Emerging Technologies for Beyond Conventional Compute Systems

by

Abdelrahman G. Qoutb

Submitted in Partial Fulfillment

of the

Requirements for the Degree
Doctor of Philosophy

Supervised by
Professor Eby G. Friedman

Department of Electrical and Computer Engineering
Arts, Sciences and Engineering
Edmund A. Hajim School of Engineering and Applied Sciences

University of Rochester
Rochester, New York
2022

Dedication

To my wife Lamiaa,

تعاهدنا على السير معًا

To my colleagues and fellow students at Fayoum University, Egypt.

Table of Contents

Biographical Sketch

Abdelrahman G. Qoutb received the B.Sc. degree in electronics and communications engineering with distinction and honors from Fayoum University, Egypt, in 2014. During his study, he was a recipient of the outstanding student academic award for four consecutive years (2011 to 2014) from Fayoum University. As the top of his class, he was appointed as a teacher assistant in the Department of Electrical Engineering in Fayoum University on April 2015. He received his Master's degree in Electrical and Computer Engineering from the University of Rochester, Rochester, NY USA in 2018. He is a Ph.D. candidate in the area of high performance VLSI/IC design at the University of Rochester, Rochester, NY USA. In the summers of 2019 and 2020, he was an intern with the device development group at Intel in Hillsboro, Oregon USA.

His current research interests include the development of nonconventional compute systems based on emerging nanoscale electronic technologies, physical design of integrated circuits, technology development, failure analysis, and design for testability.

The following publications are a result of work conducted during his doctoral study.

Patent disclosures

1. **Abdelrahman G. Qoutb** and E. G. Friedman, "Distributed Spintronic/CMOS Sensor Network for Thermal Aware Systems," U.S. Patent No. 11,378,466 B2, July 5, 2022.

2. **Abdelrahman G. Qoutb** and E. G. Friedman, "Switching of Perpendicularly Magnetized Nanomagnets With Spin-Orbit Torques in the Absence of External Magnetic Fields," U.S. Patent Application No. 16/850,173, April 2020.

Journal papers

1. **Abdelrahman G. Qoutb**, Stephen Whiteley, Jamil Kawa, and Eby G. Friedman, "Josephson Junction Stuck at Faults Detection in SFQ Circuits," *IEEE Transactions on Applied Superconductivity* (in review).

2. **Abdelrahman G. Qoutb**, Jamil Kawa, and Eby G. Friedman, "Test Modules for Enhanced Testability of Single Flux Quantum Integrated Circuits," *IEEE Transactions on Applied Superconductivity* (in review).

3. Guangchao Zhao, Zhiwei Zeng, Xingli Wang, **Abdelrahman G. Qoutb**, Philippe Coquet, and Eby G. Friedman,"Efficient Ternary Logic Circuits Optimized by Ternary Arithmetic Algorithms,", *IEEE Transactions on Circuits and Systems I* (in review).

4. **Abdelrahman G. Qoutb** and E. G. Friedman, "Double Magnetic Tunnel Junction Two Bit Memory and Nonvolatile Logic for *in situ* Computing," *Microelectronics Journal* (in press).

5. **Abdelrahman G. Qoutb** and E. G. Friedman, "Distributed Spintronic/CMOS Sensor Network for Thermal Aware Systems," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 28, No. 6, pp. 1505–1512, June 2020.

Conference papers

1. **Abdelrahman G. Qoutb** and Eby G. Friedman, "Double Magnetic Tunnel Junction- Based Nonvolatile Logic," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 311 - 315, May 2022.

2. **Abdelrahman G. Qoutb** and Eby G. Friedman, "Spintronic/CMOS-Based Thermal Sensors," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 1–5, October 2020.

3. **Abdelrahman G. Qoutb** and Eby G. Friedman, "PMTJ Temperature Sensor Utilizing VCMA," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 1–5, May 2019.

4. **Abdelrahman G. Qoutb** and Eby G. Friedman, "MTJ Magnetization Switching Mechanisms for IoT Applications," *Proceedings of the ACM Great Lakes Symposium on VLSI*, pp. 347–352, April 2018.

Acknowledgments

الْحَمْدُ لِلَّهِ الَّذِي هَدَانَا لِهَذا وَمَا كُنَّا لِنَهْتَدِيَ لَوْلَا أَنْ هَدَانَا اللَّهُ

All praise and gratitude to Allah who guided us to this and We would have never been guided if Allah had not guided us.

أَنِ ٱشْكُرْ لِى وَلِوَلِدَيْكَ

Thank Me and your parents.

All gratitude are due to Allah, God the Almighty, who created me, who guides me, who provides me with sustenance and whenever I fall ill, it is He who heals me.

As instructed by my tradition and as life has taught me, parents are always a resource for counseling and help. I cannot thank them enough for what they did for me since my birth. May Allah shower them in His mercy on the day of judgment.

PhD experience varies based on a lot of parameters and factors, but it always goes through multiple joyful and sorrowful moments. My Ph.D. experience was not just an experience of academic development, but it was a transformative experience that carefully shaped my personality and my vision to the future. The main contributor to

this experience was Prof. Eby Gershon Friedman, my academic advisor. His unique personality, wisdom, professionalism, commitment, and experience are some of his attributes that make him one of the best supervisors. Prof. Friedman's style of supervision does not include only the technical aspect of a Ph.D. but carefully direct, teach, lead by example, and he does this with each student, in a different way, based on the character of each student. I have never seen such talented skill. I look forward to a lifetime relationship with him and I will always remember his help and support. Thank you!

I would like to thank the members of my committee, Prof. Mark Bocko, Prof. Selçuk Köse, Prof. John Lambropoulos, Prof. Yonathan Shapir, and Jamil Kawa, for their valuable comments and guidance. I highly appreciate the time you invested to enhance my research and dissertation. I would also like to thank Prof. Gary Wicks for being the chairperson of my Ph.D. defense committee.

My thanks to my colleagues, past and current members of the High Performance VLSI/IC Design & Analysis Laboratory: Dr. Albert Ciprut, Dr. Kan Xu, Dr. Gleb Krylov, Dr. Rassul Bairamkulov, Tahereh Jabbari, Nurzhan Zhuldassov, Ana Mitrovic, and Andres Ayes - I enjoyed the conversations that we normally have at the lab. I was always motivated by your progress in your research.

Special appreciation goes to RuthAnn Williams for her support with the administrative tasks and her contagious cheerfulness that she always brings to the lab. I am

grateful to our Graduate Coordinator, Michele Foster, who has always been happy to help and offer advice.

My Ph.D. experience would have never existed, except for Prof. Magdy Ali El-Moursy who supervised my senior project although he was affiliated with a different organization than my academic institute. He is always supportive and helpful. It was he who introduced me to Prof. Friedman to enter the the Ph.D. graduate program. Thank you so much Prof. Magdy for being available and willingness to guide me through the years.

Throughout my life I am fortunate to have great mentors who are happy to listen, talk, support, and direct me to achieve the best. Some of these mentors are Prof. Mohammed Tharwat Hassan, Prof. Hossam A. H. Fahmy, Prof. Mohamed Wiem Mkaouer, Dr. Mahmoud Elassal, and all of my mentors at EgyptScholars Inc., especially Dr. Khaled AlAshmouny and Dr. Ahmed Naga.

Finally, I would like to thank the light of my life, my wife. All the words of gratitude and love are not enough to describe how much I feel towards you, Lamiaa. Thank you for bearing with me through hard years. Your support was crucial for me throughout my Ph.D. program. To my daughters, I love you ♡.

Abstract

The application of beyond CMOS technologies in innovative materials, memory, logic, and architectures will likely exhibit novel compute schemes and systems. This functional diversification, supported by beyond CMOS technologies, is expected to unleash a wide spectrum of novel solutions that have not been previously possible, such as reconfigurable nonvolatile logic and on-chip integrated sensor networks.

The primary objective of this dissertation is to bridge the gap among novel emerging devices, unconventional architectures, and computing schemes to support or replace conventional CMOS technologies to achieve next generation applications. In this dissertation, representative circuit and architectural advances are proposed that exploit the unique characteristics of emerging, beyond CMOS devices. The unique characteristics of these proposed systems include thermal sensitivity, non-volatility, extreme low power, and reconfigurability.

An MTJ is treated in this dissertation as an illustrative example of an emerging technology that can support beyond CMOS systems. MTJs are commonly used within

commercial systems as an embedded memory. Importantly, MTJs are compatible with CMOS fabrication processes. In this dissertation, MTJs are proposed as a solution for several different compute schemes, including self-aware computers, compute in-memory, reconfigurable logic, and distributed compute systems.

In addition to these emerging technologies, superconductive electronics is considered as a standalone replacement for conventional CMOS systems. In superconductive electronics, one important logic family is based on single flux quanta (SFQ) to encode and process data.

Although beyond CMOS devices exhibit a wide range of functions that can replace or support conventional CMOS systems, superconductive devices also exhibit reliability issues that should be identified and addressed early in the technology development process. Since these devices suffer from low yield, advanced testability methodologies that target the unusual characteristics of the technology are required. Two different approaches are described in this dissertation to enhance the testability of SFQ systems. In the first approach, circuit solutions to enhance the controllability and observability of the internal nodes within SFQ systems are presented. In the second approach, high level fault models are proposed to characterize SFQ systems and generate the required test vectors to locate and identify Josephson junction-based faults.

Contributors and Funding Sources

This work was supervised by a dissertation committee consisting of Professor Eby G. Friedman (advisor), Professor Mark Bocko, and Professor Selcuk Kose of the Department of Electrical and Computer Engineering, Professor Yonathan Shapir of the Department of Physics and Astronomy, Professor John Lambropoulos of the Department of Mechanical Engineering, Professor Gary Wicks of the Institute of Optics (committee chairman), Professor Magdy El-Moursy of Siemens EDA and Jamil Kawa of Synopsys Inc. All of the work described in the dissertation was completed independently by the student.

This research is supported in part by the National Science Foundation under Grant Nos. CCF-1716091 and 2124453, Intelligence Advanced Research Projects Activity (IARPA) under Grant No. W911NF-17-9-0001, and by grants from Cisco Systems, Qualcomm, Synopsys, and Google.

List of Tables

List of Figures

Chapter 1

Introduction

The gigascale integration era, where many billions of transistors are integrated on-chip, is marching towards terascale integration with deeply scaled device and packaging technologies. Serious challenges exist to further enhance the performance of these integrated circuits. This evolutionary development is extending integrated circuits to beyond end-of-CMOS heterogeneous applications. Novel design methodologies are under development to enhance the performance of integrated circuits across multiple abstraction levels and functions (material, digital logic, memory, and system architecture). These objectives can be achieved by addressing two different technology development paths.

The first path is the classical development of CMOS technologies by geometric scaling of transistors through the introduction of performance boosters [1] such as strain engineering, high-k gate dielectrics, metal gate, and FinFET structures, as

illustrated in Figure 1.1. In this approach, CMOS is a standalone technology where all memory and information processing devices are based on CMOS.

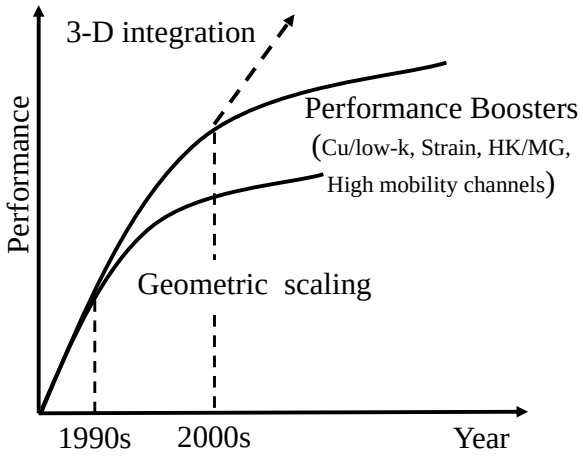


Figure 1.1: Physical and architectural paths to enhance the performance of CMOS technologies [2].

The second path is described as *beyond CMOS* where novel memory or information processing technologies incorporate emerging non-CMOS technologies. These beyond CMOS technologies can be classified as standalone integration or heterogeneous integration based on how these emerging electronic devices are combined. This beyond CMOS approach is extending microelectronics to innovative functions and applications. The primary objective of this dissertation is to bridge the gap among novel emerging devices, unconventional architectures, and computing schemes to support or replace conventional CMOS technologies to achieve next generation applications.

1.1 Assessments and benchmarking of beyond CMOS technologies

The vast majority of beyond CMOS technologies is based on innovative materials and/or different device concepts, including tunneling transistors [3], resistive memory devices [4], spintronic/magnetoelectric devices [5], and ferroelectric transistors [6]. Some of these beyond semiconductor solutions are intended to be integrated onto a silicon platform to exploit established CMOS-based infrastructures. Other technologies such as superconductive electronics are considered a promising standalone technology. Understanding the potential and limitations of beyond CMOS technologies is vital to achieve different and important applications. Multiple assessments and comprehensive benchmarking efforts are needed to determine the nature of these emerging devices to either support or replace CMOS and to quantify the performance of these emerging technologies within different computing schemes.

The process of benchmarking emerging technologies is challenging since the effort depends upon the level of abstraction and domain of functionality [7–11]. A uniform methodology for benchmarking beyond CMOS logic devices, developed at Intel [7, 10], builds Boolean benchmark circuits based on emerging beyond CMOS technologies. These benchmark circuits include sequential logic and arithmetic logic units (ALU).

The performance metrics of this study are based on estimating the standby and active power, power density, switching speed, and throughput of these circuit technologies.

In this Intel effort, a consistent, transparent, and physics-based methodology is used to characterize and compare 25 emerging logic devices, including spintronic [5], tunneling [3], ferroelectric [6], and piezoelectric devices [7]. Each of these emerging devices utilizes different switching mechanisms, such as electronic switching (charging a gate capacitor with current), ferroelectric switching (electric polarization), and voltage and current driven magnetization switching. Physics-based compact models are used to estimate the performance of the benchmark circuits. This study suggests that non-volatile spintronic logic devices exhibit design simplicity and size advantages for register and state elements while dissipating low standby power.

A recent effort to extend this beyond CMOS benchmarking methodology to Boolean and neuromorphic circuits has been completed [9]. In this study, twenty state-of-the-art emerging logic devices are considered, including tunneling FETs, ferroelectric-based FETs, other charge-based devices, spintronic devices (both current and voltage controlled), and magnetic domain wall logic devices [9]. The study suggests that spintronic devices can potentially outperform conventional CMOS devices. Voltage controlled spintronics devices are shown to be more energy efficient than current driven devices [9]. The study concludes that spintronic-based neuromorphic

computing exhibits significant performance improvements as compared to spintronic-based Boolean circuits.

In addition to these emerging technologies, superconductive electronics is considered as a standalone replacement for conventional CMOS systems. In superconductive electronics, one important logic family is based on single flux quanta (SFQ) to encode and process data. Superconductive electronics were later mapped into the aforementioned Intel study where the delay is a factor of 1,000 less than the other technologies [12, 13]. Superconductive systems have been compared to conventional Boolean compute schemes, where eleven circuits from the ISCAS'85 benchmark suite and a 32 bit RISC-V ALU are demonstrated using both a superconductive cell library [14] and the TSMC 12 nm FinFET cell library [15]. This study targets stationary systems, where superconductive technology provides a consistent energy benefit; ten times lower energy per clock cycle is achieved in comparison to the 12 nm TSMC technology [15].

Common themes have emerged from these benchmark studies:

1) None of the emerging logic technologies is projected to sustain or offer sufficient performance and energy efficiency advantages to replace CMOS in traditional Boolean circuits and von Neumann architectures [16].

2) Some beyond CMOS devices exhibit unique characteristics that enable novel circuits and architectures that are not possible with CMOS [7].

3) Novel architectures that leverage non-volatile emerging devices are increasingly important for normally off, instantly on, computing architectures [10, 17].

4) Thermal sensitivity and thermal budgets are significant challenges affecting the development of high density ICs [18].

5) Spintronics is considered a promising technology for the beyond CMOS era [8, 17, 19, 20].

6) Superconductive technology could advance high performance computing (HPC) beyond exascale levels of computation [18].

1.2 Application-specific beyond CMOS development flow

These developments in innovative materials, memory, logic, and architectures will likely exhibit novel compute schemes and applications. This functional diversification, supported by beyond CMOS technologies, is expected to unleash a wide spectrum of novel solutions that have not been previously possible, such as reconfigurable nonvolatile logic [21] and on-chip integrated sensor networks [22]. This development can be classified into two tracks, as illustrated in Figure 1.2: A standalone flow where all of the electronic devices are monolithically integrated on the same substrate; and a

hybrid integration flow where heterogeneous technologies are integrated within the same platform.

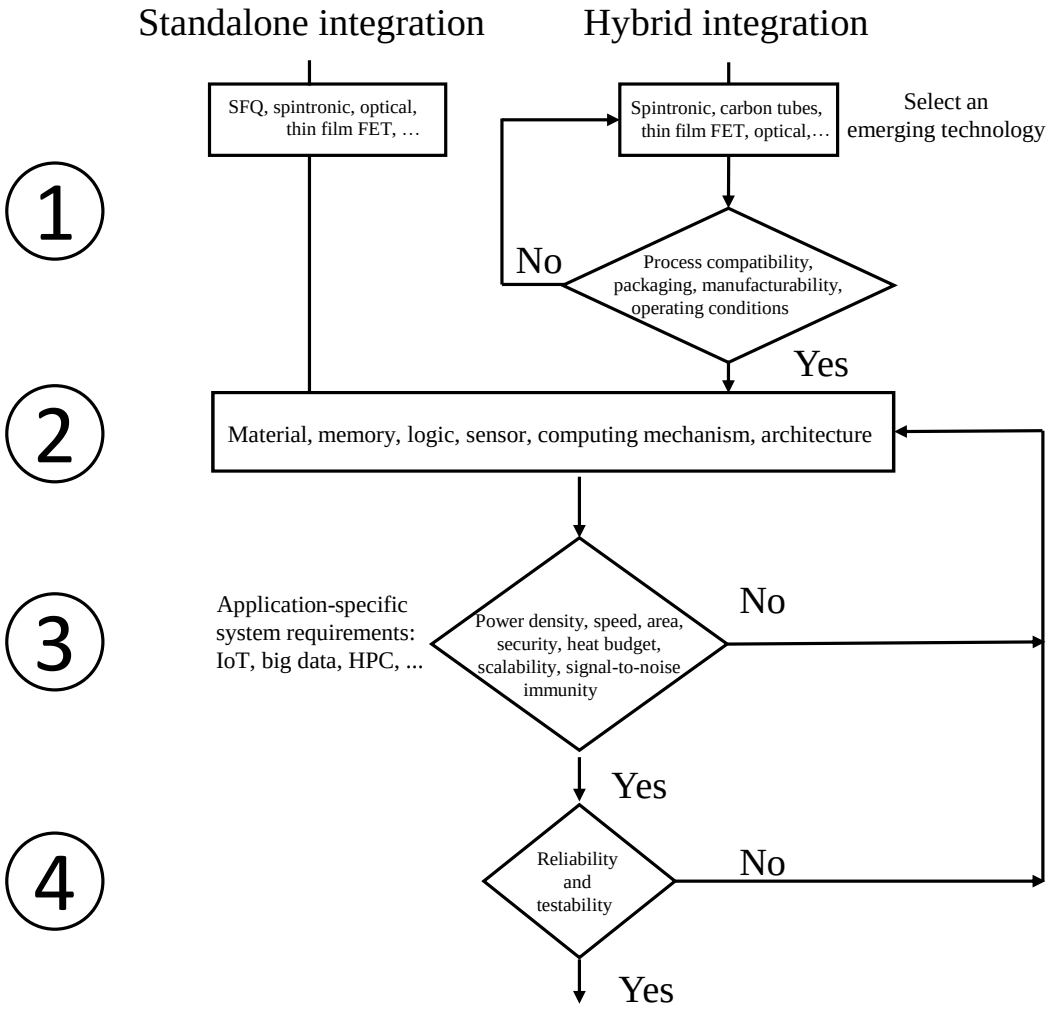


Figure 1.2: Development flow for application-specific beyond CMOS technologies to support novel compute schemes and applications, such as internet-of-things (IoT), big data, and high performance computing (HPC).

As shown in Figure 1.2, the development of next generation microelectronic technology is composed of four different iterative cycles. In the first stage, emerging

electronic technologies are heterogeneously integrated with the same manufacturing processes and share the same operating conditions. In the second stage, cooptimization of the materials, devices, and architectures is achieved by targeting specific applications and performance requirements (such as area, power density, and heat). During the final stage of this development process, the reliability of these systems can be enhanced with system-level issues such as improved testability.

Mapping an emerging technology into a specific Boolean application depends upon a set of attributes, such as the output gain, fanout, signal-to-noise immunity, power, and delay [9]. For non-Boolean devices, the size of the network, delay, noise level, and energy per network [8, 9] are the key parameters.

1.3 Beyond CMOS-based compute systems

Beyond CMOS technologies contribute to several primary areas of compute schemes, such as self-aware compute systems, reconfigurable compute in-memory systems, distributed compute systems, and high performance compute systems.

- **Distributed compute systems**: multiple nodes throughout a large network communicate and coordinate actions by exchanging messages to achieve a common objective. Each of these nodes may include a small processing unit to collect information, communicate with other nodes, and respond to stimuli. Embedded memory is common to these IoT applications. Multiple memory-based beyond CMOS

technologies are replacing CMOS-based memories at different levels of the memory hierarchy [23].

- **Self-aware compute systems**: with the growth in complexity and heterogeneity of integrated circuits, it is difficult to maintain on-chip management policies. Hence, self-aware systems need to understand, manage, and characterize the system and enable flexible runtime, speed, and voltages to develop an autonomous compute scheme. In autonomous compute systems, the system manages itself through adaptive reconfigurable logic based on a pre-programmed system set by the user [24]. This capability is achieved by on-chip distributed sensors that support self-monitoring and decision making algorithms [25]. Spintronic and memristor-based on-chip thermal sensors to support CMOS systems have previously been proposed [26–28].

- ***in situ* compute systems**: many integrated systems are data centric where processing "big data" and exascale computing with 10^{18} floating point operations per second is not achievable with conventional computing architectures [29]. In data centric architectures, data motion is greatly decreased by integrating the computational process within the storage system at different levels of the memory and storage hierarchy. The majority of emerging memristive devices (such as spintronic memristors, ferroelectric memristors, or titanium dioxide memristors) are applicable to these compute systems and exhibit promising performance [16, 18]. These systems are nonvolatile by nature,

targeting a wide spectrum of functions that utilize normally off, quickly on compute schemes [21, 30, 31].

- **High performance compute systems**: conventional Boolean compute schemes are no longer capable of delivering the computational needs of applications such as autonomous vehicles, precision medicine, and smart infrastructures [16]. Novel compute schemes such as neuromorphic computing (both artificial neural networks and spiking neural networks) can be achieved by beyond CMOS technologies, such as resistive memory (RRAM) [32, 33], phase change memory [34, 35], magnetic memory (MRAM) [28, 36, 37], ferroelectric transistors [38, 39], and Josephson junctions [40, 41]. Quantum compute schemes have attracted widespread attention by addressing computationally hard problems which require exponentially large computational resources, such as encryption and/or decryption. Quantum computers have been shown to break existing encryption methods [42], while more secure encryption systems resilient to quantum computing are under development [43]. Significant efforts in developing a quantum bit using emerging devices such as magnetic tunnel junctions (MTJ) are also under development [44], where superconductive electronics is a leading solution to control these quantum computers [45, 46].

In this dissertation, representative circuit and architectural advances are proposed that exploit the unique characteristics of emerging, beyond CMOS devices. The unique characteristics of these proposed systems include thermal sensitivity, non-volatility,

and reconfigurability. An MTJ is treated in this dissertation as an illustrative example of an emerging technology that can support beyond CMOS compute systems. MTJs are commonly used within commercial systems as an embedded memory. Importantly, MTJs are compatible with CMOS fabrication processes. In this dissertation, MTJs are proposed as a solution for several different compute schemes, including self-aware compute systems, compute in-memory, reconfigurable logic, and distributed compute systems.

The exploitation of beyond CMOS technologies in a broad variety of applications will be achieved by identifying opportunities for unconventional architectures. Although beyond CMOS devices exhibit a wide range of functions that can replace or support conventional CMOS systems, these devices also exhibit reliability issues that should be identified and addressed early in the technology development process. Since these devices suffer from low yield, advanced testability methodologies that target the unusual characteristics of emerging technologies are needed.

1.4 Reliability and testability of beyond CMOS systems

High reliability is a necessary requirement for integrated circuits. The challenge of realizing high performance with high reliability is escalating due to dimensional scaling,

novel materials and devices, hybrid integration of emerging technologies with CMOS, and operation in severe operating conditions (extreme high or cryogenic temperatures, long lifetimes, and high voltages and currents). These reliability challenges, combined with yield issues, are exacerbated by exotic manufacturing technologies.

Reliability and yield can be categorized by the failure paths (sequence of faults due to a physical failure) and failure mechanisms (physical cause of the failure). Determining the defects and faults is essential to enhance the lifetime of integrated circuits. This enhancement will be achieved by improving the fault coverage, where the system is tested to identify the characteristics of the faults, such as quantity, location, and type. Fault coverage is improved by exploiting design for testability (DFT) techniques to enhance the controllability and observability of the internal nodes within a system. An understanding of the physics of each failure mechanism and the development of effective and reliable algorithms that exploit these DFT techniques prior to fabrication are vital to the development of beyond CMOS systems.

Superconductive electronics target large scale, stationary systems where two to three orders of magnitude improvement in energy efficiency is available as compared to conventional semiconductor-based supercomputers. The challenge of achieving high performance SFQ systems with high reliability is escalating due to dimensional scaling, novel materials and devices, and operation in severe conditions (extreme cryogenic temperatures and sub-terahertz frequencies).

Advanced design for testability techniques are necessary to determine SFQ-based defects and faults and improve the ability to evaluate these faults. In this dissertation, design for testability methodologies of SFQ systems are proposed. Two different directions are described. In the first direction, embedded hardware solutions are proposed to enhance the controllability and observability of the internal nodes within an SFQ system to identify specific defects and faults. In the second direction, a methodology is described to develop a block-level fault model to produce the required test vectors to identify the type and location of certain JJ-level faults within an SFQ system.

1.5 Dissertation contributions and outline

The objective of this dissertation is to develop application-specific beyond CMOS systems, as illustrated in Figure 1.2. The contributions of this dissertation to support beyond conventional compute systems are as follows.

- A comparative study of different magnetization mechanisms to provide insight into MTJ structures that support different IoT applications and distributed compute systems.

- A thermal aware system based on a hybrid MTJ/CMOS thermal sensor to support self-aware compute systems.

- A hybrid MTJ/CMOS-based multi-bit memory cell to support a reconfigurable nonvolatile *in situ* logic compute system.

- Embedded hardware solutions, a test extraction module and hybrid test module, to enhance the testability measures of the internal nodes within an SFQ system

- A methodology to develop a block-level fault model that target JJ-based faults, stuck at a superconductive state or an open circuit state.

The internet-of-things (IoT) or the internet of everything has become a primary vehicle for connecting multiple nodes throughout a large network. Each of these nodes may include a small processing unit to collect information, communicate with other nodes, and respond to stimuli. Integrated memory is needed in these IoT applications. The form factor, initialization time, power dissipation, read/write speed, and cost are primary design criteria. An MTJ is therefore a potentially important solution for IoT applications. MTJ-based MRAM is a nonvolatile memory that operates at low latency (fast read operation), low leakage current, and high density. These capabilities support MRAM becoming a universal memory as compared to other nonvolatile memories such as e-Flash [47], phase change RAM [48], or resistive RAM [49]. The different magnetization mechanisms, physical structures, and electrical properties of MTJ-based MRAM are described in chapter 2 in terms of IoT applications. A comparative study of the magnetization mechanisms, also presented in chapter 2, provides insight into

which MTJ structures and magnetization mechanisms best support different IoT applications.

Thermal aware systems control distributed CMOS blocks based on the local temperature to enhance system speed, power, and reliability. The ultimate objective is multiple *in situ* temperature sensors, close to the CMOS device layer, distributed over the die, physically small, and leaking near zero power. MTJs provide this capability. An MTJ is a CMOS compatible device, fabricated within the metallic layers above the CMOS device layers. In chapter 3, a method for using an MTJ as a thermal sensor is presented. The method operates MTJs in an antiparallel state where an MTJ exhibits higher sensitivity to thermal variations as compared to the parallel state. The method exploits device magnetism, thermal stability, and resistance with respect to an applied voltage to sense the ambient temperature. These results are based on experimentally extracted parameters of a perpendicular and voltage controlled magnetic anisotropy MgO|CoFeB MTJ. A change in the antiparallel resistance by up to 16 Ω per degree Kelvin at a sense voltage of 0.2 volts is demonstrated.

Stacking additional systems into a compact area or scaling devices to increase the density of integration are two approaches to provide greater functional complexity. Excessive heat generated as a result of these technological advancements leads to an increase in leakage power and degradation in system reliability. Hence, a thermal aware system composed of hundreds of distributed thermal sensor nodes is an effective

solution to monitor the thermal behavior of a system. Such a system requires an efficient thermal sensor placed close to the thermal hotspots, small in size with a fast response while maintaining CMOS compatibility. Two hybrid spintronic/CMOS circuits that exhibit these traits are proposed in chapter 4. These circuits consume a low power of 11.9 μW during the on-state, linearity (R^2) of 0.96 over the industrial temperature range of operation (-40 to 125)^oC, and a sensitivity of 3.78 mV/K.

Recent developments in IC technology rely on device scaling and 3-D integration, integrating many billions of devices within a small area. These trends degrade system lifetime and reliability due to an increase in temperature caused by high power densities. Dynamically managing a system based on the local thermal characteristics is important to mitigate this issue. An on-chip thermal aware system composed of hundreds of distributed thermal sensors is proposed in chapter 5. This hybrid spintronic/CMOS thermal sensor exploits the thermal sensitivity and small area of an antiparallel magnetic tunnel junction. The sensor cell consumes as little as 500 pJ to read 1,024 thermal sensor nodes and generates a thermal map of a system composed of 32×32 thermal sensors.

In exascale computing, significant data are processed in real-time. Conventional CMOS-based computing systems follow read, compute, and write back mechanisms. This approach consumes significant power and time to compute and store data. Hence, compute in-memory systems (*in situ* computation) is an ideal platform for exascale

computation. In chapter 6, an MTJ-based multi-bit nonvolatile logic and memory cell is proposed to support *in situ* computation. The multi-level cell supports both a high speed read/write multi-bit memory cell and a nonvolatile logic gate that computes and stores input data in real-time.

Advanced testing methodologies are required to support complex digital SFQ systems. In chapter 7, two solutions are presented to enhance the testability of SFQ systems by improving the controllability and observability of the internal nodes. A test extraction module with a detection time of 7 ps and a hybrid test module with a detection time of 18 ps are presented. The proposed test modules are validated on a suite of benchmark circuits. A comparison of the effects of inserting the test modules into different benchmark circuits in terms of the overhead and testability measures is provided. The proposed test modules performing test insertion, extraction, and hybrid for the ISCAS'85 C17 benchmark circuit exhibit a power overhead of, respectively, 0.513, 0.27, and 0.78 fW/calculation. The proposed test modules significantly enhance, by more than 50%, the testability measures (controllability and observability) of the internal nodes, increasing overall fault coverage.

JJ-based fault models are proposed in chapter 8 for specific gate types. A faulty JJ has four modes of operation, stuck at superconductive, resistive, open circuit, or noisy switching. Two JJ-based fault modes are considered in chapter 8, stuck at superconductive (SC) state and stuck at open circuit state. A JJ stuck in the

superconductive state is modeled as a JJ with a high critical current, while a JJ stuck in the open circuit (OC) state is modeled as an open circuit. A high level JJ-based fault model is presented for the following RSFQ cells; JTL, splitter, DFF, OR, and AND. Test vectors to identify the type and location of a set of faults are generated based on the high level fault models. The fault coverage of the OC and SC faults and the location of each logic cell are identified; specifically, 72% of JJ-based faults (OC, SC, or both) can be detected within an SFQ system. The fault coverage of a JJ-based fault is 74% of SC faults and 70% of OC faults. While it is challenging to identify the location of OC faults within SFQ system, all SC faults within a splitter cell can be identified and the location of 18% of SC faults within an AND cell can be determined. A methodology is also proposed to develop a block-level fault model to produce the required test vectors to identify the type and location of JJ faults within SFQ systems.

Multiple technologies are currently being considered to supplement conventional CMOS circuits, targeting certain heterogeneous applications. Significant effort is required for these technologies to be more widely adopted. As discussed in chapter 9, with respect to MTJ technology; future work should include the development of analytic models, algorithms, and techniques targeting MTJ technology. This work should enhance the performance efficiency of MTJ-based memory technologies at different levels of the memory hierarchy. Additional research is necessary to further

investigate the influence of the physical structure of an MTJ on thermal sensing applications. With respect to SFQ systems, advanced SFQ defects, such as pinholes and flux trapping, need investigation to improve the quality of the fault models to enhance the fault coverage and overall testability of SFQ systems.

Chapter 2

MTJ Magnetization Switching Mechanisms (for IoT Applications)

The internet of things (IoT) or the internet of everything has become a primary vehicle for connecting multiple nodes throughout a large network. Each of these nodes may include a small processing unit to collect information, communicate with other nodes, and respond to stimuli. Integrated memory will be needed in all of these IoT applications. Form factor, initialization time, power dissipation, read/write speed, and cost are primary design criteria. Magnetic tunnel junction (MTJ)-based magnetic random access memory (MRAM) is a potentially important solution for IoT applications. MTJ-based MRAM provides a nonvolatile memory able to operate at low latency, low leakage current, and high density. These capabilities support MRAM becoming the universal memory as compared to other nonvolatile memories such as e-Flash, phase change RAM, or resistive RAM [50].

This discussion is outlined as follows. In section 2.1, the physical behavior and parameters affecting MTJ device performance are described. Different forms of magnetization operation are also discussed in section 2.1. In section 2.2, the primary magnetization mechanisms for different structures are characterized. A comparative study of the different magnetization mechanisms described in section 2.2is provided in section 2.3, emphasizing scaling, power dissipation, and circuit speed. A discussion of the effects of thermal variations on the different magnetization mechanisms for IoT applications is provided in section 2.4. Appropriate magnetization mechanisms applicable for different IoT applications are also described in section 2.4 and section 2.5. A summary is provided in section 2.6.

2.1 MTJ structures for MRAM

Tunneling magnetoresistance (TMR) was discovered in 1975 by Julliere [51] within a structure called a magnetic tunnel junction. It was not until the mid-1990s when fabrication of reliable MTJs became possible with the development of certain growth techniques and lithographic processes [52, 53]did MTJ become a commercially interesting technology. The basic MTJ structure is composed of two ferromagnetic (FM) materials separated by a nonmagnetic insulator, where the conductivity of the structure is determined by the angle between the magnetization direction of the two ferromagnetic layers. Maximum conductivity is achieved when

both magnetization directions are in parallel, and minimum conductivity when anti-parallel (AP). Slonczewski describes the conductance of the structure as a function of θ, the angle between the magnetization direction in the two FM materials, as [54]

$$G(\theta) = G_o(1 + P^2 cos\ \theta), \tag{2.1}$$

where P is the spin polarization of the ferromagnetic/barrier couple, and G_o is the conductance of the AP configuration. A primary parameter characterizing the quality of an MTJ is the TMR,

$$TMR = \frac{G_P - G_{AP}}{G_{AP}}, \tag{2.2}$$

where G_P and G_{AP} are, respectively, the conductivity of the parallel ($\theta = 0^o$) and anti-parallel ($\theta = 180^o$) configurations.

The ability to control the MTJ conductivity by the difference in magnetization angle θ is the key concept in using an MTJ as a building block for magnetic memory (such as MRAM). The FM layer is pinned while controlling the direction of magnetization of the other FM free layer (the storage layer). The FM layers are fabricated as uniaxial anisotropy layers (with a preferable single magnetization axis). The magnetization dynamics in the free layer are derived from the Landau-Liftshitz-Gilbert (LLG) equation [55, 56], as shown in (A.1).

MTJ structures are based on uniaxial anisotropy which means the magnetization direction prefers stability over a single axis (easy axis), ensuring the free layer is either in the parallel state or antiparallel state with respect to the pinned reference layer. Switching is controlled by applying a perturbing field by injecting current, applying an external magnetic field, or some other means. Uniaxial magnetic anisotropy (MA) is achieved by different MTJ structures to obtain either an in-plane MTJ (IMTJ) or perpendicular MTJ (PMTJ). Shape magnetic anisotropy is the main source for in-plane magnetic anisotropy where an elliptical MTJ is preferred as the magnetization tends to be directed along the long axis. A circular shape is preferred in the case of a perpendicular MTJ (PMTJ). Perpendicular magnetic anisotropy (PMA) in a PMTJ is due to two reasons, magneto-crystalline anisotropy of the crystalline lattice of the FM layer or by an interfacial magnetic anisotropy due to coupling at the interface between the FM layer and a neighboring layer (such as Co/Pt and Co/Pd [57]).

The pinned reference FM layer is achieved through the exchange bias effect by attaching an FM layer to an antiferromagnetic material. A synthetic antiferromagnetic structure (SAF) is preferred to achieve the pinned FM layer by using two antiparallel FM layers separated by a nonmagnetic (NM) metallic layer, reducing any stray fields affecting the free layer, as shown in Figure 2.1.

Numerous developments in MTJ structures have been achieved to enhance the TMR and other performance metrics. Different configurations used to switch the MTJ

magnetization are discussed in section 2.2. Switching by applying a magnetic field or injecting a current through the MTJ structure is described. Other forms of assistive switching, such as thermally assisted switching, are also discussed.

