May 2018

Emerging Non-Volatile Memory Technologies for Computing and Security

Rekha Govindaraj
University of South Florida, rekhag@mail.usf.edu

Emerging Non-Volatile Memory Technologies for Computing and Security

by

Rekha Govindaraj

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Computer Science and Engineering
College of Engineering
University of South Florida

Co-Major Professor: Swaroop Ghosh, Ph.D.
Co-Major Professor: Srinivas Katkoori, Ph.D.
Charles Augustine, Ph.D.
Xinming (Simon) Ou, Ph.D.
Srikant Srinivasan, Ph.D.

Date of Approval:
May 31, 2018

Keywords: Content Addressable Memory, Hardware Security, Spintronic Memory, Magnetic Tunnel Junction, Resistive Random Access Memory, Random Telegraph Noise

DEDICATION

Dedicated to my parents. “Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less” - Marie Curie (1867-1934)

ACKNOWLEDGMENTS

I would like to express my deep gratitude to Dr. Swaroop Ghosh for his indispensable support and patience in training me during my PhD journey. Advices from Dr. Ghosh have helped me tackle complex problems effortlessly by always keeping the broader picture of my research. Constant support from Dr. Ghosh through out has helped me not only be successful in my current academic research but in preparing me for future endeavors.

My special thanks to Dr. Srinivas Katkoori for accepting me as his student and supporting my research interests. I thank my co-advisors Dr. Swaroop Ghosh and Dr. Srinivas Katkoori for spending countless hours in research problem discussions, meticulously reviewing the paper drafts and this dissertation. Constructive and critical comments from Dr. Ghosh and Dr. Katkoori were crucial in having successful publications.

I acknowledge and deeply appreciate the presence and contribution of other students at Dr. Ghosh's and Dr. Katkoori's lab in this success. I like to thank University of South Florida faculty, staff and student members at the Department of Computer Science and Engineering for all the support and assistance in successfully completing my research. I also thank my cohorts at American Association of University Women (AAUW-USF) for their camaraderie at graduate school.

I sincerely thank my loving parents, amazing siblings Nagaraj and Roopa, wonderful husband Shantharaju for all their love and support for my success. I thank all my friends for their help and presence to make this journey the most cherish-able.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

With CMOS technology scaling reaching its limitations rigorous research of alternate and competent technologies is paramount to push the boundaries of computing. Spintronic and resistive memories have proven to be effective alternatives in terms of area, power and performance to CMOS because of their non-volatility, ability for logic computing and easy integration with CMOS. However, deeper investigations to understand their physical phenomenon and improve their properties such as writability, stability, reliability, endurance, uniformity with minimal device-device variations is necessary for deployment as memories in commercial applications. Application of these technologies beyond memory and logic are investigated in this thesis i.e. for security of integrated circuits and systems and special purpose memories. We proposed a spintonic based special purpose memory for search applications, present design analysis and techniques to improve the performance for larger word lengths upto 256 bits. Salient characteristics of RRAM is studied and exploited in the design of widely accepted hardware security primitives such as Physically Unclonable Function (PUF) and True Random Number Generators (TRNG). Vulnerability of these circuits to adversary attacks and countermeasures are proposed. Proposed PUF can be implemented within 1T-1R conventional memory architecture which offers area advantages compared to RRAM memory and cross bar array PUFs with huge number of challenge response pairs. Potential application of proposed strong arbiter PUF in the Internet of things is proposed and performance is evaluated theoretically with valid assumptions on the maturity of RRAM technology. Proposed TRNG effectively utilizes the random telegraph noise in RRAM current to generate random bit stream. TRNG is evaluated for sufficient randomness in the random bit stream generated. Vulnerability and countermeasures to adversary attacks are also studied. Finally, in thesis we investigated and extended the application of emerging non-volatile memory technologies for search and security in integrated circuits and systems.

CHAPTER 1 : INTRODUCTION

CMOS technology has been workhorse of design in electronics industry because of electric field controlled operation and lower power consumption of MOSFETS compared to other semiconductor technologies such as Bipolar Junction Transistors. Constant scaling of CMOS devices has been enabling the semiconductor industry to meet the consumer electronics product market requirements such as smaller size (miniaturization), light weight battery powered gadgets with enormous computing capability comparable to a desktop [4]. Scaling of CMOS itself is driven by Moore's law. Moore's law has been guiding semiconductor industry to push the boundaries of technology and process since past several decades. However, the scaling dictated by Moore's law is exponential and doesn't continue indefinitely due to several technology challenges and practical limitations [5] such as precision in photo lithographic process, electrical limitations due to Short Channel Effects (SCA), Narrow Channel Effects (NCA) and the ever changing needs by applications (memory intensive and longer standby times). This chapter presents a review of technology scaling challenges and the technologies beyond CMOS in exploration to address these challenges.

1.1 Conventional CMOS and Design Challenges

CMOS technology is benefited by the advantages of scaling in terms of performance, dynamic power and area. However, scaling below 65nm leakage power increases exponentially offering new

[1] Portions of this chapter were reprinted from Ghosh, Swaroop. "Spintronics and Security: Prospects, Vulnerabilities, Attack Models, and Preventions." Proceedings of the IEEE 104, no. 10 (2016): 1864-1893.
Permission is included in Appendix A.

challenges to continuous scaling [6]. Increase in the leakage power due to the effects of short and narrow channel which were not considered earlier become prominent in the feature size below 65nm. Power consumption of a CMOS chip has two major components namely, dynamic and leakage power. Total power in a CMOS circuit is modeled as $P = fCV^2 + VI_{leakage}$; f is the frequency of operation, C is the total load capacitance, V is operating voltage and $I_{leakage}$ is the leakage current in the active/standby mode.

In addition to Drain Induced Barrier Lowering (DIBL) and increased subthreshold leakage contributing to high leakage power, effects of Hot Carrier Injection (HCI), Time Dependent Dielectric Breakdown (TDDB) and Bias temperature Instability (BTI), high power density, increased soft error probability with scaling of threshold voltage [7],mobility degradation [8] and process variations add the reliability concerns with submicron scaling.

Aside from limitations of CMOS for logics and circuit design CMOS based memories such as Dynamic RAM (DRAM), Static RAM(SRAM) and flash memories have shown several design challenges with technology scaling [9, 10, 11, 12, 13].Also, finding a single memory technology that caters to the requirements at all levels of memory hierarchy with logic is another motivation to research emerging memory technologies. Such memory is termed as 'Universal memory' technology [14, 15, 16]. STT-RAM and RRAM technology are potential candidates as Universal memory achieved by tuning the characteristics of memory device [15, 17].

With the raising concerns for security the metrics of power, performance and area do not suffice in the IC design for modern applications. In the era of Internet besides sophisticated IC manufacturing process security at the root level in the IC manufacturing and supply chain becomes essential [18]. Various CMOS based security primitives and countermeasures have been proposed in the literature [19]. However, CMOS based designs have demonstrated vulnerability to adversary attacks such as side channel attacks[20, 21] and offer limited features for the design of security primitive. This necessitates the search of emerging technologies for hardware security applications [18].

1.2 Emerging Technologies - Beyond CMOS for Memory and Computing

In the wake of ever changing computing needs by modern applications [17] and to overcome the challenges of CMOS technology scaling several non-classical CMOS devices, memory devices and logic devices have been researched [22, 23]. Non-classical CMOS devices include ultra-thin body SOI transistor, band-engineered transistors, FinFET, vertical transistors, high-k/metal gate transistors, ballistic channel transistors [24], Double-gate transistor, and so on.Non Volatile Memory (NVM) technologies explored for both storage and computing within memory include memristor, spintric based memories such as spin valve, Magnetic RAM (MRAM), Spin-Transfer Torque RAM (STT-RAM) using Magnetic Tunnel Junction (MTJ) device and Domain Wall Memory (DWM), Restive Random Access Memory (RRAM) and variants, Phase Charge Memory (PCM), Nano ElectroMechanical (NEM) devices, and so on. MTJ and RRAM based memory cells have proven to be competent and potential replacement candidates to DRAM and SRAM currently used in computing devices [17]. Further, spintronic and RRAM devices can be fabricated in the via space of CMOS layout with fewer additional mask layers in the manufacturing process. Therefore, they are compatible with existing CMOS fabrication process technology.

Advantages of emerging NVM technologies have been exploited for various circuit and architecture design.Computing within memory is an important development with NVM solving the problem of processor-memory communication delay. Adder circuit within memory is proposed and implemented [25, 26]. Neuromorphic, quantum computing and bio-inspired computing have also been realized with emerging NVM technologies. memristor ans spintronic memorie are widely researched for logic and circuit design applications.

Memristor was discovered in 1971 and is perceived as a fourth fundamental circuit element, resistor with memory. Resistance of a memristor depends on the history of the voltage applied across its terminals.Since its discovery application of memristor for stateful logic, Boolean logic [27],[28] and implicative circuits has been extensively researched [29].Memristor based multi-state registers, non-volatile processors, memristor based neuron circuit, cross bar array circuits and neuromorphic computing have been proposed [30, 31]. memristor only logic which enables computing within cross bar memory array has been proposed in [32] [33]. These cross bar array based and memristor only based logics lack the sharing of inputs between multiple gate limiting fanout and do not implement XOR/XNOr gates. MAD memristor logic gates [34] implement XOR logic gates and can share the inputs between multiple gates.

Spintronic is another emerging technology extensively researched for its applications for computing [35]. Property of shift in DWM [36] and spin current based switching of MTJ have been greatly exploited for computing. In-memory computing [37], associative computing [38, 39], Boolean logic computing [40],reconfigurable logic based on shift based look-up table [36], big-data computing [41],spin-neurons for non-boolean computation [42], STT-RAM based processor for general purpose [43], MTJ based non-volatile logic computing [44, 45],realization of non-volatile flip flop like storage elements [46, 47, 48], crossbar computing [49, 42, 28] are worth mentioning in this context. Emerging technologies have also demonstrated enormous potential in various areas of computing.

Vision of this work is to explore and extend the application of emerging Non-Volatile Memory (NVM) technologies. We researched the application of spintronics based memories such as MTJ and DWM, and resistive memory such as hafnium oxide (HfO_x) based RRAM for associative memory and hardware security applications.

1.3 Hardware Security

In today's highly integrated circuits and systems, satisfying the functionality, frequency, and Thermal Design Power (TDP) requirements are not adequate [19]. It is essential to ensure the security and privacy of the overall system. The contemporary business model involves the *untrusted* third party in every step of Integrated Circuit (IC) manufacturing process- from design, synthesis, layout, and all the way to fabrication and packaging. The trend of integrating third-party Intellectual Property (IP) blocks into the design makes the problem more complex. Broadly, the attacks on hardware could fall under:

- *Malicious modifications* : The hardware Trojans can be inserted in the ICs which cause malfunctioning of IC or leak information for instance.
- *Cloning/Fake IC* : The adversary can imitate the design, fabricate and sell at lower price which lower the market of the genuine IC.
- *Hacking/Eavesdropping* : The adversary eavesdrops the communication channel to crack the secret key for malicious intent.
- *Side Channel Attacks* : Side channels, e.g., current, voltage are monitored to extract the secret information from the device.

- *Reverse Engineering* : IC design details are revealed by peeling off the layers of fabrication process using chemicals and mechanical methods, which reveals the secret design information.

- *IC Recycling* : Unused or barely use ICs are recycled from older PCBs and sold for reduced price compared to genuine new ICs from original manufacturer.

Furthermore, it is worth mentioning that security, trust, and authentication of an electronic system are intertwined with each other. Such an untrusted design environment results in infected hardware that in turn necessitates the authentication of the ICs in the end product. Hardware security primitives such as hardware encryption engine, Physically Unclonable Functions (PUFs), True Random Number Generators (TRNGs), recycling sensors, tamper detection sensors are promising to provide security against the threats such as Hardware Trojan insertion, IC recycling, chip cloning, data snooping and side channel attacks. Furthermore, these primitives are energy-efficient and incurs low area/design overhead. Table 1.1 summarizes the key requirements of hardware security primitives along with the respective features of emerging NVM technologies.

Table 1.1: Properties offered by emerging NVM technologies [1, 2].Hardware security primitives, key requirements and properties offered by emerging NVM technologies.

Security Primitive	Key Requirements	Features Offered by NVM technologies
PUF	high process variation nonlinearity	RRAM device to device switching variations Nonlinearity in switching resistance
TRNG	High entropy	Noise sensitivity of magnetization stochastic dynamics Telegraph noise in RRAM current
Encryption	unique and unpredictable identification key	Unique key generation from RRAM device-device variations
Recycling Sensor	Low process variation High sensitivity to usage	DW nucleation Cycles of endurance in RRAM and MTJ

It should be noted that CMOS based circuits in security applications have demonstrated the problems of area and power overhead. Further, they are sensitive to environmental fluctuations and have limited randomness and entropy offered by the silicon substrate. The emerging technologies such as magnetic, spintronic and resistive memories have shown promises in bringing an abundance of entropy and physical randomness. Unique identification keys can be generated by extracting spatial, temporal randomness and inherent entropy in a magnetic system and switching variations using custom-designed harvesting circuits. Furthermore, these technologies have also demonstrated robustness, speed and orders of magnitude energy-efficiency compared to their CMOS counterparts [2, 50, 51].In this work, salient features of HfO_x based RRAM are exploited for hardware security applications.

Contributions of this dissertation to the literature are described below.

1. Design of an associative memory cell termed Ternary Content Addressable Memory (TCAM) cell is proposed. Design selections and analysis including the selection of NMOS devices and MTJ characteristics, effect of PVT variations are presented. Process variation tolerance techniques for reliable search in longer words are also proposed [52, 53].

2. We propose an Arbiter Physically Unclonable Function (APUF) using cycle-to-cycle switching variations in resistance states of RRAM. The proposed APUF is realized within one-Transistor one-Resistor (1T-1R) conventional memory architecture with minimal invasive design changes. Proposed APUF is studied for its vulnerability to adversary attacks and design technique to improve the resiliency to machine learning based model building attacks is proposed.

3. Random Telegraph Noise (RTN) in the RRAM is exploited for designing a True random Number Generator(TRNG) [54]. Proposed TRNG uses 1T-1R RRAM cell in the bias network of Current Starved Ring Oscillator (CSRO) producing temporal variations in the frequency of CSRO. The CSRO oscillations are sampled to generate true random numbers. In addition,

proposed TRNG is configurable which can be utilized to recover from potential adversary attacks.

1.4 Organization of Thesis

Rest of this thesis is organized as follows.

Chapter 2 discusses the basics of spintronic technology and the spintronic memory device MTJ explored in this thesis.

Chapter 3 presents the fundamentals of RRAM memory and its functional operation. Specific features of RRAM exploited in hardware security applications and the details of RRAM device models used for work are also discussed.

Chapter 4 demonstrates an application of MTJ in searchable memory termed as *6T-2MTJ TCAM*. Design analysis and techniques to improve the sense margin of proposed TCAM cell are presented.

Chapter 5 presents the realization hardware security primitive TRNG using RRAM memory in the CSRO.

Chapter 6 illustrates another security primitive APUF exploiting the switching variations of RRAM. The chapter Discusses the feature of configurability and architectural changes to leverage proposed APUF to be machine learning attack resilient.Also presents potential application of APUF architecture for data attestation in the Internet of Things (IoTs).

Chapter 7 presents the conclusions and potential future works to extend and enhance the application of NVMs to in the proposed areas.

CHAPTER 2 : MAGNETIC TUNNEL JUNCTION DEVICE

This chapter discusses the basics of spintronic technology, device structure, physics and operation of spintronic based Magnetic Tunnel Junction (MTJ) memory device.

2.1 Introduction

Spintronic technology is based on magnetic properties of electron where conventional CMOS computing is based on charge of an electron. Basis of spin devices is magnet. Direction of spin associated with electron is used to encode the logic states for storage and computation. Spintronic devices retain the state due to stable magnetisation direction under the influence of no external field. This property makes the memory based on spintronic non-volatile.

Spintronic devices are realised using magnetic properties of ferromagnetic materials. Three extensively researched spintronic memories are spin valve, Spin Transfer Torque Magnetic Random-Access Memory (STTRAM) and Magnetic Random Access Memory (MRAM). Spintronic device consist of two ferromagnetic layers separated by a spacer layer [55]. The devices differentiate by type of the spacer layer in the device structure and their switching mechanism. Non-magnetic spacer layer between ferromagnetic layers create the effect of magnetoresistance termed as Giant Magnetoresistance (GMR). GMR depends on the relative magnetization orientation of the two ferromagnetic layers and the spacer material used between them (minimum of 3%-8% to maximum

[2] Portions of this chapter were reprinted from Ghosh, Swaroop. "Spintronics and Security: Prospects, Vulnerabilities, Attack Models, and Preventions." Proceedings of the IEEE 104.10 (2016): 1864-1893.
Permission is included in Appendix A.

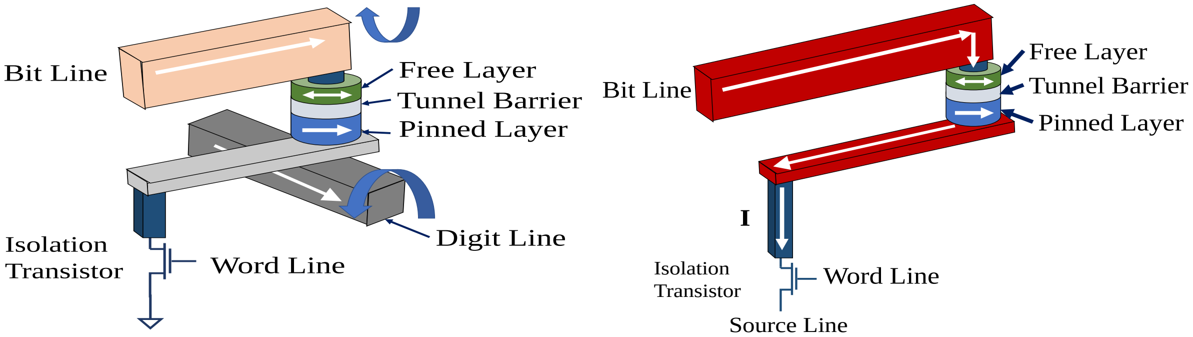
of 50%). GMR determines the ratio of the currents in low resistance state to high resistance state through spin device. Higher GMR is required for better readability of a spin device. A very thin layer of insulating, dielectric material can increase the GMR to more than 100%. The new magnetoresistance is termed as Tunnel Magneto Resistance (TMR). Details of the spintronic devices used in this thesis are explained the coming sections.

This chapter presents the details of MTJ as a spintronic device and operation. Also presents an overview of application of MTJ in memory and computing.

2.2 Magnetic Tunnel Junction

MTJ contains a free and a pinned magnetic layer separated by a thin tunneling oxide layer (a schematic is shown in Fig. 2.1a). The equivalent resistance of the MTJ stack is high (low) if the free layer magnetic orientation is anti-parallel (parallel) to that of the fixed layer. Conventionally, the high equivalent resistance is considered as data '1' and low equivalent resistance is considered as data '0'. The magnetic orientation of the MTJ free layer can be changed from Parallel (P) to Anti-Parallel (AP) (or vice versa) to that of fixed layer by either magnetic field-driven or current-driven techniques. The magnetic field-driven MTJ is the basis for MRAM technology [56] which is promising due to high-density, low standby power, and high-speed operation. On the other hand, STTRAM [57] is an energy-efficient variant of MRAM where the switching of magnetization is based on spin-polarized current.

In MRAM, MTJ lies between a pair of write lines named digit-line and bit-line (Fig. 2.1a). These lines are arranged at right angles to each other, parallel to the cell plane, and one above and one below the cell. An induced magnetic field is created by passing current through the lines.



(a) (b)

Figure 2.1: Schematic of (a) MRAM; (b) STTRAM [1].

The induced magnetic field exerts a torque on the free layer magnetic orientation causing it to flip. Therefore, the direction of write current determines the polarity of the torque and thus determines writing '0' or '1'. The isolation (access) transistor is kept off during write. However, during read, the access transistor is turned on, and a voltage is applied across the cell to sense the equivalent resistance. It should be noted that the read current is unidirectional. Other variant of MRAM is thermally Assisted MRAM (TA-MRAM). In TA-MRAM writing by a magnetic field pulse is assisted by temporary heating of the cell produced by tunneling current through the selected cell. TA-MRAM writing consumes less power compared that in filed-written MRAM. In MRAM, depending on the direction of magnetization of reference w.r.t. the plane in ferromagnetic electrodes MTJ devices are classified as In-plane Magnetic Anisotropy (IMA) and Perpendicular Magnetic Anisotropy (PMA) devices. PMA MTJ devices are preferred in the applications requiring longer retention and stability of bit information. It's noteworthy that it is difficult to grow Perpendicular-to-plane magnetized materials than in-plane magnetized materials [58]. MRAM devices offer the advantages of reliability, robustness, endurance and resistance to external radiation which are attractive in space and automotive applications [58]. First commercial MRAM chips of 1, 4, 8 and

16Mbit were developed by Everspin Technologies in 2006 followed by Thermally Assisted MRAM (TA-MRAM) chips by Crocus Technology in 2011. Latest reported maturity of MTJ fabrication technology has demonstrated TMR of upto 600% at room temperature with MgO tunnel barrier layer. TMR of 190% has been achieved with flexible MgO-barrier MTJs can be bent to various radii upto 5mm. These flexible MTJs find potential application in high performance and flexible electronics [59].

In STTRAM, each cell has one MTJ and one access transistor in series. The write operation is done by turning the access transistor ON and injecting current from the source-line to the bit-line or vice versa (Fig. 2.1b). STT-writing improves the selectivity and better scalability of cell size compared to MRAM devices. Of-plane STTRAMs require less write current than that in in-plane STTRAMs but with a trade-ff of lesser retention than in-plane counterparts. Thermal assistance can also be employed similar to MRAM. In STTRAM Joule heating produced by the write current through tunnel barrier in MTJ assists the switching of magnetization[58] . However, the read operation of STTRAM is similar to that of MRAM. The dynamics of free layer magnetization for both MRAM and STTRAM is governed by LLG equation [60, 61, 62] as follows:

$$\begin{aligned}\frac{\partial\overrightarrow{m}}{\partial t} = -\gamma\overrightarrow{m}\times(H_{eff}+\underbrace{h_{st}}_{stochastic}) - \alpha\gamma\overrightarrow{m}\times[\overrightarrow{m}\times(H_{eff}+\underbrace{h_{st}}_{stochastic})] \\ + \underbrace{\frac{I_s\hbar G(\psi)}{2e}\overrightarrow{m}\times(\overrightarrow{m}\times\overrightarrow{e_p})}_{STT}\end{aligned} \tag{2.1}$$

where, $\overrightarrow{m}$ is unit vector representing local magnetic moment, α represents the Gilbert's damping parameter, γ is gyromagnetic ratio, h_{st} is field due to stochastic noise, I_s is spin current, G(ψ) is the transmission co-efficient, $\hbar$ is reduced Plank's constant, e is charge of electron and $\overrightarrow{e_p}$ is the unit vector along fixed layer magnetization. In the above expression, $\overrightarrow{H_{eff}}$ is effective field given by, $\overrightarrow{H_{eff}} = \overrightarrow{H_a} + \overrightarrow{H_k} + \overrightarrow{H_d} + \overrightarrow{H_{ex}}$,where $\overrightarrow{H_a}$ is applied field, $\overrightarrow{H_k}$ is anisotropy field, $\overrightarrow{H_d}$ is demagnetization field, and $\overrightarrow{H_{ex}}$ is exchange field. The retention time of the MTJ, i.e., the time between which the free layer magnetization tends to flip is given by $T_{ret} = t_0 e^{\Delta}$, where t_0 attempt time (∼1ns), stability factor, $\Delta = \frac{K_u V}{k_B T}$ where K_u is the magneto-crystalline anisotropy, V is the volume of the MTJ free layer, T is the operating temperature and k_B is the Boltzmann constant. By injecting a current (I) through the MTJ having a critical current of I_{co} (where the direction of the current flips the bit), the retention time can be altered as follows [63]:

$$\Delta = \frac{K_u V}{k_B T}(1 - \frac{I}{I_{co}}) \tag{2.2}$$

The equations 2.1 and 2.2 are crucial to understanding the factors that can influence the magnetization dynamics and retention time of MTJ.

Application of STTRAM cell is investigated in this thesis. The resistance of MTJ is high when PL and FL are in antiparallel configuration whereas the resistance is low when they are parallel to each other. The value written to the STTRAM bit depends on the direction and the strength of the charge current. Minimum current required to flip the state of the MTJ in a STTRAM bit is called critical current. Bit '1' is written by passing charge current from pinned layer to fixed layer (Fig. 2.1b) and bit '0' is written by the current opposite direction [64]. Tunnel Magneto Resistance (TMR) is the ratio which determines the ratio of electrical resistances of the MTJ structure in parallel and antiparallel polarization states of FL relative to PL magnetization. If R_H (R_L) is the MTJ resistances in antiparallel (parallel) states, the TMR is defined as $TMR = \frac{(R_H - R_L)}{R_L}$. In this thesis, MTJ and the transistors in series together is referred as STTRAM while MTJ referred to the MTJ device in the memory cell. Table-2.1 summarizes the parameters of the MTJ device used in this work.

Table 2.1: Parameters of MTJ used

Parameter	Value
Saturation Magnetization (M_s)	800 Oe
Critical current density (J_{co})	3.2MA/cm^2
Uniaxial Anisotropy (K_u)	150150 erg/cc
Damping Constant	0.007
TMR	125%
Length and Width	60nmX60nm
Critical Current (I_{co})	115µA

2.3 MTJ for Memory and Computing

Advent of TMR and STT write mechanisms has enabled application of MTJ device based STTRAM to be potential universal memories. Researchers ventured to deploy spintronic devices in logic computing such as in-memory Boolean computing [65], write spin-current based logic computing [66], majority gates [67, 68] and in neural networks [69]. This section presents a review of application of MTJ in computing.

Logic gates realization proposed in [66] is dependent on the switching critical current of MTJ. Operation of NAND gate can be explained as follows. Inputs to the logic gate are stored in to parallel MTJs in terms of MTJ resistance states. MTJ storing evaluated output is connected in series with the input MTJ network. Output MTJ is initialised to a default low resistance state (R_L). When atleast one of the input MTJs are at low resistance state R_L net current flowing through the output MTJ switched it to R_H. This evaluates the output of NAND gate to logic state '1' (R_H). When the input MTJs are in R_H state net current through the output MTJ is less than the critical current and state of output MTJ evaluates to logic '0' (R_L). Similar technique is extended to realise

NOR and inverter universal gates. All Spin Logic (ASL) is another technique of realising logic gates with spintronic based devices.

Proposal of ASL in [65] uses a spin channel between input and output bits unlike in [66] where the charge current was used to evaluate the output. The device used in ASL realization is termed as asl-device. In ASL device input and output magnets are connected and interact via a spin-coherent channel. Spin coherent channel composes of material with high spin flip length, thus conserving the spin of in the channel to manipulate the output bit from the state of input bit. with information storage, communications and evaluation in spin domain ASL is faster, energy and area efficient designs for logic implementation. Also, ASL satisfies the essential characteristics for logic realization and application [65]. Functional extension of ASL gates to realise majority gates, adders, multipliers and complex logic function implementation with majority gates is proposed in [67].

A new type of spin logic device is proposed in [68] called spin torque majority gates (STMG) uses complete spin domain to for logic evaluation. STMG device consists of nanopillars and magnetization of common output free layer is switched by the net current through these nanopillars. Structure is similar to 3 terminal STTRAM. STMG can implement various complex logic functions in terms of majority gates similar to implementation proposed in [67]. However, STMG suffer from low switching speeds and high switching energy compared to ASL majority gates proposed.

Further, application of spintronic devices is explored in neural networks implementation. Deep Neural Networks (DNN) use a large number of hidden layers. DNNs find numerous applications in the fields requiring to learn complex data patterns and data prediction. Spike Neural Networks use bio-inspired design to mimic the functioning of human brain. Spike Timing Dependent synaptic Plasticity (STDP) is one of the desired properties in the synapse connecting the neurons in brain for learning. Programmability and stochastic switching of Spintronic devices by voltage pulse is used to implement synaptic plasticity [69]. MTJ along with heavy metal layer to switch MTJ by spin-Hall effect induced by the current through heavy metal structure provides separate path for programming and spike transmission in neural networks. There are numerous publications exploring the application of spintronic devices such as MTJ and DWM in neuromorphic computing and neural network implementation [70, 71, 72, 73, 74, 42]. Exhaustive discussions on this topic is out of scope of this thesis.

2.4 Summary

Basics of spintronic technology and devices researched in this thesis are discussed in this chapter. Spintronic devices are attractive alternatives to CMOS memory because of their non-volatility, scalability to sub-nm technology nodes, smaller area, low power consumption and compatibility with CMOS process technology. Further, features of spintronic devices can be exploited to realize logic and special purpose memories competent with their CMOS counterparts. Spintronic memories are potential candidates as universal memories because of their versatile and flexible characteristics with selection of materials in the stack and feature size. Extensive research and innovation is geared towards commercializing spintronic technology in computing.

CHAPTER 3 : RESISTIVE RANDOM ACCESS MEMORY[1]

3.1 Introduction

RRAM is a disruptive memory device for future NVM applications and a strong memory candidate to challenge flash memory currently used in commercial applications. RRAM based memory cell size of $6F^2$ is achieved (transistor selector device) and $4F^2$ cell size is possible with diode as memory selector device in cross-bar array structure. This offers high density memory competitive to DRAM and SRAM. It is designed by sandwiching an oxide material between two metal electrodes i.e., Top Electrode (TE) and Bottom Electrode (BE). RRAM resistive switching is primarily due to the mechanism of oxide breakdown and reoxidation which modifies a Conduction Filament (CF) in the oxide. Fig. 3.1 shows the voltage and current transfer characteristics during the SET and RESET process cycles. The minimum resistance of the filament depends on the current compliance used in the process of forming. The two states of the RRAM in low resistance and high resistance are termed as Low Resistance State (LRS) and High Resistance State (HRS). We have used the expressions from [75, 76, 77] as the basis to model the resistance of Hafnium oxide based RRAM at different voltages applied at the top electrode. The resistance switching of RRAM involves three elementary processes such as formation, SET and RESET.

[3] Portions of this chapter was reprinted from 'Govindaraj, R., & Ghosh, S. (2016, October). A strong arbiter PUF using resistive RAM within 1T-1R memory architecture. In Computer Design (ICCD), 2016 IEEE 34th International Conference on (pp. 141-148). IEEE.', 'R. Govindaraj, S. Ghosh and S. Katkoori, "CSRO-Based Reconfigurable True Random Number Generator Using RRAM," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems.doi: 10.1109/TVLSI.2018.2823274.' and 'Puglisi, Francesco Maria, Paolo Pavan, Andrea Padovani, and Luca Larcher. "A compact model of hafnium-oxide-based resistive random access memory." In Proceedings of 2013 International Conference on IC Design & Technology (ICICDT). 2013.', with permission from IEEE
Permission is included in Appendix A.

This chapter presents the overview of RRAM operation, device types and details of salient features of RRAM exploited in hardware security applications.

3.2 Types of RRAM

RRAM devices are characterized by their switching mechanism, polarity of voltage across electrodes for switching and the materials used in oxide sandwitch and electrodes. Based on polarity of SET/RESET switching voltages RRAM is categorized as unipolar and bipolar. In unipolar type RRAMs switching depends only on the magnitude of the applied voltage across electrodes and independent of the polarity. Switching is interpreted to be due to Joule heating effect in the oxide layer. In bipolar RRAM SET/RESET process is dependent on the polarity of the applied switching voltage between top and bottom electrodes. The switching is due to electrochemical migration of ions and redox reactions which depend on the polarity of operating voltage. Switching model of HfO_x based bipolar RRAM is used for our research. Details of switching mechanism is discussed below. The forming voltage is applied across the electrodes to create an electric field in the oxide material.

