Towards High-Endurance, Nonvolatile, CMOS-Compatible Ferroelectric Memories for Next-Generation Computing

Ava Tan
Sayeef Salahuddin, Ed.
Chenming Hu, Ed.
Junqiao Wu, Ed.

Electrical Engineering and Computer Sciences
University of California, Berkeley

May 1, 2022

Towards High-Endurance, Nonvolatile, CMOS-Compatible Ferroelectric Memories for
Next-Generation Computing

by

Ava Jiang Tan

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Sayeef Salahuddin, Chair
Professor Chenming Hu
Professor Junqiao Wu

Spring 2021

Towards High-Endurance, Nonvolatile, CMOS-Compatible Ferroelectric Memories for Next-Generation Computing

Abstract

Towards High-Endurance, Nonvolatile, CMOS-Compatible Ferroelectric Memories for Next-Generation Computing

by

Ava Jiang Tan

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Sayeef Salahuddin, Chair

As modern computing workloads become more and more data-centric, there has been an increasing need in recent years to develop memory solutions which can adequately provide the high performance, speed, and energy efficiencies required by data-intensive applications. Simultaneously, with the timely discovery of ferroelectricity in a well-investigated CMOS-compatible material – hafnium oxide, or HfO_2 – memory devices integrating ferroelectrics have also made a resurgence on the nonvolatile memory landscape.

This work presents the ground-up development of a novel CMOS-compatible ferroelectric oxide, doped HfO_2, and demonstrates its successful integration into ferroelectric memory capacitors (FeRAMs), transistors (FeFETs), and content addressable memory cells (FeCAMs). As HfO_2-based memories are still a nascent technology, special emphasis is placed on developing a deeper physical understanding of the various engineering challenges associated with process integration and device performance. Underlying reliability concerns related to limited cycling endurance and premature device failure are identified, and methods to mitigate some of these bottlenecks are presented and investigated. Based upon the understanding derived from identifying the physical root causes for degradation, a highly effective gate oxide engineering technique for boosting the endurance metric of the FeFET by roughly 5 orders of magnitude, which enables the highest endurance numbers reported on FeFETs with a crystalline silicon channel to date, is demonstrated. Lastly, the successful fabrication and characterization of content addressable memory cells based on the FeFET is reported. The simple 2-FeFET FeCAM cell boasts both nonvolatility and substantially smaller on-chip footprint in contrast to its existing SRAM-based CMOS counterpart.

The overall objective of this work is to provide a pathway forward for continued development on a CMOS-compatible nonvolatile memory element that can be used for embedded memory applications or for in-memory computing. The operational properties of the

doped HfO_2-based FeFET, in considering its intrinsic fast write/read speeds, low voltage requirements, and retention robustness, makes it well-suited to accommodate demanding modern computational needs by sealing the gaps between conventional memory, logic, and continued device scaling.

To Alexander Tan,

in loving memory.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to thank my advisor, Professor Sayeef Salahuddin, for his years of constant support and guidance throughout my PhD journey. His zest for scientific research and discovery has made an everlasting impact on me, and I thank him for being a huge source of motivation for me to overcome the hardships I encountered along the way and to challenge myself to be a more diligent researcher. At the same time, he has always allowed me the intellectual freedom to pursue interesting and provocative research ideas under his wing. With his supervision, I flourished as a PhD student and researcher, and I will always treasure the invaluable mentorship he has given me in the past five years.

Next, I would like to thank Professor Chenming Hu for his wisdom, encouragement, and many useful scientific and research insights over the years. It has been a privilege to work with the father of the FinFET during my years at Berkeley, and I will forever be grateful to have benefited from his mentorship and guidance. I thank him as well for serving on my qualifying exam and dissertation committees.

I would like to thank Professor Junqiao Wu in the materials science and engineering department for serving on my qualifying exam and dissertation committees, as well as Professor Tsu-Jae King Liu for serving on my qualifying exam committee. The feedback, questions and discussions they provided proved instrumental in helping me structure my research focus and dissertation content. Additionally, I would like to thank Professor Ali Javey and Professor Vivek Subramanian for serving on my prelim exam committee. I am especially grateful for my fruitful undergraduate research internship in Professor Subramanian's group in the summer of 2015, working alongside Will Scheideler, which convinced me that Berkeley would be the perfect place to go to graduate school. I would like to thank Professor Gireeja Ranade for allowing me the opportunity to be a teaching assistant for EECS 16A. It was truly not easy teaching an entirely remote course for the first time during the COVID-19 pandemic, but under her leadership I embraced the challenges of remote teaching and strove every day to provide a quality educational experience for my students.

I would like to thank Alex Yoon at Lam Research, who extended to me a fortuitous opportunity to do an internship during the summer of 2018 in his group. My teammates Zhongwei Zhu, Erica Chen, and Hwan Sung Choe were an absolute pleasure to work with, and I am grateful to say that some of the work done during this internship has found its way into this dissertation as well.

This experimental work shown in this dissertation would not have been possible without the support from the numerous hard-working and dedicated Berkeley Marvell Nanofabrication Laboratory staff and researchers. I am particularly indebted to the support provided by Bill Flounders, Dave Taosaka, Greg Mullins, Richelieu Hemphill, Kim Chan, Allison Dove, Ryan Rivers, Jay Morford, Danny Pestal, Jason Chukes, David Lo, Brian McNeil, Ben Lake, Lou Ahtty, and William King.

Some of my fondest memories that I will forever cherish of grad school are of the numerous discussions, both scientific and otherwise, that I've had with my talented labmates

and collaborators. From the very beginning, a huge thank you to Praveen Gowtham, Ko-rok Chatterjee, Justin Wong, Golnaz Karbasian, Sangwan Kim, Ajay Yadav, Niklas Roschewsky, Dominic Labanowski, Samuel Smith, Zhongyuan (John) Lu, Daewoong Kwon, Hong Zhou, Jiuren Zhou, Angada Sachid, Samuel Holladay, Yu-Hung Liao, Ming-Yen Kao, Saavan Patel, Adi Jung, Jong-Ho Bae, Li-Chen Wang, Steve Volkman, and many others for making the PhD life interesting and dynamic. I have also had the pleasure of engaging in external collaborations with Milan Pešić, Luca Larcher, Ghazal Saheli, Stephen Weeks, Jeffrey Anthis, and Dave Thompson from Applied Materials. Lastly, I thank my research mentee Neel Raman for allowing me to help him build his scientific curiosity in the lab.

Finally, I must express my gratitude to my close family, friends and loved ones, who have cheered me on every step of the way. I couldn't have asked for a better partner to embark on the graduate school journey with than my fiancé, Michael Whittaker – we have grown so much these past five years, tackling late nights and deadlines one by one, together. Armed with your persistent love and support, I was able to face every new day head-on, no matter how intimidating the laundry list of tasks at hand was. Thank you to my parents, Jean and Li-Zhe Tan, for cultivating a love for science and engineering in me from a young age, and supporting me every moment of the way through graduate school. The PhD journey was not easy, so I will forever be grateful for my strong support network that helped pull me through the tough times.

Chapter 1

Introduction

1.1   The Need for More Memory

We live in an increasingly information-centric era. Data is all around us, constantly being extracted from our actions and reintegrated into our daily decisions. It is clear that the digital age has completely transformed the way we interact with the world around us over the last decade. Increasing global connectivity and data creation afforded by Internet of Things (IoT) technologies, traditional and cloud datacenters, and edge infrastructure has created an explosion in the demand for memory storage solutions.

Annual Size of the Global Datasphere

Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018
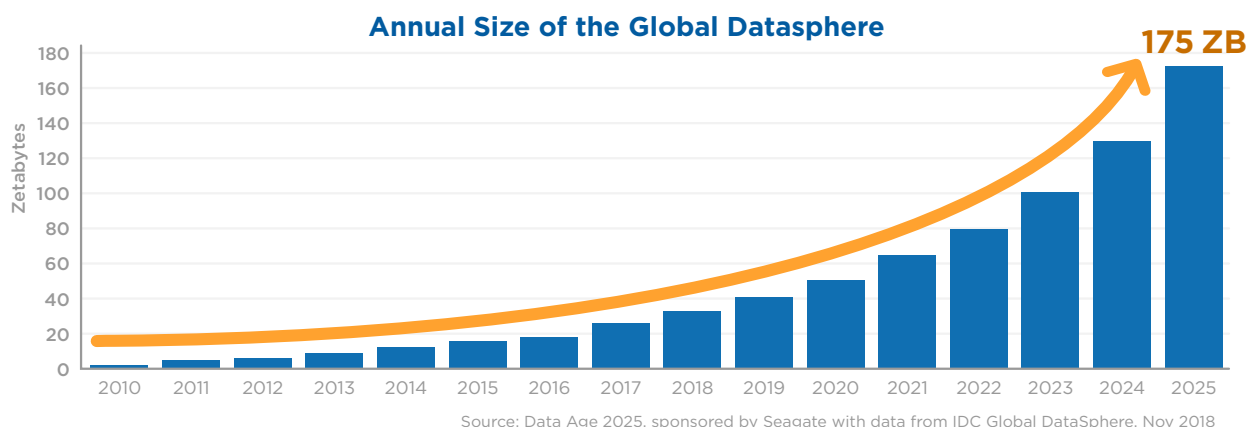
Figure 1.1: Projected size of the Global Datasphere (the total amount of data created, captured, or replicated) each year from 2010 through 2025. Figure adapted from [110].

The International Data Corporation (IDC) forecasts a growth in the Global Datasphere, which roughly represents the total summation of all data transmitted, replicated or received, of 175 zettabytes (ZB) by the year 2025. This number is equivalent to one trillion

gigabytes. To illustrate its staggering size, consider the fact that if one were to download 175 ZB of data at a typical download speed of 25 Mbps, then it would take 1.8 billion years to download the entirety of the projected 2025 Global Datasphere ([110]).

Clearly, these ever-increasing requirements for fast data access, throughput, and sheer storage size must be supplemented accordingly by disruptive technological advancements in memory solutions. For well over 30 years to date, standard CMOS memory solutions such as SRAM (static random access memory), DRAM (dynamic random access memory), and flash memory have been able to fulfill the provisional needs [25, 133]. Each of SRAM, DRAM, and flash memory technologies has secured a well-defined role within the so-called "memory hierarchy." This memory hierarchy complements the requirements of processor core in that the most frequently accessed memories are also the fastest (albeit the most expensive), designed to achieve high write and read endurance; the memories accessed less frequently are designed to hold a much higher capacity, are less expensive, and are nonvolatile in nature.

The memory hierarchy is depicted in Figure 1.2(a). It is organized in a pyramid structure according to increasing cost/speed from bottom to top, and increasing size (represented by the area of each "slice" that each memory technology occupies within the pyramid) from top to bottom. Similarly, the speed and endurance expectations of these memories is illustrated in Figure 1.2(b).
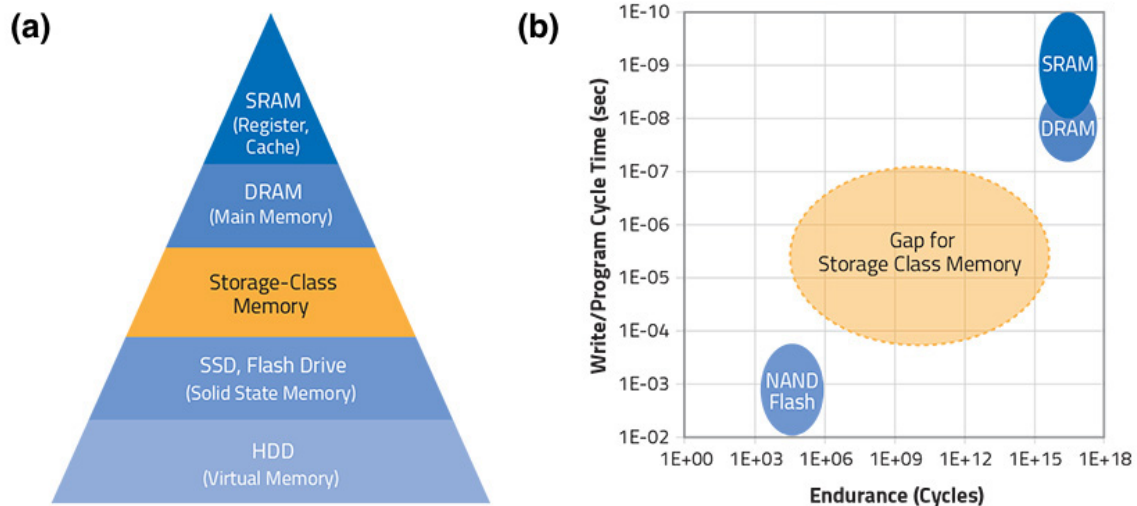
Figure 1.2: (a) The organization of the traditional memory hierarchy. Storage class memory (SCM) will be designed to sit somewhere in the middle of the hierarchy. (b) A plot of programming speed vs. endurance for several established memories as well as storage class memory. Figure adapted from [142].

A division of memories called **storage class memories** (highlighted in orange in the subfigures of Figure 1.2) encompasses many emerging memory solutions that are designed to bridge the gap in performance and cost that exists in the traditional memory hierarchy. For instance, at the time of the writing of this dissertation (May 2021), Intel's Optane 3D XPoint technology is an illustrative example of a recently commercialized memory solution designed to blur the distinctions between traditional storage and memory, which was achieved primarily through its low latency accesses that rival the latency of the system itself [34, 52]. The need for innovations in storage class memories stems from the fact that there exists a clear delineation between lower endurance, slower, but cheaper solid state memories and the higher endurance, faster, and more expensive cache memory technologies, as illustrated in Figure 1.2(b). A storage solution that can be operated at reasonable speeds but is more responsive than, for example, current NAND SSD's, can provide a cost-effective alternative to using more expensive DRAM. It is therefore anticipated that storage class memories can help to bridge this existing rift between more mature memory technologies [45].

1.2 Perspective on Emerging Nonvolatile Memory Technologies

Of all the current mainstream memory technologies, flash memories are the quintessential example of a nonvolatile memory that has a well-defined role in the memory landscape with little possibility of being supplanted in the very near future. In fact, the first nonvolatile memory element ever demonstrated – upon which modern day flash technology is based – was the floating gate transistor in 1967 by Kahng and Sze of Bell Labs [25, 59]. Flash has since become the dominant technology for storage in mobile devices and furthermore has enjoyed extended usage as a complementary storage technology for hard disk drives in large-scale and cloud computing systems [133]. The technology has witnessed tremendous growth in the past two decades especially, accounting for \$35 billion in revenue worldwide in 2016 alone [31]. That said, flash technologies have undergone and will need to continue to undergo significant changes be able to continue supporting technological demands [145]. From the development of NOR and NAND architectures to multi-bit cell storage, flash memories have already evolved substantially since their inception in 1967.

