University of California
Los Angeles

Modeling and Design of STT-MRAMs

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Electrical Engineering

by

Richard William Dorrance

2011

The thesis of Richard William Dorrance is approved.

Kang L. Wang

Chih-Kong Ken Yang

Dejan Marković, Committee Chair

University of California, Los Angeles

2011

To my parents

Table of Contents

List of Figures

List of Tables

Acknowledgments

First, I would like to thank my advisor, Professor Dejan Marković, without whom this thesis could never have been written. I am sincerely grateful for the help and support he has given me through the course of this work. I would also like to acknowledge Professors Chih-Kong Ken Yang and Kang L. Wang and Dr. Pedram Khalili. Professor Yang's knowledge and experience in designing memories was instrumental to this project. Without the help and insight of Professor Wang and Dr. Khalili I could never have comprehended the physics of spintroncs half as well as I do now.

This work could not have progressed without the infrastructure and support provided by research group. I would like to whole heartedly thank Henry Chen, Kevin Dwan, and Yuta Toriyama for their help in the editing of this manuscript. I would also like to thank the rest of the STT-RAM circuits design team. Without their hard work and dedication, this project could never have gotten off the ground. I would also like to thank graduate students Juan G. Alzate, Pramey Upadhyaya, and Mark Lewis from the STT-RAM MTJ design team. They provided an extensive number of device characterizations and simulations for the development of my MTJ macro-model.

Most of all, I would like to thank my parents, Gary Dorrance and Karen Lawrence, for all the love and support they have provided over the years. Not the least of which included the editing of early drafts of this thesis. Last, but certainly not least, I would like to thank the mighty Tyrannosaurus Rex, king of the dinosaurs. May he live on in the hopes and dreams of a man still a boy at heart.

Abstract of the Thesis

Modeling and Design of STT-MRAMs

by

Richard William Dorrance

Master of Science in Electrical Engineering

University of California, Los Angeles, 2011

Professor Dejan Marković, Chair

Spin-Torque Transfer Magnetoresistive Random Access Memory (STT-MRAM) is an emerging memory technology with the potential to become a true universal memory: the density of DRAM, the speed of SRAM, and the non-volatility of Flash. STT-MRAM uses a Magnetic Tunnel Junction (MTJ) device as a non-volatile magnetic memory storage element and the recently discovered spin-torque phenomenon to switch magnetic states. In this work, the fundamental quantum mechanical nature of the MTJ is explored to develop a highly accurate physics-based model of its spintronic operation. Innovative design-space analysis techniques are introduced to investigate existing and proposed STT-MRAM architectures. Three test chips were fabricated using these new design methodologies at 90nm, 65nm, and 45nm technology nodes. Each chip has a memory density more than two times greater and a read/write performance more than 10 times greater when compared to published state-of-the-art STT-MRAMs. Theoretical and observed scaling trends show flash-like densities, with SRAM-equivalent access times, while using 10 times less energy in more advanced technology nodes (below 32nm).

CHAPTER 1

Introduction

STT-MRAM[1] is an emerging memory technology that exploits the recently discovered phenomena of spin-torque transfer (STT) in MTJs. This chapter provides a brief motivation for STT-MRAM, as well as outlines the rest of the thesis.

1.1 Motivation for STT-MRAM

Currently, three types of memory exist, with each technology doing a single thing very well: Static RAM (SRAM), Dynamic RAM (DRAM), and Flash memory. SRAM has excellent read and write speeds, but has a very large cell size (requiring 6 or more transistors per cell). The speed of SRAM makes it ideally suited for embedded applications, particularly cache memory, where performance is more important than memory density. SRAM is volatile, but requires very little active power for data retention. DRAM is able to provide much better memory density through its use of a single transistor with a storage capacitor. However, charge tends to leak off of the capacitor, requiring a power hungry refresh cycle every few milliseconds. DRAM is typically used as the main system memory in a computer, where memory density and performance are more important than

[1] In literature, Spin-Torque Transfer Random Access Memory (STT-RAM) and Spin Random Access Memory (SPRAM) are used interchangeably with STT-MRAM. However, STT-MRAM is more common and is, therefore, used exclusively in this thesis.

	SRAM	DRAM	Flash (NOR)	Flash (NAND)	FeRAM	MRAM	PRAM	RRAM	STT-MRAM
Non-volatile	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cell Size [F^2]	50-120	6-10	10	5	15-34	16-40	6-12	6-10	6-20
Read Time [ns]	1-100	30	10	50	20-80	3-20	20-50	10-50	2-20
Write/Erase Time [ns]	1-100	15	1µs/1ms	1ms/0.1ms	50/50	3-20	50/120	10-50	2-20
Endurance	10^{16}	10^{16}	10^{5}	10^{5}	10^{12}	$>10^{15}$	10^{8}	10^{8}	$>10^{15}$
Write Power	Low	Low	Very High	Very High	Low	High	Low	Low	Low
Other Power Consumption	Leakage	Refresh	None	None	None	None	None	None	None
High Voltage Required	No	3V	6-8V	16-20V	2-3V	3V	1.5-3V	1.5-3V	<1.5V
	Existing Products						*Prototypes*		

Figure 1.1: Comparison of memory technologies (source: Wolf et al. [1]).

power consumption. Flash memory technologies are very attractive for mobile applications where non-volatility and very high densities are required. While Flash does have reasonably fast read access times, write speeds are very slow and endurance rates are very low ($< 100,000$ cycles). To optimize for power, performance, and cost, a typical system must integrate all three types of memory. STT-MRAM promises to be a "universal memory", combining all of the advantages of SRAM, DRAM, and Flash. Such a memory would eliminate the need multiple application specific memories, improving system performance and reliability, while also lowering cost and power consumption.

Magnetoresistive Random Access Memory (MRAM) is a technology that has existed in one form or another since the late 1970s [1]. MRAM is based on the concept of using the direction of magnetization to store binary information, while exploiting magnetoresistive properties for data retrieval. The mid-1990s saw a

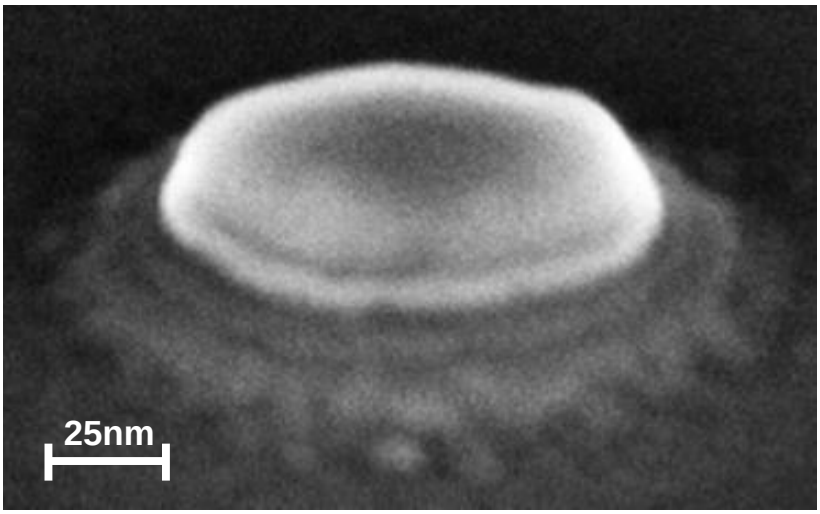
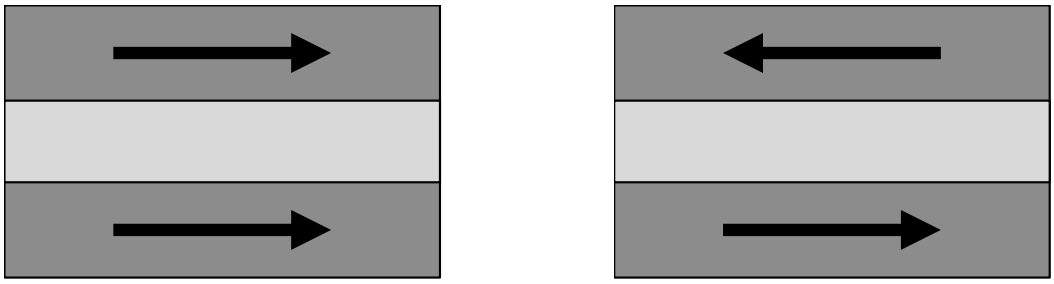


Figure 1.2: SEM photo of an MTJ, courtesy of Mark Lewis.

resurgence of interest in MRAM technologies with the discovery of room temperature tunneling magnetoresistance (TMR) in magnetic tunnel junctions (MTJs). Fig. 1.2 shows an SEM photo of a single MTJ nanopillar. The spintronic operation of the MTJ is discussed later in this work. For now, it is sufficient to understand the MTJ as a pair of ferromagnets separated by a thin insulating layer. Two possible magnetic states arise, the parallel combination of the two layers (Fig. 1.3(a)) and the antiparallel combination (Fig. 1.3(b)). The parallel configuration leads to a low resistive state (R_P), while the antiparallel configuration leads to a high resistive state (R_{AP}).

The spin-torque transfer effect was first theoretically predicted and demonstrated by J. C. Slonczewski in 1996 [2] and has formed the basis of next generation MRAMs. STT-MRAM can scale well below 65nm, while reducing writing currents by more than a hundredfold [1]. Before STT, writing currents increased exponentially with MRAM scaling, causing electromigration and power concerns that prevented scaling below 90nm. The nonvolatile nature, low power, high performance, and memory density of STT-MRAM make it an excellent candidate for



(a) Parallel **(b)** Antiparallel

Figure 1.3: MTJ ferromagnetic layers in (a) parallel and (b) antiparallel configurations.

the first commercially available universal memory. However, the lack of an accurate, compact macro-model, incorporating temperature and bias voltage effects, is the largest obstacle to the design of high performance STT-MRAMs. Without such a design tool, it is impossible to verify timing and yield or predict device behavior with scaling. Another big challenge is the integration of MTJs with CMOS. The process flow, while in principle fully compatible with CMOS, adds extra design and layout constraints. An integrated MTJ also behaves slightly differently, further stressing the need for an accurate macro-model.

1.2 Thesis Outline

Chapter 2 begins with a brief introduction to the field of spintronics, highlighting the general principles of operation and potential applications. The second half of Chapter 2 is dedicated entirely to the operation and device characteristics of the magnetic tunnel junction. Chapter 3 introduces a compact MTJ macro-model capable of accurately capturing and modeling the quantum mechanical behavior of MTJs. STT-MRAM memory cell and subarray architectures are outlined in

Chapter 4, with design-space analysis techniques introduced in Chapter 5. The analysis techniques introduced in Chapter 5 are used in Chapter 6 in the design of three separate STT-MRAM memory chips at 90nm, 65nm, and 45nm technology nodes. Finally, Chapter 7 presents ongoing and future work, and concludes the thesis.

CHAPTER 2

Magnetic Tunnel Junctions

The focus of this chapter is to introduce the MTJ device, as well as the field of spintronics. The first section provides a brief background on spintronics—its history and fundamental physical operation. Alternative devices (e.g. spin FETs, MBTs, and spin LEDs) and applications are discussed before defining the characteristics and unique properties of the MTJ device.

2.1 Introduction to Spintronics

Spintronics, the amalgamation of the words "spin" and "electronics," involves the active control and manipulation of electron spin in solid-state electronics [3]. In traditional electronic devices, information processing works on the principle of control over the flow of charge through a semiconductor material. Large scale, non-volatile memories (e.g., hard disk drives or HDDs) exploit ferromagnetism to store information by forcing the spin alignment of many electrons [4]. Spintronics, as a whole, aims to merge information processing and storage through the use of spin-polarized currents [1].

2.1.1 History

Early work into spintronics began in the mid-1930s with the discovery of unusual resistance behavior in ferromagnetic materials at extremely low temperatures [3]. Electron tunneling measurements played a key role in early experimental work, with several key experiments in the early 1970s demonstrating the viability of spin filters (discussed later). In 1975, Jullierè [5] formulated his now-famous conductance model describing the change of conductance between the parallel and antiparallel states of an MTJ. However, it wasn't until the mid-to-late 1980s that the room temperature magnetoresistive effects were discovered. Anisotropic magnetoresistive (AMR) layers were first used to construst AMR-MRAM to replace bulky and heavy plated-wire radiation-hard memories [1]. AMR was quickly replaced by the discovery of giant magnetoresistance (GMR) in 1988 [6]. Since the discovory of GMR, electron spin has formed the basis of almost all electronic information storage [7].

In the early 1990s, MTJ materials with higher TMRs (on the order of 20% at room temperature) were discovered [1]. Since then, MTJ structures (using MgO insulating barriers) with TMRs on the order of 1000% have been demonstrated at room temperature [8]. Within ten years of its discovery, spintronics has grown into a billion dollar industry, with commercial sales exceeding $3 billion in 2005 [6]. Despite these successes, spin injection from ferromagnetic layers into semiconductors remains a significant bottleneck in semiconductor-based spintronics. Recently, much emphasis has been placed in trying to induce ferromagnetism in a semiconductors to produce dilute magnetic semiconductors (DMS) [7]. DMS has the potential to improve the Curie temperature and magnetic bandgap of future spintronic devices [9].

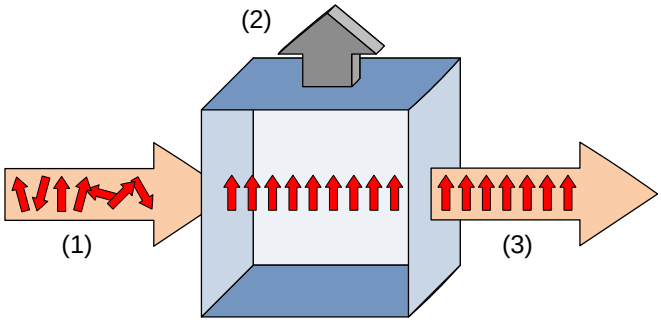
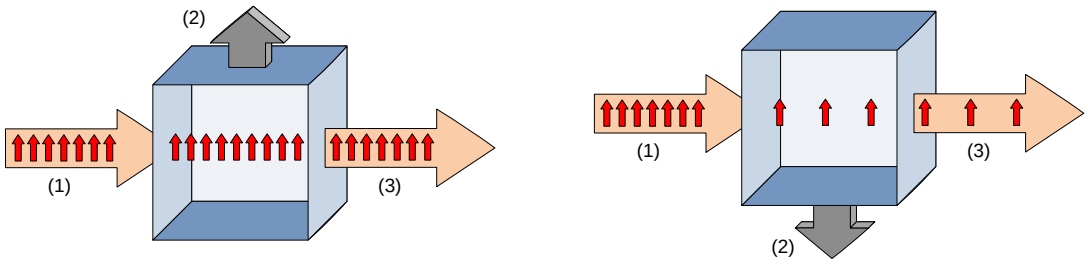


Figure 2.1: A cartoon of the operation of a spin polarizer. A spin unpolarized current enters at (1). It is then spin polarized in the direction of (2), before exiting the spin polarizer at (3).



(a) Parallel spin filter.

(b) Antiparallel spin filter.

Figure 2.2: A cartoon of the operation of a spin filter to a (a) parallel and (b) an antiparallel current. A spin-polarized current enters at (1) and filters in the direction of (2), before exiting the spin polarizer at (3).

2.1.2 Principle of Operation

Electron spin is a "pseudovector" with a fixed magnitude but a variable direction (spin polarization). The spin polarization of an electron can be made bistable by placing it in a magnetic field. In the presence of a magnetic field, only spin polarizations parallel or antiparallel to the field are possible [1]. This property introduces the concept of a spin polarizer (Fig. 2.1). A thin ferromagnetic layer can act as a spin polarizer. When a spin unpolarized current passes through the

ferromagnetic layer, it tends to become spin-polarized in the direction of magnetization [6]. Another key aspect to spintronics is the concept of a spin filter (Fig. 2.2). A spin filter will only pass a current if the two are polarized in the same direction. If the current and filter are completely antiparallel, no current is passed. Ferromagnetic films also display the properties of a spin filter [6]. A "spin valve" can be constructed by using a spin polarizer in conjunction with a spin filter [4]. By controling the angle of magentization between the polarizer and the filter, a magnetically controlled spin valve can be formed. The spin valve effect is exploited in MRAMs to use MTJs as the memory storage element [1].

2.1.3 Other Devices and Applications

Several kinds of "spin transistors" exist, including the spin field-effect transistor (spin FET), the magnetic bipolar transistor (MBT), and hot-electron spin transistors [3, 7]. Structurally similar to a MOSFET, a spin FET sandwiches the conducting channel between two ferromagnetic layers. When the ferromagnets are aligned in the parallel configuration, the spin FET behaves like a normal MOSFET. However, when configured in the antiparallel alignment, transistor will be shut off [10]. Spin FETs can be easily integrated into existing CMOS circuitry and provide much larger ON/OFF current ratios [3]. MBTs are essentially BJTs with the addition of a ferromagnetic spin injector attached to the emitter. In an MBT, the gain factor β heavily depends upon nonequilibrium spin polarization and is called magnetoamplification [11]. MBTs can be used to generate almost 100% spin-coherent currents that can be very long lived [3, 11].

Another potential application of spintronics is optics, specifically, through the use of spin light emitting diodes (spin LEDs) and spin selective Kerr rotators [7]. In a spin LED, the polarization of the light emitted is modulated through the

application of an external magnetic field [12]. Variable polarized LEDs promise to provide more energy efficient displays and significantly higher signal-to-noise ratio (SNR) in optical communications [7]. A Kerr rotator takes advantage of the magneto-optic Kerr effect (MOKE), the unique optically-reflective properties of magnetic materials, to manipulate the polarity of reflected light. Traditionally, Kerr rotators have many applications in the microscopic imaging of magnetic domains, magnetic media, and terahertz lasers [13]. A spin selective rotator, with the application of a bias voltage, can be made to reflect incident light either with or without a large Kerr rotation angle [7].

2.2 The Magnetic Tunnel Junction

This section is intended to describe the major device characteristics observed in MTJs. The science responsible for each effect, as well as their importance to the MTJ model, is discussed.

