^2\theta}{dt} \end{aligned}$$

Where $\mathcal{I}$ is the rotational inertia, and L is its angular momentum–both defined by an axis running through the center of mass in the direction of $\hat{\Gamma}$.

However, in magnetic systems relevant to MRAM technology, the magnets are mechanically fixed, and the behavior is more complicated. First, one can gain intuition from an example from classical physics, a unformly charged sphere spinning with angular velocity ω, charge Q, radius R, and mass m_s. By evaluating the vector potential $\vec{A}(\vec{x})$ via Eq. 2.2, one can find that the exact solution of the field outside the body is equal to that of an ideal

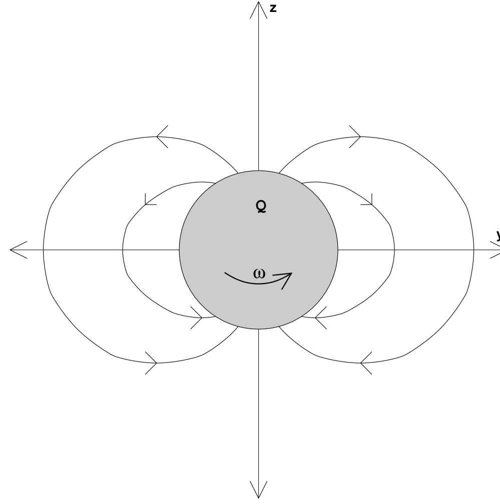


Figure 2.4: Example from classical physics: a unformly charged sphere spinning with angular velocity ω, charge Q, and mass m_s.

dipole at the origin: [4]

$$\begin{aligned}\vec{m} &= \frac{Q}{2m_s c}\,\omega\frac{2}{5}MR^2 \\ &= \gamma L\end{aligned}$$

Where γ gives the ratio of magnetic moment to angular momentum; it is called the gyromagnetic ratio. This value of $\gamma = Q/(2mc)$ holds for a variety of systems like that of a point charge in a circular orbit. This example sets the basic intuition that the magnetic dipole moment can be viewed as a proxy for the angular momentum of an electronic system.

If one had a charged sphere of this sort spinning in free space and a magnetic field was suddenly applied off axis, the dipole would not "directly" rotate towards alignment with the field. Instead the mass would "wobble" around the equilibrium axis set by the field because it's initial angular momentum is non-zero and misaligned with the axis of rotation defined

[4] In [19, p. 236] the vector potential for a charged spinning spherical shell is directly evaluated with Eq. 2.2 and shows that the field outside is the body is precisely the dipole field. The same result holds for a sphere because it can be contstructed out of a summation of concentric spherical shells. More generally, the dipole moment of an arbitrary rotationally symmetric body can be shown to have the same value of γ by building it out of rotating rings that correspond to current loops like that of Eq. 2.8; although, the solution may not be exactly the dipole field, for it may also contain higher order terms in $\frac{1}{|\vec{x}|}$.

by the applied torque.

In a similar manner, the electron has a magnetic dipole moment proportional to it's intrinsic spin angular momentum, with $\gamma = -|e|/mc$ (twice that of what is expected from classical mechanics) and the quantized angular momentum of $\pm\hbar/2$.[21] The magnetic moment of the electron must be treated with quantum mechanics. Its state can be summarized as a linear combination of two basis states along a chosen axis ($\hat{z}$ for example): a "spin up" state with a conventional (in the sense described by Eqs. 2.1 and 2.12) dipole moment with amplitude $-\mu_B$ along $\hat{z}$ and a "spin down" state with a moment of amplitude μ_B along $\hat{z}$. The value of μ_B is $|e|\hbar/2mc$. [5] This can be described by a column vector of two complex coefficients (also known as the two component spinor $|\Psi>$):

$$|\Psi> = \begin{bmatrix} c_{+z} \\ c_{-z} \end{bmatrix} \tag{2.14}$$

where the first entry gives a weighting for the spin up state and the second entry gives a weighting for the spin down state.

If the dipole moment (or equivalently the angular momentum) is measured along $\hat{z}$, [6] it will behave like the conventional dipole corresponding to spin up with probability $c^*_{+z}c_{+z} = |c_{+z}|^2$ and similarly for spin down with probability $c^*_{-z}c_{+z} = |c_{-z}|^2$. Based on this definition, the expectation of the dipole moment along $\hat{z}$ can be constructed as:

$$< \mu_z > \;=\; -\mu_B \begin{bmatrix} c^*_{+z} & c^*_{-z} \end{bmatrix} \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} c_{+z} \\ c_{-z} \end{bmatrix} \tag{2.15}$$

The inner matrix represents the operation of measuring angular momentum (or dipole mo-

[5]Note, the angular momentum and magentic moment of the electron are in *opposite* directions because the electron has negative charge.

[6]One way to "measure" the dipole moment is to pass it through a nonuniform magnetic field. The resulting force as given by Eq. 2.13 will deflect the two spin states in opposite directions. The Stern-Gerlach experiment of 1927 performed this kind of measurement on atoms of silver, whose magnetic moment and angular momentum is due to a single unpaired electron. Furthermore, "sequential" Stern-Gerlach experiments along orthogonal axes of measurement allow one to deduce the matrix representations of electron spin in this section.[21, pp. 1-10]

ment to within a proportionality factor) along $\hat{z}$. It is denoted as σ_z.

What if the angular momentum of an electron described by a column vector of basis states along $\hat{z}$ is measured along a different axis (for example $\hat{x}$)? The outcome of this experiment is given by the inner matrix in the following equation. It is denoted as σ_x.

$$< \mu_x > \; = \; -\mu_B \begin{bmatrix} c^*_{+z} & c^*_{-z} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} c_{+z} \\ c_{-z} \end{bmatrix} \tag{2.16}$$

$$< \mu_x > \; = \; -\mu_B \begin{bmatrix} c^*_{+z} & c^*_{-z} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} c_{+z} \\ c_{-z} \end{bmatrix} \tag{2.17}$$

The factorization of σ_x in Eq. 2.17 shows that it has the same eigenvalues (which correspond to measurable values of angular momentum) as σ_z, and that the matrices of eigenvectors simply perform the following change of basis:

$$\begin{bmatrix} c_{+x} \\ c_{-x} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} c_{+z} \\ c_{-z} \end{bmatrix}$$

The same interpretation of Eq. 2.14 applies to the left hand side of the above equation. Namely, if the dipole moment is measured along $\hat{x}$, it will behave like a conventional dipole $-\mu_B\hat{x}$ with probability $c^*_{+x}c_{+x} = |c_{+x}|^2$ and like a conventional dipole $+\mu_B\hat{x}$ with probability $c^*_{-x}c_{-x} = |c_{-x}|^2$.

A similar development will reveal the same properties of the matrix that represents measurement of angular momentum along $\hat{y}$:

$$\sigma_y \; = \; \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix}$$

$$< \mu_y > \; = \; -\mu_B \begin{bmatrix} c^*_{+z} & c^*_{-z} \end{bmatrix} \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix} \begin{bmatrix} c_{+z} \\ c_{-z} \end{bmatrix} \tag{2.18}$$

with $j = \sqrt{-1}$.

Finally, one can use the matrices $\{\sigma_x, \sigma_y, \sigma_z\}$ (the so-called Pauli matrices) to construct two useful mathematical representations:

1. A representation of the operator for measuring angular momentum along an arbitrary direction given by $\hat{n} = n_x\hat{x} + n_y\hat{y} + n_z\hat{z}$:

$$\sigma_n = n_x\sigma_x + n_y\sigma_y + n_z\sigma_z \tag{2.19}$$

2. A three-component cartesian coordinate representation of the electron spin:

$$< \vec{\mu} > = < \mu_x > \hat{x} + < \mu_y > \hat{y} + < \mu_z > \hat{z} \tag{2.20}$$

   If one defines

$$\hat{n} = -\frac{< \vec{\mu} >}{|< \vec{\mu} >|}$$

   with $< \vec{\mu} >$ calculated from Eq. 2.20, and then applies the operator in Eq. 2.19 to calculate $< \mu_n >$, the result will always be $-\mu_B$. Hence, Eq. 2.20 has the precise interpretation as the vector that gives the direction along which the spin magnetic moment is purely in the eigenstate corresponding to a value of $+\mu_B$.

Although the representation of the electron's magnetic moment in Eq. 2.20 is equivalent to the two component spinor in Eq. 2.14, it is not useful for quantum mechanics calculations. However, it will be useful later in analyzing the interaction of a spin polarized current with a macroscopic magnetic moment.

The change of basis property in the factorization of the σ matrices has shown that the spinor can be equivalently represented along any basis direction. By convention, the spinor is expressed in terms of basis states along $\hat{z}$. It is particularly useful to choose $\hat{z}$ such that it is in the direction of the local, externally applied magnetic field experienced by the dipole, because the time evolution is mathematically cleaner in terms of the spin up and spin down states along the axis that shares the direction of the local magnetic field. This time evolution is given by the schrodinger equation:

$$i\hbar\frac{\partial}{\partial t}|\Psi> = \mathcal{H}|\Psi> \tag{2.21}$$

Where $\mathcal{H}$ is the operator for measuring the energy of the electron. Choosing the standard basis, and recognizing that Eq. 2.12 shows that each basis state in angular momentum also has a single, unambiguous value for energy allows one to immediately write $\mathcal{H} = \mu_B H \sigma_z$:

$$i\hbar \frac{\partial}{\partial t} \begin{bmatrix} c_{+z} \\ c_{-z} \end{bmatrix} = \mu_B H \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} c_{+z} \\ c_{-z} \end{bmatrix}$$

This would not have been the case if the spinor was expressed along $\hat{x}$ and the field still applied along $\hat{z}$. The apt choice of $\hat{z}$ has resulted in a diagonal matrix, yielding two uncoupled first order differential equations which are solved to give: [21, p. 76]

$$|\Psi(t)> = \begin{bmatrix} c_{+z} \exp\left(\frac{-i\omega t}{2}\right) \\ c_{-z} \exp\left(\frac{+i\omega t}{2}\right) \end{bmatrix} \tag{2.22}$$

where $\omega = 2\mu_B H/\hbar = |e|H/m_e c$. It is insightful to construct $< \vec{\mu} >$ by Eq. 2.20 from this solution:

$$< \vec{\mu}(t) > = < \mu_\perp > \cos(\omega t + \Delta\phi)\hat{x} + < \mu_\perp > \sin(\omega t + \Delta\phi)\hat{y} + < \mu_{z0} > \hat{z} \tag{2.23}$$

where $< \mu_\perp > = -\mu_B 2|c_{+z}c_{-z}|$, $\Delta\phi = \measuredangle c_{-z} - \measuredangle c_{+z}$, and $< \mu_{z0} > = -\mu_B \left[|c_{+z}^2| - |c_{-z}^2|\right]$. Eq. 2.23 says that the $\hat{x}$ and $\hat{y}$ components of the vector spin [7] oscillate out of phase as the $\hat{z}$ component is fixed. This is exactly the precession that was anticipated from the intuition building example of a classical charged rotating body in Fig. 2.4. One must note, however, that the electron is a point particle and has no internal structure to allow the observation of a physical rotation. Yet, the expectation of its dipole moment rotates.

The discussion of real magnetic materials hereon ultimately rests on the behavior of these basic dipoles—both quantum mechanical microscopic dipoles and classical macroscopic dipoles.

[7] In this text, "vector spin" means "the direction along which the spin is purely in the +1 eigenstate."

2.2 Properties of Nanomagnets

Magnetism in macroscopic media stems from the cumulative effect of its constituitive dipoles. This phenomenon is usefully described by the magnetization vector field $\vec{M}(\vec{x})$ that gives the magnetic moment of an infinitesimal volume dV at $\vec{x}$ equal to $\vec{M}dV$. The way in which these dipoles interact with each other and externally applied fields to produce a resulting $\vec{M}$ fall into four broad categories: [23, pp. 417-484]

Diamagnetism A purely diamagnetic substance has no net magnetic moment in the absence of magnetic field. When a magnetic field is applied, the diamagnetic substance generates an *opposing* magnetic moment due to the distortion of the electron clouds within the atoms. This response of electrons by their motion is a microscopic analog of Lenz's Law—in which a current is generated in a loop to oppose the change in its enclosed magnetic flux.

Paramagnetism Paramagnetism in media results from electrons preferentially populating a lower magnetic field dependent energy state. This will result in an excess of one spin state over the other when an external magnetic field is applied. The magnetic moments of the excess unpaired spin states sum to produce a $\vec{M}$ that aligns with the applied field.

Ferromagnetism A ferromagnetic material exhibits local regions of uniform magentization $\vec{M}$ in the absense of an externally applied field. Ferromagnetism originates from the energetic favorability of aligned electron spins due to the greater tendency of like spin states to be spatially seperated. This spatial seperation minimizes energy from electrostatic repulstion. Beyond a certain tempurature T_C, ferromagnetic materials behave like paramagnets. Below this temperature, the so-called exchange interaction energy dominates the thermal disruption and the magnetization approaches a uniform saturation magnetization M_s (a material dependent parameter). Finally, the local regions of uniform magnetization, called domains, tend to be randomly oriented on a longer distance scale to minimize the energy of their dipole field interactions. For very small ferromagnets, the exchange energy dominates the conventional dipole field interaction between domains and a uniform magnetization results throughout.

Antiferromagnetism Antiferromagnetic materials originate from ferromagnetic ordering in a highly symmetric way made possible by the lattice structure. However, different subgroups of ordering tend to cancel each other and produce no net magnetic moment.

The free layer in the MTJ is a ferromagnet. Furthermore, it small enough to be approximated as a single domain with all the magnetic moments perfectly aligned. That is to

say the dipole moment of an infinitesimally small volume dV is equal to $\vec{M}_s dV$ and is the same for any location within the volume. $\vec{M}_s$ is assumed to be constant in magnitude and uniform for this monodomain approximation. Therefore, the net dipole moment of the body $\vec{m} = \vec{M}_s V$ will also be constant in magnitude.

2.2.1 The Fields and Energy of a Nanomagnet

The shape of the relevant nanomagnets in MRAM can be approximated by ellipsoids. An ellipsoid is a volume enclosed by the surface described by the loci of points satisfying:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{c^2}{a^2} = 1 \tag{2.24}$$

For the nanomagnets of interest to MRAM, the shape is an oblate ellipsoid in which the volume is "squashed" in the x-direction, and has an aspect ratio of 2:1 to 4:1 in the z-y plane with the longest axis along $\hat{z}$. Typical values of $\{a, b, c\}$ relevant to the magnets of spin transfer MRAM are $\{3, 80, 240\}[nm]$. [14] A cross-section of this oblate spheriod in the z-y plane is shown in Fig. 2.5.

For a uniformly magnetized material, the relevant maxwell equations for $\vec{H}$ reduce to:

$$\begin{aligned} \vec{\nabla} \times \vec{H} &= 0 \\ \vec{\nabla} \cdot \vec{H} &= -\vec{\nabla} \cdot \vec{M} \end{aligned}$$

In the even simpler case of a uniformly magnetized object, the second equation is zero both inside and outside the body. However, the singularity of $\vec{\nabla} \cdot \vec{M}$ imposes the following boundary conditions accross the surface of the body: [19, p. 273]

$$\begin{aligned} \left(\vec{H}_{out} - \vec{H}_{in}\right) \cdot \hat{n} &= -\left(\vec{M}_{out} - \vec{M}_{in}\right) \cdot \hat{n} \\ \left(\vec{H}_{out} - \vec{H}_{in}\right) \times \hat{n} &= 0 \end{aligned}$$

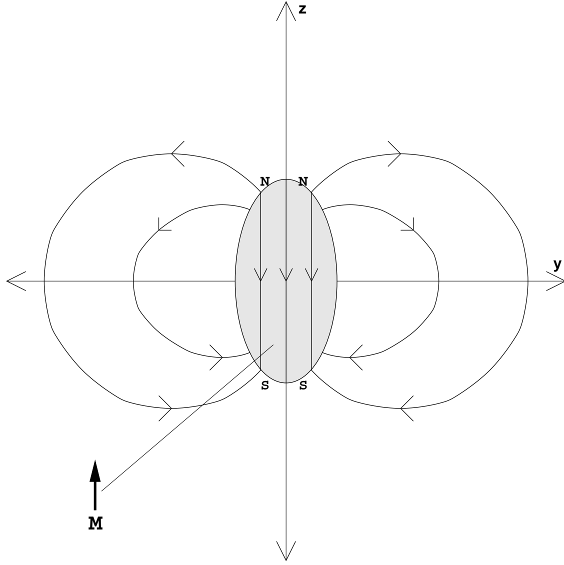


Figure 2.5: A uniformly magnetized ellipsoid with magnetic moment along the "easy" axis with the resulting $\vec{H}$ field

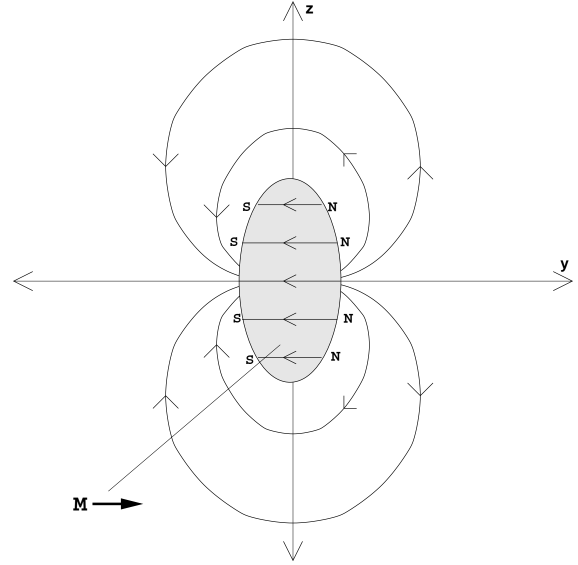


Figure 2.6: A uniformaly magnetized ellipsoid with magnetic moment along the "hard" axis with the resulting $\vec{H}$ field

where $\hat{n}$ is the local normal vector to the surface. This equation for $\vec{H}$ shows that $-\vec{\nabla}\cdot\vec{M}$ is acting as an effective magnetic charge [8] at the surface of the body that produces a "backfield" against the magnetized material (this is indicated by the "N" and "S" in Figs. 2.5 and 2.6). The solution to the above equation lends itself to electrostatics techniques and is given by: [24]

$$\vec{H}_{in} = -4\pi \left(D_a M_x \hat{x} + D_b M_y \hat{y} + D_c M_z \hat{z}\right) \tag{2.25}$$

in the interior of the magnetized body. Thus, the backfield follows $\vec{M}$ around, but more strongly in some directions. The D_ν ($\nu \in \{a, b, c\}$) demagnetization coefficients are given by:

$$D_\nu = \frac{abc}{2}\int_0^\infty \frac{ds}{(\nu^2+s)\sqrt{(a^2+s^2)(b^2+s^2)(c^2+s^2)}} \tag{2.26}$$

What's important is that $D_a + D_b + D_c = 1$ and that D_a is largest since the prolate ellipsoid is most squashed along the corresponding $\hat{x}$ direction. Outside the body, the field turns out to be that of a pure dipole with moment $\vec{m} = \vec{M}V$, where V is the volume of the body.

[8] Compare $\vec{\nabla}\cdot\vec{H} = -\vec{\nabla}\cdot\vec{M}$ to $\vec{\nabla}\cdot\vec{E} = 4\pi\rho$

The simple solutions of the field both inside and outside a uniformly magnetized ellipsoid, make this geometry useful for analytical calculations. Furthermore, it approximates actual thin film nanomagnets in magnetic tunnel junctions reasonably well. In the z-y plane the nanomagnets tend to have an elliptical outline due to the photolithographic rounding of the corners. In the vertical direction, the films are very thin so the deviation from the ellipsoidal curvature is mostly significant at the very edges. This approximation by ellipsoids has been advocated several decades ago by E. C. Stoner, "the general ellipsoidal form covers, as an approximation, almost the whole variety of possible shapes for the physical particles, or segregates, which are likely to be of physical interest." [25]

Eq. 2.12 suggests that an orientation of $\vec{M}$ with a weaker backfield from Eq. 2.25, has a lower energy configuration. For the model prolate ellipsoid, the lowest energy directions are $\pm\hat{z}$, and the lowest energy plane is the z-y plane. This lowest energy configuration of $\vec{M}$ along $+\hat{z}$ is depicted in Fig. 2.5. Infact, the $\vec{M} = \pm M\hat{z}$ correspond to the two stable energy minima in the the magnet's configuration.

The energy contribution of the demagnetizing field for the uniformly magnetized ellipsoid is calculated with: [25]

$$U_m = -\int_V \frac{1}{2}\vec{M} \cdot \vec{H}_d dV \tag{2.27}$$

$$U_m = -\frac{1}{2}\vec{M}V \cdot \vec{H}_d \tag{2.28}$$

Comparing the abvove equation with Eq. 2.12, one can see a discrepancy in the prefactor of 1/2. This is so because Eq. 2.12 gives the energy of a dipole in a uniform, *external* field derived from the conservative torque in Eq. 2.10; whereas, Eq. 2.28 gives the energy related to an assembly of dipoles $\vec{M}dV$ which reside in a self-created demagnetization field. Intuitively, one can anticipate the factor of 1/2 by recognizing that it takes no work to bring the first dipole in from infinity but it takes a full $\vec{M}dV \cdot \vec{H}_d$ amount of work to bring in the last dipole of the magnet from infinity.

The smallest energy barrier ΔU between $+\hat{z}$ and $-\hat{z}$ must occur at $\vec{M} = \pm M\hat{y}$ because

$D_a > D_b > D_c$. That is to say trading off alignment with $\hat{y}$ to increase the component along $\hat{x}$ will always make the energy of the dipole moment higher. The situation of $\vec{M} = +M\hat{y}$ is depicted in Fig. 2.6, and for this reason $\hat{y}$ is known as the "hard" axis. Recalling that the magnetization $\vec{M}$ is fixed in magnitude (because this is a saturated monodomain ferromagnet), U_m can be rewritten as:

$$\begin{aligned} U_m &= 2\pi\left(D_a M_x^2 + D_b M_y^2 + D_c M_z^2\right) V \\ M^2 &= M_z^2 + M_x^2 + M_y^2 \\ \Rightarrow U_m &= 2\pi\left((D_a - D_c)M_x^2 + (D_b - D_c)M_y^2\right) V + \text{const.} \end{aligned} \tag{2.29}$$

Using the above equation to evaluate the *difference* in energy between $\vec{M} = M\hat{y}$ and $\vec{M} = M\hat{z}$ gives:

$$\Delta U = 2\pi M^2 V (D_b - D_c) = \frac{1}{2} M V H_k = \frac{1}{2} m H_k \tag{2.30}$$

which corresponds to an energy barrier in magnetic field units: $H_k = 4\pi M(D_b - D_c)$.

The expression for U_m was determined entirely by the demagnetization field, and is known as the shape anisotropy energy. There are other sources of anisotropy from material properties based on the lattice structure of the ferromagnetic material (known as intrinsic anisotropy). These other sources of anisotropy can be treated by adding terms to U_m that are polynomials in m_x^2, m_y^2 and m_z^2. [26] [9] In practice, monodomain models for MRAM nanomagnets assume a form of U_m that is even in m_x and m_y (only two of the three components are needed since the third is given by $m^2 = m_x^2 + m_y^2 + m_z^2$), and the energy as a function of orientation is deduced by finding the appropriate constants $\{C_{j,k}\}$ such that

$$U_m = \sum_{j,k} C_{j,k} (m_x^2)^j (m_y^2)^k$$

Not surprisingly, the ellipsoid with pure shape anisotropy has only m_x^2 and m_y^2 terms.

[9] U_m is written in polynomials of m_i^2 and not simply m_i because the ellipsoid geometry must produce an energy that is an even function of the coordinaes m_i. This makes an additional assumption that planes and axes of intrinsic anisotropy do not break this symmetry.

In the presence of an applied field $\vec{H}_{ext}$, another term is added to the magnet's energy:

$$U_m = -\frac{1}{2}\vec{M}V \cdot \vec{H}_d - \vec{M}V \cdot \vec{H}_{ext} \tag{2.31}$$

As expected from Eq. 2.12, this external field does not have the prefactor of 1/2. Stoner and Wolfarth [25] have described how the magnetic moment will settle to a direction corresponding to an energy minimum in U_m, which in turn can be varied by the applied field. Suppose the field is applied purely in the z-y plane such that$\vec{H}_{ext} = H_{hard}\hat{y} + H_{easy}\hat{z}$. Then, re-writing Eq. 2.31 as a function of angular coordinates (θ, ϕ) gives: [10]

$$\begin{aligned}
U_m &= 2\pi\left((D_a - D_c)M^2\sin^2\theta\cos^2\phi + (D_b - D_c)M^2\sin^2\theta\sin^2\phi\right)V \\
&\quad - MV\sin\theta\sin\phi H_{hard} - MV\cos\theta H_{easy} \\
U_m &= \frac{1}{2}MVH_k\left(\frac{D_a - D_c}{D_b - D_c}\sin^2\theta\cos^2\phi + \sin^2\theta\sin^2\phi\right) - MV\sin\theta\sin\phi H_{hard} - MV\cos\theta H_{easy} \\
U_m &= \frac{1}{2}MVH_k\left(\left[\frac{D_a - D_c}{D_b - D_c} - 1\right]\sin^2\theta\cos^2\phi + \sin^2\theta\right) - MV\sin\theta\sin\phi H_{hard} - MV\cos\theta H_{easy} \\
U_m &= K\left(h_p\sin^2\theta\cos^2\phi + \sin^2\theta\right) - 2K\sin\theta\sin\phi h_{hard} - 2K\cos\theta h_{easy} \qquad (2.32)
\end{aligned}$$

where the units have been normalized to the energy barrier $K = \Delta U = \frac{1}{2}mH_k$, and the external field has been normalized to H_k as in [27] ($h_{easy} = H_{easy}/H_k$).

Taking the first and second derivatives of Eq. 2.32 allows one to find the locations of energy minima. There are are two distintct behaviors depending on $\vec{H}_{ext}$. For smaller values of $\vec{H}_{ext}$, two stable minima exist with an energy barrier between them. For larger values, only one stable minimum exists. Therefore, the magnet can be programmed into one stable minimum if an external field is applied beyond a certain threshold. Upon removal of the

[10] U_m is mapped to angular coordinates as follows:

$$\begin{aligned}
M_x &\rightarrow M\sin\theta\cos\phi \\
M_y &\rightarrow M\sin\theta\sin\phi \\
M_z &\rightarrow M\cos\theta
\end{aligned}$$

superthreshold field, the magnet will deterministically settle to one of the two zero-field oreintations ($\theta = 0$ or $\theta = \pi$ for the model ellipsode of Figs. 2.5 and 2.6). The solution to the boundary between having two local minima in $U_m(\theta, \phi)$ and just one local minima with an inflection point is: [26, p. 38] [28, p. 141] [11]

$$H_{hard}^{2/3} + H_{easy}^{2/3} = H_k^{2/3} \tag{2.33}$$

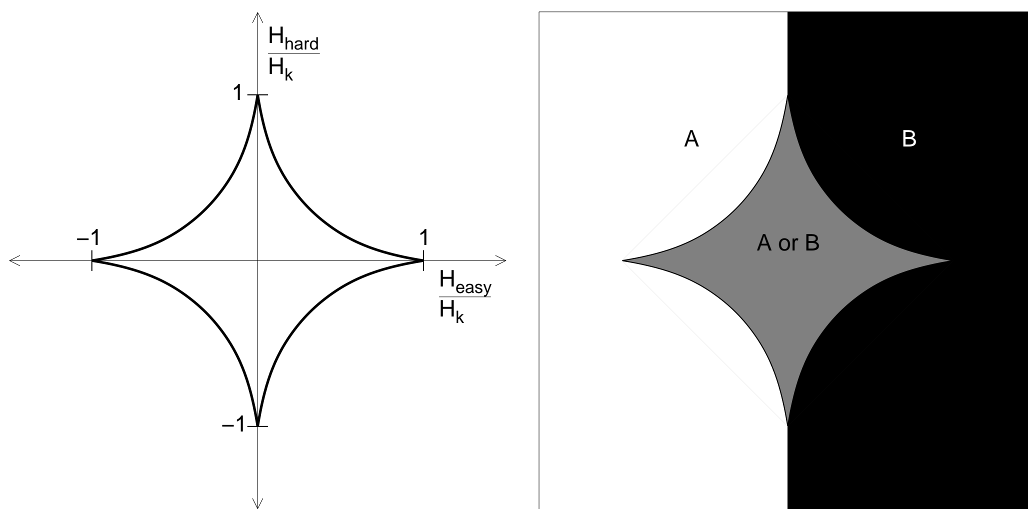
Eq. 2.33 is known as the Stoner-Wolfarth astroid. It gives the two-dimensional hysteresis in magnetic field of a monodomain nanomagnet. Shown in Fig. 2.7(a) is a plot of Eq. 2.33 and shown in Fig. 2.7(b) is a qualitative depiction of hysteretic and non-hysteretic regions. If one traces a path in H_{easy}-H_{hard} space into the black region and returns to the gray region, the magnet will be in state B. Similarly, the magnet can be programmed into state A by tracing a path in H_{easy}-H_{hard} space into the white region. State B and A represent $\theta = 0$ and $\theta = \pi$ respectively. One could have readily anticipated the threshold of H_k along the easy axis from the derivation of the energy barrier in Eq. 2.30.