A variety of emerging nonvolatile memories are currently under intense research development and consideration for future storage class memories and as potential replacement candidates for established technologies [25, 78, 133, 144]. Notable emerging nonvolatile memories under consideration include MRAM (magnetoresistive random access memory), PCM (phase change memory), ReRAM (resistive random access memory), and ferroelectric memories such as FeRAM (ferroelectric random access memory) and the FeFET (ferroelectric field effect transistor). The latter two technologies based on ferroelectric

materials will serve as the primary focus of this dissertation. The following subsections will act as a primer to explore the features, advantages, and disadvantages of the aforementioned technologies.

MRAM: Magnetoresistive Random Access Memory

Magnetoresistive random access memories are a class of nonvolatile memories that store data as magnetic states, and are read by determining the resistance of the memory (which is associated with a particular data state). The heart of any MRAM technology is the MTJ, or magnetic tunnel junction. Therefore, the mechanism for determining the state of the device – by determining resistance – is the same regardless of MRAM type; the differences arise from the write mechanisms and methods by which the data is written to the device [3]. MRAMs are therefore broadly categorized into three different types according to the switching/writing mechanism: 1) Toggle MRAM by magnetic field-induced switching; 2) STT-MRAM (spin transfer torque) by the transfer of a magnetic moment via current; and SOT-MRAM (spin orbit torque) which uses additional spin-orbit coupling to enhance the spin transfer.
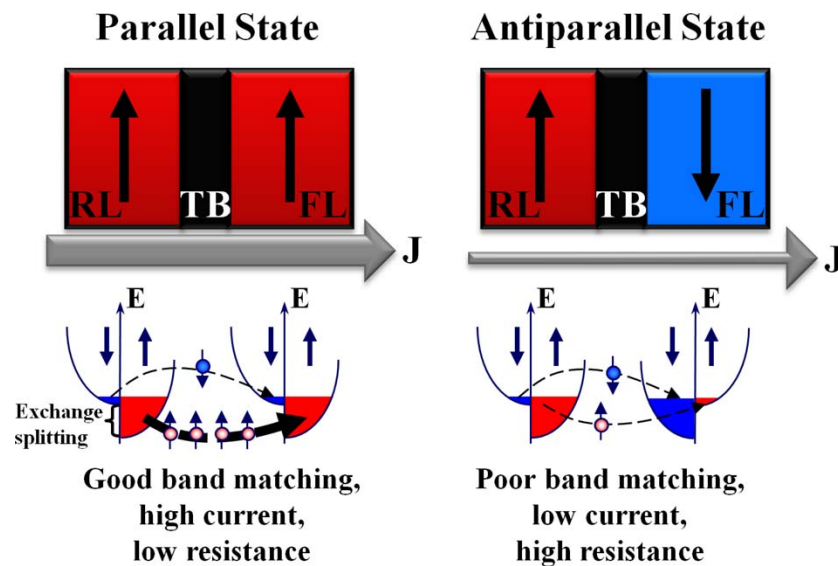
Figure 1.3: Example of a typical MTJ device consisting of two ferromagnetic layers separated by a tunnel barrier layer. Figure adapted from [3].

As shown in Figure 1.3, a tunnel barrier layer (typically MgO) is sandwiched between two ferromagnetic layers (typically Co-Fe alloys, one pinned and one free) in an MTJ device. When the magnetizations in both ferromagnetic layers are oriented in parallel, there

is good band matching, and therefore the electrons can easily tunnel through the barrier and into the majority band of the other layer, giving rise to the low-resistance state (as shown on the left of Figure 1.3). When the magnetizations are aligned to be antiparallel, the electrons instead must tunnel into the minority band, and inadequate receiving states to tunnel into and/or available carriers gives rise to the high-resistance state (as shown in the right of Figure 1.3. In effect, these two resistance states change the amount of spin-polarized quantum mechanical tunneling current which is able to flow through the MTJ device [3, 115]. This is called the tunneling magnetoresistance effect (TMR).

Toggle MRAM technology based on Savtchenko switching, by which a current-driven magnetic field modulates the magnetization of the storage layer, has been in mass production since 2006 [3, 114]. Writing is typically accomplished by flowing large, orthogonal currents in order to induce orthogonal magnetic fields at a particular location corresponding to an MRAM cell; once the required field strength for switching is reached, the magnetization of the free layer in the MTJ then switches [12, 74]. However, as the earliest writing scheme for MRAM technology, it was quickly bottlenecked by the fact that the amount of current required to write to a device is inversely proportional to the area of the storage element – hence, field assisted MRAM has been limited to 90 nm node technology and is unable to be scaled to smaller dimensions.

STT-MRAM, on the other hand, requires no generation of an external magnetic field to switch the magnetization, which greatly improves the current-limit issue experienced in toggle MRAM. In STT-MRAM, spin-polarized current is directly sent through the MTJ itself, which also simplifies the design of the MRAM cell. The transfer of the spin angular momentum then exerts a torque on the free magnetic layer, and if the torque is strong enough, it can then change the magnetization from parallel to antiparallel (or vice-versa) depending on the direction of injected current through the device. However, STT-MRAM still suffers from relatively high programming currents, and therefore high energy consumption as well [12]. Additional issues such as reduced lifetime due to breakdown of the oxide layer of the MTJ, degraded TMR under applied voltage, bit-flipping due to the fact that the write and read mechanisms use the same access path, and long write latencies for the intended application.

SOT-MRAM alleviates many issues seen in STT-MRAM technology by introducing an additional terminal to separate out the read path from the write path *and* by ensuring that the switching current does not flow through the barrier layer [123]. In this technology, the MTJ is typically fabricated directly on top of a heavy metal channel which is in contact with the free layer. Direction of current flow from source to write line determines the magnetization of the free layer and therefore the bit stored [12], and the word line accesses the cell during reads. In comparison to STT-MRAM, it is clear that this memory demonstrates faster switching capabilities. However, still the question of scalability and footprint efficiency remains, as two additional access transistors per cell must be used. As SOT-MRAM technology is still relatively in its infancy [115], only time can tell if existing shortcomings will be addressed.

PCM: Phase Change Memory

Among the most mature novel nonvolatile memory technologies is phase change memory, or PCM. It has been a key enabling technology for nanoscale, nonvolatile storage. As mentioned previously, Intel's 3D XPoint storage solution, branded as Intel Optane, was released in 2018 and is believed to use a phase-change material as the storage element [46].

The principle of operation relies on changing the nature of a chalcogenide glass sandwiched between two electrodes from either the crystalline or amorphous states by passing current through the device. The phase change material therefore can alternate between a low-resistance or high-resistance state, as illustrated in Figure 1.4, and therefore can modulate its resistivity. Therefore, by detecting the resistance of the device, one can determine the state of the memory. The SET operation involves applying a longer electrical pulse to heat most of the cell above the crystallization temperature to induce crystallization; the RESET operation involves applying a larger electrical pulse to melt the center of the cell, and must be abrupt enough to quench the material in the amorphous phase [18].
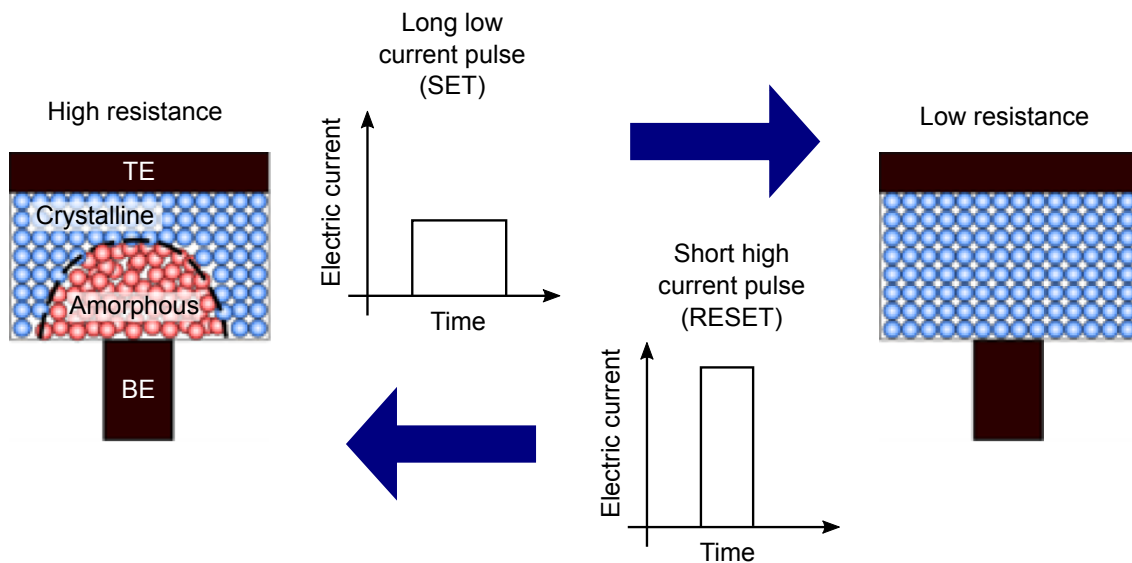
Figure 1.4: Example of a typical PCM device transitioning through the low-resistance state (crystalline) and the high-resistance state (amorphous). Adapted from [46].

Properties such as potential multi-level storage, fast access/write times, and good endurance are positive attributes to PCM technologies. However, it is clear that issues such as further scaling/integration challenges in advanced technology nodes still exist and remain to be addressed. Like MRAM, the current required to write to PCM is delivered through an access device, which could be a transistor or diode; in order for sufficient cur-

rent delivery, these access devices generally need to be much larger than the PCM element itself, which ultimately limits storage density [108]. More generally, PCM technology is still afflicted with issues of reliability/crosstalk, the power and current densities associated with the RESET mechanism, the bottlenecking SET speed (which limits the write speed of the device), and issues stemming from stochasticity [44].

ReRAM: Resistive Random Access Memory

Much like phase change memory, resistive memories also consist of a material capable of resistivity change sandwiched between two metal electrodes. They are broadly categorized into two different types: oxide RRAM and conductive bridging RAM (CBRAM). Upon the application of an electric field, in oxide RRAM, oxygen vacancies are generated and can create conductive filaments connecting one electrode to another, thus changing the device resistance. In CBRAM, the filament is formed with diffusive metal ions that are formed by the oxidation of the metal electrodes [27, 133, 144]. In both cases, the filaments which form will vary the resistance of the device and therefore differentiate between the low and high states.
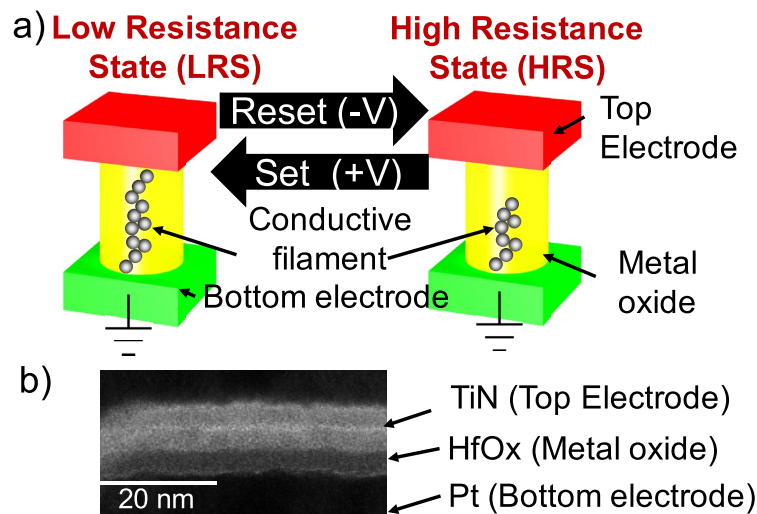
Figure 1.5: Example of a typical ReRAM device with a metal oxide layer consisting of HfO_x. (a) To switch between the low and high resistance states, a SET or RESET voltage is applied to the device to induce the formation or the dissolving of conductive filaments through the metal oxide layer. (b) A TEM showing the two-terminal ReRAM device structure with a TiN top electrode, HfO_x layer, and Pt bottom electrode. Adapted from [136].

In most oxide RRAM devices, a stress step called the *forming operation*, during which a higher voltage is applied to the device, is required to form the filament for the first time after fabrication [9, 27, 136]. The forming process essentially induces a soft breakdown of the dielectric material, which allows oxygen atoms to be kicked out of the lattice, leaving behind oxygen vacancies in the film [149]. In fact, not only does the forming operation pose a challenge as this process itself is difficult to control and requires high voltages, but even subsequent control of the filament formation from cycle to cycle and from device to device remains a challenge to be addressed [115]. The filament control issue translates to a large spread in the electrical characteristics of oxide RRAM devices, leading to reproducibility and variability concerns [133].

In comparing conductive bridging RAM to oxide RRAM, the movement of metal atoms via drift/diffusion is generally much easier in comparison to oxygen vacancies. This leads to a degradation of endurance and retention characteristics in comparison to oxide RRAM, with oxide RRAM typically outperforming CBRAM for endurance by many orders of magnitude [149]. To improve upon this issue, it is critical to be able to precisely control the lateral/vertical diffusion of metal cations in CBRAM. Though some binary oxides have shown promise for alleviating this issue, additional drawbacks such as larger reset currents and additional variability of the high-resistance state are introduced [27].

Finally, yet another consideration for two-terminal RRAM devices when integrated into crosspoint arrays is the need to develop accompanying selector devices which provide significant nonlinearity to suppress sneak currents and improve read margin. Transistors have generally been ruled out due to the need for high-temperature processing and larger area footprint; at the present moment, a two-terminal selector that will not limit the performance/reliability of the RRAM device are still under development [27].

A Comment on the Emerging Memory Landscape

At this point in time, there clearly exist many avenues of research dedicated to various resistance-based memories. Upon evaluation of the various benefits and drawbacks associated with each of the aforementioned technologies, it becomes rather apparent that we are far from developing an "ideal" or "universal" memory technology that excels above the rest. Due to the various strengths and weaknesses that both bolster and hamper each particular technology, it is probable that each solution could satisfy overlapping but shifted segments of the memory hierarchy within the space allowed for storage class memories or other applications. To first order, many solutions seem appropriate for embedded nonvolatile memories. Time will tell whether other emerging nonvolatile memories will follow in the footsteps of phase change memory/3D XPoint technology and claim a space in the storage class memory market as well.

