Low Power Machine Learning FPGA/ASIC for Edge Computing (Design of RRAM Array for ARYABHAT Analog Processor)

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT
FOR THE DEGREE OF

Master of Technology

IN THE
FACULTY OF ENGINEERING

BY

NANDYALA SRIKANTH, SR NO: 21566
SIDDHU YERIBROLU, SR NO: 21454

GUIDED BY

PROFESSOR CHETAN SINGH THAKUR

DEPARTMENT OF ELECTRONIC SYSTEMS ENGINEERING
INDIAN INSTITUTE OF SCIENCE, BANGALORE

JUNE 2024

Synopsis

The ARYABHAT Analog Processor Integration with Resistive Random-Access Memory(RRAM) project represents a significant advancement in the field of analog processing and non-volatile memory technology. ARYABHAT Analog Processor is known for its prowess in handling analog signals and data processing tasks. By integrating Resistive Random-Access Memory (RRAM) technology into this processor, the project aims to take advantage of the unique characteristics of RRAM, such as non-volatility, low power consumption, high speed, and scalability. This integration enhances data processing efficiency by allowing the processor to quickly process and store analog data without frequent data transfers, thus improving versatility, reducing latency, and potentially leading to more energy-efficient computing solutions. The project holds promise for a wide range of applications, from IoT devices to data centers, where compact, integrated, and high-performance solutions are essential.

Acknowledgement

We would like to express our deepest gratitude to Dr. Chetan Singh Thakur, our esteemed advisor, for his invaluable support and guidance throughout this thesis. His insightful feedback, constructive criticism, and constant encouragement were instrumental in shaping our research.

We extend our sincere thanks to Prof. L. Umanand and Dr. Viveka Konandur Rajanna, our reviewers, for their valuable feedback and expert supervision. Their insights and suggestions significantly improved the quality of our work.

We are also grateful to our colleagues and friends at NeuRonICS, DESE, and the Indian Institute of Science (IISc). Their constant support, brainstorming sessions, and constructive discussions created an intellectually stimulating environment that enhanced our learning experience.

We acknowledge the support and resources provided by our institution, which were essential for the smooth progress of our project.

Lastly, we extend our heartfelt gratitude to our families and friends for their unwavering support, encouragement, and love throughout this journey.

Notations

ML	Machine Learning
S-AC	Shape-based Analog Computing
CMOS	Complementary Metal-Oxide-Semiconductor
SRAM	Static Random Access Memory
RRAM	Resistive Random Access Memory
HRS	High Resistance State
LRS	Low Resistance State
CFs	Conductive Filaments
TE	Top Electrode
BE	Bottom Electrode
MIM	Metal-Insulator-Metal
1T1R	1-Transistor 1-Resistor
WL	Word Line
BL	Bit Line
SL	Source Line
DRAM	Dynamic Random Access Memory
AI	Artificial Intelligence
IoT	Internet of Things
PDK	Process Design Kit
μ**A**	Micro Ampere
nA	Nano Ampere
VWL	Word Line Voltage
VBL	Bit Line Voltage
VSL	Source Line Voltage
μ**s**	Micro Second
ns	Nano Second
DAC	Digital-to-Analog Converter
SA	Sense Amplifier

Contents

List of Figures

Chapter 1

Pre-study

1.1 Background

Analog computing techniques are appealing for implementing machine learning (ML) architectures due to their potential to achieve high computational density and energy efficiency, surpassing that of equivalent digital implementations.

Analog computing operates on continuous signals and utilizes the intrinsic properties of physical phenomena to perform computations. However, unlike digital circuits, traditional analog computing circuits face challenges in being easily mapped across different process nodes. These challenges arise due to differences in transistor biasing regimes, temperature variations, and a limited dynamic range.

To address the challenges and leverage the benefits of analog computing, an MP-based synthesis technique was developed in the analog domain.The Analog AI Chip (ARYABHAT) was developed using this synthesis technique. Figure 1.1 illustrates the CMOS implementation of the MP function. The MP Principle is a piecewise-linear (PWL) approximation to the log-sum-exponential functions commonly used in communication, optimization, and learning algorithms. Computations within the MP-based synthesis is conducted in the logarithmic domain, translating multiplications into additions, divisions into subtractions, and power computations into multiplications [1].

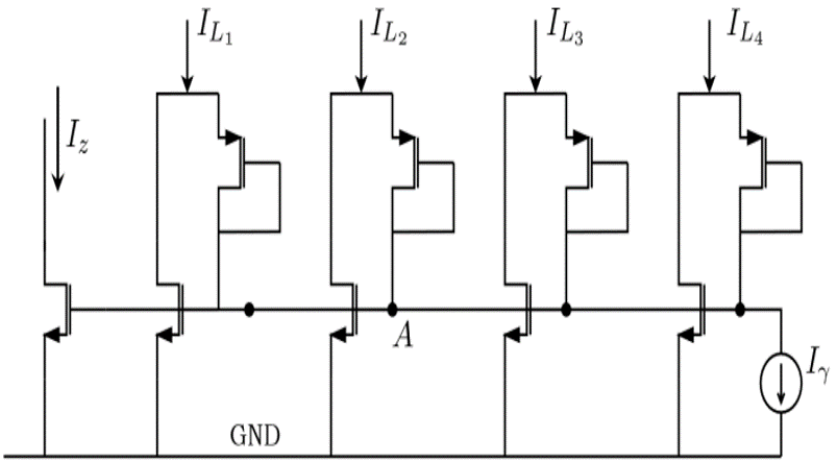


Figure 1.1: CMOS implementation of the MP function [1]

In the ARYABHAT chip, the weights are stored as binary bits in Memory Elements (SRAM) and converted into corresponding currents using the S-AC DAC. Similarly, the input data is mapped to their respective equivalent currents and fed into the core for computation. Figure 1.2 illustrates the log-binary DAC implementation using S-AC and Figure 1.3 illustrates the ARYABHAT chip [2].

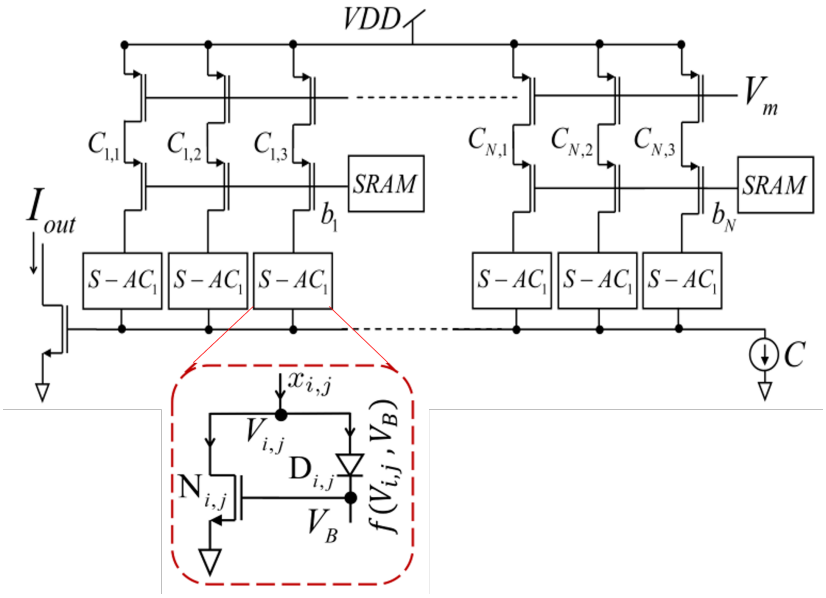


Figure 1.2: Compressive log-binary DAC implementation using S-AC [2]

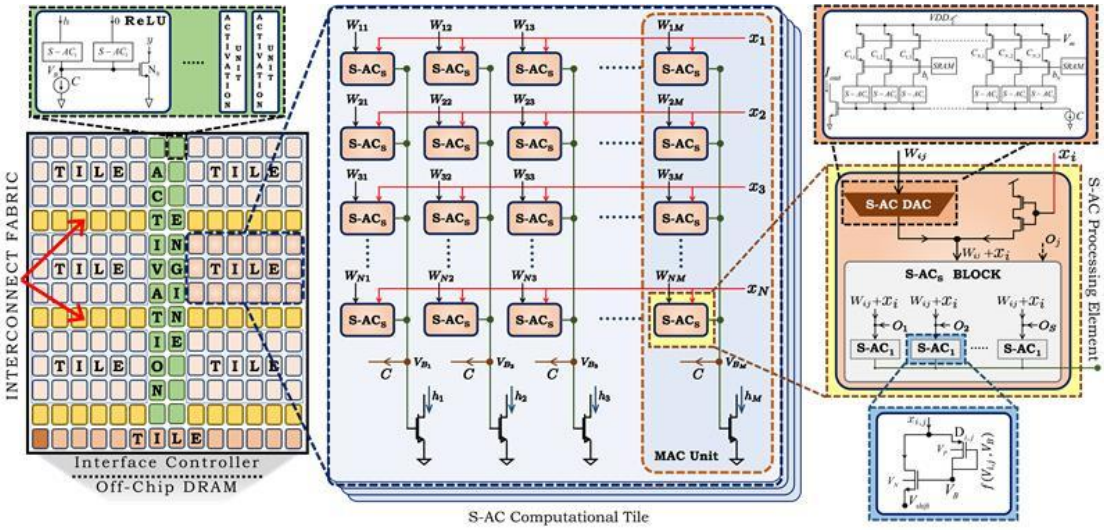


Figure 1.3: ARYABHAT chip [2]

The primary focus is on leveraging RRAM (Resistive Random-Access Memory) technology over SRAM for weight storage. This decision is primarily motivated by the fundamental distinction between the volatile nature of SRAM and the non-volatile nature of RRAM. RRAM's non-volatility ensures data retention even when power is removed, a critical requirement for edge applications. Additionally, RRAM offers the advantage of one-time write capability for machine learning tasks, while also facilitating multiple reads, further enhancing its suitability.

1.2 Functional aspects

1.2.1 Primary Function of the Proposed Project

The RRAM array developed in this project functions as a memory array. Leveraging the non-volatile nature of RRAM, data does not need to be written repeatedly; once written, it can be read multiple times. The primary function is to store data efficiently and provide it to the ARYABHAT chip accurately and reliably.

1.2.2 Secondary Functions

The secondary functions of the RRAM array may include:

- **Low Power Consumption:** Designed to consume minimal power during read and write

operations.

- **Scalability:** The array can be scaled to accommodate larger data storage requirements in future applications.

1.2.3 Integration with Larger Systems

The RRAM array is designed to work as part of a larger system, specifically interfacing with the ARYABHAT chip. It is essential that the array functions seamlessly with other components in the system, ensuring data integrity and efficient operation.

1.3 Characteristics and performance

The RRAM array is designed to store weights accurately and provide them to the ARYABHAT chip without any errors. The critical parameters to measure, monitor, and control include resistance levels (ensuring high and low resistance states are within limits), minimal read/write times, low power consumption and appropriate voltage levels on word, bit, and source lines. Performance requirements necessitate accuracy in weight storage and retrieval, fast operation, power efficiency, high endurance for numerous read/write cycles, and reliable long-term data retention. The absolute maximum ratings for voltage and current must be adhered to prevent damage and ensure reliable operation.

Key variables influencing performance include resistance drift over time, power supply stability, and the number of write/erase cycles, which can affect the endurance of the RRAM cells. Special considerations for input/output signals involve ensuring voltage levels are suitable for proper resistance state switching and precise timing for read/write pulses to avoid incomplete or incorrect data storage.

1.4 User Aspects

Assessing the technical competence of the user is crucial to ensure the usability and effectiveness of the designed RRAM array. Given the complexity of using RRAM technology and the tools involved:

- **Required Skills:** The user should have a solid understanding of semiconductor devices, memory technology, and circuit design principles. Familiarity with RRAM characteristics, such as resistance switching and endurance, is essential.
- **Tool Proficiency:** The user should be proficient in using open-source design tools like Xschem for schematic capture and Ngspice for circuit simulation. Knowledge of the SkyWater 130nm PDK and its libraries is also necessary.
- **Troubleshooting Abilities:** The user should be capable of diagnosing and troubleshooting issues that arise during the design, simulation, and testing phases. This includes interpreting simulation results and understanding potential failure mechanisms in RRAM.

1.5 Power Supply

The circuits designed for this project operate with a 1.8V supply. A power breakdown could result in several adverse effects, including the loss of data stored in the memory array, system instability or malfunction, and potential damage to the memory array or associated components. Special considerations on power consumption are essential, as the RRAM memory array is designed to function within a specific power budget. Exceeding these power limitations could lead to overheating or degradation of the RRAM cells, compromising the overall performance and reliability of the system.

1.6 Literature Survey

1.6.1 Existing Status of Technology or Algorithms

- The existing status of technology or algorithms in relation to the project includes:
  - RRAM (Resistive Random Access Memory) technology is gaining traction in non-volatile memory applications due to its low power consumption and high density characteristics.
  - Several studies have been conducted on the modeling and simulation of RRAM-based memory arrays.
  - Open-source tools such as XSchem and Ngspice have been utilized for the design and simulation of RRAM-based circuits.

  - Skywater 130nm PDK (Process Design Kit) provides a platform for implementing RRAM-based memory arrays in integrated circuits.

1.6.2 Gaps Addressed by the Proposed Project

- The proposed project addresses several gaps in the existing literature and technology, including:
  - Optimization of RRAM-based memory array designs for specific applications or performance metrics.
  - Exploration of novel techniques for enhancing the reliability and efficiency of RRAM-based memory systems.
  - Utilization of open-source tools and platforms for accessible and cost-effective development of RRAM-based products.

1.7 Product/Market Survey

The project is currently in the research phase, aiming to integrate RRAM with an analog processor (ARYABHAT) , a combination that is not yet available in the market.

1.8 User Survey

This project primarily serves electronic device manufacturers, engineers and researchers in semiconductor technology. While technical users, such as engineers and researchers, dominate the user base due to the specialized nature of RRAM technology, non-technical end-users of electronic devices incorporating RRAM may also benefit from its advancements. The technology is appreciated for its low power consumption, high density, and fast access times, making it appealing to those involved in semiconductor design and electronics.

Feedback on the proposed RRAM features is generally positive, with technical users valuing its power efficiency, density, and speed, and non-technical users noting its reliability and performance improvements. However, there are challenges in meeting all user-desired features, such as achieving extreme durability or unlimited write endurance and ultra-low power consumption

without compromising performance. Additionally, ensuring compatibility with legacy systems can be difficult due to technological constraints and cost considerations.

1.9 Wish

1.9.1 Features of the Project

The features of the project that we wish to undertake include low power consumption for energy-efficient operation, high-speed write and read access for quick data retrieval, enhanced reliability and endurance to withstand frequent write cycles, and scalability for integration into a wide range of applications and devices.

