MULTILEVEL-CELL PHASE-CHANGE MEMORY - MODELING AND RELIABILITY FRAMEWORK

THÈSE N^{O} 6801 (2016)

PRÉSENTÉE LE 14 JANVIER 2016

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE DE SYSTÈMES MICROÉLECTRONIQUES
PROGRAMME DOCTORAL EN GÉNIE ÉLECTRIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Aravinthan ATHMANATHAN

acceptée sur proposition du jury:

Prof. G. De Micheli, président du jury
Prof. Y. Leblebici, Dr M. Stanisavljevic, directeurs de thèse
Dr E. Eleftheriou, rapporteur
Dr R. Bez, rapporteur
Dr J. Brugger, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2016

All birds find shelter during a rain ...
But eagle avoids rain by flying above the clouds ...
Problems are common, but attitude makes the difference ...
— Dr. A. P. J. Abdul Kalam

To my parents and wife...

Acknowledgements

This accomplishment would not have been possible without the wonderful people who have inspired me on my journey towards the PhD. This PhD challenge would have been more of an academic title than a great adventure if not for them. I consider myself fortunate to have spent the last four years of my life in a great research atmosphere for learning, and also working with some exceptional individuals during the course of my doctoral studies.

I am greatly indebted to my supervisor, Prof. Yusuf Leblebici, who gave me the independence to pursue my ideas freely, while at the same time mentoring by example. I am especially grateful to my manager, Dr. Evangelos Eleftheriou, for giving me this wonderful opportunity to pursue my PhD thesis with IBM Research. His energy and enthusiasm have been an inspiration for me to aim ever higher in my pursuits. I am also thankful to my manager, Dr. Haralampos Pozidis, for his constant motivation and guidance throughout the course of my PhD thesis. I would also like to thank my thesis co-advisor, Dr. Milos Stanisavljevic, who has been my mentor and guide from the very beginning of my PhD studies, and has constantly provided me with help and encouragement on both professional and personal fronts. I would also like to thank my second co-advisor, Dr. Daniel Krebs, for his constant efforts in shaping me for my PhD.

Research in engineering is an extremely collaborative effort. I would like to acknowledge here my amazing colleagues and collaborators whose contributions helped in various aspects of my research.

1. Dr. Milos Stanisavljevic at IBM Research – Zurich- in the design and implementation of the various programming and readout architectures for MLC PCM.

2. Dr. Daniel Krebs at IBM Research – Zurich - in the development of the finite-element method based thermoelectric model using COMSOL Multiphysics software.

3. Dr. Haralampoz Pozidis at IBM Research – Zurich - in the design and analysis of the drift-tolerant coding and detection schemes which are used for the demonstration of reliable triple-level cell storage.

4. Dr. Nikolaos Papandreou and Mr. Urs Egger at IBM Research – Zurich - in the implementation of the embedded FPGA-based hardware characterization platform for the reliability experiments performed on the prototype PCM chip.

5. Dr. Abu Sebastian and Mr. Manuel Le Gallo at IBM Research – Zurich - in the development of the thermoelectric model by providing their valuable insights and experimental measurements.

6. Dr. Chung H. Lam and Dr. Matt Brightsky at IBM T. J. Watson Research center - for the fabrication and development of the prototype chips which were essentially used for the characterization experiments.

The IBM-SK Hynix joint-development agreement for PCRAM development, which funded a portion of my doctoral research, enabled me to collaborate with the various industrial research laboratories (IBM T. J. Watson Research in Yorktown Heights, USA, and Next Generation Memories research at SK-Hynix, South Korea), and provided a platform to exchange ideas with leading researchers from the PCM research community. I would like to thank my thesis jury members – Prof. Giovanni De Micheli, Prof. Jürgen Brugger, Dr. Roberto Bez and Dr. Evangelos Eleftheriou – for evaluating my thesis and providing valuable feedback and insights.

I would like to thank my colleagues of the Non-volatile memory systems group at IBM Research – Zurich - Thomas Parnell, Thomas Mittelholzer, Tobias Blätter and Nikolaos Papandreou - for their friendship and support, and for constantly inspiring me with their accomplishments. Thanks to Haralampos Pozidis, Milos Stanisavljevic and Charlotte Bolliger for proof-reading my thesis.

I would also thank the friends I made from IBM and EPFL – Arvind Raj, Vinodh, Viswanathan, Prasad, Nirmalraj, Jelena, Kazim, Jayant, Nicola, Rakesh and Felix - for the many enjoyable and intellectual conversations. I would also like to thank my friends Kumaran, Harish, Sailokesh, Mukund, Sindhuja, Siddharth, Ashwin Kumar, Vijayabalaji, Aishwarya, Vamsi Krishna who, despite living in far-away lands, always made me feel close to their hearts.

I am truly indebted to my loving parents for everything I achieved. I would like to thank my mom and dad, Mangalam and Athmanathan, for being everything in my life, my sister Anusha and my brother-in-law Vinodh for their love, care and support during my studies and beyond.

I am grateful to my wife Anuradha for being my soulmate, and source of my happiness. Her presence has added significant meaning to my life. Her unconditional love and care helped and motivated me to get through all tough times of my PhD thesis. I enjoy sharing my research experience with her and I dedicate this dissertation to my beloved lovely wife, Anuradha.

Finally, a special note of thanks to the beautiful town of Adliswil, in the canton of Zürich for giving me many wonderful memories, which I can cherish for the rest of my life, and for making the last four years some of the happiest of my life.

Lausanne, 14 August 2015 Aravinthan Athmanathan

Abstract

In the modern digital era of big data applications, there is an ever-increasing demand for higher memory capacity that is both reliable and cost effective. In the domain of non-volatile memory systems, Flash-based storage devices have dominated the consumer space for the past 15 years and have also entered the enterprise storage system in the past 2-3 years. However, with Flash memory devices facing serious scalability limits, there is an imminent need to explore the viability of other non-volatile memory technologies that can replace or complement Flash-based storage in the near future. Significant research efforts have been invested by various universities and research organization across the globe into realizing the so-called next-generation memories (NGMs). Phase-change memory is one such technology, which is viewed as the most promising candidate among the emerging technologies.

Phase-Change Memory (PCM) technology is not new as the principle of storing information in chalcogenide-based materials was first explored in the 1960s. However, with the predominance of charge-based memory technology (DRAM, Flash storage, etc) thriving thanks to the technological advancements of metal-oxide semiconductor field-effect transistor (MOSFET) based devices, it was not until the late 90s that renewed interest in the class of phase-change materials was ignited. This is the same genre of materials as have been widely used in laser-driven optical storage (rewritable CDs and DVDs) during the past 20 years or so. Although PCM chips have already been mass-produced by companies like Micron and Samsung, their capacity was limited as they were based on single-level cell (SLC) storage. The primary research focus of this thesis is on increasing the memory capacity by storing more than one bit of information per device, known as multilevel-cell (MLC). Achieving MLC capability is quite challenging, and we disclose some of our significant achievements in the realization of MLC PCM in the past few years.

One of the main concerns for the sustainability of PCM technology is the power-hungry RESET process. Thermoelectric physics in nano-scale PCM devices is shown to play a significant role in the programming of these devices. Interestingly, these effects result in better thermal confinement and can even pave the way for more power-efficient devices. Therefore, it is high time that the thermoelectric physics within the devices is completely understood. Reliability is the other major concern for PCM technology, especially in the realization of multilevel-cell storage at highly dense PCM arrays. Several MLC-enabling technological advancements that suppress the drift phenomenon and array variability have been explored in the recent past.

Abstract

Although a small number of demonstrations of MLC PCM feasibility have been performed, there is a need for the holistic application of enabling technologies in order to fully address the major reliability issues.

In this thesis, a comprehensive thermoelectric model is proposed for investigating the thermoelectrics physics in PCM device operation. With the increasing role of thermoelectrics at the smaller technology nodes, the proposed model provides valuable insights for a complete understanding of device operation. The model is validated by comparing the simulation results with experimental measurements. The model can also be used for fine-tuning the material properties, device design and geometry to improve the efficacy of these devices. In the second part of the thesis, the dominant reliability concerns in PCM technology are addressed, particularly focusing on MLC operation. We present a readout circuit for PCM specifically designed for drift resilience in MLC operation. Drift resilience is achieved through the use of specific non-resistance-based cell-state metrics which, in contrast to the traditional cell-state metric, i.e., the low-field electrical resistance, have built-in drift robustness. By employing novel MLC-enabling techniques, we succeeded in demonstrating for the first time, reliable 3-bits/cell memory density with a data retention of 1 week at temperatures ranging from 30°C to 80°C on devices that had been pre-cycled one million times.

Keywords: Phase-change memory, Multilevel-cell storage, Finite-element method based thermoelectric modeling of PCM, Drift resilience, Non-resistance cell-state-based readout metrics for MLC storage, Reliability and data retention analysis at elevated temperature, Reliable TLC data storage in PCM.

Zusammenfassung

Im Zeitalter von Big-Data-Anwendungen besteht ein dauernd steigender Bedarf an zuverlässigen und preisgünstigen Datenspeichern mit hoher Kapazität. Im Bereich der nichtflüchtigen Speicher dominieren Flash-basierte Halbleiterspeicher seit 15 Jahren in der Unterhaltungselektronik und bei Konsumgütern. Seit 2-3 Jahren werden sie vermehrt auch im Serverbereich eingesetzt. Jedoch stossen Flash-basierte Speicher an Grenzen bezüglich ihrer Skalierbarkeit, deshalb sind Speichertechnologien, die auf anderen Halbleitern basieren und in naher Zukunft Flash ablösen oder ergänzen können, ein sehr aktives Forschungsgebiet. Weltweiten forschen Universitäten und industrielle Forschungslabors daran, diese so genannten Next-Generation Memories (NGM) zu kommerzieller Reife zu bringen. Eine dieser neuen Technologien — und aus heutiger Sicht der erfolgversprechendste Kandidat — ist die so genannte Phase-Change Memory (PCM, bzw. Phasenwechselsspeicher) Technologie.

Das Konzept der PCM-Speichertechnologie ist grundsätzlich nicht neu, wurden doch erste Ideen zur Datenspeicherung in Chalkogeniden bereits in den 1960er Jahren erprobt. Ladungsbasierte Speichertechnologien, wie DRAM, Flash und andere, waren jedoch aufgrund des technischen Fortschritts der MOSFET-Transistoren bis in die frühen 90er Jahre vorherrschend und sehr erfolgreich. In den späten 90er Jahren stiessen dann phasenwechselnde Materialien wieder auf vermehrtes Interesse. Dies ist die gleiche Art von Materialien, wie sie für die Speicherschicht der vor etwa 20 Jahren eingeführten optischen Speichermedien (wiederbeschreibbare CDs, DVD und BluRay-Disks), die mit Laser beschrieben werden, verwendet wurden. PCM-Chips werden bereits kommerziell hergestellt, zum Beispiel von Micron und Samsung, weisen aber typischerweise eine beschränkte Speicherkapazität auf, da sie auf einem gespeicherten Datenbit pro Speicherzelle, also so genannten Single-Level Cell (SLC) Speichern, beruhen. Der Hauptfokus der vorliegenden Doktorarbeit liegt in der Erhöhung der Speicherdichte durch die Speicherung mehrerer Datenbits pro Speicherzelle, also so genannte Multi-Level Cell (MLC) Speicher. Speicherung mittels MLC ist technisch sehr anspruchsvoll, aber wir haben in den vergangenen Jahren bereits Fortschritte erzielt und wichtige Ergebnisse hinsichtlich ihrer Realisierung veröffentlicht.

Einer der Gründe für die erwähnten Bedenken gegenüber der PCM-Technologie ist der RESET-Vorgang der Speicherzellen, welcher einen hohen Energiebedarf aufweist und somit

die Lebensdauer der Speicherzellen beeinflusst. Thermoelektrische Vorgänge im Nanometer-Grössenbereich innerhalb der PCM-Zellen spielen nicht nur eine wichtige Rolle bei deren Programmierung, sondern sind auch ein wesentlicher Faktor bezüglich der Energieeffizienz. Daher ist wichtig, zu einem umfassenden Verständnis der thermoelektrisch-physikalischen Aspekte und Vorgänge in den PCM-Zellen zu gelangen. Ein weiterer kritischer Faktor in der PCM-Technologie ist die Zuverlässigkeit, speziell bei MLC-Anwendungen für hohe Speicherdichten. Verschiedene technologische Fortschritte wurden erzielt, welche die Empfindlichkeit in Bezug auf Drift und Zellvariationen verringerten. Ebenso wurde die prinzipielle Machbarkeit von MLC PCM bereits gezeigt, allerdings sind umfassende anwendungsnahe Messungen notwendig, um die kommerzielle Reife und Brauchbarkeit der MLC-PCM-Technologie zu demonstrieren.

In der vorliegenden Arbeit wird ein umfassendes, thermoelektrisches Simulationsmodell vorgestellt, das die physikalischen Vorgänge innerhalb der PCM-Zellen erklärt. Da thermoelektrische Vorgänge bei immer kleineren Technologiedimensionen eine immer wichtigere Rolle spielen, erlaubt es dieses Simulationsmodell, die Vorgänge besser zu verstehen und vorauszusagen. Das Modell und dessen Simulationsergebnisse wurden mit experimentellen Messungen verifiziert. Das Modell ermöglicht eine genau Abstimmung von Materialeigenschaften, Zellgeometrie und Zellstruktur, um die Funktionalität und die Effizienz der PCM-Zellen zu verbessern.

Wir stellen insbesondere auch einen Ausleseschaltkreis (readout circuitry) für PCM vor, der speziell hinsichtlich seiner Resilienz gegenüber Drift für MLC-Anwendungen entworfen wurde. Diese Resilienz wird durch die Verwendung einer Cell-State-Metrik erzielt, die nicht auf Basis des Widerstands beruht und somit eine inhärente Robustheit gegenüber Drift aufweist. Herkömmliche Cell-State-Metriken basieren auf dem elektrischen Widerstand bei schwachen Feldern. Im zweiten Teil der Doktorarbeit werden wichtige Aspekte, die die Zuverlässigkeit der PCM-Technologie beeinflussen, speziell auch für die MLC Anwendung, diskutiert. Durch die Anwendung neuartiger Techniken zur Ansteuerung der PCM-Zellen konnte erstmals die Zuverlässigkeit von PCM zur Speicherung von 3+ Bits/Zelle bei einer Datenerhaltung 1 Woche im erhöhten Temperaturbereich zwischen 30°C bis 80°C demonstriert werden — und zwar mit PCM Zellen, die bereits eine 1 Million Zyklen durchlaufen hatten.

Keywords: Phase-Change Memory, Multilevel-Cell storage, Finite-element method based thermoelectric modeling of PCM, Drift resilience, Non-resistance cell-state-based readout metrics for MLC storage, Reliability and data retention analysis at elevated temperature, Reliable TLC data storage in PCM.

Contents

List of Figures

List of Tables

1 Introduction

This thesis focuses on the modeling and reliability framework for multilevel cell (MLC) Phase-Change Memory (PCM) technology. The first half of the thesis proposes a comprehensive finite-element method based modeling and simulation approach to gain a complete understanding of the PCM device operation. Special focus is on modeling the thermoelectric physics in the nano-scale PCM devices, and the model is validated by comparing the results with experimentally measured data. The second half of the thesis deals with various reliability concerns affecting the realization of MLC in PCM. Novel MLC enabling technologies are presented and their performance is studied to demonstrate the viability of MLC PCM technology.

In this introductory chapter, I present a brief background about the thesis topic and the main motivation for pursuing this research topic. Specifically, the following concepts will be described: Classification of memory hierarchy, brief introduction of some of the emerging non-volatile memory technologies, need for an universal memory, basic concept and properties of PCM, MLC storage operation in PCM, state-of-the-art modeling approaches, and reliability concerns of MLC PCM. Finally, the main goal of the thesis is given, followed by my contributions to this thesis, and the organization of the thesis described.

1.1 Semiconductor Memory

Semiconductor memories are usually considered to be the most outstanding microelectronic component of a modern digital world. In the current era of big-data storage applications, modern digital systems require the capability of storing and retrieving large amounts of information (data) within a short span of time (high bandwidth). Memories are devices, circuits or systems that are capable of storing huge amounts of digital information at high density (GB/mm^2). Traditionally, these memory systems are all based on semiconductor device-based technology, and a recent survey indicates that roughly 40% of the worldwide semiconductor business is focused on storage-specific applications. Over the years, technology advancements have driven the capacity and performance of memory systems to greater heights. Also, the rapid growth and demand associated with these memories have led to the emergence of new memory technologies.

These memories are essential for data storage in all microprocessor-based applications, ranging from satellites, data centers to consumer-level electronic gadgets (such as laptops, tablets, smart-phones, digital cameras, music players, etc.). Typically, a semiconductor memory device stores digital information, either '0' or '1' in a structure that can be rapidly switched between two readily distinguishable states. Current state-of-the-art memory technologies (SRAM, DRAM, Flash memory, etc) are all based on the presence or absence of electrical charge in a tiny confined region of the memory device.

In the past decades, the semiconductor memory industry has seen regular technological advancements to smaller device dimensions, thereby closely following Moore's law. This march towards smaller dimension enables us to attain higher capacity, thereby increasing system functionality. Numerous innovations in memory technology are reflected in the continuous advancements in achieving high-density, high-data-rate and low-power memory systems, in the course of which the design rule has shown a considerable scaling from the micrometer regime down to few tens of nanometers. Such technological advancements achieved via scaling for higher densities and faster speeds, helped establish the performance standards for most of our modern-day electronic gadgets and devices.

1.2 Classification of Memory Hierarchy

The profound success of modern-day computer technology stems from the tremendous progress achieved in the storage technology space. Early computers had a few kilobytes of random-access memory. The earliest IBM PCs did not even have a hard disk. That changed with the introduction of the IBM PC-XT in 1982, with its 10-megabyte disk. By the year 2015, typical machines had 100,000 times as much disk storage, and the amount of storage has increased by a factor of 2 every couple of years for the past three decades.

In the real world, a storage system is a hierarchy of storage devices with different capacities, costs and access times. The storage hierarchy is designed to bridge the performance gap

between the fast central processing units (CPUs) and the slower memory and storage technologies, while keeping overall system costs down. A simplified memory hierarchy of current digital systems is shown in Fig. 1.1. Fast accessible memory systems are used to support the CPUs (SRAM and DRAM). To store data for longer times, slower and cheaper storage devices are used (Flash and hard-disk drives).

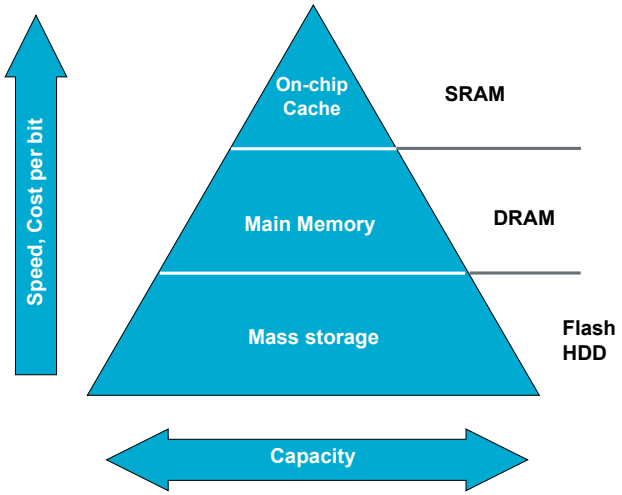


Figure 1.1: Simplified storage hierarchy of the modern digital systems.

Based on the type of data storage and the speed of data-access mechanisms, semiconductor memories can be broadly be classified into the following two main categories:

- Random-Access Memory or Volatile memory
- Read-Only Memory or Non-Volatile memory

1.2.1 Volatile Memory

Random-access memory (RAM) comes in two varieties—static and dynamic. Static RAM (SRAM) is faster and significantly more expensive than Dynamic RAM (DRAM). SRAM is mainly used as cache memories, both on and off the CPU chip. DRAM is typically used as the main memory in modern systems. Typically, a desktop system will have no more than a few megabytes of SRAM, but hundreds or even thousands of megabytes of DRAM.

1.2.1.1 Static-RAM

In Static Random-Access Memory (SRAM), the logical information is stored by setting the logic state of a bistable flip-flop. It uses four transistors, which form two cross-coupled inverters, to store 1 bit of data. This storage cell has two stable states, which are used to denote ‘0’ and ‘1’. Two additional transistors serve to control the access to a required storage cell during read

and write operations. Provided the device is powered ON, this circuit has the property that it can stay indefinitely in either of two different voltage configurations or states.

1.2.1.2 Dynamic-RAM

Dynamic random-access memory (DRAM) is a type of random-access memory that stores each bit of data in a MOS capacitor. The capacitor can be either charged or discharged, and these two states represent the two values of a bit, conventionally called '0' and '1'. As capacitors suffer from charge leakage, the stored information eventually fades over time unless the capacitor charge is refreshed periodically.

1.2.1.3 Applications

SRAMs are widely used in computers as high-performance Level1 (L1) and Level2 (L2) cache. These SRAMs typically run at the CPU clock speed, so that the access times are on the order of a few nanoseconds. The current performance of SRAM is hard to match for any other memory technology. They are also used in combination with NOR flash in consumer electronics. DRAM is a reliable and proven technology, and has been used in computers since the early 1970s. But the DRAM developers are facing their difficult times with various scaling issues associated with storage interference, device leakage, etc.

1.2.2 Read-Only Memory

Unlike the RAM, Read-Only Memory (ROM) or Non-Volatile Memory (NVM) has the ability to retain the stored information even after the electrical power is turned off. It can be broadly divided into the following categories:

- Mask-programmed ROM: The required contents of the memory are programmed during fabrication
- Programmable ROM (PROM): The required contents are written in a permanent way by burning out internal interconnections (fuses). It is a one-off procedure.
- Erasable PROM (EPROM): Data is stored as a charge on an isolated gate capacitor of a floating-gate transistor. Data can be erased by exposing the PROM to ultraviolet light.
- Electrically Erasable PROM (EEPROM): It is also based on the concept of charge storage on the floating-gate transistor. However, the contents can be erased by applying electrical pulses. These memories are very important data storage devices for mobile applications. A typical example is Flash memories.

1.2.2.1 Flash memory

Flash memory is a non-volatile memory technology that can be electrically erased and reprogrammed. In today's world, Flash memory dominates the storage space and is easily one of the most successful non-volatile memory technology to date. Flash memory has become a powerful and cost-effective solid-state storage technology widely used in mobile electronics devices and other consumer applications. Lately, Flash can also be found in enterprise storage systems, such as hybrid- and all-Flash arrays.

Flash memory is based on the floating-gate memory technology, where charges are stored at the floating gate which is electrically insulated by insulators, – i.e., a tunnel oxide and an oxide-nitride-oxide dielectric (ONO). Figure 1.2 depicts a typical Flash memory device, which is basically a floating-gate MOS transistor, i.e., a transistor with the floating gate (FG) completely surrounded by dielectrics and electrically governed by a capacitively coupled control gate (CG). Being electrically isolated, the charge injected into the FG remains there, allowing modulation of the "apparent" threshold voltage (seen from the CG) of the cell transistor [Bez et al., 2003].

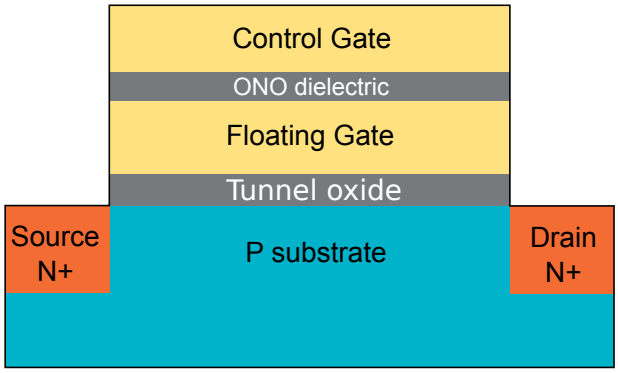


Figure 1.2: Simplified device structure of the floating-gate MOS transistor. The floating gate is electrically insulated and stores the charge for data storage.

Two major forms of Flash memory, NAND Flash and NOR Flash, have emerged as the dominant varieties mainly used in portable electronics devices. NAND Flash, which was designed with a very small cell size to enable a low cost-per-bit of stored data, has been used primarily as a high-density data storage medium for consumer devices, such as digital cameras and USB solid-state disk drives. NOR Flash has typically been used for code storage and direct execution in portable electronics devices, such as cellular phones and PDAs.

MLC & 3D NAND Flash Charges are pushed into or pulled out of the floating gate, by either carrier tunneling via the high electric field across the tunnel oxide or hot electron generation in the Si channel so that they have a large enough energy to overcome the energy barrier of the tunnel oxide. The advantage of floating-gate memory is that its simple layout pattern allows not only small cell sizes in terms of F^2, where F is the minimum lithographic feature size, but also small F, which results in low cost per bit. Multi-level capability achieved by

precisely controlling the number of charges stored in the floating gate further decreases the cost per bit [Ricco et al., 1998; Sako et al., 2015]. Consequently, NAND flash has achieved the highest density among mainstream solid-state memory devices. However, flash memory is facing serious limitations in terms of scaling challenges. Scaling of the transistor gate size decreases the number of charges that can be reliably stored in the floating gate [Kim et al., 2005]. Given the issue of increasing leakage current at smaller dimensions, the retention time of stored charges decreases drastically. This is more serious for multi-level cells realizations because of the minimal difference in the number of charges stored in each level. 3D stacking provides an alternate scaling solution and is attracting increasing attention [Goda, 2013]. In 3D NAND, the physical cell size is decoupled from the effective cell size by stacking multiple layers. This enables effective NAND size scaling without degrading cell performance and reliability. However, 3D process integration could introduce new sources of cell degradation.

1.2.3 Need for universal memory

From the memory hierarchy classification, it is evident that multiple forms of memory are currently being used to cater the specific needs of modern digital systems. However with the traditional memory technologies (DRAM, SRAM and Flash) suffering from scalability limitations, a new class of memory called 'universal memory' is gaining momentum. The expected characteristics of such a universal memory concept include high memory capacity (achievable at low cost), high operational bandwidth (high-speed read and write operations), low power consumption (both access and standby power for energy efficiency), random accessibility, non-volatile storage, excellent scaling properties down the technology node, and infinite endurance. These characteristics allow the universal memory to meet the requirements of a wide range of applications from large, expensive, enterprise-grade storage systems to low-cost, ubiquitous, consumer-grade hand-held electronic devices.

Thus, the evolving next-generation non-volatile memory technologies (NVMs) should be more scalable than conventional NAND Flash to attain the higher densities offered by future nano-scale technology nodes. There is a need for a memory that can offer better endurance cycles, data retention and I/O performance than state-of-the-art flash, to reduce costs while increasing the performance. The emergence of such non-volatile solid-state memory technology that can combine high performance, high density and low-cost could also bring some healthy changes in the memory/storage hierarchy throughout all computing platforms, ranging from mobile phones all the way up to high performance computing (HPC) applications. Also, if the cost-per-bit could be driven low enough through ultra-high memory capacity, such a memory could ultimately displace the traditional magnetic hard-disk drives (HDD) in enterprise storage server systems. Though Flash is already doing that today to a certain extent, majority of the enterprise storage is still based on HDD.

1.2.3.1 Storage Class Memory (SCM)

An interesting region to observe in the memory and storage hierarchy is the gap in terms of performance (latency) between the disks and the rest of the system, which continues to widen rapidly. Research and development efforts are underway on several novel non-volatile memory technologies that not only complement the existing memory and storage hierarchy but also reduce the distinctions between memory (fast, expensive, evanescent) and storage (slow, inexpensive, permanent). This newer class of emerging non-volatile memories that are capable of cater to future memory demands and bridging the gap between the conventional HDDs (storage) and solid-state memories are called "Storage Class Memory" (SCM).

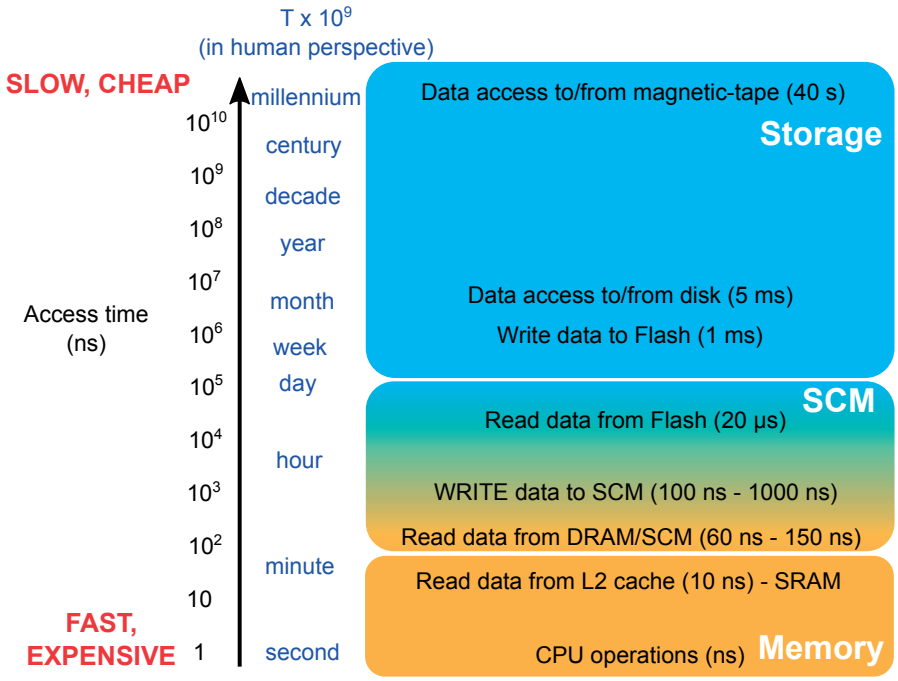


Figure 1.3: Distinction between the access times for various storage and memory technologies, both in nanoseconds and in terms of human perspective. For the latter, all times are scaled by 10^9 so that the fundamental unit of a single CPU operation is analogous to a human making a one-second decision.

SCM combines the benefits of a solid-state memory, such as high performance and robustness, with that of the archival capabilities and low cost requirements of the conventional HDD magnetic storage. Typically, SCM class of memory includes the non-volatile memory technology that could be manufactured at an extremely high effective area density, using combination of sub-lithographic patterning techniques, multilevel cell and multiple layers of devices (3D stacking).

1.3 Emerging non-volatile memory technologies

Beyond the near future, there is an uncertainty that prevails regarding most of the present day memory technologies. The uncertainty is primarily due to the scalability issues associated with these charge-based memories. Given these difficulties in scaling to future technology nodes, flash researchers are already struggling to maintain the specifications (such as write endurance, retention of heavily cycled cells and write/erase performance), let alone improve them. Although researchers have demonstrated reliable flash storage in isolated devices in the sub-20-nm regime, mass production and continued scaling are still a major concern [Goda, 2013]. At the same time, that flash is struggling to maintain the current levels of reliability and performance while increasing density, new application areas are opening up for which these specifications are just barely adequate.

Thus, there is a need for a transition from conventional charge-based memory to emerging resistive-based memory. Charge-based memories require discrete amounts of charge to induce a voltage, which is detected during readout. For example, in the non-volatile space, flash memories must precisely control the discrete charge placed on a floating gate, whereas, in volatile main memory, DRAM must not only place charge in a storage capacitor, but also mitigate sub-threshold charge leakage through the access device. Capacitors must be sufficiently large to store charge for reliable sensing, and transistors must be sufficiently large to exert effective control over the channel. Given these challenges, scaling DRAM beyond 10 nm will be increasingly difficult [Frank et al., 2001; Park, 2015].

In contrast, resistive-based memories use electrical current to induce a change in atomic structure while programming, which impacts the resistance detected during data retrieval. Resistive memories are amenable to scaling because they do not require precise charge placement and control. Programming mechanisms, such as current injection, scale with cell size. Phase-change memory (PCM), Resistive RAM (RRAM), Magnetoresistive RAM (MRAM) and ferroelectric RAM (FeRAM) are some examples of emerging resistive memory technologies. Of these, PCM is closest to realization and imminent deployment as a NOR flash replacement.

1.3.1 Ferroelectric Random Access Memory (FRAM)

Ferro electric random access memory (FRAM or FeRAM) is an emerging non-volatile memory which uses the polarization property of the ferroelectric thin-film material placed between two electrodes (Fig. 1.4). Transistors are used as access devices, and the device structure closely resembles that of the DRAM structure. However, FRAM offers non-volatile data storage and is considered to be significantly more energy efficient than the other conventional non-volatile memories. In typical devices, lead zirconate titanate ($PbZrTiO_3$, PZT) is used as the ferroelectric material, while other materials are currently being explored [Moise et al., 2002; Qazi et al., 2011].

In the presence of an electric field across the electrodes, the material will change its dipole

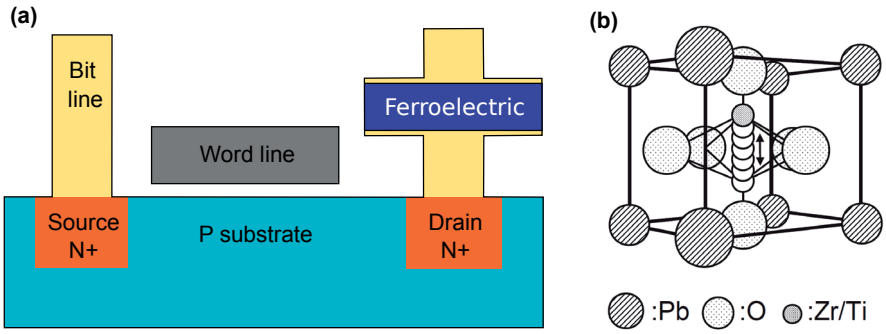


Figure 1.4: (a) Device structure of FRAM cell. (b) Atomic structure of the ferroelectric material. The center atom can be moved by applying an electric field, and the movement can be sensed by the current displacement.

direction, which can be sensed electrically. The dipole can be moved up or down by varying the electric field, which can be sensed by the displacement current due to the charge spike. It is this property that provides the non-volatility and requires only minimal power for data storage. However, one disadvantage of FRAM is its destructive read cycle. The readout process involves writing a bit to each cell; if the state of the cell changes, then a small current pulse is detected, indicating that the cell was in the OFF state. FRAM has very high endurance, fast read/write access (faster than Flash memory) and extremely low power consumption. Owing to its large cell size, it is expected to used in the products targeting some niche applications.

1.3.2 Magneto-resistive Random Access Memory (MRAM)

Magneto-resistive random-access memory (MRAM) uses the direction of the magnetization (polarization) property of the material to store information. The magneto-resistance can be used as the readout information. Figure 1.5 illustrates the device structure, consisting of two ferromagnetic layers separated by a thin tunnel barrier layer [Bette et al., 2003; Wolf et al., 2010]. The resistance of this magnetic tunnel junction (MTJ) is either low or high based on whether the relative polarization between the two magnetic layers is parallel or anti-parallel.

Some of the properties of MRAM include fast programming and readout access (in tens of nanoseconds), low operating voltage, high endurance ($\sim 10^{15}$), and good data retention characteristics. However, typical drawbacks include the large programming current and power requirement. Also, the device size is relatively large compared with other competing memory devices. In terms of scaling, MRAM requires a large peripheral circuitry to sense the relatively small resistance difference between parallel and anti-parallel states.

1.3.2.1 Spin-Transfer Torque Random Access Memory (STT-RAM)

STT-RAM is a form of MRAM that uses the spin-transfer torque to re-orient the free layer by passing a large, directional write current through the MTJ. Direct current-driven magne-

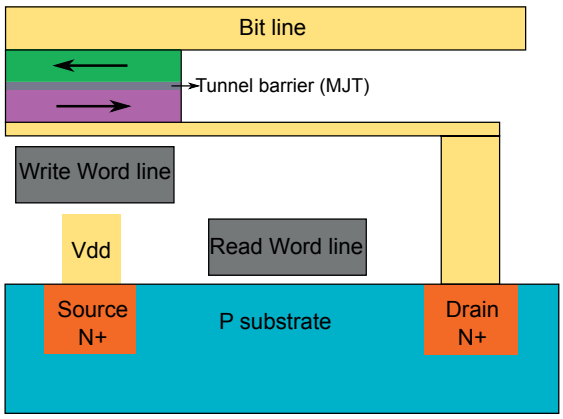


Figure 1.5: Device structure of MRAM cell.

tization in STT-RAM is expected to make it possible to decrease the programming current substantially and to continue the scaling of magnetic random access memory [Chen et al., 2010]. Considering the fact that the device size is still larger in F^2, they will be mostly relevant for embedded storage applications.

1.3.3 Resistive Random Access Memory (RRAM)

The device structure is simply an oxide material sandwiched between two metal electrodes, called the metal–insulator–metal (MIM) structure as shown in Fig. 1.6. The device structure involves various mechanisms such that they show a distinct difference in resistance upon application of a current or voltage pulse.

Numerous materials (insulators–metal oxides) have been reported to show resistance switching behavior including NiO [Baek et al., 2004], TiO_2 [Baek et al., 2004], PCMO - (Pr,Ca)MnO_3 [Zhuang et al., 2002], and STO - $SrTiO_3$:Cr [Watanabe et al., 2001; Beck et al., 2000]. Also, the switching behavior depends not only on the sandwiched oxide materials, but also on the choice of metal electrodes and their interfacial properties [Waser, 2009].

Various mechanisms are believed to be responsible for the resistance change in these materials, such as thermally induced formation and resolution of a conductive filament and ion migration combined with redox processes [Wong et al., 2012; Marinella, 2014]. RRAM Memories are still at their early research phase and therefore, it is too early to conclude whether these memories can become mainstream memory devices. However, early research results show that they have a strong potential to be low-power-operable, highly scalable and achieve ultra-high memory capacity.

The potential scalability to the nanometer regime is one of the key motivations that push the development of RRAM technology. Lee et al. successfully triggered localized switching events in NiO memory by manipulating C-AFM, thus showing that the size of CF can be less than 10 nm [Lee et al., 2009]. Apart from scaling, the RRAM technology is also capable of modulating

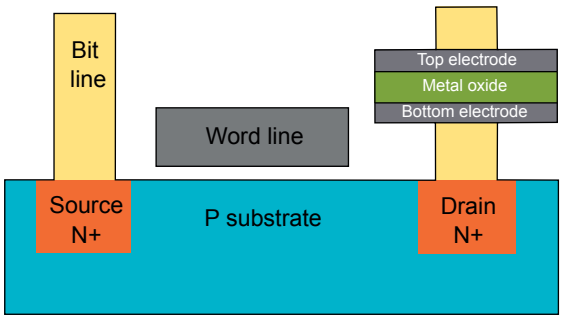


Figure 1.6: Typical device structure of RRAM memory cell. The resistive metal oxide is sandwiched between the top and the bottom electrode.

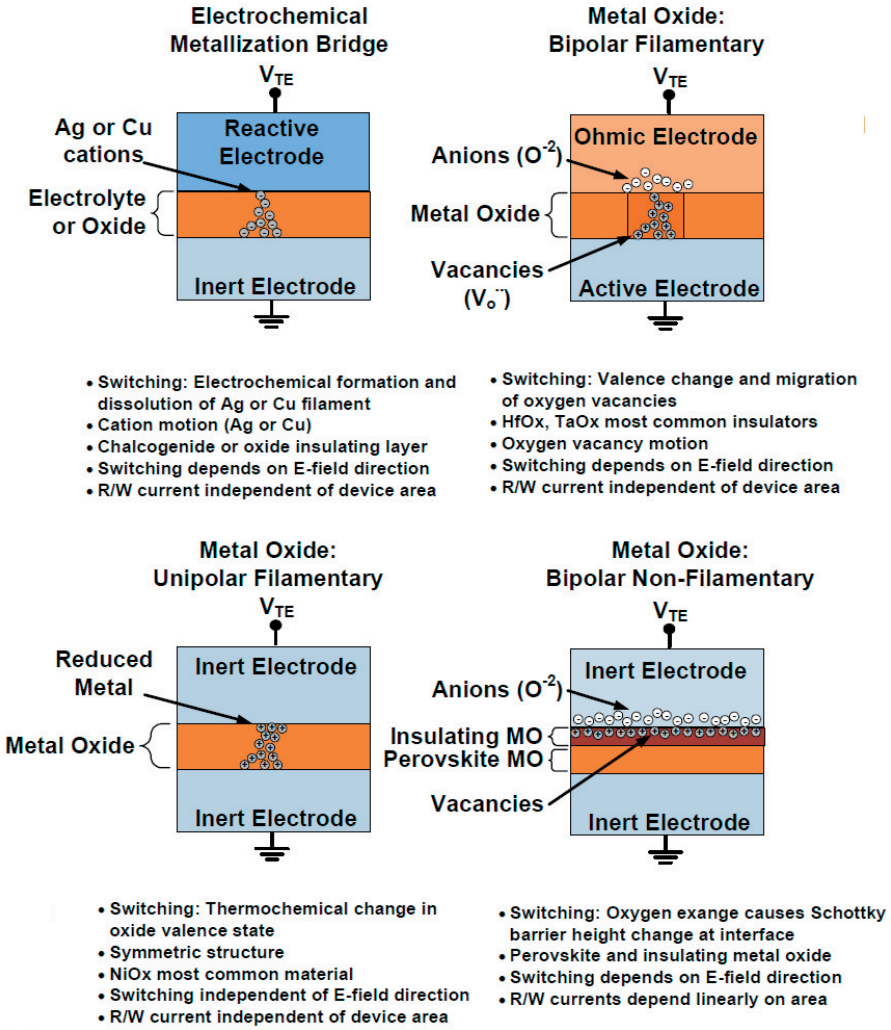


Figure 1.7: Various redox mechanisms believed to be responsible for resistive switching in the metal oxides [Marinella, 2014].

the resistance levels, favoring MLC operation. Different material and electrode combinations are currently being explored for precise controllability of the resistance to achieve multilevel operation [Terai et al., 2010]. The largest number of resistance levels reported so far are five levels, without verification for the HfO_x memory [Lee et al., 2008]. They also exhibit good endurance cycles. Lee et al. demonstrated an endurance of 10^{10} cycles using 40 ns program/erase pulses while simultaneously maintaining a resistance ratio of 20 in HfO_x-based memory devices [Lee et al., 2010]. Chip-level demonstrations have been made, confirming that these devices are CMOS-compatible [Fackenthal et al., 2014]. With all the desired capabilities and features, they are believed to have huge potential for mass production in the near future.

1.3.4 Phase-Change Random Access Memory (PCRAM)

Phase-Change Random-Access Memory (PCRAM) is one of the emerging non-volatile memory technology based on a chalcogenide alloy (combination of Germanium, Antimony and Tellurium) material, where the data is stored on the resistance of these materials [Wong et al., 2010; Burr et al., 2010]. These materials are similar to those commonly used in optical storage devices such as compact discs. PCRAM exploits the unique behavior of the phase-change materials. Electrical switching in chalcogenide glasses was first reported as early as 1960 [Ovshinsky, 1968]. However it was only in the past decade that renewed interest was shown in these materials owing to the electrical switching capability of the nano-scale volume of these materials.

These materials exhibit two easily inter changeable stable states with distinct physical properties, namely, the crystalline (low-resistiviy) and the amorphous (high-resistiviy) state. The phase transition between these two states is achieved through Joule heating, resulting from the current flowing through the material. In recent times, they have attracted significant research interest because of their desirable features.

The features of PCM are very impressive, and make it one of the top contenders among the emerging non-volatile memory technologies. The technology is sufficiently mature enough and chip-level demonstrations with large memory density have already been made [Close et al., 2013; Kim et al., 2010; Bedeschi et al., 2009]. It is one of the most promising candidates that can overcome the performance and scalability limits of the almost universal flash memories. Recent developments have shown that one can attain even higher density through the use of intermediate resistance states, known as multilevel-cell (MLC) operation [Bedeschi et al., 2009]. The modeling and reliability aspects of MLC PCM are the focus area of this thesis.

Attributes	FRAM	MRAM	RRAM	PCM
Non-Volatile	Yes	Yes	Yes	Yes
Cell Size	Large (20-40F^2)	Medium (6-12F^2)	Medium (6-12F^2)	Good (4-6F^2)
2014 latest tech. node	130 nm	90 nm	45 nm	45 nm
Write latency	≈50 ns	≈10 ns	≈20 ns	≈80-100 ns
Endurance	High (10^{12})	High (10^{12})	Low (10^5-10^6)	Medium (10^8)
Power	Low	Medium/Low	Low	Medium
Cost/GB	High	Medium	Medium	Medium
Suppliers	—	Everspin	Adesto	Micron, Samsung

Table 1.1: Basic comparison of the emerging non-volatile memory technologies [Meena et al., 2014]. The characteristics of PCM look quite promising, making it a top contender for future storage applications.

1.4 Phase-Change Memory Technology

1.4.1 Concept

Phase-Change Memory (PCM) is a class of non-volatile solid-state memory technologies, developed by exploiting the unique behavior of the chalcogenide materials [Ovshinsky, 1968]. Although the principle of reversible electrical switching in phase-change materials have been demonstrated as early as 1960, it did not kindle much interest owing to its very low crystallization speed. Renewed interest in PCM technology was triggered only after the discovery of materials such as $Ge_2Sb_2Te_5$, which are capable of fast crystallization (<100 ns). These discoveries led to the widespread use of these materials in the optical re-writable high-density data storage technologies (DVDs, CDs and Blu-Ray). Currently, electrical pulses are used to achieve memory switching in nano-scale volume of these materials to achieve high memory density.

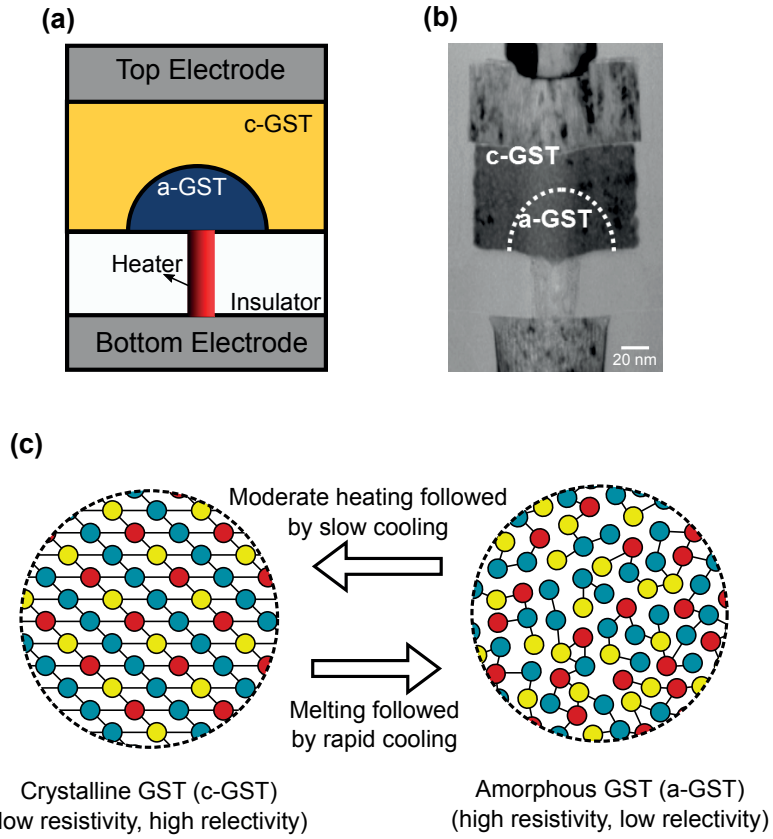


Figure 1.8: (a) The cross-section schematic of the conventional mushroom topology PCM device. The electrical current flows through the phase-change material between the top electrode and the heater. Joule heating in the phase-change material closer to the heater results in the active region. (b) TEM image of the mushroom-type device [Source: IBM T. J. Watson Research Center, USA.]. (c) Two stable states of PCM, which can be interchangeably switched by the application of suitable electrical pulses.

1.4.2 Basic characteristics of PCM

PCM technology is based on a chalcogenide alloy (typically $Ge_2Sb_2Te_5$, GST) material sandwiched between the top and the bottom electrodes. Figure 1.8 depicts a typical mushroom-topology PCM device. These materials exhibit two stable states, namely, a low-resistive crystalline state and high-resistive amorphous state. The capability to store data arises from the large resistance contrast these materials exhibit between these two states. Typically, the resistance in the crystalline state is three to four orders of magnitude lower than that of the amorphous state, implying that they have a high ON/OFF ratio of their resistance.

When an electrical pulse is applied, the Joule heating generated by the current flow can heat up the material and can bring about the phase transition. By the application of appropriate electrical pulses, the device can be set into either the low-resistance state (SET) or the high-resistance state (RESET).

1.4.3 SET and RESET Operation

The phase-transition process between the amorphous and the crystalline state is illustrated in Fig. 1.9. During the SET operation, the phase-change material is crystallized by applying an electrical pulse that heats a significant portion of the cell above its crystallization temperature. In the RESET operation, a short and high-amplitude current pulse is applied, so that the molten material quenches into the high-resistance amorphous state. The read operation can be performed by measuring the resistance at low voltage, such that the state of the device is not disturbed. These operations are summarized in Fig. 1.9. The overall operating speed of the PCM devices could be determined by taking into account the following steps:

1. **Read**
   The Read operation depends on the speed with which two (or more) resistance states can be reliably distinguished. Thus it is dominated by the readout architectures, circuit considerations such as bit-line capacitance, electrical circuit noise, etc.

2. **RESET programming**
   The high-resistance amorphous state can be attained by heating the material above its melting temperature by a short-duration electrical pulse, followed by rapid cooling. The kinetics of attaining the RESET state is quite fast. Electrical pulses as short as 400 ps have been shown to switch GST into the amorphous state [Wang et al., 2008].

3. **SET programming**
   The kinetics of crystalline formation is much slower than that of the amorphous state. Therefore, the SET operation dictates the write-speed performance of the PCM, as the required pulse width depends on the crystallization speed of the phase-change material [Krebs et al., 2008]. Thus, the crystallization process determines the achievable write bandwidth.

1.4.3.1 Crystallization temperature

A particularly important property of phase-change materials is the crystallization temperature. This is the lowest temperature at which the crystallization process becomes more favorable. It is typically measured by raising the temperature slowly while monitoring the crystallinity (either looking for X-ray diffraction from the crystalline lattice or the associated large drop in resistivity). The time and temperature required for crystallization are critical factors as they define the write bandwidth and the limits of reliable data storage.

1.4.3.2 Threshold switching voltage

The active material in an electrical PCM device needs to be switched between the low conductive amorphous and the highly conductive crystalline phase. This is done via Joule heating due to an electric current that passes through the memory cell. Because of the high conductivity

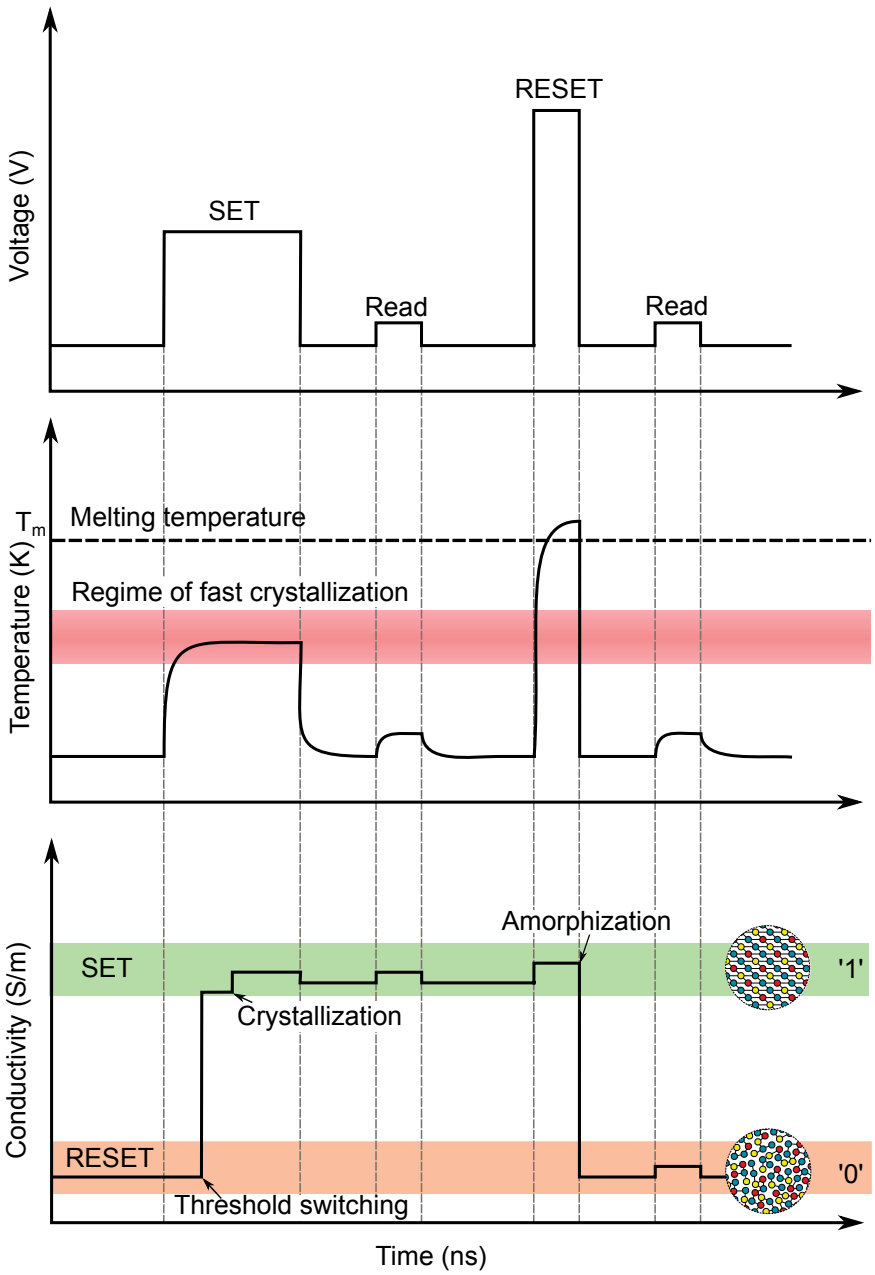


Figure 1.9: SET and RESET programming pulses for the phase transition. To SET a device into the crystalline state (green), an electrical pulse is applied such that the Joule heating power heats up the material into the regime of fast crystallization (red). To RESET the device into the amorphous state (orange), a sharp pulse of higher amplitude is applied that heats the material over the melting temperature T_m, followed by rapid cooling. To read the device content, a low-power pulse is applied that does not heat the material significantly.

in the crystalline state a sufficient amount of power (I^2R) can easily be attained via an electric current. In the high-resistive amorphous state, however a large voltage needs to be applied to have a substantial current flow for Joule heating.

Fortunately, in 1968 Ovshinsky observed that the conductivity of the amorphous state increases vastly at a certain voltage called threshold voltage [Ovshinsky, 1968]. It was also observed that the current exhibits a highly non linear dependence on the applied voltage in both the low conductive amorphous state and the high conductive crystalline state. The threshold switching phenomenon is described in detail in Section 2.3.1.2.

1.4.4 Properties of Phase-Change Memory

PCM is a non-volatile memory technology with fast read/write access, typically on the order of 100-200 ns for single-level cell (SLC) operations [Wong et al., 2010]. The crystallization process defines the speed of operation and the RESET operation, is the power-hungry process as it involves heating the phase-change material above its melting temperature.

The wide separation between the SET and RESET states in terms of resistance, ensures that there is sufficient noise margin during fast readout operations. The devices can be programmed to any desired state by application of suitable electrical pulses. The stored data can be readout without disturbing the current cell content.

PCM cells have high endurance (on the order of 10^8), allowing many switching cycles between the SET and RESET states. They also have long data retention periods. Data retention basically depends on the cell's ability to retain the amorphous state by avoiding unintended crystallization. One important feature of PCM is that data retention is independent of the endurance, which means the date retention time will be identical, irrespective of the history of programming cycles between the SET and RESET states. The typical criterion of 10 years of data retention at 85°C have easily been met by PCM technology [Pirovano et al., 2004; Gleixner et al., 2007].

The technology is CMOS-compatible, and the PCM devices are fabricated on a layer that can be introduced between the front end of line (FEOL) and the back end of line (BEOL) of the standard CMOS technology [Breitwisch et al., 2007]. Figure 1.10 shows the transmission and scanning electron microscope (TEM & SEM) images of a cross section of a PCM array, integrated in a 90-nm CMOS process flow. Also, experiments have verified that PCM is a very promising technology with respect to scalability. Xiong et al. experimentally demonstrated phase-change switching at the 10-nm scale using nanowire electrodes [Xiong et al., 2011]. This is one of the most desirable feature, while other technologies are limited by serious scalability issues (eg., Flash memory).

One of the key features of PCM that made it even more attractive among its competitors is the possibility of storing additional information in the intermediate resistive levels, typically knowns as multilevel-cell (MLC) storage. The ability to store and retain more than one bit of information per cell (MLC operation) is of significant importance, as this increases the effective memory density. Close at el. demonstrated a fully functional 256-Mcell chip with more than 2-bits/cell memory capability [Close et al., 2013]. Overall, PCM is one of the most

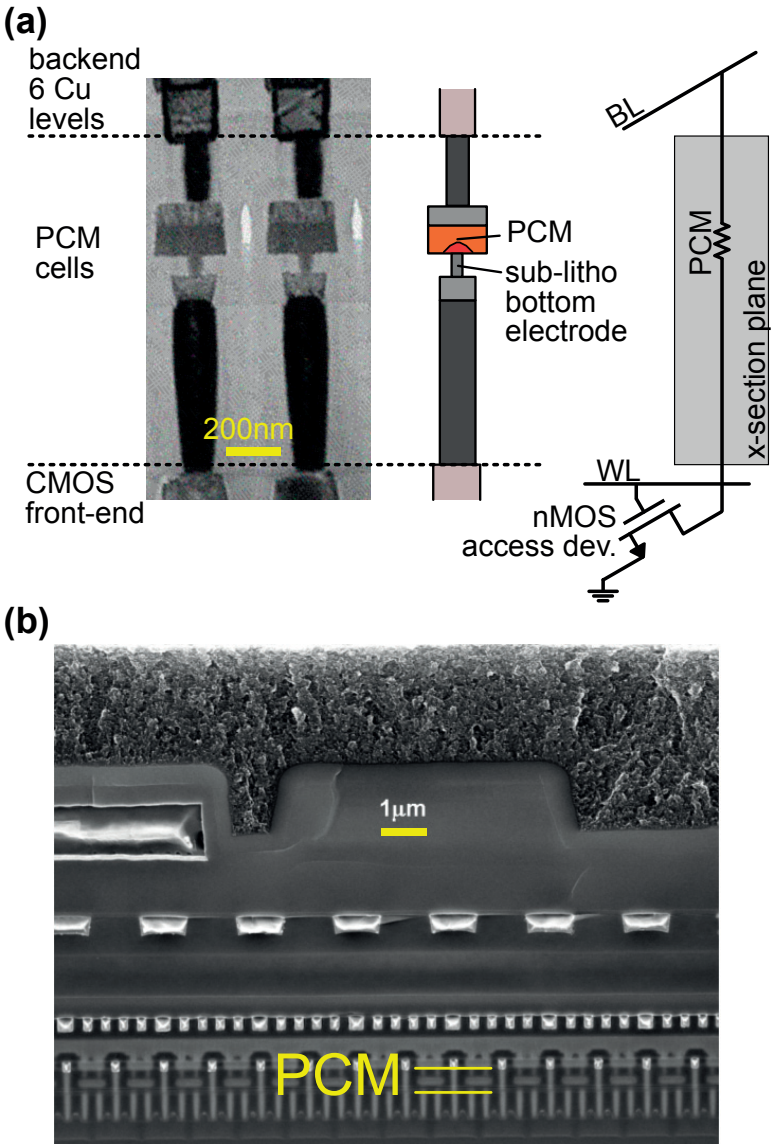


Figure 1.10: (a) Transmission electron microscopy (TEM) image of PCM cells and illustrative schematic across the bit-line (BL) direction. (b) Scanning electron microscope (SEM) image of a cross section of the PCM array, integrated in a 90-nm CMOS process flow, in the word-line (WL) direction.

promising candidates that can overcome the performance and scalability limits of the almost universal flash memories.