Another possible application of emerging NVMs lies in the space of specialized memory hardware to realize non-von Neumann computing paradigms. In traditional computer organization, data needs to be continually transferred to and from memory and the CPU via data and address busses. In modern computing systems, it is apparent that the la-

tency and energy consumption of the system is actually mostly dominated by this transfer process – *not* by the actual computation on the data itself. This problem is commonly referred to as the "von-Neumann bottleneck" [77]. Many so-called in-memory computing and brain-inspired computing paradigms are still in early development, but emerging memories are quickly proving to be an attractive solution for their respective hardware implementations.

1.3 Ferroelectricity: A Brief History and Overview

Historical Background on Ferroelectrics

The history of ferroelectric material synthesis can be traced back as far as the mid-1600's, to a small town on the southwest coast of France named La Rochelle [32, 75]. A French apothecary, Elie Seignette, developed a curious medicinal salt – sodium potassium tartrate tetrahydrate, or $NaKC_4H_4O_6 \cdot 4H_2O$ – which was then used for well over two centuries as a mild drug. It wasn't until the 1800's that Rochelle salt, or *"sel polychreste"* as Mr. Seignette called it, came under intense study for its physical properties. In 1824, David Brewster was the first to observe the property of pyroelectricity in Rochelle salt, along with a number of other crystals; shortly afterwards, in 1880, the Curie brothers Pierre and Paul-Jacques conducted the first systematic studies of the piezoelectric effect in Rochelle salt.

In 1917, during the era of World War I, the United States became involved with the development of Rochelle salt to address burgeoning needs in submarine warfare. In particular, an American physicist and electrical engineer, Walter G. Cady, spearheaded contributions to the science of piezoelectricity, laying claim to early measurements of the elastic properties, dielectric constant, and first reported observation of the Curie temperature of Rochelle salt. His work directly paved the way for the development of the piezoelectrically stabilized resonator [32].

Joseph Valasek is another important contributor to the early understanding of the behavior of Rochelle salt. From 1921 to 1924, while he was a graduate student, he conducted a study to draw parallels between the magnetic properties of ferromagnetics and the dielectric properties of Rochelle salt, and these studies would later help establish the term *ferroelectricity* to describe his observations. He published the first documented hysteresis curve of a ferroelectric material [130] and, perhaps as importantly, the temperature-dependent piezoelectric response plot for Rochelle salt, which suggests a narrow temperature range for piezoelectric activity [129].

From this point onward, additional materials were soon discovered to also possess ferroelectricity - for example, potassium dihydrogen phosphate, KH_2PO_4 (KDP), by Georg Busch and Paul Scherrer in the mid 1930's [32]. In the 1940's, many perovskite families appeared on the ferroelectrics scene. In particular, barium titanate, $BaTiO_3$, emerged as a man-made perovskite ferroelectric for use in sonar systems for submarine detection during the Second World War. In fact, the discovery of ferroelectricity in $BaTiO_3$ was

extremely important, as it proved that ferroelectricity could exist in simple oxides and didn't have to be correlated with the presence of hydrogen bonding (as was the case with other materials to date). Other material systems such as $PbZrO_3$-$PbTiO_3$ (lead zirconate titanate, or PZT) and $(Ba,Sr)TiO_3$ (barium strontium titanate, or BST), have since followed suit and been developed for applications such as transducers, accelerometers, and ultrasonic generators [1].

Initial studies to utilize ferroelectrics in nonvolatile memories began in the 1960's, but due to film quality issues, their use was largely limited until the 1980's. However, by the 1990's, ferroelectrics began to enjoy extended applications in memories, RF/microwave devices, and other kinds of sensors, actuators, and systems [76].

Description of Ferroelectric Behavior: A Primer

The "ferro" aspect of the name *ferroelectricity* is actually a common misnomer, as it arose from the early analogies drawn between the properties of ferroelectrics and ferromagnetics; the presence of iron, or Fe, in the material is not at all a requisite for its ferroelectric properties. In short, a material possesses ferroelectricity if it can maintain a nonzero, switchable electrical polarization without the presence of an electric field. Switching from one polarization state to another occurs on the application of an electric field higher than the *coercive field*, which is the minimum required field needed to initiate switching [50].
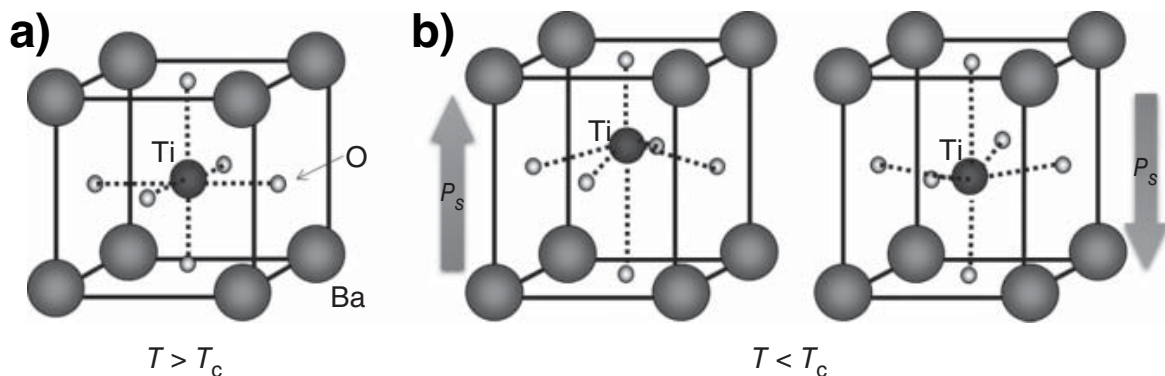
Figure 1.6: Crystal structure of barium titanate, $BaTiO_3$. (a) Above the Curie temperature, $T > T_C$, $BaTiO_3$ takes on the cubic crystalline structure, which is centrosymmetric. (b) At $T < T_C$, it is more thermodynamically favorable for $BaTiO_3$ to take on the tetragonal phase, in which the central Ti^{4+} ion shifts upward while the O^{2-} ions shift downward (or vice versa), thus causing a spontaneous polarization. Adapted from [33].

In all ferroelectric materials, there must necessarily exist space inversion symmetry breaking. In other words, the nonzero switchable spontaneous dipole moment of a ferroelectric *arises* from the absence of a center of symmetry in the crystallographic nature of

the material. However, for all ferroelectric materials, this property of noncentrosymmetry exists only below a certain critical transformation temperature known as the *Curie temperature*, or T_C; above each ferroelectric's unique Curie temperature, the material transforms into a paralectric state and loses the ability to have remanent polarization.

In 1954, A. F. Devonshire developed a phenomenological model of ferroelectrics according to the thermal, structural, and electromechanical properties of $BaTiO_3$ [50]. His work built directly upon the previous work of Landau and Ginzberg explaining phase transistions in different materials. This resulting Landau-Devonshire-Ginzberg theory of ferroelectrics to date remains one of the critical tools used to understand and model ferroelectric behavior. Devonshire chose to describe the thermodynamic state of a ferroelectric crystal in terms of the free energy:

$$U = \frac{1}{2}\alpha P^2 + \frac{1}{4}\beta P^4 + \frac{1}{6}\gamma P^6 + ... - EP \tag{1.1}$$

where α is assumed to be temperature dependent as $\alpha = \alpha_0(T - T_C)$, and the EP term describes the coupling to the applied electric field. As the free energy should not be contingent on the orientation of the polarization, the Taylor series expansion consists only of an even number of powers. Furthermore, this equation describes the behavior of the ferroelectric across all temperature ranges – even when it is in the paraelectric phase. From this free energy equation it is possible to derive the temperature-dependent relative permittivity of a ferroelectric material and many other quantities describing the ferroelectric [33].
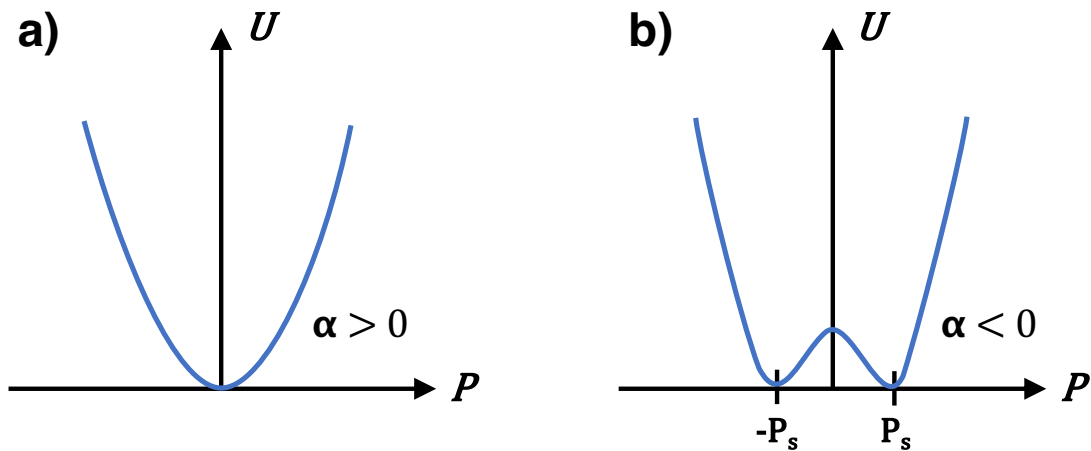
Figure 1.7: Qualitative depiction of the energy landscape for a crystal (a) with α parameter greater than 0 (where $T > T_C$), one stable minima at $P = 0$, and is therefore in the paraelectric state; and (b) with α parameter less than 0 (where $T < T_C$), with two stable minima in this case at $\pm P_S$, giving rise to its ferroelectric nature and hysteretic properties.

Following the Landau-Devonshire theory described above and based on the qualitative depictions of the associated energy landscapes for a paraelectric and ferroelectric crystal in Figure 1.7 (which excludes the effect of an external field), it is possible to understand the origin of hysteresis in ferroelectric materials. For the case where $\beta > 0$, corresponding to a second-order phase transition which entails the spontaneous polarization continuously decreasing to 0 as a function of increasing temperature, the free energy describing the material likewise evolves continuously as a function of decreasing temperature from the diagram shown in Figure 1.7(a) to 1.7(b).

As it is evident that there exist at least two minima for a ferroelectric below the Curie temperature with some energy barrier which separates them, one can now consider the effect of an applied external electric field, as depicted in Figure 1.8. This plot of measured ferroelectric polarization in response to applied electric field, or *P-E loop*, is a signature of any ferroelectric material. As the applied field is increased in magnitude beyond a certain amount required to surmount the energy barrier, the dipoles of the ferroelectric are then switched from one polarization orientation to the second. Hence, at zero applied field, the ferroelectric is expected to be sitting in one of two remanent polarization states. Note that the ferroelectric depicted in Figure 1.8 is highly idealized, and for real ferroelectrics the switching transition is not expected to be as sharp or symmetric.
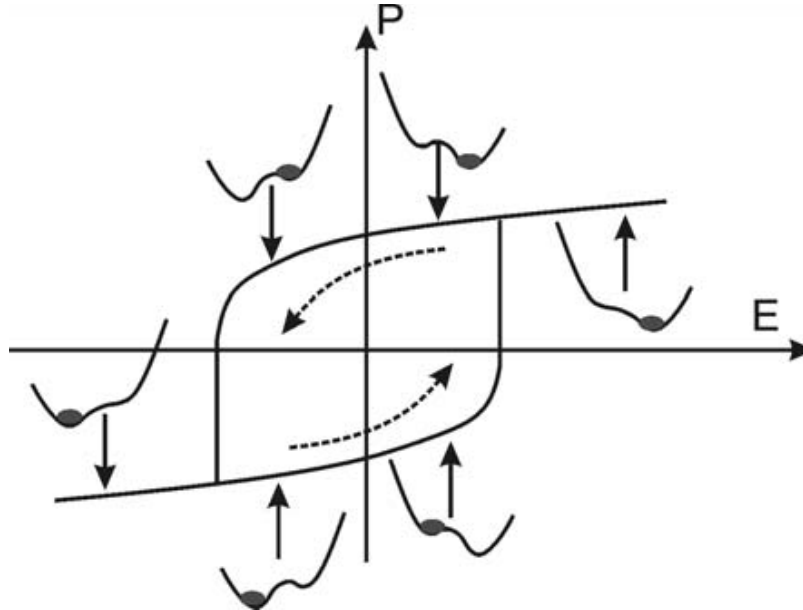
Figure 1.8: Idealized polarization vs. electric field characteristic of a ferroelectric, showing how polarization changes as a function of applied electric field E. Hysteresis arises from the switching from one stable minima to another. Figure adapted from [21].

Several chapters of this dissertation will focus on presenting and analyzing the behavior of ferroelectric materials when integrated into various device structures, and will go into detail about how ferroelectric properties are exactly captured and quantified.

1.4   Dissertation Organization and Research Objectives

This dissertation discusses the development, characterization, and applications of CMOS-compatible ferroelectric materials based on hafnium oxide (HfO_2) to realize next-generation emerging nonvolatile memories.

In **Chapter 2**, the material properties of doped HfO_2 material systems are explored, with an emphasis on zirconium and silicon doping. The influence of these dopants on the resulting ferroelectric properties are evaluated in addition to film parameters such as film thickness, annealing temperature, etc. Additional methods of materials characterization are also presented and discussed alongside electrical data.

In **Chapter 3**, the prospect of using HfO_2-based ferroelectric oxides as an insulator replacement for DRAM technologies is explored. Ferroelectric memory capacitors incorporating Si-doped HfO_2 are evaluated for their retention, imprint, and fatiguing properties. Possible mechanisms for ferroelectric capacitor device wearout are proposed.

In **Chapter 4**, the ferroelectric FET (FeFET) is introduced and analyzed. This chapter focuses entirely on the electrical characterization of the FeFET, with an emphasis on understanding the programmability of the FeFET fabricated using SOI (silicon on insulator) substrates.

In **Chapter 5**, reliability concerns surrounding the HfO_2-based FeFET are introduced and explored. Extensive time-dependent and temperature-dependent electrical characterization is employed to better understand the mechanisms driving device failure. Lastly, a technique to rapidly predict the endurance characteristics afforded by a particular gate stack design is proposed and evaluated against real device data.

In **Chapter 6**, a variety of processing techniques to improve the performance of HfO_2-based FeFET are explored and evaluated. These techniques rely heavily on interfacial oxide engineering and understanding its impact on device performance.

In **Chapter 7**, the FeFET is evaluated for prospective in-memory computing applications. A two-transistor FeFET content addressable memory cell (FeCAM cell) is demonstrated experimentally, and additional circuits and architectures based on the FeFET are also considered.

In **Chapter 8**, the main contributions of this work are summarized, and suggestions for future directions are given.