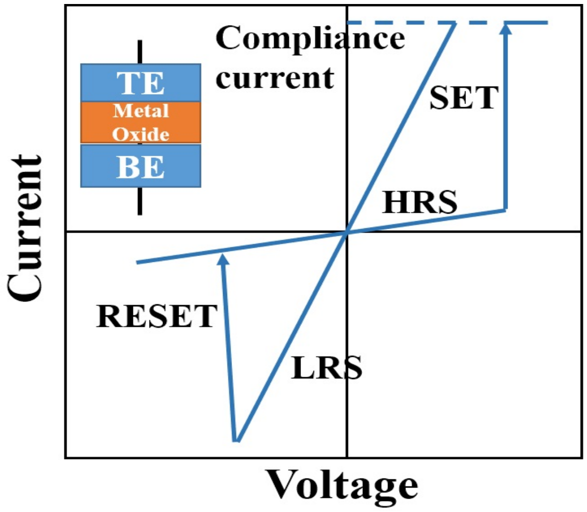


Figure 3.1: RRAM memory device and resistance transfer characteristics.

Oxygen atoms are knocked out of oxide material forming oxygen vacancies under the influence of high electric field, typically as high as 10MV/cm (Fig. 3.2). The conduction through the CF is primarily due to the transportation mechanism of electrons in these oxygen vacancies termed as Trap Assisted Tunneling (TAT). After the process of forming, the resistance of the RRAM is at the lowest (LRS). The resistance in LRS depends on the current compliance as shown in characteristic plot in Fig. 3.1. The SET process is same as forming except that only a part of CF is recovered as compared to forming process (Fig. 3.2). Also, SET is performed following a RESET process and SET voltage depends linearly on the RESET voltage [75, 78]. The process of setting state to HRS state is called RESET process. During RESET the oxygen ions drifted to the anode return to the bulk to combine with the oxygen vacancies or oxidize the metal precipitates. The rate of reoxidation depends on the magnitude of the RESET voltage [75].

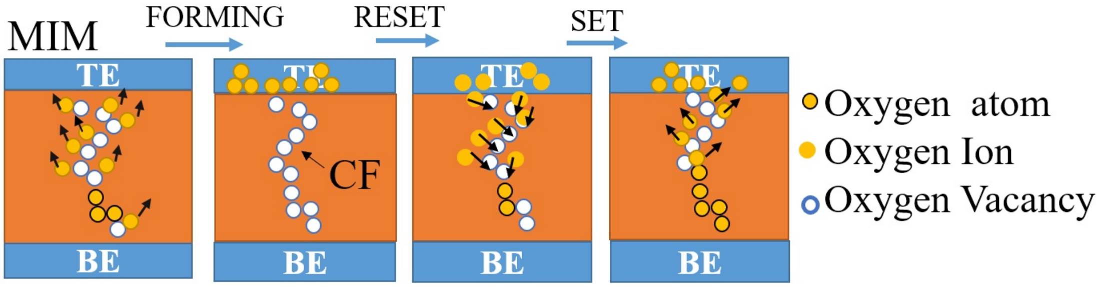


Figure 3.2: Forming, SET and RESET switching mechanism in RRAM.

RRAM is further classified as anion and cation type based on resistive switching mechanism. SET process in anion type RRAM is due to formation of oxygen vacancies in the oxide layer and RESET process is due to recombination of oxygen ions back in the oxide layer. Oxygen active electrodes or thin films oxygen reservoirs are required anion type RRAM. Performance and compatibility with CMOS fabrication technology are primary reasons for wide acceptance high-k metal-oxide materials such as HfO_x, TaO_x, AlO_x in the switching layer of anion type RRAM.

Switching process in cation type RRAM is dominated by redox reaction and migration of metal ions, formation of metallic filaments in switching layer. Conductive Bridge RAM (CBRAM) is an example of cation type RRAM. CBRAM devices use Ag or Cu oxidized electrode and an inert electrode. Switching mechanism in CBRAM can be explained as follows. First, Cu atoms oxidize and Cu^+ ions from copper electrode are injected into dielectric layer, then applied negative bias attracts the Cu^+ ions to the bottom electrode to establish a conductive path (low resistance state). On application of positive bias voltage electrochemical reaction due to Joule heating ruptures the conductive path at a maximum power dissipation point (high resistance state). High $\frac{R_{ON}}{R_{OFF}}$ ratio, longer retention (10^9) and small operating voltages are attractive features of CBRAM [79].

3.3 RRAM for Hardware Security

The resistance of the RRAM after SET and RESET follow probability distribution due to defects in the CF and the thermal voltage fluctuations. The variability in the cycle-to-cycle resistance switching which is a source of randomness can be exploited for security applications. RRAM shows the intra-device temporal variations in switching process. HRS and LRS vary cycle-to-cycle [75] and the resistance after switching depends on the generation and recombination of oxygen vacancies. This is stochastic process induced by the electric field and the temperature of the oxide under the applied switching voltage [75, 78]. The resistance switching model is based on TiN/Ti/HfO_x/TiN RRAM device having a physical oxide thickness t_{ox} of 5nm. Sources of randomness in the RRAM device and models used in this work are discussed in this section.

$$R_{set} = \rho_{Hf,CF} \times \frac{t_{ox}}{S} \tag{3.1}$$

$$R_{reset} = \rho_{Hf,CF} \times \frac{(t_{ox} - x)}{S} + R_{set} \times (\exp\left\{\frac{x}{k}\right\} - \frac{x}{t_{ox}}) \tag{3.2}$$

$$\begin{gathered} \frac{dx}{dt} = R_{set} \times C_{xv} \times |V - V_{reset}| \quad ifV > V_{reset} \\ \frac{dx}{dt} = 0 \quad otherwise \end{gathered} \tag{3.3}$$

$$I(x, V) = I_0(x) \times sinh(\frac{V}{V_0}) \tag{3.4}$$

$$I_0(x) = \frac{V_0}{R(x)} \tag{3.5}$$

$$R(x, T) = R_{set} \times \frac{(t_{ox} - x)}{t_{ox}} + R_{set} \times e^{\frac{x}{k} - 1} \times e^{\frac{E_r}{K_b \times T}} \tag{3.6}$$

where $\rho_{Hf,CF}$ is the resistivity of the CF, tox is the hafnium oxide thickness and S is the cross section of the CF. $\rho_{Hf,CF}$ depends on the current compliance used during forming (Fig.3.1). It is evident that higher current compliance induces a larger CF. V_{reset} is the function of time and ramp voltage with the peak of 1.3V is applied for RESET. C_{xv} is the proportionality coefficient. V_0 is experimentally measured quantity.x is the barrier length created by reoxidation of CF by the reset voltage. E_r is the activation energy, K_b is Boltzmann constant and T is the temperature of the device.

- *Cycle-to-cycle variation:* We have used the parameters and the equations to model cycle-to-cycle switching variations in RRAM from [75] that are calibrated with experimental data.

Current compliance of 100µA is used for modeling the SET resistance. Eqns. 3.1 to 3.6 are used to model RRAM in Verilog-A. RESET process is performed by negative ramp voltage and the differential barrier length with the voltage modeled by Eqn. 3.3. The current during RESET and SET process is given by Eqn. 3.4. RESET is a thermally activated process. The temperature increases with the electric power and overcomes the activation energy to switch the state of the device. Switching of the device at an applied RESET voltage is probabilistic activity [78, 80] model the variation in the resistance of the RRAM due to defects in the oxide material (Fig. 3.3) we assume Gaussian distribution in the SET resistance of RRAM with the variance of 0.08 [75]. The RESET resistance is calculated using Eqns. 3.4-3.6 by assuming Gaussian distribution of the proportionality coefficient C_{xv} with variance of 0.034. C_{xv} models the stochastic variation in the CF rupturing process due to recombination of oxygen vacancies with the ions [75, 81] from cycle-to-cycle. Due to exponential dependence of RESET resistance on the barrier length HRS exhibits lognormal distribution characteristics as shown in Fig. 3.3.

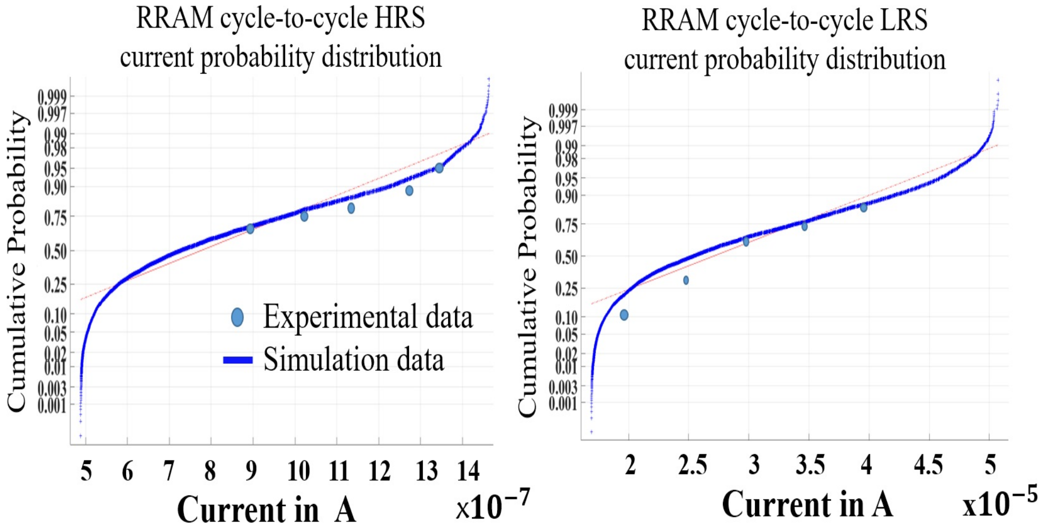


Figure 3.3: LRS and HRS distribution of RRAM. Defects in the CF, cumulative probability distribution plot from [75] and simulated data.

- *Random Telegraph Noise:* Conduction in the RRAM is explained by TAT of electrons in CF. Due to random distribution in the TAT supporting defects, the current though the RRAM shows stochastic variations with time. The phenomenon responsible for RTN is explained by the researchers [82, 76, 83] as charging and discharging of the traps at or close to the surface of the CF. Also, the frequency of trap charging increases with the bias voltage (voltage across RRAM) and temperature due to local Joule heating of CF. The trapping/emission time of the defects near the CF junction can be modeled as lognormal distribution [82, 83]. RTN results in current fluctuations through the RRAM with time. However, the relation of the trapping/emission times with the current fluctuation is still unclear and is determined to be randomly distributed. Variation of RTN current directly related to the fluctuations of current through RRAM [3]. Essentially, RTN is a multi-level low frequency noise in the RRAM of kHz range. RTN can be characterized by Factorial Hidden Markov Model (FHMM) [84] by superposing the multiple two-level RTNs. However, this doesn't provide a deterministic circuit model that could be adapted for circuit analysis. RTN being a truly random process in RRAM leading to read current fluctuations, exhibits no deterministic behavior which could be modeled without direct access to RRAM. Modeling RTN as normal distribution component in RRAM current is the simplistic model for circuit analysis. It should also be noted that the cycle-to-cycle switching parameter variations and RTN are uncorrelated but concurrent in nature [76].

In this work, multi-level RTN in HRS state (RESET) is modeled as variable current source (I_{RTN} through RRAM with 20%-30% variation in the steady current in HRS state by fitting in normal distribution curve shown in Fig. 3.4 [76, 83, 3]. This RTN model follows the RTN current measurements in [3]. The frequency of current fluctuation is affected by the temper-

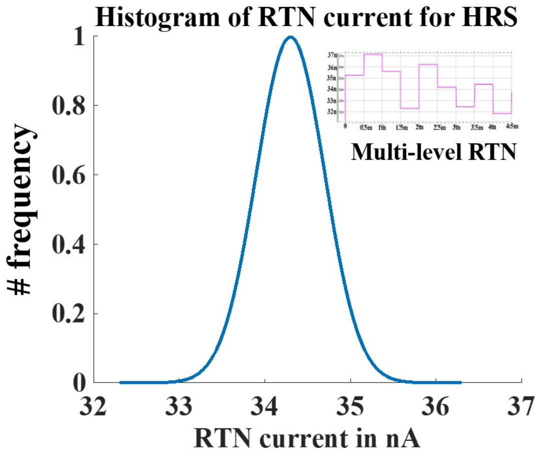


Figure 3.4: RTN current distribution in HRS state of RRAM based on measurement data [3]. I_{RTN} with the frequency of 5kHz is shown in the inset.

ature which is due to longer trapping and emission periods of electron at lower temperature compared to those at higher temperatures [83].

3.4 Summary

This chapter presented the basics of RRAM technology, types of RRAM devices. Also, variations in device characteristics that are prominent for application in hardware security are discussed. RRAM is a viable component compatible with CMOS fabrication technology and competent with CMOS DRAM in terms speed and area. However, to make RRAM practical for memory applications stability and uniformity in the switching characteristics for performance need to be improved. Phenomenon of RTN and formation of traps in the oxide layer responsible for RTN need to be completely understood and modeled accurately for study in hardware security applications.

CHAPTER 4 : 6T-2MTJ TERNARY CAM[2]

4.1 Introduction

Content Addressable Memory (CAM) finds numerous applications in pattern matching, internet data processing, packet forwarding, tag bits storage in processor cache, and as associative memory. The special functionality of the content search in CAM requires a comparison circuitry integrated with the memory cell [85]. The comparator in addition to a memory element adds area and power overhead in CAMs. The need to store and match 'don't care' requires two storage bits which further worsens the area overhead. CMOS CAM is power hungry due to power consumed in Match Line (ML), search line and leakage of the bit cell. In nanometer technologies leakage power constitutes a major fraction of the total power consumed in CAM memory [86]. Non-volatile technologies which are more area efficient than a SRAM and also can provide zero leakage are attractive in such a scenario [85, 87]. Area efficiency and, non-volatility of STTRAM-based ternary CAM is very useful for on-chip CAM applications. Numerous works have demonstrated the realization of CAM using non-volatile memory technologies like memristor [88],nano-electro-mechanical switch [89], Resistive RAM (ReRAM) [90] and spintronic elements such as Domain Wall Memory (DWM), Magnetic Tunnel Junctions (MTJ) and Spin-Torque-Transfer RAM (STTRAM) [91, 92, 93, 94]. Memristor-based NOR TCAM [88] uses a voltage divider network formed by the memristors to enable the discharging

[4] Portions of this chapter were reprinted from Govindaraj, R., & Ghosh, S. (2015, July). Design and analysis of 6-T 2-MTJ ternary content addressable memory. In Low Power Electronics and Design (ISLPED), 2015 IEEE/ACM International Symposium on (pp. 309-314). IEEE. and Govindaraj, R., & Ghosh, S. (2017). Design and Analysis of STTRAM-Based Ternary Content Addressable Memory Cell. ACM Journal on Emerging Technologies in Computing Systems (JETC), 13(4), 52.
Permission is included in Appendix A.

path to discharge the match line depending on match/miss. It incurs higher write delay of up to 800ns which hurts the table update performance in the network routers application. ReRAM-based TCAM [90] employs a special clocked self-referenced sensing scheme which complicates the memory system design due to additional reference ML required. An efficient integration of spintronic device with CMOS technology motivates to employ them in TCAM cell design. Also, the principle of bit storage and writing methodologies are different in memristor and ReRAM compared to spintronic device such as MTJ and DWM. This limits the direct extension of memristive and ReRAM TCAMs to design spintronic TCAMs.

Spintronic CAMs using MTJ and DWM suffer from issues such as larger area, unreliable write operation, high search delay and high power consumption compared to CMOS CAMs [86]. The MTJ based TCAM 4T-2MTJ design [95] is area and power efficient however it employs a technique based proportional total current drawn from the ML by bit cells in word. Further, the circuit uses both NMOS and PMOS transistors. In this technique sense margin (SM) decreases significantly as number of bits increase in the worst case search of single bit miss. So, the design is not scalable beyond a certain word length (till 144 bit is presented).DWM based TCAM [93] uses 12 transistors. Therefore, deployment of non-volatile memory in CAM needs effort to achieve smaller footprint and better performance in terms of search delay, write delay, write power and search power.

This chapter proposes a NOR type MTJ based TCAM [52] that can support wider CAM words while being tolerant to voltage and temperature variations. However, it is susceptible to poor sense margin due to process variations. Search Enable (SE) modulation to improve the sense margin under inter-die process variation is proposed.Multi-V_{TH} design to improve the sense margin and a NAND type TCAM that exploits the NOR TCAM design are also proposed. The proposed NOR type TCAM employs only 6 transistors and 2 MTJs instead of 16 transistors in CMOS TCAM, and NAND TCAM cell uses 9 transistors and 2 MTJs compared to NAND CMOS TCAM that uses 16 transistors. The following section provides overview of Content Addressable Memory(CAM).

4.2 Content Addressable Memory

CAMs can be divided into two categories depending on the number of states that can be stored in the memory cell namely, Binary CAM (BCAM) and Ternary CAM (TCAM). BCAM stores a binary bit i.e., '0' and '1' whereas TCAM can store three possible values namely, 'don't care' (X), '1', and '0'. CAMs can be further categorized into two topologies namely, NOR and NAND type (Fig. 4.1). The stored bits are compared with the data on the Search Line (SL) and its complement ($\bar{SL}$) by XOR operation with the transistor network M1, M2, M3 and M4. To store data in a TCAM cell of NOR type architecture data bit and the complement are stored in two SRAM cells (Fig. 4.1a). Don't care bit can be realized by storing '1' in both SRAM cells i.e., $D = \bar{D} = 1$. In case of match both SL-D and $\bar{SL} - \bar{D}$ paths are disconnected and the match line remains precharged. In case of miss either of the SL-D or $\bar{SL} - \bar{D}$ connect ML to ground which discharges the precharged ML. In a NAND type architecture TCAM cells are connected in series (Fig. 4.1b). Data bit D and $\bar{D}$ are derived from a single SRAM cell unlike two SRAM cells in NOR type TCAM. The stored bit is masked by using a mask bit (M) in a parallel SRAM cell. In case of match the precharged ML is connected to ground by series TCAM cells of the word by turning the NMOS transistor M1 ON. Storing the mask bit as '1' enables transistor M2 despite match or miss which implements don't care functionality. CMOS TCAM uses two SRAM cells which doubles the area overhead compared to conventional SRAM cell.

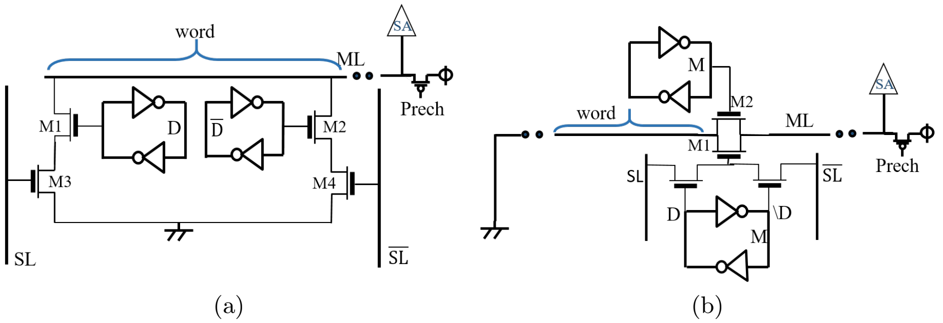


Figure 4.1: Conventional CMOS TCAM types. (a) NOR type TCAM; and, (b) Conventional NAND type TCAM. Bit line access transistors are not shown in the figure for simplicity.

4.3 Proposed TCAM Cell

In this Section first we discuss the structure of the proposed TCAM. Next we present qualitative analysis and describe read, write and search operations.

4.3.1 NOR TCAM Cell Circuit

Circuit diagram of the proposed TCAM is shown in Fig. 4.2. Two MTJs store D and $\bar{D}$ respectively. Transistors M1 and M2 form ML discharge network depending on the result of data comparison with the search lines SL and $\bar{S}L$. During search transistors M3/M5 and M4/M5 along with MTJ resistance due to TMR make a voltage divider network in which the drain voltages of M3/M4 drive the gates of discharge transistors M1/M2. The cell is designed in such a way that during match the voltage of node X1 and X2 is below the threshold voltage of M1 and M2, and the ML stays precharged. However, during a mismatch the voltage of X1 and X2 rises above the threshold voltage of M1 and M2 respectively discharging the ML. Transistor M3/M4 are the wordline (WL) selection transistors and M6 is the write access transistor that turns ON only during write (WR) operation. Transistor M6 is sized larger to allow sufficient write current. Transistor M5 is driven by Search Enable (SE) signal and, sized to limit the current through STTRAM bit for read disturb free search operation. Don't care bit can be stored in the cell by storing '1' in both D and $\bar{D}$ bits. The search bit can be masked by driving $SL = \bar{SL} = 0$ on the search lines. The Source Line (SrL) is used for two purposes namely, (a) write operation when the SrL is connected to 0 or Vdd depending on the write data to the MTJs; and, (b) search operation when SrL is driven to 0 to allow voltage division.

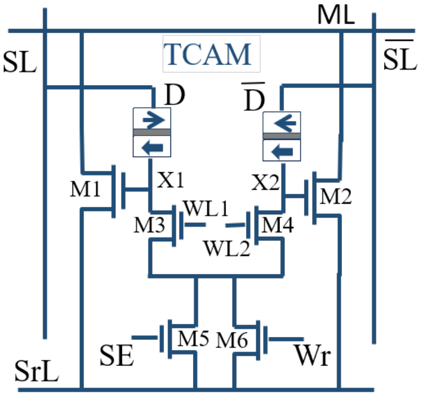


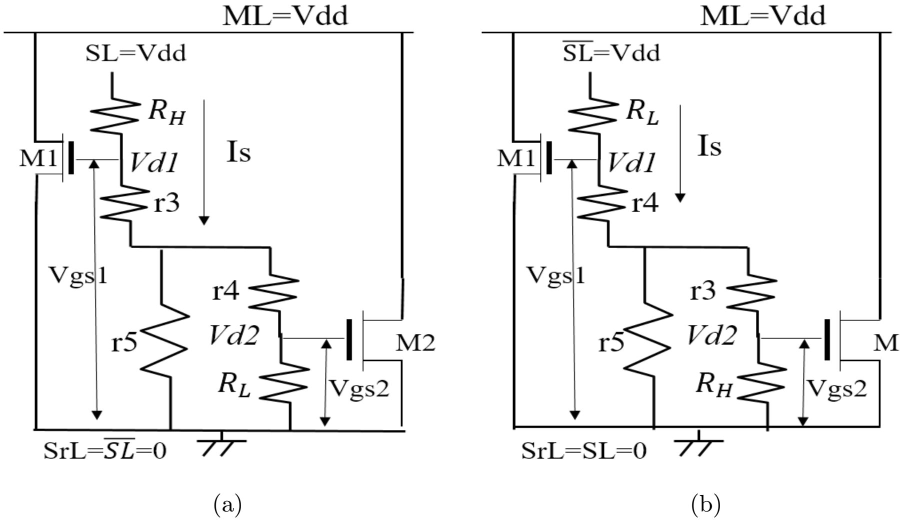
Figure 4.2: Proposed NOR type TCAM cell.

4.3.2 Qualitative Analysis of the Cell Design

There are two match cases namely, (a) $(D, \bar{D} = (SL, \bar{SL}) = (1, 0)$; and, (b) $(D, \bar{D}) = (SL, \bar{SL}) = (0, 1)$. Since both cases are identical, we will only explain the first case. For $(D, \bar{D}) = (1, 0)$ the left side MTJ is in high resistance (R_H) state whereas the right side MTJ is in low resistance (R_L) state. Since $(SL, \bar{SL}) = (1, 0)$, the voltage at node X1 is

$$V_{X1} = V_{SL} \times \frac{r}{(R_H + r)} = V_M \tag{4.1}$$

and the voltage at node X2 is voltage drop due to current flow from node X1 to $\bar{SL}$ (detailed analysis is given in Section 4.3.4). In this expression, r is the lumped ON resistance of transistors M3, M4 and M5 (Fig.4.3), and, V_{SL} is SL voltage. To keep transistor M1 OFF during match,$\bar{V_{X1}}$ should be lower than V_{TH0} (i.e., the threshold voltage of M1 and M2).



(a)

(b)

Figure 4.3: Equivalent circuit during match and mismatch. (a) Match M1 and M2 are turned ON; and, (b) Mismatch M1 and M2 are OFF as $V_{d1} < V_{th0}$

For the mismatch, there are two cases namely, (a) $(D, \bar{D}) = (1, 0)$ and $(SL, \bar{S}L) = (0, 1)$; and, (b) $(D, \bar{D}) = (0, 1)$ and $(SL, \bar{S}L) = (1, 0)$.For the first case the voltage at node X1 is

$$V_{X1} = V_{\bar{S}L} \times \frac{r}{(R_H + r_{eff})} = V_{MM1} \tag{4.2}$$

where $\bar{r_{eff}}$ is the effective resistance of $R_L, r_3, r_5, andr_4$ resistive network. Whereas, voltage at X2 is

$$V_{X2} = V_{\bar{S}L} \times \frac{r}{(R_H + r)} = V_{MM2} \tag{4.3}$$

where $V_{\bar{S}L}$ is $\bar{S}L)$ voltage. To keep transistors M1, M2 ON during mismatch, $V_{MM1}andV_{MM2}$ should be higher than V_{TH0}. Similar analysis applies to case (b). From these equations $V_{MMX} > V_{MX}$ for the two cases as $R_H > R_L$. For the design to function properly (i.e., discharge ML during mismatch at a higher speed compared to that of a match case) R_H, R_L and r should be selected such that $V_{MX} < V_{TH0} < V_{MMX}$. The following analytical equations can be used to quantify the design parameters.

$$V_{MM} = V_{dd} - I_{MM} \times R_L = V_{\bar{S}L} \times r/(R_L + r) = V_{th0} + \Delta_1 \tag{4.4}$$

$$V_M = V_{dd} - I_M \times R_H = V_{SL} \times r/(R_H + r) = V_{th0} - \Delta_2 \tag{4.5}$$

where I_{MM} and I_M are the currents drawn from SL and $\bar{S}L$ in case of mismatch and match respectively, and, Δ_1 and Δ_2 are the offset voltages with respect to V_{th0}.

Subtracting 4.4 and 4.5 and using $R_H = R_L \times (1 + TMR)$, we obtain

$$V_{MM} - V_M = V_{dd} \times \frac{r}{R_L + r} - \frac{r}{R_H + r} = \Delta_1 + \Delta_2 \tag{4.6}$$

$$V_M = V_{dd} \times \frac{r \times R_L \times TMR}{(R_L + r) \times (R_L(1 + TMR) + r)} = \Delta_1 + \Delta_2 \tag{4.7}$$

The optimization of the proposed TCAM revolves around three key requirements: (a) maximizing the difference between mismatch and match voltages i.e., $(\Delta_1 + \Delta_2)$; (b) maximizing the absolute values of offsets from V_{TH0} i.e., $|\Delta_1|$ and $|\Delta_2|$ to keep M1/M2 strongly ON or OFF as needed during mismatch and match respectively; and, (c) lowering the search current below critical write current of MTJ.

From 4.7, it can be concluded that higher TMR, higher R_H and higher r can be employed to enhance $(\Delta_1 + \Delta_2)$. Although higher r and R_L is also good for maximizing Δ_1, it minimizes Δ_2.

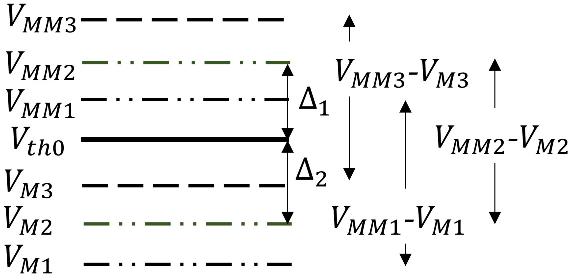
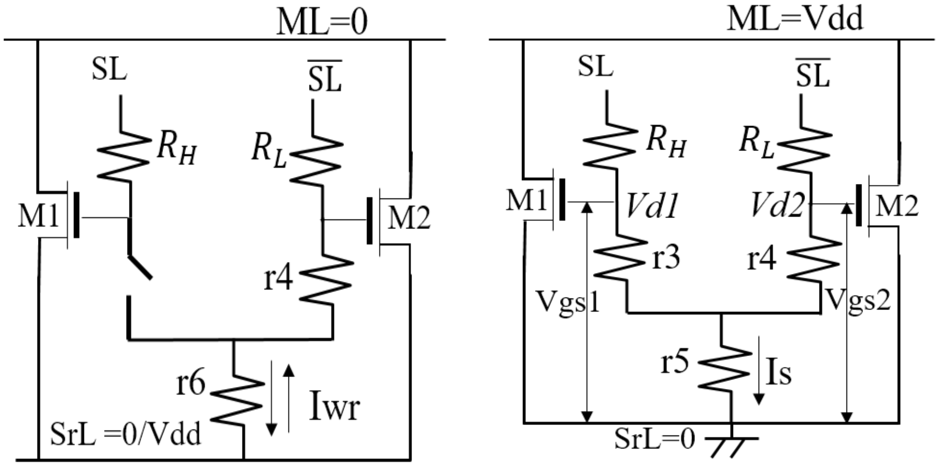


Figure 4.4: V_{gs} margin diagram illustrating best and worst V_M and V_{MM} with V_{th0}.

A lower Δ_2 can turn M1/M2 ON during match degrading the sense margin. Fig.4.4 shows pictorial representation of this situation with three operating points. The voltages V_{MM1}, V_{MM3}, V_{M1} and V_{M3} provide poor sense margin compared to V_{MM2} and V_{M2} even with same magnitude of $\Delta_1 + \Delta_2$. The ideal margin is obtained when $R_H = \infty$ and $R_L = 0$ which gives $V_{MM} = V_{dd}$ and $V_M = 0$. However, a lower R_L could be detrimental for read disturb due to high search current. High values of R_H and R_L ensure the low search line currents. This in combination with high TMR can provide better V_{gs} margin i.e., $(\Delta_1 + \Delta_2)$ with low search power consumption. The design optimization conducted accounts for the above factors.

4.3.3 Write Operation

In the proposed TCAM the search lines SL and $\bar{SL}$ are used to write data to the STTRAM bits. Table-4.1 summarizes the states of control signals in write operation. Writing '1' and '0' consume two cycles to write to the two STTRAMs while 'X' can be written in a single cycle. During write the ML precharge is disabled to avoid power consumption from the ML. This is achieved by pulling the 'precharge' signal high. NMOS transistor M6 is turned ON during write by WR signal. Note that M6 is sized to provide the drain current greater than the critical write current of the STTRAM. The state of search enable signal SE is 'Don't care' as M5 is connected parallel to M6.



(a)

(b)

Figure 4.5: Equivalent circuit during write and search operation. (a) Write operation, left side STTRAM resistance is R_H (D='1') and right side STTRAM resistance is $R_L(\bar{D}$ ='0'); and, (b) search operation.

For the analysis, we assume that SE is pulled low. The WL is turned ON only for the selected word so that the unselected cells are unaffected. The source line SrL is controlled appropriately to write a '1' or '0'. Fig. 4.5a shows the equivalent circuit of TCAM cell during write. The transistors are replaced with equivalent ON resistances. Resistors r3, r4 and, r6 are equivalent resistors of M3, M4 and, M6 respectively. The write operation is described below. In the first cycle of write operation, writing to D bit is enabled by pulling WL1 to V_{dd} and $\bar{D}$ bit path is disabled by pulling WL2 to ground. In the second cycle of write operation, writing to $\bar{D}$ bit is enabled by pulling WL2 to V_{dd} and D bit path is disabled by pulling WL1 to ground.