1.4.5 Comparison of PCM properties with those of other memory technologies

The driving factor behind the significantly increasing research efforts in exploring novel devices for data storage can be understood from the following perspectives:

- From the system point of view, the performance of the processor is increasingly limited by the access time between the processor and the memory unit. The power consumption of such memory sub systems is also critical in the design and operation of such systems.
- In the past decade, the proliferation of Flash-based memory systems in both the enterprise and consumer-grade storage space has highlighted the huge potential for high-capacity and high-speed memory technology in the memory hierarchy. The potential replacement of hard-disk drives (HDD) by Flash for selective applications supports this claim.
- As the need for high-capacity, high-performance storage systems keeps increasing because of the evolving big-data applications, even the universal Flash memories will not be able to meet the growing requirements owing to the scalability limitations of Flash.

Hence, any emerging technology should be able to address all the above-mentioned issues for it to be considered as a potential replacement for existing technologies. Therefore, the success of these emerging technologies depends on how well they perform compared with the current technology. In most cases, there should be remarkable performance improvement as the potential replacement would mean considerable modifications to the already existing storage system space. Table 1.2 summarizes the comparison of PCM with other existing memory technologies [Eilert et al., 2009]. Given that the characteristics of PCM most closely approximate that of the dynamic random access memory (DRAM) and the Flash memory, it exactly falls into the class of SCM memory technology. Therefore, it is worthwhile to briefly compare the properties of PCM with those of the currently predominant memory technologies, DRAM and Flash.

1.4.5.1 PCM vs. Flash storage

As discussed in Section 1.2.2.1, there are two kinds of Flash memories, NOR and NAND. NAND flash can be packed more densely than NOR flash, and though NOR flash offers fast random access its programming throughput is much lower than that of block-based NAND memory architectures. NAND memory is a high-density, block-based architecture with slower random access that is mainly used for mass storage applications.

The read/write endurance for both NOR and NAND is around 20 k–30 k cycles. In the sub 20 nm regime, the endurance is only around 5 k–10 k cycles. The properties of Flash are well within the reach of PCM capabilities. NOR Flash with its floating-gate technology has reached its scalability limits, mainly because of the difficulties in scaling the thickness of the tunnel oxide.

Thus PCM can be considered as a potential replacement for NOR flash in the near future. In contrast, replacing NAND flash is much harder, at least for the current trends of PCM, despite

Attributes	PCM	DRAM	NAND	HDD
Non-Volatile	Yes	Yes	Yes	Yes
Cell Area	6-8F^2	4F^2	5F^2	—
Erase Required	Bit	Bit	Block	Sector
Software	Simple	Simple	Complex	Simple
Power	≈100−500 mW/die	≈W/GB	≈ 100 mW/die	≈10 W
Write Bandwidth	1−100+ MB/s/die	≈GB/s	1−100 MB/s/die	200−400 MB/s
Write latency	≈200−1000 ns	≈20−50 ns	≈100 μs	≈10 ms
Read latency	50−100 ns	50 ns	10−25 μs	≈10 ms
Idle power	≪0.1 W	≈W/GB	≪0.1 W	≤10 W
Endurance	10^8	$\propto$	$10^4 - 10^5$	$\propto$
Data retention	Not f (cycles)	ms	f (cycles)	Not f (cycles)

Table 1.2: Comparison of PCM technology with the currently predominant memory technologies in the memory hierarchy [Eilert et al., 2009]. The characteristics of PCM most closely approximate those of dynamic random access memory (DRAM) and Flash memory.

its superiority in both endurance and read performance. The cost criterion is the biggest challenge, as NAND flash is mainly used in consumer electronic devices, where cost is of serious concern. In spite of the scalability limits, the memory density of NAND Flash can further be improved by the trap storage technology and possibly 3D integration [Choi and Park, 2012; Kim et al., 2012]. In terms of MLC realization, NAND has been shipping 2-bits/cell for years and even 3-bits/cell more recently, albeit with much lower endurance.

PCM with its wide resistance range is capable of MLC operation. PCM with 2-bits/cell have already been demonstrated, although on a smaller scale [Close et al., 2013]. Although the write operations are slow for NAND flash, impressive write data-rates can be achieved as its low write power allows the programming of many bits in parallel. Parallel programming is limited by the power requirements for PCM programming. Overall, PCM has three major advantages over flash technology:

- **Better scalability** - PCM is scalable friendly and offers better reliability at lower technology nodes. Data retention in Flash decreases significantly with the scaling of gate-oxide thickness.

- **Faster Write** - Faster programming bandwidths can be achieved in PCM than in Flash. With PCM, a cell can be written repeatedly without any intervening operations. In contrast, a programming operation in flash requires the corresponding cell to be erased before programming. This contributes to a major performance decrease for random

write operations. Typical Flash programming times are in the order of milliseconds, whereas they are on the order of few hundreds of nanoseconds in PCM.

- **Endurance** - The endurance of PCM is several orders of magnitude higher than that of flash. Also unlike the Flash, the data retention of PCM is independent of the endurance cycling.

1.4.5.2 PCM vs. DRAM

Although PCM is a non-volatile memory technology, it is worthwhile comparing it with volatile memory technologies (such as SRAM and DRAM), as PCM is considered to be one of the top contender for storage class memory. A typical SRAM cell uses six CMOS transistors to store each memory bit and occupies more than $120F^2$ in chip real estate[1] per bit which is much larger than a PCM cell [Burr et al., 2010]. SRAM cells are widely used in cache memories, where they typically run at the CPU clock speed, so that the access time must be less than 10 ns. The access time of PCM is limited by its write speed (SET pulses), which depends on the crystallization speed of the phase-change material. Although, some researchers have demonstrated the use of SET pulses shorter than 10 ns [Bruns et al., 2009], more realistic large-array demonstrations tend to require SET pulses that range from 50 to 500 ns in duration.

Both PCM and DRAM are power-hungry technologies. The power-hungry nature of DRAM is not solely due to its periodic refresh, which takes place only infrequently, while compared to the achievable data rate. Instead, the power inefficiency in DRAM is due to the simultaneous addressing of multiple banks within the chip. In DRAM, there is an inherent need to re-write the data after each read access. Thus simply by being non-volatile, PCM could potentially be an alternative, albeit at longer latency to DRAM, despite the inherently power-hungry nature of PCM write operations. In the case of stand-alone memories, the cost is directly proportional to memory cell size. State-of-the-art DRAM cells occupy $4F^2$ in chip area. Such small cell sizes have already been demonstrated in PCM using a diode as access device. PCM also competes favorably with DRAM in terms of scalability in future generations, as DRAM developers are quickly hitting various scaling limits associated with storage interference, device leakage, and challenges in integrating high aspect-ratio capacitors in tight spaces. Currently, DRAM has fallen behind NAND Flash and standard CMOS logic technologies in terms of scaling to the 32-nm technology node and preparation for the sub 20-nm node.

1.4.5.3 PCM as Storage Class Memory

Given the present scenario, PCM is the most promising NVM technology for storage class memory. It has been projected that initially the cost per bit for PCM will be much higher than that of disks and comparable to that of flash. Thus, it will be used in systems in which size (e.g.,

[1] Semiconductor device technology node is commonly described by the minimum feature size F that is available via lithographic patterning. This measure of device size is independent of the particular device technology used to fabricate the memory.

in mobile devices), performance, and/or reliability in harsh environments are paramount. Soon, the cost per bit will decrease and will approach that of disk drives. This is possible, as the memory density of PCM can be increased using the methods discussed earlier.

1.5 Multi-level cell (MLC) storage in PCM

The MLC approach is to exploit of the intrinsic capability of the memory device to store analog data in order to encode more than 1 bit of information per device. In the case of PCM technology, the large resistivity contrast exhibited by these materials between the crystalline and the amorphous state (typically 3–4 orders of magnitude, and hence the high ON/OFF ratio achieved), favors the possibility of storing information in the intermediate resistance states.

1.5.1 Factors limiting reliable MLC operation in PCM

Despite the large resistivity contrast that favors MLC storage, MLC operation in PCM is quite challenging as conventional write/read methods are not efficient enough to reliably store and retrieve the data. There are various factors that can seriously limit the number of effective intermediate levels that can be reliably stored and read back in a PCM device. Some of the prominent factors are

- resistance drift,
- device variability across the PCM array,
- intrinsic PCM noise and
- circuit noise from the programming/readout circuitry.

A detailed description of MLC realization in PCM is presented in Section 4.1. The factors affecting MLC are further investigated and analyzed in depth in Section 4.2.

1.6 Research focus area : State of the art

1.6.1 Modeling of PCM devices

Given that phase-change switching is based on the coupled electrical and thermal transport mechanisms within these materials, it is critical that these physical processes are modeled to have a precise understanding of the device operation. The model can then be used as a tool to provide numerous insights into the device, which are difficult or even impossible to obtain otherwise from direct experimental measurements. Also, given the complex procedures involved in the fabrication of device prototypes, device modeling becomes increasingly important in saving the cost and time of analyzing and understanding the characteristics of exploratory device structures.

In the past decade, various research groups across the globe came up with models, ranging from compact to complex, to understand various critical aspects of the PCM device (threshold switching model, data retention model, crystallization, structural relaxation etc.). There are numerous models for the electrical conduction and switching in the phase-change materials [Rizzi and Ielmini, 2012; Ielmini and Zhang, 2007; Sebastian et al., 2015; Ielmini et al., 2007].

The power-hungry RESET process is one of the growing concerns for the sustainability of PCM technology. The main motivation for this work comes from the unusual characteristics of these devices for opposite bias-polarity conditions. Dong-Seok et al. demonstrated the change in efficiency of these materials with n-type and p-type materials used as bottom electrodes [Dong-Seok et al., 2010]. Castro et al. experimentally found evidence of a thermoelectric Thomson effect within the PCM devices [Castro et al., 2007]. This kindled the interest in understanding the thermoelectrics in these devices as they are believed to play a significant role in the programming of these devices. Lee et al. proposed that by including the thermoelectric physics in modeling, they could observe a 16% reduction in the programming current [Lee et al., 2012]. Researchers used finite-element-based modeling approaches to study the thermoelectric influence on device operation [Faraclas et al., 2014; Ciocchini et al., 2015].

1.6.2 Reliability of MLC PCM

MLC storage in PCM is seriously hampered by the phenomenon of resistance drift and cell array variability. Various MLC enabling technologies have been explored to tackle the issue of resistance drift and variability in order to perform reliable data storage and data retention. Notably, a novel non-resistance cell-state based metric which is shown to have superior drift performance compared with the traditional low-field electrical resistance readout metric has been proposed [Sebastian et al., 2011]. Also, novel iterative schemes have been proposed that have been shown to be very effective in programming the intermediate resistance levels in the presence of cell variability [Papandreou et al., 2011].

Several MLC enabling technological advancements suppressing the drift phenomenon and array variability have been explored and their impact on the reliability and data retention has been analyzed [Pozidis et al., 2013]. Drift-tolerant coding schemes have been proposed and shown to have better drift robustness [Pozidis et al., 2015]. Such advancements renders it likely that reliable MLC storage at the highly-dense PCM arrays could be demonstrated and the viability of the MLC PCM technology can be established.

1.7 Thesis Goal

The goal of this thesis is to provide a comprehensive modeling and reliability framework for establishing the viability of MLC PCM technology.

Device modeling

1. **A comprehensive thermoelectric model:** Programming power is one of the major concern for increasing the competitiveness of the PCM technology. With many researchers reporting the remarkable positive influence of thermoelectric physics on the programming of PCM devices, I propose to have a complete finite-element method (FEM) based analysis of the thermoelectric phenomenon in nano-scale PCM devices.

   The availability of in-house macro devices (PCM devices without the access transistors) for experimental measurements is an added advantage, because such devices can be electrically characterized and the actual measurement data was used to validate the FEM model developed. Therefore, one of the thesis goal is to have a full-fledged finite-element method based model which can accurately predict the device characteristics. The model can provide numerous insights into the device operation and can also be used as a tool to fine-tune the material properties and the device design for the development of novel power-efficient PCM devices.

Metrics and Architectures for MLC PCM

2. **Drift-resilient readout metrics in MLC PCM:** Given the various factors affecting the MLC storage operation and therefore the sustainability of MLC PCM as a technology, numerous MLC enabling technologies have been investigated. Novel non-resistance-based readout metrics offers a promising scope for MLC operation [Sebastian et al., 2011].

   In order to demonstrate the versatility of such metrics, they should be extracted from dense PCM arrays. The metric-extraction scheme should therefore be implemented as a CMOS circuit architecture and co-integrated in a prototype chip with a dense PCM array. Novel programming schemes should also be explored for the integration of programming electronics in the prototype chip. Therefore, one of the goal is to design a novel read and write architecture favoring the MLC programming and the non-resistance metric extraction, in CMOS technology for integration with PCM array.

Reliability analysis of MLC PCM

3. **Reliability and data retention analysis of PCM arrays at elevated temperatures:** Once the non-resistance-based readout metrics are designed and implemented as programming and readout architectures, their performance should be investigated to determine the effectiveness of both the metrics and the circuit implementation.

In PCM, the stored levels are seriously affected by the resistance drift phenomenon. Being a stochastic process and getting accelerated by high temperatures, temperature fluctuations across the system environment also contribute to the reliability of the data retrieval process. Device variability at the array level impacts reliable storage operation by causing a broader resistance distribution. Apart from them, the reliability concerns such as endurance cycling and the performance at elevated temperatures, should all be met by the proposed metrics. Therefore, the goal of this thesis is to demonstrate reliable MLC storage operation at the PCM array-level in an ensemble of cells which have subjected to a large number of endurance cycles and elevated temperatures.

The first half of the thesis focuses on the implementation of a comprehensive thermoelectric model that can accurately capture the PCM device characteristics. The second half of the thesis focuses on a novel CMOS circuit implementation of the drift-resilient metrics in the 64-nm technology node. Moreover, also the reliability and data retention characteristics of the proposed readout metrics are demonstrated.

1.8 Summary of contributions in this thesis

I summarize the various contributions which I have made in the fulfillment of this thesis. In the preceding chapters, I have made the following contributions with respect to the modeling and reliability framework for MLC PCM:

1. **Device Modeling:**
    - In Chapter 2, I devised a comprehensive finite-element method (FEM) based thermoelectric model for capturing the thermoelectric effects in nano-scale PCM devices.
    - Various physical aspects governing the device operation were developed and implemented in COMSOL Multiphysics®.
    - In Chapter 3, I conceived the compact electro-thermal model approach to quickly evaluate the temperature distribution within the PCM arrays, mainly focusing on the thermal disturb analysis
    - I proposed novel device-design based on the numerous insights I obtained from the thermoelectric model for PCM devices.
2. **Metrics and Architectures for MLC PCM:**
    - In Chapter 4, I proposed a novel readout architecture for the extraction of drift-resilient readout metrics in MLC PCM storage.
    - I optimized the readout circuit design presented for power and speed, while still maintaining the required flexibility for exploration of the non-resistance based metric schemes.

- I characterized the prototype chip fabricated using a FPGA-based hardware characterization platform. The performance of the readout circuitry in terms of access time (450 ns) and the accuracy of the readout data (5-bit effective ADC resolution) were investigated.
- I was instrumental in the design and implementation of the iterative programming algorithms for MLC programming.
- I contributed significantly in the top-level design of the prototype chips in terms of the chip interface, design of the digital controller, array address decoders, various test point buffers, and output for monitoring the internal chip signals, etc.
- I was involved in the circuit design implementation for the eM-metric extraction.

3. **Reliability analysis of MLC PCM:**

- In Chapter 5, I investigated and also demonstrated reliable MLC storage over an ensemble of 64k cells PCM sub-array, which had been subjected to a million endurance cycles and temperature variations.
- I devised the temperature control component of the experimental characterization platform for inducing temperature variations across the prototype chip.
- I implemented various routines and procedures in the FPGA-based hardware characterization platform for the emulation and extraction of various performance based aspects of the prototype chip.
- I also demonstrated the feasibility of 3 bits/cell in PCM arrays by optimal level placement, provided that temperature variations can be kept under control.

This thesis has been performed under the framework of the IBM-SK Hynix Joint-Development Agreement (JDA) for PCRAM commercialization. This project aims to develop the first commercial MLC PCM chip with multi-gigabit capacity for mass production. Novel CMOS circuit implementation of the READ and programming architectures for MLC operation and various aspects in designing the prototype chip are all the result of the close collaborative efforts between the design engineers at IBM Research in Zurich, IBM T. J. Watson Research and SK-Hynix. The devices and prototype chips used for the experimental measurements are all fabricated by our colleagues at the IBM T. J. Watson Research center.

1.9 Thesis Organization

The remainder of the thesis is organized as follows. The first two chapters focus on device modeling aspects, whereas the last two chapters focus on a reliable MLC realization.

- **Chapter 2** presents the theory and implementation of the comprehensive thermo-electric model in COMSOL Multiphysics® software. Practical thermal and electrical

boundary conditions used to simulate the model are defined. All field- and temperature-dependent material properties used for the model simulation are described.

- **Chapters 3** describes the characterization platform used for the experimental measurements. It is then followed by the validation of the model by comparing the simulation results with measured data. Various insights from the simulated model are given, along with two novel device-design ideas for efficient PCM devices. The thermal disturb reliability factor and a compact model for the fast evaluation of thermal disturb in dense PCM arrays are presented.

- **Chapter 4** describes MLC storage operation in PCM and various factors limiting it. Various MLC enabling technologies for the realization of MLC in PCM are presented. Novel drift-resilient readout schemes and iterative programming algorithms are reported, with the majority of the chapter focusing on the circuit-level implementation of these programming and readout metrics for a prototype chip.

- **Chapter 5** deals with various reliability concerns of MLC PCM, especially at the array level. The experimental hardware platform used for the characterization of various readout metrics is presented. The influence of endurance cycles on MLC programming is investigated. Reliable MLC storage on a 64k cells PCM array that has been subjected to a million endurance cycles and a history of elevated temperature profile is demonstrated.

- **Chapter 6** summarizes the thesis contributions and the impact of the research in the field. Finally, possible future research directions are also briefly presented.

Note:

1. **Because of the diverse nature of the individual contributions of this thesis, a brief description about the background material and previous work in the respective topics will be given at the beginning of the chapters for clarity in reading.**
2. **For brevity, the term "Multilevel-Cell" will simply be referred to as "MLC" and the term "Phase-Change Memory" will simply be referred to as "PCM" in the remainder of the thesis unless stated otherwise. Also, "devices" or "cells" are used interchangeably to represent the PCM device. Similarly, "states" or "levels" are used interchangeably to represent the resistance state.**

2 Finite-element method based modeling of Phase-Change Memory devices

In this chapter, I propose a comprehensive finite-element method (FEM) based model for analyzing the influence of thermoelectric physics in phase-change memory devices. The model was implemented and simulated in a FEM-based modeling software knowns as COMSOL Multiphysics®. The novel model uses realistic field- and temperature-dependent material properties and interface conditions of the materials supported by practical thermal and electrical boundary conditions. Even though the proposed model is solely focused on implementing the mushroom-type device topology, the concepts can easily be extended to any other type of device topology (e.g., confined-cell topology, μ-trench cell topology, etc.) by simply modifying the device geometry and the corresponding material properties in the model.

This chapter is organized as follows: **Section 2.1** reviews the basic concept of finite-element method based numerical modeling and briefly describes the COMSOL simulation tool. **Section 2.2** describes the motivation for PCM device modeling. **Section 2.3** presents the theory behind PCM device operation, and the necessary modifications required to include and model the thermoelectric effects in PCM devices. **Section 2.4** describes the modeling approach used to implement and simulate the thermal and electrical transport mechanisms within the device in the COMSOL Multiphysics software. The implementation of realistic thermal and electrical boundary conditions with the modifications necessary at the interface to capture the thermoelectric effects are all discussed in this Section. Finally, **Section 2.5** deals with all relevant field- and temperature-dependent material properties required to simulate the model, and the assumed interface boundary conditions are explained.

Nomenclature

A	Area, m^2
C_p	Specific heat capacity, $J/kg \cdot K$
$\vec{E}$	Electric-field, V/m
E_a	Activation energy, eV
k_B	Boltzmann constant, eV/K
q	Electronic charge, C
L	Lorentz number, $W \cdot \Omega/K^2$
dz	Inter-trap distance, nm
m_e	Electron rest mass, eV/c^2
h	Heat transfer coefficient, $W/m^2 \cdot K$
$\vec{J}$	Current-density, A/m^2
$\vec{n}$	Unit normal area vector
$\vec{q}$	Heat flux, W/m^2
R_{th}	Thermal resistance, K/W
r_s	Radius of the system, m
S	Seebeck coefficient, V/K
P	Peltier coefficient, V
V	Voltage, V
T	Temperature, K
ΔT	Temperature difference, K
T_{amb}	Ambient temperature, K
T_m	Melting temperature of GST, K
Q_J	Volumetric Joule-heating heat contribution, W/m^3
Q_T	Volumetric Thomson heat contribution, W/m^3
Q_P	Volumetric Peltier heat contribution, W/m^3
Q_B	Volumetric Bridgman heat contribution, W/m^3
Q_{iP}	Volumetric interface Peltier heating compoent, W/m^3
t	Time, s
$\nabla(\ )$	Gradient operator
$\mathbf{F}(\ )$	Fermi integral

Greek Symbols

ϵ_r	Relative permittivity (dimensionless)
σ	Electrical conductivity, S/m
κ	Thermal conductivity, $W/m \cdot K$
κ_{ele}	Electronic contribution of the thermal conductivity, $W/m \cdot K$
κ_{pho}	Phononic contribution of the thermal conductivity, $W/m \cdot K$
μ	Mobility, $m^2/V \cdot s$
ρ	Density, kg/m^3

2.1 Semiconductor device modeling

Semiconductor devices form the backbone of emerging non-volatile memory technologies. A whole new set of novel semiconductor devices is currently being explored using exploratory semiconductor materials. Owing to the rapid development in semiconductor-based devices, researchers have shown significant interest in device modeling in the past few decades. In view of the urge to precisely understand device operation at the nano-scale level and the need to optimize the design parameters (process flow, material properties, device geometry, etc.) of such complex semiconductor device, device modeling plays a vital role in modern technology.

With the growing demand for high-performing electronic gadgets and the increasing competition, exploratory nano-scale devices should be properly modeled to understand their operation before one can commercialize (mass-scale production) products based on such devices. However, the fabrication of devices costs a fortune and takes precious time, therefore device modeling offers a powerful tool to extensively investigate new devices and to optimize the device parameters without an explicit need to fabricate them. This demand in device modeling resulted in the proliferation of a wide variety of software-based device modeling tools, such as Ansys, COMSOL, Sentaurus TCAD, etc.

The principle that underlies these device-modeling tools is that the behavior of any semiconductor device can be formulated by equations governing the fundamental physics based on which those devices operate. Once formulated, closed-form analysis of such equations might lead to solutions; however the complexity of solving them increases when these equations are multi-dimensional and non linear in nature.

2.1.1 Numerical techniques

One common approach to solve such equations is to employ numerical techniques. Although they are time-consuming compared with the closed-form approach, they yield more versatile and very accurate results. The most common numerical techniques typically applied to solve a set of partial differential equations are the finite-difference method and finite-element methods.

The goal of the numerical modeling approaches is to approximate the solutions of the partial differential equations, i.e., to find a function (or some discrete approximation to this function) which satisfies a given relationship between various of its derivatives on a given region of space and/or time, considering the boundary conditions defined along the edges of this domain.

2.1.2 Finite-difference method (FDM)

A finite-difference method (FDM) proceeds by replacing the continuous derivatives in the partial differential equations (PDE) by finite difference approximations. This gives a large algebraic system of equations to be solved in place of the differential equation, something

that can easily be solved on a computer. FDM is the oldest such method and is based on the application of a local Taylor expansion to approximate the differential equations.

The continuous derivatives of the equations are replaced by discretized standard finite-difference approximations. The discretized variables are represented by values obtained from the solution of the equations at each mesh point, except where boundary conditions determine the parameter values.

The FDM principle can be best understood with an example. Although only the FEM approach was used in the thesis, some basic FDM knowledge helps in better understanding the FEM-based approach. Consider the one-dimensional heat transfer equation,

$$\rho C_p \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(\kappa \frac{\partial T}{\partial x}\right), \tag{2.1}$$

where x is the direction along which the temperature should be calculated. If the thermal conductivity, density and heat capacity are constant over the model domain, equation (2.1) can be simplified to

$$\frac{\partial T}{\partial t} = k \frac{\partial^2 T}{\partial x^2}, \tag{2.2}$$

where $k = \frac{\kappa}{\rho C_p}$ is the thermal diffusivity expressed in m^2/s. We are interested in the temperature evolution along the x-direction versus time $T(x,t)$ which satisfies (2.2), given an initial temperature distribution (initial condition).

The first step in the FDM is to construct a grid of points over which we are interested in solving the equation. This is known as discretization and is shown in Fig. 2.1. The next step is to replace the continuous derivatives of (2.2) with their finite difference approximations. The derivative of temperature versus time can be approximated with a forward finite difference approximation as

$$\frac{\partial T}{\partial t} \approx \frac{T_i^{n+1} - T_i^n}{t^{n+1} - t^n} = \frac{T_i^{n+1} - T_i^n}{\Delta t} = \frac{T_i^{new} - T_i^{current}}{\Delta t}, \tag{2.3}$$

where T_i^n represents the temperature at the current timestep and T_i^{n+1} represents the future temperature. The subscript i denotes the location of the grid point. Both n and i are integers; n varies from 1 to n_t (total number of time steps) and i runs from 1 to n_x (total number of discretized grid points in the x-direction). The spatial derivative of the temperature can then

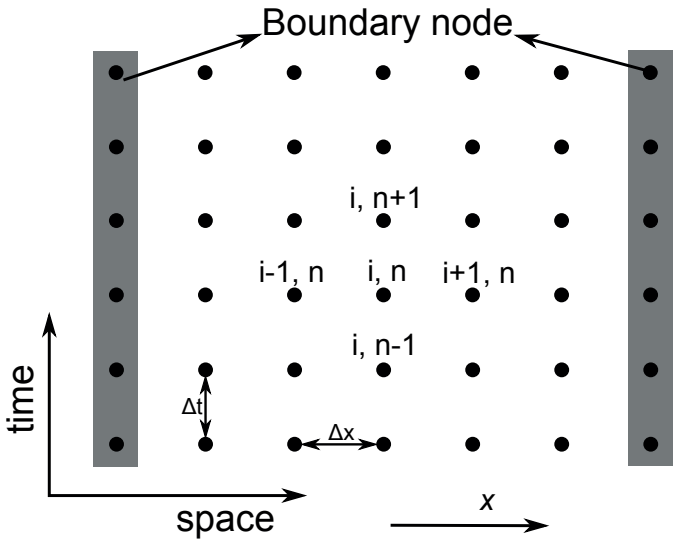


Figure 2.1: Finite difference discretization of the 1-D heat transfer equation. The finite difference method approximates the temperature at given grid points, with spacing Δx. The time evolution is also computed at given times, with a time step of Δt.

be found by a central finite difference approximation, given by

$$\frac{\partial^2 T}{\partial t^2} = \frac{\partial}{\partial x}\left(\frac{\partial T}{\partial x}\right) \approx \frac{\frac{T_{i+1}^n - T_i^n}{\Delta x} - \frac{T_i^n - T_{i-1}^n}{\Delta x}}{\Delta x} = \frac{T_{i+1}^n - 2T_i^n + T_{i-1}^n}{\Delta x^2}; \tag{2.4}$$

Substituting (2.3) and (2.4) into (2.2) results in

$$\frac{T_i^{n+1} - T_i^n}{\Delta t} = k\left(\frac{T_{i+1}^n - 2T_i^n + T_{i-1}^n}{\Delta x^2}\right); \tag{2.5}$$

By re-arranging (2.5) so that all known quantities are on the right-hand side and the unknown quantities on the left-hand side, we get

$$T_i^{n+1} = T_i^n + k\Delta t\left(\frac{T_{i+1}^n - 2T_i^n + T_{i-1}^n}{\Delta x^2}\right); \tag{2.6}$$

Since the equation at the current time step is known (initial value condition), the temperature at the boundary condition can be deduced from the boundary conditions set for the problem.

Therefore, (2.6) can be used to estimate the temperature in the subsequent time steps.

The main disadvantage of the FDM method is revealed when complicated geometries are being simulated. It is easy to see that finite difference methods run into difficulties when dealing with complex boundaries because of the need for spatial regularity of the grid. For example, using FDM to model curved surfaces will always results in an approximation of the curved surface by cubes. Increasing the mesh resolution to compensate for the surface also increases the resolution in open areas where the finer resolution is not required, leading to higher memory usage and longer computation times. If a free mesh is used, i.e., one with varying distances to neighboring points, to accommodate the irregular surface, the discretized equation loses accuracy. It is for these reasons that the FDM method is slowly getting replaced by finite-element-based methods [Frehner et al., 2008].

2.1.3 Finite-element method (FEM)

The finite-element method (FEM) provides a flexible means of investigating complex semiconductor device geometries. This method is fundamentally quite similar to the FDM approach, except that the geometric model of the device is sub-divided into smaller regions–the finite elements or domains (triangles or rectangles) rather than the regular grid points (FDM). Adjacent elements are held together by nodes. The continuous functions representing the required solution are assumed to have simple analytic forms in each element (piecewise-linear, cubic). The continuous semiconductor equations are then used to define an equivalent integral formulation. As with FDM, FEM also transforms the continuous functions of the semiconductor equations into discretized forms.

The response of each element is expressed in terms of a finite number of degrees of freedom characterized as the value of an unknown function (or functions) at a set of nodal points. Because FEM is a discretization method, the number of degrees of freedom of a FEM model is necessarily finite. They are collected in a column vector. This vector is generally called the DOF vector or state vector. Adjacent elements share the DOF at the connection nodes. Finite-element methods has become common place in recent years, and are now the basis of a multi-million dollar industry. Numerical solutions to even very complicated physical problems can now be routinely obtained using FEM analysis.

In summary, the FEM solution procedure involves the following;

- Dividing the geometrical structure into non-overlapping domains (elements with nodes) (discretization/meshing)
- Connecting (assembling) the elements at the nodes to form an approximate system of equations for the whole structure (forming element matrices)
- Solving the system of equations involving unknown quantities (dependent variables) at the nodes (e.g. temperature, displacements)

- Calculating the desired quantities (e.g., temperature, stresses, etc.,) at selected elements

2.2 Modeling of PCM devices

To gain a clear understanding of any semiconductor-based memory device, it is necessary to obtain a suitable representation of the physical process involved. In the case of PCM devices, the fundamental interplay between the thermal and the electrical transport mechanisms is essential in bringing about the phase transition process. Therefore, the electrical and thermal process associated with these materials should be understood thoroughly before making any efforts to improve the device characteristics or its performance.

With the enormous interest in PCM devices, there is a growing need to model these devices so as to investigate their properties and characteristics. As in any semiconductor device, the process and development of prototype PCM devices involve complex fabrication procedures with many iterations required to achieve the desired/intended specifications. Given the aggressive scaling trends and the focus on sub-lithographic nodes, fabricating such devices is all the more complex, and the time involved in optimizing the process flow is even more crucial. Device modeling comes to the rescue with its capability to be both cost- and time-efficient in realizing and understanding the complex physical processes in such devices. In the current trends, modeling plays an significant role in process-flow optimization, in finalizing the device geometry and in speeding up the development of prototype PCM devices. Apart from providing valuable insights into the device, which are otherwise impossible to attain through experimental measurements, device modeling can save the time substantially and cut the development costs by reducing the number of iterations required for optimizing the fabrication process.

2.2.1 Finite-Element Modeling using COMSOL Multiphysics

COMSOL Multiphysics® is a general-purpose software platform, based on advanced numerical methods, such as FEM, for the modeling and simulation of physics-based problems [1]. The software package is capable of combining any number of physical phenomena and applying them to any geometrical structure. Many standard system of equations are readily available (called modules), such as heat transfer, electrostatics, electrodynamics, etc. The modules which are mainly interesting for the PCM modeling are the *heat transfer* and *electric currents* modules.

[1] COMSOL Multiphysics, Version 4.4, Burlington MA, USA.

2.3 Thermoelectric model

2.3.1 Theory

There is an electrical and a thermal component associated with the programming of a PCM device. When an electric potential is applied across the device, the current flow results in substantial Joule heating and the generated power (I^2R) is dissipated across the device. Typically, the region where the majority of the power is dissipated is known as the active region. The location of the active region depends on the type of phase-change material, the electrodes and the device geometry. The dissipated power heats up the nanometric volume of phase-change material to very high temperatures (much above the melting temperature of the phase-change material) and hence initiates the phase-change switching process.

The applied electrical power should be increased such that the profound heating (followed by rapid cooling) results in a sufficient volume of phase-change material in the active region being switched to the amorphous phase, thereby increasing the device resistance and reducing the read current flow. The power required to create an amorphous plug such that it completely blocks the read current flow is known as the programming power or the RESET power. The required RESET power strongly depends on the device geometry and is one of the efficiency metrics of PCM devices. Figure 2.2 depicts the basic PCM operation.

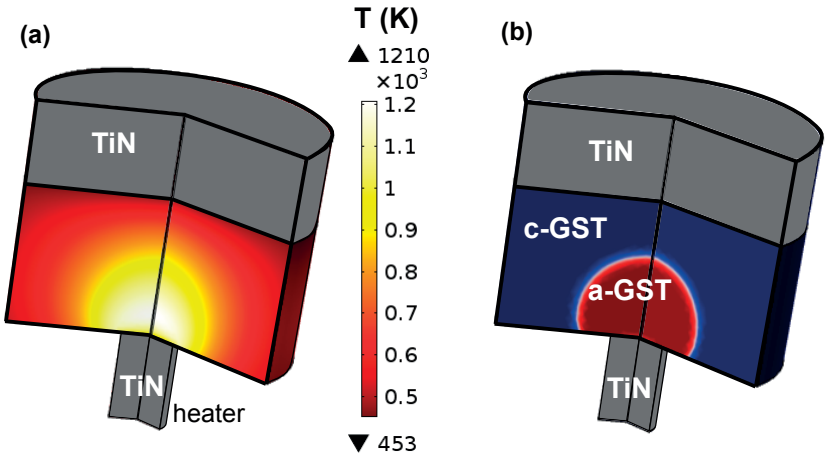


Figure 2.2: Mushroom-type device topology. (a) Thermal profile achieved within the device while a programming electrical pulse is applied. (b) Once programmed, the amorphous plug created blocks the read current flow.

2.3.1.1 Thermal system

For any applied input power, the peak temperature attained within the active region of the device is referred to as the “hotspot”. The location of the hotspot defines the switching dynamics and hence plays a critical role in determining the operating efficiency of the de-

vice. The hotspot location is greatly influenced by the device topology. For example, in a mushroom-type device topology, the hotspot is located just above the bottom electrode, whereas in a confined-cell device topology, it is located midway between the top and the bottom electrode. Apart from the device topology, other components, such as the field and temperature-dependent material properties, thermal boundary conditions (ambient temperatures), thermal conduction pathways and the interface resistances, significantly impact the hotspot location. Therefore, the temperature distribution attained within the device is critical as it provides valuable information regarding the heat loss paths and suggests ways to improve the device efficiency.

2.3.1.2 Electrical system

Based on the applied potential, the current flow through the phase-change material is determined by the electrical conductivity of the conducting medium, usually the phase-change material and the electrodes. Typically, the electrical conductivity of the electrodes is constant, whereas that of the phase-change materials brings about the progressive Joule heating, which eventually initiates the phase transition process. The electrical conductivity of these materials depends on either the applied electric field, the temperature or both. With increasing potential across the device, the electric field increases, accompanied by prominent increase in temperature (arising from the Joule heating power). At high fields, the conductance of the material increases drastically, eventually leading to threshold switching.

Threshold switching One of the interesting properties of these material, which made data storage feasible, is the phenomenon of the threshold-switching voltage. In the high-resistance regime, when the potential applied across the device is above certain voltage, V_{th}, the resistance of the device drastically drops to a low value, irrespective of the size of the amorphous region, and is even comparable to that of the crystalline resistance. This phenomenon is known as electronic threshold switching [Ovshinsky, 1968; Adler et al., 1980].

The electronic threshold switching phenomenon is the key to SET programming. When the device is in the high-resistance state, the low programming current is not sufficient to provide the Joule heating necessary to initiate the crystallization of the PCM device. The sudden drop in resistance after the threshold switching voltage enables the SET programming. Above V_{th}, the field is high enough such that the cell is in the so-called “ON state” and the corresponding resistance is denoted by “ON resistance”. Owing to its low resistance, the “ON state” is well suited for programming, whereas for the readout, the cell should be biased in the “OFF state", such that the cell content is not disturbed. Figure 2.3 shows the typical IV characteristic corresponding to a PCM device in the SET and RESET states.

The knowledge about the threshold switching voltage is critical in predicting device operation. From the reliability perspective, the read voltage should be well below V_{th}, so that the content of the device is not disturbed during readout. In contrast, the programming pulses should

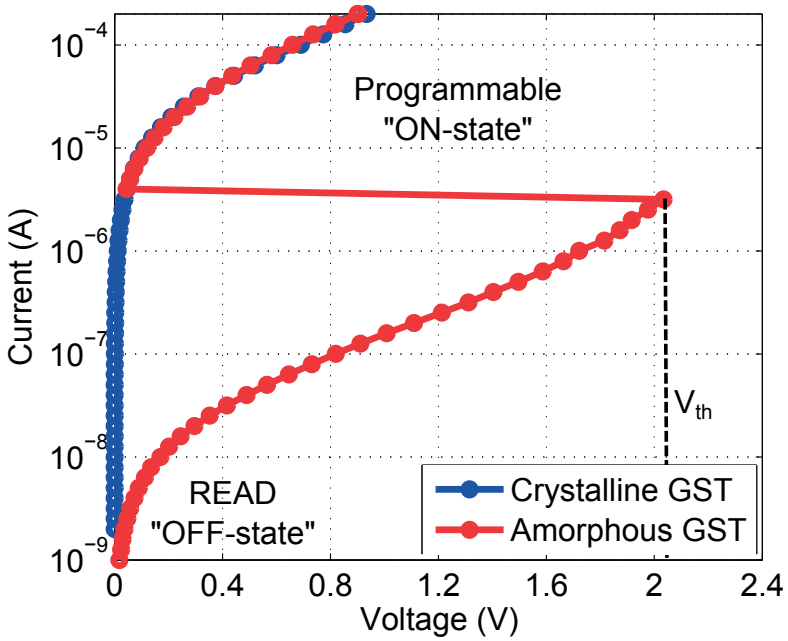


Figure 2.3: Experimental measurements of the $I-V$ characteristics of typical PCM devices. V_{th} is the threshold switching voltage, above which the resistance of the amorphous phase drops drastically.

always provide voltages that are higher than the threshold switching voltage for a successful programming of the SET and the RESET states. The other interesting thing is that V_{th} is dependent on the size of the amorphous region and that different intermediate resistive levels tend to have different values of V_{th} [Lacaita et al., 2004; Krebs et al., 2009].

The physics behind threshold switching are not fully understood, although several models have been proposed to explain the possible reason for such a behavior. Owen et al. attributed the threshold switching to the thermal runaway caused by the exponentially increasing current through the phase-change material, owing to the temperature-dependent conductivity of these materials [Owen et al., 1979]. However, Adler et al. postulated that the fundamental electronic nature of these material is responsible for the switching phenomenon [Adler et al., 1980].

More recently, Pirovano et al. proposed a band gap model for understanding the switching phenomenon, linking this phenomenon to the competing role of impact ionization and recombination via valence alternation pairs [Pirovano et al., 2004]. In another electronic model, the switching is attributed to the energy gain of the electrons in the high electrical field leading to the voltage-current instability [Ielmini, 2008].

2.3.1.3 Thermoelectric effects

Until recently, Joule heating was believed to be solely responsible for bringing about the phase transition in phase-change materials. However, some of the recent studies suggested that apart from Joule heating, also the thermoelectric heating components contribute to the switching process. Dong-Seok et al. demonstrated highly efficient PCM devices using thermoelectric effects by proper selection of the phase-change material and electrodes so as to maximize the thermoelectric contribution [Dong-Seok et al., 2010]. In contrast, Lee et al. predicted a 16% decrease in the programming current and a 44% increase in the peak temperature attained within the device by including the thermoelectric effects in the modeling of conventional PCM devices [Lee et al., 2012].

Thermoelectric effects describe the fundamental interplay between the electrical and the thermal transport process in a physical system, apart from Joule heating coupling. The two primary thermoelectric effects are known as Seebeck and Peltier effects, which when combined with conservation law yield all the remaining heating components. I will give a brief overview of the various thermoelectric effects before focusing on the implementation of the model.

Seebeck effect When a material is subjected to a temperature gradient, charge carriers in the hotter region tends to diffusively flow into cold regions. This charge flow results in an induced voltage, and this is known as the Seebeck effect. Figure 2.4 schematically represents the Seebeck effect. In an open-circuit condition, charge carriers will accumulate in the cold end, resulting in an opposing electric field. The voltage drop generated by this electric field is known as the thermoelectric voltage and is proportional to the difference in temperature and a material-dependent constant known as the Seebeck coefficient, S. It is usually expressed in V/K.

$$V_{th} = -S\Delta T \tag{2.7}$$

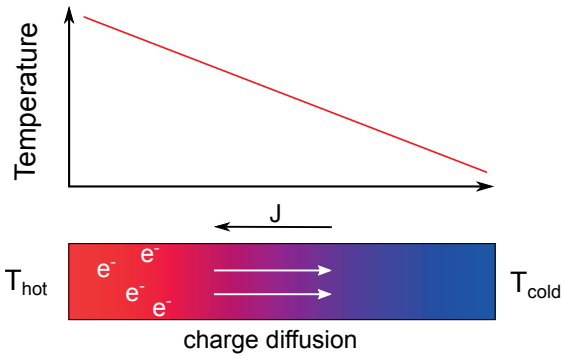


Figure 2.4: Schematic representation of the Seebeck effect. Diffusion of charge carriers based on the temperature gradient. The charge carriers flow from T_{hot} to T_{cold}.

The Seebeck coefficient is a property of the material and can either be positive or negative predominantly depending on the type of the dominant charge carriers. For an electron-dominated (n-type) material, the Seebeck coefficient is negative ($S < 0$), whereas it is positive for hole-dominated (p-type) material, ($S > 0$).

Peltier effect When an electric current flows through a junction consisting of two dissimilarly conducting media, heat is either dissipated or absorbed at the junction, depending on the direction of the current flow. This is known as the Peltier effect. Fig. 2.5 illustrates the scenario where a current flows through an interface with two different materials (n-type and p-type semiconductor). Based on the direction of the current flow, the electric charge carriers flow across the material and thereby carry some thermal energy with them away from the interface. The Peltier effect characterizes the amount of thermal energy transported by the electric-current charge carriers and is given by

$$Q_P = TSJ \tag{2.8}$$

where T is the temperature in K and J is the current density in A/m^2. The term $P = S \cdot T$ is known as the Peltier coefficient and is expressed in Volts.

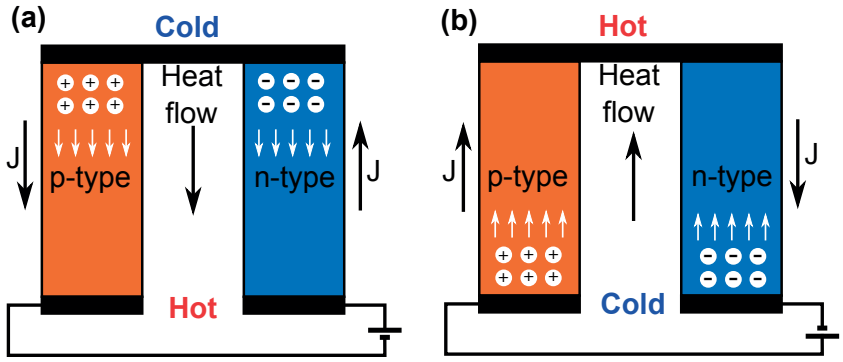


Figure 2.5: Schematic representation of the Peltier effect at the interface of two dissimilar materials. The electric charge carriers also carry some thermal energy and this results in either the evolution or absorption of heat at an interface based on the direction of the current flow.

Thomson effect The absorption or the evolution of heat along a conductor when electric current flows through it is known as Thomson effect. The heat is proportional to both the electric current and the temperature gradient. If J is the current density flowing through the conductor, the heat produced per unit volume, Q_T, is given by

$$Q_T = -\mu J \nabla T, \tag{2.9}$$

where ∇T is the temperature gradient along the conductor and μ is the proportionality constant and is known as Thomson coefficient. The Thomson coefficient is expressed in V and is

given by

$$\mu = T\frac{dS}{dT}; \tag{2.10}$$

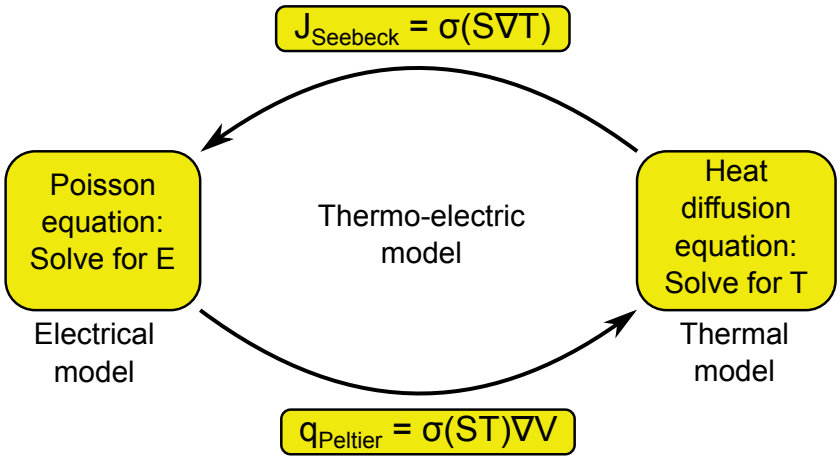


Figure 2.6: Coupled thermal and electrical model for programming a PCM device. Apart from the Joule-heating interaction between the thermal and electrical model, q_{Peltier} and J_{Seebeck} capture the thermoelectric interplay between them.

Therefore, any efforts to build a model to capture the behavior of these devices should include a thermal model for evaluating the transient temperature distribution within the device, an electrical model for predicting the potential distribution (electric field) and current density across the device, and a phase-change model for simulating the phase transition process in the phase-change material. The thermal and electrical models are coupled through the Joule heating. However, apart from the Joule heating, to include the thermoelectrics, also the additional Seebeck and Peltier effects should be included in the electrical and thermal model, respectively. The combined thermoelectric model then couples with the phase-change model through the material properties of the crystalline and amorphous states. Figure 2.6 represents the simplified block diagram of the complete thermoelectric model without the phase-change model.

All thermoelectric heating components can be captured using the Seebeck coefficient of the material. This is of great interest as in continued scaling of PCM devices, the thermoelectric effects play a significant role (in addition to the Joule heating) in bringing about the switching process, owing to the enhanced thermal confinement at smaller dimensions. Conventionally, only the Poisson equation is solved for calculating the current flow, and the heat-transfer equation is solved to calculate the heat flux through the model. According to the conservation

of energy,

$$\nabla \cdot J = 0 \tag{2.11}$$

$$\nabla \cdot q = J \cdot E, \tag{2.12}$$

where the current density, J, and the heat flux, q, can be given by

$$J = -(\sigma \cdot \nabla V) \implies J = \sigma E \tag{2.13}$$

$$q = -(\kappa \cdot \nabla T); \tag{2.14}$$

From (2.11)−(2.14), the traditional transient heat transfer equation is obtained as

$$\rho C_p \frac{\partial T}{\partial t} - \nabla \cdot (\kappa \nabla T) = \frac{J^2}{\sigma}. \tag{2.15}$$

However, to include the thermoelectric heating components, the Seebeck coefficient should be included. Therefore the current density and heat flux can be modified as

$$J = -\sigma(\nabla V + S\nabla T) \implies E = -\frac{J}{\sigma} + S\nabla T \tag{2.16}$$

$$q = -\kappa \nabla T + STJ; \tag{2.17}$$

Substituting Eq. (2.16) and (2.17) in Eq. (2.15), the heat transfer equation becomes

$$\rho C_p \frac{\partial T}{\partial t} - \nabla \cdot (\kappa \nabla T + STJ) = J \cdot \left(\frac{J}{\sigma} + S\nabla T\right); \tag{2.18}$$

By expanding and rearranging the above equation, the heat transfer equation becomes

$$\rho C_p \frac{\partial T}{\partial t} - \nabla(\kappa \nabla T) = \left(\frac{J^2}{\sigma}\right) - J\ T\ \frac{\partial S}{\partial T}\nabla T - JT\nabla S - ST\nabla J, \tag{2.19}$$

which captures all heating components, and the individual heating volumetric contribution

terms can be separated as follows:

$$\begin{aligned}
&\text{Joule heating:} \quad Q_J = \left(\frac{J^2}{\sigma}\right) \\
&\text{Peltier heating:} \quad Q_P = -JT\nabla S \\
&\text{Thomson heating:} \quad Q_T = -J\ T\ \frac{\partial S}{\partial T}\nabla T \\
&\text{Bridgman heating:} \quad Q_B = -ST\nabla J
\end{aligned} \tag{2.20}$$

(2.20) summarizes the individual volumetric heat-generation terms expressed in W/m^3. The sign represents whether a particular heat component is absorbed or evolved. Also, if the medium is homogeneous, there will be zero contribution from Peltier heating, Q_P, as the gradient of Seebeck will be zero (∇S =0).

2.3.1.4 Interface thermoelectric effects

From (2.20), it is evident that the Peltier heating component is dictated by the Seebeck gradient. Therefore, the Peltier contribution will be zero in the case of a homogeneous medium. However, at the interface between two dissimilar materials, there is a Seebeck gradient, owing to the discontinuity in the Seebeck coefficients of the materials, resulting in a Peltier heat term at the interface. At the interface between two materials, neither charge nor heat can continuously accumulate. If either of these quantities begins to accumulate at a surface, the system will eventually reach a steady-state point, where both the charge and the heat flux normal to the surface are equal on either side of the interface:

$$\hat{n}\cdot\vec{J}_L = \hat{n}\cdot\vec{J}_R \tag{2.21}$$

$$\hat{n}\cdot\vec{q}_L = \hat{n}\cdot\vec{q}_R, \tag{2.22}$$

where $\hat{n}$ is the unit normal to the interface and is defined to point from material 1 to material 2. Figure 2.7 illustrates the interface between two dissimilar materials. Re-writing these equations using Eqs. 2.16 and 2.17 and rearranging, we get

$$\hat{n}\cdot\left[\sigma_2\cdot\vec{E}_1 - \sigma_1\cdot\vec{E}_1\right] = \hat{n}\cdot\left[\sigma_2\cdot S_2\cdot\vec{\nabla}T_2 - \sigma_1\cdot S_1\cdot\vec{\nabla}T_1\right] \tag{2.23}$$

$$\hat{n}\cdot\left[\kappa_1\cdot\left(\vec{\nabla}T\right)_1 - \left[\kappa_2\cdot\left(\vec{\nabla}T\right)_2\right] = \hat{n}\cdot T\left[S_1\cdot\vec{J}_1 - S_2\cdot\vec{J}_2\right]; \tag{2.24}$$

(2.23) states that as surface charge builds up at the interface, an interfacial electric field forms that is proportional to the difference of the thermo-currents in the two materials. This

interfacial thermoelectric field is distinct from the fields generated by the materials themselves, and is simply a variant of the well-known surface charge effect from classical electrostatics.

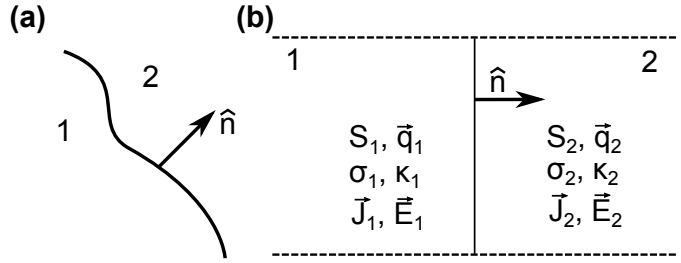


Figure 2.7: (a) Junction between two dissimilar thermoelectric materials with Seebeck coefficients S_1 and S_2 respectively. (b) Infinitesimal section of the interface. $\hat{n}$ is the vector normal to the interface.

Equation (2.24) states that there will be a heat source or sink at the interface when the Peltier heat flux normal to the surface is not balanced. This reflects the fact that electric currents in different materials carry different amounts of heat. Energy is conserved in the system by a commensurate amount of heat being absorbed from or evolved to the surrounding medium at the interface. This heat is known as the interfacial Peltier Heat, Q_{iP};

$$Q_{iP} = T\left[S_1 \cdot \vec{J}_1 - S_2 \cdot \vec{J}_2\right] \cdot \hat{n}; \tag{2.25}$$

Assuming an isotropic medium, the interface Peltier term can be simplified as

$$Q_{iP} = T\left[S_1 - S_2\right] \vec{J} \cdot \hat{n}; \tag{2.26}$$

The sign of Q_{iP} determines whether heat is absorbed or evolved at the junction. If $Q_{iP} > 0$, then there is an excess amount of Peltier heat flowing into the junction, and heat is evolved into the surroundings. In contrast, for $Q_{iP} < 0$, there is a excess of heat flowing away from the junction and heat is absorbed from the surroundings.

Given that the Peltier heating components are maximum at the interface when there is a discontinuity in the material properties (Seebeck coefficient), proper tuning of the Seebeck coefficient of the phase-change materials and electrodes can lead to efficient memory devices. Also, the distinct behavior of these devices for opposite bias-polarity conditions can be explained from the thermoelectrics described. Therefore, from a modeling perspective, the thermoelectrics phenomenon should be precisely understood and included in the PCM device modeling efforts. Recently, Faraclas et al. proposed a model to capture the individual heating thermoelectric components within a PCM device [Faraclas et al., 2014]. We developed a thermoelectric model in COMSOL to gain a proper understanding of our own devices and to accurately capture the device characteristics.

2.4 Simulation approach

2.4.1 Geometry

The mushroom-type devices, with doped-GST (d-GST) as phase-change material sandwiched between the top and bottom electrodes made of Titanium Nitride (TiN), are used in this modeling approach. The phase-change material is approximately 100 nm thick, and the bottom electrode has a diameter of 40 nm. The bottom electrode is formed by a lithographically independent, so-called "key-hole" process [Breitwisch et al., 2007], creating the keyhole with a sub-lithographic critical dimension (CD) that is essentially independent of the original feature size. The device is surrounded by an insulating layer of SiO_2. Figure 2.8 illustrates the device geometry and dimensions of the mushroom topology implemented, making use of the 2D axial symmetry geometry in COMSOL Multiphysics.

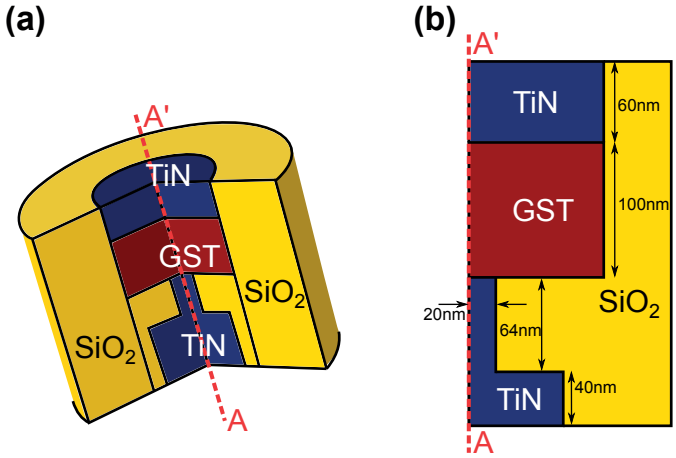


Figure 2.8: (a) PCM mushroom device topology implemented as a 2D axial symmetry geometry along AA' in COMSOL Multiphysics. (b) Cross-sectional view of the device with dimensions.

2.4.2 COMSOL implementation

To implement the thermoelectric model, modules available in the COMSOL Multiphysics® package were used. The individual domains are represented in a 2D plane and the *assembly* feature is used to build the geometry such that there exists an interface between adjacent domains. The required interface boundary conditions can then be implemented using *pair boundary* features from the COMSOL modules available.

2.4.2.1 Electrical model

The *electric currents (ec)* interface is used to compute the electric field, the current density and the voltage distributions in the conducting medium, where inductive effects are negligible. This particular physics interface solves the current conservation equation based on Ohm's law

using the scalar electric potential (V) as the dependent variable. The primary equation to be solved for the electric-current model is given by

$$E = -\nabla V, \tag{2.27}$$

$$\nabla \cdot \vec{J} = \nabla \cdot (\sigma \vec{E} + \vec{J}_{ext}) = 0; \tag{2.28}$$

Based on the applied potential, (2.27) is used to evaluate the applied electric field, which is then used to solve the current continuity equation to determine the current density according to (2.28). Apart from the current density contribution from the applied electric field, any additional external current contribution can be modeled using the $\vec{J}_{ext}$ term.

2.4.2.2 Thermal model

The *heat transfer in solids (ht)* interface is used to evaluate the heat transfer by conduction. The temperature (T) is the dependent variable, and the heat diffusion equation is solved to evaluate the temperature distribution. The thermal conductivity describes the relationship between the heat flux vector, $\vec{q}$, and the temperature gradient, ∇T, as in (2.29), which is usually referred to as Fourier's law of heat conduction:

$$q = -\kappa \nabla T + q_{ext}; \tag{2.29}$$

The volumetric heat flow term can then be calculated by

$$\rho C_p \frac{\partial T}{\partial t} - \nabla \cdot q = Q; \tag{2.30}$$

Any additional heat source can be modeled using the q_{ext} term in (2.29).

2.4.2.3 Thermoelectric components

To incorporate the thermoelectric heating effects with the electrical and the thermal model, the additional thermoelectric components, namely, J_{Seebeck} and q_{Peltier}, are added through the *external current density* feature, $\vec{J}_{ext}$, and the *additional heat flux* feature, $\vec{q}_{ext}$, to the electrical and thermal model, respectively. The interface thermoelectric component (interface Peltier effect, Q_{iP}) is modeled as *boundary heat source* feature between the GST and TiN domain interface.

2.4.3 Boundary conditions

With all the necessary equations governing the physical process on the implemented geometry defined, the practical boundary conditions should be defined at the domain boundary to

simulate the complete FEM model. The COMSOL software provides numerous boundary conditions specific for each physical interface. The boundary conditions needed for simulating the proposed model are described below.

2.4.3.1 Electrical boundary conditions

The phase-change material and electrode domains are included in the *electric currents (ec)* interface. It is therefore necessary to define the required electrical boundary condition at the corresponding domain boundary of GST and TiN. This also means that the interface between the conductors (GST, TiN) and SiO_2 should be set. The potential is applied at the top electrode, and the bottom electrode is assumed to be at electrical ground ($V = 0$). This can be implemented using the *electric potential* and *ground* features from the *ec* interface. The initial value across the GST and TiN domain is assumed to be at zero potential ($V = 0$).

All other domain boundaries of the conducting medium (GST, TiN) are assumed to be electrically insulating, such that there is no current flow across them. This can be modeled by using the *Electric insulation* feature. If $\vec{J}$ is the electric current density and $\hat{n}$ is the normal to the surface, the electrical boundary condition implemented can be expressed as

$$\hat{n} \cdot \vec{J} = 0; \tag{2.31}$$

2.4.3.2 Thermal boundary conditions

Unlike the electrical boundary condition, the thermal boundary conditions are ever more critical as they determines the thermal conduction pathways and therefore, the temperature distribution attained within the device for any given applied power. In the real-case scenario, the thermal boundary conditions of the operating device are greatly influenced by the ambient temperature and the thermal resistance of the system surroundings. Similar to the electric insulation boundary condition, the *thermal insulation* boundary condition can be implemented by

$$\hat{n} \cdot \vec{q} = 0; \tag{2.32}$$

However, it is not straight forward to capture the practical thermal boundary conditions in the finite-element model, as assuming a constant temperature (room temperature, say 300 K) at the domain boundary will over estimate the cooling effect of the system surroundings. Similarly, assuming a thermal insulation at the model boundary (2.32), will under estimate the cooling effects of the surroundings. Hence, there is a need for a dynamic thermal boundary condition that can be used to estimate the actual flow of heat flux at the boundary between the device under operation and its surroundings. Such a boundary condition can be tuned to exactly match the thermal environment of the actual experimental platform used for device measurements.