Chapter 2

Hafnium Oxide: A CMOS Compatible Ferroelectric

2.1 The Discovery of Ferroelectricity in HfO$_2$

In 2007, Intel spearheaded the first major redesign of the CMOS logic transistor since the late 1960's [13]. Until that point, the semiconductor industry had relied exclusively on the usage of silicon dioxide, or SiO$_2$, as the insulating gate dielectric in transistor technology. In order to continue shrinking the size of each generation of transistors, the principles of scaling dictated that the gate oxide thickness must also scale down in order to provide good gate control over the transistor channel. However, at that point in time, the scaling of SiO$_2$ had run into an impenetrable wall: at a mere five atoms in thickness, SiO$_2$ could not afford any further reductions, as direct tunneling of electrons through an impossibly thin dielectric would cause substantial off-current leakage and therefore wasted power. It was time for the semiconductor industry to move forward to another gate oxide material with better insulating electrical properties.

The solution required a total upheaval of existing semiconductor integration processes which involved not only the replacement of SiO$_2$ with a so-called *high-κ* dielectric, but also the replacement of polysilicon gates with metal. The high-κ used, after an investigation of a variety of different options ranging from Al$_2$O$_3$ to ZrO$_2$ to La$_2$O$_3$, was ultimately hafnium oxide.

Hafnium oxide has been known to exist in just three major crystalline phases at atmospheric pressure: the monoclinic phase (at room temperature), the tetragonal phase (at temperatures above 2050 K), and the cubic crystalline phase (at temperatures above 2803 K) [132]. Additional followup work on hafnium oxide in the following years then showed that the incorporation of dopants such as zirconium or silicon in pure hafnia could actually increase the dielectric constant of HfO$_2$ further than what would be expected from pure monoclinic phase HfO$_2$ by inducing the stabilization of the tetragonal phase at room temperature as well [15, 126].

Then, in 2011, Böscke et al. showed that the monoclinic phase formation could be greatly inhibited if a mechanical capping layer was present on the doped HfO_2 film during crystallization, leading to the formation of the *orthorhombic* crystalline phase of HfO_2 [14]. This was considered to be a breakthrough discovery, as until that point, all of the commonly reported bulk crystalline phases of HfO_2 were centrosymmetric. As discussed in the ferroelectrics behavior primer in Chapter 1 of this dissertation, the existence of ferroelectricity in a material is very closely tied to its crystalline structure; non-centrosymmetry is a requisite for properties such as ferroelectricity or piezoelectricity to exist in a material. The process of inducing the orthorhombic phase in HfO_2 is illustrated accordingly in Figure 2.1.
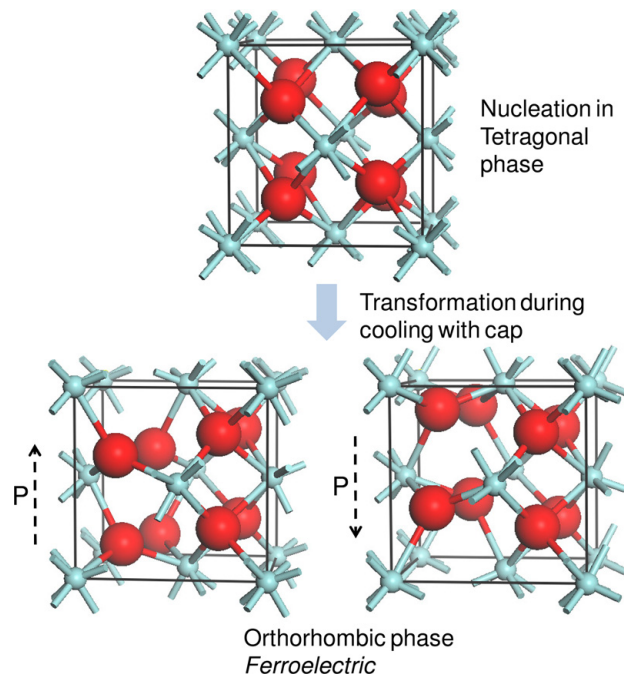
Figure 2.1: The addition of dopants to a pure hafnia film enables the formation of the tetragonal phase, which can then stabilize the transition to the previously metastable orthorhombic crystalline phase of HfO_2 which is intimately tied to ferroelectricity. Adapted from [14].

Though previously identified orthorhombic phases of HfO_2 were known to exist (space groups Pbca and Pbcm), they were identified as centrosymmetric and therefore could not be a possible origin for the ferroelectric properties. However, the diffraction pattern observed matched that of a rare orthorhombic phase with space group $Pbc2_1$ in Mg stabilized

ZrO_2. As ZrO_2 and HfO_2 are extremely similar in crystal chemistry, it is also very probable that they have the same crystalline phases.

The role of the dopant in hafnium oxide was initially found to be a critical component in inducing ferroelectric properties in the resulting film. It was discovered that the inclusion of a dopant greatly elevated the crystallization temperature and helped to promote a controlled crystallization into the orthorhombic phase. The dopant also helps to reduce the stability of the bulk monoclinic phase of HfO_2, which as mentioned previously has both a lower dielectric constant and is centrosymmetric. To date, several dopants have been reported to successfully stabilize ferroelectric properties in HfO_2. Among them include silicon, as discussed in [14, 87]; yttrium [88], aluminum [91], gadolinium [82], lanthanum [116], and many others, as shown in Figure 2.2. In fact, several other dopants yet untested are predicted to also induce ferroelectricity in hafnium oxide. It is therefore concluded that the presence of dopants influence the phase stability of various crystalline phases of HfO_2, achieving the overall goal of *destabilizing* the monoclinic crystalline phase at room temperature [86].
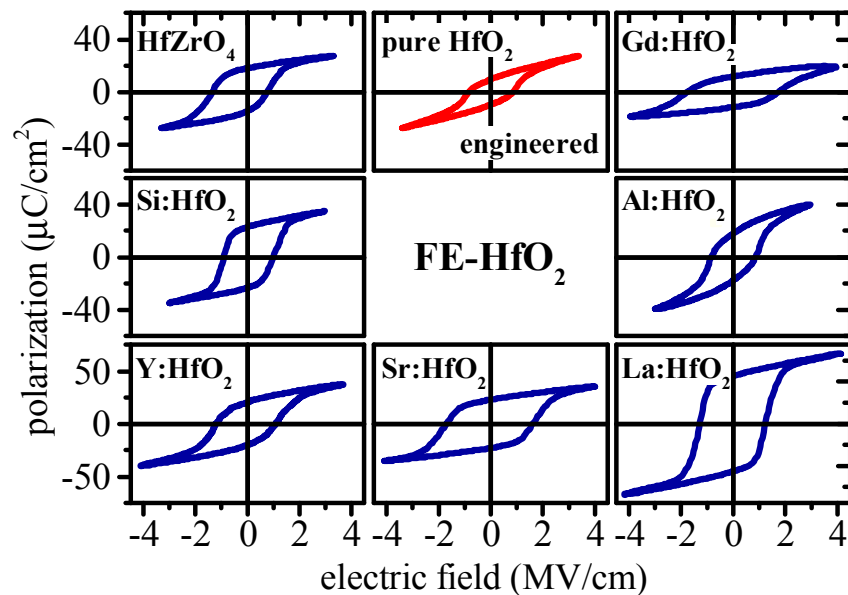
Figure 2.2: Many different dopants have been empirically found to stabilize the ferroelectric properties in HfO_2, and to varying degrees. Adapted from [89].

Following the line of reasoning presented above, it is therefore very probable that the ferroelectric properties of hafnium oxide could also be induced through other factors which similarly influence the stability of the monoclinic phase at room temperature – and that ferroelectricity in HfO_2 is not exclusively triggered by the presence of dopants. In

fact, under particular processing conditions, ferroelectric properties could also be demonstrated in pure HfO_2 completely devoid of dopants, as discussed in [107]. The authors in [107] investigated a thickness series of pure HfO_2 thin films ranging from 4 to 20 nm integrated into symmetric MFM (metal-ferroelectric-metal) capacitors with titanium nitride (TiN) as electrodes. Ferroelectric properties were distinctly identified in samples ranging from 6 to 12 nm, with the 4 nm film remaining amorphous regardless of crystallization anneal and the thicker films becoming increasingly monoclinic in phase fraction. The authors concluded that the deposition of the top TiN electrode well below the crystallization temperature of the film played a substantial role in stabilizing the ferroelectric properties once the films were subjected to the high-temperature crystallization anneal. The thicker films which showed vanishingly small ferroelectric properties were more prone to premature crystallization effects during the long thermal budget of the ALD (atomic layer deposition) process, and therefore contained a higher and higher volume fraction of the monoclinic phase. This study suggests that phase manipulation could result from grain size engineering as well as mechanical encapsulation of the film with a metal capping layer.

Clearly, it is also possible to utilize the metal capping layer as an experimental parameter in reducing the possible partial transformation of the underlying film into the monoclinic phase as well. With the presence of the capping layer, the shearing of the unit cell of the crystal is inhibited mechanically, thus leading to a possible transformation to the orthorhombic $Pbc2_1$ crystalline phase [14]. Much like the way that electrode materials influence the properties of traditional perovskite ferroelectrics, it is also expected that choice of electrode will affect factors such as interfacial oxygen vacancy levels, lattice matching, and so on in HfO_2-based ferroelectrics. Whereas the bottom electrode on which the films are grown serves as the substrate and therefore can act as a nucleation layer/template for lattice matching, the top electrode cannot affect the growth conditions but can impact the residual stress in the film as well as the thermal budget witnessed during top metal deposition [97]. To date, the most commonly investigated electrode materials are TiN and TaN [73, 90], but other materials such as Pt, Ir, Ru, Ti/Au, Ti/Pd, and W have also been investigated [19, 61, 72, 99, 113].

2.2 Investigation of Zr-Doped HfO_2

Though the binary oxides of HfO_2 and ZrO_2, in isolation, have been extensively studied, it was only in 2012 where it was discovered that the two combined over the entire compositional range could result in films that were primarily tetragonal, orthorhombic, or monoclinic in phase. This directly translates to films across the compositional spectrum possessing antiferroelectric, ferroelectric, or completely dielectric properties – depending on the percentage of ZrO_2 present, which could vary from 0% (pure hafnia) to 100% (and therefore pure zirconia) [90].

Fabrication of Zr-Doped HfO$_2$ Thin Films

The most commonly preferred technique used to grow thin hafnium oxide films is ALD, or atomic layer deposition. Atomic layer deposition is a growth technique which allows for monolayer by monolayer growth of a material, enabling the deposition of films with precise thickness control. Atomic layer deposition has become a preferred industry standard for film growth as it allows for extremely conformal thin films, therefore facilitating the realization of 3D architectures and readily scalable devices.
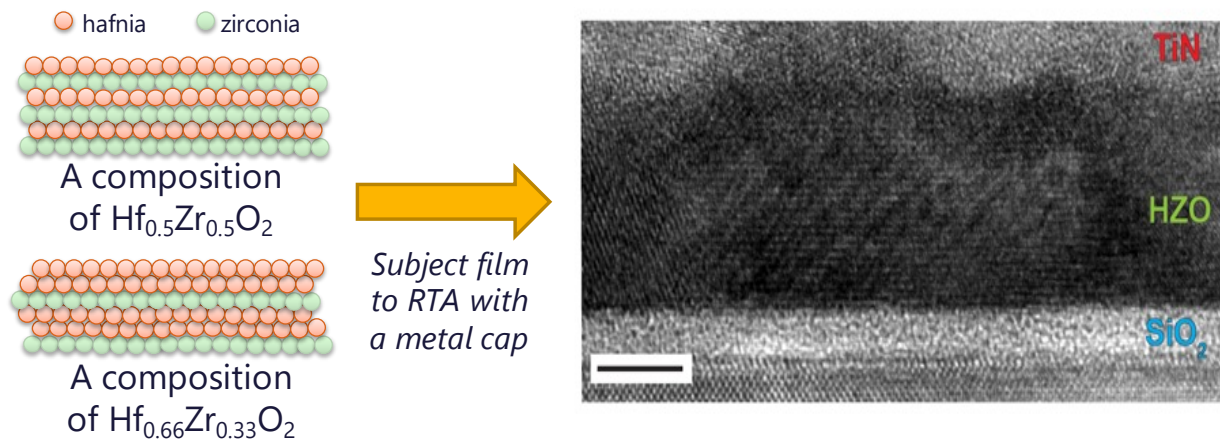
Figure 2.3: A cartoon of monolayer by monolayer growth of an ALD film is shown on the left. The stoichiometry of the resulting film is achieved simply by varying the ratio of hafnia to zirconia monolayers. After ALD growth, the film is subjected to a high temperature crystallization anneal, confined by a metal capping layer, to achieve a polycrystalline film as confirmed by TEM (scale bar = 5 nm). TEM image is adapted from [61].

Atomic layer deposition of Zr-doped HfO$_2$, otherwise referred to as HZO or Hf$_x$Zr$_{1-x}$O$_2$, is performed in a Fiji Ultratech/Cambridge Nanotech tool. The chuck, chamber, and precursor manifold are all heated to a temperature of 250 °C for deposition. The respective hafnium and zirconium precursors used are tetrakis(dimethylamino)hafnium (TDMAHf) and tetrakis(dimethylamino)zirconium precursors (TDMAZr), which are heated to 75 °C, and water vapor is used as the oxidant.

For devices characterized in the metal-ferroelectric-metal structure (MFM), the TiN substrates are first prepared by sputtering TiN on a Si wafer covered by 100 nm of thermal SiO$_2$. For devices characterized in the metal-insulator-ferroelectric-semiconductor structure (MFIS), the Si substrates (p-type, and heavily doped to 10^{19}/cm^3) are prepared by first removing the native oxide through a 1 minute diluted 10:1 HF dip, and cleaned of organic substances by a 10 minute dip in sulfuric acid heated to 120 °C. Then, an intentional thin 1.5 nm of SiO$_2$ is regrown by rapid thermal oxidation (by annealing the freshly

cleaned substrate in an RTA, or rapid thermal annealing, chamber). The oxidation condition is in 100% O_2 ambient, at 900 °C for 20 seconds.

