Purdue University Purdue e-Pubs

Open Access Dissertations

Theses and Dissertations

January 2016

Spin-Transfer-Torque (STT) Devices for On-chip Memory and Their Applications to Low-standby Power Systems

Yeongkyo Seo Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

Recommended Citation

Seo, Yeongkyo, "Spin-Transfer-Torque (STT) Devices for On-chip Memory and Their Applications to Low-standby Power Systems" (2016). *Open Access Dissertations*. 1272. https://docs.lib.purdue.edu/open_access_dissertations/1272

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Graduate School Form 30 (Revised 08/14)

PURDUE UNIVERSITY GRADUATE SCHOOL Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Yeongkyo Seo

Entitled Spin-Transfer-Torque (STT) Devices for On-chip Memory and Their Applications to Low-Standby Power Systems

For the degree of ______ Doctor of Philosophy

Is approved by the final examining committee:

KAUSHIK ROY

ANAND RAGHUNATHAN

BYUNGHOO JUNG

VIJAY RAGHUNATHAN

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

KAUSHIK ROY

Approved by Major Professor(s):

Approved by: V. Balakrishnan	07/29/2016

Head of the Department Graduate Program

Date

SPIN-TRANSFER-TORQUE (STT) DEVICES FOR ON-CHIP MEMORY AND THEIR APPLICATIONS TO LOW-STANDBY POWER SYSTEMS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Yeongkyo Seo

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2016

Purdue University

West Lafayette, Indiana

Dedicated to

My family and friend

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my advisor, Prof. Kaushik Roy for the support of my Ph.D study and research. His guidance and encouragement helped me in all the time of the research and writing of the thesis. Besides, I would like to thank to my Ph.D committee: Prof. Anand Raghunathan, Prof. Byunghoo Jung, and Prof. Vijay Raghunathan for their excellent advice and feedbacks.

My research could not have been accomplished without the support and cooperation of my lab mate and especially Dr. Xuanyao Fong, Dr. Kon-Woo Kwon, and Dr. Yusung Kim. I would sincerely like to thank them for valuable discussions, advice and collaboration.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
ABSTRACT	X
1. INTRODUCTION	1
1.1. Domain Wall Coupling-based STT-MRAM	2
1.2. Spin-Orbit Torque MRAM with supporting dual read/write ports (1R/1W)	3
1.3. Area-Efficient SOT-MRAM with a Schottky Diode	3
1.4. Shared Bit-line SOT-MRAM Structure for High Density On-chip Caches	3
1.5. Nonvolatile Flip-Flop by using Complementary Polarizer MTJ	4
1.6. Organization of Dissertation	4
2. DOMAIN WALL COUPLING-BASED STT-MRAM FOR ON-CHIP CACHE	
APPLICATIONS	6
2.1. Introduction	6
2.2. Proposed Memory Device Structure	10
2.3. Modeling and Simulatione	13
2.3.1. Micro-magnetic Simulation	14
2.3.2. NEGF based Electron Transport Simulation	16
2.4. Results and Discussion	19
2.5. Conclusion	22
3. HIGH PERFORMANCE AND ENERGY-EFFICIENT ON-CHIP CACHE USIN	١G
DUAL PORT (1R/1W) SOT-MRAM	23
3.1. Introduction	23
3.2. Review of Mutli-Port STT-MRAM	26
3.3. Design of Multi-Port SOT-MRAM	29
3.3.1. Spin-orbit Device Structure	29
3.3.2. 1R/1W SOT-MRAM Design	30
3.4. Device Modeling and Simulation Framework	32
3.4.1. LLGS based Magnetization Dynamics Simulation	33
3.4.2. NEGF based Electron Transport Simulation	35
3.5. Simulation Results and Discussions	36

	Page
3.5.1. Bit-cell Level Simulations and Results	
3.5.2. Integrated Cache Simulation	40
3.5.3. Micro-architectural Simulation and Comparison	41
3.6. Conclusion	46
4. AREA-EFFICIENT SOT-MRAM WITH A SCHOTTKY DIODE	48
4.1. Introduction	48
4.2. Proposed SOT-MRAM Structure	50
4.3. Simulation and Results	51
4.4. Conclusion	55
5. SHARED BIT-LINE SOT-MRAM STRUCTURE FOR HIGH DENSITY ON-C	CHIP
CACHES	57
5.1. Introduction	57
5.2. Shared Bit-line SOT-MRAM Structure	59
5.3. Modeling and Simulation	62
5.4. Conclusion	65
6. NONVOLATILE FLIP-FLOP BY USING COMPLEMENTARY POLARIZER	
MAGNETIC TUNNEL JUNCTION	67
6.1. Introduction	67
6.2. Review of STT-NVFF	69
6.3. Device and Proposed NVFF Structure	71
6.3.1. Device Structure of Complementary Polarizer MTJ	71
6.3.2. Proposed NVFF Structure	72
6.4. Modeling and Simulation	76
6.5. Results and Conclusion	78
6.6. Conclusion	80
6. CONCLUSION	81
LIST OF REFERENCES	84
VITA	93

LIST OF TABLES

Table2.1 Simulation Parameters of DWCSTT and STT	Page 19
2.2 iso-Write Margin VDD and Average Power per Bit	20
2.3 iso-VREAD Comparison of Sensing Margin and Sensing Power	20
2.4 iso-VREAD Comparison of Disturb Margin	
3.1 Simulation parameters of devices	34
3.2 Target specification of the bit-cell level simulation	
3.3 Bit-cell level simulation result and comparison	
3.4 Integrated cache simulation results of 4 different memories	40
3.5 Processor configuration for SimpleScalar simulation	42
3.6 Types of SPEC2000 benchmarks	45
4.1 Simulation parameters of the devices	52
4.2 Simulation results and comparison of MRAM Bit-cells	55
5.1 Biasing conditions for write and read operations in our proposed memory	60
5.2 Simulation parameters of the devices	63
5.3 Results and comparison of three different memory bit-cells	65
6.1 Iso-retention Time Simulation Parameters	
6.2 Transistor size in proposed NVFF	
6.3 Energy and delay comparison of NVFF backup operation	79
6.4 Restore operation comparison of NVFF	

LIST OF FIGURES

Figure Page
1.1 Dynamic and static power consumption trends of mobile SoC
2.1 Probability of write and disturb failures versus width of NMOS access transistor 7
2.2 Read currents distribution of single-ended sensing of standard STT- MRAM. This figure shows that, under process variation, it is apparent to sensing errors
2.3 (Left) Current flow in AP to P switching operation. (Right) Current flow and source degeneration effect in P to AP switching operation
2.4 Device structure of the Domain Wall Coupling STT-MRAM
2.5 Organization of Domain Wall Coupling STT-MRAM bit-cells in a memory array, where only 2 rows and 2 columns are shown
2.6 Applied voltages and current flow through our proposed bit-cell structure during write operation
2.7 Applied voltages and current flow through our proposed bit-cell structure during read operation
2.8 A flowchart of the simulation framework we used to evaluate Domain Wall Coupling based STT-MRAM cells
2.9 Matching micro-magnetic simulation results and the result of experimental data in terms of current density versus domain wall velocity
2.10 Micro-magnetic simulation of write operation of proposed memory. The color indicates the magnetization of the layers
2.11 The bit-cell area comparison of 1T-1MTJ STT-MRAM and DWCSTT-MRAM versus width of access transistor
2.12 (Left) Without fingered NMOS 1T-1MTJ STT-MRAM layout. (Right) With fingered NMOS 1T-1MTJ STT-MRAM layout
2.13 Layout of Domain Wall Coupling based STT-MRAM

Figure Page
3.1 (a) MTJ device structure in parallel and anti-parallel states. (b) Schematic and biasing conditions of standard STT-MRAM
3.2 Simultaneous read and write accesses of cells in two different rows of standard STT-MRAM. Due to the conflicts on the biasing condition of BL and SL, it is impossible to support dual port operation in standard STT-MRAM. WL, BL, and SL represent word-line, bit-line, and source-line, respectively
 3.3 (a) Schematic and biasing conditions of 1R/1W STT-MRAM. (b) Conceptual description of simultaneous write and read accesses in 2-by-2 array of 1R/1W STT-MRAM. WWL, RWL, WBL, RBL, and SL represent write word-line, read word-line, write bit-line, read bit-line and source-line, respectively
3.4 (a) Device structure of SOT-MRAM. (b) Flow of spin-polarized electrons during the write operation
 3.5 Current and spin-polarized electrons flowing during the AP→P switching of spin-orbit device with a spin-sink layer
3.6 (a) Bit-cell structure and the biasing conditions of the 1R/1W SOT-MRAM. (b) Simultaneous write and read accesses in a 2-by-2 array of 1R/1W SOT-MRAM 31
3.7 A flowchart of the simulation framework we used to evaluate STT-MRAMs, and SOT-MRAMs
3.8 Equivalent resistive model of spin-orbit device
3.9 Layout comparison between single-port and dual-port flavors of STT-MRAMs and SOT-MRAMs
3.10 The bit-cell area comparison of standard STT-MRAM, 1R/W STT-MRAM, and 1R/1W SOT-MRAM with changing the width of access transistor
3.11 Normalized IPC and energy measurements using SimpleScalar simulations
4.1 Bit-cell structure of (a) STT-MRAM and (b) SOT-MRAM
4.2 Proposed SOT-MRAM in (a) write operation and (b) read operation
4.3 (a) Matching experimental and SPICE simulated current of TiO _x -based Schottky diode as a function of voltage when cross-sectional area is $4\mu m^2$. (b) Experimental current density trend with varying the cross-sectional area
4.4 Layouts and schematics of STT-MRAM, SOT-MRAM, and proposed SOT-MRAM bit-cells. W _N represents the width of the transistor
5.1 1-by-2 array structure of (a) STT-MRAM and (b) SOT-MRAM

igure Pa	ge
.2 (a) SOT device structure, and (b) Direction of current flows during an anti-parallel switching operation	59
.3 1-by-2 array structure of our proposed SOT-MRAM	59
.4 Layout comparison of STT-MRAM, SOT-MRAM, and our proposed SOT-MRAM	64
.1 Power gating block diagram (a) without NVFF and (b) with NVFF	68
.2 Two step backup operations in the STT-MTJ based nonvolatile slave latch (a) in the 1 st step, (b) in the 2 nd step	69
.3 A restore operation in the STT-NVFF based nonvolatile slave latch when stored data is (a) 1, and (b) 0	a 70
.4 Device structure of the complementary polarizer MTJ	72
.5 Schematic of nonvolatile flip-flop using CPMTJ	73
.6 Timing diagram of CPMTJ based NVFF operation	73
.7 One step backup operations in the CPMTJ based nonvolatile slave latch (a) at Q='1' and (b) at Q='0'	, 74
.8 A restore operation in the STT-NVFF based nonvolatile slave latch when stored data is (a) 1, and (b) 0	a 75
.9 Layout of (top) STT-NVFF and (bottom) CP-NVFF. The dashed rectangular indicat the region of standard flip-flop	es 78

ix

ABSTRACT

Seo, Yeongkyo. Ph.D., Purdue University, August 2016. Spin-Transfer-Torque (STT) devices for on-chip memory and their applications to low-standby power systems. Major Professor: Kaushik Roy.

With the scaling of CMOS technology, the proportion of the leakage power to total power consumption increases. Leakage may account for almost half of total power consumption in high performance processors. In order to reduce the leakage power, there is an increasing interest in using nonvolatile storage devices for memory applications. Among various promising nonvolatile memory elements, spin-transfer torque magnetic RAM (STT-MRAM) is identified as one of the most attractive alternatives to conventional SRAM. However, several design challenges of STT-MRAM such as shared read and write current paths, single-ended sensing, and high dynamic power are major challenges to be overcome to make it suitable for on-chip memories. To mitigate such problems, we propose a domain wall coupling based spin-transfer torque (DWCSTT) device for on-chip caches. Our proposed DWCSTT bit-cell decouples the read and the write current paths by the electrically-insulating magnetic coupling layer so that we can separately optimize read operation without having an impact on write-ability. In addition, the complementary polarizer structure in the read path of the DWCSTT device allows DWCSTT to enable self-referenced differential sensing. DWCSTT bit-cells improve the write power consumption due to the low electrical resistance of the write current path. Furthermore, we also present three different bit-cell level design techniques of Spin-Orbit Torque MRAM (SOT-MRAM) for alleviating some of the inefficiencies of conventional magnetic memories while maintaining the advantages of spin-orbit torque (SOT) based novel switching mechanism such as low write current requirement and decoupled read

and write current path. Our proposed SOT-MRAM with supporting dual read/write ports (1R/1W) can address the issue of high-write latency of STT-MRAM by simultaneous 1R/1W accesses. Second, we propose a new type of SOT-MRAM which uses only one access transistor along with a Schottky diode in order to mitigate the area-overhead caused by two access transistors in conventional SOT-MRAM. Finally, a new design technique of SOT-MRAM is presented to improve the integration density by utilizing a shared bit-line structure.

1. INTRODUCTION

In the last few decades, battery capacity has not improved a lot as compared to the semiconductor technology. Thus, there is a need for the integrated circuits (ICs) designed for mobile applications to dissipate low power consumption for longer battery time. However, as CMOS technology has been scaled down in order to enhance performance and density, the proportion of the leakage power to total power consumption is on the increase. Leakage may account for just over a half of all the power consumption in high performance processors (Fig. 1.1) [1]-[4].



Fig. 1.1 Dynamic and static power consumption trends of mobile SoC [3], [4]

Even though conventional on-chip memories which are implemented using static random access memories (SRAMs) offer advantages such as fast read and write performance, their large leakage power dissipation has become a serious problem [4]. In order to reduce the leakage power, many design techniques such as multi-threshold CMOS, dynamic VTH design, dual-threshold voltage assignment, and forward bodybiased technique have been proposed [5]-[9]. However, these techniques cannot completely eliminate the leakage power in the SRAM because the power-supply still needs to be provided to hold the data in the memory elements. There is an increasing interest in using nonvolatile memory technologies in on-chip caches due to almost zero cell leakage. Among various promising nonvolatile memory elements, spin-transfer torque magnetic RAM (STT-MRAM) is identified as one of the most attractive alternatives to conventional SRAM because of its nonvolatility, unlimited write endurance, high-integration density and compatibility with the CMOS fabrication process [10]-[12].

1.1. Domain Wall Coupling-based STT-MRAM

Although STT-MRAM is a prospective memory technology for future on-chip caches, several design challenges of STT-MRAM such as shared read and write current paths, single-ended sensing, and high dynamic write power are major challenges to be considered for on-chip memory application [13]. In order to provide the solution to the aforementioned design issues of STT-MRAM, we propose a domain wall coupling based STT-MRAM (DWCSTT) [13]. Our proposed DWCSTT bit-cell decouples the read and the write current path by the electrically-insulating magnetic coupling layer so that read and write operations of DWCSTT may be optimized independently. Thus, the reliability of tunneling oxide is improved because large write currents never pass through the tunneling oxide. DWCSTT bit-cells also improves the write power consumption because they are able to meet critical current requirement at lower write voltage because of the low electrical resistance of the write current path. In addition, the complementary polarizer structure in the read path of the DWCSTT device allows DWCSTT to enable self-referenced differential sensing and hence, the proposed bit-cells can achieve higher read margin and lower read power than 1T-1MTJ STT-MRAM.

1.2. Spin-Orbit Torque MRAM with supporting dual read/write ports (1R/1W)

To address the issue of high-write latency and high dynamic power, we propose a Spin-Orbit Torque MRAM with supporting dual read/write ports (1R/1W) to address the issue of high-write latency [14], [15]. Our proposed dual port memory can alleviate the impact of write latency on system performance by supporting simultaneous read and write accesses. The spin-orbit device leverages the high spin current injection efficiency of spin Hall metal to achieve low critical switching current to program a magnetic tunnel junction [16]. The low write current reduces the write power consumption, and the size of the access transistors, leading to higher integration density. Furthermore, the decoupled read and write current paths of the spin-orbit device improves oxide barrier reliability, because the write current does not flow through the oxide barrier. Device, circuit, and system level co-simulations show that a 1R/1W SOT-MRAM based L2 cache can improve the performance and energy-efficiency of the computing systems compared to SRAM and standard STT-MRAM based L2 caches.

1.3. Area-Efficient SOT-MRAM with a Schottky Diode

STT-MRAM requires large amount of write current, hence, considerable MRAM research has been focused on minimizing write current [17]. To address this challenge, spin-orbit torque (SOT) based switching mechanism has been recently proposed. Despite such attribute, one of the biggest disadvantages of SOT-based memory design is that two transistors are required per a bit-cell resulting in area-overhead. Unlike conventional SOT-MRAM requiring two access transistors, our proposed MRAM uses only one access transistor along with a Schottky diode in order to achieve high integration density while maintaining the advantages of SOT-MRAM, such as low write energy and enhanced reliability of magnetic tunnel junction (MTJ) [18]. The Schottky diode is forward biased during read, whereas it is reverse biased during write to prevent sneak current paths.

1.4. Shared Bit-line SOT-MRAM Structure for High Density On-chip Caches

Furthermore, we proposes a new design technique of spin-orbit torque magnetic random access memory (SOT-MRAM) which is suitable for high density and low power

on-chip cache applications. A bit-line of the proposed memory bit-cell is shared with that of an adjacent bit-cell, hence, minimum allowable area of our proposed structure can be improved in comparison with conventional structure of SOT-MRAM by reducing the number of metals along the column direction. Furthermore, since efficient spin orbit torque based switching operation can translate to smaller size of access transistors, the proposed SOT-MRAM achieves higher integration density compared to spin-transfer torque magnetic random access memory (STT-MRAM) while maintaining the advantages of conventional SOT-MRAM such as low write energy dissipation, high readdisturb margin, and improved reliability of magnetic tunnel junction (MTJ).

1.5. Nonvolatile Flip-Flop using Complementary Polarizer MTJ

In addition, we also present new Nonvolatile flip-flop (NVFF) for fine-grain power gating architecture. NVFF using spin-transfer torque magnetic tunnel junctions (STT-MTJs) has been proposed to enable fine-grain power gating systems. However, the STT-MTJ based NVFF (STT-NVFF) may not perform fast backup and disturb-free restore operations [19]. We propose a new NVFF using complementary polarizer MTJ (CPMTJ) to alleviate these limitations [20]. Our proposed NVFF exploits the CPMTJ structure for fast and low-energy backup operation. The estimated backup delay is less than 10ns in 7nm node FinFET technology with CPMTJ size of $12nm \times 33nm$ in a rectangular shape. Furthermore, during the restore operation, CPMTJ provides guaranteed disturb-free sensing since disturb torque in CPMTJ comes from two pinned layer and they cancel each other.

1.6. Organization of Dissertation

The rest of the dissertation is organized as follows. In Chapter 2, we propose domain wall coupling based STT-MRAM (DWCSTT) for improved stability and lower read/write power. Chapter 3 shows SOT-MRAM design with dual (1R/1W) port to alleviate the long write latency of magnetic memories by taking advantages of simultaneous read and write accesses. In Chapter 4, we present a new type of SOT-

MRAM bit-cell to mitigate an area overhead of standard SOT-MRAM bit-cells. In Chapter 5, we propose an area efficient SOT-MRAM design with shared bit-line structure to improve the minimum allowable bit cell area by reducing the number of metals along column direction. Chapter 6 shows a new NVFF design using complementary polarizer MTJ for fast backup operation and disturb free restore operation. Finally, we draw the conclusions in Section 7.