Assuming a heat flow across the domain boundary, the heat flux at the boundary can be defined by

$$q = h \cdot \Delta T \tag{2.33}$$

$$\Delta T = Q \cdot R_{th} = q \cdot A \cdot R_{th}$$

$$\rightarrow h = \frac{1}{A \cdot R_{th}} \tag{2.34}$$

$$R_{th} \approx \int_{r_s}^{\infty} \rho_{th} \frac{dr}{4\pi r^2} \approx \frac{1}{4\pi r_s k_{th}}$$

$$\rightarrow h = \frac{4\pi r_s k_{th}}{4\pi r_s^2} = \frac{k_{th}}{r_s} \tag{2.35}$$

ρ_{th} and κ_{th} are the thermal resistivity and the thermal conductivity of the surrounding insulator (system). r_s is the radius of the system. h is the input parameter for the thermal boundary condition, and it can be fine-tuned to match the thermal surroundings of the actual device. Figure 2.9 depicts the assumed heat flux boundary condition for the given geometry. Figure 2.10 illustrates the variations in the thermal profile according to the thermal boundary conditions. The initial value of the dependent variable across the geometry was assumed to be at the ambient operating temperature ($T = T_{\text{amb}}$).

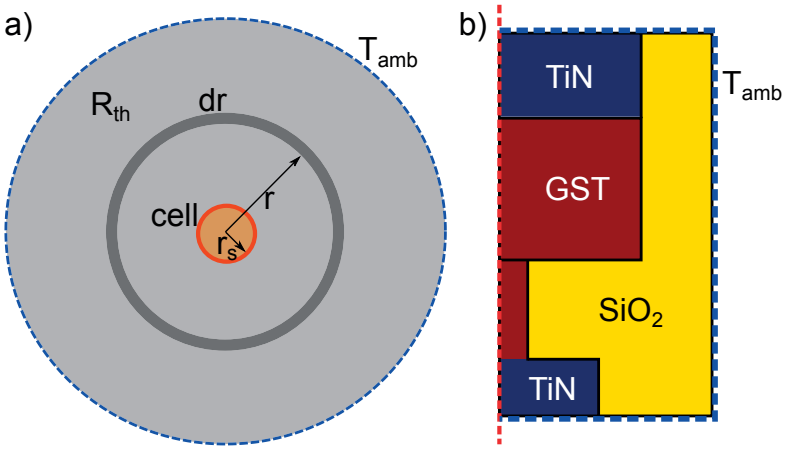


Figure 2.9: (a) Thermal system with T_{amb} as the ambient temperature of the surroundings. R_{th} is the thermal resistance of the surrounding medium. (b) Implementation of the heat flux thermal boundary condition (dotted blue line) in the geometry studied.

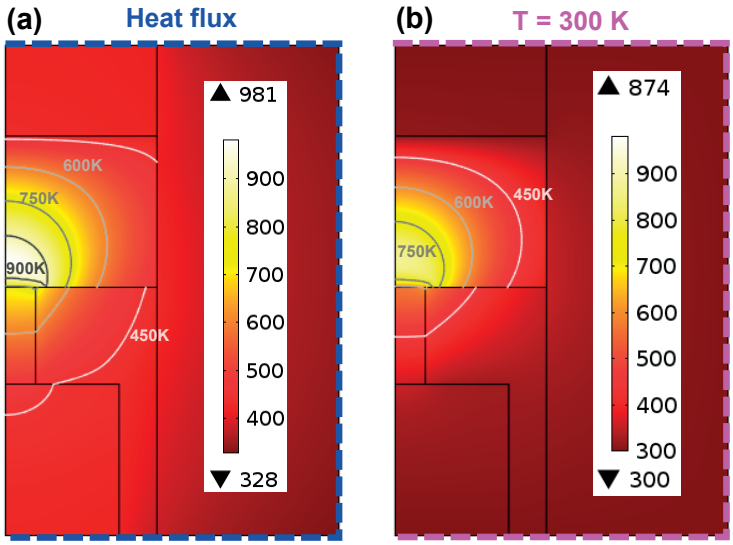


Figure 2.10: (a) Heat flux boundary condition with h calibrated to the system surroundings. (b) Constant temperature boundary condition (300 K). The thermal profile attained is different in the two cases.

2.4.3.3 Interface resistances

Apart from the thermal and electrical models and the boundary conditions described, the interface between the materials plays a vital role in heat diffusion. Experimental measurements suggested that large thermal boundary resistances exist at the interfaces between the phase-change material with the electrodes and the surrounding insulating materials [Reifenberg et al., 2010; Battaglia et al., 2010]. In the case of the mushroom-type devices considered here, a thermal boundary resistance (TBR) exists between the GST-TiN and GST-SiO_2 interface and a electrical contact resistance at the GST-TiN interface.

TBR at the interface determines the heat conduction pathways and therefore directly influences the thermal efficiency. Such interface resistances are known to have a significant impact on the programming characteristics of the device. Thermal conduction across the GST-TiN bottom electrode interface is especially critical in heat confinement in the active region. Typically, in the case of mushroom-type devices, the hotspot is located close to the bottom electrode. With its high thermal conductivity, the bottom electrode favors the easy heat flow owing to its thermally low-resistive path. However, the TBR at this particular interface helps in achieving better thermal confinement and thereby in pushing the hotspot closer to the interface. If the hotspot is located closer to the bottom electrode, less power is required to RESET the device, resulting in increased efficiency.

Recently, Reifenberg et al. studied the impact of including the thermal boundary resistance in Multiphysics modeling [Reifenberg et al., 2008]. They observed that with increasing values of TBR at the GST-TiN interface, there was a considerable reduction in the programming current.

Therefore, TBR should be precisely modeled and included in the geometries, where the TBR is comparable to bulk thermal resistances in series with the interfaces.

For the thermal model, the *pair thin thermally resistive layer* feature in the *heat transfer in solids (ht)* interface was used to model the thermal boundary resistance at the required interface. Similarly, an electrical contact resistance exists at the GST-TiN interface. In the electrical model, the *pair contact impedance* feature of the *electric currents (ec)* interface can be used to integrate the electrical contact resistance at the GST-TiN interface.

Interface Peltier effect: The interface Peltier component (2.26) at the GST-TiN interface can be implemented as an additional heat flux source at the boundary between the two domains. The *Pair boundary heat flux source* feature of the *heat transfer* interface was used to model the interface Peltier effects. The discontinuity in the Seebeck coefficient can be captured by using the *extrude* geometry feature, where the TiN boundary is projected onto the GST boundary and the difference in the Seebeck coefficient at these two boundaries can be evaluated.

2.4.4 Programming and Read model

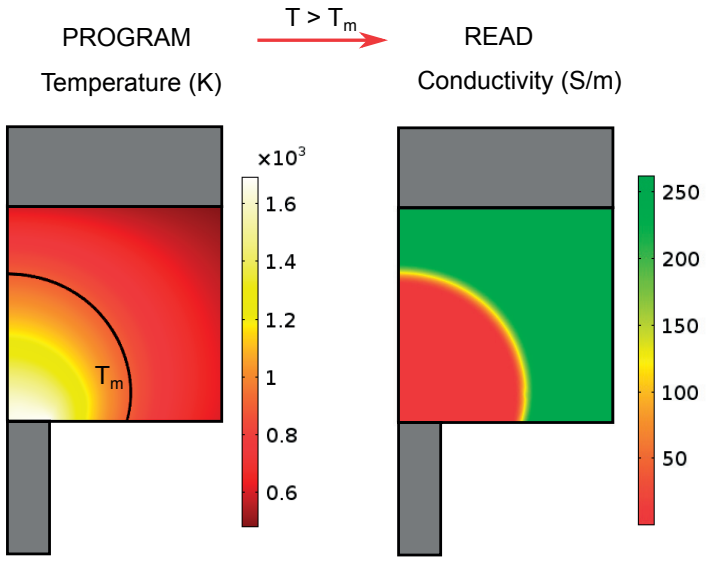


Figure 2.11: Simplified program/read model implemented in COMSOL to capture the phase-change process. Based on the thermal profile attained in the program sub-model, the conductivity is modified in the read sub-model. T_{m} is the melting temperature of GST (typically around 900 K).

To model the phase-change switching process, two separate sub-models (programming and read) were implemented. The programming model evaluates the temperature based on the applied electric pulse and the temperature-dependent material properties starting from the crystalline phase. Based on the temperature distribution of the programming model, the

necessary material properties are modified in the read model corresponding to the crystalline ($T_{\text{program}} < T_{\text{m}}$) and amorphous ($T_{\text{program}} > T_{\text{m}}$) states. Figure 2.11 illustrates the program and the read model. These two sub-models will be simulated simultaneously, and the thermal input from the programming model is continuously fed back to the read sub-model.

Instead of a complete crystallization model to capture the phase-change kinetics, a compact model to capture the transition from the crystalline to the amorphous state was implemented. As the main focus is on measuring the efficiency in terms of programming power and on investigating the thermoelectric physics rather than the phase-change kinetics, this simplified approach for the phase transition was devised to reduce the complexity of the model. For the kind of analysis performed in this thesis, I believe that the proposed way of modeling the phase-change process is accurate enough to study the device behavior, without an actual need for the crystallization model.

2.5 Material properties

With the complete model implemented in COMSOL, the final critical component required to simulate the model is the material properties. Given the significant research focus on chalcogenide materials in the past decade, extensive research has been performed, both theoretically and experimentally, in better understanding the nature of its thermal and electrical properties. We can rely on the literature for almost all of the material properties needed.

2.5.1 Electrical Conductivity

As described in Section. 2.3.1.2, one of the fundamental property of these materials, which made data storage feasible, is the electronic threshold switching phenomenon. In the past decade, several models for the electrical conductivity of GST have been proposed to describe the so-called snap-back mechanism (thrshold switching). Early models tried to explain the field-dependent property of the conductivity by calculating the number of free charge carriers via the rate equations and by considering the number of traps and generation-recombination mechanisms [Adler et al., 1980].

2.5.1.1 Poole–Frenkel condcution model

The most widely used model to describe the exponential current-voltage characteristics of amorphous semiconductors is the Poole–Frenkel (PF) conduction mechanism [Frenkel, 1938; Hill, 1971]. The Poole−Frenkel model suggests that the increase of electrical conductivity in the intense field is due to an increase in the number of free electrons and not their mobility. Poole−Frenkel-based conduction mechanism have already been verified in amorphous phase-change materials [Owen and Robertson, 1973]. Ielmini and Zhang proposed an extended version of the Poole−Frenkel effect to explain the sub-threshold conduction behavior typically

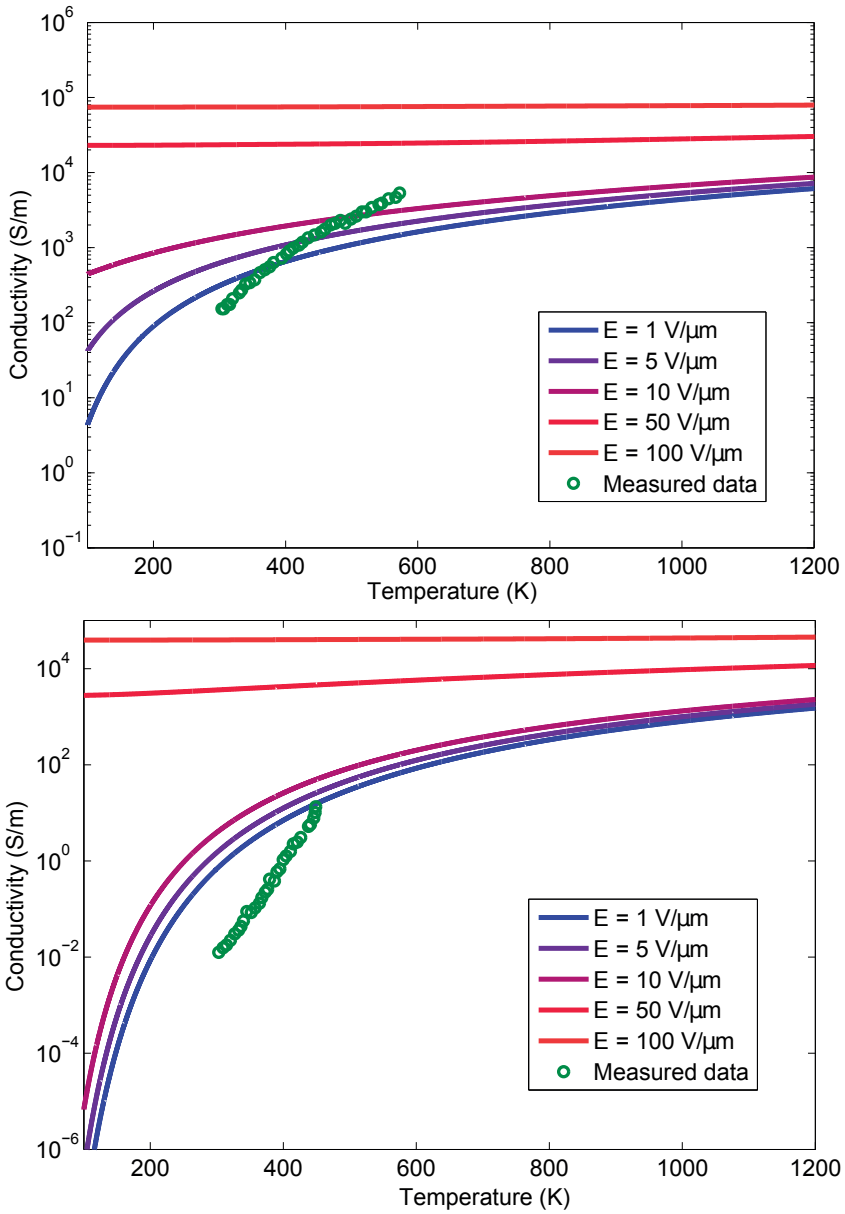


Figure 2.12: Field- and temperature-dependent electrical conductivity for the crystalline and the amorphous phase of the GST. The data points are the experimental conductivity measurements of GST thin-films from [Raoux et al., 2014].

seen in these materials [Ielmini and Zhang, 2007]. In the low-field regime, charge carrier emission from traps can be activated with an activation energy, E_a ($E_a = E_c – E_t$), corresponding to the energy level of the trap state. The number of charge carriers will then be given by

$$n_c \sim \exp\left(-\frac{E_a}{k_B T}\right); \tag{2.36}$$

Poole and Frenkel were the first to modify (2.36) by including an additional term to account for the barrier lowering of the activation energy by the applied electric field. According to the PF model, the field-dependent electrical conductivity for a high concentration of trap states can be given by

$$\sigma_{PF} \sim \exp\left(-\frac{E_a + \beta' E}{k_B T}\right), \tag{2.37}$$

where $\beta' = \frac{\Delta z}{2}$. Δz is the inter-trap distance in nm, and E is the applied electric field (V/m).

Field dependence: While implementing the conductivity function, care was taken to implement the temperature dependence with Fermi–Dirac statistics rather than with Boltzman statistics to be able to also describe the on state, in which the quasi Fermi levels are close to or even within the bands. According to the PF model, the electrical conductivity of the GST as the function of both the electric-field and temperature is given by

$$\sigma = \sigma_0 \cdot \mathbf{F}(-E_{PF}/k_B \cdot T), \tag{2.38}$$

where $\mathbf{F}(\ )$ denotes the Fermi integral. σ_0 can be approximated as

$$\sigma_0 = q \cdot \mu \cdot 4\pi \cdot N_c^{3/2} \cdot T^{3/2}, \tag{2.39}$$

where N_c is the effective density of states, q is the electronic charge (C), T the temperature (K) and μ the mobility (m^2/V·s). The effective density of states is given by

$$N_c = \left(\frac{m \cdot m_e \cdot k_B}{2 \cdot \pi^2 \cdot h_{barc}^2}\right)^{3/2}. \tag{2.40}$$

The field-induced barrier-lowering term is given by

$$E_{PF} = E_a - dz \cdot E/2, \tag{2.41}$$

where E_a is the activation energy of GST in eV and E is the applied electric-field in V/m. dz is the inter-trap distance in nm. The same conductivity model was used to represent both the crystalline and the amorphous phase of the GST using different values of the activation energy and the inter-trap distance.

Once in the on-state, the conductivity of both the amorphous and the crystalline state is similar, as can be seen from Fig. 2.12 in the high-field regime. Also at very high fields, there is no significant temperature dependence of the conductivity. In the on-state, the high current flowing through the material can quickly heat up the phase-change material, as the current is only limited by the series resistance of the device, which is typically the low-resistance bottom electrode. A further increase in the applied voltage will be sufficient for the phase-change material to reach temperatures on the order of its melting temperature.

2.5.2 Thermal conductivity

Recently, Risk et al. measured the thermal conductivities of the amorphous and crystalline phases of GST and other phase-change material using the "3ω voltage" method [Risk et al., 2009]. They observed that the phase-change material in its high electrical conductivity state (crystalline) exhibits a thermal conductivity close to that proposed by the Wiedemann–Franz value, implying that electrons dominate thermal transport at higher conductive states. In contrast, the thermal conductivity of the states with lower electrical conductivity (amorphous) substantially exceeds the value specified by Wiedemann–Franz, implying that phonons dominate thermal transport in those cases.

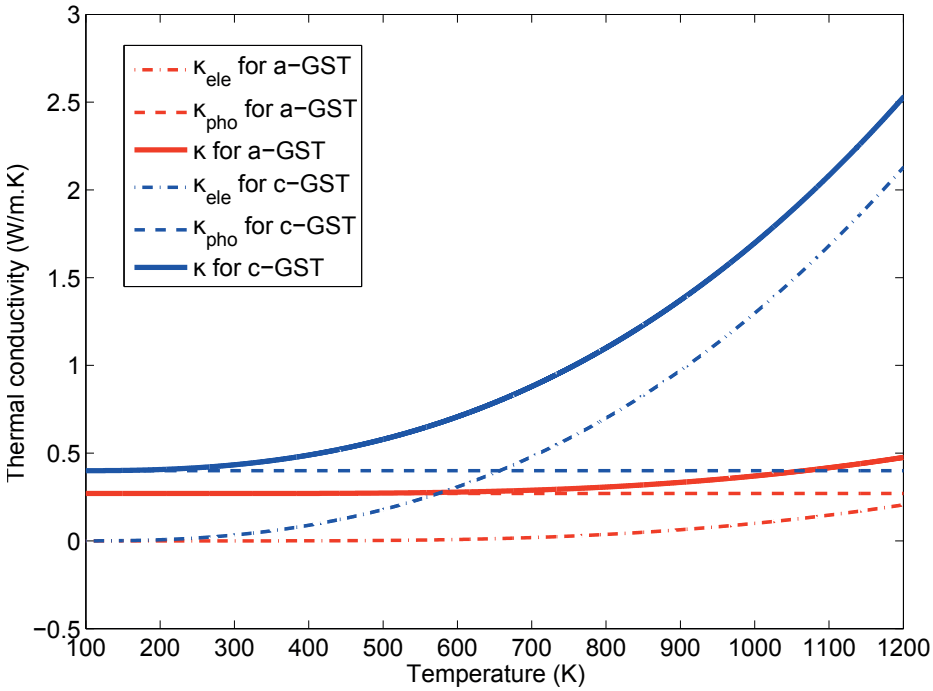


Figure 2.13: Crystalline and amorphous thermal conductivity of GST as a function of temperature (K). κ_{ele} is the electronic contribution and it follows the electrical conductivity using the Wiedemann–Franz law. κ_{pho} corresponds to the phononic contribution. Its values are taken from the literature [Lee et al., 2013].

These measurements confirm that the thermal conductivity of GST tends to have both an electronic and a phononic contribution. The electronic component correlates with the electrical conductivity in the form of the Wiedemann–Franz law. The relationship between the electrical and the thermal conductivity means that the thermal conductivity can easily be inferred from the easily measurable electrical conductivity rather than from the hard-to measure thermal conductivity. Figure 2.13 summarizes the electronic and the phononic component as a function of the temperature (K). The electonic contribution and the total thermal conductivity can

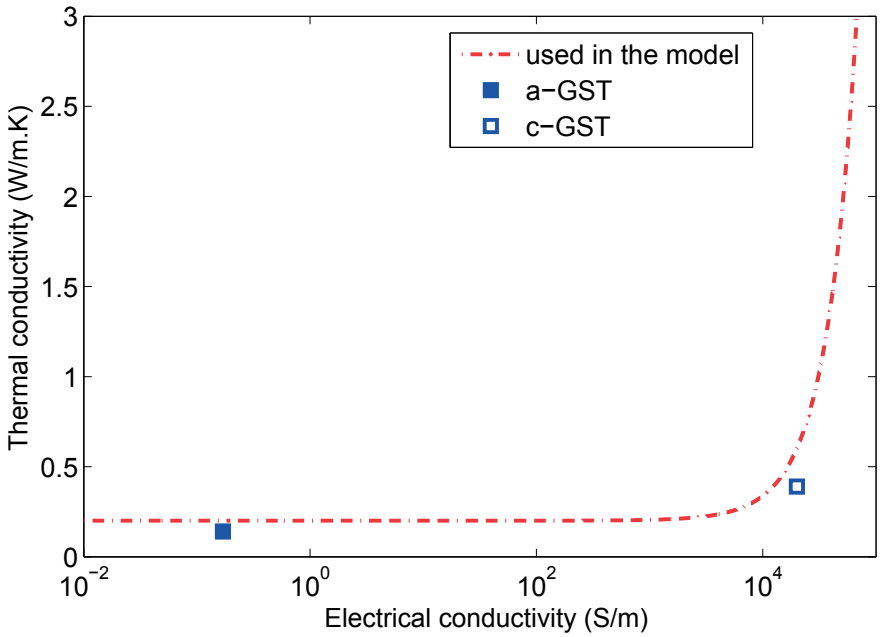


Figure 2.14: Thermal conductivity as the function of the electrical conductivity. At higher electric fields, the conductivity increases drastically and so does the thermal conductivity according to the Wiedemann–Franz law. The data points represents the thermal conductivity measurements of GST at room temperature [Risk et al., 2009].

be given by

$$\kappa_{ele}(E,T) = \sigma(E,T) \cdot L \cdot T, \tag{2.42}$$

$$\kappa = \kappa_{ele} + \kappa_{pho}, \tag{2.43}$$

where κ is the total thermal conductivity (W/m-K) of GST, σ is the electrical conductivity (S/m) of GST. σ_{ele} represents the electronic component and is the contribution from the standard Wiedemann–Franz law. κ_{pho} is the phonon-mediated contribution in terms of the thermal conductivity and is different for the crystalline and amorphous states. The corresponding values used in the simulation are 0.4 (W/m-K) and 0.27 (W/m-K) [Lee et al., 2013]. Figure 2.14 plots the total thermal conductivity as a function of the electrical conductivity (S/m).

2.5.3 Seebeck coefficient

The Seebeck coefficient of any given material measures the quantity of thermoelectric voltage induced while a temperature gradient exists across the material. The Seebeck coefficient of the material can be positive or negative depending on the type of semiconductor, p-type or n-type.

In the case of GST, the charge carriers flow in the direction of the heat flow, and hence GST has a positive Seebeck coefficient. In contrast, the TiN electrode exhibits negative Seebeck coefficient.

Lee et al. measured the phase- and temperature-dependent Seebeck coefficient of 25 nm and 125 nm-thick GST films for a range of temperatures, from room temperature to 300°C, using a silicon-on-insulator experimental structure in which a buried oxide layer induces lateral temperature fields. The temperature dependence of the Seebeck coefficient was studied and explained by using various models [Lee et al., 2012].

However, experimental measurements are available only over a short temperature range, and thus the data must be extrapolated for use in the model for the simulation. Given the model requirement to have Seebeck values for a wide range of temperatures, an alternative approach is to use the theoretical expression developed by [Mott and Davis, 1967] for the Seebeck calculation. The Seebeck expression for homogeneous, non-degenerative semiconductors is given by

$$S = k_B \cdot \left(\frac{E_{PF}}{k_B \cdot T} + \left(1 + exp\left(\frac{E_{PF}}{k_B \cdot T}\right)\right) \cdot ln\left(1 + exp\left(\frac{-E_{PF}}{k_B \cdot T}\right)\right)\right), \tag{2.44}$$

where E_{PF} is given by

$$E_{PF} = E_a - dz \cdot \frac{E}{2}, \tag{2.45}$$

where E_a is the activation energy in eV, E is the electric field in S/m and dz is the inter-trap distance in nm. The activation energy reduced by field-induced barrier lowering can be used to calculate the Seebeck coefficient. The corresponding values of the activation energy and the inter-trap distance are used in the conductivity model to calculate the Seebeck for the crystalline and the amorphous state, respectively. Figure 2.15 illustrates the field- and temperature-dependent Seebeck coefficient values for the crystalline and the amorphous state used in the simulations. The figure also compares the Seebeck values used in the model with experimental measurements from [Lee et al., 2012]. It is evident that (2.44) closely resembles the measured data and can therefore be used to evaluate the Seebeck coefficient at higher temperatures.

2.5.4 Interface resistance

2.5.4.1 Thermal boundary resistance

GST-SiO_2 thermal boundary resistance: The TBR at the interface between GST and SiO_2, typically used as insulating layer between the devices, significantly influences the heat diffusion from the active region and therefore impacts the temperature distribution attained within the device. Battaglia et al. measured the thermal boundary resistance at the GST

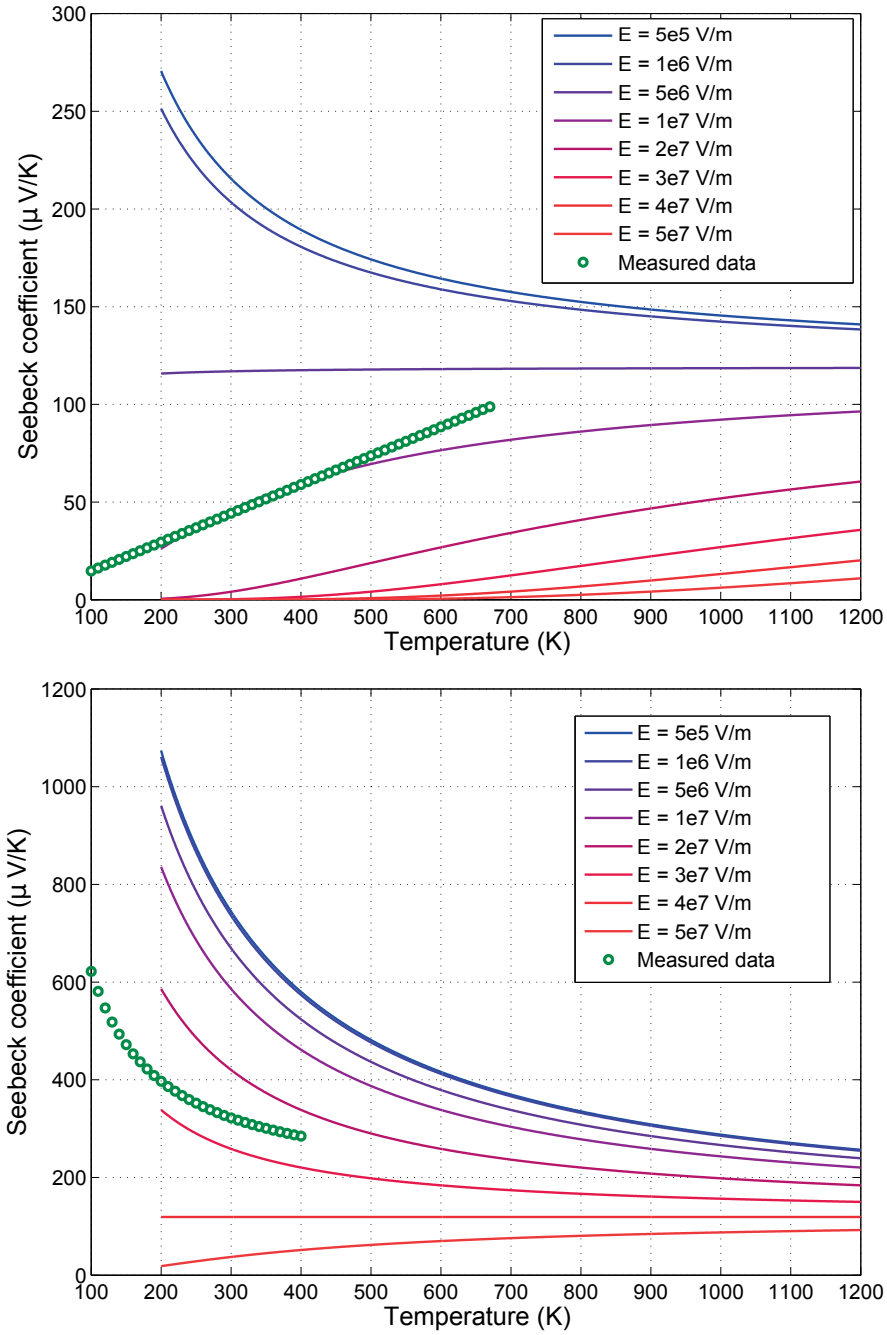


Figure 2.15: Field and temperature-dependent Seebeck coefficient of the crystalline and the amorphous phase of the GST. The same activation energy numbers were used for the crystalline and amorphous phases. Measurement data from [Lee et al., 2012] is shown for comparison.

and amorphous SiO_2 interface for wide range of temperatures from room temperature to 400°C. The measurements suggest that the thermal boundary resistance at the GST-SiO_2 inter-

face depends on the thickness of GST and it becomes even more dominant as the thickness approaches the nanometer regime [Battaglia et al., 2010].

GST-TiN thermal boundary resistance: The TBR at the GST and TiN interface is critical in determining the location of the hotspot. Reifenberg et al. used pico-second time domain thermal reflectance (TDTR) based methods to separate the TBR from the intrinsic thermal resistance and precisely measure the TBR at the GST-TiN interface for temperatures ranging from 30°C to 325°C [Reifenberg et al., 2010]. They observed that with increasing temperature, the GST-TiN interface varies from ~ 26 m^2·K/GW at room temperature to 18 m^2·K/GW at 300°C. These temperature-dependent TBR values are used at the GST-TiN interface both at the top and the bottom electrode. The values are extrapolated for the high temperatures required during the simulations.

2.5.4.2 Electrical contact resistance

Kencke et al. measured the electrical contact resistance at the interface between the bottom electrode and amorphous GST for a PCM device [Kencke et al., 2007]. The electrical contact resistance at the interface is estimated to be 2.6 x $10^{-11}\Omega$·m. This was included in the GST-TiN interface at the top and the bottom electrode.

2.5.5 Properties of Titanium nitride (TiN)

The electrical and thermal material properties of bulk TiN (top and bottom electrodes) are used for the simulation. Temperature-dependent thermal conductivity values from the literature are used [Shackelford and Alexander, Dec. 2000]. The Seebeck coefficient is estimated using the Mott equation for metals [Mott and Davis, 1967]. It is given by

$$S_{TiN} = -\frac{k_B^2 \pi^2}{3} T \left\{ \frac{d(ln(\sigma(E)))}{dE} \right\} \Bigg|_{E=E_F} ; \tag{2.46}$$

Figure 2.16 shows the temperature dependence of the thermal conductivity and the Seebeck coefficient of TiN. The other material properties required for the simulation include the specific heat capacity, the density and the relative permittivity. The values for them are all taken from the literature and summarized in Table 2.1.

2.5.6 Properties of Silicon-di-oxide (SiO_2)

In typical PCM arrays, SiO_2 is used as insulating spacer between adjacent memory devices. In our geometry, the PCM device is assumed to be completely enclosed by an SiO_2 insulating layer. The insulating medium is only included in the thermal model and hence the electrical conductivity and the Seebeck coefficient are undefined. For the thermal properties, Yamane

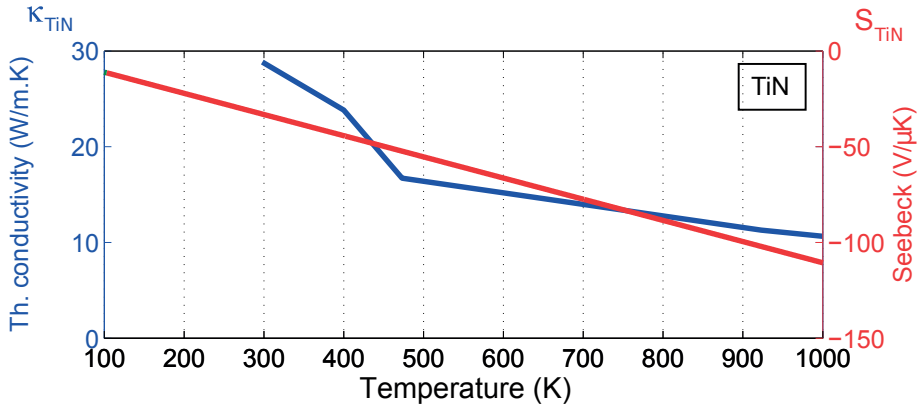


Figure 2.16: Temperature-dependence of the thermal conductivity and the Seebeck coefficient of TiN. The values are taken from [Mott and Davis, 1967] and [Shackelford and Alexander, Dec. 2000].

et al. have measured the thermal conductivity of SiO_2 thin films deposited through various fabrication procedures and the corresponding values were used for the simulation [Yamane et al., 2002]. Table. 2.1 summarizes the list of material parameters required by the model and the corresponding values used in the simulations.

Table 2.1: Material properties required for the simulation of the model. The data is obtained from [Battaglia et al., 2014], [Njoroge et al., 2002], [Shackelford and Alexander, Dec. 2000], [Shportko et al., 2008].

Material property	GST	TiN	SiO_2
Thermal conductivity (W/m-K)	$\kappa(E,T)$	$\kappa_{TiN}(T)$	1.4
Electrical conductivity (S/m)	$\sigma(E,T)$	2×10^5	10^{-14}
Seebeck coefficient (V/K)	$S_{GST}(E,T)$	$S_{TiN}(T)$	0
Density (Kg /m^3)	6220	5200	2600
Specific heat capacitance (J/Kg-K)	212	620	1190
Relative permittivity	33	1	3.9

2.6 Summary

The theory of thermoelectric physics was presented followed by the modifications of the Poisson equation and the heat diffusion equation to capture the thermoelectric physics. The typical mushroom-type device geometry was implemented in COMSOL Multiphysics®. Detailed implementation procedures for each of the required physical phenomenon were described. Finally, the practical field- and temperature-dependent material properties required

for the simulation were also presented.

Although the focus was mainly on mushroom-type devices and GST material, the same procedure can be extended to any device geometry and/or any kind of phase-change material. The validation of the proposed model by a comparison of the simulation results with experimental measurements is described in the next chapter.

Contributions:

- **Devised, developed and implemented the complete thermoelectric finite-element-based model in the COMSOL Multiphysics® software.**
- **Performed extensive investigation and analysis to define the boundary conditions of the model and to implement the interface thermoelectric effects.**
- **Field- and temperature-dependence of the material properties were thoroughly studied, and a compact model was used to define the material properties so as to maintain the simplicity of the model.**

3 Thermoelectric model validation and novel device design

In this chapter, I validate the thermoelectric model developed and implemented in the previous chapter. The main motivation behind the modeling approach is to precisely capture the switching characteristics of the device in terms of the programming power. The proposed model was simulated, and the results will now be compared with experimental measurements over a wide range of ambient operating temperatures. A brief description of the experimental platform to characterize the device is presented. Then the model is validated by the comparison of some of the interesting device characteristics, which are experimentally measured, with the simulation results. From the modeling perspective, I explain the possible reason for the distinct behavior seen in these devices for opposite bias-polarity conditions.

Given the continued scaling trends observed in PCM technology, I also briefly discuss the prominent limiting factor of device scalability, namely, the thermal disturb. The problem of thermal disturb is described and I propose a compact model for easier evaluation of the temperature distribution in highly-dense PCM arrays, which is otherwise an exhaustive and time-consuming process with FEM-based modeling. Finally, based on the numerous insights obtained from the thermoelectric model described, I propose two novel device-design ideas, that use thermoelectric effects, to improve the device efficiency in one case and to increase the memory capability in the other.

This chapter is organized as follows: **Section. 3.1** details the experimental platform used to characterize the PCM devices. **Section. 3.2** describes the key device characteristics which can be efficiently captured by the model, thereby providing valuable insight into device operation. **Section. 3.3** presents the understanding of the contrast device behavior for opposite bias-polarity. **Section. 3.4** explains the impact of critical material properties on the device operation. In **Section. 3.5**, I discuss the phenomenon of thermal disturbance and a compact model to investigate the same in highly-dense PCM arrays. Finally, **Section. 3.6** proposes a couple of novel device-design ideas which can exploit the thermoelectric effect to improve the device efficiency and memory capacity.

3.1 Characterization platform

Experimental measurements were performed on nano-scale PCM devices based on the mushroom-type device topology. The devices were fabricated in the 90 nm technology node, with the bottom electrode (heater) created via a sub-lithographic key-hole process [Breitwisch et al., 2007]. The phase-change material is doped GST (d-GST). There is no access device in series with the device, but, a resistor of approximately 5 kΩ is introduced on-chip in series with the PCM device to limit the current.

The experiments were performed on a JANIS ST-500-2-UHT nitrogen-cooled cryogenic probe station that can operate from 77 K to 400 K and offers a temperature stability of < 50 mK. Two heaters with 50 W and 25 W power and placed at different positions in the chamber allow temperature control. The temperature is measured at four points inside the chamber by calibrated Lakeshore Si DT-670B-CU-HT diodes with an accuracy of < 0.5 K. The temperature was regulated using a Lakeshore 336 Automatic Temperature Controller. To achieve higher temperatures, the sample was mounted on an invar block with two embedded tungsten heaters. The temperature was measured using a thermocouple inserted into the invar block and controlled via a Eurotherm temperature controller. The temperature on top of the PCM chip was calibrated using an Omega silicon diode sensor.

To thermally isolate the sample, a radiation shield is fixed above the sample mount and thermally connected to the nitrogen out-flux. To avoid heat exchange via convection and water condensation at low temperatures, the experiments were done under vacuum (average pressure of 10^{-5} mBar).

The chamber has six ports. Only one of them was used to connect a high-frequency Cascade Microtech Dual-Z probe. The high-frequency probe is connected with cooling braids to the Cu chuck to guarantee that the sample and the probe are at the same temperature. The sample is fixed on the Cu sample mount inside the vacuum chamber by a metal clip.

A Keithley 2400 SMU was used for DC voltage or current outputs and the measurement of the corresponding current or voltage at the sample. An Agilent 81150A Pulse Function Arbitrary Generator was used for AC voltage outputs (cell switching) and a Tektronix TDS3054B oscilloscope for AC voltage and current measurements. Mechanical relays were used to switch between AC and DC measurements. The device resistance was measured at a constant read voltage of 200 mV, and the current-voltage ($I-V$) characteristics were measured using the current output mode of the SMU from 1 nA to 1 μA.

3.2 Device characteristics

3.2.1 Model simulation

In contrast, the implemented full-fledged thermoelectric model is simulated using the field- and temperature-dependent material properties and boundary conditions described in the previous chapter. The material properties are implemented as MATLAB functions with the electric field and temperature as input parameters. Using the MATLAB *Livelink* interface feature in COMSOL, the MATLAB functions are directly used to specify the material property of the respective device domains. A time-dependent iterative solver was used to solve for the dependent variables (voltage, V and temperature, T) using the transient simulations. Voltage pulses with similar amplitude, pulse-width, leading and trailing edges corresponding to the experiments are applied to the programming model, and a constant read voltage of 200 mV was applied to measure the resistance.

Once the model is simulated, using the various post-processing features offered by COMSOL Multiphysics®, we were able to visualize the voltage and temperature distribution within the device and capture the resistance vs. programming power plots, we could even visualize the current density and individual volumetric heating contributions from each thermoelectric component, apart from the Joule heating.

3.2.2 Temperature distribution

Given that the phase-change switching phenomenon is an electro-thermal process, one of the basic efficiency measure is the temperature distribution attained within the device for the applied electrical power. Using the *2D Plot Group* and *1D Plot Group* features from the *Results* interface, we can capture the spatial and the temporal distribution of any physical quantity (temperature, voltage, current density, heat flux, electric field, etc.) for any value of power within the range of applied input power across any plane or line within the device.

From the simulations, it can be conclusively shown that the “hotspot” is located close to the bottom electrode (around 10 nm - 15 nm from the GST-bottom electrode interface). This can be attributed to the extreme asymmetry in the device geometry between the top and bottom electrodes. The substantially narrower bottom electrode (heater) results in increased current density and therefore the Joule heating power (I^2R) is maximum in the phase-change material close to the bottom electrode interface. This region, where the hotspot is located, is usually referred to as the active region.

Fig. 3.1 represents the thermal profile within the device for different ambient temperatures, but for the same applied power of 600 μW. Figure 3.2 captures the temperature profile along the symmetry axis of the device. Owing to the difference in the ambient temperature and hence the different thermal boundary conditions, a lower power is required to reach the melting temperature at the rim of the bottom electrode to create a plug, leading to a reduced

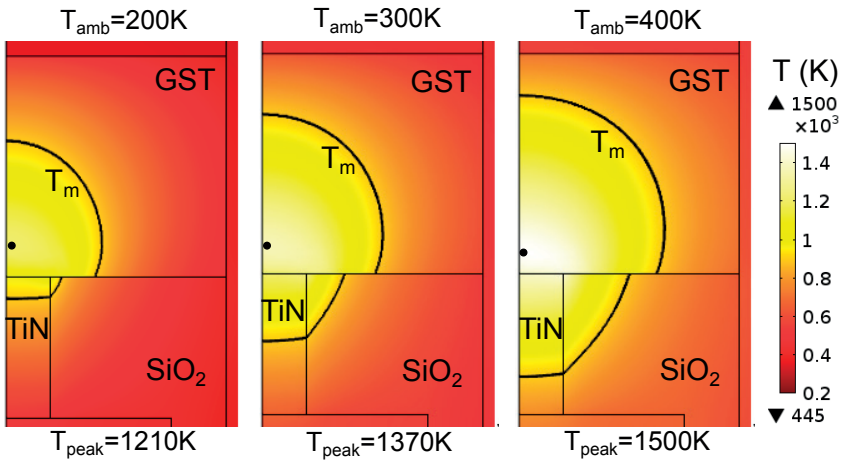


Figure 3.1: 2D thermal profile of the device for different ambient temperatures (200 K, 300 K, 400 K) for the same applied input power of 600 μW. T_{m} is the melting temperature of GST. The black dot denotes the location of the peak temperature (T_{peak}) attained.

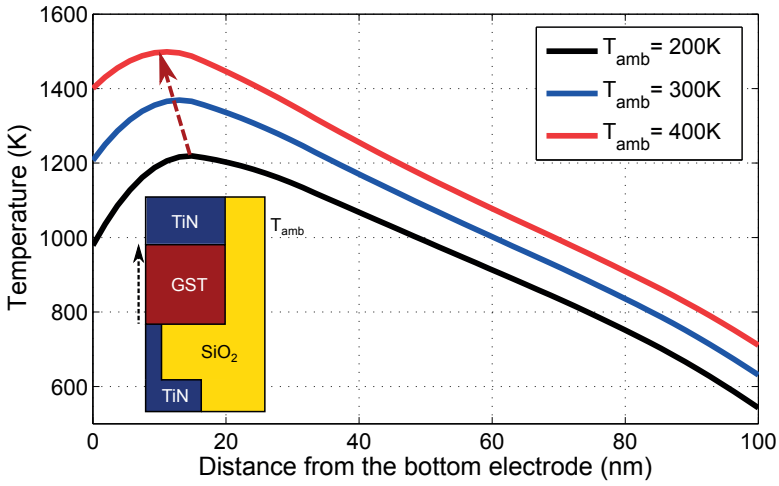


Figure 3.2: 1D temperature plot along the symmetry axis of the device (as shown in the inset) for different ambient temperatures (200 K, 300 K, 400 K) for the same applied input power of 600 μW. The dashed arrow traces the hotspot location for different ambient temperatures.

RESET current at higher ambient temperatures. However, comparing the differences in peak temperature ($\approx 290\ K$) attained in the extreme cases of 200 K and 400 K, it is evident that apart from the influence of ambient operating temperatures, also the temperature-dependent conductivity could have played a role in increasing the Joule heating component at higher ambient temperatures.

3.2.2.1 TBR influence on temperature distribution

The discontinuity in the temperature profile seen at the interface between GST-SiO_2 and GST-TiN is due to the interface thermal boundary resistance. TBR results from the difference in electronic and vibrational properties of the two dissimilar materials at their interface. When phonons or electrons attempt to travel across such interface, because of scattering only a certain fraction will succeed in transporting energy from one material to the other.

3.2.3 Thermoelectric heating components

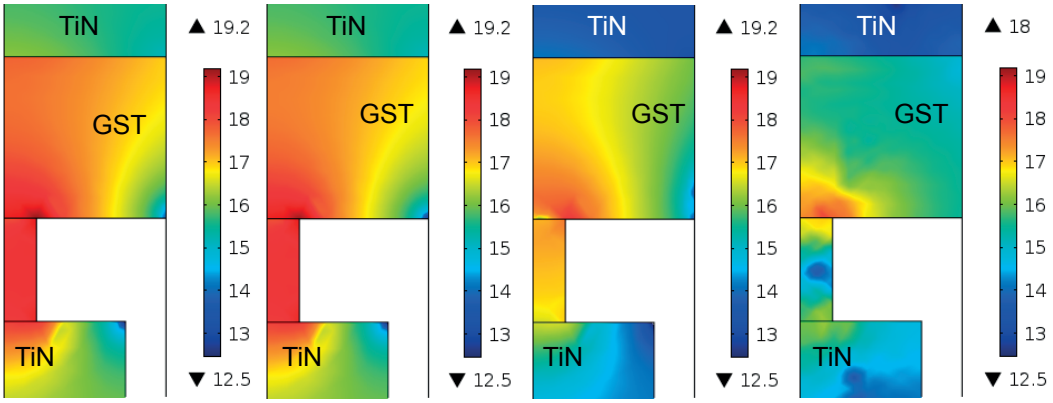


Figure 3.3: Individual volumetric heating contributions (W/m^3) from the thermoelectric heating at room temperature for an applied power of 1.1 mW expressed in log scale across the ABCD plane in Fig. 2.8. (a) Total heat, (b) Joule heating, (c) Thomson heating, and (d) Bridgman heating. The Peltier heating contribution is non-existent owing to the homogeneous GST medium. However, there will be a significant heating contribution from the interface Peltier effect at the GST-TiN interface.

The individual volumetric thermoelectric heating components as described in Section 2.3 are shown in Fig. 3.3. The Joule heating is predominant and is concentrated within the phase-change material in the region closer to the GST and the bottom electrode interface, owing to the high current density in the vicinity of the bottom electrode and the conductivity of GST. This region defines the active region, where the peak temperatures are attained.

Apart from Joule heating, the Thomson heating component is dominant and is concentrated in the active region closer to the GST-TiN interface. However, remembering from Section 2.3, Thomson heating is directly proportional to the temperature gradient and the gradient of the Seebeck coefficient with respect to temperature. Given that the peak temperature is attained in the active region, the temperature gradient (∇T) and the Seebeck gradient with respect to the temperature ($\partial S/\partial T$) are both maximum in the active region, and henceforth resulting in the Thomson heating term. In the material properties of GST, a homogeneous medium was assumed and hence there will be no contribution from the Peltier heat as the Seebeck gradient will be zero ($\nabla S = 0$). Lastly, the Bridgman heat term was proportional to the gradient in the

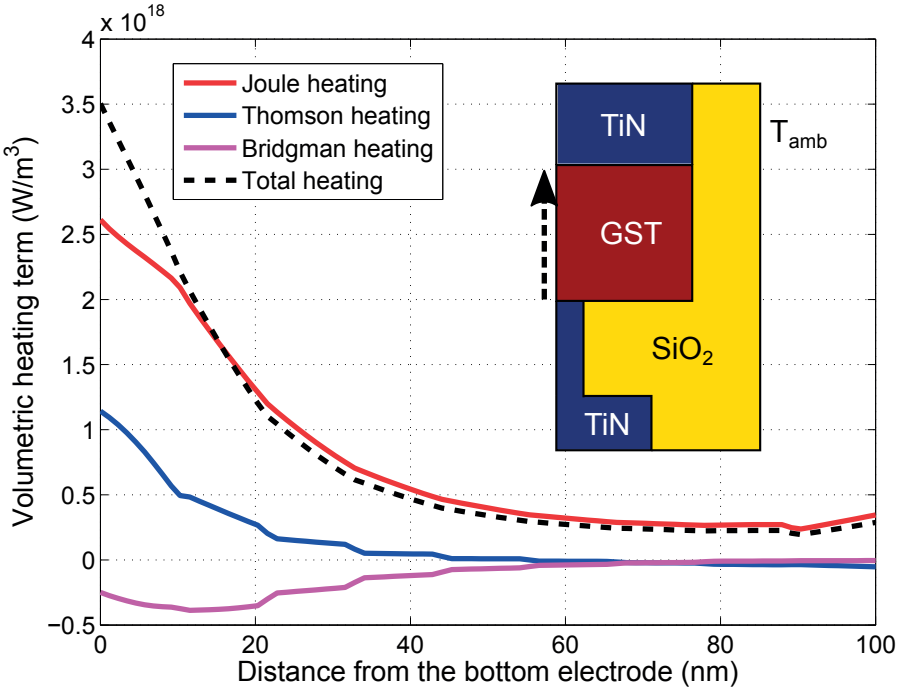


Figure 3.4: Individual volumetric heating contributions (W/m^3) from the thermoelectric heating at room temperature for positive bias with an applied power of 600 μW expressed in log scale along the symmetry axis of the device.

current density.

Fig. 3.4 depicts the individual thermoelectric heating terms along the symmetry axis of the device. It is evident that the thermoelectric heating is more concentrated in the active region. While Thomson heating adds up with Joule heating, the Bridgman component has a cooling effect. However, the overall impact of thermoelectric effect is positive as the total heating component is higher than the Joule heating term. Although, Peltier heating is non-existent in the active region, the interface Peltier effect implemented at the GST-TiN interface provides an additional Peltier heating component. Recalling from (2.24), apart from the prominent discontinuity in the Seebeck coefficient at the interface, the high temperature and current density attained in the vicinity of the interface contribute to the interface Peltier heating component.

3.2.4 Programming *I–V* characteristics

Fig. 3.5 depicts the $I-V$ characteristics while programming the device, starting from the crystalline state and for different ambient operating temperatures. The conductivity model used for the GST limits the current through the device in the sub-threshold regime. At very high fields, owing to the exponential current dependence of the PF model (threshold switch-

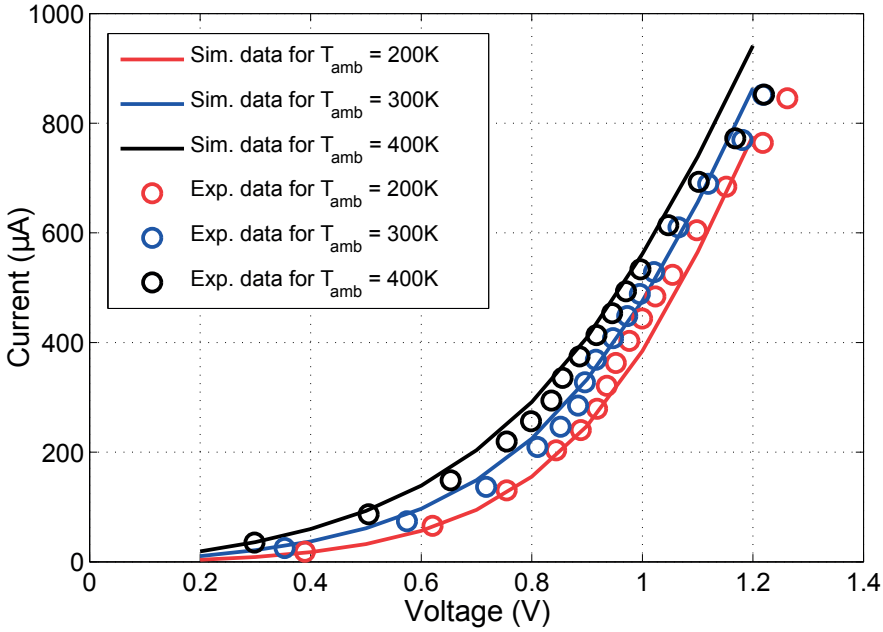


Figure 3.5: The programming $I--V$ characteristics for different ambient temperatures (200 K, 300 K, 400 K) from the simulations compared with those of the experimental measurements. The model is able to capture the sub-threshold regime accurately.

ing), the conductance of the phase-change material is high such that the resistance of the heater dominates the $I-V$ characteristics. The model is able to capture the current flowing through the device accurately. The temperature-dependent conductivity captures the exact characteristics for different ambient temperatures.

3.2.5 Resistance vs. programming power

RESET being the power-hungry process, the device efficiency is usually determined by the programming power required to switch (RESET) the device. Figure 3.6 depicts the experimental measurements and simulation results of the resistance vs. programming power plots for different ambient temperatures. It can be seen that the device is more efficient at higher ambient temperature. This is straight-forward because given the higher temperature at the device boundary, only a lower power is required to attain the melting point. However, apart from the temperature difference at the device boundary, also the location of the hotspot closer to the bottom electrode (in the case of 400 K) is due to the increased Joule and Thomson heating contributions at higher temperatures. This directly gets translated into lower programming power. Regarding the SET and RESET resistances, the conductivities of both the crystalline and the amorphous GST are field- and temperature-dependent. The higher the temperature, the higher the conductivity and hence the lower the SET and RESET resistances.

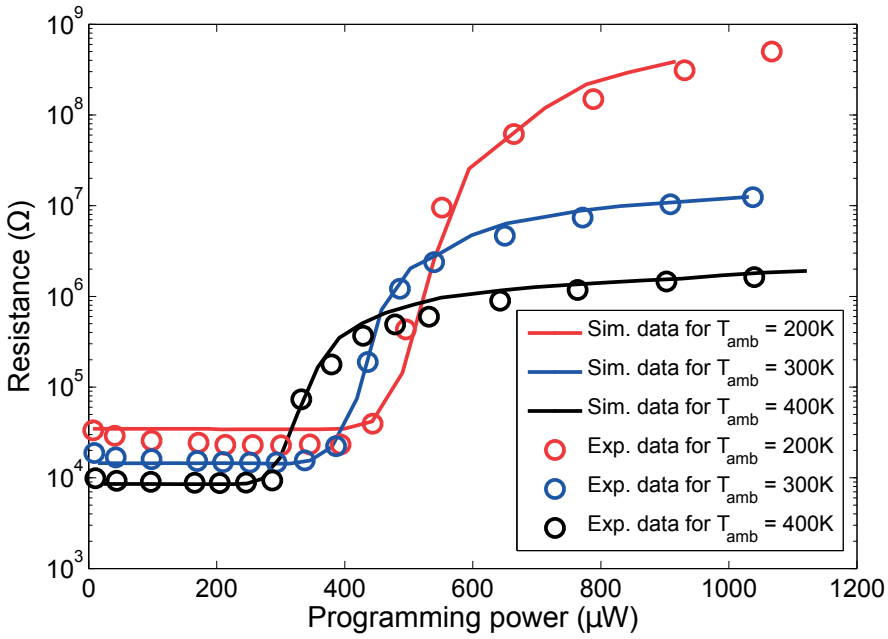


Figure 3.6: The resistance vs. programming power plots for different ambient temperatures (200 K, 300 K, 400 K) from the simulations compared with the measured data. The temperature-dependent conductivity closely captures the evolution of the SET and the RESET resistances with applied input power.

3.2.6 Experimental measurements

Figures 3.5 and 3.6 shows a complete set of $I-V$ and programming curves measured at different ambient operating temperatures ranging from 200 K to 400 K in steps of 100 K. For each point in the programming curves, the PCM device was first RESET with a 5 V voltage pulse with sharp leading and trailing edges of 16 ns. To SET the device, then a triangular pulse with a peak voltage of 3.5 V and a long trailing edge of 6 ms was applied to ensure complete crystallization at all temperatures. Next, voltage pulse of varying amplitude with a duration of 1000 ns was applied to the device to ensure steady-state melting, with sharp trailing edges of 16 ns to prevent unwanted re-crystallization. The current that flows through the PCM cell was measured and the dissipated power calculated according to

$$P_{cell} = I_{cell} \cdot V_a - R_{ser} \cdot I_{cell}^2, \tag{3.1}$$

where I_{cell} is the measured device current, V_a is the applied voltage and R_{ser} is the series resistance of the PCM device ($R_{ser} \approx 5$ kΩ). The low-field resistance was subsequently measured at room temperature at a read voltage of 200 mV. The experiments was repeated 10 times per temperature and the mean values were plotted in the Fig. 3.6.

3.3 Impact of bias-polarity on device operation

Unlike the Joule heating component, the thermoelectric heating components all are polarity-dependent and hence irreversible. This implies that the direction of current flow ($\vec{J}$) through the device greatly impacts the thermoelectric heating terms, including the interface Peltier term, and therefore the overall temperature distribution attained is different for opposite directions of current flow. With the square dependence on the current density, the Joule heating component (J^2/σ) is uninfluenced by the direction of the current flow.

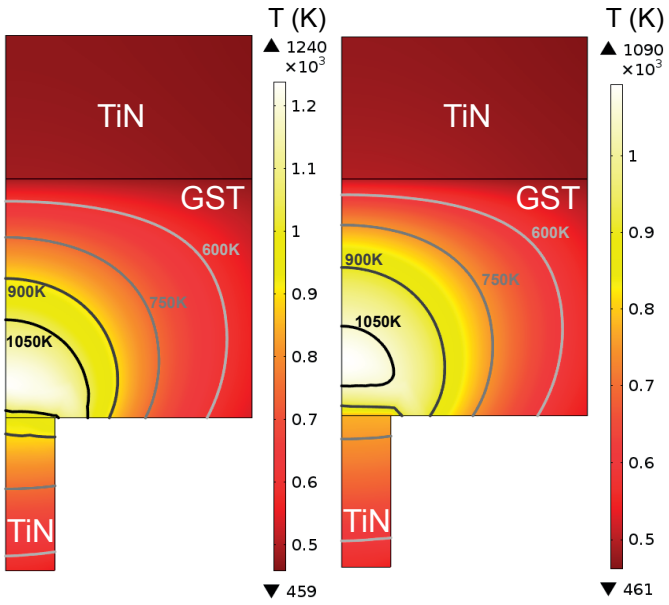


Figure 3.7: 2D thermal profile of the device for opposite bias polarity conditions. (a) Positive bias (+600 μW). (b) Negative bias (-600 μW); In the positive bias scenario, much higher temperatures are attained for the same applied input power.

Figure 3.7 illustrates the temperature distribution attained within the device for positive and negative bias. Owing to the thermoelectric effects, for the same applied input power, much higher temperatures are reached in the active region in the case of positive bias. Figure 3.8 shows the temperature difference map for the same applied power of 600 μW in the positive and the negative bias. Figure 3.9 depicts the temperature along the symmetry axis of the device. From the figures, it is apparent that for positive bias, heat is evolved at the GST-TiN bottom electrode interface and heat is absorbed (cooling) at the GST-TiN top electrode interface, and vice-versa for negative bias. Owing to the asymmetrical geometry of the device (narrower bottom electrode), the heat absorption and evolution at the top and bottom electrodes are not equal, and this explains the distinct behavior exhibited by these devices for opposite bias-polarity conditions (Fig. 3.12).

Figure 3.10 summarizes the resistance vs. programming power for the opposite bias-polarity condition. The power required to RESET the device in the case of negative polarity is much

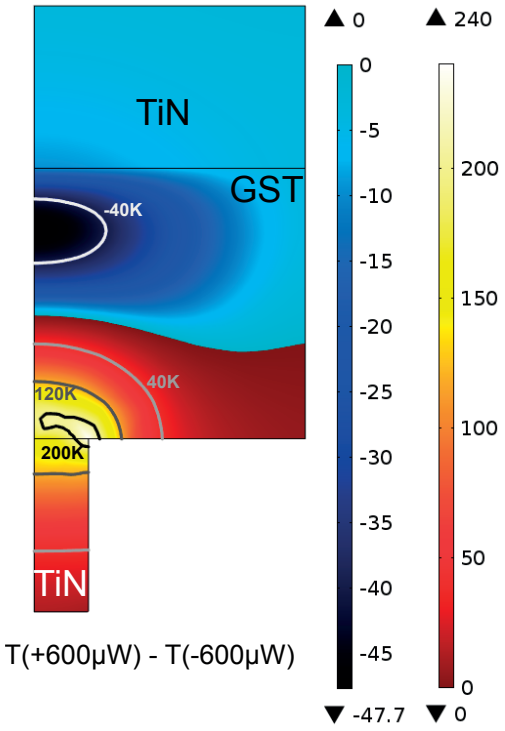


Figure 3.8: Temperature difference between the same applied power input in positive bias (+600 μW) and negative bias (-600 μW). Clearly, in the case of positive bias, higher temperatures are attained neat the GST-TiN bottom interface.