1.9.2 Dream/Wish Specifications

- The set of dream/wish specifications that we consider ideal for our project are as follows:
  - Power Consumption: Less than 1 milliwatt during active operation.
  - Write Speed: Sub-microsecond write latency and sustained write throughput.
  - Read Speed: Sub-microsecond read latency.
  - Endurance: Over 10^{15} write cycles with negligible degradation in performance.

1.9.3 Comparison Table of Specifications

Specifications	**RRAM**	**SRAM**
Power Consumption	Low	Moderate
Write Speed	Moderate	High-Speed
Endurance	Enhanced	Volatile
Read Speed	Moderate	High-Speed

Table 1.1: Comparison of Specifications

Chapter 2

Study

2.1 Introduction

Bias-scalable analog computing offers a promising approach for developing machine learning (ML) processors that can meet diverse power-performance requirements [4], [5]. In particular, ML applications vary widely in their requirements, with server workloads emphasizing heightened computational throughput for rapid training, while edge devices prioritize energy-efficient inference. A pivotal contribution in this realm is the ARYABHAT analog processor, designed to facilitate machine learning computations.

ARYABHAT Analog processor consists of analog computing circuits, achieved through the application of the margin-propagation principle extended to shape-based analog computing (S-AC). The central innovation lies in the creation of an S-AC core, encompassing a collection of near-memory compute elements. These elements incorporate: (a) non-linear activation functions, (b) inner-product compute circuits, and (c) a mixed-signal compressive memory. Notably, these components can be effectively scaled to optimize for either performance or power, all the while maintaining their fundamental functionality.

The practical viability of this approach is already substantiated through measured results derived from prototypes fabricated utilizing a 180nm CMOS process. Crucially, these prototypes manifest the robustness of computing modules against the variations stemming from transistor biasing and fluctuations in temperature. The impact of bias-scalability and computational accuracy is also explored in the context of a straightforward ML regression task. Within this

broader framework, the emergence of ARYABHAT as an analog processor tailored for machine learning computations reflects the paradigm shift towards optimizing ML processing for diverse power-performance requirements.

A notable aspect of the ARYABHAT processor is the disclosure that the SRAM cell is volatile. Also, in modern ICs for edge computing, data movement between on and off-chip memories typically consumes a large fraction of total power. Dense, non-volatile embedded memory can reduce/eliminate off-chip data movement by keeping frequently-read application data always on chip. Acknowledging these challenges, a strategic decision has been made to explore the replacement of SRAM with Resistive Random-Access Memory (RRAM).

2.2 Background of RRAM

Resistive Random-Access Memory (RRAM) has emerged as a promising non-volatile memory technology with applications in various fields, from consumer electronics to advanced computing systems. RRAM operates on the concept of resistive switching, wherein the resistance of a thin metal-oxide layer can be manipulated to store and retrieve data.

RRAM has garnered substantial attention as a highly promising contender for storage class memory applications due to its remarkable attributes, including rapid write speed, minimal power consumption, extensive scalability, three-dimensional integration capabilities, affordability, and seamless compatibility with CMOS fabrication processes [6]. The operational foundation of RRAM lies in the resistive switching phenomenon observed within a straightforward metal-insulator-metal structure.

In this report, we provide a brief overview of the materials that enable resistive switching and explain the switching mechanism behind it. Moreover, we explore a compelling application of the resistive switching device in the realm of non-volatile logic. To conclude, we offer insights into the future trajectory of RRAM.

2.3 Resistive switching materials and mechanism

2.3.1 Resistive switching materials

The design of Resistive Random Access Memory (RRAM) hinges on specific materials that exhibit a phenomenon known as resistive switching. This switching behavior is pivotal in storing and manipulating data within RRAM devices. Several materials have been explored for this purpose, each offering unique characteristics that impact the performance and reliability of RRAM.

One commonly investigated class of materials for RRAM is transition metal oxides [7]. These compounds, such as titanium dioxide (TiO_2), hafnium oxide (HfO_2), and tantalum oxide (Ta_2O_5), undergo changes in their electrical resistance when subjected to an electric field [8]. This property enables the creation of distinct resistance states, representing the ”0” and ”1” of digital data storage.

Metal-based materials like silver (Ag) and copper (Cu) have also demonstrated resistive switching behaviour. These metals, when combined with suitable insulating layers, exhibit reversible changes in conductivity under voltage bias, allowing for the creation and control of resistance states.

In addition to inorganic materials, organic materials have been explored for RRAM designs. Conductive polymers, like poly(3,4-ethylenedioxythiophene) poly(styrenesulfonate) (PEDOT:PSS), possess properties that facilitate resistive switching. Organic materials provide advantages like flexibility and compatibility with flexible electronics, broadening the scope of RRAM applications.

The choice of materials greatly influences key performance metrics of RRAM, including switching speed, energy efficiency, retention time, and endurance. Researchers aim to identify materials that strike a balance between these factors to optimize the overall functionality of RRAM devices.

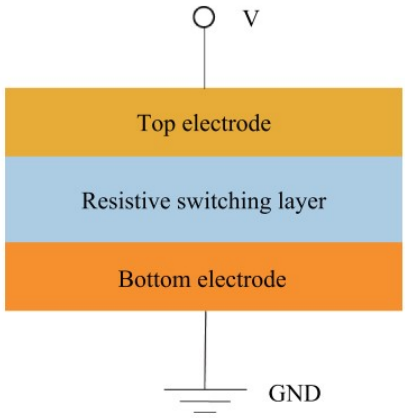


Figure 2.1: Schematic structure of RRAM Cell [3]

The structure of the RRAM device typically involves layers stacked one over the other: a layer of metal, a layer responsible for resistive switching, and another layer of metal. This arrangement is visually depicted in Figure 2.1. The layer responsible for resistive switching is nestled between the upper and lower electrodes.

The lower electrode, often made of a material like platinum (Pt), is a bit challenging to etch. In cases where a single device is considered, multiple devices can share the same lower electrode. However, when using the crossbar architecture, each device requires its own distinct lower electrode. These separate lower electrodes can be created using a sequence of processes, including lithography, physical vapor deposition (PVD), and lift-off techniques.

As for the resistive layer and the upper electrode, they are typically made of materials like metal oxides and titanium nitride (TiN). These materials are grown using techniques such as atomic layer deposition (ALD) or physical vapor deposition (PVD). Importantly, these materials are more amenable to etching after the lithography process.

As RRAM technology progresses, ongoing research focuses on refining these materials, understanding the underlying mechanisms, and developing innovative fabrication techniques to enhance the reliability, scalability, and integration potential of RRAM devices. The proper selection and engineering of resistive switching materials are pivotal in shaping the future of RRAM as a versatile and efficient memory solution.

2.3.2 Resistive Switching Mechanisms

The process of resistive switching, which involves transitioning between high resistance (HRS) and low resistance (LRS) states, is fundamental to RRAM (Resistive Random Access Memory) devices. This transition occurs through two primary processes: the ”set” and ”reset” processes. Initially, when RRAM devices are manufactured, they typically exhibit a high resistance state. To transform these devices into a low resistance state, an electro-forming process employing high voltage is often necessary.

Resistive switching in RRAM devices can be broadly classified into two types: unipolar and bipolar. In unipolar switching, as illustrated in Figure 2.2(a), the magnitude of the applied voltage dictates the switching, regardless of its polarity. Both the set and reset processes occur under the same polarity. On the other hand, in bipolar switching, depicted in Figure 2.2(b), the resistive behavior depends on the polarity of the applied voltage. The set process exclusively occurs under one polarity, while the reset process is facilitated under the opposite polarity. This distinctive behavior characterizes bipolar resistive switching.

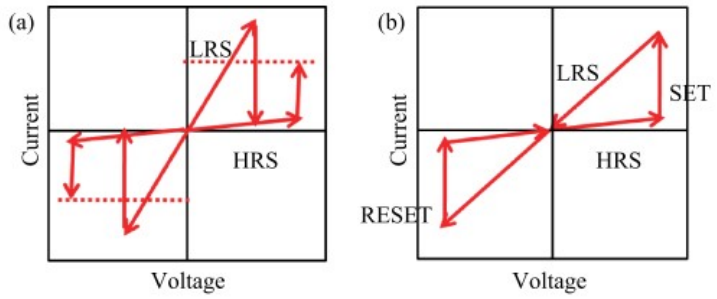


Figure 2.2: Schematic of metal oxide RRAM I-V curves showing two types of switching: (a) unipolar type and (b) bipolar type [3]

Understanding the underlying mechanisms behind resistive switching behavior has been a central focus of RRAM research. While certain aspects of the resistive switching process are still debated, the widely accepted mechanism for metal-oxide based RRAM involves the creation and rupture of conductive filaments (CFs) within the oxide layer [9]. Various factors have been proposed as the cause of these CFs, including thermal effects, ionic conduction, electronic effects, metal-insulator transitions, and ferroelectricity [10].

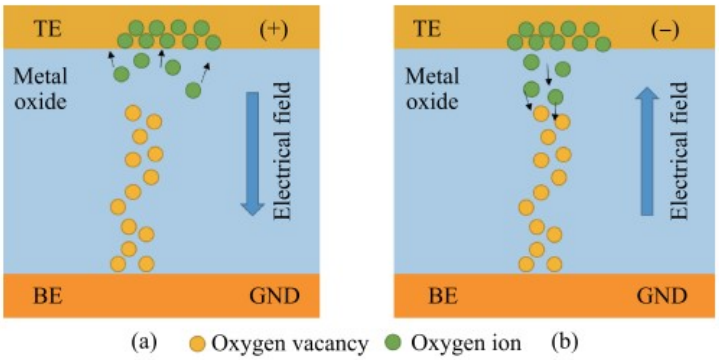


Figure 2.3: Schematic illustrations of the resistive switching processes in metal oxide-based RRAM: (a) set process and (b) reset process [3]

To provide a comprehensive explanation and guide the design of metal-oxide based RRAM devices, a unified mechanism and related models have been developed [11]. Figure 2.3 visually outlines the resistive switching process, both set and reset, within oxide-based RRAM devices. During the set process, a positive bias is applied to the top electrode (TE), generating oxygen vacancies ($V_O{}^{2+}$) in the metal oxide layer. Simultaneously, oxygen ions (O^{2-}) move towards the TE, where they are discharged or reserved in the TE metal. This accumulation of oxygen vacancies forms CFs that bridge the bottom electrode (BE) and TE, allowing electrons to move between these vacancies. Consequently, the RRAM device is in a low resistance state (LRS) after the set process. Conversely, during the reset process, a negative bias is applied to the TE, leading to the release of reserved oxygen ions, which combine with the oxygen vacancies and break the CFs. This brings the metal-oxide based RRAM device back to a high resistance state (HRS).

To provide a clear understanding, a single microscopic principle has been introduced to explain the features of both unipolar and bipolar switching in metal-oxide based RRAM. This principle establishes a connection between the type of resistive switching and the arrangement of localized oxygen vacancies within the metal oxide switching layer. This arrangement is intricately shaped by the interplay of generating and combining dissociated oxygen ions. However, ongoing discussions persist regarding the mechanism of metal-oxide based RRAM. Questions remain, such as the initiation point for the growth of conductive filaments (CFs) during the set process, as well as the precise location of their breakdown in the reset process. Additionally, understanding how these oxygen vacancies aggregate to form CFs continues to be a topic of inquiry.

2.4 RRAM Technology and 1T1R Arrays

An RRAM cell is a two-terminal device with a metal-insulator-metal (MIM) structure [12]. This setup allows conductive filaments (CFs) to form or dissolve within the cell's insulating oxide layer when voltage pulses are applied.

Most modern RRAM unit cells include an access transistor to reduce off-state current leakage, with the MIM structure typically located on the drain side of this transistor. This 1T1R unit cell configuration helps isolate individual cells for data storage applications. In this setup, the access line connected to the source of the transistor is called the source line (SL), the line connected to the top of the MIM junction is the bit line (BL), and the transistor gate is controlled by the word line (WL). The "SET" process, which lowers the cell's resistance, is done by applying a voltage pulse from BL to SL. Conversely, the "RESET" process, which increases the cell's resistance, is achieved by applying a voltage pulse from SL to BL. To read the cell's resistance, a small voltage is applied from BL to SL, and the resulting current is measured.

2.5 Opportunities and Challenges

Metal-oxide RRAM offers several advantages over other non-volatile memories, making it a strong candidate for next-generation memory systems:

- **High-Speed Operation:** Thanks to their nanoscale dimensions, metal-oxide RRAM devices provide exceptionally fast read operations.
- **Low Power Consumption:** Metal-oxide RRAM significantly reduces power consumption during read operations compared to conventional memory technologies. This makes it ideal for applications where energy efficiency and power constraints are critical.
- **High-Density Storage:** The small cell size of metal-oxide RRAM enables high-density memory configurations, allowing for more data storage within a given area. This leads to increased memory capacity and a reduced chip footprint.
- **Cost and Scalability:** Metal-oxide RRAM can be produced at a low cost using existing semiconductor manufacturing processes. Its scalability facilitates integration with complementary metal-oxide-semiconductor (CMOS) logic circuits, supporting monolithic 3-D integration and the creation of high-performance memory systems.

Metal-oxide RRAM (Resistive Random Access Memory) shows immense potential across various applications due to its unique properties:

- **Storage Class Memory:** RRAM can act as an intermediate storage class memory, bridging the gap between traditional volatile memory like DRAM and non-volatile storage such as NAND flash. With its fast access times and non-volatility, RRAM is ideal for data caching and accelerating data-intensive applications.
- **Artificial Intelligence:** The high-speed and low-power attributes of RRAM make it particularly suitable for neural network accelerators and in-memory computing architectures. These characteristics can significantly enhance the efficiency and performance of AI algorithms, facilitating real-time inference and training tasks.
- **Internet of Things (IoT):** RRAM's low power consumption, high-density storage, and compatibility with CMOS processes make it an excellent memory solution for IoT devices. It supports efficient data storage and processing in edge devices, promoting edge computing and reducing dependence on cloud services.

Despite the significant potential of RRAM as a next-generation memory technology, several challenges must be addressed for its successful implementation:

- **Variability and Endurance:** RRAM devices exhibit significant variability in their resistance switching characteristics . This variability includes differences in SET and RESET voltages, resistance levels, and switching speeds among different cells, which can impact the reliability and performance of memory arrays. Moreover, RRAM devices can degrade over time due to endurance issues, with repeated switching operations leading to resistance drift. This drift can limit the lifetime and reliability of RRAM devices.
- **Write Speed and Energy:** RRAM devices face challenges related to write speed and energy consumption.
- **Read Disturb and Write Disturb:** RRAM is susceptible to read disturb and write disturb phenomena . During read operations, unintentional switching of unselected cells can occur, leading to data corruption. Similarly, during write operations, neighboring cells may experience unwanted resistance changes, impacting data integrity. While circuit design techniques and optimization of operating conditions can mitigate these disturb effects, they remain significant challenges.