Finally, in the absence of applied field there is a thermal background energy in the form of spin waves (coherent oscillations of microscopic dipole deviation from the macroscopic $\vec{M}/M$ direction) and phonons that can cause the magnet to switch between states A and B, overcoming the energy barrier without the aid of an applied field. Encapsulating these thermal effects by adding a random field term to the dynamical equation for the magnetic moment allows the calculation of a poisson arrival rate of the unwanted thermally-induced switching: [29]

$$\begin{aligned} \lambda &= f_A \exp\left(-\frac{\Delta U}{kT}\right) \\ P_{sw} &= 1 - \exp\left(-\lambda t\right) \end{aligned} \tag{2.34}$$

[11] In both these sources, the problem was solved for uniaxial anisotropy with no easy plane anisotropy (e.g. a prolate spheroid with only shape anisotropy). However, one can argue the same result holds with $\hat{y}$ as the hard axis in the oblate ellipsoid (the shape relevant to MRAM nanomagnets), because the easy plane device's magnetic moment will rest in the z-y plane. Furthermore the application of the field will reduce the energy in the $\pm\hat{y}$ direction depending on the sign of H_{hard} and not change the location of the new minimum from that of the uniaxial case.

where f_A is the attempt frequency and can be approximated as $1GHz$ for MRAM applications. [30] To meet retention error rate equirements for a memory product, a barrier of $\Delta U \approx 60kT - 70kT$ is required. Eq. 2.34 and the expression for ΔU in Eq. 2.30 reveal the fundamental scaling challenge of conventional field-switching MRAM: the energy barrier $\Delta U = 2\pi M^2 V(D_b - D_c)$ scales directly with the cell area ($V = (\text{area}) \cdot (\text{thickness})$) with all other parameters held constant. To compensate for the decreased amount of magnetic moment, novel materials processing has to be developed to construct larger magnetization M, or more likely the aspect ratio has to be increased to boost $D_b - D_c$. Yet, either of these techniques will also increase the field switching threshold $H_k = 4\pi M(D_b - D_c)$, which in turn translates to a larger current requirement in smaller semiconductor technology nodes. This problem remains for other types of field switching schemes such as toggle switching because H_k indicates typical field strengths needed to externally control the nanomagnet.



(a) The two dimensional boundary between bistable region and monostable region in H_{easy}-H_{hard} space.

(b) A qualitative depiction of how the path through the $H_{easy} - H_{hard}$ plane determines the state of the nanomagnet at the origin.

Figure 2.7: The Stoner-Wolfarth Astroid for a monodomain magnet

2.3 Magnetization Dynamics

The macroscopic magnetic moment of a ferromagnet is a direct measure of angular momentum with a proportionality factor $\gamma = -|e|/mc$, for it is simply the vector sum of the excess electron moments in the majority spin state. Therefore, the definition of torque as the derivative of angular mometnum is applied to explain magnetization dynamics:

$$\frac{d\vec{L}}{dt} = \vec{\Gamma}$$
$$\frac{1}{\gamma}\frac{d\vec{m}}{dt} = \vec{\Gamma}$$

From Eq. 2.10, the torque from an externally applied field is simply $\vec{m} \times \vec{H}$. Supposing for a moment that the macroscopic magnetic moment $\vec{m}(t=0) = m_{0x}\hat{x} + m_{0y}\hat{y} + m_{0z}\hat{z}$ experiences only the torque from an externally applied field $\vec{H} = H\hat{z}$, the solution would be:

$$\vec{m}(t) = m_\perp \cos(\omega t + \Delta\phi)\hat{x} + m_\perp \sin(\omega t + \Delta\phi)\hat{y} + m_{z0}\hat{z}$$

where $m_\perp \cos(\Delta\phi) = m_{0x}$ and $m_\perp \sin(\Delta\phi) = m_{0y}$. This is in precise agreement with Eq. 2.23 because the ferromagnet's constituent electron dipole moments are coherently precessing. However, the demagnetization field and other anisotropy energy terms produce an additional, effective field which can be deduced from the angular gradient of U_m: [27]

$$\vec{H}_U = \frac{1}{m}\vec{\nabla}U(\theta,\phi) = \frac{1}{m}\left[\frac{1}{\sin\theta}\frac{\partial U}{\partial \phi}\hat{\phi} + \frac{\partial U}{\partial \theta}\hat{\theta}\right] \tag{2.35}$$

Finally, an empirical damping term α is added to complete the equation for magnetization dynamics, known as the Landau-Lifshitz-Gilbert (LLG) equation: [31]

$$\frac{d\vec{m}}{dt} = \gamma\vec{\Gamma} - \frac{\alpha}{m}\vec{m} \times \frac{d\vec{m}}{dt} \tag{2.36}$$
$$\frac{d\vec{m}}{dt} = \gamma\vec{m} \times \vec{H} - \frac{\alpha}{m}\vec{m} \times \frac{d\vec{m}}{dt} \tag{2.37}$$

To conceptualize the damping process, suppose $\alpha \ll 1$ so that $d\vec{m}/dt$ is basically in the direction of $\vec{m} \times H$. Therefore, the damping term will produce a vector that is perpendicular to both $\vec{m}$ and $\vec{m} \times \vec{H}$ which means the damping produces a tendency for the moment to fall into alignment with $\vec{H}$.

2.4 The MTJ structure

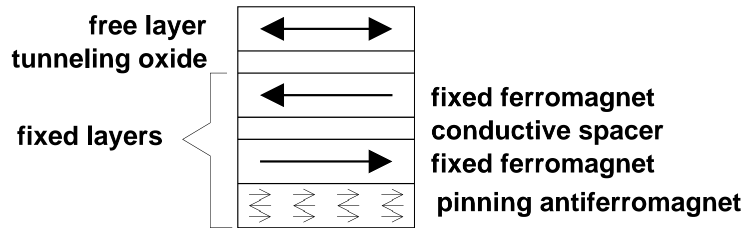


Figure 2.8: A schematic diagram of the stack of materials (Ferromagnet | Oxide | Ferromagnet | Spacer | Ferromagnet | AntiFerromagnet) that constitutes a Magnetic Tunnel Junction.

Going from top to bottom, one can understand the purpose of each layer: [7], [32]

1. The free layer stores the bit. It has two possible orientations (indicated by the double arrow): parallel or antiparallel to the fixed ferromagnet magnet immediately below it.

2. The tunneling oxide amplifies the signal in resistance that can be tuned in a wide range from 100Ω to $10k\Omega$. Without the tunneling oxide, the ferromagnetic materials would produce $1m\Omega$ to 1Ω of resistance because they are conductors.

3. The second ferromagnet is responsible for the magnetization dependent tunneling probability accross the oxide, which translates into two different resistance values when a voltage is applied accross the MTJ.

4. The third ferromagnet helps fix the second ferromagnet by coupling to it through dipole field interactions. Furthermore, this structure can be engineered to produce no net bias magnetic field in the top-most free layer. This is important for ensuring the thermal stability of the free layer and symmetric write characteristics for 1 and 0.

5. The bottom antiferromagnetic layer helps pin the bottom ferromagnet.

A key figure of merit for the read behavior of an MTJ is its magnetoresistnace ratio:

$$MR = \frac{R1 - R0}{R0} \tag{2.38}$$

where $R0$ is the lower resistance of the parallel state.

2.5 Spin Angular Momentum Transfer

Spin Angular Momentum transfer is a novel mechanism of switching the free layer in an MTJ without the application of external fields. It is based on the fact that the magnetization of a ferromagnet stems from a preferential population of spin states aligned with the macroscopic magentization. Therefore, passing a current between two ferromagnets suggests that the spin polarized currents will bring their magnetic moment with them and alter the magnetization of the other layer.

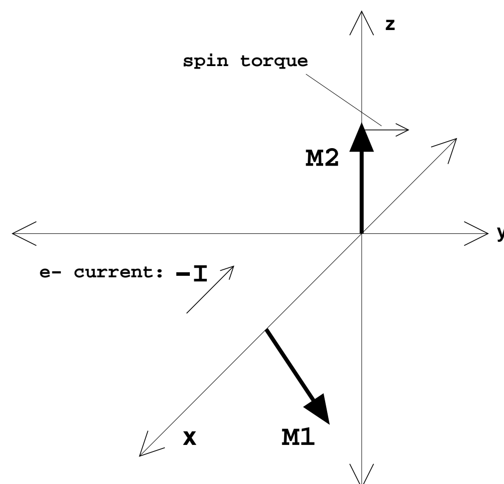


Figure 2.9: Representation of spin torque due to current between two ferromagnets

The spin torque term is readily attained from arguments based on prior developments in this chapter. Fig. 2.9 describes the coordinate setup for the calculation of the spin transfer torque term. In the figure, current is flowing from ferromagnet 1 to ferromagnet 2. Ferromagnet 1 can represent the upper ferromagnet in the fixed layer of the MTJ depicted in

Fig. 2.8 and the destination ferromagnet 2 would be the top-most free layer. In order to produce these conditions in an MTJ, a positive voltage at the top of the MTJ would be applied.

The first people to predict this effect, Slonczewski [11] and Berger [33], have described how the change in the macroscopic magnetic moment $\Delta\vec{m}_2$ of the free magnet, on average, equals the transverse component of one electron's expected spin magnetic moment $< \vec{\mu} >$ (c.f. Eq. 2.20). This is a consequence of the tendency of the spin to align with the macroscopic moment through the intra-atomic exchange interaction. Basically, this treats $< \vec{\mu} >$ as a classical vector although individual realizations of $\vec{\mu}$ will be $\pm\mu_B$ on specific directions of interaction. This treatment is justified because even the fastest spin transfer switching events reported have involved 10^6 to 10^8 electrons. [15], [12]

In order to develop an expression for $d\vec{m}_2/dt$, it is first assumed that every electron in the switching current is transmitted accross the barrier and has $< \vec{\mu} >= \mu_B\hat{n}_1$ parallel to the fixed magnet $\vec{m}_1 = m_1\hat{n}_1$, where $\hat{n}_1$ is their common unit vector. The average contribution of each electron to the change in magnetization is expressed as:

$$\begin{aligned}
\Delta\vec{m}_2 &= (\text{the projection of } < \vec{\mu} > \text{ onto a plane } \textit{normal to } \vec{m}_2) \\
&= < \vec{\mu} > \; - (\text{the projection of } < \vec{\mu} > \text{ onto } \vec{m}_2) \\
&= \mu_B \; [\hat{n}_1 - (\hat{n}_1 \cdot \hat{n}_2)\hat{n}_2] \qquad (2.39) \\
&= \mu_B \; \hat{n}_2 \times (\hat{n}_1 \times \hat{n}_2) \qquad (2.40)
\end{aligned}$$

A vector identity was applied going from Eq. 2.39 to Eq. 2.40 in anticipation of combining this expression with other torque terms in the LLG equation. Intuitively, this vector identity produces the correct magnitude with a $\mu_B \sin\theta$ term in the inner cross product, and then produces the correct direction with the outer cross product, by directing $\Delta\vec{m}_2$ such that it brings $\vec{m}_2$ into alignment with $< \vec{\mu} >$. Now $d\vec{m}_2/dt$ is simply generated from multiplication

of Eq. 2.40 by the switching current, converted from C/s to $\#e^-/s$ to μ_B/s:

$$\frac{d\vec{m}_2}{dt} = \eta \frac{I}{|e|} \mu_B \; \hat{n}_2 \times (\hat{n}_1 \times \hat{n}_2) \tag{2.41}$$

Where $0 < \eta < 1$, quantifies the deviation from the idealization that the electrical current is composed of electrons with spins exclusively aligned to $+\hat{n}_1$ and not $-\hat{n}_1$, in addition to the fact that both $+\hat{n}_1$ and $-\hat{n}_1$ electrons will have finite probabilities of transmission *and* reflection. Since both transmitted and reflected electrons impart the transverse component of $< \vec{\mu} >$, reflection of $+\hat{n}_1$ electrons boosts the effectiveness of the current I, which only counts *transmitted* electrons. Correspondingly, reflection of $-\hat{n}_1$ electrons would further deteriorate the effectiveness of the current. There are additional issues related to multiple reflections, relaxation of preferential spin states in the barrier, and interface scattering; so far, these deviations have been successfully lumped in η for the purpose of analyzing experimental results.

Now letting ferromagnet 2 represent the model uniformly magnetized oblate ellipsoid, yields the following dynamical equation upon incorporating Eq. 2.41 (the subscript 2 from m_2, the magnetic moment of the free layer, is dropped and $\hat{n}_2$ is replaced with $\hat{m}$):

$$\frac{1}{\gamma}\frac{d\vec{m}}{dt} \;=\; \vec{m} \times (\vec{H}_{ext} + \vec{H}_{eff}) + \eta \frac{I}{|e|}\frac{\mu_B}{\gamma} \; \hat{m} \times (\hat{n}_1 \times \hat{m}) - \frac{\alpha}{\gamma m}\vec{m} \times \frac{d\vec{m}}{dt} \tag{2.42}$$

where $\vec{H}_{ext}$ is the externally applied magnetic field (if any) and

$$\vec{H}_{eff} = -\frac{1}{m}\vec{\nabla}\left[K\left(h_p \sin^2\theta \cos^2\phi + \sin^2\theta\right)\right]$$

as defined by Eqs. 2.32 and 2.35. Recall that $\gamma = -|e|/(mc) = -\mu_B/(\frac{\hbar}{2}) < 0$. Furthermore, the sign of I is defined such that $I > 0$ produces a spin torque that aligns $\hat{m}$ with $+\hat{n}_1$; whereas, $I < 0$ produces a spin torque that aligns $\hat{m}$ with $-\hat{n}_1$. A current of opposite polarity produces a spin torque in the opposite direction because electrons traveling from the free layer to the fixed layer whose magnetic moments are in the opposite direction of the

destination fixed layer are more likely to be *reflected*. These reflected electrons impart their angular momentum to the free layer such that the free layer tends to become antiparallel to the fixed layer. The arguments for the torque term in Eq. 2.41 still apply to this case and therefore Eq. 2.42 is still valid.

In [27], a coupled pair of differential equation for the angular coordinates (θ, ϕ) of the magnetic moment is quoted as a result of Eq. 2.42 with an external field $\vec{H}_{ext} = H\hat{z}$. This system of differential equations is then solved for an instability condition in which the angular departure of the magnetic moment from $\hat{z}$ grows without bound. The value of the current is solved at this instability condition and is given by:

$$I_{c0} = \frac{1}{\eta}\left(\frac{2|e|}{\hbar}\right) m\alpha(H + H_k + 2\pi M) \tag{2.43}$$

The full set of vector algebra steps required to derive this equation can be found in Appendix A.2. The $2\pi M$ term reflects an increased amount of switching current due to the presence of easy plane anisotropy. The strong demagnetization field against out of plane (along $\hat{x}$) orientations constrains the magnetic moment's precession.

Furthermore, a linearized differential equation for the switching dynamics can be solved for the time it takes for θ to grow from its initial value of θ_0 (where $\theta_0 \ll 1$ to $\pi/2$, and is shown in the top expression in the following equation:

$$\tau^{-1} = \begin{cases} \frac{\eta(\mu_B/e)}{m\ln(\pi/2\theta_0)}(I - I_{c0}) & I \gg I_{c0} \\ \tau_0^{-1}\exp\left[-\frac{K}{kT}(1-h)^2\left(1 - \frac{I}{I_{c0}}\right)\right] & I \ll I_{c0} \end{cases} \tag{2.44}$$

The bottom expression reflects the fact that the magnetic moment can switch due to spin transfer at values *below* I_{c0} due to thermal agitation. This equation was developed in [12] based on similar arguments that derived Eq. 2.34. It results in an accelerated error rate with a probability of switching in time t:

$$P_{sw} = 1 - \exp\left(-\frac{t}{\tau}\right) \quad I \ll I_{c0}$$

Finally, the equation for switching above I_{c0} still has a term dependent on the initial condition. An estimate for the switching time can be produced by averging over the initial conditions with a probability weighting by the boltzmann factors:

$$P(\theta, \phi) \propto \exp\left(-U(\theta, \phi)\right)$$

with $U(\theta, \phi)$ given in Eq. 2.32. The resulting equation is:

$$\tau^{-1} = \begin{cases} \frac{\eta(\mu_B/e)}{m\pi} \frac{\ln\left(4\pi^2 h_p\right)}{\ln(K/kT)} \sqrt{\frac{1+h}{h_p}} (I - I_{c0}) & I \gg I_{c0} \\ \tau_0^{-1} \exp\left[-\frac{K}{kT}(1-h)^2 \left(1 - \frac{I}{I_{c0}}\right)\right] & I \ll I_{c0} \end{cases} \tag{2.45}$$

The theoretical predictions in the above equations are first estimates at switching behavior. The basic phenomena are (1) a probabilistic switching with an exponential increase in switching time with a linear increase in current when $I < I_{c0}$ and (2) a linear tradeoff between switching time and current when $I > I_{c0}$. The thermal effects obfuscate the ability to observe a clear value for I_{c0}. Furthermore, Eq. 2.45 includes the effect of an externally applied field although the purpose of an STS MRAM is to avoid the need for an externally applied field, because experiments that show a change in switching current threshold as a function of applied field like in [12] provide stronger evidence for the existence of a spin transfer effect over other possible explanations.

Chapter 3

Design of the 4kb Array

During the first phase of this project, a 4kb memory array with a one-transistor one-MTJ cell that supports bidirectional currents through the memory element has been developed. The fabricated array circuitry has been demonstrated on a dummy bitline of resistor cells, and, with the availability of MTJ hardware, the array can be used to explore the possibility of spin-transfer switching in MTJs.

3.1 Overview

Shown in Fig. 3.1 is the overall block diagram for the 4kb array development macro (ADM). This architecture has six notable features:

1. The array size is 64x64 with an equal number of columns (the bitline direction drawn horizontally) and rows (the wordline direction drawn vertically). The bitlines provide a current path to the MTJs and the wordlines simply drive the gates of the cell nFETs.
2. There is an additional dummy bitline of resistance cells to calibrate the support circuitry. This dummy bitline will also be used to demonstrate the circuit techniques employed to sense and bi-directionally write resistive memory elements.
3. A single sense amplifier is used because the experimental goals of this ADM require only one bit to be read at a time. Furthermore, this avoids the issue of offsets across multiple sense amplifiers.

4. There are separate write drivers (voltage source and current source type) and bitline paths for writing either a one or a zero. This separation makes bidirectional writing more flexible and helps mitigate unwanted capacitance on the master bitlines (MBLT/MBLC).

5. Magnet wire driver circuits (MAG CKTS) at the bottom two corners of the array that enable three of the bitlines with high speed magnetic fields through the memory elements.

6. There is a second collum path on the right side of the array that can be enabled with TMSENSE for I-V experimental probing through an external pad.

7. Timings are externally controlled in a robust fashion using signals like PULSEON/PULSEOFF and MAGON/MAGOFF. On-chip circuit complexity is minimized as much as possible so that test and debugging can focus on the MTJs instead of the support circuits.

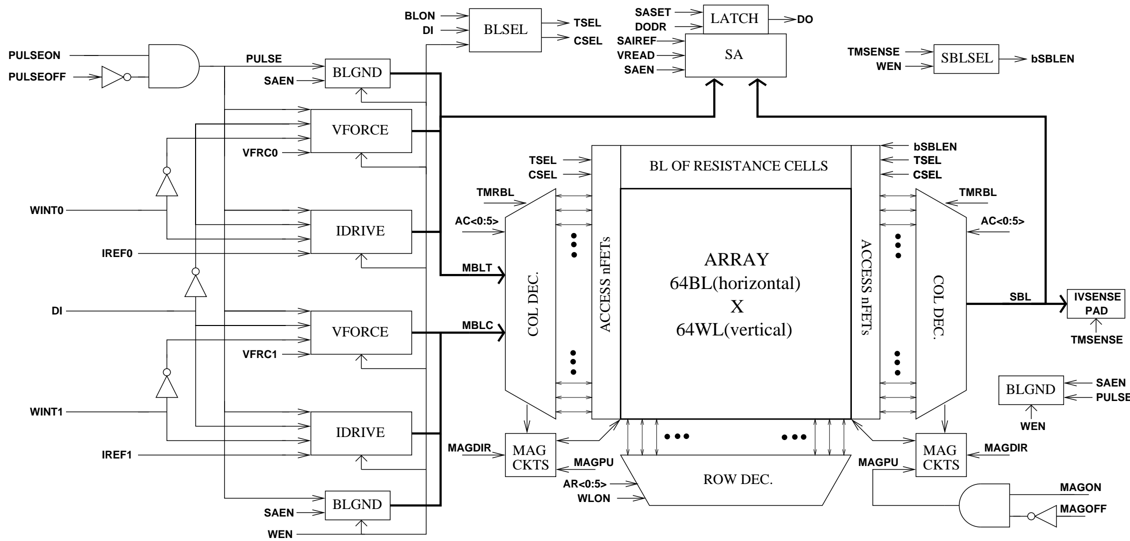


Figure 3.1: The top-level block diagram of the ADM is shown. Key features are a 64x64 array, a dummy bitline of resistance cells, a single sense amplifier, separate write drivers and bitline paths for writing a one or zero, and a second collum path on the right side of the array for IV experimental probing.

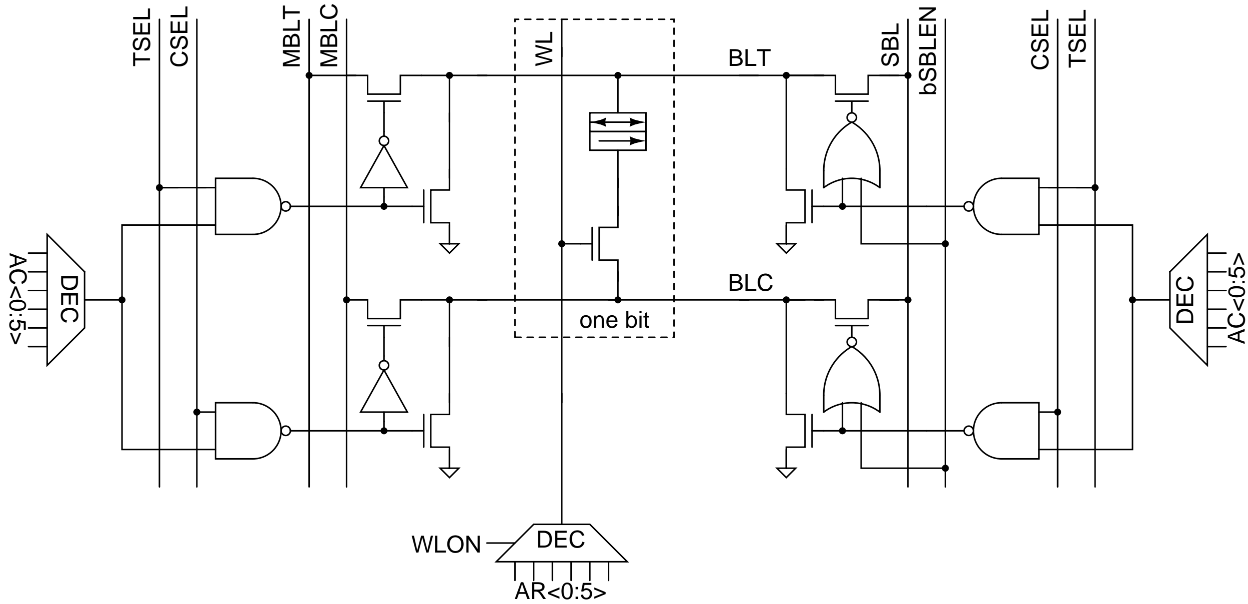


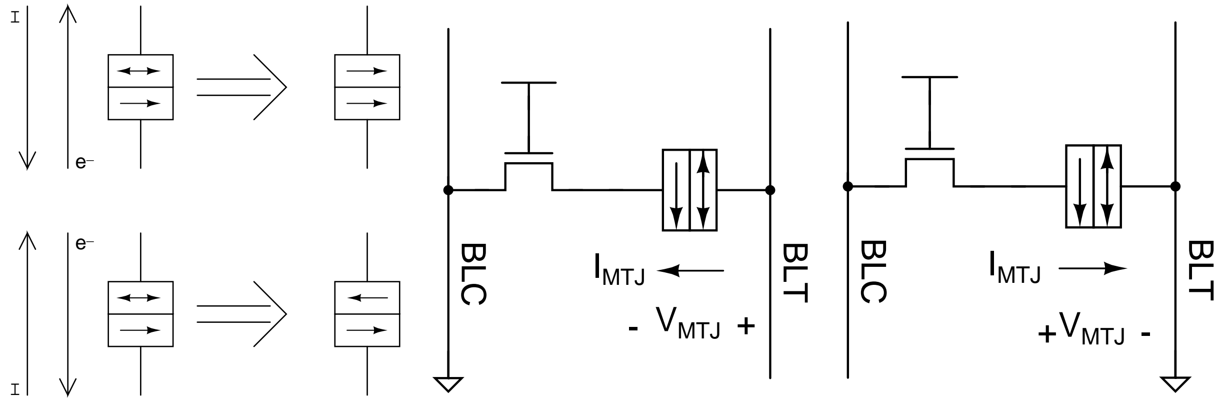
Figure 3.2: Schematic cross-section of array

3.2 The Cell

Shown in Fig. 3.3(a) is a diagram indicating the switching current directions required to write a high and low resistance state [34]. The two arrows in each double-rectangle box represent the magnetization of the two ferromagnetic layers of an MTJ.

Fig. 3.3(b) shows the electrical configuration of the memory cell when writing a low resistance state, or when reading the MTJ. Fig. 3.3(c) shows the electrical configuration of the memory cell when writing a high resistance state. An important circuit limitation with this cell choice is the reduced current capability when writing a high resistance state because the gate-to-source overdrive of the access nFET is degenerated by V_{MTJ}. A wide enough device was chosen to support the maximum desired write currents in both directions. In general, the circuits were sized to meet a broad operating window of resistance ranges (100Ω—$10k\Omega$) for the MTJs.

Shown in Fig. 3.5(a) are vertical cross-sections of the memory cell, along with the current path (indicated by a dashed line) when reading or writing a low resistance state, as schematically illustrated in Fig. 3.3(b). The cell dimensions are much larger than the features of



(a) The STS behavior of an MTJ with respect to the direction of conventional current and electron current is shown.

(b) Electrical configuration of cell when writing a low resistance state, and also when reading the state.

(c) Electrical configuration of cell when writing a high resistance memory element.

Figure 3.3: The arrows represent the magnetization of the two ferromagnetic layers that comprise an MTJ; double arrows signify the free layer. Antiparallel alignment results in a high resistance state, and parallel alignment results in a low resistance state.

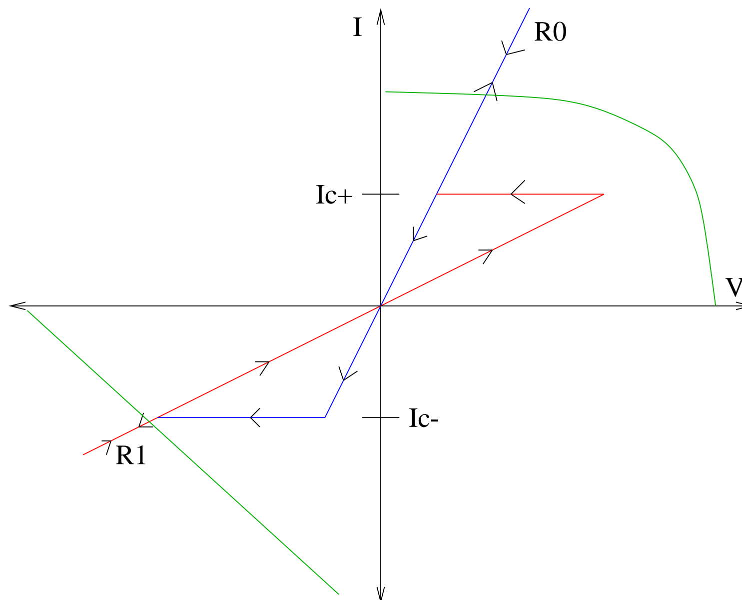
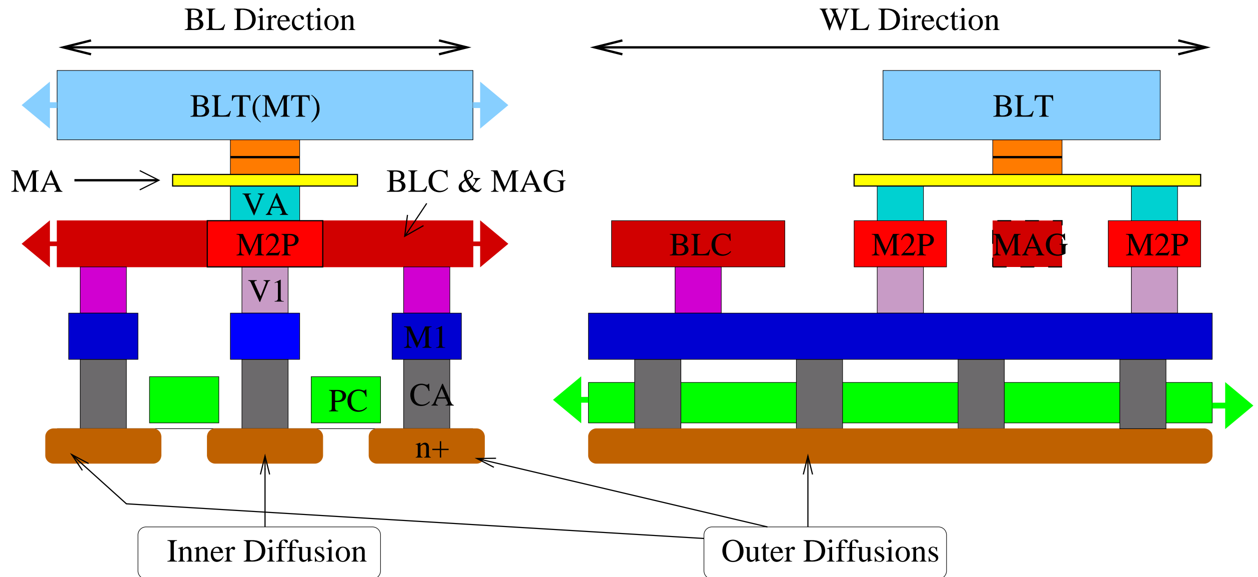


Figure 3.4: Loadline analysis of the IV hysteresis in a bidirectional cell. The FET's current capability is depicted by the green line.