2.2.1 Resistance Hysteresis

The large resistance hysteresis present in MTJs makes them very well-suited as a non-volatile memory element. The source of this hysteresis is very nicely explained by the spin-valve structure of an MTJ [3]. As mentioned before in Fig. 3.1, an MTJ consists of a thin insulating layer sandwiched between two ferromagnetic layers. The electromagnetic dynamics of the system allows for only two possible states: parallel or antiparallel [6]. The two ferromagnetic layers are magnetized in the same direction while in the parallel state and in the opposite directions while in the antiparallel state. When a current flows through the MTJ, one ferromagnetic layer acts as a spin polarizer and the other as a spin

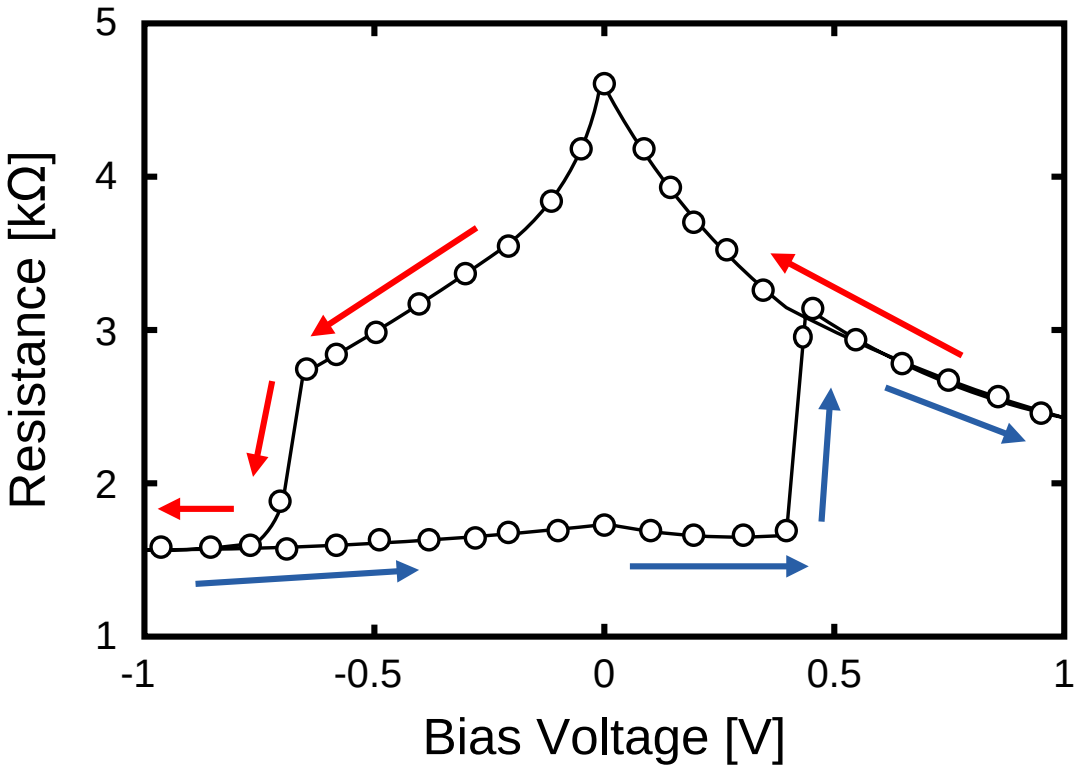


Figure 2.3: Resistance hysteresis of an MTJ. Switching from $P \rightarrow AP$ (blue arrows) and $AP \rightarrow P$ (red arrows).

filter. In the parallel state, since the two ferromagnetic layers are aligned, the current is passed undisturbed, creating a low resistive state (R_P). However, in the antiparallel state, the spin filter will block the antiparallel current generated by the polarizing layer, creating a high resistive state (R_{AP}). Tunnel magnetoresistance (TMR) is a metric for determining the efficiency of spin-valve operation in an MTJ [14]. TMR is defined as:

$$TMR = \frac{R_{AP} - R_P}{R_P}. \tag{2.1}$$

Fig. 2.3 shows a sample resistance hysteresis of an MTJ by sweeping the bias voltage. R_P and R_{AP} are clearly evident, along with several other characteristics to be discussed: critical switching currents, switching asymmetry, and the bias

voltage dependance of TMR.

2.2.2 Critical Switching Current

2.2.2.1 Asymmetric Switching Currents

It should be noted that the critical switching currents are asymmetric, with $I_C(P \rightarrow AP) > I_C(AP \rightarrow P)$ [15]. This effect was predicted by Slonczewski [2] with his discovery of the spin-torque transfer phenomena. This asymmetry is proportional to and increases linearly with TMR [16]. The simplest explanation of this behavior is that the antiparallel configuration is a lower energy state than the parallel case [3], making it is easier to switch to the antiparallel state than the parallel state. Several techniques exist to minimize the asymmetry. Lee et al. [17] were able to tune the magnetostatic offset field (using an external magnetic field) with exceptional results, reducing the asymmetric current ratio from 1.51 to 1.04. Yao et al. [18] were able to reduce the offset from 1.50 to 1.28 with the introduction of a nanocurrent-channel layer to the MTJ stack.

2.2.2.2 Switching Regimes

In MTJs, two types of magnetic switching occur due to spin-torque transfer: precessional and thermally activated switching [19, 20]. Precessional switching occurs on a nanosecond time scale, while thermally activated switching occurs at much larger time scales [14]. The transition between these two switching regions lies between 1 and 10ns, which is depicted in Fig. 2.4. The dynamics of precessional switching are well described by the Landau-Lifshitz-Gilbert equation (LLGE) [21, 22], given by:

$$\frac{\partial \vec{m}}{\partial t} = -\gamma M_S \vec{m} \times \left(\vec{h}_{eff} - \alpha \frac{\partial \vec{m}}{\partial t} \right). \tag{2.2}$$

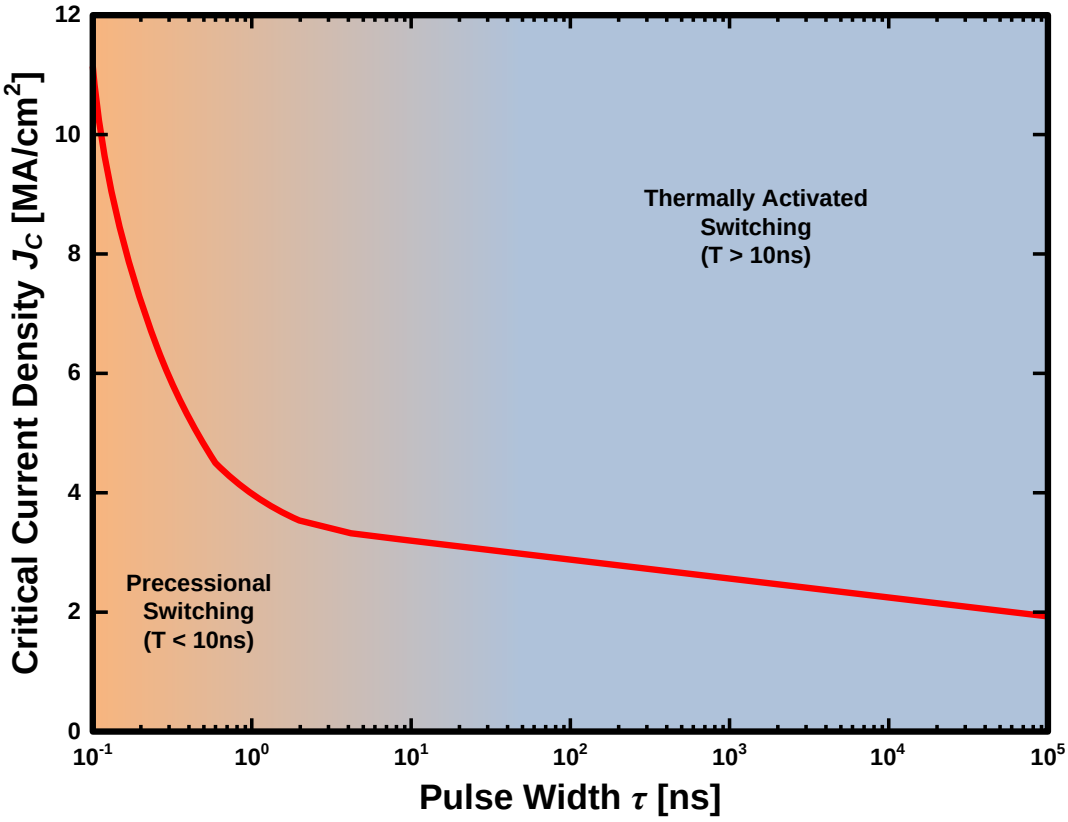


Figure 2.4: MTJ switching regimes.

Equation 2.2, with the addition of Slonczewski's spin-torque transfer term [2], will be discussed in much more detail in Chapter 3.

Switching occurs on much longer time scale when the current though the MTJ is less than the critical switching current [19]. In the thermally activated regime, the switching current is a function of pulse duration τ:

$$I_C = I_{C0}\left[1 - \frac{\ln(\tau/\tau_0)}{\Delta}\right], \tag{2.3}$$

where Δ is the thermal stability of the MTJ, τ_0 is the natural time constant, and I_{C0} is the critical switching current [23].

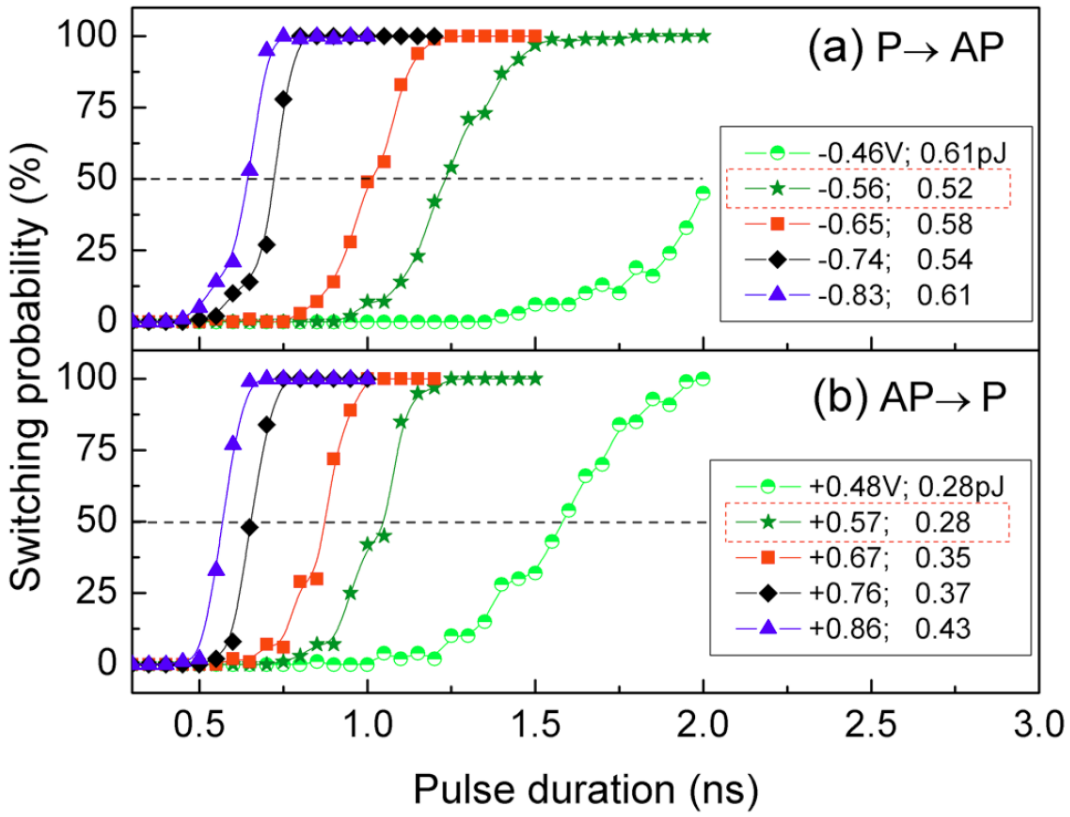


Figure 2.5: Switching probability vs. pulse duration for (a) $P \rightarrow AP$ and (b) $AP \rightarrow P$. Reproduced with permission from Zeng et al. [24].

2.2.2.3 Probabilistic Switching

Due to thermal agitation, the inital angle between the magnetizations of the fixed and free layers are in constant flux [25]. Combined with other finite temperature effects, this leads to a time-varying critical switching current [26]. This effect is very well modeled as a single critical switching current with a probabilistic distribution [27]. Fig. 2.5 shows one such measurement of the probabilistic distribution for a $135 \times 65nm^2$ CoFeB/MgO/CoFeB device. For memory applications, devices exhibiting very sharp transitions are highly desirable [28].

2.2.3 Tunnel Magnetoresistance Temperature Dependency

The sensitivity of TMR to temperature is well documented in literature [5, 29–32]. The effect at zero bias voltage is very well described by the Jullièe conductance model [29]. The Jullièe model decomposes the conductance of the MTJ into two parts: (i) G_T, the conductance due to direct elastic tunneling, and (ii) G_{SI}, the conductance due to imperfections in the insulating layer (assumed to be unpolarized). The total conductance (G), as a function of the angle θ is given by:

$$G(\theta) = G_T \{1 + P_1 P_2 \cos(\theta)\} + G_{SI}, \tag{2.4}$$

where P_1 and P_2 are the factors of spin-polarization for the two ferromagnetic layers, and $\theta = 0°$ for parallel and $\theta = 180°$ for anti-parallel magnetization. The temperature dependance of spin-polarazation has been extensively studied and shown to be:

$$P(T) = P_0 \left(1 - \alpha_{sp} T^{3/2}\right). \tag{2.5}$$

It should be noted that variations in G_T due to temperature are almost negligible, whereas $G_{SI} \propto T^{4/3}$ has been confirmed both theoretically and experimentally [33].

2.2.4 Bias Voltage Effects

The Jullièe conductance model is not perfect, being only able predict TMR at zero bias voltage [30]. Fig. 2.3 illustrates the effect of the so called "zero bias anomaly" in an MTJ structure [34]. The source of the bias voltage dependence of TMR is still not very well understood [35]. However, it is suspected that elastic currents play a role at low voltages [36] and redistribtion of the density of states at higher voltages [35]. At higher voltages, Simmons' formula can be used to

model the density of states to predict degradation of TMR to a bias voltage [31].

2.2.5 Other Important MTJ Characteristics

2.2.5.1 Self Induced Heating

Due to small device sizes and large write currents, the power density of a write operation in an MTJ can be very high. These high power densities can lead to localized heating or self induced heating in MTJs [37]. Hotspots (weak areas in the insulating barrier) and pinholes (direct contact between the magnetic layers) cause nonuniform current flow through the MTJ [38]. This leads to nonuniform heating across the tunneling barrier, affecting spin-polarization efficiency and causing inelastic electron scattering [38]. Simulations show that consecutive write opperations produce a 9-15°C increase in the temperature of the MTJ [37]. Additionally, a large number of writes followed by a read leads to degraded sensing margin. Self induced heating is exploited as the writing mechanism in Thermal Assisted Switching MRAMs (TAS-MRAMs) [39]. However, in STT-MRAMS, lower RAs are generally used to avoid self induced heating [37].

2.2.5.2 Backhopping

Backhopping is a recently discovered phenomenon, whereby increasing the bias voltage beyond the apparent switching threshold causes the MTJ to precess back and forth before settling to its original state [40]. This results in a lowered probability of switching at bias voltages beyond the threshold, causing non-monotonicity in the probability switching curves [41]. Backhopping is also much more pronounced in switching from an antiparallel to a parallel state [40,41]. This suggests that backhopping is related to the interlayer exchange coupling be-

tween the free and fixed layers. Backhopping is more pronounced on longer time scales, where self induced heating could be lowering the thermal energy barrier and causing hot-electron events [40]. Another explanation is that certain noise processes (discussed in the next section) might be responsible [41].

2.2.5.3 Noise

Many different mechanisms are responsible for noise in MTJs. Among these are thermal noise (Johnson-Nyquist), shot noise (current), flicker noise ($1/f$), random telegraph noise (RTN), and noise due to charge-trapping in the oxide barrier [28, 42–45]. Due to the strong coupling between magnetization and junction resistance in MTJs, noise in the magnetic domain is responsible for random resistance fluctuations [42]. These resistance fluctuations are responsible for $1/f$ noise as well as RTN [45]. Magnetic impurities inside the tunneling barrier are responsible for charge-trapping [42].

Thermal noise dominates at low bias voltages before quickly being overpowered by shot noise [44]. At room temperatures, shot noise typically dominates for bias voltages greater than $50mV$ [43]. The thermal noise of an MTJ is given by $S_V = 4k_BTR_{MTJ}$, where k_B is Boltzmann's constant, T is in Kelvin, and R_{MTJ} is the resistance of the MTJ [28]. Similarly, shot noise can be expressed as $S_V = 2eIR_{MTJ}^2$, where e is the charge of an electron and I is the current through the device [28].

Another significant contribution to low-frequency noise is due to domain wall hopping between pinning sites [28, 42]. These pinning sites are created by edge roughness, interface defects, bulk defects, and random film anisotropy [42]. The low-frequency noise characteristics of an MTJ can be significantly reduced by improving the smoothness of the ferromagnetic/insulator interface [45].

CHAPTER 3

Modeling MTJ Characteristics

Recent advances in MgO-based MTJs show strong potential for STT-MRAMs [46]. STT-MRAM has the potential to rival the densities of DRAM, the speed of SRAM, and is non-volatile without degrading over time like Flash [47]. The greatest hindrance in the design of STT-MRAM, and other spintronics circuits, is the lack of a compact MTJ model capable of accurately modeling temperature and voltage dependencies. Capturing these dependencies, in a compact model compatible with circuit simulators, is crucial for performing accurate Monte Carlo simulations to place yield and performance bounds on STT-MRAM. This chapter presents such a model implemented in Verilog-A. The model's simulation results were also compared to a model implemented using the LLG Micromagnetics Simulator [48] and actual device measurements from 135nm by 65nm CoFeB/MgO/CoFeB MTJs.