- **Scaling and Manufacturing Challenges:** As RRAM devices are scaled down to achieve higher density and improved performance, various manufacturing challenges arise. These include the uniform deposition of metal-oxide layers, precise control of interfaces, and consistent fabrication of high-quality devices. Achieving uniformity and reproducibility across large-scale production remains critical for manufacturing RRAM devices with high yield and reliability.

- **Reliability and Data Retention:** RRAM faces challenges in terms of long-term reliability and data retention . The resistance states of RRAM cells can drift over time due to factors such as relaxation, temperature variations, voltage stress, and aging effects. This can lead to potential data loss or errors. Ensuring robustness and data integrity over extended periods, especially in harsh operating conditions, is essential for the practical implementation of RRAM.

2.6 Applications and Future Trends

2.6.1 Neuromorphic Computing

The distinctive analog behavior and low power consumption of RRAM position it as a promising candidate for the realization of artificial neural networks and neuromorphic computing architectures. A significant avenue within this domain involves designing RRAM-based synapses that can efficiently emulate the behavior of biological synapses. This step is crucial in enabling energy-efficient and scalable neuromorphic systems, which have the potential to revolutionize pattern recognition, machine learning, and cognitive computing.

2.6.2 In-Memory Computing

The remarkable ability of RRAM to seamlessly combine storage and computation within the same device has opened up exciting possibilities in the realm of in-memory computing. By performing computations directly within the memory cells, RRAM can substantially reduce the need for data movement, thereby enhancing processing efficiency and accelerating tasks that involve complex data manipulation. This paradigm shift has the potential to reshape the landscape of data-intensive applications and algorithms.

2.6.3 Internet of Things (IoT) and Edge Devices

The inherent attributes of RRAM, including its low power consumption and non-volatile nature, align well with the requirements of Internet of Things (IoT) and edge devices. These devices often operate on limited energy budgets and demand fast wake-up times. RRAM-based memory solutions can play a pivotal role in extending the battery life of IoT devices while ensuring rapid data access and storage. Such applications encompass a wide spectrum, ranging from wearable devices and sensors to smart appliances and autonomous systems, where efficient memory management and responsiveness are paramount.

As the development of RRAM technology continues to progress, these applications and trends hold the potential to reshape various technological domains. The seamless integration of RRAM into cutting-edge computing paradigms like neuromorphic computing and in-memory processing, along with its compatibility with energy-efficient IoT and edge devices, showcases the versatile nature of RRAM and its ability to drive innovation across multiple sectors.

2.7 RRAM Design with Open-source Tools

In our RRAM design process, the integration of open-source tools significantly contributed to the efficiency and success of our project. We relied on a combination of software and resources that enabled us to tackle various aspects of the design process while ensuring accuracy and reliability.

To manage the schematic design, we utilized Xschem, an open-source schematic capture tool. Xschem facilitated the creation of detailed and precise schematics for our RRAM designs. Its intuitive interface and robust feature set allowed us to efficiently organize and manage complex circuit designs, ensuring a smooth transition from schematic to simulation.

For simulation purposes, we used Ngspice, a popular open-source circuit simulator. This tool allowed us to model and analyze the behavior of the RRAM device under different conditions and input signals. It helped us assess how the device responds to varying voltages, currents, and other parameters, thereby providing valuable insights into its functionality and performance characteristics.

A crucial element of our design approach involved utilizing the open-source Process Design Kit (PDK) named SkyWater 130nm. This PDK offers a standardized set of design rules, device models, and other essential components tailored for the specific manufacturing process. Leveraging the SkyWater PDK streamlined our design process, ensuring that our RRAM design adheres to the fabrication capabilities and constraints of the 130nm process technology.

Incorporating these open-source tools into our RRAM design process allowed us to leverage established resources, community-contributed knowledge, and proven methodologies. This approach not only enhances the accuracy and reliability of our designs but also promotes collaboration and innovation within the broader design and research community. By utilizing tools like Ngspice, Xschem, and the SkyWater PDK, we ensured a comprehensive and efficient design workflow that aligns with current industry standards and practices.

2.8 Conclusion

In recent years, RRAM has undergone rapid advancement, spanning from impressive device performance to successful demonstrations of 3D integration. This progress underscores RRAM's substantial potential in applications such as high-density memory and nonvolatile logic. However, a series of challenges persists on the path to making RRAM suitable for practical implementation. Gaining a comprehensive understanding of the microscopic conduction and switching processes is imperative to unveil the intricate nature of resistive switching phenomena. Deliberate optimization and meticulous control of RRAM device characteristics are essential to tailor them to diverse application requirements.

Selecting a high-performance device is crucial for the efficient operation of RRAM arrays. Significant technological advancements are needed in selector development to guarantee the reliable functioning of large-scale crossbar RRAM arrays. With rapid progress in the field, RRAM is anticipated to play a key role in future technologies, greatly enhancing high-density storage and nonvolatile logic applications.

Chapter 3

Design

3.1 Introduction

Metal-oxide resistive RAM (RRAM) is a cutting-edge non-volatile memory technology with the capability to store multiple bits within a single cell by adjusting the cell's resistance to various levels. This feature positions RRAM as a potential high-density embedded memory solution for a wide range of low-power and intermittent computing applications. However,to utilize RRAM effectively in these scenarios, it is crucial to minimize read/write latency and energy consumption. Achieving this requires the implementation of efficient readout and programming techniques.

3.2 RRAM Technology and 1T1R Characteristics

The RRAM component is accurately modeled using a Verilog-A file. Analysis of its characteristics reveals that RRAM has distinct set and reset voltages, as shown in Figure 3.1. From this figure, we observe that the set operation occurs at approximately 1.52V and 350μA, while the reset operation occurs at around -1.2V and 200μA. To handle the excessive current flow after reaching the set voltage, a compliance circuit is necessary. Consequently, each RRAM unit cell is equipped with an access transistor to control the excess current and mitigate off-state current leakage. Additionally, a MIM (metal-insulator-metal) structure located on the drain side of the access transistor stores information in the form of cell resistance. This 1-transistor 1-resistor (1T1R) configuration is a standard approach for data storage.

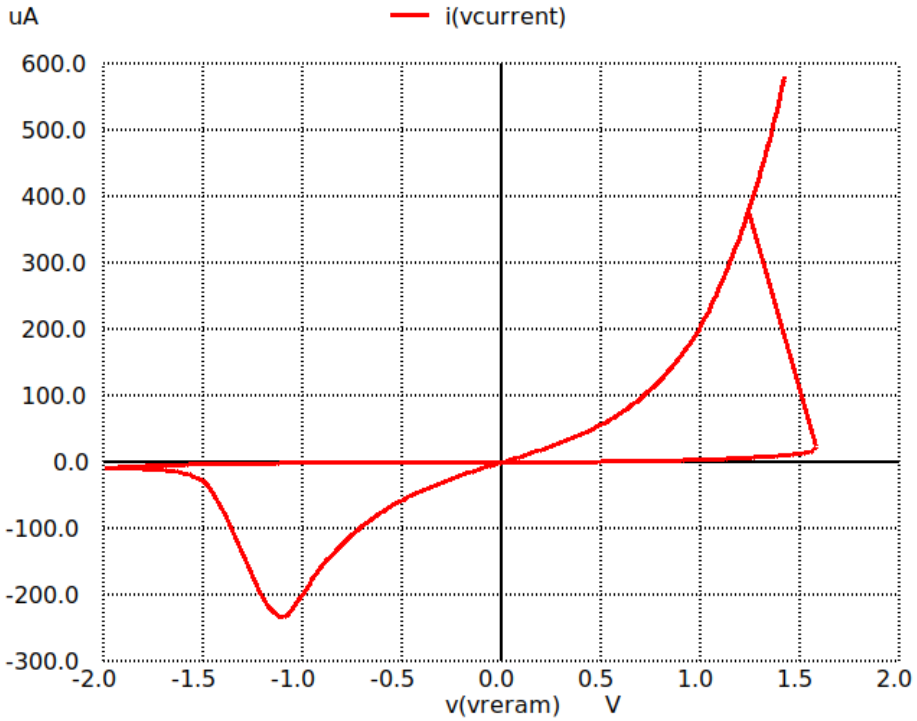


Figure 3.1: RRAM Characteristics

Figure 3.2 presents the schematic of a 1-transistor-1-resistor (1T1R) RRAM cell, highlighting its three terminals: bitline (BL), wordline (WL), and source line (SL). The voltages at these terminals serve as ”knobs” to adjust the cell resistance to a desired range. The wordline voltage (VWL) regulates the voltage across the RRAM cell and functions as a current limiter. Meanwhile, the bitline voltage (VBL) and source line voltage (VSL) manage the potential across the entire 1T1R structure, including both the transistor and the RRAM cell. The 1T transistor plays a crucial role in preventing write disturbances on unselected cells.

Each pulse can be either a SET pulse (to decrease cell resistance) or a RESET pulse (to increase cell resistance). For a SET operation, a positive bias is applied to VBL while VSL is grounded. Conversely, for a RESET operation, a positive bias is applied to VSL while VBL is grounded. Initially, the RRAM is in a High Resistance State (HRS). To transition the RRAM cell from HRS to a Low Resistance State (LRS), representing a ”1” (SET condition), a voltage of 1V is applied to the WL, while the BL voltage is varied and the SL is grounded. To perform a RESET operation, the WL voltage is set to approximately 2.5V, the BL is grounded, and the SL voltage is varied.

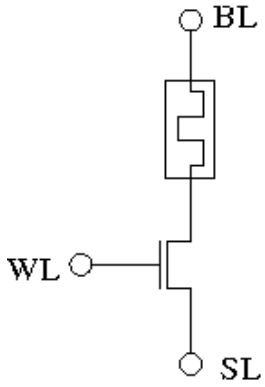


Figure 3.2: Schematic of 1T1R Cell

Figure 3.3 illustrates the characteristics of a 1-transistor-1-resistor (1T1R) RRAM cell. The voltages associated with SET, RESET, and read conditions are detailed in Table 3.1.

Operation	WL (Word Line) Voltage	BL (Bit Line) Voltage	SL (Source Line) Voltage
SET	1V	1.7V	0V
RESET	2.5V	0V	1.5V
Read	1/2.5V(LRS/HRS)	0.5V(for LRS)	0.5V(for HRS)

Table 3.1: Voltages for SET, RESET, and Read Conditions

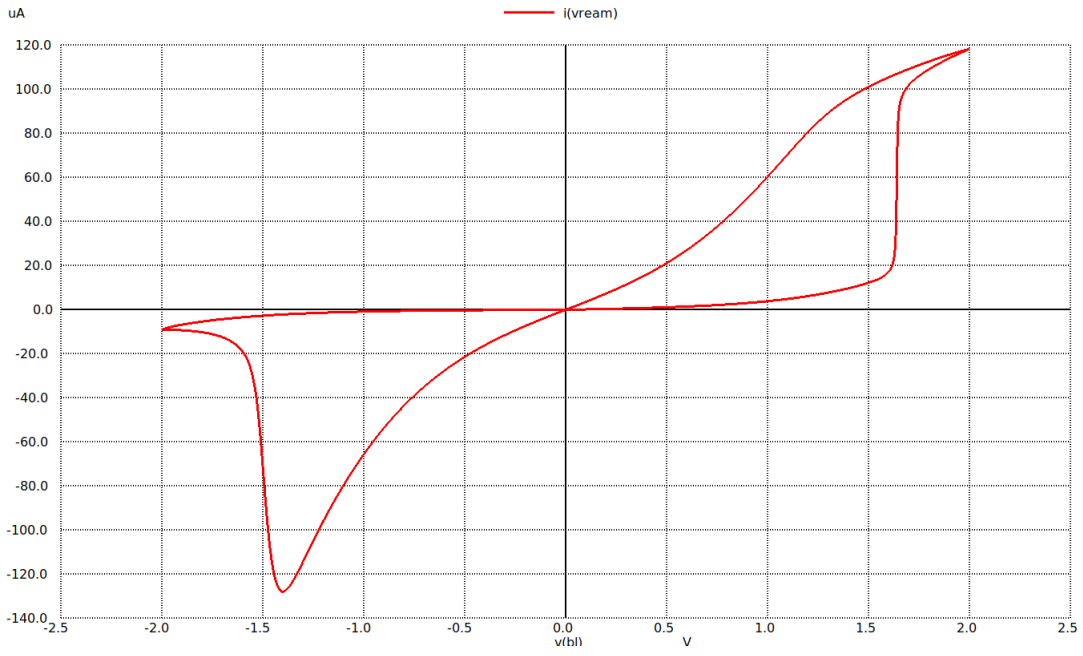


Figure 3.3: Characteristics of 1T1R Cell

Figure 3.4 illustrates the characteristics of a 1-transistor-1-resistor (1T1R) RRAM cell for various VWL values, ranging from 0.8V to 1.2V. From Figure 3.3, it is evident that the current value corresponding to the SET condition of the 1T1R cell is approximately 120μA, which is roughly one-third of the current flowing when the transistor is not present.