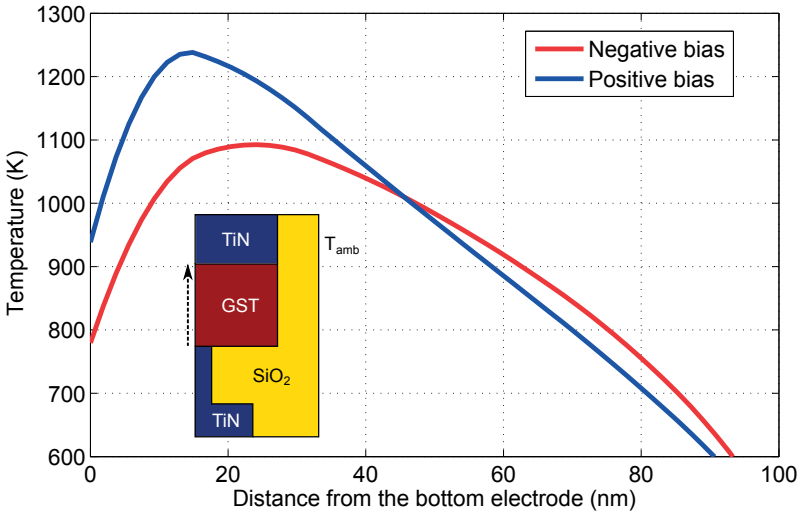


Figure 3.9: 1D temperature plot along the symmetry axis (see inset) for positive and negative bias for the same applied input power of 600 μW.

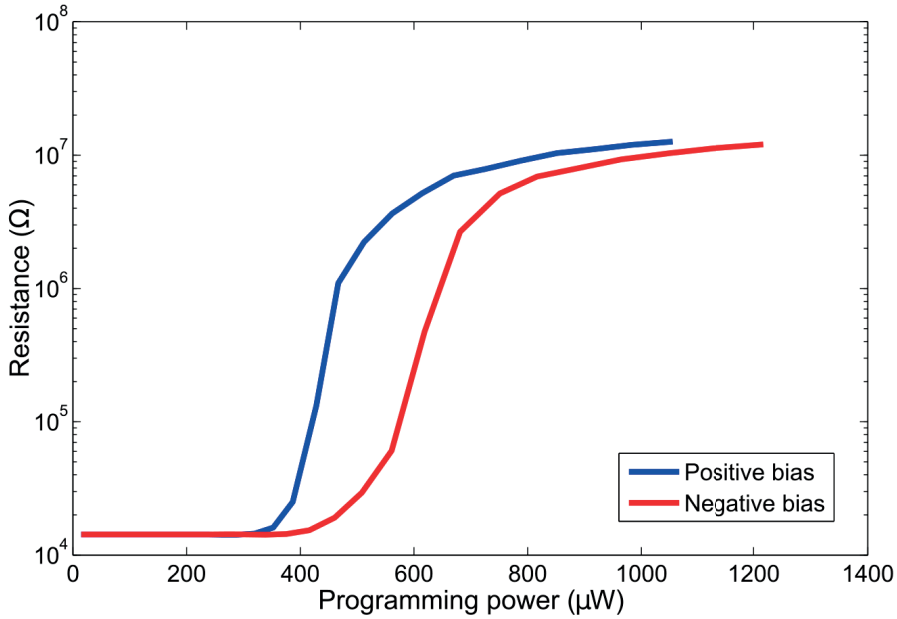


Figure 3.10: Resistance vs. programming power plot for positive and negative bias.

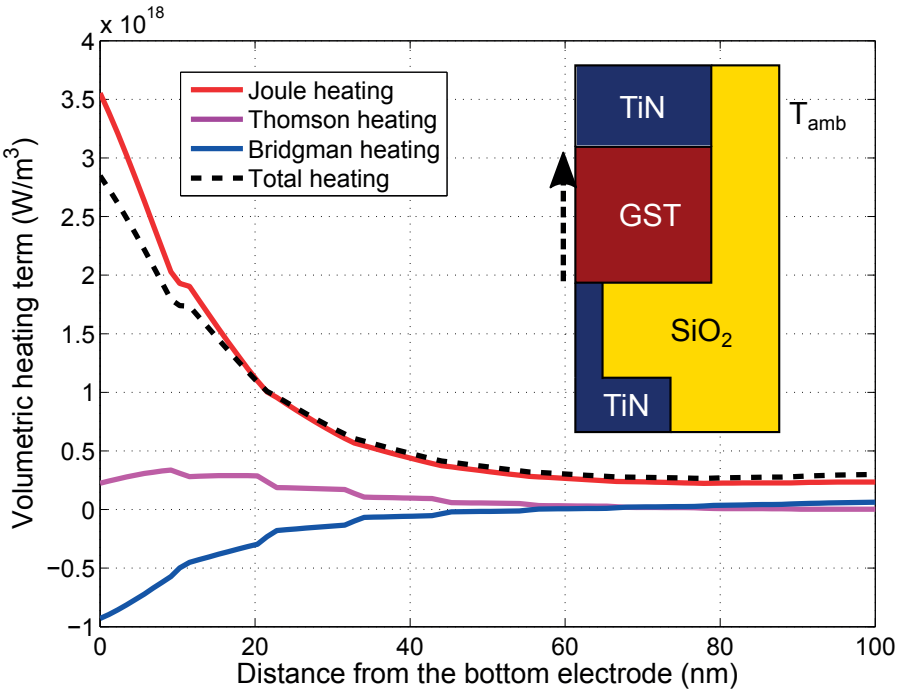


Figure 3.11: Individual volumetric heating contributions (W/m^3) for a negative bias with an applied power of 600 μW along the symmetry axis of the device.

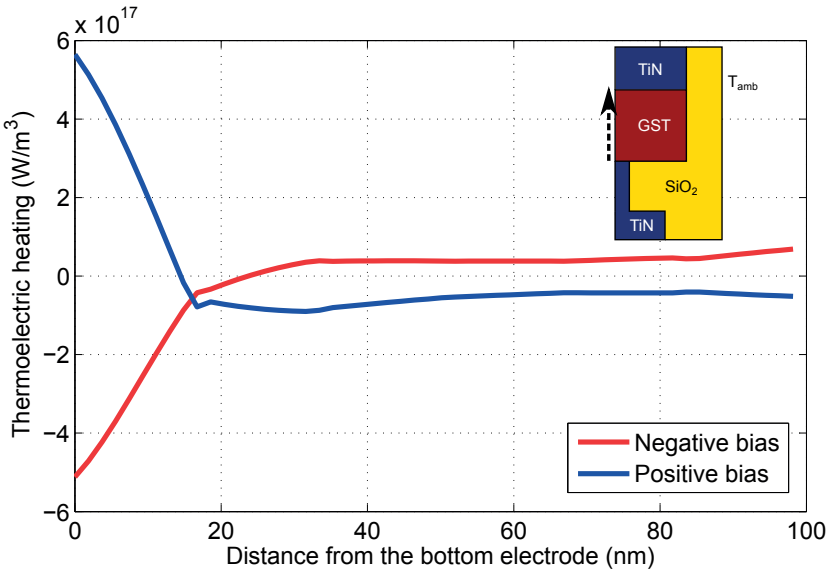


Figure 3.12: Total thermoelectric heating contribution (W/m^3) for an applied power of 600 μW in both the positive and negative bias conditions measured along the symmetry axis of the device.

higher than that of the positive bias. In the case of positive bias, the thermoelectric heating adds to Joule heating, and this helps in pushing the hotspot even closer to the bottom electrode (heater), thereby making the device more efficient. In contrast, for negative bias, the thermoelectric heating suppresses (cools down) the Joule heating at the GST-bottom electrode interface, and this moves the hotspot further away from the heater interface into the phase-change material (Fig. 3.11 and Fig. 3.12). Thus the thermoelectrics play a vital role in device operation.

3.4 Influence of Material properties on device operation

The thermoelectric model was able to accurately capture the device characteristics. The model can therefore be employed as a tool to study the impact of critical material parameters and device geometry on the device performance. The resistance vs. programming power curve is a key parameter for device efficiency. In this Section, the influence of varying the different material parameters at room temperature (300 K) was studied and is compared with that of the experimental data measured at room temperature.

3.4.1 Thermal conductivity influence

While programming, the thermal conductivity of the crystalline GST helps in better heat confinement within the active region of the device and hence directly correlates to device

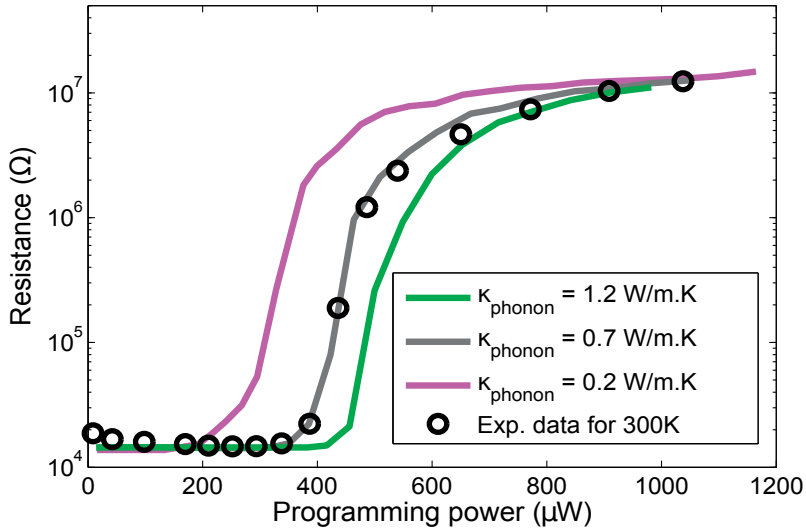


Figure 3.13: Resistance vs. programming power curve at room temperature (300 K) for different values of the phononic contribution of the thermal conductivity of crystalline GST compared with the experimental measurement at room temperature.

efficiency. As the electronic contribution follows the Wiedemann–Franz law, the resistance vs. programming power plot is studied for different values of the phononic contribution, and the results are summarized in the Fig. 3.13. To design an efficient device, the phase-change material should be chosen such that it has a lower phononic contribution in the crystalline phase.

3.4.2 Influence of electrical conductivity

The value of the activation energy used in the conductivity model for the crystalline/amorphous phase determines the SET/RESET resistance. It also defines the temperature dependence of the conductivity. Figure 3.14 and Fig. 3.15 depict the changes in the SET and RESET resistances as a function of the corresponding activation energy in eV. By proper selection of these materials, the dynamic resistance range can be increased.

3.4.3 impact of thermal boundary resistance

GST-SiO_2 thermal boundary resistance influence The thermal boundary resistance between the GST and SiO_2 influences the switching power. The higher the thermal boundary resistance between GST and SiO_2, the higher the thermal heat confinement in the active region of the phase-change material and, hence, the lower the power required to switch the device. Figure 3.16 shows the influence of $TBR_{GST\text{-}SiO_2}$ on the resistance vs. programming power measured at room temperature. Clearly, the higher the thermal boundary resistance,

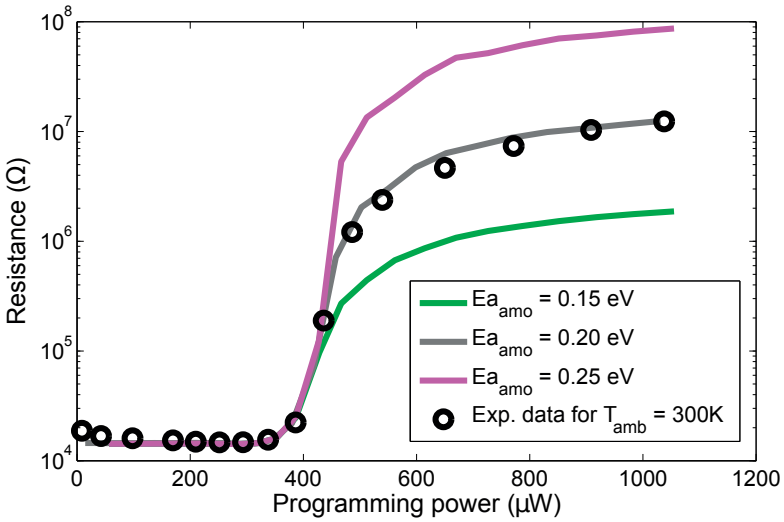


Figure 3.14: Resistance vs. programming power at room temperature (300 K) for varying values of the amorphous activation energy of GST compared with the experimental measurement at room temperature.

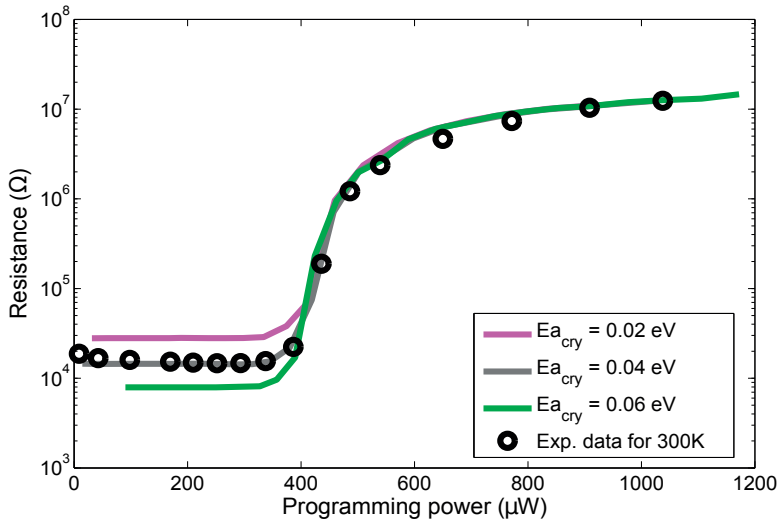


Figure 3.15: Resistance vs. programming power at room temperature (300 K) for different values of the crystalline activation energy of GST compared with the experimental measurement at room temperature.

the higher the efficiency of the cell owing to the better thermal confinement.

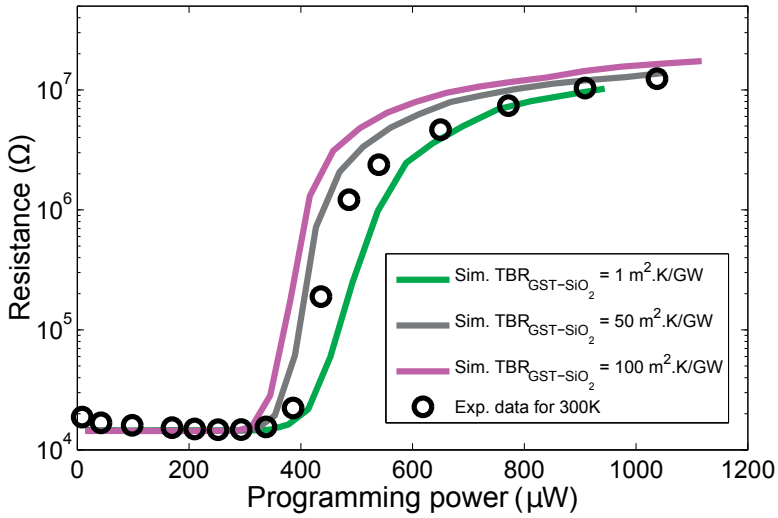


Figure 3.16: Resistance vs. programming power at room temperature (300 K) for varying values of the thermal boundary resistance between the GST and SiO_2 compared with the experimental measurement at room temperature.

3.5 Thermal disturbance in phase-change memory arrays

Given the aggressive scaling trends, particularly in a $4F^2$ array configuration, the cell pitch (distance between adjacent cells) reduces correspondingly. However, the temperature range attained within a PCM device do not scale with device dimensions as the melting and the crystallization temperatures are material properties and does not decrease with scaling. Therefore with scaling, there is a increasing possibility of unintended thermal interference between a PCM device that is being programmed and its neighboring devices. This problem is commonly referred to as the thermal disturb, thermal cross talk or program disturb. This heating of adjacent devices may result in unintended crystallization of PCM material in them thus altering the programmed state and subsequent data loss. This is a significant problem and could be a key challenge for the current scaling trends associated with PCM arrays. It has triggered significant research efforts in recent years [Kim et al., 2010; Russo et al., 2008].

Thermally activated crystallization of the high-resistive states results in a significant reduction of the resistance, leading to data retention failures in both SLC and MLC PCM. One possible cause of thermally activated crystallization is the thermal disturb. Once a cell is programmed to the RESET state, repeated programming of its adjacent cells might induce crystallization of the cell because of the heat diffused from the neighboring cells. This is also known as thermal crosstalk or proximity disturbance [Kim et al., 2010].

Thermal disturb is another reliability characteristic that is unique to PCM because of the deliberate use of heat as a programming mechanism. With the continuously decreasing cell pitch and given the large temperature gradient attained within the device while RESET

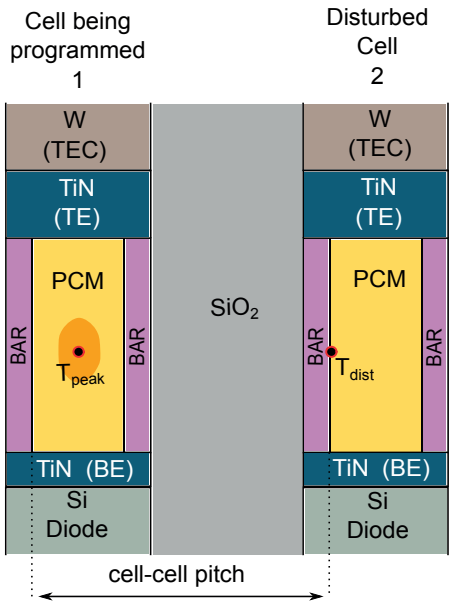


Figure 3.17: Cross-sectional view of two adjacent cells in an array. While cell 1 is being programmed the temperature also attains cell 2. T_{dist} should be well below the crystallization range of temperature for reliable storage operation. T_{peak} is the peak temperature attained in the cell being programmed.

programming, neighboring cells are prone to partial crystallization and an eventual data loss. Moreover, with increased neighbor temperature, also the drift behavior is affected by thermal disturb owing to the temperature-dependence of drift [Ielmini et al., 2008; Sebastian et al., 2015]. This leads to wider resistance distributions when the cells are thermally disturbed at an early stage after RESET programming.

The thermal disturb phenomenon has been extensively modeled and simulated. Pirovano et al. showed with thermal simulations that thermal disturb is low enough, at least at the 65 nm node, so that it does not affect data retention [Pirovano et al., 2004]. Furthermore, Russo et al. performed a similar study for nodes extending down to 16 nm and found that isotropically-scaled devices (all device dimensions scale with the technology node) are completely immune to the thermal disturb issue, whereas the thermal disturb increases with non-isotropic scaling [Russo et al., 2008].

Kim et al. showed that isotropic scaling, where the cell dimensions are scaled by the same scaling factor, causes the thermal disturb to scale proportionately by the same scaling factor [Kim et al., 2010]. Therefore, isotropic scaling makes thermal disturbance neither better nor worse than the current node. However, when cell distances are scaled more aggressively at successive technology generations to improve density and performance, thermal disturbance can be exacerbated. Therefore, any scaling optimization should involve the careful evaluation of thermal disturbances.

3.5.1 Electro-thermal model for the analysis of thermal disturbance

I present a compact electro-thermal modeling approach that serves as a simple yet powerful tool to study the problem of thermal disturb in dense PCM arrays. As described in the previous section, FEM-based modeling approaches are very popular and can provide numerous insights into PCM device operation. They are well-suited for device-level studies. However, in the case of array-level studies, they are typically time-consuming and computationally intensive, owing to the asymmetrical geometry. Hence, there is a need for a simpler compact modeling approach to study thermal disturb.

The proposed compact electro-thermal model is simple and can yield reasonably good estimates of the spatial and temporal thermal variations across a PCM cell array. Interdependent electrical and thermal sub-models constitute the compact electro-thermal model (CETM). In the CETM approach, each PCM device can be divided into elements based on its components (electrodes, phase-change material, insulating layer) and typical geometries such as cuboid, cylinder, etc. The electrical sub-model consists of a simple resistive network representing the electrical resistance of each conductive element including the "ON state" resistance of PCM. For an applied voltage, the electrical sub-model yields the current flowing through each element and hence the power (P_{th}).

The thermal sub-model uses a simple thermal-equivalent electrical circuit for each element to obtain the heat flow and temperature distribution throughout the cell at the specified node points. The well-known analogy between heat flow and electrical conduction is applied in it, with the temperature represented as voltage and the heat flow represented as electric current. The thermal and the electrical equivalence are summarized in (3.2) and (3.3). σ and κ are the electrical and the thermal conductivity, respectively. J and q represent the current density and the heat flux. V and T denote the voltage and temperature. C_{el} and C_p are the electrical and the specific heat capacitance, respectively.

$$\nabla V = -\frac{\mathbf{J}}{\sigma} \Leftrightarrow \nabla T = -\frac{\mathbf{q}}{\kappa} \tag{3.2}$$

$$C_{el}\frac{dV}{dt} = J \Leftrightarrow C_p\frac{dT}{dt} = q \tag{3.3}$$

In the thermal sub-model, each element can be modeled as a node containing thermal resistances representing its resistance to heat flow, the thermal capacitance representing its ability to store heat, and a current source (only if there is active electric power dissipation) to represent the source of heat. P_{th} represents the input power from the electrical sub-model for each element.

The electrical and the thermal sub-model corresponding to the confined cell architecture are shown in Fig. 3.18. The thermal resistances R_x, R_y and R_z represent the resistance to the heat flow in the orthogonal directions x, y and z, respectively. C_{th} defines the thermal

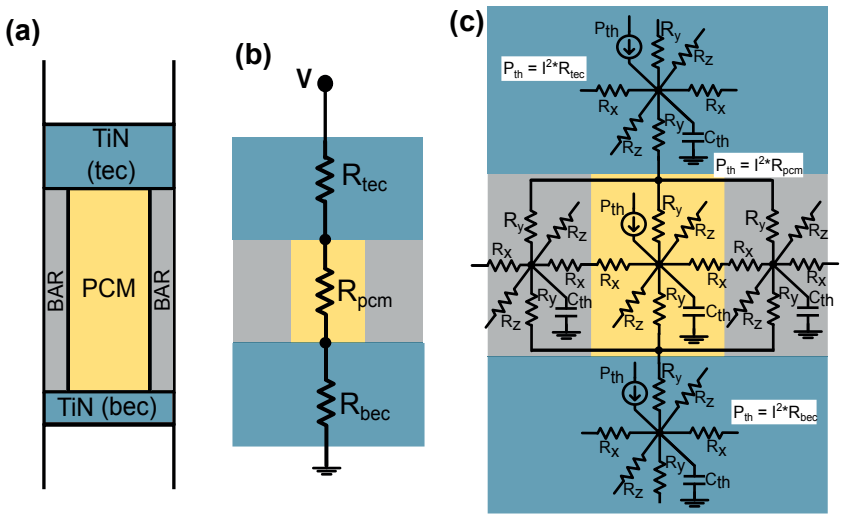


Figure 3.18: (a) Vertical cross-section of the confined cell architecture; BAR is the insulating layer surrounding the phase-change material. (b) Electrical sub-model for calculating P_{th}. R_{tec}, R_{pcm} and R_{bec} represent the electrical resistances of each element. I is the current flowing through the element. (c) Thermal sub-model circuit for a temperature profile evaluation. R_{x}, R_{y} and R_{z} represent the thermal resistances in the x, y, z orthogonal directions respectively. P_{th} is the Joule heating power input and C_{th} is the thermal capacitance for each element.

capacitance at each node (3.3). These resistances and capacitance can be calculated from the device geometry and material properties, such as the thermal conductivity and the specific heat capacitance. The temperature map from the thermal sub-model can then be used to determine the location of the "hotspot". To evaluate the temperature distribution within an array, such CETMs are repeated together with the respective nodes for the insulation barriers and the metal interconnects.

3.5.2 Confined-cell topology - case study

To validate the accuracy of the CETM approach, the thermal cross-talk analysis is performed on a representative confined-cell architecture [Kim et al., 2010]. One reason for choosing that topology is that the current generation of PCM devices which are being fabricated are all confined-cell topology devices. Realistic cell geometry parameters are obtained from the literature. For the complete switching of the device, the temperatures at the nodes (1 and 2 in Fig. 3.18) should be above the melting temperature of GST. The switching voltage for programming can be evaluated between these sub-models in a few iterations until the peak temperatures at specific nodes (based on the device geometry) reach the melting temperature. In the cell architecture studied, the peak temperature occurs in the center of the cell, spatially represented by the node of the PCM element. Although the resolution of the temperature distribution depends on the number of elements in the electrical and the thermal sub-model, there is no need to have any prior knowledge about the temperature map, as we are more

interested in the neighboring temperatures than in the exact switching power. In the case of asymmetrical cell geometries (where the peak temperature does not occur in the center as in the case of mushroom-type devices), we can break down the PCM element into a few more (for example, three or four) rather than just one element, in order to find the spatial location of the peak temperature. The CETM does not consider the phase-change kinetics or the phase dependence of the thermal conductivity. However, for the thermal disturb analysis this is not a significant drawback.

3.5.2.1 FEM vs CETM simulations

The input power is calculated in such a way that the temperature at the PCM boundaries is greater than the melting temperature of 900 K, to make sure that the cell is RESET to high resistance. The CETM can be implemented and simulated in SPICE/Spectre-like simulators. For validation purposes, the CETM model is built and simulated for an array of 9 cells (3 × 3) with a diode as access device. The results are then compared with the complete 3D FEM-based simulation results obtained in COMSOL Multiphysics® for the same geometry. Figure 3.19 summarizes the comparison results, and shows the good match between the CETM and FEM simulations. In CETM(1), the entire PCM component is modeled as a single element, whereas in CETM(2) it is divided into two elements. Clearly, CETM(2) has a higher resolution of the temperature map and marginally better accuracy than CETM(1).

3.5.3 Scaling study

To illustrate an application of the CETM approach, a scaling study on the thermal disturb effect is presented. Assuming the extreme scenario, a high RESET pulse is applied to program the cell and the thermal disturb on an adjacent cell is evaluated. T_{dist} denotes the temperature variations of the disturbed cell at the nearest boundary to the cell being programmed (inset of Fig. 3.20). T_{dist} is studied for different values of the cell pitch, with the input power normalized in each case, to obtain the same peak temperature. In order not to disturb the adjacent cell, T_{dist} should be significantly below the crystallization temperature of the phase-change material, typically in the range of 460 K-490 K. The band (marked in brown) in Fig. 3.20 denotes this critical temperature range. For the confined cell under investigation, T_{dist} exceeds this critical temperature when the cell pitch dimensions are smaller than 20 nm. This is an indication of the challenges posed by thermal disturb. To explore the possibility of bringing down this temperature for 20 nm cell pitch dimensions, T_{dist} is evaluated for various thermal conductivities of the thermal barrier layer (BAR in Fig. 3.18). This is illustrated in Fig. 3.20. The results imply that for barrier layers with thermal conductivities of less than 0.3 W/mK, T_{dist} falls below the temperature range of fast crystallization. Other approaches to mitigate the thermal cross talk problem include altering the cell and array geometry and introducing additional thermal insulation barriers. The CETM could serve as a powerful tool to study all these aspects.

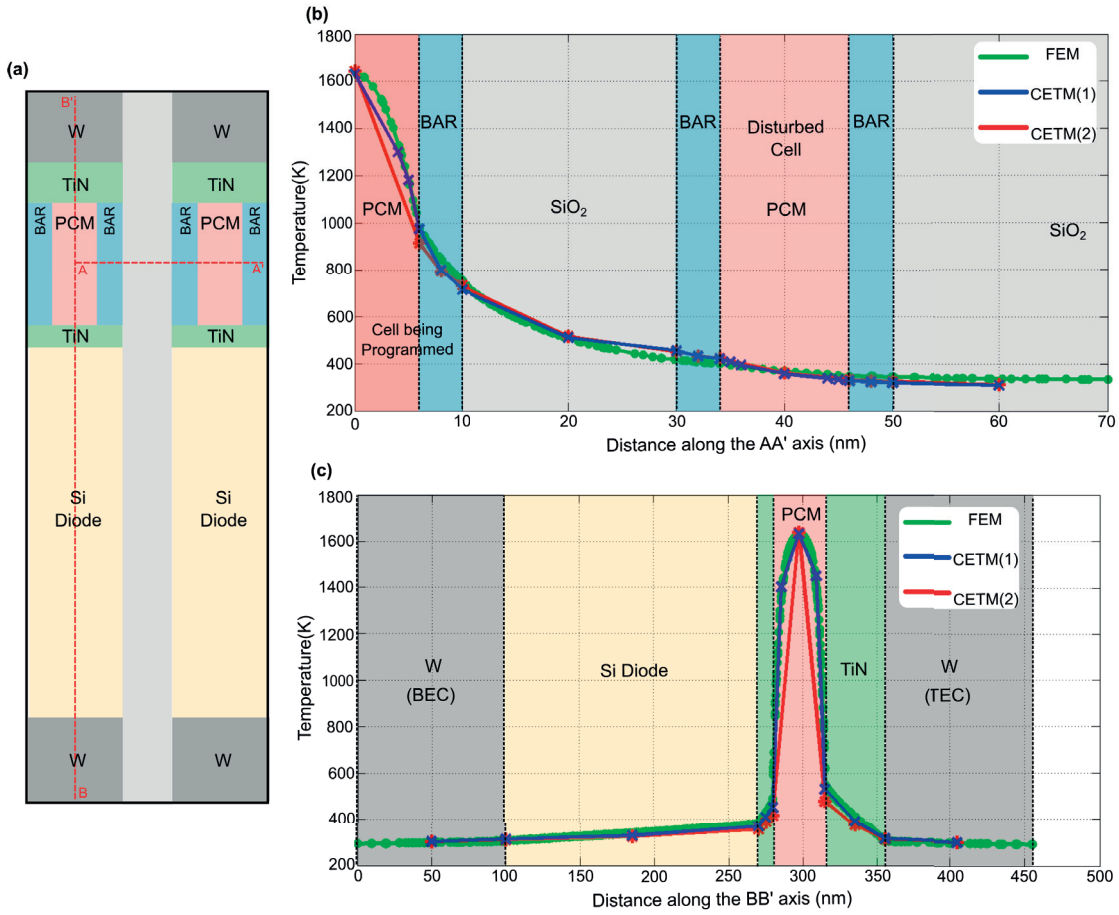


Figure 3.19: (a) Cross-sectional view of two adjacent cells with the device geometry. (b) Temperature profile along the AA' axis showing the temperature map across the cell being programmed and the disturbed cell. (c) Temperature profile along the BB' axis showing the temperature map along the vertical direction of the cell being programmed.

3.5.4 Advantages and limitations

The proposed model can be used to investigate thermal disturb issues in highly dense PCM arrays. Unlike the thermal profile within the device, in the case of thermal disturb study, we are mainly interested in the temperature at certain critical nodes of the array. The model is able to give a good estimate, with much less complexity and computation time involved. The inherent simplicity and speed of the simulation make this model a powerful tool for providing insights into the thermal characteristics of a PCM array and in particular for addressing thermal disturb. The model can also be used to evaluate PCM cell and array designs that mitigate thermal interference. The effects of altering the thermal properties of materials, the device geometry and the addition of thermal insulating barriers can readily be studied using this model. The limited resolution of the thermal profile and the inability to capture the phase-change kinetics are the major limitations of this approach.

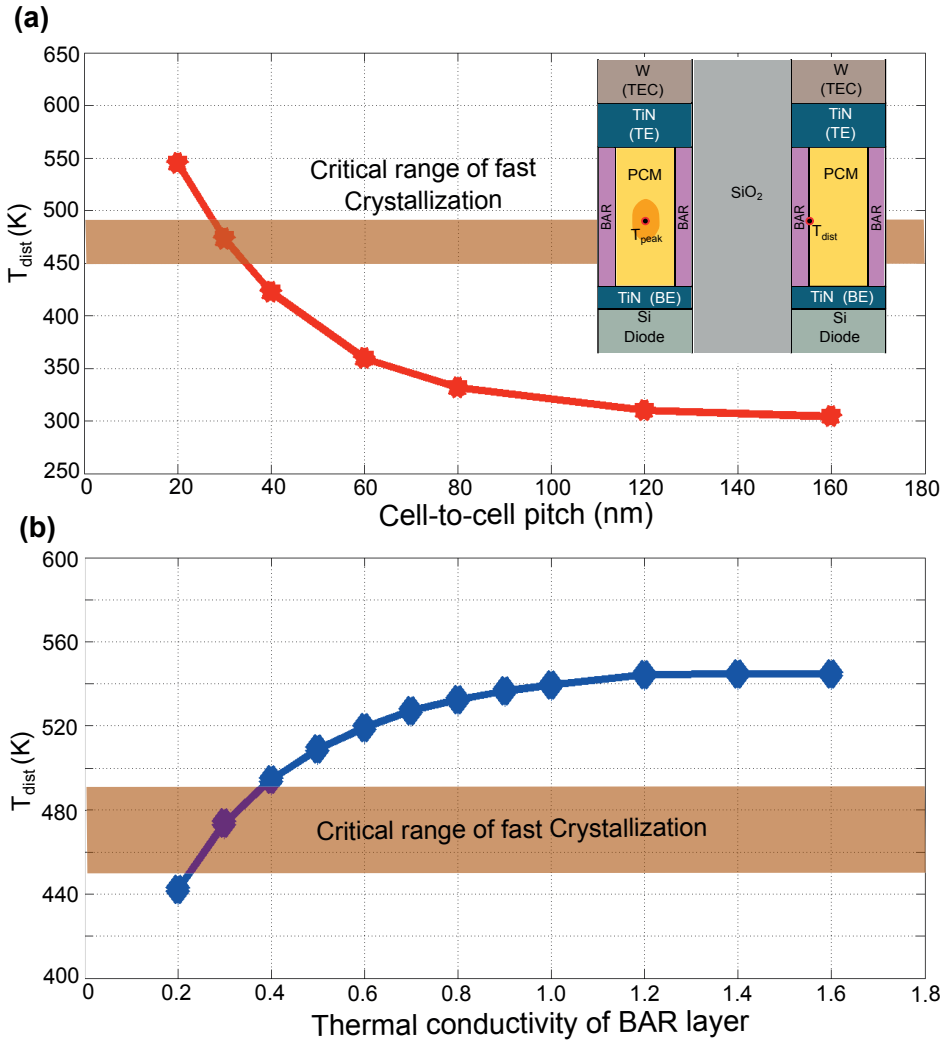


Figure 3.20: (a) The trajectory of the thermal disturb temperature, T_{dist}, for decreasing values of the cell-to-cell pitch. (b) The impact of the thermal conductivity of the insulating BAR layer on the disturb temperature T_{dist}.

3.6 Novel device design using thermoelectrics

As described in the previous section, thermal disturb seems to be an important limiting factor in the continued scalability of PCM arrays. Based on the thorough knowledge of thermoelectric physics obtained from the completely validated thermoelectric model, I propose a couple of novel device-design ideas that exploit the influence of thermoelectric effects to enhance device performance. The first idea is based on using thermoelectric effects for better thermal confinement, thereby making the device more efficient while simultaneously reducing the thermal disturb temperature. The second idea focuses on increasing the MLC capability by using the opposite-polarity bias voltages for programming the intermediate resistance levels.

3.6.1 Thermal confinement using thermoelectrics

With continued scaling in highly dense PCM arrays, thermal disturb pose a serious threat to the device pitch, especially at lower technology nodes. Given that thermoelectric effects play an essential role in device switching dynamics and are increasingly important to be considered for future device design. Moreover, it can be explored to improve device performance. The current flowing through an interface between two dissimilar materials with different Seebeck coefficients results in heat being absorbed or evolved depending on the direction of the current flow. This is known as interface Peltier effect, and is prominent at the GST-TiN bottom electrode interface in the typical mushroom-type devices.

As discussed in Section 3.3 for the polarity-based study on the mushroom cells, the thermoelectric heating assists the Joule heating in the positive-bias scenario. This means that the interface Peltier heat is evolved on the GST-TiN bottom electrode interface, whereas the heat is absorbed on the GST-TiN top electrode interface. Although the interface Peltier effect renders the device more efficient, Peltier cooling (heat absorption) is happening at the top electrode interface.

I propose a novel device design where the thermoelectric effects can be exploited to reduce the neighboring temperature while simultaneously increasing the device efficiency. The interface Peltier effect at the GST-TiN bottom electrode interface is predominant owing to the high current density and the high temperature attained in the vicinity to this interface. In contrast, because of the lower current density and lower temperature typically seen at the GST-TiN top electrode interface, the interface Peltier effect at the top interface is minimal.

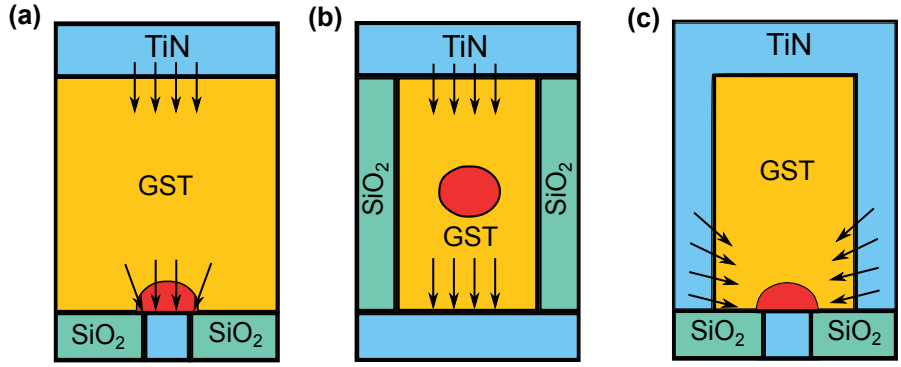


Figure 3.21: State-of-the-art device topologies of PCM devices and the proposed device design showing the typical hotspot location in the device. (a) Mushroom-device topology with hotspot located closer to the GST-bottom electrode interface. (b) Confined-cell topology with the hotspot located midway between the top and the bottom electrode. (c) Proposed device design, where the hotspot is still located close to the GST-bottom electrode interface, similarly to the mushroom-device topology.

To make the device more efficient while simultaneously reducing the neighboring temperature, we should focus on two aspects: (a) The current should flow through the GST-TiN top electrode

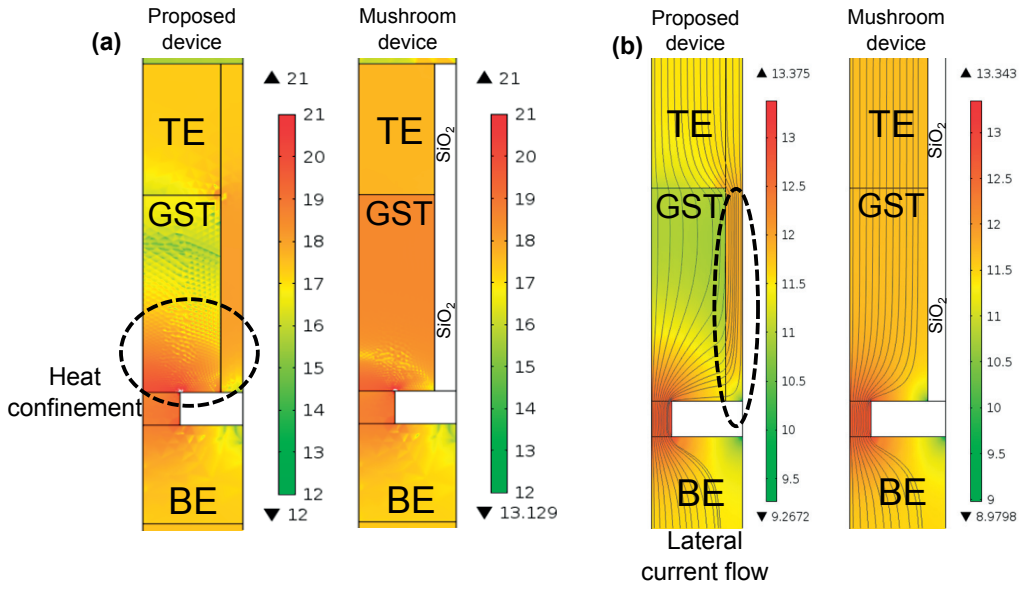


Figure 3.22: (a) Comparison of the volumetric heat-generation component across the proposed and the mushroom device expressed on logarithmic scale (W/m^3). Evidently, there is better thermal heat confinement in the new device design. (b) Comparison of the current density across the proposed and the mushroom device. Clearly in the new design, the majority of the programming current flows through the projected top electrode.

interface in lateral direction such that the cooling at this interface helps both increase the efficiency and reduce the temperatures attained in the neighboring devices (thermal disturb). (b) A higher current density should be attained at the GST-TiN top electrode interface such that there is significant interface Peltier contribution also from this interface.

Fig. 3.21 summarizes the proposed device design and compare, it with two other conventional device topologies. The new design closely resembles the mushroom topology, except that the top electrode is extended all along the length of the device into the phase-change material. The extended top-electrode serves two purposes, namely, to provide a lateral direction of the current flow at the GST-TiN top electrode interface and to increase the current density at the GST-TiN top electrode interface.

3.6.1.1 Characteristics of the proposed device

The proposed device was implemented as a 2D axial-symmetry geometry in COMSOL and the thermoelectric model developed was used to assess its performance. The model is simulated and the results compared with those of the conventional mushroom-type device. For the sake of simplicity and to have equal volumes of GST in both cases, the projected top electrode domain is set be SiO_2 for the mushroom-type device and TiN for the proposed device-design.

Figure 3.22 illustrates the volumetric heat generation across the device. Fig. 3.23 compares

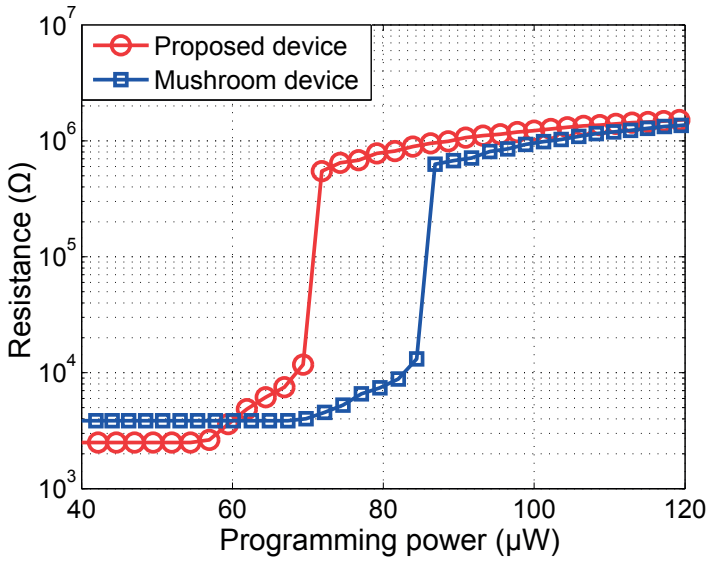


Figure 3.23: Comparison of the resistance vs. programming power plot of the mushroom type device with the proposed device. Evidently, there is a 16% reduction in the RESET programming power. The decrease in the GST volume owing to the projected electrode is responsible for the lower SET resistance.

the resistance vs. programming power curve of the mushroom device with the proposed device. Evidently, there is a 16% reduction in the power required to RESET the respective devices. The close proximity of the top and the bottom electrodes along with the better thermal confinement, owing to the extended top electrode, contributes to the significant reduction in the programming power.

3.6.1.2 Impact on thermal disturb

For the given direction of current flow, heat is absorbed at the GST-TiN top electrode interface (interface Peltier), it aids in bringing down the neighboring temperature, thereby providing a kind of thermal confinement to the active region. As a result of this, there is considerable reduction in the temperatures attained at the neighboring cells, thereby reducing the phenomenon of thermal disturb (Fig. 3.24).

3.6.2 MLC capability using thermoelectrics

MLC storage is one of the promising feature of PCM which made it attractive among its competitors. With a thorough understanding of the thermoelectric at the active region, it is clear that thermoelectric effects play a critical role in determining the hotspot location within the new device. The size and the location of the hotspot determine the attainable resistance levels and hence directly influence the MLC operation.

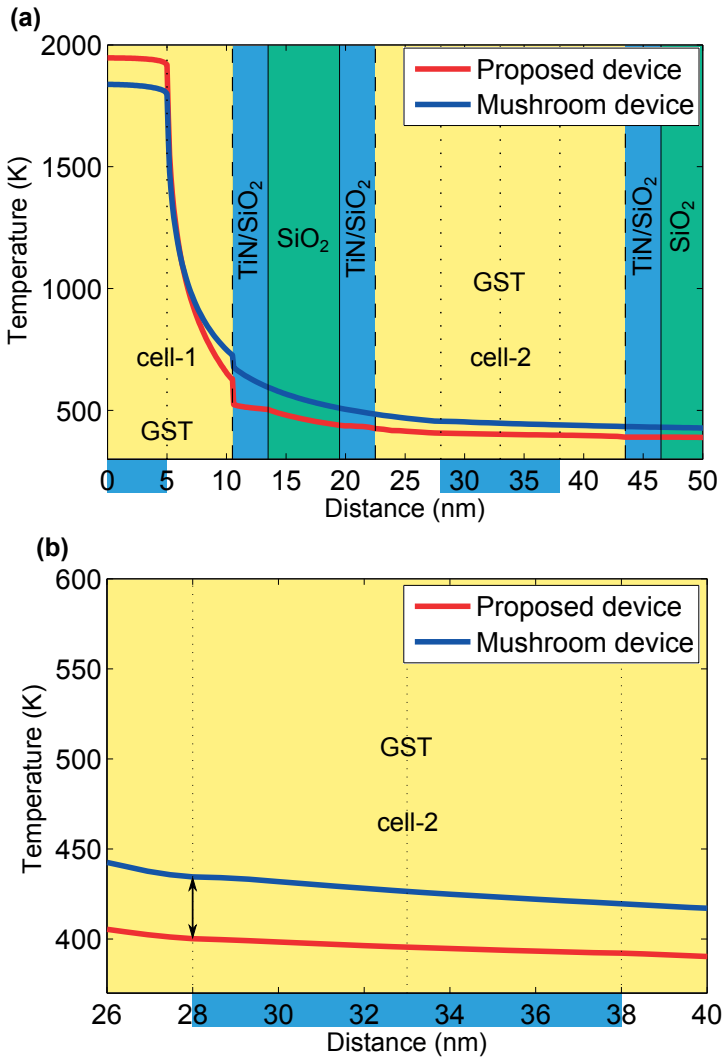


Figure 3.24: Comparison of the thermal profile attained at the neighboring cells for the mushroom device and the proposed device. From the latter, there is 6%-8% temperature reduction at the GST boundary of the adjacent device.

Recently, Kim et al. proposed a novel device-design approach to address the resistance drift phenomenon. They demonstrated a six-fold decrease in the drift coefficient values of high-resistance states by using a metallic surfactant (metal nitride) layer as an alternative read current path to the amorphous region [Kim et al., 2013]. The idea is to decouple the amorphous regions, which are prone to resistance drift, from the read current path.

The thermoelectric physics can be used to control the location of the hotspot within the phase-change material. By having a additional metallic surfactant layer to have alternate current path to determine the resistance during the readout operation. I propose a new device design as illustrated in the Fig. 3.25.

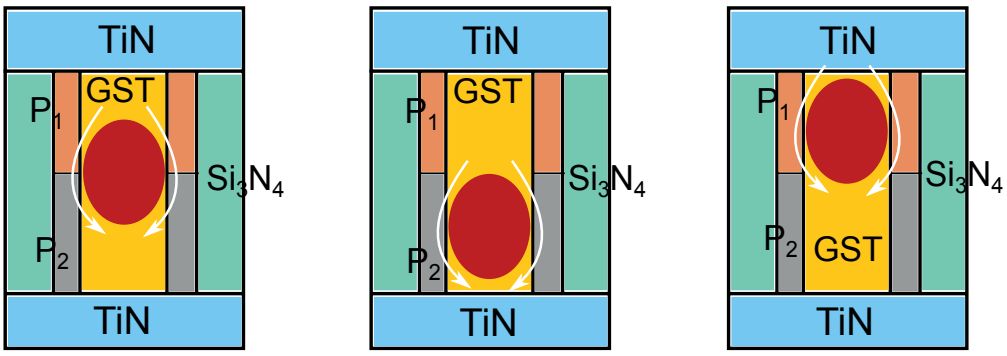


Figure 3.25: (a) Hotspot location without considering any thermoelectric effects. (b) Location of the hotspot with thermoelectric effects for positive applied bias. (c) Applied negative polarity with thermoelectric effects.

The idea is to have two different metallic projection layers with different resistivities (P1 and P2) surrounding the phase-change material towards the top and the bottom electrodes in the conventional confined-cell geometries. The projection layers, P1 and P2, are chosen such that their conductivities lie between those of the crystalline and the amorphous state ($\rho_{cry} > \rho_{P1} > \rho_{P2} > \rho_{amo}$).

For such a symmetrically confined cell, based on the polarity of the applied programming voltage, the location of hotspot will be either closer to the bottom electrode (positive bias) or the top electrode (negative polarity). The read current can therefore be decoupled from the amorphous phase-change material and made to flow through either the P1 or P2 layer, providing distinct resistances according to their conductivities. Depending on the resistivity of the projection layer (P1 and P2), different intermediate resistance states can be attained for the applied bias.

The proposed design is mainly focused on the confined-type devices for two reasons. (a) In the mushroom-type devices, although there is a difference in efficiency between programming with the positive and the negative bias, there is no significant (very minimal) movement of the hotspot. (b) The confined-type device is the new device topology, which is currently being explored owing to its simplicity and its amenability to scaling.

In a phase-change memory, currently multiple levels are stored by applying pulses of varying amplitudes (or varying duration of trailing edges), so as to program the cell at various intermediate resistive states. Unlike the mushroom-type devices, in the state-of-the-art confined-cell design, the opposite bias polarities have minimal impact on device operation owing to the symmetry of the device. However, in the novel device design, the thermoelectric effect can be used to store information in the intermediate resistance levels just by programming with the negative-bias voltages.

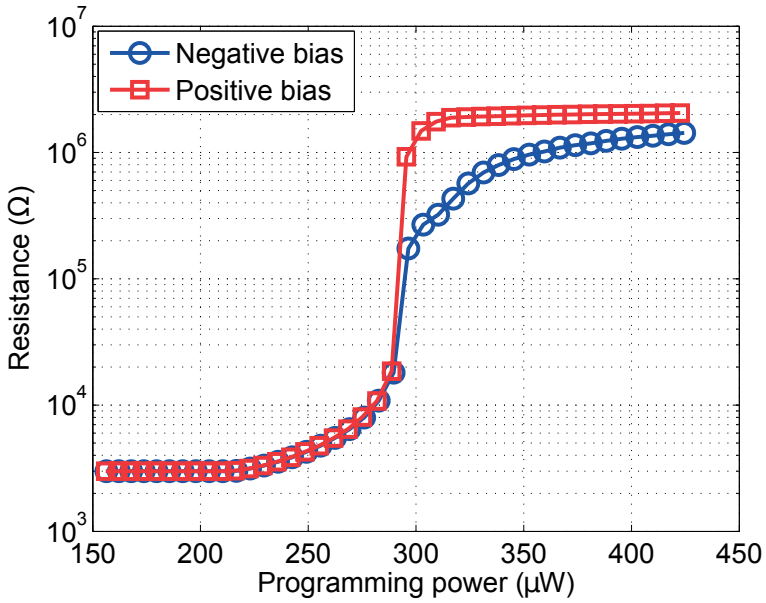


Figure 3.26: Resistance vs. programming power for different applied bias voltage. Different intermediate resistance levels are attained for the same applied programming power with opposite polarity.

3.7 Summary

The proposed comprehensive thermoelectric model has been completely validated with actual experimental device measurements. It was shown that the model can accurately capture the exact device operation under wide range of ambient operating temperatures. The model was simple to implement and could provide numerous insights for understanding the thermal and the electrical transport mechanism in the active region of the device. Therefore, the model was instrumental in studying the influence of various factors (material properties, boundary conditions, etc.) on device operation. The model can thus be used as a powerful tool to fine-tune the material properties, device geometry and design, in order to maximize device performance and improve its efficiency.

Apart from the knowledge about the thermal distribution within the device, the compact electro-thermal model is a simple but powerful tool for a quick and reasonably good estimate of the temperatures attained within the highly dense PCM array, which otherwise would be a time-consuming and complex task to perform in FEM-based simulations. The thermoelectric model provided information on the thermal profile within the device with such a precision that it provided numerous insights into the device operation. In contrast, the electro-thermal model provides fast and reasonably accurate estimates of the temperatures attained outside the device in dense memory arrays, where precision is not critical. Both models in combination can be used as a powerful tool to completely understand the thermal landscape attained within a PCM array.

At the end, I also presented two novel device designs derived from the understanding of the thermoelectric effects. I believe that ideas like these, although basic and far from actual implementation, will pave the way in the right direction for resolving future scalability issues which one can already foresee in terms of PCM scalability.

Contributions:

- **Calibration and validation of the thermoelectric model implemented by comparison of the simulation results with the experimental measurements.**
- **Based on the valuable insights obtained from the thermoelectric model, a thorough investigation on the bias-polarity-dependent behavior of the mushroom-type devices was performed.**
- **Conceived and developed the compact electro-thermal model for a quick estimate of the thermal distribution in dense PCM arrays.**
- **Proposed novel device design ideas using the thermoelectric insights into device operation to tackle the thermal disturb phenomenon and to increase the memory density through MLC capability.**

4 Multilevel-Cell Phase-Change Memory: Circuit architectures

From the aggressive, Flash-dominated non-volatile memory market, one can easily predict that all possible directions for improving the effective density will be exploited. One promising, but challenging direction is the MLC (Multi-level Cell) storage approach, where more than 1 bit of information is stored in each cell. High memory density leads to more functionality, and is currently in huge demand because of the proliferation of big-data storage applications. It has been argued that 2-bits/cell storage and read/write access times in the 100 ns to 1μs range are required to establish PCM as a tier in the memory hierarchy, somewhere between DRAM and Flash memory [Freitas and Wilcke, 2008]. Researchers have already demonstrated the feasibility of MLC in PCM technology [Close et al., 2013; Papandreou et al., 2011].

In this chapter, the process of achieving MLC capability in PCM devices and its characteristics are described. Although there is a wide contrast between the SET and RESET states, there are various factors that limit the practical realization of MLC in PCM. Resistance drift and array variability are the predominant factors affecting reliable MLC storage. Various MLC-enabling technologies are currently being explored to improve the drift resilience. One such approach is the extraction of non-resistance cell-state-based metrics. To reduce array-variability, in general, iterative programming schemes are used to program the intermediate resistance levels. Drift-immune readout metrics are described, followed by the CMOS implementation of novel readout architectures for the extraction of such metrics.

This chapter is organized as follows: **Section 4.1** explains MLC operation and characteristics in PCM. **Section 4.2** describes the various factors limiting reliable MLC operation. **Section 4.3** presents several novel technologies that enable reliable MLC storage, including the novel drift-resilient readout metrics. **Section 4.4** explains the MLC programming and readout architecture implemented in CMOS technology. Finally, in **Section 4.5.2.4**, the chip characterization results of one of the readout extraction scheme is presented.

4.1 Multilevel Cell storage

Recently, it has been demonstrated that the memory capacity/density of PCM can be increased further by storing more than one bit of information per memory device [Bedeschi et al., 2009], typically known as multi-level cell (MLC) storage. It is one of the prominent features of phase-change memory that renders PCM even more attractive among the emerging non-volatile memory technologies.

The MLC strategy is to make use of the intrinsic capability of the memory device to store analog data to encode more than 1 bit of information per device. In the case of PCM technology, the large resistivity contrast exhibited by these materials between the crystalline and the amorphous state (typically 3–4 orders of magnitude) and hence the high ON/OFF ratio achieved favor the possibility of storing information in the intermediate resistance states. Figure 4.1 illustrates the typical current vs. voltage ($I-V$) characteristics of these devices measured at room temperature. Apart from the extreme RESET (high-resistance) and SET (low-resistance) states, additional information can be stored in any of the intermediate resistance levels.

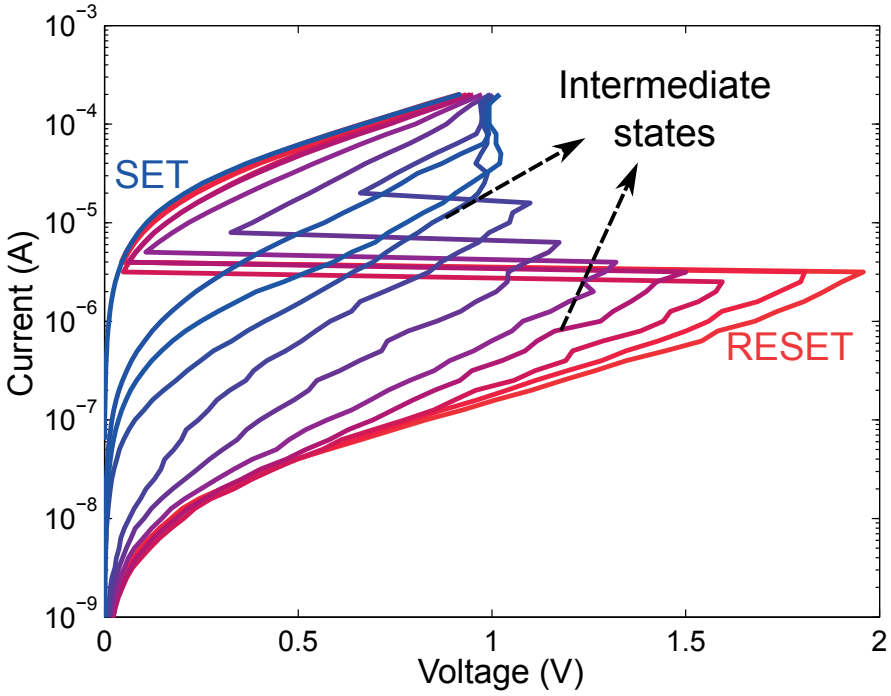


Figure 4.1: Experimentally measured current vs. voltage ($I-V$) characteristics of typical PCM devices at room temperature. Not only the extreme SET and RESET states,but also the intermediate resistance levels can be used to store information.

The intermediate resistance states can be attained by properly modulating the electrical pulses used to program the PCM. Two such pulses are shown in Figure 4.2. By carefully controlling these pulses, one can fine-tune the target analog resistive level of the device, thus paving the

way for MLC operation.

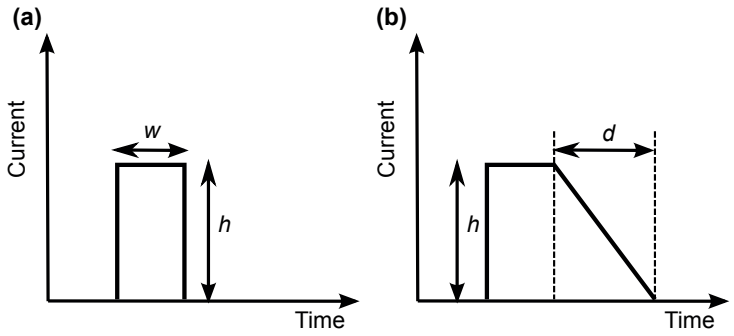


Figure 4.2: Electrical pulses for programming the intermediate resistive levels of the device. (a) Rectangular current pulses of different heights h and widths w, and (b) variable slope pulses, with different durations d of the trailing edge of a trapezoidal pulse.

The success of MLC storage is highly influenced by the resistance distribution attainable over a large ensemble of PCM devices. Figure 4.3 illustrates a conceptual probability density function of four resistive levels (2 bits per cell MLC storage). Each of these levels is associated with a resistance distribution. Ideally, these distributions should be delta functions to simplify the data retrieval process. However, if these distributions overlap, there is a non-zero probability of error when retrieving the stored data, resulting in unreliable data storage.