3.1 Modeling Dynamic Behavior

3.1.1 Magnetization Dynamics

The precessional motion of magnetization ($\vec{M}$) of the free layer of a MTJ, in the presence of an external magnetic field ($\vec{H}_{eff}$), can be very accurately modeled by the LLGE, Eq. 2.2 [23]. With the introduction of Slonczewski's spin-torque

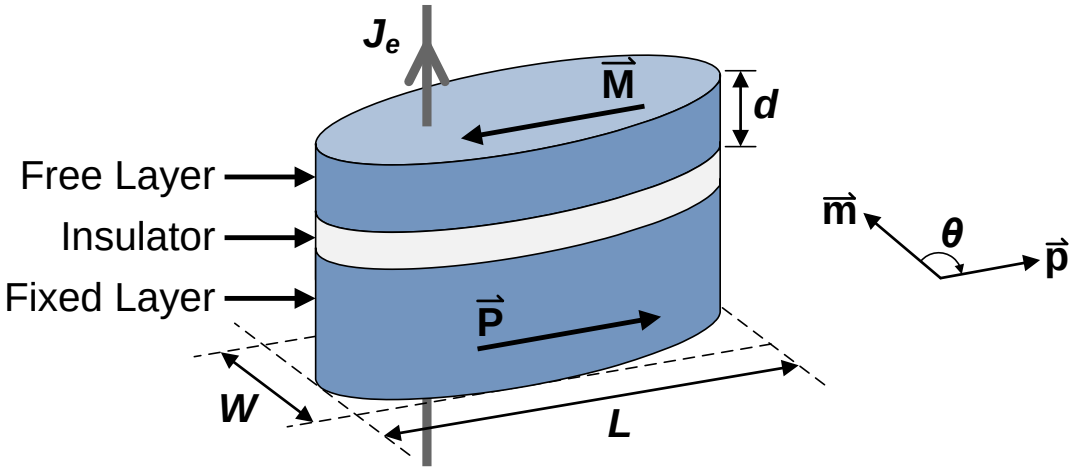


Figure 3.1: Sketch of basic MTJ structure.

transfer term [2], the normalized LLGE with STT is given by:

$$\frac{\partial \vec{m}}{\partial t} = -\gamma M_S \vec{m} \times \left(\vec{h}_{eff} + \frac{J_e}{J_C} b(\theta) (\vec{m} \times \vec{p}) - \alpha \frac{\partial \vec{m}}{\partial t} \right), \tag{3.1}$$

where $M_S = |\vec{M}|$, γ is the absolute value of the gyromagnetic ratio ($\gamma_e \mu_0$), $\vec{m}$ is the unit vector in the direction of $\vec{M}$, $\vec{p}$ is the unit vector in the direction of the magnetization of the fixed layer, $\vec{h}_{eff} = \vec{H}_{eff}/M_S$, J_e is the current density (see Fig. 3.1), θ is the angle between $\vec{m}$ and $\vec{p}$, and $\alpha > 0$ is the material-dependent Gilbert damping constant. The efficiency factor of spin-polarization ($b(\theta)$, see Fig. 3.2) is defined as:

$$b(\theta) = \left[-4 + (1+P)^3 \frac{\{3 + \cos(\theta)\}}{4P^{3/2}} \right]^{-1}, \tag{3.2}$$

where P is the percentage of electrons polarized in the $\vec{p}$ direction. The switching current density (J_C) has been modified to include thermally-activated switching [23]. For a constant pulse of duration τ, J_C is given by:

$$J_C = J_{C0} \left[1 - \frac{\ln(\tau/\tau_0)}{\Delta} \right], \tag{3.3}$$

where Δ is the thermal stability of the MTJ and $\tau_0 = (\gamma M_S)^{-1}$ is the natural time constant. Furthermore, the characteristic current density (J_{C0}) is defined

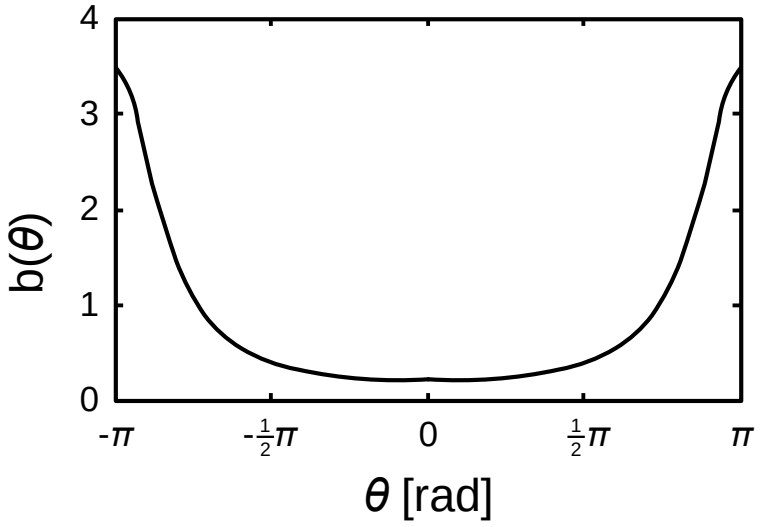


Figure 3.2: Magnitude of the efficiency factor of spin-polarization vs. θ for $P = 0.65$.

as:

$$J_{C0} = \gamma M_S \frac{eM_S d}{g_e \mu_B}, \tag{3.4}$$

where e is the absolute value of electron charge, d is the thickness of the free layer, g_e is the Landé factor for electrons, and μ_B is the Bohr magneton [49].

3.1.1.1 Effective Magnetic Field

The effective magnetic field ($\vec{H}_{eff}$) is given by:

$$\vec{H}_{eff} = \vec{H}_{ext} + \vec{H}_{dem} + \vec{H}_{an}, \tag{3.5}$$

where $\vec{H}_{ext}$ is the external applied magnetic field, $\vec{H}_{dem}$ is the demagnetization field, and $\vec{H}_{an}$ is the magnetocrystalline anisotropy field. The demagnetization field (shape anisotropy) varies with the geometry of the free layer and is modeled as $\vec{H}_{dem} = N\vec{M}$. If the free layer is assumed to be a very flat ellipsoid, the factors of the demagnetization tensor N, calculated by Osborn [50], are:

$$N_X = \frac{d}{L}\left(1 - e^2\right)^{1/2} \frac{K - E}{e^2}, \tag{3.6}$$

$$N_Y = \frac{d}{L}\frac{K - (1-e^2)\,E}{e^2\,(1-e^2)^{1/2}}, \tag{3.7}$$

$$N_Z = 1 - \frac{d}{L}\frac{E}{(1-e^2)^{1/2}}, \tag{3.8}$$

where K and E are the complete elliptic integrals of the first and second kind whose argument is:

$$e = \left(1 - W^2/L^2\right)^{1/2}. \tag{3.9}$$

3.1.1.2 Temperature Dependencies

In the dynamic equations, only the magnetization saturation (M_S) and the spin-polarization (P) vary with temperature. For temperatures below the Curie temperature (T_C), we can use the Weiss theory of ferromagnetism [51] to model:

$$M_S\,(T) = M_{S0}(1 - T/T_C)^{\beta}, \tag{3.10}$$

where M_{S0} is the magnetization saturation at absolute zero and β is the material-dependent critical exponent (see Fig. 3.3) [52]. Similarly, the temperature dependence of spin-polarazation has been extensivly studied and shown to be:

$$P\,(T) = P_0\left(1 - \alpha_{sp}T^{3/2}\right) \tag{3.11}$$

where P_0 is the spin-polarazation at absolute zero and α_{sp} is a material and geometric dependent constant [33].

3.1.2 Tunnel Magnetoresistance

Temperature variations in MTJ conductance ($G\,(\theta)$) are modeled in Shang et al. [33] by modifying the Julliere model. Julliere's model, Eq. 2.4, is reproduced

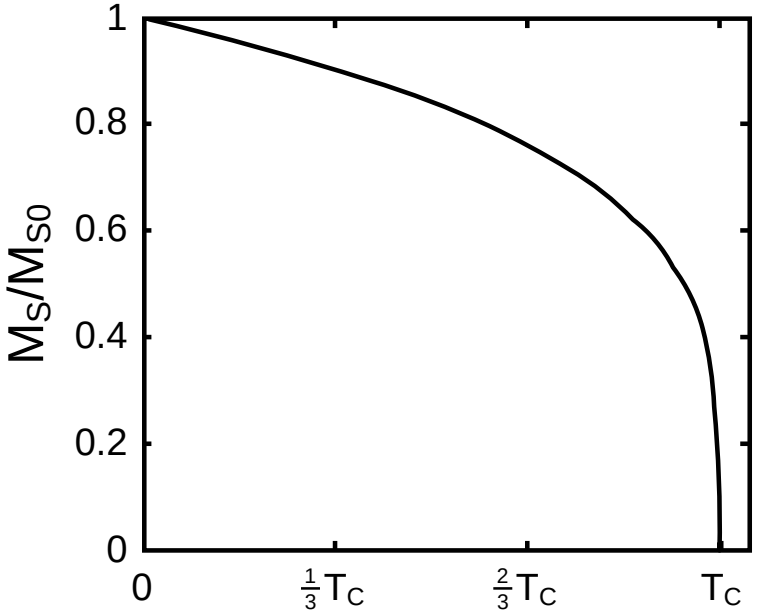


Figure 3.3: Normalized plot of magnetization saturation for generic ferromagnetic materials.

here with $P_1 = P_2 = P$:

$$G(\theta) = G_T\left\{1 + P^2\cos(\theta)\right\} + G_{SI}. \tag{3.12}$$

As a reminder, the variation of G_T due to temperature is negligible, whereas $G_{SI} \propto T^{4/3}$. Using $\theta = 0°$ for parallel magnetization and $\theta = 180°$ for anti-parallel, the tunnel magnetoresistance with zero applied bias voltage (TMR_0) can be expressed as:

$$TMR_0(T) = \frac{2P_0^2\left(1 - \alpha_{sp}T^{3/2}\right)^2}{1 - P_0^2(1 - \alpha_{sp}T^{3/2})^2 + \frac{G_{SI}(T)}{G_T(T)}}. \tag{3.13}$$

The Jullièe model fails to predict the effects of a bias voltage on TMR [31]. However, this can be rectified with the addition of a simple fitting function:

$$TMR(T,V) = \frac{TMR_0(T)}{1 + \left(\frac{V}{V_0}\right)^2}, \tag{3.14}$$

where V_0 is the voltage at which TMR is halved.

Geometric Parameters		LLGE Damping		Conductance	
W	65 [nm]	α	0.05	G_T	1.07 [mS]
L	135 [nm]	**Magnetization Saturation**		G_{SI}	0 [mS]
d	1.8 [nm]	M_{S0}	1100 [emu/cc]	**Demagnetization Tensor (Calculated)**	
t_{ox}	0.9 [nm]	T_C	1420 [K]		
Spin Polarization		β	0.4	N_X	0.0113
P_0	0.725	**TMR V_{BIAS} Fitting**		N_Y	0.0198
α_{sp}	2×10^{-5} [$K^{-3/2}$]	V_0	65 [nm]	N_Z	0.9689

Figure 3.4: Fitted MTJ parameters.

3.2 Model Verification

Implemented in Verilog-A, the compact model is comprised of two electrical terminals, an externally applied field vector, the initial state of magnetization, the demagnetization factors, and 13 device-specific parameters: 4 geometric parameters, 8 material-dependent parameters, and 1 empirically-derived parameter. The model was fitted to a 135nm by 65nm CoFeB/MgO/CoFeB MTJ (see Fig. 3.4). For validation of the model, we compare to detailed micromagnetic simulations, previously published data, and experimental results from fabricated MTJ nanopillars.

3.2.1 Comparison to Measured Devices

The ability of Eqs. 3.12 and 3.13 to accurately model the temperature dependance of R_P and TMR is well established in literature. Shang et al. [33] managed to obtain excellent fitting for Al_2O_3 based MTJs before deviating at high temperatures due to the crystallization of the amorphous insulating layer. Similarly, Kou et al. [53] and Wiśniowski et al. [55] reported extremely good fitting for MgO-

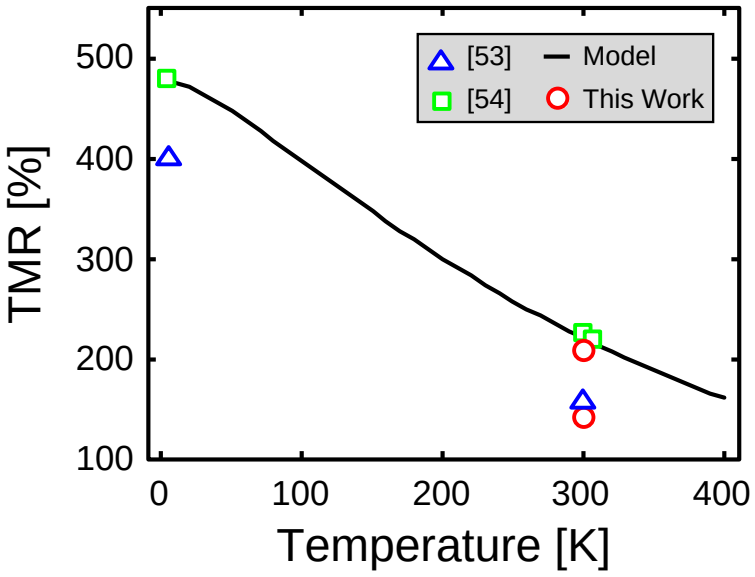


Figure 3.5: TMR vs. temperature: Verliog-A model (line), reported in [53] (triangles), reported in [54] (squares), and fabricated devices (circles).

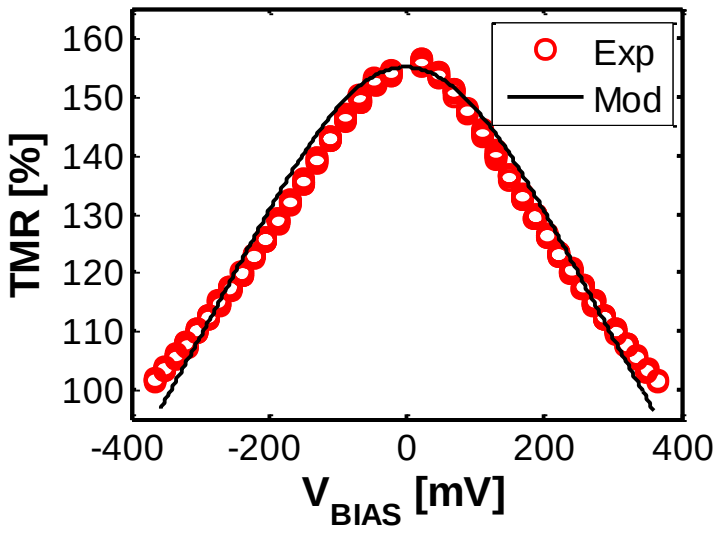


Figure 3.6: TMR vs. an applied bias voltage at 300K: Verliog-A model (black line) and fabricated devices (red circles).

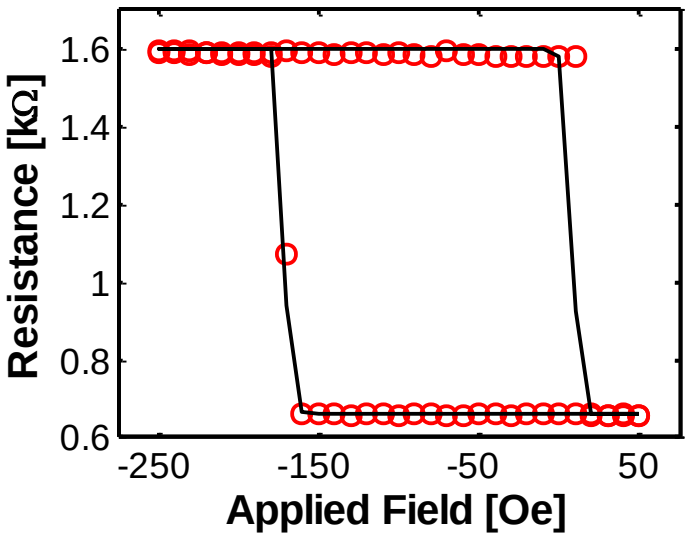


Figure 3.7: R-H hysteresis at 300K: Verliog-A model (black line) and fabricated devices (red circles).

and IrMn-based devices respectivly. Fig. 3.5 contains a loosely fitted curve of (3.13) for limited empirical data, as well as reported TMR values from literature. An excellent fitting of (2.1) to experimental data was obtained (Fig. 3.6), with an accuracy of ±3%. The steady-state accuracy of (3.1) at modeling switching thresholds for an applied external field is quite good and can be seen in Fig. 3.7.

3.2.2 Comparison to Micromagnetic Simulations

It is extremely difficult to accurately measure the switching characteristics of fabricated MTJs in the nanosecond regime. However, micromagnetic simulations are fully capable of accurately predicting their behavior [48]. As such, micromagnetic simulations were used to evaluate the switching behavior of the Verilog-A model in the nanosecond regime at different temperatures (see Fig. 3.8). Fig. 3.9 shows the time evolution of the resistance model ($R(t)$) and the micromagnetic

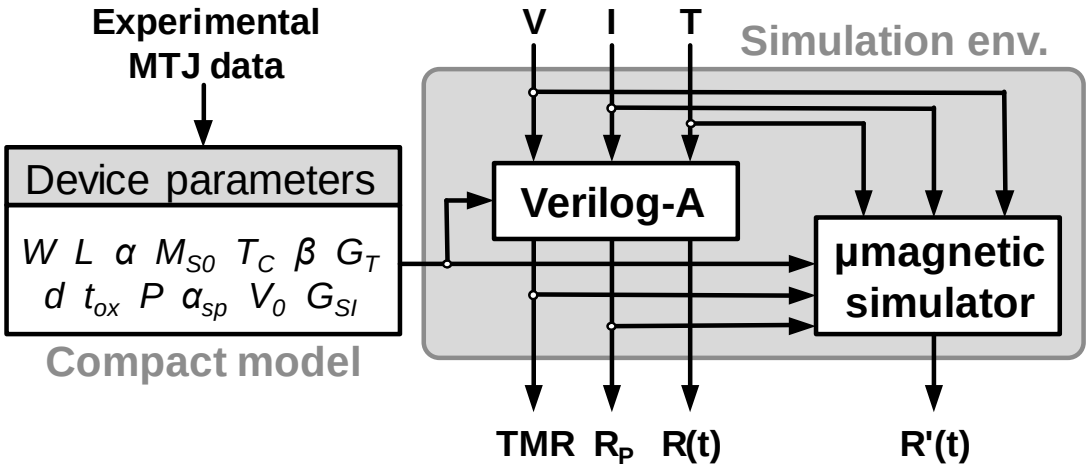


Figure 3.8: Process flow for evaluating Verilog-A model.

derived resistance ($R'(t)$) at 300K and 380K (expected operating temperature when integrated with CMOS).