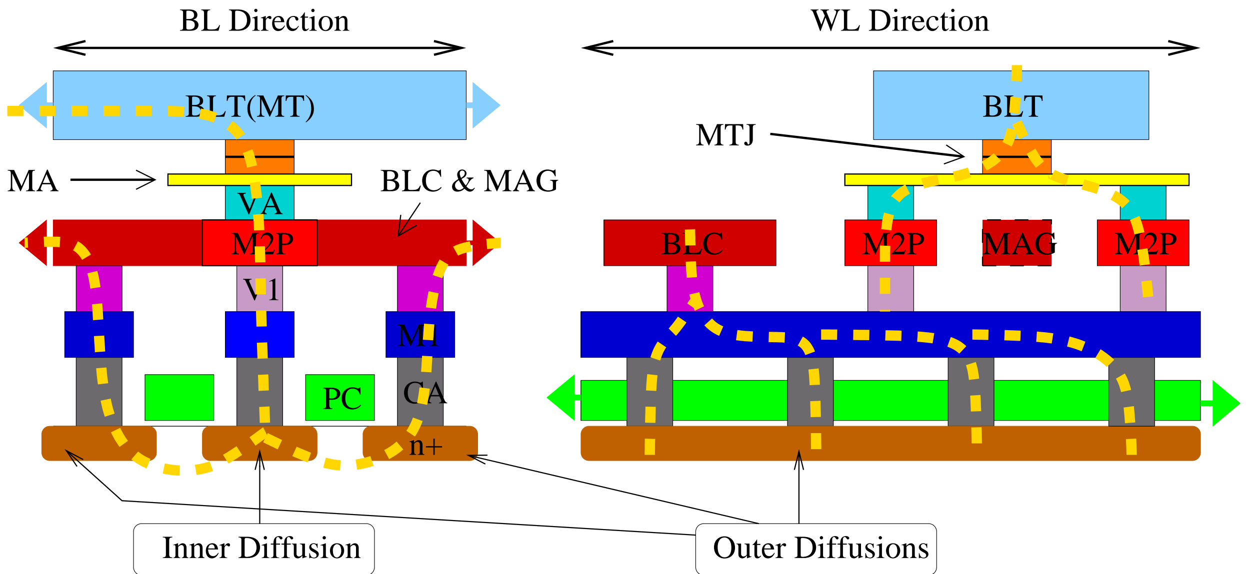
minimum wiring and the memory element because a large FET is being used to support a wide operating window for characterization purposes. To efficiently utilize chip area, the cell nFET is fingered into two parallel devices so the outer diffusions connecting to BLC can be shared between adjacent devices. This avoids the cost of additional spacing for short trench isolation between cell nFETs running along the BL direction.

The MRAM processing technology is similar to what was used by [9] which is described in more detail by [32] and [35]. A thin local interconnect (MA) is patterned as a landing area for the MTJ stack and connects to the second level of metal beneath it through a shallow via (VA). With this design the MA level allows electrical access to the MTJ while also permitting the MAG wire to run close to the MTJ and thus generate significant magnetic fields. The MTJ stack consists of nonmagnetic, ferromagnetic, and antiferromagnetic conductors in addition to an atomically thin tunneling oxide in the middle. The complex arrangement of these materials results in a fixed magnetic layer directly below the oxide and a switchable layer directly above, with a significant change in resistance values between the two states.

Because the cell dimensions are conservative, and STS switching already requires current *through* the MTJ, the array cells can be enabled with a "magnet wire" (shown as MAG in Fig. 3.5(a)) that can pulse magnetic fields through the MTJ during a write cycle in a high-speed fashion. By employing this combination of current and write fields through the MTJ, thermally assisted MRAM can be investigated as well.



(a) The vertical cross-section of the memory cell along both the bitline and wordline directions. There is an optional magnet line (shown in red) called MAG in the "WL Direction" cross-section that can be used to pulse magnetic fields through the MTJ memory element (shown in orange) to explore thermally assisted switching.



(b) Shown in the thick, dashed, gold line is the current path through the cell for reading and writing a low resistance state, as schematically illustrated in Fig. 3.3(b).

Figure 3.5: Vertical cross-section of the memory cell

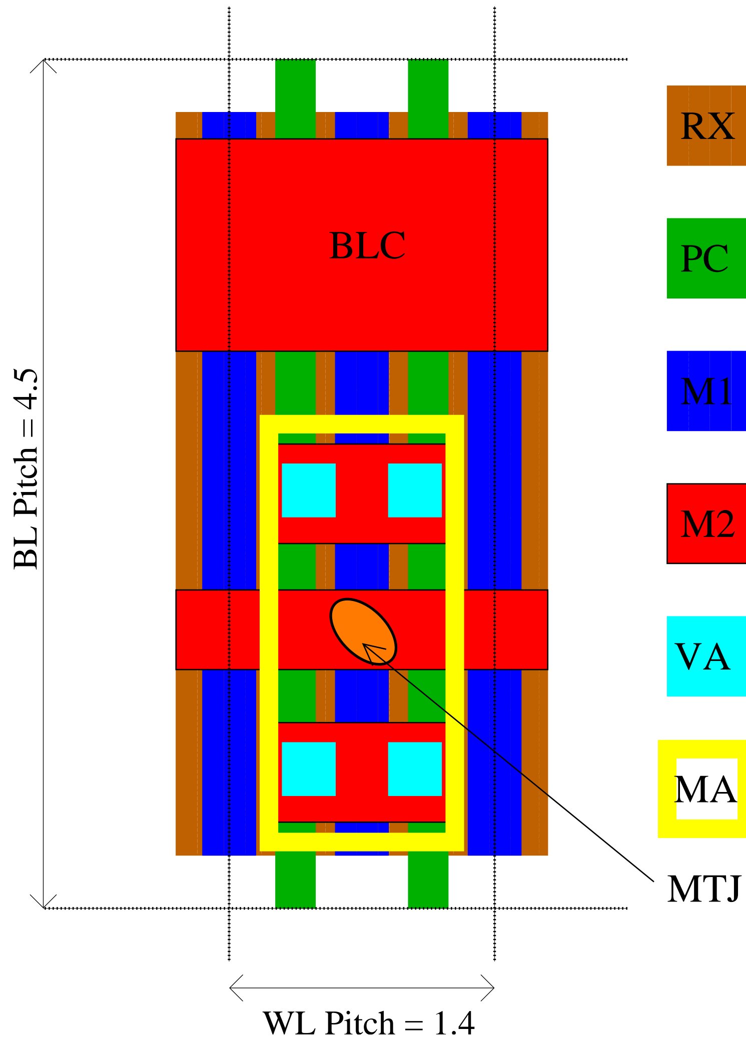


Figure 3.6: Cell Layout, M1, M2. An MT wire runs over the MA landing pad, parallel to BLC (c.f. Fig 3.5(a)). This $6.3\mu m^2$ cell has a wordline pitch of $1.4\mu m$ and a bitline pitch of $4.5\mu m$.

3.3 Row Path

The purpose of the row decoding is to turn on the desired cell's nFET on the selected wordline The cell's nFET is part of a wordline running across all bitlines that is toggled high or low with the state of the WLON signal. There are sixty-four word lines addressed from AR$< 0 : 5 >$ using one stage of two one-of-eight predecoders driving sixty-four NAND gates laid out on the wordline pitch. When WLON is low, it sets to zero all eight outputs of one of the predecoders.

Shown in Fig. 3.7 is the schematic for one of the predecoders. Given the three-bit input address $A < 2 : 0 >$, only one of the eight outputs, $DEC < n >$ is driven high where $n = A < 2 > 2^2 + A < 1 > 2 + A < 0 >$. Also, the ENABLE signal will override the address decoding. The WLON signal utilizes this input to allow the user to disable all rows in the array. By using another predecoder (without an ENABLE signal) on $A < 5 : 3 >$ and then connecting both predecoder outputs to 64 two-input AND gates such that the inputs of the nand gates are $(DECA < i >, DECB < j >) \quad \forall (i, j)$, only one out of the 64 AND gates will drive high. The AND gate with inputs $(DECA < i >, DECB < j >)$ defines row number $8i + j$.

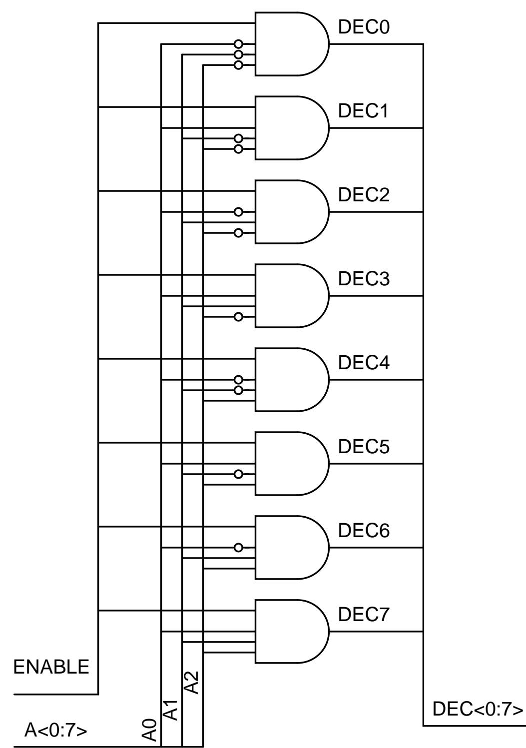


Figure 3.7: Schematic of a “one out of eight” predecoder.”

3.4 Column Path

The purpose of the column decoding is to open up a conductive path to the correct bitline. There are sixty-four normal bitline pairs (columns) and one column of resistance cells. The sixty-four columns are addressed with $AC < 5 : 0 >$ using the same logic decoding scheme for the row decoder. There is additional circuitry in the column decoding (shown in Fig. 3.2) that uses the logical address decoding to enable a connection from BLT to MBLT or BLC to MBLC.

An extra signal TMRBL is used to function as an effective seventh address bit that selects the resistance bitline regardless of the value of AC$< 0 : 5 >$. TMRBL overrides AC$< 0 : 5 >$ by setting to zero all eight outputs of one of the predecoders with the ENABLE signal (see Fig. 3.7). This column decoding scheme is also used to enable the correct magnet line from MAGW$< 1 : 3 >$ that runs under a corresponding bitline BLTC$< 61 : 63 >$, by using the predecoder output from $AC < 0 : 2 >$.

3.4.1 Control Logic

This path from MBLT or MBLC to either BLT or is toggled on and off with the state of the BLON signal, which subsequently chooses between BLT and BLC through an and operation with TSEL and CSEL (see Fig. 3.2), depending on the mode of operation. A circuit of static logic implements the following relations:

When writing a one:

$$CSEL = BLON \cdot DI \cdot WEN$$

When writing a zero or reading:

$$\begin{aligned} TSEL &= BLON \cdot (\overline{DI} \cdot WEN + \overline{WEN}) \\ &= BLON \cdot \overline{\overline{\overline{DI} \cdot WEN} \cdot WEN} \\ &= BLON \cdot \overline{(DI + \overline{WEN}) \cdot WEN} \\ &= BLON \cdot \overline{DI \cdot WEN} \end{aligned}$$

The path from SBL to either BLT or BLC is toggled on and off with the state of the bSBLEN signal. This selection process is is done through using the bSBLEN signal to enable the NOR gate that drives the access nFET on the right side of the array (see fig. 3.2). During a read, SBL is used by the feedback clamp in the sense-amplifier *regardless* of whether the user has enabled the IV experiment mode with TMSENSE. During a write, SBL is disabled to reduce capacitance unless TMSENSE is enabled to allow the user to observe the bitlines during a write. A circuit of static logic implements the following relation:

$$\begin{aligned} SBLSEL &= TMSENSE \cdot WEN + \overline{WEN} \\ bSBLSEL &= \overline{TMSENSE \cdot WEN + \overline{WEN}} \\ bSBLSEL &= \overline{TMSENSE \cdot WEN} \cdot WEN \end{aligned}$$

The only other block that TMSENSE goes to is the PADSW block that simply turns on two wide nFETs between the internal SBL node and the external IVSENSE pad.

3.4.2 MBL/SBL grounding

In order to prevent history-dependent behavior of the ADM during a read or write cycle, MBLT, MBLC, and SBL are conscientiously grounded during idle portions of the cycle time.

Suppose the architecture was much simpler, using only one MBL (and no SBL). Then, one would want MBL to float only when PULSE is high during a write cycle or when SAEN

is high during a read cycle. Namely:

$$MBLFLOAT = PULSE \cdot WEN + SAEN \cdot \overline{WEN}$$

Grounding the MBL node is accomplished by holding the gate of a pull-down nFET high. Therefore, an equivalent way to think about the MBL grounding logic is placing the signal $bMBLFLOAT \equiv \overline{MBLFLOAT}$ on the gate of a pull-down nFET.

$$\begin{aligned}
bMBLFLOAT &= \overline{PULSE \cdot WEN + SAEN \cdot \overline{WEN}} \\
bMBLFLOAT &= \overline{PULSE \cdot WEN} \cdot \overline{SAEN \cdot \overline{WEN}} \\
bMBLFLOAT &= (\overline{PULSE} + \overline{WEN}) \cdot (\overline{SAEN} + WEN) \\
bMBLFLOAT &= \overline{PULSE} \cdot \overline{SAEN} + \overline{WEN} \cdot \overline{SAEN} + \overline{PULSE} \cdot WEN
\end{aligned}$$

The expression for $bMBLFLOAT$ as a "sum of products" allows the last OR operation to be implemented by having each product drive it's own pull-down nFET. Furthermore, the term $\overline{PULSE} \cdot WEN$ can be discarded since $\overline{SAEN}$ in the first product term effectively acts like WEN in the sense that $SAEN$ is always low during a write cycle. Hence,

$$\begin{aligned}
bMBLFLOAT &= \overline{PULSE} \cdot \overline{SAEN} + \overline{WEN} \cdot \overline{SAEN} \\
bMBLFLOAT &= \overline{PULSE + SAEN} + \overline{WEN + SAEN}
\end{aligned}$$

If one were to give both WEN and $SAEN$ high at the same time, the worst that would happen is that MBL remains floating in this unspecified mode of operation. This approach requires only one stage of logic between the timing-critical signal and a grounding nFET, which makes sure that the write drivers or the sense-amp feedback amplifier do not initially face a grounded MBL.

Lastly, the actual ADM architecture has two MBLs (MBLT and MBLC) and another SBL. To keep things simple, this same grounding logic was applied to each of these three nodes. This decision has some of the nodes unnecessarily floating when they don't have to

(for example, MBLT floats even when writing a one via MBLC). However, this decision still accomplishes the fundamental goal of eliminating history-dependent operation from cycle to cycle. It also makes the SBL float at the right time so that TMSENSE can be enabled during a write cycle.

3.5 Magnet Wire

To enable magnet wire functionality, generate a MAGPU pulse from MAGON and MAGOFF in addition to the other signals specified in the write timing (see fig. 3.26). This signal should be timed like PULSE, but it may rise before/after and fall before/after PULSE. During experimentation it will be useful to sweep the relative timings of PULSE and MAGPU. This sweep can be reliably done down to $100ps$ increments.

The selection of the correct magnet wire has been described in sec 3.4. Note that the signal MAGDIR controls the direction of the magnet wire current pulse independent of DI.

The formula for the strength of the horizontal component of $\vec{H}$ is given by: [1]

$$|\vec{H}(x,y)\cdot\hat{x}| = \frac{2I}{cwh}\int_{y-h}^{y}\left[\tan^{-1}\left(\frac{x+w/2}{v}\right) - \tan^{-1}\left(\frac{x-w/2}{v}\right)\right]dv \tag{3.1}$$

with respect to the coordinate system and location of the rectangular wire shown in Fig. 3.8.

[1] This equation is derived by recognizing that the current density out of the plane of the figure is uniform accross the area of the rectangular metal line since $\vec{J} = \sigma\vec{E}$. This current is then partitioned into infinitesimally thin wires of current $dI = JdA = \frac{I}{wh}dxdy$ that each produce a radially symmetric field in the azimuthal direction:

$$\vec{\nabla}\times\vec{H} = \frac{4\pi}{c}\vec{J} \Rightarrow d\vec{H} = \frac{2dI}{cr}(-\sin\theta\hat{x} + \cos\theta\hat{y}), \qquad \sin\theta = \frac{y-y'}{r}$$

Now the integral for the horizontal (in the plane of the MTJ) component of the magnetic field can be written as:

$$\vec{H}\cdot\hat{x} = \int_A d\vec{H}\cdot\hat{x} = -\int_A \frac{2dI}{cr}\sin\theta = -\frac{2I}{cwh}\int_0^h\int_{-w/2}^{w/2}\frac{y-y'}{(y-y')^2+(x-x')^2}dx'dy'$$

The above can be reduced to Eq. 3.1 by letting $u = x - x'$ and $v = y - y'$ and using:

$$\int\frac{1}{a^2+x^2}dx = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right)$$

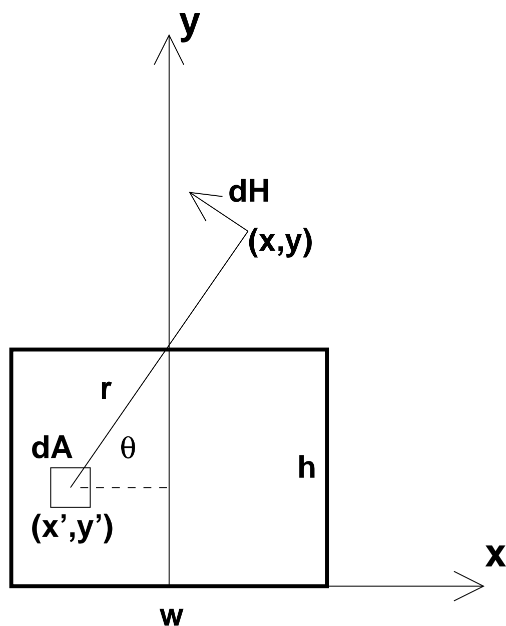


Figure 3.8: Field produced by magnet wire

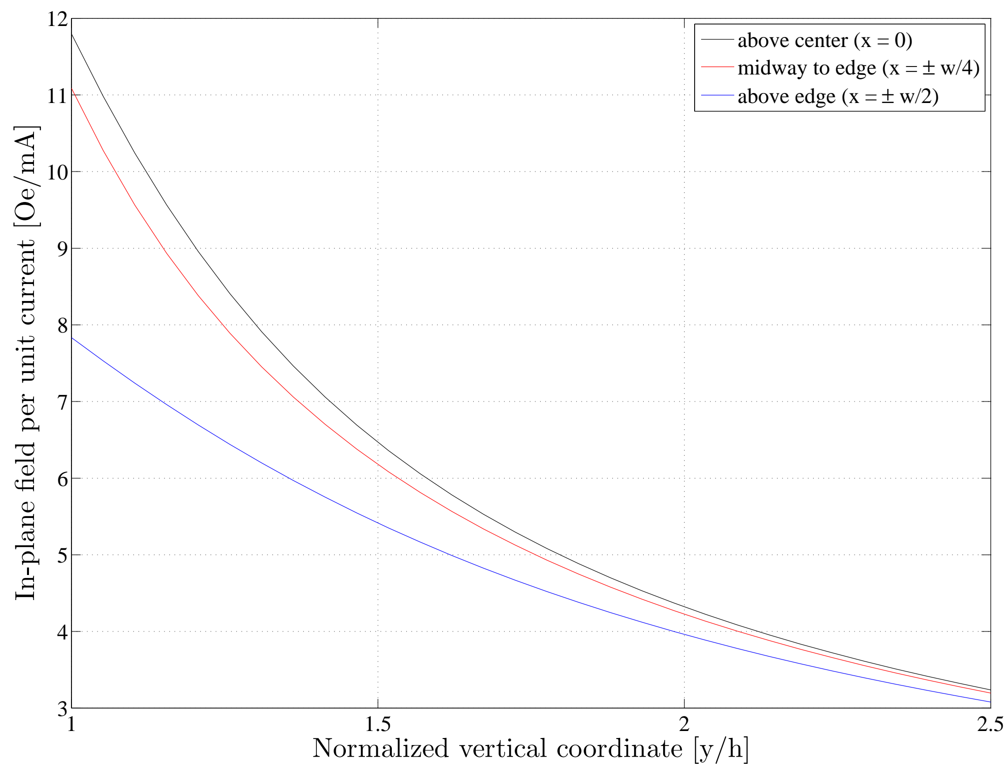


Figure 3.9: Plot of field produced by magnet wire for $1mA$ of current.

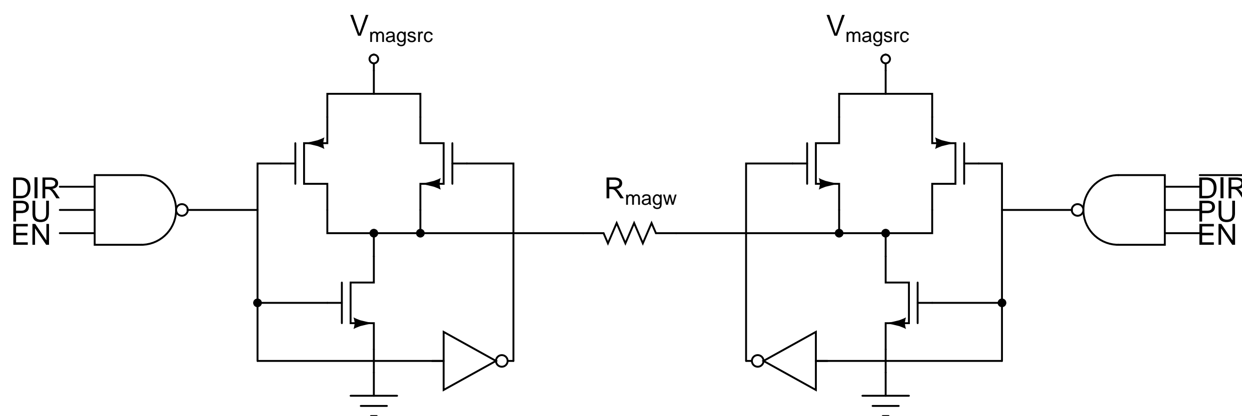


Figure 3.10: Circuits for one of three magnet wires

3.6 Sense-amplifier

The basic function of the sense-amplifier (senseamp) is to convert a signal in resistance difference to a logic high or low voltage. The speed of the sense amplifier determines the read cycle time. The upper bound on power comes from how many sense amplifiers will be needed throughout the memory. Fortunately, a single mram sense amplifier can be shared accross several bitlines allowing for increased area and power budget. The precision of the sense amplifier translates to the speed of the sense amplifier because the time is dominated by how long it takes the read current signal $I_{sig} = |I_{ref} - I_{cell}|$ to overcome offsets in the CMOS circuitry.

The senseamp is similar to the "current sensing" topology in Fig. 3.11 [9] which sets a voltage accross the memory cell and compares the resulting current to a midpoint reference current. This reference current can be generated by a parallel arrangment of memory cells written in opposite states, or it can be externally supplied as a variable input. It is important to note that with ideal device behavior and matching, the operation of the senseamp is seemingly trivial. Namely, if $I_{cell} < I_{ref}$, M1 will quickly pull up the drain of M3 resulting in a straightforward response by the comparator to fire a "1". The performance is entirely dominated by mismatch and paracitic capacitance when reading pathological memory cells in which $I_{cell} \approx I_{ref}$. Therefore, in the following analysis of the sense amplifier, one should assume that $I_{SD1} = I_{SD2}$ and that both M3 and M1 are saturated.

This 4kb array has some new challenges that requires augmentation of the read circuits for use in Spin-MRAM applications. Lower impedance memory cells in the range $100\Omega - 1k\Omega$ want to be characterized for their ability to pass higher current levels. This lower impedance weakens the ability of the source follower M3 to precisely clamp an identical read voltage on cells varying over a broad range of low resistances. The sense current would still monotonically decrease as a function of resistance, but the current seperation I_{sig} between two states will be compressed and increasingly dependent on the FET device parameters, which will vary accross chip sites. Furthermore, paracitic resistances in the column path from the column access devices and cell FET will further degrade the amount of observable

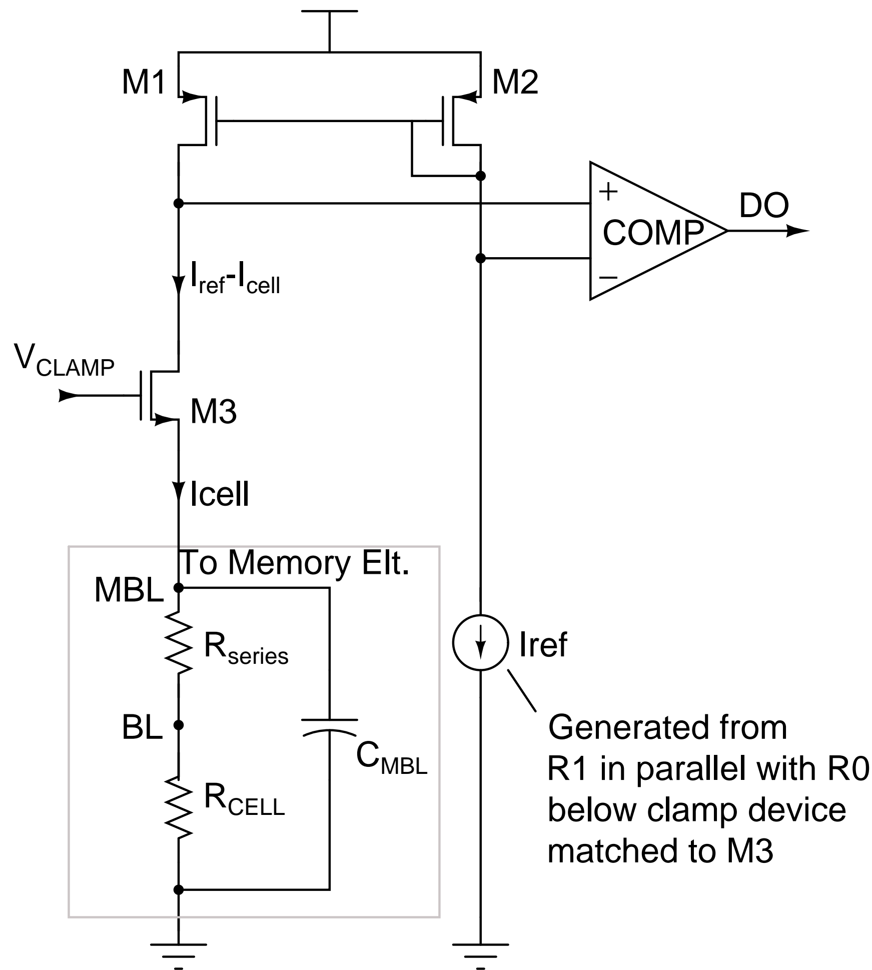


Figure 3.11: Prior sense-amplifier topology.

resistance separation. For all these reasons, prior MRAM designs have utilized MTJs with resistances well above $1k\Omega$.

Even without the lower impedance obstacle, more precise clamping voltage is desired for this experimental MRAM array for the development phase of the life cycle so that resistance distributions can be accurately extracted. This type of data gathering is needed before a product demonstrator senseamp can be optimized for a specific operating point. Also, one may want to vary the read voltage which is nominally designed in this 4kb array at what was taken to be the minimum tolerable level of $50mV$. In the experimental MJTs that switch with current through them, there is a danger that the current experienced during a read operation

may disturb the memory cell by causing unwanted switching. Recall the experimentally demonstrated effect of subthreshold spin transfer current accelerating the poisson error rate in Eq. 2.45. This equation says that the read current level must be translated back to a device engineering requirement of compensating with an increased energy barrier.

To address these challenges, a feedback amplifier is added in this 4kb array to drive the gate of a source follower nFET to precisely clamps the read voltage on the memory element (see fig. 3.12). This allows for a wider variation in magnetoresistance while still being able to clamp a uniform voltage in a high-speed ($100ns$) fashion. It will also allow more precise measurement of the conductance distribution. This design generalizes to sensing resistive memory elements, and it's higher degree of precision may allow reading multi-level bits (through multiple reference currents). To further enhance the ability to characterize the memory cell, the paracitic resistance in the column path was eliminated by sesning the voltage through the SBL path which does not share the current path from the source follower nFET by going through a second column decoder on the opposite side of the array (see Fig.3.2). Therefore, the $50mV$ reference is being truly set at BLT and not just MBL.