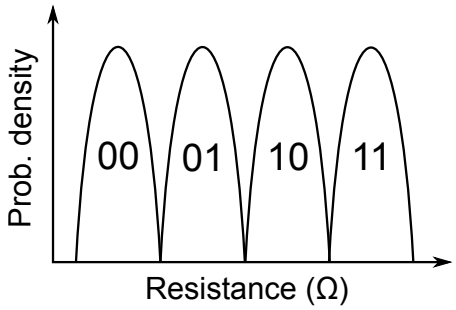


Figure 4.3: Conceptual probability density function of four resistance levels required for 2 bits per cell MLC storage. The resistance distribution should be non-overlapping for reliable readout of the stored resistive levels.

4.2 Factors limiting MLC operation in PCM

Despite the wide resistance range (between the SET and RESET states) favoring the MLC operation in PCM, there are various factors that strictly limit the number of intermediate levels that can be reliably stored in a single cell. The most critical factors are:

- Resistance drift
- Device variability across the PCM array

- Intrinsic PCM noise
- The electrical circuit noise arising from the programming/readout circuitry

Achieving MLC storage in PCM is quite challenging as it typically comes at the expense of reliability[1]. In particular, given that several physical and technological issues hinder MLC functionality in PCM, the conventional SLC write/read methods are not efficient enough to reliably store and retrieve the data. The predominant reliability concerns are the phenomenon of resistance drift and device variability [Pozidis et al., 2012]. In the remainder of this Section, I briefly elaborate on some of the above-mentioned reliability challenges.

4.2.1 Resistance Drift

Resistance drift is a physical phenomenon arising from the atomic re-arrangement within the amorphous phase of phase-change materials [Ielmini et al., 2007; Boniardi et al., 2010]. It manifests itself as random fluctuations around a monotonic increase in the resistivity of the material as a function of the time elapsed after programming. Drift causes a broadening of the resistance distributions of programmed levels over time, and thus limits the number of resistance levels that can be stored and reliably read back.

It has been experimentally verified that the high-resistive levels tend to exhibit a steady increase in resistance over time, and the evolution of resistance drift can be modeled using the power-law model [Pirovano et al., 2004]. It is given by

$$R(t) = R(t_0)\left(\frac{t}{t_0}\right)^{\nu}, \tag{4.1}$$

where $R(t)$ denotes the device resistance at time t, t_0 is some arbitrary time after programming, $t_0<t$, and ν is the drift exponent that is a characteristic of that particular resistive level [Papandreou et al., 2011; Ielmini et al., 2009]. Using (4.1), it is possible to capture the evolution of an ensemble of programmed cells over time and extract the average drift exponent (ν) for each resistance level.

The evolution of the programmed resistance levels due to drift has been investigated by monitoring the programmed array at different time instances. Figure 4.4 shows the evolution of programmed resistance states over time and the average drift exponent (ν) for each resistance level. One observation from the figure is that drift affects all stored levels, starting at very short time scales, i.e., microseconds after programming. High-resistive levels exhibit higher values of ν, confirming the fact that resistance drift is a characteristic of the amorphous phase. High-resistance states arise from a larger proportion of amorphous volume within the PCM device and are thus more prone to drift than lower-resistance states. Apart from resistance drift, other factors, like array variability and circuit noise, also influence the resistance distribution.

[1] The same is true for MLC storage in Flash memory; however, a very large ecosystem and years of development have established MLC Flash as a viable technology today.

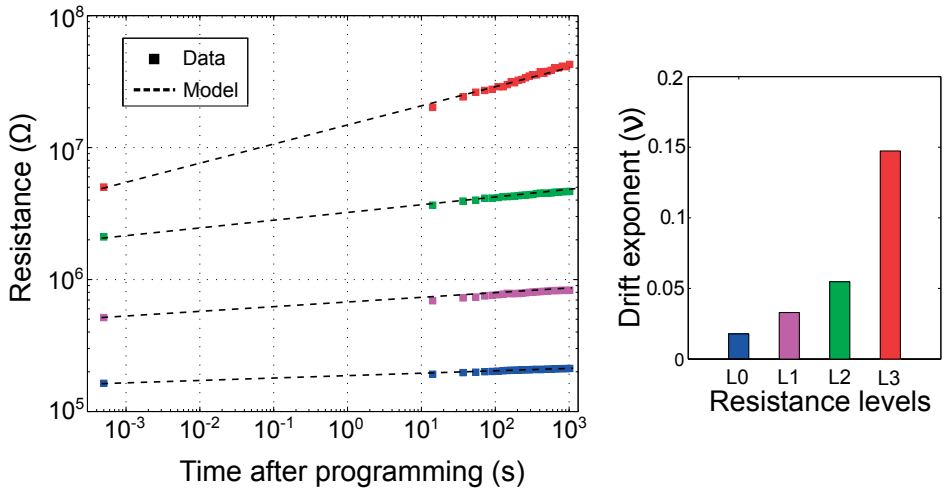


Figure 4.4: (a) Evolution of four programmed resistance levels over time at room temperature (300 K). The dashed lines depict the power-law model fit from (4.1) and the square markers correspond to experimental measurements from an actual device. (b) Drift exponent values of each of these levels. High-resistance states have a higher drift coefficient.

Therefore in the case of MLC operation, drift reduces the separation between adjacent signal levels as the stochastic fluctuations of the resistance cause a broadening of the level distributions over time, thereby compromising reliable data readout. It is detrimental in MLC storage because random fluctuations of the programmed resistance of closely-spaced levels may lead to overlap between them, which leads to decoding errors in data retrieval and eventual data loss.

4.2.1.1 Temperature dependence

Apart from the temporal dependence of resistance drift, temperature fluctuations also greatly influence the drift behavior. Researchers have recently explored and modeled the time-temperature dependence of the resistance drift using a structure relaxation model [Wimmer et al., 2014; Sebastian et al., 2015]. The temperature-dependent conductivity of phase-change materials tend to increase with increasing temperature. Higher temperatures are shown to accelerate the drift phenomenon owing to the increase in the activation energy. Therefore, with closely packed intermediate-resistive levels, temperature fluctuations further compromises reliable MLC storage. Figure 4.5 illustrates the temperature dependence of drift for a wide temperature range.

4.2.2 Variability

Another limiting factor for MLC operation is the variability (or non-uniformity) associated with an ensemble array of PCM devices. Typically, diodes or transistors are used as access

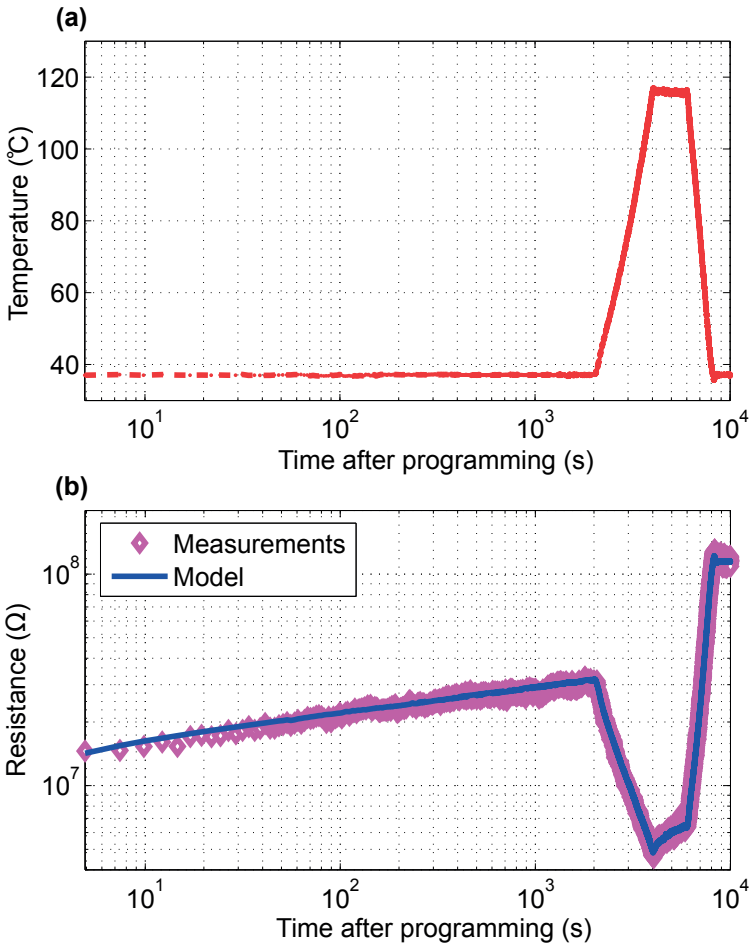


Figure 4.5: (a) The temperature variations across the PCM device ranging from 40°C to 120°C. (b) The corresponding evolution in the resistance, both measured and simulated data from the structural relaxation model proposed by [Sebastian et al., 2015].

devices, and the combination of the PCM storage element and the access device constitutes the memory device. Figure 4.6 depicts a typical array of PCM storage element with transistors as access devices. Each memory device can be addressed by the combination of bit-line (BL) and word-line (WL). Variations in the process and material parameters during device fabrication lead to variations in the critical dimension (CD) of the devices across the array. As a consequence, the same electrical stimulus might give rise to different thermal profiles within the active storage element.

Therefore, the same programming pulse applied across an ensemble of devices will result in different resistances. This leads to broadening of the programmed resistance levels. This is illustrated in Fig. 4.7, which shows the static, "single-shot" programming (resistance-current or R-I) curves for a sample of 30 cells. To program all cells at an intermediate-resistance level (e.g. 10^6 Ω), current pulses in the range between I_1 and I_2 are required.

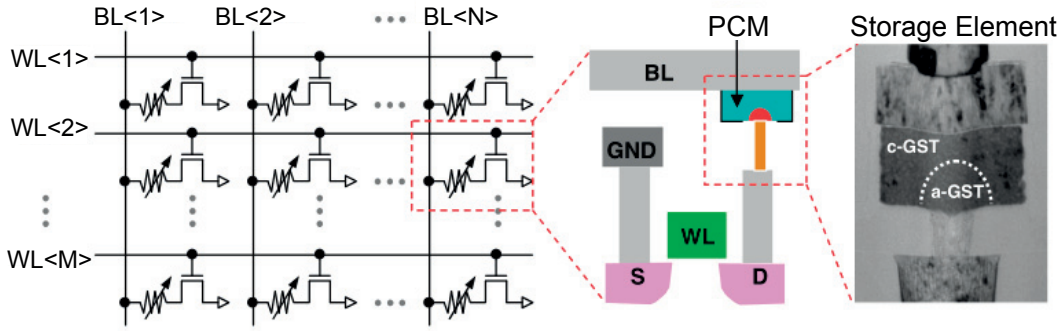


Figure 4.6: Schematic of the PCM cell array containing the storage element and the access device. Each storage element can be accessed through the corresponding word-line and bit-line. The access device is a thin oxide NFET with W/L = 240 nm/120 nm.

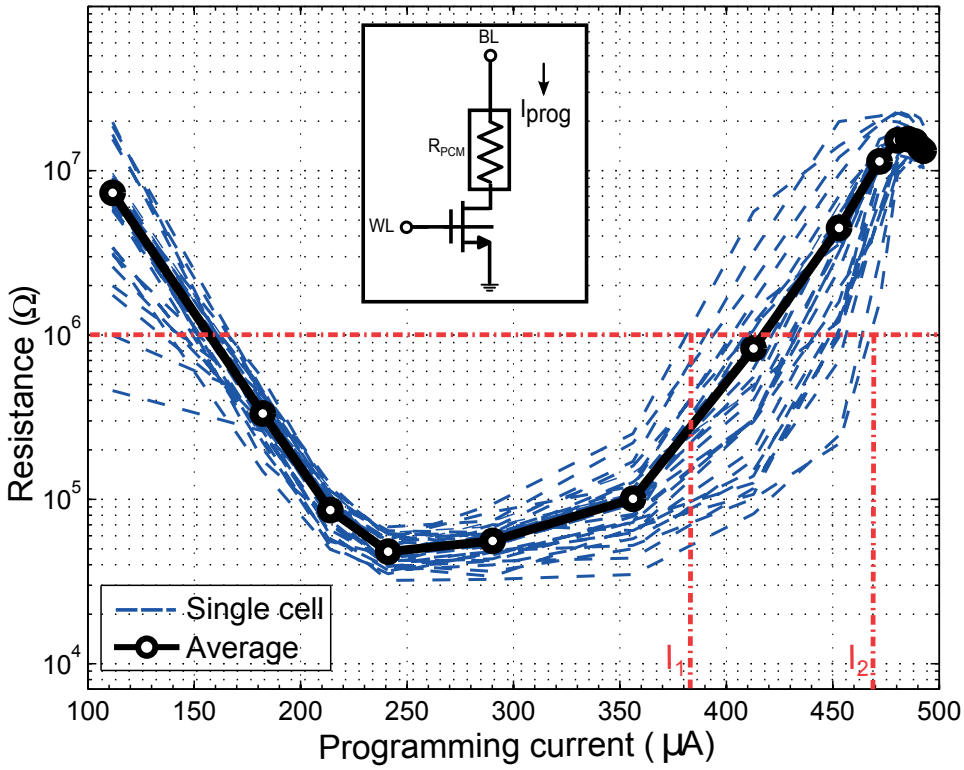


Figure 4.7: Superposition of static programming curves at room temperature for a sample of 30 cells. Given the variability, a range of currents (I_1 to I_2) is required to program all cells to an intermediate resistance of 1 MΩ.

Variability can either be inter-cell variability, arising during the fabrication process, or intra-cell variability due to repeated programming cycles. With repeated programming of a given cell, the temperature distribution attained within the device during one programming cycle may differ from that of the previous cycle. This leads to variations in the resistance levels

attained, known as intra-cell variability. However, compared with inter-cell variability effects, these variations are typically small and do not cause serious harm to MLC storage.

4.2.3 Intrinsic PCM noise

In addition to drift, there are also short-time fluctuations of the current flowing through the programmed PCM devices biased at a constant voltage. The current fluctuations exhibit abrupt changes, reminiscent of random telegraph noise and low-frequency noise. Although the origin of this excess noise is still unclear, its behavior can be modeled by assuming that each charge-carrier trap oscillates between two metastable states [Fugazza et al., 2010]. Close et al. measured the noise current as a function of the average read current from approximately 200 k readout samples. They then compared it with the noise obtained when reading out integrated poly-silicon calibration resistors under the same conditions. It was verified that the intrinsic PCM noise dominates the noise of the readout circuitry and limits the bit-error rate (BER) [Close et al., 2013].

4.3 Enabling Multiple-Level Cells in PCM

It can be inferred that resistance drift and array variability directly limit the storage capacity of PCM. A number of novel techniques have been explored for coping with the problems of resistance drift and variability. Researchers have reported various MLC enabling technologies that offer better resilience and immunity to drift and variability. Some of these techniques include changing the write target resistances to take into account the expected broadening of the resistance distributions due to drift [Kang et al., 2008]. Other schemes include compensation techniques during readout, where by pulses are used upon readout to return the device to its initial as-written resistance [Kostylev et al., 2005]. Although such techniques are shown to have better drift performance, they do not really tackle the inherent problem of drift. Some of the more active research areas focus on developing a kind of non-resistance cell-state-based metric, which has been shown to have better immunity to drift [Sebastian et al., 2011; Papandreou et al., 2011]. These topics are the focus of current research activities in the MLC PCM research community.

In this Section, I introduce three different technological approaches that have been developed to successfully implement and demonstrate reliable MLC storage and data retention on PCM cell arrays.

4.3.1 Materials and fabrication process

The type of the phase-change material and its deposition process, the device design and geometry, and the process of device fabrication are the critical parameters that define the device characteristics and performance. Some of the recent approaches are based on engineering

the properties of the phase-change material for MLC realization. Figure 4.7 depicts the typical transition between the SET and the RESET states in terms of the applied programming current. In the case of MLC operation, it is desirable to have a gradual change in the resistance, rather than a steep one, between the SET and the RESET as it makes it easier for the programming circuitry to program into the target intermediate resistance levels.

Researchers have shown that doping the GST material with nitrogen impacts the resistance vs. temperature ($R-T$) curve, resulting in a more gradual transition between the crystalline and amorphous states with respect to temperature. This increases the probability of attaining the intermediate resistance levels as target resistance level and thus enables MLC storage [Nirschl et al., 2007; Liu et al., 2004].

Another approach to enable MLC operation is through the device design, by using a stack of different phase-change materials (with different physical properties) rather than a single material as the storage element. Rao et al. demonstrated that by using a bi-layer stack of chalcogenide material (namely, GST and Sb_2Te_3), it is possible to even have a fairly stable intermediate state rather than a gradual transition (attainable through doping) between SET and RESET [Rao et al., 2007]. Zhang proposed that by having tri-layer stack of phase-change material ($Si_{16.4}Sb_{32.5}Te_{51.1}$) separated by TiN thin films acting as local heaters, different current flows through each heater result in different temperature regimes, favoring the $R-I$ curves to result in stable intermediate-resistance levels [Zhang et al., 2007]. In the above-mentioned demonstrations, the device structures relied on different thermal profile attainable within the device, owing to the differences in the properties of the stacked chalcogenide layers. This in turn paves the way for stable intermediate-resistance states.

Apart from methodologies that directly focus on enabling MLC operation, more recently, Kim et al. proposed a novel device-design approach to tackle the phenomenon of resistance drift. They demonstrated a six-fold decrease in the drift coefficient values of high-resistance states by using a metallic surfactant (metal nitride) layer as an alternative read current path to the amorphous region [Kim et al., 2013].

4.3.2 Iterative programming scheme for MLC

Although there are many process and fabrication-based demonstrations for the precise engineering of the device-design and the material properties to improve the capability of phase-change materials for MLC storage, even conventional PCM devices, like the mushroom-type device topology, can be used to program the intermediate-resistance levels using iterative programming algorithms. These procedures typically consist of applying repeated programming pulses of varying amplitude and verifying that a specific precision criterion is met, similar to those used in current Flash memories [Bez et al., 2003].

For MLC operation, the size of the amorphous region can be increased progressively by application of electrical pulses with increasing amplitude to attain any desired resistance level

between the SET and RESET states. The programming $R-I$ curve of the device defines the current required to program the device at the target resistance level. However, device variability results in wide range of resistance levels, resulting in a broader resistance distribution than when single-box type pulses are used for programming (described in the Section 4.2.2). A common approach to confine the resistance distribution of the programmed levels is to use iterative programming algorithms [Nirschl et al., 2007].

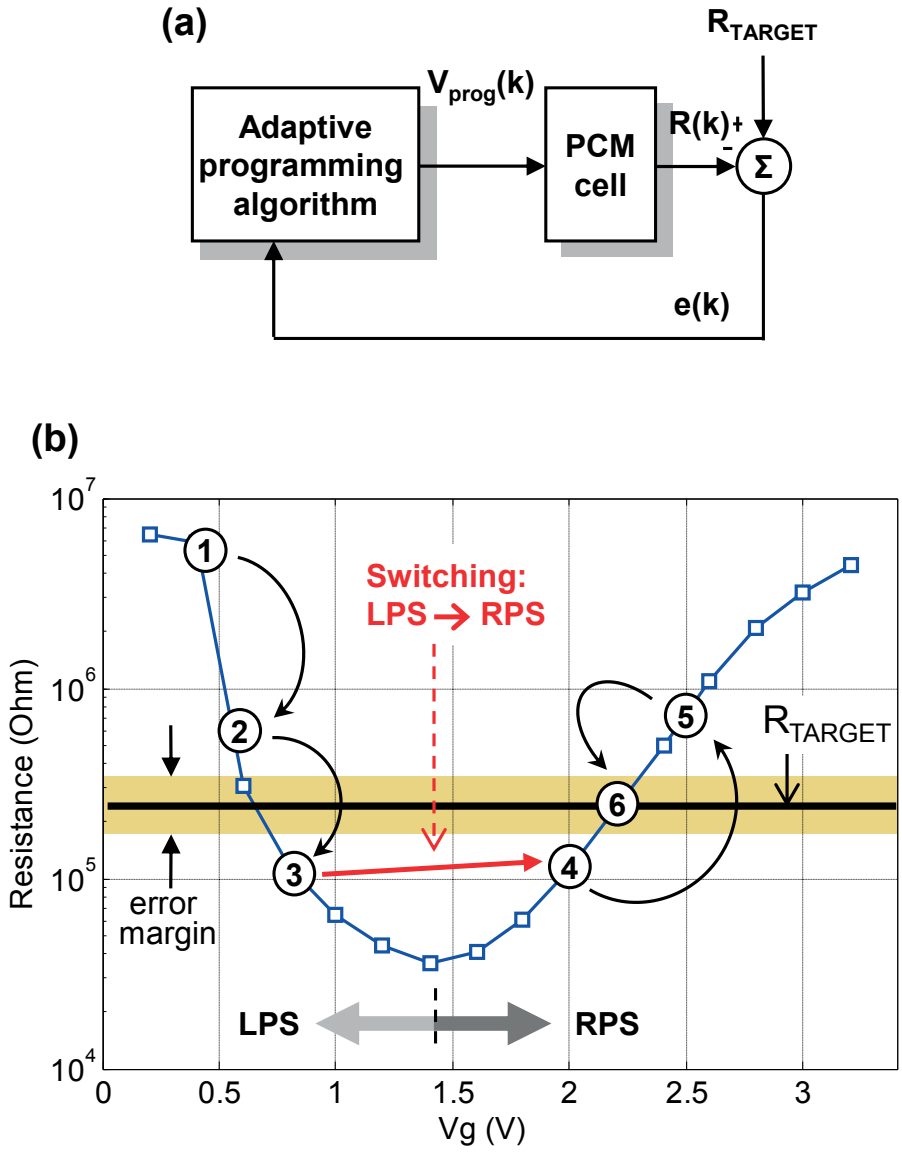


Figure 4.8: (a) Simplified block diagram of the iterative programming algorithm; Vg is the gate voltage of the access device. (b) The propagation of the iterative algorithm over the typical $R-V$ programming curve. R_{TARGET} is the target resistance and the iterative loop ends once the resistance is within the error margin or the maximum number of iterations has been attained. LPS and RPS represent the left and the right programming slope.

Bedeschi et al. proposed an iterative programming scheme where the programming begins

with a RESET pulse followed by a series of SET pulses with increasing trailing edges followed by verify pulses [Bedeschi et al., 2009]. The algorithm ends once the target resistance level has been achieved. However, by application of repeated partial-SET pulses, we can only decrease the resistance, and the cell should be melt-quenched (RESET) again to increase the resistance. Given that the RESET pulse is the power-hungry process, this approach is economical in terms of power consumption. This kind of programming approach is typically known as programming on the left-slope of the $R-I$ curve (LPS).

Alternatively, programming can also begin with SET pulse followed by partial-RESET pulses of varying amplitude until the required resistance level has been attained [Karpov and Kostylev, 2006]. This type of iterative programming scheme is known as the programming on the right-slope of the $R-I$ curve (RPS). Although the power requirements of such an approach is high owing to the power-hungry RESET pulses during each iteration, the schemes has the advantage of bi-directional flexibility (either the increase or decrease) in the resistance by proper tuning of the amplitude.

Recently, Papandreou et al. experimentally demonstrated a novel iterative programming scheme that combines the advantages of both left-slope and right-slope programming (Fig. 4.8). Programming starts with the partial-SET pulses and terminates once the target resistance level has been reached. However, if the resistance falls below the target resistance, partial-RESET pulses are used to reach the required target level. The switching between the left and the right slope is facilitated by the U-shaped programming curve and realized by application of suitable electrical pulses. The method combines the low energy dissipation of the partial-SET regime (voltage/current pulses of low amplitude) with the bi-directional flexibility of the partial-RESET regime, where resistance can be either increased or decreased with renewed melting [Papandreou et al., 2011].

Given the number of iterations it takes to program a particular resistance level, each iteration of these programming schemes should be efficient so as to maintain the required programming bandwidth. Given the repetitive program and erase cycle in a single programming operation, the devices should have very high endurance such that the iterative nature of the programming scheme has no adverse impact on device operation. Finally, irrespective of the repeated programming attempts, care must be taken that the overall energy requirements are still low as necessitated by certain low-power applications.

4.3.3 Readout metrics for MLC PCM

With the advent of various memory technologies, several readout techniques are currently being used to extract the stored data from semiconductor memory devices. The basic idea of any readout scheme is to precisely measure/sense some particular characteristic of the cell, which depends on the cell's current state (SET or RESET) and present it to the outside world in a recognized format (usually in digital form). These techniques should be accurate and fast enough as they directly impact the read bandwidth of the memory device. The measurements

should be made such that the content of the device should not be perturbed. Typically, the readout circuitry consists of control logic, address decoders and sense amplifiers.

Conventional memories consist of an array of memory devices (memory storage elements and an access circuitry) as shown in Fig. 4.6. Each memory device can be accessed by the word-line (WL), and the data transfer (for both read and write operations) takes place thorough the bit-line (BL). Usually, diodes or transistors are used as access devices to ensure that the read and write currents on bit-line are interacting with one and only one memory device at any given time. Apart from these access devices, the access circuitry consists of address decoders to access a particular memory cell and sense amplifiers to extract the stored content.

In the case of resistive memory technologies like PCM, the access circuitries should be able to accurately measure the current resistance state of the cell. Prior-art methods of determining the resistance of the memory cell and thus the data stored involve comparing a voltage or current response of the device to a reference. However given the drift and variability in MLC operation, it becomes increasingly difficult to accurately sense the resistive state by comparing the device response to a reference (voltage or current), resulting in possible bit-errors during data retrieval.

Also, care should be taken that the applied power while reading is much lower than the threshold switching power (Section 2.3.1.2). If the electrical input (current or voltage) applied to the device exceeds the threshold switching voltage or current, the resistance of the device changes drastically and hence the cell content will be disturbed. Note that the threshold switching voltage is characteristic of the resistance levels. i.e., higher/lower resistive levels tend to have higher/lower threshold switching voltages. It is clear that the threshold switching voltage is the fundamental parameter for device operation because it defines the cell readout margins.

4.3.3.1 Resistance (*R*)-Metric

The conventional approach to measure the resistance of a PCM device is by sensing the current flowing through it when a low-bias voltage (V_R) is applied across the device, also known as the "low-field electrical resistance", or simply, the resistance metric (R-metric). Given that the resistance of the device is unknown during read, the applied voltage, V_R, should be such that it is lower than the minimum threshold switching voltage of all resistance states.

Although the resistance metric is simple and fast to extract, it suffers from some serious limitations [Sebastian et al., 2011]: (a) It is highly sensitive to temporal resistance drift (described in the Section 4.2.1), especially the high-resistive amorphous states. (b) Because of device-geometry effects (especially in the case of mushroom type devices), it tends to saturate above a certain value of resistance (u_A, amorphous thickness), masking the fact that the amorphous volume increases with increasing input power, thereby reducing the available signal range. (c) It suffers from low signal-to-noise ratio (SNR) at high-resistance levels, as the cell current is

measured at a small voltage bias, because the V_R is limited by the threshold switching voltage [Close et al., 2010].

These limitations of the conventional R-metric have motivated the development of novel non-resistance cell-state-based metrics that are more amenable to MLC operation in PCM. These metrics are described in the following sections.

4.3.3.2 Drift-Tolerant M-Metric

To understand the inherent problems with the R-metric, Sebastian et al. analytically computed the $I-V$ characteristics of mushroom type devices, starting from the modified Poole–Frenkel transport model as proposed in [Ielmini and Zhang, 2007]. It was found that the computed $I-V$ characteristic can be well approximated by the $I-V$ behavior corresponding to a PCM device with cylindrical geometry described in [Sebastian et al., 2011]. Figure 4.9 depicts the cylindrical approximation of the amorphous dome.

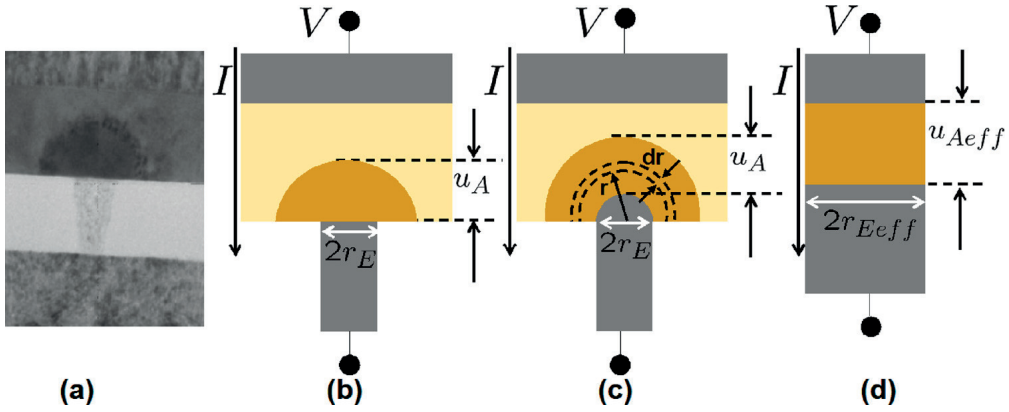


Figure 4.9: (a) Cross-sectional TEM image of a PCM mushroom cell. (b) Schematic of a PCM mushroom device topology. (c) Semi-hemispherical approximation of the amorphous-GST geometry. (d) Cylindrical approximation of the a-GST geometry.

According to the Ielmini model, the current flowing through the GST can be approximated by

$$I = 2qAN_T \frac{\Delta z}{\tau_0} e^{\frac{-(E_c-E_f)}{kT}} \sinh\left[\frac{q\Delta zE}{kT}\right] \tag{4.2}$$

where q is the electronic charge (C), A is the cross-sectional area (m^2), E is the uniform electric field (V/m), τ_0 is the characteristic attempt-to-escape time for a trapped electron, Δz is the mean inter-trap distance (nm), N_T is the trap density, k is the Boltzmann constant, T is the temperature (K), E_f is the Fermi level (eV), and E_c is the mobility edge for the conduction band (eV). To simplify the computation for the particular case of mushroom type devices, Sebastian et al. assumed a hemispherical bottom electrode and that the top electrode is infinitely wide compared with the bottom electrode. The current flowing through the cylindrical geometry

can be then approximated by

$$I = 2q\frac{\pi r_{E_{\text{eff}}}^2}{\tau_0} N_T \Delta z e^{\frac{-(E_c-E_f)}{kT}} \sinh\left[\frac{q\Delta z V}{2kTu_{A_{Eff}}}\right], \tag{4.3}$$

where $r_{E_{\text{eff}}}$ is the effective bottom electrode radius and $u_{A_{\text{eff}}}$ is an effective amorphous thickness. From (4.3), the low-field electrical resistance (for small applied voltage) is given by

$$R = \frac{kT\tau_0 u_{A_{\text{eff}}} e^{\frac{(E_c-E_f)}{kT}}}{q^2\pi r_{E_{\text{eff}}}^2 N_T \Delta z^2}; \tag{4.4}$$

It was show that, according to (4.4), for an increase in $u_{A_{\text{eff}}}$, there is also a corresponding increase in $r_{E_{\text{eff}}}$, thereby explaining the reason for the saturation of R-metric at high values of the amorphous thickness. This provides valuable insights as it explains the saturation of the low-field resistance at high programming voltages/currents routinely observed in experiments with mushroom-type devices.

In contrast, the effective amorphous thickness is mostly unchanged with respect to the actual amorphous thickness. It is therefore clear that for the mushroom-type devices, R-metric masks the fact that amorphous volume increases with increasing applied input power. In addition, note also that the resistance is measured as the resulting current when the device is biased at the low read voltage. This results in an additional $1/x$ compression, which degrades the signal-to-noise ratio (SNR) at high resistance levels [Sebastian et al., 2011].

The other critical inference from (4.4) is that the R-metric is a strong function of the activation energy ($E_a = E_c - E_f$). The temporal drift phenomenon commonly observed in these materials is attributed to fluctuations in the activation energy with time, which, according to (4.4), will affect the R-metric exponentially.

Given the limitations of R-metric, there is a need for an alternate non-resistance-based metric that has a strong dependence on the amorphous thickness, yet is less dependent on the device geometry and the activation energy. From the nature of the hyperbolic sine in (4.3), it is evident that the current flowing through the device becomes a strong function of the amorphous thickness in the high-field regime. Exploration of the high-field regime for every programmed state is an excellent approach for constructing a cell-state metric that is strongly dependent on $u_{A_{\text{eff}}}$ and weakly dependent on E_a and $r_{E_{\text{eff}}}$. The proposed metric is referred to as the M-metric.

In the so-called M-metric, the cell state is measured by current biasing and voltage sensing. One approach to extract the M-metric is to increase the read voltage substantially, rather than performing a low-bias read, until a certain pre-defined current (I_R) flows through the device. Care must be taken so that the reference current chosen is much lower than the threshold switching current. The potential across the device when the pre-defined reference current flows through the device will be the M-metric value corresponding to the resistance level.

Using (4.3) and considering the scenario in which the read voltage is varied until the device current reaches the I_R current level, the M-metric, can be expressed by

$$M(I_R) = \frac{2kTu_{A_{\text{eff}}}}{q\Delta z}\sinh^{-1}\left[\frac{I_R\tau_0 e^{\frac{(E_c-E_f)}{kT}}}{2q\pi r_{E_{\text{eff}}}^2 N_T\Delta z}\right]; \tag{4.5}$$

From (4.5), it is evident that the dependence of the M-metric on the activation energy is much weaker, and therefore it should be more robust to drift. The dependence of the M-metric on activation energy can be reduced further by using a differential variant of the M-metric, known as M_{diff} metric. Instead of a fixed target current (I_R), one could detect the difference in voltage required by the device to increase the current from a non-zero current level, I_{R0}, to the target current level, I_R ($I_{R0} < I_R$). This differential variant is likely to be even more tolerant to drift, as can be seen from an approximation of (4.5) given by

$$M(I_R) \approx \frac{2kTu_{A_{\text{eff}}}}{q\Delta z}\log\left[\frac{I_R\tau_0 e^{\frac{(E_c-E_f)}{kT}}}{q\pi r_{E_{\text{eff}}}^2 N_T\Delta z}\right] \tag{4.6}$$

$$M(I_R) \approx \frac{2kTu_{A_{\text{eff}}}}{q\Delta z}\left(\log(I_R) + \log\left[\frac{\tau_0 e^{\frac{(E_c-E_f)}{kT}}}{q\pi r_{E_{\text{eff}}}^2 N_T\Delta z}\right]\right); \tag{4.7}$$

The M_{diff} metric can be given by

$$M_{diff} = M(I_R) - M(I_{R0}) = \frac{2kTu_{A_{\text{eff}}}}{q\Delta z}\log\left(\frac{I_R}{I_{R0}}\right); \tag{4.8}$$

Theoretically, the M_{diff} metric is independent of the activation energy while preserving the linear relationship with the effective amorphous thickness. However, note that I_R and I_{R0} have to be sufficiently high for the approximation in (4.6) to be valid.

4.3.3.3 Enhanced (e)M-Metric

The R-metric suffers from temporal drift and gets saturated for high-resistive levels, whereas the M-metric, despite its wider dynamic range and drift resilience, offers lower contrast in low-resistive states because of its low and constant detection current. In addition, in the practical circuit-level M-metric implementation, the low readout current requires a longer time to achieve full settling of the voltage across the bit-line (owing to the huge parasitic capacitance at the bit-line), imposing a high latency penalty (usually on the order of few hundreds of *ns*).

Therefore, an enhanced version of the M-metric (called eM-metric) is proposed that inherits the advantages of the M-metric and in addition offers improved contrast and higher SNR in low-resistive states, in exchange for a slight decrease in dynamic range. The eM-metric

detection curve measures both low- and high-resistive states at high field, thus effectively combining the benefits of the R- and M-metrics. Figure 4.10 shows the experimentally measured $I-V$ characteristics at room temperature of a typical PCM device programmed at various intermediate resistive states and the detection schemes used to extract the various readout metrics.

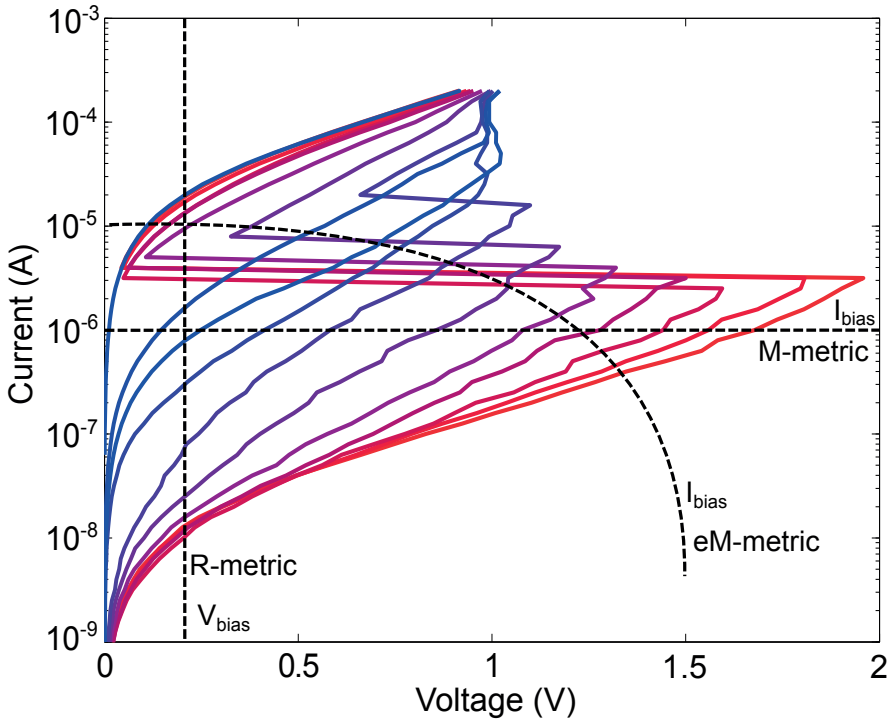


Figure 4.10: Experimentally measured $I-V$ characteristics of the PCM device at room temperature (300 K). The dashed line represents the detection curves used for various readout schemes. V_{bias} is the applied low-bias readout voltage for the R-metric readout. I_{bias} corresponds to the constant and the variable detection current used for the M- and eM-metric respectively.

Practical circuit realizations of the M- and eM-metric readout architectures are detailed in Section 4.4. Figure 4.4 and Fig. 4.11 shows the sensitivity of the R- and eM-metrics, respectively, with respect to drift, showing the metric evolution over time for four different resistance states. Evidently, the eM-metric exhibits superior immunity to drift compared with the R-metric. In particular, the typical drift exponent of the eM-metric (νeM) is an order of magnitude smaller than the typical drift exponent of the R-metric (νR).

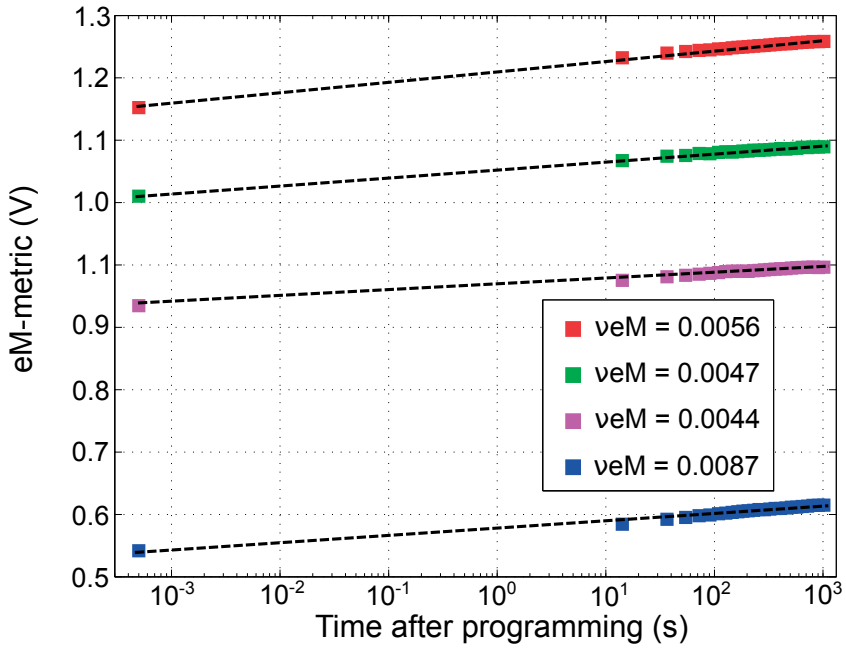


Figure 4.11: Drift evolution of the eM-metric over time of four programmed resistance levels experimentally measured at room temperature (300 K). Dashed lines depict the power-law model fit of (4.1), and the square markers represent the experimental measurements.

4.3.4 Drift-Tolerant Detection and Coding

4.3.4.1 Drift-Tolerant Detection

Traditionally, readout architectures use simple detection circuits or schemes to minimize the area overhead used for peripheral circuitry and thus maximize the storage efficiency. Even in MLC operation, which is more prone to errors than the SLC counterpart, simple detectors such as threshold comparators are typically used, and the thresholds are usually kept fixed throughout the device lifetime.

However, in the case of MLC operation, drift causes the distribution of programmed levels to gradually shift from its initial position after programming. At the same time, other factors, such as temperature fluctuations and endurance cycling, affect the level distribution. Therefore the use of fixed thresholds can lead to errors in data retrieval. In such cases, either the thresholds need to be adaptive, at the cost of increased complexity, or the data must be refreshed, at the cost of higher latency [Pozidis et al., 2013].

Therefore, to detect the stored levels reliably, appropriate level thresholds have to be placed between the distribution of adjacent levels and those thresholds have to be adjusted over time to accommodate the shift of levels due to drift or other factors. Figure 4.12 illustrates the fixed and variable threshold detection schemes for MLC readout. In practice, adjustment of the detection thresholds may be achieved by using “reference” cells, i.e., cells with known stored data, to estimate the changing resistance levels over time [Papandreou et al., 2011].

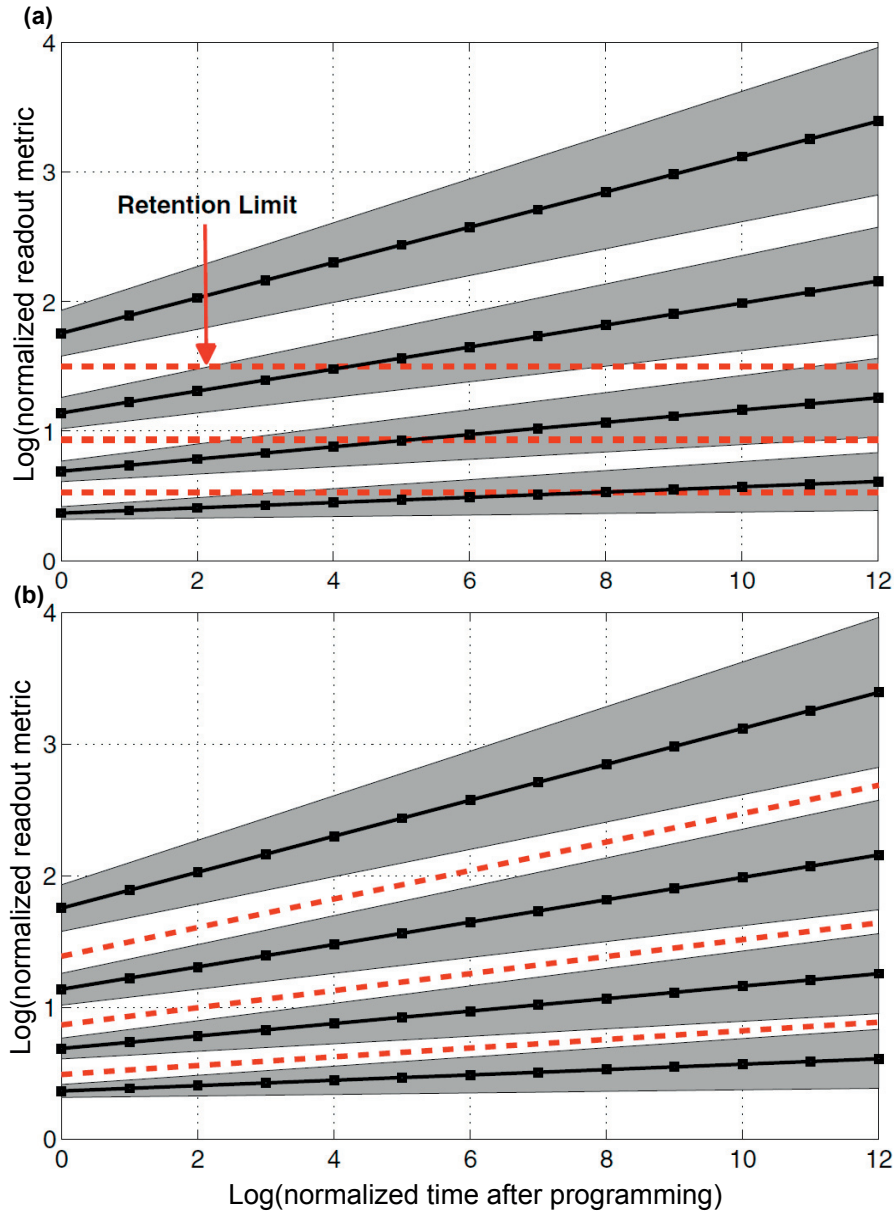


Figure 4.12: Illustration of the concept of using (a) fixed thresholds and (b) variable thresholds to detect the MLC levels. Solid lines with squares indicate trajectories of the means of four level distributions over time due to drift, whereas shaded areas correspond to $\pm 3\sigma$ variations. Dashed red lines indicate detection thresholds.

However, the use of reference cells entails a loss of storage capacity as these cells cannot be used to store user data. To make things worse, the detection performance of the reference-cell method quickly deteriorates when the stored level distributions exhibit high variance. Thus, new approaches are needed to effectively tackle the problem of data detection in MLC PCM.

4.3.4.2 Drift-Tolerant coding

Based on the observation that the direction and rate of the drift of a group of PCM cells are highly correlated, it has been shown that coding techniques relying on the relative resistance of PCM devices in a block are an effective approach to mitigate drift and other fluctuations in the read metric (e.g., temperature fluctuations) [Papandreou et al., 2011]. Coding of the levels stored in MLC PCM is a very efficient approach for increasing drift resilience. The general class of permutation modulation codes was used for this purpose [Pozidis et al., 2015]. As an example, assuming storage of 4 levels in PCM, the permutation modulation (PM) code of length $N = 20$ with level multiplicities $m_0 = 5$, $m_1 = 5$, $m_2 = 5$, $m_3 = 5$ is determined by the initial vector

$$\vec{v} = [0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,2\,2\,2\,2\,2\,3\,3\,3\,3\,3]. \tag{4.9}$$

The code then consists of all length-20 vectors (codewords) that are obtained by permutation of the components of the initial vector $\vec{v}$ [Slepian, 1965]. This code has a rate of about 1.67 bits/symbol; enumerative encoding can be used to encode user bits into codewords [Cover, 1973]. Of course, PM codes of various lengths $N \neq 20$ are possible. Longer lengths enable higher rates. However, their construction is increasingly more complex.

One big advantage of PM codes is the simplicity of their detection, as detection can be achieved by ordering the components of the readout segment of length N and then applying the inverse ordering (permutation) to the corresponding initial vector [Slepian, 1965]. Thus, no level estimation is required, in contrast to the reference-cell method. Clearly, PM codes encode information in the relative order of levels within a codeword, and thus are inherently invariant to drift, which shifts the absolute level values, but typically does not change their relative order.

4.4 Circuit architectures for MLC Programming

The programming or WRITE architectures implementing the iterative programming algorithms described in Section 4.3.2 have been designed and implemented in 64 nm CMOS technology. The typical iterative algorithm circuit consists of current or voltage digital-to-analog converters (v-DAC or i-DAC) for providing the programming current or voltage pulses. They also include a simple read circuitry for the verify part of the iterative loop.

4.4.1 Signed-error iterative programming architecture

Given the wide resistance distributions arising from cell variability, single-pulse programming methods are not suited for MLC storage. The most common solution is to employ iterative programming schemes, where a sequence of program and verify pulses is used in a closed-loop fashion to program the cell into a particular resistance level [Papandreou et al., 2011].

The performance of the iterative algorithm can be improved by programming in the current mode rather than the voltage mode. Figure 4.13 illustrates the proposed iterative programming scheme, in which a current digital-to-analog converter (i-DAC) is used to program the device in the desired intermediate resistance state. To begin the iterative loop, the current readout resistance state of the device is readout and the measured voltage across the device (V_r) is compared with the target voltage level (V_trg). Based on the error voltage ($V_\mathrm{err} = V_\mathrm{trg} - V_\mathrm{r}$), the programming current is either increased or decreased until the error falls within the desired error margin or the maximum number of iterations has been reached. The programming is done on the right slope of the programming curve (RPS), and hence the resistance can be both increased or decreased by application of suitable current pulses with increasing or decreasing amplitude, respectively.

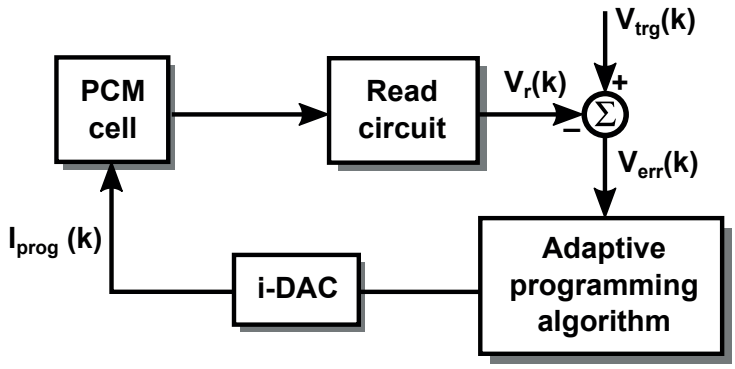


Figure 4.13: Iterative programming algorithm used in the current-mode to program the intermediate resistance states. I_prog is the programming current applied to the cell during each iteration. V_r is the readout voltage, V_trg is the voltage corresponding to the target intermediate resistance, and V_err is the error in voltage. k is the iteration number.

Figure 4.14 depicts the circuit-level schematic of the iterative algorithm. The iterative loop consists of both the program and the verity part. A low-voltage read bias is applied to the cell, and the current flowing through the device is sensed and compared with the target current. The difference current (Δi) is then converted into an equivalent voltage using an integrator. Next, the error voltage is compared with the two voltages, V_P and V_N. Based on the two-bit comparator output, the digital controller modifies the 6-bit digital code for the i-DAC input. The error voltage can also be accessed externally using the test buffer for debug purposes.

A current-steering DAC is used to force a current into the BL for current-mode programming. The programming current, applied to the PCM cell in the subsequent iteration using the i-DAC, is adapted according to the sign or the value of the error between the target level and the read value of the cell state. The programming sequence ends when the error between the target level and the programmed state is smaller than a desired error margin or when the maximum number of iterations has been reached.

If the maximum number of iterations is reached without convergence in the final iteration, a fail bit is set and propagated to a digital controller, which stops the iterative programming and

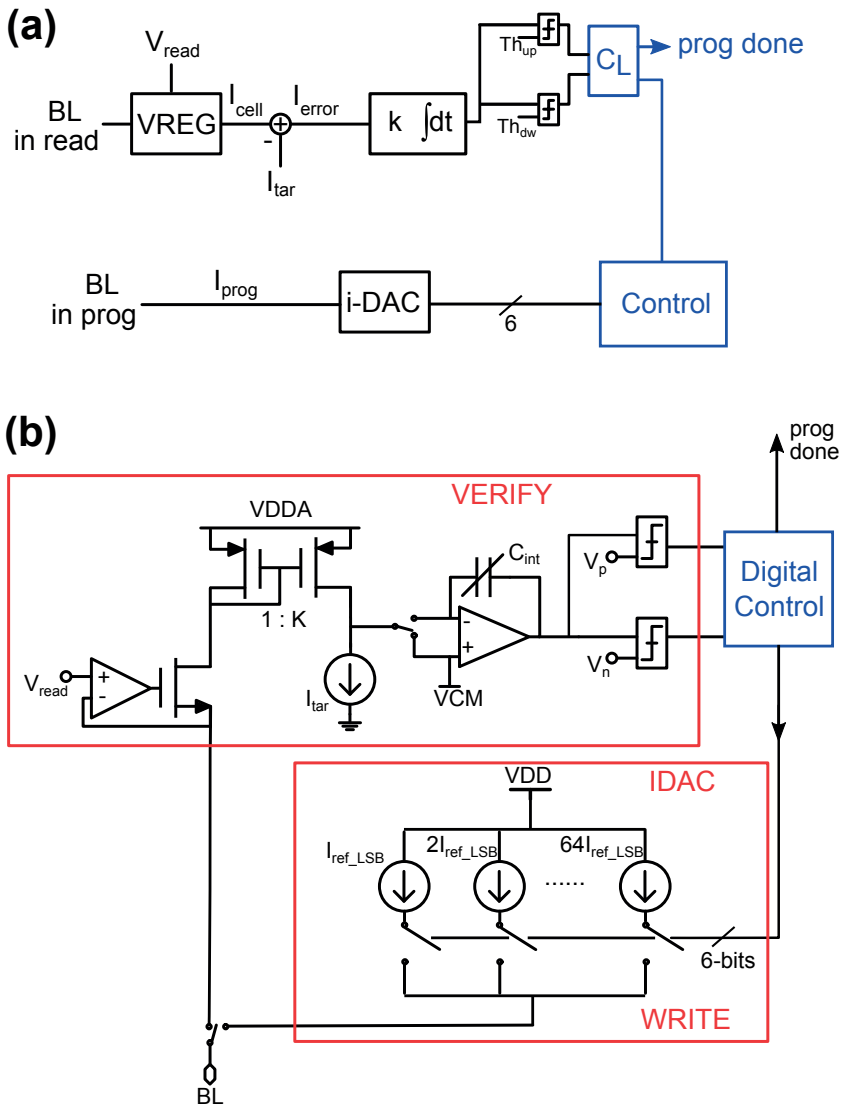


Figure 4.14: (a) Schematic block diagram of the signed-error iterative programming algorithm. (b) Circuit-level schematic for the implementation of the programming scheme. i-DAC forms the core of the programming part. The blue boxes represents the digital domain.

generates a fail signal indicating a failed programming attempt. Similar to the voltage error, the output of the i-DAC can also be accessed externally for debug and monitoring purposes. The key programmable parameters, such as the maximum number of iterations, the error margin and the resolution of the programming current, can all be set from the digital controller.

Although the fabricated prototype chip was not available for characterization at the time of this thesis write-up, the proposed iterative algorithm was emulated on the FPGA-based characterization platform, and the results are presented in the next chapter.

4.5 Circuit architecture for MLC Readout

The circuit architecture for the extraction of any readout metric should have a minimum total readout latency, as the latency directly impacts the read bandwidth of PCM. This can be achieved by the use of fast analog-to-digital converters (ADCs) for the conversion of voltage or current into digital form. In addition, the voltage across the BL during the entire readout process should always be within the allowable low-bias readout limits, as otherwise the cell content could be disturbed owing to the threshold switching.

4.5.1 *R*-metric Extraction

From the circuit perspective, extraction of the resistance metric is simple and straight forward and can easily be implemented by direct voltage-biasing the cell and sensing the current flowing through it. A typical circuit implementation for *R*-metric extraction is illustrated in Fig. 4.15. A voltage regulator can be used to apply the low-voltage bias, and the sensed current can be digitized using a fast ADC. The ADC is the critical component of the readout scheme, and its performance directly influences the read bandwidth. The readout latency is typically on the order of 100~200 ns.

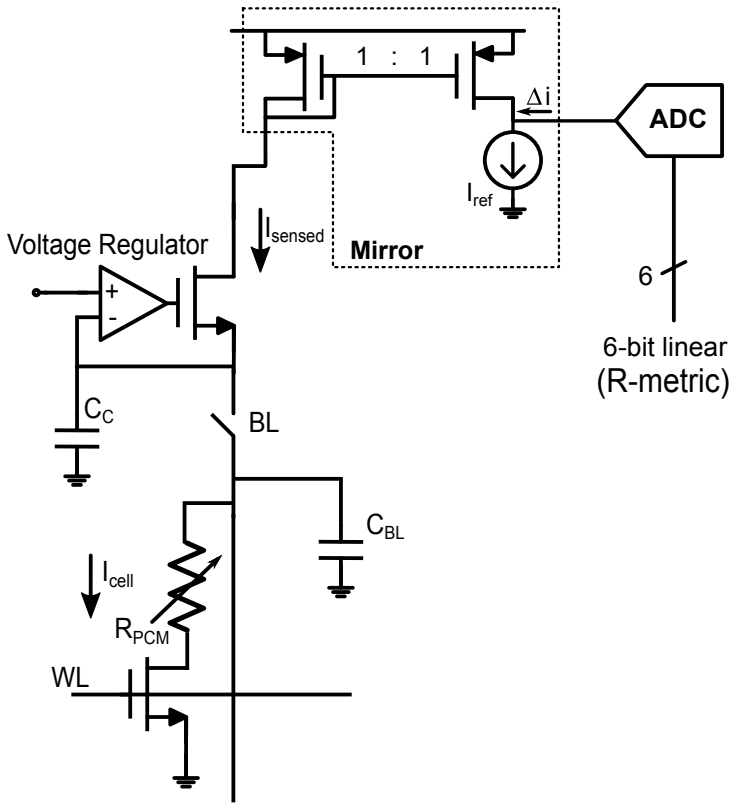


Figure 4.15: Typical MLC readout circuit implementation for extraction of the *R*-metric. Voltage regulator and the ADC are the critical components that define the readout latency of this architecture.

4.5.1.1 ADC readout architecture

In a typical MLC readout scheme for R-metric extraction, fast ADC is used to digitize the current flowing through the device for an applied low-voltage read bias. A voltage regulator is used to bias the cell and can provide voltages ranging from 150 mV – 500 mV. The sensed current, which is a characteristic of the current resistive level of the cell, is then mirrored and subtracted from a mid-scale reference current (I_{ref}). The difference current (Δi) is converted into a voltage through an integrator before getting digitized by the ADC. The core of the ADC is based on the 1.5b/stage cyclic/algorithmic architecture implemented as a switched-capacitor circuit.

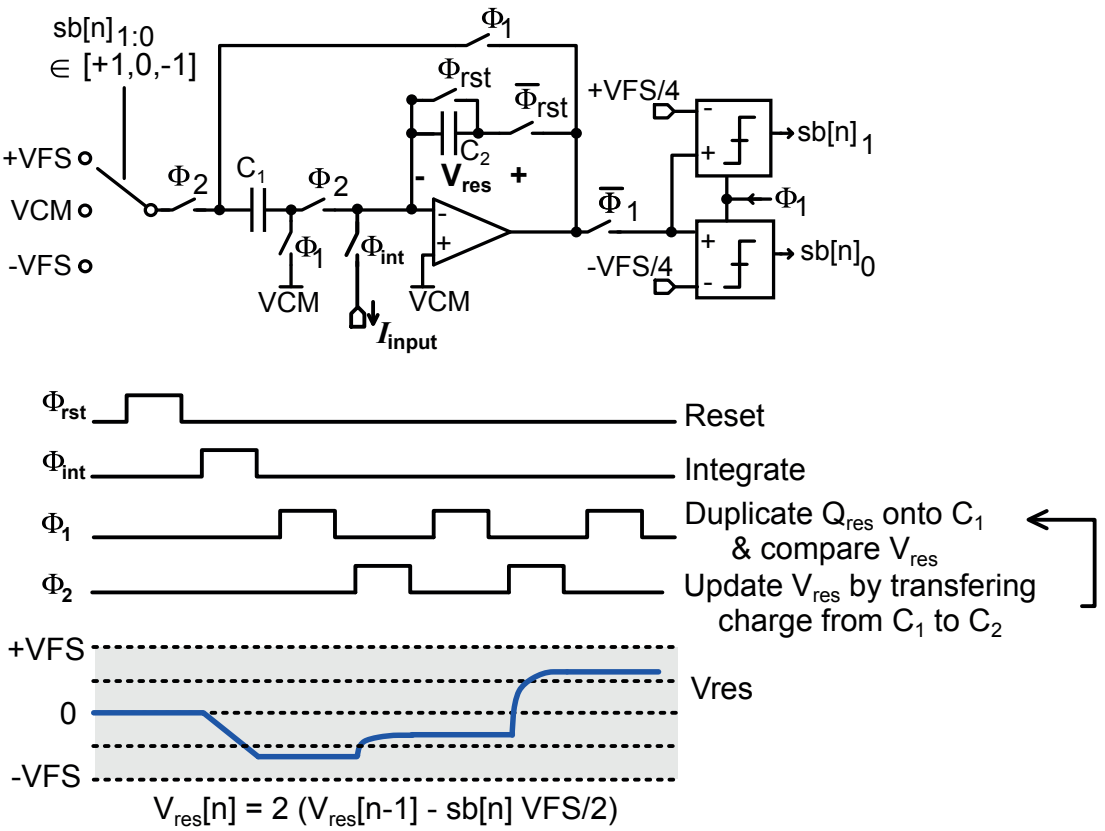


Figure 4.16: Schematic of the cyclic ADC and the timing diagram representing the ADC operation following the two non-overlapping clocks, Φ_1 and Φ_2.