2. DOMAIN WALL COUPLING-BASED STT-MRAM FOR ON-CHIP CACHE APPLICATIONS

This section proposes a domain wall coupling based magnetic device for high-speed and robust on-chip cache applications. The read and write current paths are magneticallycoupled and electrically-isolated which significantly improves the reliability of the read and write operations. Our proposed device makes use of fast and energy-efficient domain wall motion for write operation. A complementary polarizer structure is used to achieve low-power, high-speed and high sensing margin read operations. A device-to-circuits simulation framework was also developed to evaluate our proposed multi-terminal Domain Wall Coupling based spin-transfer torque magnetic random access memory (DWCSTT) cell. Compared to the conventional 1T-1MTJ STT-MRAM bit-cell, the proposed DWCSTT bit-cell achieves > 3.5x improvement in write power under iso-area and iso-write margin condition, and > 3x better sensing margin with low read power consumption, and higher read disturb margin.

2.1. Introduction

Spin-transfer torque magnetic RAM (STT-MRAM) has become a leading candidate for universal memory technology because of its ultra-low-standby power consumption, non-volatility, near unlimited endurance, and high integration density [10], [11]. However, several design challenges (such as shared read and write current paths, singleended sensing scheme, and high write power consumption [12]) need to be overcome for STT-MRAM to be suitable for on-chip cache applications.

^{© [2015]} IEEE. Reprinted, with permission, from Y. Seo, X. Fong, and K. Roy, "Domain Wall Coupling-Based STT-MRAM for On-Chip Cache Applications," *IEEE Transactions on Electron Devices*, vol.62, No. 2, pp. 554 – 560, Feb. 2015.



Fig. 2.1. Probability of write and disturb failures versus width of NMOS access transistor [12].

A significant design issue in 1T-1MTJ STT-MRAM is that electrical current flows through the magnetic tunnel junction (MTJ) during both read and write operations. However, the requirements for the magnitude of current flow during read and write operations are different [12]. The current flow during write operations (IWRITE) needs to be large to ensure successful write operations whereas the current flow during read operations (IREAD) needs to be limited to prevent accidental writing into the bit-cell (called read-disturb failure). This design issue is exacerbated by process variations. Consider for example, the impact of the width of the access transistor on the probability of failure as discussed in [12] and illustrated in Fig. 2.1. When the width of the access transistor is increased, the probability of write failure decreases whereas the probability of readdisturb failure increases. Hence, it is clear that read and write failures have conflicting requirements on the width of the access transistor. There is a similar conflict between write failures and the MTJ reliability. Large IWRITE required to mitigate write failures may degrade the reliability of the tunneling oxide in the MTJ [21]. Note that reducing the critical switching current of the MTJ (Ic) improves both its reliability and write-ability. However, doing so places a tighter constraint on IREAD and may lead to slower read operations.

Another design issue in 1T-1MTJ STT-MRAM is that read operations employ single-ended sensing scheme which may not be sufficiently robust against process variations. In STT-MRAM, data is stored as the magnetic configuration of the MTJ, which may be sensed out as the electrical resistance of the bit-cell containing the MTJ. If the MTJ is in the parallel or P state, the resistance of the bit-cell will be low. Otherwise, the MTJ is in the anti-parallel or AP state and the resistance of the bit-cell is high. In singled-ended sensing, IREAD is compared to a reference current, IREF, to determine the stored data [22]. The bit-cell resistance is high (low) if IREAD is less (more) than IREF. Under process variations, there may be bit-cells that cannot be correctly sensed as shown in Fig. 2.2. The resistance of the bit-cells on the bottom left of Fig. 2.2 is so high that they are sensed as storing the high resistance state when they are storing the low resistance state (IREAD is always smaller than IREF). Similarly, the resistance of the cells in the top right of Fig. 2.2 is so low that they are sensed as storing the low resistance state when they are storing the high resistance state (IREAD is always larger than IREF). Self-referenced single-ended sensing techniques for mitigating sensing failure of 1T-1MTJ STT-MRAM have been proposed but these schemes require long read access time [23], [24].



Fig. 2.2. Read currents distribution of single-ended sensing of standard STT- MRAM. This figure shows that, under process variation, it is apparent to sensing errors [25].

Although the standby power is reduced in STT-MRAM, the dynamic power can be significantly increased due to its high write power consumption. High critical switching current is needed to flow through the bit-cell in order to switch the free layer magnet. High write voltage is also required due to its high write current and high resistance in write path. Moreover, write power consumption may also be severely exacerbated because the access transistor is source degenerated as illustrated in Fig. 2.3 [26]. As shown in Fig. 2.3, the gate-source voltage (V_{GS}) of access transistor is close to V_{DD} during AP to P switching. The bit line and source line voltages are swapped during P to AP switching so that the current direction is reversed. The source of the access transistor is now at the terminal connected to the MTJ and is at V_{MTJ}. Hence, V_{GS} = V_{DD} - V_{MTJ} instead and the access transistor is said to be *source degenerated*. As a result, the write voltage is determined by that of P to AP switching. This may lead to excessive I_{WRITE} flow during AP to P switching [12]. This may result in excessive write energy consumption and degradation of MTJ reliability.



Fig. 2.3. (Left) Current flow in AP to P switching operation. (Right) Current flow and source degeneration effect in P to AP switching operation.

In this section, we propose a domain wall coupling based STT-MRAM (DWCSTT) for overcoming the aforementioned design issues of 1T-1MTJ STT-MRAM. As we will discuss later, our proposed DWCSTT bit-cell decouples the read and the write current path so that read and write operations of DWCSTT may be optimized independently.

Thus, the reliability of tunneling oxide is improved because large write currents never pass through the tunneling oxide. DWCSTT also exploits a complementary polarizer structure in the read port to enable self-referenced differential sensing. In addition, DWCSTT can achieve low write power consumption and avoid source degeneration of the write access transistors due to its low electrical resistance in write path.

The rest of this section is organized as follows. In Section 2.2, we discuss the design details of proposed domain wall coupling based STT-MRAM device and bit-cell. The modeling of the device and the simulation frameworks used for evaluating the bit-cell are then presented in section 2.3. In section 2.4, we analyze the performance of proposed MRAM and compare it with a state-of-the-art 1T-1MTJ STT-MRAM in terms of write power under iso-area and iso-write margin condition and read power, margin and disturb margin under iso-area and iso-V_{READ} condition. Finally, we conclude this section in section 2.5.



2.2. Proposed Memory Device Structure

Fig. 2.4. Device structure of the Domain Wall Coupling STT-MRAM.

Fig. 2.4 shows the structure of the DWCSTT storage device. Our proposed DWCSTT device looks similar to the spintronic logic device in [27] except for complementary polarizer structure in the read port [20]. The array organization of DWCSTT is shown in Fig. 2.5. The DWCSTT storage device consists of a domain wall layer in the write path

(WL, WR) and a complementary polarizer structure in read path (RL, RR, RM), which is spatially and electrically-insulated. Hence, the DWCSTT storage device has five terminals. Fig. 2.4 shows the write path of our proposed device consists of a low impedance ferromagnetic metal with perpendicular magnetic anisotropy (PMA). The two ends of the ferromagnetic metal are magnetically pinned and are magnetized in opposite directions. Hence, a narrow domain wall exists in the ferromagnetic metal somewhere between the two magnetically pinned ends.



Fig. 2.5. Organization of Domain Wall Coupling STT-MRAM bit-cells in a memory array, where only 2 rows and 2 columns are shown.

The position of the domain wall may be manipulated using current-induced domain wall motion [28], which forms the basis for the write operations in DWCSTT. The domain wall may be moved from one side of the ferromagnetic metal to the other by current injection. Consider in Fig. 2.6, when electrons are entering from WR. The spins of the electrons become polarized into the '-z' direction by the pinned layer connected to WR. As these electrons flow through the ferromagnetic metal, they exert a torque on the local moments [28] and, as a result, the domain wall moves from right to left. The

domain wall moves in the opposite direction if the flow of electrons is reversed. Hence, the domain wall moves in the direction of electron flow.

After write operation, the domain wall should remain in its programmed position so as to be stable. Notches can be engineered into the domain wall layer to improve the stability of domain wall at the location of notch, as shown in [29]. Thus, nano-scale notches can be etched at the both ends of domain wall layer (at the boundary with the pinned end) to enhance the stability of DWCSTT cells.



Fig. 2.6. Applied voltages and current flow through our proposed bit-cell structure during write operation.



Fig. 2.7. Applied voltages and current flow through our proposed bit-cell structure during read operation.

The region of the ferromagnetic metal in which the domain wall moves is magnetically coupled to a free layer in the read path. Hence, domain wall motion in the write path also programs the magnetization of the free layer in read path [27]. Note from Fig. 2.4 that the electrically-insulating coupling layer ensures that the write path is electrically decoupled from the read path.

During the read operation, current is passed from LSL (IREAD.) and from RSL (IREAD.R) into RBL by putting both LSL and RSL at VREAD, and putting RBL at GND as shown in Fig. 2.7. Depending upon the magnetization of the free layer in the read path, the currents flowing through LSL and through RSL are different. In Fig. 2.7, IREAD.L is larger than IREAD.R, if the cell stores a '0'. Alternatively, IREAD.R is larger than IREAD.L during read operation, if the cell stores a '1'. Note that the position of the domain wall in the write path is manipulated by applying a voltage between WL and WR. The free layer in the read path is magnetically coupled to the domain wall layer in the write path and hence, the magnetizations of the both layers are manipulated together during write operations. The aforementioned mechanism allows the resistance of read path in the device to be programmed to RLEFT > RRIGHT or RLEFT < RRIGHT. Because of non-volatility, removing power does not affect to the magnetization of the magnetic layers in DWCSTT and these two different resistance states in the read path are maintained.



2.3. Modeling and Simulation

Fig. 2.8. A flowchart of the simulation framework we used to evaluate Domain Wall Coupling based STT-MRAM cells.

The evaluation of DWCSTT requires 1) the modeling of magnetization dynamics using micro-magnetic simulations, as well as 2) electronic transport simulation using the Non-Equilibrium Green's Function (NEGF) formalism to model the interaction between the magnetic response of the device and its electrical characteristics. Fig. 2.8 shows the flowchart of simulation framework which we developed to evaluate DWCSTT and 1T-1MTJ STT-MRAM bit-cells. For example, micro-magnetic simulations allow us to determine the critical switching current needed for successful write operations into the DWCSTT device. The access transistors and control voltages are then chosen to ensure the DWCSTT bit-cell is successfully written into during write operations. The following sections present the modeling framework used in our evaluation and analysis of our proposed DWCSTT bit-cell.





Fig. 2.9. Matching micro-magnetic simulation results and the result of experimental data in terms of current density versus domain wall velocity

The Landau-Lifshitz-Gilbert (LLG) equation, modified to model current-induced domain wall motion [30], was solved to simulate the write operation in DWCSTT.

$$\frac{d\vec{M}}{dt} = -\gamma \vec{M} \times \vec{H} + \alpha \hat{M} \times \frac{d\vec{M}}{dt} - b_J \hat{M} \times \hat{M} \times \frac{\partial \vec{M}}{\partial x}$$
(2.1)

In Eq (2.1), *M* is the magnetization of each unit cell, *H* is the effective magnetic field that is associated with the uniaxial anisotropy field, demagnetizing field, exchange field and thermal fluctuation field, α is the Gilbert damping constant, γ is the gyromagnetic ratio, and *b_J* is the STT term, which is directly proportional to the current density [27].

The critical current density and switching time of the domain wall layer in DWCSTT and the free layer in STT are estimated using micro-magnetic simulations in Object-Oriented Micro- Magnetic Framework (OOMMF) [31]. Our micro-magnetic simulation was first calibrated to experimentally measured device data in [32]. Fig. 2.9 shows that our simulation was successfully calibrated in terms of domain wall velocity versus critical current density. We observe that there is a linear relationship between the velocity of the domain wall and the current density. If width of the write current pulse is shorter than the required time, the domain wall may stop in the middle of the write path [27]. Therefore, the current pulse must be sufficiently wide for successful write operations.



Fig. 2.10. Micro-magnetic simulation of write operation of proposed memory. The color indicates the magnetization of the layers.

Our proposed DWCSTT device consists of a free layer in the read path that is magnetically coupled to a domain wall in the write path. Current pulses move the domain wall in the write path which changes the magnetization direction of the free layer in the read path [27]. Therefore, critical requirements of this device are that read and write path should be magnetically coupled, but electrically isolated for preventing undesirable current leakage that could impact the device functionality or degrade performance [33]. The experiment in [33] shows that naturally oxidized FeCo as the electrically-insulating magnetic coupling layer for the DWCSTT device can have coupling strength that is greater than 0.35ergs/cm². Fig. 2.10 was generated from our micro-magnetic simulation using this minimum coupling strength. It shows that the free layer in the read path of our DWCSTT device is also programmed in 4 ns.

The critical switching current (Ic) that is needed to meet a particular switching time may be determined from the dimensions of the ferromagnetic layer containing the domain wall (Table. 2.1). From our assumed device dimensions, critical current density, Ic is estimated to be 41.5 μ A for 4 ns switching time from OOMMF simulations. Also, we assumed the conductivity of this layer to be 3.5 x 10⁶ S/m, which is the same as that in [32], and the resistance of the write path is calculated to be 1714 Ω .

The micro-magnetic simulation only allows us to analyze the write operations of DWCSTT bit-cells. The sensing of the data stored in the DWCSTT device requires analysis of the conversion of information from the spin-domain into the voltage domain. Just as in 1T-1MTJ STT-MRAM, the change in magnetization in our proposed DWCSTT device is sensed as the resistance of the device through its read ports. The NEGF formalism of electronic transport, which will be presented next, is used to model the DWCSTT device for read operations.

2.3.2. NEGF based Electron Transport Simulation

For calculating the amount of read current in LSL and RSL, resistance of two different kinds of read path should be obtained. Resistance of read path can be calculated by a compact model for the MTJ resistance based on the Non-Equilibrium Green's Function (NEGF) formalism for modeling electron transport [34]. Parameters of our model are used to calibrate it to experimental data reported in [35]. We then simulate read operations of our proposed DWCSTT device in HSPICE [36].

$$R \propto \left(e^{a_0 t_{MgO} + b_0} + \sum_{m=1}^{c} \left(\left(-1\right)^{m-1} V^{2m} e^{a_m t_{MgO} + b_m}\right)\right)^{-d}$$
(2.2)

$$R(\theta) = \left(\frac{1}{R_{P}}\left(\cos(\frac{\theta}{2})\right)^{2} + \frac{1}{R_{AP}}\left(\sin(\frac{\theta}{2})\right)^{2}\right)^{-1}$$
(2.3)

However, R_{MTJ} dependence on voltage applied across the MTJ (V), thickness of tunneling oxide barrier (t_{MgO}), and the magnetization directions of the magnetic layers need to be modeled. In our model, the angle dependence of R_{MTJ} is captured in Eq. (2.3) where $\theta = \cos^{-1}(\hat{m} \cdot \hat{p})$ and \hat{m} and \hat{p} are the free layer and pinned layer magnetization directions, respectively. Hence, we only need to determine R_P=R_{MTJ}(θ =0) and R_{AP}=R_{MTJ}(θ = π) from NEGF simulation. R_P and R_{AP} as functions of V and t_{MgO} are individually fitted to Eq (2.2) and Eq (2.3), where a_m, b_m, c and d are fitting parameters [34].



Fig. 2.11. The bit-cell area comparison of 1T-1MTJ STT-MRAM and DWCSTT-MRAM versus width of access transistor.

The device dimensions and simulation parameters are listed in Table 2.1. The size of our proposed memory device is limited by the spacing between metal contacts. We have also optimized the thickness of the tunneling oxides in the read path. When the thickness of the tunneling oxide is increased, the read margin is increased but read performance can degrade because of small IREAD. Thus, for sufficient read current differences with high

read margin, we choose 1.40nm in thickness of tunneling oxide with V_{READ} of 0.35V. A 45nm bulk CMOS transistor technology was also assumed for our simulations and λ is as defined in [37], commonly used for layout definition. For comparisons, 1T-1MTJ STT-MRAM bit-cells were also simulated. Parameters of MTJs used in 1T-1MTJ STT-MRAM are also calibrated to experimental data in [35]. The results of our analysis assumed that the minimum feature size of the MTJ technology is 20nm.



Fig. 2.12. (Left) Without fingered NMOS 1T-1MTJ STT-MRAM layout. (Right) With fingered NMOS 1T-1MTJ STT-MRAM layout [38].



Fig. 2.13. Layout of Domain Wall Coupling based STT-MRAM.

To perform iso-bit-cell area fair comparisons, we need to first obtain the bit-cell layouts. The layout and bit-cell area of 1T-1MTJ STT-MRAM and DWCSTT-MRAM are described in the Fig. 2.11, Fig. 2.12, and Fig. 2.13 [38], respectively. For the iso-area comparison, the width of each transistor in 1T-1MTJ STT-MARM is 1260nm and 420nm in DWCSTT bit-cell. The area of both bit-cells is $0.221\mu m^2$.

2.4. Results and Discussion

Quantity	Domain Wall Layer in DWCSTT	Free Layer in STT
Gilbert Damping, α	0.01	0.028
Sat. Magnetization, Ms	560 x 10 ³ A/m	700x10 ³ A/m
Dimension of layer (<i>W</i> _D <i>w</i> x <i>L</i> _D <i>w</i> x <i>t</i> _D <i>w</i>)	20nm x 240nm x 2nm ^a	20nm x 20nm x 2nm
Uniaxial Anisotropy, Ku	$400 \times 10^3 J/m^3$	$290 \times 10^3 \text{J/m}^3$
Polarization (PPL, PFL)	0.75 / 0.7	0.8 / 0.3
STT Fitting Parameter, Λ	1	2
Width of ATx	420nm	1260nm
IC (4ns)	41.5µA	13.0μA (Ic('0')) 19.5μA (Ic('1'))

Table 2.1 Simulation Parameters of DWCSTT and STT

^a Dimension of free layer in DWCSTT = 20nm x 240nm x 2nm

Table 2.1 lists the simulation parameters we assumed for DWCSTT and 1T-1MTJ STT-MRAM bit-cells. As shown in Table 2.1, the I_c of the DWCSTT device is higher than that of the conventional MTJ due to the size of the DWCSTT device. However, the proposed DWCSTT bit-cells are able to meet critical current requirement at lower write voltage as shown in Table 2.2 because of the low electrical resistance of the write path. Also, boosted write voltage is needed in 1T-1MTJ STT-MRAM to achieve higher write margins (write margin, WM = (I_{WRITE} - I_c) / I_c x 100%). The increase in write voltage is

Write	DWC-MRAM	STT-MRAM
Margin	(VWRITE / Power)	(VWRITE / Power)
0%	$0.104V$ / $4.404\mu W$	0.858V / 16.46µW
10%	0.115V / 5.369µW	$1.301V / 41.47 \mu W$
20%	$0.126V / 6.425 \mu W$	1.788V / 96.24µW

Table 2.2 iso-Write Margin VDD and Average Power per Bit

4ns switching time of both cases

Using 45nm bulk CMOS technology and for cell area = $0.221 \mu m^2$

tмgo=1.40nm (DWC), tмgo=1.23nm (STT)

VREAD = 0.35V	DWC-MRAM	STT-MRAM
IREF	2.423µA	7.540µA
IREAD.P	5.812µA	10.69µA
IREAD.AP	2.423µA	4.395µA
Margin	139.87%	41.71%
Power	2.882µW	5.280µW

Table 2.3 iso-VREAD Comparison of Sensing Margin and Sensing Power

much smaller to meet higher WM in DWCSTT. We observe that DWCSTT consumes less write energy than 1T-1MTJ STT-MRAM due to the lower write voltage required. Furthermore, the write current does not pass through the tunnel barriers in DWCSTT unlike in the 1T-1MTJ STT-MRAM bit-cell. Hence, the reliability of the tunnel oxide, which is crucial to the readability of the bit-cells, is improved in DWCSTT as compared to the 1T-1MTJ STT-MRAM bit-cell.