Following the deposition of $Hf_xZr_{1-x}O_2$ as described above and as depicted in Figure 2.3, 10 - 30 nm of TiN is sputtered on the device stack, which is subsequently annealed in the RTA chamber at temperatures between 500 °C - 700 °C for 30 seconds. TiN is then chemically etched away in a solution of SC1, a solution which is comprised of H_2O_2:NH_4OH:H_2O (hydrogen peroxide, ammonium hydroxide, and water) in a ratio of 1:1:10, at 50 °C. Finally, Ti/Au electrodes are patterned on the $Hf_xZr_{1-x}O_2$ stacks for electrical characterization through a combination of lithography and lift-off in acetone. The device stacks for samples grown on TiN and Si are shown in Figure 2.4(a) and (b), respectively.
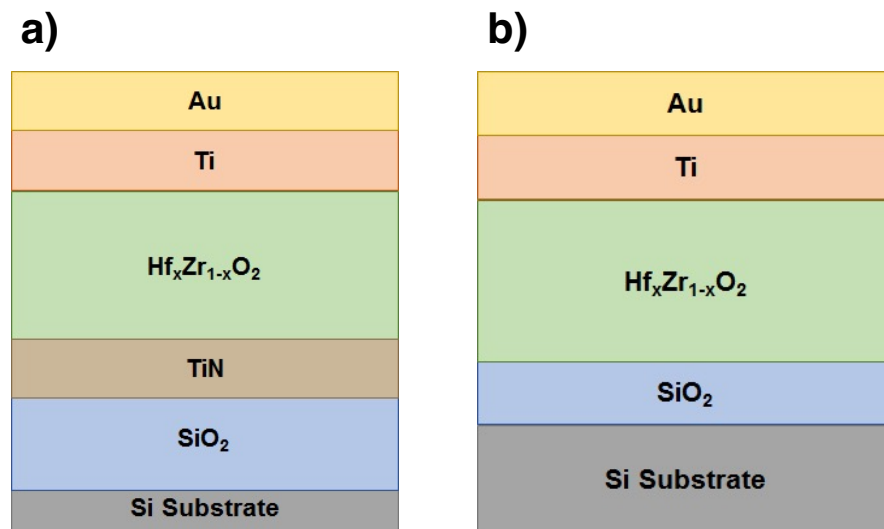
Figure 2.4: 2D cross-sectional cartoons for the device stacks characterized incorporating $Hf_xZr_{1-x}O_2$. (a) Device stack in the MFM (metal-ferroelectric metal) structure, fabricated on a SiO_2 wafer. (b) Device stack in the MFIS (metal-ferroelectric-insulator-semiconductor) structure.

Electrical and Material Properties

As shown in Figure 2.5, PE loop measurements performed on $Hf_xZr_{1-x}O_2$ films with Zr concentrations from 0% to 50% reveal their ferroelectric nature. However, Zr-rich (concentrations of greater than 50%) films on TiN substrates exhibit antiferroelectric characteristics, as seen in Figure 2.5(b).

As discussed previously, ferroelectricity in HfO_2 has been attributed to the stabilization of the non-centrosymmetric orthorhombic phase. The grazing-incidence X-ray diffraction

a)

b)
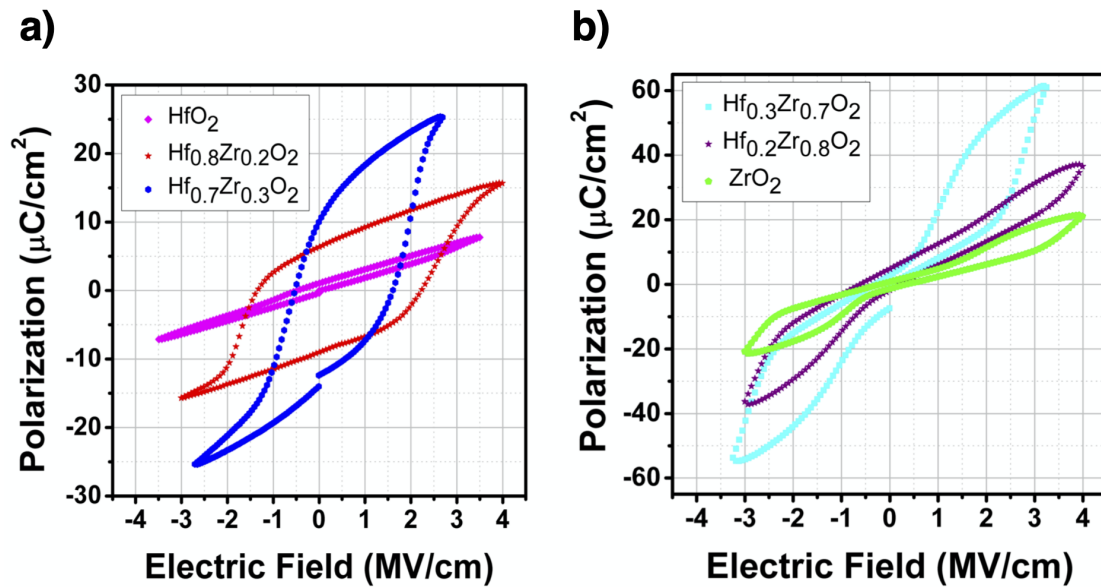
Figure 2.5: Polarization vs. electric field (PE loop) characteristics for (a) ferroelectric Hf-rich films on a bottom electrode of TiN, annealed at a temperature of 500 °C, and (b) antiferroelectric Zr-rich films on a bottom electrode of TiN, annealed at 500 °C.
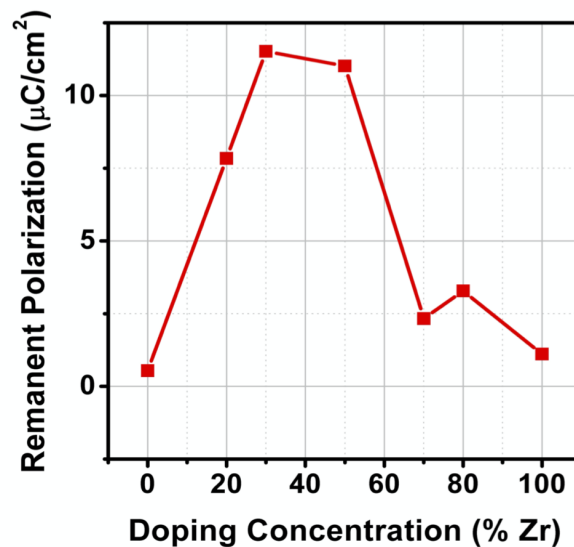
Figure 2.6: Remanent polarization vs. Zr doping concentration for $Hf_xZr_{1-x}O_2$ films on TiN, annealed at 500 °C.

(GIXRD) data from the fabricated films confirms the presence of this phase in the poly-crystalline layer, as seen in Figure 2.7(a). The stabilization of this particular phase is aided by the mechanical confinement provided by the TiN capping layer, which suppresses a transition to the monoclinic phase during cooling [86].

However, it is noteworthy that films that are Zr-rich typically appear to be more anti-ferroelectric in nature *despite* the presence of the TiN capping layer. The increased concentration of Zr induces the formation of the tetragonal phase upon cooling (Figure 2.7(b)), which is consistent with the observation that antiferroelectricity in pure ZrO_2 is linked to the presence of the tetragonal phase [111]. The remanent polarization as a function of Zr concentration is also presented in Figure 2.6 as a metric for how increasing Zr doping concentration changes the dielectric nature of the film from paraelectric to ferroelectric to antiferroelectric.
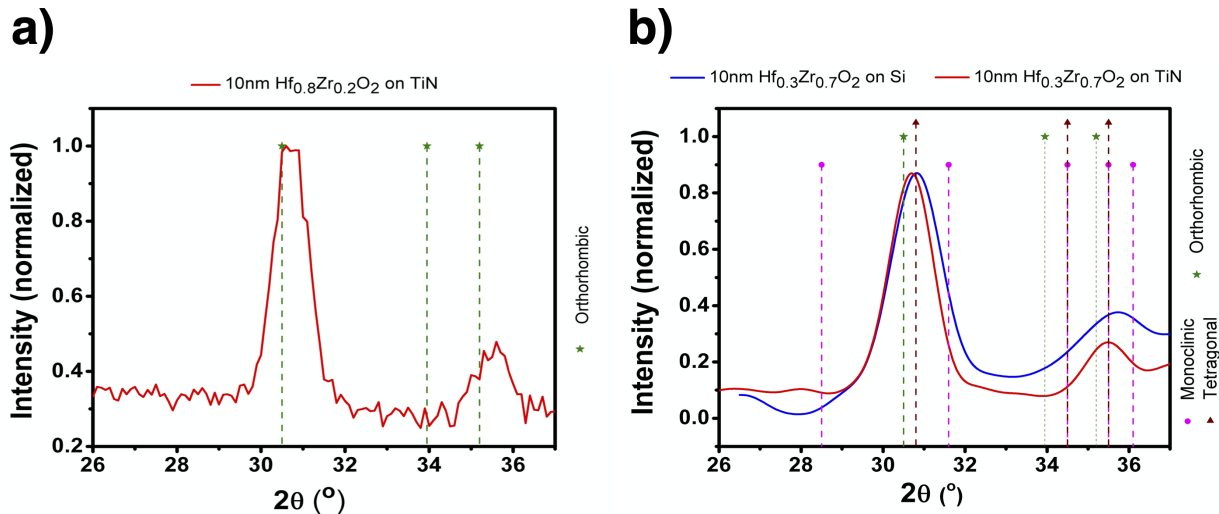
Figure 2.7: Grazing-incidence X-ray diffraction (GIXRD) data for (a) $Hf_{0.8}Zr_{0.2}O_2$ on TiN substrate and b) $Hf_{0.3}Zr_{0.7}O_2$ on both Si and TiN substrates

Annealing temperature also plays an important role in determining the resulting film properties. Higher annealing temperatures can enhance grain boundary development and diffusion of electrode materials, resulting in increased leakage pathways and the formation of so-called "dead layers" [72]. As seen in Figure 2.8(a), for an $Hf_{0.5}Zr_{0.5}O_2$ film grown on TiN substrate, the 700 °C temperature anneal results in a film with a higher coercive field. This result supports the observation that higher temperature anneals do indeed stimulate the growth of non-ferroelectric interfacial "dead layers".

In contrast to the antiferroelectric Zr-rich samples on TiN substrates discussed above, Zr-rich films on Si substrates show ferroelectric behavior at an annealing temperature of 500 °C (Figure 2.8). The underlying reason for this phenomenon is intimately linked to the
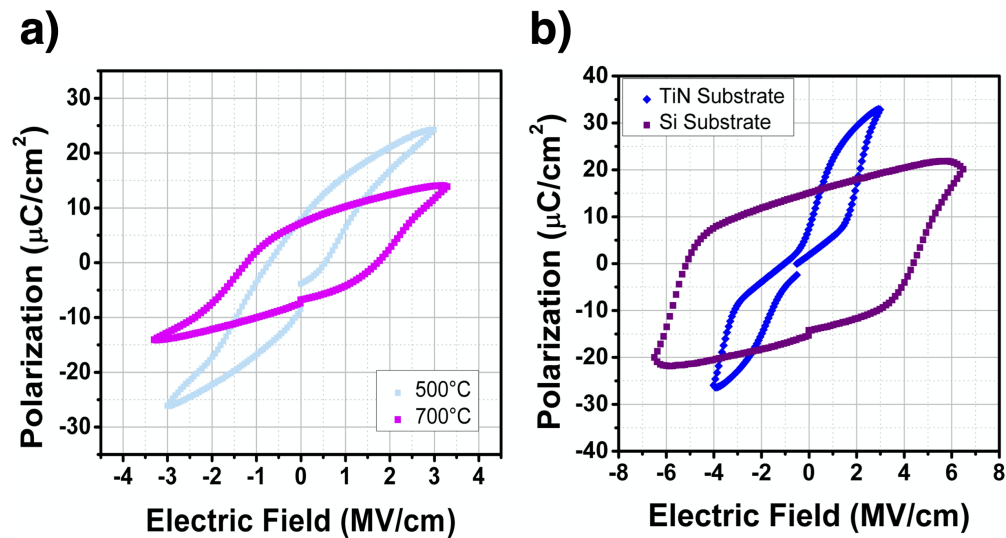
Figure 2.8: Comparison of different processing conditions. (a) For a $Hf_{0.5}Zr_{0.5}O_2$ film on TiN substrates, the ferroelectric properties can vary significantly if the film is annealed at different temperatures. (b) For a $Hf_{0.3}Zr_{0.7}O_2$ film annealed at a fixed temperature of 500 °C, it appears to be ferroelectric on a Si substrate but antiferroelectric on a TiN substrate.
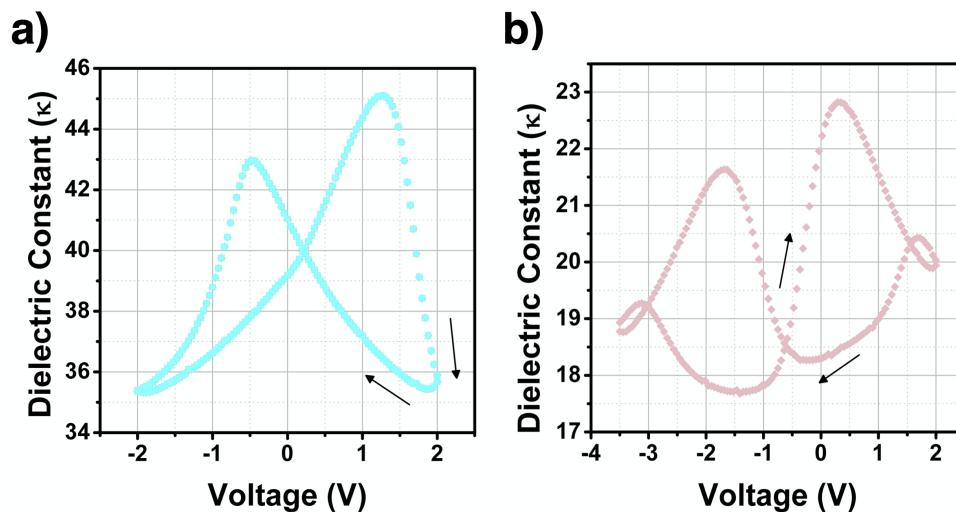
Figure 2.9: Extracted dielectric constant data from C-V measurements with an applied AC signal at 100 kHz, taken on (a) a ferroelectric $Hf_{0.7}Zr_{0.3}O_2$ film grown on TiN substrate at 500 °C, and (b) an antiferroelectric $Hf_{0.3}Zr_{0.7}O_2$ film grown on TiN substrate at 500 °C.

difference in lattice constants of the crystalline films on either Si or TiN substrates, which as discussed previously, act as a template for the deposited film to grow upon.

C-V measurements have also been performed on the fabricated structures to extract dielectric constant data. As can be seen in Figure 2.9(a) and (b), the peaks in the dielectric constant profile are indicative of ferroelectric/antiferroelectric switching. In contrast to the double-peaked "butterfly" loop characteristic exhibited by the ferroelectric material in Figure 2.9(a), the antiferroelectric material in Figure 2.9(b) displays a total of 4 switching transitions.