Figure 4.16 depicts the schematic of the cyclic ADC and explains its operation. After the integration, the ADC repeats a two-step cycle clocked by the two non-overlapping clocks Φ_1 and Φ_2 at 100 MHz derived from the 200 MHz master clock. During each iteration, the residue voltage at the output of the integrator is first quantized in three bins (centered at -VFS/2, 0, and -VFS/2) by two comparators in phase Φ_1. This generates a single signed bit $sb[n] \in \{-1, 0, +1\}$. While the comparison is taking place, the residue voltage across capacitor C_2 is duplicated and stored across the matched capacitor C_1. In the second phase, Φ_2, the quantized approximation

of the residue $sb[n]\times$ VFS/2 is subtracted, and the resulting quantization error voltage, scaled by 2, is enforced across capacitor C_2 by charge transfer:

$$V_{\text{res}}[n+1] = 2(V_{\text{res}}[n] - sb[n]\cdot V_{FS}/2) \tag{4.10}$$

$$sb[n] = \begin{cases} +1, & \text{if} \quad V_{\text{res}}[n] > +V_{FS}/4 \\ -1, & \text{if} \quad V_{\text{res}}[n] < -V_{FS}/4 \\ 0, & \text{otherwise} \end{cases}$$

The cycle is then repeated to generate the next signed bits until all bits have been collected. The cyclic ADC keeps outputting bits as long as it is enabled, possibly even beyond the target converter resolution of six bits. By adjusting the reference current and the integration time, many quantization ranges can be obtained. Apart from its low power consumption and area, the cyclic serial operation also allows the conversion to be terminated as soon as the converted value can be unambiguously matched to one of the programmable resistance bins. The total readout latency of a typical readout will be around 200 ns.

4.5.2 *M*-metric Extraction

4.5.2.1 Direct current-biasing for *M*-metric extraction

The brute-force approach to extract the M-metric by forcing the current (I_{bias}) through the device and sensing the voltage across it suffers from a severe latency penalty as illustrated in Fig. 4.17. The readout current bias (I_{bias}) should be low such that the device content is not disturbed, typically, much lower than the threshold switching current. However, a low I_{bias} requires a longer time to achieve full settling of the voltage across the bit-line (BL) because of the large parasitic capacitance associated with the BL (C_{BL}, typically around 1-1.5 pF). Hence, methods using direct current-biasing for M-metric extraction are inherently slow (time constant >1μs), heavily depend on the sensing resistance and C_{BL} and as such, are not suitable for non-volatile memory applications that generally require access times below 500 ns.

4.5.2.2 Proposed architecture for *M*-metric extraction

The main idea behind the proposed readout architecture is to acquire the metric value using a binary search algorithm in the voltage domain for a certain fixed number of iterations [Athmanathan et al., 2014]. The readout latency of M-metric extraction can be significantly improved by using this approach. The voltage across the BL (V_{BL}) is varied at each iteration until the current flowing through the cell (I_{cell}) matches the pre-defined reference current (I_{ref}). In the first iteration, an initial small voltage (V_{init}) is applied across the cell through the fast voltage regulator. In each subsequent iteration, V_{bias} is set by the v-DAC based on the outcome of the comparison of I_{cell} and I_{ref} from the preceding iteration. If the sensed cell current is larger (smaller) than the reference current, the bias voltage is decreased (increased).

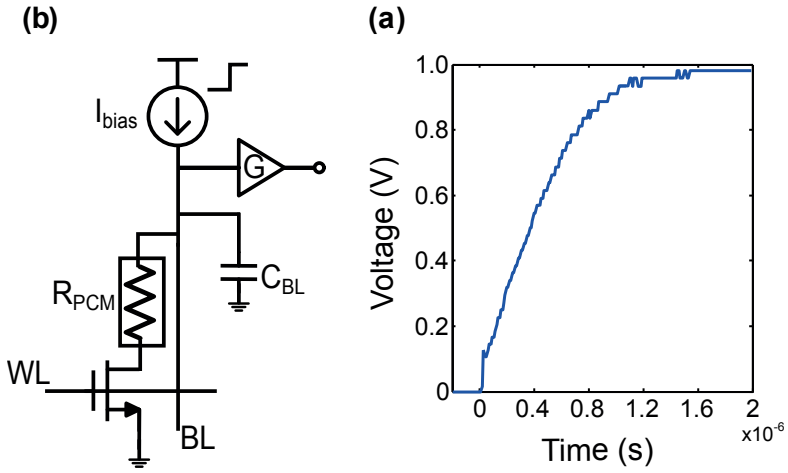


Figure 4.17: (a) Direct current-biasing a PCM device with a low read current (I_{bias}) of 1μA for M-metric readout for a device resistance of 1 MΩ. (b) Captured voltage waveform at the bit-line of the PCM array. Clearly, direct biasing schemes suffer from the inherent speed penalty owing to the huge bit-line parasitic capacitance (C_{BL}).

The voltage applied in the N^{th} iteration (V_N) is given as

$$V_N = V_{N-1} \pm \Delta V_{\text{init}}/2^{N-1}, \tag{4.11}$$

where N is the iteration number, V_{N-1} the voltage applied in the preceding iteration and ΔV_{init} the initial increment step voltage. Figure 4.18 represents the flow-chart of the algorithm implemented for M-metric extraction. The search algorithm terminates when a maximum number of iterations (N_{max}) has been reached or when the error current (I_{error}) is within the desired accuracy limits ($I_{\text{err\_mar}}$) i.e., $\Delta V_{\text{init}}/2^{N-1}$ becomes smaller than the required error margin.

In this manner, the search algorithm converges to the desired M-metric representing the current resistance state of the cell. The simplified block diagram of the readout architecture is shown in Fig. 4.19. The 1-bit result of the first iteration provides a rough estimate of the current state of the cell (range). The bias voltage in the first iteration (V_{init}) is selected in such a way that it does not disturb the content (threshold switching voltage). The V_{bias} for the second iteration is then set as $(V_{\text{init}} + V_{\text{FS}})/2$ or $V_{\text{init}}/2$, depending on result of the first iteration, where V_{FS} is the full scale voltage. The number of iterations effectively determines the resolution of the readout data.

Despite its iterative nature, this approach is significantly faster than the direct current biasing approach because of the short duration of each iteration (thanks to the fast voltage-bias settling). In addition, the total readout time is independent of the BL capacitance and the sensing resistance. The typical readout latency, excluding world-line settling time, is 450 ns [Athmanathan et al., 2014]. The analog-to-digital conversion (ADC) is directly built into the search loop, and this extraction scheme closely resembles the successive approximation

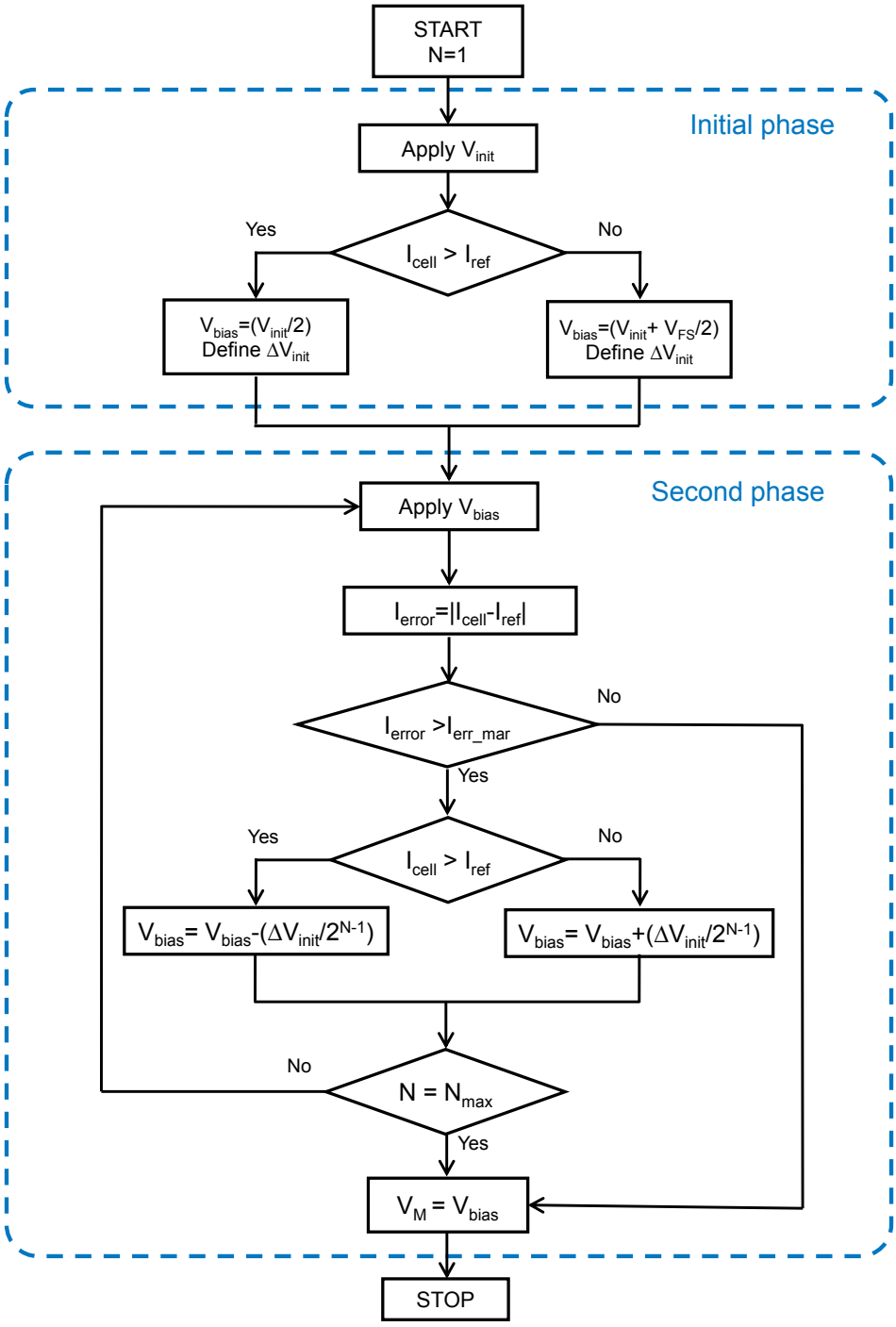


Figure 4.18: Flow chart of the algorithm implemented for M-metric extraction. N is the number of iterations and $N_{\max}$ is the maximum number of iteration. V_{bias} is the applied bias voltage and V_M is the measured M-metric.

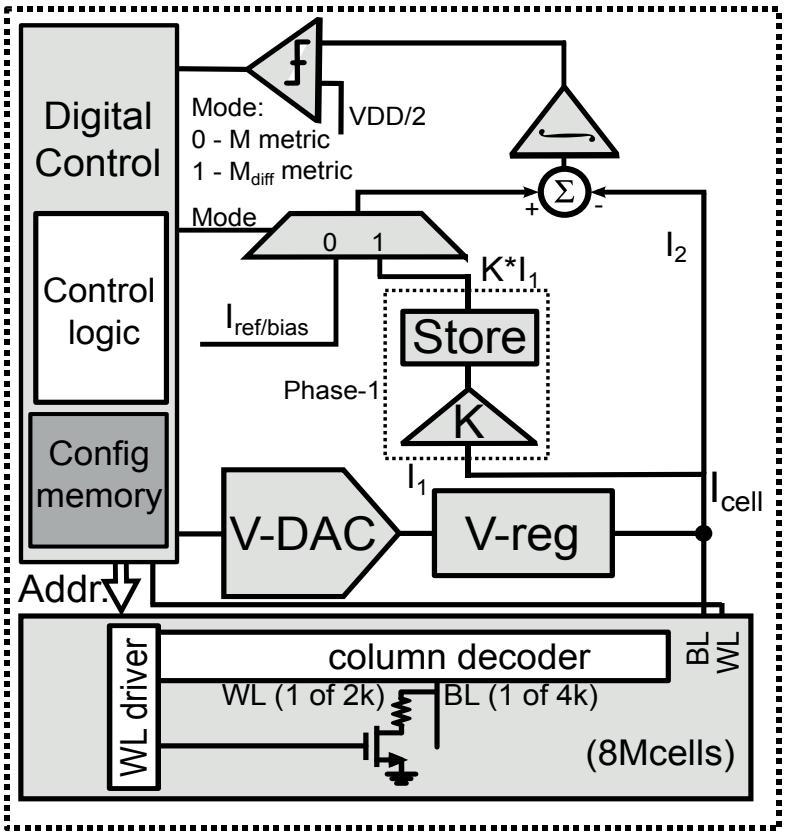


Figure 4.19: Simplified block diagram of the proposed readout architecture. The same architecture is capable of extraction of both the M-metric and the M_{diff}-metric.

register (SAR) ADC technique.

4.5.2.3 CMOS implementation of the proposed architecture for *M*-metric extraction

The proposed readout circuitry has been implemented as an analog and mixed-signal design in 64 nm CMOS technology. Unlike conventional CMOS technologies, the technology is more DRAM-like and is not optimized for typical logic or analog operations. The digital controller initiates and controls the necessary signals required for the readout operation. It also generates all corresponding reference and control signals for the analog components. Based on the comparator outputs, the controller defines the correction voltage through the v-DAC at the end of each iteration.

Compared with the direct current-biasing approach (described in Section 4.5.2.1), the key challenge of the proposed architecture is to achieve a faster settling of the BL voltage (with the target accuracy) across the PCM resistance ranging from 10 kΩ–3 MΩ. The idea is to use a fast voltage regulator to bias the BL voltage and thereby achieve faster settling at the BL. The settling time of the voltage regulator should be on the order of few tens of nanosecond

(20-30 ns), thereby enabling faster current settling at the BL. The voltage regulator can bias the device selected by the address decoders at voltages ranging from 100 mV – 1 V. Across the target resistance range, a settling time of 20 ns was achieved using the voltage regulator.

Figure 4.20 represents the circuit schematic of the voltage regulator and the measured voltage at the BL captured during a readout operation for different resistance values of 10 kΩ and 1 MΩ. The voltage regulator consists of a moderate-gain current-mirror operational transconductance amplifier (OTA) driving the NMOS voltage follower ($N1$) in closed-loop configuration. The two high-impedance nodes ($n1$ and $n2$) correspond to two poles ($p1$ and $p2$). Pole $p2$ is made dominant by adding a compensation capacitor (C_c) in parallel to the C_{BL}. The OTA voltage gain must then be set to a moderate value of 20 to limit phase degradation due to the non-dominant pole $p1$ (> 200 MHz). The impact of the load (large C_{BL} in parallel with a PCM device) on the regulator stability was analyzed. Here, the large C_c and the on-resistance of the BL selection switches strongly decouple the stability criterion from the load.

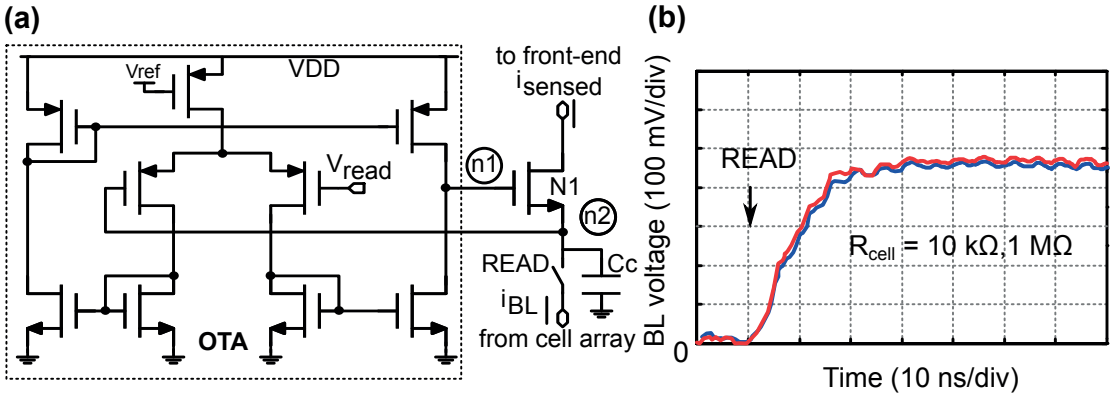


Figure 4.20: (a) Circuit schematic of the two-stage fast voltage regulator. (b) The response of the voltage regulator for device resistances of 10 kΩ and 1 MΩ.

Each readout is performed in two stages. In the first stage, a nominal voltage (around 300 mV, safe voltage range much below the threshold switching voltage) is applied across the device, and the current measured is compared with a programmable reference current, I_{ref}. The comparator output gives a coarse estimate of the resistance of the device (whether the resistance state is closer to the SET or the RESET state). In the second stage, based on the decision, the V_{bias} is applied across the device to proceed with the search algorithm. The second phase terminates after a specific number of iterations or when the required accuracy has been achieved.

Figure 4.21 illustrates the schematic of the overall readout scheme for extracting the M-metric. Once the voltage at the BL (V_{BL}) settles, the sensed current (I_{cell}, which is equal to the PCM device current) is mirrored using a variable-gain current mirror. The gain (K) of the current mirror depends on the coarse comparison at the end of the first stage and the programmable reference current (I_{ref}). The gain can be digitally varied by the controller, so that the search

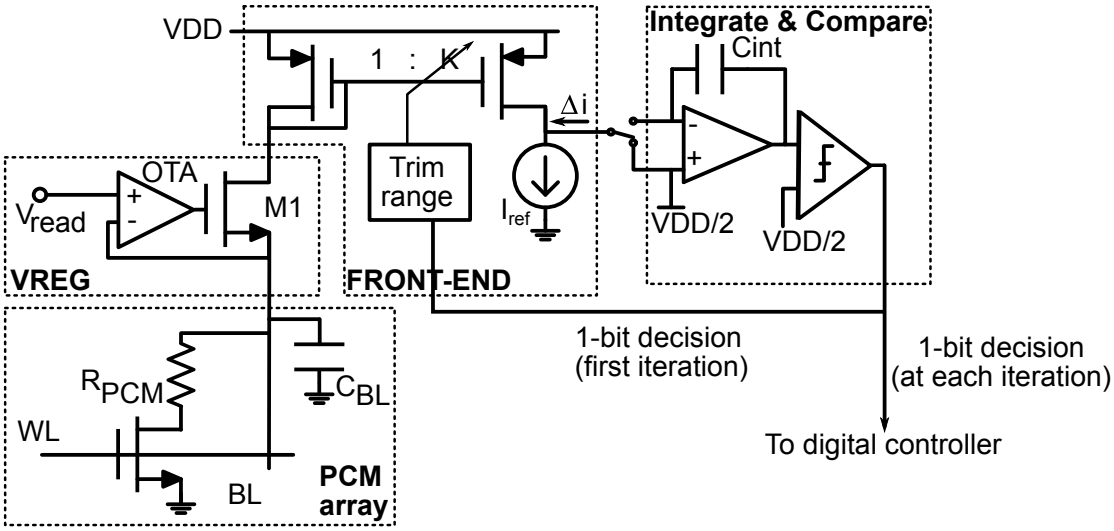


Figure 4.21: Schematic of the analog components implementing the proposed readout architecture. The fast voltage regulator is the key component in determining the readout latency.

space can be increased. If the sensed current is smaller than I_{ref}, the cell state is closer to the RESET state and hence a higher voltage can be applied across the device. In contrast, if I_{cell}>I_{ref}, the cell-state is closer to the low-resistive SET state and hence the applied voltage should be reduced. The first stage ensures that at any given iteration, the voltage applied across the device is always lower than the threshold switching voltage.

In the case of M_{diff}-metric extraction, in addition to the first stage, a factor (K, gain of the current mirror in Fig. 4.21) of the current flowing through the device, I_1, is stored in a capacitor and used as reference current ($I_{\text{ref}} = K * I_1$) for the second stage. Based on the decision at the end of first phase, the gain K should be tuned to increase the search space according to the current resistance state of the cell. The reference current (I_{ref}), either the one stored at the capacitor in the case of M_{diff}-metric or the one used externally in the case of M-metric, is then subtracted from the mirrored current.

Next, the difference current (Δi) is converted into a voltage using an integrator with adjustable integration time, prior to being digitized using a comparator. The integration time can be programmed using the digital controller. After integration, the resulting sampled voltage is compared with the common-mode voltage and a 1-bit decision is made using the comparator.

The controller computes the digital input for the voltage-DAC (v-DAC) according to the 1-bit comparator output. The v-DAC is based on a charge redistribution scheme and directly provides its output to the voltage regulator, thus completing the search loop. The full scale of v-DAC and its offsets can be digitally programmed using the controller. The v-DAC uses segmentation of the input digital code, in which the 3 least significant bits (LSBs) are encoded by a binary-weighted array of capacitors and the 3 MSBs (most significant bits) are realized by

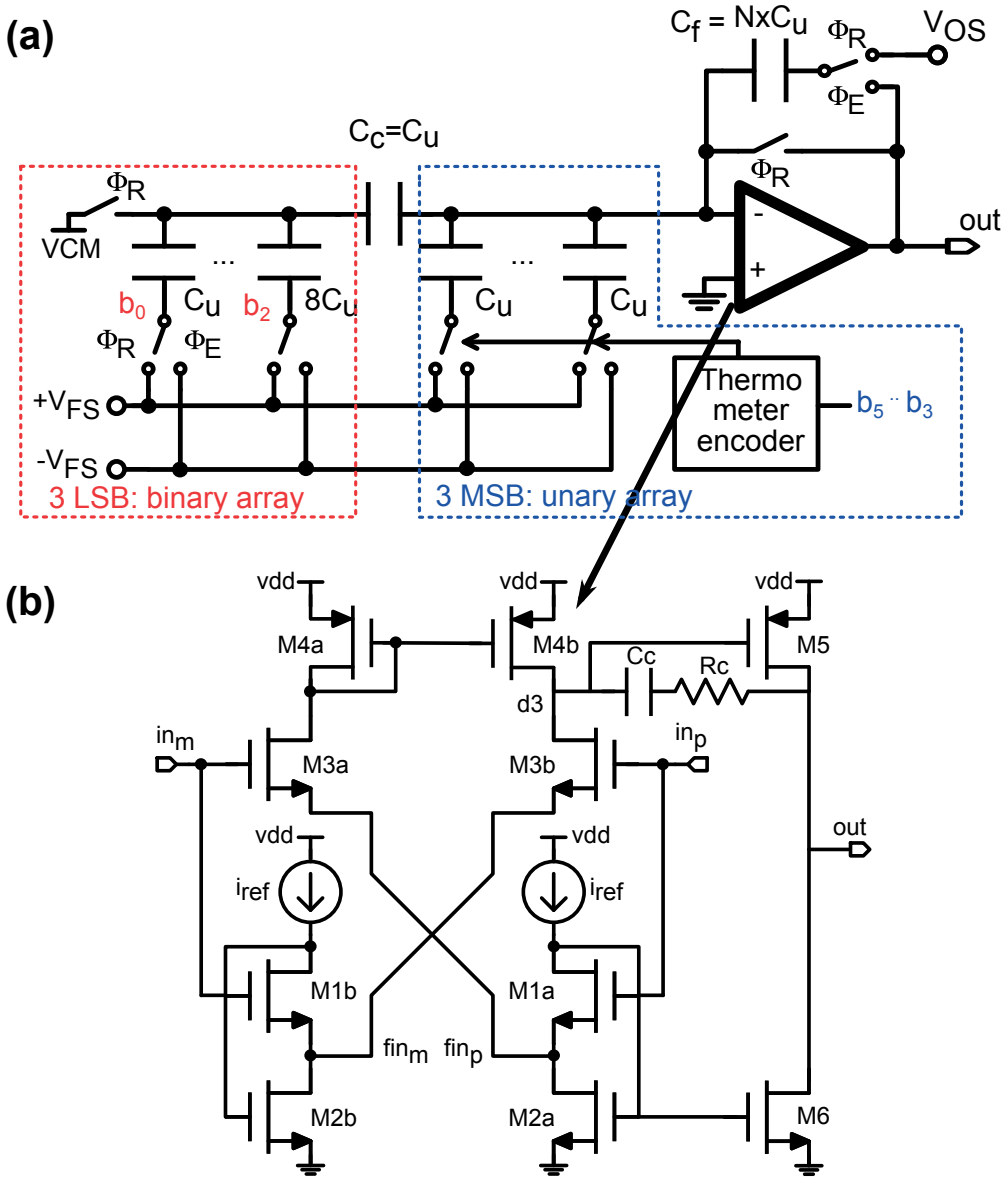


Figure 4.22: (a) Circuit schematic of the charge redistribution 6-bit voltage-output DAC. (b) Class-AB operated cross-coupled input stage OTA forming the core of the v-DAC. The speed of the voltage DAC is critical as it directly influences the readout latency. The v-DAC has a settling time of 10 ns.

an array of 8 identical elements driven from a thermometer encoder. Figure 4.22 describes the schematic of the voltage-output DAC. The OTA uses a class-AB operated cross-coupled input stage, inspired by [Peluso et al., 1997], thereby improving the slew rate without increasing the bias current consumption. The v-DAC has a settling time of 10 ns. Once the search algorithm has terminated, the voltage at the output of the v-DAC already is the M-metric value measured.

4.5.2.4 Experimental measurements

The proposed readout architecture along with the programming electronics, have been implemented as an analog and mixed-signal design in 64 nm CMOS technology . The circuit is characterized using a hardware platform designed for high-performance characterization and testing of non-volatile memories [Papandreou et al., 2013]. The characterization is facilitated by poly-silicon resistors that span the full resistance range and are integrated as a sub-array. The BL can be accessed externally through buffer amplifiers and can be used for both DC and transient measurements. In the resistance range of interest (20 kΩ - 1 MΩ), the voltage regulator achieved settling of the cell current in less than 30 ns, which allows a total iteration time of only 60 ns. The flexibility offered by the digital controller allowed us to fine-tune the parameters critical for closed-loop convergence during readout.

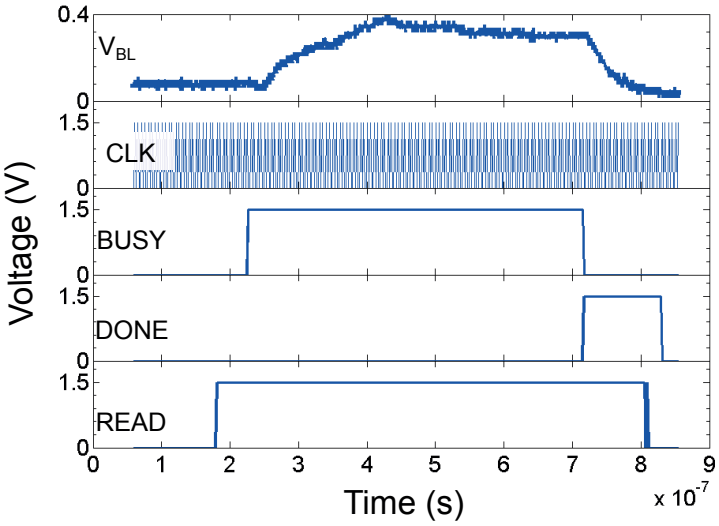


Figure 4.23: Timing waveforms captured during the READ operation.

The waveforms shown in Fig. 4.23 captured at the chip interface illustrate the timing of the read operation. The signals generated by the digital controller and the total read access time of 450 ns for 6-bit readout data have been verified. The ADC transfer curve for different values of the M-metric is shown in Fig. 4.24. The measured differential non-linearity (DNL) for the operating range of the ADC is illustrated in the inset. The increased DNL due to the extended operating ranges reduces the effective resolution of the complete readout chain to 5 bits. V_{BL} in Fig. 4.25 illustrates a binary search algorithm trajectory for four different resistances covering the range of typical PCM high-field resistance values. The chip was operated at a low frequency of 40 MHz to capture sufficient details.

Apart from the current fluctuations in the PCM cell, the noise introduced by the CMOS readout electronics is mainly due to the voltage regulator maintaining the BL at the readout voltage. The equivalent input voltage noise contributed by the voltage regulator, $V^2_{n,OTA}$, results in

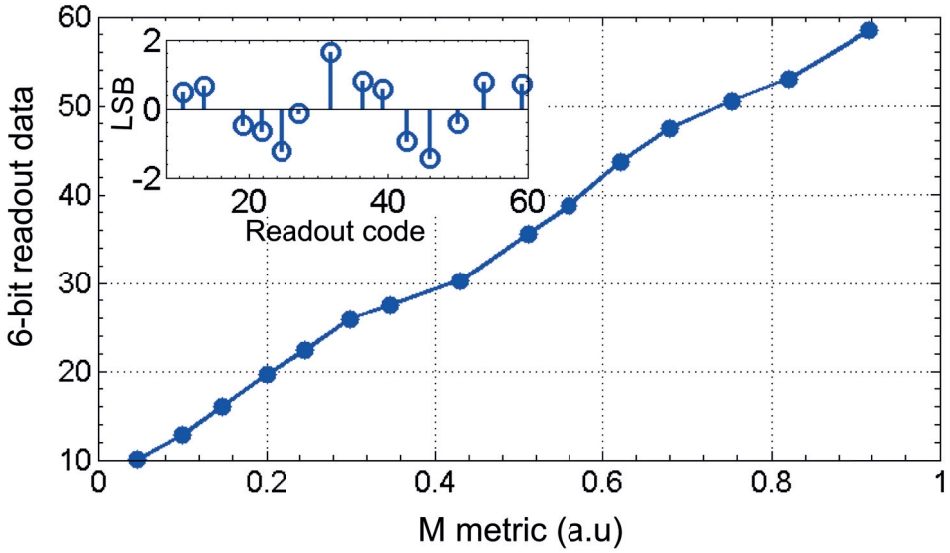


Figure 4.24: Measured ADC characteristics across the integrated poly-silicon resistor array. The inset shows the measured differential non-linearity (DNL). As the ADC operates in two operating regimes (one closer to the SET, the other to the RESET states), the effective resolution of bits achieved is 5 bits.

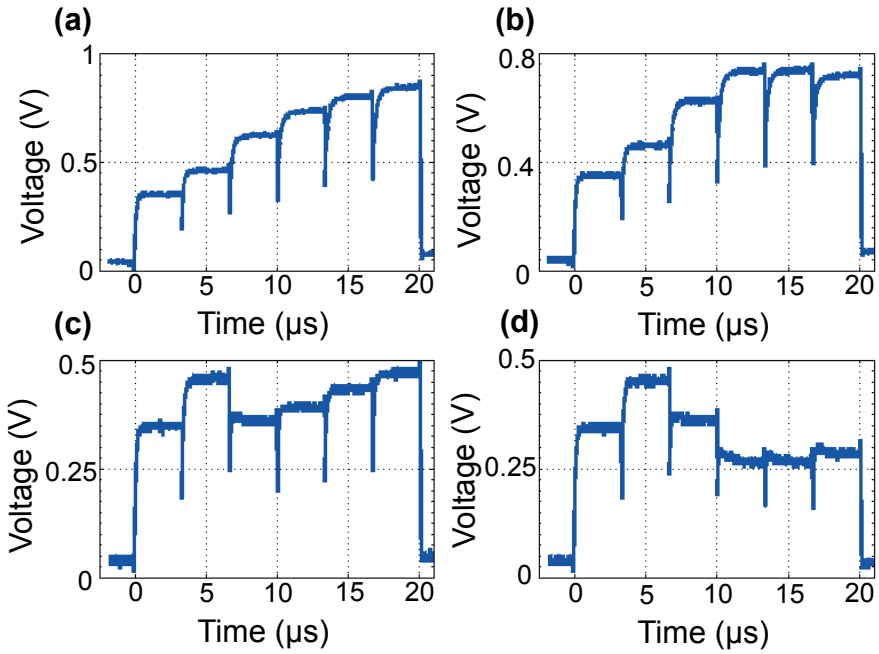


Figure 4.25: The trajectory of the bit-line voltage, V_{BL}, for different resistance values: (a) 1 MΩ, (b) 760 KΩ, (c) 420 KΩ, and (d) 240 KΩ. The voltage clearly follows a binary search algorithm for attaining the required M-metric values. The chip was operated at reduced 40 MHz clock to capture the waveforms with sufficient detail.

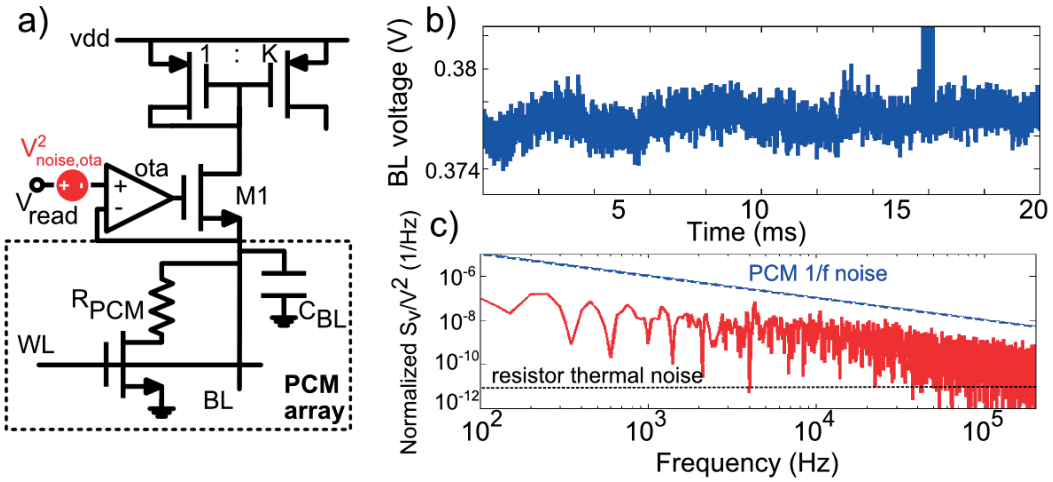


Figure 4.26: (a) Main noise source in the circuit schematic. (b) Measured voltage across BL for an intermediate level ($R \sim 150$ kΩ). (c) Normalized noise power spectral density (S_V / V^2) extracted from the captured waveform.

current fluctuations through the parallel combination of the BL capacitance and the PCM resistance (Fig. 4.26). The level of circuit noise is for all practical purposes dominated by the intrinsic PCM noise (also shown in Fig. 4.26 [Close et al., 2010]).

4.5.3 *eM*-Metric extraction

The performance of M-metric in terms of drift resilience is far better than that of the R-metric. Although the extraction of M-metric extraction can be sped up by the implementation of a binary search algorithm in the voltage space [Athmanathan et al., 2014], there is still some margin for enhancement owing to the lower and constant detection current typically used in M-metric extraction (as the detection current should be lower than the threshold switching current).

Recently, Stanisavljevic et al. proposed a practical circuit realization of the eM-metric (shown in Fig. 4.27), which consists of constant current biasing (I_0) and a resistor R_0 added in parallel to the PCM resistance, R_{PCM}, which reduces the effective resistance seen at the output node and thereby considerably reduces the settling time and read latency [Stanisavljevic et al., 2015]. The threshold detection curve is therefore defined as $I_{\mathrm{cell}} = I_0 - V_{\mathrm{cell}} / R_0$ instead of the constant detection current used in the M-metric extraction. Additionally, for low-resistive states, most of the biasing current flows through R_0, providing improved contrast and higher SNR. The scheme is the circuit-level equivalent of the metallic surfactant layer to decouple the read current from the drift-prone amorphous region [Kim et al., 2013].

Typical resistor values of R_0, are $150-200$ kΩ, yielding a read latency time (including ADC conversion) of $250-300$ ns for typical BL capacitance values of up to 1 pF. Even though the read

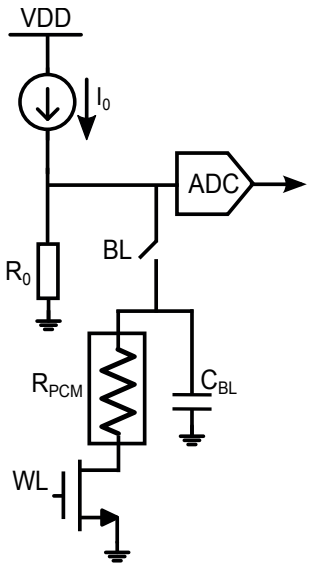


Figure 4.27: Schematic for extraction of the eM-metric. R_0 is the resistor parallel to the PCM resistance, R_{PCM}. R_0 decouples the read current from the drift-prone amorphous state.

latency is almost independent of the sensing resistance, its dependence on the BL capacitance remains owing to the direct current biasing. In addition to a lower latency than that of the iterative M-metric realization [Athmanathan et al., 2014], the eM-metric circuit realization is significantly smaller, consumes less power and has lower circuit noise because only passive elements are present in the input stage (similarly as in [Close et al., 2010], the circuit noise is for all practical purposes negligible compared with the intrinsic PCM device noise).

One disadvantage of such a configuration is the compressed signal range for high-resistive states, as the majority of the current now flows through the parallel resistor. However, this slight compression in the signal range does not compromise reliable MLC read out and data retention, as will be demonstrated in the next chapter.

4.6 Summary

The traditional low-field resistance metric is not well-suited for MLC storage operation, owing to several serious limitations. In contrast, non-resistance-based metrics, like the M-metric and eM-metric have a huge potential for reliable MLC operation. Given the advantages of the M-metrics, I strongly believe that realization of MLC PCM is possible only with the likes of such drift-resilient metrics.

The readout architecture presented demonstrated a fast and efficient implementation of drift-resilient cell-state metric suitable for MLC operation, which, for the first time, enabled the performance required by non-volatile memory applications. The programming and readout electronics were fabricated in a prototype chip as part of a Joint Development Agreement

(JDA) with SK-Hynix for the commercialization of MLC PCRAM devices. Moreover, the viability of MLC PCM has been demonstrated using the proposed readout metrics, as will be presented in the next chapter.

The readout circuit was designed and fabricated in 64 nm CMOS technology, achieving an access time of 450 ns at 6-bit raw (5-bit effective) resolution. The CMOS technology was not optimized for logic or analog circuits and was more a DRAM-like process. The circuit does not exhibit sensitivity to BL parasitics and showed low-noise properties. The readout chain is ready to be co-integrated with a 16 Mb 2x nm PCM cell array and the necessary programming electronics.

Contributions:

- **Proposed a novel readout architecture for extraction of drift-resilient readout metrics in MLC PCM storage.**
- **Characterized the prototype chip fabricated using a FPGA-based hardware characterization platform. The performance of the readout circuitry in terms of access time (450 ns) and the accuracy of the readout data (5-bit effective ADC resolution) were investigated.**
- **Design and implementation of the iterative programming algorithms for MLC programming.**
- **Contributed significantly at the top-level design of prototype chips in terms of chip interface, design of the digital controller, array address decoders, various test point buffers, and output for monitoring the internal chip signals, etc.**

5 Reliability and data retention analysis in MLC PCM

In order for any non-volatile memory (NVM) to be considered a viable technology, its reliability should be verified at the array level. In particular, properties such as high endurance and at least moderate data retention are considered essential. PCM possesses highly desirable features and has reached an advanced level of maturity through intensive research and development in the past decade. Multilevel-cell (MLC) capability is not only desirable as it reduces the effective cost per storage capacity, but a necessary feature for the competitiveness of PCM against the incumbent technologies, namely, DRAM and Flash memory. However, as described in the previous chapter, MLC storage is seriously challenged by the phenomena of resistance drift and array variability.

In this chapter, I present some of the critical reliability concerns in PCM, especially for MLC storage operation. I summarize the performance of various readout metrics, introduced in the previous chapter, against the challenges of resistance drift and variability. I demonstrate the viability of MLC storage in PCM through reliable storage and moderate data retention achieved for 2 bits/cell, on a 64k cell sub-array, which is pre-cycled for a million (10^6) cycles (SET/RESET endurance cycles). The data-retrieval process takes place at regular time intervals, with the temperature varying from 30°C to 80°C. Under similar operating conditions, I also show the feasibility of 3 bits/cell (triple-level cell, TLC) storage in PCM, for the first time ever.

This chapter is organized as follows: **Section. 5.1** explains the major reliability criteria especially for MLC operation in PCM. **Section. 5.2** describes the FPGA-based experimental hardware characterization test platform used for the reliability and data retention experiments. **Section. 5.3** summarizes the influence of endurance cycling on the MLC functionality. **Section. 5.4** presents the reliability and data retention analysis results from the experiments performed on a 64k cells and finally, in **Section. 5.6**, the conclusion from the reliability and data retention point of view for MLC operation are provided.

5.1 Reliability concerns in MLC PCM

Endurance cycling, data retention and data disturb are the main reliability factors in PCM. Endurance cycling refers to the number of repeated programming cycles that can be reliably performed in a memory device. Data retention is the capability of the device to retain the stored data over time. Data disturb refers to the unintended distortion of the stored data in a cell either by programming neighboring cells (thermal disturb) or by repeated readouts of the same cell (read disturb).

5.1.1 Endurance cycling

Endurance cycling is an important reliability criterion for any memory technology, as higher endurance broadens the application area to areas where frequent read/write is required. One of the main reasons for the pronounced interest in PCM is its high endurance, which is measured to be on the order of 10^8 cycles at the array level [Gleixner et al., 2007]. Figure 5.1 shows the typical SET/RESET cycling endurance of cells that were experimentally studied in this thesis. For comparison, MLC Flash memory is limited to only 10^4 endurance cycles, and thus necessitates the application of complex wear-leveling schemes to make it suitable for enterprise applications.

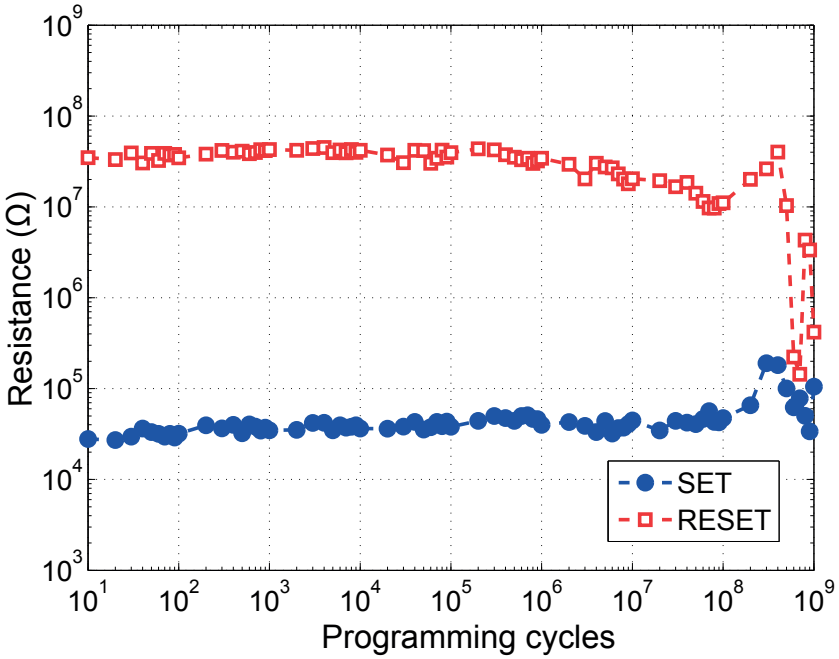


Figure 5.1: Endurance cycling at room temperature of a typical PCM device studied in this thesis.

Endurance cycling is a characteristic of the phase-change material. The two main failure modes commonly observed in PCM are the stuck-at-SET and stuck-at-RESET failures. Pedersen et al. demonstrated that the phase transition between the crystalline and the amorphous

state is accompanied by a considerable change in density [Pedersen et al., 2001]. They experimentally measured that GST thin films show a 6.5 % thickness decrease upon crystallization. With repeated cycling between the crystalline and amorphous states, the corresponding density variations result in mechanical stress. Eventually, this leads to void formation within the active region of the device, resulting in stuck-at-RESET failures. However, Chen et al. proposed that by adding suitable amount of doping material to GST, they are able to delay the void formation process, thereby significantly improving the endurance to more than 10^9 cycles (only 10^6 cycles in the case of undoped-GST) [Chen et al., 2009].

The other failure mode, stuck-at-SET, is usually attributed to the elemental segregation that these materials exhibit with repeated cycling. Raoux et al. observed that repeated cycling leads to Sb enrichment at the bottom electrode [Raoux et al., 2008]. Sb-rich materials tend to have lower crystallization temperatures. This results in the crystallization of the active region at much lower temperatures than the actual crystallization temperature corresponding to the original material composition, thus causing stuck-at-SET failures.

5.1.2 Data retention

The other critical reliability measure for any non-volatile memory is its capability to retain the stored information, known as data retention. The typical data retention criterion for enterprise-level storage applications is 3 months at 40 °C [1]. Significant research efforts have been performed to estimate the actual data retention of PCM. In contrast to the crystalline phase, the amorphous phase is meta-stable and hence determines the retention time. Typical data retention measurements involve measuring the RESET resistance of the device over time. When the resistance falls below a certain threshold (typically, a resistance reduction by a factor of two under isothermal conditions with no applied bias), the device is assumed to suffer a retention failure. These measurements are also performed at various higher temperatures (around 160°C) to accelerate the crystallization process and hence the subsequent retention failure. From the high-temperature measurements and a reasonable activation model, researchers can predict the retention performance at elevated temperatures. Using measurements from a mushroom-type device, Pirovano et al. showed an Arrhenius behavior with an activation energy of 2.6 eV, extrapolating to a data retention capability of 10 years at 110°C [Pirovano et al., 2004].

Later Redaelli et al. proposed a percolation model for data retention on 180 nm μ-trench PCM cells and analyzed this model using a temperature-dependent percolation effect. They showed that repeated retention measurements on the same device resulted in widely varying retention times, demonstrating the stochastic component of GST crystallization. They attributed these variations to the random spatial configuration of the as-nucleated grains in the amorphous region [Redaelli et al., 2005]. When these percolation paths appear through an amorphous layer, the resistance drops significantly, resulting in retention failure.

[1] Solid-State Drive Requirements and Endurance Test Method. JEDEC Standard, JESD218A, 2011.

Russo et al. showed that using an Arrhenius extrapolation would be pessimistic in the data retention estimation as the energy barrier for nucleation is even larger at higher temperatures. This reduces the driving force for nucleation, causing the temperature dependence of nucleation to be less than Arrhenian [Russo et al., 2007]. A number of groups have also investigated other PCM materials and doping as a way to increase the activation energy for crystallization, so as to increase the retention time [Matsuzaki et al., 2005; Kim and Ahn, 2005; Oh et al., 2006].

Apart from spontaneous crystallization, which does not pose any serious threat to meeting the desired data retention criterion, the other major concerns for data retention are resistance drift (in the case of MLC storage operation), thermal disturb and read disturb.

5.1.2.1 Resistance drift

All the data retention analysis described above are based on SLC (single-level cells, 1-bit/cell) operation in PCM. The scenario is entirely different in the case of MLC operation, as resistance drift poses a far more serious threat to data retention of the intermediate resistance levels (than the crystallization), resulting in data loss. As a consequence, it might not be possible to attain the typical data retention criteria for the intermediate resistive levels in MLC operation. Thus, MLC PCM may not be suitable for typical storage applications. However, moderate data retention between that of DRAM (64 ms) and Flash, if achieved, would open up a range of new storage-class memory applications for MLC PCM [Freitas and Wilcke, 2008].

The phenomenon of drift results in the steady resistance increase of the high-resistive amorphous state over time. In the case of SLC PCM, the drift process increases the margin between the ON/OFF (SET/RESET) states and hence does not cause any harm in data retention and in fact may aid the readout process. However, in the case of MLC, drift is detrimental. It has been shown that high temperatures accelerate the drift phenomenon, and hence, the stability of the stored levels at elevated temperatures is a critical reliability concern for MLC operation [Sebastian et al., 2015]. Therefore in MLC PCM, retention failures should be addressed such that moderate data retention is achieved. Various MLC-enabling technologies for PCM described in Section 4.3 should therefore be designed by taking the reliability and data retention issue into consideration.

5.1.3 Thermal or program disturb

Thermal disturb is another reliability characteristic that is unique to PCM because of the latter's deliberate use of heat for the programming mechanism. Thermally activated crystallization of the high-resistive states results in a significant reduction of the resistance, leading to data failures in both SLC and MLC PCM. One possible cause of the crystallization is thermal disturb. When a cell has been programmed to the RESET state, repeated programming of the adjacent cells might induce the crystallization of the cell because of the heat diffused from the neighboring cells (described in Section 3.5). However, this has been proved not to be critical

at the 45 nm technology node, and has also been projected, through models and simulations, to not pose a serious threat at future, sub-20 nm technology nodes [Russo et al., 2008].

5.1.4 Read disturb

Apart from spontaneous crystallization, the other mechanism that can influence crystallization and induce data loss is repeated read operations. During every read operation, the small read current flowing through the device can induce a localized heating that can accelerate the spontaneous transition from amorphous to crystalline, thus inducing a premature fault. As the crystallization process of the amorphous phase is promoted at higher temperatures, these failure mechanisms are aggravated at elevated temperatures [Sebastian et al., 2015].

Given the typical reliability concerns for MLC operation in PCM, it is interesting to study each of the MLC-enabling technologies (described in Section 4.3) offering resilience to drift and variability in terms of the MLC reliability criteria, such as temperature fluctuations and endurance cycling.

5.2 Experimental Characterization of MLC PCM

The choice of phase-change material, the design and fabrication of the memory device, including the process definition and the material deposition technique, are key components in demonstrating the reliability and endurance performance for MLC operation. Mushroom-type memory devices with doped-GST as phase-change material are used for the experiments. The bottom electrode is formed by a lithographically independent, so-called “key-hole” process [Breitwisch et al., 2007], creating the keyhole with a sub-lithographic CD that is essentially independent of the original feature size. The keyhole is subsequently filled with TiN, which acts as the heater (bottom electrode).

The PCM devices are integrated into a prototype chip serving as the characterization test-vehicle in 90 nm IBM9SF CMOS technology. Our experimental hardware platform is built around the prototype PCM chip with a 2×2 M cell array with a 4-bank interleaved architecture (Fig. 5.2). The memory array size is 2 ×1000 μm × 800 μm. In addition to the PCM cell array, the prototype chip contains the necessary circuitry for cell addressing, on-chip ADC for cell readout and voltage- or current-mode cell programming [Close et al., 2010]. Table 5.1 summarizes the specifications of the PCM prototype chip.

5.2.1 Characterization Platform

Figure 5.3 illustrates the schematic block diagram of the versatile hardware platform designed for high-performance characterization and testing of non-volatile memories. The platform is based on a reconfigurable hardware/software architecture with a high degree of multi-domain testing and data-acquisition capabilities and is closely based on the one proposed

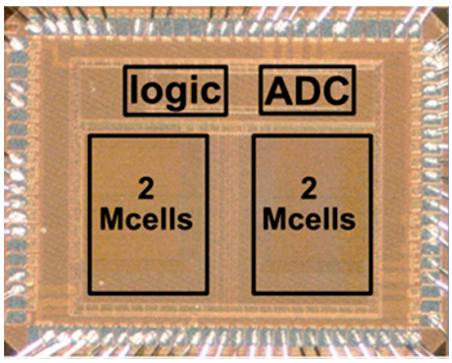


Figure 5.2: Die micrograph of the prototype memory chip used in the experiment. The PCM chip consists of a 2×2 M cell array with a 4-bank interleaved architecture.

Table 5.1: Specifications of the PCM prototype chip

CMOS baseline	90-nm, 6 metal levels of Cu
Access device	NMOS
Memory element	Doped $Ge_2Sb_2Te_5$ (d-GST)
Array size	2×2 M cells
Supply voltage	2.5 V
ADC resolution	8 bits
Readout latency	500 ns

in [Papandreou et al., 2013]. It provides a valuable tool for statistical characterization of the solid-state memory channel for new and emerging non-volatile memories. It consists of the following main units:

- A high-performance analog-front-end (AFE) board that contains a number of digital-to-analog and analog-to-digital convertors (DACs and ADCs) along with discrete electronics, such as power supplies, voltage, and current reference sources, etc.
- A FPGA board that implements the data acquisition and the digital logic to interface with the memory device under test, with all the electronics of the AFE board.
- A second FPGA board with an embedded processor and Ethernet connection that implements the overall system control and data management as well as the interface with the data processing unit.
- A temperature control unit consisting of a Eurotherm temperature controller, a temperature sensor, a heating resistor and the power supply.
- A workstation with MATLAB running a lab automation test suite for executing numerous experimental scenarios and providing real-time data processing.

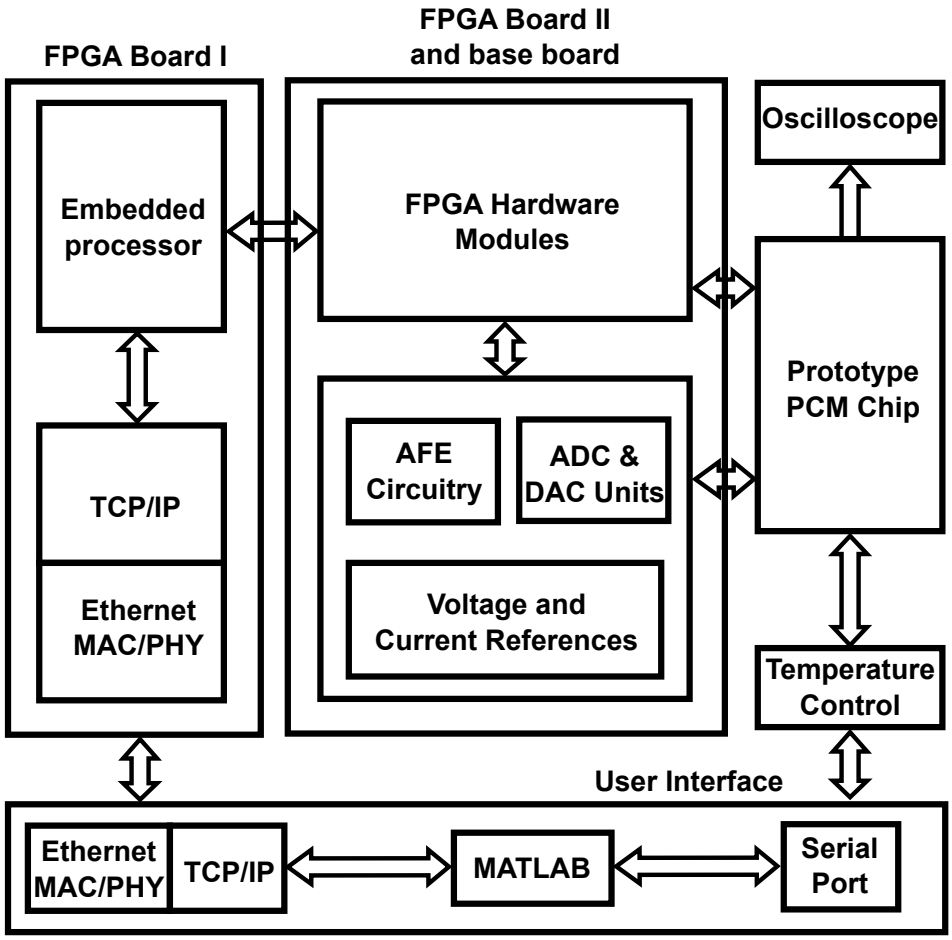


Figure 5.3: Schematic block diagram of the complete hardware characterization platform.

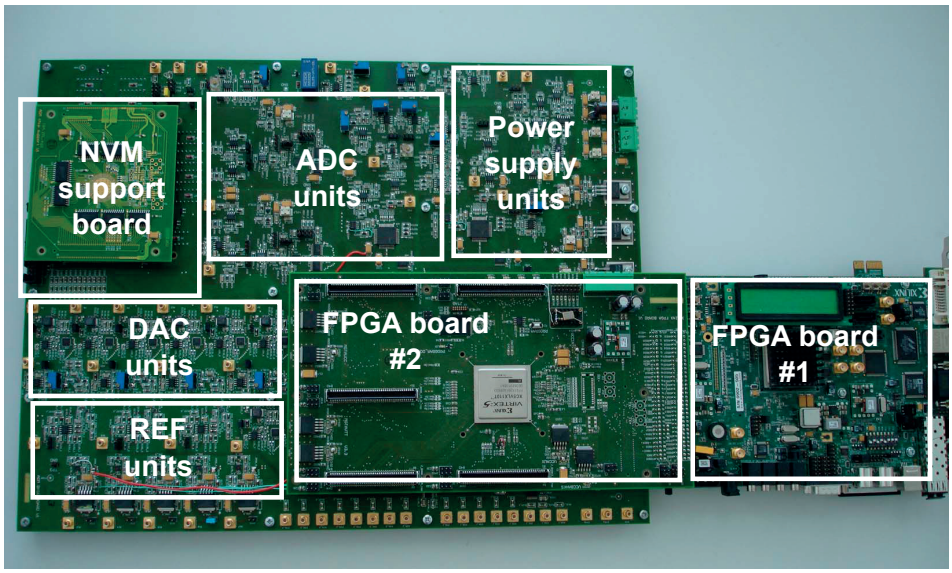


Figure 5.4: Photograph showing the experimental hardware characterization platform.

The prototype PCM chip is connected to the hardware characterization platform via a specially designed chip support board that offers numerous test-point buffers enabling external access to the chip's critical internal signals for debugging and monitoring purposes. The lab automation test suite also provides the necessary interface to the platform's temperature control unit for variable high-temperature characterization experiments. The temperature of the device under test is measured using the temperature sensor attached underneath the chip. A heater with 20 W power is also mounted along side the temperature sensor. The PID temperature controller monitors the input from a temperature sensor and provides a power output through the power supply to the heating element. The overall chip temperature can be increased by applying power to the heater.

5.2.2 Experimental Procedure

Figure 5.5 depicts the flow chart of the experimental procedure that was designed to assess the MLC reliability of the PCM prototype chip in terms of data retention and cycling endurance under variable time and temperature conditions. A sub-array of 64 k pristine memory cells is selected and subjected to repeated SET/RESET programming cycles. During the endurance cycling, single RESET/SET pulses are used to repeatedly program and erase the cells. The cycling RESET pulse is a box-type pulse of maximum power and a width of 150 ns. The cycling SET pulse is of moderate power with a trailing edge of 2.4 μs. At regular intervals in a logarithmic fashion, i.e., every 10^4 cycles, 10^5 cycles, etc., endurance cycling stops and the cells are analyzed in terms of MLC programming and data retention under variable time and temperature profiles.

More specifically, the sub-array is programmed into four distinct resistance levels, namely, $L0$ (SET), $L1$, $L2$ and $L3$ (RESET). The two intermediate levels, $L1$ and $L2$, are programmed using the iterative programming algorithm. The maximum number of iterations allowed during the iterative programming operation was set to 20. The two corner levels, $L0$ and $L3$, are programmed by single SET and RESET pulses, respectively. The stored levels are based on random user patterns encoded by the scheme described in Section 4.3.4.

After the successful programming of the cells under test, the stored data in the sub-array is read back at regular intervals to evaluate the drift performance of the various read-out metrics described in Section 4.3.3. To assess the bit-error performance under practical operating conditions and to study the drift behavior at elevated temperatures, a profile in which the temperature is varied between 30 °C and 80 °C was used. The temperature profile over time is illustrated in Fig. 5.11(a). After the retention measurements, the cycling continues to the next cycling point. This procedure continues for a total of 1 million (10^6) cycles.

The two intermediate levels ($L1$ and $L2$) are optimally placed for each of the R- and eM-metrics before programming in order to maximize retention for the given temperature profile and readout metric.