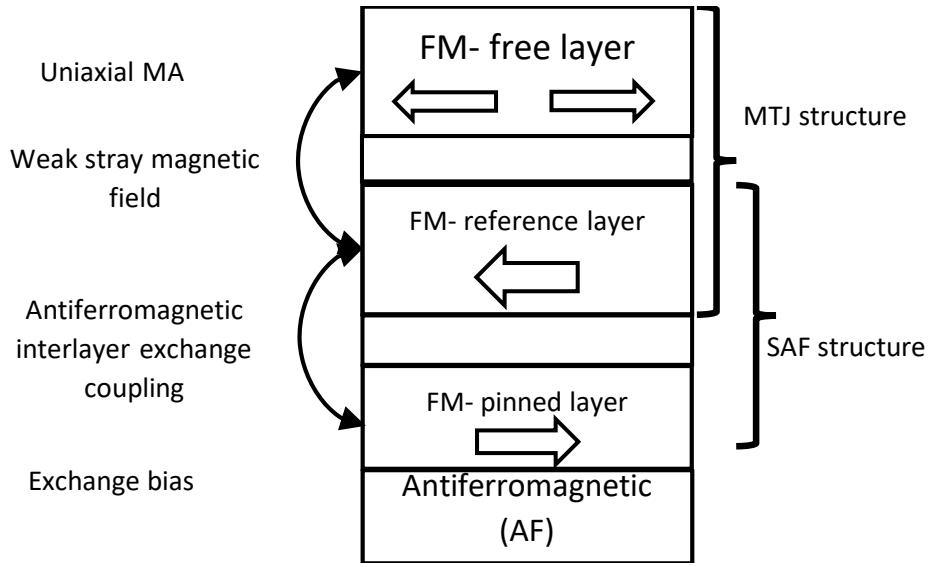


Figure 2.1: Stacking structure of IMTJ with SAF

2.2 MTJ-based MRAM magnetization mechanisms

Several magnetization mechanisms of MTJ-based MRAM have been developed. Controlling the magnetization direction of the free layer can be achieved by applying an external magnetic field (induced by current passing through a nearby wire) [58], injecting current through an MTJ structure to cause thermal perturbations to affect device magnetization properties (thermally assisted) [59], injecting greater current to induce spin transfer torque to the magnetization layer [60], or by adding an additional

perpendicular magnetization layer within an IMTJ to enhance STT performance to reduce the write current [61]. Recently developed magnetization mechanisms in MRAM are also discussed in this section.

2.2.1 Field-induced magnetic switching MRAM (FIMS-MRAM)

The magnetic field required to switch an MTJ is governed by the relationship, $\mu_o \mathrm{H}_{write} = 2K/M_s$, where K is a constant characterizing the magnetic anisotropy of the MTJ. The Stoner-Wohlfarth MRAM (ST-MRAM) was the first demonstrated MTJ-based MRAM, based on Stoner-Wohlfarth (SW) switching, as shown in Figure 2.2 [62]. Two orthogonal magnetic fields - one along the easy axis - are applied to an MTJ structure. The total field applied over the x and y axis, respectively, $\mathrm{H_x}$ and $\mathrm{H_y}$, based on (5), sets the lower limit of the write field,

$$\mathrm{H_x}^{2/3} + \mathrm{H_y}^{2/3} = \left(\frac{2K_u}{M_s}\right)^{2/3}, \tag{2.3}$$

where K_u is the magnetic anisotropy constant and M_s is the saturated magnetization. The upper limit of the switching operation is bounded by the MTJ shape, where the easy axis field is lower than the anisotropy field, $\mathrm{H}_K = 2K_u/M_s$. An improved field written mechanism has been achieved by Savtchenko [63], where two orthogonal fields are oriented at an angle of 45^o to the MTJ easy axis. The write operation is

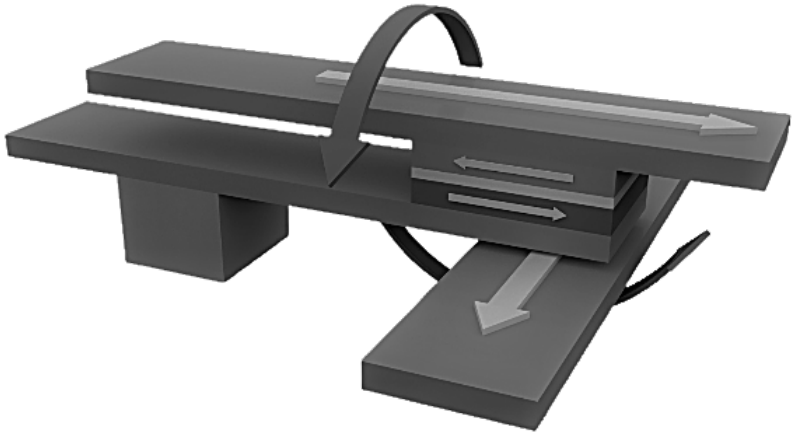


Figure 2.2: IMTJ writing using FIMS

performed in a toggle form to maintain full magnetization operation, hence called "Toggle MRAM." The spin flip flop field $\mathrm{H_{spinflop}}$ is

$$\mu_o M_s \mathrm{H_{spin_flop}} = 2\sqrt{K_{eff}\left(\frac{A}{t} + K_{eff}\right)}, \tag{2.4}$$

assuming the two FM layers are identical with a thickness t (a symmetric MTJ). K_{eff} is the effective anisotropy, and A is the interfacial coupling through the spacer (the tunnel barrier).

2.2.2 Spin transfer torque MRAM (STT-MRAM)

In STT-MRAM, switching the magnetization of the free layer is achieved by passing a polarized electric current through an MTJ structure, where the polarized moving electrons exert a torque on the magnetic storage layer. When a normal unpolarized

current passes through an FM layer, the current becomes polarized by the FM layer with a spin polarization factor [54]. Based on whether the current direction is into or out of the MTJ structure (the pinned layer), respectively, an antiparallel or parallel configuration is achieved, as shown in Figure 2.3.

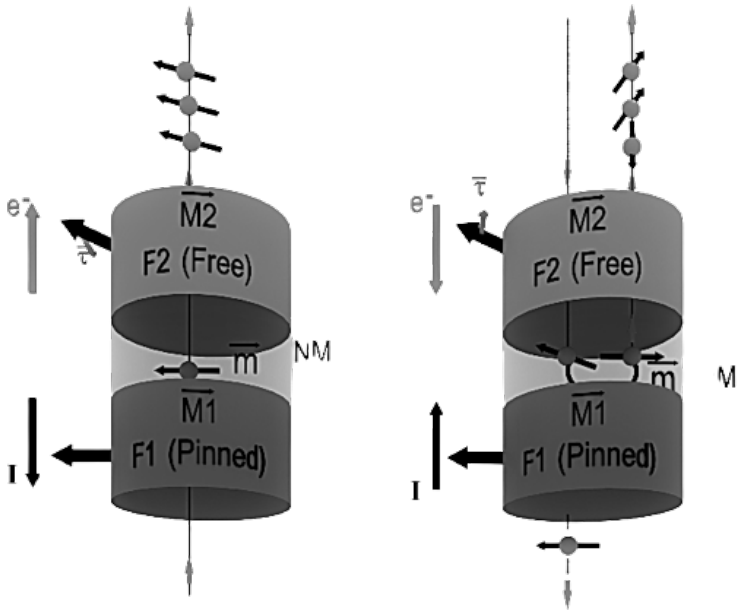


Figure 2.3: IMTJ writing using FIMS

The switching current should be greater than a critical magnitude, where the level depends upon the device dimensions and materials. For an IMTJ, the critical current density J_{co} to switch the FM layer is described by (2.5). The critical current is affected by thermal variations and depends upon the width of the current pulse. The critical

current $J_c(t)$ is described by (2.6),

$$J_{co} = \frac{2e\alpha\mu_o M_s t_{FL}(H_{K\|} + (\frac{M_s}{2}))}{\hbar\eta}, \tag{2.5}$$

$$J_c(t) = J_{co}\left[1 - \frac{K_B T}{K_u V_{FL}} ln\left(\frac{t}{\tau_o}\right)\right], \tag{2.6}$$

where e is the electron charge, μ_o is the permeability of free space, t_{FL} is the thickness of the FM, $H_{K\|}$ is the in-plane magnetic anisotropy field, $\hbar$ is the reduced plank constant, η is the spin transfer efficiency, t is the pulse width, K_B is the Boltzmann constant, T is the temperature, K_u is the uniaxial magnetic anisotropy constant, V_{FL} is the volume of the FM layer, and τ_o 1 is the inverse of the frequency factor or the attempt frequency of the thermal reaction. For an IMTJ, the magnetization configuration does not efficiently support both thermal stability and writability [64]. Other configurations to reduce the critical current in an IMTJ are

- Dual MTJ: where the storage FM layer is sandwiched between two pinned FM layers, hence the current passing through the structure exerts a larger torque (2x), reducing the critical current to switch the layer [65].

- Perpendicular polarizer: where a perpendicular polarizer is stacked within an MTJ [66], enhancing the precessional switching. The storage layer acquires an out-of-plane component due to the added magnetic precessional movement.

- Reduced demagnetized field: where either volume or interfacial perpendicular magnetic anisotropy is added - by changing the barrier and FM materials - decreasing the demagnetizing field effect, thereby reducing the critical current [67].

MTJ structures that exhibit magnetic anisotropy normal to the surface are faster, dissipate less power, and are higher density than an IMTJ due to a lower critical current and the circular shape of the magnetic cell. A PMTJ-based MRAM is therefore more appropriate for IoT applications than an IMTJ. The critical current density J_{co} for a PMTJ is described by (9) where $H_{K\perp}$is the perpendicular-to-plane magnetic anisotropy,

$$J_{co} = \frac{2e\alpha\mu_o M_s t_{FL}(H_{K\perp} - M_s)}{\hbar\eta}. \tag{2.7}$$

A PMTJ suffers from a higher Gilbert damping factor α than an IMTJ. In addition, it is difficult to fabricate the crystalline PMTJ structure to produce perpendicular magnetic anisotropy. The critical procedures of annealing and oxidation are required to grow a PMTJ. That leads to a preferential choice between IMTJ and PMTJ based

on the manufacturing difficulty. A comparative study between IMTJ and PMTJ in terms of IoT applications is provided in section 2.3.

2.2.3 Thermally assisted MRAM (TA-MRAM)

In TA-MRAM, thermal variations are induced by passing current through an MTJ and/or using a high thermal conductivity material. The magnetic anisotropy constant is lower at higher temperatures which affects the magnetic field writing to the storage layer. Accordingly, a thermal assist mechanism is integrated within the FIMS-MRAM structure, achieving higher performance [68]. The same concept can be applied to an STT-MRAM, where the same current line is used for heating and switching the cell. The write operation is composed of multiple stages. The first stage heats the cell by passing a current for a specific duration. A lower level of current is applied for a different duration to switch the state. The cell is left to cool, storing the written state [69].

2.3 Comparative study

A PMA-based STT-MRAM structure exhibits better performance, as listed in Table 2.1, due to a higher density and reduced power [70]. Research is on-going for determining the optimum device characteristics to provide acceptable performance as a replacement memory technology. Research on different structures, materials, and

mechanisms suggests that a PMTJ-based MRAM is more applicable for those IoT applications that require higher operating speed.

Table 2.1: Advantages and disadvantages of different MTJ structures

Technique	Advantages	Disadvantages
IMA	Controlled retention time	Density Power consumption
Interfacial PMA	Higher density Optimizing I_c and α	Low retention time
Crystalline PMA	Higher density	Largerα Low retention time
SAF pinned layer	Symmetric MTJ switching	Increases height of MTJ stack
Dual MgO/Freelayer interface	Enhance interfacial PMA Reducesα	Increases MTJ resistance (R—MTJ)
Dual tunnel barrier with dual PL	Reduces I_c and α	Increases R—MTJ Increases height of MTJ stack
Tilted magnetic anisotropy(MA)	Reduces I_c	Reduces TMR ratio Difficult to fabricate
Orthogonal pinned layer	Reduces I_c without affecting TMR ratio	Increases height of MTJ stack Difficult to fabricate

A comparison of magnetization mechanisms for MTJ-based MRAM is summarized in Table 2.2 [50,71]. A FIMS-MRAM is based on an IMTJ, requiring large cells (which suffers from scaling and bit selectivity due to stray field disturbance from the writing magnetic field from neighboring cells). FIMS consumes large power (due to the large

write current to induce the write magnetic field) but exhibits long retention times and thermal stability.

Toggle based MRAM solves the selectivity problem but can place the memory bit in an undefined state - requiring the read operation to be performed before the write operation. A toggle MRAM is predicted to not be highly scalable below 90 nm [50]. FIMS is a robust technology, exhibiting high reliability, endurance, and resistance to radiation, making it a good candidate for automotive, sensor-based weather forecasting, and IoT applications [64, 72].

Table 2.2: Performance of different magnetization mechanisms

	Scalability	**Endurance**	**Write time**	**Write current**
FIMS	Poor	1016	>10 ns	∼10 mA
Toggle	Good	1015	>30 ns	>30 mA
TAS	Good	1012	>20 ns	∼1 mA
STT	Very good	1016	<5 ns	∼100 mu A
TAS+STT	Best	1012	<8 ns	∼100 mu A

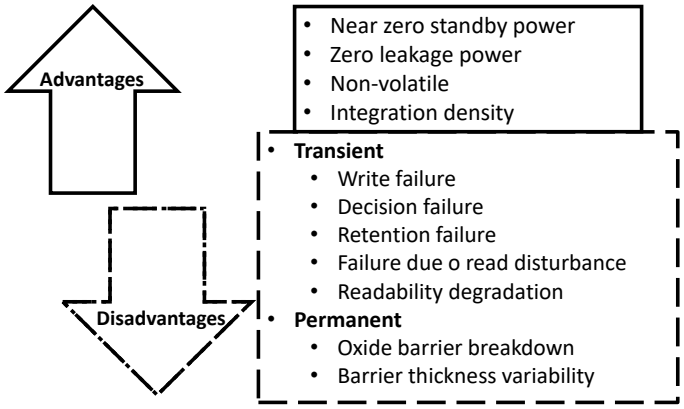


Figure 2.4: Advantages and reliability challenges of STT-MRAM

The advantages and disadvantages of STT-MRAM are illustrated in Figure 2.4 [73], where the reliability is improved by optimizing the read/write currents, utilizing novel FM and barrier materials, or applying different magnetization mechanisms such as VCMA or SOT. Spin orbit torque MRAM (SOT-MRAM) shows promising potential due to the separate read and write paths, supporting symmetric switching.

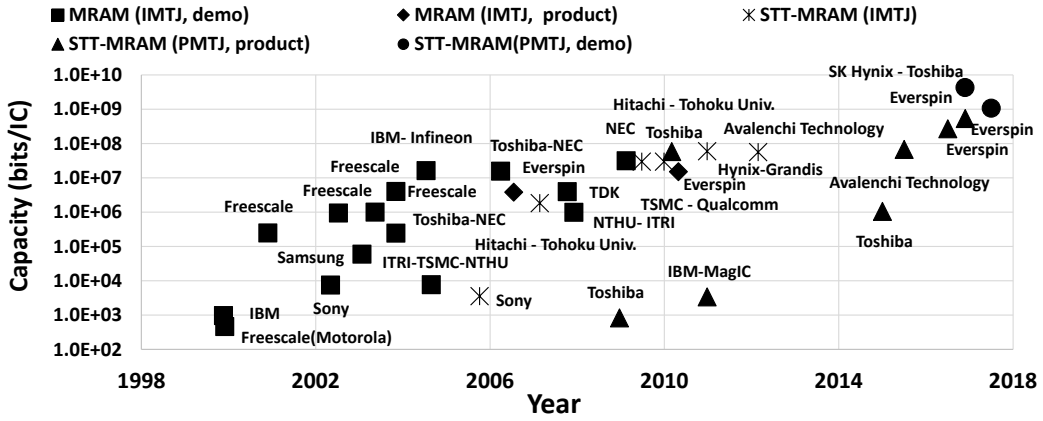


Figure 2.5: MRAM development trend

The thermally assisted switching (TAS) mechanism requires less writing power and is higher speed than FIMS, although difficult to scale. STT-based mechanisms exhibit better performance for both IMTJ and PMTJ [50, 70, 73, 74]. TAS-based MRAM suffers from the same problems as FIMS. Merging TAS with STT-MRAM exhibits better scalability and performance as compared to FIMS, TAS, and STT-MRAM [50, 71]. TAS+STT-MRAM is a good candidate for IoT applications which require high speed while dissipating low power.

2.4 MTJ-based MRAM for different temperatures

The operation of an MTJ-based MRAM as a nonvolatile memory is composed of three states. The write state with one magnetization mechanism, the retention state, describing how long the memory can maintain the written information, and the read state. The key objective in MTJ-based MRAM is to maintain low read/write/idle power consumption, high retention time, a wide operating temperature range, and low read and write delays. Each of these objectives are affected by the thermal stability factor Δ, as expressed in (10),

$$\Delta = \frac{\Delta E}{K_B T} = \left. \frac{\mu_o {M_s}^2 {t_{FL}}^2 (A_R - 1) w}{K_B T} \right|_{IMTJ} =$$

$$\left. \frac{[(K_v - (1/4)\mu_o (3N_z - 1){M_s}^2) t_{FL} + K_s] \frac{\pi}{4} w^2}{K_B T} \right|_{PMTJ} \tag{2.8}$$

where $\Delta E = K_{eff} V_{FL}$is the barrier height of the magnetic material. For an elliptically shaped FM IMTJ layer, A_R is the aspect ratio of the ellipse with width w. A circular FM PMTJ layer has a diameter w, K_v is the PMA constant, K_s is the surface energy, and N_z is the perpendicular-to-plane demagnetization coefficient.

The thermal stability factor decreases with scalability, as shown in (10), directly affecting the retention time and power consumption. The integration density, error rate, and thermal stability of MTJs are presented in [75], as shown in Figure 2.6, describing the critical integration level while maintaining an acceptable error rate.

Thermal variations affect the electrical device characteristics (K_{eff}, J_{co}, TMR) in a stochastic manner. TMR was discovered in a Fe-GeO-Co structure at 4.2 K [51]. By the mid-1990s, a TMR ratio of 10 to 70% at room temperature was achieved [52, 53]. Later, after 2004, giant TMR exhibited a ratio of 250% - reaching up to 600% - at room temperature by using a monocrystalline magnesium oxide (MgO) barrier [76–78]. These developments achieve a PMTJ with perpendicular anisotropy with low switching current and high thermal stability [71, 79] which is appropriate for ultra-low power sensor nodes used in IoT applications. Toggle MRAM exhibits a wide range of operating temperatures, from 0 to 70^{o}C for commercial applications and -40 to 125^{o}C for automotive and military applications[68, 80]. These different operating temperatures are particularly relevant for IoT-based sensor nodes and processing units located in extreme environmental conditions.

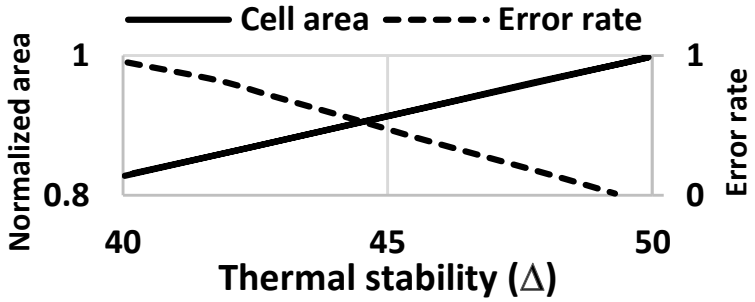


Figure 2.6: Impact of thermal stability on cell area and error rate

2.5 MTJ-Based MRAM for different IoT applications

Wireless sensor networks are a form of IoT, where the sensors are located at far distances and operate autonomously and for a long time. The nodes are expected to survive extreme environmental conditions while conserving energy. The sensor node behaves as an embedded system where an operating system is located within local memory. More than ten year retention time is a useful capability for those memories supporting IoT systems. Retention time, power consumption, and thermal stability are directly proportional to the current levels (which is proportional to the size of the MTJ), as described in Table 2.3 for STT-MRAM [81].

Table 2.3: Retention time versus write latency for STT-MRAM

Retention Time	10 ms	1 sec	10 years
Write Latency (ns)	3	6	11
$I_c(\mu$ **A**)	61	82	114

MRAM with different magnetization mechanisms has been integrated into commercial applications, as shown in Figure 2.5 [79]. Everspin Technologies recently released a 1 Gb DDR4 Spin Torque MRAM. GlobalFoundries and eVaderis announced the development of an ultra-low power microcontroller based on an embedded magnetoresistive non-volatile memory. MRAM provides a fast write-read and low power consumption with no static power which positions this technology for many IoT applications. For

those IoT applications which require intermittent access with fast working memory, TAS with STT-MRAM is a good candidate, while greater development is needed to enhance read access and read failure rates.

Toggle MRAM is a good candidate for those IoT applications requiring long periods of standby current as this technology is non-volatile with fast write times. TAS with STT-MRAM within a PMTJ structure is a better choice for ultra-low power, battery operated IoT nodes, as this application requires zero standby energy with a fast power-up time.

2.6 Summary

MTJ-based MRAM is an effective candidate technology for the specialized memory needed in IoT applications. PMTJ structures exhibit higher density and lower power than IMTJ due to the circular shape and perpendicular interfacial or magnetocrystalline anisotropy. STT-MRAM assisted with TAS provides enhanced speed and power as compared to FIMS, toggle, TAS, and STT-MRAM. The ability to maintain a ten year retention time operating at extremely low energy makes MRAM highly appropriate for IoT applications, particularly for sensor nodes located in harsh environmental conditions.

Chapter 3

PMTJ Temperature Sensor utilizing VCMA

The need for thermal aware systems is increasing. Systems able to measure the local ambient to affect system function are becoming increasingly desirable [82]. A thermal sensor near the CMOS device layer and sufficiently accurate to capture the local temperature is required. Materials with a high thermal conductivity and thermal stability are therefore desirable.

A thermal sensor is a two terminal device that produces an output current or voltage in relation to the ambient temperature. A thermal sensor should have low thermal mass and a fast response time. Integrated thermal sensors are monolithically fabricated in a semiconductor technology. These integrated thermal sensors are often in the form of bipolar junction transistors (BJT) where the relationship between the base-emitter voltage and collector-emitter current is a function of the ambient

temperature. Recent developments have enabled the fabrication of stacked BTJs with a precision amplifier to enhance sensor accuracy [83].

A magnetic tunnel junction (MTJ) is a three layer structure composed of two ferromagnetic (FM) metallic layers separated by a nonmagnetic insulator layer. The tunneling magneto-resistance characterizes the change in device resistance based on the difference in orientation of the magnetization direction between the two FM layers [51]. The MTJ fabrication process is also CMOS compatible, fabricated within the IC metal layers.

MTJs are good candidates to become an effective temperature sensor since the MTJ resistance depends upon the ambient temperature [27, 84]. The device is small, CMOS compatible, close to the device layer, and leaks almost zero power. A network of MTJs, distributed over the die, can support local monitoring of the thermal characteristics across a system.

A mechanism is described here for using an MTJ in the antiparallel state as a thermal sensor based on a change in resistance with temperature. The antiparallel (AP) state is preferable for thermal sensing applications, as the parallel (P) state is almost independent of the ambient temperature. The chapter is outlined as follows. In section 3.1, the effects of temperature and voltage on the MTJ magnetic and electrical properties are discussed. A procedure for using an MTJ as a thermal sensor is described in section 3.2. A case study of an MTJ-based thermal sensor is also described. In

the case study, an electrical model is presented which is based on experimentally extracted parameters, as described in section 3.3. An MTJ-based thermal sensor circuit is presented in section 3.4. The chapter is summarized in section 3.5.

3.1 Influence of temperature and voltage on MTJ

MTJs are used as a memory element by pinning the magnetization of one of the FM layers while the other FM layer is the free magnetization layer. Both FM layers exhibit uniaxial magnetic anisotropy (MA) behavior, where a uniaxial preferably magnetization direction exists for the material magnetization. Two popular forms of FM materials used within an MTJ are in-plane MTJ (IMTJ) or out-of-plane – perpendicular MTJ (PMTJ). IMTJs are elliptical structures where the shape anisotropy dominates the anisotropy behavior, resulting in magnetization along the long axis. PMTJs, alternatively, exhibit out-of-plane MA due to other forms of anisotropy, e.g., magneto-crystalline anisotropy or interfacial magnetic anisotropy [57].

An MTJ is used as a memory element by changing the state of the MTJ between parallel and antiparallel states, where both FM layers share the same orientation or are 180^o out-of-phase. Changing the state of an MTJ is achieved by applying a torque with an energy greater than the system magnetization energy. The spontaneous magnetization energy is in the form of multiple magnetic anisotropies that exist within the FM system.

MTJs can be modeled as a bistable system, with two possible stable states, parallel and antiparallel. The energy separation between these states is controlled by the effective anisotropy energy ΔE. The system magnetization energy is a function of the device shape and physical dimensions in addition to the FM and tunneling (insulator) materials. The effective anisotropy energy ΔE is

$$\begin{aligned} \Delta E = K_{eff}\nu_{FM} = E_{Bulk\ anistropy} + \\ E_{Interface\ anistropy} + E_{Demagnetization} + E_{VCMA}, \end{aligned} \tag{3.1}$$

where K_{eff} is the effective anisotropy constant, and ν_{FM} is the volume of the FM layer. ΔE can be approximated based on the MTJ structure, and E_{VCMA} is the energy maintained by the voltage controlled magnetic anisotropy (VCMA). Different forms of magnetic anisotropy can exist within the FM layers.

Recent studies have promoted PMTJs over IMTJs for high performance memory applications due to scalability, fast switching, and low power consumption for both writing and reading [85]. For a PMTJ, the perpendicular magnetic anisotropy (PMA) mechanism exists as either interfacial, bulk PMA, or both, based on the device material and shape. The anisotropy energy for a PMTJ without the voltage dependent term can be approximated as [64]

$$\Delta E|_{PMTJ} \approx [K_V + \frac{2K_i}{t_{FM}} - \frac{1}{2}\mu_0(N_Z - N_{XY}){M_S}^2]\nu_{FM}, \tag{3.2}$$

where K_V is the volume anisotropy constant, K_i is the surface interfacial anisotropy constant, t_{FM} is the thickness of the FM layer, μ_0 is the permeability of free space, N_Z is the demagnetization tensor in the $\hat{z}$ (perpendicular) direction, M_S is the saturation magnetization, and ν_{FM} is the volume of the FM layer.

3.1.1 Influence of voltage on MTJ

Some MTJ structures exhibit VCMA, where the applied voltage affects the magnetic anistropy of the MTJ. The effect of voltage on the MTJ can be modeled as a magnetic field $\vec{H}_{VCMA} = [2\zeta_{VCMA}V/\mu_0 M_S t_{ox} t_{FL}]\ \hat{e}_n$ normal to the MTJ structure, where ζ_{VCMA} is the VCMA coefficient, and V is the applied voltage [86]. The voltage dependent term of the anisotropy energy is modeled as [87]

$$\Delta E|_{VCMA} = \Delta E(0) - \frac{2\zeta_{VCMA} V A_{MTJ}}{t_{ox}}, \tag{3.3}$$

where $\Delta E(0)$ is the anisotropy energy with zero applied voltage, and A_{MTJ} is the surface area of the MTJ.

The tunneling magneto-resistance (TMR) characterizes the quality of the electrical response of the MTJ, where $TMR = (R_{AP} - R_P)/R_P$, R_{AP} is the antiparallel state resistance, and R_P is the parallel state resistance. TMR exhibits a voltage dependence,

as shown in [88]

$$TMR(V) = \frac{TMR(0)}{1 + (V/V_h)^2}, \tag{3.4}$$

where $TMR(0)$ is the TMR at a zero applied voltage. According to the Juliere model, $TMR(0) = 2P_1P_2/(1 - P_1P_2)$, where P_1 and P_2 are the spin polarization percentage of the two FM layers, and V_h is the voltage at which TMR is halved [51].

3.1.2 Influence of temperature on MTJ

R_P is independent of thermal variations, as experimentally described in [89]. The antiparallel resistance, however, decreases with an increase in temperature. The temperature influences most of the MTJ parameters, including the spin polarization $P(T)$, saturation magnetization $M_S(T)$, and all of the magnetic anisotropic constants $K(T)$, which affect the MTJ antiparallel resistance. $R_{AP} = 1/(G_{AP} \times A_{MTJ})$ where G_{AP} is the conductance of the AP state. The dependence of G_{AP} on temperature is

$$G_{AP}(T) = G_T \left[1 - P_1(T)P_2(T)\right] + G_{SI}, \tag{3.5}$$

where $G_T = G_0 \left(\sin(CT)/CT\right)$ is the elastic spin dependent term, G_0 is the parallel state conductance $G_0 = \left(3.16 \times 10^{10}\sqrt{\phi_B}/t_{ox}\right) exp(-1.025 \times \sqrt{\phi_B} \times t_{ox})$ at zero voltage and zero temperature, T is the ambient temperature, ϕ_B is the average tunneling barrier height (in eV), t_{ox} is the thickness of the insulator barrier layer, and

$C = 1.387 \times 10^{-4} t_{ox}/\sqrt{\phi_B}$ is a material dependent parameter [90]. $G_{SI} = ST^{4/3}$ is the inelastic spin independent conductance, and S is a fitting parameter. The dependence of the spin polarization on temperature can be fitted as [91, 92]

$$P(T) = P(0)\left[1 - \beta_P T^{\alpha_P}\right], \tag{3.6}$$

where β_P and α are fitting parameters related to the device dimensions and material properties.

The dependence of other MTJ parameters on temperature is primarily modeled by fitting expressions [93],

$$M_S(T) = M_S(0)\left[1 - \left(\frac{T}{T^*}\right)^{\beta_M}\right], \tag{3.7}$$

$$K_i(T) = K_i(0)\left(\frac{M_S(T)}{M_S(0)}\right)^{\beta_K}, \tag{3.8}$$

$$\zeta_{VCMA}(T) = \zeta_{VCMA}(0)\left(\frac{M_S(T)}{M_S(0)}\right)^{\beta_\zeta}, \tag{3.9}$$

where $P(0)$, $M_S(0)$, $K_i(0)$, and $\zeta_{VCMA}(0)$ are measured at zero temperature, and β and α are fitting factors for each parameter.

3.1.3 Combined influence of temperature and voltage on MTJ

Using (4), (5), and (6), $TMR(T,V)$ can be described [94] by (10), and $R_{AP}(T,V)$ by (11).

$$TMR(T,V) = [\frac{TMR(0)}{1+(V/V_h)^2}]\times \frac{2\left(1-\beta_P T^{\alpha}\right)^2}{TMR(0)\left[1-\left(1-\beta_P T^{\alpha}\right)^2\right]+\left[1+TMR(0)\right]G_{SI}/G_T+2}, \tag{3.10}$$

$$R_{AP}(T,V) = R_P(TMR(T,V)+1). \tag{3.11}$$

Using (2), (7), (8), and (9), the system anisotropy energy for a VCMA controlled PMTJ, neglecting the volume magnetic anisotropy term, is approximated as

$$\Delta E(T,V)|_{PMTJ} \approx [\frac{2K_i}{t_{FM}} - \frac{1}{2}\mu_0(N_Z - N_{XY}){M_S}^2 - \frac{2\zeta_{VCMA}VA_{MTJ}}{t_{ox}}]\nu_{FM}. \tag{3.12}$$

As previously mentioned, an energy barrier exists between the P and AP state. The device switches from the AP state to the P state when applying an energy that exceeds the built-in energy (related to the critical switching energy). The critical switching voltage, determined from (12), is

$$V_{C0}(T)\big|_{PMTJ} \approx \frac{t_{ox}t_{FM}}{\zeta_{VCMA}(T)}\left[\frac{K_i(T)}{t_{FM}} - \frac{1}{4}\mu_0(N_z - N_{x,y})M_S(T)^2\right]. \tag{3.13}$$

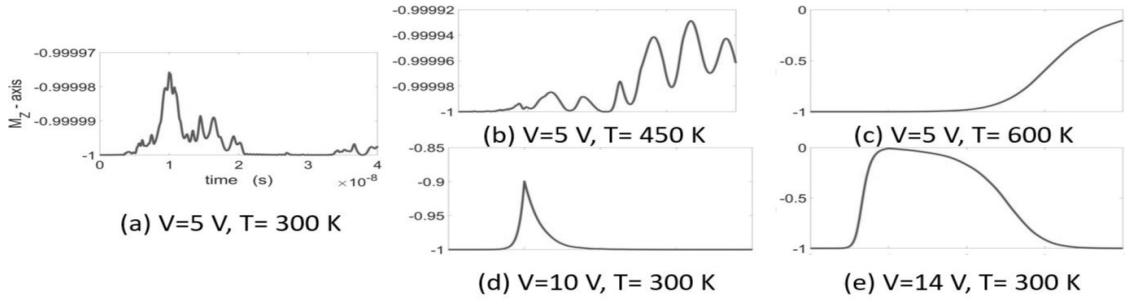


Figure 3.1: Magnetization behavior of MTJ in Z-axis under different sense voltages V and ambient temperatures T, (a) V = 5 volts and ambient temperature T of 300 K, (b) V = 5 volts and T = 450 K, (c) V = 5 volts and T = 600 K, (d) V = 10 volts and T = 300 K, and (e) V =14 volts and T = 300 K

3.2 MTJ-based thermal sensor

To sense the MTJ resistance, a voltage pulse is applied across the MTJ. To accurately measure the temperature, the change in the AP resistance due to an applied voltage and temperature needs to be determined. Equation (3.14) describes the maximum rate of change in the resistance with respect to temperature at the sensing voltage,

$$\max_{V_{min} \leq V \leq V_{max}} \left. \frac{\partial R_{AP}(T, V_i)}{\partial T} \right|_{V=V_i}, \tag{3.14}$$

where V_{min} and V_{max} are, respectively, the minimum and maximum voltage applied across an MTJ without changing the AP state.

The thermal stability Δ determines the limits of the applied voltage and the range of temperature where the device can stably operate. Δ is the ratio of the MTJ system magnetization energy and the energy perturbation to the system, which is a function

of temperature and applied voltage,

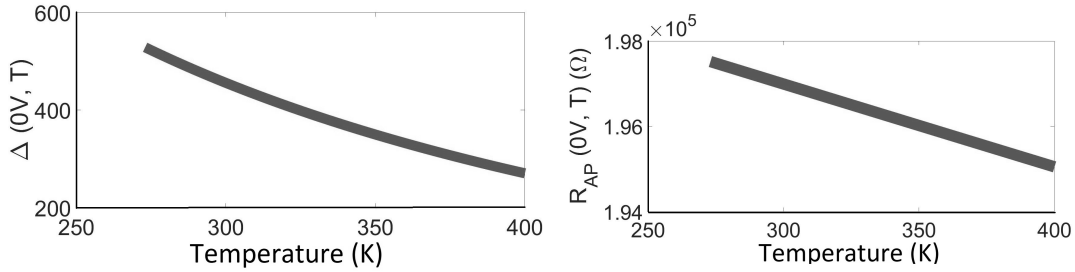
$$\Delta(T,V) = \frac{\Delta E(T,V)|_{MTJ}}{K_B T} = \frac{K_{eff}(T,V)\nu_{FM}}{K_B T}, \tag{3.15}$$

where K_B is the Boltzmann constant. The bias point (the voltage pulse) can be determined from (13).

3.3 Case study

The physical parameters used in this work are based on PMA and VCMA nanoscale MgO|CoFeB based MTJs [93, 94]. These physical parameters are listed in Table 3.1.

At a specific sensing voltage, the thermal stability of the device decreases with increasing temperature. The change in the thermal stability and antiparallel resistance when no sensing voltage is applied is depicted in Figure 3.2.



(a) (b)

Figure 3.2: Change in (a) thermal stability, and (b) antiparallel resistance with respect to temperature with no applied sense voltage

Table 3.1: MTJ physical parameters

Parameters	**Description**	**Value**
w_{FL}	FM width = radius	20 nm
t_{FL}	FM thickness	1.5 nm
t_{ox}	Barrier thickness	1.1 nm
Φ_{BL}	Barrier height	0.39 eV
V_h	Voltage @ half TMR	0.5 V
S	Spin independent conductance factor	1.1×10^{-12}
β_P	Fitting parameter for P	2.07×10^{-5}
α_P	Fitting parameter for P	2.3
β_M	Fitting parameter for M_S	1.5
T^*	Fitting parameter	1120 K
β_{Ki}	Fitting parameter	2.3
$\beta_{\zeta VCMA}$	Fitting parameter	2.83
N_z	Demagnetization tensor factor in Z	0.9343
N_{xy}	Demagnetization tensor factor in XY	0.015
K_{i0}	Interfacial MA at 0 K	$2.02 \times 10^{-3} J/m^2$
M_{S0}	Saturation magnetization at 0 K	$1457 \times 10^3 A/m$
TMR_0	TMR at 0 K	3
ξ_{VCMA0}	VCMA factor at 0 K	48.9×10^{-15} J/(V.m)

To characterize a read disturbance, the critical switching voltage, determined from (13), is 25.8 volts. The sensing voltage should therefore be less than half of that value, around 13 volts. In practical memory applications, to decrease the critical switching voltage, an embedded in-plane magnetic field bias is used [86, 95], or a polarizer layer is added to the MTJ structure [96]. Note that the device only operates due to VCMA. Other perturbation torques exerted by spin polarized electrons, such as spin transfer torque, are not considered.

An MTJ is evaluated under different scenarios to characterize the effect of the sense voltage and ambient temperature on the MTJ read disturbance. The magnetization behavior as a function of sense voltage with variable amplitude, pulse width of 10 ns, and ambient temperature of 300 K is illustrated in Figure 3.1a. The device exhibits a small disturbance but maintains the write state at a voltage below the predetermined critical switching voltage. As the sense voltage approaches the critical switching voltage, the read disturbance increases, as illustrated in Figures 3.1(b) and 3.1(c).

The effect of thermal stability on the read disturbance and hence device sensitivity has also been evaluated. The device is evaluated with the same sense voltage, pulse width, but different ambient temperatures, as illustrated in Figures 3.1(d) and 3.1(e). $\Delta(300, 5)$ is approximately 365, meaning the device energy is greater than the energy disturbance. In the second scenario, the same device is evaluated with the same sense voltage but with an ambient temperature of 600 K. Assuming the device maintains the same physical behavior at this high temperature, $\Delta(600, 5)$ is approximately 85. The device switches to the parallel state as the steady state is approached.

The simulations are based on a macrospin compact model with a 10 A/m external magnetic field in the $\vec{x}$ direction to evaluate the switching behavior. The simulation results for different scenarios of ambient temperature and voltage pulse amplitude are shown in Figure 3.1. Both the ambient temperature and/or the amplitude of the

sense voltage are able to switch the device into the parallel state. The parallel state should be avoided to ensure the device remains sensitive to the ambient temperature.

The behavior of R_{AP} with respect to temperature and sense voltage is illustrated in Figure 3.3. At a specific sense voltage, the resistance decreases with respect to the ambient temperature. The rate of change in the resistance with respect to the sense voltage is illustrated in Figure 3.4. Operating an MTJ as a thermal sensor requires a sense voltage that can bias the device to the maximum rate of change in resistance with respect to the ambient temperature.