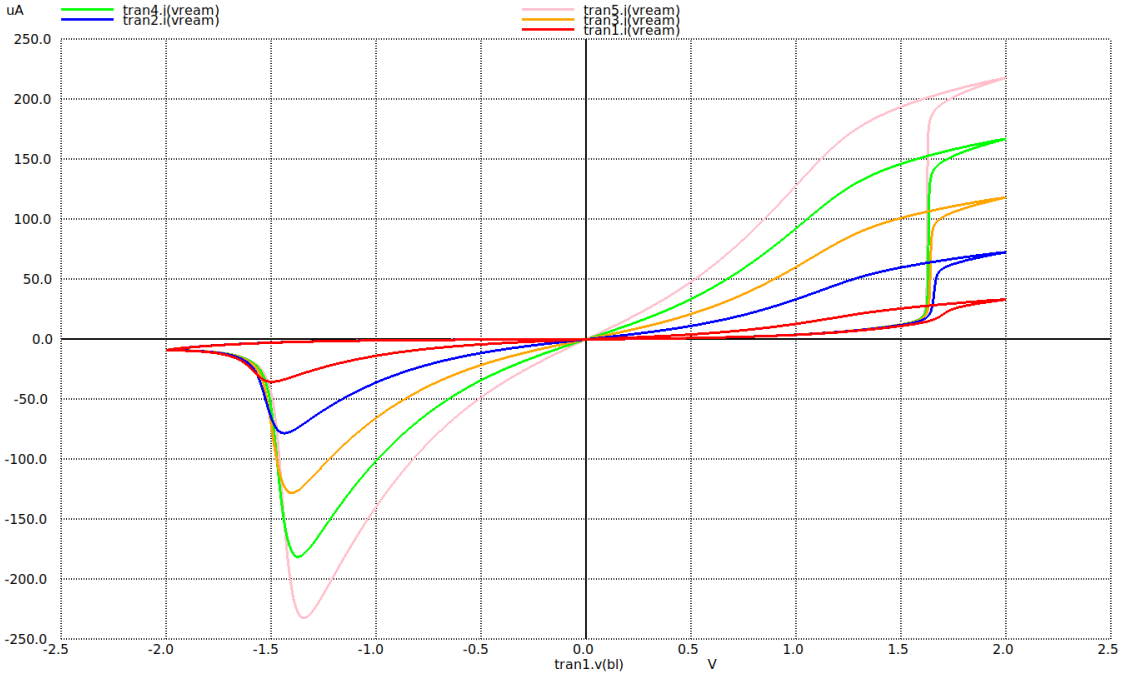


Figure 3.4: Characteristics of 1T1R Cell for different VWL(0.8V to 1.2V)

3.3 RRAM : A Non-Volatile Memory

A memory is said to be non-volatile when it retains stored data even when the power supply is turned off. Unlike volatile memory, which loses its content when power is lost, non-volatile memory ensures data persistence and is used for long-term storage. Examples of non-volatile memory include flash memory, ROM (Read-Only Memory), EEPROM (Electrically Erasable Programmable Read-Only Memory), and emerging technologies like RRAM (Resistive Random Access Memory).

Figure 3.5 illustrates the non-volatile characteristics of RRAM. During the period from 2 to 3μs, we set the RRAM cell by applying 1.7V to VBL, 0V to VSL, and 1V to VWL. We then read the RRAM cell immediately between 3 and 4μs by applying 0.5V to VBL, 0V to VSL, and 1V to VWL, observing a current of approximately 18μA corresponding to VBL. The RRAM

cell was powered off between 4 and 5μs. When we read the RRAM cell again between 5 and 6μs with the same voltages (0.5V to VBL, 0V to VSL, and 1V to VWL), the current remained around 18μA. We repeated this cycle, powering off the cell between 6 and 7μs and reading it between 7 and 8μs, consistently observing the same current. This behavior confirms the non-volatile nature of RRAM, demonstrating that once the RRAM is set to write a ”1” or ”0”, it retains the value even when the power is turned off.

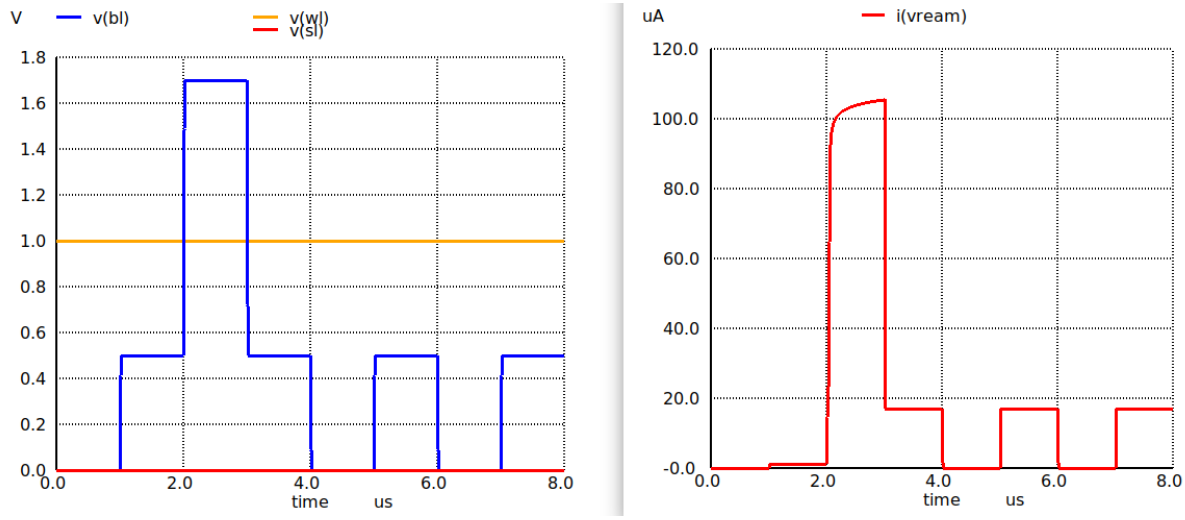


Figure 3.5: RRAM : A Non-Volatile Memory

3.4 Sensing Circuit 1 : The Initial Design

The ARYABHAT Analog Processor operates in the sub-threshold region, where currents are in the range of nA. However, we have observed that the currents for RRAM are in the range of μA. Therefore, RRAM cannot be directly used to pump currents in this processor. To address this issue, we need to read the RRAM current and compare it with a reference current using a current comparator. Based on the comparison, we can latch the result as either ”0” or ”1”, corresponding to the reset or set state of the RRAM cell. After latching to either ”0” or ”1”, a DAC interface can be used to convert the current to the nA range required by the ARYABHAT Analog Processor. Figure 3.6 shows the proposed block diagram for sensing the RRAM current.

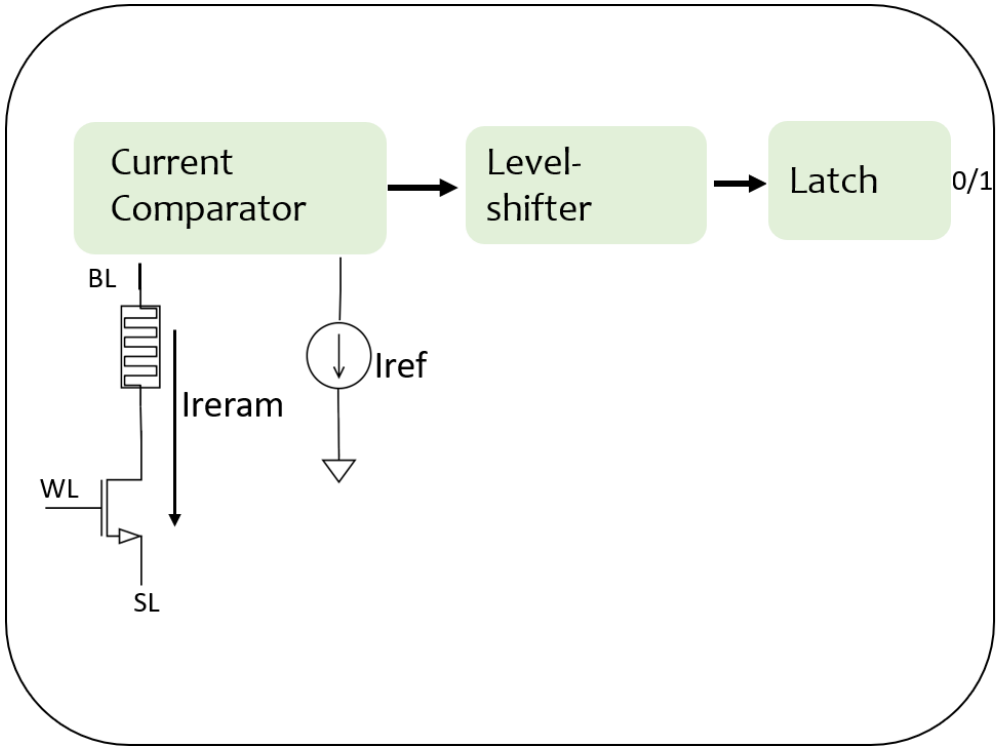


Figure 3.6: Proposed block level to sense the RRAM Current

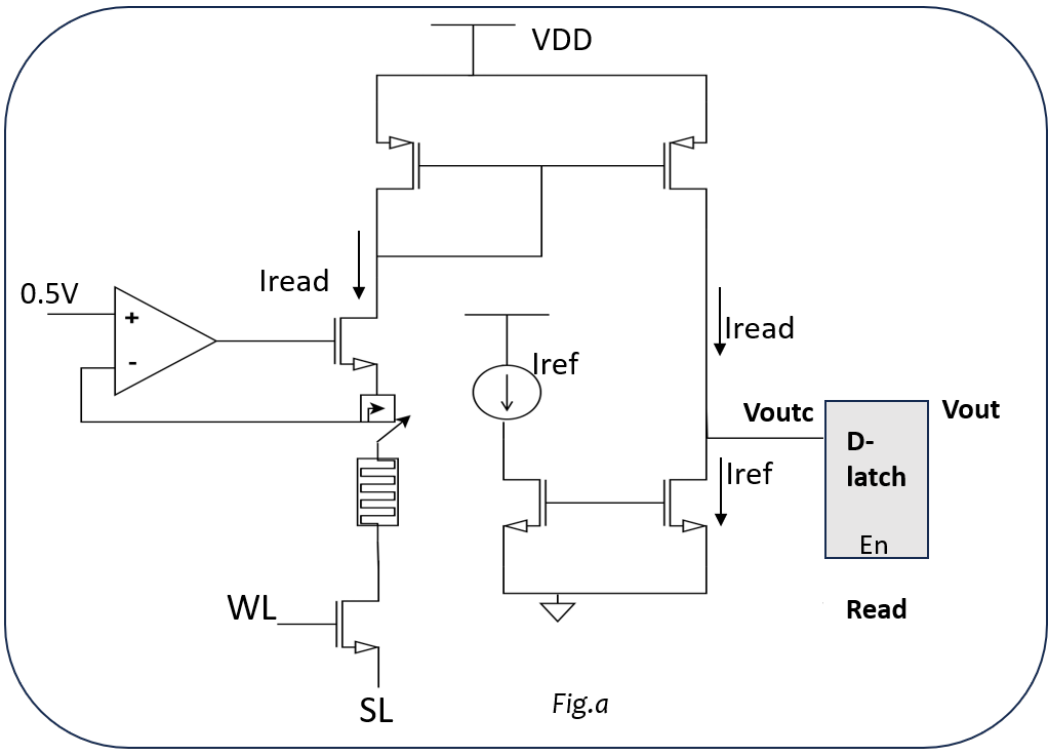


Figure 3.7: Circuit for reading the RRAM Current

As shown in Figure 3.6, it is not straightforward to mirror the RRAM current because the knobs of the 1T1R are not accessible. Therefore, we used the circuit depicted in Figure 3.7. This circuit employs a MUX with a select line connected to the Read signal. Based on the Read signal, the RRAM's TE is either connected to the BL voltage or the read voltage. An Op-amp, common to the entire array, is used to connect the read voltage to the TE of the RRAM. This setup allows us to read the current and compare it with a reference current (typically chosen between the set current and reset current). The result is then latched to either "1" or "0" based on whether the RRAM is set or reset, respectively.

Figure 3.8 presents the simulation results where we observed the RRAM currents during the read operations. In these simulations, a BL pulse was applied to set the RRAM (writing a "1"), and an SL pulse was applied to reset the RRAM (writing a "0"). Figure 3.9 presents the simulation results, showing the read pulses to select 0.5V and the corresponding voltages at the comparator node, V(Voutc), where the reference current and RRAM current are compared.

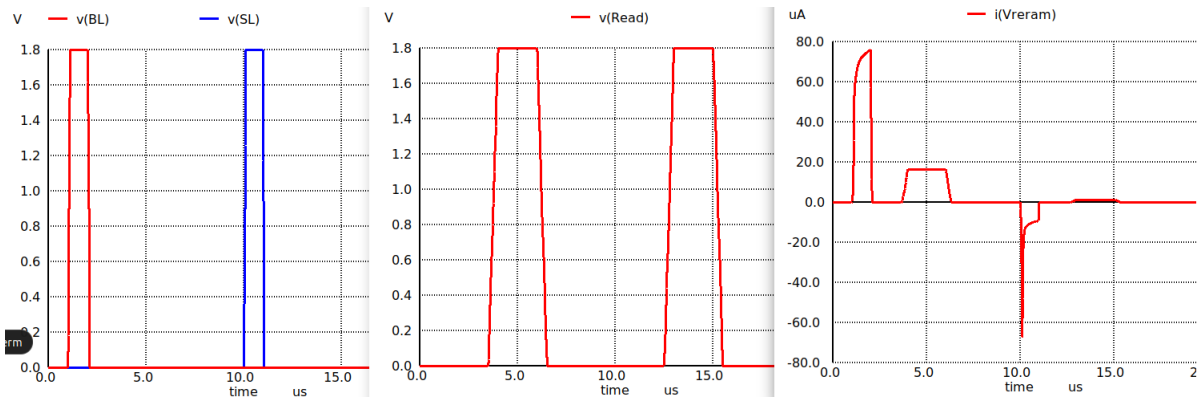


Figure 3.8: i) BL and SL pulses for writing ii) Read pulse to select 0.5V and to read currents iii) RRAM Current

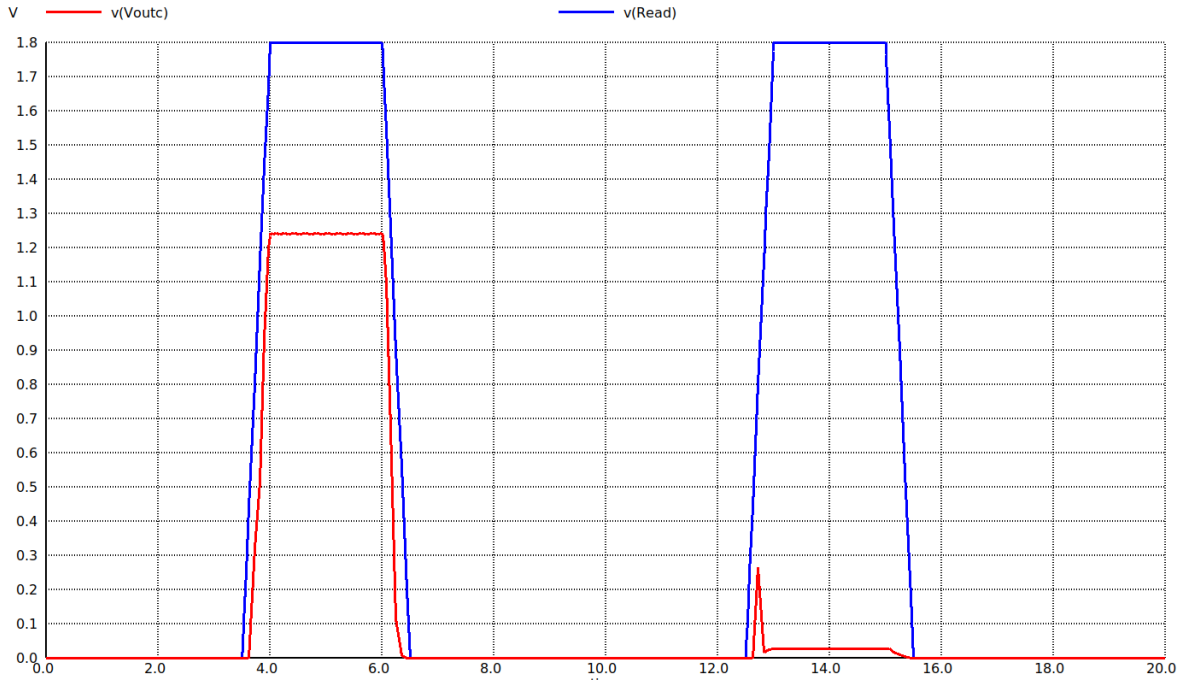


Figure 3.9: i) Read pulse to select 0.5V and to read currents ii) Node voltage at the Comparator

To ensure proper latching of the comparator output, it's crucial to generate enable pulses (Readp and Readpb) accurately during the read operation. To achieve this, we employed the circuit depicted in Figure 3.10, which is designed to be common to the entire array. This circuit facilitates the generation of latch enabling pulses using the Read pulse. Additionally, we took measures to ensure that the enable pulses are generated for only a short duration, thus effectively reducing power consumption. Figure 3.11 showcases the simulation results, providing a visual representation of the latch enable pulses in relation to the comparator output.