Table 4.1: States of control signals NOR TCAM memory write operations. WR=V_{dd}, SE='X'

	Write D-bit (WL1= V_{dd}, WL2=0)			Write $\bar{D}$-bit (WL1=0, WL2=V_{dd})		
	SL	$\bar{S}L$	SrL	SL	$\bar{S}L$	SrL
Write '1'	0	X	V_{dd}	X	V_{dd}	0
Write '0'	V_{dd}	X	0	X	0	V_{dd}
Write 'X'	0	0	V_{dd}	WL1=WL2=V_{dd}		

- Writing '1': In the first cycle, SrL is pulled high and SL line is pulled to ground. The write current flows from SL writing antiparallel state to the STTRAM storing bit D. There is no current through the other STTRAM bit as the WL2 control signal is grounded. In the second cycle the SrL is pulled low, $\bar{SL}$ is pulled to V_{dd} and WL2 is pulled high which programs the other STTRAM storing $\bar{D}$ to parallel state. There is no current through the other STTRAM bit as WL1 is grounded.

- Writing '0': In the first cycle, the SL is pulled high and the SrL line is pulled low. This cycle writes parallel magnetization state to STTRAM storing D bit. In the second cycle, the SrL is pulled high while $\bar{SL}$ is at 0, which programs the $\bar{D}$ bit to antiparallel state.

- Writing 'X': The 'X' state can be stored by writing logic 1 to both D and $\bar{D}$. The SrL is pulled to V_{dd} and the search lines SL and $\bar{SL}$ are pulled low. The current flows through both the STTRAMs storing antiparallel states to D and $\bar{D}$.

4.3.4 Search Operation

Search is a single cycle operation in CAM. The ML is precharged to V_{dd} and WR is pulled to ground. The SrL is pulled to ground throughout the search operation. Next SE and WL is pulled high to enable the conducting path through M5 and M3/M4 (4.2). Either V_{MM} or V_M voltage is developed depending on the match or mismatch respectively at the gate of M1/M2. The search lines SL is pulled to V_{dd} and $\bar{SL}$ is pulled low to search a bit '1'. Similarly, SL is pulled low and, $\bar{SL}$ is pulled to V_{dd} to search for bit '0'. Both SL and $\bar{SL}$ are pulled low to search 'X'. Circuit operation in match and mismatch cases are discussed below. Fig. 4.5b shows the equivalent circuit during search operation.

- Match: Let (D,$\bar{D}$) = (SL, $\bar{D}$) = (1, 0). Voltages V_{X1} and V_{X2} at the nodes X1 and X2 (Fig.4.2) are given by,

$$V_{X1} = V_{dh} = V_{dd} \times \frac{(r3 + (r5 \parallel (r4 + R_L))}{R_H + r3 + (r5 \parallel (r4 + R_L))} \tag{4.8}$$

$$V_{X2} = V_{dl} = V_{dd} \times \frac{((r5 \parallel (r4 + R_L)) \times R_L)}{(R_H + r3 + (r5 \parallel (r4 + R_L)))(r4 + R_L)} \tag{4.9}$$

  Note that V_{X2} is less than V_{X1} and appears due to the potential across r5 which results in a current though R_L even when $\bar{SL}$=0 (Fig. 4.3a). The transistors M3 and M5 are sized such that $V_{X1} < V_{th0}$. So M1/M2 are turned OFF and the ML remains precharged. The other match case i.e., (D,$\bar{D}$) = (SL,$\bar{D}$) = (0, 1) is similar.

- Mismatch: Mismatch: Let (D,$\bar{D}$) = (1, 0) and (SL,$\bar{SL}$) = (0, 1). Then,

$$V_{X1} = V_{dl} = V_{dd} \times \frac{(r4 + (r5 \parallel (r3 + R_H))}{R_L + r4 + (r5 \parallel (r3 + R_H))} \tag{4.10}$$

$$V_{X2} = V_{dh} = V_{dd} \times \frac{((r5 \parallel (r4 + R_H)) \times R_H)}{(R_L + r3 + (r5 \parallel (r4 + R_H)))(r4 + R_H)} \tag{4.11}$$

Table 4.2: States of control signals NOR TCAM memory search operations. WL1=WL2=V_{dd}, WR=0, SE=V_{dd}

Operation	SL	$\bar{SL}$	SrL
Search '1'	V_{dd}	0	0
Search '0'	0	V_{dd}	0
Mask search	0	0	0

  $V_{dl}(miss) > V_{th0} > V_{dh}(match)$. Under these conditions (Fig. 4.3b) both M1 and M2 are turned ON to discharge the precharged ML which provides better sense margin. Fig.4.6 illustrates the ML voltages during search operation for TCAM of varied word sizes namely 1, 16, 128 and 256-bit for match and mismatch. Predictive 22nm model is used for simulations [96]. The waveforms correspond to the worst case sense margin i.e., single miss in the whole word. The rate of discharge of ML line in match case increases with the word size due to the more number of cells leaking the ML current through weakly driven M1/M2. This in turn limits the sense margin for larger word sizes. The equations V_{X1} and V_{X2} can be also be derived for the transistor device parameters by replacing the voltage across transistor by $V_d s$ and drain current of the transistors by the current through MTJ as below. The objective is to optimize the sense margin, search power and limit the drain current below the critical current of the MTJ.I_{d3}, I_{d4}, I_{d5} are drain currents and $V_{ds3}, V_{ds4}, V_{ds5}$ are the drain to source voltages of transistors M_3, M_4, M_5 respectively. I_{c0} is the STTRAM critical current.

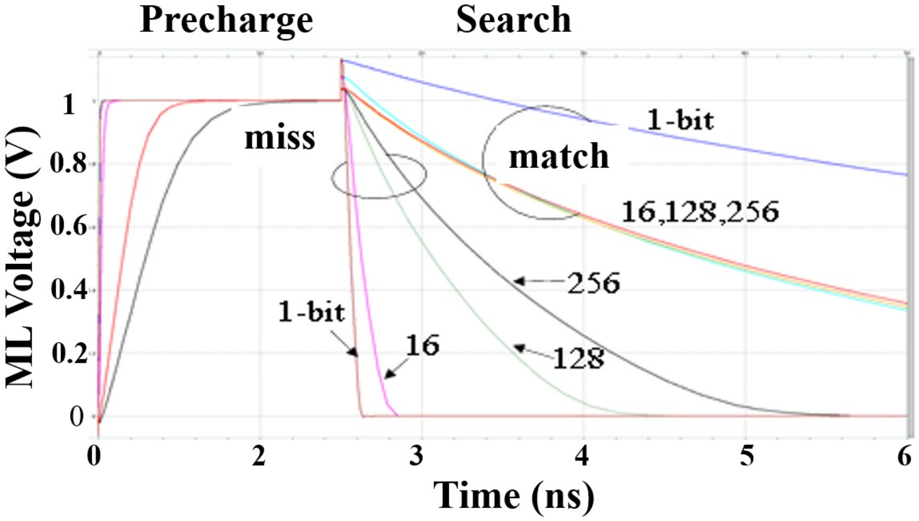


Figure 4.6: Waveform showing the search operation. Match line voltages during mismatch and match cases for 1, 16, 128 and 256-bit word sizes are shown.

- Match: Match: Let (D, $\bar{D}$) = (SL, $\bar{SL}$) = (1, 0). Voltages V_{X1} and V_{X2} at the nodes X_1 and X_2 (Fig. 4.2) are given by,

$$V_{X1} = V_{dh} = V_{dd} - (I_{d3} \times R_H); V_{X2} = V_{dl} = I_{d4} \times R_L \tag{4.12}$$

  $I_{d3} = I_{d4} + I_{d5}; V_{ds5} = V_{ds4} + I_{d4} \times R_L; I_{d4} \ll I_{c0}$ of the MTJs; $V_{X1} > V_X \ll V_{TH1}$, V_{TH2}

- Mismatch: Let (D,$\bar{D}$) = (1, 0) and (SL,$\bar{SL}$) = (0, 1). Then,

$$V_{X1} = V_{dh} = V_{dd} - (R_H; V_{X2} = V_{dl} = I_{d3} \times R_H \tag{4.13}$$

  $I_{d4} = I_{d3} + I_{d5}; V_{ds5} = V_{ds3} + I_{d3} * R_L; I_{d3} \ll I_{c0}$ of the MTJs; $V_{X1} > V_{X2} \gg V_{TH1}, V_{TH2}$ Drain currents and V_{ds} of transistors M_3, M_4, M_5 are different in case 1 and case 2. However, deriving analytical expressions for transistor parameters from above expressions is straightforward but tedious with short channel transistor equations. We have adopted simulation based approach to minimize such efforts.

4.4 Proposed NAND Type TCAM Cell

Two types of TCAM topology are traditionally investigated in the literature i.e. NAND and NOR. Typically, NAND topology TCAM is faster compared to NOR topology TCAM in full CMOS realization. We investigate NOR and NAND topology TCAM realization using STTRAM which completes the study of STTRAM based TCAM design. In this section, we propose a NAND type TCAM cell using STTRAM. Fig. 4.7 shows the circuit diagram of NAND type TCAM cell along with the match line structure. The cell consists of 2 PMOS, 7 NMOS transistors and 2 MTJs.

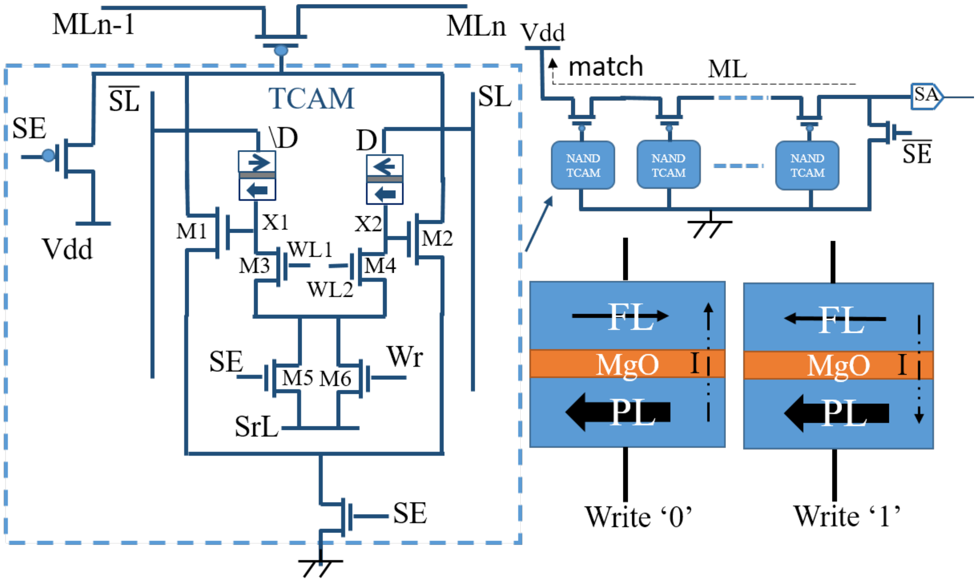


Figure 4.7: Proposed NAND configuration TCAM cell (Bit 0 stored).

We use the complementary method i.e., bit '1' is encoded by parallel magnetic spins and bit '0' is encoded as antiparallel states of relative magnetic spins, to realize NAND type TCAM in this work. Six NMOS transistors M1-M6 are sized for reliable search and write operation same as NOR type TCAM explained earlier. In other words, the design analysis of NOR TCAM cell embedded in the NAND type TCAM cell remain similar to that of a NOR type cell. For successful search operation in NAND TCAM data bit '1' is encoded as parallel state and bit '0' as antiparallel state of the MTJ respectively. Three additional transistors (M7, M8, M9) are added on the basic NOR TCAM cell to realize the NAND type TCAM cell. These additional transistors are of minimum size. For search operation, initially the match line is predischarged and the chain of PMOS transistors (M9 in Fig. 4.7) of individual TCAM cell connects the match line to V_{dd} only in case of a complete match. Gates of PMOS transistors in the chain are precharged initially such that the chain is completely disconnected from V_{dd}. The search data and search enable signals are asserted after predischarging of the match line. In case of a match the respective gate voltage is pulled low by the

NMOS transistors M1/M2 in the search circuit. Thus, the match line is pulled high to complete V_{dd} in case of complete match of a word. Table-4.3 and Table-4.4 summarizes the control signals in write and search operation of NAND TCAM respectively. Writing is performed by injecting current more than the critical current from the search lines and source line (SrL). SE signal is 'X' and Wr is pulled high throughout the write operation. Transistor M6 is sized to carry the current higher than the critical current of the MTJs for programmability. WL1 and WL2 are pulled to V_{dd} to enable writing 'D' bit and '$\bar{D}$' bit respectively in the first and second cycles of write respectively.

Table 4.3: States of control signals NAND TCAM memory write operations. WR=V_{dd}, SE='X'

	Write D-bit (WL1= V_{dd}, WL2=0)			Write $\bar{D}$-bit (WL1=0, WL2=V_{dd})		
	SL	$\bar{SL}$	SrL	SL	$\bar{SL}$	SrL
Write '1'	V_{dd}	X	0	X	0	V_{dd}
Write '0'	0	X	V_{dd}	X	V_{dd}	0
Write 'X'	V_{dd}	V_{dd}	0	WL1=WL2=V_{dd}		

Table 4.4: States of control signals NAND TCAM memory search operations. WL1=WL2=V_{dd}, WR=0, SE=V_{dd}

Operation	SL	$\bar{SL}$	SrL
Search '1'	V_{dd}	0	0
Search '0'	0	V_{dd}	0
Mask search	0	0	0

- Write '1': Bit '1' can be stored in two cycles of write i.e., by writing parallel state to the MTJ storing D-bit and anti-parallel state to the MTJ storing $\bar{D}$ bit. In the first cycle, WL1 and SL are driven to V_{dd} and SrL, WL2 are pulled low which results in current flow from free layer to fixed layer writing parallel state to the MTJ storing 'D' bit. In the second cycle, WL2, SrL are precharged to V_{dd} while WL1 and $\bar{SL}$ are pulled low. This results in the current flow from the PL to FL storing antiparallel state in '$\bar{D}$' bit.

- Write '0' : Bit '0' is stored by writing antiparallel state in 'D' bit and parallel state in '$\bar{D}$' bit. In the first cycle SrL and WL1 are precharged to V_{dd} while SL is pulled to ground to write antiparallel state to 'D'. In the second cycle WL2 and $\bar{SL}$ are pulled to V_{dd} while SrL is pulled low in order to write parallel state to the '$\bar{D}$'.

- Write 'X': Don't care bit is written by storing parallel state in both the MTJs 'D' and '$\bar{D}$' which results in match case for both search bits '0' and '1'. Writing 'X' can be performed in a single cycle by pulling WL1, WL2, SrL to V_{dd} and SL, $\bar{SL}$ are pulled low to ground which result in writing '0' (parallel state) simultaneously to both the MTJs.

Design analysis of NOR type cell design is presented in the next section. We analyse the design parameter selections (MTJ resistance, TMR, search transistors sizing) for successful search operation in TCAM. Design analysis for write operation remains same as in a conventional STTRAM cell which we have excluded in our work.

4.5 NOR TCAM Cell Design Analysis

In this section, first we present the methodologies to determine the sizing and MTJ resistance for reliable operation of the proposed NOR TCAM. We consider broad range of word sizes in the analysis. In the proposed design, the parameters are optimized for sense margin and search power. Parameters for write current transistor M6 is chosen to drive sufficient write current. All the other parameters in the cell design are optimized for search operation.

4.5.1 Selection of R_L and NMOS Device

The low MTJ resistance and sizing of transistor M5 are chosen to keep the search current below the critical current while providing a sufficient V_{gs} to drive M1/M2 in order to differentiate the miss and match cases. Other than keeping the search current below critical current, limiting the search current is crucial to keep the search power as low as possible while achieving reliable sense margin. Moreover, ensuring the highest search current (through the lowest resistance) yields reliable design parameters i.e. total search current in the TCAM cell is less than the critical current under all PV conditions. The high MTJ resistance is determined by the TMR. The transistor M6 is sized to provide write current greater than critical current through the STTRAM during write operation. We simulated a range of RL (5k to 9k) with fixed TMR of 100%. The trend is shown in the Fig. 4.8 for a 16-bit word. It can be observed from the plot that high resistance values with smaller NMOS widths provide good sense margin (close to $\frac{V_{dd}}{2}$) with lower STTRAM current from the search line. Based on this, $R_L = 8k\Omega$ is selected for the proposed design. The STTRAM current during mismatch is also plotted. Note that mismatch current is always greater than the match current therefore we consider it for estimating the worst case read disturb during search operation.

Width of NMOS devices M3/M4 and M5 are important parameters to ensure low search current and reduce the power dissipated from the search lines. Plot in the Fig. 4.8 shows the distribution of STTRAM current for various widths of the NMOS device M5 with different R_L values. Smaller width of NMOS offer high resistance, reduces search current (good for lower read disturb and power) and improves the sense margin (following the discussion in Section 4.3.2). However minimum sized transistor can be susceptible to manufacturing process variations. We selected 50nm for M5

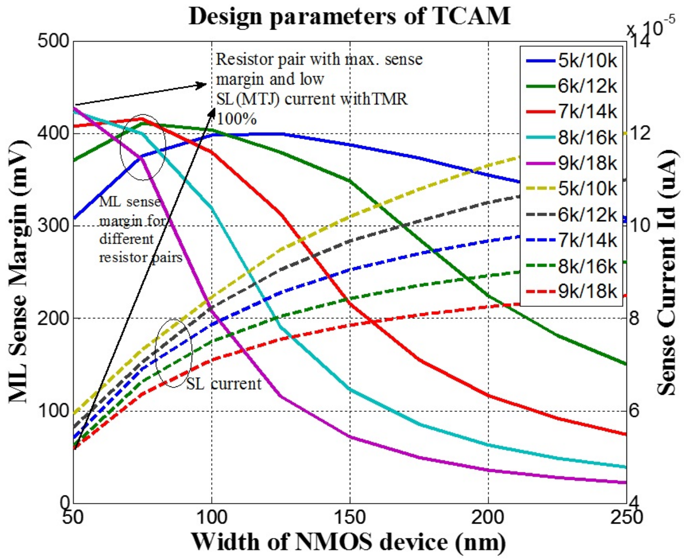


Figure 4.8: Width of M5 v/s SM and STTRAM current from SL for various R_L.

width for the low search current. Further, two transistors of 100nm width in series can be used to minimize the process tolerance. It can be observed from the plot that miss case current is highly dependent on width of M5 NMOS device and remains almost same for different R_L values. High R_L is selected to keep the TMR within practical limits 100-150% [97]. To determine the optimal size of transistors M3/M4 we swept the size and observed the sense margin and sense current for 50nm M5 width (Fig. 4.9). It is evident from the plot that the sense margin increases sharply from 50nm till 200nm. After 200nm improvement in the sense margin saturates. Also, the search current increases by approximately 10X with increase in the width by 25nm. Therefore, we selected the width of M3/M4 to be 200nm.

4.5.2 Impact of TMR on Sense Margin

Fig. 4.10 shows the trend of match current and sense margin versus width of NMOS M5 for different TMR values. The R_L of MTJ is fixed to 8K (as decided in Section 4.5.1) for this analysis,

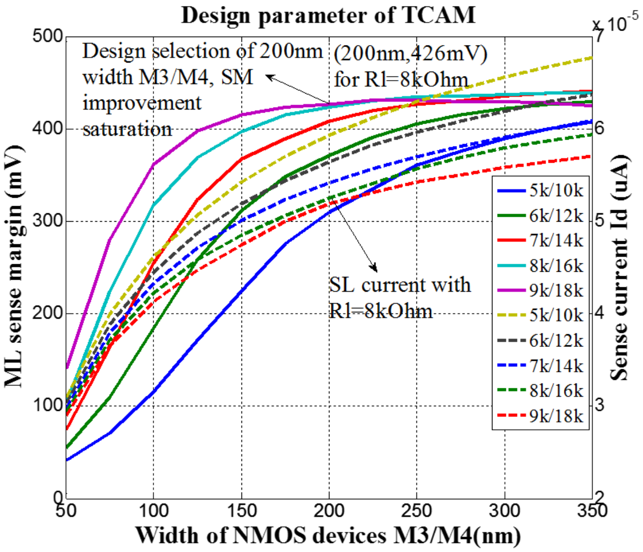


Figure 4.9: Width of M3/M4 v/s sense margin and search current for various R_L.

TMR and R_H are selected for low match case search current and higher sense margin. It can be seen that higher TMR ensures better sense margin and low STTRAM match current with fixed R_L. It can be seen from the plot that the NMOS width does not affect the STTRAM current compared to that in the miss case because the MTJ high resistance R_H dominates the effective NMOS resistance of M3/M4-M5. This also results in low drain voltage at M3/M4 compared to that in the mismatch case. So, the width of NMOS is selected based on the mismatch current drawn from the SL while the TMR is chosen to satisfy the match case conditions. It can be noted that the sense margin benefit of TMR greater than 125% saturates. Hence, we have used TMR=125% that provides less than 45µA of match current with a sense margin close to 500mV.

Resistance of MTJ is shown to depend on oxide thickness and surface area of free layer [87]. Therefore, by tuning these parameters it is possible to obtain MTJ resistance of $R_L = 8k\Omega$. Similarly, it has been experimentally shown that TMR could be improved up to 236% [87]. This ensures the realization of TMR=125% in the design.

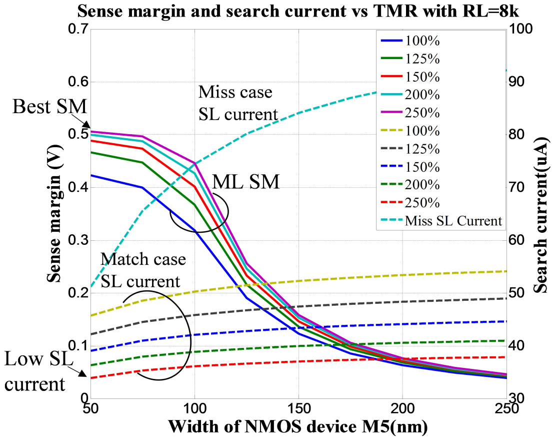


Figure 4.10: ML sense margin and search current with width of NMOS M5 and TMR.

4.6 Simulation Results of NOR TCAM Cell

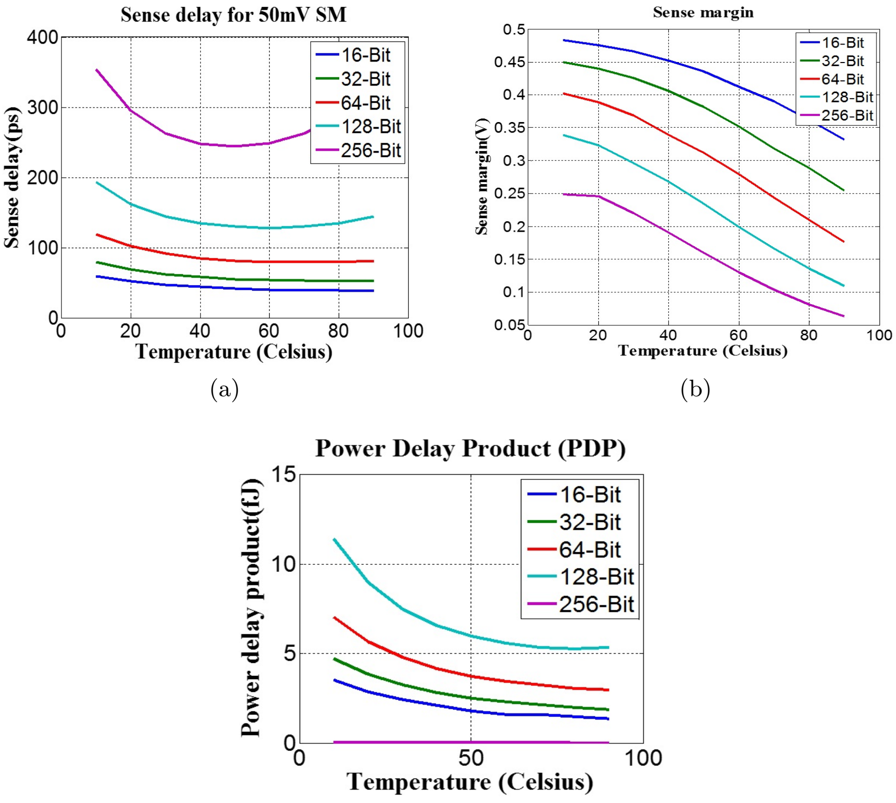
In this section, we present analysis of the proposed TCAM with respect to temperature, voltage and process-variations. We also propose modulating search enable signal and threshold voltage to improve robustness.

4.6.1 Setup

We used TMR=125% with $R_L = 8k\Omega$, 50nm M5 transistor and 200nm M3/M4 transistors (as discussed in Section 4.5). MTJ models from [64] is used with 60nmx60nmx3nm free layer dimension and 0.876nm oxide (MgO) thickness for design simulations. Word size of 16, 32, 64, 128, and 256-bit is simulated to analyze the design with respect to process, temperature and voltage variations.

4.6.2 Temperature Variation Analysis

Thermal fluctuations result in the critical current and switching time variations in the MTJ which is modeled in the effective magnetic field in LLG equations [98]. The worst-case sense margin, search delay (for 50mV sense margin development) and the Power Delay Product (PDP) per bit search from 10^{o}C to 90^{o}C are shown in Fig. 4.11 for different word sizes.



(a)

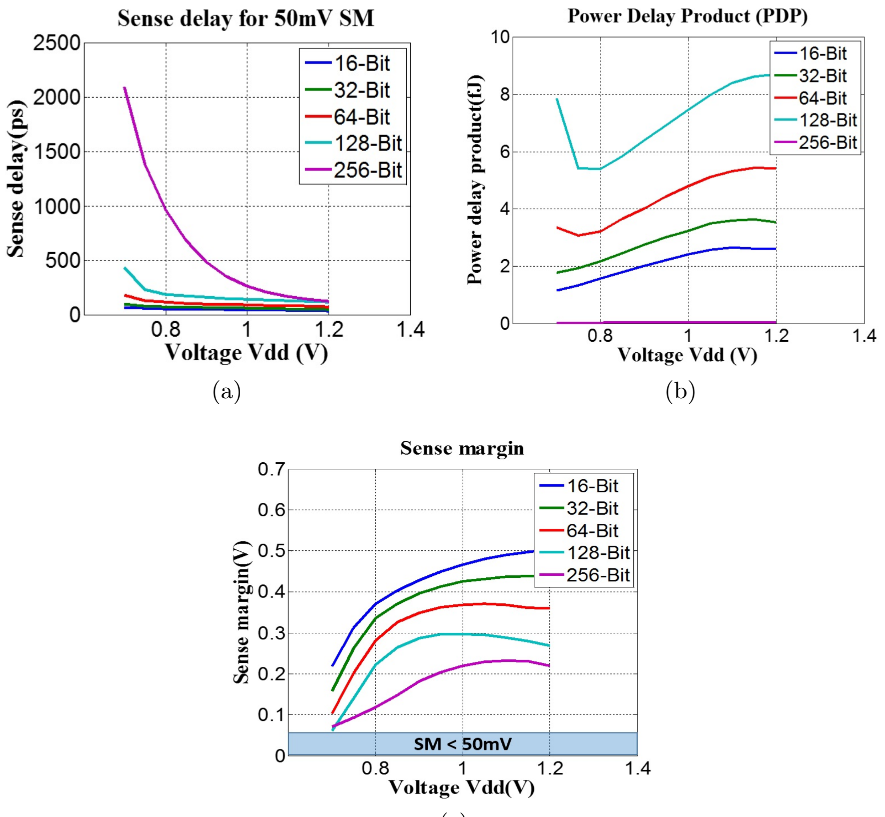
(b)

(c)

Figure 4.11: (a) Sense margin; (b) Search delay; and, (c) PDP v/s temperature.

A single bit mismatch is considered for sense margin and search delay as it is the worst-case condition. The search delay increases proportionally as the word size due to increment in ML interconnect capacitance. As the temperature increases, the rate of ML discharge increases due to lowering of threshold voltage of the discharge transistors M1/M2. Sense margin decreases with temperature due to ML discharge through subthreshold leakage current of discharge transistors in the match case. Therefore, the search delay (for 50mV sense margin) increases with the temperature. The PDP is proportional to the change in search delay while the operating voltage and the search line current are similar across different temperatures. From Fig. 4.11a it is evident that we obtain a reliable sense margin of greater than 50mV across the range of temperature till 256-bit word size.

4.6.3 Voltage Scaling



(a)

(b)

(c)

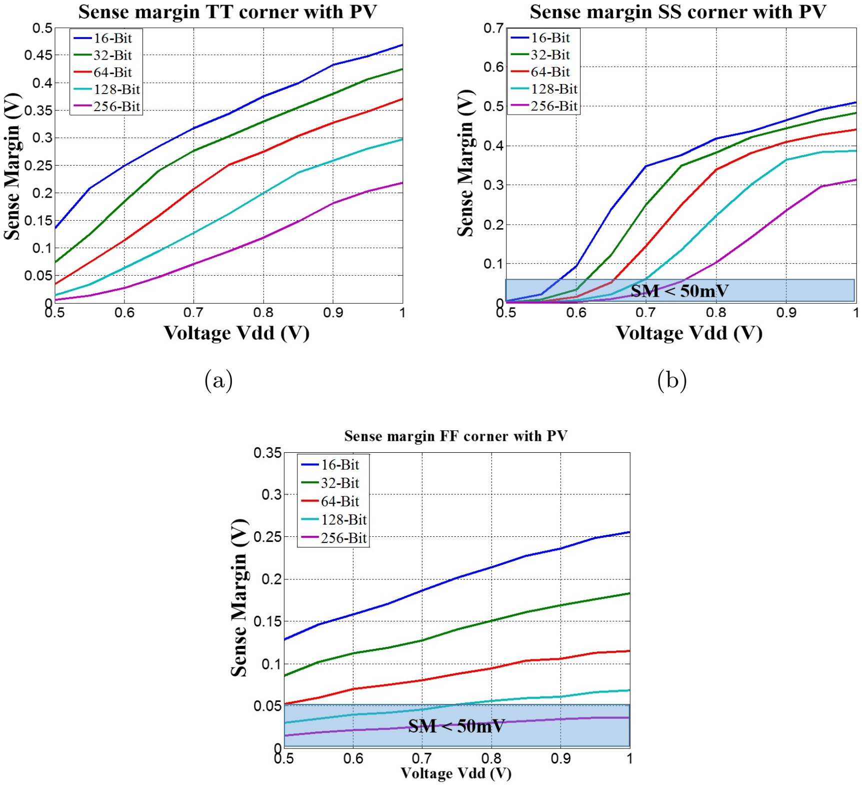
Figure 4.12: Voltage scaling from 0.7V to 1.2V. (a) Sense margin; (b) search delay in logarithmic scale; and, (c) PDP.

For this study, the operating voltage is varied from 0.7V to 1.2V to observe the sensitivity of sense margin, search delay and PDP per bit search (Fig. 4.12). A 50mV sense margin development time is used to measure the search delay. Below 0.7V the sense margin of 256-bit CAM word is less than 50mV. Sense margin and search delay are sensitive to V_{dd} due to lowering of gate voltage of M1/M2 while their threshold voltages remain fixed. At lower voltages, the M1/M2 transistors fail to turn ON or weakly conduct even during mismatch degrading the sense margin (especially for wider words). Search delay for a 256-bit TCAM word varies from 124ps at 1.2V to 2.098ns at of 0.7V (search delay is plotted in log_{10} scale). The increase in the search delay results in sharp increase in the PDP at 0.7V.