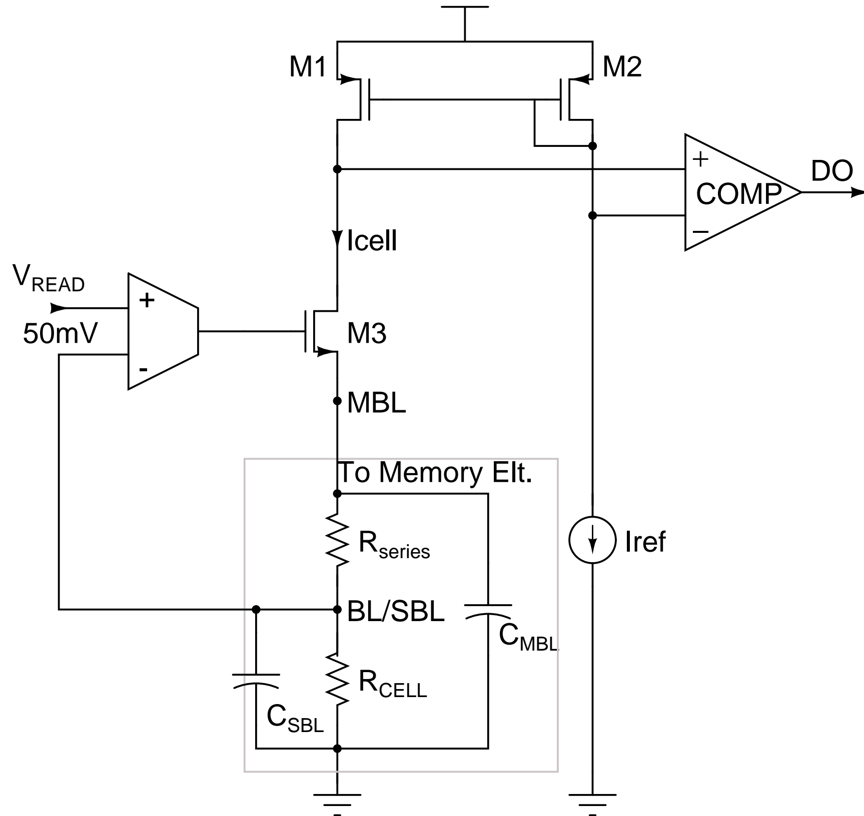


Figure 3.12: Sense-amplifier topology.

3.6.1 Sizing of mirrors and source follower clamp device

The first stage of gain in the read path is the pfet current mirror load (M1-M2 in Fig. 3.12). The voltage gain is simply $ro1$ in parallel with the source-degenerated output impedance of M3: [2]

$$\begin{aligned} R_o &= ro1||(R_S + (1 + g_{s3}R_S)r_{o3}) \qquad (3.2) \\ R_S &= R_{series} + R_{cell} \end{aligned}$$

First the pFETs were sized. Because the value of I_{cell} is not well controlled (since a

[2]Whenever a numeric subscript i is used with small signal variables like $\{r_{oi}, g_{mi}, g_{mbi}, g_{si}, C_{gdi}, C_{gsi}\}$, the the variable corresponds to the MOSFET with the same numeric suffix Mi. Also $g_s = g_m + g_{mb}$. The definitions of small signal models and the variable conventions can be found in [36].

wide range of R_{cell} needs to be supported), the output resistance of M1 cannot be optimized for a single value of quiescent current. The figure of merit is not $dV/dI = R_o$ but rather $dV/(dI/I) = RoI_{SD1}$, i.e. the voltage gain for a *fractional* current change. In addition, a large device while still preserving a sizable gate overdrive over this range is desired to minimize the mismatch. A plot of the resutling R_oI_{SD1} from the chosen device size is shown in Fig. 3.13. One can expect that a 10% change in current will produce atleast one V_{dd} of swing in an incremental sense.

Next, the source follower nFET was sized so as to not detract from the output impedance of the pfet and to have a low enough output impedance to drive the cell resistance precisely. Because of the feedback amplifier, a relatively small device could be used. The lower bound on the size of M3 was not the incremental output impedance but rather the required voltage range on its gate over the range of read current values. This translates into the output voltage swing requirement of the feedback amplifier.

Finally, the required voltage swing on the drain of M1 is much less than V_{dd} because it goes through the second gain stage of the comparator. The true "work" done by I_{sig} is to overcome the offset of the comparator and the mismatch in M1 and M2.

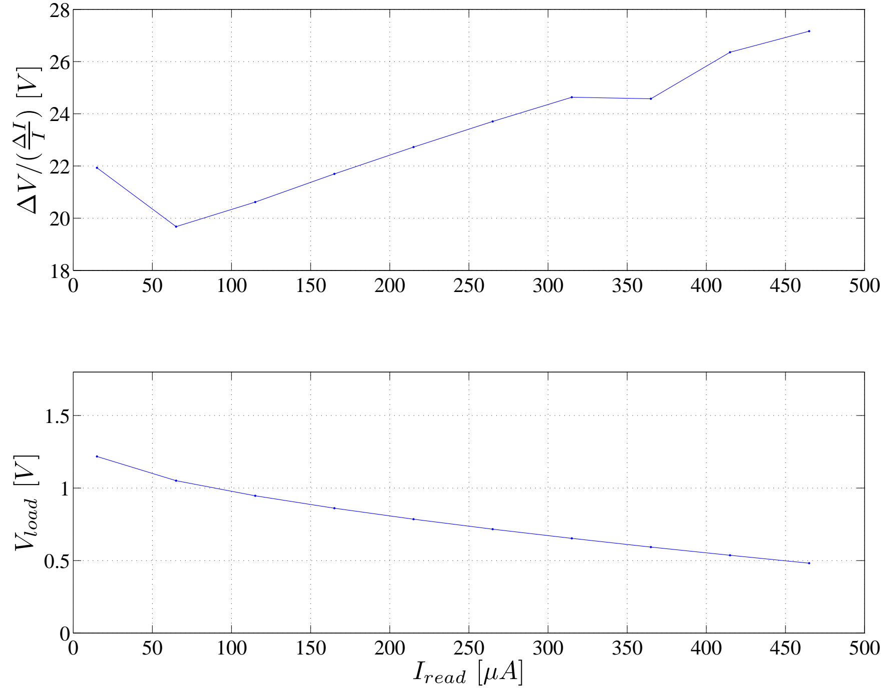


Figure 3.13: Voltage gain for a fractional input of current at various values of quiescent read current. On the bottom plot is the resulting voltage at the gate/drain of M2.

3.6.2 Design of transconductance amplifier

To meet the low input common mode requirements, input pfets were used. To meet the large swing and $V_{dd}/2$ nominal output value to drive the gate of the source follower nFET, a folded cascode buffer into a cascode mirror load was used. These DC requirements set the basic topology of the chosen amplifer in addition to the precision requirement for minimal offset. This topology has only three critical pairs that need to be matched: M1-M2, M9-M10 for both small signal and operating point considerations, and M3-M4 for large signal operating point cosiderations (quiescent currents need to match). For such a low common mode input and compartively higher output, this topology has the least possible number of matching critical pairs.

Looking at the small signal circuit in Fig. 3.15, M5 buffers the current prouced by M1 and similarly M6 buffers the current produced by M2. The pfet cascode mirror reflects M5's current at V_{out} and the resulting summation of current from M5 and M6 passes produces a

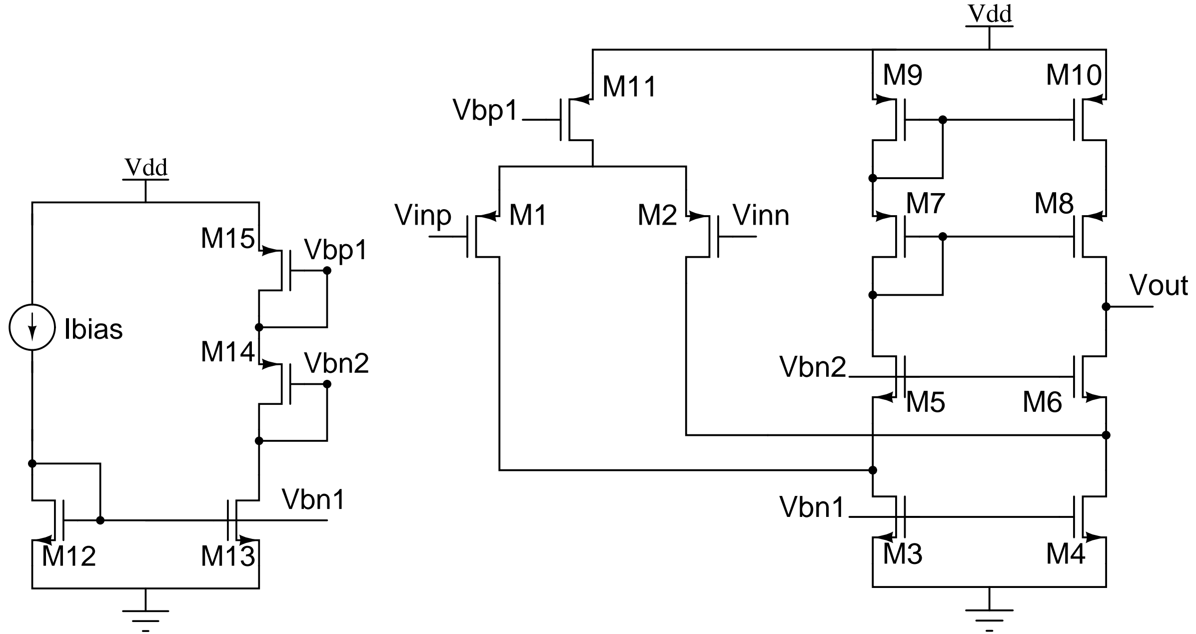


Figure 3.14: Schematic of transconductance amplifier

high gain at the high output impedance at node of V_{out}. Before analyzing the small signal circuit in more detail, the biasing of M3-M4 and M11 comes from straight forward current mirroring of I_{bias}. The sizes of the devices were made large to enhance matching, and the levels of current were chosen to run at a significant amount of gate overdrive to further enhance matching. The biasing of the cascoding nFETs was a little more subtle. V_{bn2} needs to be low enough to keep M2-M1 in saturation, especially at very low input common mode voltages but it is also desirable to center the quiescent value of V_{out} in the range of expected gate voltages for the clamping nFET in Fig. 3.12 to reduce error from finite gain. To reconcile this, the biasing $pFET$ transistor M14 matches M17 in size and bias current and was designed such that V_{bn2} is centered in the expected range of required output voltages. This biasing guarantees that the amplifier is centered in the middle of the output swing and that M5, M6, M7, and M8 are firmly in saturation because for all of them, their gate-source voltage equals their drain-source voltage.

Now, the detailed analysis of the small signal circuit of the sense amplifier will yield more insight into device sizing and biasing. Ultimately, one would like the circuit to simply

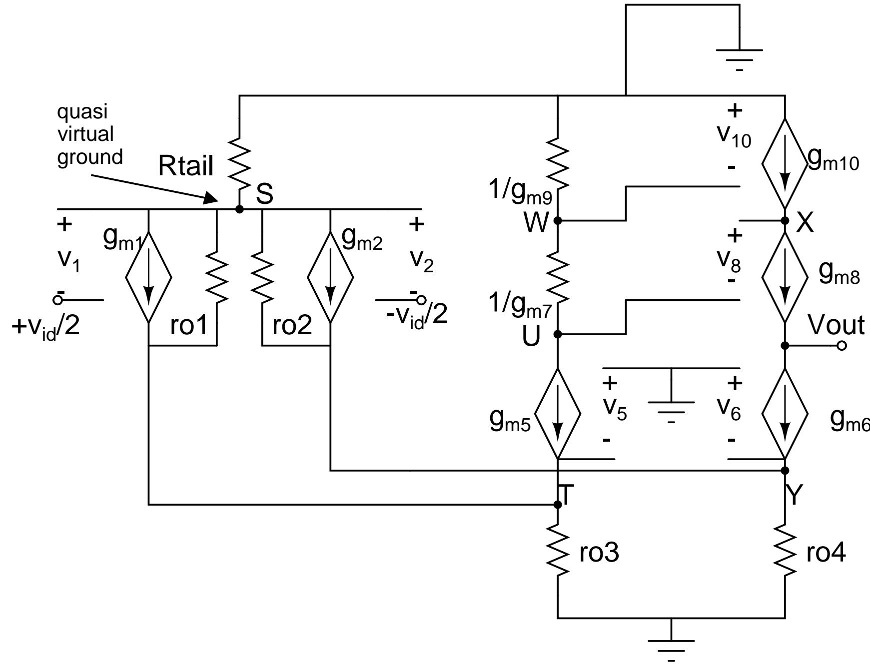


Figure 3.15: Full small-signal schematic of transconductance amplifier

behave like a high output impedance R_o transconductance amplifier G_m with gain $G_m R_o$. A differential input on the gates of M1 and M2 is assumed. Shown in Fig. 3.17 is a small signal schematic of the amplifier with only the most salient feature of the FETs shown (although all three of g_m, g_{mb}, and r_o will be included in the calculations).

Looking at Fig. 3.17, the pfet mirror M7-M9-M10-M8 asymmetrically presents a low impedance to the drain of M5 and a very high impedance to the drain of M6. This breaks the symmetry needed to argue that node S is a virtual ground to employ differential half-circuit techniques. However, if the output node is connected to a low impedance source and one is interested in the short circuit output current, half-circuit techniques will yield accurate results.

To begin the analysis, a two port current amplifier model will be determined for the M7-M9-M10-M8 cascode mirror with parameters R_{im}, A_{im}, and R_{om}. By drawing a test

current i_t *from* node U and solving for the resulting $-v_t$, the input resistance can be solved:

$$\begin{aligned}
v_W - v_t &= (i_t - g_{mb} v_W)/g_{m7} \\
v_W &= -i_t/g_{m9} \\
(-\frac{i_t}{g_{m9}} - v_t) &= i_t(1 + \frac{g_{mb7}}{g_{m9}})/g_{m7} \\
-v_t &= i_t \left[\frac{1}{g_{m7}} + \frac{1}{g_{m9}} + \frac{g_{mb7}}{g_{m7} g_{m9}} \right] \\
\Rightarrow R_{im} &= \left[\frac{1}{g_{m7}} + \frac{1}{g_{m9}} + \frac{g_{mb7}}{g_{m7} g_{m9}} \right] \qquad (3.3)
\end{aligned}$$

This low value of input resistance will be easily driven by the cascode nFET M5.

Next, the current gain is solved by recognizing that the current divider formed by $1/g_{m9}$ and r_{o9} will result in the current mirrored by M10 to equal:

$$i_{10} = \frac{r_{o9}}{r_{o9} + 1/g_{m9}} i_{in}$$

However, not all of i_{10} makes it to the output node. The error term is due to the current shunted accross r_{o10}:

$$i_e = (i_{10} - i_{g8}) \frac{r_{o8}||\frac{1}{g_{s8}}}{r_{o10} + r_{o8}||\frac{1}{g_{s8}}}$$

where i_{g8} is the component of the tranconductance current in M8 due to the voltage swing on node U.

$$i_{g8} = i_{in} g_{m8} R_{im}$$

Therefore the short circuit output current $i_{10} - i_e$ is

$$i_{sc} = \frac{r_{o9}}{r_{o9} + \frac{1}{g_{m9}}} \frac{r_{o10}}{r_{o10} + r_{o8}||\frac{1}{g_{s8}}} i_{in} + \frac{r_{o8}||\frac{1}{g_{s8}}}{r_{o10} + r_{o8}||\frac{1}{g_{s8}}} g_{m8} R_{im} i_{in}$$

and

$$\begin{aligned} A_i &= \left[\frac{r_{o9}}{r_{o9} + \frac{1}{g_{m9}}} \frac{r_{o10}}{r_{o10} + r_{o8} || \frac{1}{g_{s8}}} + \frac{r_{o8} || \frac{1}{g_{s8}}}{r_{o10} + r_{o8} || \frac{1}{g_{s8}}} g_{m8} R_{im} \right] \\ A_i &\approx 1 \end{aligned} \tag{3.4}$$

Lastly, the output resistance is a common source pFET with source degeneration $R_S = r_{o10}$ (see Eq. 3.2):

$$R_{om} = r_{o10} + (1 + g_{s8} r_{o10}) r_{o8} \tag{3.5}$$

With the pFET current mirror load abstracted into a two port network, the analysis can be further clarified by formulating a Thevinin equivalent for the input stage: M11-M1-M2-M3-M4. [3] By taking nodes T and Y to be the two terminals of the Thevinin equivalent, the subcircuit is perfectly symmetric about node S and half-circuit analysis can be employed with node S grounded to give the Thevenin voltage as simply the voltage gain of an input nFET loaded by the two output resistances:

$$V_{th} = g_{m1,2} (r_{o1,2} || r_{o3,4}) v_{id} \tag{3.6}$$

where $r_{o1,2}$ reads as r_{o1} or r_{o2} because they are equal.

The output resistance can be quickly found by decomposing the test voltage source v_t applied to node T with respect to node Y into $+v_t/2$ at node T (with respect to ground) and $-v_t/2$ at node U (with respect to ground). Under this decomposition, node S is still a virtual ground and therefore the transconductance generators of both input FETs are off. The resulting Thevinin resistance is:

$$R_{th} = 2(r_{o1,2} || r_{o3,4}) \tag{3.7}$$

At this point, the small signal circuit in Fig. 3.15 is reduced to Fig. 3.16. The output

[3] These calculations are in the spirit of the analysis by [37] of the standard 5 transistor ordinary transconducatnce amplifier (OTA) with a single-ended output.

resistance of the amplifier can now be determined by applying a test voltage v_t at the output and determining the resulting current, which is equal to $-(i_p + i_n)$. First, disregard the current source $A_i i_{im}$ and accompanying output resistance R_{om}. The current into the drain of the cascode nFET M6, is simply v_t divided by the output resistance looking into the drain of M6.

$$-i_n = \frac{v_t}{R_S + (1 + g_{s6} R_S) r_{06}}$$

This resistance is another source degenerated common base output resistance with

$$R_S = R_{th} + R_{i5}$$

where R_{i5} is the resistance looing into the source of cascoding nFET M5. Determining the expression for R_{i5} utilizes Fig. 3.17, which illustrates the effect of R_{im} to produce a negative feedback in current accross r_{o5}.

$$\begin{aligned} |\frac{i_{im}}{v_t}| &= \frac{1/r_{o5}}{1 + R_{im}/r_{o5}} + g_{s5} \frac{1}{1 + R_{im}/r_{o5}} \\ |\frac{i_{im}}{v_t}| &= \frac{1 + g_{s5} r_{o5}}{r_{o5} + R_{im}} \\ \Rightarrow R_{i5} = &= \frac{r_{o5} + R_{im}}{1 + g_{s5} r_{o5}} \end{aligned} \tag{3.8}$$

Now reconnect the $A_i i_{im}$ current source. By tracing clockwise from M6, one can see that $i_{im} = i_n$ so now the test voltage source has to support an additional current load of $A_i(-i_n)$. The component of the output resistance due to this current and the original $-i_n$ into M6 is:

$$\begin{aligned} R'_o &= \frac{v_t}{(-i_n) + (-i_p)} \\ &= \frac{1}{1 + A_i} \frac{v_t}{-i_n} \\ &= \frac{1}{1 + A_i} [R_S + (1 + g_{s6} R_S) r_{o6}] \end{aligned}$$

Finally the output resistance of the current mirror R_{om} is added in parallel to the above

expression to yield a final result for the output resistance of the amplifier:

$$R_o = \left(\frac{1}{1 + A_i} \left[R_{th} + R_{i5} + \left(1 + g_{s6} \left[R_{th} + R_{i5} \right] \right) r_{o6} \right] \right) || R_{om} \tag{3.9}$$

The last parameter to calculate is the overall transconductance G_m by shorting the output to ground and finding the resulting short circuit current. Having calculated R_{i5} in the previous step, the small signal schematic reduces to Fig. 3.18, from which the i_n component of the output current is read off from the feedback loop:

$$i_n = V_{th} \frac{g_{s6} + 1/r_{o6}}{1 + (R_{th} + R_{i5})(g_{s6} + 1/r_{o6})}$$

Adding in the current from the pFET cascode mirror gives:

$$\begin{aligned} i_{out} &= i_n + i_p \\ &= i_n(1 + A_i) \\ &= (1 + A_i) g_{m1,2} (r_{o1,2} || r_{o3,4}) v_{id} \frac{g_{s6} + 1/r_{o6}}{1 + (R_{th} + R_{i5})(g_{s6} + 1/r_{o6})} \end{aligned} \tag{3.10}$$

and the overall transconductance is:

$$G_m = (1 + A_i) g_{m1,2} (r_{o1,2} || r_{o3,4}) \frac{g_{s6} + 1/r_{o6}}{1 + (2(r_{o1,2} || r_{o3,4}) + R_{i5})(g_{s6} + 1/r_{o6})} \tag{3.11}$$

Taking the appropriate limits, one can see how this expression approaches $G_m = g_{m1,2}$. The gain, which translates to accuracy in clamping voltage is $G_m R_o$.

Having analyzed the small signal behavior of the amplifier, one now has a complete picture of the sizing and bias requirements.

Table 3.1: Small signal design considerations.

Variable	Expression	Design goal
R_{im}	$Eq.$ 3.3	small
A_i	$Eq.$ 3.4	unity
R_{om}	$Eq.$ 3.5	maximize
V_{th}	$Eq.$ 3.6	big
R_{th}	$Eq.$ 3.7	big
R_{i5}	$Eq.$ 3.8	small
R_o	$Eq.$ 3.9	big
G_m	$Eq.$ 3.11	big

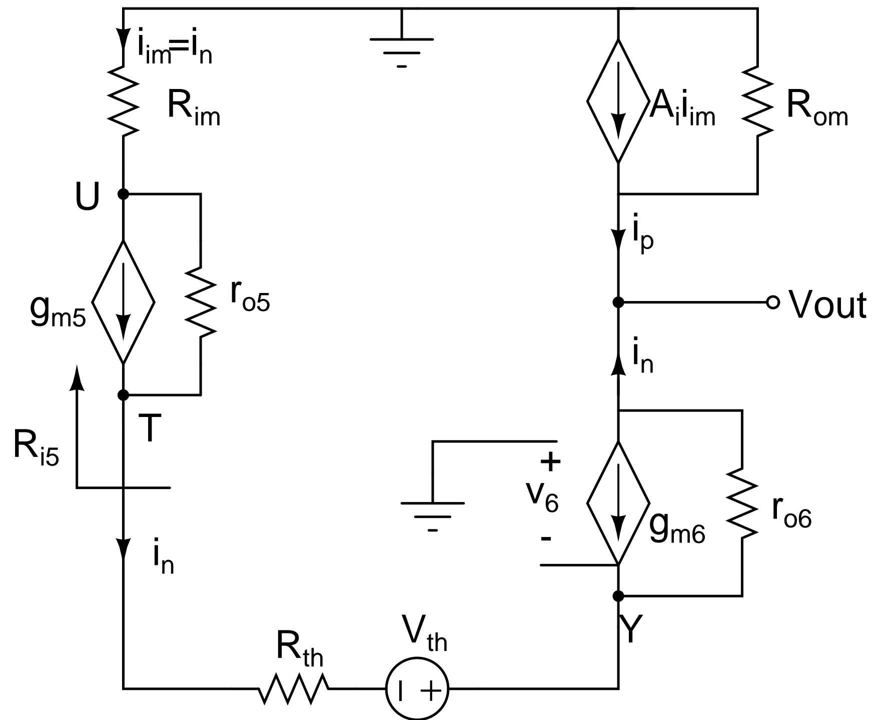


Figure 3.16: Simplified small-signal schematic of transconductance amplifier

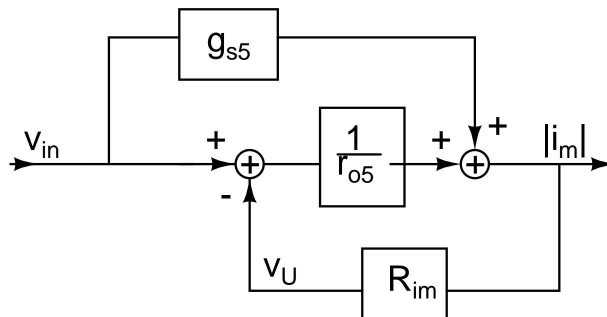


Figure 3.17: Feedback loop for M5

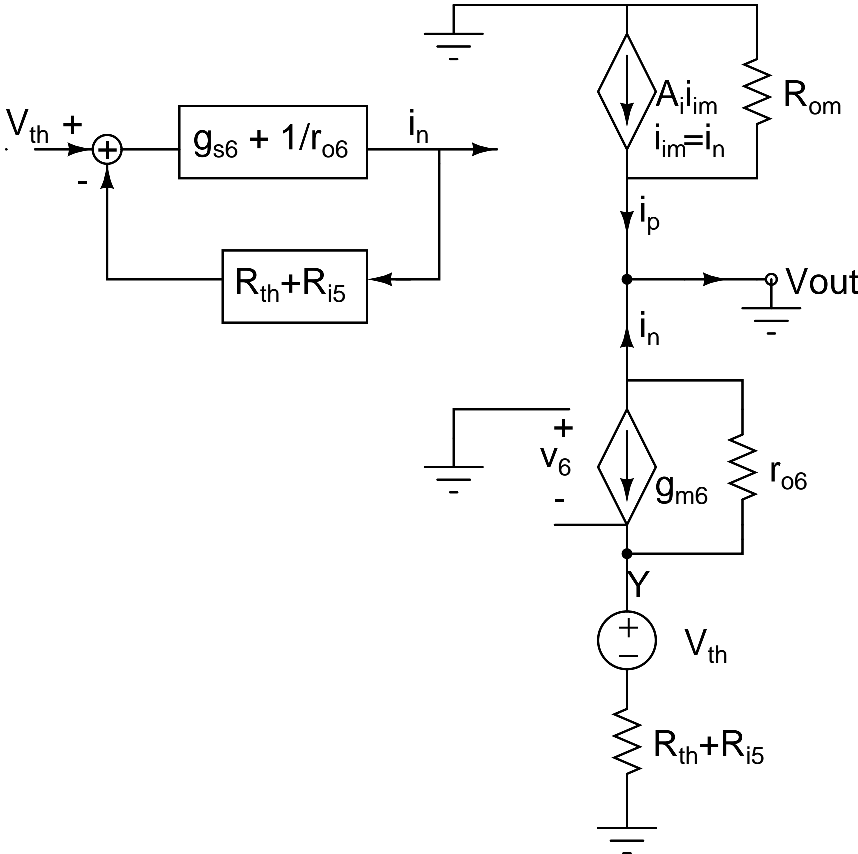


Figure 3.18: Small-signal schematic to calculate overal G_m

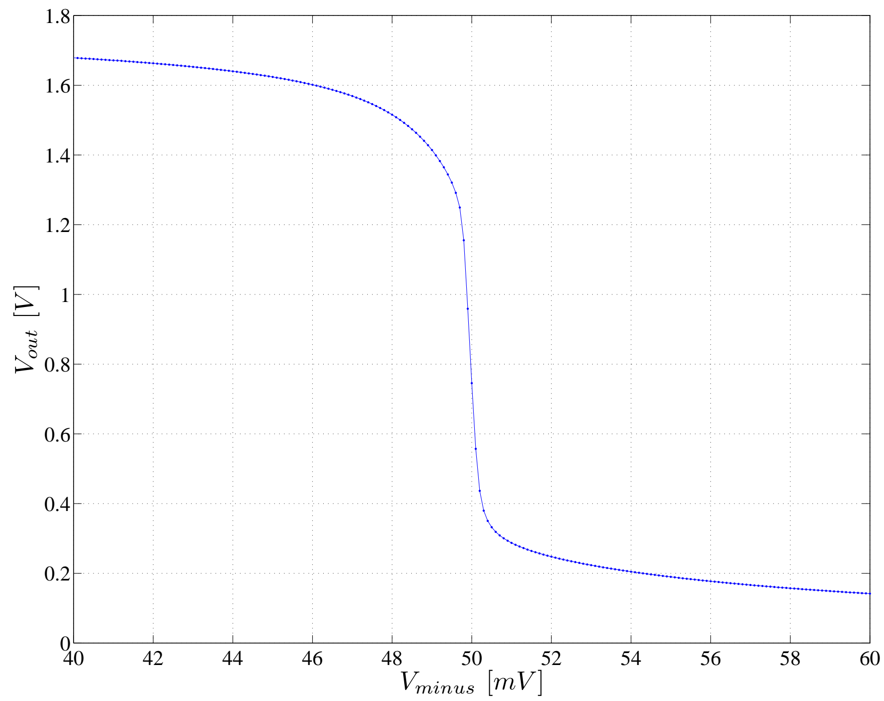


Figure 3.19: Simulation of VTC of feedback amplifier with chosen sizings and bias.