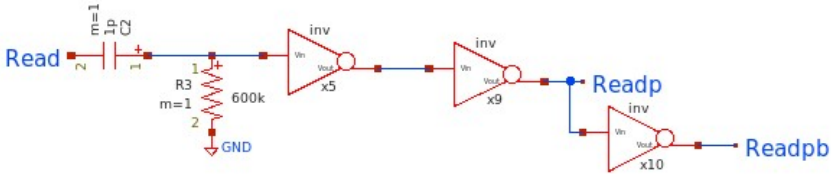


Figure 3.10: Circuit to generate Latch enable pulses w.r.t comparator output

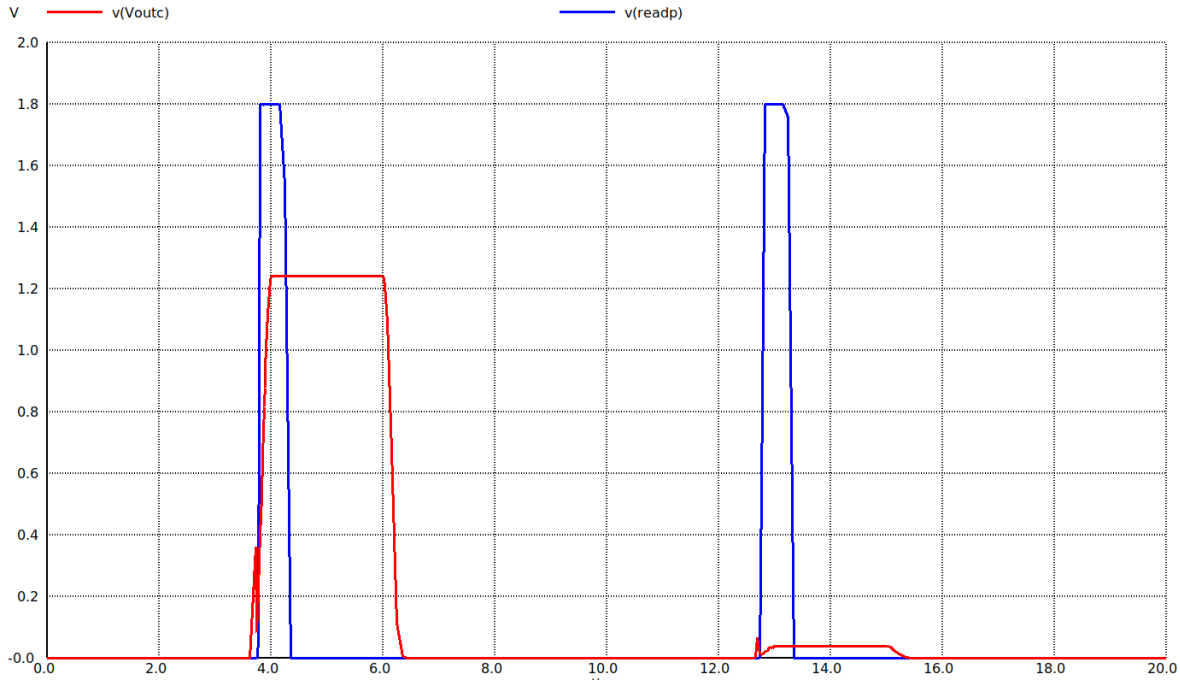


Figure 3.11: Latch enable pulses w.r.t comparator output

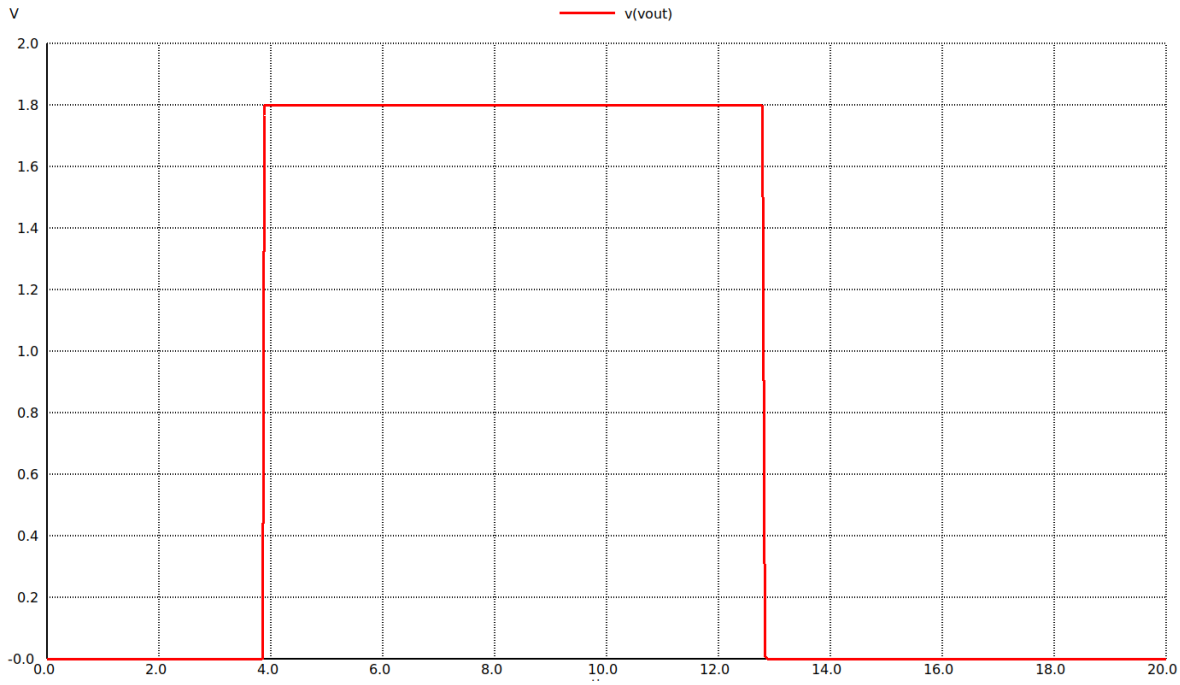


Figure 3.12: Latch Output

Figure 3.12 presents the final latch output. It is clear that when the RRAM is in the Set state and a read pulse is applied, the latch output remains stable at 1.8V (Vdd), representing a ”1”.

Conversely, when the RRAM is in the Reset state and a read pulse is applied, the latch output remains stable at 0V, representing a ”0”.

However, the circuit shown in Figure 3.8 lacks robustness. Mirroring the current twice can lead to incorrect data readings. To address these issues, a differential circuit is proposed in the next section, ensuring greater robustness and reliability.

3.5 Sensing Circuit 2 : Proposed Differential Circuit

3.5.1 Reading Mechanism

Figure 3.14 illustrates the read path and circuit structure of the proposed sensing circuit [13] [14]. In this figure, M1 and M2 are designed to provide precharge current, while M3 to M6 are used for switching the sensed currents. A CMOS latch is utilized to further amplify the sensing result and store the sense amplifier (SA) output. The detailed read operation of the circuit can be divided into three phases: (i) Precharge phase, (ii) Amplification phase, and (iii) Latching phase. Figures 3.15 to 3.17 show the circuits corresponding to the Precharge, Amplification, and Latching phases, respectively. The latch block used in the Figures 3.15 to 3.17 is shown in Figure 3.13.

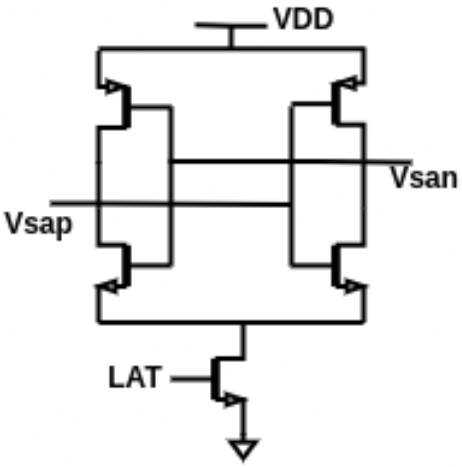


Figure 3.13: Schematic of the Latch block

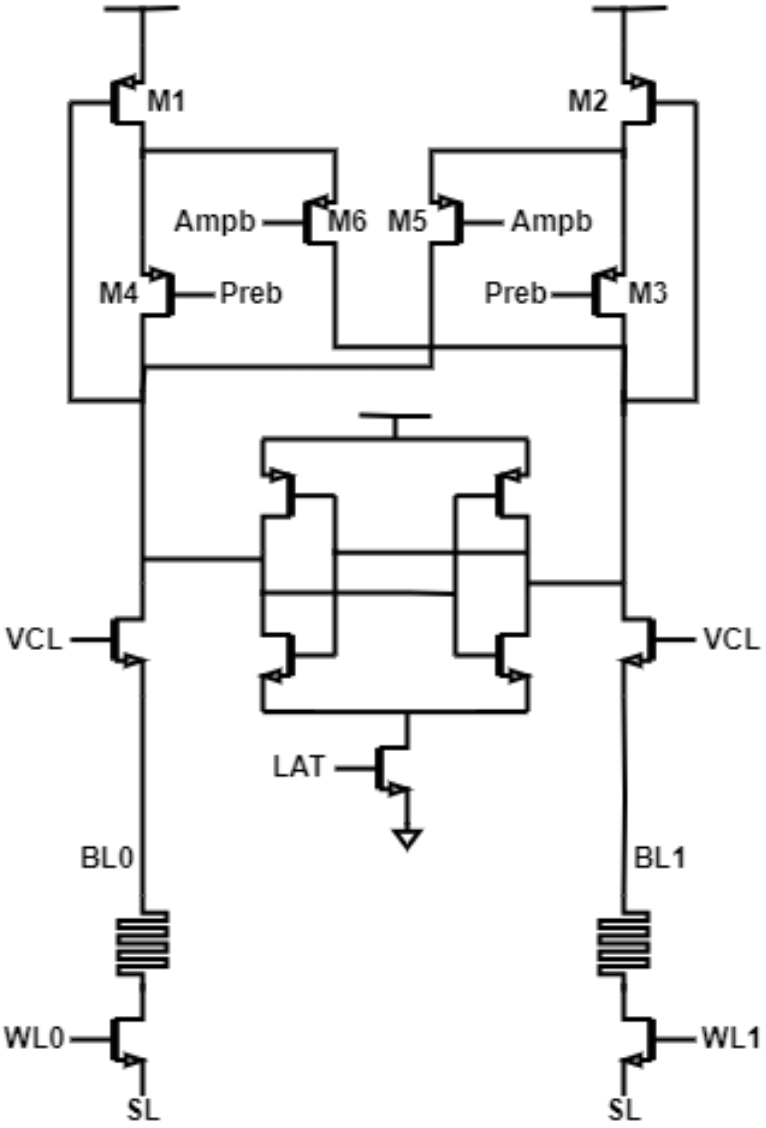


Figure 3.14: Sense Amplifier

Figure 3.15 shows the precharge phase and its corresponding switch connections. During this phase, switches S1 and S2 are turned on. After sufficient precharge time, the diode-connected transistors M1 and M2 facilitate the currents flowing according to the state of the cells, irrespective of any mismatch or process variations between M1 and M2. If the RRAM cell corresponding to WL0 has a current of I_{set} and the RRAM cell corresponding to WL1 has a current of I_{reset}, then during the precharge phase, the current across M1 will be I_{set} and the current across M2 will be I_{reset}.

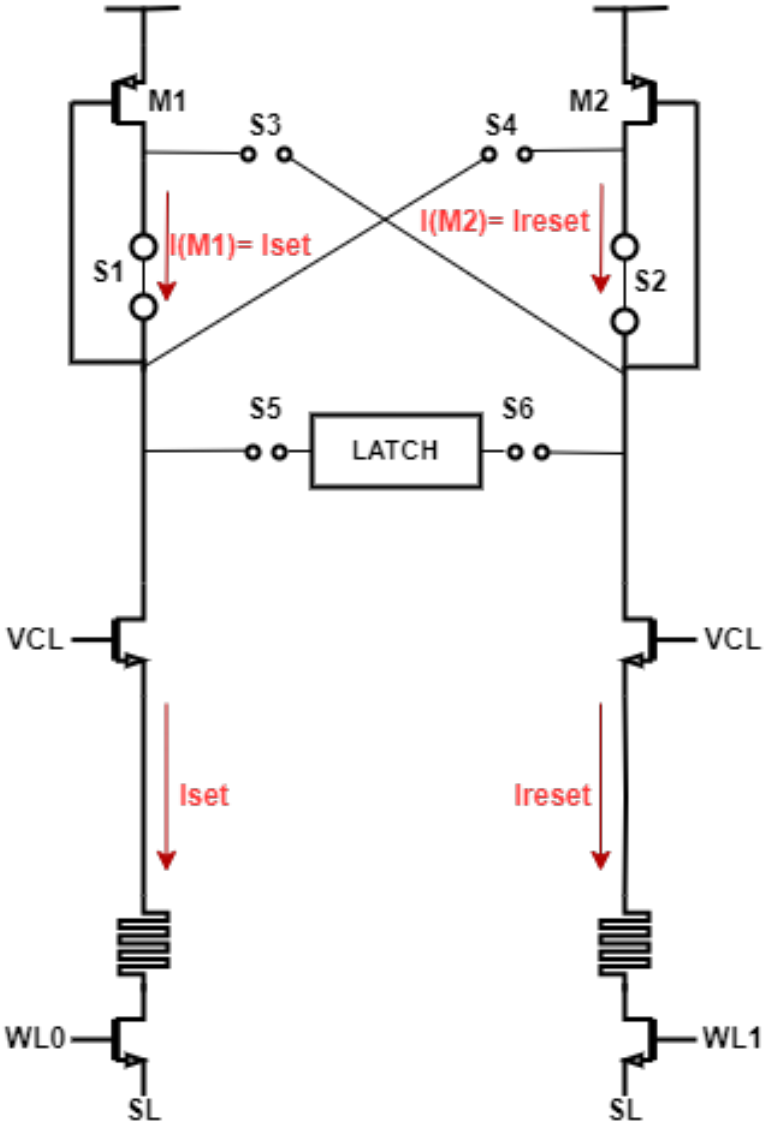


Figure 3.15: Phase 1 :- Precharge

Figure 3.16 shows the amplification phase and its corresponding switch connections. At the beginning of the amplification phase, S1 and S2 are turned off, while S3 and S4 are simultaneously turned on. The voltages retained on the parasitic capacitances of M1 and M2 are minimally affected by the switching. Thus,, I_{M1} (I_{M2}) remains close to I_{set} (I_{reset}) at the start of this phase. Next, node Q (Q1) is charged by I_{M2} (I_{M1}) but discharged by I_{set} (I_{reset}). Consequently,V_Q (V_{Q1}) is shifted by the net current I_{M2}-I_{set} (I_{M1}-I_{reset}). The sensing margin is doubled compared to the previous circuit shown in Figure 3.8 (Section 3.4), as we are now comparing the difference between I_{set} and I_{reset} instead of some reference value [(I_{set}+ I_{reset})/2], as explained in Section 3.4.