The comparison in the Table 2.3 shows that the sensing margin (SM = $(I_{READ} - I_{REF}) / I_{REF} \times 100\%$) of DWCSTT is better than that of 1T-1MTJ STT-MRAM. This is due to the self-referenced differential sensing scheme used for the read operation in DWCSTT, which is different from the single-ended sensing scheme used in 1T-1MTJ STT-MRAM. Furthermore, the thickness of the tunneling oxide in read path in DWCSTT may be optimized to improve read operations without impacting write performance. This cannot be done in 1T-1MTJ STT-MRAM. Hence, a smaller read current flows through the

DWCSTT bit-cell. Together with the use of self-referenced differential sensing scheme, DWCSTT is able to achieve much better read performance and lower read power dissipation than 1T-1MTJ STT-MRAM. A latch based sense amplifier [39] may also be used for sensing DWCSTT bit-cell resistance to achieve much faster sensing delay than the singled-ended sensing scheme used for 1T-1MTJ STT-MRAM. The transient HSPICE simulation with parasitic capacitances was used to evaluate the array level read operation of DWCSTT with the sense amplifier [39]. In the DWCSTT based 16k cache block (256 x 64), the SLs (LSL and RSL) and BL length, and our assumed wire capacitance are 118 μ m and 0.1fF/ μ m, respectively. Our simulation results show that the latch-based sense amplifier can achieve high speed read operation that is higher than 1.5GHz sensing.

Table 2.4 iso-VREAD Comparison of Disturb Margin

Disturb Margin	DWC-MRAM	STT-MRAM
IMargin	$> 10.488 \mu A^{a}$	8.810µA

^a Critical current of read path in DWC-MRAM is obtained by $V_{BL} = 3V$.

In Table 2.4, we compare the read-disturb margin ($RM = I_C - I_{READ}$) of DWCSTT and 1T-1MTJ STT-MRAM. In the DWCSTT device, the read current path is different from the write current path. Furthermore, two different spin-polarized currents flow into the free layer. Our simulations show that the free layer in the read path is not programmed even when V_{LSL} and V_{RSL} are at 3V. The DWCSTT device is almost disturb-free because of three reasons. First, the read current needed to accidentally write into the DWCSTT device is higher than that of 1T-1MTJ STT-MRAM. This is because the larger dimensions of the free layer in the read path of the DWCSTT device as compared to that in the 1T-1MTJ STT-MRAM bit-cell. Second, the torque exerted by the read current flowing through one pinned layer in the DWCSTT device is cancelled by that exerted by the read current flowing through the other pinned layer. Finally, our simulations indicate that using a latch based sense amplifier for DWCSTT read operations enables sub-

nanosecond sensing delays. The current needed to cause disturb failure within subnanosecond delay is > 15 times larger than the current flowing through the DWCSTT device during read operations. Hence, the disturb margin of DWCSTT is significantly higher than that of 1T-1MTJ STT-MRAM.

2.5. Conclusion

We proposed a five-terminal Domain Wall Coupling based STT-MRAM for on-chip cache applications. It utilizes a domain wall motion layer and a complementary polarizer structure to achieve energy efficiency, high performance and high disturb, sensing and write margins. The design requirements of DWCSTT bit-cell is significantly relaxed compared to 1T-1MTJ STT-MRAM because the read and write current paths are decoupled. The use of a low resistance write path allows the proposed DWCSTT bit-cell to mitigate source degeneration of the write access transistor, which also reduces write energy consumption as compared to 1T-1MTJ STT-MRAM. Furthermore, the complementary polarizer structure in the read path of the DWCSTT device allows DWCSTT to achieve higher read margin and lower read power than 1T-1MTJ STT-MRAM. The decoupled read and write paths also ensure that large write currents never flow through the tunnel oxide in the DWCSTT device. Hence, the reliability of the tunnel oxide, which is crucial for the readability of MTJs and our DWCSTT device, is improved as compared to 1T-1MTJ STT-MRAM. Furthermore, the latch-based sense amplifier may be used in DWCSTT to achieve high speed read operation. Thus, as compared with 1T-1MTJ STT-MRAM, our proposed DWCSTT is more suitable for robust high performance low power on-chip cache applications.

3. HIGH PERFORMANCE AND ENERGY-EFFICIENT ON-CHIP CACHE USING DUAL PORT (1R/1W) SOT-MRAM

This section proposes a dual (1R/1W) port spin-orbit torque magnetic random access memory (1R/1W SOT-MRAM) for energy efficient on-chip cache applications. Our proposed dual port memory can alleviate the impact of write latency on system performance by supporting simultaneous read and write accesses. The spin-orbit device leverages the high spin current injection efficiency to achieve low critical switching current to program a magnetic tunnel junction. The low write current reduces the write power consumption, and the size of the access transistors, leading to higher integration density. Furthermore, the decoupled read and write current paths of the spin-orbit device improves oxide barrier reliability, because the write current does not flow through the oxide barrier. Device, circuit, and system level co-simulations show that a 1R/1W SOT-MRAM based L2 cache can improve the performance and energy-efficiency of the computing systems compared to SRAM and standard STT-MRAM based L2 caches.

3.1. Introduction

Spin-transfer torque magnetic random access memory (STT-MRAM) has recently gained significant attention as a potential candidate for on-chip memories due to its desirable features such as non-volatility, high integration density, and compatibility with the CMOS fabrication process [11], [40]. The non-volatile nature of STT-MRAM enables zero leakage power consumption in un-accessed bit-cells. Owing to its small bit-cell footprint, STT-MRAM is also capable of achieving $> 2 \times$ higher integration density in comparison to CMOS static RAM (SRAM) [11], [41]. An increase in the capacity of on-

^{© [2016]} IEEE. Reprinted, with permission, from Y. Seo, K-W. Kwon, X. Fong, and K. Roy, "High Performance and Energy-Efficient On-Chip Cache Using Dual Port (1R/1W) Spin-Orbit Torque MRAM," IEEE *J. Emerging and Selected Topics in Circuits and Systems*, Apr. 2016.
chip memory leads to a reduction in the number of accesses to the off-chip memory [42]. Since off-chip memory accesses require long latency and large energy consumption, the system energy efficiency and performance can be improved in STT-MRAM based caches due to its large capacity. Moreover, STT-MRAM is compatible with the CMOS fabrication process [43]. These advantages make STT-MRAM technology a viable option to replace SRAMs in the on-chip cache hierarchy.



Fig. 3.1. (a) MTJ device structure in parallel and anti-parallel states. (b) Schematic and biasing conditions of standard STT-MRAM

A standard STT-MRAM bit-cell is composed of a magnetic tunnel junction (MTJ) shown in Fig. 3.1(a), and an NMOS access transistor connected as shown in Fig. 3.1(b). The MTJ is the storage element consisting of a pinned layer and a free layer, sandwiching a tunneling oxide barrier. The magnetization of the pinned layer (PL) is magnetically pinned, whereas that of the free layer (FL) can be changed by injecting an electrical current. Typically, the FL magnetization is bi-stable, either parallel (P) or anti-parallel (AP) with respect to the PL magnetization [44]. In order to write the P state, bit-line (BL) is set to V_{DD} , the source-line (SL) to G_{ND} , and the word-line (WL) asserted high such that the current flows from FL to PL. Similarly, the AP state can be written by reversing the direction of current flow (i.e., BL to G_{ND} and SL to V_{DD}). For read operation, a small read voltage is applied to the BL and the SL is grounded so that the read current flows from

BL to SL. Since the resistance of an MTJ is high in the AP state and low in the P state, read operations are performed by sensing the MTJ resistance.

In spite of the aforementioned advantages of STT-MRAM, the long write latency (>10ns) and high write current (>2MA/cm2) requirements need to be addressed to improve STT-MRAM for on-chip cache applications [45]-[47]. Recent work has shown that write latency longer than 10ns may degrade system performance because read requests from the processor are blocked during write operations [48], [49]. One possible solution is to design the STT-MRAM bit-cell with multi-port capability such as 1-read/1write STT-MRAM (1R/1W STT-MRAM) shown in Fig. 3.3(a) [50]. As a result, read and write operations can occur simultaneously and hence, the impact of a slow write operation is effectively mitigated. Note, however, that multi-port capability reduces the achievable memory density because an additional transistor is required. Moreover, the high critical write current density still remains a drawback of STT-MRAMs. The high critical write current density leads to large write dynamic energy consumption and negatively impacts memory density (due to the need for wider access transistor width) and MTJ reliability. Furthermore, as we will discuss later, during write "0", the access transistor is under high stress condition because of large VGS, which can degrade the reliability of the access transistor.

In this work, we propose a 1R/1W MRAM based on spin-orbit torque (1R/1W SOT-MRAM) for high-performance and energy-efficient on-chip cache memory applications. Our proposed memory can take advantage of simultaneous read and write operation without incurring area overhead compared to single-port SOT-MRAM. Because of the high efficiency of spin current generation (>100%) via the spin Hall effect, SOT-MRAM requires a lower write current compared to STT-MRAM. As a consequence, our proposed design can mitigate the degradation of memory density (by reducing the width of access transistors) and MTJ reliability (by separating read and write current paths of the spinorbit device), and reduce write energy dissipation. In addition, the proposed memory can avoid the reliability issue associated with the access transistor of 1R/1W STT-MRAM, as mentioned earlier. The rest of this section is organized as follows. Section 3.2 introduces the fundamentals of multi-port STT-MRAM. Section 3.3 describes the design of our proposed 1R/1W SOT-MRAM. The modeling of the devices and the simulation frameworks are presented in Section 3.4. Section 3.5 compares different types of on-chip memories and describes the impacts of caches on overall system performance and energy consumption using a system-level simulator. Finally, we draw the conclusion of work in Section 3.6.



3.2. Review of Multi-Port STT-MRAM

Fig. 3.2. Simultaneous read and write accesses of cells in two different rows of standard STT-MRAM. Due to the conflicts on the biasing condition of BL and SL, it is impossible to support dual port operation in standard STT-MRAM. WL, BL, and SL represent word-line, bit-line, and source-line, respectively.

Fig. 3.2 illustrates the design of a 2-by-1 array of standard STT-MRAM. The word line, WL connects the bit-cells along the row, while the bit-cells along the column are connected to one set of bit line, BL and source line, SL. As Fig. 3.2 shows, simultaneous read and write accesses, which are the benefits of the dual port operation are not supported in standard STT-MRAM. When data "1" needs to be written to a bit-cell on the

top row, SL is driven to the write voltage level (V_{DD}) and BL is grounded. However, if the other bit-cell in the same column is accessed for the read operation at the same time, SL should be biased at G_{ND} and BL at V_{RD} . Hence, due to the conflicts on the biasing condition of BL and SL, read and write operations to different bit-cells in the same column cannot occur simultaneously. Consequently, when a standard STT-MRAM array receives read requests during a write operation, the read requests need to wait until the write operation is completed, before being serviced.



Fig. 3.3. (a) Schematic and biasing conditions of 1R/1W STT-MRAM. (b) Conceptual description of simultaneous write and read accesses in 2-by-2 array of 1R/1W STT-MRAM. WWL, RWL, WBL, RBL, and SL represent write word-line, read word-line, write bit-line, read bit-line and source-line, respectively.

In order to alleviate the memory access conflicts, STT-MRAM with separate read and write ports (1R/1W STT-MRAM) was proposed in [50]. To support multiple port operation, an extra access transistor is added to the bit-cell as shown in Fig. 3.3(a). Compared to the standard STT-MRAM, one additional transistor M2 associated with an extra word-line (read-word-line (RWL) in Fig. 3.3(a)) and an additional data access connection (read-bit-line (RBL) in Fig. 3.3(a)) are inserted to separate the write port from the read port. Hence, two pairs of BL and access transistors ((M1, WBL), (M2, RBL)) are exclusively used for write and read operations, respectively.

A write operation in 1R/1W STT-MRAM occurs by applying appropriate voltages listed in the table in the Fig. 3.3(a) to WBL, WWL and SL. To write "1", WBL is set to V_{DD} , SL to G_{ND} , and WWL is asserted high to turn on the write access transistor (M1). Then, the write current flows from WBL to SL. To write "0", the direction of the write current is reversed by applying a negative voltage (V_{WN}) to the WBL, and keeping the WWL at V_{DD} . A read operation is performed by turning on the read access transistor M2, and charging RBL to V_{RD} . Then, the read current flows through the MTJ and the transistor M2. The write access transistor M1 does not need to be activated during the read operation. Thus, as shown in Fig. 3.3(b), it is possible to perform read and write operations simultaneously in 1R/1W STT-MRAM through the separated read and write ports.

However, under the biasing condition for write "0", the gate-source voltage (V_{GS}) of the write access transistor can be higher than V_{DD} (V_{GS} = V_{GATE} - V_{SOURCE} = V_{DD} - V_{WN} > V_{DD}). Hence, transistor M1 is under a high stress condition because of large (excessive) gate-source voltage (V_{GS}). This can result in reliability degradation of the transistor due to bias temperature instability (BTI) [51]. The benefits of multi-port operation can be achieved by paying significant area overhead compared to standard STT-MRAM due to an additional transistor requirement. Furthermore, high write current requirement of MTJ not only leads to large write dynamic energy dissipation but also degrades the MTJ reliability and memory density. High write current flowing through the oxide barrier in the MTJ may give rise to high voltage across the MTJ and may result in oxide breakdown [52]. In order to provide high write current of MTJ, the width of the access transistor should be enlarged, however, this may reduce the memory density.

3.3. Design of Multi-Port SOT-MRAM

In order to address the aforementioned issues of 1R/1W STT-MRAM, we propose SOT-MRAM with supporting multiple port operation (a preliminary version appeared in [14]). In this section, we first describe the details of the spin-orbit device. Then, we present the design of our proposed 1R/1W SOT-MRAM and discuss its benefits.

3.3.1. Spin-orbit Device Structure



Fig. 3.4. (a) Device structure of SOT-MRAM. (b) Flow of spin-polarized electrons during the write operation.



Fig. 3.5. Current and spin-polarized electrons flowing during the AP \rightarrow P switching of spin-orbit device with a spin-sink layer.

The device structure of SOT-MRAM is composed of an MTJ and a spin Hall metal (SHM), which is a nonmagnetic conductor with large spin-orbit interaction, as shown in Fig. 3.4(a) [16], [53]. An in-plane magnetic anisotropy (IMA) free layer is on the top of the spin Hall metal (which can be tungsten experimentally demonstrated in [16]) with a large spin Hall angle. During the write operation, a current is passed through the SHM to flip the magnetization of the free layer of the MTJ. When charge current (I_C) flows from node W1 to W2, spin-polarized electrons oriented at -y and +y directions are injected to

+z and –z surfaces of the SHM, respectively, due to the spin Hall effect (SHE) [53], [54] and/or a field-like torque on the free layer caused by Rashba effect [55], [56]. As Fig. 3.5 shows, the flow of –y directed spin-polarized electrons (Is) exerts spin-transfer torque on the free layer of MTJ, and anti-parallelizes a free layer with the pinned layer. Reversing the flow of charge current parallelizes the magnetization of FL with that of PL instead.

Note that the spin current injection efficiency (defined as $(I_S \div I_C) \times 100\%$) can be higher than 100%, which improves the energy efficiency of the write operation in SOT-MRAM compared to STT-MRAM. This is because an electron flowing through the SHM can repeatedly scatter at the surface of SHM. As a result, multiple units of angular momentum can be transferred as illustrated in Fig. 3.4(b) [57], [58]. However, the SHM cannot be too thin. Otherwise, the spin current injection efficiency may degrade due to spin diffusion, which is caused by the spin accumulation at the –z surface of the SHM. Note, the spin current injection efficiency of the device can be enhanced by adding an anti-ferromagnetic spin-sink layer (SSL) to reduce the effect of spin accumulation on the free surface of the SHM (Fig. 3.5) [59].

A read operation of the spin-orbit device is performed by passing a small read current from node R1 to node W2 to sense the MTJ resistance. Note that the read and write current paths of the spin-orbit device are separate and hence, we can optimize the read and write operations independently. Because the MTJ is not in the write current path, the thickness of the oxide barrier of the spin-orbit device can be chosen to improve the tunneling magnetoresistance (TMR), and hence, the read operation, without affecting its write-ability. Consequently, our proposed 1R/1W SOT-MRAM can achieve low read power consumption as we shall see later.

3.3.2. 1R/1W SOT-MRAM Design

The bit-cell structure and the biasing conditions of 1R/1W SOT-MRAM, which consists of a spin-orbit device and two access transistors, are shown in Fig. 3.6(a). Two pairs of word-lines and the bit-lines ((RWL, RBL), (WWL, WBL)) are required: one for read operation and the other for write operation. Thus, read and write operations to different bit-cells can occur simultaneously as Fig. 3.6(b) shows.

In order to write "1", a positive voltage ($V_{WP} > G_{ND}$) is applied to WBL whereas SL is connected to G_{ND} . The write access transistor (M1) is then turned on so that write current flows from WBL to SL. The direction of the write current flow is reversed to write "0" by applying a negative voltage to WBL ($V_{WN} < G_{ND}$) and keeping SL at G_{ND} . During read operation, RBL, SL and WL are biased at V_{RD} , G_{ND} , and V_{DD} , respectively, to inject read current from RBL to SL. Due to separate BL and WL for read and write operations in 1R/1W SOT-MRAM, read and write operations to two different rows in the same column can be performed simultaneously as shown in Fig. 3.6(b).



Fig. 3.6. (a) Bit-cell structure and the biasing conditions of the 1R/1W SOT-MRAM. (b) Simultaneous write and read accesses in a 2-by-2 array of 1R/1W SOT-MRAM.

Our proposed SOT-MRAM has several virtues. The memory bit-cell can support dual 1R/1W port which can mitigate the performance penalty caused by long write operations. A lowered critical switching current of spin-orbit device can alleviate the issue of density degradation as well as high write power consumption in 1R/1W STT-MRAM. Unlike 1R/1W STT-MRAM, our proposed memory bit-cell can provide sufficient write current with smaller access transistor due to smaller write current requirement. Thus, even though two access transistors are required, 1R/1W SOT-MRAM can achieve relatively small bit-cell area. Moreover, as mentioned earlier, because the write current does not flow through the MTJ, the reliability of the oxide barrier (MgO layer in Fig. 3.4 (a)) in the spin-orbit device improves.

However, the application of negative WBL can increase the overdrive voltage (V_{GS}) of the write access transistor to be higher than V_{DD} , which may degrade its reliability. Note, however, the WWL voltage may be reduced to VWL (< V_{DD}) to mitigate this issue. Because of its low critical current (and low resistance of SHM), our proposed memory can supply sufficient write current at the reduced WWL voltage condition. This technique cannot be used in 1R/1W STT-MRAM. If we reduce the WWL voltage in 1R/1W STT-MRAM, WBL voltage has to be increased to sustain the high critical current of MTJ, and hence, V_{GS} of 1R/1W STT-MRAM will exceed V_{DD} .

3.4. Device Modeling and Simulation Framework



Fig. 3.7. A flowchart of the simulation framework we used to evaluate STT-MRAMs, and SOT-MRAMs.