4.6.4 Process Variation Analysis

For process variation analysis, we have considered FF, SS and TT corners. We have modeled the process variation in transistors by widely accepted technique of lumping the variation in channel length, oxide thickness, flat band conditions etc. into threshold voltage of the transistor [99]. The SS (FF) is simulated by adding (subtracting) 150mV from nominal threshold voltage. Process variation in the MTJ device is modeled by considering the effects of variation in the MTJ surface area and oxide thickness [98]. We have considered process variability in MTJ by varying the MTJ set resistance R_L as normal distribution with mean of $8k\Omega$ and sigma $\pm 500\Omega$ and TMR variation of 0.1% (variation in surface area and oxide thickness). The worst-case sense margin is plotted for different supply voltages at TT, SS and FF corners (Fig. 4.13). It can be observed that the design can provide a reliable sense margin of above 50mV at all corners till 0.75V for 128-bit words or less. The poor sense margin at lower voltages is linked with poor V_{gs} across M1/M2 that keeps the ML precharged even in mismatch conditions.

The 256-bit word fails to provide adequate sense margin in FF corner at 1V. This is primarily due to poor Δ_2 (as shown in Fig. 4.4) when V_{TH0} moves down coupled with leakage from the match bits. Therefore, match bits leak in case of mismatch degrade the sense margin. We propose threshold voltage modulation and Search Enable (SE) voltage boosting or underdrive to improve sense margin for 256-bit word. Furthermore, these techniques will not worsen the reliability of the NMOS device since thicker oxide (associated with high V_{th}) and lower gate voltage are expected to be better for reliability such as hot-carrier degradation, NBTI and TDDB.



(a)

(b)

(c)

Figure 4.13: Distribution of sense margin in (a) TT; (b) SS; and, (c) FF corners.

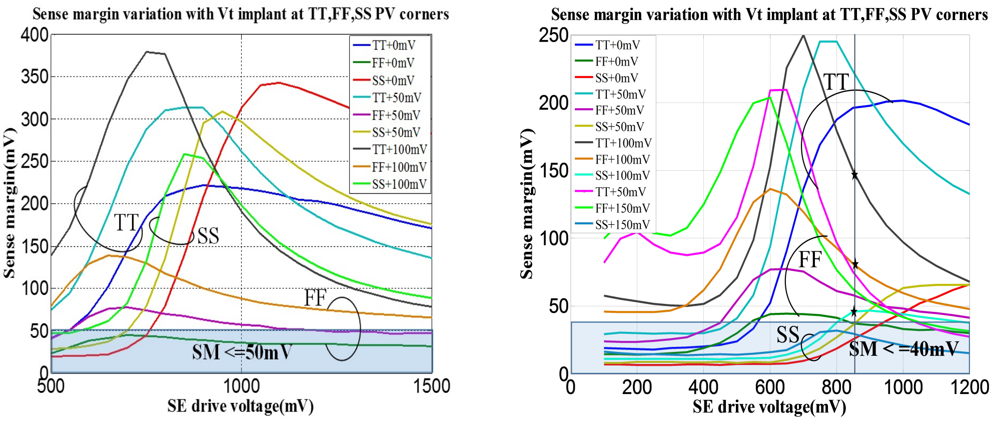
4.7 V_{TH} and SE Modulation for Sense Margin Improvement

In order to solve the poor sense margin, we propose to modulate V_{TH0} , Δ_1 and Δ_2 by exploring threshold voltage modulation of transistor M1/M2 (to tune V_{TH0}) and SE voltage modulation (to tune Δ_1 and Δ_2). Fig. 4.14a shows the results at 1V for the three PV corners for 256-bit word at different SE signal voltages, and, 0mV, 50mV and 100mV higher V_{TH}. Change in the gate drive of M3/M4 changes their ON resistance and results in corresponding change in Δ_1

and Δ_2. It can be noted that optimum choice of SE can improve the sense margin. Moreover, repositioning of V_{TH0} can improve the sense margin further. Fig. 4.14b illustrates the sense margin across three PV corners with V_{TH} implants at 850mV supply voltage. It can be noted that V_{TH} modulation can improve the worst-case sense margin significantly (FF and SS in this case) even though the sense margin in TT corner is degraded. The improvement results from decreased match case current through M1/M2 at SS and the reverse effect in miss case at FF. At the same time, lower SE increases the resistance of M3/M4 which in turn increases Δ_2. As expected, the sense margin in FF with V_{TH} implant is comparable to TT corner without implant. With 100mV V_{TH} implant the design can provide a reliable sense margin of above 40mV in all the PV corners even without SE modulation. A 150mV SE under-drive can improve the sense margin at TT to more than 120mV and a 250mV SE under-drive can improve the sense margin at FF to more than 50mV. So, we employ positive V_{TH} implant of 100mV and gate control signal SE under drive by 150mV to improve performance across all the PV corners. Positive V_{TH} implant is realized by thicker oxide and gate under drive below highest V_g of technology node improves the device reliability while improving the performance.

4.8 NAND Type TCAM Cell Simulation Results

We simulated single bit, 8-bits, 16-bits, 32-bits and 64-bits NAND type TCAM words and the waveform illustrating the match and miss case states of the match line is shown in the Fig. 4.15. It can be seen from the figure that NAND type TCAM can provide up to 500mV of sense margin from 1-bit TCAM simulation. It is also observed that the sense margin decreases as the number of bits in the word increase due to the charge sharing of the match line from intermediate nodes between the bits of a word. We have measured the miss case match line voltage with the miss on



(a) (b)

Figure 4.14: Maximum SM variation with V_{TH} implant. V_{TH} implant of 0mV, 50mV and, 100mV at (a) $V_{dd} = 1V$; and, (b) $V_{dd} = 850mV$.

the farthest bit from the sense amplifiers end to consider the worst-case scenario. The sense margin measured for 16, 32 and 64 bits TCAM words are 147mV, 90.5mV and 33.4mV under nominal conditions. The search delay measured for a minimum SM of 50mV for 16 and 32 bit TCAM words are 1.83ns, 2.72ns respectively. 64-bits word has search delay of 5.45ns for 30mV SM. Search delay is measured as the time required to develop required sense margin on the matchline from the time WL crosses 0.5×V_{dd}. The sense margin can be improved by adding a capacitor to the match line which makes it harder for the match line to get charged by the intermediate node voltages with stray charges in case of a miss. This technique increases the match line power due to increase in match line capacitance.

The search power in a NAND type TCAM cell at 0.8V and 1V supply voltage are tabulated in Table-4.5. The power consumption in NAND TCAM is higher than the proposed NOR TCAM. This is due to additional logic around the NOR type TCAM cell in the design realization. The search delay and sense margin plot for different word length of NAND type TCAM is as shown in the Fig.4.16. It can be concluded from the plot that the search delay increases by two-fold with the

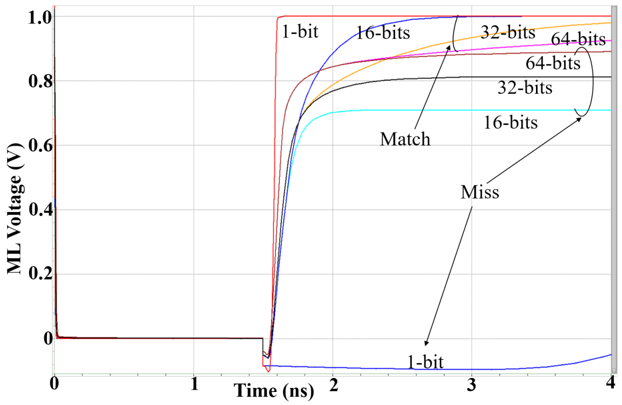


Figure 4.15: Match and miss case match line voltages.

number of bits in the TCAM word (word length). The maximum sense margin decreases greatly for larger word length beyond 64 bits. Maximum sense margin for 64-bit word is 33.4mV with search delay of 5.8ns which is due to larger resistance offered by the PMOS transistor chain in the match line. We have retained the size of PMOS in the match line same for different word lengths for simplicity of analysis and also to alleviate the area overhead in larger words.

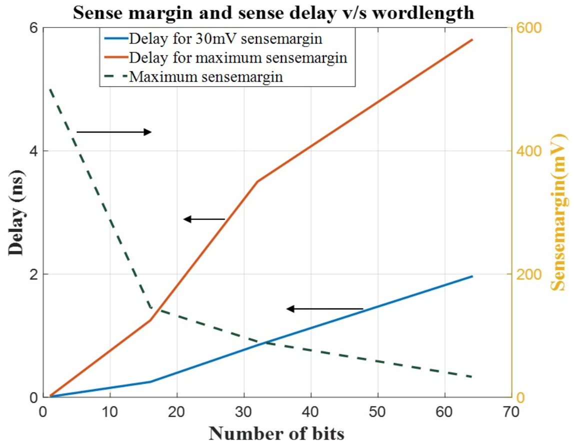


Figure 4.16: Search delay and maximum SM of NAND type TCAM v/s word length.

4.9 Comparative Analysis

In this section, we present the comparative analysis of the proposed NOR and NAND TCAM with respect to CMOS CAM and other spintronic CAMs from literature.

4.9.1 Comparison with CMOS TCAM

Conventional TCAM cell consists of 16 transistors while the proposed NOR type TCAM consists of only 6 NMOS transistors and 2 MTJ bits which 63.5% reduction in the number of transistors. For power comparison, we implemented the CMOS TCAM and simulated using 22nm predictive model. The leakage power of the proposed TCAM is zero as the power supply can be completely shut off during the sleep while SRAM TCAM consumes a considerable amount of standby power. In the mostly OFF applications such as Internet-of-Things and smartphone the proposed TCAM could be very attractive compared to CMOS CAM. The search power consumption of the proposed TCAM is higher compared to conventional CMOS because of the search line current (∼51µA in case of a mismatch at 1V) drawn to generate a secondary voltage at the drain terminals of M3/M4 which enables the discharge transistors of ML. The search line current can be reduced by selecting MTJ with high RL and high TMR. The power consumption during search operation of '1' and '0' bits at 0.8V in STTRAM based TCAM is observed to be up to 8% higher in the worst case (successful search of '1') compared to NOR type CMOS TCAM. The power consumption of NOR type CMOS TCAM and the proposed spintronic TCAM are tabulated in Table-4.6. The NAND type TCAM consumes 2-3% more power and 50% more number of transistors (6T v/s 9T) compared to the proposed NOR type TCAM.

Table 4.5: Power (in µW) comparison of CMOS and proposed TCAMs

	V_{dd}(V)	Match	Miss	Search 'X' from Sl=$\bar{SL}$=0	Search 'X' with D=$\bar{D}$=1
CMOS	0.8	0.3	2.03	1.03	0.2403
Proposed NOR	0.8	24.84	23.8	0.6	22.39
Proposed NOR	1	43.07	53.3	1.02	41.25
Proposed NAND	0.8	24.5	25.08	20.26	29.08
Proposed NAND	1	54.5	43.75	46.4 (SL=SL=1)	63.9

4.9.2 Comparison with Spintronic CAMs

We compared the proposed TCAM cell performance with the other spintronic TCAM structures proposed so far (Table-4.6). The proposed NOR type TCAM draws 51µA (39µA) from the search line during mismatch (match) which is significantly energy-efficient than domain wall memory (DWM) CAM [93]. NOR type TCAM is 33.3% less number of transistors compared to the MTJ TCAM [91] and 50% (25%) less number of transistors than DWM TCAMs [92, 93]. The proposed NAND type TCAM has 12.5% additional transistors compared to the MTJ TCAM [91] and 33.3% (44.4%) less number of transistors than DWM TCAM [92, 93]. The BCAM proposed in [91] requires additional circuitry (NMOS transistor and a MTJ) to configure as a TCAM. In the proposed TCAM data can be written to the bit cell by conventional current induced magnetization technique [64] and controlling the source line. Therefore, it eliminates the need of external writing circuitry. The NMOS transistor M6 (driven by 'WR' signal) provides the additional current required for write. This is unlike in [91] which does not provide methods for memory write. The TCAM cells [91, 94] both MTJs are integrated into search circuit in series which makes the write operation more complex and erroneous. DWM CAMs [92, 93] use domain wall motion based write and MTJ sense circuit based search which adds area overhead and complexity in memory design. MTJ based CAM proposed in [95] also uses 4 transistors (NMOS and PMOS) and 2 MTJs, sensing is based on the amount of current drawn from the match line by different low and high state resistance offered by the MTJ. The technique fails as the number of bits in a word increases (up to 144-bit word). The memory cell has low tolerance to variations in temperature and low V_{TH} process corners due to leakage in the diode connected NMOS transistor. With larger word capacitance of the ML increases while the differential current remains the same and thus affecting the ML sense margin available. The situation becomes worse with process variation of diode connected NMOS transistor. Also, 2T-2MTJ [95], accumulator to store segmented search results and segment activation. The additional circuits incur delay and power overhead in the scheme. Overall the technique is not efficient in terms of delay and power compared to other spintronic CAMs. Though search delay is mentioned in Table-4.6 for different word lengths, the search delay reported for proposed TCAM is for larger word (256- bits). Also, it is only 7.5 times the search delay in a single bit search of [94]. The proposed TCAM search delay differentiates by its smaller value for a larger word. Network IP address is 128 bits in IPv6 protocol [100]. For 128-bits search delay is less than 250ps which can theoretically support 3GHz-4GHz search speed for application in routers.

Table 4.6: Comparison with other spintronic CAMs

Parameter	MTJ-based CAM [91]	DWM CAM [92])	DWM CAM [93]	MTJ-based CAM[94]	Proposed NOR CAM	Proposed NAND CAM
Technology	180nm	90nm	65nm	32nm	22nm	22nm
CAM type	BCAM/TCAM	TCAM	TCAM	BCAM	TCAM	TCAM
CAM topology	NOR	NOR	NOR	NOR and NAND	NOR	NAND
Search energy/bit	NA	2.9fJ	12fJ	2.82fJ	4.7fJ	6.39fJ

Area	6T-2MTJ(BCAM)
	8T-2MTJ(TCAM)
		12T-2MTJ	12T-2MTJ	6T-2MTJ (NOR)
				5T-2MTJ(NAND)
					6T-2MTJ	9T-2MTJ

Search delay	3.3ns (32-bit)	5ns (128-bits)	2ns (8-bit)	34.4ps (1-bit)	263ps (256-bit)	2.72ns (64-bit)
Ext. write circuit	NA	Yes	Yes	No	No	No

We have used 60nmX60nm IMA MTJ model which shows the write latency of 4ns with write energy 0.69pJbit.

4.10 Summary

The chapter proposed a spintronic TCAM which is promising for zero standby leakage and uses less number of transistors. We conducted detailed analysis in the presence of process, voltage and temperature variations for wide range of word sizes. The proposed design operates with reliable sense margin up to 128-bit word size till 0.7V. We also propose threshold voltage modulation and search enable underdrive to improve sense margin for 256-bit word. The proposed TCAM has 62.5% reduced number of transistors compared to conventional CMOS TCAM and 33-50% lesser number of transistors compared to other spintronic CAMs. The worst case active leakage power of the NOR TCAM cell is measured to be 0.38nW. We also propose 9T-2MTJ NAND type TCAM cell which has 43.75% reduction in number of transistors compared to conventional TCAM cell. Proposed NOR TCAM cell has better performance and power metrics compared to NAND TCAM cell. Our study revealed that NOR TCAM using the proposed approach is better than the NAND TCAM in area, delay and power. Therefore, it makes practical sense to employ the NOR TCAM in search applications.

CHAPTER 5 : CSRO BASED TRUE RANDOM NUMBER GENERATOR

5.1 Introduction

Information security is one of the primary concerns with the growth of internet and cloud storage. Data encryption and cryptography are reliable techniques for protecting the data over communication channel (network and storage). Random Number Generator (RNG) is an integral part of cryptography algorithms in encryption engines [101]. Data and system security depends on the randomness of the bit stream generated by RNG [102, 103]. Entropy of the source is instrumental in ensuring the security of the encrypted data. RNGs also find numerous applications other than cryptography such as gaming, gambling, industrial testing and labeling, Monte Carlo simulations, password generation and so on. Software based encryption engines depend on the random number generated by the computer which is only pseudo random due to deterministic algorithms used for generating random number from an initial seed value. Hardware RNG exploit the randomness in physical processes such as, electronic noise, quantum processes, chaotic light emission etc. to generate a continuous stream of random numbers. Although CMOS-based solutions [101, 104, 105] are promising they offer limited security-specific properties such as process variations, noise and chaos. Emerging technologies such as, spintronics [106, 107], memristor [108], and RRAM [109, 110, 111] have demonstrated significant promise because in addition to low-power, high-density

[6] Portions of this chapter were reprinted from R. Govindaraj, S. Ghosh and S. Katkoori, "CSRO-Based Reconfigurable True Random Number Generator Using RRAM," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems.doi: 10.1109/TVLSI.2018.2823274.
Permission is included in Appendix A.

and high speed they also offer new sources of randomness, and easy integration with CMOS [112]. We exploit inherent noise sources of RRAM to design a TRNG.

In this chapter, we explore RRAM technology and features such as, cycle-to-cycle variations and Random Telegraph Noise (RTN) for TRNG design. We make following contributions in this work:

- We propose a high speed (kHz-MHz) Current Starved Ring Oscillator (CSRO) based TRNG using RRAM [54]. We evaluate the proposed TRNG using NIST test suite.
- We propose a methodology to reconfigure the TRNG when entropy reduces over time, and to recover from non-invasive adversary attacks such as exploiting temperature sensitivity of RTN.
- We discuss the security vulnerabilities of RRAM based TRNGs and potential countermeasures in the proposed TRNG.

The remainder of the chapter is organized as follows. Section 5.2 provides the background of TRNG, and RRAM-based TRNG. Section 5.3 and Section 5.4 describe the design and simulation of the proposed TRNG. Section 5.5 discusses potential adversary attacks on RRAM based TRNG and countermeasures. The chapter is summarized in Section 5.6.

5.2 Background and Related Work

In this section we discuss the background of TRNG, and RRAM based designs from the literature in this section.

RNGs are broadly categorized into two basic types based on the quality (in terms of randomness) and the method of bit stream generation, namely, Pseudo RNG (PRNG) and TRNG. In PRNGs, bit stream is not completely random as the algorithm is deterministic except the seed value [[101]. More secure data encryption algorithms require fully random and non-deterministic method of generation. Such streams are generated using TRNGs. Several TRNGs have been proposed in the literature [101, 104, 105, 106, 107, 108] based on randomness in electrical noise, thermal noise, and oscillator based RNGs such as, Free-Running Oscillator, Fibonacci RO (FIRO), Galois RO (GARO) and so on. Noise based RNGs (Fig. 5.1) post-process the noise from the analog source (resistance, voltage source, temperature) to generate random numbers for a digital system. Amplifying tiny noise voltage or converting noise from physical environment to a digital signal often requires multiple stages of processing [101, 104] which depreciates the randomness from the source. Furthermore, the TRNGs which employ the analog parts are weak due to their vulnerability to various adversary attacks.

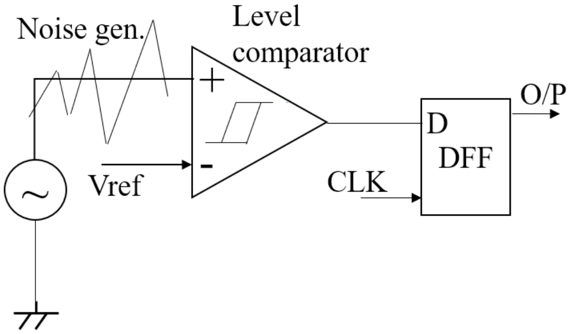


Figure 5.1: Noise based RNG.

Emerging technologies such as spintronics and RRAM [106, 107, 108, 109, 81, 54] which are compatible with CMOS technology [112] and provide rich sources of entropy on-chip are attractive in such scenario. However, the resistance range of spintronic device is limited and the speed of RRAM based TRNGs is as low as few kHz due to their dependency on programming speed of RRAM. A high-speed TRNG is proposed in [110] which employs the RRAM RTN noise. The principle is to utilize the differential change in the bias voltage to modify the sampling frequency. The distinction between [110] and proposed work are as follows:

- RTN of RRAM in the cell of a memory array modulates the bias voltage of a Voltage Controlled Oscillator (VCO). However, the bitline interconnect noise could be large enough to suppress the effect of RTN on voltage differential eventually affecting the available entropy. Furthermore, on-chip noise from pseudorandom source such as power supply, temperature, crosstalk [104] can overpower RTN noise of RRAM which is as small as in the range of nanoampere when bias voltage is generated from a cell in large memory array. The proposed design incorporates a dedicated RRAM in TRNG circuit which preserves the entropy of RRAM RTN.

- Once the adversary can predict stable frequency of the faster clock (by the method of frequency injection when memory and digital supply rails are accessible) [113], random number samples could be predicted for various sampling frequencies under such weak frequency modulation method. In the proposed design power supply of the TRNG can be isolated and placed such that it is not accessible externally. This prevents from the possibility of frequency injection attack.

- TRNG in [110] employs multiplexers to drive each of the inverters in the VCO for frequency trimming which adds considerable area overhead.

- The peak-to-peak (p-p) amplitude of current variations due to RTN is a Figure of Merit (FoM) for a RRAM in storage application. Considering the FoM of RRAM to be used in memory array, it is not feasible to use the RTN of RRAM from a memory cell as source of entropy [3] because of their contrary FoM requirements. RO is placed as boundary circuit in the memory architecture. RTN being a noise voltage of less than 100nA, the method uses a bias current of greater than 50µA to generate bias voltage differential of ∼200mV. The proposed TRNG can operate with bias voltage differential as low as 0.7mV without any current source for biasing. Further, having a µA range of current source in the bias circuit also increases the power dissipated in the bias circuit compared to the proposed TRNG. Reconfiguration of the faster clock oscillator and using a dedicated RRAM cell within bias voltage circuit is essential for a robust design under these circumstances. The proposed method provides two levels of recovery from external adversary attacks by configuring the TRNG through programming RRAM (SET/RESET), and by tuning the sampling frequency to obtain good statistical properties of the generated bit stream.

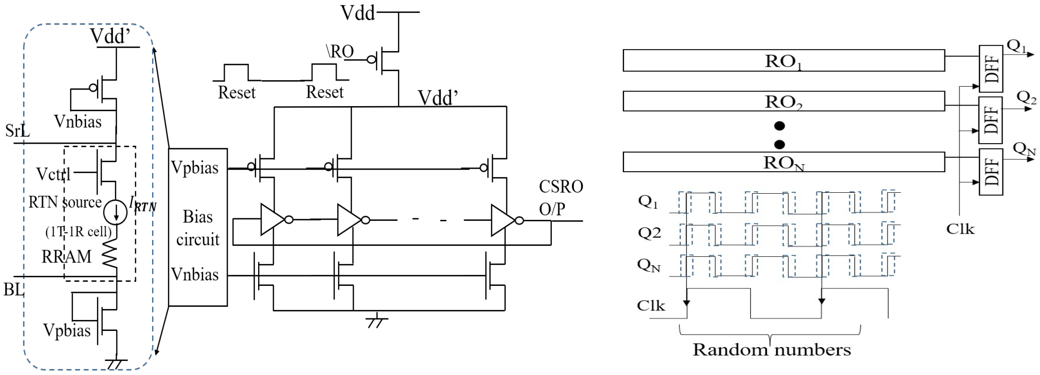
Therefore, the proposed design is more effective in exploiting the entropy of the RRAM device for TRNG application. We discuss various potential adversary attacks on TRNG in Section 5.5.

5.3 Proposed RRAM TRNG

In this section, we describe the proposed TRNG and perform qualitative and quantitative analysis.

5.3.1 Details of the Proposed TRNG

The proposed TRNG based on CSRO is as shown in Fig. 5.2a. Delay of the inverters in CSRO can be controlled to adjust the frequency of oscillations. Principle of delay control is based on current starving of the inverters by controlling the gate voltage of the additional control transistors [114] stacked in NMOS and PMOS network respectively (Fig. 5.2a). Gate voltages of these two series transistors is derived from a bias circuit. In the proposed TRNG, we embed RRAM in the bias circuit to control the gate bias voltages randomly as dictated by the RTN and cycle-to-cycle switching variations of RRAM. The bias circuit is shown in the inset of Fig. 5.2a . It consists of RRAM and access transistor (1T-1R structure) for programming the RRAM as required. NMOS is sized to carry the compliance current of RRAM. V_{ctrl} of the access transistor can be connected to constant voltage greater than threshold voltage of NMOS device during normal TRNG operation. Frequency of programming depends on the quality of bit stream generated with time and RRAM switching speed. In 22nm technology, the width of the diode connected transistors ($W_p = 400nm, W_n = 200nm$) in the bias network are chosen to keep the voltage across the RRAM below 300mV under the highest HRS of the RRAM (3MΩ). The operating voltage of the TRNG is 1V. We assume worst case conditions for voltage drop estimation across RRAM for process variation tolerance. The bias voltages V_{nbias} vary in proportion to current through the RRAM. V_{pbias} varies in complementary slope with respect to RRAM current and V_{nbias} i.e.$V_{pbias} = V_{dd} - (V_{RRAM} + V_{ds}(V_{ctrl}) + V_{nbias})$.



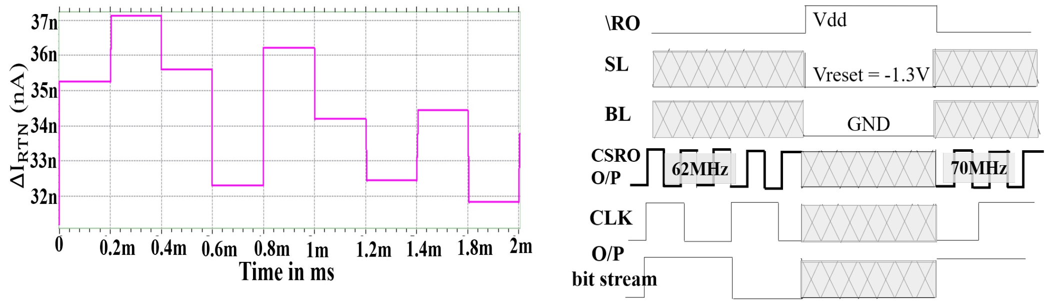
(a) (b)

Figure 5.2: (a) Proposed TRNG based on CSRO; and, (b) Illustration of N-bit TRNG.

Variation of CSRO frequency with the current through RRAM is explained as follows. When the RTN current increases the current though the bias network, V_{ds} of diode connected NMOS increases proportionally increasing V_{pbias} node voltage. At the same time V_{nbias} decreases proportionally. Because of increasing the PMOS gate voltage and decreasing NMOS gate voltage in the inverter stack, delay of the inverters increase. Consequently, the frequency of the oscillator decreases. Thus, current variations in RRAM due to RTN induce respective differential change in V_{pbias} and V_{nbias}. Differential change in bias voltages in turn change the delay of the inverter chain and thus, the frequency of the CSRO. It should be noted that the direction of inverter delay differential depends on the net effect of strength of PMOS and NMOS delay control transistors and bias voltages. In this work, we have used 2:1 ratio for PMOS to NMOS sizing. The speed of inverters varies in the direction of V_{pbias}. These variations are stochastic in nature and, thus data sampled by the sampling clock is random due to stochastic variations in operation of CSRO. The output of multiple ROs are provided to D-Flip Flops which sample the outputs using a sampling clock as shown in Fig. 5.2b.

5.3.2 Sampling Frequency

Sampling frequency determines the rate of generation of random numbers. Minimum sampling frequency is dictated by the frequency of oscillations generated by CSRO. Sampling frequency must be selected at least half of that of CSRO oscillations to avoid the duplication of the bits in the random bit stream. Theoretically, sampling frequency upto several MHz can be selected for CSRO oscillations greater than 10MHz. We have selected 6MHz of sampling frequency for the CSRO oscillations in ∼60MHz-70MHz range. Sampling frequency can be selected during the time of design by estimating the frequency of CSRO from the initial delay of the inverters. Sampling frequency can also be selected dynamically to improve the statistical properties of the bit stream [115]. A technique based on Built In Self Test (BIST) is proposed to measure the statistical properties of RO based TRNGs in [115]. However, it requires on chip clock generator with dynamically adjustable frequency and BIST with logic for testing statistical properties which adds to the design complexity and additional cost. Frequency of random bit stream is theoretically limited by the number of inverter stages in the CSRO and frequency of the various sources of entropy in the TRNG. Circuit and device noise depends on bias voltage, temperature, junction capacitance of MOS devices and scales proportionally with the number of stages in a single ended CSRO [116]. To achieve synergistic effect of circuit noise and RTN (Fig. 5.3a) and high speed generation of random numbers we limit the sampling frequency in the range of MHz.



(a)

(b)

Figure 5.3: RTN current ΔI and reconfiguration of TRNG.(a)ΔI through RRAM due to RTN with 5kHz of frequency; and, (b) reconfiguration of TRNG.

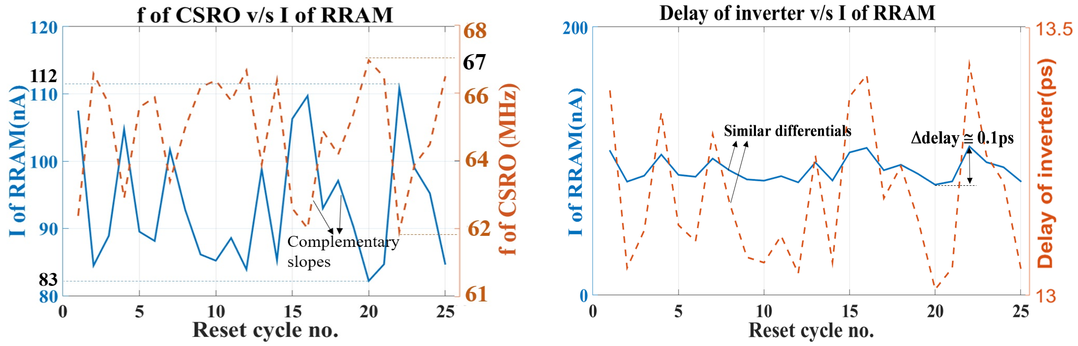
5.3.3 Configurability

Frequency of CSRO can be dynamically configured by altering the parameters in the bias circuit, which varies the current starved by the delay inverters. For this purpose, we embed 1T-1R cell in the bias circuit. Due to exponential dependency of HRS current on the barrier length, a small change in the barrier length manifests as significant change in the resistance unlike in LRS where current is linearly dependent on the barrier length. HRS exhibits the higher cycle to cycle variability and RTN compared to LRS [76, 83] .Therefore, in the design we RESET the RRAM for reconfiguration. It should be noted that the cycle-to-cycle switching parameter variations and RTN are uncorrelated but concurrent in nature [76].