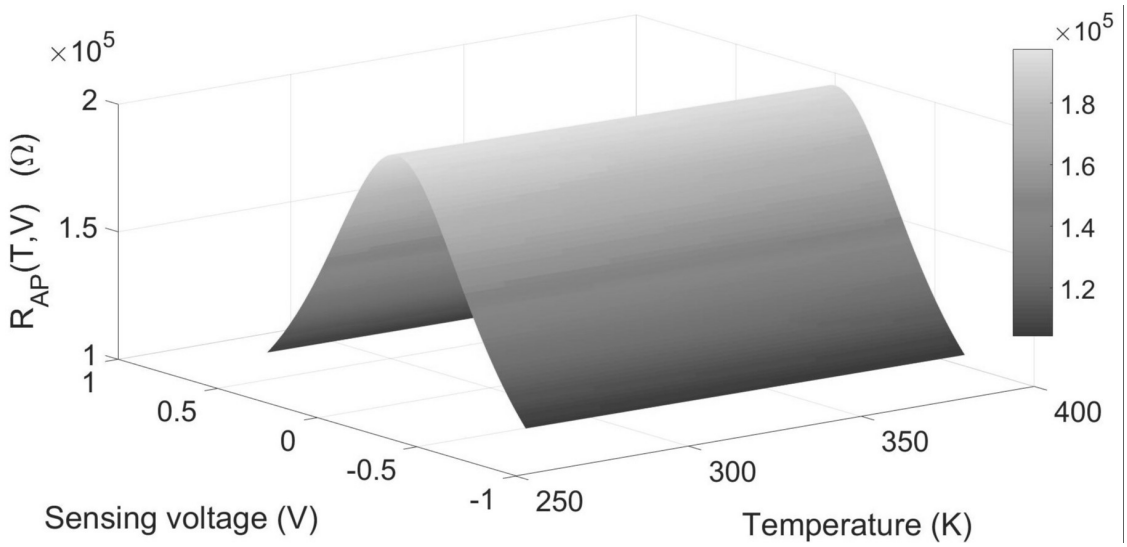


Figure 3.3: Change in the antiparallel resistance with respect to temperature and voltage

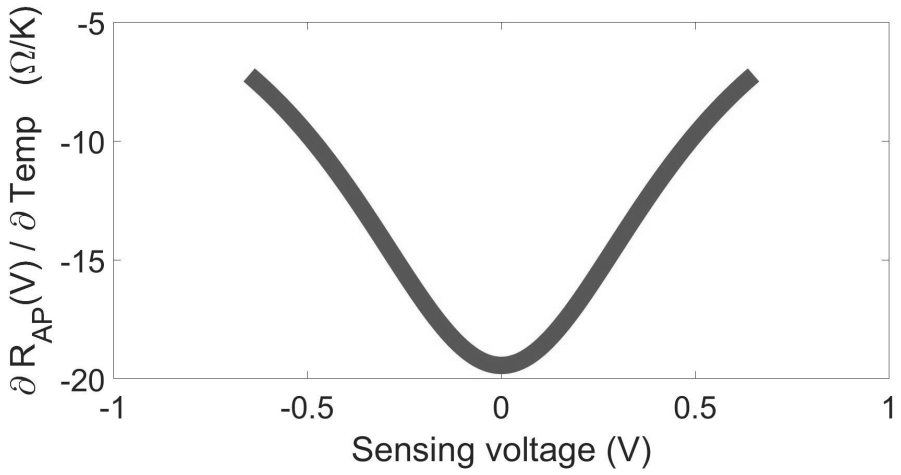
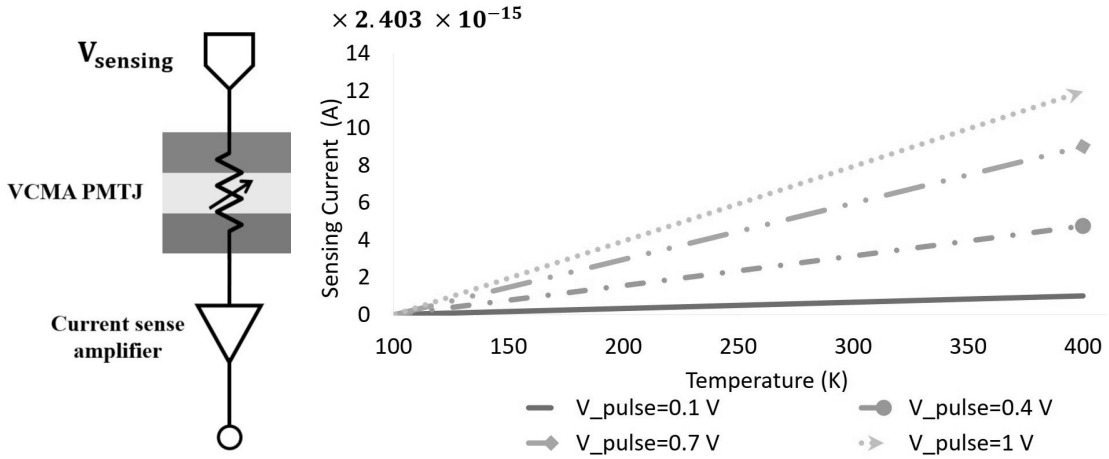


Figure 3.4: Rate of change in the antiparallel resistance with respect to temperature and MTJ sense voltage

3.4 Thermal sensor

The thermal sensor is shown in Figure 3.5(a). The current amplifier enhances the sensitivity and amplifies any changes in the sense currents. R_{AP} can be measured as $R_{AP} = V_{sensing}/I_{sensing}$.

VCMA-based MTJs can be used as a thermal sensor by applying a voltage pulse, and measuring the device resistance. The rate of change in the MTJ resistance with respect to temperature at around 0.2 volts is approximately 16 Ω per degree Kelvin, as illustrated in Figure 3.4. For multiple sense voltages, the rate of change in the normalized output current at an ambient temperature is shown in Figure 3.5(b). For typical integrated sensors, a nominal output of 298 μA at 300 K with a change of 1 μA per degree Kelvin is exhibited.



(a) (b)

Figure 3.5: Proposed thermal sensor, (a) thermal sensor circuit, and (b) thermal sense current for different sense voltages

An MTJ can be integrated above the CMOS device layer, providing high accuracy and local thermal access. An array of MTJs can be placed in series to increase the sensitivity of the sensor. Adding a distributed network of MTJs above the CMOS circuit blocks supports local monitoring of the system temperature. In addition, MTJ leaks almost zero current. The cost in standby power to support a thermal aware network is therefore low as compared to semiconductor-based thermal sensing techniques.

3.5 Summary

A method is proposed for using an MTJ as a thermal sensor. The technique considers the effects of temperature and voltage on the thermal stability and resistance

of an MTJ. Physical parameters, based on experimentally fabricated devices, are used to characterize the behavior of the MTJ. The MTJ can operate as a thermal sensor with a change in device resistance of up to 16 Ω per degree Kelvin at a sense voltage of 0.2 volts. A network of distributed MTJs can be used to efficiently monitor complex integrated systems to dynamically control local operation based on thermal and power constraints.

Chapter 4

Spintronic/CMOS-Based Thermal Sensors

Conventional methods for allocating a small number of integrated thermal sensors is insufficient to fully monitor the thermal behavior of a large scale system [97, 98]. CMOS-based thermal sensors exhibit low sensitivity and an exponential relationship with temperature, complicating the process of estimating the ambient temperature [99, 100]. Hence, to increase sensor accuracy and sensitivity, additional computational blocks, such as amplifiers, A/D converters, and look-up tables, should be added to determine the precise temperature [101, 102]. This complexity makes it difficult to distribute hundreds of thermal sensors across an integrated system.

In this chapter, hybrid spintronic/CMOS based thermal sensors are proposed as the backbone of a next generation thermal aware system. The proposed thermal sensing circuits exhibit small area, high sensitivity, and low power. These circuits operate over a wide temperature range with high sensitivity and linearity, making

the proposed hybrid spintronic/CMOS circuits competitive with CMOS-only thermal sensors.

The chapter is organized as follows. The thermal influence on spintronic and CMOS devices is described in Section 4.1. The proposed thermal sensors are reviewed in Section 4.2. A comparison between the spintronic/CMOS-based thermal sensors and CMOS-only thermal sensors is presented in Section 4.3, followed by the summary in Section 4.4.

4.1 Temperature effects on the resistance of MTJ and CMOS devices

Resistance describes the movement of electrons from one atom to another under the influence of an electric field. The movement of electrons is characterized by interactions and scattering within the device material, which causes the devices and interconnects to heat up. The heat generated from these interactions changes the ambient temperature, causing the atoms to vibrate at a higher rate. The higher the temperature, the more violently the atoms vibrate, affecting the resistance of the electronic devices.

To simplify the influence of temperature on the resistance of an electronic device, two types of devices are considered as thermal resistors with a temperature coefficient of resistance TCR, as described by the following expression,

$$\frac{R - R_{Ref}}{T - T_{Ref}} = \frac{dR}{dT} = TCR \times R_{Ref}, \tag{4.1}$$

where T_{Ref} is a reference temperature at 0^oC, and R_{Ref} is the resistance at T_{Ref}. TCR describes an absolute measure of the relative change of a thermal resistor to a change in temperature.

The conductance of an MTJ is composed of two primary components, a spin-dependent (elastic) conductance and a spin independent (inelastic) conductance. Both conductance components exist in the parallel and antiparallel states but with different contributions. The MTJ conductance is due to a mixture of different mechanisms such as hopping dependent tunneling, magnon assisted tunneling, phonon assisted tunneling, and direct spin polarized tunneling [90,103,104]. An increase in temperature causes an increase in the number of impurities which enhances the hopping mechanism, thereby raising the hopping conductance [105]. The temperature influences the magnetoresistance of the ferromagnetic layers of the MTJ and the magnetism of the interface between the ferromagnetic layers and the insulator [105–107]. Most of the conductance in the MTJ parallel state is direct spin polarized conductance, which is less affected by temperature than other types of conductance mechanisms.

The antiparallel resistance of an MTJ is more sensitive to temperature than the parallel state, since most of the conductance mechanisms contributing to the antiparallel state are an inelastic conductance which exhibits a high temperature dependence [107–109]. A compact physical model of an MTJ based on experimentally fabricated devices is described in this chapter to characterize the electrical and magnetic behavior of an MTJ [93, 94, 109]. As an example, an MTJ with a sense voltage of 0.8 volts exhibits a $TCR|_{MTJ}$ of $-8 \times 10^{-5} 1/^{o}C$ and a linearity R^2 of 0.99999.

The temperature dependence of a CMOS transistor depends upon whether the device is operating in the linear or saturation region. The linearity and sensitivity of the change in drain-to-source resistance of a CMOS transistor to temperature are illustrated in Figure 4.1. As an example, a CMOS transistor operating in saturation with $V_{GS} = 0.45$ volts and $V_{DS} = 0.45$ volts exhibits a $TCR|_{Transistor}$ of $53 \times 10^{-4} 1/^{o}C$ and a linearity of 0.9992.

A saturated transistor features a high linearity of up to 1 and a sensitivity approaching 600 ohms/K while an MTJ exhibits a negative sensitivity approaching -4 ohms/K but with higher linearity. The sensitivity of an MTJ can be controlled by changing the size, bias point, or material [109, 110]. In this chapter, an MTJ is biased with a CMOS transistor, where the thermal influence of both devices compensates each other to achieve a thermal sensor with high sensitivity and linearity. The proposed circuits are presented in the following section.

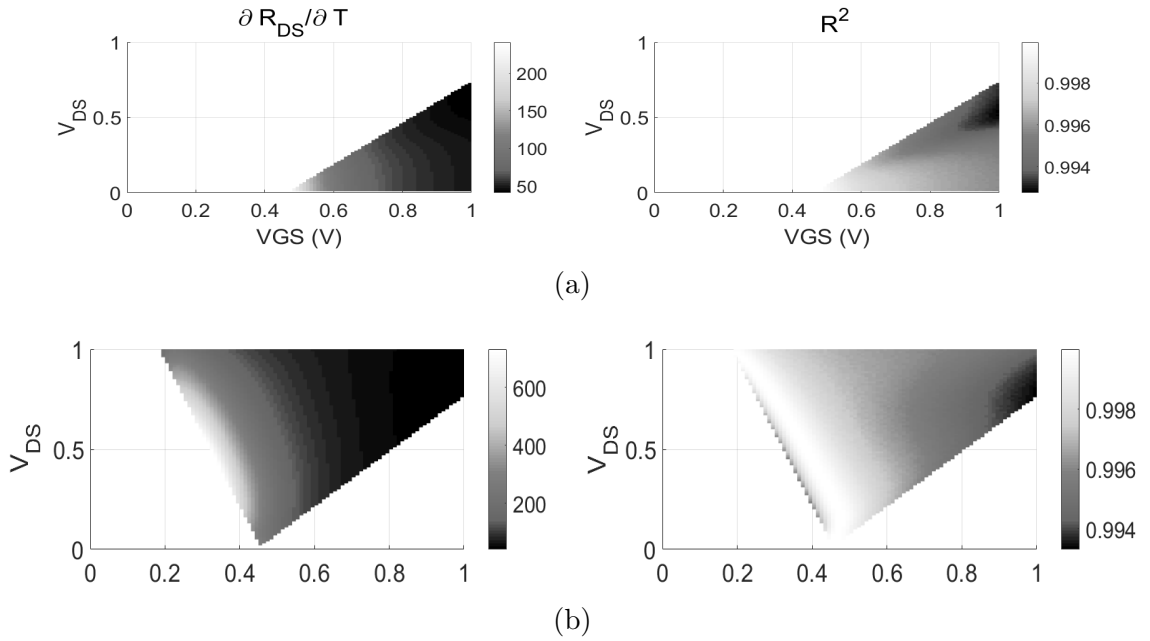


Figure 4.1: Linearity and sensitivity of a CMOS transistor at different bias conditions, a) linear region, and b) saturation region

4.2 Proposed spintronic/CMOS-based thermal sensors

MTJ devices are combined with CMOS to provide an efficient temperature sensor. The sensing technique considers the effects of temperature and sense voltage on the thermal stability and resistance of an MTJ. An MTJ exhibits higher sensitivity in the AP state than in the P state [92, 103, 111]. Hence, an MTJ is designed to operate as a thermal sensor in the stable AP state despite fluctuations in operating temperature and supply voltage.

The thermal stability Δ of an MTJ determines the limits of the applied voltage and range of temperature over which the device can stably operate without switching [109]. Δ, as shown in Equation 3.15, is the ratio of the magnetization energy of an MTJ and the thermal perturbation to the system, which is a function of temperature and applied voltage. Fluctuations in the sense voltage when switching an MTJ is related to the critical switching voltage of an MTJ as discussed in subsection 3.1.3,

CMOS-only thermal sensors exhibit an exponential relationship with temperature which complicates the measurement process. To exploit the capabilities of both an MTJ and CMOS transistor to sense temperature, both of these devices are used within the same circuit. Two different MTJ/CMOS-based thermal sensing circuits are proposed in this chapter, as shown in Figure 4.2. These circuits are described in terms of the thermal sensitivity and linearity characteristics in, respectively, Sections 4.2.1 and 4.2.2.

4.2.1 Circuit I, Hybrid-I MTJ/transistor

Circuit I, Hybrid-I is composed of a transistor and an MTJ, where the output voltage V_{out} (the drain-to-source voltage) exhibits a linear relationship between the output voltage of the circuit and the temperature. Note the decrease in the MTJ resistance and increase in the transistor resistance with temperature; both devices

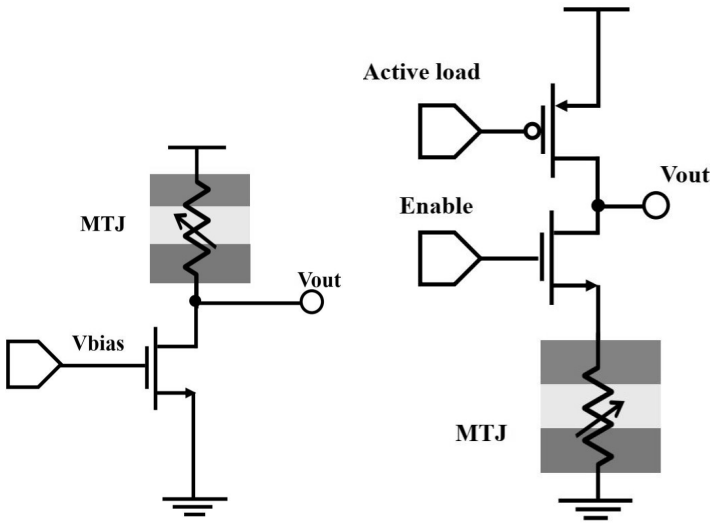


Figure 4.2: Proposed CMOS/MTJ thermal sensors, a) Hybrid-I MTJ/transistor, and b) Hybrid-II MTJ/transistor with an active load

compensate each other and hence the output voltage exhibits a linear relationship with temperature.

The output voltage V_{out} of Circuit I is described by the following expression,

$$V_{out1} = \frac{V_{DD} R_{DS}(T)}{R_{DS}(T) + R_{MTJ}(T)}, \tag{4.2}$$

where R_{DS} and R_{MTJ} are, respectively, the drain-to-source resistance of the transistor and the resistance of the MTJ. As previously mentioned, a change in the resistance of both the transistor and MTJ with temperature compensates each other, making the output voltage directly proportional to the temperature.

The linearity and sensitivity of the output voltage of Hybrid-I under different bias conditions are illustrated in Figure 4.3. The circuit achieves a sensitivity of up to 1.2 mV/K and a linearity of almost 1. These simulations are based on the predictive transistor model (PTM) for CMOS transistors [112]. The CMOS transistors are 32 nm $\times$ 16 nm [112], and the 16 nm PTM model parameters are based on BSIM-CMG [112].

Although Hybrid-I exhibits high linearity and sensitivity, note the circuit behavior under thermal or bias fluctuations. The MTJ in Hybrid-I provides a stable feedback system, since any decrease in R_{MTJ} with an increase in temperature will raise the drain voltage, maintaining the transistor within the saturation region. As previously mentioned, operating a transistor in the saturation region enhances the linearity and sensitivity of the circuit. Hybrid-I cannot however maintain stable operation under fluctuations in the supply voltage. Hybrid-II is proposed to overcome this disadvantage, as described in the following subsection.

4.2.2 Circuit II, Hybrid-II

Circuit II, Hybrid-II is composed of an NMOS transistor, PMOS transistor, and an MTJ. The PMOS transistor is biased at a DC operating point, labeled as Active load, as illustrated in Figure 4.2-b. Hybrid-II provides greater control on the bias voltage by changing the state of the active load. In addition, the NMOS, PMOS, and MTJ

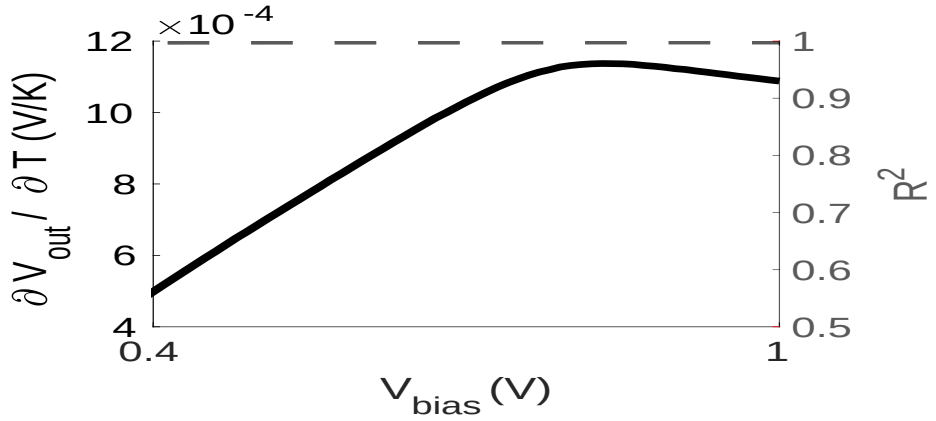


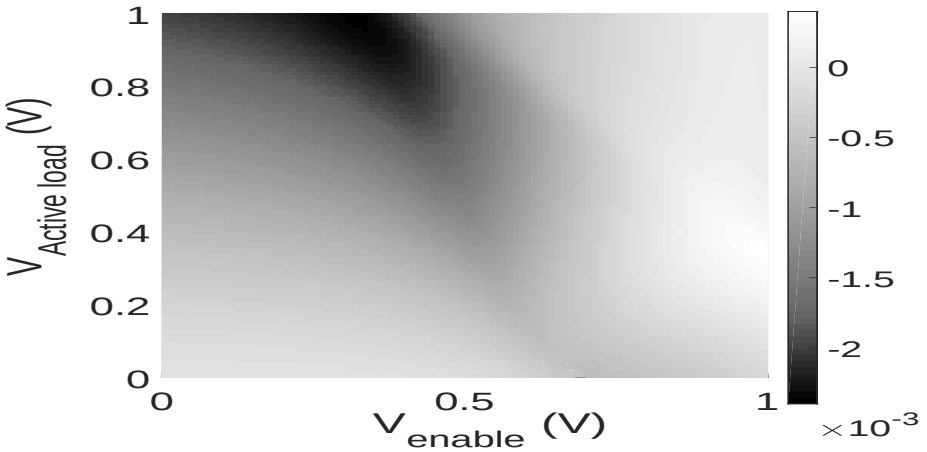
Figure 4.3: Thermal performance of the Hybrid-I circuit; sensitivity $\partial V_{out}/\partial T$ (solid line) and linearity R^2 (dotted line)

devices contribute to the relationship between the output voltage and temperature, which affects both the linearity and sensitivity of the circuit. The resistance of both the NMOS and PMOS transistors increases with temperature, while the MTJ resistance decreases. The nonlinear increase in the resistance of the NMOS and PMOS devices compensates the linear decrease in the resistance of the MTJ, maintaining a linear relationship with temperature, as expressed by (4.3). The output voltage of the circuit is described by the following expression,

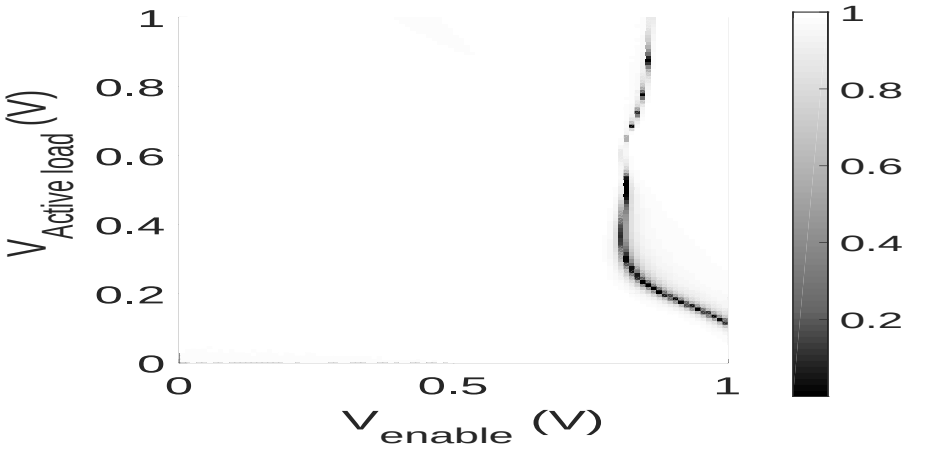
$$V_{out2} = \frac{V_{DD} R_{DS1}(T)}{R_{DS1}(T) + R_{DS2}(T) + R_{MTJ}(T)}. \tag{4.3}$$

The variation in linearity and sensitivity of Hybrid-II under different bias conditions is illustrated in Figure 4.4. The circuit exhibits a sensitivity of up to 2.2 mV/K with a linearity approaching 1. Over a wide range of bias conditions, Hybrid-II maintains high

linearity and sensitivity, making the circuit a good candidate to overcome fluctuations in the bias voltage. The dark area in Figure 4.4-a describes where the circuit exhibits the highest sensitivity and linearity.



(a)



(b)

Figure 4.4: Thermal performance of the Hybrid-II circuit, a) sensitivity $\partial V_{out}/\partial T$, and b) linearity R^2. The dark area in (a) describes where the circuit exhibits the highest sensitivity and linearity.

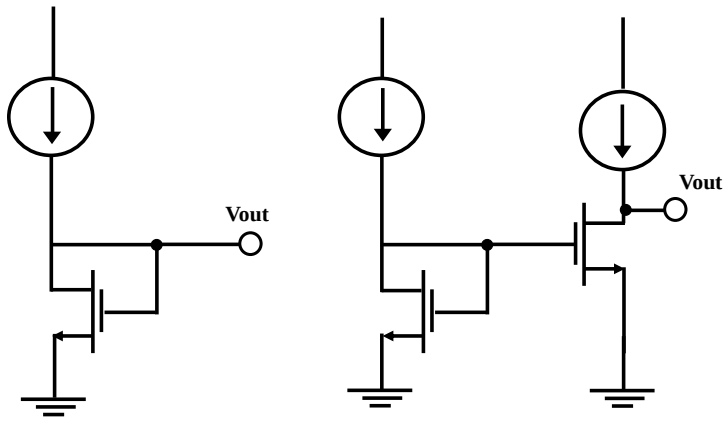
To provide further insight into the superiority of the proposed circuits, a comparison between four thermal sensors (a diode connected transistor, two paired transistors, Hybrid-I, and Hybrid-II) is characterized in terms of the sensitivity, linearity, power consumption, and area, as presented in Section 4.3.

4.3 Comparison of thermal sensors

The proposed circuits benefit from the influence of temperature on the transistor parameters (such as the threshold voltage, mobility, saturation velocity, gate tunneling current, tunneling and recombination current, drain induced barrier lowering, impact ionization, and body effect) and the MTJ antiparallel resistance. The proposed MTJ/CMOS-based thermal sensors are illustrated in Fig. 4.2 while the CMOS-only circuits are illustrated in Figure 4.5. A comparison of the four different circuits clarifies the advantages of the hybrid thermal sensor composed of an MTJ with CMOS.

Circuits CMOS-I and CMOS-II are CMOS-only thermal sensors, where CMOS-I is a diode connected thermal sensor biased by a current source, and CMOS-II is the same as CMOS-I followed by a common source amplifier. A comparison of these sensors is listed in Table 4.1. The CMOS transistors are sized the same (32 nm $\times$ 16 nm) and biased at the same current (17 μA) to establish a fair comparison.

For the two CMOS thermal sensors, these circuits exhibit good sensitivity with reasonable linearity. In Hybrid-II, the two CMOS transistors and MTJ behave



(a) (b)

Figure 4.5: CMOS-only sensors, a) CMOS-I, diode connected transistor, and b) CMOS-II, two paired transistors

Table 4.1: Comparison of the proposed temperature sensor and conventional CMOS sensors in terms of sensitivity, linearity, power consumption, and area

		CMOS-I	**CMOS-II**	**Hybrid-I**	**Hybrid-II**
Sensitivity (mV/K)	**Commercial (0 to 85)**	0.51	0.51	0.4	1.91
	Industrial (-40 to 100)	1.03	1.03	0.64	3.78
	Automotive (-40 to 125)	1.08	1.08	0.77	3.97
	Military (-55 to 125)	1.35	1.35	0.81	4.8
Linearity	**Commercial (0 to 85)**	0.985	0.985	1	0.983
	Industrial (-40 to 100)	0.953	0.953	0.999	0.96
	Automotive (-40 to 125)	0.941	0.941	0.999	0.947
	Military (-55 to 125)	0.919	0.919	0.996	0.936
Power Consumption at 27^{o}C (μW)		40	80	18	11.9
Area (μm^2)		4X	8X	1X	2X

as temperature sensor elements. Hybrid-II exhibits a higher thermal sensitivity than Hybrid-I. In terms of power consumption, Hybrid-II exhibits the lowest power consumption. The MTJ/CMOS thermal sensor requires less area since no current

source is required. CMOS-II exhibits a higher linearity than CMOS-I but requires more area. Based on Table 4.1, Hybrid-II provides appropriate capabilities for a system requiring a large number of on-chip distributed temperature sensors.

4.4 Summary

Two hybrid spintronic/MTJ thermal sensors are proposed in this chapter. These circuits are based on a magnetic tunnel junction which exhibits a thermal sensing capability with a linearity up to 0.983 and a thermal sensitivity of 1.91 mV/K over a wide range of operational temperatures while consuming low power (32 μW). Incorporating an MTJ with a CMOS transistor exhibits a sensitivity of 3.78 mV/K and a linearity approaching 1 while consuming only 11.9 μW during the on-state over a temperature range of -40 to 125^oC. The proposed hybrid spintronic/CMOS temperature sensors are appropriate within a next generation thermal aware system composed of hundreds of on-chip distributed thermal sensor nodes.

Chapter 5

Distributed Spintronic/CMOS Sensor Network for Thermal Aware Systems

A thermal aware system can be achieved by distributing a large number of on-chip thermal sensors. These on-chip thermal sensors should be small in size, low power, high speed, temperature sensitive, and accurate over a wide temperature range. The on-chip thermal sensors should be appropriately placed to capture local hot spots. The location of the thermal sensors depends upon the sensor characteristics, system requirements, IC package, and cooling techniques [113].

A small number of thermal sensor nodes are typically located around an IC, particularly near potential hot spots to support a thermal aware system. For instance, Intel utilizes one thermal sensor per core in the Xeon 5400 series [97], while 25 thermal sensors are embedded within the IBM POWER6 processor [98]. The use of a few thermal sensors, however, limits the ability to fully monitor the significant spatial

and dynamic temperature variations across an integrated system [114]. Thermal aware systems manage the locally distributed thermal sensor nodes around an IC, dynamically controlling the system workload [114–116]. These systems, however, utilize a software-based management system which do not respond to individual thermal sensor nodes. In addition, the response time of these software solutions is long and consumes significant power; hence hardware solutions are desirable. In this chapter, an integrated system to support a thermal aware capability, shown in Figure 5.1, is proposed, where multiple thermal sensor nodes are distributed across an IC.

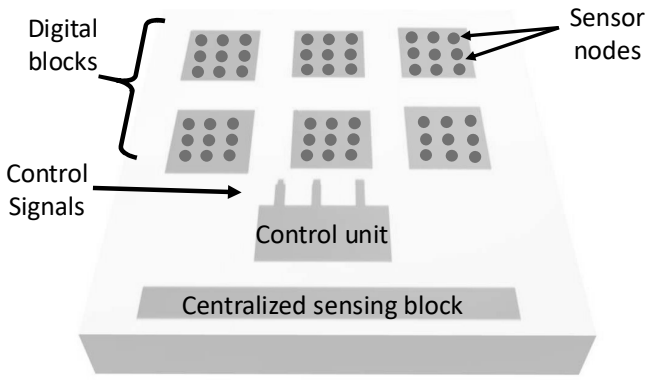


Figure 5.1: Distributed thermal network system

The distributed thermal sensor nodes communicate with a centralized sensing unit which collects temperature information from the individual sensor nodes, producing a thermal map of the system. A hybrid spintronic/CMOS-based analog thermal sensor is proposed here where the high temperature sensitivity of the magnetic tunnel junction (MTJ) antiparallel resistance is exploited. The sensor output is compared with a

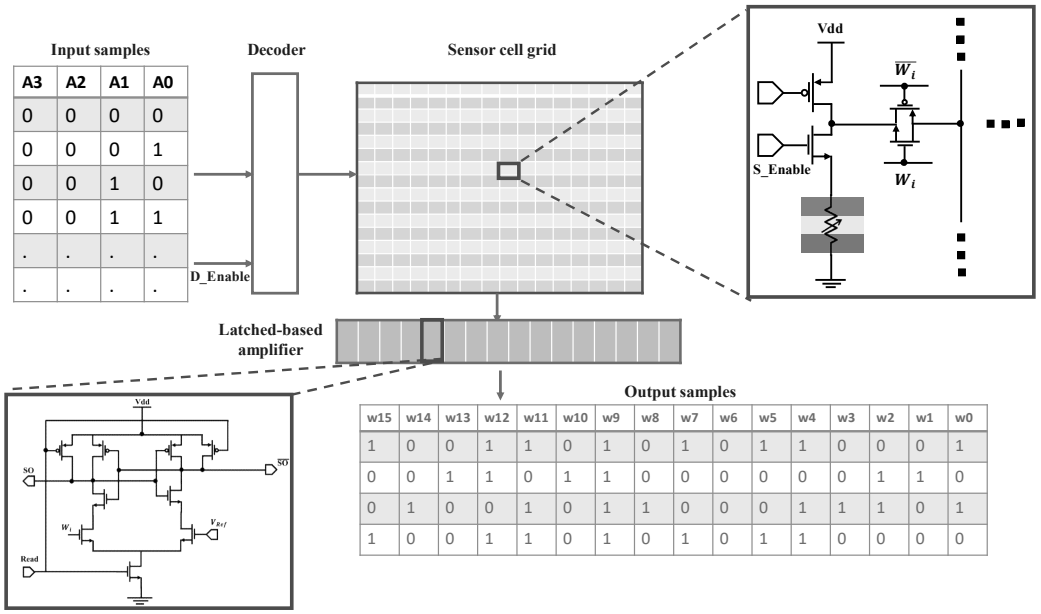


Figure 5.2: Proposed thermal aware system. The system input chooses the row being read through a decoder. The decoder enables the transmission gate of the sensor cell to the read line. The read lines are connected to a latched-based amplifier which produces the system output.

reference source, as shown in Figure 5.3 [117]. The analog thermal sensor behaves as a threshold temperature-based sensor, triggering a signal if the temperature (or voltage) exceeds a certain reference temperature (or voltage).

Several papers discuss thermal sensors using spintronic technology [27, 118]. In [27], a patent describes the use of an MTJ as a thermal sensor by sensing the change in the resistance of an MTJ to temperature. [27] does not describe a thermal sensor, guidelines for using an MTJ as a thermal sensor, or the distinctive behavior of the P and AP resistance of an MTJ to temperature. In [118], the influence of temperature on the probability of device switching is noted. Sensing a change in the switching

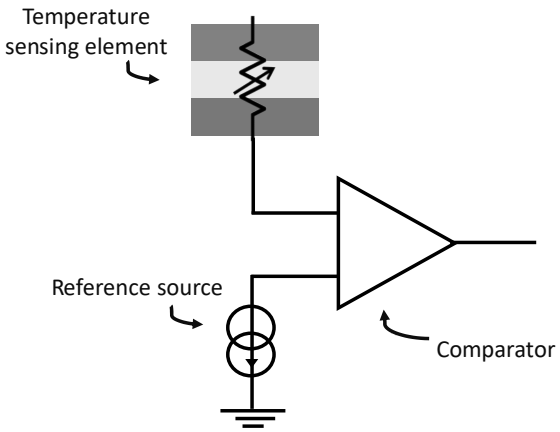


Figure 5.3: On-chip analog thermal sensor

probability requires additional circuitry. In this work, the temperature is measured by a change in the antiparallel resistance.

The proposed system includes a network of thermal sensor nodes distributed around an IC and additional circuitry, described in section 5.1, that manages and controls the sensor signals and hence the system performance, as schematically shown in Figure 5.1. The chapter is organized as follows. The proposed thermal aware system is described in section 5.1, where the system architecture and circuit requirements are discussed. Simulation results are presented in section 5.2 followed by the summary in section 5.3.

5.1 Distributed thermal network

The proposed thermal aware system is a network of thermal sensor nodes communicating with a control unit that collects temperature data and produces a thermal

map. This thermal network provides the monitored system with dynamic real-time thermal information. The proposed system architecture, read and data signaling, and related circuitry are discussed below. The system components are described in subsection 5.1.1, the system signaling is illustrated in subsection 5.1.2, and the fabrication characteristics of the system are reviewed in subsection 5.1.3.

5.1.1 System architecture

The proposed system architecture, shown in Figure 5.2, is managed as a memory grid, where the sensor nodes are organized in a grid-based topology. To read a system of $m \times n$ sensor nodes with m columns and n rows, a $log_2\ n - to - n$ decoder and m amplifiers are required. The input to the system decoder identifies the row being read. Each row shares the same enable signal, while each column shares the same bit line. The enable signal, generated from the system decoder, passes the sensor node voltage to the bit line and is read through a sense amplifier. The proposed sense amplifier is latch-based, composed of two inverters controlled by a *Read* signal. The sensor node voltage is compared with a reference voltage that sets a threshold temperature. The system output is in a binary format indicating whether the state of the sensor node is either below or above a threshold voltage.

5.1.2 System read and data signaling

The sequence of operations is as follows. During each read cycle, the enable signal controls the decoder to individually select one row. The output of each cycle is a vector of m sensor node reads. During each cycle, one row is read, and n cycles are required to read n rows. The system input is generated from a counter, and the system output is stored within a memory.

An example of the data signal waveform of a 4×4 data signal is shown in Figure 5.4. The decoder and sensor nodes are enabled by the Enable signal, where the decoder input data are annotated as $A0$ and $A1$. The output of the decoder enables the individual transmission gates. Each transmission gate connects the associated sensor node output to the bit line. The Read signal enables the sense amplifier to latch a bit line. In comparison with a reference voltage, the amplifier output is latched to either high or low. The output signals, $w0$, $w1$, $w2$, and $w3$, indicate the temperature status. By turning the Enable signal off, the system saves energy by isolating the power from the sensor nodes and decoder.

The output of a distributed thermal network composed of 16×16 sensor nodes is illustrated in Figure 5.5b. The binary thermal map, shown in Figure 5.5a, reflects the location of the individual sensor nodes. The thermal map indicates if the temperature is above or below a predefined threshold temperature and hence determines in real-time the location of the critical hot spots.