Being based on a macrospin model, $R(t)$ does not account for non-uniformities in the free layer magnetization during switching. Despite this, the pre-switching oscillations and underdamped behavior of $R'(t)$ are still observed to a point in $R(t)$. This effect is captured by the shape anisotropy modeled by the demagnetization tensor (Eqs. 3.6, 3.7, and 3.8). Also, $R(t)$ manages to track the switching delay of $R'(t)$ across a wide range of temperatures and pulse shapes.

3.3 Statistical Characterization of MTJ Devices

3.3.1 MTJ Device Variability

While statistical variation of CMOS is generally well understood, similar characteristics for MTJs have not been well documented. This section uses a combination of fundamental equations and measured device characteristics to model the statistical behavior of MTJs. Figure 3.10(a) contains a plot of measured R_P

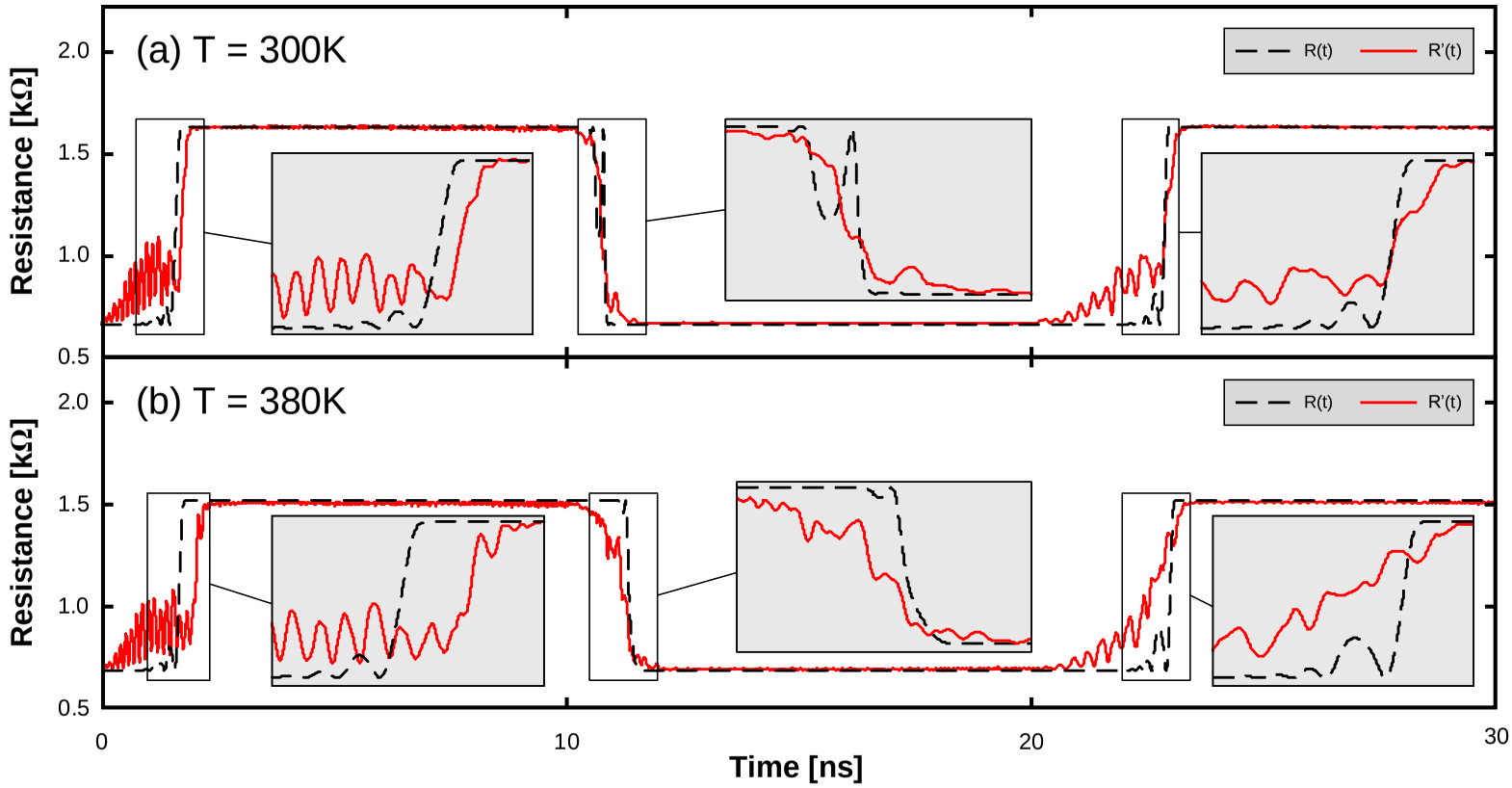


Figure 3.9: Resistance vs. time for an applied ±2V 100MHz square-wave at (a) 300K and (b) 380K. Verliog-A model ($R(t)$) is the black, dashed line and micromagnetic simulations ($R'(t)$) in the red, solid line.

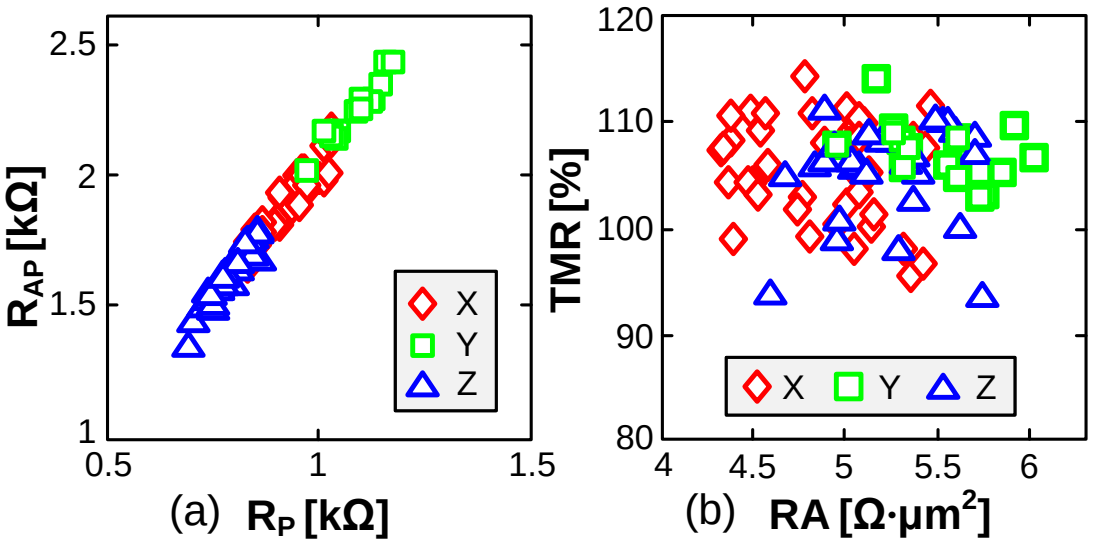


Figure 3.10: Measured (a) R_{AP} vs. R_P and (b) TMR vs. RA for MTJ nanopillars measuring $150 \times 45nm^2$ (X), $130 \times 50nm^2$ (Y), and $170 \times 45nm^2$ (Z)

	X	Y	Z
TMR [%]	105.7	107.3	105.3
σ_{TMR} [%]	4.7	2.7	4.6
RA $[\Omega \cdot \mu m^2]$	4.88	5.51	5.22
σ_{RA} $[\Omega \cdot \mu m^2]$	0.342	0.297	0.311

Table 3.1: Measured device statistics.

vs. R_{AP} for 105 MTJ nanopillars of varying size and target RAs. Variations in resistance and TMR are due to a combination of lithographic variations in the physical dimensions of the nanopillar and minute variations in the thicknesses of the up to 20 different layers in state-of-the-art MTJ processes [56]. The cumulative effects of these variations on RA and TMR can be easily measured [57], as shown in Fig. 3.10(b) and Table 3.1.

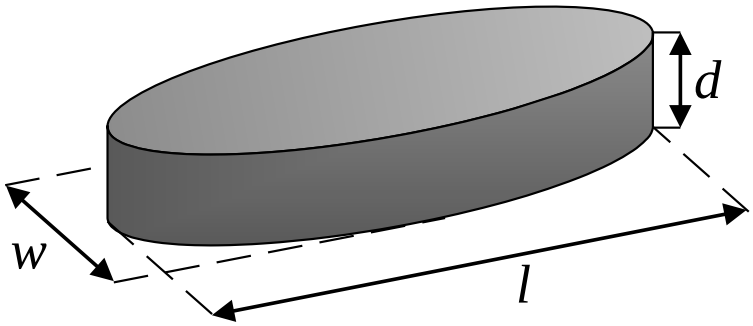


Figure 3.11: MTJ free layer dimensions.

3.3.2 Scaling of MTJ Current and Resistance

It would be useful to understand how the resistance and switching current of an MTJ changes as the device is scaled. A precessional-based switching model, modified to include thermally-activated switching, was used as a starting point. The switching current of an MTJ is given by:

$$I_C = I_{C0}\left[1 - \frac{\ln\left(\tau/\tau_0\right)}{\Delta}\right], \tag{3.15}$$

with the critical switching current (I_{C0}) given by:

$$I_{C0} = \frac{\alpha 4\pi e}{\eta\hbar} M_S^2 V, \tag{3.16}$$

where k_B is Boltzmann's constant, T is the absolute temperature in Kelvin, H_K is the out-of-plane uniaxial anisotropy, and E is the energy of anisotropy [58,59].

For an MTJ with free layer dimentions $l > w >> d$ [60], as shown in Fig. 3.11, the thermal stability of an MTJ is approximately:

$$\Delta = \frac{E}{k_B T} = \frac{H_K M_S}{2k_B T} V \approx d\left(\frac{1}{w} - \frac{1}{l}\right)\frac{M_S^2}{k_B T} V. \tag{3.17}$$

Dimensional scaling is performed to maintain a constant Δ in order to ensure the long-term non-volatility of the MRAM. If dimensions l and w of the MTJ are scaled by a factor λ to manipulate I_{C0} and $R_{P/AP}$, then to keep Δ constant, d must scale by $\lambda^{-1/2}$. This results in $I_{C0} \propto lwd \rightarrow \lambda^{3/2}$ and $R_{P/AP} \propto l^{-1}w^{-1} \rightarrow \lambda^{-2}$.

CHAPTER 4

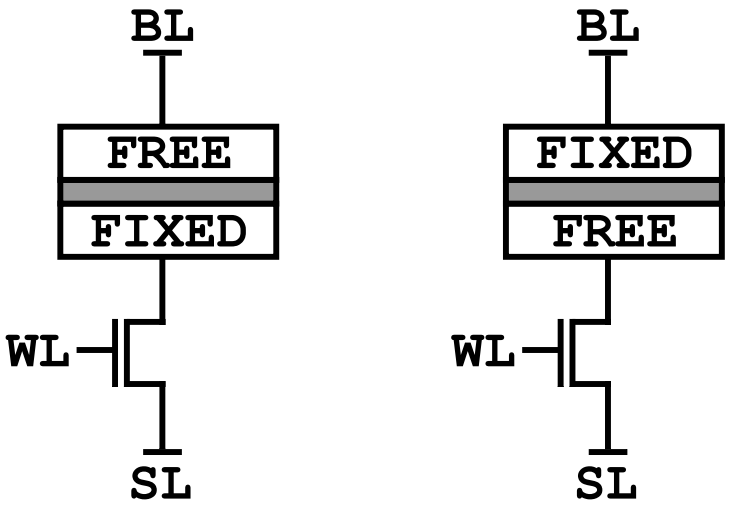
Memory Architectures

Several different types of memory architectures exist for STT-MRAMs. At the cell level, many architectures are tailored to certain MTJ characteristics, more specifically to the ratio of the critical writing currents $I_C(P \rightarrow AP)$ and $I_C(AP \rightarrow P)$. Other cell architectures attempt to exploit the different thresholds between reading and writing current to increase effective memory density. At the array level, several different subarraying techniques are employed to maximize performance and minimize area.

4.1 Cell Architectures

4.1.1 1T-1MTJ

There are two widely used 1T-1MTJ STT-MRAM cell architectures, the "conventional" cell (Fig. 4.1(a)) and the "reverse" cell (Fig. 4.1(b)) [61]. The "conventional" architecture gets its name from that fact that most MTJ are deposited with the fixed layer on the bottom. A smooth deposition surface is required to form a high quality pinning layer capable of generating the fixed layer [62]. The surface roughness introduced by various film deposition steps generally makes depositing the pinning layer on the top of the MTJ stack impractical. This means that it is easier to connect the fixed layer of the MTJ to the access transistor and



(a) Conventional **(b)** Reversed

Figure 4.1: 1T-1MTJ memory cell architectures.

the free layer to the bitline.

The "reverse" structure is built exactly as it sounds, with the fixed and free layers connected in the reverse fashion of the "conventional" cell. The "reverse" architecture attempts to match the inherent driving current asymmetry of the access transistor to the asymmetric switching currents of the MTJ [16]. It is widely assumed that the trade-off between these two architectures depends solely upon the ratio of the critical writing currents $I_C(P \rightarrow AP)$ and $I_C(AP \rightarrow P)$. If this ratio is greater than 1, a "reverse-connected" architecture should be used. Otherwise, a "conventional" architecture should be used. However, this is not exactly true as this assumption fails to take into account both the V_{GS} and V_{DS} operating points of the access transistors during the write opperation. For state-of-the-art technology nodes with supply voltages below $1V$, the cutoff point between these architectures is a writing current ratio closer to 1.5.

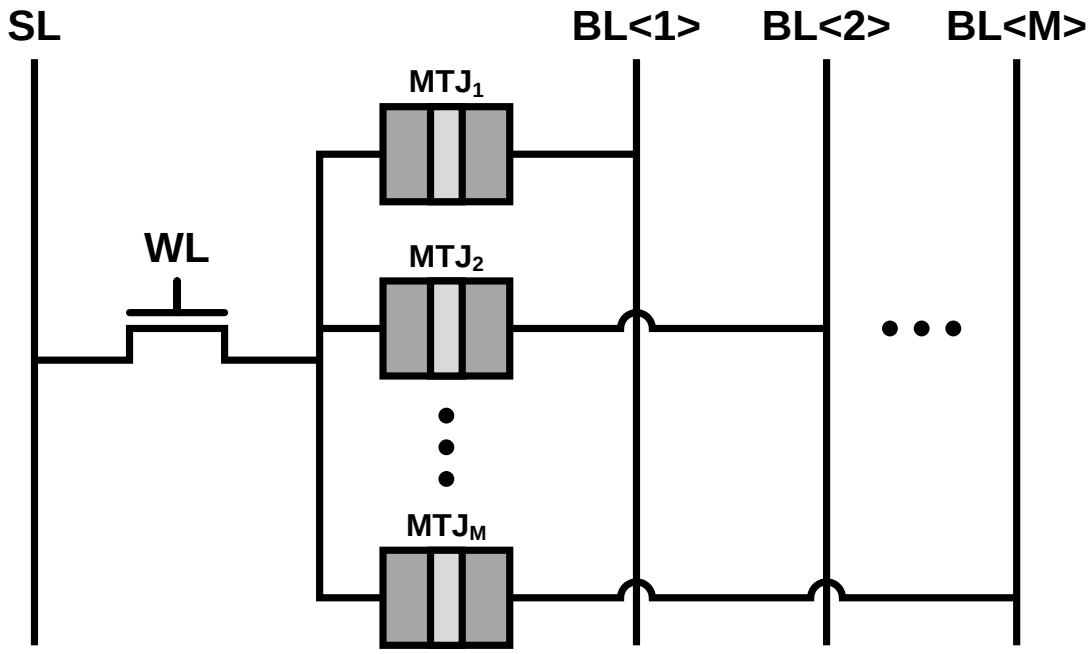


Figure 4.2: Shared memory cell architecture with M MTJs per transistor.

4.1.2 Shared

Access device "sharing" is one potential technique for increasing cell density. As shown in Fig. 4.2, one access transistor is connected to multiple MTJ devices, with additional bitlines to support independant access. This also allows the access transistor to be sized up to provide higher write current while maintaining the same overall memory density. However, there are several shortcomings associated with this technique which will be discussed in future sections.

4.1.3 Stacked

MTJ device "stacking" is another potential technique to increase cell density. "Stacking" works by connecting several different types of MTJ in series with one access transistor, shown in Fig. 4.3, in a similar fashion to multi-bit Flash cells.

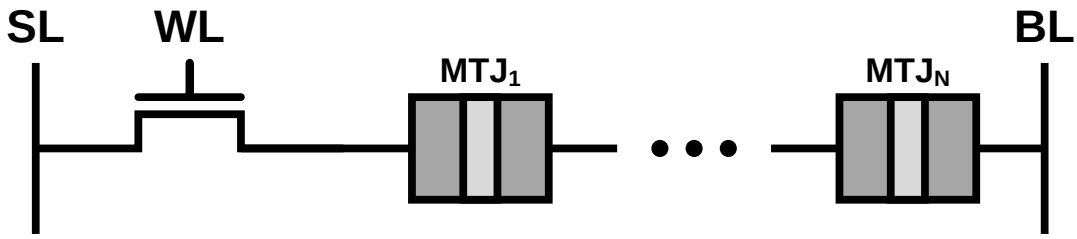


Figure 4.3: Stacked memory cell architecture with N MTJs per transistor.

To ensure functionality, the resistance and critical writing current of each MTJ need to be sufficiently different. Reading and writing to a cell would require multiple cycles, one for each bit.

4.2 Subarraying

4.2.1 1T-1MTJ

As stated before, subarraying is necessary for larger memories. Single, large memory arrays are slow and require additional buffering to drive very long wires. Breaking it up into several smaller subarrays allows the memory to operate faster and share peripheral circuitry. For 1T-1MTJ cell architectures, the number of cells per bitline is limited by the capacitance of the access transistor and the MTJ itself. Generally, a single bitline can support no more than 256 cells.