Though Zr-doped HfO_2 remains a heavily studied materials system due to the fine-grained control achievable over its ferroelectric/antiferroelectric/dielectric properties, thanks to the fact that HfO_2 and ZrO_2 form solid solutions across the entire compositional range, its one fatal drawback is its limited thermal processing range. As discussed in the previous subsection, annealing temperatures must be kept in the vicinity of 500 °C, in order to avoid the creation of additional "dead layers", increased leakage through the film due to diffusion of electrode materials and/or increased crystallization, etc. Therefore, it is highly desirable to investigate other doped HfO_2 systems which may allow for an increased thermal budget to accommodate typical CMOS back end of the line (BEOL) processing, or any additional post-processing steps which occur at elevated temperatures for extended periods of time.

2.3 Investigation of Si-Doped HfO_2

The process flow to realize the MFM devices to characterize Si-doped HfO_2 is shown in Figure 2.10, which comprises a simple 4-step process.

Form back electrode of ~10 nm Iridium by sputtering

Deposit Si-doped HfO_2 by ALD

Form top electrode of ~10 nm Iridium by sputtering through a shadow mask

Anneal stack using RTP tool

10 nm Ir   10 nm Ir

Si:HfO$_2$

10 nm Iridium
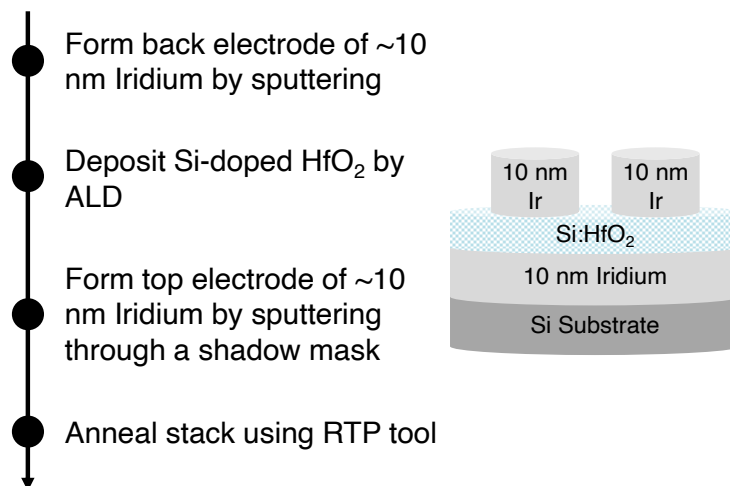
Si Substrate

Figure 2.10: Fabrication process flow to realize Si-doped MFM capacitors for study.

Silicon substrates were first sputtered with a blanket layer of 10 nm iridium for the back electrode. The ferroelectric HfO_2 layer was grown by ALD using alternating growth cycles of HfO_2 and SiO_2, using ozone as the oxidant, appropriately chosen to achieve specific thicknesses and doping concentrations. Film thicknesses from between 4 nm to 8 nm were grown, targeting a silicon doping concentration in the range of 5-6 mol% Si to maximize the ferroelectric properties of the resulting film [72]. Next, 10 nm iridium was sputtered on top of the stack through a shadow mask with 200 μm diameter holes to form an array of capacitors. The samples were then diced and annealed under varying conditions.
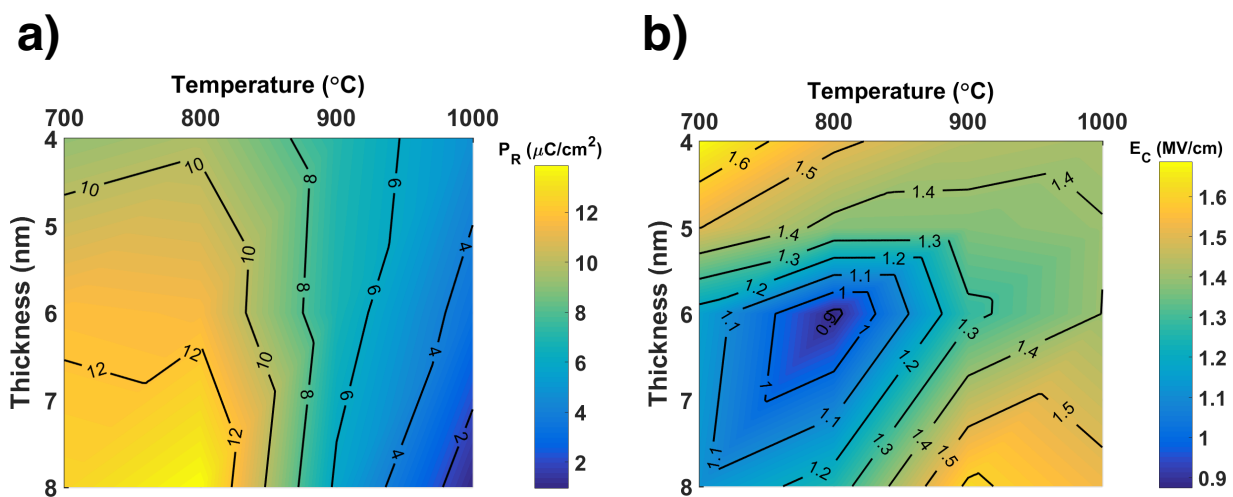
Figure 2.11: (a) 2D contour map showing the dependence of remanent polarization on annealing temperature and Si-doped HfO_2 film thickness. (b) 2D contour map for coercive field dependence on annealing temperature and HfO_2 film thickness.

All of the samples with various film thicknesses were separately annealed at temperatures ranging from $600 - 1000\,°C$, and afterwards were measured with a Radiant Precision LC II Ferroelectric Tester. From the measured PE loops, the P_R (remanent polarization at zero applied voltage) and E_C (coercive field at which ferroelectric switching initiates) were extracted. Both of these ferroelectric parameters are plotted as a function of anneal temperature and film thickness in 2D contour plots shown in Figure 2.11. In general, it appears that remanent polarization can be maximized by choosing an anneal temperature on the lower end of the range investigated, and choosing a film thickness on the higher end. However, for coercive field strength, the point where it appears to be minimized is somewhere in the middle of the thickness and temperature ranges investigated. HfO_2 is well-known to possess an extraordinarily high coercive field in comparison to more traditional ferroelectrics [86], and therefore it will also be imperative to understand where in the process parameter tradeoff space, for example, the lowest coercive field strength

can be achieved in order to realize low-voltage operation of a device which integrates this material.
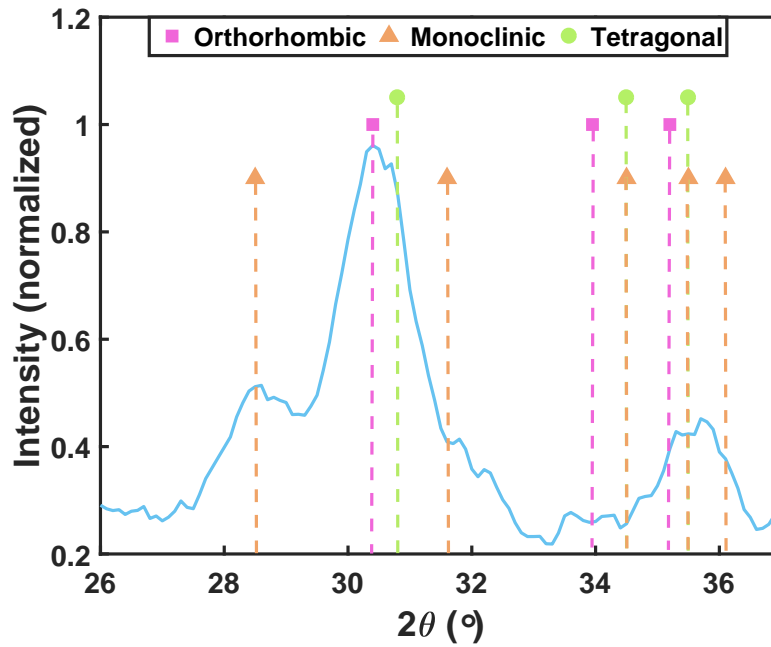
Figure 2.12: GIXRD spectra of a 4 nm Si-doped HfO_2 film, showing strong presence of the orthorhombic crystalline phase.

Film crystallinity is verified through X-ray diffraction (XRD) as shown in Figure 2.12. The representative 4 nm film shows its strongest intensity strong peak around 30.5°, which corresponds to the orthorhombic crystalline phase. However, as expected, other crystalline phases are also clearly present in the spectra due to the polycrystalline nature of thin HfO_2 – namely, the tetragonal and monoclinic phases.

The polycrystalline nature of the film can also be verified in Figure 2.13, which shows a TEM image of a ferroelectric HfO_2 capacitor device with a ferroelectric layer thickness of roughly 8-9 nm. Due to the use of an imperfect shadow mask to deposit the Ir top electrode to complete the capacitor structure, there exists some variation in the thickness of the top electrode. The measured thickness of the ferroelectric HfO_2 layer is also impacted by the presence of a slight interfacial oxide layer between the top Ir electrode and the HfO_2 film. The presence of this interfacial layer will be discussed in further detail in Chapter 3, which will delve into the electrical characterization of the Si-doped HfO_2 capacitors for memory applications.

Lastly, the measured mol % of Si doping in the characterized Si-doped HfO_2 films as determined by XPS (x-ray photoelectron spectroscopy) are tabulated in Table 2.1. For virtually all other dopants in HfO_2 aside from Zr, it is important to control the doping
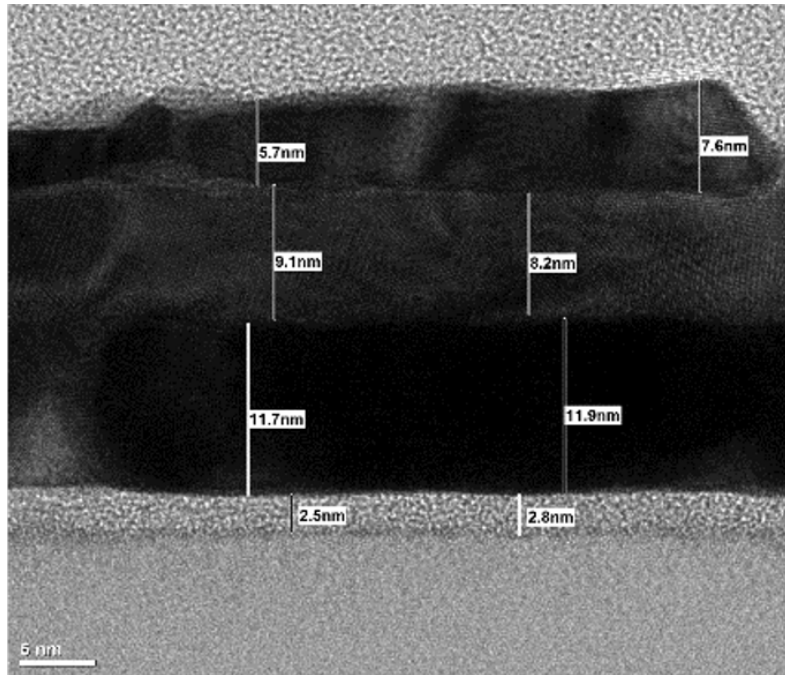
Figure 2.13: TEM image of a representative 8-9 nm thick Si-doped HfO_2 memory capacitor with iridium electrodes. TEM scale bar is 5 nm.

Sample Thickness	Measured mol % of Si Doping
4 nm	5.77%
5 nm	4.42%
6 nm	4.69%
8 nm	5.35%

Table 2.1: XPS data showing mol % concentration of Si in ferroelectric HfO_2 (an indicator of ferroelectricity).

percentage to within a small mol % of the total film composition in order to achieve the desired ferroelectric properties. For a Si-doped HfO_2 film on the order of 9 nm or less in thickness, this range appears to be between 3-7 mol %, with peak remanent polarization values achieved close to 4-5 mol % [147].

In comparing the results on Si-doped HfO_2 to Zr-doped HfO_2, the biggest advantage is perhaps the elevated thermal budget, allowing for more flexibility in processing. In terms of scalability and ease of doping, Zr-doped HfO_2 has the edge, with a 1:1 ALD cycle ratio of hafnia to zirconia ostensibly allowing for extreme ease of doping at highly scaled thickness regimes. Indeed, recent work has shown the scalability of Zr-doped HfO_2 to

the sub-2 nm regime [24]. The ability to scale becomes substantially more difficult for HfO_2 systems which require no more than just a few mol % doping in order to achieve ferroelectricity.

Chapter 3

Ferroelectric HfO$_2$-based RAM (FeRAM)

3.1 DRAM and its Technology Challenges

In the mid-1960's, IBM scientist Robert Dennard came up with the basis of the DRAM (dynamic random access memory) cell, which consists of one capacitor as a storage element and a transistor which controls the amount of charge stored on the capacitor [135]. This memory cell is often referred to as the "1T1C" or one-transistor, one-capacitor cell architecture, as shown in Figure 3.1.
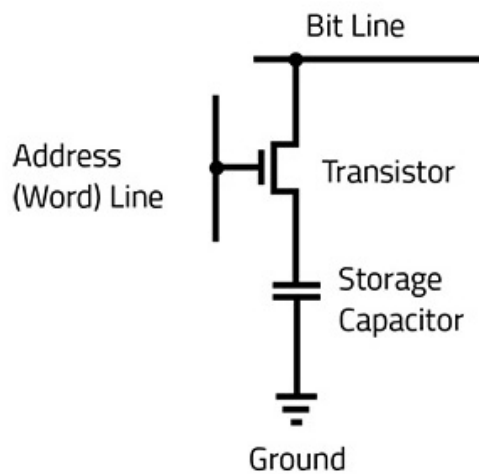
Figure 3.1: A one-transistor, one-capacitor (1T1C) DRAM memory cell. Figure adapted from [143].

The word line, or address line, is connected to the gate of the access transistor, and the bit line is connected to the source of the transistor. When data is written to or read from the storage capacitor, the word line is activated to provide access to the capacitor. The bit line is then used to either sense the amount of charge written in the storage capacitor during a read operation, or to provide a voltage when a new value is to be written to the capacitor [42]. The different current levels read out on the bit line correspond directly the amount of charge stored on the capacitor itself, thus allowing the encoding of a binary logic state 1 or 0. Therefore, it is apparent that the capacitor must be large enough to able to store a measurable amount of charge in order to distinguish between two written states.