In order to evaluate of standard STT-MRAM, 1R/1W STT- MRAM, single-port SOT-MRAM and 1R/1W SOT-MRAM, the device modeling and simulations used herein are composed of two components: 1) modeling the switching dynamics of the MTJ free layer require solving the Landau-Lifshitz-Gilbert-Slonczewski equation; and 2) electronic transport simulation using the Non-Equilibrium Green's Function formalism, which captures the interaction between the magnetic state of the device and its electrical characteristic (Fig. 3.7). The Landau-Lifshitz-Gilbert-Slonczewski equation solver, with the spin current injection efficiency of the spin-orbit device, allows us to determine the device's critical switching current that is needed for successful write operations. The Non-Equilibrium Green's Function formalism is used to model the electron transport and determines the resistance of the MTJ in parallel (P) and anti-parallel (AP) states. Then, we can obtain the power consumption and operation margins of spin-based memory bitcells. The following sub-sections present more details of the simulation framework.

3.4.1. LLGS based Magnetization Dynamics Simulation

The critical current needed for successful write operations within a given time requirement may be determined by modeling the FL magnetization dynamics using the Landau-Lifshitz-Gilbert-Slonczewski (LLGS) equation [12]. Since the size of FL is sufficiently small, it may be approximated as a mono-domain magnet [60], [61]. The LLGS equation is given as follows:

$$\frac{d\hat{m}}{dt} = -|\gamma|(\hat{m} \times \vec{H}) + \alpha(\hat{m} \times \frac{d\hat{m}}{dt}) + STT$$
(3.1)

$$STT = \frac{\gamma \cdot \hat{h} \cdot g(\hat{m} \cdot \hat{p}) \cdot J_{MTJ}}{2 \cdot q \cdot M_{S} \cdot t_{FL}} (\hat{m} \times \hat{m} \times \hat{p})$$
(3.2)

 \vec{H} is the effective magnetic field, γ is gyromagnetic ratio, and α is the Gilbert damping constant. STT in Eq. (3.1) and Eq. (3.2) models the spin-transfer torque exerted by the flow of electrons. J_{MTJ} is the current density flowing through the FL of MTJ, is the reduced Planck's constant, q is the charge of an electron, is the unit vector describing the magnetization direction of the PL, M_S is the saturation magnetization of FL, and t_{FL} is the

thickness of FL. By solving the LLGS equation solver, we obtain the amount of spin current (I_S) needed for switching the MTJ free layer magnet.

$$I_{S} = \frac{A_{MTJ}}{A_{SHM}} \cdot \theta_{SH} \cdot I_{C}$$
(3.3)

The critical switching current (I_C) of spin-orbit device flowing through the SHM is determined from the estimated spin current (I_S) by LLGS equation solver and spin current injection efficiency. The spin current injection efficiency describing the relationship between I_C and I_S is shown in Eq. (3.3), where θ_{SH} is the spin Hall angle of SHM, t_{SHM} is the thickness of SHM, A_{MTJ} (= $\pi/4 \times W_{FL} \times L_{FL}$) and A_{SHM} (= W_{MTJ} × t_{MTJ}) is the crosssectional area of MTJ and SHM, respectively [58], [62]. Even though θ_{SH} is smaller than 1 in our assumed device material, we can achieve >100% spin current injection efficiency (= (I_S÷I_C) × 100%), by designing A_{SHM} to be much smaller than A_{MTJ}.

The simulation parameters and device dimensions are listed in Table 3.1. Tungsten (W) experimentally validated in [16] is assumed to be used for SHM in our simulation.

Design Parameter	Spin-orbit Device	Standard MTJ	
Gilbert Damping, α	0.0122	0.007	
Sat. Magnetization, Ms	$1000 x 10^3 A/m$	$1000 x 10^3 A/m$	
Dimension of free layer	105 mm x 40 mm x 2 mm	105nm x 40nm x 2nm ²	
(WFL X LFL X tFL)	105mm x 40mm x 2nm ⁻		
Dimension of SHM	105mm x 20mm x 2mm		
(WSHM X LSHM X tSHM)		—	
Spin Hall Angle, θ_{SH}	0.3 (W)	-	
SHM Resistivity	200μΩ·cm	-	
MgO thickness, <i>tMgO</i>	1.20nm	1.08nm	
Spin Current	4710/	_	
Injection Efficiency	4/170	—	
Critical Current (10ns)	53µA	179µA	

Table 3.1 Simulation parameters of devices

^{1,2} MTJ free layer is the elliptical shape.

The critical current of the MTJ with our assumed simulation parameters is estimated to be 179 μ A for 10ns switching time from LLGS equation solver. On the other hand, the critical current of the spin-orbit device is calculated to be 53 μ A, which is much lower than that of standard MTJ. In addition, we can obtain the magnetic energy barrier (56k_BT) of the in-plane MTJ by using our assumed device parameters and equations shown in [63].

3.4.2. NEGF based Electronic Transport Simulation

For calculating the amount of current flowing during the read operation of spinbased memories, the resistance of the oxide barrier in the MTJ should be obtained. The resistance of the MTJ which is dependent on the angle between FL and PL, MTJ voltage, and the device dimension has to be modeled to capture the behavior of the MTJ in the circuit [64]. We simulated electronic transport in the MTJ using the Non-Equilibrium Green's Function (NEGF) based simulation framework proposed in [31] to capture the relationship between the MTJ resistance and its magnetic configuration. MTJ characteristics obtained using our spin-dependent transport solver were calibrated to experimentally measured MTJ characteristics in [34] as shown in [52].



Fig. 3.8. Equivalent resistive model of spin-orbit device.

In addition, the write current path in the spin-orbit device is composed of SHM, hence, we need to obtain the resistance of the SHM to determine the power consumption and operation margin of SOT-MRAM. The resistance of the SHM is calculated to be 762 Ω using the dimensions and resistivity of SHM from Table 3.1. By using the SHM resistance and the calibrated MTJ model, we obtain the equivalent circuit of the spin-orbit

device as shown in Fig. 3.8. Then, the equivalent resistive model is utilized together with commercial 45nm CMOS transistor for bit-cell level simulations.

3.5. Simulation Results and Discussions

		Target Spec
Write Margin	= (Iwrite – Ic) / Ic × 100%	20%
Read-disturb Margin	= (IC – IREAD) / IC × 100%	50%
Sensing Margin	= (IREAD – IREF) / IREF × 100%	30%
Swi	10ns	

Table 3.2 Target specification of the bit-cell level simulation

In order to compare our proposed 1R/1W SOT-MRAM based cache with other memory technologies, a multi-level approach is used. At first, we analyze the behavior of a single memory bit-cell. Then, the results are used to determine the integrated cache results. Finally, we evaluate the system performance and energy consumption of five different memory based caches using the aforementioned simulations.

3.5.1. Bit-cell Level Simulations and Results

In order to perform fair comparison, the bit-cells of four different magnetic memories (standard STT-MRAM, 1R/1W STT-MRAM, single-port SOT-MRAM, and 1R/1W SOT- MRAM) were designed to fulfill the same requirements shown in Table 3.2. The amount of reference current is determined by $I_{REF} = 0.5 \times (I_{READ_P} + I_{READ_AP})$. We compared the magnetic memories using proper bit-cell layouts and HSPICE simulations in a 45nm CMOS technology (nominal V_{DD}: 1V) [36].

The bit-cell layout of single-port and dual-port flavors of STT-MRAMs and SOT-MRAMs, shown in Fig. 3.9, were estimated using λ -based layout rules (λ : half of the minimum feature size, F) [38], [58]. In a standard STT-MRAM, when the transistor width is small, a 1-finger layout offers a smaller bit-cell area compared to a 2-finger

layout due to its small metal pitch limited area. On the other hand, when the width of the access transistor is larger than 14λ , a 2-finger layout is used to achieve optimal bit-cell area. Since SL is shared between neighboring cells in the same column, the redundant active region spacing is reduced.



Fig. 3.9. Layout comparison between single-port and dual-port flavors of STT-MRAMs and SOT-MRAMs.



Fig. 3.10. The bit-cell area comparison of standard STT-MRAM, 1R/W STT-MRAM, and 1R/1W SOT-MRAM with changing the width of access transistor.

	Standard STT-	1R/1W STT-	Single-port	1R/1W SOT-
	MRAM	MRAM	SOT-MRAM	MRAM
ATx width	500nm	500nm (Write) 320nm (Read)	180nm (Write) 180nm (Read)	180nm (Write) 180nm (Read)
Area	0.099µm ²	$0.179 \mu m^2$	$0.110 \mu m^2$	$0.110 \mu m^2$
Write Voltage	1V	$1V(V_{WP}) / -0.7V(V_{WN})^{-1}$	0.2V	0.3V(Vwp) / -0.2V(Vwn) ²
Read Voltage	0.2V	0.2V	0.2V	0.2V
Avg. write power	$266.44 \mu W$	$206.87 \mu W$	15.11µW	17.80µW
Avg. read power	30.71µW	$24.87 \mu W$	10.86µW	10.86µW
Aspect ratio of area	1.032	0.571	1.917	1.917

Table 3.3 Bit-cell level simulation result and comparison

¹ WWL voltage = 1.0V

² WWL voltage = 0.8V

The minimum bit-cell area of 1R/1W STT-MRAM, which is determined by the minimum metal pitch, is the same as that of a 2-finger layout of standard STT-MRAM as shown in Fig. 3.9. In 1R/1W STT-MRAM, an additional BL is required to separate the BLs for read and write operations. To eliminate the area overhead caused by the extra BL, a shared SL technique is utilized [65]. During both read and write operations, the SL is biased to ground, thus, the SL metal layer need not be routed in the row-direction (SL can be directly connected to the ground rail) [26]. By routing the SL in the column direction, the SLs of each cell in the same row are shared so that the number of metals per column is identical with that found in standard STT-MRAM. However, when the transistor width is larger than 12λ , 1R/1W STT-MRAM incurs significant area penalty compared to standard STT-MRAM. In contrast with a two-fingered transistor in standard STT-MRAM, two separate transistors are exclusively used for read and write operations in 1R/1W STT-MRAM such that each transistor separately satisfies the requirement of the width. As a result, 1R/1W STT-MRAM cell area is more sensitive to the access transistor size than the area of standard STT-MRAM cell.

In contrast, our proposed 1R/1W SOT-MRAM can take advantage of simultaneous read and write operations without having any area overhead compared with single-port SOT-MRAM. An additional transistor is not required to separate the read and write ports, but one extra BL is needed to isolate the read and write ports. The shared SL technique, which is also utilized in 1R/1W STT-MRAM, is used to reduce the area overhead caused by the additional BL [65].

The transistor width and bit-cell area used in our simulation is shown in Table 3.3 and Fig. 3.10. The width of the access transistor, which is a crucial parameter in determining the bit-cell area, is properly chosen to sustain a write current higher than the critical switching current. Due to the high write current of the MTJ, the width of an access transistor is large in standard STT-MRAM and 1R/1W STT-MRAM. Even though a large transistor is used, a small bit-cell area can still be accomplished in the standard STT-MRAM since only one access transistor is required. However, due to an additional transistor requirement, the significant area overhead is incurred in 1R/1W STT-MRAM. On the other hand, although two separate transistors are required, SOT-MRAMs can achieve small bit-cell area due to smaller widths of transistors.

Owing to small write current requirement, single-port SOT-MRAM and our proposed MRAM can achieve > 10 times improvement in write power consumption compared with STT-MRAMs. In addition, the reliability problem of the write access transistor in 1R/1W STT-MRAM caused by large V_{GS} does not occur in 1R/1W SOT-MRAM. The application of negative voltage can lead to high V_{GS} during the write "0" operation. As shown in in Table 3.3, during the write "0", the V_{GS} of 1R/1W STT-MRAM (= V_{GATE} - V_{SOURCE} = 1 - (-0.7) = 1.7) is much higher than the nominal V_{DD} (= 1V). By contrast, using reduced word-line voltage, the V_{GS} of our proposed memory (= $V_{GATE} - V_{SOURCE} = 0.8 - (-0.2) = 1.0$) does not exceed the nominal V_{DD}.

The comparison of read operations was based on DC read current flowing at iso- V_{RD} ($V_{RD} = 0.2V$) condition. In standard STT-MRAM, an access transistor is used for both read and write operations. On the other hand, read and write access transistors are separate in the 1R/1W STT-MRAM, hence, we can shrink the width of the 1R/1W STT-MRAM read access transistor while still meeting the aforementioned design

	SRAM	STT- MRAM	1R/1W STT- MRAM	SOT- MRAM	1R/1W SOT- MRAM
Capacity Cache area	512kB 2.372mm ²	2MB 2.483mm ²	1MB 2.191mm ²	2MB 2.744mm ²	2MB 2.744mm ²
Read delay Write delay	0.9ns	2.1ns 11.0ns	2.0ns 10.9ns	1.9ns 10.8ns	1.9ns 10.8ns
Read dynamic energy	0 10n I	0.79nJ	0.32nJ	0.57nJ	0.43nJ
Write dynamic energy	0.1911	1.79nJ	1.35nJ	0.19nJ	0.20nJ
Leakage power	1103mW	49mW	49mW	50mW	50mW

Table 3.4 Integrated cache simulation results of 4 different memories

requirements. This contributes to improving the read power consumption. As discussed earlier, because of the separate read and write current paths of the spin-orbit device, we can determine the thickness of the oxide barrier just for improving the read operation without impacting the writ-ability. Therefore, the SOT-MRAM bit-cells can further improve the read power consumption by using a thicker oxide barrier.

3.5.2. Integrated Cache Simulation

As we have noted, long write latency and high write power, compared to conventional on-chip memories such as SRAM, precludes spin-based memories from being implemented in L1 cache which requires fast and frequent write operations. However, in low-level caches, high density and low leakage power consumption make spin-based memories attractive [66]. Thus, spin-based memories are assumed to be used in the L2 cache in our simulation.

In order to estimate energy consumption, access time, and area of spin-based L2 caches, a modified version of the CACTI 6.5 simulator is used [67]. In our system level simulation, we compared the SRAM based L2 cache and spin-based L2 cache while maintaining similar cache area. Note, the capacity of the each cache is determined based

on iso-area constraint. Table 3.3 shows that the bit-cell area of standard STT-MRAM, single-port SOT-MRAM and 1R/1W SOT-MRAM is approximately 2 times smaller than that of 1R/1W STT- MRAM, so the standard STT-MRAM and 1R/1W SOT- MRAM can provide a higher cache capacity. A conventional thin-cell layout of the 6T-SRAM bit-cell (cell area: $44\lambda 2$, aspect ratio: 2.75) was designed to perform detailed comparisons with spin-based caches [66]. Table 3.4 shows the cache capacity and area (similar area) for SRAM, standard STT-MRAM, 1R/1W STT-MRAM, single-port SOT-MRAM, and 1R/1W SOT- MRAM based L2 caches.

Owing to the long write latency of magnetic storage elements, standard STT-MRAM, 1R/1W STT-MRAM, single-port SOT-MRAM, and 1R/1W SOT-MRAM have much longer write delays compared to the SRAM. The write dynamic energy consumption is also higher in STT-MRAMs, whereas the high write energy issue is mitigated in SOT-MRAMs due to the high spin current injection efficiency of the spin-orbit device. But, the write energy improvement of our proposed memory is lower in the integrated cache simulation compared to that in the bit-cell level simulation because of energy consumption for charging and discharging BLs and WLs.

On the other hand, spin-based caches consume much less leakage power than SRAM-based caches. One of the distinct benefits of spin-based memories, compared to SRAM, is their non-volatility. When the magnetic energy barrier is higher than $50k_BT$, MTJ has more than 10 years retention time so that it can be regarded as non-volatile storage elements [68]. The stored information is kept without any external power, hence, power does not need to be supplied in spin-based memories when it is in the standby mode. The leakage power for peripheral circuits is the major portion of total leakage power dissipation in spin-based caches. By contrast, even if unused, SRAM needs to have power supplied to maintain stored data so that, it has high cell leakage power.

3.5.3. Micro-architectural Simulation and Comparison

The overall system performance and energy consumption of 1R/1W SOT-MRAM based caches are compared to that of SRAM, standard STT-MRAM, 1R/1W STT-MRAM, and single-port SOT-MRAM based caches using a modified version of the SimpleScalar architectural simulator with various types of SPEC2000 benchmarks [69].

Table 3.5 lists the processor configuration parameters used in our simulation, where spinbased memories are used to realize the L2 cache.

The system performance is affected by the read and write speeds of the L2 cache, and the number of L2 cache misses. The write latency of STT-MRAM and SOT-MRAM

	Characteristics	
Processer Core	out-of-order processor, issue width-4	
Functional Units	Integer – 4 ALUs, 2 multipliers Floating Points – 2 ALUs, 1 multipliers	
L1 Data Cache	32kB, 2-way set associative, 64byte cache line size	
L1 Instruction Cache	32kB, 2-way set associative, 64byte cache block size	
L2 Unified Cache	4-way set associative, 64byte cache block size	

Table 3.5 Processor configuration for SimpleScalar simulation

is large and leads to significant performance degradation in write-intensive applications. This is because, during the write operation, the read requests are stalled until the long write operation is finished. On the other hand, in 1R/1W ports memories, the read and write operations are performed at the same time through the separate read and write ports so that read requests do not need to be stalled while waiting for the write operation to finish.

The L2 cache miss also has a decisive effect on the overall system performance. When a L2 cache miss occurs, the required data needs to be fetched from off-chip memory. Access to off-chip memory requires a long access time so that cache misses may significantly degrade the overall system performance. The misses in the cache tend to decrease with increasing the size of the cache. Thus, high capacity L2 caches such as standard STT-MRAM, single-port SOT-MRAM, and 1R/1W SOT-MRAM can reduce the performance degradation caused by the long latency of off-chip memory access.

In order to analyze the energy efficiency of L2 cache, we measured the total energy consumption, including leakage energy, read energy, write energy, and off-chip access energy caused by L2 cache misses. The total energy dissipation of L2 cache is determined using the following model [70]:

Because of high spin injection efficiency, low write energy consumption can be achieved in SOT-MRAMs, whereas STT-MRAMs dissipate high dynamic write energy. However, because of the non-volatility of magnetic storage devices, spin based caches consume much lower leakage power than a SRAM based cache. As a result, even though its dynamic energy consumption is low, the SRAM-based cache may show the largest energy consumption compared to spin-based caches because the leakage energy is predominant in scaled CMOS technology [48].

When performing a total energy comparison, the miss energy should be considered to take into account the energy penalty caused by L2 cache misses. Since the cache miss rate may decrease in a higher capacity cache, the large size of caches (standard STT-MRAM, single-port SOT-MRAM and 1R/1W SOT-MRAM) can reduce the energy penalty caused by cache misses. The miss energy is associated with energy dissipation for accessing the off-chip memory. The off-chip energy is highly dependent on the actual system configuration, hence, it is quite difficult to determine the miss energy [70]. In our simulation, we have chosen the off-chip energy of a particular system configuration–"DRAM interface energy + DRAM access energy" from [71], [72].

The performance and energy measurements of the system level simulation show different trends depending on the key features of the benchmarks. As shown in Table 3.6, we categorize SPEC2000 benchmarks into the following 4 types based upon L2 cache miss rate sensitivity to the cache size, and L2 cache write intensity. We can observe write intensity and miss rate sensitivity of each benchmark by using the number of write

operations per 1K instructions (WPKI) and the number of miss change per 1K instructions (MCKI), respectively. In our simulation, we distinguish "high" or "low" of write intensity and miss rate sensitivity when WPKI and MCKI is higher or lower than 6 and 0.25, respectively.