4.2.2 Shared Architectures

As mentioned earlier, there are several shortcomings to a "shared" MTJ architecture. During the write operation, there are multiple parasitic current paths that siphon current from the device being written to, forcing the access device to be sized up. These parasitic currents also have the potential to flip cells not being

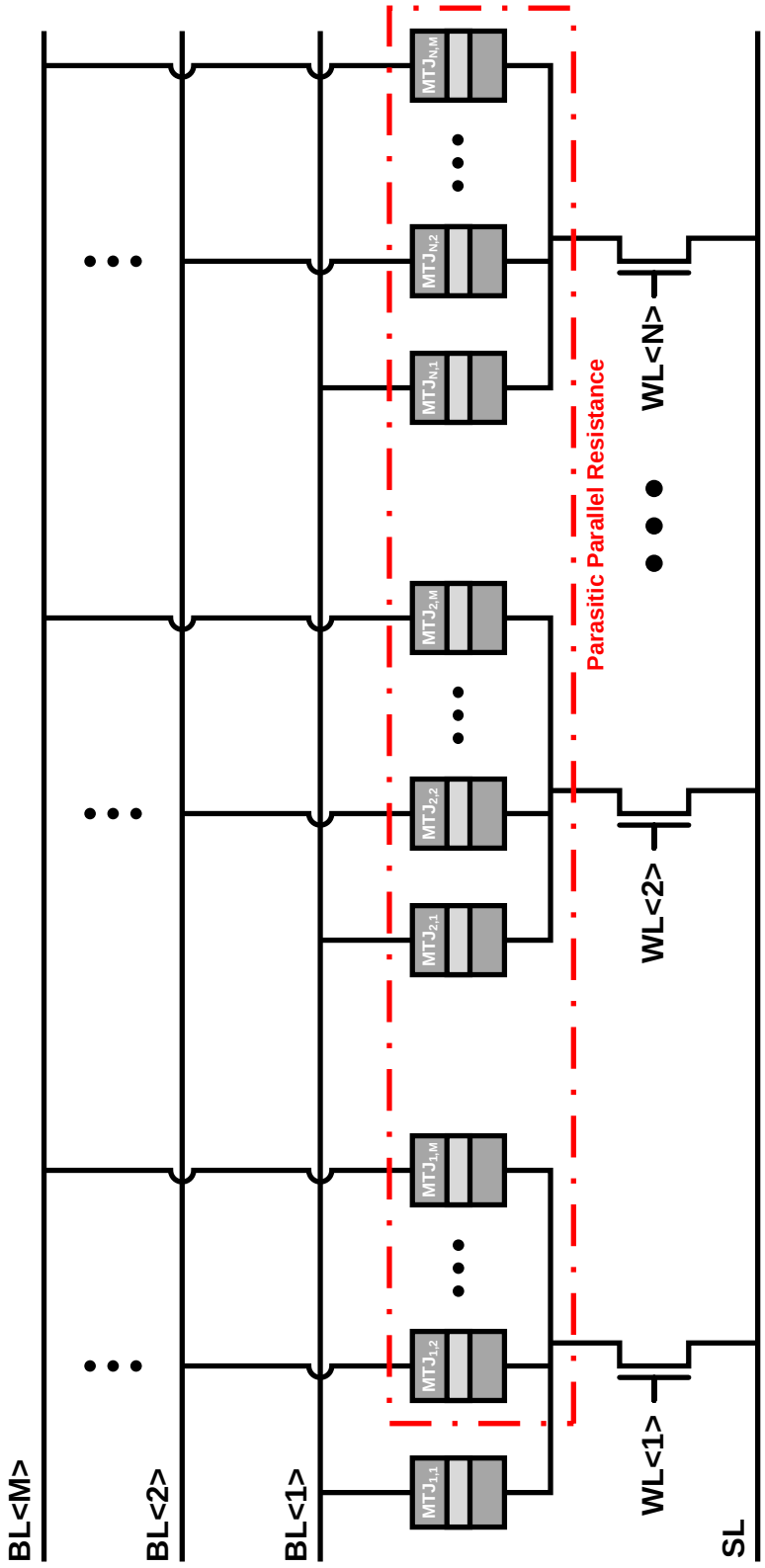


Figure 4.4: Shared architecture with M MTJs per transistor and N wordlines per subarray.

accessed. When reading, these parasitic paths lower the effective TMR that can be observed.

For M MTJs per access device and N wordlines per subarray, the effective TMR (TMR_{eff}) can be expressed as:

$$TMR_{eff} = \frac{2\left(N+M-1\right)-NM}{NM+\left(N-1\right)\left(M-1\right)\cdot TMR}\cdot TMR. \tag{4.1}$$

The worst case TMR degradation can be easily derived by finding the largest possible R_P and the smallest possible R_{AP}. If $R_{\|}$ is the parasitic parallel resistance shown in Fig. 4.4, then $R_{P,MAX}$ can be calculated as $R_P \parallel R_{\|}$, where:

$$R_{\|} = \frac{N+M-1}{\left(N-1\right)\left(M-1\right)}\cdot\left(1+TMR\right)\cdot R_P \tag{4.2}$$

for the case of all parasitic resistances in the antiparallel state. Similarly, $R_{AP,MIN}$ can be calculated as $R_P\cdot\left(1+TMR\right) \parallel R_{\|}$, where:

$$R_{\|} = \frac{N+M-1}{\left(N-1\right)\left(M-1\right)}\cdot R_P \tag{4.3}$$

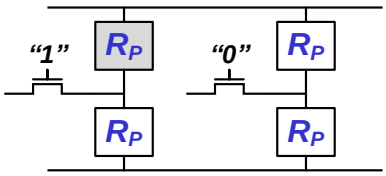
for the case of all parasitic resistances in the parallel state. For $M > 2$, TMR_{eff} is negative, limiting sharing to only two MTJs per access device.

In order to quantify the impact of sharing on the writing operation, the maximum allowable disturbance current a parasitic device can handle before a significant probability of switching occurs must first be calculated. For $M = 2$, since the ability to read is already limited, the four worst corner cases are shown in Fig. 4.5. For the corner case of Fig. 4.5(a), we require:

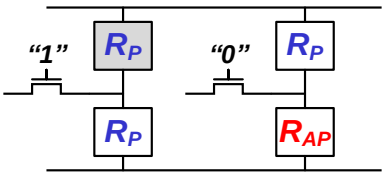
$$R_P\cdot I_{WRITE}\left(P\rightarrow AP\right) - R_P\cdot I_{READ}\left(P\rightarrow AP\right) \leq 2R_P\cdot I_{READ}\left(P\rightarrow AP\right). \tag{4.4}$$

Solving:

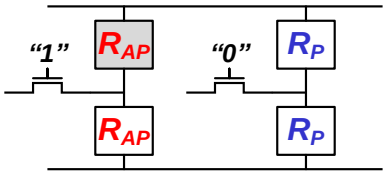
$$\frac{I_{READ}\left(P\rightarrow AP\right)}{I_{WRITE}\left(P\rightarrow AP\right)} \geq \frac{1}{3}. \tag{4.5}$$



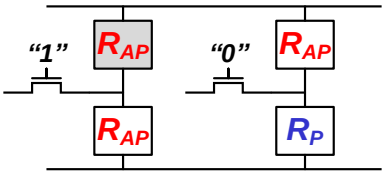
(a) Writing R_P with series R_P and R_P.



(b) Writing R_P with series R_P and R_{AP}.



(c) Writing R_{AP} with series R_P and R_P.



(d) Writing R_{AP} with series R_{AP} and R_P.

Figure 4.5: Worst-case writing configurations for sharing.

Similarly, for the other three cases, $I_{READ}/I_{WRITE} \geq 1/3$. This means that sharing can only be successfully implemented if the MTJ can tolerate a reading current greater than one third of the writing current without a significant probability of flipping.

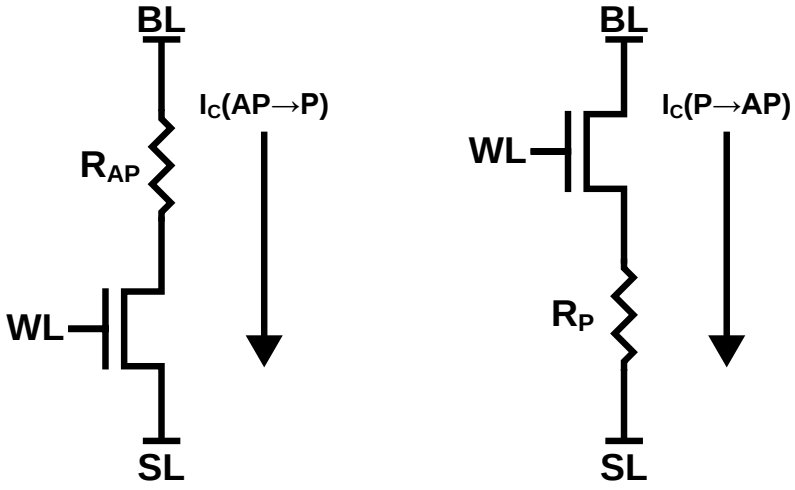
CHAPTER 5

Design-Space Analysis

In order to design an STT-MRAM with adequate design margin for high yield, one must consider all the implications of MTJ/CMOS integration. The design depends considerably on the underlying transistor technology, since a given CMOS technology constrains the design space due to the overhead and impact of the access transistor in each memory cell. The feasibility and yield of the memory are also heavily dependent upon the variation of the MTJs [63]. Previous work, such as Raychowdhury et al. [64], has failed to address these dependencies and provide the necessary framework for large-scale design. This chapter introduces the concept of the memory cell design space and a sensitivity analysis to optimize yield, power, and density for an STT-MRAM.

5.1 Defining the Design Space

The analysis in this chapter is done for a conventional 1T-1MTJ cell architecture as shown in Fig. 5.1, but can be generalized to any cell architecture. The writing currents for flipping the cell resistance are defined as $I_C(P \rightarrow AP)$ and $I_C(AP \rightarrow P)$. The design space of a single STT-MRAM memory cell can be illustrated using an R_{AP} vs. R_P plot as is shown in Fig. 5.2. The feasibility region is indicated by the shaded region. It contains all points (R_P, R_{AP}) in the design space so that a memory cell made with such an MTJ is functional. In the



(a) $I_C(AP \rightarrow P)$. **(b)** $I_C(P \rightarrow AP)$.

Figure 5.1: 1T-1MTJ cell architecture showing MTJ switching current for (a) AP to P and (b) P to AP.

design space, the two lower bounds are set by the read margin of the cell, while the two upper bounds are set by the write margin of the cell.

Lower bound $R_{P,MIN}$ is dependent on the implementation of the sense amplifier, and represents the minimum resistance required for reliable circuit operation. Additionally, $R_{AP,MIN}$ is determined by TMR_{MIN} (Fig. 5.3), the minimum TMR required for the read amplifier to differentiate between R_P and R_{AP}. Regardless of the specifics of the implementation, all sense amplifiers can be classified as either a voltage- or current-sensing topology. For a generic current-sensing read circuit, TMR_{MIN} can be expressed as:

$$TMR_{MIN} = \frac{2\Delta I_{ref}/I_{ref}}{1 - \Delta I_{ref}/I_{ref}}. \tag{5.1}$$

For I_{ref} flowing through the reference resistance R_{ref}, $I_{ref} + \Delta I_{ref,1}$ flows through R_P and $I_{ref} - \Delta I_{ref,2}$ through R_{AP}. When $\Delta I_{ref,1} = \Delta I_{ref,2} = \Delta I_{ref}$, TMR_{MIN} is minimized. Under this condition, $R_{ref} = 2(R_P \parallel R_{AP})$ and we can express

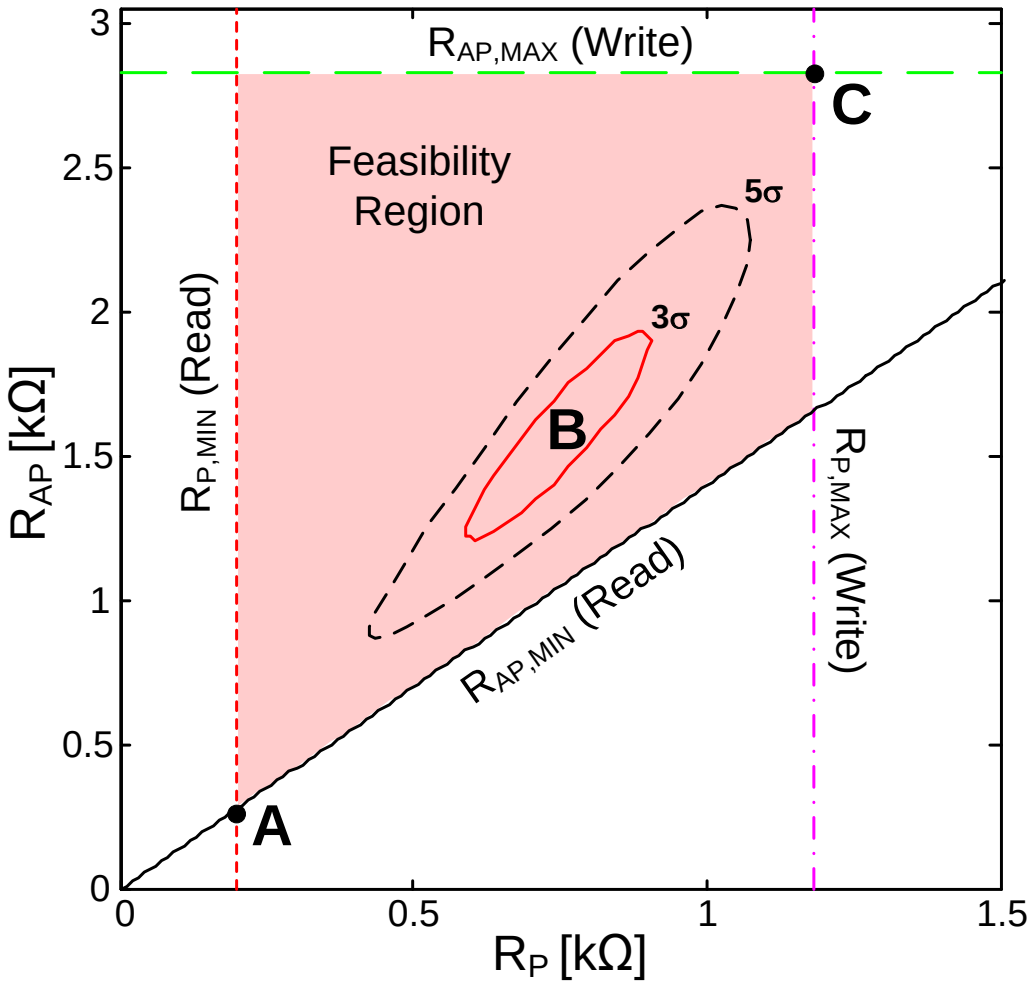


Figure 5.2: Design space, in a 65nm process, for $W_N = 2.0\mu m$, $I_C(P \rightarrow AP) = 500\mu A$, $I_C(AP \rightarrow P) = 375\mu A$, with an overlay of device X from Table 3.1.

TMR_{MIN} as a function of the normalized fractional sensing current ($\Delta I_{ref}/I_{ref}$). In Eq. 5.1, ΔI_{ref} must be chosen so that the read amplifier correctly evaluates across all transistor PVT variations. Similarly, TMR_{MIN} can be derived for voltage-sensing topologies. It should be noted that the lower bounds $R_{P,MIN}$ and $R_{AP,MIN}$, while critical to the readability of the cell, are almost completely independent of the MTJs used. The only requirement is that sensing time and current (I_{ref}) be small enough so as not to disturb the cell during the read

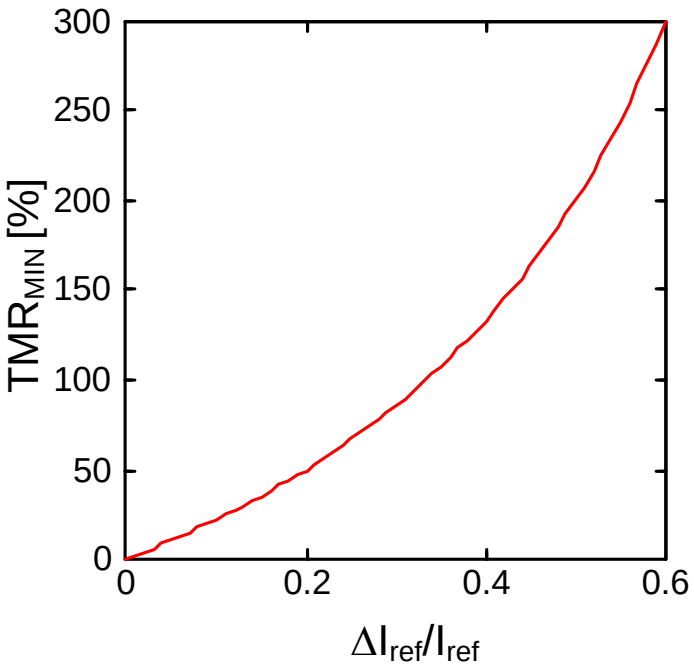
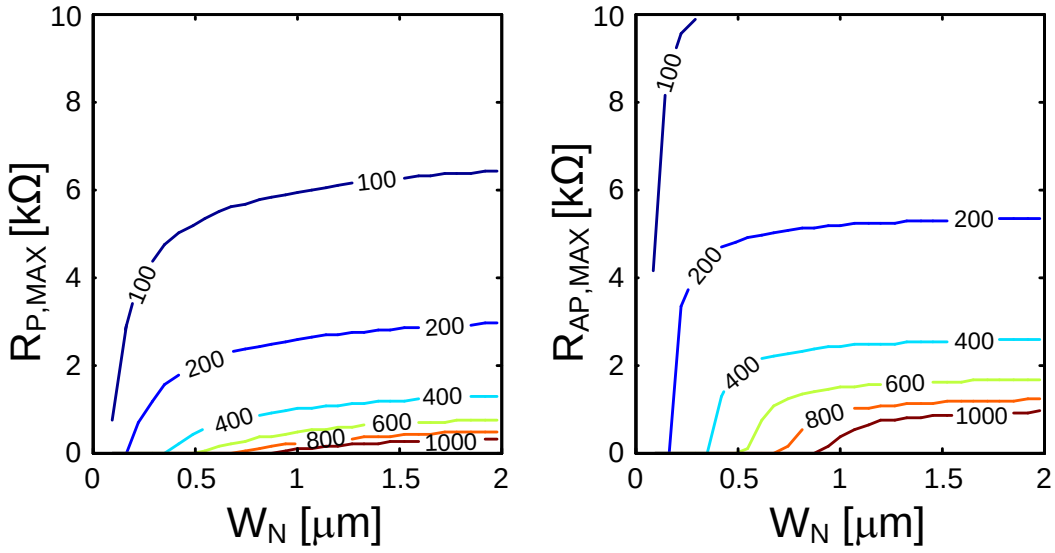


Figure 5.3: Design space lower bound TMR_{MIN} vs. $\Delta I_{ref}/I_{ref}$ for a current-sensing read circuit with ideal reference resistance $2(R_P \parallel R_{AP})$.

operation.