Table 5.1: Comparative analysis of TRNG

	Methodology	Source of entropy	Speed	Advantages	Drawbacks
Spin Dice [106] Perturb and tracking [107]	RESET and probabilistic switching voltage for programming. And [107] eliminates reset every cycle conditionally from the previous o/p sample. 3 phases: RESET, perturb and read	Probabilistic switching of MTJ	MHz	Ultra-low voltage switching operation of MTJ.	Speed is limited by RESET [103], probabilistic switching time; Delay, area and power overhead of tracking system [105].
Balatti et al.[109]	Stochastic SET process by a random pulse with median of SET voltage distribution	Stochastic switching process	kHz-MHz	Broader Resistance distribution compared to MTJ.	Accurate switching voltage control, and Slow switching limits the speed.
Balatti et al. [81]	Probabilistic switching of a pair of RRAM (series/parallel). 3 phases: set, reset and read.	RRAM probabilistic switching	~0.16kHz	No biasing of random bit	Slow switching speed. Requires analog parts: Comparator
Yang et al.[110]	RRAM cell in bias circuit of sampling frequency oscillator.	RTN of RRAM	MHz-GHz	Sample frequency generator based TRNG. Biasing circuit uses a current source of >50µA. Requires ~200mV of bias voltage differential.	FoM of RRAM requirements are in contrary for storage and TRNG applications. Interface noise and routing congestions would mitigate the effect of RRAM RTN.
Proposed TRNG	RRAM current makes the current through the bias network of CSRO which in turn generates phase noise and jitter in the oscillations.	RTN, electronic noise in RO and cycle-cycle reset switching variations	MHz-GHz	High speed not limited by RRAM switching speed.Simple CSRO based design. No analog parts	Requires power gating; and larger area compared other RRAM based TRNGs.

For programming the RRAM (Fig. 5.3b), we halt the operation of CSRO by power gating PMOS transistor connected to power supply. Power gating transistor is driven by a pulse with pulse of width equal to the write time. The NMOS transistor controlled by V_{ctrl} in the regular operation is connected to V_{dd} or a constant voltage. RRAM is RESET by applying V_{reset} ramp voltage of -1.3V across the electrodes from SL and BL signals. RRAM demonstrates switching time of ~10ns which adds penalty of one cycle with CSRO frequency of upto 100MHz in the worst possible scenario of write conditions. The primary advantage of the proposed TRNG over other RRAM based TRNGs is high speed generation [109, 81] of random bit stream and the frequency of TRNG is independent of the write time of RRAM [81].Table-5.1 presents the comparative analysis of the proposed TRNG with other spintronic and RRAM TRNGs. By choosing a reset voltage at probabilistic switching voltage and using probability switching model of RRAM, the entropy can be further improved. The reconfiguration feature can also be exploited to recover from adversary attacks by generating new random numbers. However, this requires additional circuitry to detect the adversary attacks and



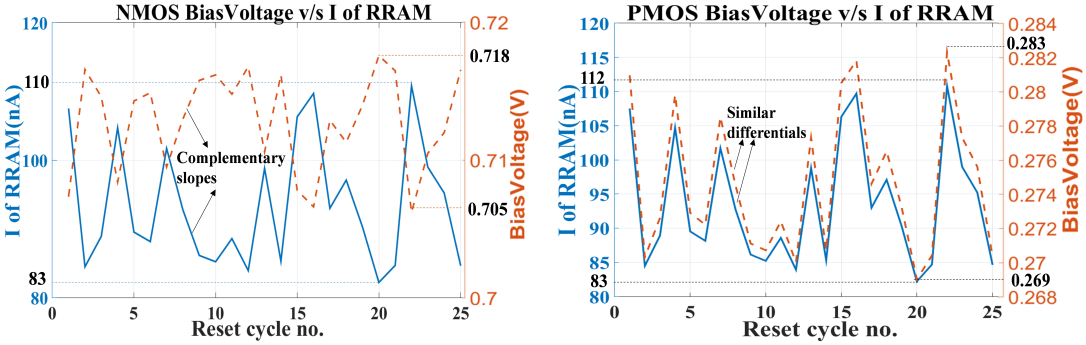
(a)

(b)

Figure 5.4: Variations CSRO parameters (configurability) with RRAM current. (a) Frequency; and, (b) inverter delay over 25 reset cycles.

activate the write operation of RRAM. The TRNG is reconfigured in regular intervals under default conditions without any assistance to detect adversary attacks for simplicity of the solution.

By applying the probabilistic switching voltage instead of RESET voltage -1.3V the RRAM undergoes probabilistic switching. The RRAM remains RESET or changes to SET state by the applied switching voltage [77, 80]. This kind of switching could improve the randomness in the oscillator frequency further after configuration. This is out of the scope of this work and, will be explored in our future work.



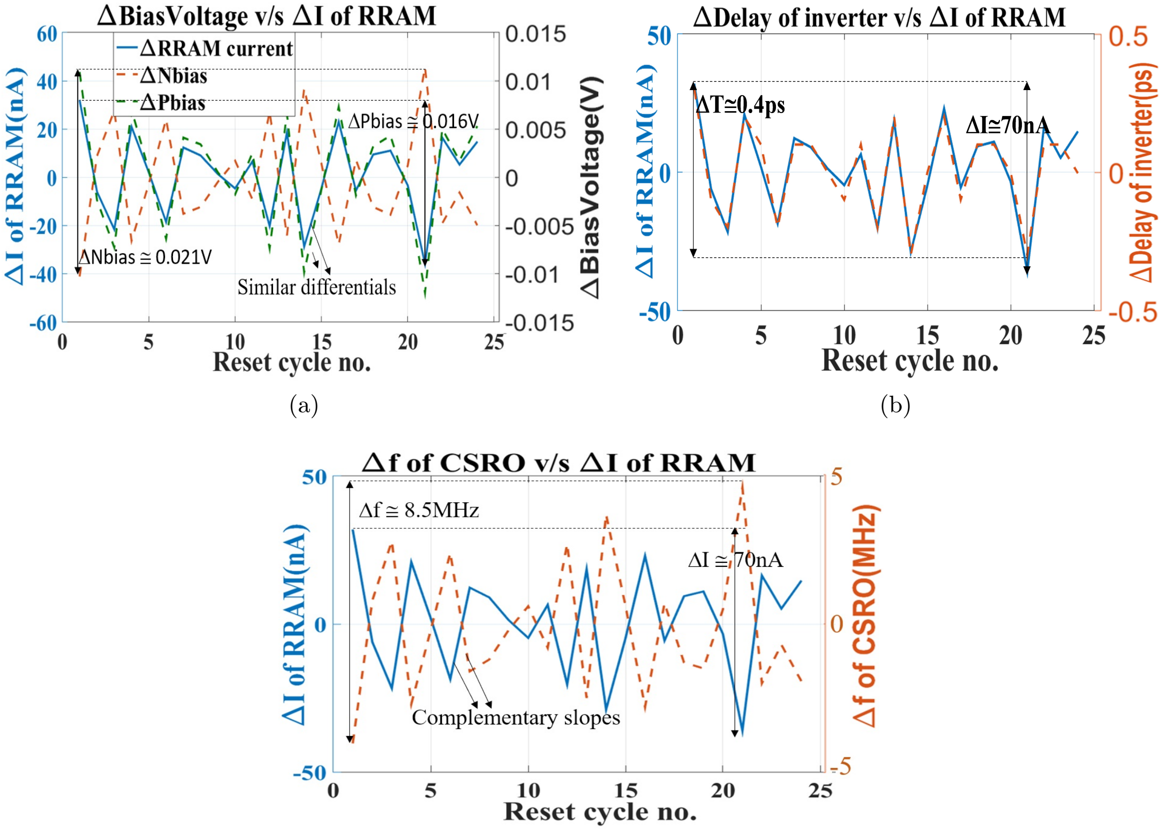
(a) scale=0.4

(b) scale=0.4

Figure 5.5: Variations of voltage across NMOS and PMOS with RRAM current.

Frequency of programming pulse is at least few 1000 times slower than frequency of CSRO oscillations. Typically, TRNG is configured after generating a few sets of random numbers. Within a single configuration cycle circuit noise and RTN acts as source of randomness to generate jitter in oscillations. Between different configuration cycles RRAM switching parameters' variation and respective RTN synergistically contribute to entropy in the TRNG system.

5.4 Simulation Results



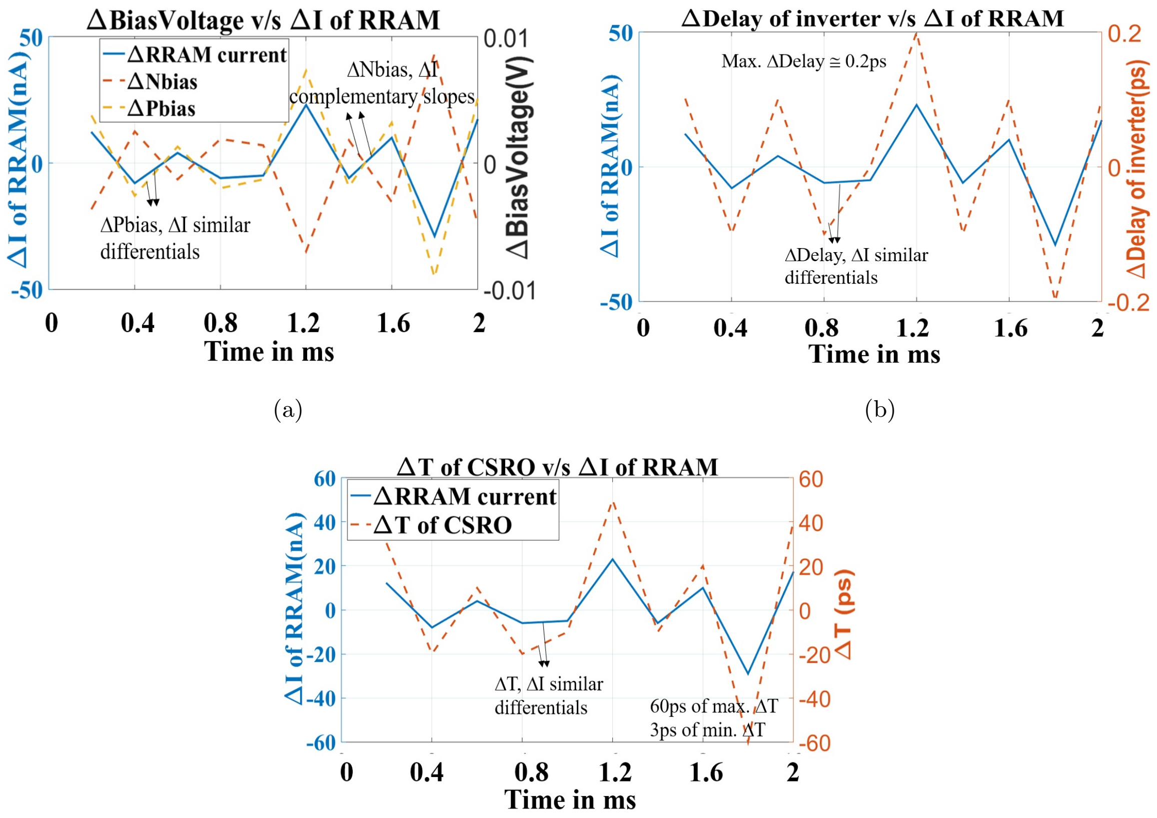
(a)

(b)

(c)

Figure 5.6: Differential changes due to change in RRAM current with configuration. (a) bias voltages; (b) delay of inverter; (c) frequency of oscillations.

We present the simulation results of proposed TRNG using 22nm PTM models of MOS transistors and Verilog-A model of RRAM (Chapter 3). Fig. 5.4a shows the frequency of a CSRO and RRAM current at different RESET cycles reconfiguring the resistance of RRAM. Frequency of CSRO changes in the range of 62MHz-67MHz over 25 cycles of configuration in unpredictable random steps. Also, frequency of CSRO varies in complementary slope (± and vice versa) with respect to current through RRAM. NMOS and PMOS bias voltages (Fig. 5.5a, and Fig. 5.5b) undergo differential change in each of the configuration. The current through the RRAM in different RESET cycles which varies from ∼83nA to ∼112nA in random steps which induces respective differential change in the delay of inverter. Delay of the inverter varies in tandem with the current though the RRAM/bias network (Fig. 5.4b).



(a)

(b)

(c)

Figure 5.7: Differential changes due to RTN at 30^oC. RTN though RRAM resulting differential change in (a) bias voltages; (b) delay of inverters; and, (c) jitter (ΔT) in oscillations.

It can also be observed that the PMOS and NMOS bias voltage differentials vary in complementary slopes (± and vice-versa) with each other. PMOS bias voltage varies proportional to current through the RRAM (Fig. 5.6a). The bias voltages demonstrate a differential change of few mV (21mV and 16mV) which induce a proportional change in the delay of the inverters and frequency of CSRO. Fig. 5.6b illustrate the cycle to cycle differential change in the delay of inverters, and it varies in the direction of V_{pbias}. Fig. 5.6c shows the frequency of CSRO varying with current through the RRAM. As the current through RRAM decreases frequency of CSRO increases and vice-versa. RTN in the RRAM induces jitter in the oscillations which leads to randomness in the bit stream sampled from the CSRO oscillations. To illustrate the effect of RTN on jitter in CSRO oscillations, we have plotted the differential in current through the RRAM and bias voltages with time (Fig. 5.7a) time period of the oscillations with time. Also, differential change in delay of the inverter is plotted due to respective change in the current through RRAM (Fig. 5.7b). Delay of inverter changes in the range of ±10fs to ±200fs exhibiting a maximum differential change of 200fs. Jitter in the range as low as 3ps to as high as 60ps is observed (Fig. 5.7c). This additional jitter due to RTN in the RRAM acts as source of randomness to produce the random bit stream when

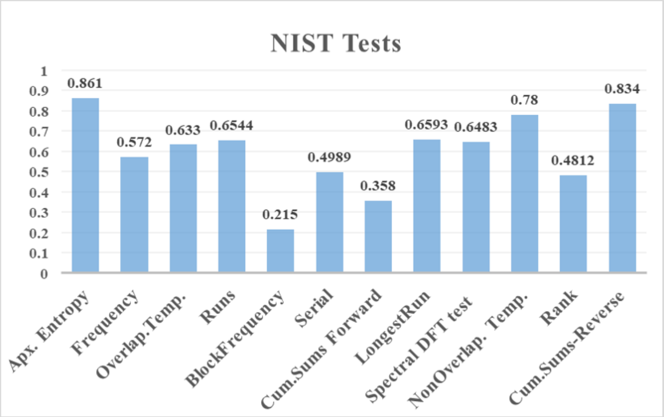
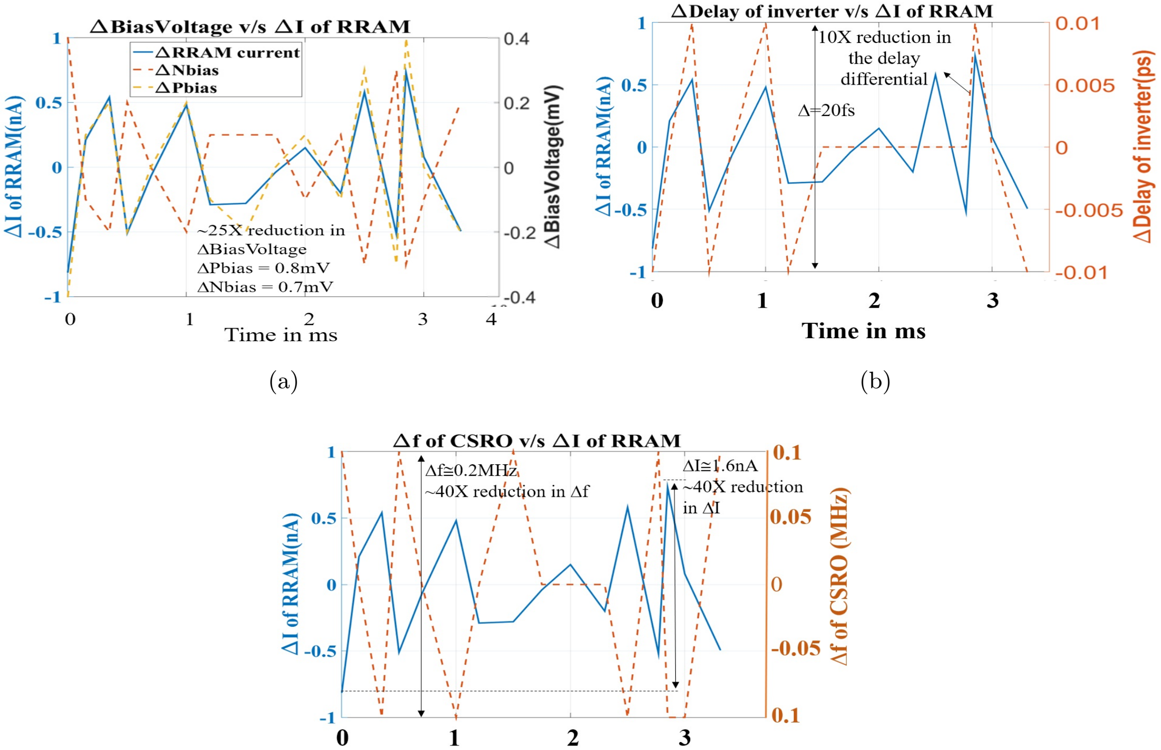


Figure 5.8: NIST test results on bit stream from 4-bit TRNG.

CSRO oscillation is sampled by a clock of stable frequency. We present NIST test results of a 4-bit TRNG with 2500 random data samples (10000 bits in a stream) to validate the randomness of the data generated. From the Fig.5.8 it can be noted that the p-value in the NIST tests is greater than 0.01 which indicates sufficient randomness in the generated bit stream.

5.5 Adversary Attacks on TRNGs

In this section, we discuss the adversary attacks on the RO based TRNGs and RRAM based TRNGs. We also discuss the robustness of the proposed design against these attacks.



(a)

(b)

(c)

Figure 5.9: Differential change due to RTN at 5^{o}C. RTN though RRAM resulting differential change in: (a) bias voltages;(b) delay of inverters; and, (c) frequency (Δf) in oscillations.

5.5.1 Background of Attacks and Prevention

Several adversary attacks like frequency injection attack, attack over the network [117], electromagnetic waves emission based [118, 119], fault attacks [113, 118, 119, 120, 121] have been investigated in the literature. Researchers have also proposed techniques such as error correction, induction of non-linearity in the response [118], attacks detection from frequency, bit stream monitoring and recovery using RC filters closer to the power supply [120, 121] to safeguard against these adversary attacks. Attacks such as frequency injection attacks from the power rails could be avoided by keeping the power rails not accessible to the adversary externally. This can be achieved by deriving the voltage from a dedicated on-chip power supply [122]. In this chapter, we focus our discussion on potential attacks on RRAM based TRNGs.

5.5.2 Vulnerabilities of RRAM-based Design

RRAM based TRNGs are vulnerable to the adversary attacks due to the sensitivity of RRAM characteristics to temperature and voltage. RTN of RRAM is associated with the charge and discharge time of traps in the CF. The frequency of charge and discharge is dependent on the joule heating of the CF and the ambient temperature. Fig. 5.9 illustrates the simulation results for the effect of RTN at 5^oC. At cooler temperature, the charge and discharge time of the electrons in the traps are longer which reduces the rate of change of RRAM current [83]. Hence, change in the RRAM current varies at the rate of few Hz (1-4 times/second). Very few traps available are responsible for RTN decreasing the variation range of RRAM current in few nA (ΔI of RRAM is 0.5nA-1nA) as shown in Fig. 5.9. Differential change in the bias voltages, delay and frequency is reduced by ∼25X, ∼10X and ∼40X respectively compared to the variations at 30^oC. There are flat regions in the frequency plot where the differential change in the frequency is almost zero due to degradation in the entropy available at cooler temperature. Attack model is discussed in the next subsection.

5.5.3 Attack Model: Colling and Model Building

Entropy decreases (due to RTN) considerably at low temperature [83] which affects the quality of random numbers generated. This makes the underlying cryptographic system vulnerable to adversary attacks [123, 102, 103]. Adversary can cool the chip by nitrous oxide and control the temperature of the chip which would eventually affect the entropy of the RRAM based TRNGs. Although the vulnerability to Machine Learning (ML) based model building attacks on TRNGs is still unproven, TRNGs could be vulnerable to model based attacks similar to Physically Unclonable Functions (PUF) [124].

5.5.4 Countermeasures: Temperature Sensing and Configurability

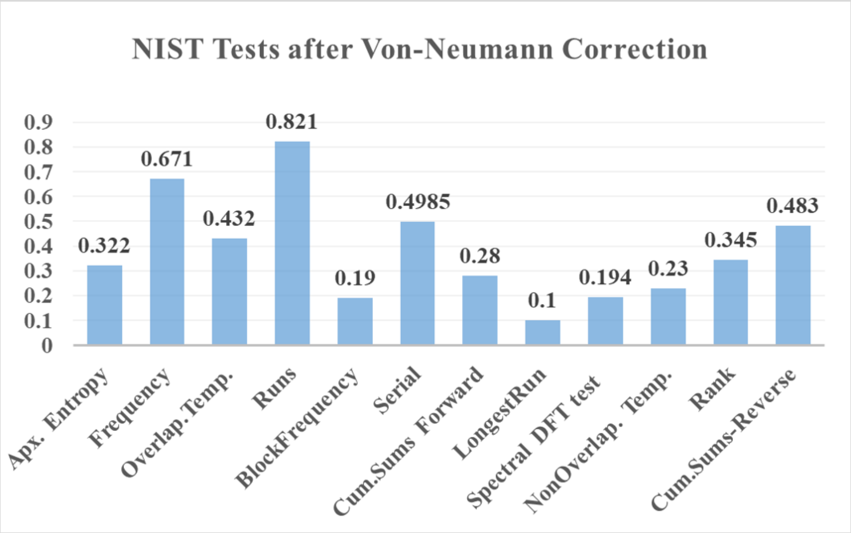


Figure 5.10: NIST tests on 4-bit TRNG within RESET cycle with zero RTN. Von Neumann correction is applied to TRNG stream.

By using an on-chip temperature sensor to sense the ambient temperature and configuring the TRNG at an adaptive frequency depending on the temperature could safeguard against temperature-based attacks. Diode temperature sensor proposed in [125] can be employed for on chip temperature sensing. It should also be noted that RRAM demonstrates lower durability of few million cycles compared to other non-volatile memories [81] which affects the productive lifetime of TRNG in the security chip. Reconfiguration at regular intervals after generating a few random bit streams makes it almost impossible for the adversary to predict the new CSRO frequency. This can be used to safeguard against the model building attacks. Further, RRAM switching speed adds to speed overhead in such scenario. In the applications where the speed of TRNG renders it useless, Von Neumann correction technique [126] is employed within RESET cycle to compensate for the reduction in the entropy with configuration frequency of kHz. We applied Von Neuman correction on the bit stream generated after setting the effect of RTN to zero (no RTN current source) in current fluctuations of RRAM on 4-bit TRNG of 2500 length (Fig. 5.10).

5.6 Summary

The chapter proposed a high speed (kHz-MHz), reconfigurable CSRO based TRNG for on chip applications. It exploits the RTN low frequency noise in RRAM and cycle to cycle switching parameter variations as the source of entropy. We propose a technique to reconfigure the system to recover against adversary attacks. Configurability makes the model building and ML attacks harder. The 4-bit random data stream is validated successfully for sufficient randomness using NIST test suite. The speed of the designed TRNG is 6MHz. Energy/bit for random bit generation is 22.8fJ.

CHAPTER 6 : RRAM ARBITER PUF WITHIN 1T-1R MEMORY ARCHITECTURE

6.1 Introduction

PUF is one of the widely accepted hardware security primitives that finds application in authentication as well as random number generation. It generates a secured key by the physical nature of an electronic system. Physical structure of every electronic system is unique due to inherent differences during manufacturing by the same process technology [127]. Several PUFs based on CMOS [128, 127, 129], memristor [130] and spintronic technologies [131, 50, 51] have been proposed in the literature. The CMOS PUFs include SRAM based memory PUF, arbiter PUF and ring oscillator based PUFs [128]. Resistive Random Access Memory (RRAM) is another non-volatile memory technology [132, 80] based on binary metal electrodes that have been explored for memory and cross-bar array based PUFs [80, 133, 134, 135, 136] . The responses are generated by comparing the resistance of memory bits from two symmetric columns of an array. The existing RRAM PUFs are weak due to linear number of CRPs. Extension to strong PUFs (such as arbiter PUF) while staying within the array structure is a non-trivial problem due to requirements such as arbiter circuits and multiplexers.

[5] Portions of this chapter were reprinted from Govindaraj, R., & Ghosh, S. (2016, October). A strong arbiter PUF using resistive RAM within 1T-1R memory architecture. In Computer Design (ICCD), 2016 IEEE 34th International Conference on (pp. 141-148). IEEE.
Permission is included in Appendix A.

In this chapter, we propose design and application of RRAM based arbiter PUF for the first time. We make following contributions in our work:

- We propose a strong arbiter PUF architecture with low area overhead.
- We employ sense amplifier in the RRAM subarray as arbiter in the proposed architecture.
- We implement the proposed arbiter PUF with minimally invasive changes in the pre-existing RRAM memory subarray which can potentially reduce the design time significantly.
- We evaluate the proposed Arbiter PUF (APUF) characteristics in terms of inter- and intra-HD for various number of stages.
- We propose an APUF architecture resilient to Machine Learning (ML) based model building attacks with few additional modifications.
- We also investigate the application of proposed PUF architecture in data attestation and signature in IoT devices.

The remainder of this chapter is organized as follows. Section 6.2 provides the background of PUF, and RRAM-based PUF. Section 6.3 describes the proposed PUF architecture and Section 6.4 presents simulation results. Section 6.5 discusses the potential adversary attacks on proposed APUF with results of ML based model building and side channel attack based on power information. In Section 6.6, the proposed APUF architecture is leveraged for ML attack resilient design. Section 6.7 presents the application of APUF for data attestation in IoTs when integrated with a Logic in Memory (LiM) encryption platform. Finally, Section 6.8 summarizes the chapter.

6.2 Background

6.2.1 Background of PUFs

PUFs are promising security primitives that find applications in authentication and key generation for secure operation. PUFs can be categorized based on the circuit topologies and the characteristics such as CRPs. Based on circuit topology, PUFs can be categorized as memory PUFs and delay based PUFs [128, 127, 129] whereas based on the characteristic property PUFs can be categorized as weak and strong PUFs.

Memory PUFs exploit the random initialization of the SRAM bits in an array due to process variations in CMOS memory cell. The address bits are used as challenge and the bits read from the SRAM array with the address is the response of the PUF. The number of CRPs are limited by the number of address bits of the memory. APUF (Fig. 6.1) comprises of two symmetric electrical paths and the delay difference between the paths is digitized by an arbiter. The arbiter output is the response of an APUF. The symmetric paths consist of gates and multiplexer circuits. Multiplexer select lines are used as challenge bits. At every stage multiplexer select lines connect the two incoming paths to one of the outputs. The randomness in the path delay due to process variation makes the delay parameters unclonable which makes an APUF stronger. Further, with exponential increase in the number of CRPs and multiple delay paths while employing XOR gate as arbiter makes APUF more secure and resilient against potential adversary attacks such model building.

6.2.2 Background of RRAM PUFs

The resistance of the RRAM after SET and RESET follow probability distribution due to defects in the CF and the thermal voltage fluctuations. The variability in the cycle-to-cycle

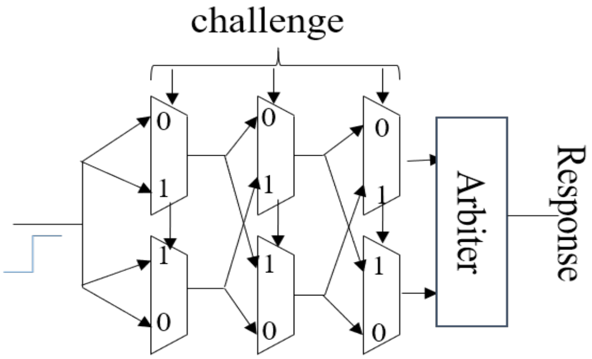


Figure 6.1: CMOS Arbiter PUF.

resistance switching which is a source of randomness can be exploited for security applications. RRAM PUFs are proposed based on 1T-1R bit cell and crossbar array. In the 1T-1R architecture, the input to the row decoder and column multiplexer is used as a challenge to select and read two cells of a row randomly. The response is generated by a current sense amplifier by comparing the resistance of two cells. This forms a weaker PUF for array size of $N \times N$, with $N \times N \times log_2 N$ number of CRPs [135]. Cross bar array PUFs become unreliable as the technology node shifts below 50nm due to interconnect resistance of crossbar array.

6.3 Proposed PUF Architecture

In this section, we propose an APUF using regular 1T-1R RRAM array architecture. This is a stronger PUF with exponential number of CRPs. The motivation for the proposed architecture is to achieve desirable PUF characteristics for hardware security while improving the area efficiency of APUF staying with 1T-1R memory architecture. We provide the details of the PUF architecture and performance analysis as well for various number of stages and reconfigurability.

6.3.1 Summary of Proposed PUF Architecture

Fig.6.2 shows the proposed the architecture of 1T-1R bit cell based APUF. The design is obtained by making minimally invasive changes to the existing RRAM memory array. The column circuitry is modified by including two 2:1 multiplexers (inset in Fig. 6.2). The MUXes connect two selected bit cells from a global column (GC2) to the other bit cells selected from another global column (GC1). Two bit cells selected in a GC are from two different sectors by Word Line (WL), and column select (Y_sel) signal selects one of the local columns in each of the sectors. Source Lines (SL) from GC2 and Bit Lines (BL) from GC1 of the selected bit cells are selectively connected through the MUXs. Select signal to the set of MUXs in each of the GC forms challenge bit in association with the address bits (i.e., sector select and Y_sel) used for bit addressing in every global column. The scheme effectively pairs up any two bit cells of GC1 with other bit cells of GC2 and the connecting path is controlled by the challenge bits (which could be address bits and MUX selects). This type of architecture provides an exponential increase in the number of CRPs with size of memory array.

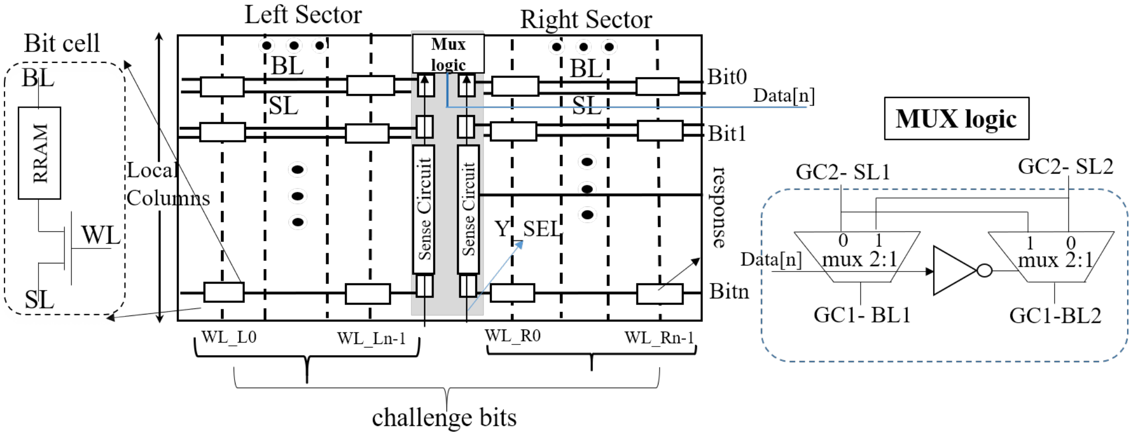


Figure 6.2: High level architecture of proposed APUF.