Relative IPC Measurement

Fig. 3.11. Normalized IPC and energy measurements using SimpleScalar simulations.

	Benchmarks	L2 Cache Miss Rate Sensitivity to the Size	L2 Cache Write Intensity
Туре І	AMMP, APSI, MGRID, GCC	High	High
Type II	BZIP, GARBEL, PERLBMK	High	Low
Type III	APPLU, SWIM	Low	High
Type IV	MCF, EQUAKE, LUCAS	Low	Low

Table 3.6 Types of SPEC2000 benchmarks

The results of system level simulations are presented in Fig. 3.11. First, we observe that high Instruction per Cycle (IPC) is accomplished in the high capacity caches such as standard STT-MRAM, single-port SOT-MRAM, and 1R/1W SOT- MRAM in Type 1 benchmarks. This is because the L2 cache miss rate is significantly reduced in high capacity caches due to their high L2 cache miss rate sensitivity to the cache size. Additionally, 1R/1W SOT-MRAM can further enhance the IPC compared to the standard STT-MRAM by taking advantage of 1R/1W port operation because of their high write intensity. Second, Type II benchmarks have a high L2 cache miss rate sensitivity, while their L2 cache write intensity is low so that the benefits of simultaneous read and write accesses in multi-port memories may be small. Thus, in these benchmarks, larger size of cache leads to better performance. Third, in Type III benchmarks, the cache miss rate sensitivity is small, resulting in negligible performance improvements stemming from the use of larger cache sizes, whereas dual-port memories can take advantage of simultaneous read and write accesses to mitigate the negative performance implications of high write intensities. Consequently, 1R/1W STT-MRAM and 1R/1W SOT-MRAM can achieve higher IPC compared to single-port memories. Finally, both L2 cache miss rate sensitivity and L2 cache write intensity of Type IV benchmarks are small, thus, the IPC measurements of five different kinds of L2 cache are almost identical.

The relative energy measurement of the system level simulation is also shown in Fig. 3.11. As mentioned earlier, the static energy is the dominant portion of the overall energy dissipation so that, in all of the benchmarks, the total energy consumption of SRAM is much larger than that of spin-based memories due to its significantly higher cell leakage

power. Among spin-based caches, high capacity caches considerably reduce the energy penalty caused by off-chip memory accesses in applications that have high L2 cache miss rate sensitivity. Therefore, standard STT-MRAM, single-port SOT-MRAM, and 1R/1W SOT-MRAM may achieve lower energy consumption in Type I and Type II benchmarks (especially, AMMP, BZIP, and GALGEL). On the other hand, L2 cache miss sensitivity is small in Type III and Type IV benchmarks so that dynamic energy is a major portion of total energy dissipation. For this reason, single-port SOT-MRAM and 1R/1W SOT-MRAM can achieve highest energy efficiency in Type III and Type IV benchmarks compared to STT-MRAMs because of its low write dynamic energy consumption.

With respect to average performance and energy consumption of different types of benchmarks, our proposed 1R/1W SOT-MRAM based L2 cache achieves the highest IPC and lowest energy consumption compared to SRAM, standard STT-MRAM, and 1R/1W STT-MRAM based L2 caches under similar cache area condition. Because of its small bit-cell area, our proposed memory achieves a high cache capacity, leading to a reduction in the performance and energy penalties caused by L2 cache misses. The separate read and write port of 1R/1W SOT-MRAM enables simultaneous read and write accesses, hence, the performance degradation caused by the long write latency of spin-based memories is alleviated in our proposed memory. Indeed, the high spin current injection efficiency of the spin-orbit device enhances energy-efficiency of the write operation.

3.6. Conclusion

We proposed an SOT-MRAM with supporting dual 1R/1W ports for simultaneous read and write accesses. This may mitigate the problem of higher write latency associated with STT-MRAMs, leading to higher performance. Our proposed 1R/1W SOT-MRAM has low write current requirements due to its high spin current injection efficiency. This leads to relatively small bit-cell area and low dynamic write power consumption. Furthermore, the separate read and write current paths of the spin-orbit device can improve the MTJ reliability because the write current does not flow through the tunneling barrier. Additionally, although the large V_{GS} of 1R/1W STT-MRAM may impact the reliability of the write access transistor, a reduced word-line voltage technique is effective in alleviating the transistor reliability problem in 1R/1W SOT-MRAM. Our results demonstrate that 1R/1W SOT-MRAM based L2 cache offers great benefits for high performance and energy-efficient on-chip cache applications.

4. AREA-EFFICIENT SOT-MRAM WITH A SCHOTTKY DIODE

This section presents a spin-orbit torque magnetic random access memory (SOT-MRAM) for high-density, reliable and energy-efficient on-chip memory application. Unlike conventional SOT-MRAM requiring two access transistors, the proposed MRAM uses only one access transistor along with a Schottky diode in order to achieve high integration density while maintaining the advantages of SOT-MRAM, such as low write energy and enhanced reliability of magnetic tunnel junction (MTJ). The Schottky diode is forward biased during read, whereas it is reverse biased during write to prevent sneak current paths. Our simulation results show that the proposed MRAM can achieve 30% and 50% reduction in bit-cell area in comparison to conventional STT-MRAM and SOT-MRAM, respectively, and ~2.5× improvement in write power compared to STT-MRAM.

4.1. Introduction

Spin-transfer torque magnetic random access memory (STT-MRAM) has attracted significant attention for on-chip memory applications due to desirable features such as non-volatility, high integration density, and CMOS process compatibility [3],[11],[52],[73]. Despite such attributes, the amount of current to perform a write operation at reasonable speed is large [3], leading to high energy overhead and severe stress on the oxide layer in magnetic tunnel junction (MTJ) [52].

The aforementioned high write current issue can be addressed with spin-orbit torque (SOT) based switching mechanism [14],[16],[53],[58],[74]-[76]. As shown in Fig. 4.1(b),

^{© [2016]} IEEE. Reprinted, with permission, from Y. Seo, K-W. Kwon, and K. Roy, "Area-Efficient SOT-MRAM with a Schottky Diode", *IEEE Electron Device Letters*, Jun. 2016.

write operation in SOT-MRAM is performed by applying charge current via nonmagnetic heavy metal (HM). Spin current, transverse to the charge current is generated and injected into the MTJ's free layer (FL) in direct contact with HM, thus exerting a spin-transfer torque on the FL for magnetization switching [53]. Note, SOT based write operation can be more energy-efficient than conventional STT since a single electron passing through the HM can transfer multiple units of angular momentum [53], [58]. It should be also noted that the reliability associated with the oxide layer can be improved because SOT-MRAM does not require any current flow through the oxide during write. Moreover, decoupled write and read current paths enable separate optimization of SOT device for write and read operations, respectively [74], [77]. However, these advantages are achieved at the expense of an additional transistor. The two-transistor based SOT-MRAM poses a challenge to high-density memory design.



Fig. 4.1. Bit-cell structure of (a) STT-MRAM and (b) SOT-MRAM.

In this section, we propose a new type of SOT-MRAM for high-density memory application. The proposed MRAM does not require an access transistor on the read current path. Instead, a Schottky diode is used based on the fact that read current path is unidirectional. The diode in the proposed MRAM passes unidirectional current through the MTJ during read operation, while it prevents sneak current paths during the write operation. In comparison with conventional SOT-MRAM, our proposed MRAM achieves higher cell density because of reduction in the number of access transistors. Even when

compared to in-plane magnetic anisotropy (IMA) based STT-MRAM, the proposed MRAM is capable of higher density since its lower write current requirement using SOT translates to the smaller access transistor size. In addition, our proposed design has the aforementioned advantages of SOT-MRAM such as low write energy consumption, enhanced MTJ reliability and decoupled write and read current paths.



4.2. Proposed SOT-MRAM Structure

Fig. 4.2. Proposed SOT-MRAM in (a) write operation and (b) read operation.

The proposed memory bit-cell is composed of a write access transistor, a Schottky diode and an SOT device as shown in Fig. 4.2. As will be explained, the Schottky diode is preferentially biased during access by applying appropriate voltages to bit-line (BL), source-line (SL) and two word-lines (WWL and RWL).

In order to write '1' in the bit-cell, BL is set to a positive voltage (V_W), SL to G_{ND} , and WWL is asserted high to turn on the write access transistor. Charge current flows from BL to SL through the HM. Note from Fig. 4.2(a) that there also exists a voltage difference between BL and RWL. However, the diode is reverse biased and thus prevents sneak current flow. The electrons entering from SL are spin-polarized to -y direction at the top surface of HM and exert STT to anti-parallelize the FL magnetization with respect to reference layer (RL) magnetization. To write '0', the direction of the charge current is

reversed by applying G_{ND} to BL and V_W to SL. Then spin-polarized electrons oriented at +y direction exert STT to switch the FL magnetization parallel to the RL magnetization.

The amount of spin current (Is) generated by passing the charge current (Ic) is calculated as follows:

$$I_{S} = \frac{A_{MTJ}}{A_{HM}} \cdot \theta_{SH} \cdot (1 - \operatorname{sech}(\frac{t_{HM}}{\lambda_{HM}})) \cdot I_{C}$$
(4.1)

where θ_{SH} is the spin Hall angle, λ_{sf} is the spin flip length, t_{HM} is the thickness of HM, and A_{HM} and A_{MTJ} are the cross-sectional areas of the HM and the MTJ, respectively [58]. It is noteworthy that Is can be larger than Ic if A_{HM} is much smaller than A_{MTJ} . Using the parameters shown in Table 4.1, we obtain a spin injection efficiency (defined as Is/Ic ×100%) of 258%. In addition to such high write energy efficiency, the proposed bit-cell enhances the MTJ reliability since it does not pass current through the oxide layer during the write operation. Moreover, the oxide thickness (t_{MgO}) can be optimized solely for read operation without impacting write-ability. Note, in the case of STT-MRAM, the oxide thickness trades off read-ability and write-ability due to the shared read and write current paths. In general, the maximum thickness of MgO is limited by the write operation in such devices [14].

In order to perform read operation, a positive read voltage (V_R) is applied to RWL while BL, SL, and WWL are grounded, as shown in Fig. 4.2(b). Under such biasing condition, the Schottky diode is forward-biased, and hence, a unidirectional read current flows from RWL to SL through the MTJ. Since the MTJ resistance in the parallel (P) state is lower than that in the anti-parallel (AP) state, the read current in the P state is higher than that in the AP state. A current mode sense amplifier can sense such difference in read current to determine the state of the MTJ.

4.3. Simulation and Results

Our simulation framework consists of 1) Landau-Lifshitz-Gilbert-Slonczewski (LLGS) equation solver for magnetization dynamics, 2) Non-Equilibrium Green's Function (NEGF) formalism to obtain the MTJ resistance [34], 3) Verilog-A model to

Design Parameter	SOT-Device	STT-Device	
Gilbert Damping, α	0.0122	0.007	
Sat. Magnetization, M _S	1200x10 ³ A/m	1200x10 ³ A/m	
Dimension of free layer	$105nm \times 40nm \times 2nm^{-1}$	$105nm \times 40nm \times 2nm^2$	
$(W_{FL} extbf{X} L_{FL} extbf{X} t_{FL})$	1031111 X 401111 X 21111	10311111 X 401111 X 21111 ²	
Dimension of HM	105nm x 80nm x 2nm	_	
$(W_{HM} \times L_{HM} \times t_{HM})$		_	
Spin Hall Angle (W)	0.3	-	
HM Resistivity	200µΩ·cm	-	
Spin flip length, λsf	1.4nm	-	
MgO thickness, <i>t_{MgO}</i>	1.45nm	1.05nm	
I_C (15ns switching)	115µA	197µA	

Table 4.1 Simulation parameters of the devices

^{1, 2} MTJ free layer is the elliptical shape.

obtain the voltage dependent current of the Schottky diode, and 4) SPICE circuit simulations of the bit-cell.

In order to analyze switching dynamics of MTJ in the bit-cell, we use the LLGS equation solver where the switching time is determined by Is and the device parameters given in Table 4.1. The MTJ switching time can be further determined by Ic using the relation of Is and Ic described in (1). Note from Table 4.1, when compared to the STT device, the SOT device requires lower Ic for the same switching time owing to higher spin injection efficiency (258%). In addition, the magnetic energy barrier of our IMA MTJ is assumed to be ~80k_BT, determined by the method shown in [63]. The energy barrier is good enough for almost fault-free operation even under high temperature condition [78].

For SPICE circuit simulations, the voltage dependent current of the Schottky diode and the resistances of the HM and the MTJ are required. The diode behavior is simulated using a Verilog-A compact model in which the voltage dependent current values closely match with the experimental results on TiO_x-based Schottky diode with a cross-sectional area of $4\mu m^2$, published in [79] (See Fig. 4.3). Using the experimental current density trend with varying cross-sectional area (Fig. 4.3 (b)) [79], we can extrapolate the amount of current flow for the proposed device dimensions. The resistance of HM is calculated using experimental values of resistivity published in [16]. In addition, the resistance of the MTJ in the P and the AP state is obtained by using an NEGF based electron transport simulation framework [34]. The voltage dependent resistances of the MTJ along with the diode model are used with a commercial 45nm transistor model to evaluate its read and write operation.



Fig. 4.3. (a) Matching experimental and SPICE simulated current of TiO_X-based Schottky diode as a function of voltage when cross-sectional area is $4\mu m^2$. (b) Experimental current density trend with varying the cross-sectional area [79].

For area and power comparison of our proposed memory to conventional STT-MRAM and SOT-MRAM, three different bit-cells are designed to meet the following specifications–15ns switching time, 10% write margin (defined as $(I_W-I_C)/I_C$) and 30% sensing margin (defined as $(I_R-I_{REF})/I_{REF}$) (the reference current (I_{REF}) is calculated as

 $(I_{R_AP}+I_{R_P})/2)$. The layouts of STT-MRAM, SOT-MRAM, and the proposed MRAM are shown in Fig. 4.4 using λ -based design rules [38]. In Table 4.2, we compare the area of the three different memory bit-cell layouts. Since the area of MRAM cell is mainly determined by the number of the access transistors, conventional two transistor-based SOT-MRAM has a larger area than STT-MRAM whereas our proposed SOT-MRAM can mitigate the area overhead because of single transistor-based design. Furthermore, our proposed SOT-MRAM can be more area-efficient than STT-MRAM since, unlike STT-MRAM, smaller write current leads to the smaller access transistor. Because two access transistors in conventional SOT-MRAM are exclusively used for write and read operations, one might think that the integration density of SOT-MRAM can be improved by reducing the size of the read access transistor (which provides a small amount of read current). However, because the transistor width is in the metal pitch limited region, reduction in the transistor size below the metal spacing limit does not affect the overall bit-cell area.



Fig. 4.4. Layouts and schematics of STT-MRAM, SOT-MRAM, and proposed SOT-MRAM bit-cells. W_N represents the width of the transistor.

	STT-MRAM	SOT-MRAM	Proposed SOT-MRAM
Tx Width (W _N)	410nm	180nm (Write) 180nm (Read)	180nm
Area	$0.0848 \mu m^2$	$0.1104 \mu m^2$	$0.0552 \mu m^2$
Write Voltage	1.0V	0.7V	0.7V
Read Voltage	0.2V	0.2V	0.8V
Write Power	249.24µW	100.60µW	100.60µW
Read Power	33.54µW	1.22µW	10.50µW
Read-disturb Margin	71%	>95%	>95%

Table 4.2 Simulation results and comparison of MRAM Bit-cells

In terms of power and reliability, our proposed SOT-MRAM exhibits distinct advantages over STT-MRAM. First, our proposed memory can achieve ~2.5X improvement in write power compared to STT-MRAM on account of lower write current. Second, as mentioned earlier, oxide layer reliability in STT-MRAM can be alleviated since no current flows through the MTJ during the write operations in SOT-MRAM. Finally, our proposed SOT-MRAM achieves lower read power and higher read-disturb margin (defined by $(I_C-I_R)/I_C$) than STT-MRAM. This is because write and read current paths of the SOT device are electrically separated so that thicker oxide layer can be used for improving disturb margin and read power. However, the read power of our proposed MRAM is expected to be larger than that of conventional SOT-MRAM since the read voltage had to be increased to overcome the diode turn-on voltage.

4.4. Conclusion

In this section, we proposed an area-efficient SOT-MRAM with a Schottky diode. In our proposed memory, the Schottky diode is introduced in the read current path (instead of a read access transistor) to achieve lesser area than two transistor-based SOT-MRAM. The low write current in SOT device further improves the area by utilizing a small access transistor while simultaneously achieving low write power. Moreover, the reliability of the oxide layer is enhanced since high write current does not flow through the MTJ. We believe that our proposed MRAM can be a promising candidate for future high density, reliable and low power on-chip memories.

5. SHARED BIT-LINE SOT-MRAM STRUCTURE FOR HIGH DENSITY ON-CHIP CACHES

This section proposes a new design technique of spin-orbit torque magnetic random access memory (SOT-MRAM) which is suitable for high density and low power on-chip cache applications. A bit-line of the proposed memory bit-cell is shared with that of an adjacent bit-cell, hence, minimum allowable area of our proposed structure can be improved in comparison with conventional structure of SOT-MRAM by reducing the number of metals along the column direction. Furthermore, since the efficient spin orbit torque based switching operation can translate to smaller size of access transistors, the proposed SOT-MRAM achieves higher integration density compared to spin-transfer torque magnetic random access memory (STT-MRAM) while maintaining the advantages of SOT-MRAM such as low write energy dissipation, high read-disturb margin, and improved reliability of magnetic tunnel junction (MTJ). Compared to the conventional SOT-MRAM bit-cells, our proposed SOT-MRAM bit-cell can have 20% reduction in bit-cell area. In addition, the proposed memory achieves > 6× lower write power and higher read-disturb margin than STT-MRAM.

5.1. Introduction

Spin-transfer torque magnetic random access memory (STT-MRAM) is regarded as possible next-generation on-chip memory due to its non-volatility, high integration density, and CMOS process compatibility [11]. However, STT-MRAM requires large amount of write current, hence, considerable MRAM research has been focused on

Y. Seo, and K. Roy. *Shared Bit-line SOT-MRAM Structure for High Density On-chip Caches*. Submitted to the IEEE Magnetics Letters, 2016.

minimizing write current [17]. To address this challenge, spin-orbit torque (SOT) based novel switching mechanism has been recently proposed [15], [16], [53]. By a flow of charge current through a heavy metal (HM), large amount of spin current is generated in the direction transverse to the charge current so that a low current switching operation is enabled in the current-induced SOT mechanism [16].



Fig. 5.1 1-by-2 array structure of (a) STT-MRAM and (b) SOT-MRAM.

Despite such attribute, one of the biggest disadvantages of SOT-based memory design is that two transistors are required per a bit-cell resulting in area-overhead [15]. As shown in Fig. 5.1 (b), the write and read current path of SOT-MRAM is decoupled such that each current path requires each access transistor to control the two separate current paths. Thus, SOT-MRAM may not be a promising alternative for high density on-chip memory. In SOT-MRAM, there is a possibility of area improvement by reducing the size of write access transistor aggressively (due to small write current). However, there is a limitation for improving the bit-cell area by reducing the width of write access transistor. In conventional SOT-MRAM, two metals per a bit-cell is routed along the column,

hence, reduction in the write transistor size below a space for 2 metal lines does not affect to the bit-cell area [15], [38].

In this section, we propose a high-density SOT-MRAM design by using a shared bitline structure. Read-bit-line (RBL) of the proposed MRAM cell is shared with that of a neighboring cell such that we can reduce the number of metals along column direction to 3 metal lines per 2 bit-cells. This results in 20% improvement in minimum allowable bitcell area compared to conventional SOT-MRAM. Moreover, our proposed design still has the advantages of SOT device such as low write energy consumption and decoupled write and read current path.