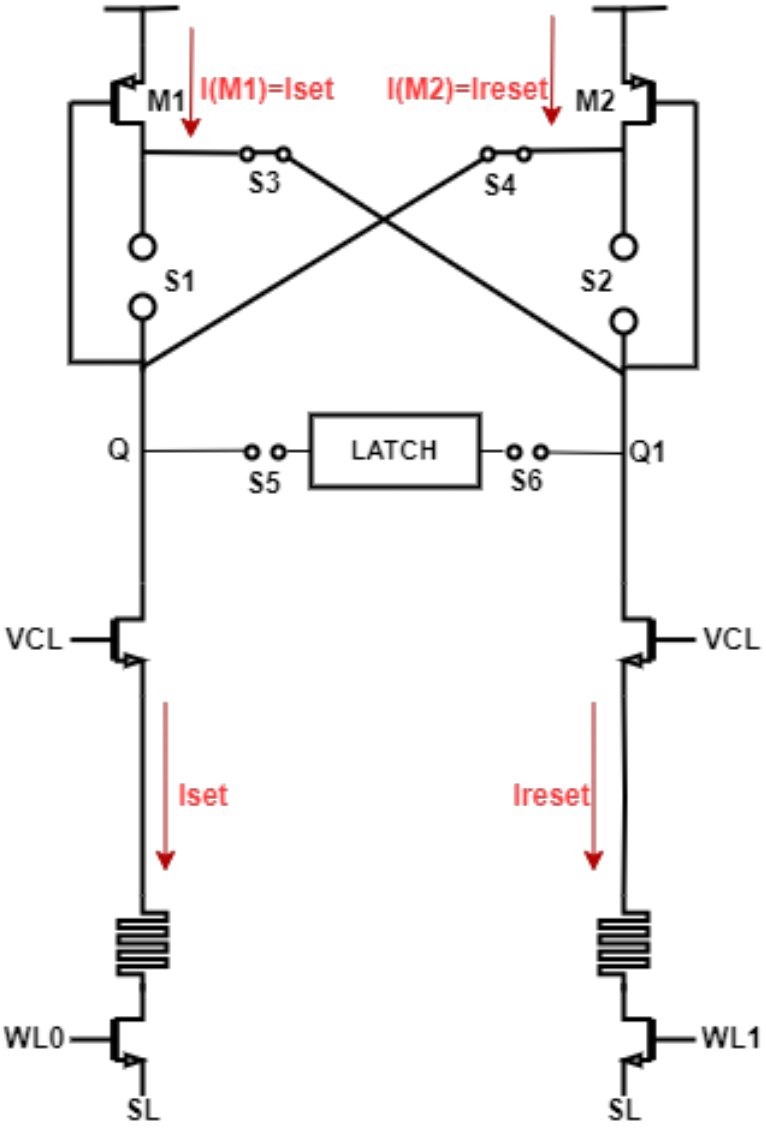


Figure 3.16: Phase 2 :- Amplification

Figure 3.17 shows the latching phase and its corresponding switch connections. During this phase, switches S5 and S6 are turned on to enable the CMOS latch. The CMOS latch requires a very short latching time to further amplify the voltage difference between nodes Q and Q1 ($\Delta V = |V_Q\text{-}V_{Q1}|$) and generate the sensing output.

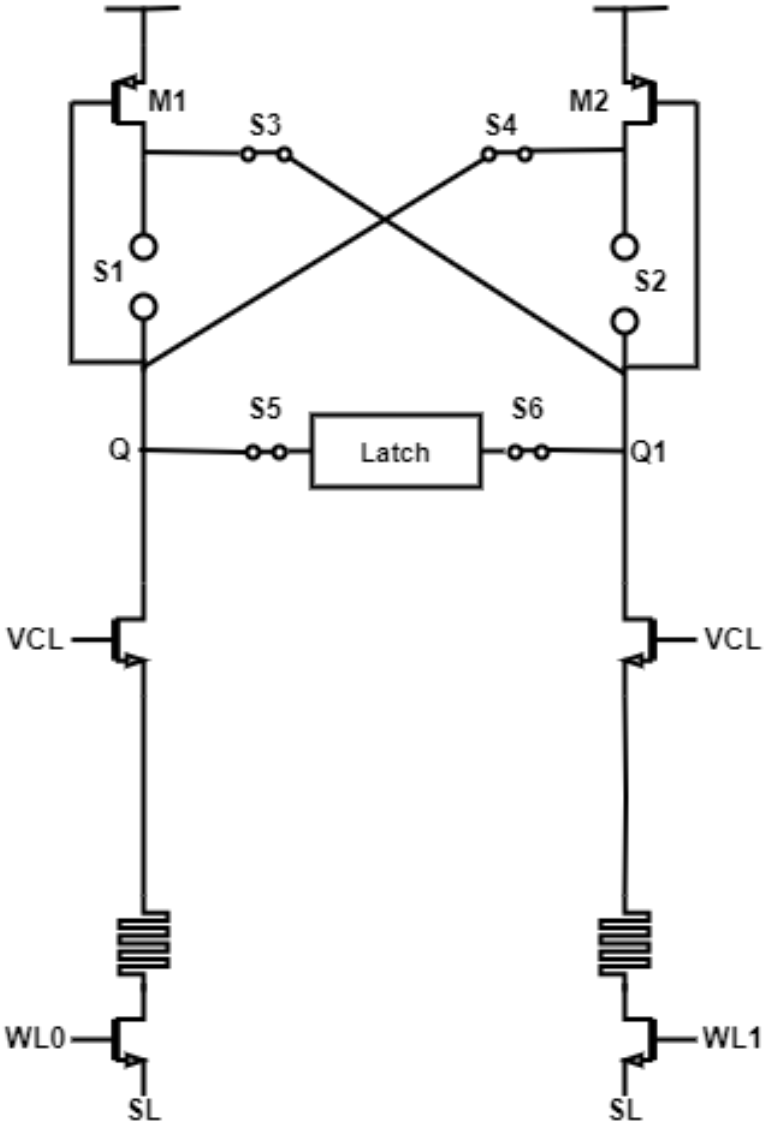


Figure 3.17: Phase 3 :- Latching

The total read sensing time t_{SEN} can be expressed as:

$$t_{SEN} = t_{PRE} + t_{AMP} + t_{LAT} \tag{3.1}$$

Where t_{PRE}, t_{AMP} and t_{LAT} represent the time durations of the precharge, amplification, and latching phases, respectively.

3.5.2 Writing Mechanism

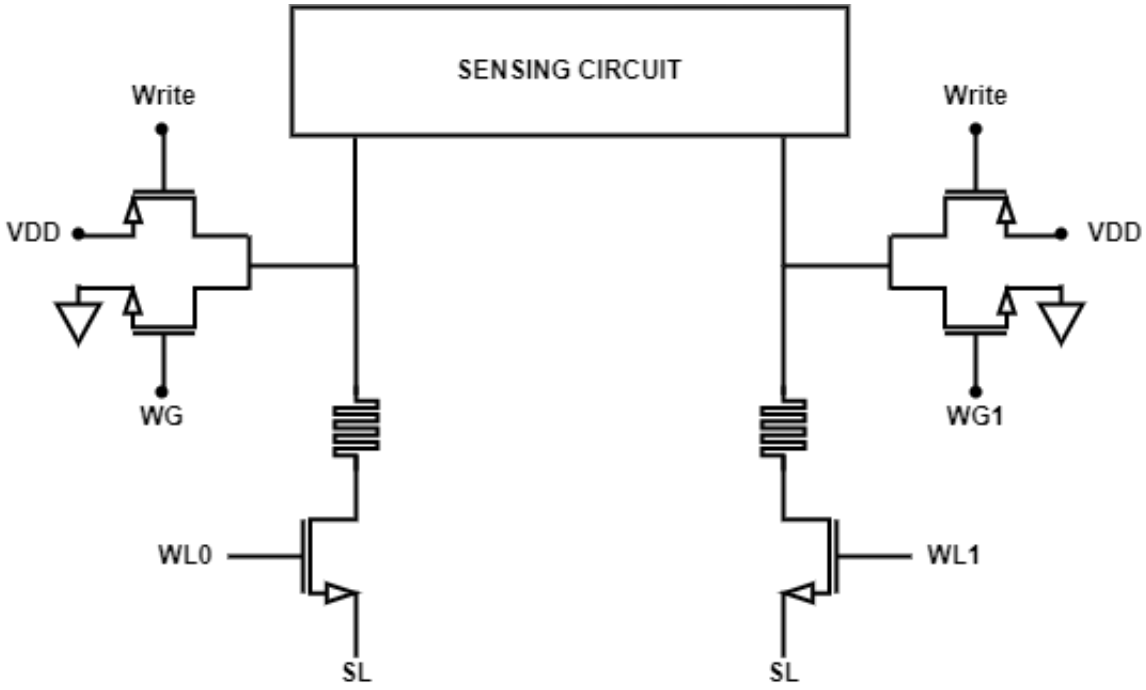


Figure 3.18: Write Circuit

Figure 3.18 illustrates the write path and circuit structure of the proposed writing circuit. The write signal is active low, so it is activated by setting it low when we intend to set the RRAM cell. In contrast, WG and WG1 are active high signals used to reset the RRAM cell. In the above circuit, setting an RRAM cell is a single-cycle operation where the cell is directly set. However, resetting an RRAM cell involves a two-cycle operation: first, the cell is set, and then it is reset. The writing process is completely isolated from the reading (sensing) circuit by switches positioned between them.

Chapter 4

Engineering & Results

4.1 Single bit cell using RRAM

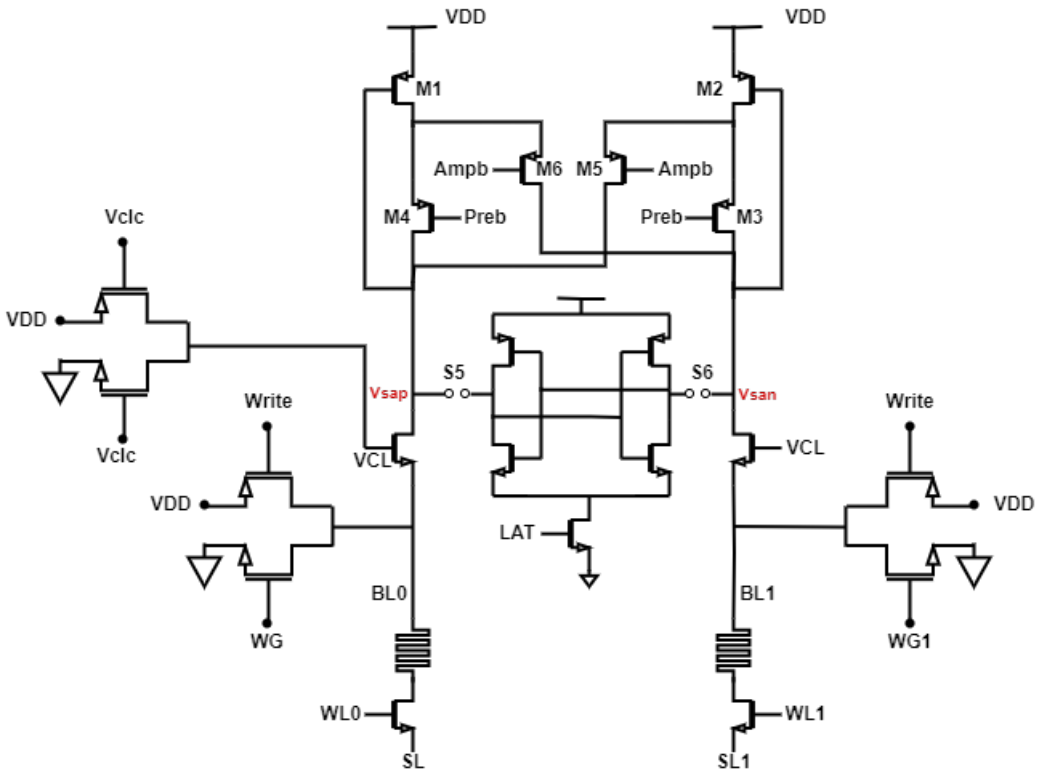


Figure 4.1: Single bit cell using RRAM

Figure 4.1 shows the overall circuit diagram of the single bit cell, which we proposed in Section 3.5. We used Xschem and NgSpice for the schematic design and to verify the simulations. To set the RRAM cell corresponding to WL0 and reset the RRAM cell corresponding to WL1, as discussed in Section 3.5, setting is a single-cycle operation while resetting is a two-cycle operation, where the cell is initially set and then reset. Additionally, during the write operations, the read (sensing) circuit is completely isolated from the write circuit by the NMOS pass transistors in between, which can be clearly observed in the circuit shown in Figure 4.1.