The upper bounds, $R_{P,MAX}$ and $R_{AP,MAX}$, are the maximum allowable resistances such that the access transistor, in a 1T-1MTJ configuration, is still able to provide the minimum critical writing currents $I_C(P \rightarrow AP)$ and $I_C(AP \rightarrow P)$. These upper bounds are consequently very sensitive to the specific characteristics of the MTJ device used. Transistor-level simulations are used to determine the relationship between R_{MAX}, I_C, and cell size (transistor width W_N) for a technology. Fig. 5.4 shows an example of such a simulation in a 65nm process. Using the conventional configuration from Fig. 5.1, W_N is swept along with R_{MAX}. The contours of the simulated current are shown.

Fig. 5.2 shows a specific MTJ cell and its associated statistical variation (the concentric ovals around point B) overlaid on the design space. The design-space



(a) $R_{P,MAX}$ vs. W_N.

(b) $R_{AP,MAX}$ vs. W_N.

Figure 5.4: (a) $R_{P,MAX}$ and (b) $R_{AP,MAX}$ at nominal V_{DD} for a 65nm process (I_C contours are in μA).

margin (DSM) can be defined as the number of σ's of MTJ variation before crossing any of the previously defined bounds. Defining DSM in terms of σ simplifies feasibility characterization to a single variable and thus allows yield to be quickly calculated. To a first order, 3σ, 4σ, 5σ, and 6σ of design margin roughly correspond to being able to reliably produce 1kbit, 32kbit, 4Mbit, and 1Gbit memory arrays.

5.2 Sensitivity Analysis and Design Example

Many variables, at both the circuit and device levels, affect the design space. In order to optimize a variable for a target memory specification, it must be determined how such a variable impacts the design space. This section introduces a design-space sensitivity (DSS) as a metric to quantify the behavior of the change

in design space as a function of various design parameters (V_{DD}, λ, J_C, RA, TMR, W_N, etc.).

5.2.1 Design-Space Sensitivity Analysis

First consider the points A, B, and C in Fig. 5.2. Points A and C correspond to the corner values of R_P and R_{AP} in the feasible design space. Point B represents the nominal MTJ at the center of the MTJ device distribution. For a positive design margin to exist, point B must fall somewhere between points A and C.

A "better" design space can be achieved from altering a design parameter, if a larger distribution of the MTJs (the number of σ) falls within the feasible region. Note that the improved design space is not simply increasing the area of the feasibility region, since the motion of point B must be considered as well. Recall that point A depends only slightly on the MTJ parameters. Therefore, the improvement (or deterioration) of the design space depends mostly on the change in DSM between points B and C as a function of a particular design variable.

Therefore, the design-space sensitivity to the parameter X is defined as:

$$DSS(X) = \frac{\partial(\frac{R_C - R_B}{\sigma})_{P/AP}}{\partial X}, \tag{5.2}$$

where R_B and R_C are taken as either R_P or R_{AP} at points B and C, thus defining the DSS along each dimension of the design space. $\frac{R_C - R_B}{\sigma}$ is the normalized distance between points B and C in the design space along the $R_{P/AP}$ dimension. Intuitively, the $DSS(X)$ describes the instantaneous rate of change in DSM to a particular design parameter X. The derivative loses positional information, and so the DSS is used in conjunction with the original plot of the design space to determine the benefit of tuning the design parameter X. For both the R_P and

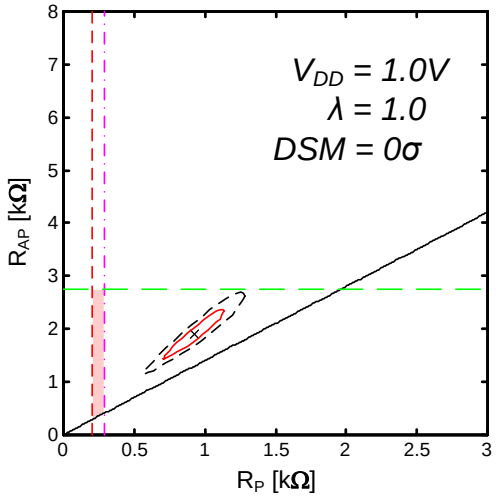
R_{AP} dimensions, if $DSS(X) > 0$, then the DSM is improved by increasing X, and if $DSS(X) < 0$, then DSM is improved by decreasing X. When the design-space sensitivities for the two dimensions conflict, the size of the design space in each dimension should then be taken into account.

5.2.2 Design Example

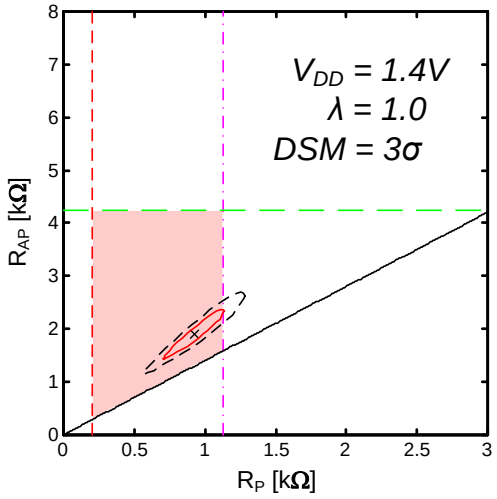
In this section, the sensitivity analysis was used to design a 4Mbit STT-MRAM with a $30F^2$ cell size (comparable to eDRAM) in a 65nm technology. Device X from Table 3.1, with $I_C(P \rightarrow AP) = 450\mu A$ and $I_C(AP \rightarrow P) = 300\mu A$, is the nominal MTJ and can be scaled by λ. Also, approximately 5σ of design margin is required for reasonable yield.

Fig. 5.5(a) shows the design space for a nominal $V_{DD} = 1.0V$ and $\lambda = 1.0$. The inner red oval is the 3σ variation of the MTJ, while the dashed, black oval represents the 5σ variation of the MTJ. Clearly, with nominal V_{DD} and λ, the memory is not functional. Fig. 5.6 shows that the design space is much more sensitive to V_{DD} than it is to λ. Therefore, we choose to scale V_{DD} to $1.4V$. Fig. 5.5(b) shows the new design space, with the 3σ bound at the edge of the design boundary.

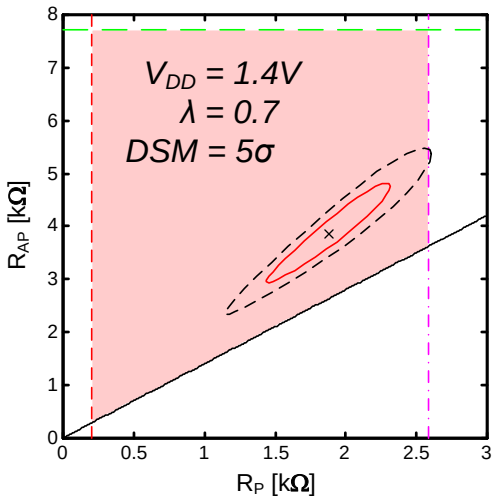
Scaling V_{DD} alone proves insufficient to meet the 5σ design margin required, and so λ was simultaneously scaled. Fig. 5.6(b) shows that scaling λ results in conflicting DSS. The R_{AP} margin improves more by scaling λ up, while the R_P margin improves by scaling λ down. However, Fig. 5.5(b) indicates that R_{AP} has considerable margin, and we can trade off some of that margin for improved margin in R_P. Therefore, we choose to scale λ down to 0.7. As we can see in Fig. 5.5(c), the desired 5σ bound on MTJ variation is essentially enclosed within the design space.



(a) $V_{DD} = 1.0V$ and $\lambda = 1.0$. 0σ design margin.

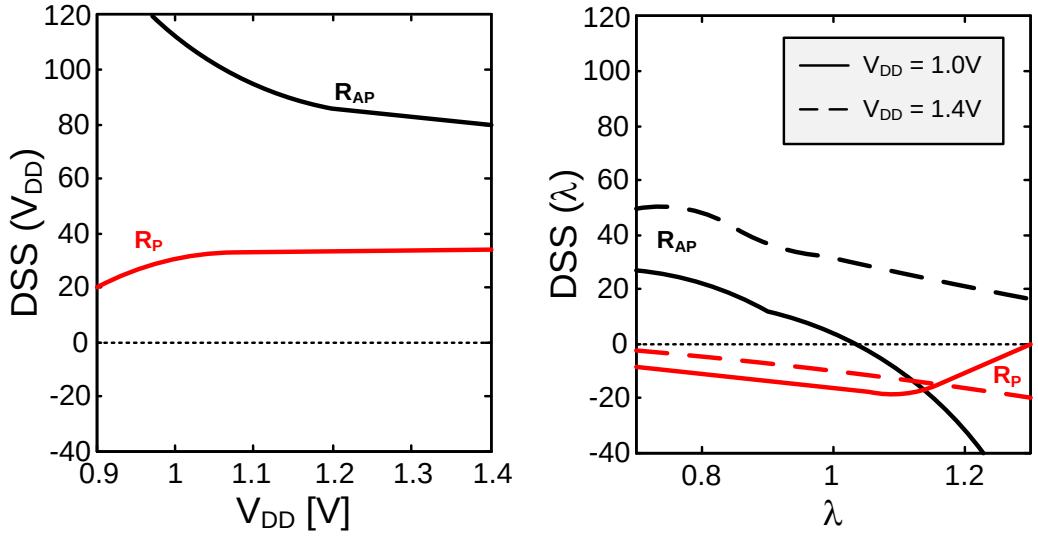


(b) $V_{DD} = 1.4V$ and $\lambda = 1.0$. 3σ design margin.



(c) $V_{DD} = 1.4V$ and $\lambda = 0.7$. 5σ design margin.

Figure 5.5: Design space, in a 65nm process, for a $30F^2$ cell ($W_N = 0.65\mu m$) for device X from Table 3.1: $I_C(P \rightarrow AP) = 450\mu A$, $I_C(AP \rightarrow P) = 300\mu A$. Inner red oval represents 3σ of MTJ device variation. Dashed, black oval corresponds to 5σ of MTJ variation.



(a) Sensitivity to V_{DD} vs. V_{DD}.

(b) Sensitivity to λ vs. λ.

Figure 5.6: Design-space sensitivity of parameters (a) V_{DD} and (b) λ in a 65nm technology.

5.3 Future Scalability

Scalability is an important feature for the success of a memory technology. Fig. 5.7 shows how scaling of the transistor technology impacts the design margin when using an MTJ that has a current density (J_C) of 6×10^6 A/cm^2 for 10ns P to AP switching and RA of $5\Omega \cdot \mu m^2$. The figure shows that SRAM equivalent sizes ($120F^2$) scale well, with a design margin more than sufficient to construct gigabit memories ($> 6\sigma$). However, as the cell size is decreased, the design margin begins to degrade below 45nm. The design margin practically disappears once an eDRAM-equivalent cell size ($30F^2$) is reached. However, if the MTJ current density scales with technology, also shown in Fig. 5.7, then this trend is reversed. By scaling J_C by as little as 4.5%, a constant design margin can be achieved between each technology node below 45nm.

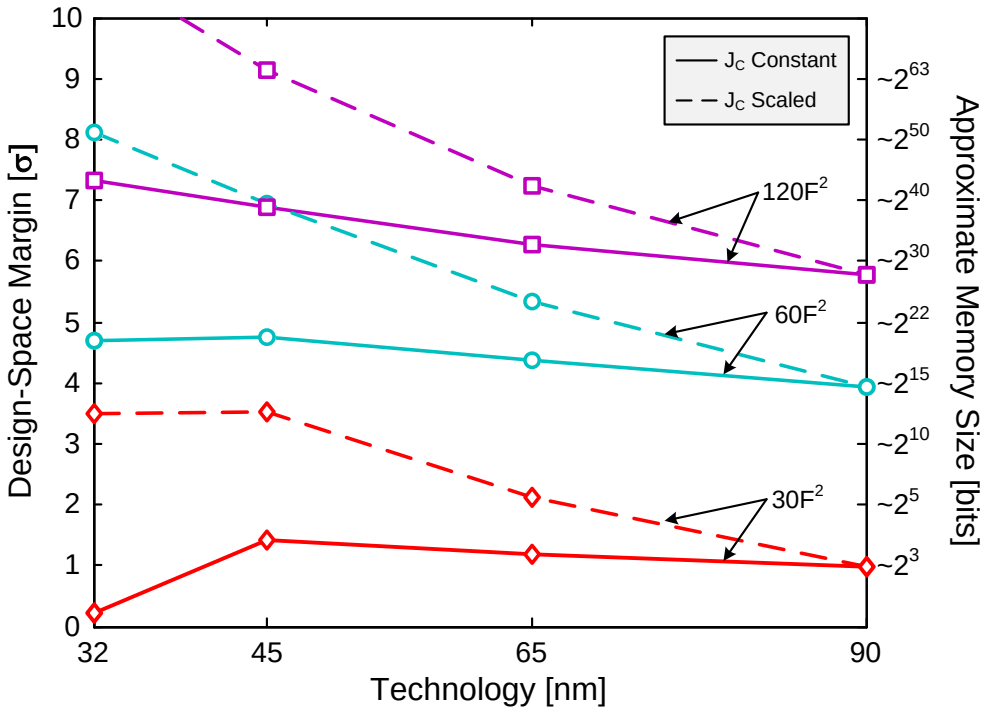


Figure 5.7: Design margin vs. technology node. For constant $J_c(P \rightarrow AP) = 6 \times 10^6$ A/cm^2 and $J_c(P \rightarrow AP)$ scaling by 4.5% each technology generation.

Tables 5.1, 5.2, and 5.3 contain critical switching current densities for Flash-, eDRAM-, and SRAM-equivalent cell sizes for RAs of 5, 10, and $15\Omega \cdot \mu m^2$, respectively. All table values correspond to 5σ of design margin and sub-10ns switching times. It should be noted that while larger RAs require smaller current densities (in order to meet voltage headroom constraints), they scale much better between successive technology nodes. Also, Flash-like cell sizes ($6F^2$) require current densities below 3×10^6 A/cm^2 in current technologies and less than 2×10^6 A/cm^2 in upcoming 22nm and 16nm technology nodes. Aggressive scaling of MTJ currents will be required to achieve Flash-like densities in future technologies.

Table 5.1: $J_C(P \rightarrow AP)$ for an RA of 5 $\Omega \cdot \mu m^2$

Equivalent Cell Size	J_C vs. Technology Node (10^6 A/cm^2)					
	90nm	65nm	45nm	32nm	22nm‡	16nm‡
FLASH ($6F^2$)	2.72	2.62	2.65	2.20	2.04	1.86
eDRAM ($30F^2$)	4.60	4.58	4.57	4.31	4.23	4.12
SRAM ($120F^2$)	6.67	6.76	6.86	7.07	7.17	7.30

‡Predicted

Table 5.2: $J_C(P \rightarrow AP)$ for an RA of 10 $\Omega \cdot \mu m^2$

Equivalent Cell Size	J_C vs. Technology Node (10^6 A/cm^2)					
	90nm	65nm	45nm	32nm	22nm‡	16nm‡
FLASH ($6F^2$)	2.22	2.14	2.19	1.82	1.69	1.56
eDRAM ($30F^2$)	3.27	3.28	3.33	3.20	3.18	3.16
SRAM ($120F^2$)	4.27	4.33	4.44	4.59	4.68	4.78

‡Predicted

Table 5.3: $J_C(P \rightarrow AP)$ for an RA of 15 $\Omega \cdot \mu m^2$

Equivalent Cell Size	J_C vs. Technology Node (10^6 A/cm^2)					
	90nm	65nm	45nm	32nm	22nm‡	16nm‡
FLASH ($6F^2$)	1.83	1.81	1.82	1.67	1.62	1.56
eDRAM ($30F^2$)	2.56	2.59	2.63	2.58	2.59	2.61
SRAM ($120F^2$)	3.17	3.21	3.33	3.47	3.55	3.65

‡Predicted

CHAPTER 6

Memory Design

In this chapter, the design flow of three test chips implemented in 90nm, 65nm, and 45nm processes is described. Several architectures from Chapter 4 were selected for testing. Each design was subjected to the analysis outlined in Chapter 5 in order to optimize read/write performance, memory density, and energy considerations. Each chip was designed to operate with MTJs fabricated by UCLA's Western Institute of Nanoelectronics (WIN). MTJ device specifications are detailed in [65], [66] and [67]. However, manufacturing requirements and restrictions made integration of MTJs with CMOS not possible on the test chips. A brief explanation is provided with a discussion of the process flow for MTJ/CMOS integration before the chip design is described.