5.2. Shared Bit-line SOT-MRAM Structure



Fig. 5.2. (a) SOT device structure, and (b) Direction of current flows during an antiparallel switching operation.



Fig. 5.3. 1-by-2 array structure of our proposed SOT-MRAM.
Fig. 5.2 illustrates the structure of the SOT device used as storage element in our proposed memory. The storage device based on current-induced SOT is composed of an MTJ, an HM, and a spin-sink layer (SSL) where the HM is directly contacted with the free layer of the MTJ and the SSL [58], [59]. The SOT device has two main advantages over the MTJ. A first major benefit is that SOT-based switching mechanism can achieve lower switching current than MTJ due to the high spin current injection efficiency. Because one electron travelling through the HM can transfer multiple units of angular momentum, SOT device, an SSL is used to achieve even higher spin current injection efficiency by reducing the backflow of spin currents caused by the effect of spin accumulation at the bottom surface of HM [59].

$$I_{S} = \frac{A_{MTJ}}{A_{HM}} \cdot \theta_{SH} \cdot I_{C}$$
(5.1)

	Odd cells		Even cells		
	Write	Read	Write	Read	
WL_A	0	VDD	VwL	0	
WL_B	VwL	0	0	VDD	
S_RBL	0	VR	0	VR	
WBL[1]	Vwp(Vwn)	0	0	0	
WBL[2]	0	0	Vwp(Vwn)	0	
SL	0	0	0	0	

Table 5.1. Biasing conditions for write and read operations in our proposed memory.

Eq (1) describes the spin current injection efficiency defined as the ratio of the injected spin current to the MTJ and the charge current following through the HM, where θ_{SH} is a spin Hall angle, and A_{HM} and A_{MTJ} are the cross-sectional areas of HM and MTJ, respectively [58]. Typically, the HM has much smaller cross-sectional area than the MTJ so that the SOT device can achieve high spin current injection efficiency (471% using the device parameters given in Table. 5.2).

A second major advantage of the device is that the SOT device can perform switching operations by having a current flow through the HM, and not the MTJ. For example, for parallelizing the FL with regard to a reference layer (RL), charge current flows from T1 to T2, and hence, electrons entering HM are spin-polarized to +y direction and injected to the FL of MTJ. On the other hand, anti-parallel state of the MTJ can be written by reversing the direction of the charge current. Since write operation can be performed without any current directly flowing through the MTJ, the memory not only mitigates MTJ reliability issue caused by high write current through the MTJ but also allows separate optimization for write and for read due to decoupled write and read current paths [15], [77].

Fig. 5.3 and Table 5.1 show shared bit-line memory structure using the SOT device and the biasing conditions for write and for read in our proposed memory. Unlike conventional SOT-MRAM, bit-lines for write and read (write-bit-line (WBL) and readbit-line (RBL)) are separated and routed to the column direction, and a RBL of an even cell is shared with that of the neighboring odd cell. Thus, the number of the metals along column direction in our proposed memory is reduced to 3 metal lines per 2 bit-cells in order to improve the integration density. In addition, we can prevent any increase in the number of metals along the column direction by routing the SL to row direction.

Let us consider the write and the read operations of our proposed memory bit-cell. Please note that, during write operations, the biasing condition for odd cells and that for even cells are different. This is due to the fact that a word-line (WL) for a write access transistor in odd cells (WL_B) and that in even cells (WL_A) are different. During the write operation of a memory cell in an odd row, odd cell's write-bit-line (WBL[1]) is set to write voltage (V_{WP} or V_{WN}, depending upon the desired data), SL to G_{ND}, and WL_B is asserted high. For instance, to write '1', positive write voltage (V_{WP}) is applied to WBL[1] such that write current is flowing from WBL[1] to SL, whereas, in order to write '0', the direction of write current is reversed by applying negative write voltage (V_{WN}) to WBL[1]. In addition, during the operation of write '0', negative voltage is applied to WBL[1] and WL_B kept high. Note, however, that under such a condition V_{GS} can be larger than V_{DD}, resulting in reliability issues such as bias-temperature instability (BTI)

on the transistor. The possible transistor reliability issue can be alleviated by reducing the WL voltage [15]. In the proposed memory, owing to efficient SOT-based write operation, sufficient write current can be provided even under reduced gate voltage, and hence, V_{GS} of write access transistor need not exceed V_{DD} . Moreover, during write operation of odd cells, possible sneak current flowing to the even cells can be prevented by keeping WL_A to G_{ND} . On the other hand, a write operation of the even cell can be performed by setting write voltage to the WBL[2] and turning on the WL_A. Note, WL_B is turned off to avoid unwanted current flowing to the odd cell.

The biasing conditions for an odd cell's read operation and that for an even cell's read operation are also different since different read-word-lines are utilized for the cell in odd rows and even rows. For example, in order to sense an odd-row cell, a read access transistor of the odd cell is turned on and a write access transistor of the cell is turned off by applying V_{DD} and G_{ND} to WL A and WL B, respectively. Then, the read voltage (V_R) is set to the shared read-bit-line (S RBL) such that a read current is flowing from S RBL to the SL. The amount of read current is determined by the resistance of the storage device (The MTJ resistance is low when the MTJ is in the parallel state, while the MTJ resistance is high when the MTJ is in anti-parallel state) so that we can sense the stored data by comparing the amount of read current with a reference current. On the other hand, the read operation of an even-row cell is done by asserting WL B high to turn on the read access transistor and applying a read voltage to S RBL. Note that the read operations of the odd and even cells cannot be performed at the same time since the bitcells in odd rows and even rows require different biasing conditions. However, typically in on-chip cache applications, simultaneous accesses to the odd-row cells and even-row cells does not occur because bit-interleaving is widely used in the on-chip memory arrays to achieve high error tolerance with conventional error-correction schemes [66], [80].

5.3. Modeling and Simulation

Our simulation framework is composed of 1) Landau- Lifshitz-Gilbert-Slonczewski (LLGS) equation solver for the modeling of magnetization dynamics, 2) Non-Equilibrium Green's Function (NEGF) formalism to obtain the MTJ resistance [34], and 3) SPICE circuit simulations of the bit-cells. The LLGS equation solver with the device parameters shown in Table 5.2 was used to model the magnetization dynamics of storage devices and to obtain critical switching current. Please note from Section 5.2, the spin current injection efficiency, the relation

Design Parameter	STT Device	SOT Device	
Gilbert Damping, α	0.007	0.0122	
Sat. Magnetization, MS	1000x10 ³ A/m	1000x10 ³ A/m	
Dimension of free layer	$120mm = 40mm = 1.5mm^2$		
(WFL x LFL x tFL)	120nm x 40nm x 1.5nm -	120nm x 40nm x 1.5nm ²	
Dimension of SHM		120nm x 80nm x 2nm	
(WSHM x LSHM x tSHM)	—		
SHM Resistivity	_	200μΩ·cm	
Spin Hall Angle, θSH	_	0.3	
MgO thickness, <i>t_{MgO}</i>	1.15nm	1.30nm	
Critical Current (10ns)	155µA	47μΑ	

Table 5.2. Simulation parameters of the devices

^{1,2} MTJ free layer has the elliptical shape.

between I_S and I_C which is described in (1), should be considered when determining the critical current of SOT device. As shown in Table 5.2, compared to STT device, SOT device can perform free layer switching operation under lower write current owing to its high spin injection efficiency (471%).

For SPICE circuit simulations, the resistance values of the HM and the MTJ are required. The resistance of HM is calculated using experimental values of its resistivity published in [16] and the dimension of HM. In addition, the resistance of MTJ in P and AP state is obtained by using a NEGF based electron transport simulation framework [34]. Our NEGF based simulation framework was successfully calibrated the experimental data shown in [52]. The aforementioned SOT device model is used with a commercial 45nm transistor model to form the memory bit-cell structure and simulated to evaluate its write and read operations.

For area and power comparison of our proposed memory to conventional STT-MRAM and SOT-MRAM, three different bit-cells are designed to meet the following specification: 10ns switching time, 5% write margin (defined as $(I_W-I_C)/I_C$) and 35% sensing margin (defined as $(I_R-I_{REF})/I_{REF}$). Fig. 5.4 shows the layouts of STT-MRAM, SOT-MRAM and the proposed MRAM designed by using λ -based design rules [15], [38]. As shown in Table 5.3, conventional SOT-MRAM has larger area than STT-MRAM due to the need for an extra more transistor whereas our proposed SOT-MRAM can mitigate the area overhead. As mentioned earlier, a RBL is shared between two adjacent bit-cells so that the number of metals along the column direction is reduced. Therefore, the bit-cell area of the proposed SOT-MRAM can be further optimized with aggressively reduced write access transistor (by taking an advantage of the small write current requirement of the SOT device). On the other hand, in conventional SOT-MRAM, two metals per a cell are routed along the column direction so that the area optimization is limited by the space for the two metals. Moreover, note, even though two transistors are required for our proposed design, the memory can be more area-efficient than standard STT-MRAM.



Fig. 5.4. Layout comparison of STT-MRAM, SOT-MRAM, and our proposed SOT-MRAM.

	STT-MRAM	SOT-MRAM	Proposed SOT-MRAM
Ty Width (W)	110nm	180nm (Write)	80nm (Write)
$1 \times \text{wrdun}(\text{w}_{\text{N}})$	4401111	180nm (Read)	180nm (Read)
Area	$0.0896 \mu m^2$	$0.1104 \mu m^2$	$0.0874 \mu m^2$
Write Voltage	1.00V	0.12V	$0.96V$ / -0.23V 1
Read Voltage	0.20V	0.20V	0.20V
Write Power	208.90µW	6.06µW	29.62µW
Read Power	22.86µW	7.08µW	7.08µW
Read-disturb Margin	77%	89%	89%

Table 5.3. Results and comparison of three different memory bit-cells.

¹ WL voltage for the write operation is 0.77V.

In terms of power and robustness, the proposed SOT-MRAM shows distinct advantages over STT-MRAM. First, our proposed memory achieves > $6\times$ lower write power than STT-MRAM. Second, the proposed SOT-MRAM can improve read power and read-disturb margin (defined as (I_R-I_C)/I_C) compared to STT-MRAM. This is due to the fact that, in SOT device, thicker oxide can be used for improving read power and the disturb margin without affecting the writ-ability of the cell since the write and the read current paths of the device are electrically separated. Finally, as previously mentioned, the reliability issue on oxide layer associated with STT-MRAM may be mitigated in SOT-MRAM since large write current does not flow through the MTJ. However, the write power of our proposed MRAM is expected to be somewhat larger than that of conventional SOT-MRAM. This is because higher write voltage is required in our proposed memory in order to provide the write current with smaller transistor.

5.4. Conclusion

In this section, we proposed a high density SOT-MRAM by using shared BL structure. In our proposed memory, RBL of a bit-cell is shared with a neighboring cell so that the proposed MRAM has improved the integration density. Since the number of metals along column direction is reduced, the bit-cell area can be further improved by

reducing the size of the write access transistor. Due to the need for lower write current in SOT device, the proposed SOT-MRAM can provide sufficient write current with a small sized write access transistor. Moreover, separate write and read currents paths of SOT device allows us to mitigate oxide reliability and to improve disturb failures during read operation. Consequently, our proposed MRAM bit-cell can be a prospective candidate for future high density, low power, and robust on-chip memories.

6. NONVOLATILE FLIP-FLOP BY USING COMPLEMENTARY POLARIZER MTJ

Nonvolatile flip-flop (NVFF) using spin-transfer torque magnetic tunnel junctions (STT-MTJs) has been proposed to enable fine-grain power gating systems. However, the STT-MTJ based NVFF (STT-NVFF) may not perform fast backup and disturb-free restore operations. We propose a new NVFF using complementary polarizer MTJ (CPMTJ) to alleviate these limitations. Our proposed NVFF exploits the CPMTJ structure for fast and low-energy backup operation. The estimated backup delay is less than 10ns in 7nm node FinFET technology with CPMTJ size of 12nm \times 33nm in a rectangular shape. Furthermore, during the restore operation, CPMTJ provides guaranteed disturb-free sensing since disturb torque in CPMTJ comes from two pinned layer and is cancelled each other. The simulation results show > 2x improvement in the backup delay with higher restore-disturb margin compared with the STT-NVFF.

6.1. Introduction

As CMOS technology scales down, the leakage power consumption during the standby mode increases because of lowered threshold voltage. To reduce the leakage power consumption during sleep mode, transistor stacking with single-threshold CMOS, multi-threshold CMOS, and reverse- bias control have been proposed [5], [7], [9]. However, these techniques do not completely eliminate the leakage power because the power still needs to be provided to hold the data in the flip-flop during the sleep mode. Thus, power gating architectures (shutting down the unused blocks, thereby reducing

Y. Seo, and K. Roy. *Fast and Disturb-Free Nonvolatile Flip-Flop using Complementary Polarizer MTJ*. Submitted to the IEEE Trans. VLSI Systems, 2016.

leakage power) has received increasing attention for the past several years. Fig. 6.1 (a) shows a block diagram of a typical power gating system to achieve reduced leakage power where circuit blocks are turned off during the sleep mode.



Fig. 6.1. Power gating block diagram (a) without NVFF and (b) with NVFF.

Note that power gating may require additional power and delay overhead during the backup and the restore operation. Hence, a judicious action is required to activate finegrained power-gating to amortize the power and delay overhead. Power-gating requires the flip-flop data to be stored before power-off and to be restored right after the virtual VDD is charged up. The retention registers, which typically have an auxiliary register that is slower but has less leakage current, are used for storing the contents of the flip-flops in the functional blocks during power gating [81]. However, as shown in Fig. 6.1 (a), expensive power and delay overhead are expected due to data transfer to the retention registers through the long-distance buses. In order to reduce such power and delay overhead, nonvolatile flip-flops (NVFFs) have been widely investigated (Fig. 6.1 (b)) [19], [76], [82]. NVFF acts as a standard flip-flop during the active mode. When the sleep mode is activated, NVFF saves its current logic state into its nonvolatile storage elements before the power supply is gated. When the power is restored, the data in the storage elements is restored to the NVFF so that the processor can resume from the state prior to power gating.

STT-MTJ has been considered as a promising nonvolatile storage element for NVFF due to long data retention time, excellent endurance, and CMOS process compatibility [82]. The cross-coupled inverter in the nonvolatile slave latch is utilized for storing the data to STT-MTJ and restoring the data from STT-MTJ. For backing up data into STT-MTJs, spin-polarized electrons are passed through it to program the STT-MTJ state. However, as we will discuss later, two steps are needed to perform the backup operation in the STT-MTJ based NVFF, and furthermore, fast and low energy backup operation is difficult to achieve. Indeed, during the restore operation, the MTJ state can be potentially disturbed by a restore current.

In this section, we propose a complementary polarizer MTJ (CPMTJ) based NVFF for fast backup and robust restore operations. Due to one step backup operation, the delay and energy consumption of the backup operation can be improved compared with that in STT-MTJ based NVFF. Indeed, CPMTJ used as a storage element in the proposed NVFF can inherently provide truly disturb-free restore operation since disturb torque from two pinned layers is cancelled each other in the CPMTJ.

6.2. Review of STT-NVFF



Fig. 6.2. Two step backup operations in the STT-MTJ based nonvolatile slave latch (a) in the 1st step, (b) in the 2nd step.

The STT-MTJ based NVFF (STT-NVFF) is composed of two STT-MTJs, and four additional transistors – two access transistors (N2, N3), an equalizing transistor (P1) and a footer transistor (N1) with standard flip-flop [19]. The STT-MTJ is the nonvolatile storage element which consists of a free layer (FL), pinned layer (PL) sandwiching a tunneling barrier (MgO). The PL magnetization is fixed whereas the magnetization of FL is stable at either parallel (P) state or anti-parallel (AP) state with regard to the PL magnetization.



Fig. 6.3. A restore operation in the STT-NVFF based nonvolatile slave latch when stored data is (a) 1, and (b) 0.

A backup operation is performed by passing a current through the STT-MTJ that exceeds a critical current for a minimum length of time. However, in the STT-NVFF, it is difficult to achieve high speed backup operation because of two reasons. First, two steps are needed to perform the backup operation in the STT-NVFF [19]. In the STT-NVFF, two separate STT-MTJs (STT-L and STT-R) are connected to the nonvolatile slave latch. One of two STT-MTJs is needed to be switched to the AP state and another STT-MTJ is switched to the P state during the backup operation. In Fig. 6.2(a), for example, when CTRL=1 during the first step of the backup operation, a current flows from FL to PL of STT-L so that STT-L is switched to the P state. In the second step of the backup operation, as shown in Fig. 6.2(b), STT-R is switched to the AP state by flowing a current from PL to FL of STT-R when CTRL=0. Second, the speed of backup operation in STT-

NVFF is degraded by the asymmetry in the latency of the first step and second step backup operation. A longer time is needed to anti-parallelize the FL with PL in the second step compared with parallelizing the FL with PL in the first step. This is due to the fact that electrons entering the FL for anti-parallelizing the MTJ state are not wellpolarized compared with that entering the PL for parallelizing the MTJ state [39]. Moreover, slow anti-parallel switching operation is aggravated by the sourcedegeneration effect [26]. As shown in Fig 6.2(b), the access transistor is sourcedegenerated by the STT-MTJ in the second step. Because of the source-degeneration effect, the gate-source voltage (V_{GS}) of N3 is less than V_{DD} , leading to a weak current driving capability of the access transistor. Therefore, the overall speed of backup operation in STT-NVFF is severely degraded by the slow anti-parallelizing operation.

During a restore operation, power supply line is pulled up, CTRL is set to G_{ND} , EQ to G_{ND} , and AC is asserted high to turn on the access transistors. Then, small currents flow from QT (I_{RES_L}) and from QC (I_{RES_R}) to CTRL (Fig. 6.3). Since the STT-MTJ in the P state has a lower resistance compared to when it is in the AP state, a restore operation can be performed by sensing the resistance of STT-MTJ. However, during the restore operation, a MTJ storing a P state can be potentially disturbed by the restore current which tries to anti-parallelize the FL [83]. If the MTJ is accidentally overwritten during the restore operation, there is a possibility of incorrect information being restored.

6.3. Device and Proposed NVFF Structure

In this section, we first describe the details of the CPMTJ device. Then, we present the design of CP-NVFF and discuss its benefits.

6.3.1. Device Structure of Complementary Polarizer MTJ

The device structure of CPMTJ with perpendicular magnetic anisotropy (PMA) is shown in Fig. 6.4 [39]. This three-terminal device is composed of two pinned layers (PLs) and a free layer (FL). The magnetizations of PL1 and PL2 are opposite in direction and are fixed. However, the magnetization of FL may be pointing in either the +z or the – z direction. The magnetization of FL is switched by passing electrons from only one of the two PLs into the FL. When charge current flows from T1 to T2, the FL is switched to the +z direction, whereas FL is switched to the -z direction when the charge current flows from T1 to T3. A key advantage of the CPMTJ is that only AP to P switching occurs during programming [39]. Since electrons entering the PL during AP to P switching, are well polarized, the delay needed to align a FL with a PL is shorter than that needed to anti-align the FL with the same PL so that CP-NVFF can achieve high speed backup operation.



Fig. 6.4. Device structure of the complementary polarizer MTJ.