3.6.3 Analysis of loop dynamics

The big win on precision by sensing the resistance through the SBL wire which carries no DC current (and hence no voltage offset from BLT) has the reprecussion of the loop dynamics depending on the path from one end of the array to the other. Conservative, slow, dominant pole compensation was utilized to make sure that the complicated impedance of the collumn path does not make the amplifier unstable. As an added difficulty, the loop transmission strongly depends on the resistance value of the memory cell (which sets the bias current and in turn transconductance of the source follower nFET).

The dominant pole is given by

$$\tau_1 = \frac{g_m}{C_{COMP}}$$

and the unwanted pole from the array is

$$\tau_2 = \left[\frac{1 + r_{op}/r_{on}}{g_{s1}} || R_{cell}\right] (C_{MBL} + C_{SBL})$$

The resistance for τ_2 comes from a calculation similar to that of Eq. 3.8.

Given the gain from the + terminal to the breakpoint at the - terminal, the location of τ_1 was chosen by adjusting C_{COMP} so that there would be less than 10% overshoot over the operating range of cell resistances. Shown in Fig. 3.21 is a bode plot of the loop transmission; a phase margin of 75° is achieved.

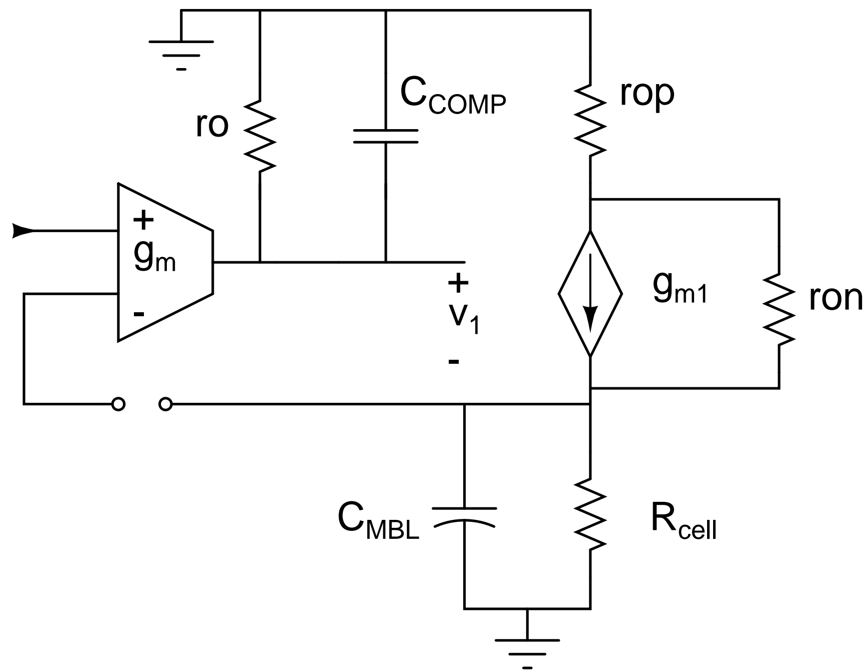


Figure 3.20: Small signal circuit for stability analysis

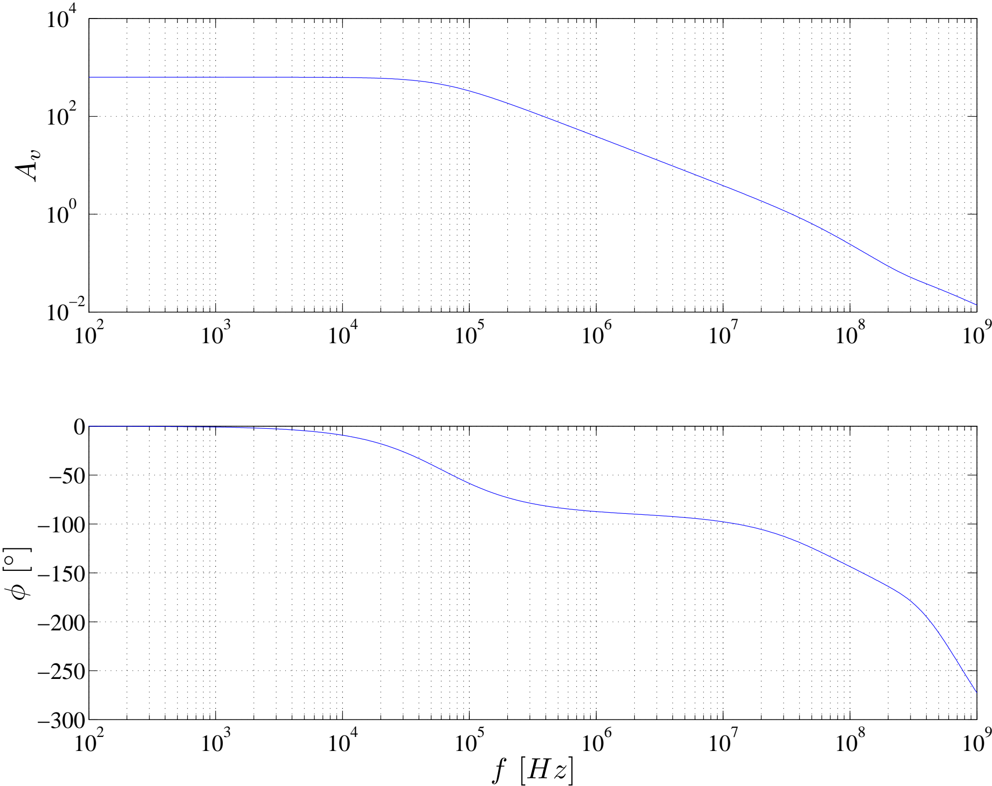


Figure 3.21: Bode plot of sense-amplifier loop transmission

3.6.4 Transient operation

The read time is dominated by the slew rate at the output of the transconductance amplifier as it ramps the gate of the source follower nFET up to its appropriate value. Note, the second order transfer function from the AC analysis will not give the correct rise time because of the large period of slewing.

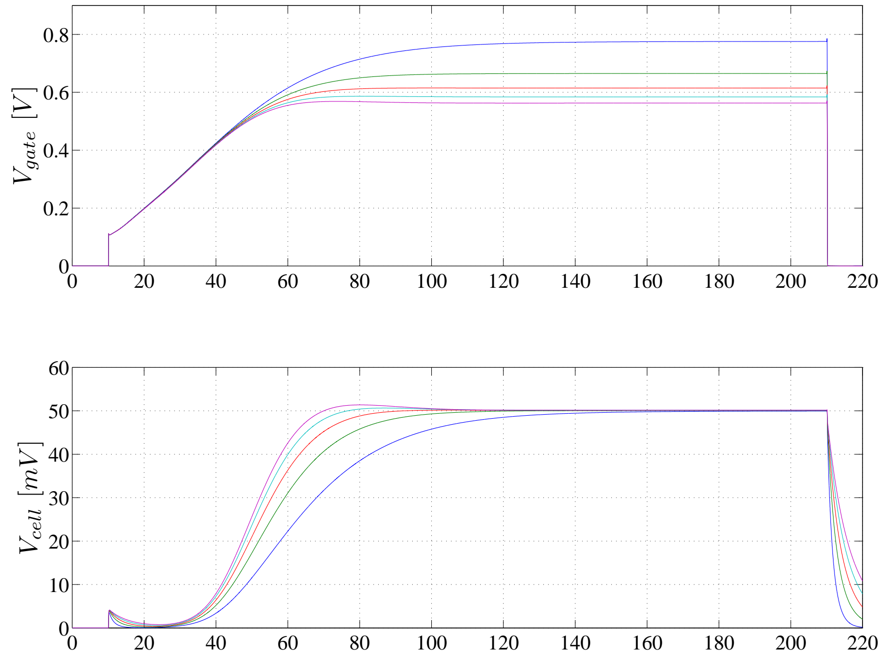


Figure 3.22: Transient wave form of sense-amplifier

To ensure that the initialization of driving data off chip does not disrupt the nodes at the comparator input, the data from the comparator is sent through two level-sensitive latches in a conservative strategy that guarantees no data corruption to occur when the off chip driver potentially disrupts the current mirror load in the sense-amplifier. The driver is not tristate, it is always actively driving the data from the second latch. A seperate data input pad is used for the write drivers.

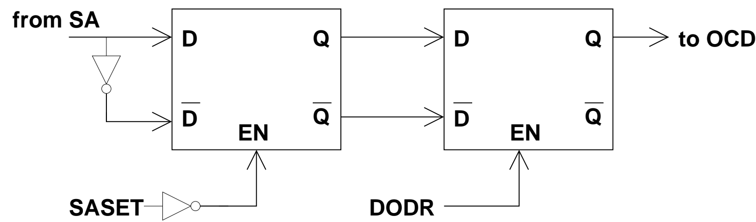


Figure 3.23: The outpout of the SA goes through two latches

3.7 Write Drivers

There are two kinds of write drivers:

1. There is a VFORCE write driver that simply opens up MBLT or MBLC (depending on DI) to a pad that will be connected to a low impedance voltage source. This will be useful for exploring the fastest possible write pulses.

2. There is an IDRIVE write driver that mirrors a reference current (and scales it by a factor of six) into the selected cell through the MBL → BL path. This writing mode will be useful for experiments that control the current level with precision, in order to explore the nature of the resistance hysteresis with current as the independent variable. Furthermore, the switching threshold is conceptually clearer in terms of a threshold current.

All write drivers are enabled with the PULSE signal.

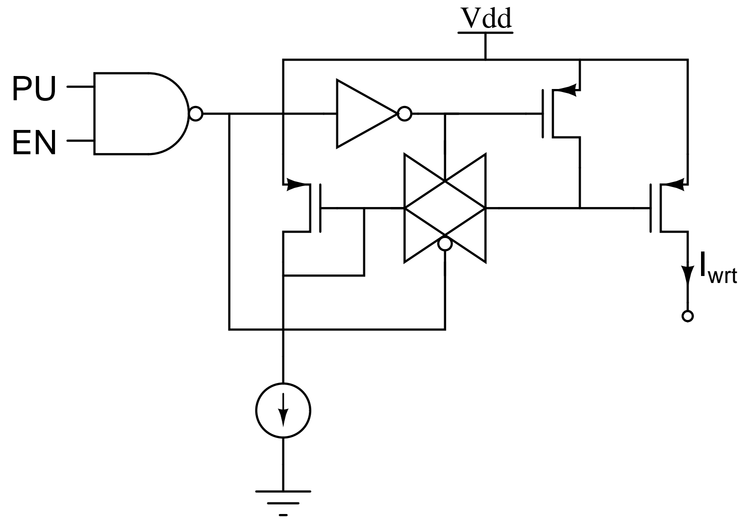


Figure 3.24: Schematic of current driver

3.8 Operation

Table 3.2: Functional description of normal operation

AR$<0:5>$	AC$<0:5>$	OEN	DI	WEN	WINT$<0:1>$	Description
X	X	X	X	1	00	write "0"s and "1"s using VFORCE
X	X	X	X	1	01	write "0"s using VFORCE and "1"s using IDRIVE
X	X	X	X	1	10	write "0"s using IDRIVE and "1"s using VFORCE
X	X	X	X	1	11	write "0"s and "1"s using IDRIVE
X	X	1	X	0	X	read location AR$<0:5>$, AC$<0:5>$ to D0

"X" signifies that the signal can be any legitimate digital value.

3.8.1 PULSE timing

To precisely control pulse width from the tester, the PULSE signal is generated from the difference in arrival times of PULSEON and PULSEOFF as shown in fig. 3.25.

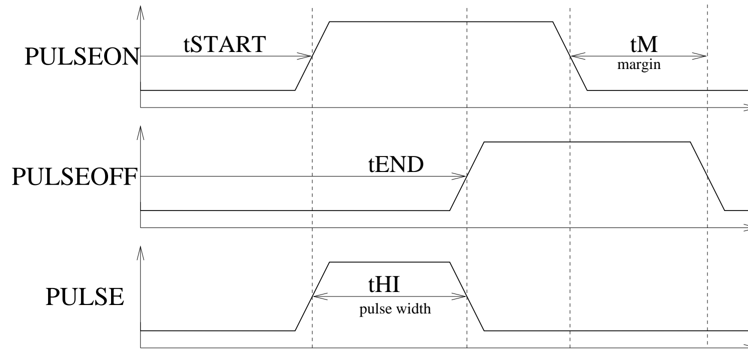


Figure 3.25: Timings for PULSE signal: As a consequence of the logic used in Fig. 3.1, PULSEON must fall before PULSEOFF to prevent a second pulse. This scheme is also used to generate the MAGPU signal from MAGON and MAGOFF.

3.8.2 Standard Write

All other signals not described for a standard write should be strictly zero.

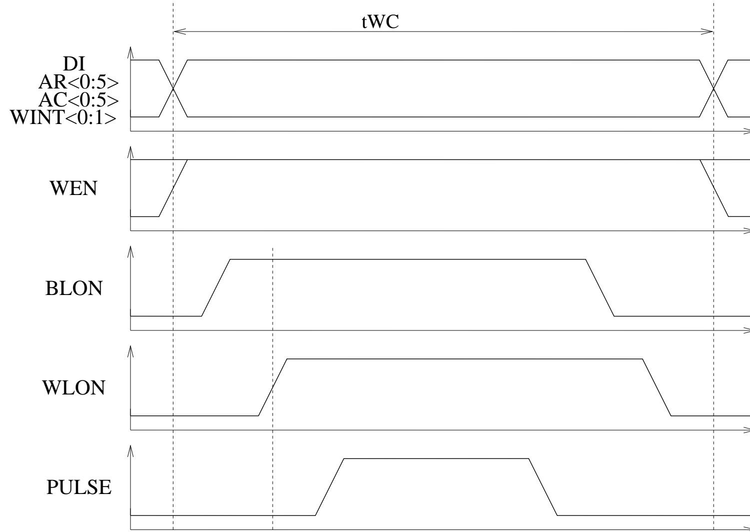


Figure 3.26: Timing diagram for write cycle

Table 3.3: Timing values for a write cycle.

Signal	**Algebraic Constraints**	
	tU	tD
$BLON$	> 0	$\leq tWC$
$WLON$	> 0	$\leq tWC$
$PULSEON$	$> \max(tU_{WLON}, tU_{BLON})$	$< tD_{PULSEOFF}$
$PULSEOFF$	$> tU_{PULSEON}$	$\leq tWC$

- The beginning of the write cycle (the point at which $t = 0$), may occur no sooner than the rising edge of WEN. The time for which WEN must be held high, defines the write cycle time tWC.
- All other signals should realize their values at the beginning of the cycle.
- tU and tD stand for the times at which the signal rises "up" and falls "down."

3.8.3 Standard Read

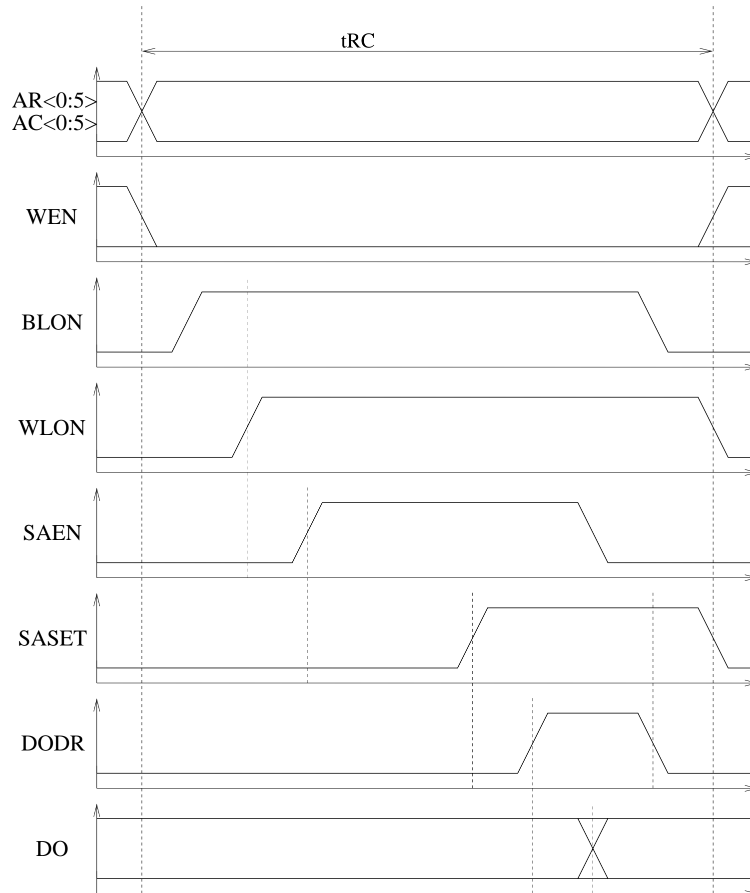


Figure 3.27: Timing diagram for read cycle

All other signals not described for a standard read should be strictly zero.

Table 3.4: Timing values for a read cycle.

Signal	Algebraic Constraints	
	tU	tD
$BLON$	> 0	$\leq tRC$
$WLON$	> 0	$\leq tRC$
$SAEN$	$> \max(tU_{WLON}, tU_{BLON})$	$\leq tRC$
$SASET$	$> tU_{SAEN}$	$\leq tRC$
$DODR$	$> tU_{SASET}$	$< tD_{SASET}$

- The beginning of the read cycle (the point at which $t = 0$), may occur no sooner than the falling edge of WEN. The time for which WEN must be held low, defines the read cycle time tRC, which is nominally $150ns$.
- All other signals should realize their values at the beginning of the cycle.
- tU and tD stand for the times at which the signal rises "up" and falls "down."

3.8.4 TMRBL

Holding the "TMRBL" input high has the ADM address to a fake bitline of cells that have resistors with preset values or chains of MTJs instead of the standard, single MTJ cell. This will allow debugging and demonstration that the support circuitry works. In fact, a partially fabricated wafer only up through the second level of metal will be fully functional on the resistance bitline.

The chosen resistance values in units of Ω are:

$$0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1k, 1.5k, 3k, 5k, 10k, \infty$$

This pattern of sixteen resistances is repeated four times along the resistance bitline.

3.8.5 TMSENSE

Holding the "TMSENSE" input high enables the column decoder on the right side of the array (see Fig. 3.1) to open up a low impedance path from BLT or BLC (see Fig. 3.2) to an external IVSENSE pad via the net SBL. This command can be issued during any mode of operation, and should not be destructive or disruptive because SBL is part of a "four point probe" configuration that does not draw any DC current. This will allow measuring the voltage across the memory cell with no other contributing $I \cdot R$ drops. However, the larger capacitance from the external pads and extra wiring may significantly affect transient operation (slowing it down to the ms-μs time scale).

3.9 Layout Floorplan

Figure 3.28: ADM floor plan for major core circuits

Chapter 4

Testing

After the completion of the design phase, thorough testing of the array was conducted. [1] This ADM functions as a digital memory open to variable write and read conditions. The simplest type of test involves writing the entire array to 0 and subsequently reading the array for a 0, and also writing the entire array to a 1 and subsequently reading the entire contents of the array for a 1.

A wide variety of experiments can be tailored to the specific failure mechanism under investigation. For example, if one wanted to explore whether transient current spikes and stray magnetic field from high bitline currents disturbed (randomly switched) unselected memory cells on the same column as the active cell, one could attack this question in three ways:

1. Checkerboard test: Write a pattern of alternating 1 and 0 values to adjacent cells on a column, then read back the data and verify the observation of a 1 and 0 pattern.

2. At each address, write a 1. Then give several dummy write pulses in which the wordline is not activated (i.e. the cell FET is not turned on to pevent write current) but the bitline is still driven. Next read for a 1. Finally, increment the address and repeat.

[1] Multiple instances of the fabricated 4kb design have been tested. The data reported in this chapter comes exclusively for one test site so that measurements from different experiments can be more effectively compared to each other.

3. Write the whole address space to a 1 before reading back the whole address space for a 1 (checking for failures), write the whole address space to a 0, read the whole address space for a 0 (checking for failures). In this version, the write 1, read 1, write 0, read 0 sequence is the outer loop and the address traversal is the inner loop. Compare the results of this so-called "march" pattern [2] to a non-disturbing test in which the address traversal is the outer loop and each cell experiences the full write 1, read 1, write 0, read 0 sequence. The greater degree of failure in the march pattern will correlate to the amount of disturbance.

This example highlights some of the techniques of memory testing; the goal is to finding a set of conditions that maximize the bit yield, and characterizing the failure mechanisms that prevent 100% yield. The fundamental challenge in this process is seperating failure mechanisms. Did a bit fail because it was disturbed, because not enough write current was used, because it is stuck in one state, or because the address decoder is bad?

At any point in the test flow, the state of the 4kb MRAM array can be compactly recorded in a conductance distribution extracted by repeatedly performing reads on the entire array while incrementing the sense amplifier reference current for each successive read test. Once a clear understanding of the resistance distribution in both states is attained, and if the current version of the hardware has a clear seperation, the sense amplifier reference current can be fixed to an optimum midpoint value. This will enable faster digital experiments in which write tests need to be followed by only one read cycle test to extract the digital contents of the memory. In general, by examining the consistency between write commands and read data, the degree of operational failure can be measured. This information can be used to infer parameters like minimum write time, minimum read time, and distribution of MTJ resistance values. These techniques have been used by [9] and [10] in the development of MRAM.

In a secondary mode of operation, this array can allow individual electrical access to each cell through the external IVSENSE pad. The TMSENSE signal will enable the measurement of the resulting voltage directly across the memory cell because the path through the sense bitline (SBL) requires no DC current (see Figs. 3.1 and 3.2). A common experiment is to

[2] A wide variety of patterns and algorithm time order of growth are described in [1, pp. 158-194] in addition to providing an overview of RAM testing.

sweep the incremental resistance as a function of bias current through the device. To perform a quasi-static dI/dV measurement, a DC bias current plus a small AC current is applied to the cell through a write driver. The ratio of the AC voltage to the AC current will give an incremental resistance measurement for the given bias current. As the bias current is swept past MTJ switching thresholds, a sharp jump in incremental resistance will be observed. These techniques have been used by [12]. Furthermore, dynamic, large signal I-V behavior can be observed through the IVSENSE pad during any mode of operation—with the caveat of significantly larger capacitance than what would be present during normal operation. Therefore, if bit failmaps from the standard digital operation of the array reveal particular bits of interest, this secondary operation will allow full access to a particular sample.

These data will allow one to study process variations across the array, wafer (wich has multiple arrays), and multiple wafers. Additionally, the analysis of the data from MTJ hardware can lead to a circuit designer's black-box abstraction for an STS or TAS MTJ. Essentially, numerical values for current thresholds in both directions, resistance values, and the relationship between write current and dynamic performance can be provided to complete qualitative pictures like Fig. 3.3.

4.1 Test Setup

In order to conduct the experiments described in the previous section, a wide variety of AC and DC signals need to be provided to the memory array with stringent timing requirements. Furthermore, in order to check the data read out and record failures at specific locations, additional hardware and experimental techniques need to be employed—beyond merely providing the correct stimuli to the ADM. To test the the 4kb array, a custom memory tester developed by IBM, the Memory Analysis and Characterization Exerciser (MACE) Tester, was employed. It is essentially custom digital hardware for minimally constrained memory testing. A high-level block diagram can be found in Fig. 4.1.

At the heart of the MACE tester is the pattern sequence, the set of step by step instruc-

tions (pattern lines) that describe various facets of an test. A given pattern line contains information relevant to:

- The address space, and how the different address values are traversed utilizing three 16-bit synchronous counters whose carry-outs may or may not be chosen to feed into each other's carry-ins (in addition to the option of using the cycle clock). Extra counter bits that don't correspond to addresses, can be used as loop variables.
- The data to be written during the current write cycle, or the data to be expected during the current read cycle. Also the data IO circuits are configured in tristate for reading the device under test (DUT).
- Which set of signal timings (i.e. locations of leading and trailing edges of pulses) to use for the current cycle.
- The next pattern line: This could be the same line or a previous line if a loop is taking place; this could be the next line if a loop has terminated and it is time to branch; or this could be an arbitrary line specified by the user.

Additional modules configure the address counting, pulse timings, analog up and down levels for input signals, power supply voltages, data drivers, data receiver comparators and loads, reference voltages, and reference currents. All configurations, including the contents of the pattern memory are programmed by a PC into local memories on the various boards of the tester. This locally configured architecture, with the PC's role isolated from real-time operation, is required for high speed operation in which the DUT must receive all of the appropriate stimuli within its specified operating cycle time. This architecture also makes it very easy to to sweep any variable of interest—timing edges, cycle time, power supplies, and references—between successive iterations of the same test sequence.

In addition to to providing a broad set of stimuli within one cycle of DUT operation, the second fundamental challenge of the memory tester is to read out the data and present it in compact representations to the user. Reading out data is accomplished by comparing the data from the DUT output drivers to the expected data provided in the current pattern line. The result of the comparison (fail = 0 for a match, fail = 1 for a mismatch) is stored in a "Catch RAM" that mimicks the address space of the DUT. During a given test, there

may be several events in which read out data is compared to expected data, but the catch ram has only one bit per address location so the result stored in the catch ram is the "OR" operation of all comparisons during read tests in the experiment. In the semantics of this chapter, a test is the execution of a basic unit of write and read cycles during which the DUT is both stimulated and read. An experiment consists of several tests repeated multiple times, between which the stimuli are altered and data are gathered.

At the end of a test, the contents of the catch RAM are read out in a serial fashion (using the same address counter configuration as the experiment). The serial data stream is then converted by the PC into a two dimensional representation of which bits passed and which bit failed on the memory chip. The arrangement of these fail maps reflect the physical structure of the array design. The fail map often contains too much information to parse, especially for an experiment that runs multiple tests. As a compromise, the result of a fail map readout is often consolidated into a single numerical value which is the sum of all the fails for the given iteration of the given test.

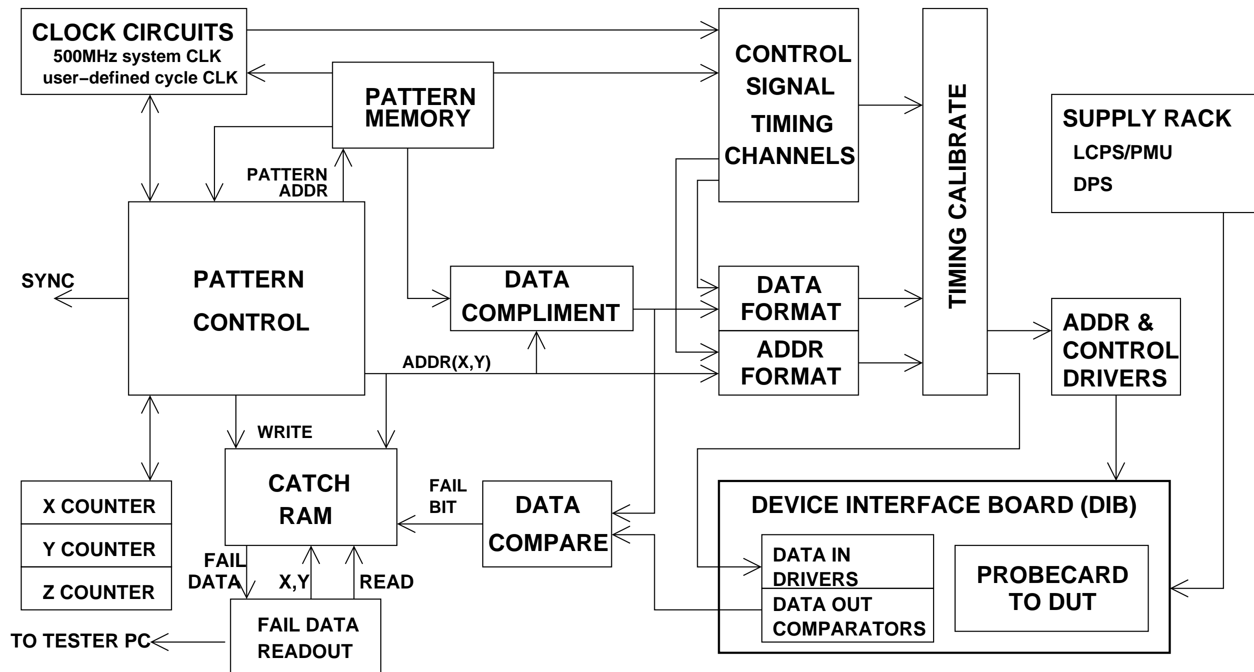


Figure 4.1: Shown is a block diagram of the MACE tester, derived from discussions with [38].

4.2 Experimental Results on Resistance Bitline

The MACE tester was first used to demonstrate functionality of the array circuits on the dummy bitline of resistor cells. The column address was over-ridden with the TMRBL signal and the row addresss was counted from 0 to 63.

4.2.1 Write Pulses

For a fixed column (set by TMRL = 1 to override column decoder) and write voltage (or current) magnitudes:

1. Loop the row address AR$< 5 : 0 >$ from 0 to 63
   (a) write 0 with intensity VFRC$< 0 >$ (or VIREF$< 0 >$)
   (b) write 1 with intensity VFRC$< 1 >$ (or VIREF$< 1 >$)

Shown in Figs. 4.2 and 4.3 are the waveforms observed on the IVSENSE pad for the first 16 row addresses. Each cell exhibits the correct pair of write pulses, for both the 1 and 0 directions (different magnitudes were chosen for 1 and 0 for clarity). In Fig. 4.2, the expected linearly increasing voltage with linearly increasing resistance is observed from current pulses at a fixed amplitude. In Fig. 4.3, sharper edge voltage pulses are seen accross the memory cells. The impedance of the driving voltage source is limited by the column select nfet, so the voltage divider effect is more pronounced for the low resistance cells.