6.1 MTJ/CMOS Integration

As mentioned before, MTJs are well suited for integration into a commercial CMOS process flow. In this flow, the deposition of the insulating oxide barrier is critical to the performance of the MTJ. If the layer is too thin ($< 0.7nm$) the MTJ does not exhibit any TMR, due to the formation of pin holes and soft points shorting the barrier. If the layer is too thick ($> 2.5nm$), then the resistance of the device is too large [68]. The deposition surface also needs to be very smooth, whereas typical Al interconnects (with a $\langle 111 \rangle$ texture) are far too

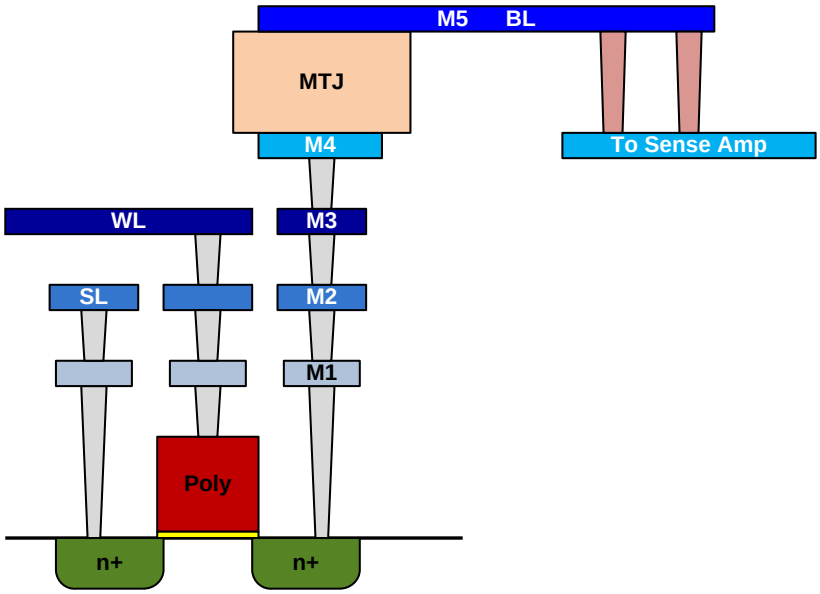


Figure 6.1: MTJ/CMOS integration at M4.

rough. However, the Cu interconnects available in the thin metal layers of modern state-of-the-art fabrication process are ideal for MTJ deposition [69]. MTJs are typically integrated after the thin Cu layer, usually M4 in most processes. Fig. 6.1 shows the side view of a typical 1T-1MTJ with full integration at M4. MTJ pillar dimensions down to 30nm can be accomplished with e-beam lithography, focused ion beam etching, or double patterning [70].

Unfortunately, it was not possible to integrate the MTJs from [65], [66] and [67] into the test chips due to fabrication restrictions. There are 45nm wafers ready for MTJ deposition once the fabrication flow is finalized. Meanwhile, the characterizations of test chips are done through simulation or with the addition of dummy resistor arrays on CMOS.

6.2 Test Chips

6.2.1 90nm Bulk CMOS

Fig. 6.2 shows a block diagram of the layout of our 90nm bulk CMOS test chip. The design work was performed for a conventional cell architecture and was intended to test the integration process flow. With a total of 6kbit, the memory had two 1kbit memory arrays, using RVT transistors with a cell size of $55F^2$, and two 2kbit memory arrays, using LVT transistors with a cell size of $30F^2$. For purposes of comparison, an SRAM cell in this technology is approximately $75F^2$. Included in the design are two resistor arrays with values ranging from a few hundred ohms to several kilohms (RSEL<16:0> and RSEL<33:17> in Fig. 6.2), representing a range of TMR from 0% to 1000%. Simulated and measured results for the read performance of R_P (logical 0) are shown in Table 6.1. Similar results for the read performance of R_{AP} (logical 1) are shown in Table 6.2. The drive current of the $50F^2$ cell was also measured to be approximately $300\mu A$. Estimates show that this level of drive current should easily allow for thermally activated switching with write times on the order of 10 to 20 nanoseconds.

6.2.2 65nm Bulk CMOS

In the 65nm process, the memory array was increased to 16kbit and included three cell sizes: $28F^2$, $35F^2$, and $50F^2$. Again, the design work assumed a conventional cell architecture. A "short-pulse" reading scheme was also introduced with a bidirectional write driver (Fig. 6.3) to improve read/write performance. Short-pulse reading works by delivering a large, but very short, current pulse to the MTJ and latching in its value. A dual-wordline voltage boosting scheme (dual-boosting) was implemented that allowed the drive current to be increased

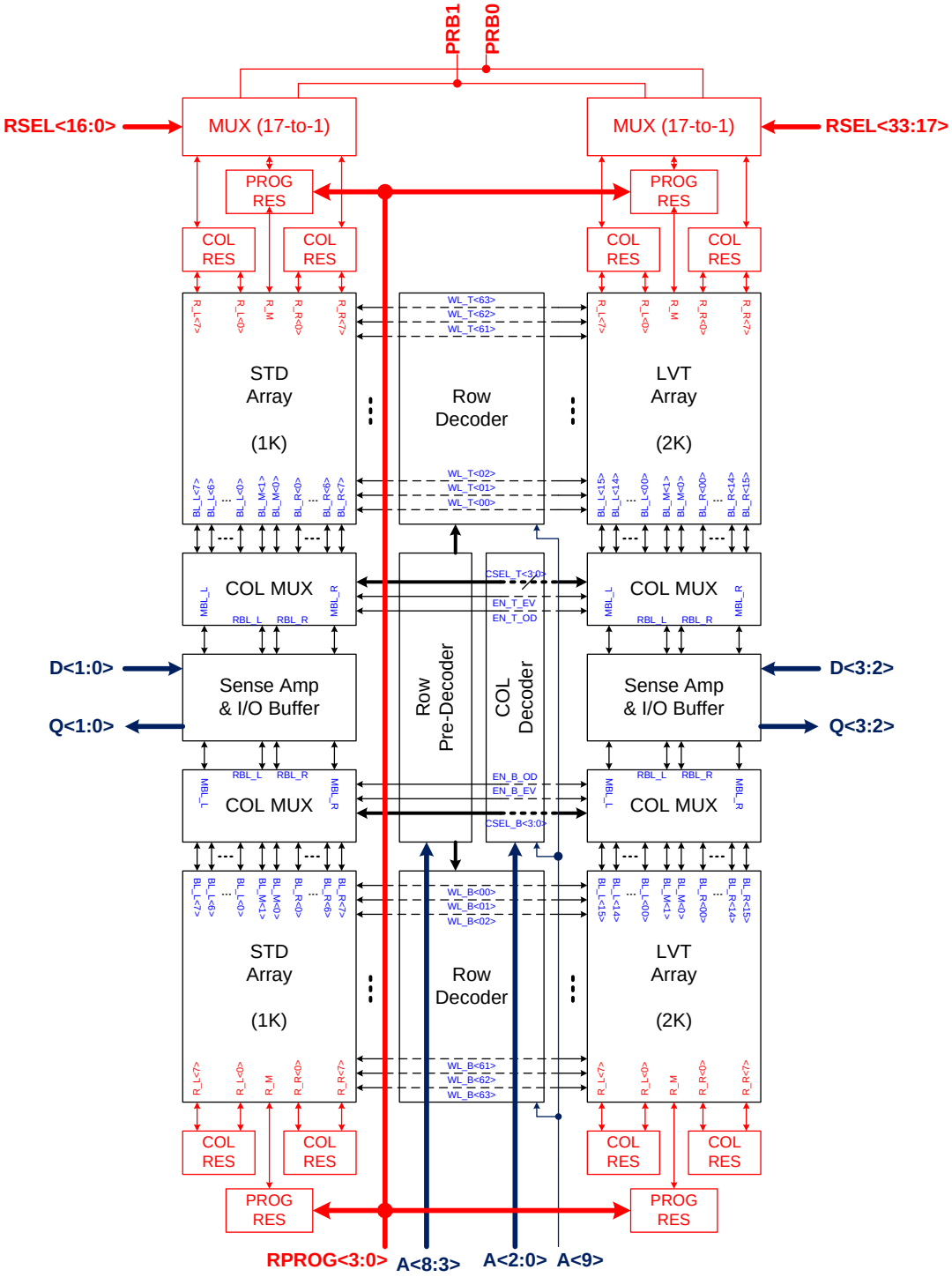


Figure 6.2: Block diagram of the 90nm test chip.

Table 6.1: Time to read R_P (90nm)

R_P [Ω]	TMR [%]	Simulated [ns]	Measured [ns]	Cell Size[F^2]
500	100	3.77	7.20	30
500	100	3.17	5.20	55
670	50	7.30	8.50	30
670	50	5.18	10.2	55
670	200	3.04	4.20	30
670	200	2.69	4.80	55

Table 6.2: Time to read R_{AP} (90nm)

R_P [Ω]	TMR [%]	Simulated [ns]	Measured [ns]	Cell Size[F^2]
500	100	2.00	2.10	30 & 55
670	50	2.60	2.70	30 & 55
670	200	2.37	2.70	30 & 50

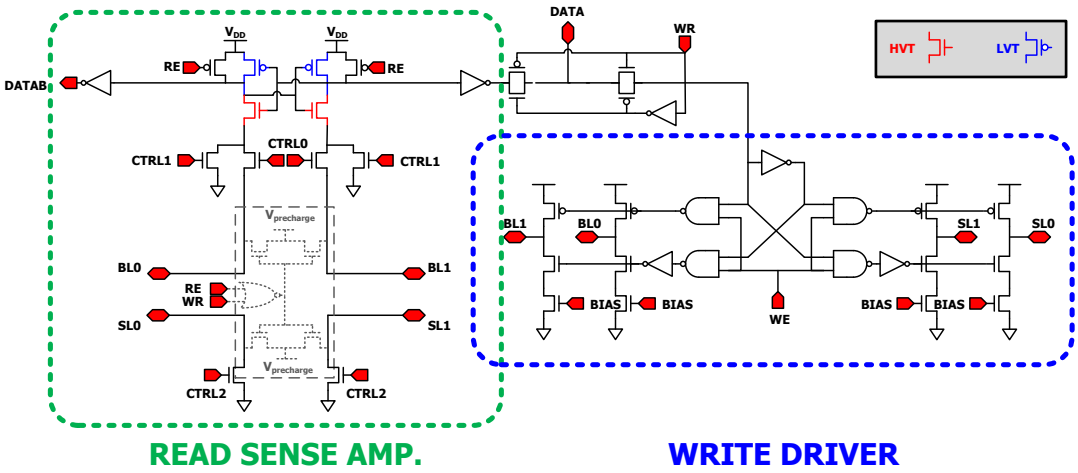


Figure 6.3: Read/Write driver for short-pulse reading.

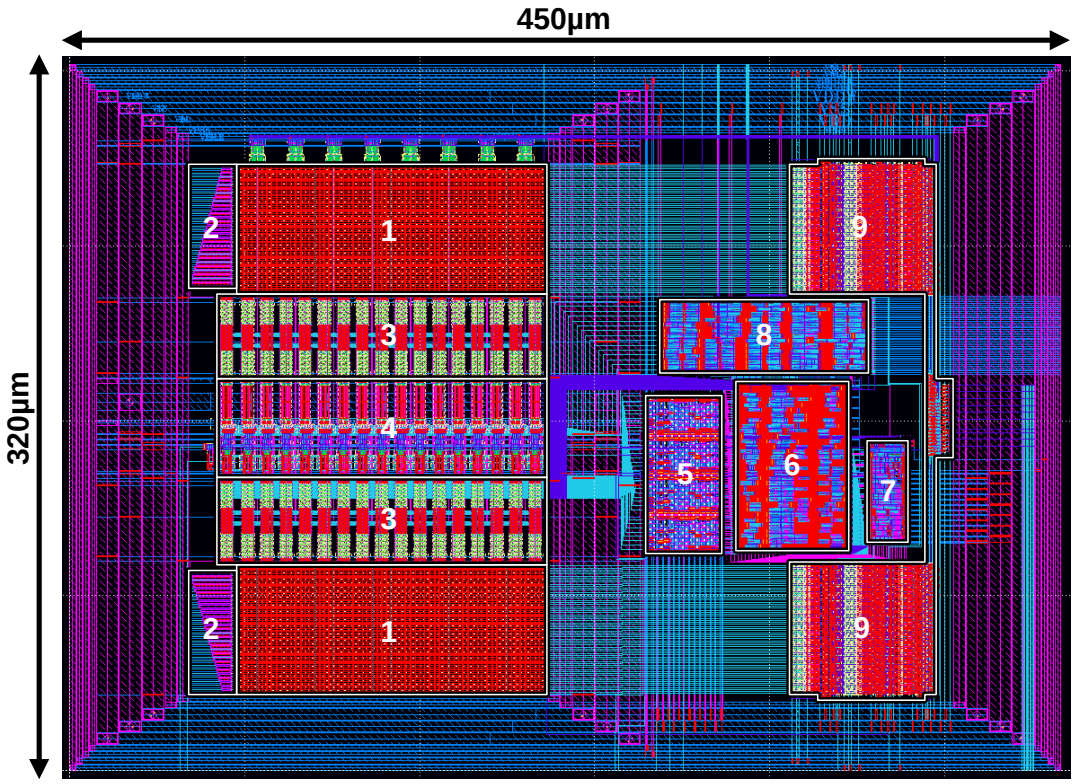


Figure 6.4: Cadence layout of the 65nm test chip: (1) 8kbit memory array, (2) resistor array, (3) muxes, (4) read/write circuitry, (5) read delay measurement circuitry, (6) configuration scan chain, (7) pulse generator, (8) data scan chain, and (9) decoder.

by 70% in our smallest cell sizes and 130% in our largest cells. The resistor arrays from the 90nm design were kept for CMOS characterization. Fig. 6.4 shows a layout of the 65nm design.

6.2.3 45nm SOI CMOS

The 45nm design moved the bulk CMOS to an SOI process. The memory size was increased to 32kbit and the same memory architecture was kept from our 65nm design (dual-boosting with short-pulse reading). It was possible to decrease the cell sizes to $17F^2$, $25F^2$, and $40F^2$, while still maintaining the same

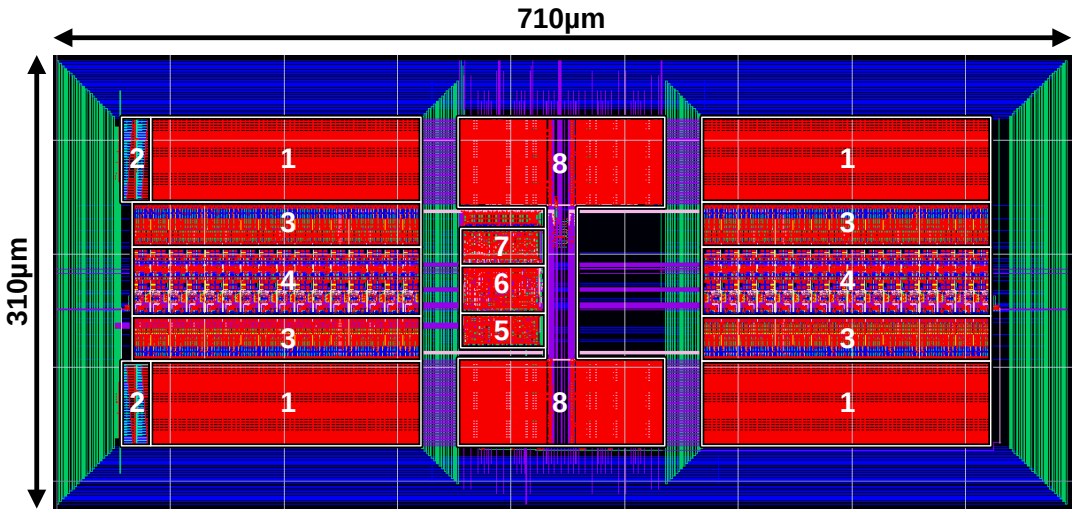


Figure 6.5: Cadence layout of the 45nm test chip: (1) 8kbit memory array, (2) resistor array, (3) muxes, (4) read/write circuitry, (5) configuration scan chain, (6) pulse generator and read delay measurement circuitry, (7) data scan chain, and (8) decoder.

drive current as the 65nm design. Overall, large improvements to the memory layout and organization were made. Fig. 6.5 shows a layout of the 45nm design.

6.2.4 Design Comparison

A comparison of the 90nm, 65nm and 45nm chips presented in this work is done with state-of-the-art STT-MRAMS from [61], [71], [72], and [73].

Year	2011	2011	2010	2009 [61]	2009 [71]	2009 [72]	2009 [73]
Designer	This Work	This Work	This Work	Qualcomm	NEC	Fujitsu & UT	Toshiba
Power Supply [V]	1.1/1.4	1.2/1.6	1.2	1.1/1.8	1/1.5	1.2/3.3	1.2
Architecture	Conventional (Dual-Boosted)	Conventional (Dual-Boosted)	Conventional	Reversed	2T-1MTJ (Boosted)	Conventional	2T-1MTJ
Memory Size	32kbit	16kbit	6kbit	32Mbit	32Mbit	16kbit	64Mbit
Process [nm]	45	65	90	45	90	130	65
Cell Size [F^2]	17, 25, 40	28, 35, 50	30, 60	11	70	140	36
Read Time	1-3ns	3-5ns	3-10ns	<100ns	60ns	8ns	11ns
Write Time	3-5ns	3-5ns	10-20ns	10ns-1ms	91ns	9-10ns	30ns

Figure 6.6: Design comparison of STT-MRAMs.

CHAPTER 7

Conclusion

The potential for STT-MRAM to become a true universal memory has been advanced by fabricating three test chips that demonstrate increased memory densities and performance when compared to published state-of-the-art STT-MRAMs. The 45nm test chip has a memory density 3.8 times greater on average when compared to published state-of-the-art STT-MRAMs. It has read access times ranging from 5 to 100 times faster, and is capable of delivering over $500\mu A$ of drive current to write in less than 5ns. While MTJ/CMOS integration was not accomplished, for reasons outside of our control, the chip represents the next step forward in making STT-MRAM a true universal memory. Analysis of the future scalability of MTJs shows that with a 22nm technology node, STT-MRAM will be able to achieve Flash-like densities, with sub-nanosecond performance, while using ten times less power.

7.1 Summary of Research Contributions

Specific accomplishments of this research are:

- Development of a physics-based MTJ macro-model capable of accurately modeling and predicting device behavior across temperature. The model was implemented in Verilog-A.

- Development of a design-space analysis toolset specifically for STT-MRAMs. These tools allowed an accurate and fair assessment of different memory architectures at the same time in order to maximize design margin, and minimize power and area for a fixed level of performance.
- Design of three STT-MRAM test chips:
  - 90nm bulk CMOS: 6kbit memory array, 2.5x density improvement over SRAM, 2-10ns read, $200\mu A$ write current.
  - 65nm bulk CMOS: 16kbit memory array, 4.3x density improvement over SRAM, new high-speed read/write architecture, $400\mu A$ write current.
  - 45nm SOI CMOS: 32kbit memory array, 7.1x density improvement over SRAM, average 3.8x density improvement over state-of-the-art STT-MRAMs, $500\mu A$ write current.