Furthermore, one caveat of the DRAM cell is that the imperfect access transistor provides a leakage path through which charge can "leak" away from the storage capacitor. This means that, in order to maintain the written state of the DRAM capacitor, the DRAM memory must be periodically refreshed to maintain the stored data. Therefore, unlike the other nonvolatile memories introduced in Chapter 1 of this dissertation, DRAM is a kind of *volatile* memory solution, in which the stored data will be lost when the system is powered off and dynamic refreshing (and hence, the origin of the name **dynamic**) cannot be performed.
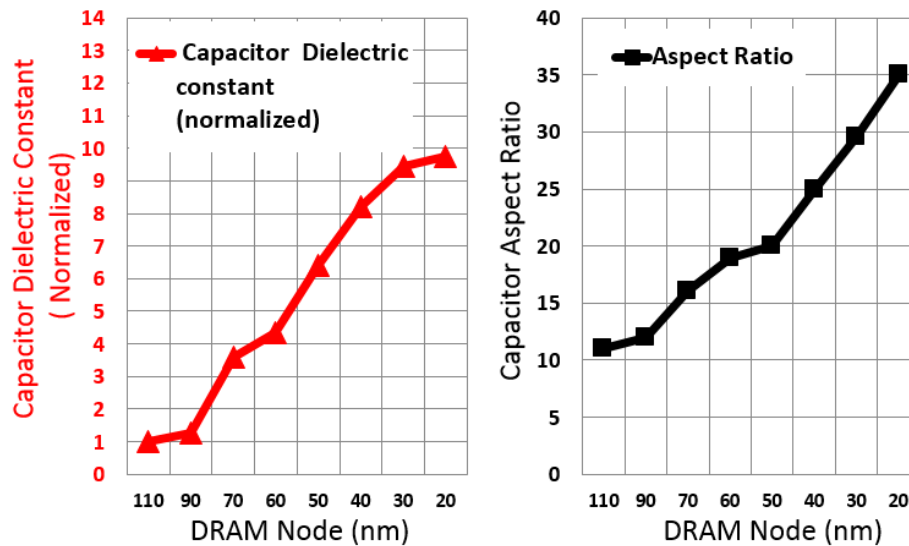
Figure 3.2: Scaling trends in DRAM technology across several technology generations. On the left, the increasing dielectric constant κ trend enables greater charge storage for a given unit area. On the right, increasing aspect ratios point to deeper and deeper DRAM trench capacitors, pointing to increased vertical scaling. Figure adapted from [35].

Today, DRAM remains the standard technology used for main memory in computers. It has provided superior storage densities over other potential solutions, due to its compact architecture as described above, and is low cost. The need for increased stor-

age density has prompted DRAM technologies to continue scaling in the vertical, rather than lateral, direction on chip in order to increase the capacitor size (and therefore storage capacity). DRAM's scaling path has therefore involved more and more aggressive lithography requirements as feature sizes scale down to be able to make deep, yet narrow trench capacitors. Simultaneously, there have been efforts with each technology generation to increase the κ-value of the dielectric used in the storage capacitor in order to improve charge storage in DRAM. These aforementioned trends are captured in Figure 3.2.

Historically, DRAM has been able to surmount both the device and process integration challenges at every new technology node, for well over 50 years to date [58]. However, it has become apparent in recent years that DRAM will run into substantial difficulties to overcome scaling below the 10 nm node and to continue delivering on future performance requirements [38, 120]. In order to continue patterning DRAM capacitors with smaller and smaller feature sizes, the industry eventually must adopt advanced EUV, or extreme ultraviolet lithography, to pattern feature sizes at 13 nm resolutions. However, other technical challenges may impede ultimate DRAM scaling even before the adoption of EUV patterning becomes required [35, 38].

FeRAM as a Nonvolatile DRAM Option

FeRAM (ferroelectric random access memory) technology based on established perovskites such as PZT (lead zirconate titanate), SBT (strontium bismuth titanate), and others have been in production for many years and are well-established in niche applications [80]. FeRAM technology bears a very similar resemblance to DRAM technology, requiring one access transistor and a (ferroelectric) storage capacitor as well. However, the primary reason for their lackluster widespread adoption stems primarily from the fact that conventional perovskites suffer from thickness scaling issues – and worse yet, lack of available deposition techniques for these materials for 3D nanostructures [43]. These issues have forced the lateral footprint of the ferroelectric capacitor to remain rather large due to the inability to scale PZT, for example, to below 30 nm in thickness, and therefore providing little benefit in comparison to standard planar approaches [83]. Scaling of FeRAMs was therefore largely limited to the 130 nm technology node.

With the advent of ferroelectric HfO$_2$ in recent years, FeRAMs can now be revisited for a direct DRAM comparison; to first order, both the scaling and deposition issues are resolved through the usage of HfO$_2$, which enables highly conformal ALD ferroelectric films to be grown, and ferroelectric properties to persist in the sub-10 nm regime. Furthermore, compared to standalone DRAM, it is possible to reduce the capacitor area by a factor of 8 due to the higher polarization charge density achievable by ferroelectric HfO$_2$, thereby reducing as well the associated cost per bit of storage [122].

In the following sections, the performance of Si-doped HfO$_2$ integrated into an FeRAM capacitor is explored and evaluated.

3.2 Si-doped HfO$_2$ Memory Capacitors

Details relating to the fabrication and materials characterization of Si-doped HfO$_2$ thin films can be found in Chapter 2. This section will focus on the electrical characterization of Si-doped HfO$_2$ capacitors for memory applications. Standard memory tests such as fatigue, retention, and imprint were conducted on the samples.

Endurance Testing

For every film thickness investigated from 4 to 8 nm, endurance/fatigue testing was conducted. All films are first given 10^4 cycles to "wakeup" the film at an appropriate voltage required to saturate the remanent polarization P_R fully to allow the ferroelectrics to reach their final characteristics. The wakeup effect is a well-documented effect in ferroelectrics based on HfO$_2$, and it specifically refers to the increase in remanent polarization with field cycling until the onset of fatigue [62, 69, 151]. The 10^4 cycles are counted into the fatigue limit accordingly.

Then, each film was cycled with a 10 kHz bipolar voltage pulse train of ± 2.5 MV/cm until failure, or 10^9 total stress cycles was reached. The magnitude of voltage stressing was calculated according to the total desired field stress of ± 2.5 MV/cm and the thickness of the ferroelectric film in TEM.
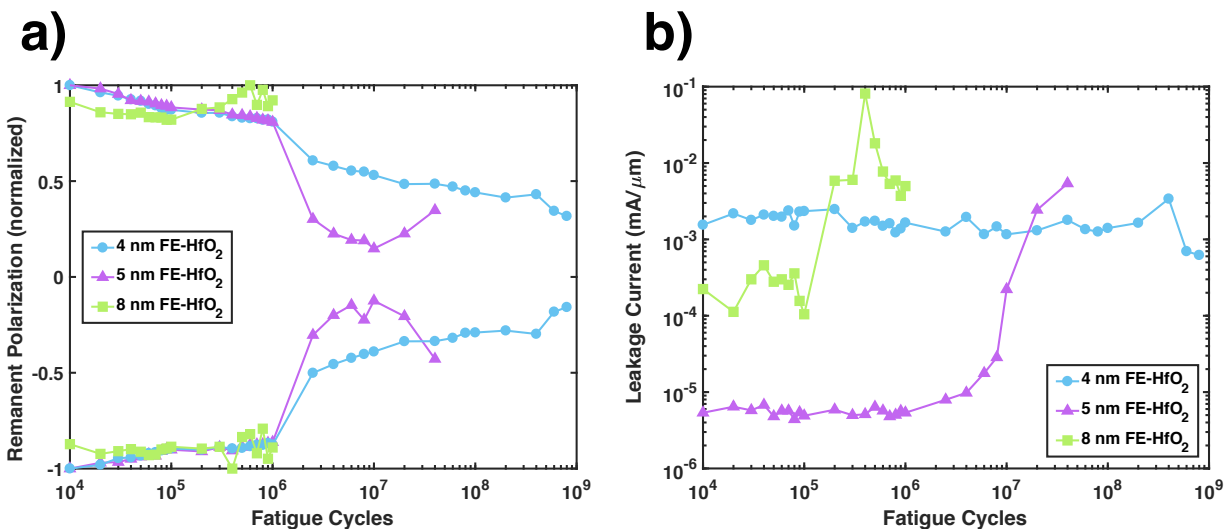
Figure 3.3: Endurance characteristics for Si-doped HfO$_2$ films of varying thicknesses. (a) Normalized remanent polarization for 4, 5, and 8 nm ferroelectric memory devices vs. fatigue cycles. (b) Leakage current through the 3 respective films vs. fatigue cycles to show correlation between film leakage over time and onset of device failure.

As seen in Figure 3.3(a), the write endurance appears to be inversely proportional to film thickness, with the 4 nm film witnessing a gradual decrease in remanent polarization over its lifetime and the 8 nm film failing abruptly after 10^6 cycles. Notably, this behavior is correlated with the magnitude of the leakage current through the film, as shown in Figure 3.3(b). Though the thinnest film starts off with the highest leakage current, during the cycling process, the magnitude of this current remains more or less constant. In contrast, the thicker 5 nm and 8 nm films witness cycling breakdown events which strongly correlate with a several-orders-of-magnitude jump in leakage current through the film. Device failure for the thicker films, in this scenario, is tied to the onset of hard breakdown, which causes irreversible damage to the insulating properties of the ferroelectric oxide.

a)   **b)**
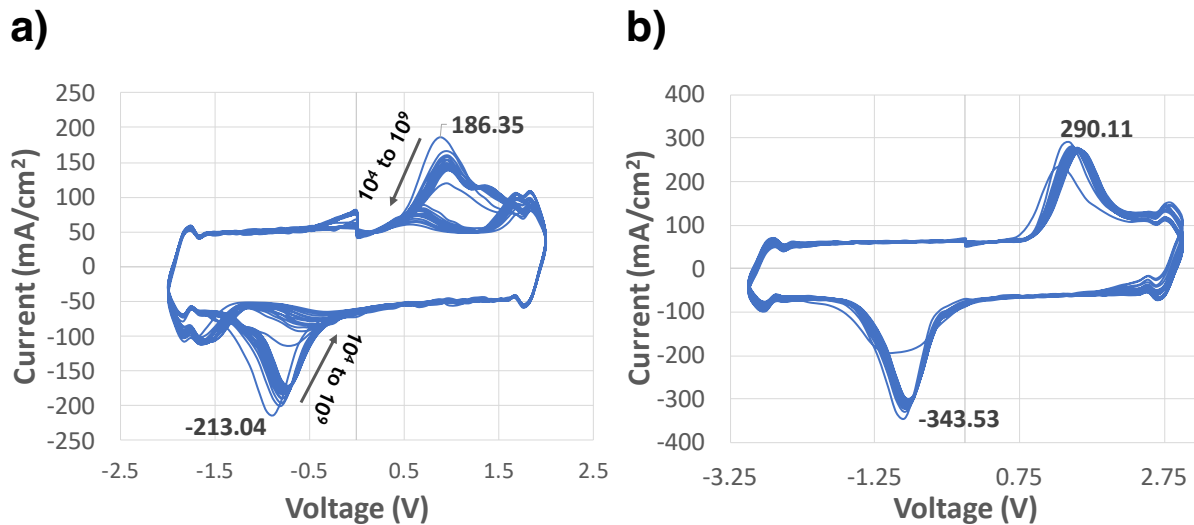
Figure 3.4: Switching current evolution and hysteresis for two Si-HfO₂ films. (a) Current through a 4 nm thick film, showing a switching peak which diminishes in magnitude as the film is continually cycled. (b) Current through a 8 nm thick film, showing that the peak switching current stays mostly constant up to hard breakdown.

These aforementioned observations, coupled with the fact that the magnitude of the peak switching current for the 8 nm film is $\sim 343\,\mathrm{mA/cm^2}$ vs. $\sim 213\,\mathrm{mA/cm^2}$ for the 4 nm film, suggests that higher switching currents cause greater damage over time and thereby trigger the onset of early device failure. The switching current evolution for both films are plotted in Figure 3.4. There is a strong correlation between peak switching current and film thickness – namely, it is apparent that the initial peak switching current through the ferroelectric increases as a function of film thickness.

Furthermore, in comparing the data from Figure 3.4(a) and (b), it is apparent that the switching behavior *evolution* changes dramatically as a function of film thickness. It is observed that for the 4 nm film, the peak switching current at the coercive voltage at a mag-

nitude of around ~ 200 mA/cm^2 actually diminishes slowly over time, becoming smaller in magnitude and broader as well. This is a clear signature of domain switching suppression, and points to the existence of a broad range of switching fields in the areas of the capacitor which are fatigued [36]. However, for the 8 nm film, this behavior of switching evolution is not witnessed before the device ultimately breaks down. According to [36], it is highly probable that the formation of non-ferroelectric interfacial layers at the electrodes as the film is cycled can effectively screen the external electric field, thereby reducing the actual electric field applied to the ferroelectric layer (and causing incomplete polarization switching). It is also possible that dielectric degradation and consequent charge trapping can also cause the ferroelectric to witness a lower applied field [105]. In comparing the films of different thicknesses, it appears that the influence of this interfacial layer on endurance degradation increases substantially with reduced film thickness. Further studies should focus on examining TEM cross-sections of devices prior to and after cycling to confirm the formation of the interfacial "dead layer" and correlate its presence to the endurance characteristics.

Elevated Temperature Retention/Imprint Testing

Retention tests were conducted using the methods as described in [84] and subjecting the programmed samples to an elevated temperature of 100 °C.

Same State + (data "0") = 2_1 - 1_1
New Same State + (data "0") = 2_3 - 1_2
Opposite State + (data "1") = 1_4 - 2_4

Same State - (data "1") = 4_1 - 3_1
New Same State - (data "1") = 4_3 - 3_2
Opposite State - (data "0") = 3_4 - 4_4
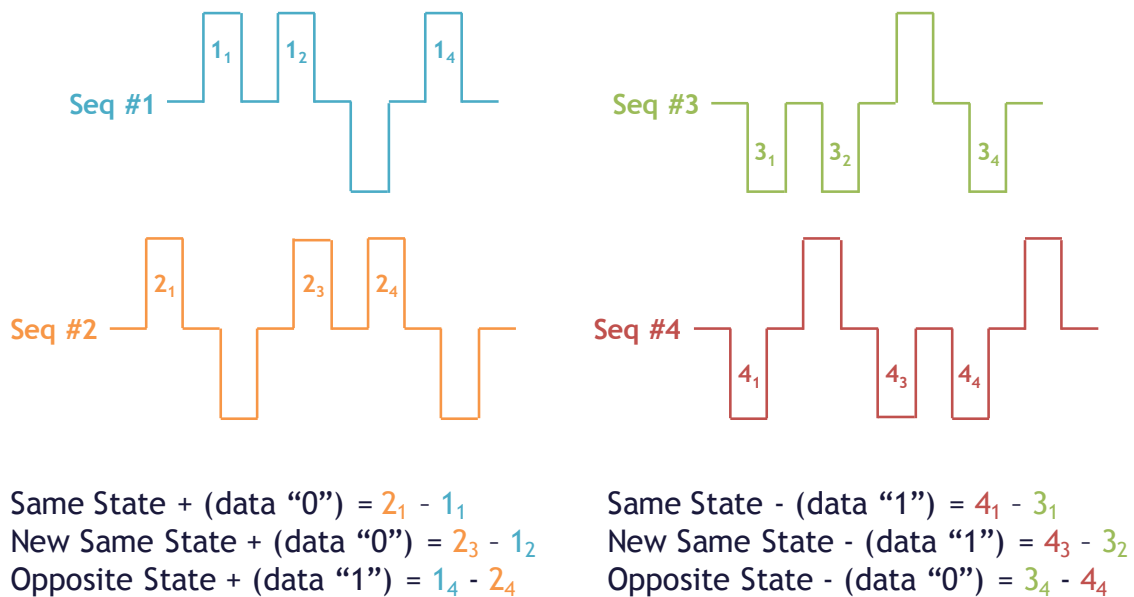
Figure 3.5: Voltage pulse sequences applied to 4 ferroelectric capacitors with identical characteristics in order to determine same state, new same state, and opposite state retention.