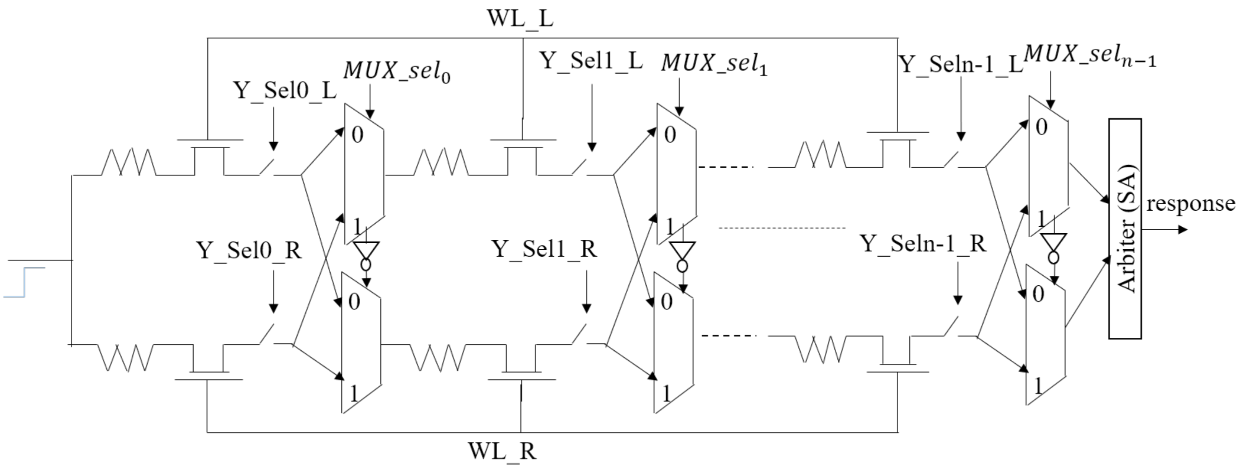


Figure 6.3: Conceptual schematic of the proposed APUF.

We repurpose sense amplifier as arbiter in the architecture. Two symmetric paths of APUF are connected as two inputs of a differential sense amplifier. This is evident from conceptual schematic of the PUF shown in Fig. 6.3. Instead of measuring delay between signals racing in two paths to determine the response we measure the voltage difference at a given sense time. This enables the implementation of the PUF architecture from a conventional RRAM memory array minimally invasive. Multiplexers of the arbiter path are placed in the column area of the memory subarray (Figs. 6.2 and 6.3). The variability in sense amplifier adds another source of entropy in the PUF response.

6.3.2 Implementation of APUF

Fig. 6.2 illustrates the implementation of APUF and Fig. 6.3 depicts the conceptual view of RRAM APUF. RRAM bit cells are connected in daisy chain fashion selectively between two symmetric paths. Inset in the right-hand side of Fig. 6.2 shows the MUX circuit placed in the column area of PUF architecture. The state of select signals controls the selected bit cell connection between the two symmetric paths. An arbiter is used to produce final response depending on the delay difference between the signals arriving from these two paths. Process variations in the access transistor, MUX circuit and RRAM bit cell are the sources of entropy in such an APUF. We use the access transistors of minimum size. MUX circuit transistors are chosen for optimal speed of the delay paths and smaller ON resistance. Smaller ON resistance ensures that process variation in RRAM RESET resistance is not suppressed in path delay contribution.

6.3.3 Implementation of Challenges

Fig. 6.2 demonstrates the APUF implementation along with the challenge bits. Bit cell addressing from the row decoder (WL) and column decoder (Y_sel) along with MUX select lines are used as challenge bits. Such an implementation offers an exponential growth in the number of CRPs with the size of memory array and arbiter stages. Number of arbiter stages should be optimized for larger number of CRPs.

6.3.4 Implementation of Arbiter

We repurpose sense amplifier present in a conventional memory architecture as arbiter. The implementation is minimally invasive in the memory architecture. Sense amplifier is designed to measure the available sense margin between the two symmetric paths. The method is based on the fact that the delay difference between the signals on two symmetric paths is proportional to the voltage difference at any instant of time before settling. Multiple sense amplifiers are connected in parallel to suppress the effect of noise on sense margin.

6.3.5 Number of CRPs

The number of MUXes in the paths of APUF varies proportionally on the number of global columns in the memory subarray. For N number of global columns, 'M' number of local columns for a memory of word length 'W', the number of CRPs are $(2^{N}) \times (2^{M}) \times (2^{W})$. The combination synthesizes a delay path with 'N' 1T-1R bit cells and MUXes in two symmetric delay paths. The area overhead of the proposed PUF is very minimal due the addition of only two MUXes and a single select line in each of the GCs. APUF architecture with large number of CRPs yields a stronger PUF compared to RRAM memory PUFs proposed in the literature. By choosing 'X' number of GCs out of 'N' number of GCs in each arbiter path one can have '$\frac{N}{X}$' number of symmetric paths. The number of CRPs in such a case would be $2^{\frac{N}{X}} \times 2^{X} \times 2^{M} \times 2^{W}$ for a memory of word length 'W' with 'M' number of local columns. The number of CRPs grow exponentially with the size of memory subarray and number of bit cells in an arbiter path.

6.3.6 Configurability

The number of multiplexer stages in each path of APUF can be dynamically configured. This is accomplished by using DEMUX at every stage of MUX. 2:1 DEMUX is used to selects one of the two paths, either to generate final response with sense amplifier or direct it next stage of APUF. Fig. 6.4 shows the DeMUX stage integrated with the multiplexer stages in the architecture to facilitate configurability. DeMUX_sel signal is used for configuration at each stage. When DeMUX_sel is '0' the path breaks bypassing the further stages. A1 and A2 are connected to the arbiter. When DeMUX_sel is '1' outputs from the MUX is connected to on-coming path. It should be noted that A1, A2 inputs of the arbiter should be multiplexed at the input of single arbiter or exclusive arbiters should be employed after each of DeMUX stage. However, sense amplifiers in each of the GCs are utilized to generate response by connecting to DeMUX in respective GC. Therefore, arbiter is readily available at every stage of MUX and DeMUX. Further, configurability also increases the number of CRPs as discussed in the next subsection.

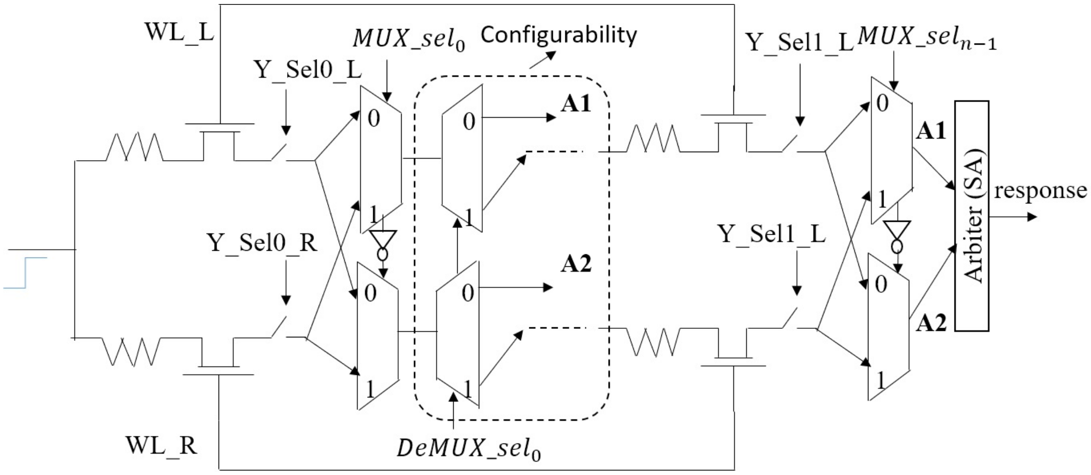


Figure 6.4: Configurability for the number of stages in proposed APUF.

6.3.7 CRPs with Configurability

Configurability in the proposed architecture offers various advantages along with the ability to dynamically select the number of APUF stages. To completely exploit the rich features in the proposed architecture, DeMUXs can be utilized to increase the number of CRPs exponentially yielding a much stronger APUF. For instance, 'N' bit arbiter with DeMUX in each of the GCs, CRPs available from the configuration for 'N-1', 'N-2'..... 5,4.. so on could be combined together to obtain $2^N + 2^{N-1} + 2^{N-2} + 2^4$ number of CRPs. The total number of bits in a challenge in such case is '$M_N + D_N$' where M_N is the number of multiplexers in the arbiter and D_N is the number of DeMUXs. D_N bits remain constant while M_N bits follow regular binary number pattern for a fixed configuration. For example, in a 16-stage APUF for up to 8-stage configuration total number of CRPs is '$2^8+2^9+2^{10}+.....2^{16}$'. D_N bits in the first configuration are '1111111100000000' likewise in the last configuration for all 16-stages in the APUF D_N bits are '1111111111111111'. By approximating summation in the equation, we get $2^{GC} \times 2^{LC} \times \sum_{K=4}^{N} 2^K$. Number of CRPs is twice compared to $2^{GC} \times 2^{LC} \times 2^N$ with non-configurable PUF stages. Total number of CRPs with minimum of 4 stages in APUF configuration with 8 GCs and 8 LCs in the memory array are as tabulated in Table 6.1.

Table 6.1: Number of CRPs with and without configurability.

Number of PUF Stages 'N'	No. of CRPs with Configurable Stages ($2^{GC} \times 2^{LC} \times \sum_{K=4}^{N} 2^K$)	No. of CRPs with Non-configurable PUF Stages ($2^{GC} \times 2^{LC} \times 2^N$)
8	32505856	16777216
12	535822336	268435456
16	8588886016	4.295×10^9
20	1.37438×10^{11}	6.872×10^{10}
32	5.6295×10^{14}	2.815×10^{14}
40	1.44115×10^{17}	7.206×10^{16}
52	5.90296×10^{20}	2.951×10^{20}
64	2.41785×10^{24}	1.209×10^{24}

6.4 Simulation Results

We present simulation results of the proposed APUF with 8 GCs. We have used 65nm predictive technology models [96] and Verilog-A model of RRAM for simulations. All bit cells are initially set by forming process. Then a RESET voltage of -1.3V is applied at the BL of RRAM while the SLs are connected to ground. HRS and LRS are modeled as explained in Chapter 3. The LRS is assumed to have gaussian distribution with mean 3.65kΩ and variance of 0.034 [75]. C_{xv} is a model parameter and is assumed to have gaussian distribution with variance of 0.08 in compliance with experimental results [75] of hafnium oxide based RRAM. We evaluate PUF for three metrics namely, uniqueness, reliability, and uniformity. It should be noted that the responses are not uniformly distributed unlike in a RRAM based memory PUFs proposed so far [136]. This is because the delay is used as response in a path composed of multiple 1T-1R bit cells and MUXes. The address bits used to select the bit cells in each of the global columns and subarrays is used as challenge along with the multiplexer select line inputs.

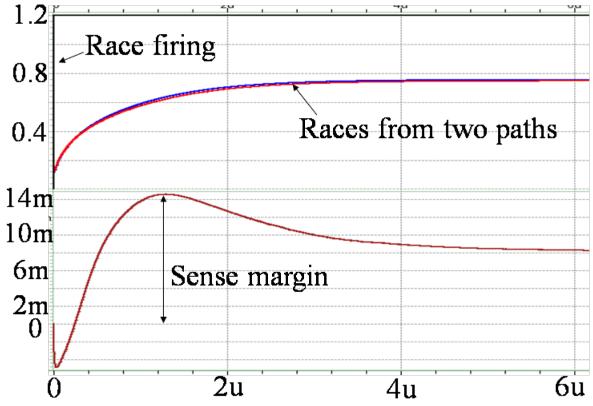
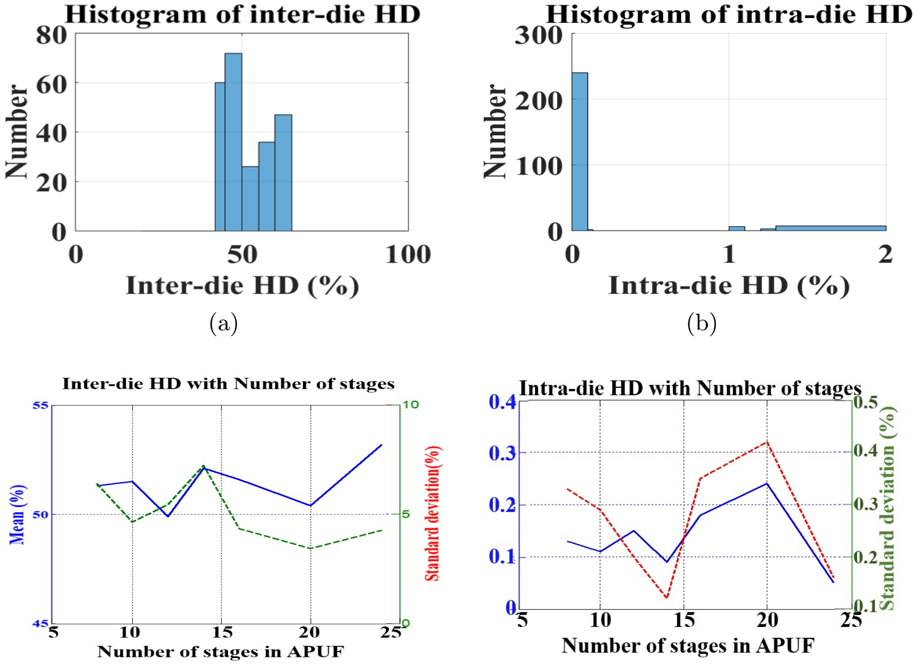


Figure 6.5: Race firing and sense margin development at differential SA.

Fig. 6.5 shows the firing of race and the race signals from two paths in the simulation waveform. Race signals from paths are fed to the differential sense amplifier which resolves to generate a response. The sense margin is the voltage difference between the race signals is also shown in the diagram. A peak sense margin of 14.5mV is developed when the sense amplifier can be fired. Approximately 10mV of continuous difference between the signals of the race is due to potential drop difference across the resistance of RRAMs in the path.

- *Uniqueness*: Uniqueness in the PUF response enables the identification of different chips uniquely. Uniqueness is measured by inter-die HD. 50% of inter-die HD indicates better uniqueness of PUF response [137]. Fig. 6.6a shows the plot of percentage inter-HD of the proposed APUF. We have measured inter-die HD by varying the threshold voltage of access transistor by ±10% and process variation in RRAM array as explained in Chapter 3. It is evident from the graph that inter-HD is close to 50% with the mean of 51.3% which indicates desired quantity of uniqueness for practical applications. Fig. 6.6c shows the mean and SD of inter-die HD for APUF with various APUF, which demonstrates desirable inter-die properties with mean in the range of 49%-53% and SD range of 3.45%-7.23%.



(a)

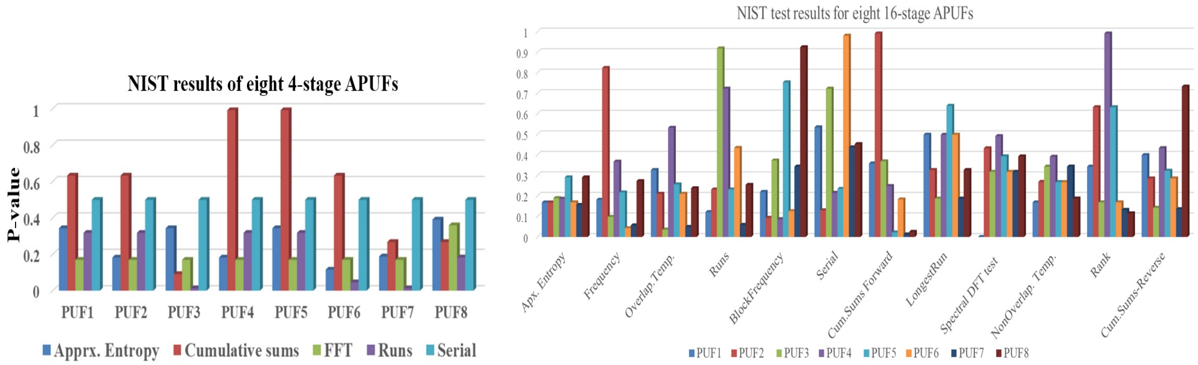
(b)

(c)

(d)

Figure 6.6: Histogram of the PUF responses such as inter-die HD and intra-die HD. (a) & (b) Inter-die HD and intra-die HD for 8 stage APUF;and, (c) & (d) Inter-HD and intra-HD of APUF with various number of stages mean and standard deviation.

- *Reliability*: : Reliability is the measure of the dependency of PUF response to the intra chip parameter variations such as voltage and temperature. The reliability of a PUF can be measured by its intra-die HD which should be close to 0% for all the possible challenges of a PUF for all responses [137]. Intra-die HD is measured by 'XOR'ing the responses of the PUF at various conditions of voltage (±10% variations) and temperature (-10^{o}C to 90^{o}C). In our measurements HD is close to 0% for most of the challenges and is less than 2% with the mean of 0.13% (Fig. 6.6b). Fig. 6.6d illustrates the intra-die HD of APUF with various stages. It can be observed that most of the responses have zero intra-die HD with mean in the range of 0.05%-0.24% and SD in the range of 0.12%-0.42% for all the responses.



(a)

(b)

Figure 6.7: NIST results of eight: (a) 4-stage; and (b) 16-stage APUFs. Passed all the tests with P-value > 0.01.

- *Uniformity*: For uniformity in the PUF response, the probability of 1s and 0s in the response for possible challenges should be 50%. We evaluate the uniformity by the frequency metric in the NIST benchmark for all the possible 256 CRPS of 8 RRAM bits in an APUF. The test showed 50-53% of probability of 1s and 0s with block frequency test, which guarantees a desired uniformity in the PUF responses. NIST test results for eight different 4-stage and 16-stage PUFs are shown in the graph Fig. 6.6. Entropy test on the responses show p-value greater than 0.01 which ensures randomness. 16-stage PUF is chosen to have sufficient length of bitstream to apply all the tests in NIST test suite [137].

Number of CRPs with respect to the number of PUF stages is shown in Fig. 6.8. It can be observed that exponential number of CRPs can be obtained from the proposed PUF by altering the number of global columns employed in the path. Number of CRPs in the proposed grow exponentially with the number of stages in APUF (Fig. 6.8a) and can be increased further by exploiting the feature of configurability. The number of CRPs with minimum of 4 stages in APUF configuration with 8 GCs and 8 LCs in the memory array is shown in Fig. 6.8b.

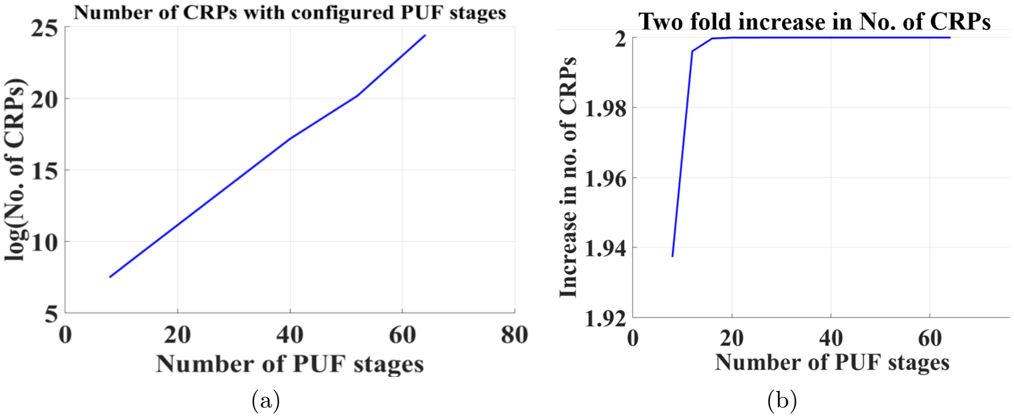


Figure 6.8: Number of CRPs with and without configurability. (a) Number of CRPs with respect to number of stages; and, (b) Two-fold increase in the number of CRPs with configurable PUF stages.

6.5 Attacks on Proposed APUF

In this section, we discuss the adversary attacks on APUF and present the results of vulnerability analysis on the proposed APUF architecture. Model building attacks based on ML algorithms, side channel attack, hybrid ML and side channel attacks [124, 138, 21, 139, 140] have been proposed in the literature. We investigate ML algorithms for model building and side channel attack based on power information of the APUF in this chapter.

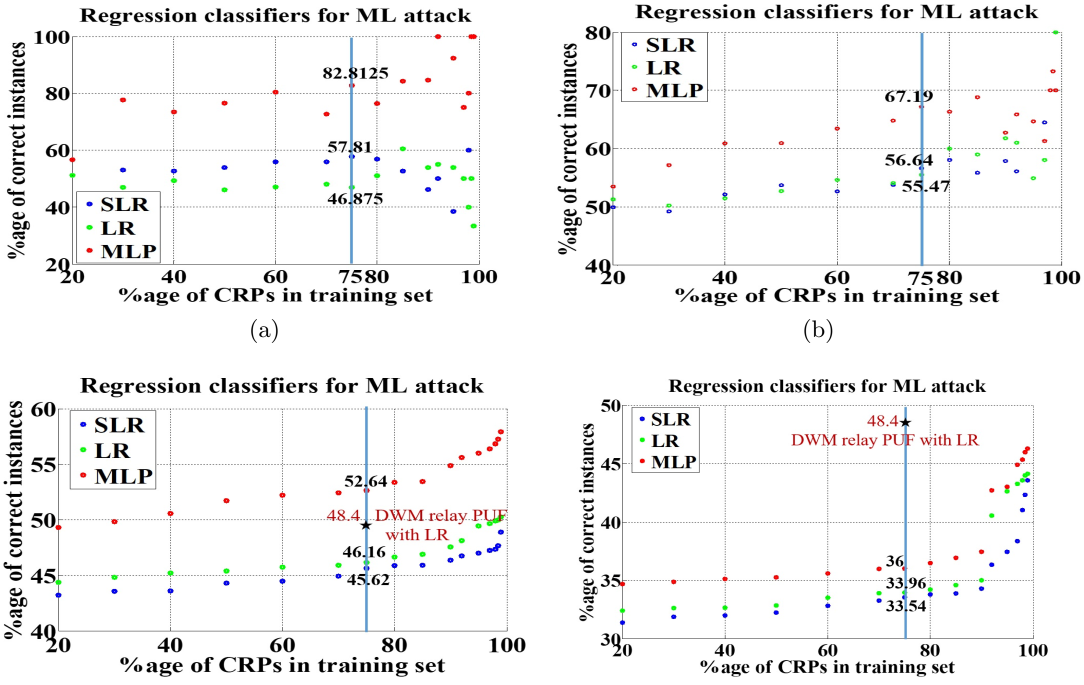
6.5.1 ML Based Model Building Attacks

ML attacks are based on using computer algorithms to model the behavior of PUF. Output of PUF is a binary response for a given challenge which is solved as a classification problem in ML. ML algorithms for PUF modeling use the classifiers with supervised learning. A percentage of CRPs are used to train the ML classifier and the model developed on the training data set is used to predict the remaining responses of PUF [124]. Logistic Regression (LR) and Support Vector Machine are two machine learning frameworks investigated extensively in PUF modeling attacks.

We investigate LR based ML attacks using data mining tool from the University of Waikato [141, 142], Waikato Environment for Knowledge Analysis (WEKA). We use WEKA to model the proposed APUF using ML algorithms. We also investigate Multi-Layer Perceptron (MLP) which is another LR classifier for ML attack on the proposed PUF. The ability of MLP neural network algorithm to model non-linear behavior is the motivating factor for this study [143]. Performance of ML attack is measured as percentage of correctly predicted instances with given percentage of training samples, termed as success rate in the rest of this chapter.

Fig. 6.9 shows the results of using Simple Logistic Regression (SLR), LR and MLP data mining algorithms for model building of the APUF with 8, 10, 16 and 24 stages. SLR is a simple variant of LR algorithm where the predicted variable takes only two values (0/1, TRUE/FALSE in PUF) [144]. We measure success rate with various percentage of test instances from 10% to 99%. success rate of various ML attacks is compared at 75% of test instances to establish comparison with other APUFs in the literature. SLR algorithm performed better than LR algorithm for smaller training samples. With 75% training set the correctly predicted instances were as low as 46.87% in LR while in SLR and MLP the correctly predicted instances 57.81% and 82.81% respectively

(Fig. 6.9a). Percentage of correctly predicted instances in 8-stage APUF using LR (46.87%) is smaller than that in SRAM (65.6%) and DWM PUFs (48.4%) proposed in literature [50]. Also, the percentage of instances correctly predicted vary in non-linear fashion with the percentage of training samples (Fig. 6.9) which could be associated with the non-linear behavior of proposed APUF.



(a)

(b)

(c)

(d)

Figure 6.9: Performance of regression classifiers on proposed APUF. SLR, LR, and MLP classifiers for machine learning attack on RRAM APUF: (a) 8-stage; (b) 10-stage; (c) 16-stage; and, (d) 24-stage.

Observing the performance of LR and SLR algorithms motivates us to use MLP which can be used to build models predicting non-linearity. MLP is a logistic regression based multi-layer neural network algorithm with non-linear activation functions used in the hidden layers [143]. MLP performed well in predicting non-linear behavior of RRAM APUF and yields more than 72% of correctly predicted instances in most of the percentage splits of training and test samples from 30-85% which demonstrates the weakness of APUF to model building attacks by choosing a suitable

ML algorithm [124]. However, to get 100% of correct instances using MLP we had to use training set size of 92%. ML attacks SLR, LR and MLP performed well on 10-stage APUF (Fig. 6.9b). With 75% of training set 56.64%, 55.47% and 67.19% of test instances were predicted correctly with SLR, LR and MLP regression classifiers respectively.

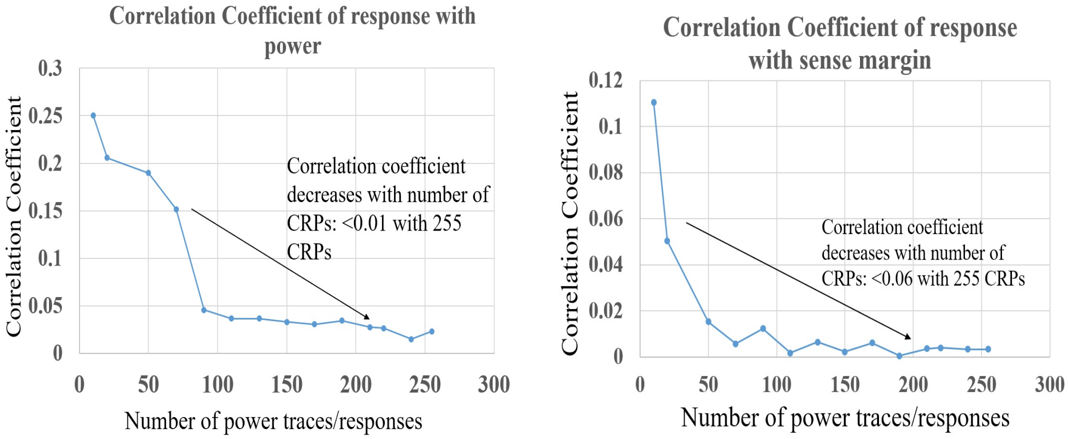
For 16-stage and 24-stage even with MLP only 52.64% and 36% of the instances were correctly predicted with 75% training and test sample split (Figs. 6.9c and 6.9d). The improvement in the resilience of APUF is due to increased number of samples with non-linearity to be predicted with fewer training samples. Given millions (2^{24}) of CRPs with 24-stage APUF, adversary will be able to efficiently observe only smaller percentage of samples for training. Therefore, limiting the training samples to less than 50% at most 35% of the test samples were predicted correct with LR, SLR and MLP. With longer paths and larger CRPs proposed APUF is robust to ML attacks compared to SRAM and, DWM PUFs. CMOS APUFs are vulnerable to ML attacks; with <20% training set in 64-stage APUF Support Vector Machine and Artificial Neural Network (with multilevel hidden layer suitable to binary classification problem in non-linear data set) provide success rate of greater than 65% successfully [138]. We leverage the proposed architecture to improve the resilience of RRAM APUF with fewer stages against ML attacks (Section 6.6).

6.5.2 Analysis of Side Channel Attacks

APUF is a strong PUF with large number of CRPs. APUFs can be designed to safeguard against the ML attacks [128] which can be achieved by selection of an arbiter function to minimize the correlation between the CRPs and path delay. Side channel attacks [21] are based on analyzing correlation between the dynamic powers consumed by the PUF with respective CRPs. The correlation coefficient indicates vulnerability of the circuit to power analysis attacks such as side channel attack [139]. We calculate the correlation coefficient by calculating the power consumed for each of 256 CRPs in an 8-stage APUF. correlation coefficient between the logic values (R) and power (P) is calculated by using the equation:

$$Correlation - Coefficient(R, P) = \frac{E[(R - \mu_R)(P - \mu_P)]}{\sigma_R \times \sigma_P} \tag{6.1}$$

where, μ_R and μ_P are the mean of R and P, σ_R and σ_P are the standard deviation of R and P respectively. The correlation coefficient between CRPs and power gets closer to zero with the number of CRPs (Fig. 6.10a) which indicates that there exists no strong correlation exists between the response bits and the power drawn by the APUF circuit. The correlation coefficient decreases with number of CRPs examined in the proposed PUF. The proposed APUF is resilient against the side channel attack. We also calculate the correlation coefficient between the sense margin of all the CRPs and the power for generating the PUF response for 8-stage APUF. The correlation coefficient plot is shown in Fig. 6.10b. This indicates no strong correlation between the challenges and sense



(a)

(b)

Figure 6.10: Correlation coefficient with number of CRPs. CC decreases with the larger number of CRPs.

margin which is amplified by the sense amplifier to generate the response. Hence, it is not feasible to model the PUF response from a known set of CRPs by the method of side channel attack. Next section proposes an APUF architecture resilient to ML attacks based on classic XOR arbiter circuit.

6.6 Proposed Architecture for Resiliency to ML Attacks

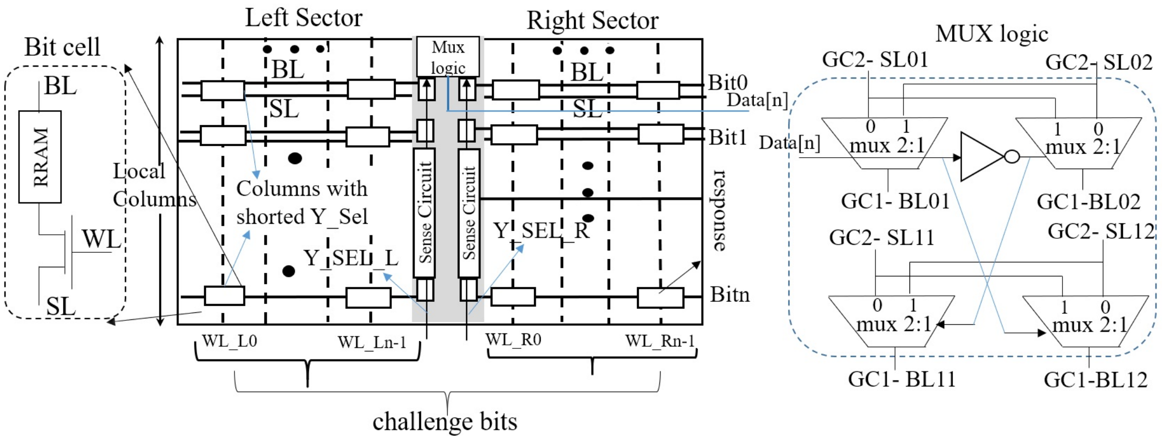
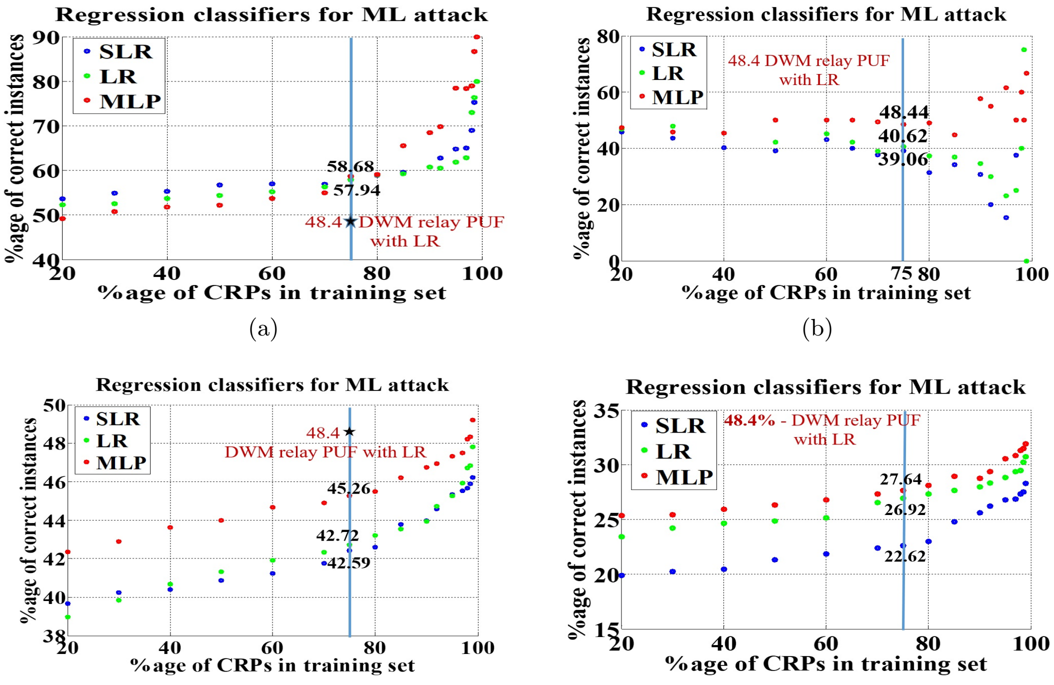


Figure 6.11: RRAM APUF architecture resilient to ML attacks. Two local columns are selected simultaneously from two sectors.