7.2 Future Work

Additional steps required to demonstrate that STT-MRAM can be developed into a true universal memory are:

- Extend the compact MTJ model to include localized heating effects, the backhopping phenomenon, and noise models.
- Explore stacked memory architecture to improve cell density.
- Continue to refine the DSM and DSS of the design-space analysis tools for integration with circuit simulators.
- Finalize the fabrication flow for MTJ/CMOS integration, including running performance and yield characterizations on an integrated 45nm test chip.

References

[1] S. A. Wolf *et al.*, "The Promise of Nanomagnetics and Spintronics for Future Logic and Universal Memory," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2155 –2168, December 2010. 2, 3, 6, 7, 8, 9

[2] J. C. Slonczewski, "Current-Driven Excitation of Magnetic Multilayers," *Journal of Magnetism and Magnetic Materials*, vol. 159, no. 1-2, pp. L1–L7, 1996. 3, 12, 13, 19

[3] I. Žutić, J. Fabian, and S. Das Sarma, "Spintronics: Fundamentals and Applications," *Rev. Mod. Phys.*, vol. 76, no. 2, pp. 323–410, April 2004. 6, 7, 9, 10, 12

[4] Y. C. Tao and J. G. Hu, "Superconducting Spintronics: Spin-Polarized Transport in Superconducting Junctions with Ferromagnetic Semiconducting Contact," *Journal of Applied Physics*, vol. 107, no. 4, 2010. 6, 9

[5] M. Julliere, "Tunneling Between Ferromagnetic Films," *Physics Letters A*, vol. 54, no. 3, pp. 225–226, 1975. 7, 15

[6] M. E. Flatte, "Spintronics," *IEEE Transactions on Electron Devices*, vol. 54, no. 5, pp. 907–920, May 2007. 7, 9, 10

[7] S. J. Pearton *et al.*, "Spintronics Device Concepts," *Circuits, Devices and Systems, IEE Proceedings*, vol. 152, no. 4, pp. 312–322, August 2005. 7, 9, 10

[8] J. Mathon and A. Umerski, "Theory of Runneling Magnetoresistance of an Epitaxial Fe/MgO/Fe(001) Junction," *Phys. Rev. B*, vol. 63, no. 22, p. 220403, May 2001. 7

[9] S. Das Sarma, J. Fabian, X. Hu, and I. Zutic, "Issues, Concepts, and Challenges in Spintronics," in *IEEE 58th DRC (Device Research Conference) Conference Digest*, June 2000, pp. 95–98. 7

[10] S. Kos *et al.*, "Modeling Spin-Polarized Electron Transport in Semiconductors for Spintronics Applications," *Computing in Science Engineering*, vol. 9, no. 5, pp. 46–52, September-October 2007. 9

[11] M. E. Flatté, Z. G. Yu, E. Johnston-Halperin, and D. D. Awschalom, "Theory of Semiconductor Magnetic Bipolar Transistors," *Applied Physics Letters*, vol. 82, no. 26, pp. 4740–4742, 2003. 9

[12] S. Chakrabarti *et al.*, "Spin-Polarized Light-Emitting Diodes with Mn-Doped InAs Quantum Dot Nanomagnets as a Spin Aligner," *Nano Letters*, vol. 5, no. 2, pp. 209–212, 2005. 10

[13] P. Bruno, Y. Suzuki, and C. Chappert, "Magneto-optical Kerr Effect in a Paramagnetic Overlayer on a Ferromagnetic Substrate: A Spin-Polarized Quantum Size Effect," *Phys. Rev. B*, vol. 53, no. 14, pp. 9214–9220, April 1996. 10

[14] J. Z. Sun, "Spin Angular Momentum Transfer in Current-Perpendicular Nanomagnetic Junctions," *IBM Journal of Research and Development*, vol. 50, no. 1, pp. 81–100, January 2006. 11, 12

[15] Z. Diao *et al.*, "Spin Transfer Switching in Dual MgO Magnetic Tunnel Junctions," *Applied Physics Letters*, vol. 90, no. 13, p. 132508, 2007. 12

[16] W. Zhu, H. Li, Y. Chen, and X. Wang, "Current Switching in MgO-Based Magnetic Tunneling Junctions," *Magnetics, IEEE Transactions on*, vol. 47, no. 1, pp. 156–160, January 2011. 12, 31

[17] K. Lee and S. Kang, "Control of Switching Current Asymmetry by Magnetostatic Field in MgO-Based Magnetic Tunnel Junctions," *Electron Device Letters, IEEE*, vol. 30, no. 12, pp. 1353–1355, December 2009. 12

[18] X. Yao, H. Meng, Y. Zhang, and J.-P. Wang, "Improved Current Switching Symmetry of Magnetic Tunneling Junction and Giant Magnetoresistance Devices with Nano-Current-Channel Structure," *Journal of Applied Physics*, vol. 103, no. 7, p. 07A717, 2008. 12

[19] R. H. Koch, J. A. Katine, and J. Z. Sun, "Time-Resolved Reversal of Spin-Transfer Switching in a Nanomagnet," *Phys. Rev. Lett.*, vol. 92, no. 8, p. 088302, February 2004. 12, 13

[20] H. W. Schumacher *et al.*, "Precessional Switching of the Magnetization in Microscopic Magnetic Tunnel Junctions (Invited)," *Journal of Applied Physics*, vol. 93, no. 10, pp. 7290–7294, 2003. 12

[21] T. L. Gilbert, "A Phenomenological Theory of Damping in Ferromagnetic Materials," *IEEE Transactions on Magnetics*, vol. 40, no. 6, pp. 3443–3449, November 2004. 12

[22] L. Landau and E. Lifshitz, "On the Theory of the Dispersion of Magnetic Permeability in Ferromagnetic Bodies," *Physikalische zeitschrift der Sowjetunion*, vol. 8, pp. 153–169, 1935. 12

[23] T. Moriyama *et al.*, "Tunnel Magnetoresistance and Spin Torque Switching in MgO-based Magnetic Tunnel Junctions with a Co/Ni Multilayer Electrode," *Applied Physics Letters*, vol. 97, no. 7, p. 072513, 2010. 13, 18, 19

[24] Z. M. Zeng *et al.*, "Effect of resistance-area product on spin-transfer switching in mgo-based magnetic tunnel junction memory cells," *Applied Physics Letters*, vol. 98, no. 7, p. 072512, 2011. 14

[25] Y. Higo *et al.*, "Thermal Activation Effect on Spin Transfer Switching in Magnetic Tunnel Junctions," *Applied Physics Letters*, vol. 87, no. 8, p. 082502, 2005. 14

[26] M. Pakala *et al.*, "Critical Current Distribution in Spin-Transfer-Switched Magnetic Tunnel Junctions," *Journal of Applied Physics*, vol. 98, no. 5, p. 056107, 2005. 14

[27] T. Aoki, Y. Ando, M. Oogane, and H. Naganuma, "Reproducible Trajectory on Subnanosecond Spin-Torque Magnetization Switching Under a Zero-Bias Field for MgO-Based Ferromagnetic Tunnel Junctions," *Applied Physics Letters*, vol. 96, no. 14, p. 142502, 2010. 14

[28] E. R. Nowak, M. B. Weissman, and S. S. P. Parkin, "Electrical Noise in Hysteretic Ferromagnet–Insulator–Ferromagnet Tunnel Junctions," *Applied Physics Letters*, vol. 74, no. 4, pp. 600–602, 1999. 14, 17

[29] S. X. Huang, T. Y. Chen, and C. L. Chien, "Spin Polarization of Amorphous CoFeB Determined by Point-Contact Andreev Reflection," *Applied Physics Letters*, vol. 92, no. 24, p. 242509, 2008. 15

[30] J. S. Moodera and G. Mathon, "Spin Polarized Tunneling in Ferromagnetic Junctions," *Journal of Magnetism and Magnetic Materials*, vol. 200, no. 1-3, pp. 248 – 273, 1999. 15

[31] Y. Lu *et al.*, "Bias Voltage and Temperature Dependence of Magnetotunneling Effect," *Journal of Applied Physics*, vol. 83, no. 11, pp. 6515–6517, 1998. 15, 16, 22

[32] X. Liu *et al.*, "Thermal Stability of Magnetic Tunneling Junctions with MgO Barriers for High Temperature Spintronics," *Applied Physics Letters*, vol. 89, no. 2, p. 023504, 2006. 15

[33] C. H. Shang, J. Nowak, R. Jansen, and J. S. Moodera, "Temperature Dependence of Magnetoresistance and Surface Magnetization in Ferromagnetic

Tunnel Junctions," *Phys. Rev. B*, vol. 58, no. 6, pp. R2917–R2920, August 1998. 15, 21, 23

[34] J. G. Simmons, "Generalized Formula for the Electric Tunnel Effect between Similar Electrodes Separated by a Thin Insulating Film," *Journal of Applied Physics*, vol. 34, no. 6, pp. 1793–1803, 1963. 15

[35] T. Zhu, X. Xiang, and J. Q. Xiao, "Bias Dependence of Tunneling Magnetoresistance on Ferromagnetic Electrode Thickness," *Applied Physics Letters*, vol. 82, no. 16, pp. 2676–2678, 2003. 15

[36] G. G. Cabrera and N. García, "Low Voltage I–V Characteristics in Magnetic Tunneling Junctions," *Applied Physics Letters*, vol. 80, no. 10, pp. 1782–1784, 2002. 15

[37] S. Chatterjee, S. Salahuddin, S. Kumar, and S. Mukhopadhyay, "Analysis of Thermal Behaviors of Spin-Torque-Transfer RAM: A Simulation Study," in *Low-Power Electronics and Design (ISLPED), 2010 ACM/IEEE International Symposium on*, August 2010, pp. 13–18. 16

[38] Y. Zhang *et al.*, "Micromagnetic Study of Hotspot and Thermal Effects on Spin-Transfer Switching in Magnetic Tunnel Junctions," *Journal of Applied Physics*, vol. 101, no. 10, p. 103905, 2007. 16

[39] S. Chaudhuri *et al.*, "Design of TAS-MRAM Prototype for NV Embedded Memory Applications," in *Memory Workshop (IMW), 2010 IEEE International*, May 2010, pp. 1 –4. 16

[40] J. Z. Sun *et al.*, "High-Bias Backhopping in Nanosecond Time-Domain Spin-Torque Switches of MgO-based Magnetic Tunnel Junctions," *Journal of Applied Physics*, vol. 105, no. 7, p. 07D109, 2009. 16, 17

[41] T. Min *et al.*, "Back-Hopping after Spin Torque Transfer Induced Magnetization Switching in Magnetic Tunneling Junction Cells," *Journal of Applied Physics*, vol. 105, no. 7, p. 07D126, 2009. 16, 17

[42] S. Ingvarsson *et al.*, "Low-Frequency Magnetic Noise in Micron-Scale Magnetic Tunnel Junctions," *Phys. Rev. Lett.*, vol. 85, no. 15, pp. 3289–3292, Oct 2000. 17

[43] K. B. Klaassen, J. C. L. van Peppen, and X. Xing, "Noise in Magnetic Tunnel Junction Devices," *Journal of Applied Physics*, vol. 93, no. 10, pp. 8573–8575, 2003. 17

[44] K. Shimazawa *et al.*, "Frequency Response of Common Lead and Shield Type Magnetic Tunneling Junction Head," *Magnetics, IEEE Transactions on*, vol. 37, no. 4, pp. 1684–1686, July 2001. 17

[45] A. F. M. Nor *et al.*, "Low-Frequency Noise in MgO Magnetic Tunnel Junctions," *Journal of Applied Physics*, vol. 99, no. 8, p. 08T306, 2006. 17

[46] Y. Huai, "Spin-Transfer Torque MRAM (STT-MRAM): Challenges and Prospects," *AAPPS Bulletin*, vol. 18, no. 6, pp. 33–40, December 2008. 18

[47] B. F. Cockburn, "The Emergence of High-Density Semiconductor-Compatible Spintronic Memory," in *International Conference on MEMS, NANO and Smart Systems, 2003. Proceedings.*, July 2003, pp. 321–326. 18

[48] M. R. Scheinfein, "LLG Micromagnetics Simulator." [Online]. Available: http://llgmicro.home.mindspring.com 18, 25

[49] J. Z. Sun, "Spin-Current Interaction with a Monodomain Magnetic Body: A Model Study," *Phys. Rev. B*, vol. 62, no. 1, pp. 570–578, July 2000. 20

[50] J. A. Osborn, "Demagnetizing Factors of the General Ellipsoid," *Phys. Rev.*, vol. 67, no. 11-12, pp. 351–357, June 1945. 20

[51] P. Weiss, "L'hypothèse du Champ Moléculaire et la Propriété Ferromagnétique," *J. Phys. Theor. Appl.*, vol. 6, no. 1, pp. 661–690, 1907. 21

[52] A. Raghunathan, Y. Melikhov, J. E. Snyder, and D. C. Jiles, "Modeling the Temperature Dependence of Hysteresis Based on Jiles-Atherton Theory," *IEEE Transactions on Magnetics*, vol. 45, no. 10, pp. 3954–3957, October 2009. 21

[53] X. Kou, J. Schmalhorst, A. Thomas, and G. Reiss, "Temperature Dependence of the Resistance of Magnetic Tunnel Junctions with MgO Barrier," *Applied Physics Letters*, vol. 88, no. 21, p. 212115, 2006. 23, 24

[54] P. Padhan *et al.*, "Frequency-Dependent Magnetoresistance and Magnetocapacitance Properties of Magnetic Tunnel Junctions with MgO Tunnel Barrier," *Applied Physics Letters*, vol. 90, no. 14, p. 142105, 2007. 24

[55] P. Wiśniowski *et al.*, "Temperature Dependence of Tunnel Magnetoresistance and Magnetization of IrMn Based MTJ," *Physica Status Solidi*, vol. 201, pp. 1648–1652, 2004. 23

[56] S. R. Min *et al.*, "Etch Characteristics of Magnetic Tunnel Junction Stack with Nanometer-Sized Patterns for Magnetic Random Access Memory," *Thin Solid Films, Proceedings of the International Symposium on Dry Process, 2006. (DPS 2006).*, vol. 516, no. 11, pp. 3507–3511, November 2008. 28

[57] R. Beach *et al.*, "A Statistical Study of Magnetic Tunnel Junctions for High-Density Spin Torque Transfer-MRAM (STT-MRAM)," in *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, December 2008, pp. 1–4. 28

[58] V. Korenivski and R. Leuschner, "Thermally Activated Switching in Nanoscale Magnetic Tunnel Junctions," *IEEE Transactions on Magnetics*, vol. 46, no. 6, pp. 2101–2103, June 2010. 29

[59] J. Sun and D. Ralph, "Magnetoresistance and Spin-Transfer Torque in Magnetic Tunnel Junctions," *Journal of Magnetism and Magnetic Materials*, vol. 320, no. 7, pp. 1227–1237, 2008. 29

[60] H. Chang and J. Burns, "Demagnetizing and Stray Fields of Elliptical Films," *Journal of Applied Physics*, vol. 37, no. 8, pp. 3240–3245, July 1966. 29

[61] C. J. Lin *et al.*, "45nm Low Power CMOS Logic Compatible Embedded STT MRAM Utilizing a Reverse-Connection 1T/1MTJ Cell," in *Electron Devices Meeting (IEDM), 2009 IEEE International*, December 2009, pp. 1–4. 30, 54

[62] T. Takenaga *et al.*, "Control of Pinned Layer Magnetization Direction in Spin-Valve-Type Magnetic Tunnel Junction with an IrMn Layer," *Journal of Applied Physics*, vol. 95, no. 11, pp. 6795–6797, 2004. 30

[63] M. Motoyoshi *et al.*, "A study for 0.18 μm high-density MRAM," in *Symposium on VLSI Technology, 2004. Digest of Technical Papers.*, June 2004, pp. 22–23. 37

[64] A. Raychowdhury, D. Somasekhar, T. Karnik, and V. De, "Design Space and Scalability Exploration of 1T-1STT MTJ Memory Arrays in the Presence of Variability and Disturbances," in *Electron Devices Meeting (IEDM), 2009 IEEE International*, December 2009, pp. 1–4. 37

[65] P. Amiri *et al.*, "Low Write-Energy Magnetic Tunnel Junctions for High-Speed Spin-Transfer-Torque MRAM," *Electron Device Letters, IEEE*, vol. 32, no. 1, pp. 57–59, January 2011. 48, 49

[66] P. K. Amiri *et al.*, "Switching Current Reduction Using Perpendicular Anisotropy in CoFeB-MgO Magnetic Tunnel Junctions," *Applied Physics Letters*, vol. 98, no. 11, p. 112507, 2011. 48, 49

[67] G. E. Rowlands *et al.*, "Deep Subnanosecond Spin Torque Switching in Magnetic Tunnel Junctions with Combined In-Plane and Perpendicular Polarizers," *Applied Physics Letters*, vol. 98, no. 10, p. 102509, 2011. 48, 49

[68] W. Zhao *et al.*, "New Non-Volatile Logic Based on Spin-MTJ," *physica status solidi (a)*, vol. 205, no. 6, pp. 1373–1377, 2008. 48

[69] X. F. Han and A. C. C. Yu, "Patterned Magnetic Tunnel Junctions with Al Conduction Layers: Fabrication and Reduction of Pinhole Effect," *Journal of Applied Physics*, vol. 95, no. 2, pp. 764–766, 2004. 49

[70] S. Isogami, M. Tsunoda, and M. Takahashi, "30-nm Scale Fabrication of Magnetic Tunnel Junctions using EB Assisted CVD Hard Masks," *Magnetics, IEEE Transactions on*, vol. 41, no. 10, pp. 3607–3609, October 2005. 49

[71] R. Nebashi *et al.*, "A 90nm 12ns 32Mb 2T1MTJ MRAM," in *ISSCC 2009*, February 2009, pp. 462–463, 463a. 54

[72] D. Halupka *et al.*, "Negative-Resistance Read and Write Schemes for STT-MRAM in 0.13 μm CMOS," in *ISSCC 2010*, February 2010, pp. 256–257. 54

[73] K. Tsuchida *et al.*, "A 64Mb MRAM with Clamped-Reference and Adequate-Reference Schemes," in *ISSCC 2010*, February 2010, pp. 258–259. 54