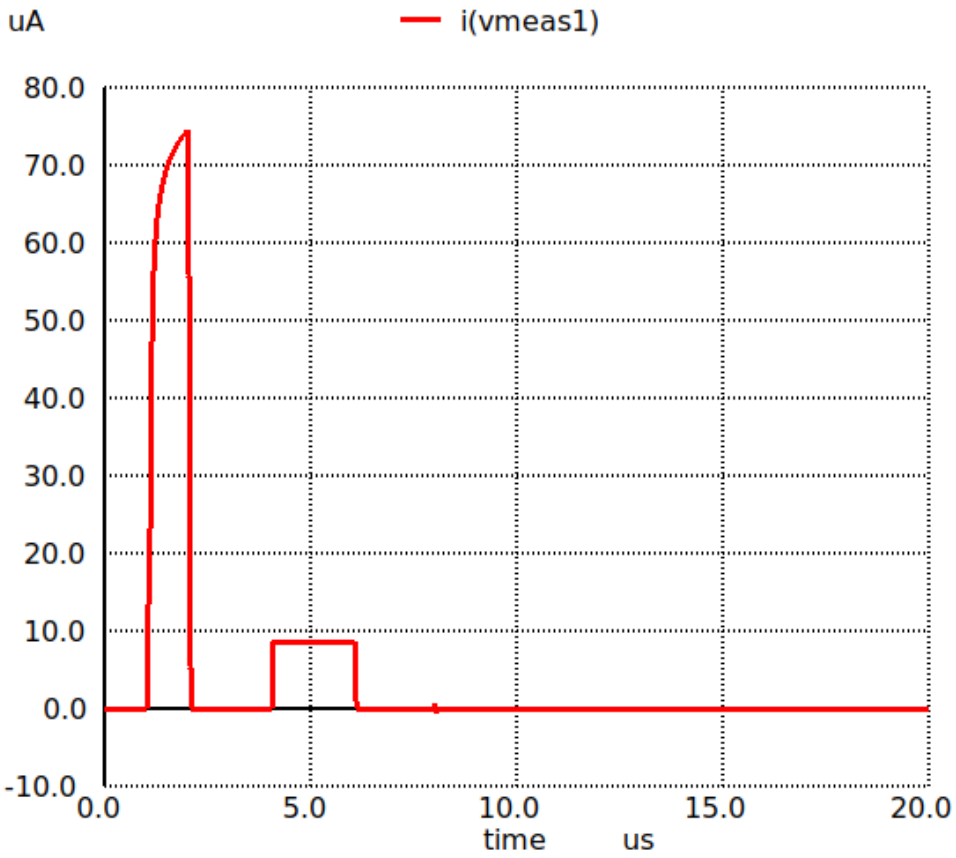
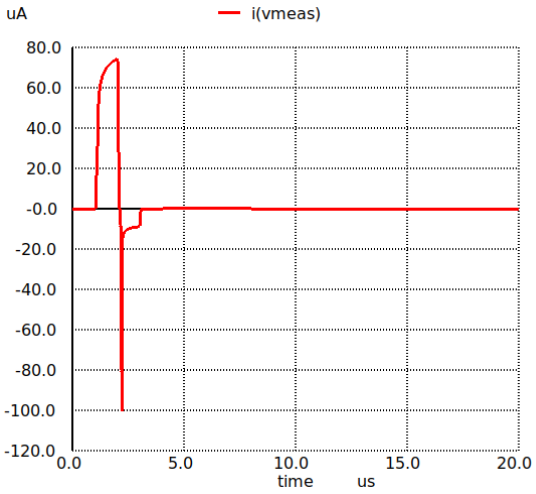
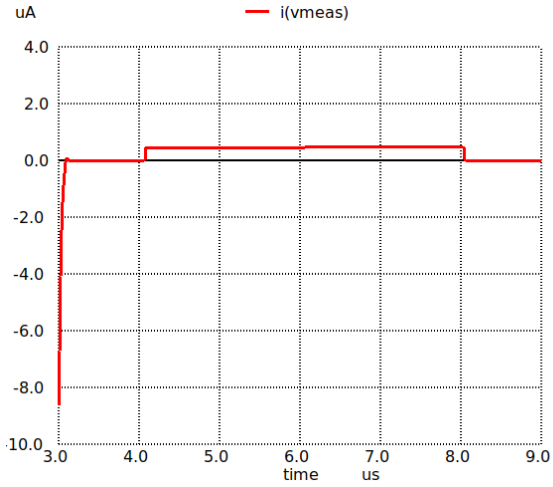


Figure 4.2: Set the RRAM Cell

Figure 4.2 illustrates the set currents during the writing and reading operations of the RRAM cell corresponding to WL0. The write signal is made active low while setting the RRAM cell, and the WG signal is also kept active low during this process. The current across the cell was measured to be 75μA during the writing operation and approximately 8μA during the reading operation. As discussed in Section 3.5, the reading process is completely isolated from the write operation and is divided into three phases: Precharge, Amplification, and Latching.



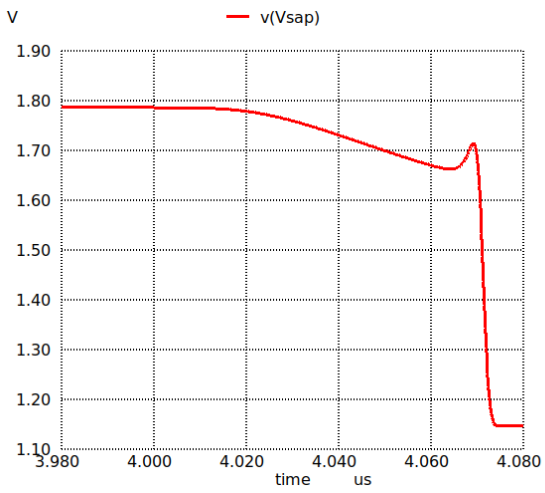
(a) Reset the RRAM Cell



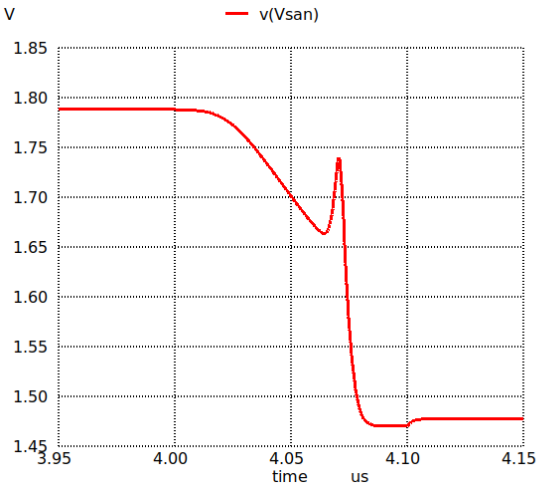
(b) Read current across the cell

Figure 4.3: Currents across the cell during Reset

Figure 4.3(a) and 4.3(b) illustrate the reset currents during the writing and reading operations of the RRAM cell corresponding to WL1. As discussed in Section 3.5, the reset process involves two cycles, where the RRAM cell corresponding to WL1 is first set and then reset. The write signal and WG1 signal must be made active low while setting the RRAM cell, and then change the write and WG1 signals to active high to reset the cell corresponding to WL1. In Figure 4.3, it is evident that we set the cell first and then perform the reset operation before reading the cell. The currents across the cell were measured to be 75μA, -100μA, and 0.4μA respectively during the writing, resetting, and reading operations.



(a) Precharge phase time for Set



(b) Precharge phase time for Reset

Figure 4.4: Precharge phase time

Figure 4.4 illustrates the time taken for the precharge phase. As discussed in Section 3.5, the reading process is divided into three phases, with the first phase being the precharge phase. During this phase, the currents across the diode-connected PMOS transistors M11 and M13 in the circuit shown in Figure 4.1 attempt to mimic the currents of the RRAM cells corresponding to WL0 and WL1. From Figures 4.4(a) and 4.4(b), it is clear that the precharge phase time is around 70 ns for the set condition and 0.1 μs for the reset condition. It is evident that, as the current is lower across the RRAM cell in the reset condition, it takes more time to mimic compared to the RRAM cell in the set state. Therefore, in the worst case, the precharge phase needs to be at least 0.1 μs.

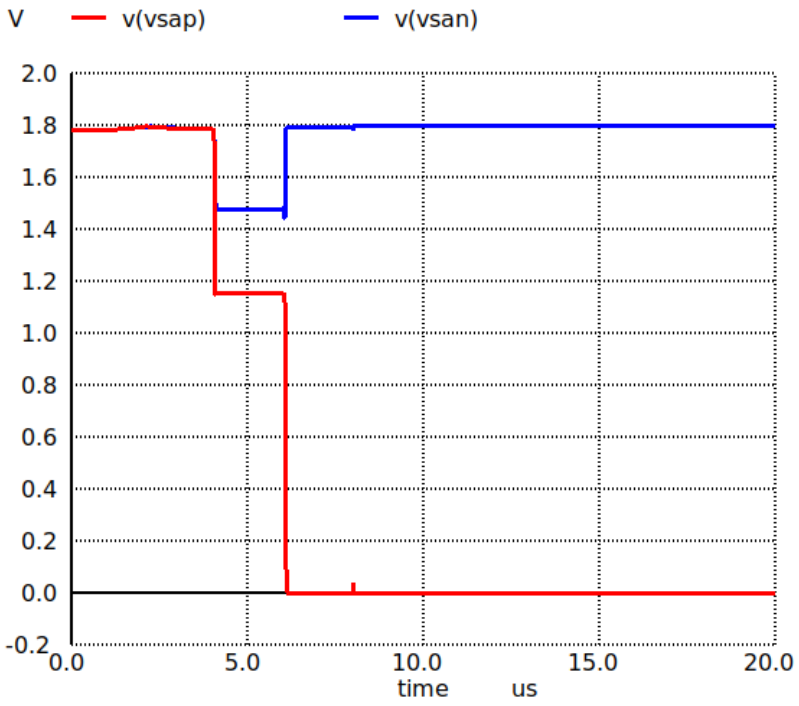


Figure 4.5: Voltages at the Nodes after Amplification

The second phase of the reading operation is the Amplification phase. During this phase, the voltages across the parasitic capacitances of M11 and M13 ensure that the currents flowing through them remain almost the same as those in the precharge phase, which correspond to the currents flowing through the RRAM cells associated with WL0 and WL1. Figure 4.5 shows the voltages at nodes V_{sap} and V_{san} as depicted in the circuit in Figure 4.1.

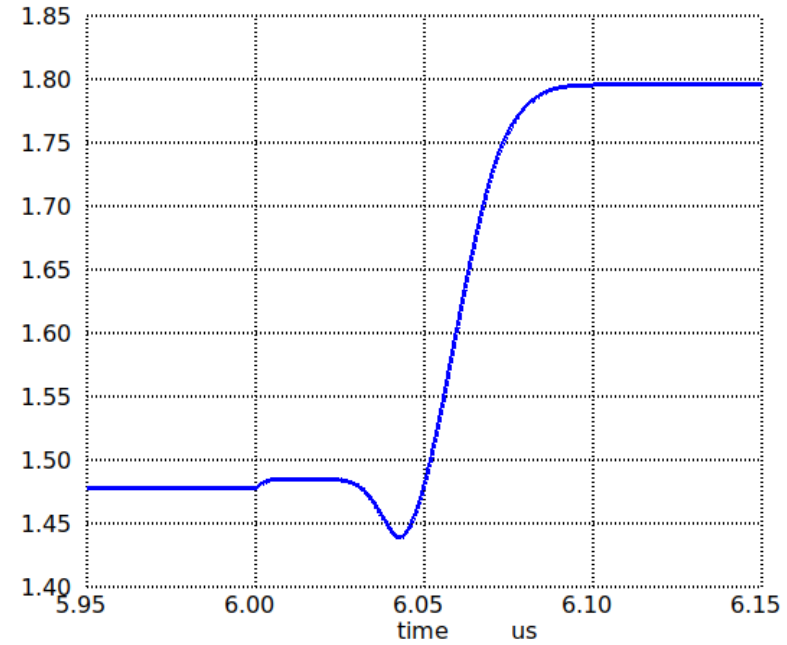


Figure 4.6: Amplification of 1

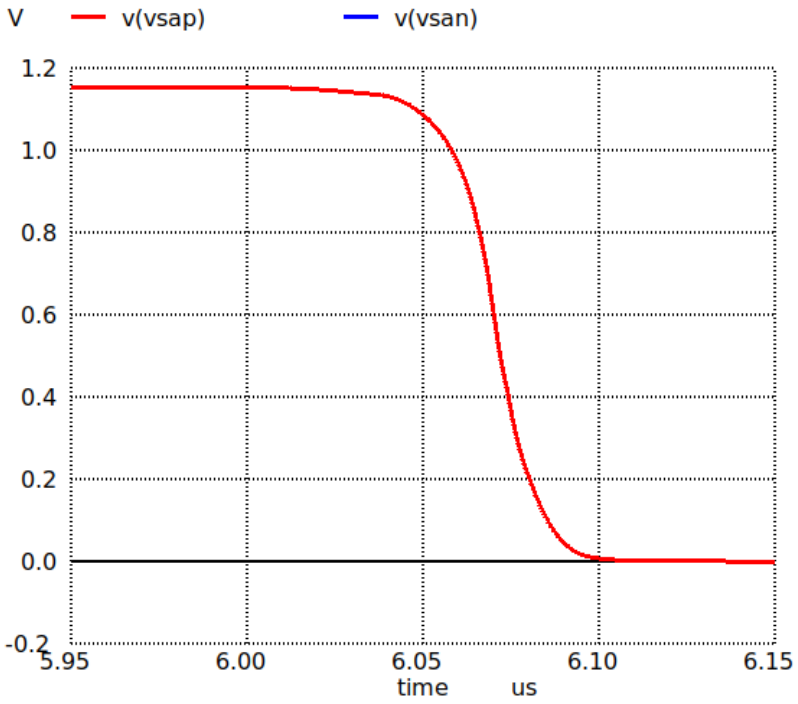


Figure 4.7: Amplification of 0

Figure 4.6 illustrates the amplification of 1 at node V_{san} and the amplification of 0 at node V_{sap}. At node V_{san}, the amplification of 1 occurs in 0.1 μs, and at node V_{sap}, the amplification of

0 also occurs in 0.1 μs. Therefore, the minimum time required for the amplification phase is 0.1 μs. The third phase of the reading operation is the Latching phase. During this phase, the voltage difference between the nodes V_{san} and V_{sap} is used to generate the sensing output in a very short latching time, as the nodes are already close to V_{DD} and 0V. Additionally, the read path to the RRAM cells is disconnected during this phase to reduce power consumption.