To improve the resilience against ML based model building attacks we employ the classic approach of using XOR operation to generate final PUF response. XOR APUFs are simplest realization of ML attack resilient design of APUFs [128, 145]. We leverage the proposed architecture to generate APUF response by 'XOR'ing the responses from multiple APUFs. RRAM cells in the paths of multiple APUFs are selected simultaneously by shorting their 'Y_sel' signals. Fig. 6.11 shows the architecture of RRAM APUF leveraged for ML attack resilience with XOR of responses from two APUFs. We select two cells from same row and different local columns of a GC to establish four paths for two APUFs. Responses from the two APUFs are XOR'ed to generate final response. MUX selects for the two MUXes are generated from a single MUX (Data[n]) select signal by complementing it. This minimizes the routing complexity of interface signals and simplifies implementation from a conventional 1T-1R memory architecture.

The technique can be extended by using different Y_sel lines to select more than two local columns simultaneously. 'XOR' of all the responses (more than 2 APUFs) is calculated at the final stage to generate the PUF response. However, it adds the overhead of additional interconnects and input ports in the 1T-1R array. This is due to separate MUX selects and Y_sel signals required for each of the APUFs.



(a)

(b)

(c)

(d)

Figure 6.12: ML attack on proposed ML resilient architecture. Performance of regression classifiers SLR, LR and MLP for ML attack on proposed ML resilient architecture with (a) 8-stage; (b) 10-stage; (c) 16-stage; and, (d) 24-stage.

Fig. 6.12 presents the results of regression classifiers SLR, LR and MLP for ML attack on the proposed ML resilient architecture for 8-stage, 10-stage, 16-stage and 24-stage APUF. PUF response is produced by XORing the response from two APUFs. Significant reduction in the number of correctly predicted instances is observed from the plots. With 75% of the test vectors only less than 50% of the instances were predicted correctly in 10-stage PUF. In 8-stage APUF less than 60% with (less compared to SRAM PUFs 65.6%) of instances were predicted correctly using LR and SLR classifiers. In ten or more number of stages of APUF success rate of ML algorithms including MLP lowers below that in DWM PUF using LR (48.4%) with 75% of training data. This demonstrates appreciable improvement in resilience to ML attacks with the proposed 'XOR' based APUF architecture.

The proposed APUF is strong and can be leveraged to be ML attack resilient with minimal implementation, and area overhead from 1T-1R memory architecture. Table-6.2 summarizes the comparative analysis of the proposed PUF with other RRAM based PUFs in the literature. We present an application of the proposed APUF for data attestation in the IoTs in the next section along with the qualitative analysis of the proposed attestation technique.

Table 6.2: Comparison of RRAM based PUFs

	[133]	[134]	[135]	Proposed APUF
Topology	1T-1R Memory PUF	1T-1R Memory PUF	Cross bar array	APUF with 1T-1R memory
No. of CRPs	Quadratic: $N \times N \times log_2N$	Quadratic: $N \times N \times log_2N$	Linear: $C_n^2 \times N \times log_2N$ C_n: No. of cols.	Exponential: $2^{GC} \times 2^{LC} \times 2^N$ GC: No. of GCs, LC: No. of LCs N: No. of MUX stages
% Inter HD	50% with SD=3.2%	NA	49-50%	51.3% mean; SD=0.33%
% Intra HD	0%	NA	1.7-2.3%	0.13% mean; SD=0.33%
ML attack success rate	NA	NA	NA	<=36% without XOR; <28% with XOR given 75% training set

6.7 Application in Hardware Attestation

6.7.1 Basics of Hardware Attestation

IoT is a system consisting of various computing and non-computing, living and non-living things connected to interact with each other. In such an environment, establishing the integrity of each of the connected objects is a challenge [146]. Attestation is a method of establishing the trust and integrity of a remote device. Attestation ensures the security by establishing trust in operations performed on remote device. Various techniques based on software, hardware and hybrid have been proposed in the literature [146, 147, 148, 149, 150, 151, 152]. Sensor nodes in an IoT system are light weight with minimal or almost no software application layer. Integrity of the sensor hardware is an important requirement for establishing trust in its data. The sensor nodes in IoT system where hardware plays vital role in sensing and sending the data to the base station, hardware implementation of attestation algorithm is viable. Unlike in a computing device running various software applications in which software needs attestation, in an IoT system data sent from the sensor node should be attested by its hardware.

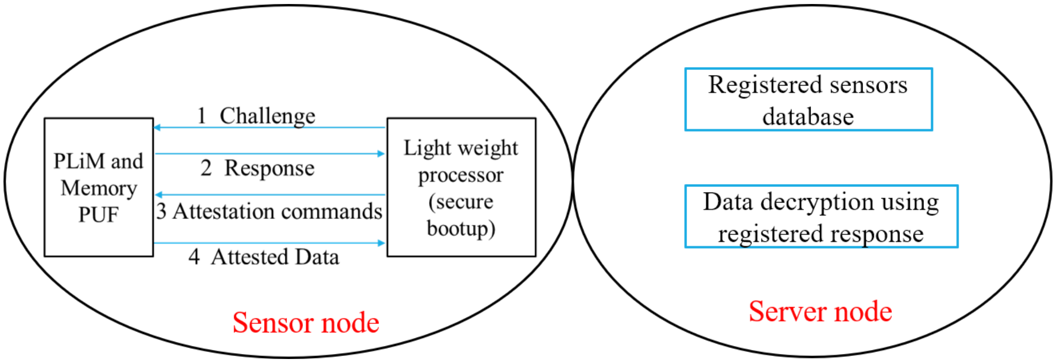

PLiM and Memory PUF
1 Challenge
2 Response
3 Attestation commands
4 Attested Data
Light weight processor (secure bootup)
Sensor node
Registered sensors database
Data decryption using registered response
Server node

Figure 6.13: Sensor nodes and server node in IoT network.

Proposed APUF demonstrates good statistical properties and resilience to adversary attacks. We propose a method of attestation using the proposed PUF architecture. This is achieved by integrating a data encryption algorithm that uses the key generated from the APUF. A light weight attestation module implementation is proposed. Attestation in light weight sensor nodes consist of a sensor which is registered with the base station prior to its deployment. Therefore, base station is aware of hardware and in specific PUF CRPs of a registered sensor node. The proposed PUF has large number of CRPs, only a few CRPs for each of the registered sensor nodes are used for attestation. Challenge is a public key and respective response is the secret key used for encryption (Fig. 6.13). We will discuss the implementation of encryption hardware and data attestation in the coming subsections.

6.7.2 Implementation of Encryption Hardware

Design for data attestation uses a Programmable Logic in Memory (PLiM) computer [153]. In [153], a technique is proposed where the computation is done within RRAM memory. It employs a light weight finite state machine for instruction execution on the data stored in memory. The principle of computation within memory is based on the RESET and SET operation of RRAM. With '1' stored in the RRAM, it switches to '0' or remains '1' depending on the polarity of the voltage applied across its terminals. SET state is read as '1' and RESET state is read as '0'. Combining the operations in the two cases: a) switching when '1' is stored; and b) switching when '0' is stored, operation on memory location can be written as $Z_n = A.Z + B'.Z + B'.A$. Where, Z is the initial value stored in the memory. Z_n is the value stored in the memory location after operation with A and B are the signals applied to top and bottom electrodes respectively. A' and B' are complement of A and B respectively. For the computation A, B and Z are stored in memory locations initially. Instructions are executed in terms of read and write operations initiated from an external Finite State Machine (FSM) based light weight processor. Implementation of PRESENT encryption algorithm in the PLiM computer is also presented in [153]. Data is encrypted with the response as key stored in the user register using PRESENT or by simple XOR operation.

6.7.3 Data Attestation Sensor Node

Sensor node is registered with the base station during installation. The base station selects a set of CRPs to be used for data attestation from the sensor node. Likewise, each of the sensor nodes in the network are registered with a selected number of CRPs for attestation of the data sent. Sensor nodes encrypt using the attestation algorithm with Response of its registered challenge and records the challenge. While sending the data, the sensor node sends challenge along with encrypted data. Different challenges are used in random order to encrypt different data blocks. Base station when it receives the encrypted data with the challenge, it looks up for the respective response from its database and decrypt with response as secret key to read the data transferred.

6.7.4 Performance Analysis of Proposed Attestation Technique

The write time of 1ns and write energy of 0.1fJ/bit is assumed for the write performance of RRAM with maturity of RRAM technology [153]. We measure energy/bit of key generation from PUF. Proposed APUF can generate a bit of response every 0.8us. Speed of 80-bit key generation is 156.25kbps by using 10 APUFs to generate 8-bits of 80-bit key in parallel. Each of the ten PUFs generate 8-bit response. The responses are appended to form an 80-bit key for attestation. Energy/bit generation of APUF response is 50fJ. For 80bit key generation 4pJ of energy is consumed. Attesting a 64-bits block of data with 80-bit key consumes ~5.88pJ of energy. Total energy for attestation of a block of data is 9.88pJ for 64-bits data block. Proposed architecture with PRESENT encryption together can offer a speed of 120.7kbps [153].

6.8 Summary

Thew chapter presented a 1T-1R RRAM APUF using a hafnium oxide based RRAM. Multiplexers in the symmetric paths of APUF could be placed in the column area of the memory subarray. Sense circuits in the conventional memory architecture is employed as arbiter. Overall, the implementation of the proposed PUF is minimally invasive from a 1T-1R memory subarray. The proposed PUF response is evaluated by systematic PUF evaluation methodology demonstrates 0% intra HD for most of CRPs with the mean of 0.13% and inter HD of mean 51.3% with sufficient randomness in the response. Number of CRPs in proposed APUF increase exponentially with the array size and number of global columns in the subarray. The proposed APUF is strong and resilient against possible adversary attacks compared to RRAM memory PUFs proposed in the literature. A potential application for data attestation in IoTs is also presented. Speed of 120.7kbps can be achieved with 9.88pJ of total energy for 64-bits block data attestation.

CHAPTER 7 : SUMMARY

A crucial and interesting topic of broadening the application of emerging NVM technologies is researched in this thesis. MTJ device is proposed for applications as associative memory. Interesting features of RRAM that are concerning in employing technology as a conventional memory are studied and exploited effectively in the design of hardware security primitives such as TRNG and APUF.Study of potential applications and design trade-offs using emerging NVM outside conventional application as storage is the primary contribution in this thesis.

Associative memory cell 6T-2MTJ TCAM proposed finds numerous applications in the areas computing and communication (network routers and search engines). Proposed TCAM cell is viable compared to its silicon and other NVM (DWM and memristor) counterparts in terms of area,speed and negligible standby power. Proposed TCAM is appropriate in the applications requiring longer standby times and longer word lengths upto 256 bits with fewer memory updates.

Another NVM technology RRAM is explored for its applications in hardware security. Salient features of RRAM such as switching variation and RTN are extremely important in designing hardware security primitives. Compatibility with CMOS, high switching speed and low power operation of RRAM enables chip architecture for application in IoTs with CMOS processor controlling the computation within memory. Within memory encryption and attestation is a viable solution due to enhanced security against invasive and non-invasive attacks, area and power efficiency.

Future work that extends the application of NVM technologies:

1. Crypto processor within 1T-1R memory architecture

   Implementation of APUF within 1T-1R RRAM memory architecture is proposed and evaluated in 6. Also, an application for data attestation is presented. The architectural design technique can be extended to realise advanced cryptographic and encryption algorithms with an on-chip encryption key and random seed value generation. The proposed architecture employs a light weight processor external to memory for instruction execution in terms of memory read and write cycles. Architectural optimization by reducing the number read/write cycles can be explored in future work.

2. RRAM APUF against ML attacks

   In this work, we have leveraged the proposed APUF within 1T-1R memory architecture to be ML attack resilient using a classical technique of 'XOR' operation on PUF responses. However, ML techniques are in extensive research and have proven promising in predicting complex data patterns. Advanced architectural and APUF design techniques with different types of arbiters and RRAM memory architectures should be investigated in future work. Effects of RRAM aging on PUF response, characteristics and compensation techniques are of future scope in APUF design.

REFERENCES

[1] Swaroop Ghosh. Spintronics and security: Prospects, vulnerabilities, attack models, and preventions. *Proceedings of the IEEE*, 104(10):1864–1893, 2016.

[2] Rekha Govindaraj and Swaroop Ghosh. A strong arbiter puf using resistive ram within 1t-1r memory architecture. In *Computer Design (ICCD), 2016 IEEE 34th International Conference on*, pages 141–148. IEEE, 2016.

[3] D Veksler, G Bersuker, B Chakrabarti, E Vogel, S Deora, K Matthews, DC Gilmer, H-F Li, S Gausepohl, and PD Kirsch. Methodology for the statistical evaluation of the effect of random telegraph noise (rtn) on rram characteristics. In *Electron Devices Meeting (IEDM), 2012 IEEE International*, pages 9–6. IEEE, 2012.

[4] Bijan Davari, Robert H Dennard, and Ghavam G Shahidi. Cmos scaling for high performance and low power-the next ten years. *Proceedings of the IEEE*, 83(4):595–606, 1995.

[5] M. Mitchell Waldrop. Nature news feature. http://www.nature.com/news/the-chips-are-down-for-moore-s-law-1.19338. Accessed: 2017-11-03.

[6] John M Shalf and Robert Leland. Computing beyond moore's law. *Computer*, 48(12):14–23, 2015.

[7] Yuan Taur, Douglas A Buchanan, Wei Chen, David J Frank, Khalid E Ismail, Shih-Hsien Lo, George A Sai-Halasz, Raman G Viswanathan, H-JC Wann, Shalom J Wind, et al. Cmos scaling into the nanometer regime. *Proceedings of the IEEE*, 85(4):486–504, 1997.

[8] Antoine Cros, Krunoslav Romanjek, Dominique Fleury, Samuel Harrison, Robin Cerutti, Philippe Coronel, Benjamin Dumont, Arnaud Pouydebasque, Romain Wacquez, Blandine Duriez, et al. Unexpected mobility degradation for very short devices: A new challenge for cmos scaling. In *Electron Devices Meeting, 2006. IEDM'06. International*, pages 1–4. IEEE, 2006.

[9] Kihwan Choi. Nand flash memory. *Samsung Electronics Co., Ltd.*, 2010.

[10] Lei Wang, CiHui Yang, Jing Wen, and Shan Gai. Emerging nonvolatile memories to go beyond scaling limits of conventional cmos nanodevices. *Journal of Nanomaterials*, 2014:235, 2014.

[11] Sung-Kye Park. Technology scaling challenge and future prospects of dram and nand flash memory. In *Memory Workshop (IMW), 2015 IEEE International*, pages 1–4. IEEE, 2015.

[12] Meng-Fan Chang, Pi-Feng Chiu, and Shyh-Shyuan Sheu. Circuit design challenges in embedded memory and resistive ram (rram) for mobile soc and 3d-ic. In *Proceedings of the 16th Asia and South Pacific Design Automation Conference*, pages 197–203. IEEE Press, 2011.

[13] Benjamin C Lee, Engin Ipek, Onur Mutlu, and Doug Burger. Architecting phase change memory as a scalable dram alternative. In *ACM SIGARCH Computer Architecture News*, volume 37, pages 2–13. ACM, 2009.

[14] H-S Philip Wong and Sayeef Salahuddin. Memory leads the way to better computing. *Nature nanotechnology*, 10(3):191, 2015.

[15] Anurag Nigam, Clinton W Smullen, Vidyabhushan Mohan, Eugene Chen, Sudhanva Gurumurthi, and Mircea R Stan. Delivering on the promise of universal memory for spin-transfer torque ram (stt-ram). In *Low Power Electronics and Design (ISLPED) 2011 International Symposium on*, pages 121–126. IEEE, 2011.

[16] Jeffrey S Vetter and Sparsh Mittal. Opportunities for nonvolatile memory systems in extreme-scale high-performance computing. *Computing in Science & Engineering*, 17(2):73–82, 2015.

[17] Jagan Singh Meena, Simon Min Sze, Umesh Chand, and Tseung-Yuen Tseng. Overview of emerging nonvolatile memory technologies. *Nanoscale research letters*, 9(1):526, 2014.

[18] Mark Tehranipoor, Domenic Forte, Garrett S Rose, and Swarup Bhunia. *Security Opportunities in Nano Devices and Emerging Technologies*. CRC Press, 2017.

[19] Masoud Rostami, Farinaz Koushanfar, and Ramesh Karri. A primer on hardware security: Models, methods, and metrics. *Proceedings of the IEEE*, 102(8):1283–1295, 2014.

[20] Kaveh Shamsi, Wujie Wen, and Yier Jin. Hardware security challenges beyond cmos: Attacks and remedies. In *VLSI (ISVLSI), 2016 IEEE Computer Society Annual Symposium on*, pages 200–205. IEEE, 2016.

[21] Jeroen Delvaux and Ingrid Verbauwhede. Side channel modeling attacks on 65nm arbiter pufs exploiting cmos device noise. In *Hardware-Oriented Security and Trust (HOST), 2013 IEEE International Symposium on*, pages 137–142. IEEE, 2013.

[22] James A Hutchby, George I Bourianoff, Victor V Zhirnov, and Joe E Brewer. Extending the road beyond cmos. *IEEE Circuits and Devices Magazine*, 18(2):28–41, 2002.

[23] Kelin J Kuhn. Cmos scaling for the 22nm node and beyond: Device physics and technology. In *VLSI Technology, Systems and Applications (VLSI-TSA), 2011 International Symposium on*, pages 1–2. IEEE, 2011.

[24] Suman Datta. Recent advances in high performance cmos transistors: From planar to non-planar. *The Electrochemical Society Interface*, 22(1):41–46, 2013.

[25] Hoang Anh Du Nguyen, Lei Xie, Mottaqiallah Taouil, Razvan Nane, Said Hamdioui, and Koen Bertels. Computation-in-memory based parallel adder. In *Nanoscale Architectures (NANOARCH), 2015 IEEE/ACM International Symposium on*, pages 57–62. IEEE, 2015.

[26] Hoang Anh Du Nguyen, Lei Xie, Mottaqiallah Taouil, Razvan Nane, Said Hamdioui, and Koen Bertels. On the implementation of computation-in-memory parallel adder. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 25(8):2206–2219, 2017.

[27] Lei Xie, Hoang Anh Du Nguyen, Mottaqiallah Taouil, Said Hamdioui, and Koen Bertels. Boolean logic gate exploration for memristor crossbar. In *Design and Technology of Integrated Systems in Nanoscale Era (DTIS), 2016 International Conference on*, pages 1–6. IEEE, 2016.

[28] Lei Xie, Hoang Anh Du Nguyen, Mottaqiallah Taouil, Said Hamdioui, and Koen Bertels. Fast boolean logic mapped on memristor crossbar. In *Computer Design (ICCD), 2015 33rd IEEE International Conference on*, pages 335–342. IEEE, 2015.

[29] Ioannis Vourkas and Georgios Ch Sirakoulis. Emerging memristor-based logic circuit design approaches: A review. *IEEE Circuits and Systems Magazine*, 16(3):15–30, 2016.

[30] Said Hamdioui, Shahar Kvatinsky, Gert Cauwenberghs, Lei Xie, Nimrod Wald, Siddharth Joshi, Hesham Mostafa Elsayed, Henk Corporaal, and Koen Bertels. Memristor for computing: Myth or reality? In *Proceedings of the Conference on Design, Automation & Test in Europe*, pages 722–731. European Design and Automation Association, 2017.

[31] J Joshua Yang, Dmitri B Strukov, and Duncan R Stewart. Memristive devices for computing. *Nature nanotechnology*, 8(1):13, 2013.

[32] Shahar Kvatinsky, Dmitry Belousov, Slavik Liman, Guy Satat, Nimrod Wald, Eby G Friedman, Avinoam Kolodny, and Uri C Weiser. MagicâĂŤmemristor-aided logic. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 61(11):895–899, 2014.

[33] Shahar Kvatinsky, Nimrod Wald, Guy Satat, Avinoam Kolodny, Uri C Weiser, and Eby G Friedman. MrlâĂŤmemristor ratioed logic. In *Cellular Nanoscale Networks and Their Applications (CNNA), 2012 13th International Workshop on*, pages 1–6. IEEE, 2012.

[34] Lauren Guckert and Earl E Swartzlander. Mad gatesâĂŤmemristor logic design using driver circuitry. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 64(2):171–175, 2017.

[35] Weisheng Zhao and Guillaume Prenat. *Spintronics-based computing*. Springer, 2015.

[36] Weisheng Zhao, Dafine Ravelosona, J Klein, and Claude Chappert. Domain wall shift register-based reconfigurable logic. *IEEE Transactions on Magnetics*, 47(10):2966–2969, 2011.

[37] Yuhao Wang, Hao Yu, Leibin Ni, Guang-Bin Huang, Mei Yan, Chuliang Weng, Wei Yang, and Junfeng Zhao. An energy-efficient nonvolatile in-memory computing architecture for extreme learning machine by domain-wall nanowire devices. *IEEE Transactions on Nanotechnology*, 14(6):998–1012, 2015.

[38] Sanjukta Bhanja, DK Karunaratne, Ravi Panchumarthy, Srinath Rajaram, and Sudeep Sarkar. Non-boolean computing with nanomagnets for computer vision applications. *Nature nanotechnology*, 11(2):177, 2016.

[39] Mrigank Sharad, Deliang Fan, Kyle Aitken, and Kaushik Roy. Energy-efficient non-boolean computing with spin neurons and resistive memory. *IEEE Transactions on Nanotechnology*, 13(1):23–34, 2014.

[40] Dan A Allwood, Gang Xiong, CC Faulkner, D Atkinson, D Petit, and RP Cowburn. Magnetic domain-wall logic. *Science*, 309(5741):1688–1692, 2005.

[41] Yuhao Wang and Hao Yu. An ultralow-power memory-based big-data computing platform by nonvolatile domain-wall nanowire devices. In *Low Power Electronics and Design (ISLPED), 2013 IEEE International Symposium on*, pages 329–334. IEEE, 2013.

[42] Mrigank Sharad, Deliang Fan, and Kaushik Roy. Ultra low power associative computing with spin neurons and resistive crossbar memory. In *Proceedings of the 50th Annual Design Automation Conference*, page 107. ACM, 2013.

[43] Xiaochen Guo, Engin Ipek, and Tolga Soyata. Resistive computation: avoiding the power wall with low-leakage, stt-mram based computing. In *ACM SIGARCH Computer Architecture News*, volume 38, pages 371–382. ACM, 2010.

[44] Weisheng Zhao, Eric Belhaire, Claude Chappert, François Jacquet, and Pascale Mazoyer. New non-volatile logic based on spin-mtj. *physica status solidi (a)*, 205(6):1373–1377, 2008.

[45] Shoun Matsunaga, Jun Hayakawa, Shoji Ikeda, Katsuya Miura, Tetsuo Endoh, Hideo Ohno, and Takahiro Hanyu. Mtj-based nonvolatile logic-in-memory circuit, future prospects and issues. In *Proceedings of the Conference on Design, Automation and Test in Europe*, pages 433–435. European Design and Automation Association, 2009.

[46] Kyungho Ryu, Jisu Kim, Jiwan Jung, Jung Pill Kim, Seung H Kang, and Seong-Ook Jung. A magnetic tunnel junction based zero standby leakage current retention flip-flop. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 20(11):2044–2053, 2012.

[47] Weisheng Zhao, Eric Belhaire, and Claude Chappert. Spin-mtj based non-volatile flip-flop. In *Nanotechnology, 2007. IEEE-NANO 2007. 7th IEEE Conference on*, pages 399–402. IEEE, 2007.

[48] Anirudh Srikant Iyengar, Swaroop Ghosh, and Jae-Won Jang. Mtj-based state retentive flip-flop with enhanced-scan capability to sustain sudden power failure. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 62(8):2062–2068, 2015.

[49] Dwaipayan Chakraborty, Sunny Raj, Julio Cesar Gutierrez, Troyle Thomas, and Sumit Kumar Jha. In-memory execution of compute kernels using flow-based memristive crossbar computing. In *Rebooting Computing (ICRC), 2017 IEEE International Conference on*, pages 1–6. IEEE, 2017.

[50] Anirudh Iyengar, Swaroop Ghosh, Kenneth Ramclam, Jae-Won Jang, and Cheng-Wei Lin. Spintronic pufs for security, trust, and authentication. *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, 13(1):4, 2016.

[51] Swaroop Ghosh and Rekha Govindaraj. Spintronics for associative computation and hardware security. In *Circuits and Systems (MWSCAS), 2015 IEEE 58th International Midwest Symposium on*, pages 1–4. IEEE, 2015.

[52] Rekha Govindaraj and Swaroop Ghosh. Design and analysis of 6-t 2-mtj ternary content addressable memory. In *Low Power Electronics and Design (ISLPED), 2015 IEEE/ACM International Symposium on*, pages 309–314. IEEE, 2015.

[53] Rekha Govindaraj and Swaroop Ghosh. Design and analysis of sttram-based ternary content addressable memory cell. *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, 13(4):52, 2017.

[54] R. Govindaraj, S. Ghosh, and S. Katkoori. Csro-based reconfigurable true random number generator using rram. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, pages 1–10, 2018. ISSN 1063-8210. doi: 10.1109/TVLSI.2018.2823274.

[55] Linda Engelbrecht. *Modeling spintronics devices in Verilog-A for use with industry-standard simulation tools*. Oregon State University, 2011.

[56] ND Rizzo, D Houssameddine, J Janesky, R Whig, FB Mancoff, ML Schneider, M DeHerrera, JJ Sun, K Nagel, S Deshpande, et al. A fully functional 64 mb ddr3 st-mram built on 90 nm cmos technology. *IEEE Transactions on Magnetics*, 49(7):4441–4446, 2013.

[57] Mark H Kryder and Chang Soo Kim. After hard drivesâĂŤwhat comes next? *IEEE Transactions on Magnetics*, 45(10):3406–3413, 2009.

[58] Bernard Dieny, Ronald B Goldfarb, and Kyung-Jin Lee. *Introduction to magnetic random-access memory*. John Wiley & Sons, 2016.

[59] Jun-Yang Chen, Yong-Chang Lau, JMD Coey, Mo Li, and Jian-Ping Wang. High performance mgo-barrier magnetic tunnel junctions for flexible and wearable spintronic applications. *Scientific Reports*, 7:42001, 2017.

[60] LALE Landau and Evgeny Lifshitz. On the theory of the dispersion of magnetic permeability in ferromagnetic bodies. *Phys. Z. Sowjetunion*, 8(153):101–114, 1935.

[61] Jianwei Zhang, Peter M Levy, Shufeng Zhang, and Vladimir Antropov. Identification of transverse spin currents in noncollinear magnetic structures. *Physical review letters*, 93(25): 256602, 2004.

[62] André Thiaville and Yoshinobu Nakatani. Domain-wall dynamics in nanowiresand nanostrips. In *Spin dynamics in confined magnetic structures III*, pages 161–205. Springer, 2006.

[63] Teruya Shinjo. *Nanomagnetism and spintronics*, chapter 3. Elsevier, Amsterdam, 1 edition, 2013.

[64] Xuanyao Fong, Sri Harsha Choday, Panagopoulos Georgios, Charles Augustine, and Kaushik Roy. Spice models for magnetic tunnel junctions based on monodomain approximation. 2013.

[65] Behtash Behin-Aein, Deepanjan Datta, Sayeef Salahuddin, and Supriyo Datta. Proposal for an all-spin logic device with built-in memory. *Nature nanotechnology*, 5(4):266–270, 2010.

[66] Shruti Patil, Andrew Lyle, Jonathan Harms, David J Lilja, and Jian-Ping Wang. Spintronic logic gates for spintronic data using magnetic tunnel junctions. In *Computer Design (ICCD), 2010 IEEE International Conference on*, pages 125–131. IEEE, 2010.

[67] Charles Augustine. *Spintronic memory and logic: From atoms to systems*. PhD thesis, Purdue University, 2011.

[68] Dmitri E Nikonov, George I Bourianoff, and Tahir Ghani. Proposal of a spin torque majority gate logic. *IEEE Electron Device Letters*, 32(8):1128–1130, 2011.

[69] Gopalakrishnan Srinivasan, Abhronil Sengupta, and Kaushik Roy. Magnetic tunnel junction based long-term short-term stochastic synapse for a spiking neural network with on-chip stdp learning. *Scientific reports*, 6:29545, 2016.

[70] Angik Sarkar, Behtash Behin-Aein, Srikant Srinivasan, and Supriyo Datta. All spin logic device as a compact artificial neuron. In *Device Research Conference (DRC), 2012 70th Annual*, pages 99–100. IEEE, 2012.

[71] Shankar Ganesh Ramasubramanian, Rangharajan Venkatesan, Mrigank Sharad, Kaushik Roy, and Anand Raghunathan. Spindle: Spintronic deep learning engine for large-scale neuromorphic computing. In *Proceedings of the 2014 international symposium on Low power electronics and design*, pages 15–20. ACM, 2014.

[72] Jacob Torrejon, Mathieu Riou, Flavio Abreu Araujo, Sumito Tsunegi, Guru Khalsa, Damien Querlioz, Paolo Bortolotti, Vincent Cros, Kay Yakushiji, Akio Fukushima, et al. Neuromorphic computing with nanoscale spintronic oscillators. *Nature*, 547(7664):428, 2017.

[73] Julie Grollier, Damien Querlioz, and Mark D Stiles. Spintronic nanodevices for bioinspired computing. *Proceedings of the IEEE*, 104(10):2024–2039, 2016.

[74] Abhronil Sengupta and Kaushik Roy. Spin-transfer torque magnetic neuron for low power neuromorphic computing. In *Neural Networks (IJCNN), 2015 International Joint Conference on*, pages 1–7. IEEE, 2015.