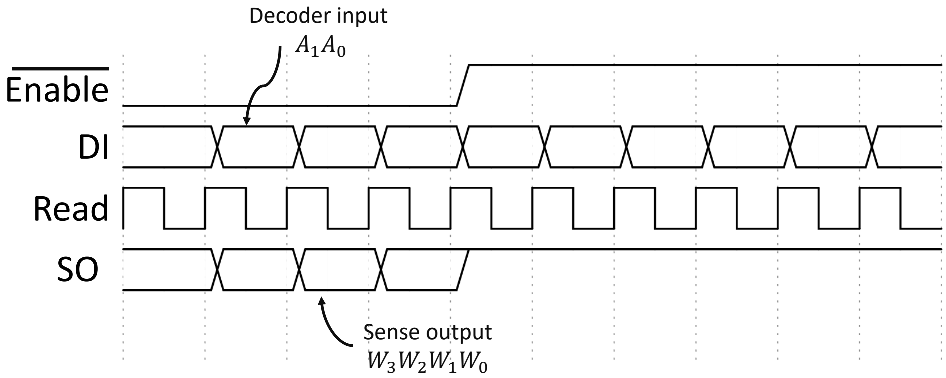
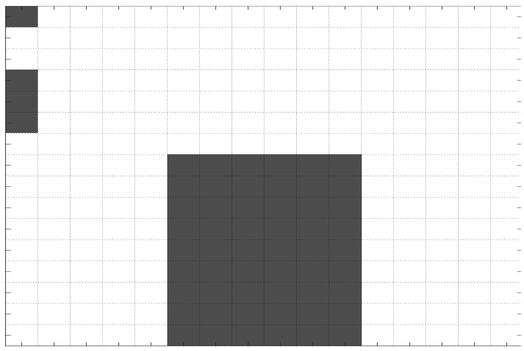


Figure 5.4: System waveforms

1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0

(a)



(b)

Figure 5.5: 16 × 16 thermal map, (a) output sensor node readings, and (b) thermal map. The dark areas represent nodes with a temperature above the temperature threshold.

5.1.3 System characteristics

The proposed system incorporates hybrid spintronic/CMOS devices. The spintronic circuit is based on a magnetic tunnel junction. An MTJ is a structure composed of two ferromagnetic layers separated by an insulator barrier [51]. The resistance of the device is controlled by the difference in the magnetization angle between the two layers. The device exhibits two stable states, a parallel (P) state (where the two layers

are magnetized in the same direction) and an antiparallel (AP) state [54]. The MTJ is combined with CMOS to provide an efficient temperature sensing element. An MTJ/CMOS-based thermal sensor exhibits small size, low power, high linearity, and high sensitivity [119]. These capabilities support a thermal aware system composed of hundreds of distributed thermal sensor nodes.

The MTJ is integrated between the metallic layers above the CMOS device layers, as shown in Figure 5.6, making this structure a good candidate for a local, distributed thermal sensor. MTJ fabrication is sufficiently mature for different technology platforms such as bulk-CMOS, FDSOI-CMOS, and FINFET CMOS [120]. Intel [121], GlobalFoundries [122], Samsung [123], and other large foundries are integrating MTJ technology with CMOS at different technology nodes. These advancements in fabrication can produce high quality MTJs for thermal sensing applications. In addition, MTJ memory can operate over a wide range of temperatures, -40^{o}C to 125^{o}C, in a stable manner for commercial, automotive, and military applications [72]. The ability of MTJ technology to be integrated with CMOS, operate over a wide, stable temperature range, and exhibit almost zero leakage current in the off state, with higher temperature sensitivity than conventional CMOS devices suggests an MTJ/CMOS temperature sensor is an effective candidate for next generation thermal aware systems [119].

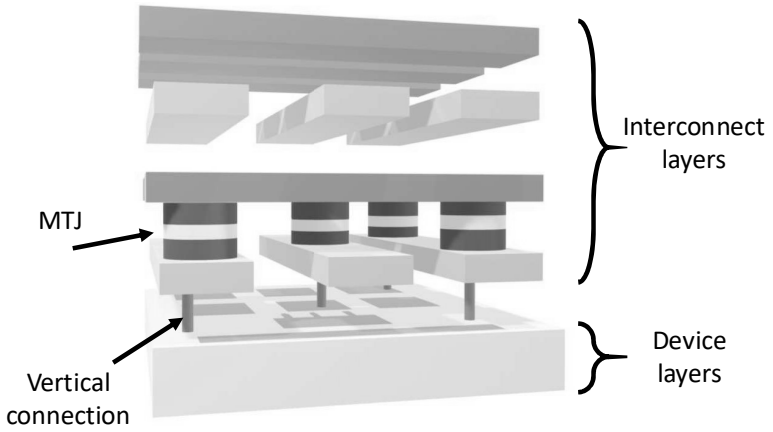


Figure 5.6: MTJ, interconnect, and device layers

The use of an MTJ as a thermal sensor is supported by the high thermal sensitivity of the MTJ antiparallel resistance. The resistance of an MTJ changes almost linearly with high sensitivity with temperature in the antiparallel state (as compared to the parallel state) [92,103,110,111]. The sensitivity of an MTJ to temperature is proposed in multiple MTJ structures such as, CoFeB/Al-O/CoFeB [92,111], Fe/MgO/Fe [110], and CoFeB/MgO/CoFeB [103]. The thermal sensitivity of the MTJ antiparallel resistance depends upon the device material structure, dimensions, and applied sense voltage.

The proposed temperature sensor cell is discussed in the following section. The physical, magnetic, and electrical behavior of an MTJ in addition to the proposed thermal sensor are reviewed. A comparison between the proposed temperature sensor and conventional CMOS sensors in terms of sensitivity, linearity, power consumption, and area is also provided.

5.2 Simulation results

The system operation works as follows. The sense amplifier sets the sensor node voltage. Based on the grid size and number of nodes, preamplifier stages or buffers increase the current, enhance the sensitivity, and isolate the sensor node signal. The signal path of the sensor node to the output, shown in Figure 5.7, is used to characterize system performance. The system characteristics are listed in Table 5.1, where the power consumption includes the energy consumed in the sensor nodes, buffers, inverters, amplifiers, and decoder. The delay of the read operation is the time required to read each of the rows.

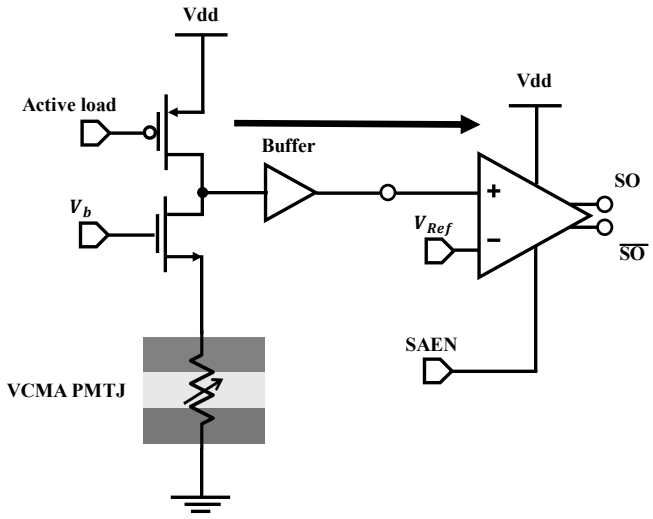


Figure 5.7: Sensor signal path

A read pulse of 1 ns is used to produce an output decision of one sensor node. The comparator delay is 0.03 ns. The accuracy of the system temperature is ±3 K for a reference voltage with an accuracy of ±1 mv. The area of each sensor node is

32 nm × 64 nm where the MTJ layer is between the second and third interconnect layer, as shown in Figure 5.6. The average sense current of a thermal sensor node is 11 μA. The sensor nodes need to be calibrated prior to use due to the influence of manufacturing process variations. Different calibration schemes of multiple on-chip thermal sensors have been proposed [117]. The design, management, and control of these thermal sensors are the foci of this chapter. The ability to fabricate an MTJ with a different antiparallel resistance (thermal sensitivity) has been achieved [110], and additional research is required to enhance the sensitivity of an MTJ to thermal and process variations.

Table 5.1: Characteristics of the proposed distributed thermal network for different grid sizes

System size	**Energy consumption (pJ)**	**Relative path delay to read the grid w.r.t. 4 × 4**	**System size #**	
			Transistors	**MTJs**
4 × 4	1.32	1x	90	16
8 × 8	8.96	2x	304	64
16 × 16	65.50	4x	1,120	256
32 × 32	499	8x	4,980	1024

An example of the system output at three different reference voltages, 300 mV, 304 mV, and 306 mV, mapped to, respectively, threshold temperatures of 332 K, 343 K, and 350 K is shown in Figure 5.8. A multiplexer can be added to switch the reference signal between different voltages to vary the threshold temperature of the sensor nodes.

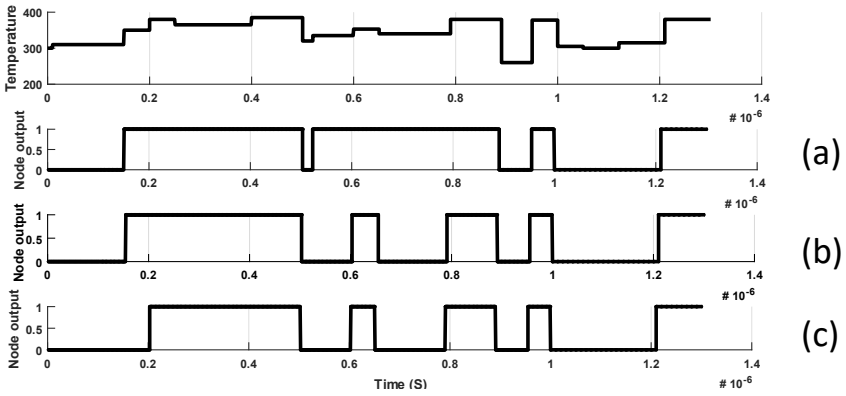


Figure 5.8: System output, (a) V_{ref} = 300 mV which maps to a threshold temperature of T = 332 K, (b) V_{ref} = 304 mV and T= 343 K, and (c) V_{ref}= 306 mV and T = 350 K

A comparison between the proposed hybrid CMOS/MTJ thermal sensor and [118] in terms of system requirements, sensing scheme, energy, read accuracy, and temperature range is listed in Table 5.2. The primary purpose of this chapter is to describe a hybrid MTJ/CMOS-based thermal sensor and a related thermal aware system. A distributed thermal sensor network able to provide an updated spacial and temporal thermal map in real-time is also described in this chapter.

The proposed system provides flexibility in choosing a threshold temperature. The system can also support a multi-threshold sensing scheme. This capability can be achieved by multiplexing the reference voltage. At each reference voltage, the system identifies whether the temperature at a sensor node is above or below a certain threshold temperature. As an example, with two different reference voltages, the system could identify the temperature at a sensor node within three different temperature regions (below T1, between T1 and T2, or above T2) [124].

Table 5.2: Comparison between the proposed CMOS/MTJ thermal sensor and [118]

	Proposed thermal aware system	**[118]**
Sensing Scheme	Change in AP resistance to temperature	Change in probability of switching
Sensor node	One MTJ, two transistors, V_{dd}	Two MTJs, two transistors, V_{dd}, I_{bias}
System Requirement	Latch-based amplifier	Circuit to map probability of switching an MTJ to temperature
Energy	0.5 nJ (To read network of 32 × 32 cells)	8.5 nJ
Accuracy	3 K	1 K
Output	1 or 0 indicating above or below threshold temperature	Local temperature

With hundreds of on-chip thermal sensor nodes distributed across a system, the ability to monitor local heat (characterizing the generated heat and thermal paths) is achieved. This capability for real-time spacial and temporal sensing provides significant information characterizing the thermal behavior which can be used to mitigate on-chip heat generation and distribution issues.

5.3 Summary

The need for a thermal aware system increases with device scaling and the size of the integrated system. A thermal aware system is proposed where a grid structure is composed of individual thermal sensor cells. The sensor nodes are based on hybrid spintronic/CMOS technology, where the antiparallel resistance of a magnetic tunnel

junction exhibits a thermal linearity of 0.9 and thermal sensitivity of 4.8 mv/K over a temperature range of -55 ^{o}C to 125 ^{o}C. A system of 1,045 thermal sensors distributed in a 32 $\times$ 32 grid structure consumes approximately 500 pJ. This low energy and high sensitivity are appropriate for next generation thermal aware systems.

Chapter 6

Double Magnetic Tunnel Junction Multi-Bit Memory Logic for *in situ* Nonvolatile Computing

Many integrated systems have become data centric, where a huge amount of data are collected and processed in real-time. Processing "big data" and exascale computing with 10^{18} floating point operations per second is not achievable with conventional computing architectures. Conventional von Neumann architectures, where the memory is separate from the processing elements, struggle despite advanced memory solutions. In data centric architectures, data motion is greatly decreased by integrating the computational process within the storage system at different levels of the memory and storage hierarchy. This capability for *in situ* computation can be achieved by considering an emerging memory technology that exhibits two modes of operation within the same platform, memory mode and compute mode.

The separation between memory and computing expends a significant amount of energy and space. These systems are volatile and leak significant current. Non-volatile memory (NVM) has been proposed to replace CMOS memory within different parts of the memory hierarchy. Some of these NVM solutions support in-memory computing, such as memristor-based logic [30] and spintronic-based compute-in-memory [21]. Molecular memristors (e.g., titanium dioxide memristors) exhibit a low endurance rate (up to 10^{10} cycles [125]) as compared to the high endurance characteristic of spintronic systems (10^{15} write cycles [126]). This higher endurance makes spintronic memristors a more effective solution for compute in-memory applications.

Magnetic random access memory (MRAM) is a spintronic NVM, considered as a possible solution at all memory hierarchies. This breadth is supported by the development of multiple magnetic memory technologies, serving each level of the memory hierarchy, such as magnetic tunnel junctions (MTJ), domain wall motion devices, spin orbital torque (SOT) MTJ, magnetic skyrmions, and topological insulator/ferromagnetic memory [127]. MRAM exhibits a retention time of almost ten years with high endurance rates, high speed, small size, and CMOS compatibility. While most of these spintronic MRAM solutions target in-memory computing [127], these systems are large in size and only support a one bit memory cell. In this chapter, a two bit memory cell with a nonvolatile AND, OR, and NOT logic gate is proposed.

Most MRAM solutions are based on a perpendicular MTJ (PMTJ) with spin transfer torque (STT). This technology has attracted considerable attention due to the high endurance rate, fast switching, CMOS compatibility, simple device structure, and ability to scale to sub-10 nm dimensions [128]. STT MRAM, however, requires a high critical current density to switch a device. Consequently, the memory cell is scaled to satisfy both density and power demands. These PMTJ devices, however, suffer from aging and low endurance. Hence, a multi-level cell is proposed here to further increase the memory density. A double magnetic tunnel junction (DMTJ) is a multi-level STT PMTJ cell composed of two serially connected PMTJ devices. The DMTJ is manufactured in a vertical stack, with an area comparable to a single PMTJ. The DMTJ device exhibits four stable resistance states, 00, 01, 10, and 11, where the most significant bit (MSB) represents the resistance state of the larger PMTJ device, and the least significant bit (LSB) represents the resistance state of the smaller PMTJ. 0 and 1 represent the resistance state of a PMTJ device in, respectively, the parallel and antiparallel state.

The contribution of this chapter lies in two aspects. First, a write circuit for the DMTJ-based STT PMTJ is proposed. Second, a nonvolatile AND, OR, and NOT logic gate based on the multi-level MTJ is presented. This logic gate is described by a state diagram and the physical operation of the DMTJ device. In this chapter, the DMTJ structure is combined with CMOS to provide a hybrid multi-bit memory cell

and a nonvolatile logic element. The chapter is organized as follows. In section 6.1, background on recent compute in-memory solutions utilizing different nonvolatile memory technologies is discussed. The structure and physical model of the DMTJ device are described in section 6.2. The DMTJ-based multi-bit memory cell is presented in section 6.3, where a hybrid CMOS/DMTJ read/write circuit is also proposed. The DMTJ-based nonvolatile AND, OR, and NOT gate is described in section 6.4. Simulation results are presented in section 6.5. A comparison between the proposed work and earlier approaches is offered in section 6.6. The chapter is summarized in section 6.7.

6.1 NVM-based Logic

NVM is based on an emerging memristive device that exhibits a hysteresis characteristic, acting as a state machine. The current state of a memristive device is maintained unless a perturbation (e.g., voltage, current, magnetic field, or electric field) is applied [58]. Multiple logic functions can be demonstrated by the state diagram of one or multiple connected memristive devices. Several NVM-based logic in-memory systems have previously been proposed [21, 30, 31]. Memristors and spin orbital torque (SOT) devices are often considered as a base element for logic in-memory systems. At least two memristors and up to three clock cycles are required in memristor-based in-memory systems to deliver a functionally complete logical operation [31]. Similar to

the system proposed here, these memristive systems require multiple cycles to perform a logical operation.

MRAM solutions, particularly SOT devices, exhibit a high endurance rate. SOT devices are more frequently considered for logic in-memory solutions than STT devices due to the higher endurance rate of SOT devices and the decoupled read and write paths [21]. Similar to the proposed system, SOT-based logic requires the initial state to be written [129] and the output described by a nonvolatile resistance state which is read by a standard memory read operation [129].

These compute in-memory systems are based on a single bit memory cell. Multi-level STT MTJs (e.g., DMTJ) improve the density of STT MRAM and reduce the cost per bit [130]. Limited work exists on DMTJ STT-based logic in-memory systems. The DMTJ STT in-memory system proposed in [131] is volatile. This structure is achieved by adding circuitry to the sense scheme to support the logical AND, OR, and XOR operation [131]. The drawbacks of this earlier system lie in the need to initially store the input within the DMTJ before calculation, while the output is volatile and not stored. Hence, this system requires additional time and power to perform a nonvolatile logical operation. Alternatively, fully nonvolatile logic based on the DMTJ is proposed in this chapter. A write circuit for the DMTJ-based multi-bit memory cell is also described. A functionally complete multi-bit memory cell and

nonvolatile logical AND, OR, and NOT operations are demonstrated for a 32 nm CMOS technology node.

6.2 Multi-level STT-MRAM cell

An STT PMTJ multi-level cell is non-volatile [132] and exhibits a fast read/write time. The simple cell structure requires only two additional mask steps to integrate the STT storage elements into a logic compatible CMOS process [133]. In this chapter, a DMTJ, composed of two serially connected PMTJs, provides both a multi-bit memory cell and a non-volatile logic cell.

A compact model capturing the static and dynamic behavior of a perpendicular magnetic anisotropy MTJ, including the influence of the device dimensions and temperature on the perpendicular magnetic anisotropy of a PMTJ, is described in [134,135]. A version of this macrospin model is used to support the hybrid MTJ/CMOS circuit design and simulation process. The model considers the influence of the current and temperature on the device tunneling magnetoresistance, layer spin polarization, saturation magnetization, and device interfacial and bulk magnetic anisotropy constant [134, 135].

A DMTJ is a multi-level STT PMTJ cell composed of two serially connected PMTJ devices. The two PMTJs share the same characteristics but with different diameters, D_{MTJ1} and D_{MTJ2}. A DMTJ is a two terminal device modeled as two

variable resistors connected in series, as shown in Figure 6.1. Each PMTJ exhibits two stable resistance states, parallel (0) and antiparallel (1), whereas the DMTJ exhibits four different resistance states, represented as R_{00}, R_{01}, R_{10}, and R_{11}. The MSB represents the state of the large PMTJ, PMTJ2, and the LSB represents the state of the small PMTJ, PMTJ1. In the R_{01} state, PMTJ2 operates in the parallel state and PMTJ1 operates in the antiparallel state. Since PMTJ2 has a larger diameter than PMTJ1, PMTJ2 exhibits a lower resistance than PMTJ1. Consequently, PMTJ2 requires a greater current than PMTJ1 to switch between the parallel and antiparallel states.

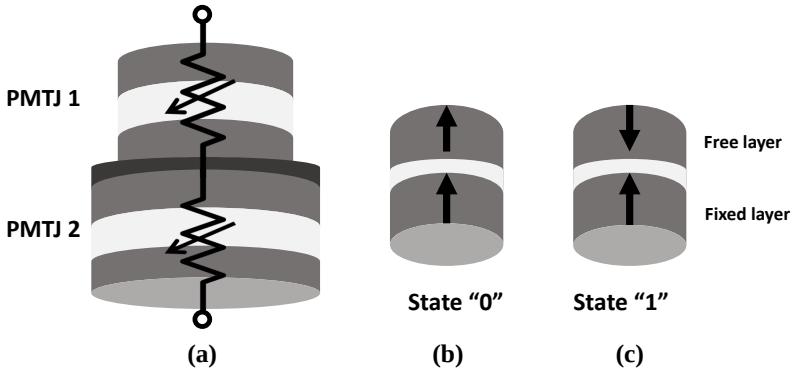


Figure 6.1: Multi-level STT MRAM cell composed of two serially connected PMTJs, (a) two PMTJs connected in series modeled as a variable resistance based on the state of operation, (b) state "0" when the magnetization state of the free and reference layer is in parallel, and (c) state "1" when the magnetization state is in the antiparallel state.

The resistance-current characteristic of a DMTJ based on two serially connected STT PMTJs is shown in Figure 6.2, indicating the four resistance states of the DMTJ and the critical current at which each PMTJ switches. The performance and reliability

of a DMTJ cell are sensitive to CMOS and MTJ device variations and thermal induced randomness. To avoid a read failure (an overlapping distribution of resistances), the size of each PMTJ within the DMTJ is critical. The distribution density of the four resistance states in a DMTJ, shown in Figure 6.3, is based on Monte Carlo simulations of process and temperature variations within a DMTJ device.

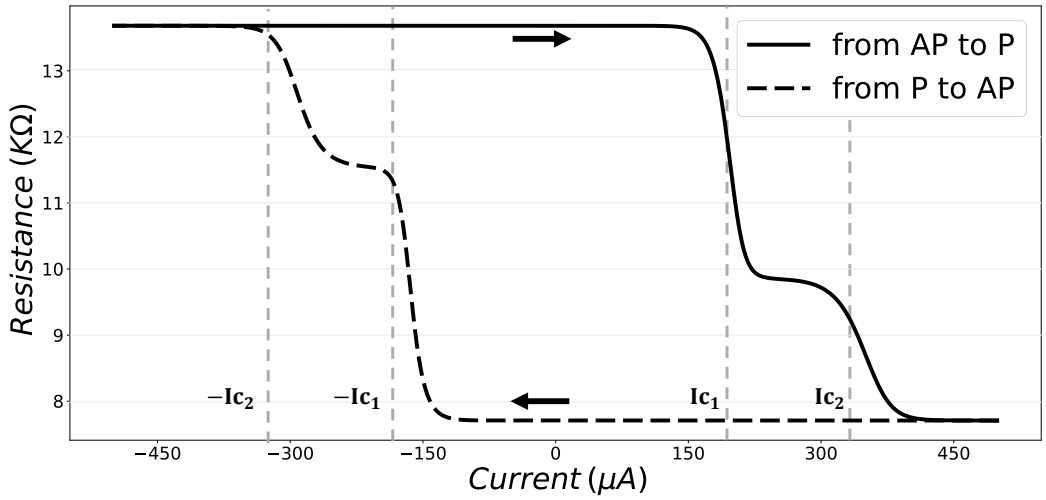


Figure 6.2: AP-P transition and P-AP transition of a DMTJ. The vertical axis is the resistance of the DMTJ at 0 volts, and the horizontal axis is the current to switch a DMTJ

A DMTJ cell has previously been demonstrated with four distinctive resistive states with successful read and write operations [136, 137]. The DMTJ is applicable to a wide variety of applications, supporting high density and low power cache memory [130]. In a previous study [138], a systematic analysis of the sources of variations in a DMTJ STT MRAM and the reliability of the read and write operations are described.

This study concludes that a DMTJ with series connected PMTJs exhibits higher read and write reliability than parallel connected PMTs [138].

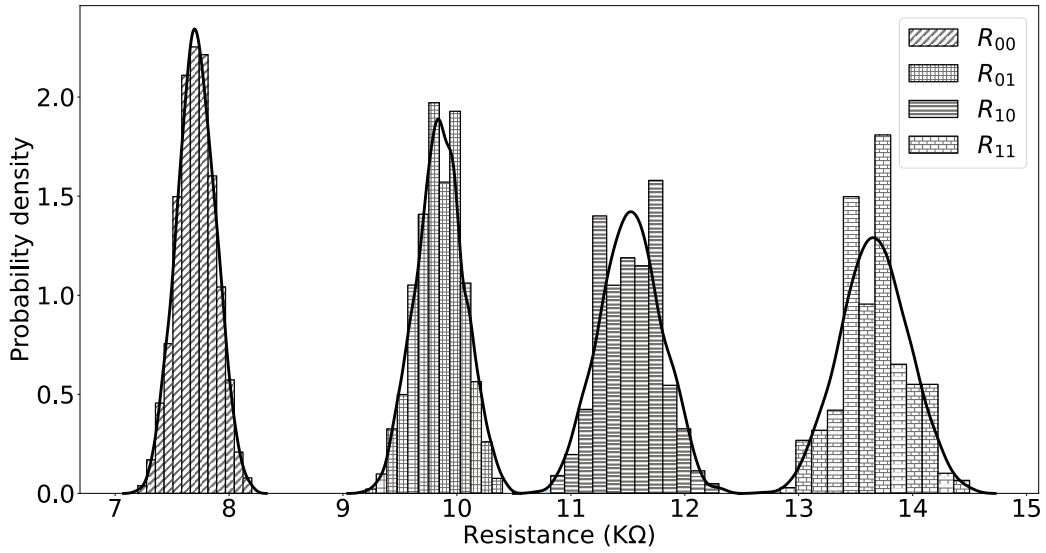


Figure 6.3: Monte Carlo simulation of the four state resistance distributions of a DMTJ with process and temperature variations.

A state flow diagram of a DMTJ is shown in Figure 6.4, where $+I_{c2}$ and $-I_{c2}$, are, respectively, the current to switch the PMTJ2 between the parallel and antiparallel state, and I_{c1} is the current to switch PMTJ1 between the parallel and antiparallel state. I_{c2} is larger than I_{c1}. In the following section, a multi-bit hybrid MTJ/CMOS memory cell is proposed. The circuit requires a two step write scheme and a one step read scheme.

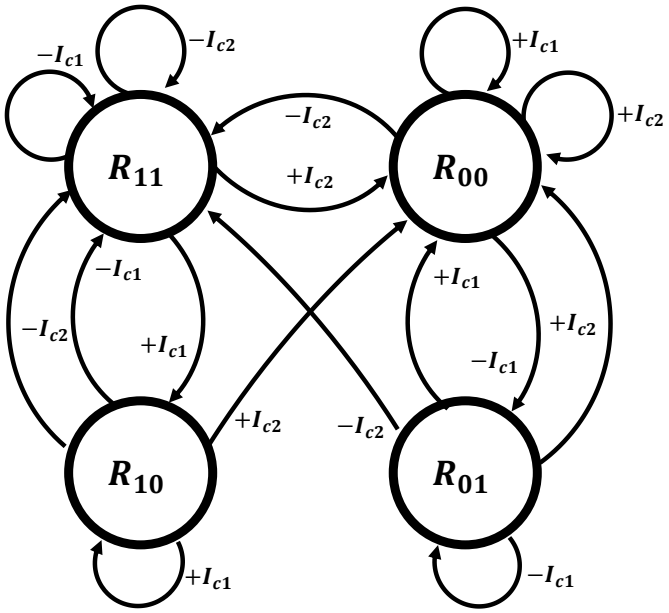


Figure 6.4: State flow diagram of a DMTJ with two serially connected PMTJs. The device provides four resistance states and switches between the states based on the applied current.

6.3 DMTJ STT PMTJ as multi-bit memory cell

To employ a DMTJ as a multi-bit memory cell, the cell should supply bidirectional current and exhibit two different current levels. A circuit commonly known as an H-bridge [139] (see Figure 6.5), was adopted and modified to achieve the write capability of the DMTJ. When switch S_0 is closed, the current supplied to the DMTJ is in the opposite direction than when switch S_1 is closed. The H-bridge has been previously adopted to switch an STT-based MTJ, while in this chapter, an H-bridge circuit is modified to switch a DMTJ, as discussed in subsection 6.3.1.

Based on the current resistance state, a DMTJ transitions to another state in one step except for the transition between the R_{01} and R_{10} resistance states, as illustrated in the state diagram of a DMTJ shown in Figure 6.4. A reset step is initially required, followed by a write step. Since no sense step is used before the write operation, a one step process writes the R_{00} or R_{11} state by applying, respectively, a high positive current pulse or negative current pulse. The two step write process consists of a *reset* step and a *write* step to produce the other two resistance states of the DMTJ. The write and read operations of a DMTJ are described in the following subsections.

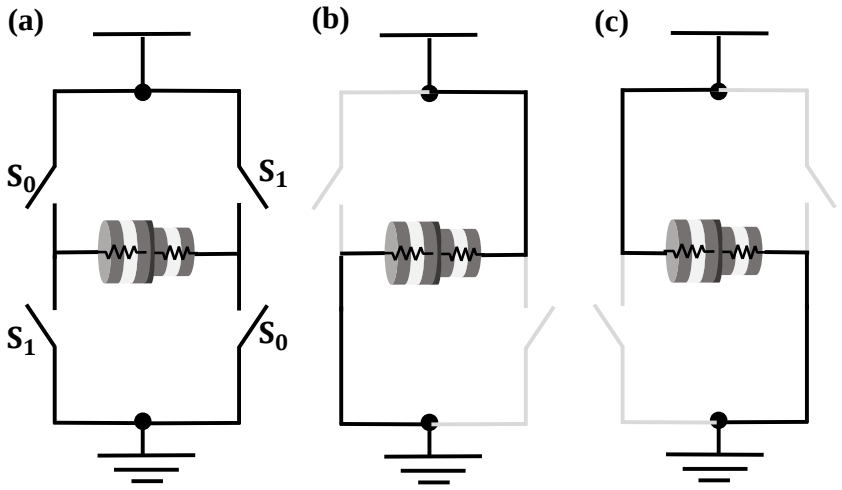


Figure 6.5: H-bridge circuit, (a) switches S_0 and S_1 control the direction of the current within the DMTJ, (b) S_1 is closed, and (c) S_0 is closed.

6.3.1 Write technique

A notation is used to represent the current supplied to a DMTJ, as follows: I_{00} represents $+I_{c2}$, I_{01} represents $+I_{c1}$, I_{10} represents $-I_{c1}$, and I_{11} represents $-I_{c2}$. The

MSB is the direction of the current, positive if zero and negative if one, and the LSB represents the current magnitude, where zero is less current and one is more current. The hybrid CMOS/MTJ multi-bit memory circuit is shown in Figure 6.6 where I_1 and I_0 represent, respectively, the MSB and LSB of each current notation. As an example, if I_{00} is the applied circuit, $I_0 = 0$ and $I_1 = 0$. Hence, the top left CMOS network operates with the bottom right CMOS network. En_W enables the write operation, and En_S enables the sense operation.

Each switch of the write circuit shown in Figure 6.5 utilizes a CMOS network to source the critical current to switch a DMTJ. The right side of the branch includes only one transistor; therefore, if $I_1 = 1$, two CMOS transistors are connected in series with the DMTJ. These transistors are small, providing low current to the DMTJ, sufficient to switch the small PMTJ. If both $I_1 = 1$ and $I_0 = 1$, both sides of the CMOS network produce a large current. Note that the left side transistors within the CMOS network are larger than the right side transistors to source this larger current. The size of the CMOS networks are critically dependent on the four resistance states within the DMTJ and the critical current to switch the PMTJs within the DMTJ.

6.3.2 Read technique

In the proposed multi-bit memory cell, a one step read scheme senses the resistance of the DMTJ [136]. The sense circuit is illustrated in Figure 6.7 where the voltage

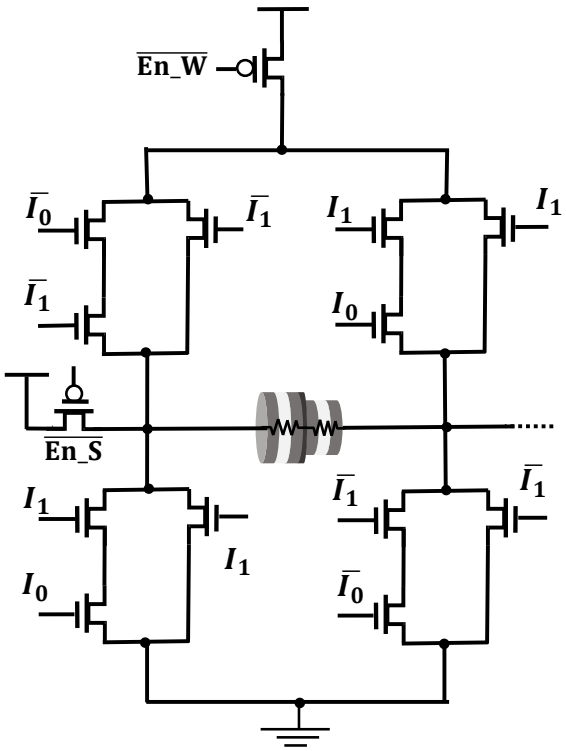


Figure 6.6: Multi-bit DMTJ memory cell based on a multi-level STT PMTJ. En_W enables the write operation, En_S enables the sense operation, and I_0 and I_1 are, respectively, the magnitude and direction of the current supplied to the DMTJ.

across the DMTJ is compared with three different voltage references. The read circuit has three sense amplifiers to simultaneously compare the selected DMTJ cell voltage with the voltage references followed by an encoder that identifies the state of the MSB and the LSB. The sense current is sufficiently small to not switch the DMTJ.

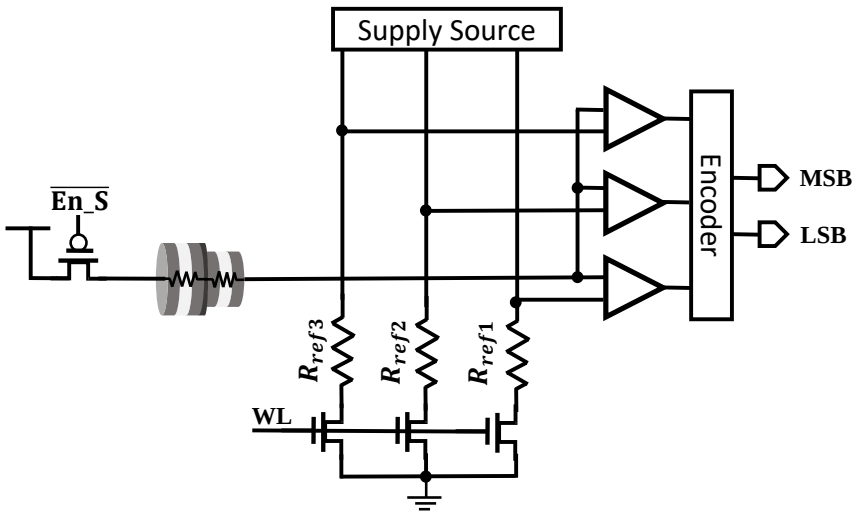


Figure 6.7: One step read scheme for a four level cell memory [136] indicating the output of the DMTJ-based nonvolatile AND, OR, and NOT gate

6.4 DMTJ STT PMTJ as AND, OR, and NOT logic gate

A DMTJ operates as a non-volatile logic cell in addition to a multi-bit memory cell. The behavior of a DMTJ operating as a nonvolatile logic element is illustrated in the state flow diagram shown in Figure 6.4. The DMTJ is treated as a state machine with two one bit inputs and four states. The two one bit inputs, I_1 and I_0, refer, respectively, to the direction and magnitude of the supplied current. The four states refer to the four resistance states of a DMTJ. The truth table of a DMTJ, as a state machine, is listed in Table 6.1. S_1 and S_0 represent, respectively, the present state of the large PMTJ and small PMTJ, and $S_1^{'}$ and $S_0^{'}$ represent, respectively, the future state of the large PMTJ and small PMTJ.

Table 6.1: Truth table of a DMTJ as a state machine

S_1	S_0	I_1	I_0	$S_1^{'}$	$S_0^{'}$
0	0	0	0	0	0
0	0	0	1	0	0
0	0	1	0	0	1
0	0	1	1	1	1
0	1	0	0	0	0
0	1	0	1	0	0
0	1	1	0	0	1
0	1	1	1	1	1
1	0	0	0	0	0
1	0	0	1	1	0
1	0	1	0	1	1
1	0	1	1	1	1
1	1	0	0	0	0
1	1	0	1	1	0
1	1	1	0	1	1
1	1	1	1	1	1

The boolean representation of the future state bits, $S_0^{'}$ and $S_1^{'}$, in terms of the current state of a DMTJ and input current bits, illustrated in Figure 6.8, is respectively,

$$S_0^{'} = I_1 \tag{6.1}$$

$$S_1^{'} = I_1 I_0 + S_1 I_0 + S_1 I_1. \tag{6.2}$$

A DMTJ transitions into a non-volatile state once a current is applied. The future memory state of a DMTJ is represented by $S_0^{'}$ and $S_1^{'}$, where $S_0^{'}$ and $S_1^{'}$ are, respectively, the future state of the small PMTJ and large PMTJ. Based on (1), $S_0^{'}$ cannot represent a logical operation of the input bits. However, as described by (2), $S_1^{'}$

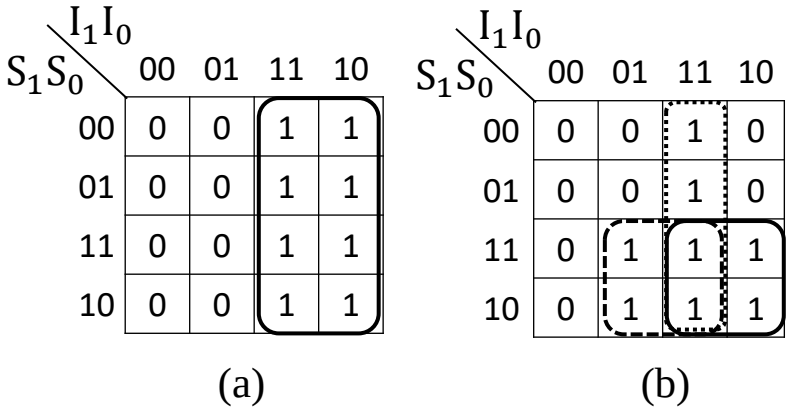


Figure 6.8: Karnaugh map of the future state bits of a DMTJ, (a) $S_0^{'}$, and (b) $S_1^{'}$

is the logical computation of the input bits which can produce a logic family through the following relationship: When $S_1 = 0$, $S_1^{'} = I_1 I_0$; hence, the future memory state of the large PMTJ is the AND operation between the two input bits, I_1 and I_0. When $S_1 = 1$, $S_1^{'} = I_1 I_0 + I_0 + I_1 S_1^{'}$ is therefore the OR output of I_1 and I_0.

Accordingly, a DMTJ operates as an AND gate when the device is initially reset to the state where $S_1 = 0$, such as R_{00} or R_{01}. The input to the AND gate is I_1 and I_0, and the output is stored in the large PMTJ. To operate a DMTJ as an AND gate, the DMTJ is reset to the R_{00} state since only one pulse is required to set the PMTJs into this state.