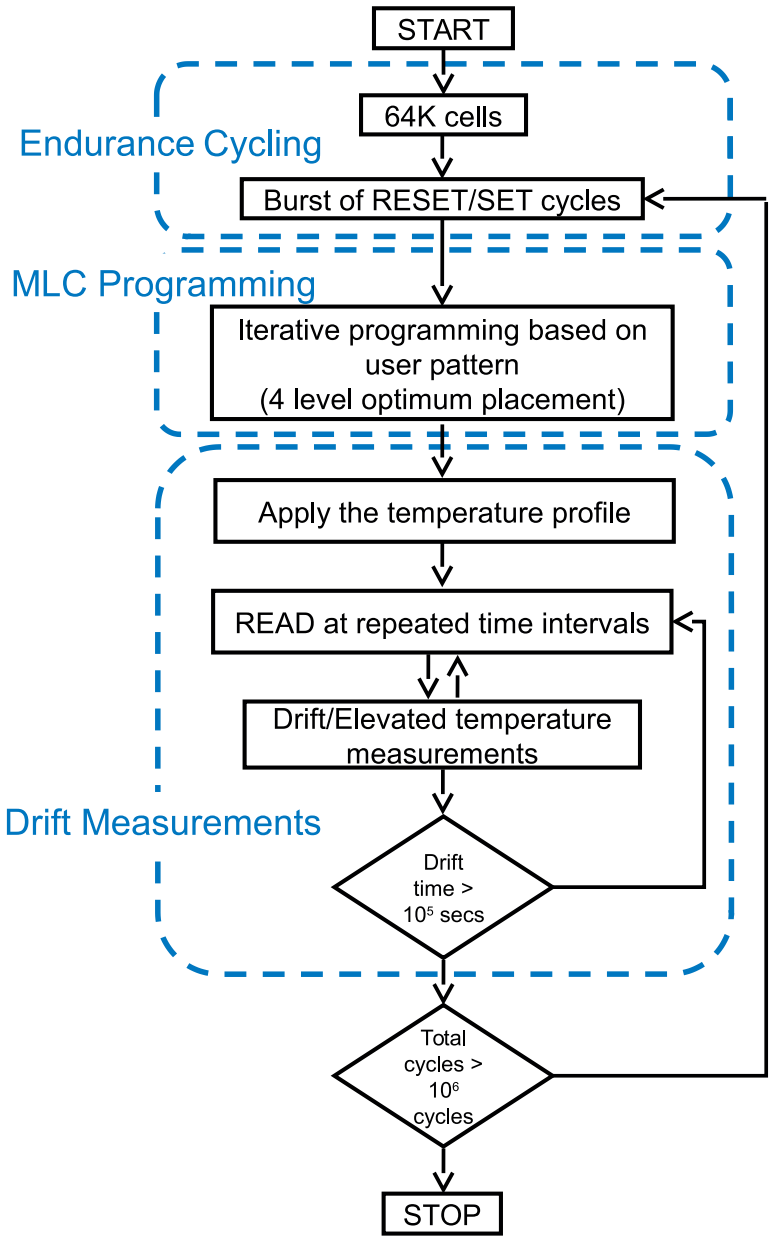


Figure 5.5: Flowchart of the experimental procedure. The cells are cycled repeatedly in the Endurance cycling stage. Four levels are optimally placed using the iterative programming algorithm. The stored levels are repeatedly read at time intervals and at the desired temperature.

5.3 Impact of endurance cycling on MLC performance

The phase-change material, the device-type and the fabrication process in particular, are responsible for the remarkable stability that these devices exhibit as a function of repeated programming and erase cycles. This is illustrated in Fig. 5.1, which shows the measured resistance of the PCM devices used for this experiment, after application of consecutive RESET

and SET programming pulses. It can be observed that in both states the resistance is stable over more than 10^8 programming cycles. However, for MLC operation, in addition to SET and RESET states, it is of particular significance to assess the effect of cycling on the characteristics of the intermediate resistance states. This is particularly important for MLC programming because large variations on the programming curve will curtail the ability of the iterative programming algorithm to accurately program intermediate resistance levels to the target resistance levels.

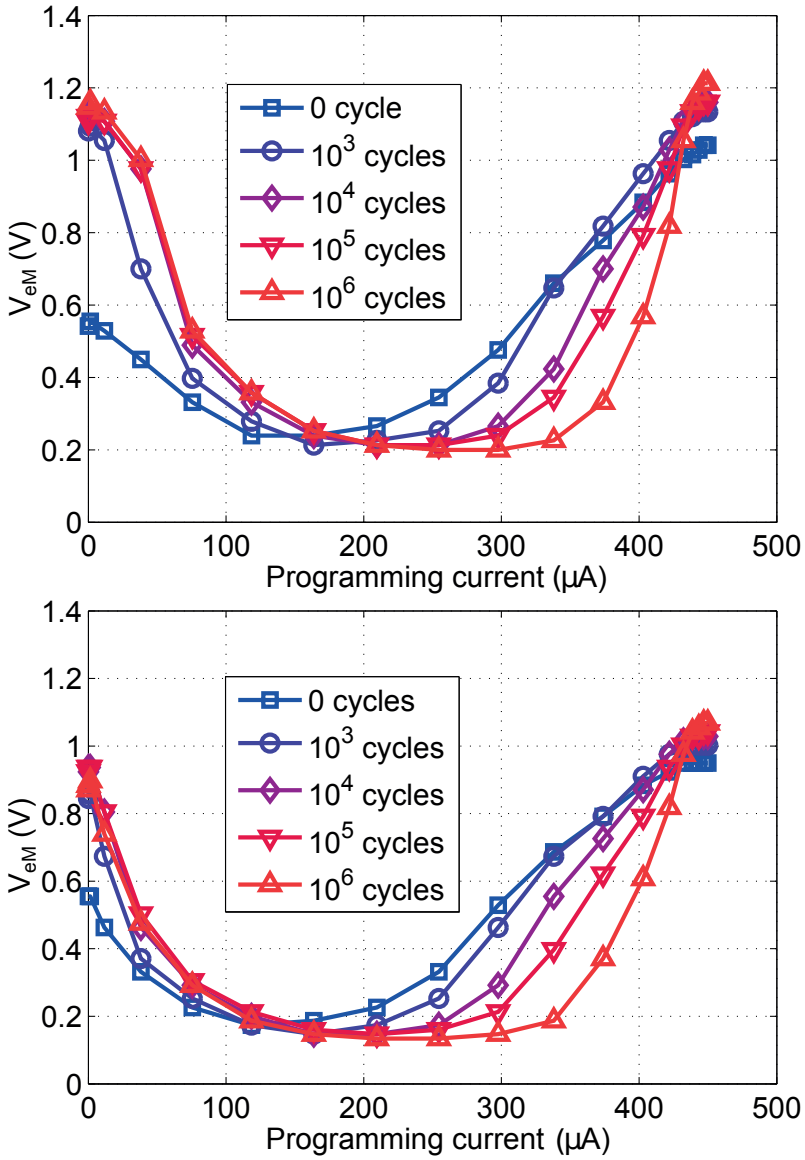


Figure 5.6: The evolution of the $eM - I$ programming curves with respect to the number of SET/RESET endurance cycles for two different ambient operating temperatures (a) 30 °C and (b) 80 °C. The measured data corresponds to the average of 64 kcells. The difference in the eM-metric values is due to the decrease in the resistance values at higher temperatures.

5.3.1 Evolution of the *eM-I* Programming Curve

Figure 5.6 depicts the measured static, "single-shot" programming $eM-I$ curves averaged over 64k cells as a function of the endurance cycles at two different ambient operating temperatures, 30 °C and 80 °C. With repeated programming cycles, the programming $eM-I$ curve moves to the right and also widens a bit. In other words, with repeated cycling, a slightly higher power is required to RESET the cells, but at the same time, the dynamic signal range improves. For MLC programming, the right slope of the $eM-I$ curve is used by the iterative programming algorithm to program the cell in the desired intermediate resistance level. However, endurance cycling causes the right slope of the programming curve to become increasingly steeper, making it more difficult to attain the target resistance levels while programming the intermediate states.

5.3.2 *eM*-metric evolution

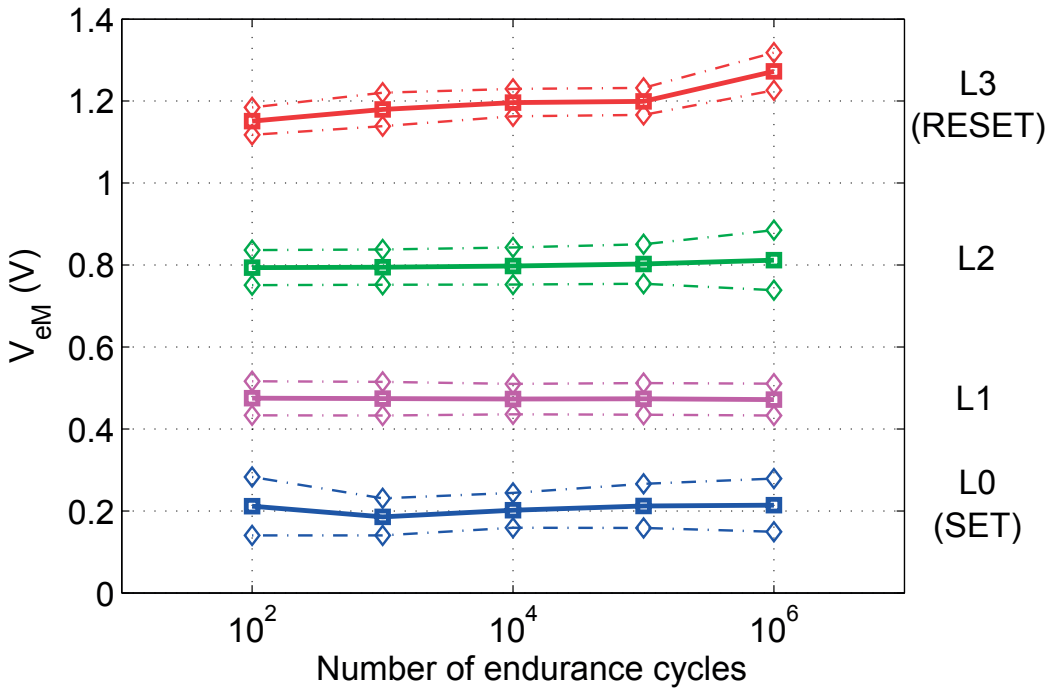


Figure 5.7: Evolution of eM-metric values for four levels as a function of SET/RESET endurance cycles, measured at room temperature (30 °C). $L0$ and $L3$ are the extreme SET and RESET levels. $L1$ and $L2$ are the intermediate levels for MLC programming. The solid line shows the mean value of 64k cells; while the dash-dotted line represent $\pm 3\sigma$ variation.

The mean value and $\pm 3\sigma$ variation of the eM-metric values, read 100 s after programming for the four programmed levels as a function of endurance cycles, are summarized in Fig. 5.7. Although endurance cycling affects the steepness of the right slope of the $eM-I$ programming curves (see Fig. 5.6), the two intermediate levels ($L1$ and $L2$) are programmed in closed-loop

using the iterative programming algorithm described in Section 4.3.2, keeping their mean value constant and independent of the number of cycles. However, with endurance cycling, the variability increases at 1 million (10^6) cycles, especially for the $L2$ intermediate state (Fig. 5.7). In addition, with endurance cycling, the SET and RESET mean values change similarly as in the $eM-I$ programming curves and therefore the resistance margin between them increases. In a way, this increased SET-RESET resistance range effectively compensates for the increase in variance at 10^6 cycles, suggesting that endurance cycling compromises neither level placement nor data retention.

5.4 MLC Reliability and data retention analysis

In this section, the data retention capability of the various readout schemes, presented in Section 4.3.3, is analyzed using the measured data from the experimental platform. Firstly, the results of iterative algorithm for programming the two intermediate levels on a 64 kcell sub-array that has been pre-cycled 1 million (10^6) times are shown. The influence of temperature on the characteristics of the device at the array level was studied. To maximize data retention, two intermediate levels should be optimally placed during programming. By fitting the previously measured experimental data, the mean and variances of each intermediate level at arbitrary time instances can be obtained. Then, by assuming Gaussian statistics of the level distributions, the optimal placement of two intermediate levels that minimizes the bit-error rate (BER) is calculated iteratively for each of the readout metrics (the optimal level placement is metric dependent). Following that, the reliable readout and data retention of the stored data were analyzed.

5.4.1 MLC Storage operation (2 bits/cell)

According to the experimental procedure of Fig. 5.5, after the cells have been pre-cycled a million times, two intermediate resistance levels are optimally placed between the extreme SET and RESET states. The performance of the iterative programming algorithm used for programming the intermediate resistance levels is presented in Fig. 5.8. Figure 5.8(a) shows the convergence distribution of the optimally placed intermediate resistance levels, $L1$ and $L2$. On average, these two levels require 3.3 and 6.4 iterations to converge to the desired target resistance level, respectively. Irrespective of the million endurance cycles, the iterative algorithm achieves convergence of 99% of the cells in fewer than 14 iterations. Figure 5.8(b) summarizes the eM-metric value distribution for all four levels programmed, over the ensemble of 64k cells, 500 μs after programming. Clearly, the levels are tightly distributed and well separated, making them a suitable choice for performing the drift evolution experiment.

Detailed $I-V$ characteristics for the entire sub-array showing all four programmed levels, 100 s after programming, are shown in Fig. 5.9. The $I-V$ points are collected below the threshold power (constant 1.5 μW power; dashed curve in Fig. 5.9) to prevent potential disturbance of the resistance states during readout. The solid line in Fig. 5.9 captures the detection curve

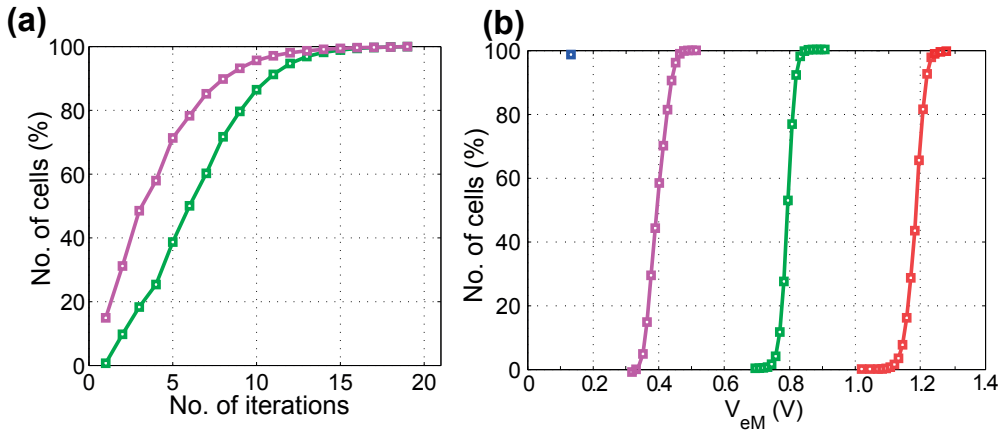


Figure 5.8: (a) Convergence – number of iterations required to program the two intermediate resistance levels using the iterative programming algorithm. (b) Distribution of all four programmed levels in eM-metric.

used for extraction of the eM-metric, with an equivalent readout circuit biasing current of 4.5 μA and parallel resistor of 300 kΩ. As can be observed, the levels are very well separated in the eM-metric measured 100 s after programming.

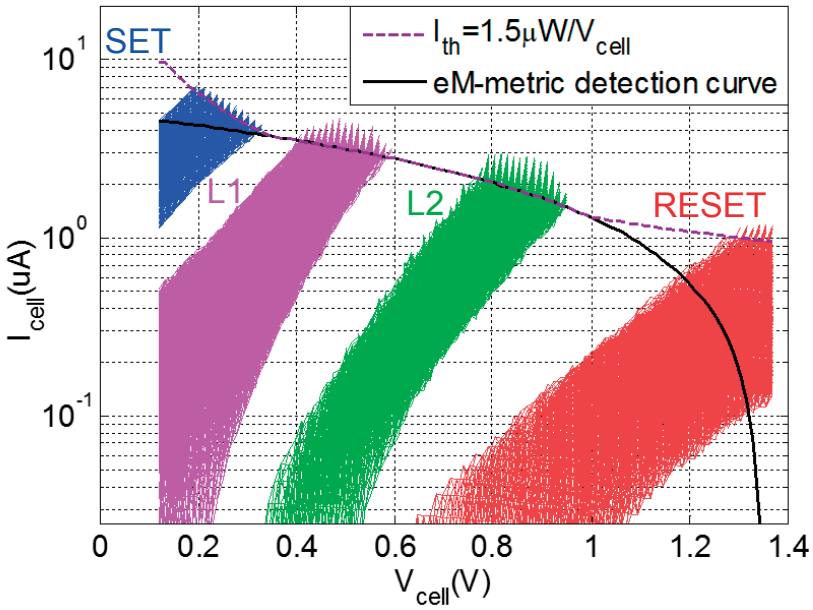


Figure 5.9: $I-V$ curves of the four programmed resistance levels for 64k cells measured at room temperature, 100 s after programming. The solid black line captures the readout detection curve used for eM-metric extraction. The dashed purple line represents the iso-power curve used for collecting the $I-V$ points without disturbing the cell content (threshold current protection).

5.4.2 Influence of Temperature

The impact of temperature fluctuations on the device characteristics and performance at the array level has been studied. This is the first time such a study has been performed in MLC PCM. The available dynamic range as well as the eM-metric values are reduced at higher temperature (i.e., at 80 °C). Moreover, the eM-metric value variations at higher temperature are not uniform across all levels. This is evident from Fig. 5.10, where the distribution of the four levels is shown 500 μs and 500 s after programming at both 30 °C and 80 °C. High-resistive states are prone to a significant reduction in the eM-metric value compared with low-resistive states. This is also clearly visible from the measured $eM - I$ curves in Fig. 5.6. This behavior is expected owing to the temperature-dependent conductivity of the phase-change materials and therefore the conductivity increases at elevated temperatures [Raoux et al., 2009].

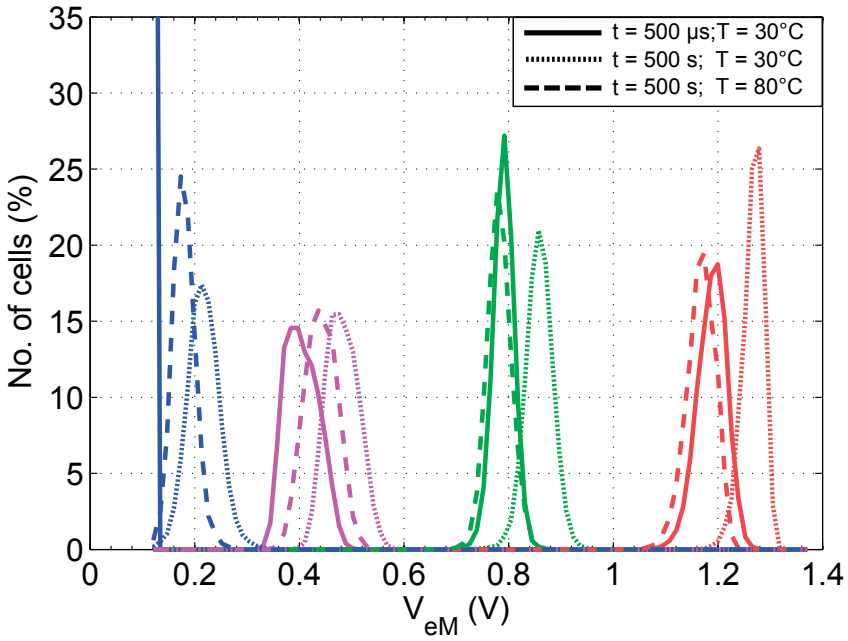
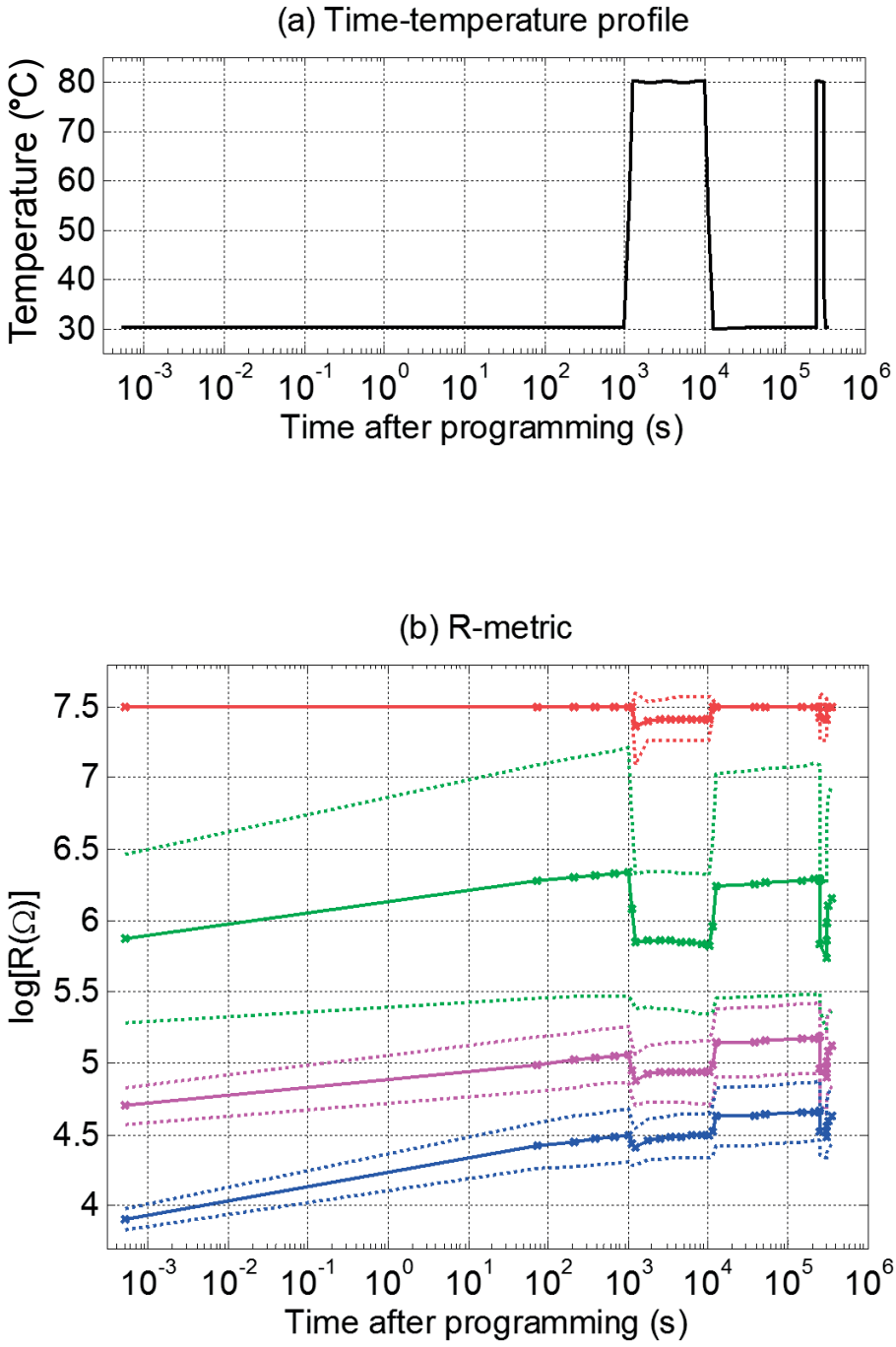


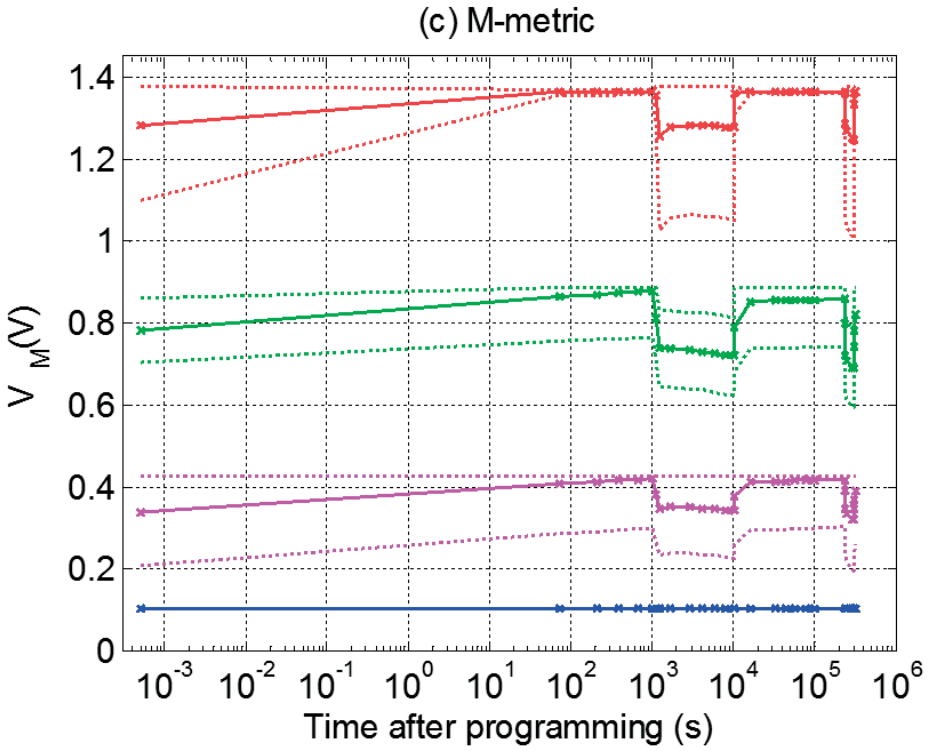
Figure 5.10: Histogram of the four programmed levels read 500 μs and 500 s after programming measured at 30 °C and 80 °C.

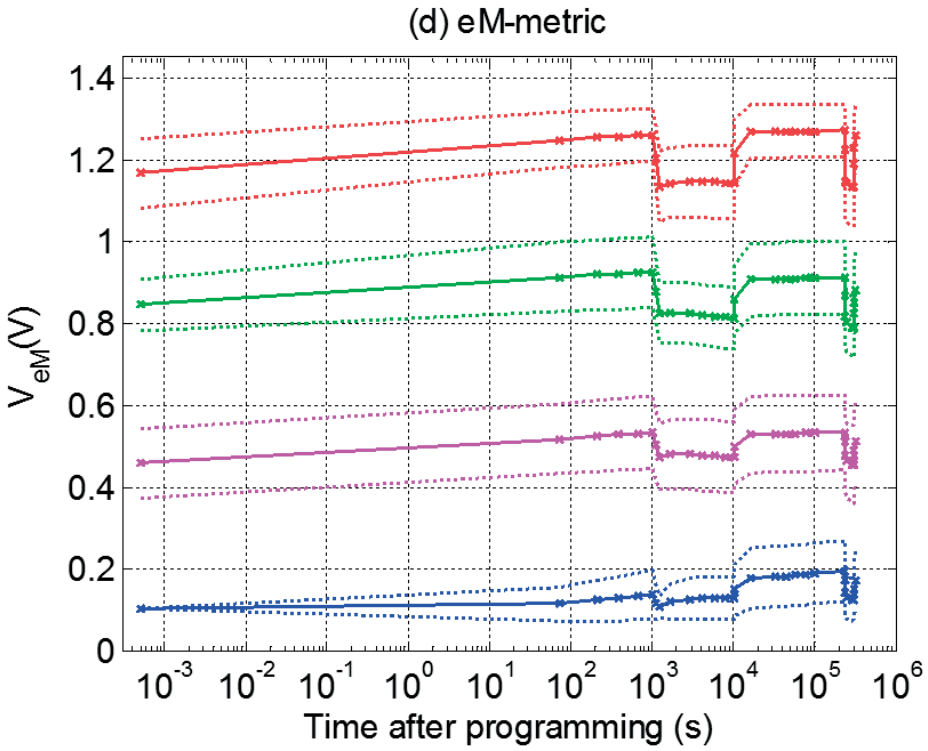
5.4.3 MLC data Retention analysis

The time evolution of four optimally placed levels for each metric (R, M and eM) after 10^6 endurance cycles is shown in Fig. 5.11. The corresponding time-temperature profile is depicted in Fig. 5.11(a). We have designed this profile such that it induces significant stress in the PCM cells, representing a worst-case situation for data retention. This is exemplified by (a) two temperature ramps to 80 °C, lasting for 3 h and 16 h, respectively, and (b) a total of 4 days of data retention. The retention performance of the M-metric and eM-metric appears significantly better than that of the R-metric, as these two metrics offer better contrast between the stored

Figure 5.11: (a) Time-temperature profile of the temperature variations across the chip. Drift evolution of the four individually optimally placed levels for (b) the R-metric, (c) the M-metric and (d) the eM-metric, measured at regular time intervals and at the temperature corresponding to the time-temperature profile. The solid line and the dotted line represents the mean value and $\pm 3\sigma$ bounds of the respective metrics. The performance of both the M-metric and eM-metric is superior to that of the R-metric.




(c) M-metric
V M(V)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
10-3 10-2 10-1 100 101 102 103 104 105 106
Time after programming (s)


(d) eM-metric
V eM(V)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
10-3 10-2 10-1 100 101 102 103 104 105 106
Time after programming (s)

levels, therefore facilitating the data-retrieval process (Fig. 5.11(b)-(d)).

Owing to the low-voltage bias, attaining 2 bits/cell, using the resistance metric is quite challenging because the dynamic range is significantly smaller than those of the M and eM-metrics. This is mainly due to the saturation issues associated with the R-metric as discussed in Section 4.3.3.1. The M and eM-metrics exhibit rather similar time and temperature evolutions. Therefore, to quantify the retention performance, only the R and eM-metrics are compared in terms of the bit-error rate (BER) in the next section.

5.4.3.1 BER Performance

Figure 5.12(a) shows the bit-error-rate (BER) evolution of the R-metric value as a function of the time elapsed after programming. Two data-detection schemes are compared: the reference-cell scheme and the permutation modulation (PM) code and detection method, both described in Section 4.3.4, and denoted as "REF-CELL" and "CODE", respectively. For fairness of comparison, the fraction of reference cells used for the level estimation is chosen equal to the redundancy of the PM code of length 20, i.e., $\frac{2-1.67}{2} \approx 0.17$.

It can be observed that the BER of the R-metric already crosses the 10^{-4} mark 1000 sec after programming, i.e., during the time when the levels drift at room temperature. Furthermore, although the BER temporarily improves when the temperature ramps to 80 °C, the R-metric generally exhibits a high sensitivity to temperature changes. The direct dependence on the activation energy makes it vulnerable to temperature fluctuations and resistance drift. This may be inferred from the characteristic behavior of the resistance metric with temperature fluctuations, illustrated in Fig. 5.11(b). However, the SNR of the resistance metric decreases steadily as a function of time at constant temperature. This leads to the monotonic increase of the BER with time in stretches of constant temperature, both at 30 °C (until 1000 s and between 10^4 and 10^5 s), and at 80 °C (between 10^3 and 10^4 s). Also, the "reference cell" scheme is clearly unable to detect the data reliably.

Similarly, Fig. 5.12(b) depicts the BER evolution of the readout eM-metric with time after programming. Evidently, the performance of the eM-metric is largely superior to that of the resistance metric, both for the reference cell and the coded schemes. Specifically, the BER performance of the eM-metric when the modulation code is used is extremely robust, staying below 7.7×10^{-5} even after four days of data retention at adverse ambient temperature conditions.

To put this error rate into perspective, a standard Bose–Chaudhuri–Hocquenghem (BCH) error-correction code operating on a user segment of 512 bits and being able to correct 7 bits, would be able to bring a raw BER of approx. 1.5×10^{-4} at its input down to 10^{-15} at its output. BCH codes of such complexity can be decoded in tens of nanoseconds in modern FPGAs, and are therefore suitable for low-latency memory applications. This proves the viability of reliable 2 bits/cell data storage in PCM cell arrays after extensive endurance cycling, as well as reliably,

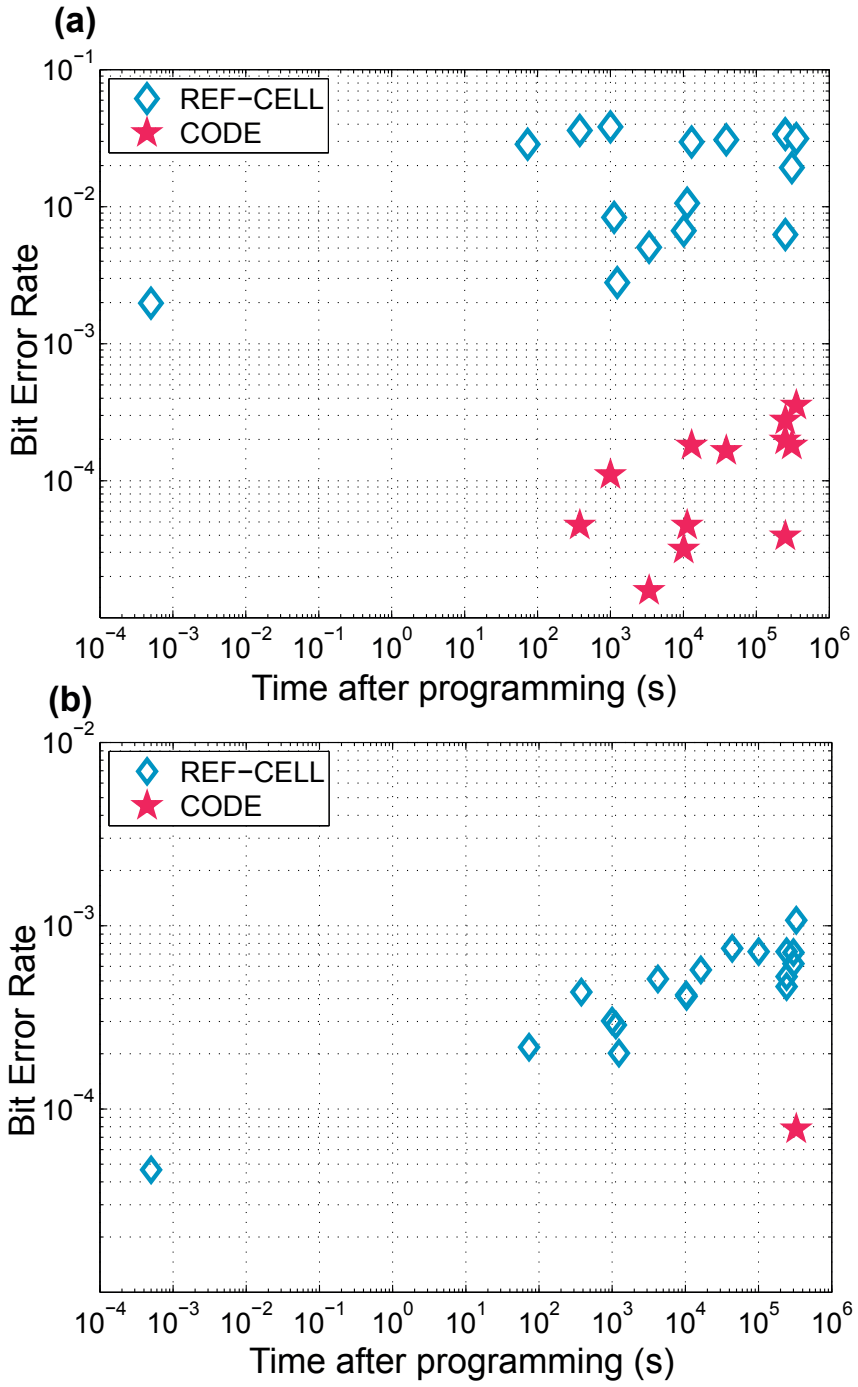


Figure 5.12: BER performance of readout metric measured at 340,000 sec (approx. 4 days) after programming: (a) R-metric; (b) eM-metric; Detection schemes based on the reference cell method and permutation modulation coding scheme are compared. The performance of the eM-metric is extremely robust, staying below 7.7×10^{-5} even after four days of data retention at adverse ambient temperature conditions.

moderate data retention in the presence of temperature fluctuations (30°C – 80°C), in line with storage-class memory requirements.

5.5 Triple-level cell (TLC) storage in PCM

Given the superior performance of the M-metric and eM-metric over the R-metric with very encouraging MLC data retention results, we were motivated to explore also the possibility of increasing the memory density to 3 bits/cell, also known as Triple-Level Cell (TLC) storage. We repeated the same experimental procedure to store eight programmed levels (3 bits/cell) on cells that had a history of a million cycles, to investigate the data retention performance of these metrics for TLC storage operation.

5.5.1 Feasibility of 3 bits/cell (TLC) programming

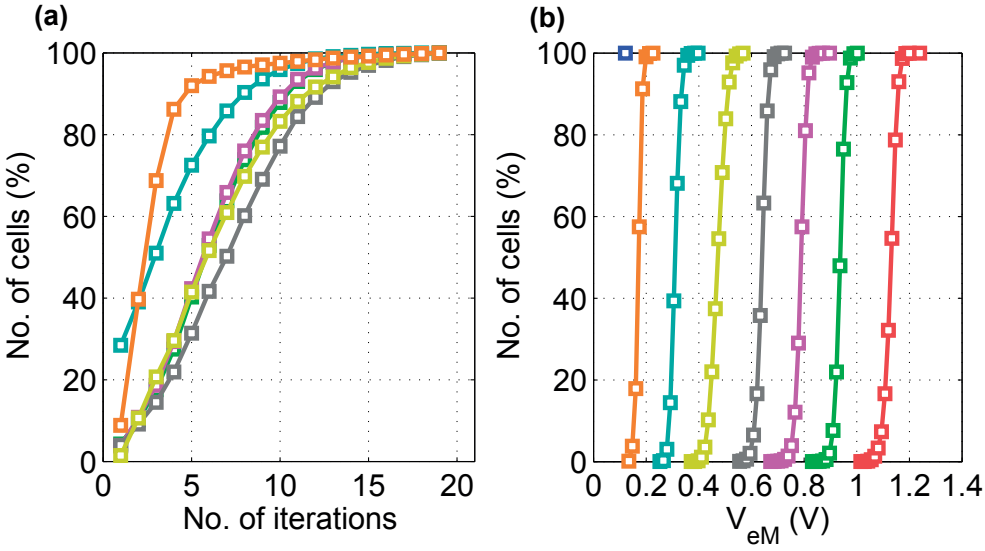


Figure 5.13: (a) Convergence – number of iterations required to program the six intermediate resistance levels using the iterative programming algorithm. (b) Distribution of all eight programmed levels in eM-metric.

The read and programming circuitry can support more than 2 bits/cell operation as the resolution of the ADC used in the readout circuitry is 8 bits. A potentially limiting factor in terms of the programming electronics for TLC operation is the convergence of the iterative programming algorithm with reduced target bin width, which is essential to accommodate 8 levels (3 bits/cell). Figure 5.13(a) demonstrates the convergence of the iterative algorithm for the programming of six intermediate eM-metric levels apart from the corner SET and RESET states. The distribution of the levels just after programming is shown in Fig. 5.13(b). It is evident that the iterative programming algorithm appears to be capable of accurately programming the six intermediate levels. Figure 5.14 depicts the $I-V$ curves of all eight programmed resistance levels, measured 100 s after programming.

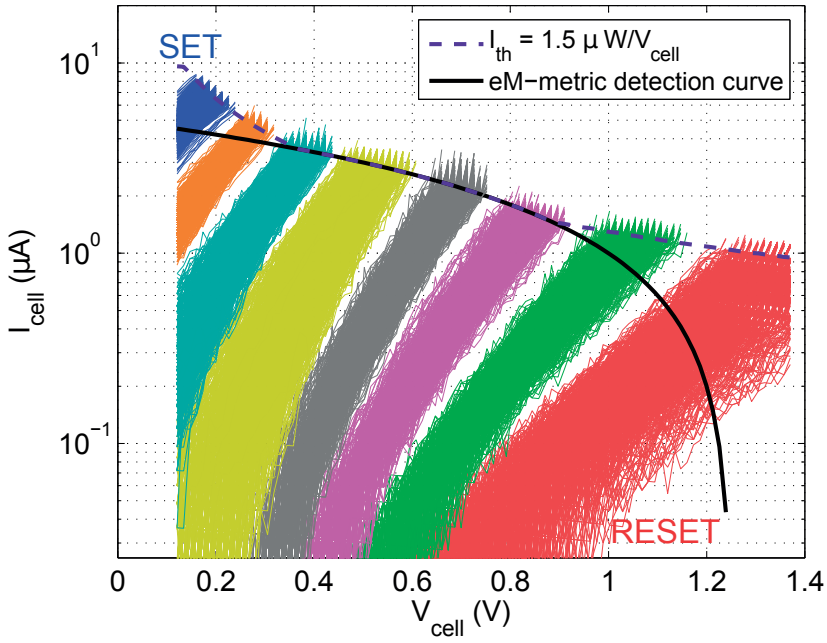


Figure 5.14: $I-V$ curves of the eight programmed resistance levels for 64k cells measured at room temperature, 100 s after programming. The solid black line captures the readout detection curve used for eM-metric extraction. The dashed purple line represents the iso-power curve used for collecting the $I-V$ points without disturbing the cell content (threshold current protection).

5.5.2 *eM*-metric evolution with endurance cycling

The mean value and $\pm 3\sigma$ variation of the eM-metric values read 100 s after programming for all eight programmed levels over 64k cells as a function of endurance cycling are summarized in Fig. 5.15. With endurance cycling, there is a marginal increase in the SET-RESET range, whereas the variance remains fairly constant, apart from a slight increase in the high-resistance states at 10^6 cycles, owing to the steepness of the right slope of the programming curve. However, it can be seen that by judicious level placement, it is possible to spread the levels to account for the variance, suggesting that endurance cycling will not compromise TLC data storage.

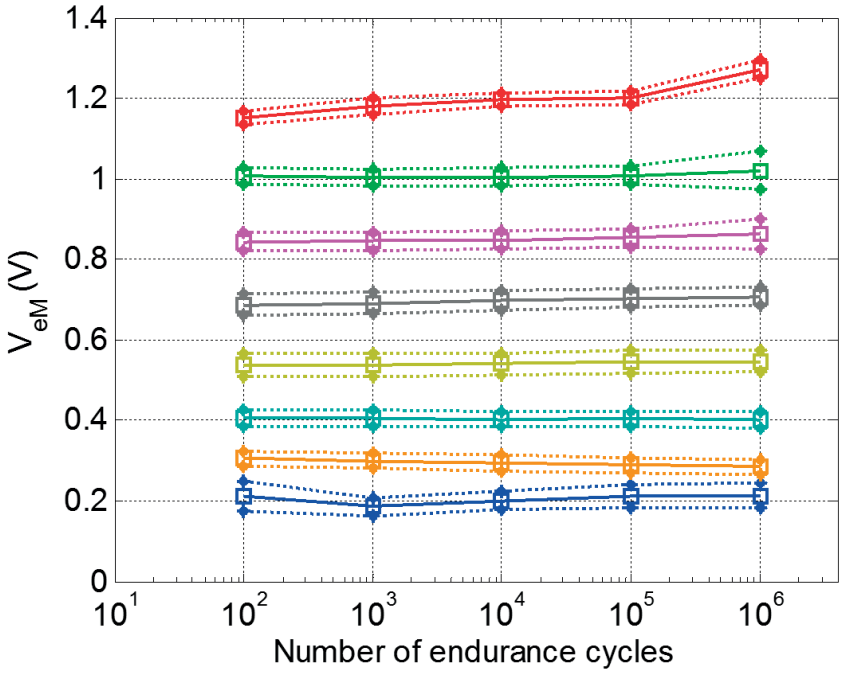


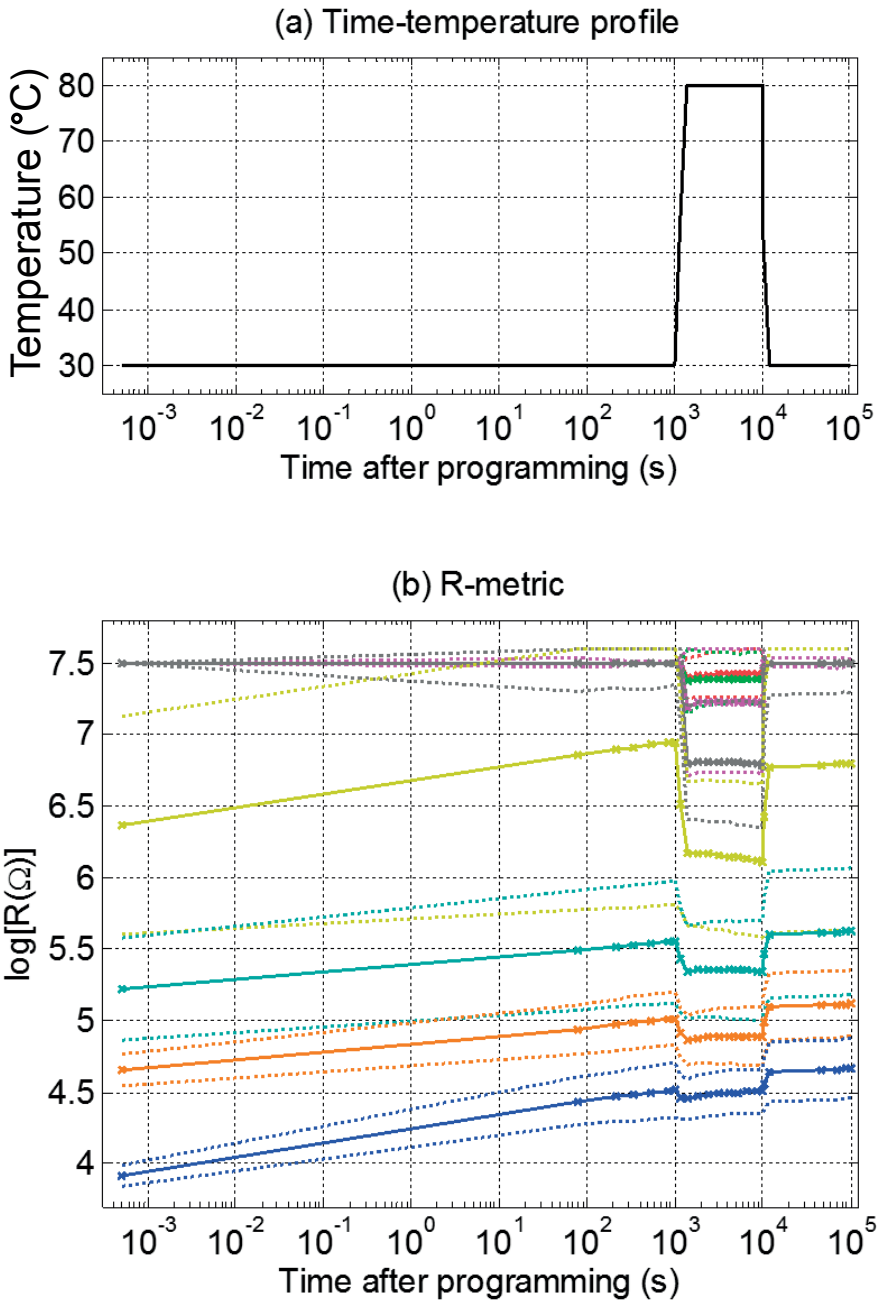
Figure 5.15: eM-metric values for the eight programmed levels as a function of the SET/RESET endurance cycles measured at room temperature. The solid and dotted lines represent the mean values and $\pm 3\sigma$ variations of the eM-metric values, respectively.

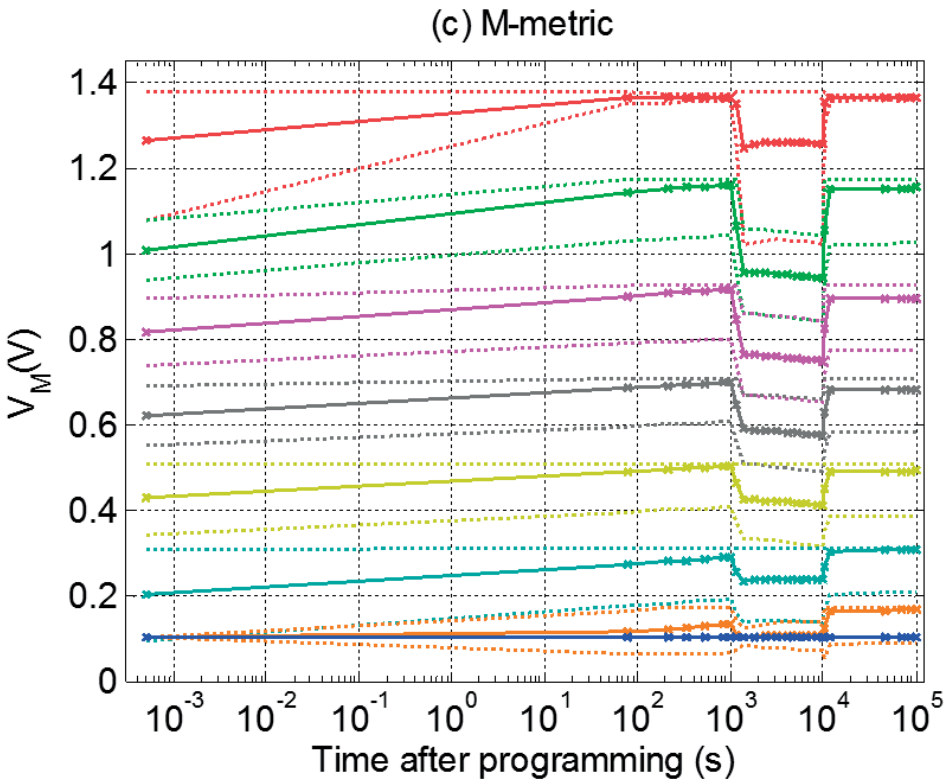
5.5.3 TLC data retention analysis

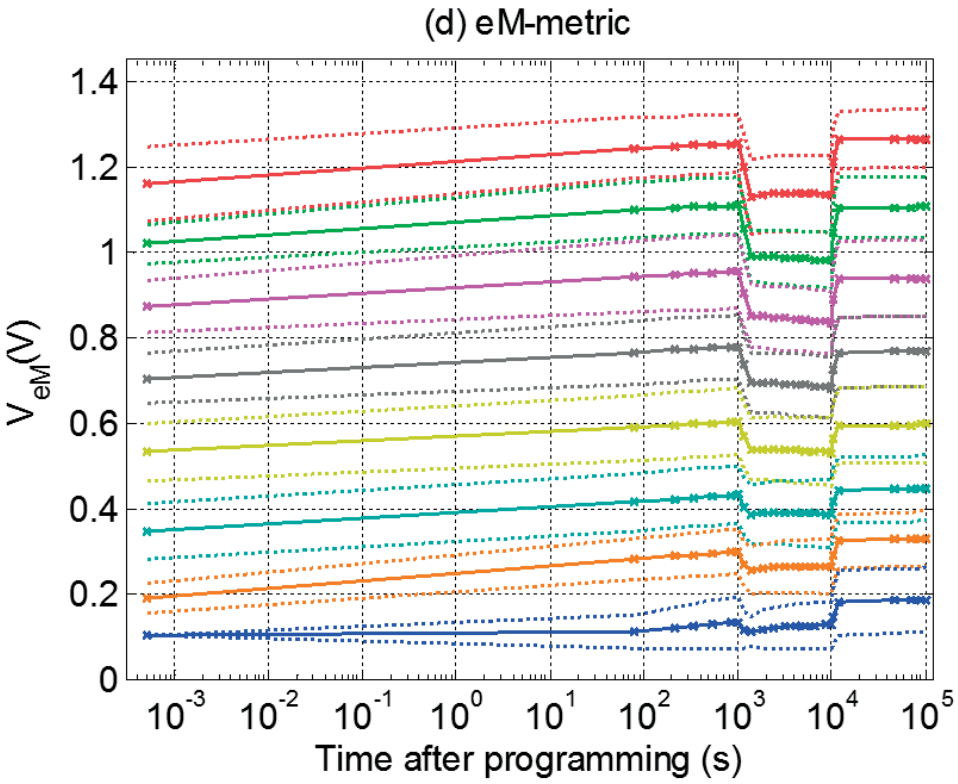
In Fig. 5.16, we compare the drift performance of the three metrics (R, M and eM) in terms of their feasibility for reliable TLC storage and data retention. With level overlap at the high-resistance states, the R-metric appears unsuitable for TLC storage. Although the M-metric is able to capture all programmed levels, the constant low-detection current leads to overlap between the SET and the intermediate states closer to the SET state.

In contrast, the eM-metric exhibits good potential for reliable detection of all eight levels over time. Even though the dynamic range of the eM-metric is reduced compared with the M-metric (compressed in the high-resistive states), the eM-metric exhibits remarkable potential for data retention under temperature variations for TLC storage because of the higher contrast it offers between the SET and the low-resistive intermediate states, and its higher SNR.

Figure 5.16: (a) Time-temperature profile of the temperature variations across the chip. Trajectories of the eight individually optimally placed levels for (b) the R-metric; (c) the M-metric and (d) the eM-metric, measured at regular time intervals and temperatures corresponding to the time-temperature profile. The solid and dotted lines represent the mean values and $\pm 3\sigma$ bounds of the respective metrics. Clearly, the performance of eM-metric is significantly better than the other two metrics. Owing to the constant low current in the M-metric, there is not enough contrast between the lower SET-like states.




(c) M-metric
1.4
1.2
1
0.8
0.6
0.4
0.2
0
V_M(V)
10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4} 10^{5}
Time after programming (s)


(d) eM-metric
1.4
1.2
1
0.8
0.6
0.4
0.2
0
V_{eM}(V)
10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4} 10^{5}
Time after programming (s)

5.5.4 BER performance

The BER performance of eM-metric readout for 3 bits/cell as a function of time after programming is shown in Fig. 5.17. The variation in the temperature across the device is shown in Fig. 5.16(a). The BER (using the PM code detector) remains at very acceptable levels (around 10^{-4}) until about 5000 s after programming, in the middle of the 80 °C stretch. Afterwards, the BER increases monotonically, as expected from the compression of the level placement shown in Fig. 5.16(d). However, the BER stabilizes again after 10^4 s, i.e., when the temperature drops again to 30 °C. This suggests that at room temperature drift does not affect TLC retention with the eM-metric. The BER performance of the reference-cell scheme is unacceptable, already exceeding 10^{-2} only 1 ms after level programming. Figure 5.18 illustrates the eM-metric distribution of the eight programmed levels drifting over time.

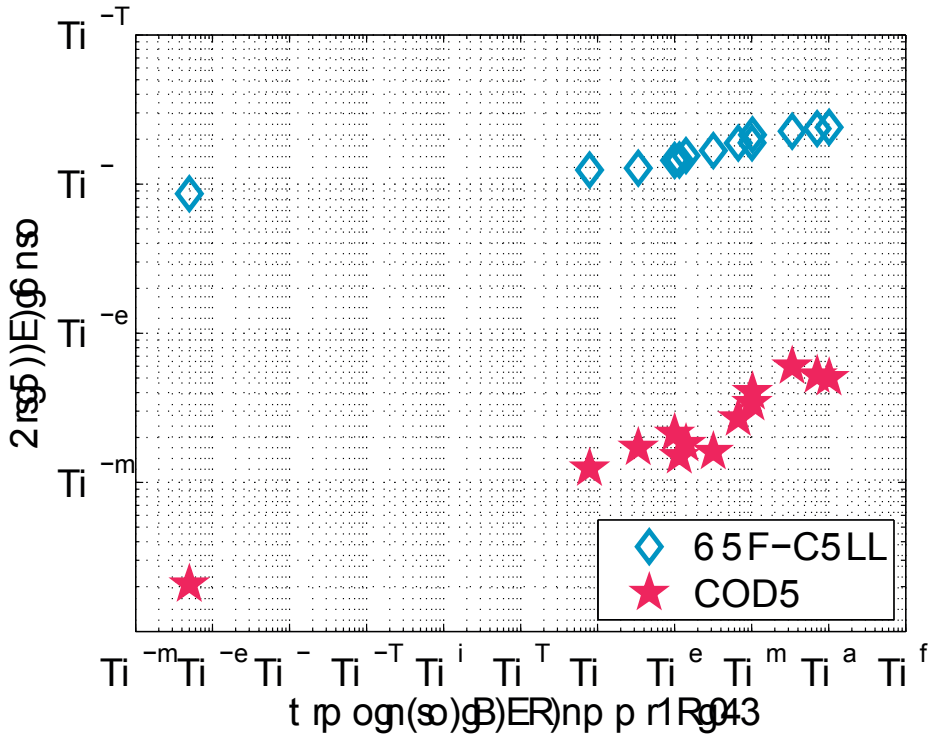


Figure 5.17: BER performance of the eM-metric readout for TLC storage in PCM after 10^6 endurance cycles.

To summarize, it appears that 3 bits/cell storage and moderate data retention in PCM are well within reach, in particular if ambient temperature fluctuations can be controlled. This is only possible with a judicious combination of the eM-metric, with its robustness to drift and temperature fluctuations, and the proposed modulation coding scheme, offering enhanced resiliency to level variations.

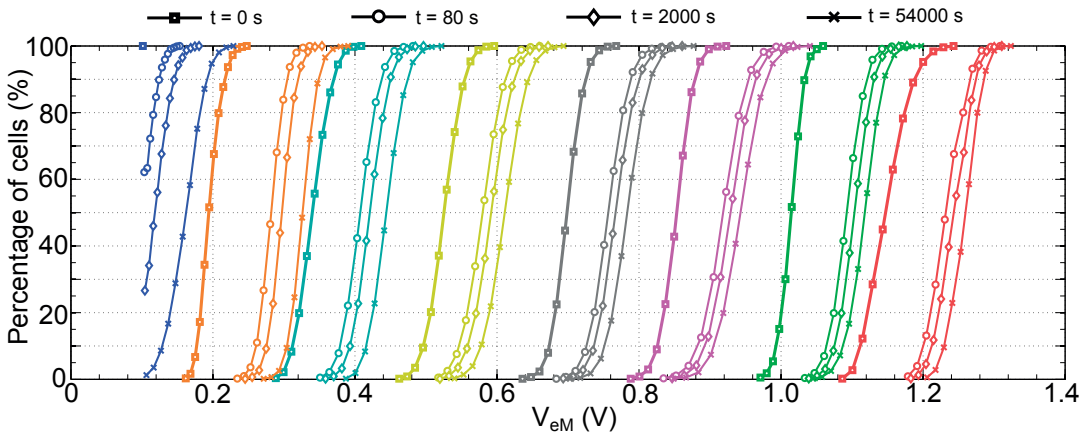


Figure 5.18: Drift evolution and distribution of the eight programmed levels after programming. There is a clear distinction between the adjacent levels in terms of V_{eM} and hence its more suitable to TLC storage operation.

5.6 Summary

Phase-change memory exhibits a large resistance margin between its two stable states. Moreover, this margin is constantly maintained or even marginally increased after repeated endurance cycling. Although this looks promising for achieving reliable MLC storage in PCM, the drift and variability associated with PCM cells pose a serious challenge for reliable MLC storage. However, we demonstrated that a powerful combination of drift-immune non-resistance-based metrics and drift-tolerant coding and detection schemes enable reliable storage of four levels per cell (2 bits/cell), with adequate data retention at elevated temperature, even after 10^6 endurance cycles. We also showed that it is even possible to reliably store and retrieve 3 bits/cell in PCM arrays, provided the temperature variations can be controlled.

Contributions:

- **Investigated and also demonstrated reliable MLC storage over a PCM sub-array with an ensemble of 64k cells, which had been subjected to one million endurance cycles and temperature variations.**
- **Devised the temperature control component of the experimental characterization platform for inducing temperature variations across the prototype chip.**
- **Implemented various routines and procedures in the FPGA-based hardware characterization platform for the emulation and extraction of various performance-based aspects of the prototype chip.**
- **Demonstrated the feasibility of storing 3 bits/cell in PCM arrays by optimal level placement, provided that temperature variations can be kept under control.**

6 Conclusion and future work

From a modeling perspective, all important aspects of FEM based modeling approach of thermoelectric physics within PCM devices have been thoroughly investigated by this thesis. Similarly, from a reliability perspective, a complete analysis has been performed experimentally on arrays of PCM cells (64k cells), that has been cycled over a million times and have experienced a history of temperature variations (30°C to 80°C). The results demonstrate the feasibility of MLC PCM chips commercialization in the near future. In this final chapter, a brief summary of thesis major contributions, the impact of this work in the PCM research community, and most promising future directions of research are presented.

6.1 Summary of Achievements

The accomplishments of this thesis include a complete modeling and reliability framework for MLC PCM technology. The first part of the thesis focused on the various device modeling aspects in the implementation of a comprehensive thermoelectric model that can accurately capture the PCM device characteristics. The second part of the thesis focused on a novel CMOS circuit implementations of new drift-resilient readout metrics in the 64 nm technology node. Furthermore, the reliability of MLC and TLC PCM has been thoroughly investigated and demonstrated in the framework of the proposed readout metrics, as well as novel drift-tolerant coding schemes.