The CPMTJ presents the different amount of resistance based on the magnetization of FL with respect to the magnetization of two PLs. When the magnetization of FL is in the +z direction, the magnetizations of the FL and PL2 are anti-parallel (AP) and the resistance between FL and PL2 via the tunneling oxide is high. On the other hand, the magnetizations of FL and PL1 are parallel (P) and the resistance between FL and PL1 via the tunneling oxide is low. When the magnetization of FL is in the -z direction instead, the resistance between FL and PL1 is high and that between FL and PL2 is low. So, we can sense the stored data by the relative resistance of two current paths.

6.3.2. Proposed NVFF Structure

The schematic of our proposed NVFF is shown in Fig. 6.5. The nonvolatile slave latch includes a CPMTJ and four additional transistors (N1, N2, N3, and P1). This

nonvolatile slave latch is connected with a master latch to form a master-slave flip-flop. During active mode, the nonvolatile flip-flop acts as standard master-slave flip-flop (EQ = V_{DD} and AC = CTRL = G_{ND}). The nonvolatile slave latch does not access the nonvolatile storage element during the active mode which may help to alleviate performance degradation during normal flip-flop operations.



Fig. 6.5. Schematic of nonvolatile flip-flop using CPMTJ.



Fig. 6.6. Timing diagram of CPMTJ based NVFF operation

The timing diagram in Fig. 6.6 shows how the signals EQ, AC, and CTRL are sequenced during the backup and restore operation. To perform the backup operation, the AC signal is asserted high and the EQ signal is held high to save the flip-flop state into the CPMTJ. When the CTRL signal is active, a charge current flows from the one of the two PLs to the FL due to the voltage difference between the QT or QC and CTRL. For example, FL magnetization of CPMTJ is switched to +z direction when NVFF stores '0' as shown in Fig. 6.7(a) whereas when the output of the NVFF is '1' in Fig. 6.7(b), FL magnetization of CPMTJ is switched to -z direction. Hence, a high speed backup operation is enabled by possible one step backup operation. In addition, the access transistors (N2 and N3 in Fig. 6.5) are not source-degenerated by the magnetic storage device during the backup operation. Note that the available write current is reduced by the source-degeneration effect during P to AP switching in the STT-NVFF (Fig. 6.2(b)) [26], and may degrade the delay of the backup operation. Furthermore, as explained earlier, our proposed NVFF can further improve the performance of backup operation.



Fig. 6.7. One step backup operations in the CPMTJ based nonvolatile slave latch (a) at Q='1', and (b) at Q='0'.



Fig. 6.8. A restore operation in the STT-NVFF based nonvolatile slave latch when stored data is (a) 1, and (b) 0.

When exiting the sleep mode, the power supply line is pulled up while the EQ signals remain low to initiate the restore operation. At first, the equalizing PMOS is activated by the low EQ signal and the nodes QT and QC in the slave latch are initialized to the same voltage (Fig. 6.6). Next, when the AC signal goes high, the current flowing through each branch of the CPMTJ depends on the magnetization of the FL. For example, as Fig. 6.8(a) shows when the FL magnetization is in the -z direction, I_{RES_R} is larger than I_{RES_L} whereas, in Fig. 6.8(b), I_{RES_L} is larger than I_{RES_R} when FL magnetization is in the +z direction. Then, a small voltage difference is developed between QC and QT of the nonvolatile latch. When the EQ signal goes to V_{DD} , the equalizing PMOS is turned off and the footer NMOS is turned on. The small voltage difference between QC and QT in the slave latch is restored to V_{DD} due to the cross-coupled inverter action (i.e., the voltage of QT and QC in the slave latch will go to V_{DD} and G_{ND} or G_{ND} and V_{DD} , respectively). Lastly, the AC signal is then turned off and the proposed NVFF goes into the active mode.

Note the fact that there is a possibility that the magnetization of FL may be accidentally overwritten because a restore current is flowing through the FL of CPMTJ, resulting in restore disturb failure. However, because of two reasons, our proposed NVFF

can achieve almost disturb-free restore operation. First, a current needed to accidentally flip the CPMTJ is higher than that of MTJ since CPMTJ has larger dimensions of the FL

Energy Barrier (EA)	40 <i>k</i> _B <i>T</i> at T=300K
Sat. Magnetization, Ms	860x10 ³ A/m
Gilbert Damping, α	0.014
Polarization (PPL, PFL)	0.8 / 0.3
STT Fitting Parameter, Λ	2
Free Layer Size (CPMTJ)	12nm X 33nm X 1.5nm
Free Layer Size (STT-MTJ)	12nm X 12nm X 1.5nm
MgO thickness	1.07nm
CMOS Technology	7nm node Tied-gate FinFET

Table 6.1 Iso-retention Time Simulation Parameters

Table 6.2 Transistor size in proposed NVFF

Transistor	P1	N1	N2 / N3	Other PFET	Other NFET
Number of Fin	3	5	2	3	2

as compared to that in the MTJ. Second, the disturb torque exerted by the current flowing through one PL in the CPMTJ device is cancelled by that exerted by the current flowing through the other PL [39]. Therefore, as compared with STT-NVFF where the disturb torque comes from only one PL, our proposed CP-NVFF achieves much higher disturb margin during restore operations.

6.4. Modeling and Simulation

We implemented the simulation framework proposed in [39] to analyze the NVFFs. Comparison of our proposed CP-NVFF to the STT-NVFF requires the modeling of 1) electronic transport simulation using the Non-Equilibrium Green's Function (NEGF) formalism which captures the interaction between the magnetic state of the device and its electrical characteristic, and 2) Object-Oriented Micro-Magnetics Framework (OOMMF) which can model the magnetization dynamics of FL of STT-MTJ and CPMTJ. For calculating the amount of current flowing during the backup and restore operation, a resistance of the oxide barrier in STT-MTJs and CPMTJ should be obtained. Using a spin-dependent transport solver with the Non-Equilibrium Green's Function (NEGF) formalism, the relationship between the resistance of STT-MTJ and CPMTJ, and their magnetic configurations is determined [39]. MTJ characteristics, obtained using our spin-dependent transport solver, are calibrated to experimentally measured MTJ characteristics in [52] to make our analysis as realistic as possible. We then simulate the backup and restore operations of STT-NVFF and our proposed NVFF in HSPICE circuit simulation with 7nm node Tied-gate FinFET transistor model (nominal V_{DD} : 0.7V) [84]. The device dimension and the number of fins for each transistor in STT-NVFF and CP-NVFF are listed in Table 6.1 and Table 6.2, respectively.

Then, OOMMF based magnetization dynamics simulation is utilized to obtain the switching time to successfully program the CPMTJ and STT-MTJ devices with our assumed simulation parameters in Table 6.1 and the current during backup operation [31]. Thus, the delay of the backup operation in STT-NVFF and CP-NVFF is determined using the micro-magnetic simulations. Note that the flow of spin-polarized electrons may not be uniform in CPMTJ and hence, the effect of the localized current injection in CPMTJ device has to be taken into account in the OOMMF simulation

The layouts of STT-NVFF and CP-NVFF are first determined so that we can compare the area efficiency of NVFFs. Note that we have assumed 7nm node FinFET technology [84] for our simulations. The layouts for STT- NVFF and CP-NVFF (Fig. 6.9) are developed using the FinFET based design rules described in [85], [86]. Typically, the area of NVFFs is dominated by the size of each transistor so that we can perform a detailed comparison of different NVFFs under iso-area condition by using the same number of fins in each transistor. However, as shown in Fig. 6.9, their layout areas are 37% bigger than that of standard Flip-Flop (FF) since the additional transistors for the backup and restore operation are introduced in STT-NVFF and CP-NVFF. The delay and energy dissipation of the backup operation are shown in Table 6.3. We have performed three different process, voltage, and temperature (PVT) corner simulations (worst case: $SS/0.9V_{DD}/125^{\circ}C$, typical case: $TT/V_{DD}/25^{\circ}C$, and best case: $FF/1.1V_{DD}/-45^{\circ}C$) to

evaluate NVFFs in various environments. Under the same amount of current condition, when the temperature increases, the switching time of spintronic devices reduces due to the thermal fluctuation [87]. Note, however, the backup delay in the worst corner (high temperature case) is longer than that in the typical and the best corner cases since the amount of current during the backup operation is degraded by the transistors in the worst condition.





Fig. 6.9. Layout of (top) STT-NVFF and (bottom) CP-NVFF. The dashed rectangular indicates the region of standard flip-flop.

	Worst Corner		Typical Corner		Best Corner	
Parameter	STT- NVFF	CP- NVFF	STT- NVFF	CP- NVFF	STT- NVFF	CP- NVFF
Delay	25.7ns ^a	8.1ns	23.9ns ^b	6.6ns	17.7ns ^c	5.6ns
Energy	120.25fJ	84.23fJ	114.86ns	81.64fJ	96.59fJ	80.05fJ

Table 6.3 Energy and delay comparison of NVFF backup operation

^{a, b, c} The backup delay in STT-NVFF is the sum of AP to P switching delay in the 1st step and P to AP switching delay in the 2nd step.

Parameter	STT-NVFF		CP-NVFF	
Restore-disturb Margin	3.1µA		23.1µA	
Delay	2		ns	
Energy	6.88fJ ^a 9.2		8fJ ^b	11.35fJ °

Table 6.4 Restore operation comparison of NVFF

^a Worst corner, ^b Typical corner, ^c Best corner.

The complementary PLs in CPMTJ need to be separated by a certain distance due to the layout rules. For connecting with the PLs, the size of the FL in CPMTJ needs to be extended so that more time is required to switch CPMTJ compared to STT-MTJ. For instance, the switching time for CPMTJ (6ns at the typical corner) is $\sim 3 \times$ longer than AP to P switching time for MTJ (2ns at the typical corner). However, during the backup operation, CP-NVFF has > 2x less delay than STT-NVFF in all three corner cases because of three reasons. First, STT-NVFF needs two step backup operations, while, as discussed earlier, the backup operation in CP-NVFF can be performed in a step. Second, unlike CP-NVFF, slow P to AP switching operation occurs in STT-NVFF. Lastly, slow anti-parallelizing operation in STT-NVFF is severely deteriorated by the effect of source degeneration. Due to the source-degeneration effect, the backup current during the second step (8.1µA at the typical corner) becomes smaller than that during the first step (12.4µA at the typical corner). Therefore, the overall latency of backup operation in STT- NVFF is aggravated by the slow anti-parallelizing operation. Furthermore, due to shorter backup delay, CP-NVFF can achieve lower energy dissipation.

Finally, Table 6.4 shows the delay, energy, and disturb margin in CP-NVFF and STT-NVFF during the restore operation. The MgO thickness of spintronic devices and the size of the each transistor in both NVFFs are the same to evaluate the backup speed and disturb margin under the iso-restore delay/energy condition. In addition, in order to compare the disturb failure, we defined the restore-disturb margin as I_{C_2ns} - I_{READ} under the condition (FF/1.1V_{DD}/125°C) most vulnerable to the restore- disturb failure. As shown in Table 6.4, the disturb margin for CP-NVFF is much bigger than that for STT-NVFF. Due to its larger volume, CP-NVFF requires more current to accidentally flip the data during the restore operation. Furthermore, the disturb margin of CP-NVFF is expected to be even better. During the restore operation, two restore currents flow simultaneously. Note, hence, the torque exerted by one read current may be canceled by the torque by the other read current so that our proposed NVFF can perform near disturb-free restore operation.

6.6. Conclusion

We propose a complementary polarizer MTJ based NVFF which has less delay and energy consumption during the backup operation compared with STT-NVFF. The possible one step backup operation in CP-NVFF can reduce the delay and energy consumption of the backup operation. Avoiding the anti-parallelizing operation (present in STT-MTJ based design) can further improve the performance and energy consumption of backup operation. Besides, our proposed scheme can mitigate the restore disturbance problem in STT-NVFF. Therefore, we believe that the proposed NVFF is suitable for the fast and robust fine-grained power gating systems.

7. CONCLUSIONS

Integrated circuits designed for mobile applications should meet stringent power dissipation requirements due to limited battery capacity. Lowering the supply voltage by technology scaling is one of the most effective ways to improve the power consumption. However, scaling down of transistor size leads to the substantial increase in the leakage current. In order to mitigate high standby power consumption of conventional CMOS technology, a number of nonvolatile storage technologies are being investigated. Among several new technologies, spin-transfer torque devices are promising candidate for future on-chip memories. In this work, we proposed new spin-based devices for on-chip caches and their application for low standby power system.

In Chapter 2, we presented a new spin-transfer torque device – Domain Wall Coupling-based STT-MRAM (DWCSTT). STT-MRAM is a promising candidate for onchip memory due to its low leakage power consumption, nonvolatility, and high integration density. However, several drawbacks of STT-MRAM (such as shared read and write current paths, single-ended sensing scheme, and high write power consumption) need to be overcome to make it suitable for on-chip cache applications. However, our proposed spin-based device utilizes a domain wall motion layer and a complementary polarizer structure to achieve energy efficiency, high performance and high disturb, sensing and write margins. The design requirements of DWCSTT bit-cell is significantly relaxed compared to 1T-1MTJ STT-MRAM because the read and write current paths are decoupled. The use of a low resistance write path allows the proposed DWCSTT bit-cell to mitigate source degeneration of the write access transistor, which also reduces write energy consumption as compared to 1T-1MTJ STT-MRAM. Furthermore, the complementary polarizer structure in the read path of the DWCSTT device allows DWCSTT to achieve higher read margin and lower read power than 1T- 1MTJ STT-MRAM. Compared to the conventional 1T-1MTJ STT-MRAM bit-cell, the proposed DWCSTT bit-cell achieves low write power consumption under iso-area and iso-write margin condition, and better sensing margin with low read power consumption, and higher read disturb margin.

In Chapter 3, we showed dual 1R/1W port spin-orbit torque based MRAM for onchip cache application. High write latency and write energy requirements need to be overcome for STT-MRAM to be suitable for on-chip memories. One way to address the issue of high-write latency is to have separate read and write ports (1R/1W) to each bitcell of STT-MRAM. However, an additional transistor is required in order to separate the read and the write ports of 1R/1W STT-MRAM. Moreover, high write energy still remains an issue for on-chip memory application. However, our proposed memory utilizes multiple ports for simultaneous accesses of cache memory, alleviating memory access conflict, while trying to hide the high write latency of STT-MRAM. The dualported memory can be implemented without any additional area overhead compared to the single-ported spin-orbit memory. In addition, high spin injection efficiency of the spin-orbit devices in our assumed device dimension and spin Hall angle enables low power write operation. Separate optimization for read and write can be performed in SOT-MRAM leading to enhancing the read operation. Compared to the standard STT-MRAM bit-cell, the 1R/1W SOT-MRAM bit-cell can achieve lower power consumption and higher IPC.

Chapter 4 presents a new type of SOT-MRAM for high-density, reliable and energyefficient on-chip memory application. Unlike conventional SOT-MRAM which requires two access transistors, the proposed MRAM uses only one access transistor along with a Schottky diode to achieve high integration density while maintaining the advantages of SOT-MRAM, such as low write energy and improved reliability of MTJ. The Schottky diode is forward biased during a read operation, whereas it is reverse biased during a write operation to prevent sneak current paths. Therefore, the proposed MRAM can achieve lower bit-cell area in comparison to conventional STT-MRAM and SOT-MRAM, and improvement in write power compared to STT-MRAM. Chapter 5 proposes a new design technique of SOT-MRAM for high density and low power on-chip cache applications. A bit-line of the proposed memory bit-cell is shared with that of an adjacent bit-cell, hence, minimum allowable area of our proposed structure can be improved in comparison with conventional structure of SOT-MRAM by reducing the number of metals along the column direction. Furthermore, since the efficient spin orbit torque based switching operation can translate to smaller size of access transistors, the proposed SOT-MRAM achieves higher integration density compared to STT-MRAM while maintaining the advantages of conventional SOT-MRAM such as low write energy dissipation, high read-disturb margin, and enhanced reliability of MTJ. Our proposed SOT-MRAM bit-cell can achieve higher integration density than conventional SOT-MRAM bit-cells. In addition, the proposed MRAM improves write power and read-disturb margin compared to STT-MRAM.

In Chapter 6, we propose a new NVFF using CPMTJ to alleviate limitations of STT-NVFF. Our proposed NVFF exploits the CPMTJ structure for fast and low-energy backup operation. The estimated backup delay of CP-NVFF is lesser than that of conventional STT-NVFF. Furthermore, during the restore operation, CPMTJ provides guaranteed disturb-free sensing since disturb torque in CPMTJ comes from two pinned layer and is cancelled each other. LIST OF REFERENCES

LIST OF REFERENCES

- [1] K. Roy, S. Mukhopadhyay and H. Mahmoodi-Meimand, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits", in *Proceedings of IEEE*, New York, vol. 91, no. 2, pp. 305-327, February, 2003.
- [2] Semiconductor Engineering Magazine (2014) [Online] http://semiengineering.com/as-nodes-advance-so-must-power-analysis/
- [3] ITRS (2011) [Online] <u>http://www.itrs.net/links/2011itrs/home2011.htm.</u>
- [4] Kwon K-W. (2015). Nonvolatile Cache and Flip-Flop Design for Low Standby Leakage SoC (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses. (Accession No. 3736250)
- [5] S. Shigematsu, S. Mutoh, Y. Matsuya, and J. Yamada, "A 1-V High-Speed MTCMOS Circuit Scheme for Power-Down Application", in *IEEE Symp. VLSI Circuits*, pp. 12-126. Jun. 1995
- [6] C. H. Kim, and K. Roy, "Dynamic VTH Scaling Scheme for Active Leakage Power Reduction", in *Proc. Design, Automation & Test in Europe Conf.*, pp. 163-167, Mar. 2002
- [7] M. C. Johnson, D. Somasekhar, L-Y. Chiou, and K. Roy, "Leakage Control with Efficient Use of Transistor Stacks in Single Threshold CMOS", in *IEEE Trans. VLSI Systems*, vol. 10, no. 1, pp. 1-5, Feb, 2002.
- [8] T. Kuroda *et al.*, "A 0.9-V, 150-MHz, 10-mW, 4 mm², 2-D Discrete Cosine Transform Core Processor with Variable Threshold-Voltage (V_T) Scheme", in *IEEE J. Solid State Circuits*, vol. 31, no. 11 pp. 1770-1779, Nov. 2002.
- [9] L. T. Clark, M. Morrow, and W. Brown, "Reverse-Body Bias and Supply Collapse for Low Effective Standby Power", in *IEEE Trans. VLSI Systems*, vol. 12, no. 9, pp. 947-956, Sep. 2004.
- [10]K. Lee, and S. H. Kang, "Development of Embedded STT-MRAM for Mobile System-on-Chips", in *IEEE Trans. Magnetics*, vol. 47, no. 1, pp. 131-136, Dec 2010.