To configure a DMTJ as an OR gate, the DMTJ device is initially reset to the state where $S_1 = 1$ such as R_{11} or R_{10}. The output of the OR operation is stored within the large PMTJ. The R_{11} state is chosen as a *reset* state when using a DMTJ as an OR gate since the R_{11} state can be written by the one pulse write scheme.

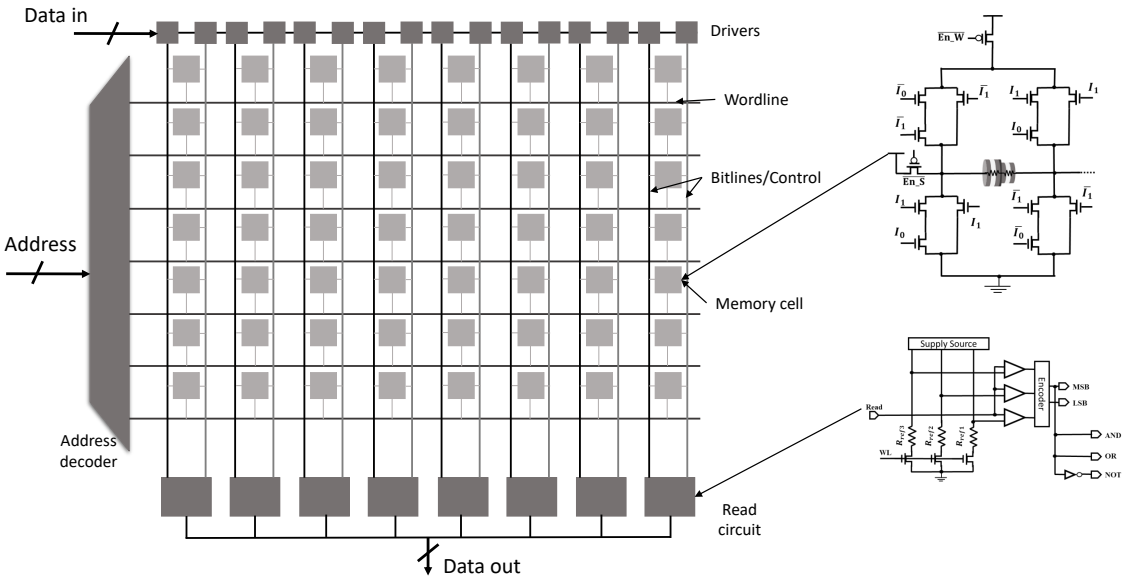


Figure 6.9: Proposed DMTJ-based multi-bit memory cell that supports the compute in-memory paradigm. The system input chooses the row being read/write/calculate through a decoder. The decoder enables the cell to perform the write/read/calculate operation. The read lines are connected to a read scheme which produces the system output.

The DMTJ can be realized as a NOT gate by storing the input signal within a DMTJ; hence, the output of the NOT gate can be achieved by a CMOS inverter, as shown in Figure 6.7. I_1 is the input to the NOT gate. To operate the DMTJ as a NOT gate, the device is initially reset to the R_{11} state, followed by a *calculate* state where the input is applied to the I_1 node and I_0 is set to zero. Based on (2), where $S_1 = 1$ and $I_0 = 0$, $S_1^{'} = I_1$. The output of the NOT gate is as shown in Figure 6.7.

The primary advantage of the proposed system lies in the ability to perform a logical operation and store the result in real-time as a non-volatile memory state. The proposed logical operation of a DMTJ is supported by the same memory system

without requiring any additional circuitry. The operational mechanism and simulation of a DMTJ behaving as a multi-bit memory cell and a nonvolatile logic element are described in the following section.

6.5 Operational mechanism and simulation results

A DMTJ-based multi-bit memory element and a nonvolatile logic element are proposed here, where a hybrid CMOS/DMTJ architecture that supports a read/write/nonvolatile logic operation is presented (see Figure 6.9). In the proposed work, a DMTJ is treated as a state machine, where the next state of a DMTJ is based on the input current and the present state of the DMTJ. The input current is based on the input bits and the output is stored within the DMTJ in a nonvolatile state. The operation of a DMTJ as a memory element and a nonvolatile logic element is based on a macrospin model of a DMTJ composed of two serially connected STT PMTJs with a diameter of, respectively, 30 nm and 40 nm. A 32 nm predictive technology model (PTM) is used to characterize the CMOS transistors [112]. The CMOS transistors are sized to provide sufficient current to switch each PMTJ based on inputs I_1 and I_0. The size of the CMOS transistors and the diameter of the two PMTJs are critical since the greater the size of a PMTJ, the lower the resistance and the larger critical current required to switch the DMTJ cell.

The operational mechanism of the proposed DMTJ-based multi-bit/nonvolatile logic system is listed in Table 6.2. Based on the functional operation (memory write or logical AND, OR, or NOT), a one-step or two-step scheme is followed. The output of the logical operation is collected at the MSB port (stored in the status of PMTJ1 - see Figure 6.9).

Table 6.2: Operational mechanism of the proposed DMTJ-based multi-bit/nonvolatile logic system

Operation		**Input**		**Output**
		Step 1	**Step 2**	
Memory	Write 00	$I_1I_0 = 00$	N.A	N.A
	Write 01	$I_1I_0 = 00$	$I_1I_0 = 10$	
	Write 10	$I_1I_0 = 11$	$I_1I_0 = 01$	
	Write 11	$I_1I_0 = 11$	N.A	
Logical AND	C = A.B	$I_1I_0 = 00$	$I_0 = A\,, I_1 = B$	C = S_1
Logical OR	C = A + B	$I_1I_0 = 11$	$I_0 = A\,, I_1 = B$	C = S_1
Logical NOT	C = $\bar{B}$	$I_1I_0 = 11$	$I_0 = 0\,, I_1 = B$	C =$\overline{S_1}$

A waveform of a DMTJ-based two bit memory cell is illustrated in Figure 6.10 where the write operation is controlled by the enable write signal En_W. En_W supports two modes of operation. When $En_W = 1$, the circuit operates in the write mode and a current is produced. When $En_W = 0$, the circuit operates in the *hold* mode with no supply current, hence the DMTJ is set in a stable state. The pulse width of the *write* and *hold* time signals is carefully chosen. Limitations on the *write* and *hold* time are

governed by the influence of the spin transfer torque on the PMTJ. The wider the pulse of the supply current, the less critical current required to switch the PMTJ. The limitations of the write pulse width are due to the worst case write scenario (the write signal only switches the smaller PMTJ, not the larger PMTJ). The width of the minimum write pulse is chosen to switch the smaller PMTJ, while the width of the maximum write pulse is set to not switch the larger PMTJ.

The direction and magnitude of the supplied current are, respectively, controlled by I_1 and I_0. Z_0 and Z_1 are, respectively, the orientation of the perpendicular magnetization of PMTJ1 and PMTJ2 within the DMTJ. The pinned ferromagnetic layer of the DMTJ is magnetized in the positive z direction (pointing up). As an example, if Z_0=1, then PMTJ1 is in the parallel state with S_0=0. The input signals to write the four resistance states within a DMTJ is shown in Figure 6.10. A minimum and maximum write pulse of, respectively, 25 ns and 35 ns is required to write a state within a DMTJ. A minimum hold time of 10 ns is required once each *write* state is set to ensure the device remains in a stable state.

Operating a DMTJ as a non-volatile logic element is achieved in two steps. A *reset* state is initially written into the DMTJ based on the logical operation, the R_{00} state for an AND operation and the R_{11} state for an OR operation. This step is followed by the *calculate* state, where the logical operation is computed in real-time, and the

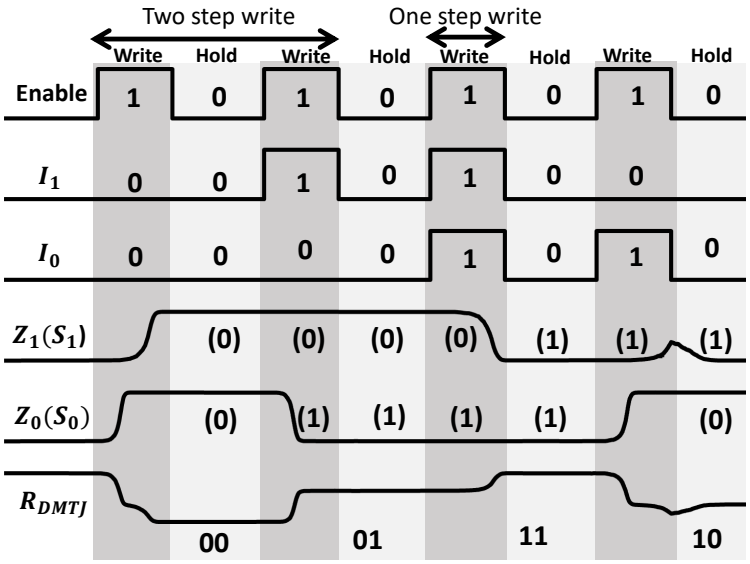


Figure 6.10: Waveform of a DMTJ behaving as a two bit memory element, where Enable is the control signal, I_0 and I_1 are the inputs, Z_0 and Z_1 are, respectively, the perpendicular magnetization of the small PMTJ and large PMTJ, S_0 and S_1 are, respectively, the corresponding state of the small PMTJ and large PMTJ, and R_{DMTJ} is the change in the resistance of the DMTJ

non-volatile state is stored within the DMTJ. A waveform of a DMTJ-based AND gate and OR gate is shown, respectively, in Figures 6.11(a) and 6.11(b).

A critical current of approximately 100 μA and 140 μA is required to switch, respectively, the smaller and larger STT PMTJs. Significant development has recently been achieved to decrease the critical current required to switch an STT PMTJ. These approaches primarily use 1) low damping materials for the thick free layers [140], and 2) new composites (such as MgO_xN_{1-x} [141]) as a tunnel barrier to overcome the inter-layer diffusion [141]. These developments in STT PMTJ memory are producing higher switching speeds, greater endurance, and lower critical currents [140, 141].

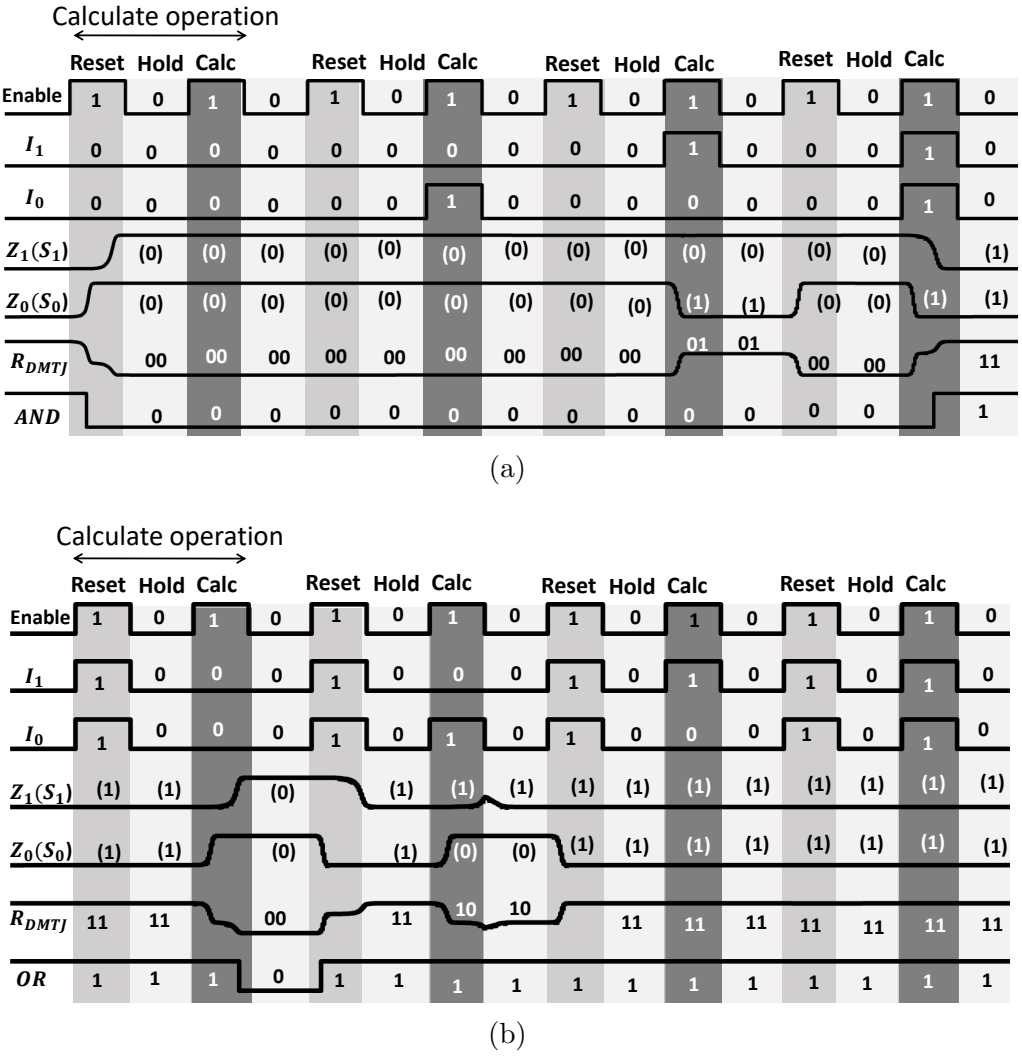


Figure 6.11: Operation of (a) a nonvolatile DMTJ AND gate and (b) a nonvolatile DMTJ OR gate

6.6 Comparison with state of the art memristive-based compute *in-Memory* systems

The proposed DMTJ-based hybrid multi-bit and compute in-memory system is different from alternative memristive-based compute in-memory systems such as a

single STT MTJ MRAM-based system [142], SOT-MRAM-based system [129], and memristor-based systems [31]. In single STT MRAM-based compute in-memory systems [142], where each memory cell stores a single bit, logic circuitry is needed after the sense amplifier to perform the logical operation, producing a volatile output. A comparison of the delay and system requirements of the proposed hybrid multi-bit memory and logic cell with recent *in situ* MRAM-based in-memory computing schemes is listed in Table 6.3. The memristor-based in-memory compute system, described in [31] with Ta/GeTe/Ag memristors with an area of the GeTe functional layer of $10 \times 10\ \mu m^2$, requires two cycles (100 ns) to perform a nonvolatile AND operation, while the proposed DMTJ system requires 70 ns. The SOT-based logic in-memory system proposed in [129] requires approximately 10 ns for a nonvolatile AND operation assuming a 40 nm CMOS technology with a voltage-gated spin hall effect MRAM cell. The length of each row in the SOT crossbar array is however limited by the resistance of the heavy metal layer that supports spin orbital torque interactions within multiple SOT devices, leading to different write conditions for the SOT cells along the heavy metal layer.

6.7 Summary

A double magnetic tunnel junction (DMTJ) is a device composed of two serially connected perpendicular magnetic tunnel junctions with different diameters, providing

Table 6.3: Comparison of nonvolatile AND in-memory compute systems

Technology	**Device Requirement**	**# of Steps**	**Delay**	**Process**
DMTJ	One DMTJ (Two serially connected PMTJs)	Two	70 ns	32 nm CMOS
Memristor [31]	Two serially connected memristors with opposite polarities	Two	100 ns	-
SOT [129]	One voltage-gated spin hall effect device	Two	$\sim$ 10 ns	40 nm CMOS

four different resistance states. In this work, the DMTJ is proposed both as a multi-bit memory element and as a nonvolatile logic element. A hybrid CMOS/DMTJ circuit supports a read/write/nonvolatile logic operation. The DMTJ behaves as a multi-bit memory element through a two step write mechanism producing two resistance states, a one pulse write mechanism for the two resistance states and a one step read mechanism. The DMTJ also behaves as a nonvolatile logic element. The DMTJ is reset to an initial state followed by a calculate state. The output of the logical operation is stored as a nonvolatile state within the DMTJ. The DMTJ behaves as a nonvolatile AND, OR, and NOT gate with, for a 32 nm CMOS technology node, a delay of 70 ns. The multi-bit memory cell exhibits an access time of 35 ns with a one step write scheme to write either the R_{00} state or R_{11} state, and 70 ns for a two step write scheme to write either the R_{01} state or R_{10} state.

Chapter 7

Test Modules for Enhanced Testability of Single Flux Quantum Integrated Circuits

Superconductive electronics is a promising technology with important characteristics, such as low energy per operation [143], lossless interconnects at DC, zero static power, operation at clock frequencies exceeding 100 GHz [144], and a natural interface with quantum computing systems [145]. Single flux quantum (SFQ) logic is a superconductive logic family for low power, high performance cryogenic computing [146].

High reliability is a necessary requirement for superconductive integrated systems. The complexity of SFQ circuits has reached 800,000 Josephson junctions, operating at subterahertz clock frequencies [147]. The challenge of achieving high performance with high reliability is escalating due to dimensional scaling, novel materials and devices,

and operation in severe conditions (extreme cryogenic temperatures and sub-terahertz frequencies).

These reliability challenges, combined with yield issues, are exacerbated by exotic manufacturing technologies. Reliability and yield can be categorized by the failure paths (sequence of faults due to a physical failure) and failure mechanisms (physical cause of the failure). Determining the defects and faults is essential to enhance the lifetime of superconductive systems. This capability is achieved by improving the fault coverage, where the system is evaluated to identify the characteristics of the faults, such as the quantity, location, and type.

Fault coverage is improved by exploiting design for testability (DFT) techniques to enhance the controllability and observability of the internal nodes within a system. An understanding of the physics of each failure mechanism and the development of effective and reliable algorithms that exploit these DFT techniques prior to fabrication are vital to the development of superconductive systems.

A methodology is proposed here to include DFT within SFQ systems, a topic currently in an embryonic stage. This objective is achieved by enhancing the controllability and observability of the internal nodes within an SFQ system to identify specific defects and faults. This capability can be accomplished by exploiting embedded hardware solutions such as test insertion and/or test extraction.

Several significant differences exist between conventional CMOS logic and SFQ logic. SFQ logic operates at sub-terahertz clock frequencies in a cryogenic environment. An SFQ signal is represented by the existence of an SFQ pulse. The following additional differences prevent the use of standard CMOS-based DFT techniques [148],

1) observation of an internal node within a CMOS system is achieved by direct probing, while in SFQ systems, a test extraction module is required to non-destructively readout the signal of an internal node,

2) SFQ logic gates are inherently clocked and latched within at least one storage loop, where several clock cycles are required to produce an output [149]. Additional information, such as the number of cycles, is required by a test controller in SFQ systems.

3) limited fan-out of SFQ gates and flip flops [150]. A splitter is required to provide an additional output [150]. A test technique is proposed here to enhance the controllability and observability of the internal nodes within SFQ systems.

The primary contributions described in this chapter lie in two areas. One aspect is a circuit solution to support DFT in SFQ systems, where two test modules are presented. The proposed test modules are a test extraction module that observes the internal nodes of an SFQ system, and a hybrid test module to observe and control the internal nodes. A second aspect is a methodology for evaluating the tradeoffs of

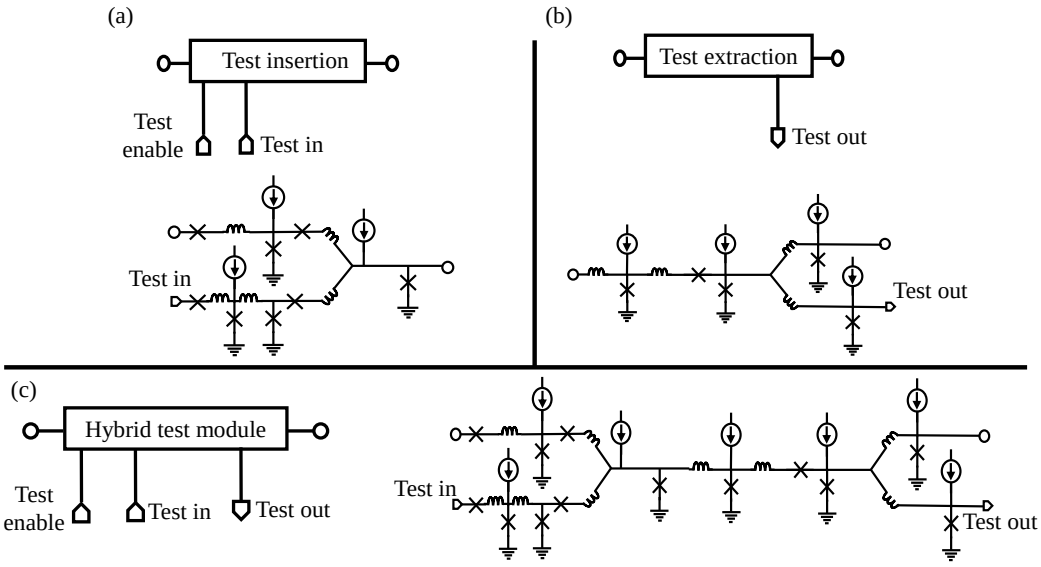


Figure 7.1: Circuit- and block-level diagram of DFT approaches for SFQ systems, (a) test insertion module [151], (b) test extraction module, and (c) hybrid test module.

inserting these circuit solutions on the observability and controllability of the internal nodes and hence the system testability.

This chapter is organized as follows. In section 7.1, the proposed test modules are described. The effects of incorporating the proposed test modules on certain test measures are discussed in section 7.2. A methodology and related tradeoffs that consider these test modules in terms of power, area, detection speed, and overall testability are presented in section 7.3. The chapter is summarized in section 7.4.

7.1 Test Modules

An SFQ-based test point insertion module to enhance the controllability at any node has previously been proposed [151, 152]. The test point insertion module consists of blocking gates and a confluence buffer (CB) at each bit along a data path. The blocking gates are supplied by a clock signal. This clock signal is gated by a non-destructive read out T flip flop controlled by a test controller. The test insertion module selects between the input test signal and a data signal. For example, applying clocked blocking gates to insert test points within a 64 bit register requires 35% fewer Josephson junctions as compared to using multiplexers [151]. This advantage increases with current controlled blocking gates. This test insertion module supports both set/scan chains and test point insertion. These techniques can be applied to evaluate the fault characteristics of SFQ systems and to demonstrate built-in self-test (BIST) of SFQ compatible memory systems [152].

To support DFT in SFQ, a test extraction module and a hybrid test module are proposed to enhance the observability and controllability of SFQ systems, as shown in Figure 7.1. The proposed test extraction module consists of a transmission line and a splitter that generates in real-time a copy of the propagated data. An additional hybrid module to improve the controllability and observability at the same node is also proposed.

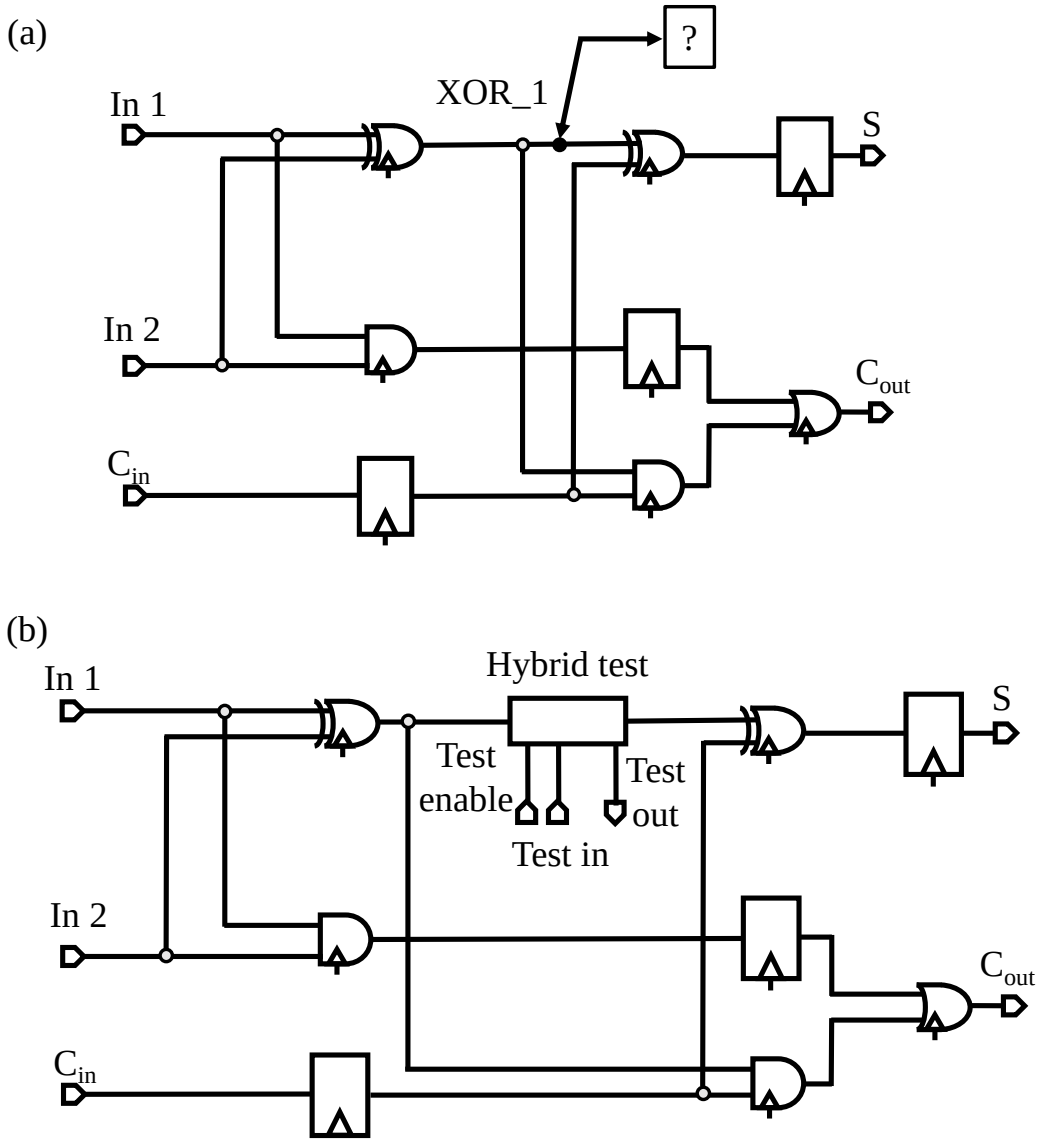


Figure 7.2: Test circuit to validate the functionality of the proposed test modules, a) SFQ-based single bit full adder where the node under test (XOR_1) is the target node being observed/controled, and b) proposed hybrid test module inserted at the target node being observed/controled. Note that the hollow circle indicates a splitter cell.

As an example, the proposed hybrid test module is inserted to control or observe the output of the XOR gate in a single bit full adder, as illustrated in Figure 7.2. The

functionality of the proposed test modules is validated on a single bit full adder, as illustrated in Figure 7.3. The detection speed, area overhead, and performance of the proposed test modules on a circuit under test are discussed in Section III. These results validate the feasibility of these approaches to enhance the testability of SFQ systems.

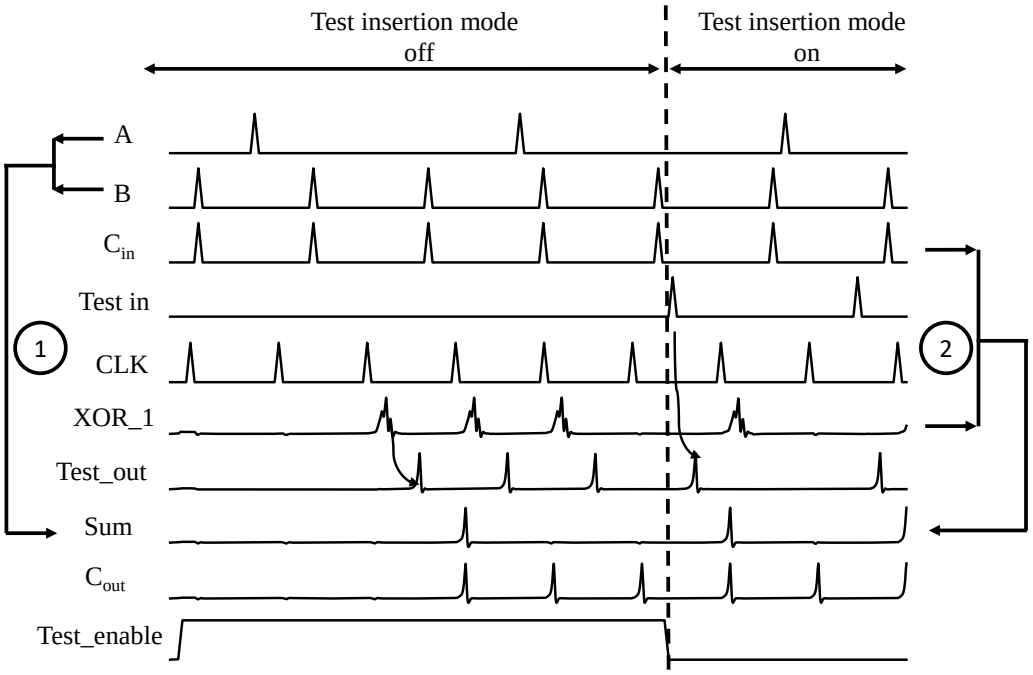


Figure 7.3: Operation of proposed hybrid test module located at the XOR_1 node of a full adder, as illustrated in Figure 7.2. The proposed hybrid test module operates as both a test insertion module and test extraction module with two modes of operation. *Test insertion mode off* when the signal at the XOR_1 node (the input to the test module) is produced at the output. In this case, the Sum output is the sum operation of the A and B signals. *Test insertion mode on*, where the $Test\_in$ signal is passed to the output of the module. Hence, the Sum output is the XOR operation of the $Test\_in$ signal and the C_{in} signal. The $Test\_out$ signal is a real-time copy of the output of the test module. $Test\_enable$ switches between the two test modes.

7.2 Test Measures

To quantify the influence of the proposed test modules on testability quality measures, the Sandia controllability observability analysis program (SCOAP) algorithm is used [153]. SCOAP analyzes and quantifies the difficulty to control or observe internal nodes within a circuit, guides test generation, estimates fault coverage, and determines the test vector length. SCOAP measures the combinational circuits as follows. A combinational controllability of zero ($CC0$) describes the difficulty to set an internal node to logic 0 (ranging from 1 to ∞). A combinational controllability of one ($CC1$) measures the difficulty of setting an internal node to logic 1 (ranging from 1 to ∞). A combinational observability (CO) measures the difficulty in observing an internal node (ranging from 0 to ∞). Higher values of $CC0$, $CC1$, and CO indicate greater difficulty in controlling or observing an internal node.

Benchmark circuits such as ISCAS'85 C17 are used to explore and validate the proposed test modules to enhance the controllability and observability of the internal nodes within an SFQ system. The SCOAP testability and controllability measures of the internal nodes are determined before and after inserting the proposed test modules, as illustrated in Figure 7.4 and listed in Table 7.1. The SCOAP measures are described as ($CC0$,$CC1$)CO. As shown in Figure 7.4(a), before inserting a test module, the SCOAP measures at node X are (2,5) 6 with $CC0$=2, $CC1$=5, and CO=6.

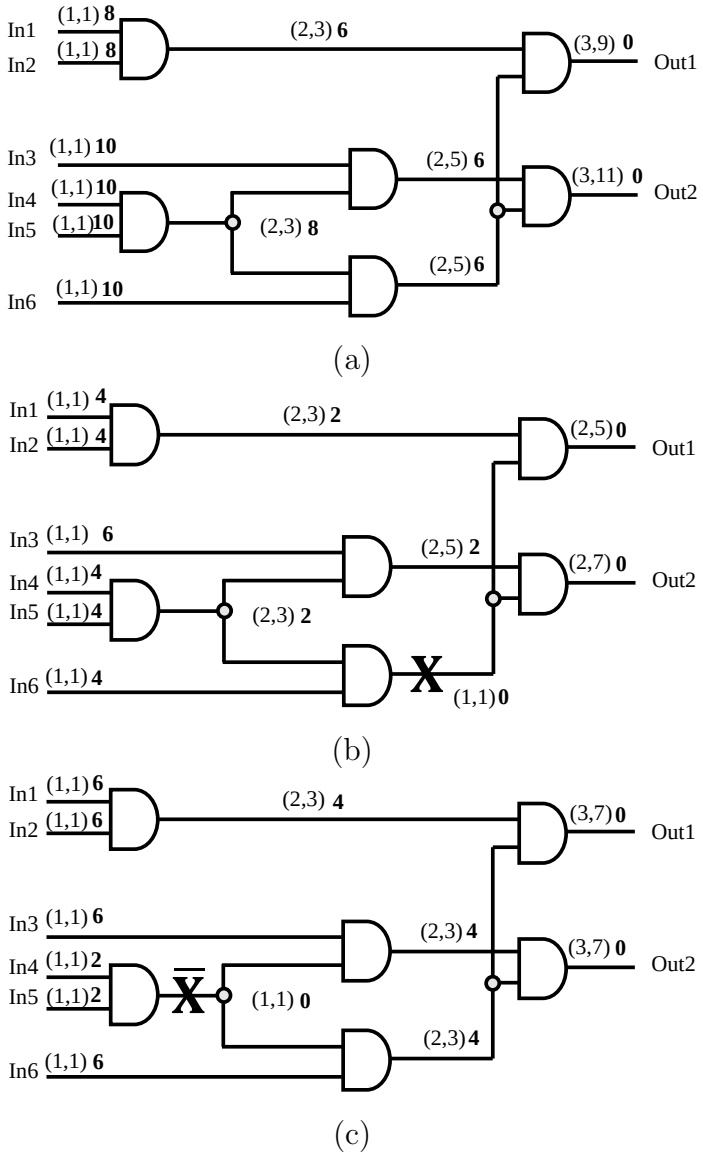


Figure 7.4: SCOAP testability evaluation of ISCAS'85 C17 benchmark circuit, (a) before insertion of the proposed test modules, and (b) after insertion of the hybrid test module at node X (the output of the second level AND gate) and at node $\overline{X}$ (the output of the first level AND gate).

The test point location process is determined, as follows:

1) evaluate the global test measures and identify those nodes with the lowest testability

characteristics. Nodes with high fanout are preferable.

2) insert the test module and determine the updated testability characteristics.

3) iterate between steps (1) and (2) to identify the optimal node (the test node which produces the greatest enhancement of the testability characteristics upon insertion of a test module).

4) Based on the testability requirements of the system, a methodology is developed to identify the number of test modules and target testability characteristics.

The process of selecting the optimal test point location/s is a computational complex problem, where a common algorithm to achieve this objective, for example, is the iterative test point insertion algorithm [154, 155]. In this algorithm, each test point location is individually evaluated to determine the optimal location(s).

As an example, before inserting a test module within benchmark circuit C17, it is important to evaluate the global test coverage to determine any testability bottlenecks. As shown in Figures 7.4(b) and 7.4(c), nodes X and $\overline{X}$ exhibit the worst testability characteristics and highest fanout. Nodes X and $\overline{X}$ are set as the nodes under test. The proposed hybrid test module is individually inserted at each of these nodes, and the SCOAP testability measures are recalculated.

SCOAP is used to analyze the structure of the circuit under test to guide the DFT insertion process. The location of the inserted modules is chosen to enhance the testability characteristics of the overall system. Due to the difference between

the testability measures in the circuits shown in Figures 7.4(b) and 7.4(c), node X is chosen for test point insertion over node $\overline{X}$. Inserting a test module at node X enhances the controllability and observability characteristics at most of the internal nodes of the circuit under test, as illustrated in Figure 7.4(b).

7.3 Methodology of Incorporating Test Modules

The objective of a methodology to incorporate the proposed test modules is to identify the number, type, and location of each of the test modules within an SFQ system. Each test module influences the testability characteristics differently, affecting the power, area, and delay overhead. As an example, a comparison of the effects of inserting only one test module (test insertion, test extraction, or hybrid test module) at node X in the C17 benchmark circuit (shown in Figure 7.4), is listed in Table 7.1. The overhead of inserting each of these modules is listed in Table 7.2.

After inserting a test extraction module at node X in the C17 benchmark circuit, as illustrated in Figure 7.4, the sum of the combinational observability measure of all nodes ($\sum CO$) is enhanced by 36%. Inserting one test insertion module enhances both the sum of the combinational controllability to logic 1 ($\sum CC1$) and to logic 0 ($\sum CC0$) by, respectively, 50% and 15%, and $\sum CO$ by 41%. Inserting one hybrid test module improves $\sum CC1$ and $\sum CC0$ by, respectively, 50% and 15% and $\sum CO$ by 61%.

Table 7.1: Comparison of the effects of inserting the proposed test modules into the ISCAS'85 C17 benchmark circuit in terms of SCOAP testability measures.

	CC0 max	**CC1 max**	**CO max**	$\sum$ **CC0**	$\sum$ **CC1**	$\sum$ **CO**
W/O test modules	3	11	10	20	42	82
Test extraction module	3	11	10	20	42	52
Test insertion module [9]	2	5	10	17	28	48
Hybrid test module	2	5	7	17	30	32

Table 7.2: Comparison of the overhead of inserting the proposed test modules on the area (number of resistively shunted JJs, inductors, and power resistors), power dissipation, detection time, and delay (for a 10 KA/cm^2 process technology).

	Area overhead			**Detection time (ps)**	**Advantage**	**Delay overhead τ_{Delay} (ps)**	**Power overhead (fW)/calculation**
	JJ	**L**	**R**				
Test extraction module	5	4	4	7	Enhanced observability	0 Since $T_{Delay} + T_{combinational} < T_{CLK}$	0.513
Test insertion module [9]	8	5	3	NA	Enhanced controllability and observability		Normal: 0.275 Test: 0.27
Hybrid test module	13	9	7	18			Normal:0.788 Test: 0.781

SFQ logic gates are inherently clocked and latched [156]. Moreover, each logic stage between sequentially-adjacent registers may require several clock cycles to produce an output [157]. The proposed test extraction module has no effect on system speed as long as $T_{Delay} + T_{Combinational} < T$, where T_{Delay} is the delay of the test extraction module, $T_{Combinational}$ is the delay of the circuitry between the two registers/logic cells

connected to the test extraction module, and T is the clock period [158]. The detection time of the proposed test extraction module is 7 ps, and 18 ps for the proposed hybrid test module (in a 10 KA/cm^2 process technology [159, 160]).