The total read sensing time t_{SEN} can be expressed as:

$$t_{SEN} = t_{PRE} + t_{AMP} + t_{LAT} \tag{4.1}$$

where t_{PRE}, t_{AMP} are 0.1μs, 0.1μs respectively and the latching time is negliglible.

$$t_{SEN} = 0.1 + 0.1 + 0$$
$$t_{SEN} = 0.2\ \mu\text{s}.$$

So the total read sensing time of the single bit RRAM cell is **0.2 μs**.

4.1.1 Energy and power across reading phases:

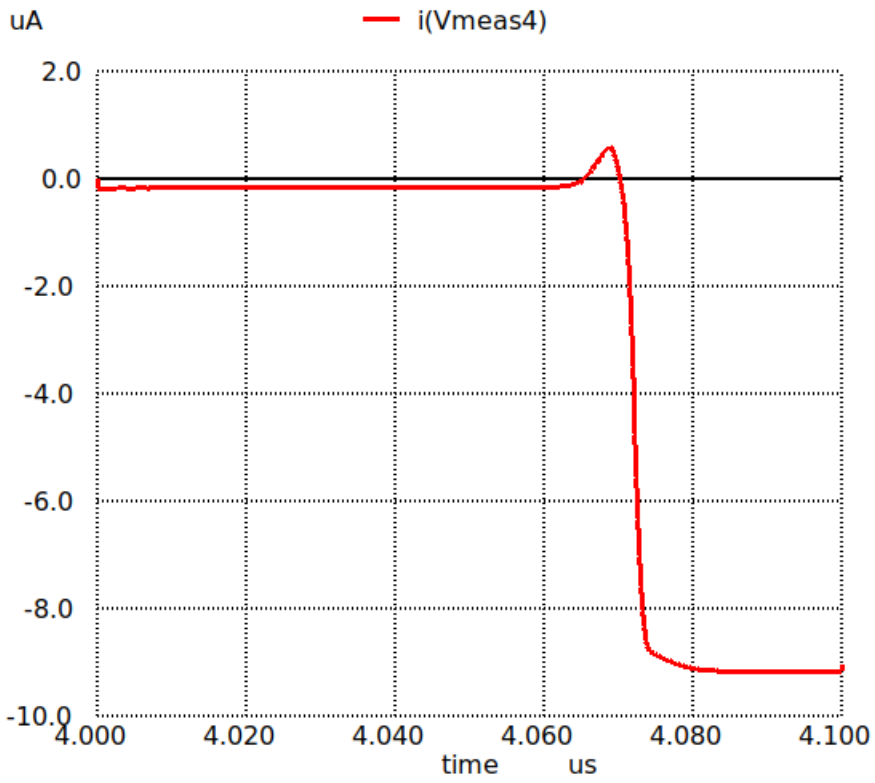


Figure 4.8: Current-Time Graph during Precharge phase

In Figure 4.8, the current vs. time graph depict the behavior during the precharge phase of the reading operation. It is observed that during this phase, the energy consumption is approximately 1.609 pJ. As previously indicated, in the worst-case scenario, the precharge phase must last at least 0.1 μs.

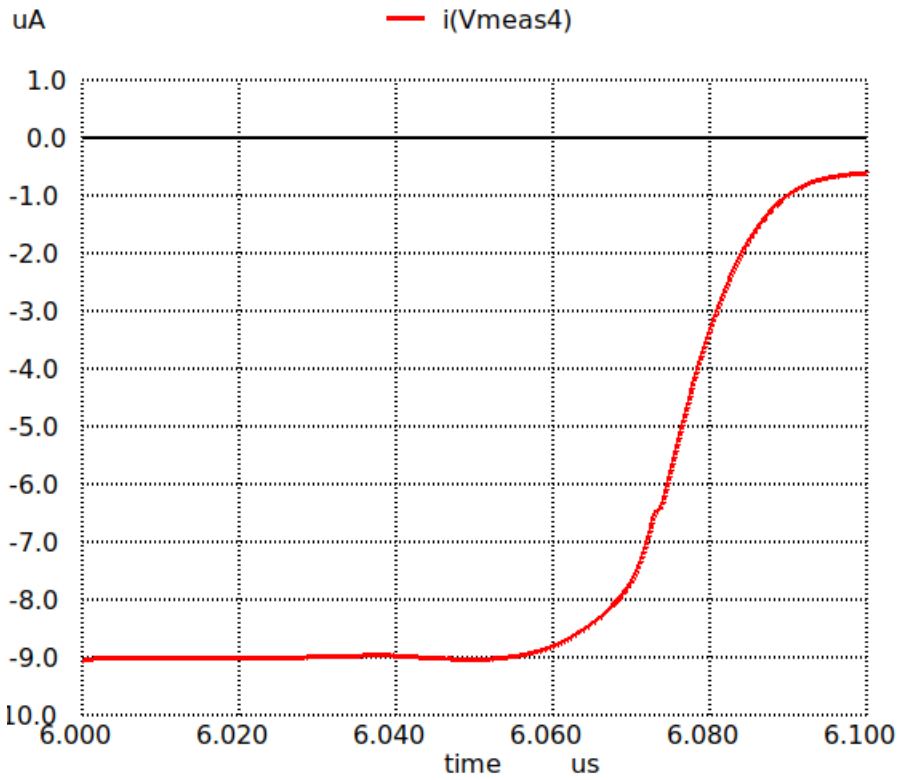


Figure 4.9: Current-Time Graph during Amplification phase

Figure 4.9 presents the current vs. time graph, illustrating the behavior observed during the amplification phase of the reading operation. It is noted that this phase incurs an energy consumption of approximately 2.447 pJ. As emphasized earlier, ensuring a minimum duration of 0.1 μs for the amplification phase is crucial under worst-case conditions.

Figure 4.10 depicts the current vs. time graph, demonstrating the behavior observed during the latching phase of the reading operation. This phase is associated with a power consumption of approximately 25.2 nW.

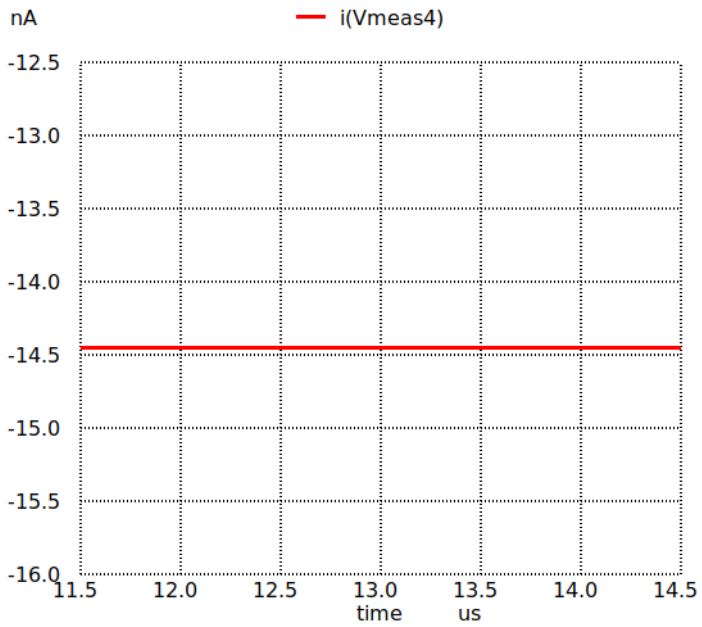


Figure 4.10: Current-Time Graph during Latching phase

4.2 RRAM Array

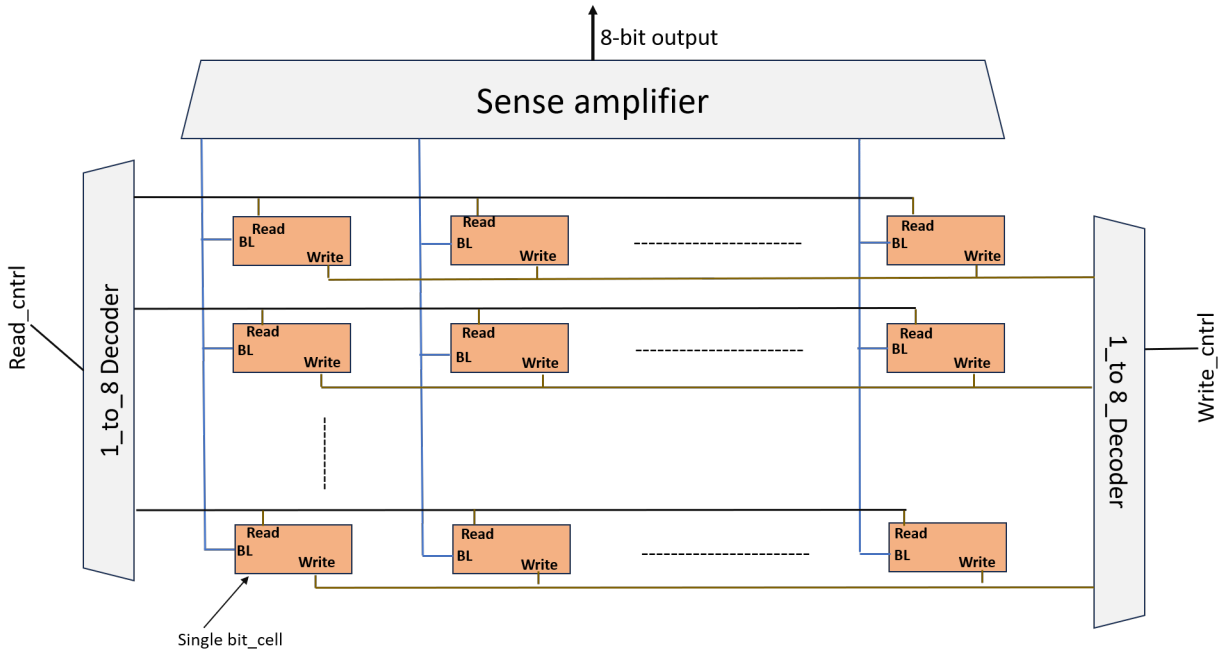


Figure 4.11: RRAM array architecture

The single-bit cell created in Section 4.1 can be expanded to form an 8x8 RRAM array architecture, as depicted in Figure 4.11. A common sense amplifier circuit can be employed for each column of the array. To read a specific row of the array, a read control circuit can be utilized, incorporating a demultiplexer to select the desired row. Similarly, for writing to the circuit, a write control circuit can be implemented, although manual writing is feasible since RRAM is non-volatile and requires writing only once.

Chapter 5

Concluding Remarks

5.1 Suggestions for Next Generation

- **Optimized Sense Amplifier:** Explore further optimization techniques for the sense amplifier to enhance the speed and accuracy of the reading phase.
- **Multibit RRAM Features:** Investigate the implementation of multibit features in RRAM cells to increase storage density and efficiency. The RRAM Array developed can be integrated with the S-AC DAC circuit of the ARYABHAT chip. This integration allows for the conversion of weights stored in the RRAM array into currents using the S-AC DAC. These currents, along with the input currents, can collectively be fed to the MP circuits.

5.2 Future Scope

The RRAM-based memory array designed using the open-source PDK and tools holds significant potential for various applications in the semiconductor industry. With continuous advancements in RRAM technology, the future scope for this product is promising. Potential areas for exploration include:

- **Market Impact:** The RRAM-based memory array has the potential to revolutionize memory storage solutions, offering higher density, lower power consumption, and faster access times compared to traditional memory technologies.
- **Advanced Applications:** Further research and development can lead to the integration of

RRAM-based memory arrays in advanced computing systems, IoT devices, and artificial intelligence applications, driving innovation in various industries.

- **Commercialization:** Collaborate with industry partners to explore opportunities for commercializing the RRAM-based memory array, ensuring widespread adoption and market penetration.

Bibliography

[1] M. Gu and S. Chakrabartty, "Synthesis of bias-scalable cmos analog computational circuits using margin propagation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 70, no. 3, pp. 1277–1290, 2023.

[2] P. Kumar, A. Nandi, S. Chakrabartty, and C. S. Thakur, "Process, bias, and temperature scalable cmos analog computing circuits for machine learning," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 70, no. 6, pp. 3429–3443, 2023.

[3] D. Zhu, Y. Li, W. Shen, Z. Zhou, L. Liu, and X. Zhang, "Resistive random access memory and its applications in storage and nonvolatile logic," *Journal of Microelectronics and Applications*, vol. 78, no. 4, pp. 123–134, 2023.

[4] K. Freund, "Ibm research says analog ai will be 100x more efficient," 2021, accessed: 2024-06-06. [Online]. Available: https://www.forbes.com/sites/karlfreund/2021/09/23/ibm-research-says-analog-ai-will-be-100x-more-efficient-yes-100x/?sh=61b5e23b129b

[5] C. S. Thakur, R. Wang, T. J. Hamilton, J. Tapson, and A. V. Schaik, "A low power trainable neuromorphic integrated circuit that is tolerant to device mismatch," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 2, pp. 211–221, Feb 2016.

[6] Y. C., W. J. J., and H. G. et al., "Physical mechanism and performance factors of metal oxide based resistive switching memory: a review," *J Mater Sci Technol*, vol. 32, p. 1, 2016.

[7] B. A., B. J. G., and G. C. et al., "Reproducible switching effect in thin oxide films for memory applications," *Appl Phys Lett*, vol. 139, p. 141, 2000.

[8] G. B. and et al., "Complementary metal oxide semiconductor compatible hf-based resistive random access memory with ultralow switching current/power," *Jpn J Appl Phys*, vol. 51, p. 04DD08, 2012.

[9] G. B., Y. S., and X. N. et al., "Oxide-based rram switching mechanism: a new ion-transport-recombination model," in *IEDM Tech Dig*, 2008, p. 4796751.

[10] R. M. J., I. I. H., and S. M. J., "Nonvolatile memory with multilevel switching: a basic model," *Phys Rev Lett*, vol. 92, p. 178302, 2004.

[11] G. B., S. B., and Z. H. W. et al., "Unified physical model of bipolar oxide-based resistive switching memory," *IEEE Electron Device Lett*, vol. 30, p. 1326, 2009.

[12] H. S. P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F. T. Chen, and M.-J. Tsai, "Metal-oxide rram," *Proceedings of the IEEE*, vol. 100, no. 6, pp. 1951–1970, 2012.

[13] L. R. U. et al., "Ember: A 100 mhz, 0.86 mm2, multiple-bits-per-cell rram macro in 40 nm cmos with compact peripherals and 1.0 pj/bit read circuitry," in *Proc. IEEE 49th Eur. Solid State Circuits Conf. (ESSCIRC)*, Sep 2023, pp. 469–472.

[14] W. Zhu, J. Jiang, H. Zhang, J. Xiao, G. Yang, and S. Zou, "Offset-cancelling current-mode sense amplifier for fast-read eflash memory," *Electronics Letters*, vol. 55, no. 7, pp. 376–378, April 2019, vol. 55 No. 7.