[75] Francesco Maria Puglisi, Paolo Pavan, Andrea Padovani, and Luca Larcher. A compact model of hafnium-oxide-based resistive random access memory. In *IC Design & Technology (ICICDT), 2013 International Conference on*, pages 85–88. IEEE, 2013.

[76] FM Puglisi, L Larcher, P Pavan, A Padovani, and G Bersuker. Instability of hfo 2 rram devices: Comparing rtn and cycling variability. In *Reliability Physics Symposium, 2014 IEEE International*, pages MY–5. IEEE, 2014.

[77] Francesco M Puglisi, Luca Larcher, Gennadi Bersuker, Andrea Padovani, and Paolo Pavan. An empirical model for rram resistance in low-and high-resistance states. *IEEE Electron Device Letters*, 34(3):387–389, 2013.

[78] Shimeng Yu, Ximeng Guan, and H-S Philip Wong. On the switching parameter variation of metal oxide rramâĂŤpart ii: Model corroboration and device design strategy. *IEEE Transactions on Electron Devices*, 59(4):1183–1188, 2012.

[79] Ting-Chang Chang, Kuan-Chang Chang, Tsung-Ming Tsai, Tian-Jian Chu, and Simon M Sze. Resistance random access memory. *Materials Today*, 19(5):254–264, 2016.

[80] An Chen and Ming-Ren Lin. Reset switching probability of resistive switching devices. *IEEE Electron Device Letters*, 32(5):590–592, 2011.

[81] Simone Balatti, Stefano Ambrogio, Roberto Carboni, Valerio Milo, Zhongqiang Wang, Alessandro Calderoni, Nirmal Ramaswamy, and Daniele Ielmini. Physical unbiased generation of random numbers with coupled resistive switching devices. *IEEE Transactions on Electron Devices*, 63(5):2029–2035, 2016.

[82] Simone Balatti, Stefano Ambrogio, Antonio Cubeta, Alessandro Calderoni, Nirmal Ramaswamy, and Daniele Ielmini. Voltage-dependent random telegraph noise (rtn) in hfo x resistive ram. In *Reliability Physics Symposium, 2014 IEEE International*, pages MY–4. IEEE, 2014.

[83] Yuan Heng Tseng, Wen Chao Shen, and Chrong Jung Lin. Modeling of electron conduction in contact resistive random access memory devices as random telegraph noise. *Journal of applied physics*, 111(7):073701, 2012.

[84] Francesco Maria Puglisi and Paolo Pavan. Factorial hidden markov model analysis of random telegraph noise in resistive random access memories. *ECTI Transactions on Electrical Engineering, Electronics, and Communications*, 12(1):24–29, 2014.

[85] Kostas Pagiamtzis and Ali Sheikholeslami. Content-addressable memory (cam) circuits and architectures: A tutorial and survey. *IEEE Journal of Solid-State Circuits*, 41(3):712–727, 2006.

[86] Robert Karam, Ruchir Puri, Swaroop Ghosh, and Swarup Bhunia. Emerging trends in design and applications of memory-based computing and content-addressable memories. *Proceedings of the IEEE*, 103(8):1311–1330, 2015.

[87] Weifeng Shen, Dipanjan Mazumdar, Xiaojing Zou, Xiaoyong Liu, BD Schrag, and Gang Xiao. Effect of film roughness in mgo-based magnetic tunnel junctions. *Applied physics letters*, 88 (18):182508, 2006.

[88] Pilin Junsangsri and Fabrizio Lombardi. A memristor-based tcam (ternary content addressable memory) cell: design and evaluation. In *Proceedings of the great lakes symposium on VLSI*, pages 311–314. ACM, 2012.

[89] Azam Seyedi, Vasileios Karakostas, Stefan Cosemans, Adrian Cristal, Mario Nemirovsky, and Osman Unsal. Nemscam: A novel cam cell based on nano-electro-mechanical switch and cmos for energy efficient tlbs. In *Nanoscale Architectures (NANOARCH), 2015 IEEE/ACM International Symposium on*, pages 51–56. IEEE, 2015.

[90] Jing Li, Robert K Montoye, Masatoshi Ishii, and Leland Chang. 1 mb 0.41 μm^2 2t-2r cell nonvolatile tcam with two-bit encoding and clocked self-referenced sensing. *IEEE Journal of Solid-State Circuits*, 49(4):896–907, 2014.

[91] Wei Xu, Tong Zhang, and Yiran Chen. Spin-transfer torque magnetoresistive content addressable memory (cam) cell structure design with enhanced search noise margin. In *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, pages 1898–1901. IEEE, 2008.

[92] R Nebashi, N Sakimura, Y Tsuji, S Fukami, H Honjo, S Saito, S Miura, N Ishiwata, K Kinoshita, T Hanyu, et al. A content addressable memory using magnetic domain wall motion cells. In *VLSI Circuits (VLSIC), 2011 Symposium on*, pages 300–301. IEEE, 2011.

[93] Yue Zhang, Weisheng Zhao, Jacques-Olivier Klein, Dafiné Ravelsona, and Claude Chappert. Ultra-high density content addressable memory based on current induced domain wall motion in magnetic track. *IEEE Transactions on Magnetics*, 48(11):3219–3222, 2012.

[94] Ke Chen, Jie Han, and Fabrizio Lombardi. Design and evaluation of two mtj-based content addressable non-volatile memory cells. In *Nanotechnology (IEEE-NANO), 2013 13th IEEE Conference on*, pages 707–712. IEEE, 2013.

[95] Shoun Matsunaga, Sadahiko Miura, Hiroaki Honjou, Keizo Kinoshita, Shoji Ikeda, Tetsuo Endoh, Hideo Ohno, and Takahiro Hanyu. A 3.14 um 2 4t-2mtj-cell fully parallel tcam based on nonvolatile logic-in-memory architecture. In *VLSI Circuits (VLSIC), 2012 Symposium on*, pages 44–45. IEEE, 2012.

[96] YU Cao, T Sato, D Sylvester, M Orshansky, and C Hu. Predictive technology model. *Internet: http://ptm. asu. edu*, 2002.

[97] Ki Chul Chun, Hui Zhao, Jonathan D Harms, Tae-Hyoung Kim, Jian-Ping Wang, and Chris H Kim. A scaling roadmap and performance evaluation of in-plane and perpendicular mtj based stt-mrams for high-density cache memory. *IEEE Journal of Solid-State Circuits*, 48(2):598–610, 2013.

[98] Xiaobin Wang. *Metallic Spintronic Devices*. CRC Press, 2014.

[99] Swaroop Ghosh, Saibal Mukhopadhyay, Keejong Kim, and Kaushik Roy. Self-calibration technique for reduction of hold failures in low-power nano-scaled sram. In *Proceedings of the 43rd annual Design Automation Conference*, pages 971–976. ACM, 2006.

[100] Rekha Govindaraj, Indranil Sengupta, and Santanu Chattopadhyay. An efficient technique for longest prefix matching in network routers. *Progress in VLSI Design and Test*, pages 317–326, 2012.

[101] Mario Stipčević and Çetin Kaya Koç. True random number generators. In *Open Problems in Mathematics and Computational Science*, pages 275–315. Springer, 2014.

[102] Atakan Arslan, Süleyman Kardas, Sultan Aldirmaz, and Sarp Ertürk. Are rngs achilles' heel of rfid security and privacy protocols? *IACR Cryptology ePrint Archive*, 2016:1130, 2016.

[103] Yuan Ma, Jingqiang Lin, and Jiwu Jing. On the entropy of oscillator-based true random number generators. In *CryptographersâĂŹ Track at the RSA Conference*, pages 165–180. Springer, 2017.

[104] Benjamin Jun and Paul Kocher. The intel random number generator. *Cryptography Research Inc. white paper*, 1999.

[105] Yingjie Lao, Qianying Tang, Chris H Kim, and Keshab K Parhi. Beat frequency detector–based high-speed true random number generators: Statistical modeling and analysis. *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, 13(1):9, 2016.

[106] Akio Fukushima, Takayuki Seki, Kay Yakushiji, Hitoshi Kubota, Hiroshi Imamura, Shinji Yuasa, and Koji Ando. Spin dice: A scalable truly random number generator based on spintronics. *Applied Physics Express*, 7(8):083001, 2014.

[107] Won Ho Choi, Yang Lv, Jongyeon Kim, Abhishek Deshpande, Gyuseong Kang, Jian-Ping Wang, and Chris H Kim. A magnetic tunnel junction based true random number generator with conditional perturb and real-time output probability tracking. In *Electron Devices Meeting (IEDM), 2014 IEEE International*, pages 12–5. IEEE, 2014.

[108] Yandan Wang, Wei Wen, Hai Li, and Miao Hu. A novel true random number generator design leveraging emerging memristor technology. In *Proceedings of the 25th edition on Great Lakes Symposium on VLSI*, pages 271–276. ACM, 2015.

[109] Simone Balatti, Stefano Ambrogio, Zhongqiang Wang, and Daniele Ielmini. True random number generation by variability of resistive switching in oxide-based devices. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 5(2):214–221, 2015.

[110] Jianguo Yang, Juan Xu, Bo Wang, Xiaoyong Xue, Ryan Huang, Qingtian Zhou, Jingang Wu, and Yinyin Lin. A low cost and high reliability true random number generator based on resistive random access memory. In *ASIC (ASICON), 2015 IEEE 11th International Conference on*, pages 1–4. IEEE, 2015.

[111] Robert Karam, Rui Liu, Pai-Yu Chen, Shimeng Yu, and Swarup Bhunia. Security primitive design with nanoscale devices: a case study with resistive ram. In *Great Lakes Symposium on VLSI, 2016 International*, pages 299–304. IEEE, 2016.

[112] Gang Niu, Hee-Dong Kim, Robin Roelofs, Eduardo Perez, Markus Andreas Schubert, Peter Zaumseil, Ioan Costina, and Christian Wenger. Material insights of hfo2-based integrated 1-transistor-1-resistor resistive random access memory devices processed by batch atomic layer deposition. *Scientific reports*, 6:28155, 2016.

[113] A Theodore Markettos and Simon W Moore. The frequency injection attack on ring-oscillator-based true random number generators. In *CHES*, volume 5747, pages 317–331. Springer, 2009.

[114] GS Jovanović and MK Stojčev. Current starved delay element with symmetric load. *International journal of electronics*, 93(03):167–175, 2006.

[115] Siam U Hussain, Mehrdad Majzoobi, and Farinaz Koushanfar. A built-in-self-test scheme for online evaluation of physical unclonable functions and true random number generators. *IEEE Transactions on Multi-Scale Computing Systems*, 2(1):2–16, 2016.

[116] Ali Hajimiri, Sotirios Limotyrakis, and Thomas H Lee. Phase noise in multi-gigahertz cmos ring oscillators. In *Custom Integrated Circuits Conference, 1998. Proceedings of the IEEE 1998*, pages 49–52. IEEE, 1998.

[117] Anju P Johnson, Rajat Subhra Chakraborty, and Debdeep Mukhopadhyay. A novel attack on a fpga based true random number generator. In *Proceedings of the WESS'15: Workshop on Embedded Systems Security*, page 6. ACM, 2015.

[118] Pierre Bayon, Lilian Bossuet, Alain Aubert, Viktor Fischer, François Poucheret, Bruno Robisson, and Philippe Maurine. Contactless electromagnetic active attack on ring oscillator based true random number generator. *COSADE*, 7275:151–166, 2012.

[119] Pierre Bayon, Lilian Bossuet, Alain Aubert, and Viktor Fischer. Fault model of electromagnetic attacks targeting ring oscillator-based true random number generators. *Journal of Cryptographic Engineering*, 6(1):61–74, 2016.

[120] Berk Sunar, William J Martin, and Douglas R Stinson. A provably secure true random number generator with built-in tolerance to active attacks. *IEEE Transactions on computers*, 56(1), 2007.

[121] Eberhard Böhl and Markus Ihle. A fault attack robust trng. In *On-Line Testing Symposium (IOLTS), 2012 IEEE 18th International*, pages 114–117. IEEE, 2012.

[122] Selçuk Köse and Eby G Friedman. Distributed power network co-design with on-chip power supplies and decoupling capacitors. In *Proceedings of the System Level Interconnect Prediction Workshop*, page 13. IEEE Press, 2011.

[123] Scott Durrant. Random numbers in data security systems. *hup: Z/wu-w. inte*, 1, 1999.

[124] Ulrich Rührmair, Frank Sehnke, Jan Sölter, Gideon Dror, Srinivas Devadas, and Jürgen Schmidhuber. Modeling attacks on physical unclonable functions. In *Proceedings of the 17th ACM conference on Computer and communications security*, pages 237–249. ACM, 2010.

[125] Mohtashim Mansoor, Ibraheem Haneef, Suhail Akhtar, Andrea De Luca, and Florin Udrea. Silicon diode temperature sensorsâĂŤa review of applications. *Sensors and Actuators A: Physical*, 232:63–74, 2015.

[126] Siew-Hwee Kwok, Yen-Ling Ee, Guanhan Chew, Kanghong Zheng, Khoongming Khoo, and Chik-How Tan. A comparison of post-processing techniques for biased random number generators. In *IFIP International Workshop on Information Security Theory and Practices*, pages 175–190. Springer, 2011.

[127] G Edward Suh and Srinivas Devadas. Physical unclonable functions for device authentication and secret key generation. In *Proceedings of the 44th annual design automation conference*, pages 9–14. ACM, 2007.

[128] Charles Herder, Meng-Day Yu, Farinaz Koushanfar, and Srinivas Devadas. Physical unclonable functions and applications: A tutorial. *Proceedings of the IEEE*, 102(8):1126–1141, 2014.

[129] Roel Maes and Ingrid Verbauwhede. Physically unclonable functions: A study on the state of the art and future research directions. In *Towards Hardware-Intrinsic Security*, pages 3–37. Springer, 2010.

[130] Anas Mazady, Md Tauhidur Rahman, Domenic Forte, and Mehdi Anwar. Memristor pufâĂŤa security primitive: Theory and experiment. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 5(2):222–229, 2015.

[131] An Chen, X Sharon Hu, Yier Jin, Michael Niemier, and Xunzhao Yin. Using emerging technologies for hardware security beyond pufs. In *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2016*, pages 1544–1549. IEEE, 2016.

[132] H-S Philip Wong, Heng-Yuan Lee, Shimeng Yu, Yu-Sheng Chen, Yi Wu, Pang-Shiu Chen, Byoungil Lee, Frederick T Chen, and Ming-Jinn Tsai. Metal–oxide rram. *Proceedings of the IEEE*, 100(6):1951–1970, 2012.

[133] An Chen. Utilizing the variability of resistive random access memory to implement reconfigurable physical unclonable functions. *IEEE Electron Device Letters*, 36(2):138–140, 2015.

[134] Le Zhang, Xuanyao Fong, Chip-Hong Chang, Zhi Hui Kong, and Kaushik Roy. Feasibility study of emerging non-volatilememory based physical unclonable functions. In *Memory Workshop (IMW), 2014 IEEE 6th International*, pages 1–4. IEEE, 2014.

[135] Pai-Yu Chen, Runchen Fang, Rui Liu, Chaitali Chakrabarti, Yu Cao, and Shimeng Yu. Exploiting resistive cross-point array for compact design of physical unclonable function. In *Hardware Oriented Security and Trust (HOST), 2015 IEEE International Symposium on*, pages 26–31. IEEE, 2015.

[136] A Chen. Reconfigurable physical unclonable function based on probabilistic switching of rram. *Electronics Letters*, 51(8):615–617, 2015.

[137] Abhranil Maiti, Vikash Gunreddy, and Patrick Schaumont. A systematic method to evaluate and compare the performance of physical unclonable functions. In *Embedded systems design with FPGAs*, pages 245–267. Springer, 2013.

[138] Gabriel Hospodar, Roel Maes, and Ingrid Verbauwhede. Machine learning attacks on 65nm arbiter pufs: Accurate modeling poses strict bounds on usability. In *Information Forensics and Security (WIFS), 2012 IEEE International Workshop on*, pages 37–42. IEEE, 2012.

[139] Stefan Mangard, Elisabeth Oswald, and Thomas Popp. *Power analysis attacks: Revealing the secrets of smart cards*, volume 31. Springer Science & Business Media, 2008.

[140] Xiaolin Xu and Wayne Burleson. Hybrid side-channel/machine-learning attacks on pufs: a new threat? In *Proceedings of the conference on Design, Automation & Test in Europe*, page 349. European Design and Automation Association, 2014.

[141] Wikipedia. Wekawikipedia, . URL \url{http://www.cs.waikato.ac.nz/ml/weka}. Accessed: 2017-11-09.

[142] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[143] Wikipedia. Mlpwikipedia, . URL \url{http://deeplearning.net/tutorial/mlp.html}. Accessed: 2017-11-09.

[144] Wikipedia. Slrwikipedia, . URL \url{http://www.biostathandbook.com/simplelogistic.html}. Accessed: 2017-11-09.

[145] Fatemeh Ganji, Shahin Tajik, and Jean-Pierre Seifert. Why attackers win: on the learnability of xor arbiter pufs. In *International Conference on Trust and Trustworthy Computing*, pages 22–39. Springer, 2015.

[146] Ihtesham Haider, Michael Höberl, and Bernhard Rinner. Trusted sensors for participatory sensing and iot applications based on physically unclonable functions. In *Proceedings of the 2nd ACM International Workshop on IoT Privacy, Trust, and Security*, pages 14–21. ACM, 2016.

[147] Arvind Seshadri, Adrian Perrig, Leendert Van Doorn, and Pradeep Khosla. Swatt: Software-based attestation for embedded devices. In *Security and Privacy, 2004. Proceedings. 2004 IEEE Symposium on*, pages 272–282. IEEE, 2004.

[148] Ittai Anati, Shay Gueron, Simon Johnson, and Vincent Scarlata. Innovative technology for cpu based attestation and sealing. In *Proceedings of the 2nd international workshop on hardware and architectural support for security and privacy*, volume 13, 2013.

[149] Hailun Tan, Wen Hu, and Sanjay Jha. A remote attestation protocol with trusted platform modules (tpms) in wireless sensor networks. *Security and Communication Networks*, 8(13): 2171–2188, 2015.

[150] Uthpala Subodhani Premarathne, Ibrahim Khalil, and Mohammed Atiquzzaman. Secure and reliable surveillance over cognitive radio sensor networks in smart grid. *Pervasive and Mobile Computing*, 22:3–15, 2015.

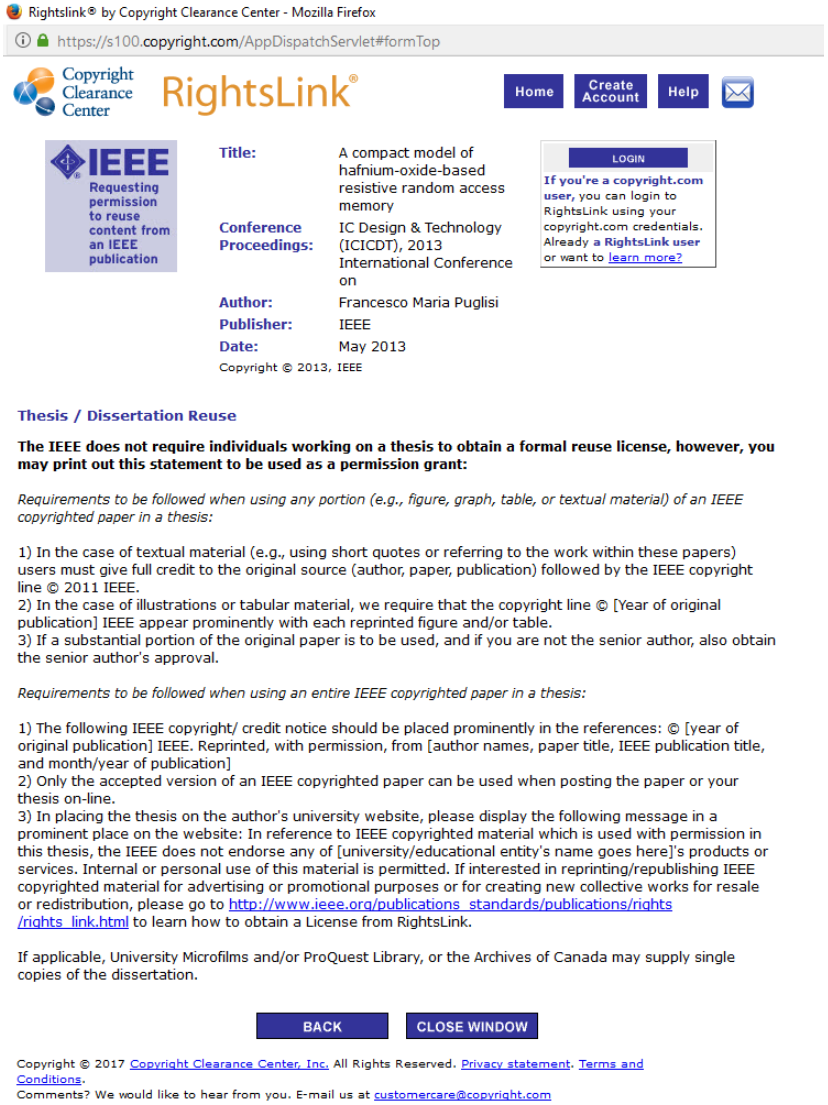
[151] Rodrigo Vieira Steiner and Emil Lupu. Attestation in wireless sensor networks: A survey. *ACM Computing Surveys (CSUR)*, 49(3):51, 2016.

[152] Tigist Abera, N Asokan, Lucas Davi, Farinaz Koushanfar, Andrew Paverd, Ahmad-Reza Sadeghi, and Gene Tsudik. Things, trouble, trust: on building trust in iot systems. In *Proceedings of the 53rd Annual Design Automation Conference*, page 121. ACM, 2016.

[153] Pierre-Emmanuel Gaillardon, Luca Amarú, Anne Siemon, Eike Linn, Rainer Waser, Anupam Chattopadhyay, and Giovanni De Micheli. The programmable logic-in-memory (plim) computer. In *Proceedings of the 2016 Conference on Design, Automation & Test in Europe*, pages 427–432. EDA Consortium, 2016.

APPENDIX A: COPYRIGHT CLEARANCE FORMS

Below is the copyright reuse permission from IEEE for contents of Chapter 1 and RRAM model equations used in Chapter 3.


Rightslink® by Copyright Clearance Center - Mozilla Firefox
https://s100.copyright.com/AppDispatchServlet#formTop
Copyright Clearance Center
RightsLink®
Home
Create Account
Help
IEEE
Requesting permission to reuse content from an IEEE publication
LOGIN
BACK
CLOSE WINDOW

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already a RightsLink user or want to learn more?

Title:	A compact model of hafnium-oxide-based resistive random access memory
Conference Proceedings:	IC Design & Technology (ICICDT), 2013 International Conference on
Author:	Francesco Maria Puglisi
Publisher:	IEEE
Date:	May 2013

Copyright © 2013, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

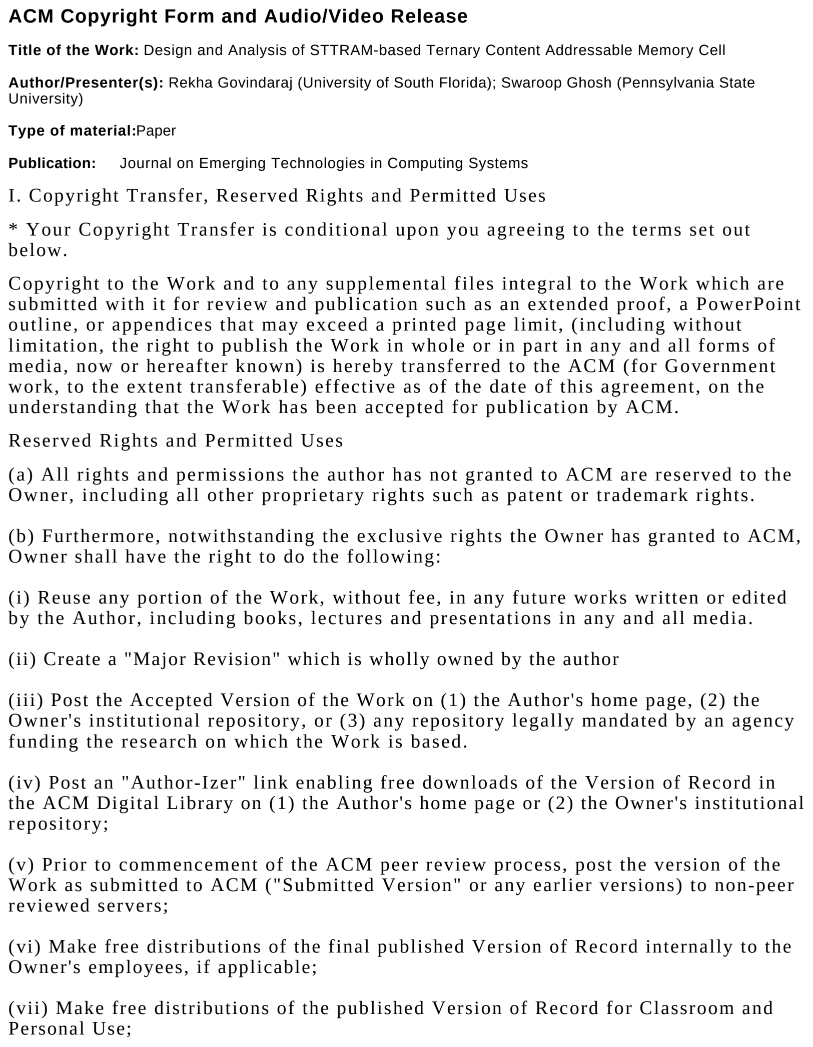
1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

Below is the permission for contents of Chapter 2 and Chapter 4.



ACM Copyright Form and Audio/Video Release

Title of the Work: Design and Analysis of STTRAM-based Ternary Content Addressable Memory Cell

Author/Presenter(s): Rekha Govindaraj (University of South Florida); Swaroop Ghosh (Pennsylvania State University)

Type of material: Paper

Publication: Journal on Emerging Technologies in Computing Systems

I. Copyright Transfer, Reserved Rights and Permitted Uses

* Your Copyright Transfer is conditional upon you agreeing to the terms set out below.

Copyright to the Work and to any supplemental files integral to the Work which are submitted with it for review and publication such as an extended proof, a PowerPoint outline, or appendices that may exceed a printed page limit, (including without limitation, the right to publish the Work in whole or in part in any and all forms of media, now or hereafter known) is hereby transferred to the ACM (for Government work, to the extent transferable) effective as of the date of this agreement, on the understanding that the Work has been accepted for publication by ACM.

Reserved Rights and Permitted Uses

(a) All rights and permissions the author has not granted to ACM are reserved to the Owner, including all other proprietary rights such as patent or trademark rights.

(b) Furthermore, notwithstanding the exclusive rights the Owner has granted to ACM, Owner shall have the right to do the following:

(i) Reuse any portion of the Work, without fee, in any future works written or edited by the Author, including books, lectures and presentations in any and all media.

(ii) Create a "Major Revision" which is wholly owned by the author

(iii) Post the Accepted Version of the Work on (1) the Author's home page, (2) the Owner's institutional repository, or (3) any repository legally mandated by an agency funding the research on which the Work is based.

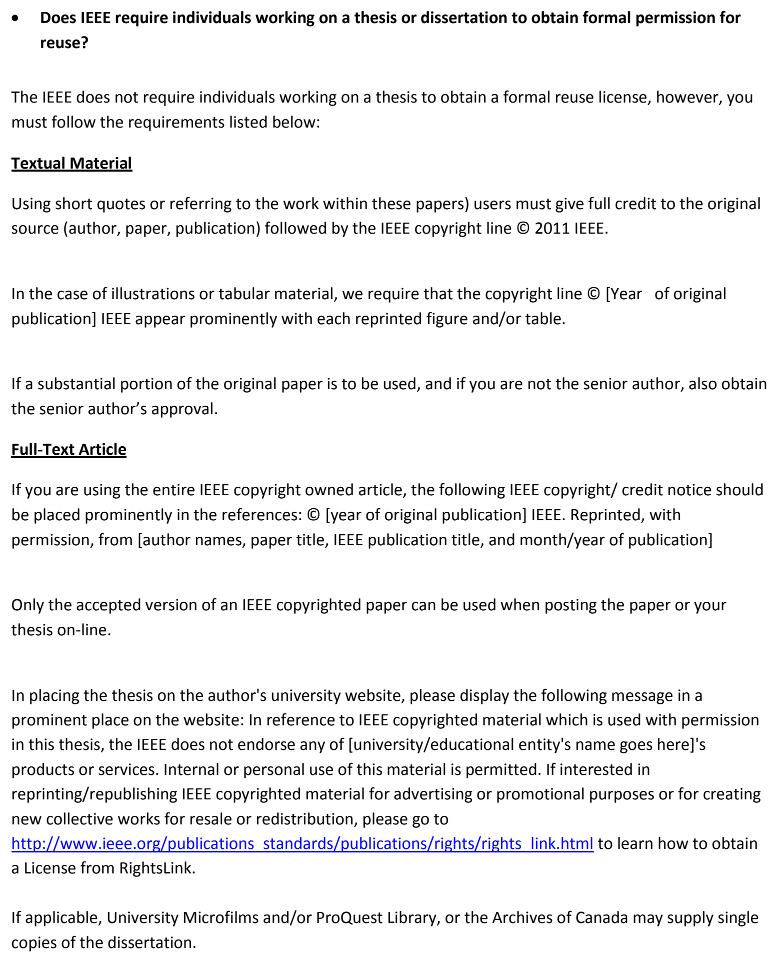
(iv) Post an "Author-Izer" link enabling free downloads of the Version of Record in the ACM Digital Library on (1) the Author's home page or (2) the Owner's institutional repository;

(v) Prior to commencement of the ACM peer review process, post the version of the Work as submitted to ACM ("Submitted Version" or any earlier versions) to non-peer reviewed servers;

(vi) Make free distributions of the final published Version of Record internally to the Owner's employees, if applicable;

(vii) Make free distributions of the published Version of Record for Classroom and Personal Use;

Below is the copyright reuse permission from IEEE for the contents of Chapter 5 and Chapter 6.



- **Does IEEE require individuals working on a thesis or dissertation to obtain formal permission for reuse?**

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you must follow the requirements listed below:

Textual Material

Using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.

In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.

If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Full-Text Article

If you are using the entire IEEE copyright owned article, the following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]

Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.

In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

ABOUT THE AUTHOR

Rekha received Bachelor of Engineering (Hons.) from Visweswaraya Technological University (2009), Master of Technology from Indian Institute of technology, Kharagpur (2012).Currently Rekha is a doctorate student at University of South Florida at LOGICS lab under the supervision of Dr.Swaroop Ghosh and Dr.Srinivas Katkoori. She worked as System on Chip design engineer with Qualcomm Inc. for two years (2012-2014) prior to starting her Doctorate degree. She also worked with custom circuits design team at Apple Inc. during Summer 2017. Her primary research interests include low power VLSI circuits and systems design, emerging non-volatile memory technologies, and hardware security. She has authored and coauthored several conference papers, journal articles and a book chapter. She holds a US patent (US Patent 9,543,013). She is a student member of National Academy of Inventors (NAI) USF chapter, and Florida Gamma chapter of Tau Beta Pi. She is awarded 'Provost's Award' recognizing her outstanding teaching as a Graduate Teaching Assistant at USF in STEM category (Spring 2017). Rekha is a recipient of 'The Spirit of Innovation Award' for the academic year 2017-18 from the College of Engineering, University of South Florida. She aspires to continue her research providing simple and elegant solutions to complex problems in engineering.