As listed in Table 7.2, the proposed test extraction module exhibits a power overhead of 0.513 fW/calculation. The test insertion module exhibits 0.275 fW/calculation in normal mode - test insertion mode off - and 0.275 fW/calculation in test mode. The hybrid test module exhibits 0.788 fW/calculation in normal mode and 0.78 fW/calculation in test mode (for a 10 KA/cm^2 process technology).

The structure of the circuit under test determines the influence of inserting a test module [152, 161]. Certain parameters, such as the number of nodes, number of logic gates, the number of logic levels from the primary inputs (for controllability) or primary outputs (for observability), and fanout of each internal node, determine whether inserting a test module is an effective solution.

Multiple benchmark circuits have been analyzed to evaluate the effectiveness of the proposed modules for different circuit structures. One hybrid test module is inserted at a node in several benchmark circuits; ISCAS'85 circuits, C17 and C432, and 74X-series circuits, 74182 and 74283. In the data listed in Table 7.3, one hybrid test module is inserted at one of the target nodes. This node is selected after identifying the critical signal path with the most number of nodes with high fanout and worst testability characteristics (as previously discussed in section 7.2). Inserting a hybrid test module

into C17 and C432 is more effective in improving testability than into 74182 and 74283 since C17 and C432 contain a greater number of high fanout nodes and long signal paths.

Pseudocode describing the methodology and tradeoffs of inserting test modules into an SFQ system is shown in Algorithm 1. The algorithm is composed of three steps. First, all internal nodes are scanned to determine the nodes with high testability measures (possible test points). Second, one test module is inserted at each possible test point followed by evaluating the testability characteristics of the entire system. The overhead of these test modules may not exceed the target performance limits. Finally, a test module is inserted at the internal node with the poorest testability characteristics. These three steps are repeated until the target testability characteristics are achieved or the system exceeds the target performance requirements.

Table 7.3: Comparison of the effects of inserting a hybrid test module into ISCAS'85 C17 and C432 benchmark circuits, and 74X-series circuits 74182 and 74283 in terms of the per cent enhancement of the SCOAP testability measures before and after insertion of the hybrid test module.

	CC0 max	CC1 max	CO max	$\sum CC0$	$\sum CC1$	$\sum CO$
C17	33%	55%	30%	15%	29%	61%
C432	37%	40%	32%	25%	22%	24%
74182	0	11%	0	5%	5%	11%
74283	0	13%	0	3%	7%	7%

Algorithm 1 Pseudocode of Algorithm for Inserting Test Modules into an SFQ System

Input: Number of nodes N, location of each node ($Nloc$), target testability characteristics $Tcon$ and Tob, and performance overhead limit $Over_{limit}$

Output: Number n, type $Ttype$, location $Tloc$ of test modules

Evaluate testability characteristics of all internal nodes $Ncon$ and Nob
for $k \leftarrow 1$ to N **do**
    if $(Ncon_k \geq Tcon)||(Nob_k \geq Tob)$ **then**
        $Pn \leftarrow Nloc_k$ ▷ Possible nodes for testpoints
    else
        exit
    end if
end for
for $k \leftarrow 1$ to size of Pn **do**
    Insert test module
    Evaluate testability characteristics Con_k and Obs_k
    Evaluate performance overhead $Overhead$
    Over power flag $OF = 1$
    if $(Overhead > Over_{limit})$ **then**
        continue
    end if
    $OF = 0$
    Pre-final node $PF_k \leftarrow (Nloc(Pn_k), Con_k, Obs_k)$
end for
if $(OF = 1)$ **then**
    exit
end if
final node $FN \leftarrow PF(i)$ with i is the index of lowest $Con(PF)$ and $Obs(PF)$
$n = n + 1$
$Tloc_n = Nloc(FN)$
go to Step 1

7.4 Summary

Advanced testing methodologies are required to support complex digital SFQ systems. In this chapter, two solutions are presented to enhance the testability of SFQ systems by improving the controllability and observability of the internal nodes. A test extraction module with a detection time of 7 ps and a hybrid test module with a detection time of 18 ps are presented. The proposed test modules are validated on a suite of benchmark circuits. A comparison of the effects of inserting the test modules into different benchmark circuits in terms of the overhead and testability measures is provided. The proposed test modules (test insertion, extraction, and hybrid) for the ISCAS'85 C17 benchmark circuit exhibit a power overhead of, respectively, 0.513, 0.27, and 0.78 fW/calculation. The proposed test modules significantly enhance, by more than 50%, the testability measures (controllability and observability) of the internal nodes, increasing overall fault coverage.

Chapter 8

Josephson Junction Stuck at Fault Detection in SFQ Circuits

The challenge of achieving high performance with high reliability is escalating due to dimensional scaling, novel materials and devices, and operation in severe conditions (such as extreme cryogenic temperatures and sub-terahertz frequencies). These reliability challenges, combined with yield issues, are exacerbated by exotic manufacturing technologies.

Single flux quantum (SFQ) logic is a superconductive technology for low power, high performance cryogenic computing. The development of SFQ technology has enabled complex integrated circuits achieving over 11,000 JJs for digital single processors [162] and similar complexity prototype RSFQ microprocessors [163]. SFQ circuits with a regular layout structure such as an AC biased SFQ shift register have reached 800,000 JJs [147], operating at subterahertz clock frequencies. The achievable frequencies and cryogenic environment make SFQ circuits difficult to control via

external probing. Prototype evaluation of these circuits, therefore, requires advanced testing methodologies.

Reliability and yield can be categorized by the failure paths and failure mechanisms. Determining the defects and faults is essential to enhancing the lifetime and testability of superconductive systems. This capability is achieved by improving the fault coverage, where the system is evaluated to identify the characteristics of the faults, such as the quantity, location, and type. Understanding the behavior of each failure mechanism and the development of effective and reliable methodologies that exploit design for testability (DFT) techniques prior to fabrication are vital to the development of testable superconductive systems.

The building elements of SFQ systems are JJs, resistors, inductors, and interconnects. In this chapter, high level JJ-based fault models are proposed, and the required test vectors are described to detect the location and type of these faults [146, 151].

Several significant differences exist between conventional transistor-based CMOS fault models and JJ-based SFQ fault models beyond sub-terahertz clock frequencies and the cryogenic environment. An SFQ signal is represented by the existence of an SFQ pulse not as a voltage level (as in CMOS). In both CMOS and SFQ device-based faults, it is challenging to identify the location and type of faults within a system. The following additional differences prevent the use of standard CMOS-based DFT techniques [148, 151]:

1) Only two states exist in SFQ logic, zero (the absence of a pulse) or one (existence of a pulse). In CMOS logic, output states such as zero, one, and floating may exist.
2) SFQ logic gates are inherently clocked and latched within at least one storage loop, where several clock cycles are required to produce an output [149]. Unlike CMOS, in SFQ systems, additional information, such as the number of cycles, is required by a test controller.
3) Limited fan-out of SFQ gates and flip flops [150]. Splitters are required to provide additional outputs [150].

A methodology is proposed here to include DFT within SFQ systems. To the author's knowledge, this work is the first to describe JJ-based stuck at faults. This objective is achieved by developing high level fault models that target JJ-based faults, such as stuck at a superconductive state or an open circuit state. These fault models can be exploited to develop a fault simulation algorithm. The required test vectors to identify the type and location of these sets of faults are generated based on a high level fault model. A summary of the quality measures of each fault model is discussed in this chapter. The fault coverage of open circuit and short-circuit faults and the location of each logic cell are also identified.

Potential faults within SFQ systems are categorized by device-based faults, fabrication-based faults, and DC bias network faults, as shown in Figure 8.1. JJ-based fault

models are evaluated to generate the required test vectors to determine the location and/or type of defects.

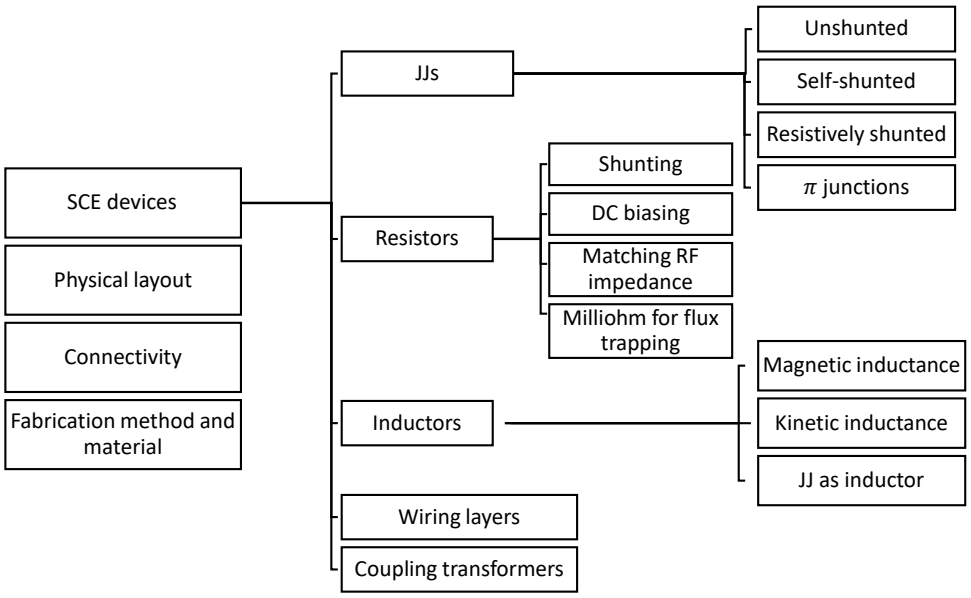


Figure 8.1: SFQ fault mechanisms. Component-based faults are attached to a specific SFQ component. High level models detect the faults associated with resistively shunted JJs. Other SFQ fault mechanisms include faults associated with the physical layout, faults due to connectivity between the devices, and faults due to limitations in the manufacturing process.

The primary contribution of this chapter is a test methodology for SFQ systems. High-level JJ-based fault models are developed followed by a methodology for developing a fault model to target a specific block or type of fault. The chapter is organized as follows. JJ-based fault mechanisms and related fault models are discussed in section 8.1. These proposed JJ-based fault models are validated in section 8.2. The required test vectors to detect and allocate JJ-based faults within an SFQ system are presented in section 8.3. The fault coverage of the proposed models are presented in

section 8.4. A methodology to develop a block level JJ-based faul model to generate the required test vectors is proposed in section 8.5. The chapter is summarized in section 8.6.

8.1 JJ-based High Level Fault Models

Multiple types of JJs exist within SFQ systems, such as unshunted, self-shunted, resistively shunted, and π junctions. Resistively shunted JJs are the most advanced fault type within state-of-the-art high performance SFQ circuits. Faults associated with resistively shunted JJs are the focus of this chapter.

A faulty resistively shunted JJ has four modes of operation, stuck at superconductive, stuck at resistive, open circuit, and noisy switching. Further simplifications are necessary to develop JJ-based fault models that support complex testability mechanisms processing millions of JJs.

The JJ stuck at fault model is the most general fault model [164]. The great majority of physical failures results in stuck at shorts and opens [164]. In this chapter, two JJ-based fault modes are considered, stuck at superconductive (SC) state and stuck at open circuit (OC) state.

Multiple physical defects can lead to a JJ stuck in the superconductive state, such as a JJ with a higher critical current than the expected value. Typical margins for a bias network are 20% to 30% of the critical current [165, 166]. If the critical current of

a JJ is above this margin, the device behaves as a stuck at SC state. A JJ stuck in the open circuit state can occur if a break exists in the tunneling barrier or interconnect. In this chapter, a JJ stuck at the SC state is modeled as a JJ with a high critical current (5 mA) to ensure the operation is in the stuck at SC state, while a JJ stuck at the OC state is modeled as an open circuit.

8.1.1 Fault Simulation and Analysis

To develop a high-level fault model of a logic cell with a faulty JJ, the response of a circuit is considered to be only due to one fault type in a single JJ at a time. The logic cells within an SFQ cell library are based on the configuration shown in Figure 8.2. A JJ-based fault model is presented for the following SFQ cells; JTL, splitter, DFF, OR, and AND gates.

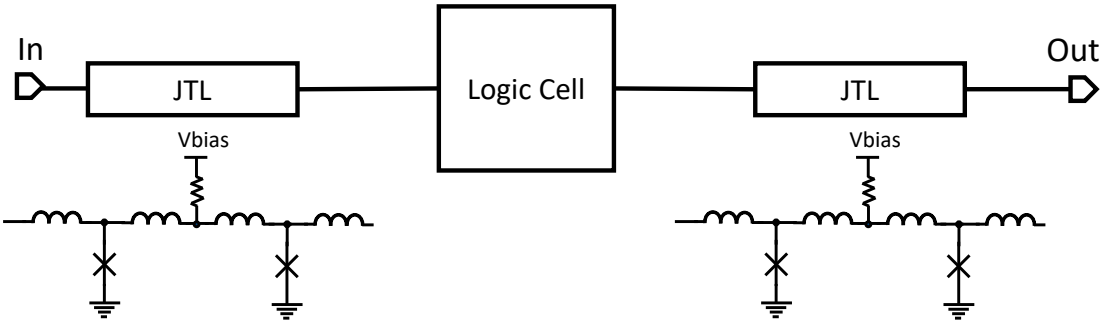


Figure 8.2: Configuration of the cell under test where a Josephson transmission line is placed at the primary inputs and outputs of the logic cell under test.

The simulation environment to model JJ-based faults within a Josephson transmission line (JTL) is illustrated in Figure 8.3(a). The output of a JTL due to a JJ stuck

at SC state is zero. The output of a JTL with a JJ stuck at OC state is challenging to detect. This output exhibits a small delay from a reference JTL without faults and unexpected pulses, as shown in the circled areas depicted in Figure 8.3(b).

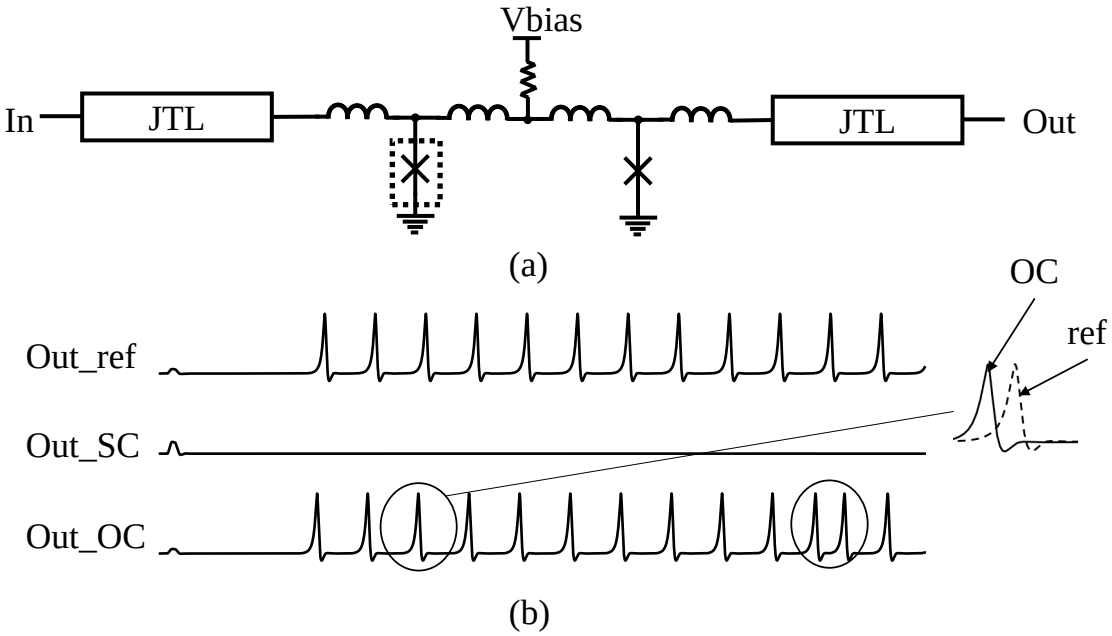


Figure 8.3: JTL faults, (a) model of a JTL, and (b) simulation of a JJ stuck at SC or OC state, indicating additional failure behaviors (circled). The squared JJ is the faulty JJ. *Out_SC* is the output of a JTL with a reference cell (without faults). *Out_SC* and *Out_OC* are, respectively, the output of a JTL with a JJ stuck at SC state and OC state.

Two types of JJs exist within a splitter, the driver JJ and the branch JJ. The simulation environment to model JJ-based faults within a splitter cell is shown in Figure 8.4. The faulty output of a splitter with the driver JJ stuck in the OC state (see Figure 8.4(a)) is similar to a reference output (without JJ faults), as illustrated in Figure 8.4(a). In this condition, a stuck at OC fault in the driver JJ is undetectable. The output of a splitter cell with a branch JJ stuck in the SC state depends upon

the location of the faulty JJ, as shown in Figure 8.4(b). The output of a splitter cell, attached to a faulty JJ (stuck in the SC state), exhibits a single pulse followed by zeros, as depicted in Figure 8.4(b).

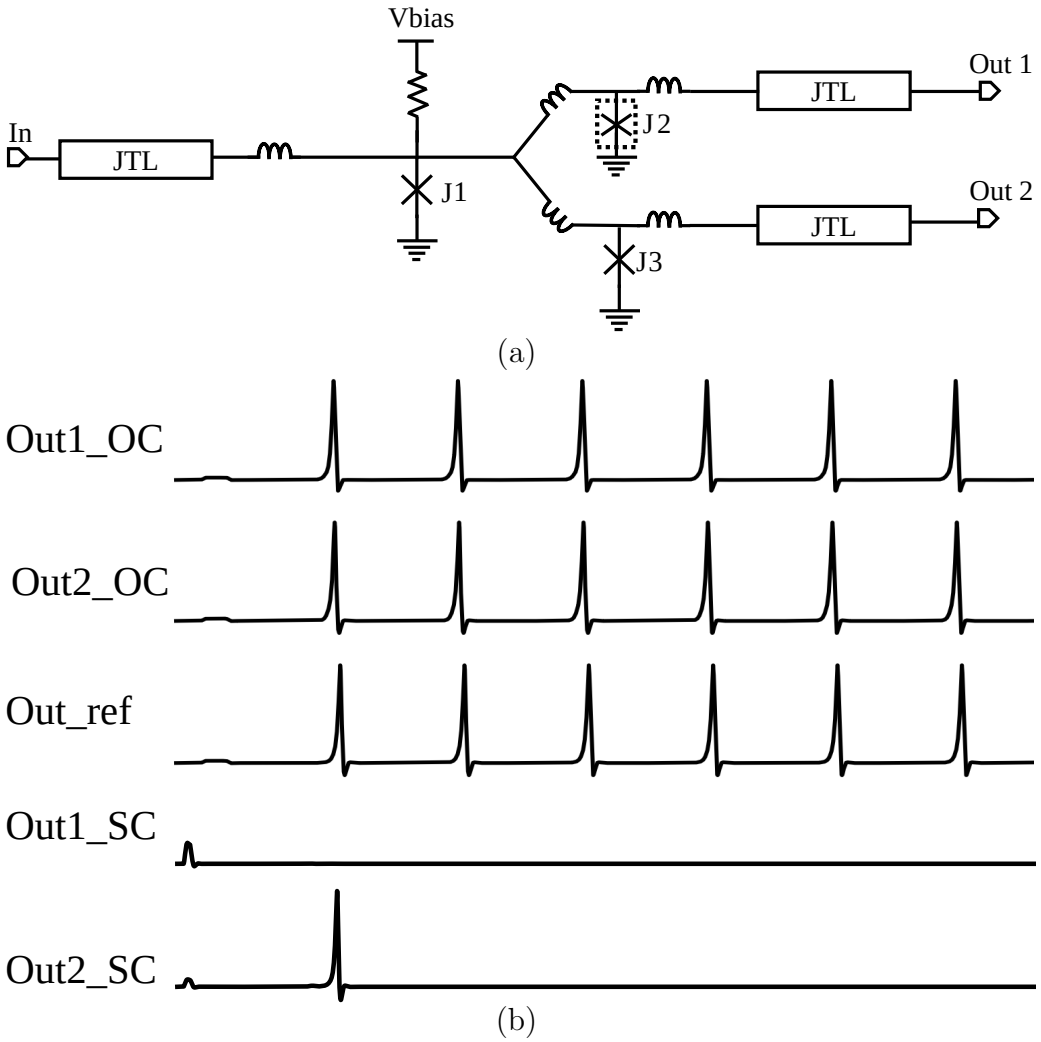


Figure 8.4: Splitter faults, (a) splitter cell, and (b) simulation of J2 stuck at SC or OC state.

The same procedure is followed to model other SFQ cells, such as DFF, AND, and OR gates. These logic cells have multi-input and multi-output ports including a clock

signal. To provide a fault model for these gates, the cell under test is analyzed for all input conditions with the clock signal enabled or disabled. Testing all input cases is essential to determine the location and type of faulty JJs within a cell.

At least one storage loop exists in these logic gates. The location of a JJ within these logic gates sets the dependence of the cell function on the clock signal or latching operation. For the example shown in Figure 8.5(a), a fault in J2 influences the clock signal, while a fault in J1 or J3 affects the storage loop. At least two test vectors are required to detect faults within these clocked logic gates, as discussed in section 7.2.

As previously discussed in section 7.1, high level fault models are based on different cell behaviors caused by JJ-based faults. These fault models are independent of technology and/or manufacturing process. As listed in Table 8.2, the output of each logic cell due to a fault in a JJ is compared to a reference cell (without faults). The faulty output is highlighted as a gray cell.

As shown in Figure 8.5(b), an OR cell is composed of 8 JJs with four control JJs and one DFF, while, as illustrated in Figure 8.5(c), an AND cell is composed of 11 JJs, forming two DFFs and three control JJs. To provide a high level JJ-based fault model of both an OR cell and an AND cell, the output is evaluated with and without the clock signal. As an example, when the clock signal is on, a clock pulse is inserted after each input. High level JJ-based fault models of an OR gate and an AND gate are, respectively, listed in Tables 8.1(a) and 8.1(b). Note that a fault in the JJs forming

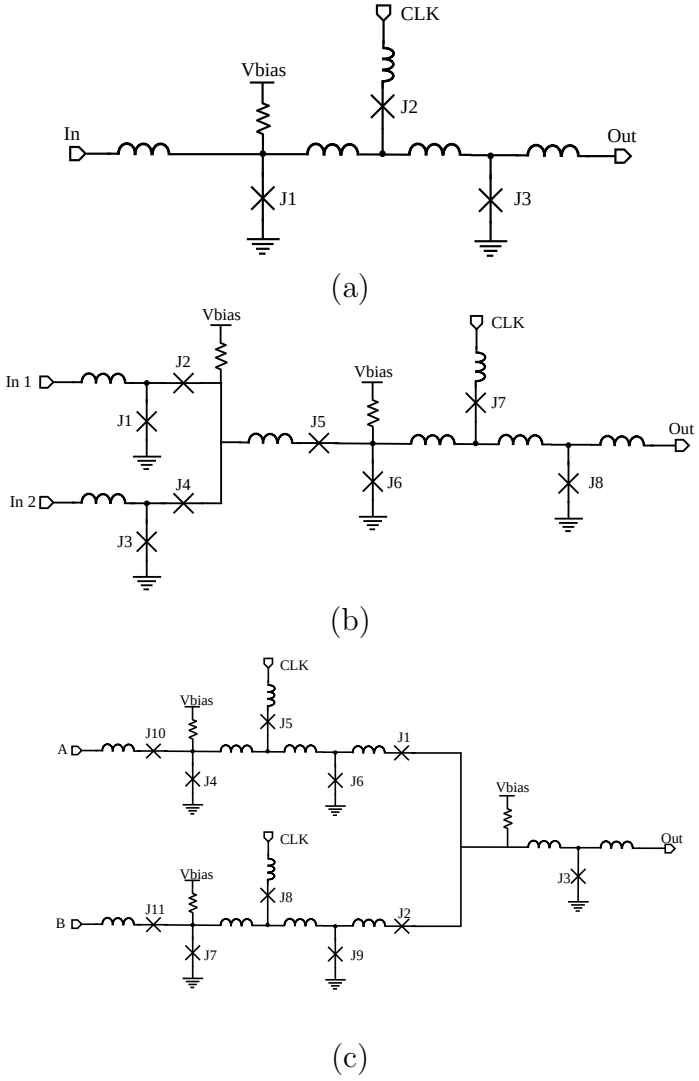


Figure 8.5: Circuit structure of (a) DFF, (b) OR cell, and (c) AND cell.

the DFFs within the OR and AND gates may exhibit a different behavior when the clock signal is on or when it is off. However, a fault in the control JJs can only be identified when the clock signal is on.

Table 8.1: High level JJ-based fault models, (a) OR cell, and (b) AND cell. The faulty output is highlighted as a gray cell.

A	B	Ref		J1		J2		J3		J4		J5		J6_w/o _clck		J6_w _clck		J7_w/o _clck		J7_w _clck		J8_w/o _clck		J8_w _clck	
		W_clck	W/o_clck	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC
0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	1	1	1	0	1	0	1	1	0	1	1	0	1	0	0	0	0	1	1	0	1	0
1	0	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	0	0	1	1	0	1	0
1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	0	0	1	1	0	1	0

(a)

A	B	Out_ref		J1		J2		J3		J10		J11		J4w/o _clck		J4_w _clck		J5_w/o _clck		J5_w _clck		J6_w/o _clck		J6_w _clck		J7w/o _clck		J7_w _clck		J8_w/o _clck		J8_w _clck		J9_w/o _clck		J9_w _clck	
		W _clck	W/o _clck	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	0	0	1	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0

(b)

8.2 Validation of Proposed JJ-based Fault Models

The benchmark circuit shown in Figure 8.7 is used here to validate the proposed JJ-based fault models. This validation process identifies the logic paths of JJ-based faults within an SFQ system. As an example, one fault is inserted at J2 within the AND_1 gate. As listed in Table 8.1(b), when J2 is stuck in the SC state within an AND gate, the faulty output is one when A=1 and B=0. As shown in Figure 8.6(a), the faulty and reference AND gates produce a correct output when A=1 and B=1. When A=1 and B=0, output AND_1 is one when J2 is stuck in the SC state. This faulty output propagates to the second stage of the benchmark circuit, as shown in Figure 8.6(b). *Out_Ref* is the result of an AND operation between AND_1 and C_1.

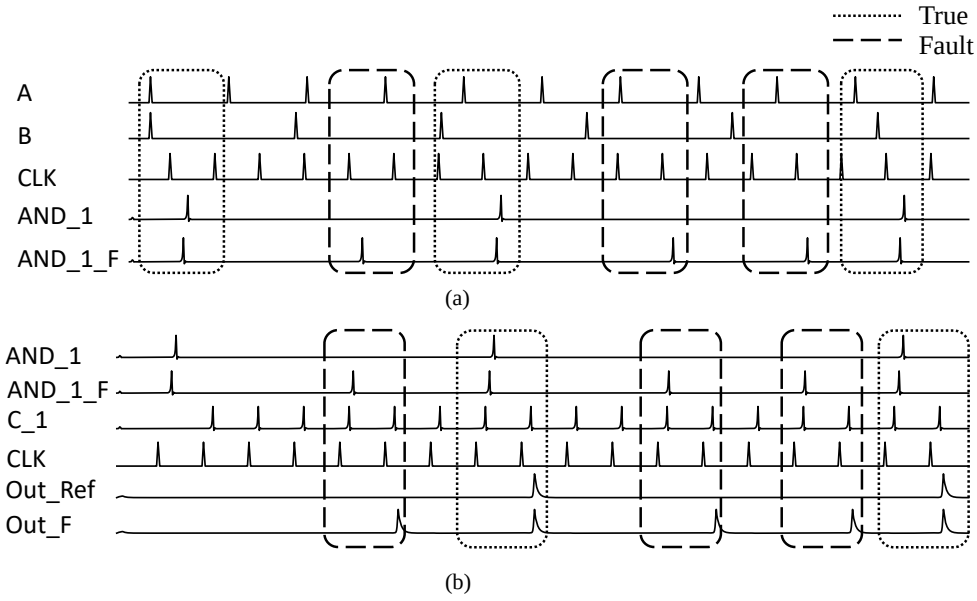


Figure 8.6: Validation of the proposed JJ-based fault models, where a single JJ fault is inserted at AND_1 gate. J2 is stuck at SC, as shown in Figure 8.7. The output of the reference cell without any faults, (a) at the first stage, faulty output AND_1_F is one when either A=1 and B=1 or A=1 and B=0, while the true operation AND_1 is one only when A=1 and B=1. (b) At the second stage, where no faults exist, but the faulty output of the first stage propagates to the second stage. This behavior exemplifies that JJ-based stuck at faults are localized faults that only affect the operation of a specific cell (the cell with the faulty JJ).

The second stage AND gate is free of faults. Hence, the source of the fault is only from AND_1.

Based on these validation results; 1) the proposed JJ-based high level fault models describe the operation of a cell with JJs stuck at OC or SC state. 2) The influence of JJ-based stuck at faults is localized within the gate where the fault exists. Hence, a JJ stuck at fault in a specific gate does not cause additional faults in other gates. Based on these remarks, the proposed technique of developing high-level fault models

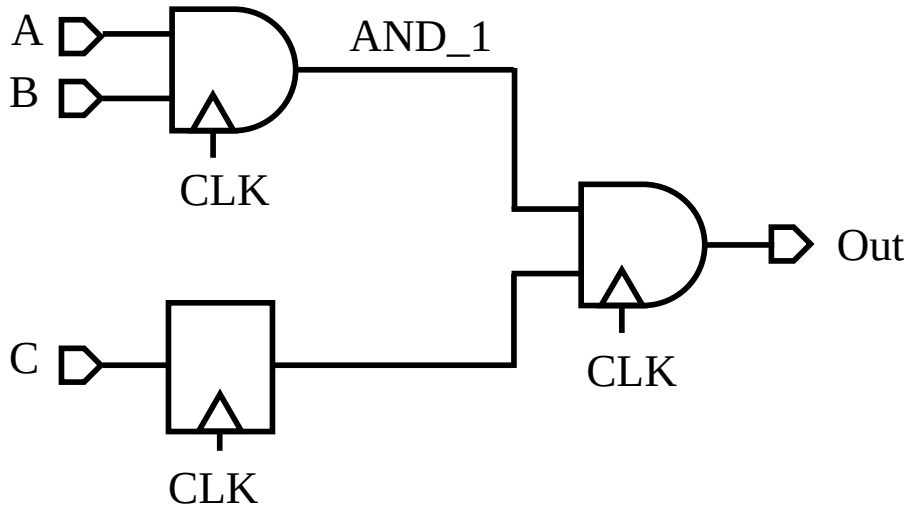


Figure 8.7: Benchmark circuit to evaluate JJ-based fault models. A single JJ fault is inserted in a two level cell. The output is compared with the predicted output based on the fault model.

can abstract JJ-based faults from a gate-level model to a block-level model. This technique can be used to develop a fault simulation tool for JJ-based stuck at faults.

8.3 Test Vector Generation

The required test vectors that identify the location and/or type of JJ-based fault are based on the proposed fault models. As an example, consider the JJ-based fault model of a JTL, as listed in Table 8.2(a). OC faults cannot be identified since the output of a faulty JTL with JJ1 stuck at OC state is similar to a reference cell. The output of a faulty JTL with a JJ stuck in the SC state can be identified by detecting zero when a one is inserted at the JTL. Another example, listed in Table 8.2(b), is the

location of stuck at SC faults which can be identified by detecting the two outputs of a splitter. If inserting a one into a splitter and detecting a zero at the two outputs, stuck at SC is identified at the driver JJ. If inserting a one into a splitter and detecting a one in one branch and a zero in the other, a stuck at SC is detected at the JJ located in the other branch.

Table 8.2: High level JJ-based fault models, (a) JTL, (b) splitter, and (c) DFF. The faulty output is highlighted as a gray cell. Two test vectors are required to detect the fault if J2 or J3 is stuck at SC. Detecting JJ-based faults within a DFF is achieved by applying up to three test vectors to set or reset the stored value within the storage loop of a DFF, regardless of the initial condition before testing.

(a)

In	Out		
	Ref	OC	SC
0	0	0	0
1	1	1	0

(b)

	In	Ref		OC		SC	
		Out1	Out2	Out1	Out2	Out1	Out2
J1	0	0	0	0	0	0	0
	1	1	1	1	1	0	0
J2	0	0	0	0	0	0	0
	11	11	11	11	11	00	10
J3	0	0	0	0	0	0	0
	10	11	11	11	11	10	00

(c)

Input	CLK	00	1	010	101
	In	11	0	101	010
Ref		00	0	010	001
J1	OC	00	0	010	001
	SC	00	0	000	000
J2	OC	00	0	000	000
	SC	00	0	010	001
J3	OC	01	0	010	001
	SC	00	0	000	000

A complete list of test vectors to detect the location and/or type of JJ-based fault within an SFQ system is listed in Table 8.3. Based on this list, different types of test vectors can be applied, as follows.

1) Test vectors can be applied to detect a specific fault at a specific location. For example, detecting if J1 (at a specific location) is stuck at SC (for a specific fault) within a splitter cell or if J3 is stuck at OC within a DFF cell.

2) Test vectors can be applied to detect if a specific location has a JJ-based fault (stuck at SC or OC); for example, detecting if J4 is stuck at SC or OC fault within an OR gate by applying A=10 and B=00.

3) Test vectors can be applied to detect if a JJ-based stuck at fault occurs within a cell (without identifying the type or location of the fault); for example, applying A=X and B=X to an OR cell, where X means any value.

8.4 Fault Coverage of JJ-based Faults

A summary of JJ-based faults within an SFQ system is listed in Table 8.4. Considering specific logic cells, 100% of OC faults and 64% of SC faults can be detected within an AND cell. 100% of JJs stuck at SC state within a splitter cell can be identified and located. It is, however, challenging to determine the type and/or location of all JJ stuck faults within JTL, DFF, and OR cells.

Table 8.3: Test vectors to detect the location and/or type of JJ-based faults within an SFQ cell. The JJ labels are illustrated in the circuit structures shown in Figures 8.3(a), 8.4(a), 8.5(a), 8.5(b), and 8.5(c).

Cell	Detected faults	In Test vector	Ref/ Ref1	Out/Out1	Ref2	Out2
JTL	J1 SC or J2 SC	1	1	0		
Splitter	J1 SC	11	1	00	1	00
	J2 SC	11	11	00	11	10
	J3 SC	11	11	10	11	00
DFF	J1 SC, J2 OC, or J3 SC	CLK 010 IN 101	01	00		
	J3 OC	CLK 00 In 11	00	01		
OR	J1 SC or J4 SC	A 1, B 0	1	0		
	J2 OC or J4 OC	A 0, B 0	0	1		
	J2 SC or J3 SC	A 0, B 1	1	0		
	J6 OC or J8 OC	CLK 0 A 1, B 0	0	1		
	J5 OC, J6 SC, J7 OC, and J8 SC	A 1, B 1	1	0		
AND	J1 SC	A 01, B 10	00	10		
	J2 SC	A 01, B 10	00	01		
	J3 OC	A 01, B 10	00	11		
	J1 OC, J2 OC, J3 SC, J4 OC, J4 SC, J5 OC, J6 OC, J6 SC, J7 OC, J7 SC, J8 OC, J9 OC, J9 SC, J10 OC, or J11 OC	A 1, B 1	1	0		

A total of 72% of JJ faults can be identified within an SFQ cell library composed of JTL, splitter, DFF, OR, and AND cells. Only 70% of OC faults can be identified, and 74% stuck at SC state faults can be determined. Only the location of 19% of stuck at SC state faults can be identified, while the location of 7% stuck at OC faults can be determined.

These numbers do not directly reflect a fault coverage of a cell but do reflect an estimate of the fault coverage of JJ-based faults. As an example, for the benchmark circuit presented in Figure 8.7, the circuit is composed of two AND gates and one

DFF with a total of 25 JJs with a possibility of 50 JJ-based faults. Assuming the ability to detect and observe the primary input and output of all of the logic gates, the following fault coverage can be achieved; 80% of JJ-based faults, 96% of OC faults, and 64% of SC faults.

Table 8.4: Summary of the fault coverage of JJ-based faults within multiple SFQ cells where the number of JJs within each cell, total number of JJ faults that may exist, number of total faults that can be detected, number of only OC or only SC faults that can be detected, and number of specific OC or SC fault that can be detected at a specific location

Cell	**#JJs**	**#Faults**	**Detected faults**			**Detected location**	
			#	**OC**	**SC**	**OC**	**SC**
JTL	2	4	2	0	2	0	0
Splitter	3	6	3	0	3	0	3
DFF	3	6	4	2	2	1	0
OR	8	16	12	6	6	0	0
AND	11	22	18	11	7	1	2
Total	27	54	39	19	20	2	5
Per cent			72%	70%	74%	7%	19%

These low fault coverage results are due to the response of SFQ cells to stuck OC faults. In most cases, JJs stuck at OC fault pass an SFQ pulse without interruption, such as the JJs within a JTL, as illustrated in Figure 8.3(b). In these scenarios, it is challenging to detect stuck at OC faults through the proposed high level functional model.

One method to enhance the fault coverage of stuck at OC faults can be achieved by applying a test methodology to detect and distinguish a transition failure between a faulty output of a stuck at OC fault and a reference cell. As illustrated in Figure 8.3(b), a small delay is detected between the faulty output of a JTL with a stuck at OC fault and a reference cell. Depending upon the location of a JJ, a different delay is detected. Transition faults have been widely used to model stuck-open faults for determining transistor-based faults in CMOS systems [167]. These on-chip transition-based testing mechanisms are challenging to design, require significant power, and increase the test time [167].

It is difficult to identify stuck at faults at the device level. In CMOS systems, gate-level or node-level stuck at faults are preferred over transistor stuck at faults since it is easier to detect and locate these faults with high fault coverage [168]. CMOS-based node or gate stuck-at faults affect several transistor terminals at the same time [164]. High fault coverage is to be expected. For most CMOS stuck at fault models, the coverage of a transistor-level stuck-at fault is significantly less than the fault model of gate-level stuck at faults [168]. A higher coverage is obtained for node stuck-at faults than gate stuck-at faults.

8.5 JJ-based Targeted Testing

A test vector generation algorithm is required to target a specific cell and fault. The objective of this algorithm is to provide a block-level fault model to generate test vectors to detect faults and determine the fault coverage within a block, as illustrated in Figure 8.8.