To further exercise the write current levels, an elevated-nonstandard power supply can be used. The waveforms corresponding to this enhanced functionality can be seein in Figs. 4.4-4.5.

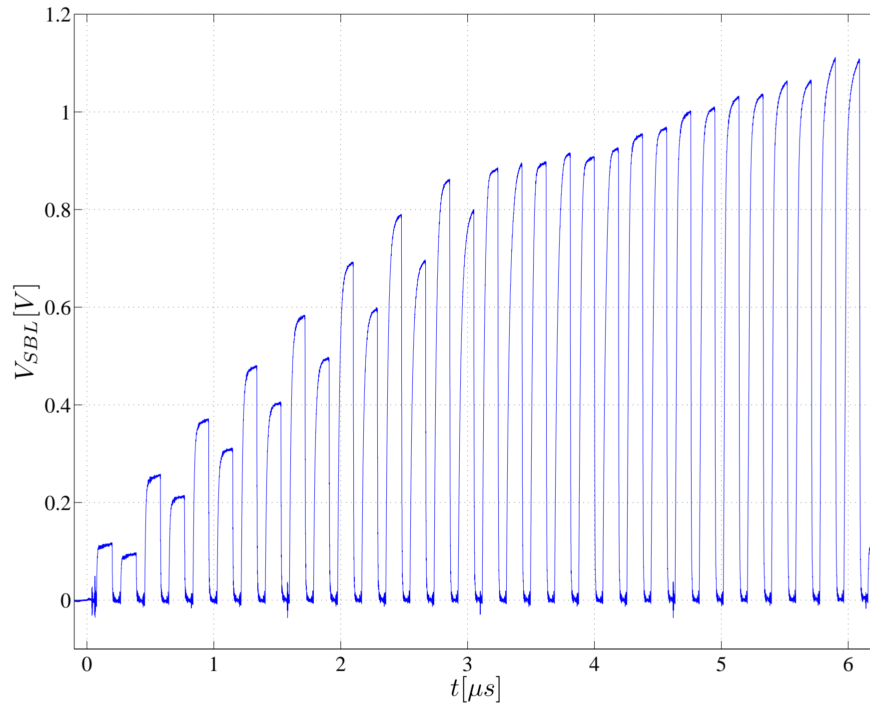


Figure 4.2: Current write pulses at $V_{dd} = 1.8V$

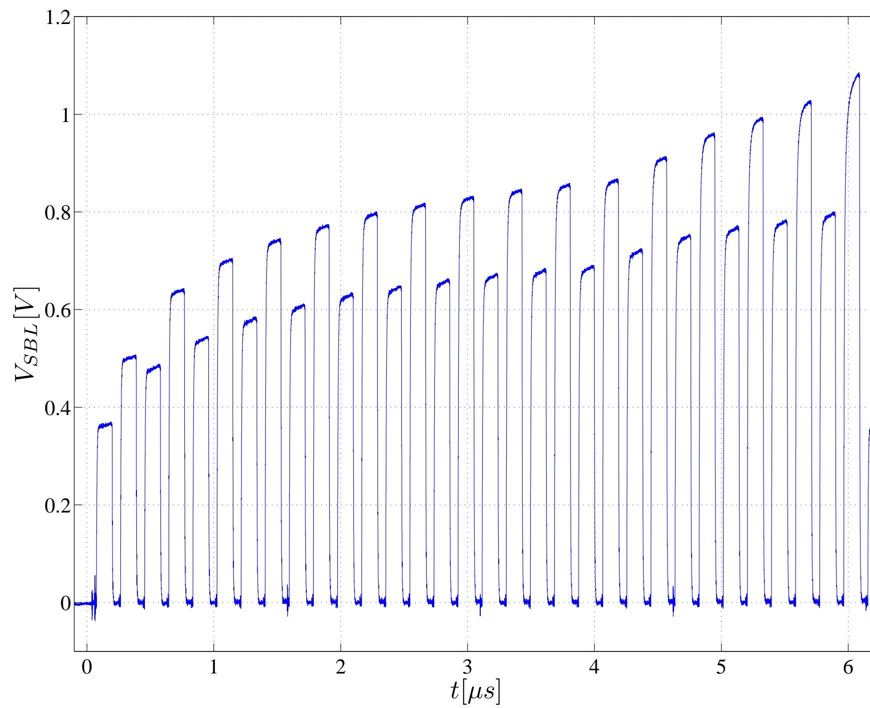


Figure 4.3: Voltage write pulses at $V_{dd} = 1.8V$

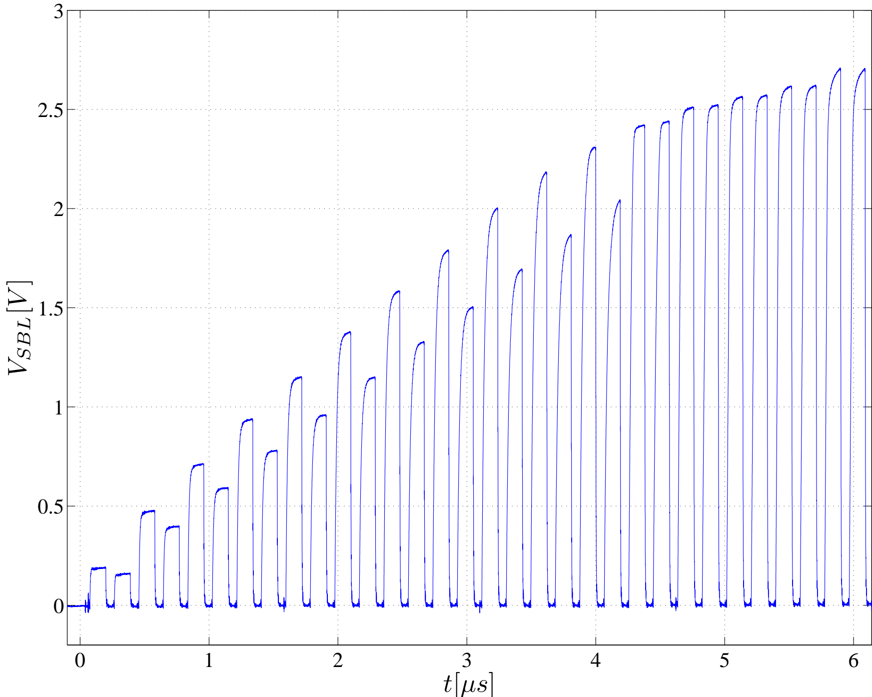


Figure 4.4: Current write pulses at $V_{dd} = 3.6V$

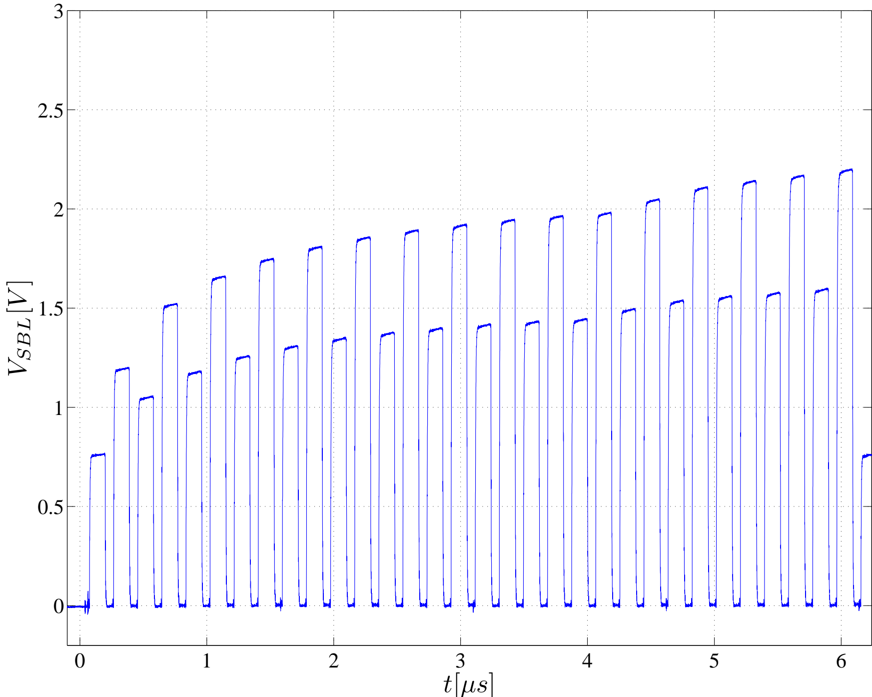


Figure 4.5: Voltage write pulses at $V_{dd} = 3.6V$

4.2.2 Read Pulses

Shown in the following figures are the read pulses over the same sixteen resistance bitline cells for an intermediate level of reference current. One must note that the feedback loop dynamics are significantly altered by connecting the SBL to the IVSENSE Pad. A capacitance of $20pF$ was extracted by driving the probecard trace that leads to the IVSENSE pad with a voltage pulse in series with a precisely measured resistor. The response at the IVSENSE pad (with TMSENSE disabled) corresponded to a first-order exponential rise (initial slope was non-zero, and there was no ringing) so the capacitance was extracted from the RC time constant. Therefore, enabling this IVSENSE pad increases the SBL-MBL capacitance by an additional $20pF$ which is the dominant contribution to the total capacitance.

The waveforms in Figs. 4.6 through 4.13 show that the sense-amplifier successfully clamps a stable, uniform read voltage. Recall that the resistance bitline contains several resistance cells above the specified range of $1k\Omega$ so the large degree of ringing is not unexpected. The ringing can be mitigated by reducing the sense amplifier bias current at the expense of longer read cycle time. Also note that an offset of $5mV$ and $10mV$ is observed. This is due to a finite impedance between the ground node in the core circuits and ground on the tester probe card. The background currents are responsible for producing this deviation.

Shown in the figures are two data sets, one for $50mV$ read voltage and another for $100mV$ read voltage. Although, all experiments have been conducted at $50mV$ of read voltage with no problems reading out conductance distributions, it is expected that reading at $100mV$ will be more robust against pickup and ripple on ground wiring, especially when reading resistances on the threshold between 1 and 0 ($I_{cell} \approx I_{ref}$). Also shown are waveforms of the pulses using a high bandwidth oscilloscope (that has less vertical resolution as a tradeoff). These waveforms can be compared to the expected dynamics of the sense amplifier.

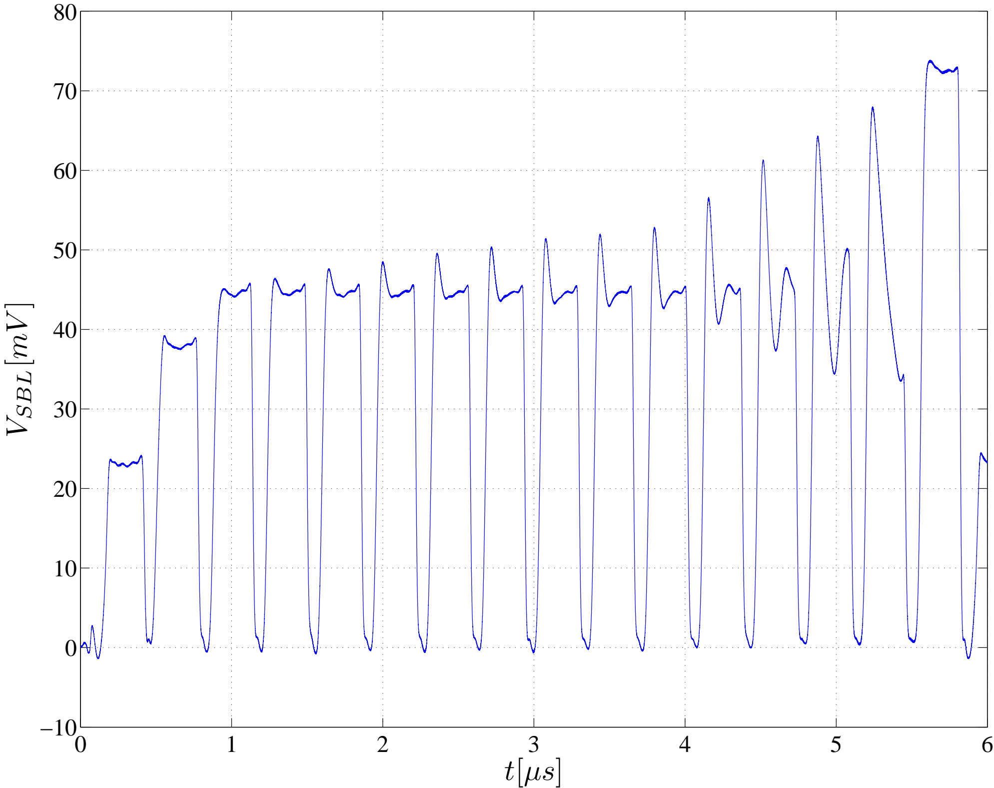


Figure 4.6: $50mV$ read pulses with high resolution scope probe.

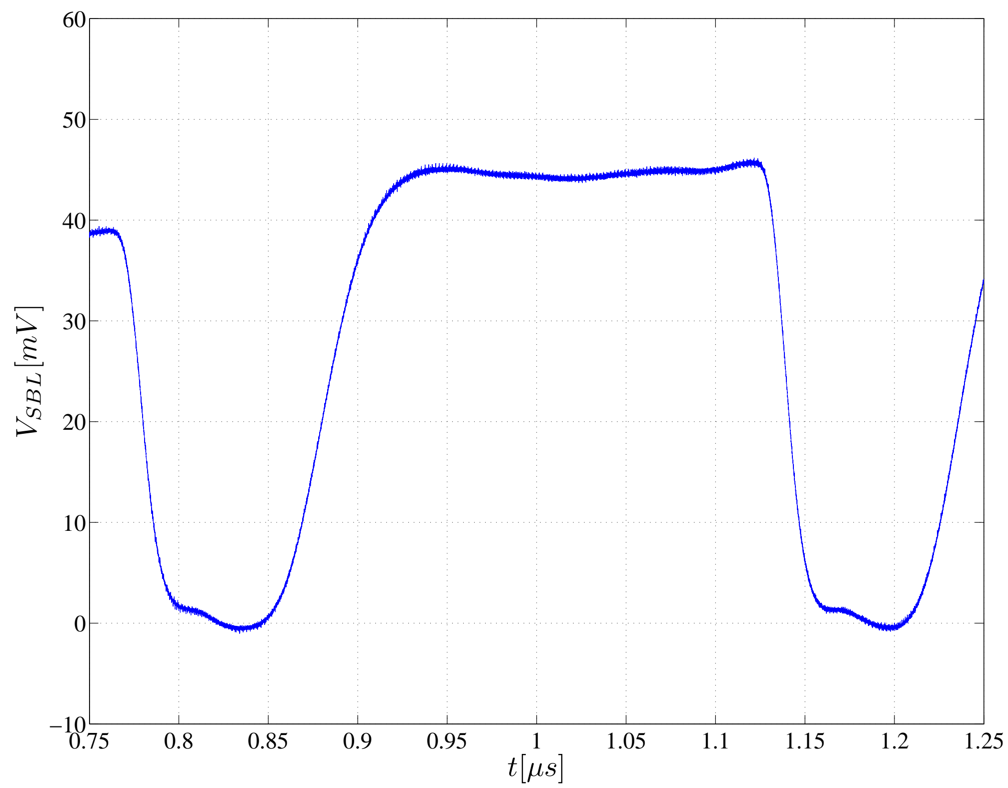


Figure 4.7: Close examination of a $50mV$ read pulses for resistance value in the middle of the expected operating range.

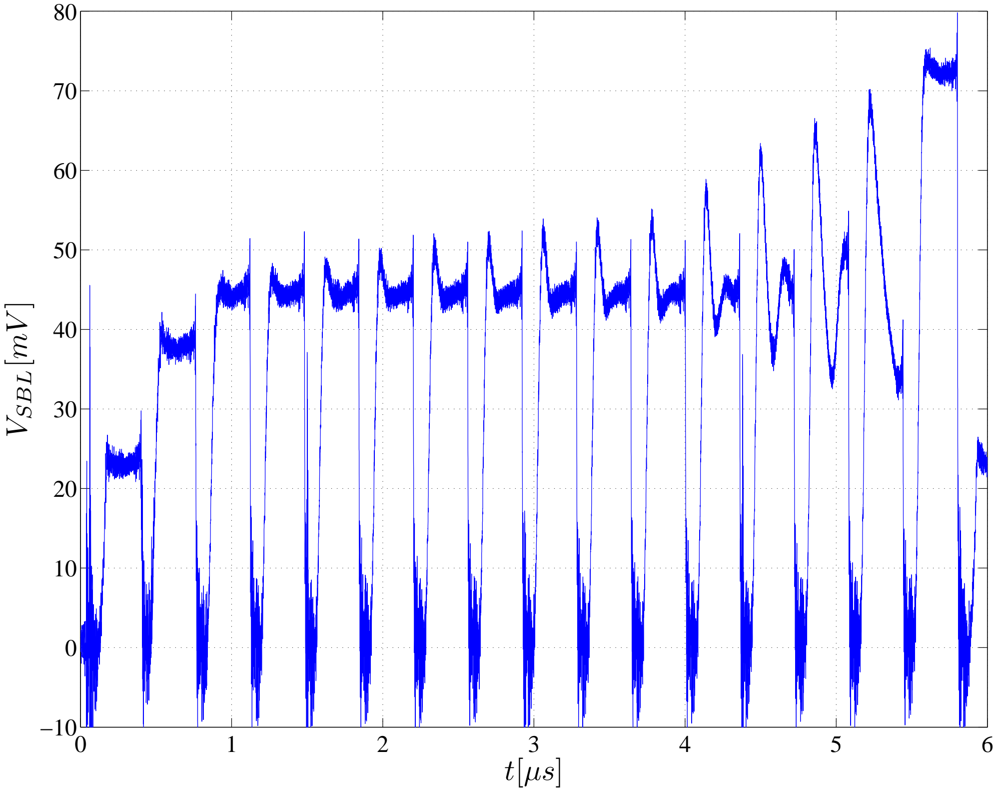


Figure 4.8: $50mV$ read pulses with high bandwidth scope probe.

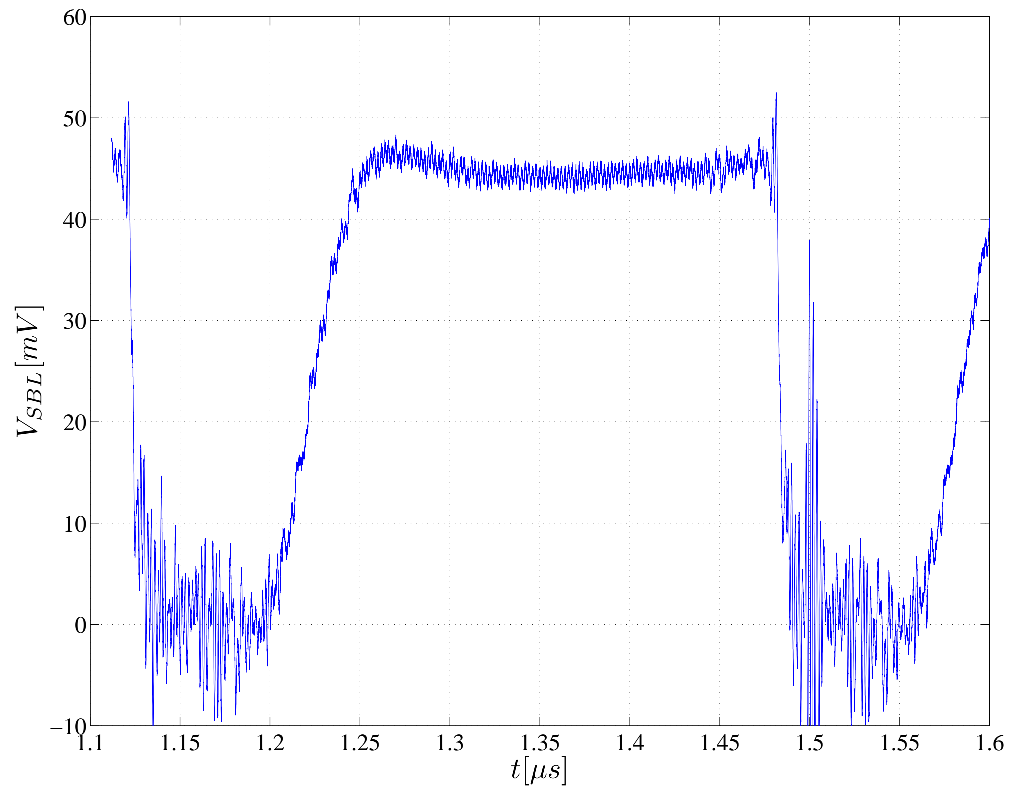


Figure 4.9: Close examination of a $50mV$ read pulses for resistance value in the middle of the expected operating range (high bandwidth scope).

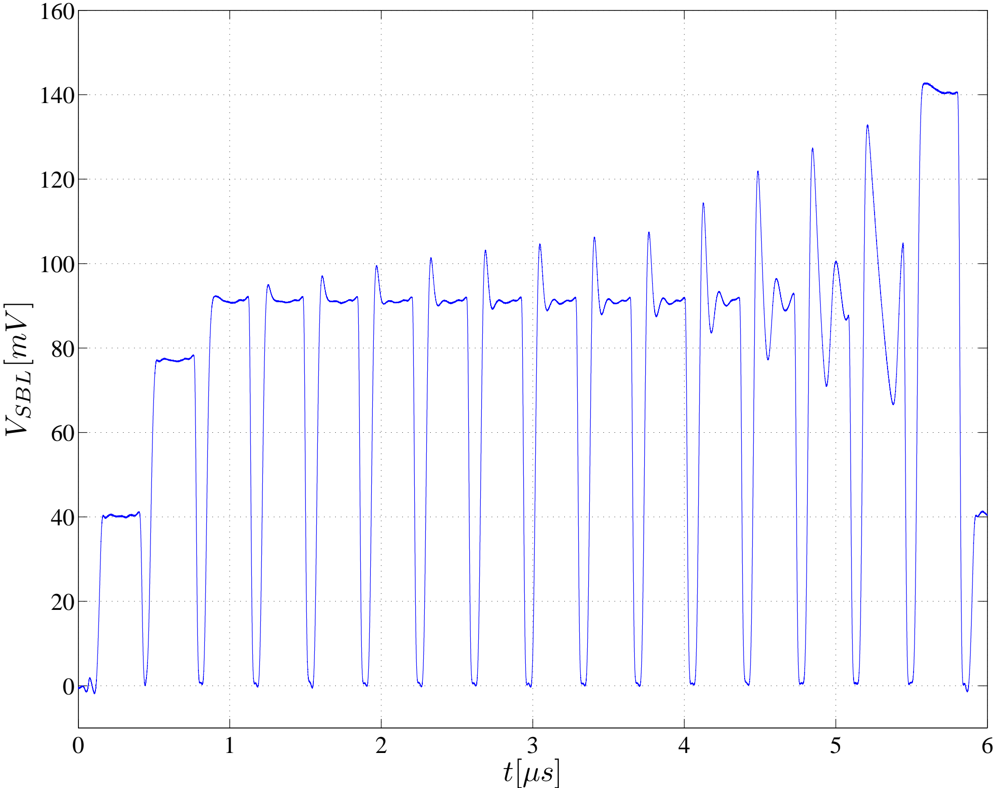


Figure 4.10: $100mV$ read pulses with high resolution scope probe.

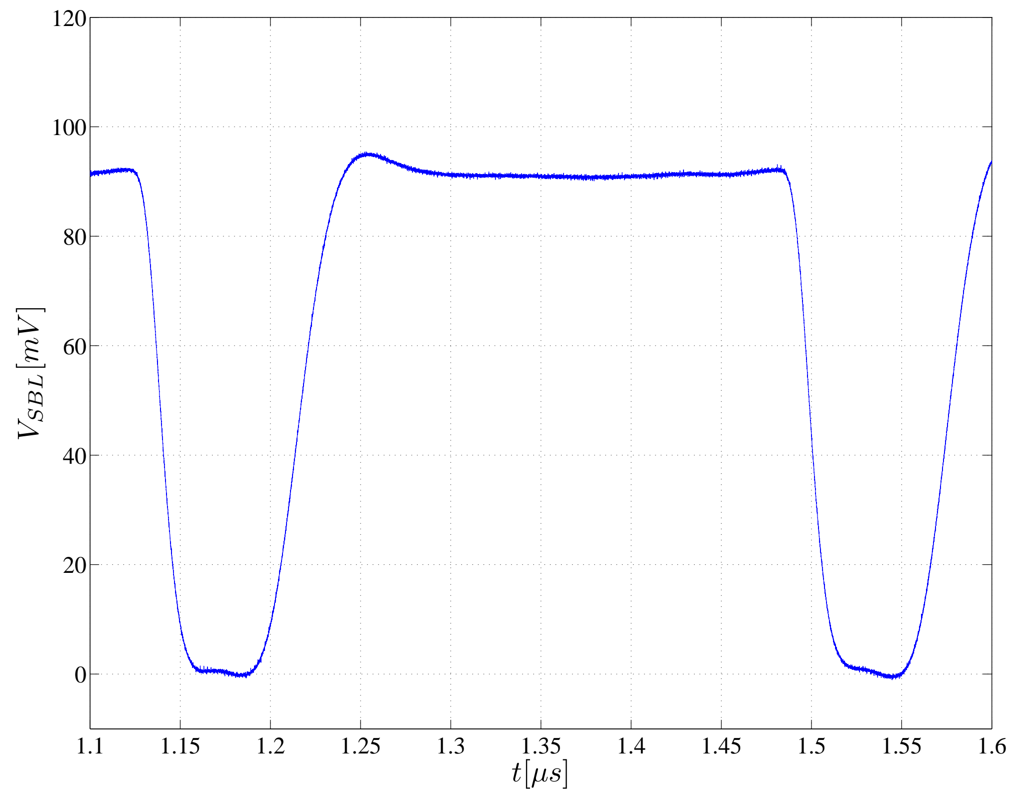


Figure 4.11: Close examination of a $100mV$ read pulses for resistance value in the middle of the expected operating range.

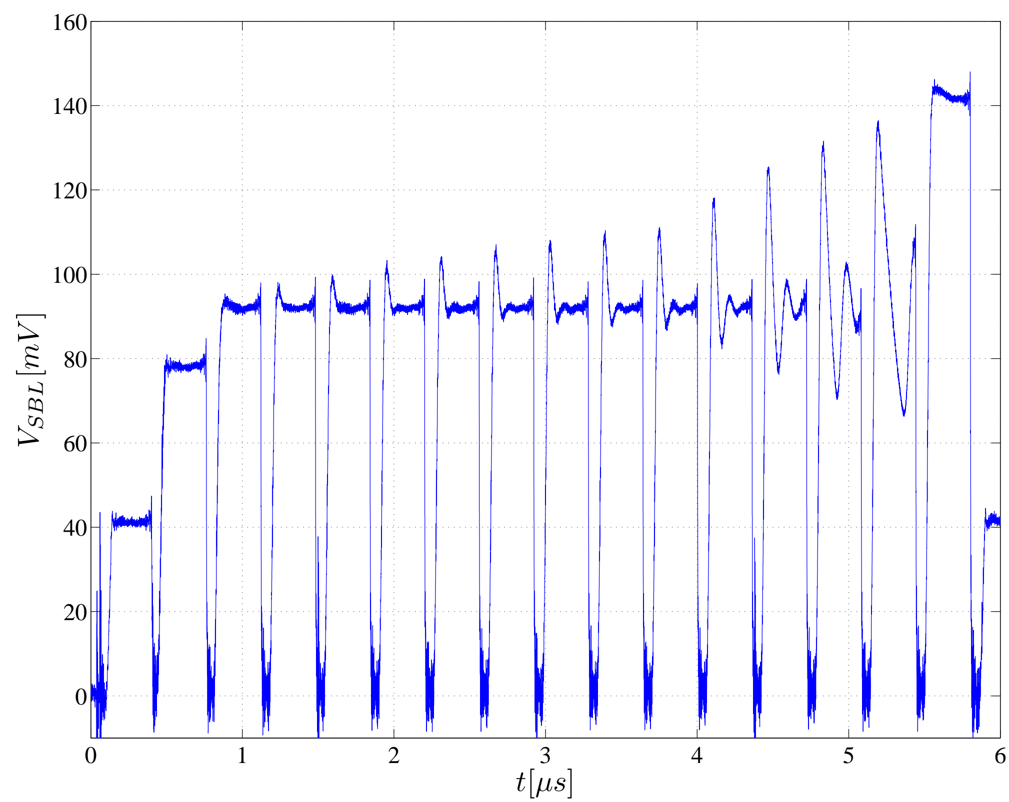


Figure 4.12: $100mV$ read pulses with high bandwidth scope probe.