Device modeling

1. **A comprehensive thermoelectric model:** I proposed a comprehensive thermoelectric model for the complete understanding of thermoelectric effects in PCM devices. Specifically, the thermoelectric physics of the nano-scale PCM devices has been modeled and simulated to gain a thorough knowledge of the thermal and electrical transport mechanisms within the device. Novel strategies were used to model and implement various physical aspects of the device in COMSOL Multiphysics® [**?**]. Realistic field- and temperature-dependent material properties were used in the model simulation. The

model was then validated by comparing the simulation results with experimental data measured on a wide range of ambient operating temperatures, from 200 K to 400 K. The simulation results showed good match with the experimental measurements, thereby verifying the efficacy of the model. The proposed model is instrumental for a comprehensive understanding of the device operation and dynamics, and hence provides valuable feedback for fine tuning the material properties and device design so as to enhance and improve the device efficiency and performance.

In addition, from the insights provided by the model, we were able to acquire the possible reason behind the distinct behavior exhibited by these devices for opposite-bias polarity conditions. This paves the way for future power-efficient device designs, where the thermoelectric contributions can be further enhanced by appropriate phase-change materials and electrodes selection or by modifying the device geometry.

I also proposed a simple and compact model to quickly assess the thermal disturb, with reasonably good accuracy, in highly dense PCM arrays. Given that these measurements are impossible to obtain from actual operating devices, the model offers an additional important benefit in providing reasonable estimates in the significantly smaller time frame (compared to the more complex and time consuming FEM model). The thermoelectric model for the thermal profile within the device and the electro-thermal model for the temperature distribution outside the device, both models when coupled together can give a very good estimate of the overall temperature distribution in the memory array. Given the importance of the knowledge of temperature distribution in technologies like PCM as it strongly impacts the area and power requirements, these models provide very valuable and much needed insights.

I also proposed novel device-design ideas targeting better performances compared to the conventional devices, which came as a result of the knowledge obtained from the model. Though these ideas are still very far from potential implementation, they point us to the right direction for addressing the problems of scalability and power issues and thereby making PCM technology more amenable.

Metrics and Architectures for MLC PCM

2. **Drift-resilient readout metrics in MLC PCM:** I propose a novel readout circuit architecture for the extraction of drift-resilient non-resistance cell-state-based metrics for reliable MLC storage in PCM. Resistance drift seriously limits the number of intermediate resistance states that can be packed reliably into a single cell. Various non-resistance-based drift-immune metrics have been shown to have better drift performance. However, most of these demonstrations were based on measurements from macro devices. A need for an efficient readout architecture which can extract this new breed of metrics and can be co-integrated with the array of PCM devices in a prototype chip clearly exist. The proposed novel readout scheme satisfies this purpose.

   The proposed architecture was implemented in 64 nm CMOS technology, and the

fabricated prototype chip was characterized with integrated poly-silicon resistor array, confirming the desired operation of the readout scheme. This part of the thesis was done in collaboration with SK-Hynix as part of a Joint-Development-Agreement (JDA) for MLC-PCRAM development and commercialization. The readout scheme presented demonstrated a fast and efficient extraction of drift-resilient metric suited for MLC operation, enabling, for the first time, the performance required by non-volatile memory applications. A readout access time of 450 ns at 6-bit raw (5-bit effective) resolution was achieved. The circuit exhibits low-noise characteristics and no sensitivity to bit-line parasitics. The readout chain is ready for co-integration with a multi-gigabit 2x nm PCM cell array together with the necessary programming electronics.

Being an exploratory prototype chip and an eventual platform for exploring such novel readout metrics, various flexibility knobs were added to the readout architecture for a large degree of extensive investigation of these metric schemes. Although this flexibilities made the implementation very complex, the readout scheme are versatile enough so that many possible combinations of the metrics can be extracted using the circuitry. Apart from that, I was also instrumental in the implementation of the enhanced version of the drift-resilient metric, which was demonstrated to have even better drift performance.

Reliability analysis of MLC PCM

3. **Reliability analysis of PCM arrays:** I demonstrated reliable MLC storage and data-retention based on various MLC readout metrics on an ensemble of 64k cells PCM arrays, which were cycled for a million (10^6) endurance cycles and have experienced a history of temperature variations ranging from 30°C to 80°C. Various technological advancements are currently being deployed to enable reliable MLC storage in PCM technology. It is of a great importance that their performance is thoroughly studied and investigated. Given the temperature-dependence of resistance drift, temperature variations and repeated programming and erase cycles can have an adverse impact on data retention, thereby preventing reliable MLC storage.

   Time-temperature drift evolution of the various readout metrics were presented and their data retention analyzed. The impact of endurance on the programming $eM-I$ curves and the evolution of the intermediate resistive-levels with respect to endurance cycling were presented, for the first time. Permutation codes were used to encode the data and iterative programming schemes to store the encoded data in the PCM sub-array. For the data retention analysis, the stored information is repeatedly read back at regular intervals of time while the chip undergoes temperature variations. The performance of various readout metrics were compared and their bit-error rate (BER) analyzed. It has been demonstrated that with optimal level-placement, reliable MLC operation is achievable, with the performance of non-resistance based metrics being an order of magnitude better than that of the low-field electrical resistance metric.

   Finally, the feasibility of reliable 3-bits/cell (Triple-level cell, TLC) storage with a data

retention of approximately 1 day was also presented. The demonstration is the first of its kind and provides valuable insights into and information on the data retention behavior of the readout metrics in PCM arrays, which is necessary for establishing the viability of MLC PCM technology in the near future.

6.2 Summary of contributions in this thesis

I summarize the various contributions which I have made in the fulfillment of this thesis. In the preceding chapters, I have made the following contributions with respect to the modeling and reliability framework for MLC PCM:

1. **Device Modeling:**
    - In Chapter 2, I devised a comprehensive FEM based thermoelectric model for capturing the thermoelectric effects in nano-scale PCM devices.
    - Various physical aspects governing the device operation were developed and implemented in COMSOL Multiphysics®.
    - In Chapter 3, I conceived the compact electro-thermal model approach to quickly evaluate the temperature distribution within the PCM arrays, mainly focusing on the thermal disturb analysis
    - I proposed novel device-design based on the numerous insights I obtained from the thermoelectric model for PCM devices.
2. **Metrics and Architectures for MLC PCM:**
    - In Chapter 4, I proposed a novel readout architecture for the extraction of drift-resilient readout metrics in MLC PCM storage.
    - I optimized the readout circuit design presented for power and speed, while still maintaining the required flexibility for exploration of the non-resistance based metric schemes.
    - I characterized the prototype chip fabricated using a FPGA-based hardware characterization platform. The performance of the readout circuitry in terms of access time (450 ns) and the accuracy of the readout data (5-bit effective ADC resolution) were investigated.
    - I was instrumental in the design and implementation of the iterative programming algorithms for MLC programming.
    - I contributed significantly in the top-level design of the prototype chips in terms of the chip interface, design of the digital controller, array address decoders, various test point buffers, and output for monitoring the internal chip signals, etc.
    - I was involved in the circuit design implementation for the eM-metric extraction.

3. **Reliability analysis of MLC PCM:**

    - In Chapter 5, I investigated and also demonstrated reliable MLC storage over an ensemble of 64k cells PCM sub-array, which had been subjected to a million endurance cycles and temperature variations.
    - I devised the temperature control component of the experimental characterization platform for inducing temperature variations across the prototype chip.
    - I implemented various routines and procedures in the FPGA-based hardware characterization platform for the emulation and extraction of various performance based aspects of the prototype chip.
    - I also demonstrated the feasibility of 3 bits/cell in PCM arrays by optimal level placement, provided that temperature variations can be kept under control.

This thesis has been performed under the framework of the IBM-SK Hynix Joint-Development Agreement (JDA) for PCRAM commercialization. This project aims to develop the first commercial MLC PCM chip with multi-gigabit capacity for mass production. Novel CMOS circuit implementation of the READ and programming architectures for MLC operation and various aspects in designing the prototype chip are all the result of the close collaborative efforts between the design engineers at IBM Research in Zurich, IBM T. J. Watson Research and SK-Hynix. The devices and prototype chips used for the experimental measurements are all fabricated by our colleagues at the IBM T. J. Watson Research center.

6.3 Impact of this Research

Publications: The work done in this thesis has resulted in various research publications. The novel readout architecture proposed in Chapter 4 was presented at the *Asian Solid-State Circuits Conference (A-SSCC), Taiwan 2014* [Athmanathan et al., 2014]. The comprehensive thermoelectric model described in Chapter 2 and Chapter 3 was accepted and will be presented at the *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), USA 2015* [Athmanathan et al., 2015]. The MLC and TLC reliability and data retention demonstrations in Chapter 5 has been accepted for publications in the December 2015 issue of the *IEEE Journal on Emerging and Selected Topics in Circuits and Systems, to be published in Dec. 2015* [Athmanathan et al., 2015]. The readout scheme for M-metric extraction and the compact electro-thermal model for thermal disturb studies were well received at the *European\Phase Change and Ovonics Symposia in Zürich 2011 & Berlin 2013*, respectively. The work on thermoelectric modeling was presented as invited talk at the *International Workshop on Semiconductor Memory Modeling, Italy 2013*.

In addition to the above, many other research papers directly based on the contributions of this thesis have presented at major conference and published. For instance, I closely worked with my mentor Dr. Milos Stanisavljevic at IBM Research - Zurich in developing

the enhanced version of the drift-resilient metric for reliable storage and data retention at elevated temperatures in Chapter 5, which was presented at the *International Reliability Physics Symposium (IRPS), USA, 2015* [Stanisavljevic et al., 2015]. A slightly modified version of the readout architecture proposed in Chapter 4 was accepted for presentation at the *Custom Integrated Circuits Conference (CICC), USA, 2015* [Cheon et al., 2015].

Patents: A patent application was filed for the device design for thermoelectric heat confinement under the title *MEMORY DEVICE AND METHOD FOR THERMOELECTRIC HEAT CONFINEMENT*. The idea of MLC capability using thermoelectric was published as an IBM Research report.

Drift-resilient readout metrics: A special mention must be made about the impact of the work on the drift-resilient metrics in this thesis. The proposed novel readout circuit architectures for the extraction of the drift-immune readout metrics was used as the readout scheme for the multi-gigabit PCM chip integrated at 2x nm technology node with 64 nm technology for the programming and readout electronics. This is the first high density MLC PCM chip in the industry, fabricated at state-of-the-art technology node.

6.4 Future Work

The following research directions may be explored in the future to build upon the work performed during the course of this thesis:

1. **Enhancements to the thermoelectric model:** The proposed model is capable of capturing the efficiency of the device in terms of the heating contributions from the thermoelectric physics for any input power applied. The model can be enhanced further by coupling it with a crystallization model to model and capture the phase-change switching kinetics. The model can then be used to study the crystal growth velocity, re-crystallization, etc.

2. **Device design and material study:** The proposed model can be used to study the performance and characteristics of various device topologies and different combinations of phase-change materials and electrodes, which can be employed to have PCM devices with more desirable features. As presented in the section on the novel device-design ideas, different electrode types and shapes can be explored to increase the significance of thermoelectric physics in an effort to reduce the power required for the RESET operation, thereby achieving more power-efficient devices.

3. **Design optimization for area and power:** The proposed readout scheme is capable of the extraction drift-resilient metrics. Given that the circuitry is designed for a prototype chip, the design was both flexible and complex, and therefore the critical design

components, like the area, power and speed, are not fully optimized. Once the desired features of the readout schemes are finalized based on the performance of the PCM devices, there might be still some scope for design optimization and improvement in terms of speed and power.

4. **Exploration of the M_{diff} metric:** The proposed circuitry offers a flexibility that different sets of drift-resilient readout metrics can be extracted using this architecture. One particular drift resilient metric is the differential variant of the M-metric. Owing to the delay in the fabrication process of PCM devices at the 2x nm technology, the characterization results presented are only based on the integrated on-chip poly-resistors. Once the devices fabricated, the same set of experiments should be repeated to measure the PCM device characteristics, including the performance of the M_{diff} metric.

5. **Improving TLC storage by temperature compensation:** It can be observed that reliable TLC storage is well within the reach of the proposed eM-metric scheme. However, there exists a strong temperature-dependence of these metrics, and this might be a concern for closely packed intermediate levels. One way of reducing the fluctuations in the readout metric is by including a temperature-compensation scheme in the readout architecture. By sensing the current temperature of the chip during readout, the detection thresholds can be modified, thus minimizing the corresponding variations in the readout metric values.

Bibliography

D. Adler, M. S. Shur, M. Silver, and S. R. Ovshinsky. Threshold switching in chalcogenide-glass thin films. *Journal of Applied Physics*, 51(6):3289–3309, 1980. doi: 10.1063/1.328036.

A. Athmanathan, M. Stanisavljevic, J. Cheon, S. Kang, C. Ahn, J. Yoon, M. Shin, T. Kim, N. Papandreou, H. Pozidis, and E. Eleftheriou. A 6-bit drift-resilient readout scheme for multi-level phase-change memory. In *IEEE Asian Solid-State Circuits Conference (A-SSCC)*, pages 137–140, Nov 2014. doi: 10.1109/ASSCC.2014.7008879.

A. Athmanathan, D. Krebs, A. Sebastian, M. Le Gallo, H. Pozidis, and E. Eleftheriou. A finite-element model for thermoelectric effects in phase-change memory devices. In *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, page to be published, Sep 2015.

A. Athmanathan, M. Stanisavljevic, N. Papandreou, H. Pozidis, and E. Eleftheriou. Multilevel-cells phase-change memory - a viable technology. In *IEEE Journal on Emerging and Selected Topics in Circuits and Systems (JETCAS)*, page to be published, Dec 2015.

I.G. Baek, M.S. Lee, S. Seo, M.-J. Lee, D.H. Seo, D.-S. Suh, J.C. Park, S.O. Park, T.I. Kim, I.K. Yoo, U-in Chung, and J.T. Moon. Highly scalable nonvolatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses. In *Electron Devices Meeting, 2004. IEDM Technical Digest. IEEE International*, pages 587–590, Dec 2004. doi: 10.1109/IEDM.2004.1419228.

J. Battaglia, A. Kusiak, A. Saci, R. Fallica, A. Lamperti, and C. Wiemer. Effect of a thin ti interfacial layer on the thermal resistance of ge2sb2te5-tin stack. *Applied Physics Letters*, 105(12):121903, 2014. doi: http://dx.doi.org/10.1063/1.4896325.

J.-L. Battaglia, A. Kusiak, V. Schick, A. Cappella, C. Wiemer, M. Longo, and E. Varesi. Thermal characterization of the sio2-ge2sb2te5 interface from room temperature up to 400°c. *Journal of Applied Physics*, 107(4):044314, 2010. doi: http://dx.doi.org/10.1063/1.3284084.

A. Beck, J. G. Bednorz, Ch. Gerber, C. Rossel, and D. Widmer. Reproducible switching effect in thin oxide films for memory applications. *Applied Physics Letters*, 77(1):139–141, 2000. doi: http://dx.doi.org/10.1063/1.126902.

F. Bedeschi, R. Fackenthal, C. Resta, E.M. Donze, M. Jagasivamani, E.C. Buda, F. Pellizzer, D.W. Chow, A. Cabrini, G. Calvi, R. Faravelli, A. Fantini, G. Torelli, D. Mills, R. Gastaldi, and G. Casagrande. A bipolar-selected phase change memory featuring multi-level cell storage. *IEEE Journal of Solid-State Circuits*, 44(1):217–227, Jan 2009. ISSN 0018-9200. doi: 10.1109/JSSC.2008.2006439.

F. Bedeschi, R. Fackenthal, C. Resta, E.M. Donze, M. Jagasivamani, E.C. Buda, F. Pellizzer, D.W. Chow, A. Cabrini, G. Calvi, R. Faravelli, A. Fantini, G. Torelli, D. Mills, R. Gastaldi, and G. Casagrande. A bipolar-selected phase change memory featuring multi-level cell storage. *Solid-State Circuits, IEEE Journal of*, 44(1):217–227, Jan 2009. ISSN 0018-9200. doi: 10.1109/JSSC.2008.2006439.

A. Bette, J. DeBrosse, D. Gogl, H. Hoenigschmid, R. Robertazzi, C. Arndt, D. Braun, D. Casarotto, R. Havreluk, S. Lammers, W. Obermaier, W. Reohr, H. Viehmann, W.J. Gallagher, and G. Muller. A high-speed 128 kbit mram core for future universal memory applications. In *VLSI Circuits, 2003. Digest of Technical Papers. 2003 Symposium on*, pages 217–220, June 2003. doi: 10.1109/VLSIC.2003.1221207.

R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti. Introduction to flash memory. *Proceedings of the IEEE*, 91(4):489–502, April 2003. ISSN 0018-9219. doi: 10.1109/JPROC.2003.811702.

M. Boniardi, D. Ielmini, S. Lavizzari, A.L. Lacaita, A. Redaelli, and A. Pirovano. Statistics of resistance drift due to structural relaxation in phase-change memory arrays. *Electron Devices, IEEE Transactions on*, 57(10):2690–2696, Oct 2010. ISSN 0018-9383. doi: 10.1109/TED.2010.2058771.

M. Breitwisch, T. Nirschl, C.F. Chen, Y. Zhu, M-H Lee, M. Lamorey, G.W. Burr, E. Joseph, A. Schrott, J.B. Philipp, R. Cheek, T.D. Happ, S.-H. Chen, S. Zaidi, P. Flaitz, J. Bruley, R. Dasaka, B. Rajendran, S. Rossnage, M. Yang, Y.C. Chen, R. Bergmann, H. L Lung, and C. Lam. Novel lithography-independent pore phase change memory. In *IEEE Symposium on VLSI Technology*, pages 100–101, June 2007. doi: 10.1109/VLSIT.2007.4339743.

G. Bruns, P. Merkelbach, C. Schlockermann, M. Salinga, M. Wuttig, T. D. Happ, J. B. Philipp, and M. Kund. Nanosecond switching in gete phase change memory cells. *Applied Physics Letters*, 95(4):043108, 2009. doi: http://dx.doi.org/10.1063/1.3191670.

Geoffrey W. Burr, M. J. Breitwisch, M. Franceschini, D. Garetto, K. Gopalakrishnan, B. Jackson, B. Kurdi, C. Lam, L. A. Lastras, A. Padilla, B. Rajendran, S. Raoux, and R. S. Shenoy. Phase change memory technology. *Journal of Vacuum Science & Technology B*, 28(2):223–262, 2010. doi: http://dx.doi.org/10.1116/1.3301579. URL http://scitation.aip.org/content/avs/journal/jvstb/28/2/10.1116/1.3301579.

D.T. Castro, L. Goux, G.A.M. Hurkx, K. Attenborough, R. Delhougne, J. Lisoni, F.J. Jedema, M.A.A. 't Zandt, R.A.M. Wolters, D.J. Gravesteijn, M. Verheijen, M. Kaiser, and R.G.R. Weemaes. Evidence of the thermo-electric thomson effect and influence on the program conditions

and cell optimization in phase-change memory cells. In *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pages 315–318, Dec 2007. doi: 10.1109/IEDM.2007.4418934.

Chieh-Fang Chen, A. Schrott, M.H. Lee, S. Raoux, Y.H. Shih, M. Breitwisch, F.H. Baumann, E.K. Lai, T.M. Shaw, P. Flaitz, R. Cheek, E.A. Joseph, S.-H. Chen, B. Rajendran, H.L. Lung, and C. Lam. Endurance improvement of ge2sb2te5-based phase change memory. In *Memory Workshop, 2009. IMW '09. IEEE International*, pages 1–2, May 2009. doi: 10.1109/IMW.2009.5090589.

E. Chen, D. Apalkov, Z. Diao, A. Driskill-Smith, D. Druist, D. Lottis, V. Nikitin, X. Tang, S. Watts, S. Wang, S.A. Wolf, A.W. Ghosh, J.W. Lu, S.J. Poon, M. Stan, W.H. Butler, S. Gupta, C.K.A. Mewes, T. Mewes, and P.B. Visscher. Advances and future prospects of spin-transfer torque random access memory. *Magnetics, IEEE Transactions on*, 46(6):1873–1878, June 2010. ISSN 0018-9464. doi: 10.1109/TMAG.2010.2042041.

J. Cheon, I. Lee, C. Ahn, M. Stanisavljevic, A. Athmanathan, N. Papandreou, H. Pozidis, E. Eleftheriou, M. Shin, T. Kim, J. Kang, and J. Chun. Non-resistance metric based read scheme for multi-level pcram in 25nm technology. In *IEEE Custom Integrated Circuits Conference (CICC)*, page to be published, Sep 2015.

E. Choi and S. Park. Device considerations for high density and highly reliable 3d nand flash cell in near future. In *Electron Devices Meeting (IEDM), 2012 IEEE International*, pages 9.4.1–9.4.4, Dec 2012. doi: 10.1109/IEDM.2012.6479011.

N. Ciocchini, M. Laudato, A. Leone, P. Fantini, A.L. Lacaita, and D. Ielmini. Universal thermoelectric characteristic in phase change memories. In *Memory Workshop (IMW), 2015 IEEE International*, pages 1–4, May 2015. doi: 10.1109/IMW.2015.7150311.

G. F. Close, U. Frey, M. Breitwisch, H. L Lung, C. Lam, C. Hagleitner, and E. Eleftheriou. Device, circuit and system-level analysis of noise in multi-bit phase-change memory. In *Electron Devices Meeting (IEDM), 2010 IEEE International*, pages 29.5.1–29.5.4, Dec 2010. doi: 10.1109/IEDM.2010.5703445.

G.F. Close, U. Frey, J. Morrish, R. Jordan, S. Lewis, T. Maffitt, M. BrightSky, C. Hagleitner, C. Lam, and E. Eleftheriou. A 256-mcell phase-change memory chip operating at 2+ bit/cell. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 60(6):1521–1533, June 2013. ISSN 1549-8328. doi: 10.1109/TCSI.2012.2220459.

T.M. Cover. Enumerative source encoding. *IEEE Transactions on Information Theory*, 19(1): 73–77, 1973.

S. Dong-Seok, Kim Cheolkyu, Kijoon H. P. Kim, K. Youn-Seon, L. Tae-Yon, K. Yoonho, P. Tae Sang, Y. Young-Gui, Im Jino, and I. Jisoon. Thermoelectric heating of ge2sb2te5 in phase change memory devices. *Applied Physics Letters*, 96(12):123115, 2010. doi: http://dx.doi.org/10.1063/1.3259649.

S. Eilert, M. Leinwander, and G. Crisenza. Phase change memory: A new memory enables new memory usage models. In *Memory Workshop, 2009. IMW '09. IEEE International*, pages 1–2, May 2009. doi: 10.1109/IMW.2009.5090604.

R. Fackenthal, M. Kitagawa, W. Otsuka, K. Prall, D. Mills, K. Tsutsui, J. Javanifard, K. Tedrow, T. Tsushima, Y. Shibahara, and G. Hush. 19.7 a 16gb reram with 200mb/s write and 1gb/s read in 27nm technology. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, pages 338–339, Feb 2014. doi: 10.1109/ISSCC.2014.6757460.

A. Faraclas, G. Bakan, L. Adnane, F. Dirisaglik, N.E. Williams, A. Gokirmak, and H. Silva. Modeling of thermoelectric effects in phase change memory cells. *IEEE Transactions on Electron Devices*, 61(2):372–378, Feb 2014. ISSN 0018-9383. doi: 10.1109/TED.2013.2296305.

D.J. Frank, R.H. Dennard, E. Nowak, P.M. Solomon, Yuan Taur, and Hen-Sum Philip Wong. Device scaling limits of si mosfets and their application dependencies. *Proceedings of the IEEE*, 89(3):259–288, Mar 2001. ISSN 0018-9219. doi: 10.1109/5.915374.

M. Frehner, S. M. Schmalholz, E. H. Saenger, and H. Steeb. Comparison of finite difference and finite element methods for simulating two-dimensional scattering of elastic waves. *Physics of the Earth and Planetary Interiors*, 171(1–4):112 – 121, 2008. ISSN 0031-9201. doi: http://dx.doi.org/10.1016/j.pepi.2008.07.003. Recent Advances in Computational Geodynamics: Theory, Numerics and Applications.

R.F. Freitas and W.W. Wilcke. Storage-class memory: The next storage system technology. *IBM Journal of Research and Development*, 52(4.5):439–447, July 2008. ISSN 0018-8646. doi: 10.1147/rd.524.0439.

J. Frenkel. On pre-breakdown phenomena in insulators and electronic semi-conductors. *Phys. Rev.*, 54:647–648, Oct 1938. doi: 10.1103/PhysRev.54.647.

D. Fugazza, D. Ielmini, G. Montemurro, and A.L. Lacaita. Temperature- and time-dependent conduction controlled by activation energy in pcm. In *Electron Devices Meeting (IEDM), 2010 IEEE International*, pages 29.3.1–29.3.4, Dec 2010. doi: 10.1109/IEDM.2010.5703443.

B. Gleixner, A. Pirovano, J. Sarkar, F. Ottogalli, E. Tortorelli, M. Tosi, and R. Bez. Data retention characterization of phase-change memory arrays. In *IEEE International Reliability Physics Symposium (IRPS)*, pages 542–546, April 2007. doi: 10.1109/RELPHY.2007.369948.

A. Goda. Opportunities and challenges of 3d nand scaling. In *VLSI Technology, Systems, and Applications (VLSI-TSA), 2013 International Symposium on*, pages 1–2, April 2013. doi: 10.1109/VLSI-TSA.2013.6545625.

A. Goda. Recent progress and future directions in nand flash scaling. In *Non-Volatile Memory Technology Symposium (NVMTS), 2013 13th*, pages 1–4, Aug 2013. doi: 10.1109/NVMTS.2013.6851055.

Robert M. Hill. Poole-frenkel conduction in amorphous solids. *Philosophical Magazine*, 23(181):59–86, 1971. doi: 10.1080/14786437108216365. URL http://dx.doi.org/10.1080/14786437108216365.

D. Ielmini. Threshold switching mechanism by high-field energy gain in the hopping transport of chalcogenide glasses. *Phys. Rev. B*, 78:035308, Jul 2008. doi: 10.1103/PhysRevB.78.035308.

D. Ielmini and Y. Zhang. Analytical model for subthreshold conduction and threshold switching in chalcogenide-based memory devices. *Journal of Applied Physics*, 102(5):054517, 2007. doi: http://dx.doi.org/10.1063/1.2773688.

D. Ielmini, A.L. Lacaita, and D. Mantegazza. Recovery and drift dynamics of resistance and threshold voltages in phase-change memories. *Electron Devices, IEEE Transactions on*, 54 (2):308–315, Feb 2007. ISSN 0018-9383. doi: 10.1109/TED.2006.888752.

D. Ielmini, S. Lavizzari, D. Sharma, and A.L. Lacaita. Physical interpretation, modeling and impact on phase change memory (pcm) reliability of resistance drift due to chalcogenide structural relaxation. In *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pages 939–942, Dec 2007. doi: 10.1109/IEDM.2007.4419107.

D. Ielmini, S. Lavizzari, D. Sharma, and A. L. Lacaita. Temperature acceleration of structural relaxation in amorphous ge2sb2te5. *Applied Physics Letters*, 92(19):193511, 2008. doi: http://dx.doi.org/10.1063/1.2930680.

D. Ielmini, D. Sharma, S. Lavizzari, and A.L. Lacaita. Reliability impact of chalcogenide-structure relaxation in phase-change memory (PCM) cells—part i: Experimental study. *IEEE Transactions on Electron Devices*, 56(5):1070–1077, 2009. doi: 10.1109/TED.2009.2016397. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4804709.

D-H. Kang, J.-H. Lee, J.H. Kong, D. Ha, J. Yu, C.Y. Um, J.H. Park, F. Yeung, J.H. Kim, W.I. Park, Y.J. Jeon, M.K. Lee, J.H. Park, Y.J. Song, J.H. Oh, H.S. Jeong, and H.S. Jeong. Two-bit cell operation in diode-switch phase change memory cells with 90nm technology. In *VLSI Technology, 2008 Symposium on*, pages 98–99, June 2008. doi: 10.1109/VLSIT.2008.4588577.

I.V. Karpov and S.A. Kostylev. Set to reset programming in phase change memories. *Electron Device Letters, IEEE*, 27(10):808–810, Oct 2006. ISSN 0741-3106. doi: 10.1109/LED.2006.882527.

D.L. Kencke, I.V. Karpov, B.G. Johnson, Sean Jong Lee, DerChang Kau, S.J. Hudgens, John P. Reifenberg, S.D. Savransky, Jingyan Zhang, M.D. Giles, and G. Spadini. The role of interfaces in damascene phase-change memory. In *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pages 323–326, Dec 2007. doi: 10.1109/IEDM.2007.4418936.

I. S. Kim et al. High performance PRAM cell scalable to sub-20nm technology with below $4F^2$ cell size, extendable to dram applications. In *VLSI Technology (VLSIT), 2010 Symposium on*, pages 203–204, 2010. doi: 10.1109/VLSIT.2010.5556228.

K. Kim and S. J. Ahn. Reliability investigations for manufacturable high density pram. In *Reliability Physics Symposium, 2005. Proceedings. 43rd Annual. 2005 IEEE International*, pages 157–162, April 2005. doi: 10.1109/RELPHY.2005.1493077.

K. Kim, J. H. Choi, J. Choi, and H. Jeong. The future prospect of nonvolatile memory. In *VLSI Technology, 2005. (VLSI-TSA-Tech). 2005 IEEE VLSI-TSA International Symposium on*, pages 88–94, April 2005. doi: 10.1109/VTSA.2005.1497091.

S. Kim, N. Sosa, M. BrightSky, D. Mori, W. Kim, Y. Zhu, K. Suu, and C. Lam. A phase change memory cell with metallic surfactant layer as a resistance drift stabilizer. In *IEEE International Electron Devices Meeting (IEDM)*, pages 30.7.1–30.7.4, Dec 2013. doi: 10.1109/IEDM.2013.6724727.

S. Kim, N. Sosa, M. BrightSky, D. Mori, W. Kim, Y. Zhu, K. Suu, and C. Lam. A phase change memory cell with metallic surfactant layer as a resistance drift stabilizer. In *Electron Devices Meeting (IEDM), 2013 IEEE International*, pages 30.7.1–30.7.4, Dec 2013. doi: 10.1109/IEDM.2013.6724727.

S. B. Kim, B. Lee, M. Asheghi, G.A.M. Hurkx, J. Reifenberg, K. Goodson, and H.-S.P. Wong. Thermal disturbance and its impact on reliability of phase-change memory studied by the micro-thermal stage. In *Reliability Physics Symposium (IRPS), 2010 IEEE International*, pages 99–103, May 2010. doi: 10.1109/IRPS.2010.5488847.

Y. Kim, J. Yun, S. H. Park, W. Kim, J. Y. Seo, M. Kang, K. Ryoo, J. Oh, J. Lee, H. Shin, and B. Park. Three-dimensional nand flash architecture design based on single-crystalline stacked array. *Electron Devices, IEEE Transactions on*, 59(1):35–45, Jan 2012. ISSN 0018-9383. doi: 10.1109/TED.2011.2170841.

S.A. Kostylev, W. Czubatyj, and T. Lowrey. Method of eliminating drift in phase-change memory, July 2005. US Patent 6,914,801.

D. Krebs, S. Raoux, C. T. Rettner, R. M. Shelby, G. W. Burr, and M. Wuttig. Set characteristics of phase change bridge devices. In *Symposium G – Phase-Change Materials for Reconfigurable Electronics and Memory Applications*, volume 1072 of *MRS Proceedings*, 2008. doi: 10.1557/PROC-1072-G06-07. URL http://journals.cambridge.org/article_S1946427400030438.

D. Krebs, S. Raoux, C. T. Rettner, G. W. Burr, M. Salinga, and M. Wuttig. Threshold field of phase change memory materials measured using phase change bridge devices. *Applied Physics Letters*, 95(8):082101, 2009. doi: http://dx.doi.org/10.1063/1.3210792.

A.L. Lacaita, A. Redaelli, D. Ielmini, F. Pellizzer, A. Pirovano, A. Benvenuti, and R. Bez. Electrothermal and phase-change dynamics in chalcogenide-based memories. In *Electron Devices Meeting, 2004. IEDM Technical Digest. IEEE International*, pages 911–914, Dec 2004. doi: 10.1109/IEDM.2004.1419330.

H.Y. Lee, P.S. Chen, T.Y. Wu, Y.S. Chen, C.C. Wang, P.J. Tzeng, C.H. Lin, F. Chen, C.H. Lien, and M.J. Tsai. Low power and high speed bipolar switching with a thin reactive ti buffer layer in

robust hfo2 based rram. In *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, pages 1–4, Dec 2008. doi: 10.1109/IEDM.2008.4796677.

H.Y. Lee, Y.S. Chen, P.S. Chen, P.Y. Gu, Y.Y. Hsu, S.M. Wang, W.H. Liu, C.H. Tsai, S.S. Sheu, P.-C. Chiang, W.P. Lin, C.H. Lin, W.S. Chen, F.T. Chen, C.H. Lien, and M. Tsai. Evidence and solution of over-reset problem for hfox based resistive memory with sub-ns switching speed and high endurance. In *Electron Devices Meeting (IEDM), 2010 IEEE International*, pages 19.7.1–19.7.4, Dec 2010. doi: 10.1109/IEDM.2010.5703395.

J. Lee, M. Asheghi, and K. E. Goodson. Impact of thermoelectric phenomena on phase-change memory performance metrics and scaling. *Nanotechnology*, 23(20):205201, 2012.

J. Lee, T. Kodama, Y. Won, M. Asheghi, and K. E. Goodson. Phase purity and the thermoelectric properties of $ge_2sb_2te_5$ films down to 25 nm thickness. *Journal of Applied Physics*, 112(1): 014902, 2012. doi: http://dx.doi.org/10.1063/1.4731252.

J. Lee, E. Bozorg-Grayeli, S. B. Kim, M. Asheghi, H.-S. Philip Wong, and K. E. Goodson. Phonon and electron transport through ge2sb2te5 films and interfaces bounded by metals. *Applied Physics Letters*, 102(19):191911, 2013. doi: http://dx.doi.org/10.1063/1.4807141.

M. Lee, S. Han, S. Jeon, B. Park, B. Kang, S. Ahn, K. Kim, C. Lee, C. Kim, I. Yoo, D. H. Seo, X. Li, J. Park, J. Lee, and Y. Park. Electrical manipulation of nanofilaments in transition-metal oxides for resistance-based memory. *Nano Letters*, 9(4):1476–1481, 2009. doi: 10.1021/nl803387q. URL http://dx.doi.org/10.1021/nl803387q. PMID: 19296606.

B. Liu, T. Zhang, J. Xia, Z. Song, S. Feng, and B. Chen. Nitrogen-implanted ge 2 sb 2 te 5 film used as multilevel storage media for phase change random access memory. *Semiconductor Science and Technology*, 19(6):L61, 2004.

M.J. Marinella. Emerging resistive switching memory technologies: Overview and current status. In *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on*, pages 830–833, June 2014. doi: 10.1109/ISCAS.2014.6865264.

N. Matsuzaki, K. Kurotsuchi, Y. Matsui, O. Tonomura, N. Yamamoto, Y. Fujisaki, N. Kitai, R. Takemura, K. Osada, S. Hanzawa, H. Moriya, T. Iwasaki, T. Kawahara, N. Takaura, M. Terao, M. Matsuoka, and M. Moniwa. Oxygen-doped gesbte phase-change memory cells featuring 1.5 v/100-/spl mu/a standard 0.13/spl mu/m cmos operations. In *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, pages 738–741, Dec 2005. doi: 10.1109/IEDM.2005.1609459.

J. Meena, S. Sze, U. Chand, and T. Tseng. Overview of emerging nonvolatile memory technologies. *Nanoscale Research Letters*, 9(1):526, 2014. ISSN 1931-7573. doi: 10.1186/1556-276X-9-526.

T.S. Moise, S.R. Summerfelt, H. McAdams, S. Aggarwal, K.R. Udayakumar, F.G. Celii, J.S. Martin, G. Xing, L. Hall, K.J. Taylor, T. Hurd, J. Rodriguez, K. Remack, M.D. Khan, K. Boku, G. Stacey,

M. Yao, M.G. Albrecht, E. Zielinski, M. Thakre, S. Kuchimanchi, A. Thomas, B. McKee, J. Rickes, A. Wang, J. Grace, J. Fong, D. Lee, C. Pietrzyk, R. Lanham, S.R. Gilbert, D. Taylor, J. Amano, R. Bailey, F. Chu, G. Fox, S. Sun, and T. Davenport. Demonstration of a 4 mb, high density ferroelectric memory embedded within a 130 nm, 5 lm cu/fsg logic process. In *Electron Devices Meeting, 2002. IEDM '02. International*, pages 535–538, Dec 2002. doi: 10.1109/IEDM.2002.1175897.

N. F. Mott and E. A. Davis. *Electronic Process in Non-Crystalline Materials*. Oxford University Press, 1967.

T. Nirschl, J.B. Philipp, T.D. Happ, G.W. Burr, B. Rajendran, M-H Lee, A. Schrott, M. Yang, M. Breitwisch, C.F. Chen, E. Joseph, M. Lamorey, R. Cheek, S.-H. Chen, S. Zaidi, S. Raoux, Y.C. Chen, Y. Zhu, R. Bergmann, H. L Lung, and C. Lam. Write strategies for 2 and 4-bit multi-level phase-change memory. In *IEEE International Electron Devices Meeting (IEDM)*, pages 461–464, Dec 2007. doi: 10.1109/IEDM.2007.4418973.

W. K. Njoroge, H. Wöltgens, and M. Wuttig. Density changes upon crystallization of ge2sb2.04te4.74 films. *Journal of Vacuum Science & Technology A*, 20(1):230–233, 2002. doi: http://dx.doi.org/10.1116/1.1430249.

J.H. Oh, J.H. Park, Y.S. Lim, H.S. Lim, Y.T. Oh, J.S. Kim, J.M. Shin, Y.J. Song, K.C. Ryoo, D.W. Lim, S.S. Park, J.I. Kim, J.H. Kim, J. Yu, F. Yeung, C.W. Jeong, J.H. Kong, D.H. Kang, G.H. Koh, G.T. Jeong, H.S. Jeong, and Kinam Kim. Full integration of highly manufacturable 512mb pram based on 90nm technology. In *IEEE International Electron Devices Meeting (IEDM)*, pages 1–4, Dec 2006. doi: 10.1109/IEDM.2006.346905.

S. R. Ovshinsky. Reversible electrical switching phenomena in disordered structures. *Phys. Rev. Lett.*, 21:1450–1453, Nov 1968. doi: 10.1103/PhysRevLett.21.1450.

A.E. Owen and J. M. Robertson. Electronic conduction and switching in chalcogenide glasses. *Electron Devices, IEEE Transactions on*, 20(2):105–122, Feb 1973. ISSN 0018-9383. doi: 10.1109/T-ED.1973.17617.

A.E. Owen, J.M. Robertson, and C. Main. The threshold characteristics of chalcogenide-glass memory switches. *Journal of Non-Crystalline Solids*, 32(1–3):29 – 52, 1979. ISSN 0022-3093. doi: http://dx.doi.org/10.1016/0022-3093(79)90063-2. Electronic Properties and Structure of Amorphous Solids.

N. Papandreou, H. Pozidis, T. Mittelholzer, G. F. Close, M. Breitwisch, C. Lam, and E. Eleftheriou. Drift-tolerant multilevel phase-change memory. In *IEEE International Memory Workshop (IMW)*, pages 1–4, May 2011. doi: 10.1109/IMW.2011.5873231.

N. Papandreou, H. Pozidis, A. Pantazi, A. Sebastian, M. Breitwisch, C. Lam, and E. Eleftheriou. Programming algorithms for multilevel phase-change memory. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 329–332, May 2011. doi: 10.1109/ISCAS.2011.5937569.

N. Papandreou, A. Sebastian, A. Pantazi, M. Breitwisch, C. Lam, H. Pozidis, and E. Eleftheriou. Drift-resilient cell-state metric for multilevel phase-change memory. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, pages 3.5.1–3.5.4, Dec 2011. doi: 10.1109/IEDM.2011.6131482.

N. Papandreou, T. Antonakopoulos, U. Egger, A. Palli, H. Pozidis, and E. Eleftheriou. A versatile platform for characterization of solid-state memory channels. In *Digital Signal Processing (DSP), 2013 18th International Conference on*, pages 1–5, July 2013. doi: 10.1109/ICDSP.2013.6622745.

S. Park. Technology scaling challenge and future prospects of dram and nand flash memory. In *Memory Workshop (IMW), 2015 IEEE International*, pages 1–4, May 2015. doi: 10.1109/IMW.2015.7150307.

T. P. Leervad Pedersen, J. Kalb, W. K. Njoroge, D. Wamwangi, M. Wuttig, and F. Spaepen. Mechanical stresses upon crystallization in phase change materials. *Applied Physics Letters*, 79(22):3597–3599, 2001. doi: http://dx.doi.org/10.1063/1.1415419.

V. Peluso, M.S.J. Steyaert, and Willy Sansen. A 1.5-v-100- mu;w delta; sigma; modulator with 12-b dynamic range using the switched-opamp technique. *Solid-State Circuits, IEEE Journal of*, 32(7):943–952, Jul 1997. ISSN 0018-9200. doi: 10.1109/4.597284.

A. Pirovano, A.L. Lacaita, A. Benvenuti, F. Pellizzer, and R. Bez. Electronic switching in phase-change memories. *Electron Devices, IEEE Transactions on*, 51(3):452–459, March 2004. ISSN 0018-9383. doi: 10.1109/TED.2003.823243.

A. Pirovano, A.L. Lacaita, F. Pellizzer, S.A. Kostylev, A. Benvenuti, and R. Bez. Low-field amorphous state resistance and threshold voltage drift in chalcogenide materials. *IEEE Transactions on Electron Devices*, 51(5):714–719, May 2004. ISSN 0018-9383. doi: 10.1109/TED.2004.825805.

A. Pirovano, A. Redaelli, F. Pellizzer, F. Ottogalli, M. Tosi, D. Ielmini, A.L. Lacaita, and R. Bez. Reliability study of phase-change nonvolatile memories. *Device and Materials Reliability, IEEE Transactions on*, 4(3):422–427, Sept 2004. ISSN 1530-4388. doi: 10.1109/TDMR.2004.836724.

H. Pozidis, N. Papandreou, A. Sebastian, T. Mittelholzer, M. BrightSky, C. Lam, and E. Eleftheriou. A framework for reliability assessment in multilevel phase-change memory. In *IEEE International Memory Workshop (IMW)*, pages 1–4, May 2012. doi: 10.1109/IMW.2012.6213671.

H. Pozidis, N. Papandreou, A. Sebastian, T. Mittelholzer, M. BrightSky, C. Lam, and E. Eleftheriou. Reliable MLc data storage and retention in phase-change memory after endurance cycling. In *IEEE International Memory Workshop (IMW)*, pages 100–103, 2013. doi: 10.1109/IMW.2013.6582108. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6582108.

H. Pozidis, T. Mittelholzer, N. Papandreou, T. Parnell, and M. Stanisavljevic. Phase change memory reliability: A signal processing and coding perspective. *IEEE Transactions on Magnetics*, 51(4):1–7, April 2015. ISSN 0018-9464. doi: 10.1109/TMAG.2014.2357176.

M. Qazi, M. Clinton, S. Bartling, and A.P. Chandrakasan. A low-voltage 1mb feram in 0.13μm cmos featuring time-to-digital sensing for expanded operating margin in scaled cmos. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, pages 208–210, Feb 2011. doi: 10.1109/ISSCC.2011.5746285.

F. Rao, Z. Song, M. Zhong, L. Wu, G. Feng, B. Liu, S. Feng, and B. Chen. Multilevel data storage characteristics of phase change memory cell with doublelayer chalcogenide films (ge 2 sb 2 te 5 and sb 2 te 3). *Japanese Journal of Applied Physics*, 46(1L):L25, 2007.

S. Raoux, R. M. Shelby, J. Jordan-Sweet, B. Munoz, M. Salinga, Y. Chen, Y. Shih, E. Lai, and M. Lee. Phase change materials and their application to random access memory technology. *Microelectronic Engineering*, 85(12):2330 – 2333, 2008. ISSN 0167-9317. doi: http://dx.doi.org/10.1016/j.mee.2008.08.004. Materials and Emerging Technologies for Non-Volatile-Memory DevicesProceedings of the Symposium H the {EMRS} 2008 Spring Meeting.

S. Raoux, H. Cheng, B. Munoz, and J. Jean. Crystallization characteristics of ge-sb and ge-te phase change materials. pages 91–98, September 2009.

S. Raoux, F. Xiong, M. Wuttig, and E. Pop. Phase change materials and phase change memory. *MRS Bulletin*, 39:703–710, 8 2014. ISSN 1938-1425. doi: 10.1557/mrs.2014.139.

A. Redaelli, D. Ielmini, A.L. Lacaita, F. Pellizzer, A. Pirovano, and R. Bez. Impact of crystallization statistics on data retention for phase change memories. In *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, pages 742–745, Dec 2005. doi: 10.1109/IEDM.2005.1609460.

J. P. Reifenberg, C. Kuo-wei, M. A. Panzer, S. B. Kim, J.A. Rowlette, M. Asheghi, H.-S.P. Wong, and K. E. Goodson. Thermal boundary resistance measurements for phase-change memory devices. *Electron Device Letters, IEEE*, 31(1):56–58, Jan 2010. ISSN 0741-3106. doi: 10.1109/LED.2009.2035139.

John P. Reifenberg, D.L. Kencke, and Kenneth E. Goodson. The impact of thermal boundary resistance in phase-change memory devices. *Electron Device Letters, IEEE*, 29(10):1112–1114, Oct 2008. ISSN 0741-3106. doi: 10.1109/LED.2008.2003012.

B. Ricco, G. Torelli, M. Lanzoni, A. Manstretta, H.E. Maes, D. Montanari, and A. Modelli. Nonvolatile multilevel memories for digital applications. *Proceedings of the IEEE*, 86(12):2399–2423, Dec 1998. ISSN 0018-9219. doi: 10.1109/5.735448.

W. P. Risk, C. T. Rettner, and S. Raoux. Thermal conductivities and phase transition temperatures of various phase-change materials measured by the 3ω method. *Applied Physics Letters*, 94(10):101906, 2009. doi: http://dx.doi.org/10.1063/1.3097353.

M. Rizzi and D. Ielmini. Energy landscape model of conduction and switching in phase change memories. In *Electron Devices Meeting (IEDM), 2012 IEEE International*, pages 26.2.1–26.2.4, Dec 2012. doi: 10.1109/IEDM.2012.6479106.

U. Russo, D. Ielmini, and A.L. Lacaita. Analytical modeling of chalcogenide crystallization for pcm data-retention extrapolation. *Electron Devices, IEEE Transactions on*, 54(10):2769–2777, Oct 2007. ISSN 0018-9383. doi: 10.1109/TED.2007.904976.

U. Russo, D. Ielmini, A. Redaelli, and A.L. Lacaita. Modeling of programming and read performance in phase-change memories x2014;part ii: Program disturb and mixed-scaling approach. *Electron Devices, IEEE Transactions on*, 55(2):515–522, Feb 2008. ISSN 0018-9383. doi: 10.1109/TED.2007.913573.

U. Russo, D. Ielmini, A. Redaelli, and A.L. Lacaita. Modeling of programming and read performance in phase-change memories x2014;part ii: Program disturb and mixed-scaling approach. *Electron Devices, IEEE Transactions on*, 55(2):515–522, Feb 2008. ISSN 0018-9383. doi: 10.1109/TED.2007.913573.

M. Sako, Y. Watanabe, T. Nakajima, J. Sato, K. Muraoka, M. Fujiu, F. Kouno, M. Nakagawa, M. Masuda, K. Kato, Y. Terada, Y. Shimizu, M. Honma, A. Imamoto, T. Araya, H. Konno, T. Okanaga, T. Fujimura, Xiaoqing Wang, M. Muramoto, M. Kamoshida, M. Kohno, Y. Suzuki, T. Hashiguchi, T. Kobayashi, M. Yamaoka, and R. Yamashita. 7.1 a low-power 64gb mlc nand-flash memory in 15nm cmos technology. In *Solid- State Circuits Conference - (ISSCC), 2015 IEEE International*, pages 1–3, Feb 2015. doi: 10.1109/ISSCC.2015.7062959.

A. Sebastian, N. Papandreou, A. Pantazi, H. Pozidis, and E. Eleftheriou. Non-resistance-based cell-state metric for phase-change memory. *Journal of Applied Physics*, 110(8):084505–084505–6, Oct 2011. ISSN 0021-8979. doi: 10.1063/1.3653279.

A. Sebastian, D. Krebs, M. Le Gallo, H. Pozidis, and E. Eleftheriou. A collective relaxation model for resistance drift in phase change memory cells. In *IEEE International Reliability Physics Symposium (IRPS)*, pages MY.5.1–MY.5.6, April 2015. doi: 10.1109/IRPS.2015.7112808.

J. F. Shackelford and W. Alexander. *CRC Materials Science and Engineering Handbook, Third Edition*. CRC press, Boca Raton, FL, USA, Dec. 2000.

K. Shportko, S. Kremers, M. Woda, D. Lencer, J. Robertson, and M. Wuttig. Resonant bonding in crystalline phase-change materials. *Nat Mater*, 7(8):653–658, Aug 2008. ISSN 1476-1122. doi: 10.1038/nmat2226.

D. Slepian. Permutation modulation. *Proceedings of the IEEE*, 53(3):228–236, 1965. doi: 10.1109/PROC.1965.3680. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1445610.

M. Stanisavljevic, A. Athmanathan, N. Papandreou, H. Pozidis, and E. Eleftheriou. Phase-change memory: Feasibility of reliable multilevel-cell storage and retention at elevated

temperatures. In *Reliability Physics Symposium (IRPS), 2015 IEEE International*, pages 5B.6.1–5B.6.6, April 2015. doi: 10.1109/IRPS.2015.7112747.

M. Terai, Y. Sakotsubo, S. Kotsuji, and H. Hada. Resistance controllability of Ta_2 O_5/TiO_2 stack reram for low-voltage and multilevel operation. *Electron Device Letters, IEEE*, 31(3):204–206, March 2010. ISSN 0741-3106. doi: 10.1109/LED.2009.2039021.

W. J. Wang, L. P. Shi, R. Zhao, K. G. Lim, H. K. Lee, T. C. Chong, and Y. H. Wu. Fast phase transitions induced by picosecond electrical pulses on phase change memory cells. *Applied Physics Letters*, 93(4):043121, 2008. doi: http://dx.doi.org/10.1063/1.2963196.

R. Waser. Resistive non-volatile memory devices (invited paper). *Microelectronic Engineering*, 86(7–9):1925 – 1928, 2009. ISSN 0167-9317. doi: http://dx.doi.org/10.1016/j.mee.2009.03.132. {INFOS} 2009.

Y. Watanabe, J. G. Bednorz, A. Bietsch, Ch. Gerber, D. Widmer, A. Beck, and S. J. Wind. Current-driven insulator–conductor transition and nonvolatile memory in chromium-doped srtio3 single crystals. *Applied Physics Letters*, 78(23):3738–3740, 2001. doi: http://dx.doi.org/10.1063/1.1377617.

M. Wimmer, M. Kaes, C. Dellen, and M. Salinga. Role of activation energy in resistance drift of amorphous phase change materials. *Frontiers in Physics*, 2(75), 2014. ISSN 2296-424X. doi: 10.3389/fphy.2014.00075. URL http://www.frontiersin.org/condensed_matter_physics/10.3389/fphy.2014.00075/abstract.

Stuart A. Wolf, Jiwei Lu, M.R. Stan, E. Chen, and D.M. Treger. The promise of nanomagnetics and spintronics for future logic and universal memory. *Proceedings of the IEEE*, 98(12):2155–2168, Dec 2010. ISSN 0018-9219. doi: 10.1109/JPROC.2010.2064150.

H.-S.P. Wong, S. Raoux, S. B. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, and K. E. Goodson. Phase change memory. *Proceedings of the IEEE*, 98(12):2201–2227, 2010. ISSN 0018-9219. doi: 10.1109/JPROC.2010.2070050.

H.-S.P. Wong, Heng-Yuan Lee, Shimeng Yu, Yu-Sheng Chen, Yi Wu, Pang-Shiu Chen, Byoungil Lee, F.T. Chen, and Ming-Jinn Tsai. Metal x2013;oxide rram. *Proceedings of the IEEE*, 100(6):1951–1970, June 2012. ISSN 0018-9219. doi: 10.1109/JPROC.2012.2190369.

F. Xiong, A. D. Liao, D. Estrada, and E. Pop. Low-power switching of phase-change materials with carbon nanotube electrodes. *Science*, 332(6029):568–570, 2011. doi: 10.1126/science.1201938. URL http://www.sciencemag.org/content/332/6029/568.abstract.

T. Yamane, N. Nagai, S. Katayama, and M. Todoki. Measurement of thermal conductivity of silicon dioxide thin films using a 3ω method. *Journal of Applied Physics*, 91(12):9772–9776, 2002. doi: http://dx.doi.org/10.1063/1.1481958.

Y. Zhang, J. Feng, Y. Zhang, Z. Zhang, Y. Lin, T. Tang, B. Cai, and B. Chen. Multi-bit storage in reset process of phase change access memory (pram). *physica status solidi (RRL) – Rapid Research Letters*, 1(1):R28–R30, 2007. ISSN 1862-6270. doi: 10.1002/pssr.200600020.

W.W. Zhuang, W. Pan, B.D. Ulrich, J.J. Lee, L. Stecker, A. Burmaster, D.R. Evans, S.T. Hsu, M. Tajiri, A. Shimaoka, K. Inoue, T. Naka, N. Awaya, A. Sakiyama, Y. Wang, S.Q. Liu, N.J. Wu, and A. Ignatiev. Novel colossal magnetoresistive thin film nonvolatile resistance random access memory (rram). In *Electron Devices Meeting, 2002. IEDM '02. International*, pages 193–196, Dec 2002. doi: 10.1109/IEDM.2002.1175811.

Aravinthan Athmanathan
Albisstrasse 45
8134 Adliswil
Switzerland

+41-78-792-3456 (mobile)
e-mail:aat@zurich.ibm.com
http://researcher.ibm.com/person/zurich-aat

Research Interests

- Emerging non-volatile memory technologies with primary research focus on Phase-Change Random Access Memory (PCRAM)
- READ/WRITE architectures & analog-mixed signal design for Multi-Level Cells PCRAM
- Finite-element-modeling of exploratory memory devices & electrical characterization of PCRAM devices and arrays

Summary

- Pre-doctoral fellow in Non-Volatile Memory Systems Group at IBM Research in Zurich, Switzerland (2011 - present)
- More than 5 articles published in top journals and conferences and 1 patent filed
- M.Sc. in Electrical & Electronics Engineering, EPFL, Switzerland (2009 - 2011)
- ASIC R&D Engineer at IBM Systems & Technology Group, Bangalore, India (2007 - 2009)

Work Experience

- **IBM Research - Zurich** — Rüschlikon, Switzerland
  Pre-doctoral fellow, Non-Volatile Memory Systems Group — *Oct. 2011 - Present*
  · Finite-Element Modeling (FEM) of thermal and electrical transport in nano-scale devices and thermo-electric physics in phase-change memory devices
  · Electrical characterization of Phase-Change Memory arrays at elevated temperatures for Multi-Level Cell reliability and data retention of PCM using embedded FPGA based hardware characterization platform
  · Analog-mixed signal circuit design of programming and sensing schemes/architectures of readout metrics for Multilple-level cells Phase-change memory
  · Actively involved in the IBM-SK Hynix Joint Development Program for PCRAM development and worked closely with the SK-Hynix next generation memory design & development team in Icheon, South Korea
- **IBM Research - Zurich** — Rüschlikon, Switzerland
  Research Intern, Storage technologies — *Oct. 2010 - Sep. 2011*
  · Design of a novel readout metric scheme for Multi-Level Cell Phase-Change Memory in 90 nm CMOS technology
  · Digital design, synthesis and implementation of READ/WRITE digital controller for Phase-Change Memory
- **IBM ASIC Design Centre** — Bangalore, India
  R & D Engineer, Systems & Technology Group — *Jul. 2007 - Sep. 2009*
  · Complete ASIC backend physical design-flow with focus on floorplanning, clock tree synthesis, Place & Route and timing closure on ASIC chips from clients like Cisco, Huawei and Hitachi

Education

- **École Polytechnique Fédérale de Lausanne (EPFL)** — Lausanne, Switzerland
  Ph.D., Electrical Engineering — *Oct. 2011 - present*
  · Thesis: A modeling and reliability frame-work for multi-level cells (MLC) Phase-Change Memories
  · Advisors: Dr. Milos Stanisavljevic (IBM Research - Zurich) & Prof. Yusuf Leblebici (LSM, EPFL)
  · READ/WRITE circuit design, Reliability and modeling of multi-bit per cell phase change memories
- **École Polytechnique Fédérale de Lausanne (EPFL)** — Lausanne, Switzerland
  M.Sc., Electrical and Electronic Engineering **(CGPA - 5.15/6)** — *Sep. 2009 - May 2011*
  · Thesis: Novel readout scheme for Multi-Level Cells - Phase Change Memory

· Advisors: Dr. Gael Close (IBM Research - Zurich) & Prof. Yusuf Leblebici (LSM, EPFL)
· Readout architecture design of a novel cell-state voltage-based metric scheme in 90 nm Phase-Change memories

- **Madras Institute of Technology, Anna university** Chennai, India
  B.Tech., Electronics and Communications Engineering **(CGPA - 8.6/10)** *Jul. 2003 - Jun. 2007*
  · Thesis: Design of Two-stage CMOS Operational amplifier
  · Advisor: Prof. Ganesh Madhan
  · Specialized in Analog IC design and VLSI circuits

Technical Skills

- IC Design : Cadence ICFB, Encounter, Virtuoso, SPICE, Cadence ncsim, Modelsim
- Modeling & Simualtion : MATLAB, COMSOL Multiphysics
- Hardware Description Langauges : Verilog, VHDL, Verilog-AMS, VHDL-AMS
- Software Language : C, C++, perl, UNIX

Main publications

- **Journals**

  Multi-Level Phase-Change Memory - A Viable Technology
  Aravinthan Athmanathan, Milos Stanisavljevic, Nikolaos Papandreou, Haris Pozidis and Evangelos Eleftheriou
  accepted at IEEE JETCAS Journal, December 2015

  A Comprehensive Finite-Element Thermo-electric model for Phase-Change Memory devices
  Aravinthan Athmanathan, Daniel Krebs, Abu Sebastian, Haris Pozidis and Evangelos Eleftheriou
  submitted to Journal of Applied Physics, 2015

- **Conferences**

  A 6-bit drift-resilient readout scheme for multi-level Phase-Change Memory
  Aravinthan Athmanathan,Milos Stanisavljevic, Nikolaos Papandreou, Haris Pozidis and Evangelos Eleftheriou
  in Asian Solid-State Circuits Conference (A-SSCC) 2014, Taipei, Taiwan

  A Finite-element Thermo-electric model for Phase-Change Memory devices
  Aravinthan Athmanathan, Daniel Krebs, Abu Sebastian, Haris Pozidis and Evangelos Eleftheriou
  to be presented at SISPAD 2015, Washington DC, USA

  Phase-Change Memory: Feasibility of Reliable Multilevel-cell Storage and Retention at Elevated Temperatures
  Milos Stanisavljevic, **Aravinthan Athmanathan**, Nikolaos Papandreou, Haris Pozidis and Evangelos Eleftheriou
  in International Physics Reliability Symposium (IRPS) 2015, California, USA

  Scalibility of Phase-Change Memory: a modeling perspective **(Invited)**
  Aravinthan Athmanathan, Daniel Krebs, Abu Sebastian, Haris Pozidis and Evangelos Eleftheriou
  in International Workshop on Simulation and Modeling of Memory devices (IWSMM) 2013, Milano, Italy

Patents filed

- **A. Athmanathan** & D. Krebs, Memory Device and Method for thermoelectric heat confinement, Aug 2014.

Personal Information

- Nationality : Indian
- Work permit : Valid work permit-B
- Availability : October 2015
- Date of birth : 7^{th} October 1985
- Languages : English (fluent), Tamil (native), German (beginner)
- References : Available on request