- [11]K. C. Chun, H. Zhao, J. D. Harms, T. Kim, J. Wang and C. H. Kim, "A Scaling Roadmap and Performance Evaluation in In-Plane and Perpendicular MTJ Based STT-MRAMs for High-Density Cache Memory", in *IEEE J. Solid State Circuits*, vol. 48, no. 2, pp. 598-610, Feb. 2013.
- [12]X. Fong, Y. Kim, S. H. Choday, and K. Roy, "Failure Mitigation Techniques for 1T-1MTJ Spin-Transfer Torque MRAM Bit-cells", in *IEEE Trans. VLSI Systems*, vol. 22, no. 2, pp. 384-395, Jul. 2013.
- [13] Y. Seo, X. Fong, and K. Roy, "Domain Wall Coupling-Based STT-MRAM for On-Chip Cache Applications" in *IEEE Transactions on Electron Devices*, vol.62, no. 2, pp. 554 – 560, Feb. 2015.
- [14] Y. Seo, X. Fong, K-W. Kwon, and K. Roy, "Spin-Hall Magnetic Random-Access Memory With Dual Read/Write Ports for On-Chip Caches", in *IEEE Magnetics Letters*, vol. 6, pp. 3000204, Apr. 2015.
- [15] Y. Seo, K-W. Kwon, X. Fong, and K. Roy, "High Performance and Energy-Efficient On-Chip Cache Using Dual Port (1R/1W) Spin-Orbit Torque MRAM", in *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, pp. 1-12, Apr. 2016.
- [16]C-F. Pai, L. Liu, Y. Li, H. W. Tseng, D. C. Ralph, and R. A. Buhrman, "Spin Transfer Torque Devices utilizing the Giant Spin Hall Effect of Tungsten," in *Applied Physics Lett.*, vol. 101, no. 12, pp. 122404, Sep. 2012.
- [17] J. Kim, W. Tuohy, C. Ma, W. Choi, I. Ahmed, D. Lilja, and C.H. Kim, "Spin-Hall Effect MRAM Based Cache Memory: A Feasibility Study", in *Device Research Conf.*, pp. 117-118, Jun. 2015.
- [18] Y. Seo, K-W. Kwon, and K. Roy, "Area-Efficient SOT-MRAM with a Schottky Diode", in *IEEE Electron Device Letters*, Jun. 2016.
- [19]S. Yamamoto and S. Sugahara, "Nonvolatile Delay Flip-Flop based on Spintransistor Architecture", in *Jpn. J. Applied Physics*, vol. 49, no. 9, pp. 090204, Sep. 2010.
- [20]X. Fong and K. Roy, "Complementary Polarizers STT-MRAM (CPSTT) for On-Chip Caches", in *IEEE Electron Device Lett.*, vol. 34, no. 2, pp. 232-234, Jan. 2013.

- [21]Y. Kim, S. H. Choday, and K. Roy, "DSH-MRAM: Differential Spin Hall MRAM for On-Chip Memories," in *IEEE Electron Device Lett.* vol. 34, no. 10, pp. 1259-1261, Oct. 2013.
- [22] W. S. Zhao, Y. Zhang, T. Devolder, J. Q. Klein, D. Ravelosona, C. Chappert, and P. Mazoyer, "Failure and Reliability Analysis of STT-MRAM", in *Microelectronic Reliability*, vol. 52, no. 9-10, pp. 1710-1723, Oct. 2012.
- [23]G. Jeong, W. Cho, S.Ahn, H. Jeong, G. Koh, Y. Hwang, and K. Kim, "A 0.24-μm 2.0-V 1T-1MTJ 16-kb Nonvolatile Magnetoresistance RAM with Self-Reference Sensing Scheme", in *IEEE J. Solid State Circuits*, vol. 38, no. 11, pp. 1906-1910, Nov. 2003.
- [24] Y. Chen, H. Li, X. Wang, W. Zhu, W. Xu, and T. Zhang, "A Nondestructive Self-Reference Scheme for Spin-Transfer Torque Random Access Memory (STT-RAM)", in *Proc. Design, Automation & Test in Europe Conf.*, pp. 148-153, Mar. 2010.
- [25]X. Fong and K. Roy, "Robust Low-power Multi-terminal STT-MRAM", in *Proc. Non-volatile Memory Technology Symp.*, Aug. 2013.
- [26] D. Lee, S. K. Gupta, and K. Roy, "High-Performance Low-Energy STT-MRAM based on Balanced Write Scheme", in Proc. ACM/IEEE Int. Symp. on Low Power Electronics and Design, pp. 9-14, Jul. 2012.
- [27]D. Bromberg, D. Morris, L. Pileggi and J. Zhu, "Novel STT-MTJ Device Enabling All-Metallic Logic Circuits", in *IEEE Trans. Magnetics*, vol. 48, no.117, pp. 3215-3218, Nov. 2012.
- [28]S. Fukami, N. Ishiwata, N. Kasai, M. Yamanouchi, H. Sato, S. Ikeda, and H. Ohno, "Scalability Prospect of Three-Terminal Magnetic Domain- Wall Motion Device", in *IEEE Trans. Magnetics*, vol. 48, no. 7, pp. 2152-2157, Jul. 2012.
- [29]D. Lacour, J. A. Katine, L. Folks, T. Block, J. R. Childress, M. J. Carey, and B. A Gurney, "Experimental Evidence of Multiple Stable Locations for a Domain Wall Trapped by a Submicron Notch", in *Applied Physics Lett.*, vol. 84, no11, pp. 1910-1912, Mar. 2004.
- [30]Z. Li, and S. Zhang, "Domain-wall dynamics and spin-wave excitations with spintransfer torques", in *Phys. Rev. Lett.*, vol. 92, pp. 207203, May. 2004.

[31]OOMMF. [Online]. Available: <u>http://math.nist.gov/oommf</u>

- [32]K. Ueda *et al.*, "Temperature Dependence of Carrier Spin Polarization Determined from Current-Induced Domain Wall Motion in Co / Ni Nanowire", in *Applied Physics Lett.*, vol. 100, pp. 202407-1-202407-3, May. 2012.
- [33] V. Sokalski, D. M. Bromberg, D. Morris, M. T. Moneck, E. Yang, L. Pileggi, and J. Zhu, "Naturally Oxidized FeCo as a Magnetic Coupling Layer for Electrically Isolated Read/Write Paths in mLogic", in *IEEE Trans. Magnetics*, vol. 49, no. 7, pp. 4351-4354, Jul. 2013.
- [34]X. Fong, S. K. Gupta, N. N. Mojumder, S. H. Choday, C. Augustin, and K. Roy, "KNACK: A Hybrid Spin-Charge Mixed-Mode Simulator for Evaluating Different Genres of Spin Transfer Torque MRAM Bit-cells", in *Proc. Int. Conf. Simul. Semicond. Process. Devices*, pp. 51-54, Sep. 2011.
- [35]T. Kishi *et al.*, "Low-current and fast switching of a perpendicular TMR for high speed and high density spin-transfer-torque MRAM", in *Proc. IEEE Int. Electron Dev. Meeting*, pp. 12.6.1-12.6.4, Dec. 2008.
- [36]HSPICE. [Online], Available : http://www.synopsys.com/tools/verification/amsverification/circuitsimulation/hspice/
- [37]C. Mead and L. Conway, Introduction to VLSI Systems. Reading, MA: Addison-Wesley, 1980.
- [38]S. K. Gupta, S. P. Park, N. N. Mojumder and K. Roy, "Layout-aware Optimization of STT-MRAMs", in *Proc. Design, Automation & Test in Europe Conf.*, pp. 1455-1458, Mar. 2012.
- [39]X. Fong, R. Venkatesan, A. Raghunathan, and K. Roy, "Non-Volatile Complementary Polarizer Spin-Transfer Torque On-Chip Caches: A Device/Circuit/Systems Perspective", in *IEEE Trans. Magnetics*, vol. 50, no. 10, pp.3400611, Oct. 2014.
- [40] C. Augustine, N Mojumder, X. Fong, H. Choday, S. P. Park, and K. Roy, "STT-MRAMs for Future Universal Memories: Perspective and Prospective," in *Proc. Int. Conf. on Microelectronics*, pp. 13-16, May. 2012.

- [41]S. Natarajan, S. Chung, L. Paris, and A. Keshavarzi, "Searching for the Dream Embedded Memory," in *IEEE Solid-State Circuits Magazine*, vol. 1, no. 3, pp. 33-44, Aug. 2009.
- [42] D. Kaseridis, M. F. Iqbal, and L. K. John, "Cache Friendliness-Aware Management of Shared Last-Level Caches for High Performance Multi-Core Systems," in *IEEE Trans. Computers*, vol. 63, no. 4, pp. 874-887, Apr. 2014.
- [43]S. Matsunaga, J. Hayakawa, S. Ikeda, K. Miura, T. Endoh, H. Ohno, and T. Hanya, "MTJ-based Nonvolatile Logic-in-Memory Circuit, Future Prospects and Issues," in *Proc. Design, Automation & Test in Europe Conf.*, pp. 433-435, Apr. 2009.
- [44] J. Li, P. Ndai, A. Goel, S. Salahuddin, and K. Roy, "Design Paradigm for Robust Spin-Torque Transfer RAM (STT-MRAM) From Circuit/Architecture Perspective," in *IEEE Trans. VLSI Systems*, vol. 18, no. 12, pp. 1710-1723, Dec. 2010.
- [45]X. Wang, W. Zhu, M. Siegert, and D. Dimitrov, "Spin Torque Induced Magnetization Switching Variations," in *IEEE Trans. Magnetics*, vol. 45, no. 4, pp. 2038-2041, Apr. 2009.
- [46] Y-T Cui, G. Finocchio, C. Wang, J. A. Katine, R. A Buhrman, and D. C. Ralph, "Single-Shot Time-Domain Studies of Spin-Torque-Driven Switching in Magnetic Tunnel Junctions," in *Physical Review Lett.*, vol. 104, no. 9, pp. 097201, Mar. 2010.
- [47]K. Itoh, "Embedded Memories: Progress and a Look into the Future," in *IEEE Design & Test of Computers*, vol. 28, no. 1, pp. 10-13, Jan. 2011.
- [48]G. Sun, X. Dong, Y. Xie, J. Li, and Y. Chen, "A Novel Architecture of the 3D Stacked MRAM L2 Cache for CMPs," in *Proc. IEEE Int. Symp. on High Performance Computer Architecture*, pp. 239-249, Feb. 2009.
- [49]K-W. Kwon, S. H. Choday, Y. Kim, and K. Roy, "AWARE (Asymmetric Write Architecture With Redundant Blocks): A High Write Speed STT-MRAM Cache Architecture," in *IEEE Trans. VLSI Systems*, vol. 22, no. 4, pp. 712-720, Apr. 2014.
- [50]X. Bi, M. A. Weldon, and H. Li, "STT-RAM Designs Supporting Dual-port Accesses", in *Proc. Design, Automation & Test in Europe Conf.*, pp. 1530-1591, Mar. 2013.

- [51]T. Pompl, C. Schlunder, M. Hommel, H. Nielen, and J. Schneider, "Practical Aspects of Reliability Analysis for IC Designs," in *Proc. ACM/IEEE Design Automation Conf.*, pp.193-198, May. 2006.
- [52]C. J. Lin, et al., "45nm Low Power CMOS Logic Compatible Embedded STT MRAM Utilizing a Reverse-Connection 1T/1MTJ Cell," in Proc. IEEE Int. Electron Dev. Meeting, pp. 11.6.1-11.6.4, Dec. 2009.
- [53]L. Liu, C-F. Pai, Y. Li, H. W. Tseng, D. C Ralph, and R. A. Buhrman, "Spin-Torque Switching with the Giant Spin Hall Effect of Tantalum," Science, vol. 336, no. 6081, pp. 555-558, May 2012.
- [54] T. Jungwirth, J. Wunderlich, and K. Olejnik, "Spin Hall effect devices," in *Nature Materials*, vol. 11, no. 5, pp. 382-390, Apr. 2012.
- [55] P. Gambardella and I. M. Miron, "Current-induced spin-orbit torques," in *Phil. Trans. R. Soc. A*, vol. 369, pp. 3175-3179, Aug 2011
- [56]I. M. Miron, K. Garello, G. Gaudin, *et al.*, "Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection," in *Nature*, vol. 476, pp. 189–193, Aug. 2011.
- [57] A. Hoffmann, "Spin Hall effects in metals," in *IEEE Trans. Magnetics*, vol. 49, no. 10, pp. 5172-5193, Oct. 2013.
- [58]Y. Kim, X. Fong, K-W. Kwon, M-C. Chen, and K. Roy, "Multilevel Spin-Orbit Torque MRAMs," in *IEEE Trans. Electron Devices*, vol. 62, no. 2, pp. 561-568, Feb. 2015.
- [59]H. Ulrichs, V. E. Demidov, S. O. Demokritov, W. L. Lim, J. Melander, N. Ebrahim-Zadeh, and S. Urazhdin, "Optimization of Pt-based spin-Hall-effect spintronic devices," in *Applied Physics Lett.*, vol. 102, no. 13, pp. 132402, Apr. 2013.
- [60] J. C. Slonczewski, "Current-driven excitation of magnetic multilayers," in J. Magnetism and Magnetic Materials, vol. 159, no. 1-2, pp. L1-L7, Jun. 1996.
- [61] J. Xiao, A. Zangwill, and M. D. Stiles, "Macrospin models of spin transfer dynamics," in *Physical Review B*, vol. 72, no. 1, pp. 014446, Jul. 2005.

- [62]S. Datta, S. Salahuddin, and B. Behin-Aein, "Non-volatile Spin Switch for Boolean and non-Boolean logic", in *Applied Physics Lett.*, vol. 101, no. 25, pp. 252411, Dec. 2012.
- [63] J. K. Han, J. H. NamKoong, and S. H. Lim, "A new analytical/numerical combined method for the calculation of the magnetic energy barrier in a nanostructured synthetic antiferromagnet," in *J. Physics D: Applied Physics*, vol. 41, no. 23, pp. 232005, Nov. 2008.
- [64] Y. Shang, W. Fei, and H. Yu, "Analysis and Modeling of Internal State Variables for Dynamic Effects of Nonvolatile Memory Devices," in *IEEE Trans. Circuits and Systems I: Regular Papers*, vol. 59, no. 9, pp. 1906-1918, Sep. 2012.
- [65]B. Zhao, J. Yang, Y. Zhang, Y. Chen, and H. Li, "Architecting a Common-Source-Line Array for Bipolar Non-Volatile Memory Devices," in *Proc. Design, Automation & Test in Europe Conf.*, pp. 1451-1454, Mar. 2012.
- [66] S. P. Park, S. Gupta, N. Mojumder, A. Raghunathan, and K. Roy, "Future Cache Design using STT MRAMs for Improved Energy Efficiency: Devices, Circuits and Architecture," in *Proc. ACM/EDAC/IEEE Design Automation Conf.*, pp.492-497, Jun. 2012.
- [67] N. Muralimanohar, R. Balasubramomian, and N. Joippi, "Optimizing NUCA Organizations and Wiring Alternatives for Large Caches with CACTI 6.0," in *Proc. IEEE/ACM Int. Symp. on Microarchitecture*, pp. 3-14, Dec. 2009.
- [68]D. C. Worledge *et al.*, "Switching distributions and write reliability of perpendicular spin torque MRAM," in *Proc. IEEE Int. Electron Dev. Meeting*, pp. 12.5.1-12.5.4, Dec. 2010.
- [69] T. Austin, E. Larson, and D. Ernst, "SimpleScalar: An infrastructure for computer system modeling," in *IEEE Trans. Computers*, vol. 35, no. 2, pp. 59-67, Feb. 2002.
- [70] A. Gordon-Ross, F. Vahid, and N. D. Dutt, "Fast Configurable-Cache Tuning With a Unified L2 Cache" in *IEEE Trans. VLSI Systems*, vol. 17, no. 1, pp. 80-91, Jan. 2009.
- [71] S. W. Keckler, W. J. Dally, B. Khailany, M. Garland, and D. Glasco, "GPUs and the Future of Parallel Computing," in *IEEE Micro*, vol. 31, no. 5, pp. 7-17, Oct. 2011.

- [72] T. Vogelsang, "Understanding the Energy Consumption of Dynamic Random Access Memories," in *Proc. IEEE/ACM Int. Symp. on Microarchitecture*, pp. 363-374, Dec. 2010.
- [73]K.-W. Kwon, X. Fong, P. Wijesinghe, P. Panda, and K. Roy, "High-Density and Robust STT-MRAM Array Through Device/ Circuit/Architecture Interactions", *IEEE Trans. Nanotechnology*, vol. 14, no. 6, pp. 1024-1034, Jul. 2015.
- [74]T. Gosavi, S. Manipatruni, D. Nikonov, I. A. Young, and S. Bhave (Jul, 2015). "Experimental Demonstration of Efficient Spin-Orbit Torque Switching of an MTJ with sub-100 ns Pulses,"
- [75] A. Makarov, T. Windbacher, V. Sverdlov, and S. Selberherr, "SOT-MRAM based on 1Transistor-1MTJ-cell structure," *Proc. Non-Volatile Memory Tech. Symp.*, pp. 1-4, Oct. 2015.
- [76]K.-W. Kwon, S. H. Choday, Y. Kim, X. Fong, S. P. Park, and K. Roy, "SHE-NVFF: Spin Hall effect-based nonvolatile flip-flop for power gating architecture" in *IEEE Electron Device Letter*, vol. 35, no. 4, pp. 488-490, Apr. 2014.
- [77] N. N. Mojumder, S. K. Gupta, S. H. Choday, D. E. Nikonov, and K. Roy, "A three terminal dual pillar STT-MRAM for high-performance robust memory applications," *IEEE Trans. Electron. Devices*, vol. 58, no. 5, pp. 1508–1516, May 2011.
- [78]X. Wang, Metallic Spintronic Devices. Reading, MA: CRC Press, 2014. ISBN: 978-1-4665-8844-8
- [79] Y. T. Li, S. B. Long, H. B. Lv, Q. Liu, M. Wang, H. W. Xie, K. W. Zhang, X. Y. Wang, and M. Liu, "Novel Self-compliance Bipolar 1D1R Memory Device for High-density RRAM Application." *IEEE Int. Memory Workshop*, pp. 184-187, May. 2013.
- [80] J. Maiz, S. Hareland, K. Zhang, and P. Armstrong, "Characterization of multi-bit soft error events in advanced SRAMs" *IEEE Electron Devices Meeting*, pp. 21.4.1-21.4.4, Dec. 2003.
- [81] M. Keating, et al., Low Power Methodology Manual for System-on-Chip Design. Reading, MA: Springer, 2007.
- [82] N. Sakimura, T. Sugibayashi, R. Nebashi, and N. Kasai, "Nonvolatile Magnetic Flip-Flop for Standby-Power-Free SoCs," *IEEE J. Solid State Circuits*, vol. 44, no. 8, pp. 2244-2250, Aug. 2009.

- [83] J. P. Kim, et al., "45nm 1Mb Embedded STT-MRAM with design techniques to minimize read-disturbance," Proc. IEEE Symp. on VLSI Circuits, pp. 296-297, Jun. 2011.
- [84]S. Sinha, et al., "Exploring Sub-20nm FinFET Design with Predictive Technology Models," Proc. ACM/EDAC/IEEE Design, Automation Conf., pp. 283-288, Jun. 2012.
- [85]S. A. Tawfik and V. Kursun, "Low-Power and Compact Sequential Circuits with Independent-Gate FinFETs," *IEEE Trans. Electron Devices*, vol. 55, no.1, pp. 60-70, Jan. 2008.
- [86]K. G. Anil, et al., "Layout Density Analysis of FinFETs," in Proc. European Solid-State Device Research Conf., pp. 139-142, Sep. 2003.
- [87]S. Chatterjee, S. Salahuddin, S. Kumar, and S. Mukhopadhyay, "Impact of Self-Heating on Reliability of a Spin-Torque-Transfer RAM Cell," *IEEE Trans. Electron Devices*, vol. 59, no. 3, pp. 791-799, Mar, 2012.

VITA

VITA

Yeongkyo Seo received B.E. degree in Electrical Engineering from Korea University, Seoul, Korea in Aug 2011. He is currently pursuing Ph.D degree in Electrical and Computer Engineering at Purdue University, West Lafayette, IN, USA. His research interests include device and circuit co-design for energy-efficient systems.