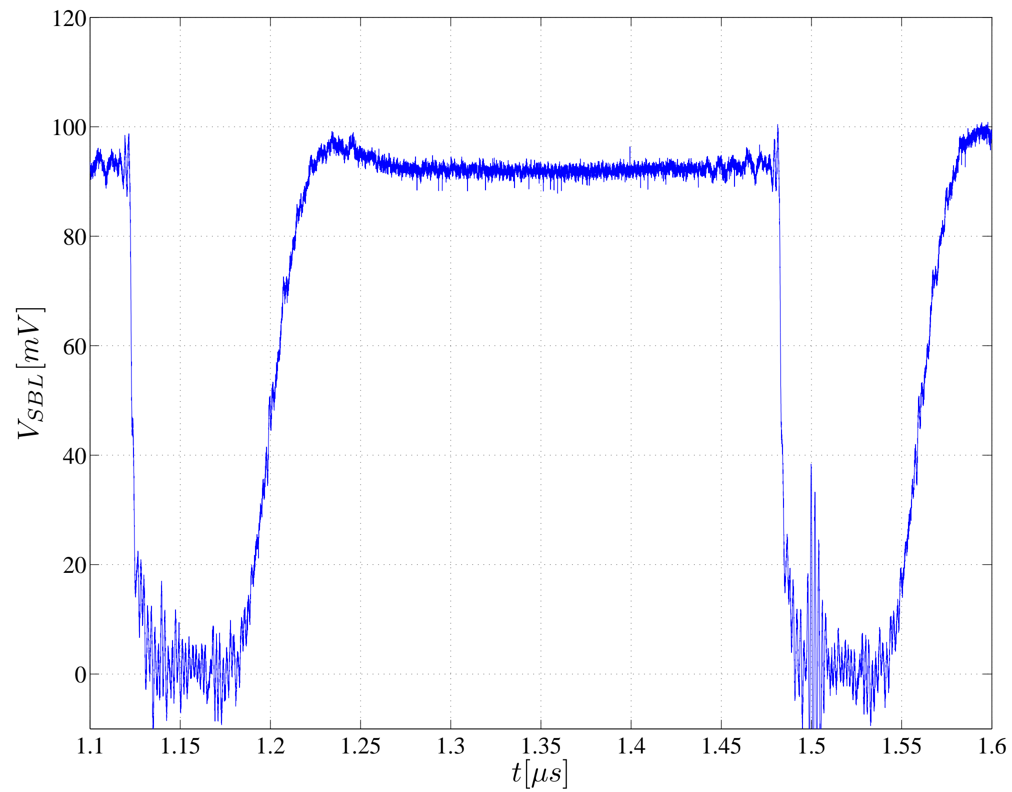


Figure 4.13: Close examination of a $100mV$ read pulses for resistance value in the middle of the expected operating range (high bandwidth scope).

Finally, in Fig. 4.14, one can see that the current mirror load and comparator gain stages are translating the signal current into the correct data value. Shown on the lower plot in the figure is the data output pin, which goes high for the latter segment of the 16 resistance value sequence and goes low when low resistances are again encountered.

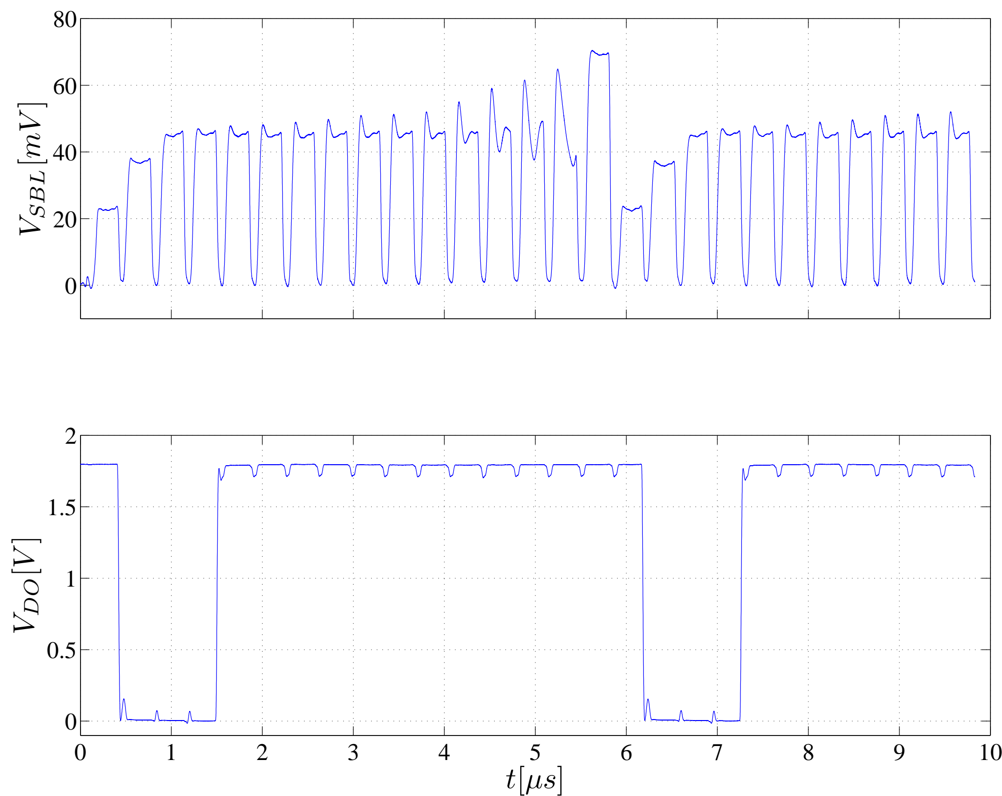


Figure 4.14: The digital output correctly reads the resistance of the memory cell.

4.2.3 Senseamp Reference Sweep

To analyze the data, the distribution of cell conductances is extracted. [3] First, a description of the test pattern for extracting a conductance distribution:

1. For each $I_{ref} \in [0, I_{max}]$

[3]Since the read voltage is fixed as the current is swept, the conductance of the memory cell is directly measured. Of course, the distributions can be mapped to to resistances through $R = 1/G$.

(a) For each address in the array

(b) Read for a 0 (R0): If cell current is greater than I_{ref}, record the fail bit for the current address as 0; else, 1. A failure means that the circuits read a 1 instead of the expect data input of 0.[4]

(c) Add up all the fail bits and record the result with current value of I_{ref}

As this conductance distribution experiment is executed, each individual cell exhibits the failcount plot shown in Fig. 4.15. Namely, there is a value of the sense amp reference current $I_{ref} = V_r G_{cell}$ beyond which the cell is read as a 1 and below which the cell is read as a 0. A 1 is defined as a large resistance, low conductance state.

However, this experiment does not have access *separately* to these individual failcount plots; instead, the superposition of 4096 of these individual failcount plots is measured, as shown in Fig. 4.16. If this test were performed on hardware just received from fabrication, one would expect to see two distinct transition regions seperated by a flat region, indicating a mix of 1s and 0s.

The full array failcount plot is a cumulative distribution from which a histogram of conductance values can be inferred by taking the difference in failcounts between each current step:

$$h(I_k) \equiv \#(\frac{I_k}{V_r} < G_{cell} < \frac{I_{k+1}}{V_r}) = FC(I_{k+1}) - FC(I_k) \tag{4.1}$$

A successful senseamp reference current sweep experiment was conducted on the parametric resistance bitline of 4kb array and the cumulative distribution was recorded. Recall that in 3.8.4 a wide range of resistance was used (repeated here):

$$0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1k, 1.5k, 3k, 5k, 10k, \infty$$

The actual (low bias) resistances were extracted by forcing an external current through the

[4] This test is equivalent to reading the digital data output of each cell, and then summing over all cells, for a given value of I_{ref}. The language of "read 0" is employed because of how the pattern is specified in the tester, which, in turn, enables more intricate test patterns. Such patterns write and read cells multiple times and consolidate the data comparison to a single fail bit that indicates success (fail bit = 0) or failure (fail bit = 1).

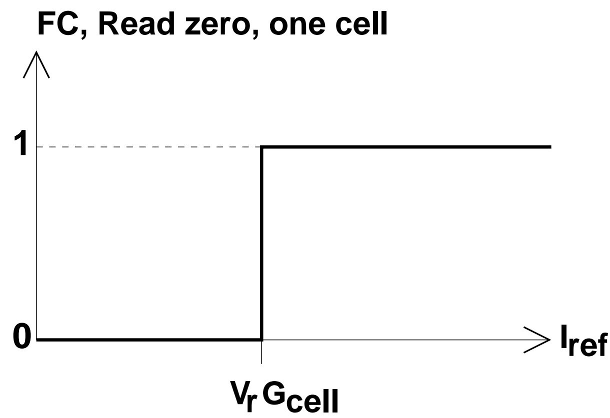


Figure 4.15: Read 0 failcount plot for a single cell.

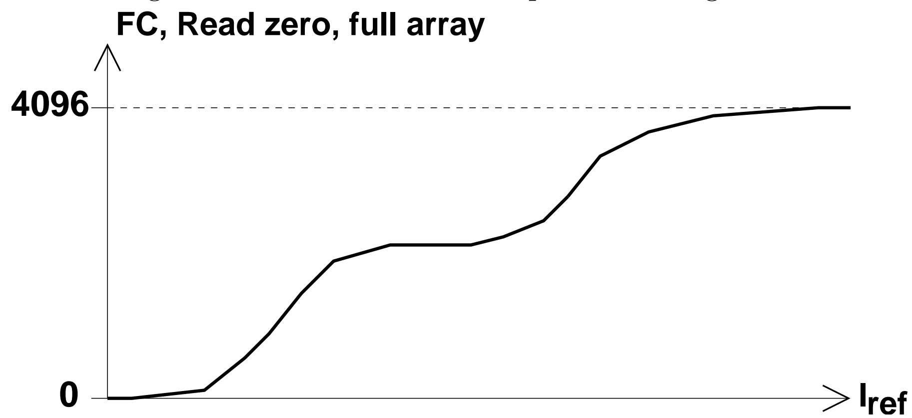


Figure 4.16: Read 0 failcount plot for full array. Two distinct transition regions suggest the presence of bits in two different resistance states.

VFORCE< 0 > path and measuring the DC voltage on the IVSENSE pad with a multimeter:

$$98, 230, 330, 430, 530, 630, 730, 830, 930, 1.03k, 1.13k, 1.77k, 3.15k, 5.09k, 9.89k, \infty$$

As seen from the 98Ω measurement, the low bias resistance of the FET in series with the grounding FETs on both sides of the array at the column decoders is 100Ω. This value corresponds to a zero resistance value for the memory element. The polysilicon resistors have some offset and variation unique to this site.

These measured resistor values can be compared with the results from the reference current sweep. In Fig. 4.17 a cumulative distribution is shown, but the horizontal axis was mapped from I_{ref} (which is the true independent variable) to V_{read}/I_{ref} so that the results can be readily compared. It can be seen that the 16 distinct resistance values are detected. Note this sweep is on the full 64 bit long column so intermediate values occur in the transitio nregions due to mismatch between the four replications of the 16 resistance pattern. The transition regions are slightly offset to the positive direction, this mostly stems from $-5mV$ offset in the read voltage observed on the read pulses in Sec. 4.2.2 effectively boosting the resistance value by 10%. There are also three other issues relating to ground shift from quiescent current of the core circuits, out of specification resistors above $1k\Omega$ being read, and the limited precision of the input reference from the tester. All things considered, this resistance distribution demonstrates the ability to clearly extract a monotonically increasing failcount over the specified resistance range of $100 - 1k$ which corresponds to $200 - 1.1k$ with the cell FET resistance included.

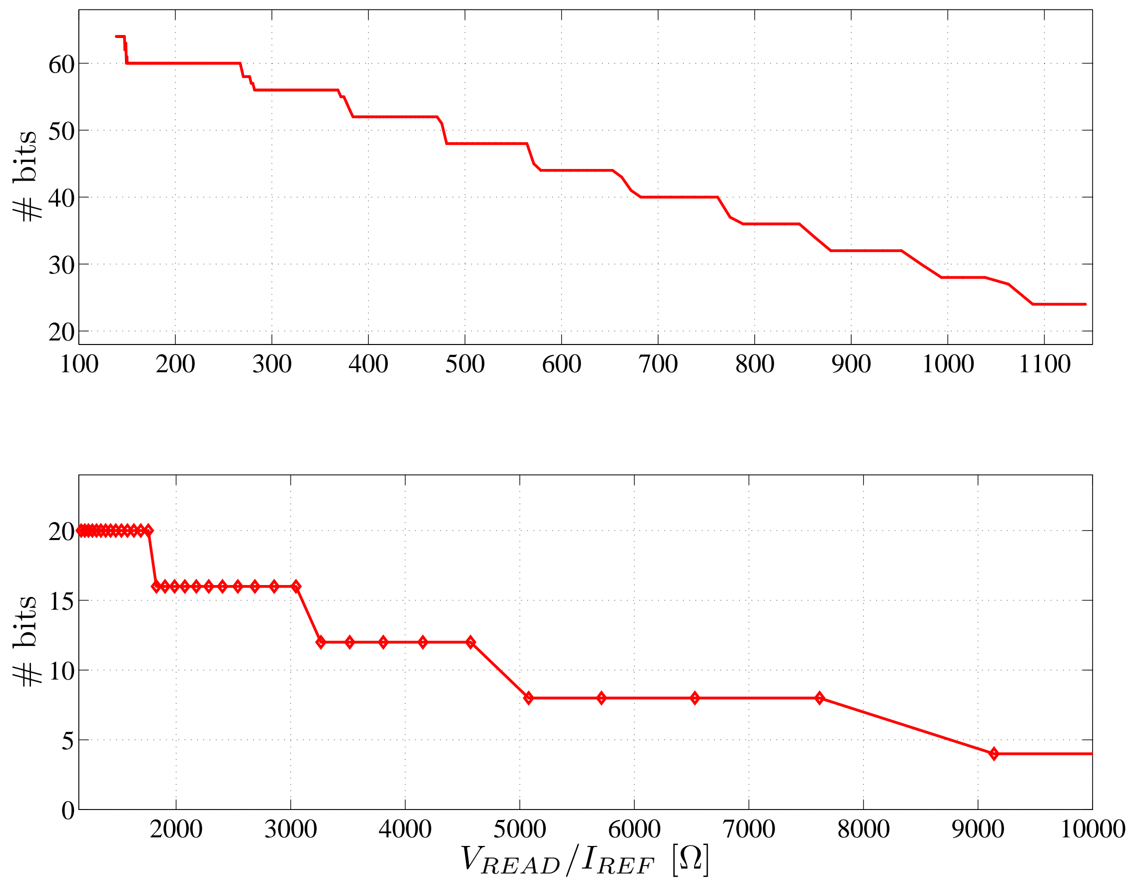


Figure 4.17: Successful extraction of resistance distribution on RBL.

4.3 Description of Initial Test Plan

Having verified functionality of the array circuits on the resistance bitline, the test plan presented in this section was developed to find a "sweet spot" of the write and read conditions that yield the highest bit functionality when first receiving a wafer with MTJ hardware. Presented here is an overview of the approach.

The important parameters are:

Parameter	Description
$V1$	write 1 voltage magnitude
$I1$	write 1 current magnitude
$T1$	write 1 pulse duration
$V0$	write 0 voltage magnitude
$I0$	write 0 current magnitude
$T0$	write 0 pulse duration
V_r	read voltage
I_{ref}	SA reference current

At first pass, only $V1$, $V0$, and I_{ref} will be swept. The pulse duration will be fixed to $T1 = T0 \approx 450ns$; current-mode writing ($I1$, $I0$) will not employed; and the read voltage will be fixed to $V_r = 50mV$. Another key challenge is to not break down the devices, in this vein the low read voltage was selected and low write pulse magnitudes will be attempted first.

In principle, it is possible to gather all the data from which the desired information can be extracted by parking at one address and then conducting detailed sweeps of the three variables one one MTJ at a time. However, this is prohibitively cumbersome in both the gathering and parsing of data for 4096 cells. To overcome this complexity, high-speed test patterns apply read and write cycles throughout the whole address space during the sweep of the three variables.

Having established the technique for extracting a conductance distribution in section 4.2.3, the effect of applying write pulses can now be observed as shifts in this distribution. The following test pattern will be applied for various values of both $V1$ and $V0$ to find the smallest

possible magnitudes that switch the maximum number of bits:

1. For each $I_{ref} \in [0, I_{max}]$
   (a) For each address in the array
      i. Write a 1 (W1): Pulse STS current in the appropriate direction through the cell, applying a voltage $V1$ across it for a duration of $T1$.
      ii. Read for a 0 (R0): If cell current is greater than I_{ref}, record the fail bit for the current address as 0; else, 1.
   (b) Add up all the fail bits and record the result along with the current value of I_{ref}

The above pattern will extract the $FC_{W1R0}(I)$ distribution, and a similar test pattern will be applied for write 0 (W0) to extract a $FC_{W0R0}(I)$ distribution. Note how a read 0 (R0) command is executed regardless of the write data in order to extract the resistance distributions in a consistent manner.

If some of the bits are switching for the specific values of $V1$ and $V0$ chosen, then two shifted FC curves should be observed as in Fig. 4.19. A measure of yielding bits YC—the number of bits that switch in both directions *and* are readable by the current value of I_{ref}—can be extracted:

$$YC(I) \equiv \#(V_r G_1 < I < V_r G_0) = FC_{W1R0}(I) - FC_{W0R0}(I) \tag{4.2}$$

Fig. 4.19 also shows as $V0$ is swept upwards from 0, $YC(I_{ref})$ will tend to increase at larger values of I_{ref}. Similarly, as $V1$ is swept upwards from 0, $YC(I_{ref})$ will tend to increase at smaller values of I_{ref}. The fundamental goal is to find a set of conditions such that $YC = 4096$.

Deviations from this maximum can result from either non-switching bits or the existence of multiple good bits whose read current windows (RCWs) do not overlap. The RCW is defined as $V_r G_0 - V_r G_1$ and is shown in Fig. 4.18, which plots the failcount for W1R0 and W0R0 of a *single* bit. In this formalism, a defective bit (e.g. open or short) has $RCW = 0$, or equivalently $G1 = G0$. If the working bits have a reasonably controlled average read current

window, $< RCW >$, in terms of the *separation* in I_{ref} between the two states (regardless of absolute location in I_{ref}), then the number of working bits can be estimated as follows:

$$\#(\text{good bits}) = \frac{1}{< RCW >} \int_0^\infty YC(I)dI \tag{4.3}$$

$< RCW >$ can be estimated from $h_{W1R0}(I)$ and $h_{W0R0}(I)$ as appropriately defined by Eq. 4.1:

$$< RCW >= \mathrm{argmax}(\mathrm{h_{W0R0}(I)}) - \mathrm{argmax}(\mathrm{h_{W1R0}(I)}) \tag{4.4}$$

Finally, once this test is conducted and a concrete sense of the resistance distributions of the two distinct states is achieved, the read reference can be fixed to an optimum midpoint. This will free up testing time and data processing resources to sweep additional variables of interest.

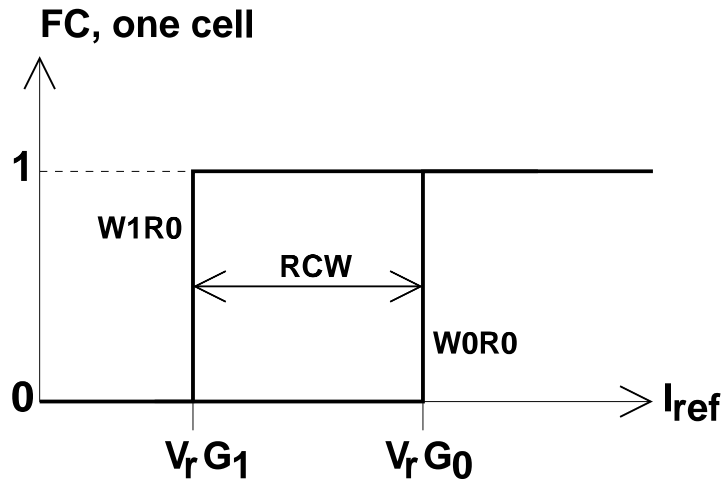


Figure 4.18: Single cell failcount curves for both W1R0 and W0R0 test patterns.

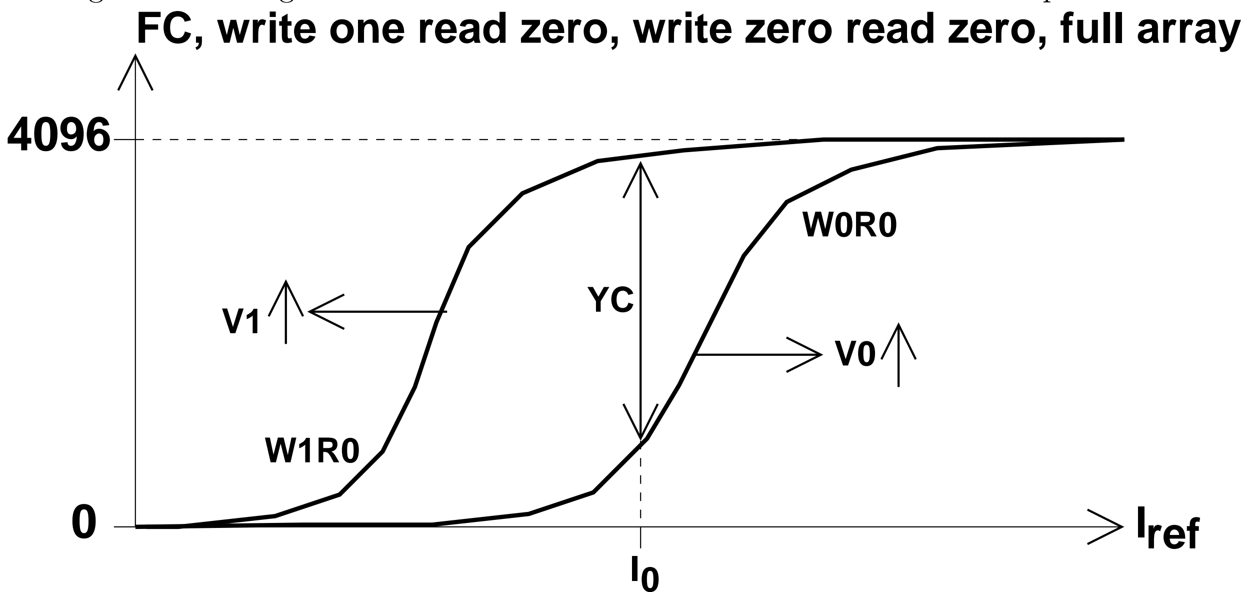


Figure 4.19: Full array failcount curves for both W1R0 and W0R0 test patterns.

4.4 Results from Field Switching

The methodology outlined in the previous section was applied to a first iteration of integrated magnetic hardware. The results here are based on utilization of the magnet wire to switch the bits in the reduced magnet wire address space in a blanket fashion. This will allow an understanding of the field characteristics of the MTJs in addition to the nature of the resistance distribution.

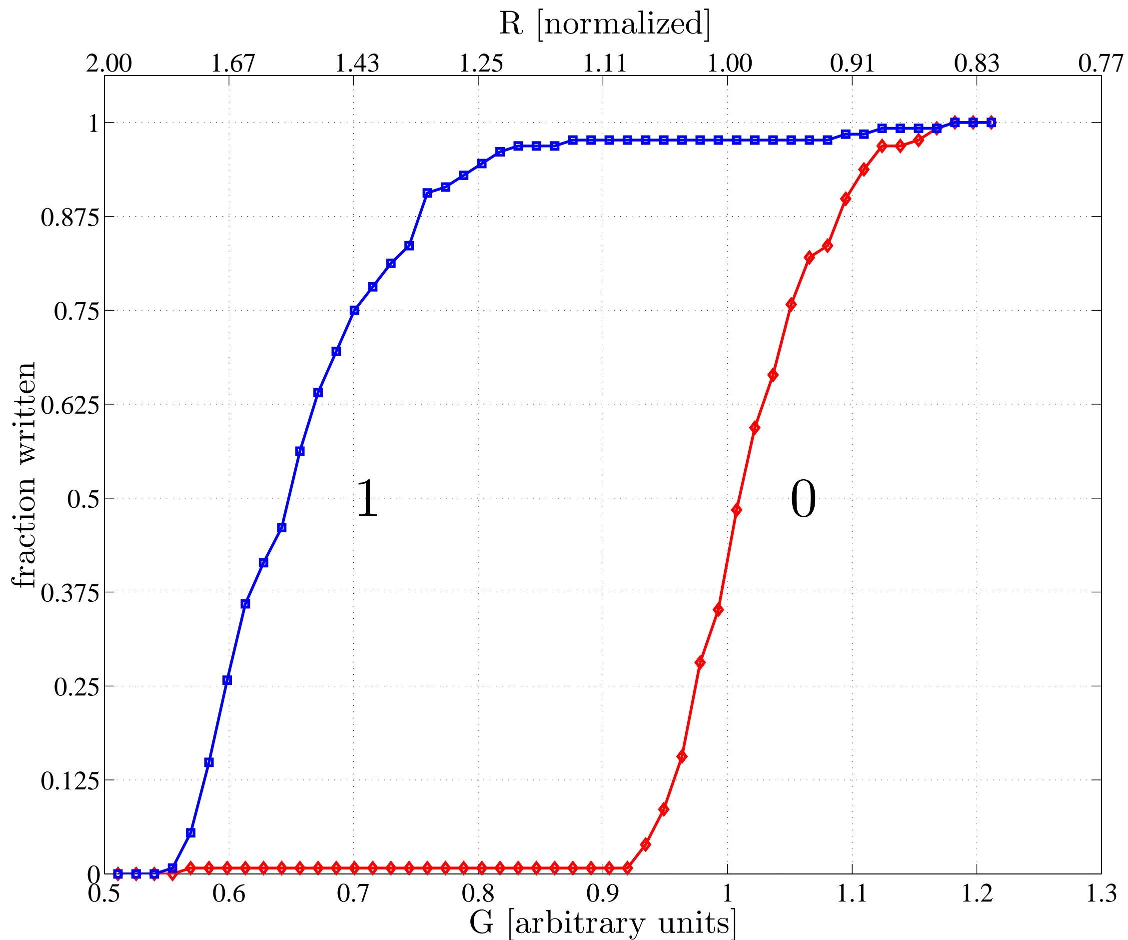


Figure 4.20: Extraction of cumulative distributions in conductance after applying magnet wire pulses in the "1" direction and then in the "0" direction.

Fig. 4.20 was obtained by writing all the bits to 0 with maximum field intensity and then extracting a conductance distribution by running a read reference sweep experiment.

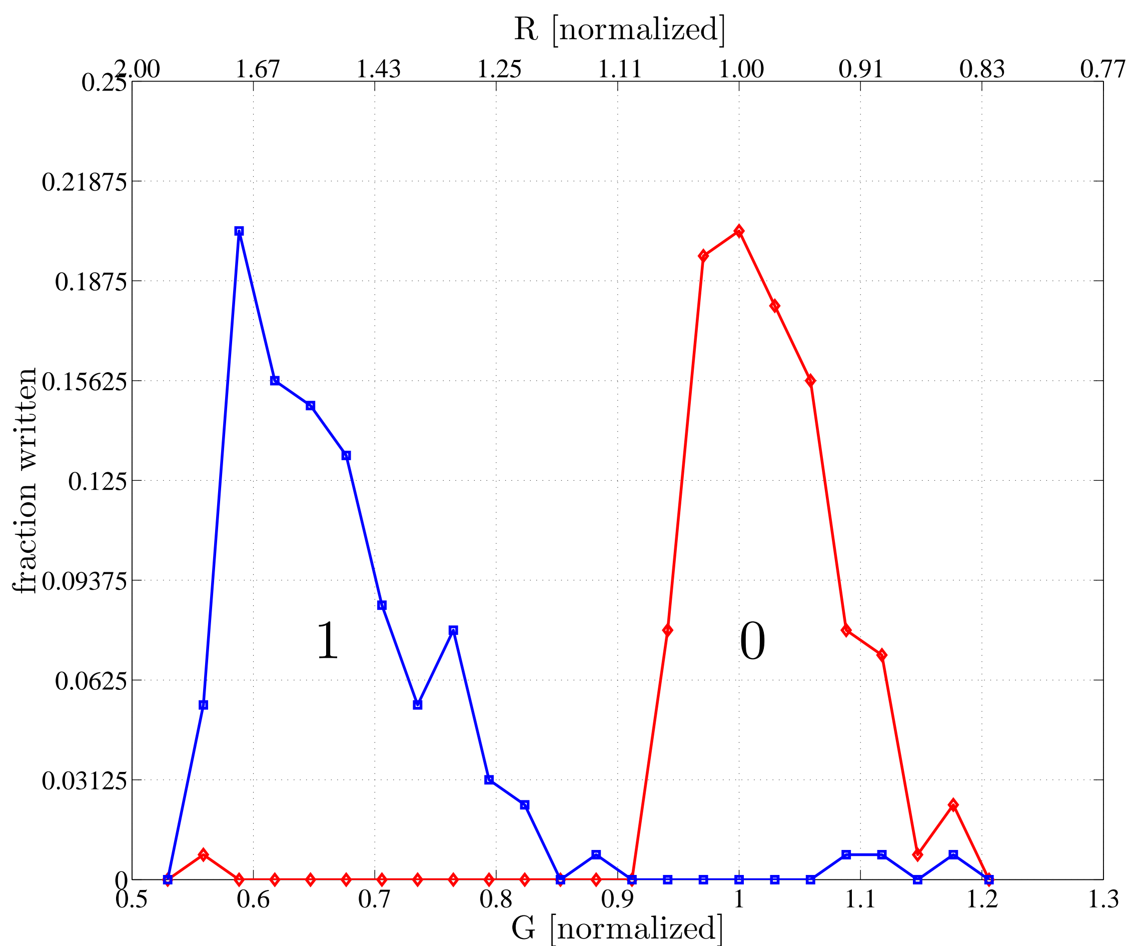


Figure 4.21: The corresponding histograms after W1 and W0 from the cumulative distributions in Fig. 4.20

Subsequently, all the bits were written to 1 with maximum field intensity and then another conductance distribution was extracted.