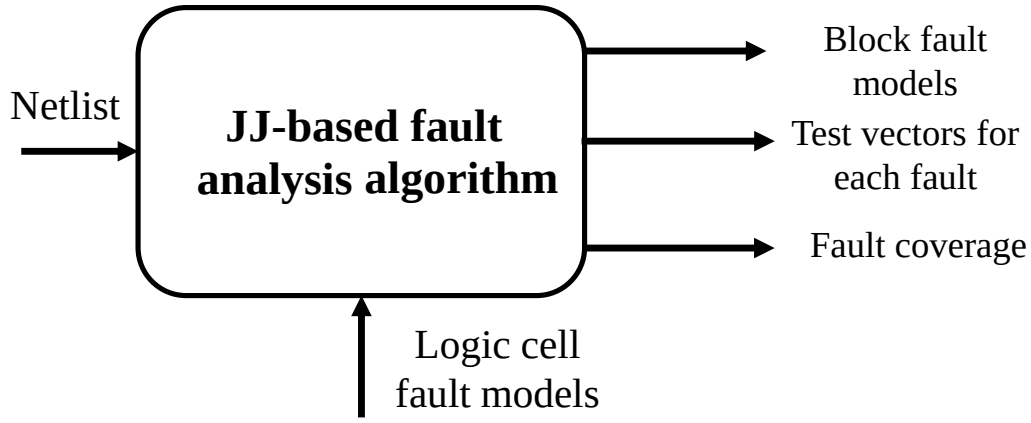


Figure 8.8: Block diagram of proposed algorithm to generate test vectors to identify JJ-based faults within an SFQ system

To increase the fault coverage of JJ-based faults within an SFQ system, a test methodology is necessary that generates the required test vectors to detect only stuck at SC faults or only OC faults. As an example, within an OR cell, J1, J2, J3, and J4 stuck at SC faults can be detected by applying two test vectors (A=10, B=01) with a clock pulse applied after each of the test vectors.

Pseudocode describing generating test vectors to target a specific fault type or specific fault location is shown in Algorithm 1. The proposed algorithm generates the

Algorithm 2 Pseudocode of Algorithm for Generating Test Vectors to Identify JJ-Based Faults

Input: Number of gates N, number of input conditions C, JJ-based fault models of all logic cells LM with JJ number JJn under JJ-based fault f of 0 for SC or 1 for OC fault, target testing requirements

Output: Block JJ-based fault model BM, test vector of target testing $Test\_Vector$, undetected faults DF, and detected faults DF

Evaluate the behavior of the netlist with reference cells under all input conditions Out_Ref
for $k \leftarrow 1$ to N **do** ▷ Evaluate each logic cell
  for $i \leftarrow 1$ to JJn **do** ▷ Evaluate each JJ-based fault within the logic cell
    for $j \leftarrow 1$ to C **do** ▷ Evaluate each input condition
      Evaluate the behavior of the netlist with LM_k under a fault in JJ i with fault type f under input condition j,
      $BM \leftarrow (Out, k, i, f, j)$
      if $(Out) = Out\_Ref$ **then**
        $UF \leftarrow (k, i, f, j)$ ▷ Undetected fault in K cell, i JJ, fault f, and input condition i
        $uf = uf + 1$
      else
        $DF \leftarrow (Out, k, i, f, j)$ ▷ Detected fault in K cell, i JJ, fault f, and input condition i
        $df = df + 1$
      end if
    end for
  end for
end for
Extract DF with $f = 0$ to group stuck at SC faults BM_{SC}
Extract DF with $f = 1$ to group stuck at SC faults BM_{OC}
Extract DF regardless f to group a stuck at a fault BM_U
Generate the test vector for each group $Test\_Vector_{SC}$, $Test\_Vector_{OC}$, $Test\_Vector_U$

required test vectors, as follows.

1) The block under test is evaluated for all possible fault scenarios. A fault is inserted. Each JJ is modeled as either stuck at SC or OC. A stuck at OC JJ is modeled as an OC, while a stuck at SC JJ is modeled as a JJ with a high critical current (such as 5 mA).

2) The output of each fault scenario is compared to a reference output (where no fault is inserted). Undetected faults are those faults where the faulty output is similar to the reference output.

3) All identical faulty outputs due to the same input combination(s) are grouped together. These faults share the same test vectors.

8.6 Summary

Advanced testing methodologies are required to support complex digital SFQ systems. In this chapter, JJ-based fault models are proposed for specific gate types. A faulty JJ has four modes of operation, stuck at superconductive, resistive, open circuit, or noisy switching. Two JJ-based fault modes are considered in this chapter, stuck at superconductive state and stuck at open circuit state. A JJ stuck in the superconductive state is modeled as a JJ with a high critical current, while a JJ stuck in the open circuit state is modeled as an open circuit. A high level JJ-based fault model is presented for the following RSFQ cells; JTL, splitter, DFF, OR, and AND.

Test vectors to identify the type and location of a set of faults are generated based on the high level fault models. The fault coverage of the OC and SC faults and the location of each logic cell are identified; specifically, 72% of JJ-based faults (OC, SC, or both) can be detected within an SFQ system. The fault coverage of a JJ-based fault is 74% of SC faults and 70% of OC faults. While it is challenging to identify the location of OC faults within SFQ system, all SC faults within a splitter cell can be identified and the location of 18% of SC faults within an AND cell can be determined. A methodology is also proposed to develop a block-level fault model to produce the required test vectors to identify the type and location of JJ faults within SFQ systems.

Chapter 9

Future Work

Multiple technologies are currently being considered to supplement conventional CMOS circuits, targeting certain heterogeneous applications. These technologies are at different stages of maturity (research, development, production). As an example, spintronic technology is currently in production as a nonvolatile memory while in development for sensor applications [169] and in research for use within compute applications [170]. Superconductive technologies are mature at the production stage for certain magnetic sense applications while in development for standard computing applications [171] and in research for quantum computing applications [171]. Significant effort is required for these technologies to be more widely adopted, as discussed in this chapter.

Numerous MTJ structures and switching mechanisms have recently been developed to support storage, computation, and sensory applications [127, 172, 173]. As discussed in section 9.1, future work should include the development of analytic models,

algorithms, and techniques targeting MTJ technology to enhance the performance efficiency of MTJ-based memory technologies at different levels of the memory hierarchy.

A basic structure of an MTJ is composed of two ferromagnetic layers separated by an insulating tunneling barrier. Additional layers have recently been added to further enhance the operation of an MTJ as a storage element. As an example, a capping layer is placed above the MTJ and buffer layers before the tunneling barrier [174]. Each of these layers affects the thermal dependence of the performance of the device. Additional research is necessary to further investigate the influence of the physical structure of an MTJ on thermal sensing applications, as discussed in section 9.2.

Superconductive electronics has the potential to vastly improve both the speed and power efficiency of stationary compute systems; particularly, data centers and supercomputers. Advanced testing methodologies are necessary to support complex SFQ systems. Additional research is required to complement the research results described in Chapters 7 and 8 in supporting DFT in SFQ systems. Advanced SFQ defects, such as pinholes and flux trapping, need to be investigated to improve the quality of the fault models to enhance the fault coverage and overall testability of SFQ systems, as discussed in section 9.3

9.1 MTJ-based Memory Hierarchy

Magnetic tunnel junctions are becoming a mainstream technology in support of modern storage systems. This improvement is due to the ability to maintain a magnetized state over long periods of time while scaling the technology. Electrically controlled MTJ devices are potentially suitable for a variety of applications, such as a replacement for random access memory (both DRAM and SRAM) due to low standby power and as a high speed write buffer for hard disk and solid-state drivers. Hybrid CMOS-MTJ technologies offer enhanced functionality as compared to CMOS. A combination of CMOS logic and memory with advanced magnetic technologies will improve the performance of a broad range of applications. Magnetic random access memory (MRAM) based on MTJ has been proposed as a universal memory for processing [127]. MRAM is expected to enhance the computational power, access time, and processing speed of data centers and supercomputers.

The objective of this research path is to enable the development of future heterogeneous systems for a wide variety of applications including the internet of things, deep learning, and big data, where high processing speed operating on significant amounts of data is essential. MTJ devices are both two terminal and three terminal structures with in-plane or perpendicular magnetization anisotropy and different magnetization mechanisms such as STT, VCMA, thermally assisted STT, SOT, hybrid VCMA/STT,

and hybrid STT/SOT [175]. Compact models that include advanced nanoscale effects, such as self-heating, failure mechanisms, and spin asymmetry torque, are required.

The focus of this research should be algorithms, methodologies, and test circuits to decide which MTJ technology to support different storage systems. This objective requires enhanced compact models for different MTJ devices (characterizing the electrical, thermal, and magnetic behavior). A methodology to integrate both the performance requirements of the storage systems and the advantages and drawbacks of MTJ technologies is desirable. Storage systems such as embedded memory, cache memory, and primary memory should be considered in this research path. The performance metrics should include power, speed, area, endurance rate, retention time, and compatibility with CMOS back-end-of-line processes.

9.2 MTJ Structures for Thermal Sensing

As previously discussed in chapter 4, a hybrid MTJ/CMOS-based thermal sensor can be achieved by considering the influence of temperature on both the threshold voltage of the transistor and the antiparallel resistance of the MTJ. Advanced MTJ structures are required to further enhance the thermal sensitivity and other performance metrics of MTJ-based thermal sensors.

The electrical conductance of an MTJ is composed of two components; a spin dependent (elastic) term due to the thermal excitation of the spin polarized electrons,

and a spin independent term (inelastic) due to scattering by defects and impurity states [90, 103, 104]. The spin independent conductance component is more sensitive to temperature than the spin dependent component. The antiparallel resistance of an MTJ decreases with increasing temperature because most of the contribution to the electrical conductance is due to the spin independent component. Hence, additional research is essential to determine the relative contribution of the spin dependent and spin independent conductance components to the total conductance of an MTJ in both the parallel and antiparallel states.

As shown in Figure 9.1(a), the structure of an MTJ is composed of two ferromagnetic layers (fixed and free) separated by a tunneling barrier. A cladding/capping layer is placed above the MTJ to provide a protection layer to the MTJ structure. Additional layers can be added later, as shown in Figure 9.1(b), to further improve the performance of an MTJ as a nonvolatile storage element. Objectives of this research path are to provide guidelines and recommendations for adopting the structure of an MTJ to serve thermal sensing applications.

Popular tunneling barriers for an MTJ are Al_2O_3 and MgO [127]. The tunneling barrier has a strong influence on both conductance components (elastic and inelastic) [103,104]. The tunneling barrier material and thickness determine the spin polarization and damping factor of an MTJ. A change in the spin polarization factor α_p of an MTJ increases the thermal sensitivity of an MTJ, as illustrated in Figure 9.2. Other

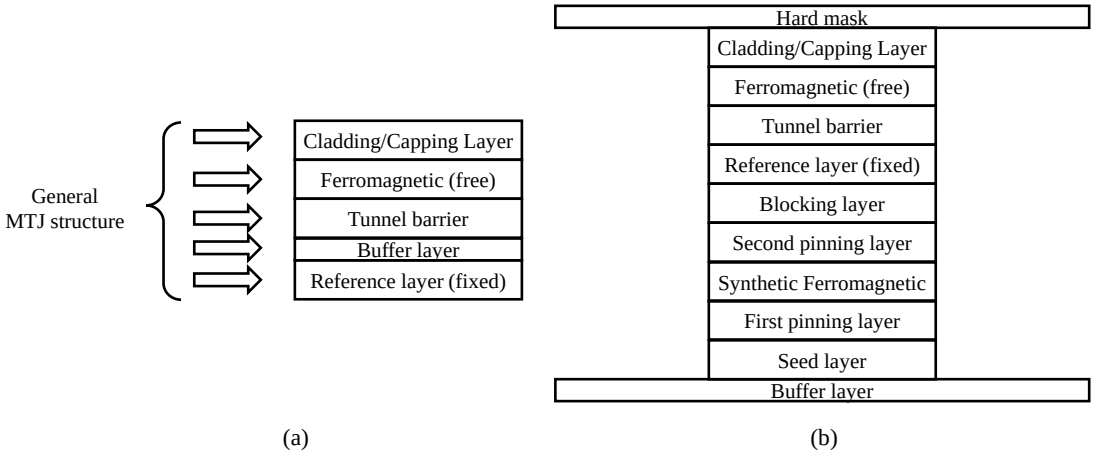


Figure 9.1: The structure of an MTJ, (a) basic structure is composed of a top cladding layer to protect the device and two ferromagnetic layers separated by a tunneling barrier, and (b) advanced structure of an MTJ by adding layers to pin the fixed ferromagnetic layer, enhancing the thermal stability of the device.

factors that affect the thermal behavior of an MTJ include the annealing temperature during manufacturing [176].

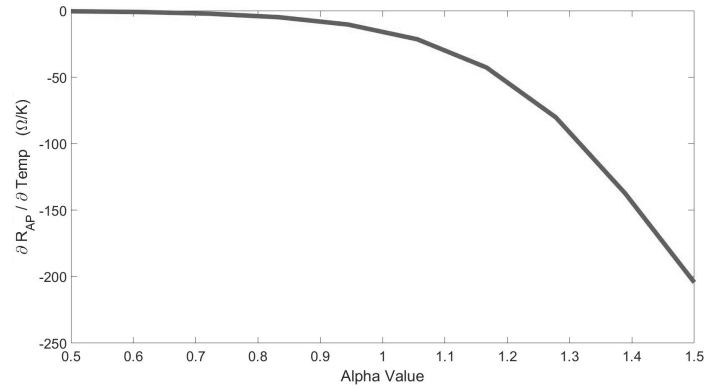


Figure 9.2: Effect of spin polarization in MTJ on the thermal sensitivity. For the same sense voltage, the higher the spin polarization parameter α, the higher the thermal sensitivity

9.3 Fault Models for SFQ Systems

The development of SFQ-based fault models is an essential step to achieve large scale integration of superconductive systems. Sophisticated failure mechanisms in SFQ systems such as pinhole defects and flux trapping are challenging to model and characterize. Both mechanisms strongly influence the overall behavior and performance of SFQ systems [177].

Hence, future research is required to model the influence of pinholes within a JJ on the variability of the device characteristics and the switching performance of a JJ, as described in subsection 9.3.1. Additional research is also required to model and locate flux trapping within SFQ systems, as discussed in subsection 9.3.2.

9.3.1 Fault models of pinholes

Pinholes can be modeled as a two level system defect (a system with a probability of existence in two different states) [178]. Pinholes exist in the tunneling barrier of a JJ due to the amorphous structure of the insulating barrier or atom dislocation, such as the movement of the atomic position of the oxygen in the oxide forming the barrier within a JJ [177, 179].

Pinholes can also exist due to the absence of the insulating oxide layer of a JJ. Low density pinholes within the oxide barrier affects the critical currents whose magnitude is far greater than the intended current.

Pinholes cause fluctuations in the critical current of a JJ due to deformation of the barrier within a JJ. Pinholes may occur during the manufacturing process or during normal operation of a JJ. Disorder and critical current variability in JJs due to the existence of pinholes have previously been studied [180]. Hence, the development of high-level fault models is essential to determine the influence of pinholes within a JJ on SFQ systems. These fault models can be used to characterize the effects of variability during switching of a JJ and also the influence of pinholes on circuit performance.

9.3.2 Fault models of flux trapping

Flux trapping dramatically affects the operation of superconductive electronic circuits [181]. A fault model describing the probability of flux being trapped at specific locations within a layout is critical to enhance the reliability and lifetime of SFQ systems.

Existing algorithms such as 3D-MLSI [182] can be used to extract the self- and mutual inductances around holes in superconductive films. Additional work exists on evaluating flux trapping within certain interconnect and layout structures [181, 183]. In these studies, flux trapping is incorporated into compact models for use in circuit simulation. A model describing the influence of a flux trapping fault at a specific location on the operation and overall performance of nearby SFQ circuits is however

missing. Hence, an essential research path is to develop a fault model to determine the influence of flux trapping on JJ circuit behavior and to develop techniques for detecting the location of flux trapping faults.

Chapter 10

Conclusions

Two different technological development paths are considered to enhance the performance of integrated circuits across multiple abstraction levels and functions (material, digital logic, memory, and system architecture). The first path is the classical development of CMOS technologies by geometric scaling. Scaling has significantly slowed. The demand for functionally diverse heterogeneous integrated circuits remains and is increasing. Hence, the second path is beyond CMOS where microelectronics is extended to emerging technologies to achieve next generation applications. These new applications can be realized by bridging the gap among novel emerging devices, unconventional architectures, and advanced computing schemes to support or replace conventional CMOS technologies.

Beyond CMOS technologies exploit innovative materials and novel device structures. Some of these beyond CMOS solutions are intended to be integrated onto a silicon platform to exploit established CMOS-based infrastructures. Other technologies, such

as superconductive electronics, are considered a promising standalone technology. Although beyond CMOS devices exhibit a wide range of functions that can replace or support conventional CMOS systems, these devices also exhibit reliability issues that should be identified and addressed early in the technology development process.

To achieve next generation applications using beyond CMOS technologies, two research questions are highlighted in this dissertation. 1) How can emerging technologies support beyond CMOS compute systems? 2) How to identify, address, and solve reliability issues in these technologies?

In this dissertation, spintronic technology based on magnetic tunnel junctions is investigated to address the first question. Due to the compatibility with CMOS fabrication, design simplicity, size advantage, nonvolatility, and low standby power, spintronic technology is under development to support beyond CMOS compute systems. Superconductive single flux quantum logic based on Josephson junctions is a solution to the second question. Due to the unique challenges that exist in SFQ systems, such as high clock frequencies and cryogenic environment, design for testability methodologies are necessary for SFQ systems.

The unique characteristics of MTJs are considered in this dissertation to develop non-conventional hybrid MTJ/CMOS compute systems, including self-aware compute systems, compute in-memory, reconfigurable logic, and distributed compute systems.

A thermal aware system is achieved by distributing a large number of on-chip thermal sensors. These on-chip thermal sensors should be small in size, low power, high speed, temperature sensitive, and accurate over a wide temperature range. A hybrid MTJ/CMOS-based analog thermal sensor is proposed here where the high temperature sensitivity of the MTJ antiparallel resistance is exploited. The proposed thermal aware system is a network of thermal sensor nodes communicating with a control unit that collects temperature data and produces a thermal map. This thermal network provides the monitored system with dynamic real-time thermal information. The proposed system provides flexibility in choosing a threshold temperature. The system can also support multi-threshold sensing schemes. This capability can be achieved by multiplexing a reference voltage. At each reference voltage, the system identifies whether the temperature at a sensor node is above or below a certain threshold temperature.

Many integrated systems have become data centric, where a huge amount of data are collected and processed in real-time. In data centric architectures, data motion is greatly decreased by integrating the computational process within the storage system at different levels of the memory hierarchy. This capability for *in situ* computation can be achieved by the proposed double MTJ (DMTJ) circuit topology. A multi-bit nonvolatile AND, OR, and NOT logic gate and memory cell is available within DMTJ. DMTJ exhibits a state transition diagram between four resistance states. A write cell

is required to transition the states within a DMTJ to provide a two step memory and logic function, composed of an initial reset step followed by a calculate or memory step.

Superconductive electronics target large scale, stationary systems where two to three orders of magnitude improvement in energy efficiency is available as compared to conventional semiconductor-based supercomputers. The challenge of achieving high performance with high reliability SFQ systems is escalating due to dimensional scaling, novel materials and devices, and operation in severe conditions (extreme cryogenic temperatures and sub-terahertz frequencies). Advanced design for testability techniques are necessary to determine SFQ-based defects and faults and improve the ability to evaluate these faults.

Embedded hardware solutions, such as a test extraction module and a hybrid test module, are proposed to enhance the controllability and observability of the internal nodes within an SFQ system to identify specific defects and faults. An algorithm is also proposed to describe the methodology and tradeoffs of inserting test modules into SFQ systems. These test modules significantly enhance by more than 50% the testability measures (controllability and observability) of the internal nodes, increasing overall fault coverage.

A methodology is described here to include DFT within SFQ systems. This objective is achieved by developing high level fault models that target JJ-based

faults, stuck at a superconductive state or an open circuit state. A JJ stuck in the superconductive state is modeled as a JJ with a high critical current, while a JJ stuck in the open circuit state is modeled as an open circuit. A high level JJ-based fault model is presented for the following SFQ cells; JTL, splitter, DFF, OR, and AND. A methodology is further proposed to develop a block-level fault model to produce the required test vectors to identify the type and location of certain JJ faults within an SFQ system.

The topics presented in this dissertation are intended to propose different solutions to apply beyond CMOS technologies to support advanced compute systems, while providing physical perspective and engineering insight into the many challenges faced by these technologies. Design methodologies and circuit techniques targeting the unique physical properties of these emerging technologies are necessary to develop the next generation of application-specific compute systems.

Bibliography

[1] S. Salahuddin, K. Ni, and S. Datta, "The Era of Hyper-Scaling in Electronics," *Nature Electronics*, Vol. 1, No. 8, pp. 442–450, August 2018.

[2] A. Sheibanyrad, F. Pétrot, and A. Jantsch, *3D Integration for NoC-Based SoC Architectures*, Springer, 2011.

[3] A. C. Seabaugh and Q. Zhang, "Low-Voltage Tunnel Transistors for Beyond CMOS Logic," *Proceedings of the IEEE*, Vol. 98, No. 12, pp. 2095–2110, December 2010.

[4] H. D. Nguyen, J. Yu, L. Xie, M. Taouil, S. Hamdioui, and D. Fey, "Memristive Devices for Computing: Beyond CMOS and Beyond Von Neumann," *Proceedings of the IEEE International Conference on Very Large Scale Integration*. pp. 1–10, October 2017.

[5] H. Liu, S. Manipatruni, D. H. Morris, K. Vaidyanathan, D. E. Nikonov, T. Karnik, and I. A. Young, "Synchronous Circuit Design With Beyond-CMOS Magnetoelectric Spin–Orbit Devices Toward 100-mV Logic," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, Vol. 5, No. 1, pp. 1–9, June 2019.

[6] Z. Luo, M. Yang, Y. Liu, and M. Alexe, "Emerging Opportunities for 2D Semiconductor/Ferroelectric Transistor-Structure Devices," *Advanced Materials*, Vol. 33, No. 12, pp. 2005620.1–2005620.26, March 2021.

[7] D. E. Nikonov and I. A. Young, "Benchmarking of Beyond-CMOS Exploratory Devices for Logic Integrated Circuits," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, Vol. 1, pp. 3–11, April 2015.

[8] C. Pan and A. Naeemi, "Non-Boolean Computing Benchmarking for Beyond-CMOS Devices Based on Cellular Neural Network," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, Vol. 2, No. August 2016, pp. 36–43, December 2016.

[9] C. Pan and A. Naeemi, "An Expanded Benchmarking of Beyond-CMOS Devices Based on Boolean and Neuromorphic Representative Circuits," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, Vol. 3, pp. 101–110, December 2017.

[10] D. E. Nikonov and I. A. Young, "Uniform Methodology for Benchmarking Beyond-CMOS Logic Devices," *Proceedings of the IEEE International Electron Devices Meeting*, pp. 25.4.1–25.4.4, December 2012.

[11] H. H. Radamson, H. Zhu, Z. Wu, X. He, H. Lin, J. Liu, J. Xiang, Z. Kong, W. Xiong, J. Li, H. Cui, J. Gao, H. Yang, Y. Du, B. Xu, B. Li, X. Zhao, J. Yu, Y. Dong, and G. Wang, "State of the Art and Future Perspectives in Advanced CMOS Technology," *Nanomaterials*, Vol. 10, No. 8, pp. 1555.1–1555.86, August 2020.

[12] D. S. Holmes and E. P. DeBenedictis, "Superconductor Electronics and the International Roadmap for Devices and Systems," *Proceedings of the IEEE International Superconductive Electronics Conference*, pp. 1–3, June 2017.

[13] M. Dorojevets, Z. Chen, C. L. Ayala, and A. K. Kasperek, "Towards 32-bit Energy-Efficient Superconductor RQL Processors: The Cell-Level Design and Analysis of Key Processing and On-Chip Storage Units," *IEEE Transactions on Applied Superconductivity*, Vol. 25, No. 3, pp. 1–8, June 2015.

[14] N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Adiabatic Quantum-Flux-Parametron Cell Library Adopting Minimalist Design," *Journal of Applied Physics*, Vol. 117, No. 17, pp. 173912.1–173912.7, May 2015.

[15] O. Chen, R. Cai, Y. Wang, F. Ke, T. Yamae, R. Saito, N. Takeuchi, and N. Yoshikawa, "Adiabatic Quantum-Flux-Parametron: Towards Building Extremely Energy-Efficient Circuits and Systems," *Scientific Reports*, Vol. 9, No. 1, pp. 1–10, July 2019.

[16] A. Chen, S. Datta, X. S. Hu, M. T. Niemier, T. S. Rosing, and J. J. Yang, "A Survey on Architecture Advances Enabled by Emerging Beyond-CMOS Technologies," *IEEE Design and Test*, Vol. 36, No. 3, pp. 46–68, June 2019.

[17] B. Dieny, I. L. Prejbeanu, K. Garello, P. Gambardella, P. Freitas, R. Lehndorff, W. Raberg, U. Ebels, S. O. Demokritov, J. Akerman, A. Deac, P. Pirro, C. Adelmann, A. Anane, A. V. Chumak, A. Hiroata, S. Mangin, M. C. Onbasli, M. d. Aquino, G. Prenat, G. Finocchio, L. L. Diaz, R. Chantrell, O. C. Fesenko, and P. Bortolotti, "Opportunities and Challenges for Spintronics in the Microelectronic Industry," *Nature Electronics*, Vol. 3, No. 8, pp. 446–459, August 2019.

[18] J. M. Shalf and R. Leland, "Computing Beyond Moore's Law," *Computer*, Vol. 48, No. 12, pp. 14–23, December 2015.

[19] E. Y. Vedmedenko, R. K. Kawakami, D. D. Sheka, P. Gambardella, A. Kirilyuk, A. Hirohata, C. Binek, O. Chubykalo-Fesenko, S. Sanvito, B. J. Kirby, J. Grollier, K. Everschor-Sitte, T. Kampfrath, C. Y. You, and A. Berger, "The 2020 Magnetism Roadmap," *Journal of Physics D: Applied Physics*, Vol. 53, No. 45, pp. 453001, August 2020.

[20] P. Barla, V. K. Joshi, and S. Bhat, "Spintronic Devices: A Promising Alternative to CMOS Devices," *Journal of Computational Electronics*, Vol. 20, No. 2, pp. 805–837, April 2021.

[21] R. Zand, A. Roohi, D. Fan, and R. F. DeMara, "Energy-Efficient Nonvolatile Reconfigurable Logic Using Spin Hall Effect-Based Lookup Tables," *IEEE Transactions on Nanotechnology*, Vol. 16, No. 1, pp. 32–43, January 2017.

[22] T. Hanyu, "Standby-Power-Free Integrated Circuits Using MTJ-Based VLSI Computing for IoT Applications," *Proceedings of the IEEE Berkeley Symposium on Energy Efficient Electronic Systems & Steep Transistors Workshop*, pp. 1–3, October 2017.

[23] T. Endoh, H. Koike, S. Ikeda, T. Hanyu, and H. Ohno, "An Overview of Nonvolatile Emerging Memories— Spintronics for Working Memories," *IEEE*

Journal on Emerging and Selected Topics in Circuits and Systems, Vol. 6, No. 2, pp. 109–119, June 2016.

[24] P. Dehraj and A. Sharma, "A Review on Architecture and Models for Autonomic Software Systems," *The Journal of Supercomputing*, Vol. 77, No. 1, pp. 388–417, January 2021.

[25] J. Henkel, "Self-Awareness in Systems on Chip, Part II," *IEEE Design and Test*, Vol. 35, No. 5, pp. 4–4, October 2018.

[26] C. E. Merkel and D. Kudithipudi, "Towards Thermal Profiling in CMOS/Memristor Hybrid RRAM Architectures," *Proceedings of the IEEE International Conference on VLSI Design*, pp. 167–172, January 2012.

[27] Y. S. Chung, R. W. Baird, and M. A. Dulram, *Magnetic Tunnel Junction Temperature Sensors and Methods*, U.S. Patent No. 7,510,883, March 31, 2009.

[28] A. Sengupta and K. Roy, "Encoding Neural and Synaptic Functionalities in Electron Spin: A Pathway to Efficient Heuromorphic Computing," *Applied Physics Reviews*, Vol. 4, No. 4, pp. 041105.1–041105.23, December 2017.

[29] Y. Zhang, J. Ren, J. Liu, C. Xu, H. Guo, and Y. Liu, "A Survey on Emerging Computing Paradigms for Big Data," *Chinese Journal of Electronics*, Vol. 26, No. 1, pp. 1–12, January 2017.

[30] S. Kvatinsky, G. Satat, N. Wald, E. G. Friedman, A. Kolodny, and U. C. Weiser, "Memristor-Based Material Implication (IMPLY) Logic: Design Principles and Methodologies," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 22, No. 10, pp. 2054–2066, October 2013.

[31] Y. Zhou, Y. Li, L. Xu, S. Zhong, H. Sun, and X. Miao, "16 Boolean Logics in Three Steps With Two Anti-Serially Connected Memristors," *Applied Physics Letters*, Vol. 106, No. 23, pp. 233502.1–233502.4, June 2015.

[32] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The Missing Memristor Found," *Nature*, Vol. 453, No. 7191, pp. 80–83, May 2008.

[33] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov, "Training and Operation of an Integrated Neuromorphic Network Based on Metal-Oxide Memristors," *Nature*, Vol. 521, No. 7550, pp. 61–64, May 2015.

[34] S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, C. di Nolfo, S. Sidler, M. Giordano, M. Bodini, N. C. P. Farinha, B. Killeen, C. Cheng, Y. Jaoudi, and G. W. Burr, "Equivalent-Accuracy Accelerated Neural-Network Training Using Analogue Memory," *Nature*, Vol. 558, No. 7708, pp. 60–67, June 2018.

[35] S. R. Nandakumar, M. Le Gallo, I. Boybat, B. Rajendran, A. Sebastian, and E. Eleftheriou, "A Phase-Change Memory Model for Neuromorphic Computing," *Journal of Applied Physics*, Vol. 124, No. 15, pp. 152135.1–152135.11, October 2018.

[36] S. G. Ramasubramanian, R. Venkatesan, M. Sharad, K. Roy, and A. Raghunathan, "SPINDLE: SPINtronic Deep Learning Engine for Large-Scale Neuromorphic Computing," *Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design*, pp. 15–20, August 2014.

[37] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, and J. Grollier, "Neuromorphic Computing with Nanoscale Spintronic Oscillators," *Nature*, Vol. 547, No. 7664, pp. 428–431, July 2017.

[38] Y. van de Burgt, E. Lubberman, E. J. Fuller, S. T. Keene, G. C. Faria, S. Agarwal, M. J. Marinella, A. Alec Talin, and A. Salleo, "A Non-Volatile Organic Electrochemical Device as a Low-Voltage Artificial Synapse for Neuromorphic Computing," *Nature Materials*, Vol. 16, No. 4, pp. 414–418, April 2017.

[39] M. Jerry, P.-Y. Chen, J. Zhang, P. Sharma, K. Ni, S. Yu, and S. Datta, "Ferroelectric FET Analog Synapse for Acceleration of Deep Neural Network Training," *Proceedings of the IEEE International Electron Devices Meeting*, pp. 6.2.1–6.2.4, December 2017.

[40] R. Cheng, U. S. Goteti, and M. C. Hamilton, "Superconducting Neuromorphic Computing using Quantum Phase-Slip Junctions," *IEEE Transactions on Applied Superconductivity*, Vol. 29, No. 5, pp. 1–5, August 2019.

[41] R. Cheng, U. S. Goteti, and M. C. Hamilton, "High-Speed and Low-Power Superconducting Neuromorphic Circuits Based on Quantum Phase-Slip Junctions," *IEEE Transactions on Applied Superconductivity*, Vol. 31, No. 5, pp. 1–8, August 2021.

[42] W. Buchanan and A. Woodward, "Will Quantum Computers Be the End of Public Key Encryption?," *Journal of Cyber Security Technology*, Vol. 1, No. 1, pp. 1–22, January 2017.

[43] K. Fisher, A. Broadbent, L. K. Shalm, Z. Yan, J. Lavoie, R. Prevedel, T. Jennewein, and K. J. Resch, "Quantum Computing on Encrypted Data," *Nature Communications*, Vol. 5, pp. 4074.1–4074.7, September 2014.

[44] S. Bhatti, R. Sbiaa, A. Hirohata, H. Ohno, S. Fukami, and S. Piramanayagam, "Spintronics Based Random Access Memory: A Review," *Materials Today*, Vol. 20, No. 9, pp. 530–548, November 2017.

[45] D. Saida, Y. Yamanashi, M. Hidaka, F. Hirayama, K. Imafuku, S. Nagasawa, and S. Kawabata, "Experimental Demonstrations of Native Implementation of Boolean Logic Hamiltonian in a Superconducting Quantum Annealer," *IEEE Transactions on Quantum Engineering*, Vol. 2, pp. 1–8, August 2021.

[46] W. D. Kalfus, D. F. Lee, G. J. Ribeill, S. D. Fallek, A. Wagner, B. Donovan, D. Riste, and T. A. Ohki, "High-Fidelity Control of Superconducting Qubits Using Direct Microwave Synthesis in Higher Nyquist Zones," *IEEE Transactions on Quantum Engineering*, Vol. 1, pp. 1–12, January 2020.

[47] C. Monzio Compagnoni and A. S. Spinelli, "Reliability of NAND Flash Arrays: A Review of What the 2-D-to-3-D Transition Meant," *IEEE Transactions on Electron Devices*, Vol. 66, No. 11, pp. 4504–4516, November 2019.

[48] P. Fantini, "Phase Change Memory Applications: The History, the Present and the Future," *Journal of Physics D: Applied Physics*, Vol. 53, No. 28, pp. 283002, May 2020.

[49] Y. Chen, "ReRAM: History, Status, and Future," *IEEE Transactions on Electron Devices*, Vol. 67, No. 4, pp. 1420–1433, April 2020.

[50] S. Senni, L. Torres, G. Sassatelli, A. Gamatie, and B. Mussard, "Exploring MRAM Technologies for Energy Efficient Systems-On-Chip," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, Vol. 6, No. 3, pp. 279–292, September 2016.

[51] M. Julliere, "Tunneling between Ferromagnetic Films," *Physics Letters A*, Vol. 54, No. 3, pp. 225–226, September 1975.

[52] T. Miyazaki and N. Tezuka, "Giant Magnetic Tunneling Effect in Fe/Al2O3/Fe Junction," *Journal of Magnetism and Magnetic Materials*, Vol. 139, No. 3, pp. 94–97, January 1995.

[53] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, "Large Magnetoresistance at Room Temperature in Ferromagnetic Thin Film Tunnel Junctions," *Physical Review Letters*, Vol. 74, No. 16, pp. 3273–3276, April 1995.

[54] J. C. Slonczewski, "Current-Driven Excitation of Magnetic Multilayers," *Journal of Magnetism and Magnetic Materials*, Vol. 159, No. 1-2, pp. L1—-L7, June 1996.

[55] T. L. Gilbert and H. Ekstein, "Basis of the Domain Structure Variational Principle," *The Bulletin of the American Physical Society*, Vol. 1, pp. 25, November 1956.

[56] T. Gilbert, "Classics in Magnetics A Phenomenological Theory of Damping in Ferromagnetic Materials," *IEEE Transactions on Magnetics*, Vol. 40, No. 6, pp. 3443–3449, November 2004.

[57] K. Yakushiji, T. Saruya, H. Kubota, A. Fukushima, T. Nagahama, S. Yuasa, and K. Ando, "Ultrathin Co/Pt and Co/Pd Superlattice Films for MgO-Based Perpendicular Magnetic Tunnel Junctions," *Applied Physics Letters*, Vol. 97, No. 23, pp. 232508.1–232508.3, December 2010.

[58] R. Patel, X. Guo, Q. Guo, E. Ipek, and E. G. Friedman, "Reducing Switching Latency and Energy in STT-MRAM Caches with Field-Assisted Writing," *IEEE*

Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 24, No. 1, pp. 129–138, January 2016.

[59] S. Bandiera and B. Dieny, "Thermally Assisted MRAM," *Handbook of Spintronics*, Vol. 19, No. 16, pp. 1065–1100, April 2015.

[60] A. V. Khvalkovskiy, D. Apalkov, S. Watts, R. Chepulskii, R. S. Beach, A. Ong, X. Tang, A. Driskill-Smith, W. H. Butler, P. B. Visscher, D. Lottis, E. Chen, V. Nikitin, and M. Krounbi, "Basic Principles of STT-MRAM Cell Operation in Memory Arrays," *Journal of Physics D: Applied Physics*, Vol. 46, No. 8, pp. 074001.1–074001.21, February 2013.

[61] H. Liu, D. Bedau, D. Backes, J. A. Katine, J. Langer, and A. D. Kent, "Ultrafast Switching in Magnetic Tunnel Junction Based Orthogonal Spin Transfer Devices," *Applied Physics Letters*, Vol. 97, No. 24, pp. 242510.1–242510.3, December 2010.

[62] E. C. Stoner and E. P. Wohlfarth, "A Mechanism of Magnetic Hysteresis in Heterogeneous Alloys," *IEEE Transactions on Magnetics*, Vol. 27, No. 4, pp. 3475–3518, May 1991.

[63] L. Savtchenko, B. N. Engel, N. D. Rizzo, M. F. Deherrera, and J. A. Janesky, *Method of Writing to Scalable Magnetoresistance Random Access Memory Element*, U.S. Patent No. 6,545,906, October 2003.

[64] B. Dieny, R. Goldfarb, and K. Lee, *Introduction to Magnetic Random-Access Memory*, Wiley, 2016.

[65] Z. Diao, A. Panchula, Y. Ding, M. Pakala, S. Wang, Z. Li, D. Apalkov, H. Nagai, A. Driskill-Smith, L. C. Wang, E. Chen, and Y. Huai, "Spin Transfer Switching in Dual MgO Magnetic Tunnel Junctions," *Applied Physics Letters*, Vol. 90, No. 13, pp. 132508.1–132508.3, March 2007.

[66] D. Houssameddine, U. Ebels, B. Delaët, B. Rodmacq, I. Firastrau, F. Ponthenier, M. Brunet, C. Thirion, J. P. Michel, L. Prejbeanu-Buda, M. C. Cyrille, O. Redon, and B. Dieny, "Spin-Torque Oscillator using a Perpendicular Polarizer and a Planar Free Layer," *Nature Materials*, Vol. 6, No. 6, pp. 447–453, June 2007.