In addtion to revealing the nature of 1 and 0 distributions, the histograms in Fig. 4.21 also suggest an optimal read reference current. This variable can be fixed allowing the sweep of other variables of interest. In fact, the optimum reference I_{ref} = argmax(YC(I)) was extracted from the data in Fig. 4.21 and used to perform a sweep of the magnitude of write field for both the 1 and 0 directions.

Shown in Fig. 4.22 is a result of the following experiment conducted at a fixed I_{ref} on the reduced address space of three magnet wire enabled bitlines:

1. For each $V_1 \in [0, V_{max}]$
   (a) For each $V_0 \in [0, V_{max}]$
      i. Write a (W1): Apply a pulse of magnetic field by pulsing current on the magnet wire in the 1's direction on the entire magnet wrie address space.
      ii. Read 1 (R1): Read all the bits on the magnet wire address space. For each memory cell, record a pass if the data output is a 1.
      iii. Write a 0 (W0): Apply a pulse of magnetic field by pulsing current on the magnet wire in the 0's direction on the entire magnet wrie address space.
      iv. Read 0 (R0): Read all the bits on the magnet wire address space. For each memory cell, record a pass if the data output is a 0.
      v. Sum the total number of fails in the Catch RAM and record the value along with the current values of (V_0, V_1).

The data in Fig. 4.22 allow the observation and verification of several important features:

- At low values of either Vmag1 and Vmag0, all bits fail and at high values all most all bits pass (there is a floor of 4% in the lightly shaded region).
- It is easier to write a 1 than it is to write a 0. This indicates the presence of an offset field in the free layer due to the stack of magnetic materials beneath it. The wider transition region for one switching direction also reflects the fact that the two points along which the astroid boundary is crossed are not symmetric (see Fig. 2.7).
- The anisotropy field term H_k (see Eq. 2.30) can be estimated by translating V_{mag} at the switching threshold back into current values. And then translating the current back into magnetic field with Eq. 3.1.

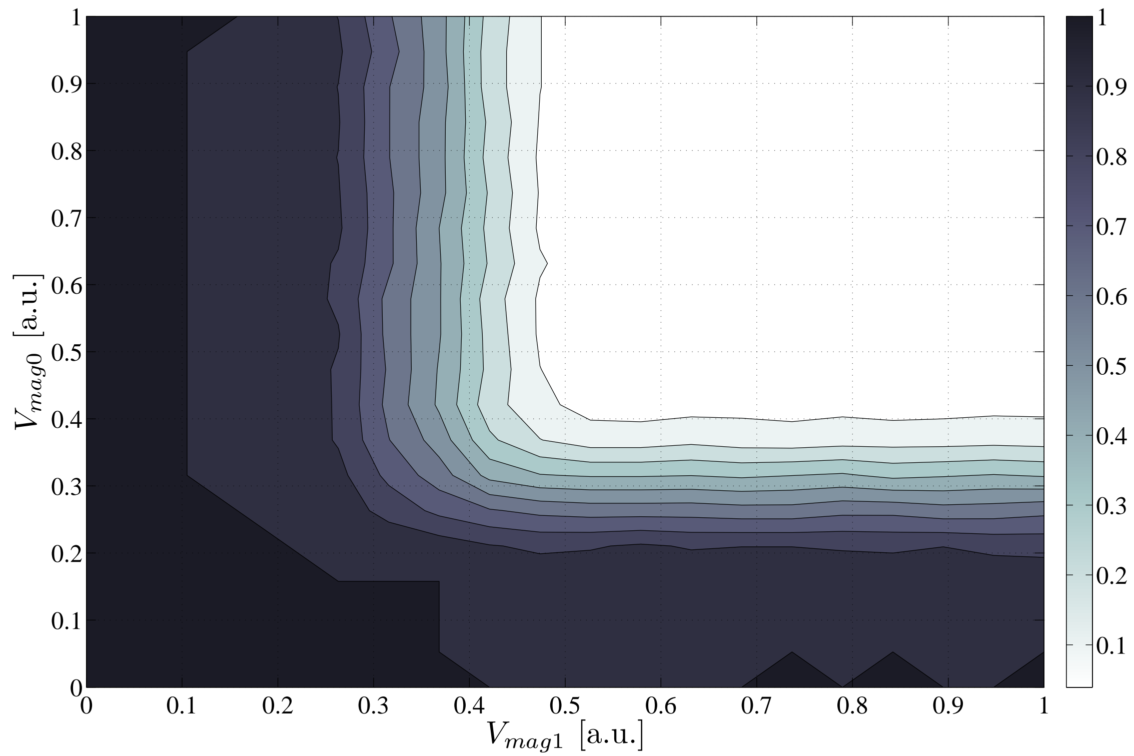


Figure 4.22: A two dimensional sweep of switching field strength in both directions. The shading correspond to the fraction of bits that fail to switch in both directions at a given value of (V_{mag1}, V_{mag0}).

Chapter 5

Conclusion

The goal of producing a 4kb array and a test environment to conduct experiments on less-understood mechanisms of MRAM based on perpendicular switching current has been achieved.

5.1 Summary of Contributions

The specific contributions made in this work are:

- A 4kb array design with digital circuits for data, control, and address decoding signals.
- A flexible suite of voltage, current, and magnetic field write drivers.
- The ability to observe real time electrical behavior of MTJs in the TMSENSE mode.
- A higher precision sense-amplifier that supports a broad range of low resistance MTJs.
- A test methodology for exercising MTJs switched with current through them in a conservative manner to avoid damage in the process of characterization.
- Initial data that demonstrates the existence of a a field hysteresis and attractive resistance change in experimental MTJs—in addition to some statistical information on endurance.

The 4kb array is poised to continue testing future iterations of magnetic hardware to find working spin-transfer memory elements and then extract their properties in detail.

5.2 Future Work

To make a viable MRAM product, the MTJs in a given array must all fit within a common window of write pulse duration, write current magnitude, and resistance seperation from a locally defined midpoint at a success rate of roughly 99.999% to 99.9999%.

In order to quantify the deviation of present Spin-MRAM technology from this goal, array data on spin transfer MTJs is essential. Also, Spin-MRAM brings with it unique considerations apart from process variation related to the probabilistic nature of write operations and disturbances during read operations—all on a single bit. Detailed measurements of switching probability at various pulse magnitudes and durations are required to refine the theoretical estimates in Eq. 2.45.

For the circuit design, it will be a challenge to develop high speed sense amplifiers for MTJ resistance values that are comparable to paracitic FET resistances in the read path.

By developing Spin-MRAM technology based on switching the MTJ with bidirectional current through the memory cell, the scalling problems for MRAM can be addressed.

Appendix A

Electromagnetics Reference

A.1 Maxwell's Equations

Maxwell's equations (CGS) units:

$$\vec{\nabla} \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t} \tag{A.1}$$

$$\vec{\nabla} \times \vec{B} = \frac{1}{c}\frac{\partial \vec{E}}{\partial t} + \frac{4\pi}{c}\vec{J} \tag{A.2}$$

$$\vec{\nabla} \cdot \vec{E} = 4\pi\rho \tag{A.3}$$

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{A.4}$$

Relation between $\vec{H}$ and $\vec{B}$:

$$\vec{B} = \vec{H} + 4\pi\vec{M} \tag{A.5}$$

$$\vec{B} = \mu\vec{H} \text{ (linear media only)} \tag{A.6}$$

A.2 Derivation of Spin-Transfer Switching Dynamics for a Mono-domain model

The goal of this section is to work through in full detail the steps from Eq. 1 to Eq. 11 in [27] and articulate the relevance of the linearized small angle dynamical equation to the MRAM circuit designer.

Equation Eq. 2.35 is a special case (that discards the component of the field parallel to $\vec{m}$) of a generalized definition of magnetic field: [29], [31]

$$\vec{H} = -\left[\frac{\partial U}{\partial m_x}\hat{x} + \frac{\partial U}{\partial m_y}\hat{y} + \frac{\partial U}{\partial m_z}\hat{z}\right] \tag{A.7}$$

First, the definition on $U_m(\theta, \phi)$ in Eq. 2.32 is translated back into $U_m(m_x, m_z)$:

$$U_m = \frac{H_k}{2m}\left(h_p m_x^2 - m_z^2 - 2mm_z h_{easy}\right)$$

Note, (1) $\sin^2(\theta)$ was replaced with $-\cos^2\theta$ (the additional constant of 1 was discarded because it will not contribute to the gradient of U_m), (2) it is assumed the applied field $\vec{H}_{ext} = H_{easy}\hat{z}$ is purely along the easy axis, and (3) the energy barrier K was replaced with its defining expression $(1/2)mH_k$. Applying the equation for magnetic field on $U_m(m_x, m_z)$ gives:

$$\begin{aligned} \vec{H} &= H_k\left[-h_p\frac{m_x}{m}\hat{x} + (\frac{m_z}{m} + h_{easy})\hat{z}\right] \\ \vec{H} &= H_k\left[-h_p\cos\beta\hat{x} + (\cos\theta + h_{easy})\hat{z}\right] \end{aligned} \tag{A.8}$$

where θ is the angular departure from $+\hat{z}$ and β is the angular departure of the magnetic moment $\vec{m}$ from $+\hat{x}$ ($\cos\beta = \sin\theta\cos\phi$). Now the full dynamical equation for a monodomain nanomagnet with spin polarized current given in Eq. 2.42 (and repeated here) as:

$$\frac{1}{\gamma}\frac{d\vec{m}}{dt} = \vec{m}\times(\vec{H}_{ext} + \vec{H}_{eff}) + \eta\frac{I}{|e|}\frac{\mu_B}{\gamma}\ \hat{m}\times(\hat{n}_1\times\hat{m}) - \frac{\alpha}{\gamma m}\vec{m}\times\frac{d\vec{m}}{dt}$$

can be rewritten (assuming the spin current is polarized along $\hat{n}_1 = -\hat{z}$ and invoking $\mu_B/\gamma = -\frac{\hbar}{2}$) as

$$\frac{1}{\gamma H_k}\frac{d\hat{m}}{dt} = (\cos\theta + h_{easy})\hat{m}\times\hat{z} - (h_p\cos\beta)\hat{m}\times\hat{x} + h_s\hat{m}\times(\hat{z}\times\hat{m}) - \frac{\alpha}{\gamma H_k}\hat{m}\times\frac{d\hat{m}}{dt} \quad (A.9)$$

where

$$h_s = \eta\frac{I}{|e|}\frac{\hbar}{2}\frac{1}{mH_k}$$

is the spin torque magnitude in normalized magnetic field units ($h_s > 0$ will align $\hat{m}$ with $-\hat{z}$ and $h_s < 0$ will align $\hat{m}$ with $+\hat{z}$). Also note that the equation is written in terms of $\hat{m} = \vec{m}/m$ as a result of dividing both sides by mH_k.

The damping term proportional to α in Eq. A.9 makes this an implicit equation for $d\hat{m}/dt$. First take the cross product of $\hat{m}$ with both sides of that equation:

$$\begin{aligned}\frac{1}{\gamma H_k}\hat{m}\times\frac{d\hat{m}}{dt} &= (\cos\theta + h_{easy})\hat{m}\times(\hat{m}\times\hat{z}) - (h_p\cos\beta)\hat{m}\times(\hat{m}\times\hat{x}) \\ &\quad + h_s\hat{m}\times\hat{z} + \frac{\alpha}{\gamma H_k}\frac{d\hat{m}}{dt}\end{aligned} \quad (A.10)$$

where the vector identity

$$\vec{A}\times(\vec{B}\times\vec{C}) = (\vec{A}\cdot\vec{C})\vec{B} - (\vec{A}\cdot\vec{B})\vec{C}$$

was first applied to derive the vector term proportional to h_s:

$$\begin{aligned}\hat{m}\times(\hat{z}\times\hat{m}) &= (\hat{m}\cdot\hat{m})\hat{z} - (\hat{m}\cdot\hat{z})\hat{m} \\ \Rightarrow \hat{m}\times[\hat{m}\times(\hat{z}\times\hat{m})] &= \hat{m}\times\hat{z}\end{aligned}$$

and then the vector identity was applied to derive the $d\hat{m}/dt$ term:

$$\begin{aligned}\hat{m} \times (\frac{\alpha}{\gamma H_k}\hat{m} \times \frac{d\hat{m}}{dt}) &= \frac{\alpha}{\gamma H_k}\left[(\hat{m} \cdot \frac{d\hat{m}}{dt})\hat{m} - (\hat{m} \cdot \hat{m})\frac{d\hat{m}}{dt}\right] \\ &= -\frac{\alpha}{\gamma H_k}\frac{d\hat{m}}{dt} \end{aligned} \tag{A.11}$$

where it is noted that $\hat{m}$ and $\frac{d\hat{m}}{dt}$ are orthogonal (take the dot product of $\hat{m}$ with Eq. A.9), consistent with the construction of $\vec{m}$ representing a monodomain, saturated ferromagnet (i.e. constant magnitude).

Now substitue Eq. A.10 back into Eq. A.9

$$\begin{aligned}\frac{1}{\gamma H_k}\frac{d\hat{m}}{dt} &= (\cos\theta + h_{easy})\hat{m} \times \hat{z} - (h_p \cos\beta)\hat{m} \times \hat{x} + h_s \hat{m} \times (\hat{z} \times \hat{m}) \\ &\quad -\alpha\left[(\cos\theta + h_{easy})\hat{m} \times (\hat{m} \times \hat{z}) - (h_p\cos\beta)\hat{m} \times (\hat{m} \times \hat{x}) + h_s\hat{m} \times \hat{z} + \frac{\alpha}{\gamma H_k}\frac{d\hat{m}}{dt}\right]\end{aligned}$$

finally yielding an explicit expression for $d\hat{m}/dt$:

$$\begin{aligned}\frac{1+\alpha^2}{\gamma H_k}\frac{d\hat{m}}{dt} &= (\cos\theta + h_{easy} - \alpha h_s)\hat{m} \times \hat{z} - (\alpha\cos\theta + \alpha h_{easy} + h_s)\hat{m} \times (\hat{m} \times \hat{z}) \\ &\quad -(h_p\cos\beta)\hat{m} \times \hat{x} + \alpha h_p \cos\beta \hat{m} \times (\hat{m} \times \hat{x}) \end{aligned} \tag{A.12}$$

It is useful to choose θ and ϕ as the state variables for the system and to introduce $d\tau = dt\gamma H_k/(1+\alpha^2)$ as a natural time unit. Note, the fact that $\gamma < 0$ has non-trivial implications on the interpretation of the following equations, which will be reconciled upon obtaining the state evolution matrix of the small-angle linearized system.

$$\frac{d\theta}{d\tau} = \frac{d\hat{m}}{d\tau} \cdot \hat{\theta} \tag{A.13}$$

$$\frac{d\phi}{d\tau} = \frac{1}{\sin\theta}\frac{d\hat{m}}{d\tau} \cdot \hat{\phi} \tag{A.14}$$

In order to apply these two equations to the simplified result of the LLG equation, it is useful

to first tabulate the following relations between position[1] vectors:

$$\begin{aligned} \hat{m} \times \hat{z} &= -\sin\theta\hat{\phi} \\ \hat{m} \times (\hat{m} \times \hat{z}) &= \sin\theta\hat{\theta} \\ \hat{m} \times \hat{x} &= \cos\theta\cos\phi\hat{\phi} + \sin\phi\hat{\theta} \\ \hat{m} \times (\hat{m} \times \hat{x}) &= -\cos\theta\cos\phi\hat{\theta} + \sin\phi\hat{\phi} \end{aligned} \tag{A.15}$$

which result in

$$\begin{aligned} \frac{d\hat{m}}{d\tau} &= \left[-\alpha\sin\theta\cos\theta - \alpha h_{easy}\sin\theta - h_s\sin\theta - h_p\cos\beta(\sin\phi + \alpha\cos\theta\cos\phi)\right]\hat{\theta} \\ &\quad \left[-\sin\theta\cos\theta - h_{easy}\sin\theta + \alpha h_s\sin\theta + h_p\cos\beta(\alpha\sin\phi - \cos\theta\cos\phi)\right]\hat{\phi} \end{aligned} \tag{A.16}$$

At this point, the relations in Eqns. A.13, A.14, A.16 produce a nonlinear system of two differential equations that can be numerically solved for the magnetization dynamics for various values of spin torque h_s and easy-axis field h_{easy}. However, the MRAM circuit designer is primarily interested in how to get $\hat{m}$ from $\theta = 0$ to $\theta = \pi$, not the full solution to the trajectory of $\hat{m}$. A small angle linearized differential equation provides the necessary insight (take $\cos\theta \approx 1$ and $\sin\theta \approx \theta$, $\cos\beta = \sin\theta\cos\phi \approx \theta\cos\phi$):

$$\begin{aligned} \frac{d\hat{m}}{d\tau} &\approx \left[-\alpha\theta - \alpha h_{easy}\theta - h_s\theta - h_p\theta\cos\phi(\sin\phi + \alpha\cos\phi)\right]\hat{\theta} \\ &\quad + \left[-\theta - h_{easy}\theta + \alpha h_s\theta + h_p\theta\cos\phi(\alpha\sin\phi - \cos\phi)\right]\hat{\phi} \end{aligned}$$

Applying Eqns. A.13, A.14 to the above small-angle equation gives:

$$\begin{bmatrix} \dot{\theta} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} -\alpha\theta - \alpha h_{easy}\theta - h_s\theta - h_p\theta\cos\phi(\sin\phi + \alpha\cos\phi) \\ -1 - h_{easy} + \alpha h_s + h_p\cos\phi(\alpha\sin\phi - \cos\phi) \end{bmatrix} \tag{A.17}$$

[1] $\hat{m}$ is analagous to $\hat{r}$ except the position coordinates are in the nanomagnet's (m_x, m_y, m_z) (or equivalently (θ, ϕ)) phase space.

This equation can be analytically solved in the small-angle limit by rewriting it in terms of components of $\hat{m}$: $n_x = m_x/m = \sin\theta\cos\phi \approx \theta\cos\phi$ and $n_y = m_y/m = \sin\theta\sin\phi \approx \theta\sin\phi$. This is accomplished by writing down:

$$\dot{n}_x = \dot{\theta}\cos\phi - \theta\sin\phi\dot{\phi}$$
$$\dot{n}_y = \dot{\theta}\sin\phi + \theta\cos\phi\dot{\phi}$$

and using Eq. A.17 to substitute for the derivatives on the right hand side. After simplification, a linear differential equation is achieved:

$$\begin{aligned} \begin{bmatrix} \dot{n}_x \\ \dot{n}_y \end{bmatrix} &= \begin{bmatrix} -[\alpha(1+h_{easy}) + h_s + \alpha h_p] & (1+h_{easy} - \alpha h_s) \\ -[1+h_{easy} - \alpha h_s + h_p] & -(\alpha(1+h_{easy}) + h_s) \end{bmatrix} \begin{bmatrix} n_x \\ n_y \end{bmatrix} \\ \dot{\bar{n}} &= A\bar{n} \end{aligned}$$

Since $d\tau = dt\gamma H_k/(1+\alpha^2) < 0$, one must reverse the sign of h_s and the overall sign of the off-diagonal terms in the matrix A in order to appropriately interpret the dynamics as moving *forward* in time (one can deduce these requirements by examining how terms get affected by $\gamma \to -|\gamma|$ in Eq. 2.42). This gives the correct state evolution matrix A' as:

$$A' = \begin{bmatrix} -[\alpha(1+h_{easy}) - h_s + \alpha h_p] & -(1+h_{easy} + \alpha h_s) \\ [1+h_{easy} + \alpha h_s + h_p] & -[\alpha(1+h_{easy}) - h_s] \end{bmatrix} \tag{A.18}$$

In order for this system to be stable, the real parts of the eigenvalues of A' have to be negative which requires $\mathrm{tr(A')} < 0$ and $\mathrm{det(A')} > 0$. The trace is:

$$\mathrm{tr(A')} = -2\left[\alpha(1+\mathrm{h_{easy}}) - \mathrm{h_s} + \frac{1}{2}\alpha\mathrm{h_p}\right]$$

which must be negative:

$$\begin{aligned} \mathrm{tr}(\mathrm{A}) &< 0 \\ \Rightarrow h_s &< (\alpha(1+h_{easy}) + \frac{1}{2}\alpha h_p) \end{aligned} \tag{A.19}$$

The above equation gives the value of spin torque $h_{sc} = \alpha(1 + h_{easy} + \frac{1}{2}h_p)$ beyond which the dynamics of a magnetic moment oriented along $+\hat{z}$ become unstable.[2] Translating back to current gives:

$$\begin{aligned} I_{c0} &= \frac{1}{\eta}\frac{|e|}{\frac{\hbar}{2}}mH_k h_{sc} \\ &= \frac{1}{\eta}\frac{2|e|}{\hbar}m\alpha(H_k + H_{easy} + 2\pi M(D_a - D_b)) \\ &= \frac{1}{\eta}\frac{2|e|}{\hbar}m\alpha(H_k + H_{easy} + 2\pi M) \end{aligned}$$

Where it is recalled from Eq. 2.32 that $h_p = (D_a - D_b)/(D_b - D_c)$ and $H_k = 4\pi M(D_b - D_c)$, and that $D_a - D_b \approx 1$ for the geometry of the free layer of the MTJ with a very small width along the x direction (and therefore, strong demagnetization field along the x direction). The previous equation correspons to Eq. 2.43 and completes the derivation.

This derivation also highlights a linearized differential equation $\dot{\bar{n}} = A'\bar{n}$ that can be analytically solved to deduce the switching time by determining how long it takes a moment initially at (θ_0, ϕ_0) to reach $\theta = \pi/2$. If one evaluates the switching times in this manner for various initial conditions and compares the results with numerical solutions from the full, non-linear differential equation provided in [27], one will see good enough agreement for the purpose of estimating required current levels for a target switching time.

[2] One can verify that the condition on $\det(\mathrm{A}')$ is met by substituting $h_s = h_{sc}$ and obtaining a positive value for the determinant.

Bibliography

[1] A. K. Sharma, *Advanced Semiconductor Memories.* Wiley-Interscience and IEEE Press, 2003.

[2] T. H. Ning, “Silicon technology—emerging trends from a system application perspective,” in *2003 International Symposium on VLSI Technology, Systems, and Applications*, Oct. 2003, pp. 6–8.

[3] S. Tehrani *et al.*, “Magnetoresistive random access memory using magnetic tunnel junctions,” in *Proceedings of the IEEE*, vol. 91, May 2003, pp. 703–714.

[4] M. Julliere, “Tunneling between ferromagnetic films,” *Physics Letters A*, vol. 54, pp. 225–226, Sept. 1975.

[5] J. Slonczewski, “Conductance and exchange coupling of two ferromagnets seperated by a tunneling barrier,” *Physical Review B*, vol. 139, pp. 6995–7002, Apr. 1989.

[6] S. Parkin *et al.*, “Giant tunneling magnetoresistance at room temperature with MgO (100) tunnel barriers,” *nature materials*, vol. 3, Dec. 2004.

[7] W. J. Gallagher and S. S. P. Parkin, “Development of the magnetic tunnel junction MRAM at IBM: from first junctions to a 16-Mb MRAM demonstrator chip,” *IBM Journal of Research and Development*, vol. 50, Jan. 2006.

[8] T. M. Maffitt *et al.*, “Design considerations for MRAM,” *IBM Journal of Research and Development*, vol. 50, Jan. 2006.

[9] D. Gogl, C. Arndt, J. C. Barwin, A. Bette, J. DeBrosse, E. Gow, H. Hoenigschmid, S. Lammers, M. Lamorey, Y. Lu, T. Maffitt, K. Maloney, W. Obermaier, A. Sturm, H. Viehmann, D. Willmott, M. Wood, W. J. Gallagher, G. Mueller, and A. R. Sitaram, “A 16-Mb MRAM featuring bootstrapped write drivers,” *IEEE J. Solid-State Circuits*, vol. 40, pp. 902–908, Apr. 2005.

[10] S. Tehrani *et al.*, “A 4-Mb toggle MRAM based on a novel bit and switching method,” *IEEE Trans. Magn.*, vol. 41, pp. 132–136, Jan. 2005.

[11] J. Slonczewski, "Current-driven excitation of magnetic multilayers," *Journal of Magnetism and Magnetic Materials*, vol. 159, pp. 1–7, June 1996.

[12] J. Sun, T. Kuan, J. Katine, and R. Koch, "Spin angular momentum transfer in a current-perpendicular spin-valve nanomagnet," in *Organic Photonic Materials and Devices VI. Proceedings of the SPIE*, vol. 5359, July 2004, pp. 445–455.

[13] H. Meng and J. Wang, "Spin transfer effect in magnetic tunnel junction with a nano-current-channel layer in free layer," *IEEE Trans. Magn.*, vol. 41, pp. 2612–2614, Oct. 2005.

[14] J. Hayakawa *et al.*, "Current-driven magnetization switching in cofeb/mg0/cofeb magnetic tunnel junctions," *Japanese Journal of Applied Physics*, vol. 44, no. 41, 2005.

[15] M. Hosomi *et al.*, "A novel nonvolatile memory with spin torque transfer magnetization switching: Spin-RAM," in *Electron Devices Meeting, 2005. IEEE International*, Washington, DC, Dec. 2005.

[16] P. L. Abraham, David W. Trouilloud, "Thermally-assisted magnetic random access memory (MRAM)," U.S. Patent 6,385,082, May, 2002. [Online]. Available: http://www.freepatentsonline.com/6385082.html

[17] I. L. Prejbeanu *et al.*, "Thermally assisted switching in exchange-biased storage layer magnetic tunnel junctions," *IEEE Trans. Magn.*, vol. 40, pp. 2625–2627, July 2004.

[18] E. M. Purcell, *Electricity and Magnetism*, 2nd ed. McGraw-Hill, 1985.

[19] D. J. Griffiths, *Introduction to Electrodynamics*, 3rd ed. Prentice Hall, 1998.

[20] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. Wiley, 1998.

[21] J. J. Sakurai, *Modern Quantum Mechanics*, 2nd ed. Addison Wesley, 1994.

[22] H. Rogers, *Multivariable Calculus*, 3rd ed. Prentice Hall, 1998.

[23] C. Kittel, *Introduction to Solid State Physics*, 7th ed. Wiley, 1996.

[24] E. C. Stoner, "The demagnetizing factors for ellipsoids," in *Philosophical Magazine*, ser. 7, Dec. 1945, vol. 36, no. 264, pp. 803–821.

[25] E. C. Stoner and E. P. Wohlfarth, "A mechanism of magnetic hysteresis in heterogeneous alloys," *IEEE Trans. Magn.*, vol. 27, no. 4, pp. 3475–3518, July 1991.

[26] M. d'Aquino, "Nonlinear magnetization dynamics in thin-films and nanoparticles," Ph.D. dissertation, Universitá Degli Studi di Napoli Federico II, Dec. 2004.

[27] J. Z. Sun, "Spin-current interaction with a monodomain magnetic body: A model study," *Physical Review B*, vol. 62, no. 1, pp. 570–578, July 2000.

[28] L. Landau, E. Lifshitz, and L. Pitaevskii, *Electrodynamics of Continuous Media*, 2nd ed. Elsevier, 1982.

[29] W. F. Brown, "Thermal fluctuations of a single-domain particle," *Physical Review*, vol. 130, no. 5, pp. 1677–1961, June 1963.

[30] M. Igarashi, F. Akagi, K. Yoshida, and Y. Natakani, "Effect of angle dependent attempt frequency on arrhenius-neel thermal decay in thin film media," *IEEE Trans. Magn.*, vol. 36, pp. 2459–2461, Sept. 2000.

[31] T. L. Gilbert, "A phenomenological theory of damping in ferromagnetic materials," *IEEE Trans. Magn.*, vol. 40, no. 6, pp. 3443–3449, Nov. 2004.

[32] A. R. Sitaram *et al.*, "A 0.18μm logic-based MRAM technology for high performance nonvolatile memory applications," in *2003 Symposium on VLSI Technology Digest of Technical Papers*, June 2003, pp. 15–16.

[33] L. Berger, "Emission of spin waves by a magnetic multilayer traversed by a current," *Physical Review B*, vol. 54, no. 13, pp. 9353–9358, Oct. 1996.

[34] M. Covington *et al.*, "Magnetization dynamics driven by spin momentum transfer," in *Fluctuations and Noise in Materials II. Proceedings of the SPIE*, vol. 5843, 2005, pp. 11–22.

[35] M. C. Gaidis *et al.*, "Two-level BEOL processing for rapid iteration in MRAM development," *IBM Journal of Research and Development*, vol. 50, Jan. 2006.

[36] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 4th ed. Wiley, 2001.

[37] B. Razavi, *Analysis and Design of Analog CMOS Integrated Circuits.* McGraw-Hill, 2001.

[38] J. Parenteau, "Personal discussion," IBM, Aug. 2006.