[67] S. Mangin, D. Ravelosona, J. Katine, M. Carey, B. Terris, and E. E. Fullerton, "Current-Induced Magnetization Reversal in Nanopillars with Perpendicular Anisotropy," *Nature Materials*, Vol. 5, No. 3, pp. 210–215, March 2006.

[68] J. Wang and P. P. Freitas, "Low-Current Blocking Temperature Writing of Double Barrier Magnetic Random Access Memory Cells," *Applied Physics Letters*, Vol. 84, No. 6, pp. 945–947, February 2004.

[69] E. Gapihan, R. C. Sousa, J. Hérault, C. Papusoi, M. T. Delaye, B. Dieny, I. L. Prejbeanu, C. Ducruet, C. Portemont, K. MacKay, and J. P. Nozières, "FeMn Exchange Biased Storage Layer for Thermally Assisted MRAM," *IEEE Transactions on Magnetics*, Vol. 46, No. 6, pp. 2486–2488, June 2010.

[70] X. Fong, Y. Kim, R. Venkatesan, S. H. Choday, A. Raghunathan, and K. Roy, "Spin-Transfer Torque Memories: Devices, Circuits, and Systems," *Proceedings of the IEEE*, Vol. 104, No. 7, pp. 1449–1488, July 2016.

[71] H. Cai, Y. Wang, L. A. De Barros Naviner, J. Yang, and W. Zhao, "Exploring Hybrid STT-MTJ/CMOS Energy Solution in Near-/Sub-Threshold Regime for IoT Applications," *IEEE Transactions on Magnetics*, Vol. 54, No. 2, pp. 1–9, February 2018.

[72] J. Heidecker, "MRAM Technology Status: NASA Electronic Parts and Packaging (NEPP) Program Office of Safety and Mission Assurance," Technical Report, Jet Propulsion Laboratory, Pasadena, California, February 2013.

[73] S. Salehi, D. Fan, and R. F. Demara, "Survey of STT-MRAM Cell Design Strategies," *ACM Journal on Emerging Technologies in Computing Systems*, Vol. 13, No. 3, pp. 1–16, April 2017.

[74] R. De Rose, G. Carangelo, M. Lanuzza, F. Crupi, G. Finocchio, and M. Carpentieri, "Impact of Voltage Scaling on STT-MRAMs Through a Variability-Aware Simulation Framework," *Proceedings of the IEEE International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design*, pp. 1–4, June 2017.

[75] B. Del Bel, J. Kim, C. H. Kim, and S. S. Sapatnekar, "Improving STT-MRAM Density Through Multibit Error Correction," *Proceedings of the IEEE Design, Automation & Test in Europe Conference & Exhibition*, pp. 1–6, March 2014.

[76] S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. M. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, and H. Ohno, "Tunnel Magnetoresistance of 604% at 300 K by Suppression of Ta Diffusion in CoFeBMgOCoFeB Pseudo-Spin-Valves Annealed at High Temperature," *Applied Physics Letters*, Vol. 93, No. 8, pp. 082508.1–082508.3, August 2008.

[77] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, "Giant Room-Temperature Magnetoresistance in Single-Crystal Fe/MgO/Fe Magnetic Tunnel Junctions," *Nature Materials*, Vol. 3, No. 12, pp. 868–871, December 2004.

[78] S. S. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and S. H. Yang, "Giant Tunnelling Magnetoresistance at Room Temperature with MgO (100) Tunnel Barriers," *Nature Materials*, Vol. 3, No. 12, pp. 862–867, December 2004.

[79] S. Ikeda, H. Sato, M. Yamanouchi, H. Gan, K. Miura, K. Mizunuma, S. Kanai, S. Fukami, F. Matsukura, N. Kasai, and H. Ohno, "Recent Progress of Perpendicular Anisotropy Magnetic Tunnel Junctions for Nonvolatile VLSI," *Spin*, Vol. 02, No. 03, pp. 1240003.1–1240003.12, September 2012.

[80] A. E. Council, "Failure Mechanism Based Stress Test Qualification for Integrated Circuits," Technical Report, Automotive Electronics Council, Automotive Electronics Council, May 2014.

[81] A. Jog, A. K. Mishra, C. Xu, Y. Xie, V. Narayanan, R. Iyer, and C. R. Das, "Cache Revive: Architecting Volatile STT-RAM Caches for Enhanced Performance in CMPs," *Proceedings of the ACM/IEEE Design Automation Conference*, pp. 243–252, June 2012.

[82] J. Zhou, T. Wei, M. Chen, J. Yan, X. S. Hu, and Y. Ma, "Thermal-Aware Task Scheduling for Energy Minimization in Heterogeneous Real-Time MPSoC Systems," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 35, No. 8, pp. 1269–1282, August 2016.

[83] T. Kuglestadt, "Semiconductor Temperature Sensors Challenge Precision RTDs and Thermistors in Building Automation," *Texas Instruments: Application Report: SNAA267–04*, pp. 2–10, 2015.

[84] S. Schafer, D. Apalkov, A. V. Khvalkovskiy, V. Nikitin, R. Beach, and Z. Duan, *Method and System for Determining Temperature Using a Magnetic Junction*, U.S. Patent No. 10,297,300, April 2016.

[85] A. G. Qoutb and E. G. Friedman, "MTJ Magnetization Switching Mechanisms for IoT Applications," *Proceedings of the ACM/IEEE on Great Lakes Symposium on VLSI*, pp. 347–352, April 2018.

[86] J. G. A. Vinasco, *Voltage-Controlled Magnetic Dynamics in Nanoscale Magnetic Tunnel Junctions*, University of California, Los Angeles, March 2014.

[87] W. Kang, Y. Ran, Y. Zhang, W. Lv, and W. Zhao, "Modeling and Exploration of the Voltage-Controlled Magnetic Anisotropy Effect for the Next-Generation Low-Power and High-Speed MRAM Applications," *IEEE Transactions on Nanotechnology*, Vol. 16, No. 3, pp. 387–395, May 2017.

[88] S. Zhang, P. M. Levy, A. C. Marley, and S. S. Parkin, "Quenching of Magnetoresistance by Hot Electrons in Magnetic Tunnel Junctions," *Physical Review Letters*, Vol. 79, No. 19, pp. 3744–3747, November 1997.

[89] Y. Wang, H. Cai, L. A. Naviner, Y. Zhang, J. O. Klein, and W. S. Zhao, "Compact Thermal Modeling of Spin Transfer Torque Magnetic Tunnel Junction," *Microelectronics Reliability*, Vol. 55, No. 9-10, pp. 1649–1653, August 2015.

[90] W. F. Brinkman, R. C. Dynes, and J. M. Rowell, "Tunneling Conductance of Asymmetrical Barriers," *Journal of Applied Physics*, Vol. 41, No. 5, pp. 1915–1921, April 1970.

[91] C. H. Shang, J. Nowak, R. Jansen, and J. S. Moodera, "Temperature Dependence of Magnetoresistance and Surface Magnetization in Ferromagnetic Tunnel Junctions," *Physical Review B - Condensed Matter and Materials Physics*, Vol. 58, No. 6, pp. R2917–R2920, August 1998.

[92] L. Yuan, S. H. Liou, and D. Wang, "Temperature Dependence of Magnetoresistance in Magnetic Tunnel Junctions with Different Free Layer Structures," *Physical Review B*, Vol. 73, No. 13, pp. 134403.1–134403.8, April 2006.

[93] J. G. Alzate, P. K. Amiri, G. Yu, P. Upadhyaya, J. A. Katine, J. Langer, B. Ocker, I. N. Krivorotov, and K. L. Wang, "Temperature Dependence of the Voltage-Controlled Perpendicular Anisotropy in Nanoscale MgO|CoFeB|Ta Magnetic Tunnel Junctions," *Applied Physics Letters*, Vol. 104, No. 11, pp. 112410.1–112410.5, March 2014.

[94] L. Zhang, Y. Cheng, W. Kang, L. Torres, Y. Zhang, A. Todri-Sanial, and W. Zhao, "Addressing the Thermal Issues of STT-MRAM From Compact Modeling to Design Techniques," *IEEE Transactions on Nanotechnology*, Vol. 17, No. 2, pp. 345–352, March 2018.

[95] M. Kazemi, E. Ipek, and E. G. Friedman, "Adaptive Compact Magnetic Tunnel Junction Model," *IEEE Transactions on Electron Devices*, Vol. 61, No. 11, pp. 3883–3891, June 2014.

[96] C. Y. You, "Reduced Spin Transfer Torque Switching Current Density with Non-Collinear Polarizer Layer Magnetization in Magnetic Multilayer Systems," *Applied Physics Letters*, Vol. 100, No. 25, pp. 1–5, June 2012.

[97] M. Mansoor, I. Haneef, S. Akhtar, A. De Luca, and F. Udrea, "Silicon Diode Temperature Sensors - A Review of Applications," *Sensors and Actuators, A: Physical*, Vol. 232, pp. 63–74, August 2015.

[98] M. S. Floyd, S. Ghiasi, T. W. Keller, K. Rajamani, F. L. Rawson, J. C. Rubio, and M. S. Ware, "System Power Management Support in the IBM POWER6 Microprocessor," *IBM Journal of Research and Development*, Vol. 51, No. 6, pp. 733–746, November 2007.

[99] J. Nilsson, J. Borg, and J. Johansson, "High-Temperature Characterization and Analysis of Leakage-Current-Compensated, Low-Power Bandgap Temperature Sensors," *Analog Integrated Circuits and Signal Processing*, Vol. 93, No. 1, pp. 137–147, October 2017.

[100] M. Malits, I. Brouk, and Y. Nemirovsky, "Study of CMOS-SOI Integrated Temperature Sensing Circuits for On-Chip Temperature Monitoring," *Sensors*, Vol. 18, No. 5, pp. 1629.1–1629.14, May 2018.

[101] A. Bakker, J. H. Huijsing, and J. Huijsing, *High-Accuracy CMOS Smart Temperature Sensors*, Vol. 595, Springer Science & Business Media, 2000.

[102] A. P. Brokaw, "A Temperature Sensor with Single Resistor Set-point Programming," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 334–335, February 1996.

[103] A. A. Khan, J. Schmalhorst, G. Reiss, G. Eilers, M. Münzenberg, H. Schuhmann, and M. Seibt, "Elastic and Inelastic Conductance in Co-Fe-B/MgO/Co-Fe-B Magnetic Tunnel Junctions," *Physical Review B*, Vol. 82, No. 6, pp. 064416.1–064416.8, August 2010.

[104] J. Teixeira, J. Ventura, J. Araujo, J. Sousa, P. Wisniowski, and P. Freitas, "Tunneling Processes in Thin MgO Magnetic Junctions," *Applied Physics Letters*, Vol. 96, No. 26, pp. 262506.1–262506.3, June 2010.

[105] C. Bellouard, Y. Lu, A. Duluard, B. Negulescu, C. Senet, N. Maloufi, M. Hehn, and C. Tiusan, "Symmetry-State Features in a Global Analysis of the Temperature-Dependent Spin Transport in Fe/MgO/Fe Junctions," *Physical Review B*, Vol. 98, No. 14, pp. 144437.1–144437.8, October 2018.

[106] J. G. Simmons, "Generalized Thermal J-V Characteristic for the Electric Tunnel Effect," *Journal of Applied Physics*, Vol. 35, No. 9, pp. 2655–2658, February 1964.

[107] L. Yuan, S.-H. Liou, and D. Wang, "Temperature Dependence of Magnetoresistance in Magnetic Tunnel Junctions with Different Free Layer Structures," *Physical Review B*, Vol. 73, No. 13, pp. 134403.1–134403.8, April 2006.

[108] B. Oliver and J. Nowak, "Temperature and Bias Dependence of Dynamic Conductance—Low Resistive Magnetic Tunnel Junctions," *Journal of Applied Physics*, Vol. 95, No. 2, pp. 546–550, January 2004.

[109] A. G. Qoutb and E. G. Friedman, "PMTJ Temperature Sensor Utilizing VCMA," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 1–5, May 2019.

[110] Q. L. Ma, S. G. Wang, J. Zhang, Y. Wang, R. C. Ward, C. Wang, A. Kohn, X. G. Zhang, and X. F. Han, "Temperature Dependence of Resistance in Epitaxial Fe/MgO/Fe Magnetic Tunnel Junctions," *Applied Physics Letters*, Vol. 95, No. 5, pp. 4–6, July 2009.

[111] Z. Zeng, Y. Wang, X. Han, W. Zhan, and Z. Zhang, "Bias Voltage and Temperature Dependence of Magneto-Electric Properties in Double-Barrier Magnetic Tunnel Junction with Amorphous Co-Fe-B Electrodes," *The European Physical Journal B-Condensed Matter and Complex Systems*, Vol. 52, No. 2, pp. 205–208, July 2006.

[112] "Predictive Technology Model (PTM)," 2008. [Online]. Available: http://ptm.asu.edu/.

[113] M. Malits and Y. Nemirovsky, "Nanometric Integrated Temperature and Thermal Sensors in CMOS-SOI Technology," *Sensors*, Vol. 17, No. 8, pp. 1739.1–1739.12, July 2017.

[114] H. F. Sheikh, I. Ahmad, Z. Wang, and S. Ranka, "An Overview and Classification of Thermal-Aware Scheduling Techniques for Multi-Core Processing Systems," *Sustainable Computing: Informatics and Systems*, Vol. 2, No. 3, pp. 151–169, September 2012.

[115] A. Das, G. V. Merrett, M. Tribastone, and B. M. Al-Hashimi, "Workload Change Point Detection for Runtime Thermal Management of Embedded Systems," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 35, No. 8, pp. 1358–1371, August 2016.

[116] K.-C. Chen, Y.-H. Chen, and Y.-P. Lin, "Thermal Sensor Allocation and Full-System Temperature Characterization for Thermal-Aware Mesh-based NoC System by using Compressive Sensing Technique," *Proceedings of the IEEE International Symposium on VLSI Design, Automation and Test*, pp. 1–4, April 2017.

[117] C. Yao, K. K. Saluja, and P. Ramanathan, "Calibrating On-Chip Thermal Sensors in Integrated Circuits: A Design-for-Calibration Approach," *Journal of Electronic Testing: Theory and Applications*, Vol. 27, No. 6, pp. 711–721, June 2011.

[118] A. Sengupta, C. M. Liyanagedera, B. Jung, and K. Roy, "Magnetic Tunnel Junction as an On-Chip Temperature Sensor," *Scientific Reports*, Vol. 7, No. 1, pp. 11764.1–11764.8, December 2017.

[119] A. G. Qoutb and E. G. Friedman, "Spintronic/CMOS-Based Thermal Sensors," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 1–5, May 2020.

[120] Y. K. Lee, Y. Song, J. Kim, S. Oh, B.-J. Bae, S. Lee, J. Lee, U. Pi, B. Seo, H. Jung, K. Lee, H. Shin, H. Jung, M. Pyo, A. Antonyan, D. Lee, S. Hwang, D. Jang, Y. Ji, S. Lee, J. Lim, K.-H. Koh, K. Hwang, H. Hong, K. Park, G. Jeong, J. S. Yoon, and E. Jung, "Embedded STT-MRAM in 28-nm FDSOI Logic Process for Industrial MCU/IoT Application," *Proceedings of the IEEE Symposium on VLSI Technology*, pp. 181–182, June 2018.

[121] H.-J. Lee, S. Rami, S. Ravikumar, V. Neeli, K. Phoa, B. Sell, and Y. Zhang, "Intel 22nm FinFET (22FFL) Process Technology for RF and mm Wave Applications and Circuit Design Optimization for FinFET Technology," *Proceedings of the IEEE International Electron Devices Meeting*, pp. 14.1.1–14.1.4, December 2018.

[122] J. Lim, N. Raghavan, A. Padovani, J. Kwon, K. Yamane, H. Yang, V. Naik, L. Larcher, K. Lee, and K. Pey, "Investigating the Statistical-Physical Nature of MgO Dielectric Breakdown in STT-MRAM at Different Operating Conditions," *Proceedings of the IEEE International Electron Devices Meeting*, pp. 25.3.1–25.3.4, December 2018.

[123] Y. J. Song, J. H. Lee, S. H. Han, H. C. Shin, K. H. Lee, K. Suh, D. E. Jeong, G. H. Koh, S. C. Oh, J. H. Park, S. O. Park, B. J. Bae, O. I. Kwon, K. H. Hwang, B. Seo, Y. Lee, S. H. Hwang, D. S. Lee, Y. Ji, K. Park, G. T. Jeong, H. S. Hong, K. P. Lee, H. K. Kang, and E. S. Jung, "Demonstration of Highly Manufacturable STT-MRAM Embedded in 28nm Logic," *Proceedings of the IEEE International Electron Devices Meeting*, pp. 18.2.1–18.2.4, December 2018.

[124] A. G. Qoutb and E. G. Friedman, "Distributed Spintronic/CMOS Sensor Network for Thermal-Aware Systems," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 28, No. 6, pp. 1505–1512, June 2020.

[125] K. M. Kim, J. J. Yang, J. P. Strachan, E. M. Grafals, N. Ge, N. D. Melendez, Z. Li, and R. S. Williams, "Voltage Divider Effect for the Improvement of Variability and Endurance of TaO_x Memristor," *Scientific Reports*, Vol. 6, No. 1, pp. 1–6, February 2016.

[126] J. J. Kan, C. Park, C. Ching, J. Ahn, Y. Xie, M. Pakala, and S. H. Kang, "A Study on Practically Unlimited Endurance of STT-MRAM," *IEEE Transactions on Electron Devices*, Vol. 64, No. 9, pp. 3639–3646, August 2017.

[127] X. Fong, Y. Kim, R. Venkatesan, S. H. Choday, A. Raghunathan, and K. Roy, "Spin-Transfer Torque Memories: Devices, Circuits, and Systems," *Proceedings of the IEEE*, Vol. 104, No. 7, pp. 1449–1488, April 2016.

[128] K. Watanabe, B. Jinnai, S. Fukami, H. Sato, and H. Ohno, "Shape Anisotropy Revisited in Single-Digit Nanometer Magnetic Tunnel Junctions," *Nature Communications*, Vol. 9, No. 1, pp. 1–6, February 2018.

[129] H. Zhang, W. Kang, B. Wu, P. Ouyang, E. Deng, Y. Zhang, and W. Zhao, "Spintronic Processing Unit Within Voltage-Gated Spin Hall Effect MRAMs," *IEEE Transactions on Nanotechnology*, Vol. 18, pp. 473–483, May 2019.

[130] X. Bi, M. Mao, D. Wang, and H. H. Li, "Cross-Layer Optimization for Multilevel Cell STT-RAM Caches," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 25, No. 6, pp. 1807–1820, February 2017.

[131] Y. Pan, P. Ouyang, Y. Zhao, W. Kang, S. Yin, Y. Zhang, W. Zhao, and S. Wei, "A Multilevel Cell STT-MRAM-based Computing In-Memory Accelerator for Binary Convolutional Neural Network," *IEEE Transactions on Magnetics*, Vol. 54, No. 11, pp. 1–5, July 2018.

[132] K. Lee, J. H. Bak, Y. J. Kim, C. K. Kim, A. Antonyan, D. H. Chang, S. H. Hwang, G. W. Lee, N. Y. Ji, W. J. Kim, J. H. Lee, B. J. Bae, J. H. Park, I. H. Kim, B. Y. Seo, S. H. Han, Y. Ji, H. T. Jung, S. O. Park, O. I. Kwon, J. W.

Kye, Y. D. Kim, S. W. Pae, Y. J. Song, G. T. Jeong, K. H. Hwang, G. H. Koh, H. K. Kang, and E. S. Jung, "1Gbit High Density Embedded STT-MRAM in 28nm FDSOI Technology," *Proceedings of the IEEE International Electron Devices Meeting*, pp. 2.2.1–2.2.4, December 2019.

[133] Y. Chih, Y. Shih, C. Lee, Y. Chang, P. Lee, H. Lin, Y. Chen, C. Lo, M. Shih, K. Shen, H. Chuang, and T. J. Chang, "13.3 A 22nm 32Mb Embedded STT-MRAM with 10ns Read Speed, 1M Cycle Write Endurance, 10 Years Retention at 150°C and High Immunity to Magnetic Field Interference," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 222–224, February 2020.

[134] H. Wang, W. Kang, Y. Zhang, and W. Zhao, "Modeling and Evaluation of Sub-10-nm Shape Perpendicular Magnetic Anisotropy Magnetic Tunnel Junctions," *IEEE Transactions on Electron Devices*, Vol. 65, No. 12, pp. 5537–5544, November 2018.

[135] S. Prajapati, S. Verma, A. A. Kulkarni, and B. K. Kaushik, "Modeling of a Magnetic Tunnel Junction for a Multilevel STT-MRAM Cell," *IEEE Transactions on Nanotechnology*, Vol. 18, pp. 1005–1014, October 2018.

[136] M. Aoki, H. Noshiro, K. Tsunoda, Y. Iba, A. Hatada, M. Nakabayashi, A. Takahashi, C. Yoshida, Y. Yamazaki, T. Takenaga, and T. Sugii, "Novel Highly Scalable Multi-Level Cell for STT-MRAM with Stacked Perpendicular MTJs," *Proceedings of the IEEE Symposium on VLSI Technology*, pp. T134–T135, June 2013.

[137] K. Tsunoda, M. Aoki, H. Noshiro, T. Takenaga, C. Yoshida, Y. Yamazaki, A. Takahashi, Y. Iba, A. Hatada, M. Nakabayashi, and T. Sugii, "Highly Manufacturable Multi-Level Perpendicular MTJ with a Single Top-Pinned Layer and Multiple Barrier/Free Layers," *Proceedings of the IEEE International Electron Devices Meeting*, pp. 3.3.1–3.3.4, December 2013.

[138] Y. Zhang, L. Zhang, W. Wen, G. Sun, and Y. Chen, "Multi-Level Cell STT-RAM: Is It Realistic or Just a Dream?," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 526–532, November 2012.

[139] A. Skirda, E. Tokhtuev, C. Owen, and V. Slobodyan, *Driving Circuit for Powering a Bi-Directional Load*, U.S. Patent No. 8,730,701, August 2010.

[140] N. Perrissin, S. Lequeux, N. Strelkov, A. Chavent, L. Vila, L. D. Buda-Prejbeanu, S. Auffret, R. C. Sousa, I. L. Prejbeanu, and B. Dieny, "A Highly Thermally Stable Sub-20 nm Magnetic Random-Access Memory Based on Perpendicular Shape Anisotropy," *Nanoscale*, Vol. 10, No. 25, pp. 12187–12195, May 2018.

[141] M. G. Moinuddin, A. H. Lone, S. Shringi, S. Srinivasan, and S. K. Sharma, "Low-Current-Density Magnetic Tunnel Junctions for STT-RAM Application Using MgO_x N_{1-x} ($x = 0.57$) Tunnel Barrier," *IEEE Transactions on Electron Devices*, Vol. 67, No. 1, pp. 125–132, January 2020.

[142] S. Jain, A. Ranjan, K. Roy, and A. Raghunathan, "Computing in Memory With Spin-Transfer Torque Magnetic RAM," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 26, No. 3, pp. 470–483, March 2017.

[143] O. A. Mukhanov, "Energy-Efficient Single Flux Quantum Technology," *IEEE Transactions on Applied Superconductivity*, Vol. 21, No. 3, pp. 760–769, June 2011.

[144] W. Chen, A. Rylyakov, V. Patel, J. Lukens, and K. Likharev, "Rapid Single Flux Quantum T-Flip Flop Operating Up to 770 GHz," *IEEE Transactions on Appiled Superconductivity*, Vol. 9, No. 2, pp. 3212–3215, June 1999.

[145] R. McDermott, M. G. Vavilov, B. L. T. Plourde, F. K. Wilhelm, P. J. Liebermann, O. A. Mukhanov, and T. A. Ohki, "Quantum–Classical Interface Based on Single Flux Quantum Digital Logic," *Quantum Science and Technology*, Vol. 3, No. 2, pp. 024004.1–024004.18, April 2018.

[146] G. Krylov and E. G. Friedman, *Single Flux Quantum Integrated Circuit Design*, Springer, 2022.

[147] V. K. Semenov, Y. A. Polyakov, and S. K. Tolpygo, "AC-Biased Shift Registers as Fabrication Process Benchmark Circuits and Flux Trapping Diagnostic Tool," *IEEE Transactions on Applied Superconductivity*, Vol. 27, No. 4, pp. 1–9, June 2017.

[148] M. Bushnell and V. Agrawal, *Essentials of Electronic Testing for Digital, Memory and Mixed-Signal VLSI Circuits*, Vol. 17, Springer Science & Business Media, December 2004.

[149] G. Krylov and E. G. Friedman, "Globally Asynchronous, Locally Synchronous Clocking and Shared Interconnect for Large-Scale SFQ Systems," *IEEE Transactions on Applied Superconductivity*, Vol. 29, No. 5, pp. 1–5, August 2019.

[150] T. Jabbari, G. Krylov, J. Kawa, and E. G. Friedman, "Splitter Trees in Single Flux Quantum Circuits," *IEEE Transactions on Applied Superconductivity*, Vol. 31, No. 5, pp. 1–6, August 2021.

[151] G. Krylov and E. G. Friedman, "Design for Testability of SFQ Circuits," *IEEE Transactions on Applied Superconductivity*, Vol. 27, No. 8, pp. 1–7, December 2017.

[152] G. Krylov and E. G. Friedman, "Test Point Insertion for RSFQ Circuits," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 1–4, May 2017.

[153] L. H. Goldstein and E. L. Thigpen, "SCOAP: Sandia Controllability/Observability Analysis Program," *Proceedings of the ACM/IEEE Design Automation Conference*, pp. 190–196, June 1980.

[154] H.-C. Tsai, K. Cheng, C.-J. Lin, and S. Bhawmik, "A Hybrid Algorithm for Test Point Selection for Scan-Based BIST," *Proceedings of the ACM/IEEE Design Automation Conference*, pp. 478–483, June 1997.

[155] Y. Ma, H. Ren, B. Khailany, H. Sikka, L. Luo, K. Natarajan, and B. Yu, "High Performance Graph ConvolutionaI Networks with Applications in Testability Analysis," *Proceedings of the ACM/IEEE Design Automation Conference*, pp. 1–6, June 2019.

[156] K. Gaj, E. G. Friedman, and M. J. Feldman, "Timing of Multi-Gigahertz Rapid Single Flux Quantum Digital Circuits," *Journal of VLSI Signal Processing Systems*, Vol. 16, No. 2, pp. 247–276, July 1997.

[157] E. G. Friedman, "Clock Distribution Networks in Synchronous Digital Integrated Circuits," *Proceedings of the IEEE*, Vol. 89, No. 5, pp. 665–692, May 2001.

[158] J. L. Neves and E. G. Friedman, "Design Methodology for Synthesizing Clock Distribution Networks Exploiting Nonzero Localized Clock Skew," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 4, No. 2, pp. 286–291, June 1996.

[159] S. K. Tolpygo, V. Bolkhovsky, T. J. Weir, A. Wynn, D. E. Oates, L. M. Johnson, and M. A. Gouker, "Advanced Fabrication Processes for Superconducting Very Large-Scale Integrated Circuits," *IEEE Transactions on Applied Superconductivity*, Vol. 26, pp. 1–10, September 2016.

[160] S. K. Tolpygo, V. Bolkhovsky, T. J. Weir, C. J. Galbraith, L. M. Johnson, M. A. Gouker, and V. K. Semenov, "Inductance of Circuit Structures for MIT LL Superconductor Electronics Fabrication Process With 8 Niobium Layers," *IEEE Transactions on Applied Superconductivity*, Vol. 25, pp. 1–5, August 2015.

[161] N. Touba and E. McCluskey, "Test Point Insertion Based on Path Tracing," *Proceedings of IEEE VLSI Test Symposium*, pp. 2–8, May 1996.

[162] H.-D. Hahlbohm and H. Lübbig, Eds., *SQUID '85 Superconducting Quantum Interference Devices and their Applications*, De Gruyter, 1986.

[163] O. A. Mukhanov, D. Kirichenko, I. V. Vernik, T. V. Filippov, A. Kirichenko, R. Webber, V. Dotsenko, A. Talalaevskii, J. C. Tang, A. Sahu, P. Shevchenko, R. Miller, S. B. Kaplan, S. Sarwana, and D. Gupta, "Superconductor Digital-RF Receiver Systems," *IEICE Transactions on Electronics*, Vol. E91-C, No. 3, pp. 306–317, March 2008.

[164] J. Galiay, Y. Crouzet, and M. Vergniault, "Physical Versus Logical Fault Models MOS LSI Circuits: Impact on their Testability," *IEEE Transactions on Computers*, Vol. 29, No. 06, pp. 527–531, June 1980.

[165] A. F. Kirichenko, I. V. Vernik, M. Y. Kamkar, J. Walter, M. Miller, L. R. Albu, and O. A. Mukhanov, "ERSFQ 8-Bit Parallel Arithmetic Logic Unit," *IEEE Transactions on Applied Superconductivity*, Vol. 29, No. 5, pp. 1–7, August 2019.

[166] G. Krylov and E. G. Friedman, "Design Methodology for Distributed Large-Scale ERSFQ Bias Networks," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 28, No. 11, pp. 2438–2447, November 2020.

[167] S. D. Millman and E. J. McCluskey, "Detecting Stuck-Open Faults with Stuck-At Test Sets," *Proceedings of the IEEE Custom Integrated Circuits Conference*, pp. 22.3/1–22.3/4, May 1989.

[168] P. Liden and P. Dahlgren, "Coverage of Transistor-Level and Gate-level Stuck-at-Faults in CMOS Checkers," *Proceedings of the IEEE International Symposium on Circuits and Systems*, Vol. 3, pp. 2124–2127, April 1995.

[169] D. Apalkov, B. Dieny, and J. M. Slaughter, "Magnetoresistive Random Access Memory," *Proceedings of the IEEE*, Vol. 104, No. 10, pp. 1796–1830, October 2016.

[170] M. Wang, Y. Zhang, X. Zhao, and W. Zhao, "Tunnel Junction with Perpendicular Magnetic Anisotropy: Status and Challenges," *Micromachines*, Vol. 6, No. 8, pp. 1023–1045, August 2015.

[171] S. Anders, M. Blamire, F.-I. Buchholz, D.-G. Crété, R. Cristiano, P. Febvre, L. Fritzsch, A. Herr, E. Il'ichev, J. Kohlmann, J. Kunert, H.-G. Meyer, J. Niemeyer, T. Ortlepp, H. Rogalla, T. Schurig, M. Siegel, R. Stolz, E. Tarte, H. ter Brake, H. Toepfer, J.-C. Villegier, A. Zagoskin, and A. Zorin, "European Roadmap on Superconductive Electronics – Status and Perspectives," *Physica C: Superconductivity*, Vol. 470, No. 23-24, pp. 2079–2126, December 2010.

[172] S. Lee and K. Lee, "Emerging Three-Terminal Magnetic Memory Devices," *Proceedings of the IEEE*, Vol. 104, No. 10, pp. 1831–1843, October 2016.

[173] T. Hanyu, T. Endoh, D. Suzuki, H. Koike, Y. Ma, N. Onizawa, M. Natsui, S. Ikeda, and H. Ohno, "Standby-Power-Free Integrated Circuits Using MTJ-Based VLSI Computing," *Proceedings of the IEEE*, Vol. 104, No. 10, pp. 1844–1863, September 2016.

[174] M. Bersweiler, E. Enobio, S. Fukami, H. Sato, and H. Ohno, "An Effect of Capping-Layer Material on Interfacial Anisotropy and Thermal Stability Factor

of MgO/CoFeB/Ta/CoFeB/MgO/Capping-Layer structure," *Applied Physics Letters*, Vol. 113, No. 17, pp. 172401, October 2018.

[175] B. Dieny, R. B. Goldfarb, and K.-J. Lee, *Introduction to Magnetic Random-Access Memory*, John Wiley & Sons, 2016.

[176] K. Watanabe, S. Fukami, H. Sato, S. Ikeda, F. Matsukura, and H. Ohno, "Annealing Temperature Dependence of Magnetic Properties of CoFeB/MgO Stacks on Different Buffer Layers," *Japanese Journal of Applied Physics*, Vol. 56, No. 8, pp. 0802B2, June 2017.

[177] U. S. Goteti, M. Denton, K. Krause, A. Stephen, J. A. Sellers, S. Sullivan, M. C. Hamilton, A. Wynn, and S. K. Tolpygo, "Reliability Studies of Nb/AlOx/Al/Nb Josephson Junctions Through Accelerated-Life Electrical Stress Testing," *IEEE Transactions on Applied Superconductivity*, Vol. 29, No. 5, pp. 1–7, March 2019.

[178] A. Kleinsasser, W. Mallison, R. Miller, and G. Arnold, "Electrical Characterization of Nb/Al-oxide/Nb Josephson Junctions with High Critical Current Densities," *IEEE Transactions on Applied Superconductivity*, Vol. 5, No. 2, pp. 2735–2738, June 1995.

[179] T. C. DuBois, M. C. Per, S. P. Russo, and J. H. Cole, "Delocalized Oxygen As the Origin of Two-Level Defects in Josephson Junctions," *Physical Review Letters*, Vol. 110, No. 7, pp. 077002, February 2013.

[180] M. A. Sulangi, T. Weingartner, N. Pokhrel, E. Patrick, M. Law, and P. Hirschfeld, "Disorder and Critical Current Variability in Josephson Junctions," *Journal of Applied Physics*, Vol. 127, No. 3, pp. 033901, January 2020.

[181] K. Jackman and C. J. Fourie, "Flux Trapping Experiments to Verify Simulation Models," *Superconductor Science and Technology*, Vol. 33, No. 10, pp. 105001, August 2020.

[182] M. Khapaev Jr, "Extraction of Inductances of Plane Thin Film Superconducting Circuits," *Superconductor Science and Technology*, Vol. 10, No. 6, pp. 389, June 1997.

[183] C. J. Fourie and K. Jackman, "Experimental Verification of Moat Design and Flux Trapping Analysis," *IEEE Transactions on Applied Superconductivity*, Vol. 31, No. 5, pp. 1–7, January 2021.

Appendix A

MTJ macrospin model

A macrospin compact model which characterizes a voltage controlled magnetic anisotropy (VCMA) MgO—CoFeB perpendicular MTJ is described here [109]. The model considers the dynamic response of the device magnetic and electrical performance. The magnetization dynamics of the free ferromagnetic (FM) layer are described by the modified Landau-Lifshitz-Gilbert equation. The expression describes the dynamic magnetic behavior of the FM layer as

$$\frac{\partial \vec{M}}{\partial t} = -\frac{\gamma \mu_o}{1+\alpha^2}[\vec{M} \times \vec{H}_{eff} + \alpha\ \vec{M} \times \frac{\partial \vec{M}}{\partial t}] + \gamma\ \sum \vec{\tau}_i, \tag{A.1}$$

where $\vec{M}$ is the normalized free layer magnetization, t is the time variable, $\vec{H_{eff}}$ is the effective magnetic field expressed in A/m, γ is the electron gyromagnetic ratio, $\gamma \approx -2\pi \times 27.99$ GHz/T, μ_o is the permeability of free space, α is the Gilbert damping factor, and $\vec{\tau}_i$ is the applied torque due to other perturbations such as current which exerts a spin transfer torque [55].

The macrospin model is developed in association with the static and dynamic micromagnetic analysis of the system energy. The applied effective magnetic field to the free layer $\vec{H_{eff}}$ is

$$\vec{H}_{eff} = \vec{H}_{UA} - \vec{H}_{dem} + \vec{H}_c + \vec{H}_{ext} - \vec{H}_{VCMA} + \vec{H}_{th}, \tag{A.2}$$

where $\vec{H_{UA}}$ is the uniaxial anisotropy field sometimes defined as $\vec{H_K}$, $\vec{H_{dem}}$ is the demagnetization field, $\vec{H_c}$ is the coupling field due to the other FM layer, $\vec{H_{ext}}$ is the applied external magnetic field, $\vec{H_{VCMA}}$ is due to VCMA, and $\vec{H_{th}}$ is the stochastic magnetic field due to thermal variations.

The MTJ antiparallel conductance is modeled as [90]

$$G_{AP}(T) = G_T \left[1 - P_1(T)P_2(T)\right] + G_{SI}, \tag{A.3}$$

where $G_T = G_0 \left(\sin(CT)/CT\right)$ is the thermal smearing factor, G_0 is the parallel state conductance $G_0 = \left(3.16 \times 10^{10}\sqrt{\phi_B}/t_{ox}\right) exp(-1.025 \times \sqrt{\phi_B} \times t_{ox})$ at zero voltage and zero temperature, T is the ambient temperature, ϕ_B is the average tunneling barrier height (in eV), t_{ox} is the thickness of the insulator barrier layer, and $C = 1.387 \times 10^{-4} t_{ox}/\sqrt{\phi_B}$ is a material dependent parameter [90]. $G_{SI} = ST^{4/3}$ is the inelastic spin independent conductance, and S is a fitting parameter. P_1 and P_2 are the spin polarization percentage of the two FM layers. The dependence of the

spin polarization on temperature can be fitted as [91, 92]

$$P(T) = P(0)\left[1 - \beta_P T^{\alpha_P}\right], \tag{A.4}$$

where β_P and α_P are fitting parameters related to the device dimensions and material properties.

The physical parameters are based on perpendicular magnetic anisotropy and VCMA MgO—CoFeB [93, 94, 109]. The experimentally extracted model parameters are listed in Table A.1.

Table A.1: MTJ physical parameters

Parameters	**Description**	**Value**
w_{FL}	FM width = radius	20 nm
t_{FL}	FM thickness	1.5 nm
t_{ox}	Barrier thickness	1.1 nm
Φ_{BL}	Barrier height	0.39 eV
V_h	Voltage @ half TMR	0.5 V
S	Spin independent conductance factor	1.1×10^{-12}
β_P	Fitting parameter for P	2.07×10^{-5}
α_P	Fitting parameter for P	2.3
β_M	Fitting parameter for M_S	1.5
T^*	Fitting parameter	1120 K
β_{Ki}	Fitting parameter	2.3
$\beta_{\zeta VCMA}$	Fitting parameter	2.83
N_z	Demagnetization tensor factor in Z	0.9343
N_{xy}	Demagnetization tensor factor in XY	0.015
K_{i0}	Interfacial MA at 0 K	$2.02 \times 10^{-3} J/m^2$
M_{S0}	Saturation magnetization at 0 K	$1457 \times 10^3 A/m$
TMR_0	TMR at 0 K	3
ξ_{VCMA0}	VCMA factor at 0 K	48.9×10^{-15} J/(V.m)