In order to measure retention for a ferroelectric capacitor, it is necessary to switch the capacitor first and measure the resulting switched polarization – and it is important to note that this total switched charge varies according to the previously written state, hence why there exist 6 total metrics to quantify endurance, as shown in Figure 3.5. The test consists of applying four distinct pulse sequences to four different ferroelectric capacitors with identical ferroelectric properties at various time intervals throughout the test, and with each measurement, integrating the current response during each of the corresponding pulse trains (and therefore calculating the total *associated charge* for each numbered pulse). The charge calculated will correspond to a purely dielectric response if the previous applied pulse was of the same polarity (for example, pulse 1_2), and will likewise correspond to a ferroelectric switching current if the previous applied pulse is of opposite polarity (for example, pulse 1_4). Once the switching currents are integrated accordingly, the total charge corresponding to the retention of the same state, new same state, and opposite state (and their binary counterparts) can therefore be calculated using the simple subtractions outlined in Figure 3.5 and then plotted as a function of time.
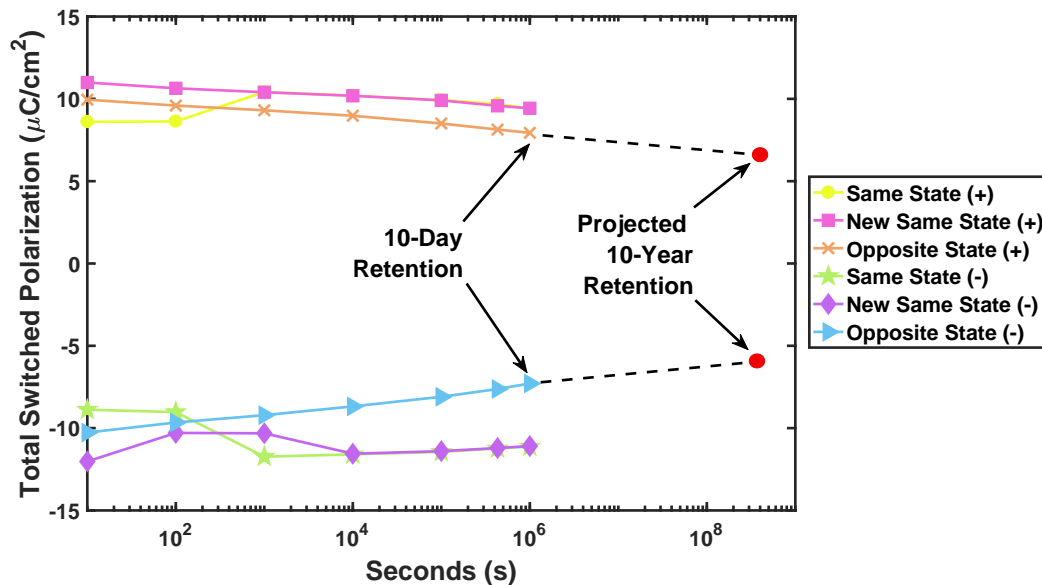
Figure 3.6: Voltage pulse sequences applied to 4 ferroelectric capacitors with identical characteristics in order to determine same state, new same state, and opposite state retention.

Retention testing on a representative 8 nm ferroelectric film is shown in Figure 3.6. After 10 days at 100 °C, the film retains most of its polarization, and is projected to retain more than 50% of the initial opposite state switched polarization (which appears to be the limiting state due to imprint) after 10 years. Imprint, therefore, is one additional memory

test that should be conducted on ferroelectric capacitors in order to fully understand the retention behavior of the state which first limits the retention.

a)

b)
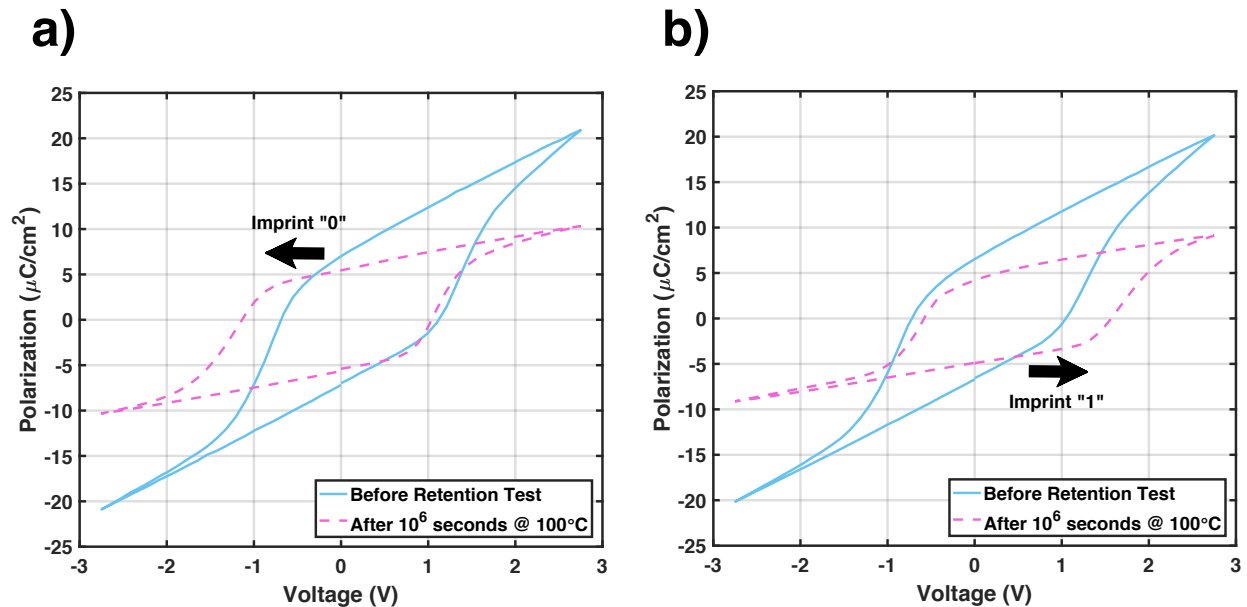
Figure 3.7: (a) A device showing imprint after being programmed with an assigned bit "0" for three weeks at an elevated temperature of 100 °C. Imprint is evident by shifts in $\pm E_C$. (b) A device showing imprint after being programmed with an assigned bit "1" for three weeks at 100 °C.

The imprint characteristics of each programmed capacitor for the "0" and "1" states were assessed by comparing the PE loop characteristics of the device after 3 weeks at 100 °C to the fresh device prior to the thermal testing. As seen in Figure 3.7(a) and (b), a clear shifting to the left is visible for the capacitor programmed in the "0" state, and a shifting to the right is visible for the capacitor programmed in the "1" state. The imprinting effect causes the required coercive voltages for a given capacitor to become more and more asymmetric over time due to the shifting of the entire hysteresis loop on the x-axis, which also negatively impacts the total switched charge associated with the state opposite to the one which was written. If the total switched charge associated with the opposite state decreases, then the margin for state determination for the state in question also decreases, thus limiting the overall opposite state retention.

Future studies on Si-doped HfO_2 memory capacitors should involve the improvement of endurance, as both the read and write operations will require switching the state of the capacitor. Furthermore, if Si-doped capacitors are intended to act as a nonvolatile DRAM replacement, the endurance requirements will necessarily need to be even more stringent in order to satisfy the high access requirements for main memory.

Chapter 4

Electrical Characterization of the Ferroelectric Transistor (FeFET)

4.1   Principles of FeFET Operation

In short, a ferroelectric field effect transistor (or FeFET) can be best described as a standard MOSFET, with the gate insulator replaced by a ferroelectric oxide instead of a standard dielectric material.
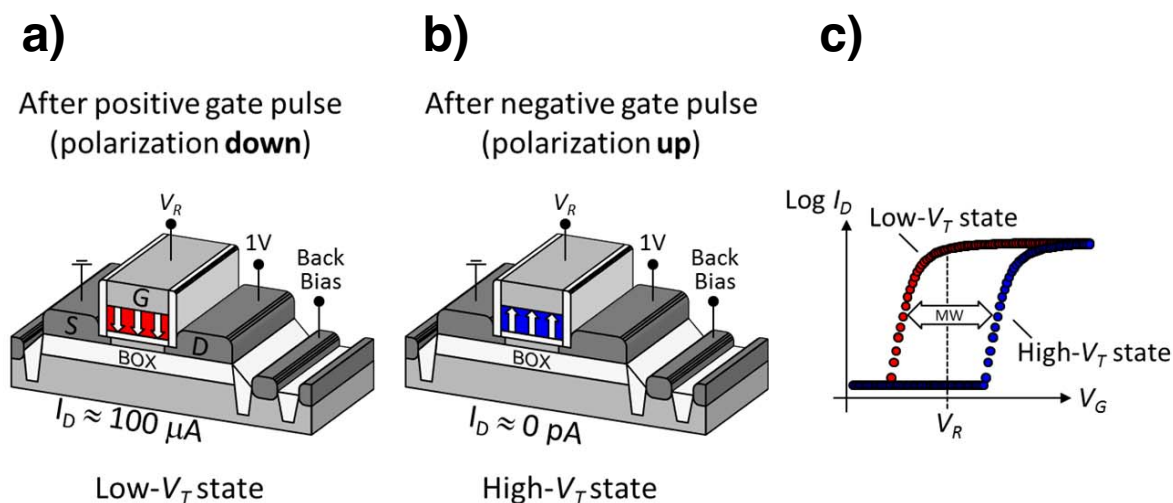
Figure 4.1: (a) After a positive gate bias, the FeFET's polarization points towards the channel, putting the device in the low V_T state. (b) After a negative gate bias, the FeFET's polarization points towards the gate, putting the device in the high V_T state. (c) The shifts in V_T correspond to hysteresis as seen in the device $I_D V_G$ characteristics. Figure adapted from [40].

Whereas a conventional MOSFET has a single threshold voltage which dictates the on-set of channel inversion – which is when the channel begins to fill with carriers that conduct current – the FeFET's threshold voltage is *polarization-dependent* and shifts according to the magnitude, duration, and polarity of the previously applied gate voltage. The effect of polarization switching and threshold voltage shifting is depicted in Figure 4.1. The Fe-FET's variable threshold voltage gives rise to *hysteresis* in the device's $I_D V_G$ characteristics, which corresponds to the memory window (MW) achievable. As shown in Figure 4.1(c), when measuring the drain current of the device at a fixed read voltage, V_R, the level of the readout drain current I_D is indicative of whether a positive or a negative gate bias was last seen by the device.
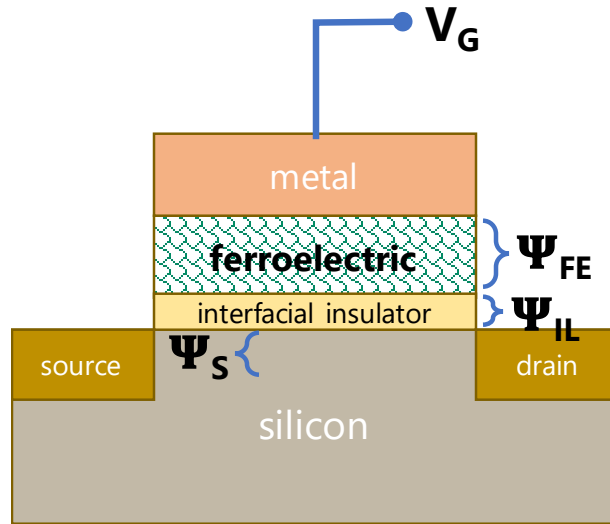
Figure 4.2: Gate stack of a generic FeFET incorporating a dielectric interfacial layer and a ferroelectric oxide. For a given applied gate voltage, the surface potential Ψ_S, the interfacial oxide voltage Ψ_{IL}, and the ferroelectric oxide voltage Ψ_{FE} must be taken into account to properly model the hysteretic threshold voltage behavior.

The FeFET's variable threshold voltage can also be understood following a simplistic model of voltage distribution across the entire gate stack with an applied V_G and its effects on ferroelectric polarization [81]. Assuming no additional nonideal sources of charge contribution (i.e, from trapping at material interfaces or within the bulk of any oxides depicted in Figure 4.2), the applied gate voltage can be expressed as:

$$V_G = \Psi_S + \Psi_{IL} + \Psi_{FE} + \Phi_{MS} \tag{4.1}$$

where Ψ_S represents the surface potential of the semiconductor channel, Ψ_{IL} represents the voltage across the dielectric interfacial layer, Ψ_{FE} represents the voltage across the

ferroelectric layer, and Φ_{MS} is the workfunction difference between the gate metal and the semiconductor. The voltage drop across the interfacial layer is a direct function of the semiconductor's surface potential, as described by standard MOSFET operational theory. The voltage drop across the ferroelectric can be expressed as:

$$\Psi_{FE} = \frac{t_{FE}}{\kappa_{FE}} \times [Q_S(\Psi_s) - P(E_{FE})] \tag{4.2}$$

where t_{FE} represents the thickness of the ferroelectric, κ_{FE} represents its dielectric constant, and Q_s the induced charge in the semiconductor as a function of semiconductor surface potential. Equation 4.2 indicates that the voltage across the ferroelectric is a function of both surface potential *and* the history-dependent behavior of the ferroelectric, which manifests itself as the polarization of the ferroelectric, $P(E_{FE})$. There are many methods to capture the polarization P as a function of the previously applied voltage/electric field, E_{FE}. The most commonly used models include the Priesach model as well as a more physical model based on the single domain Landau model, which was introduced in Chapter 1 of this dissertation.
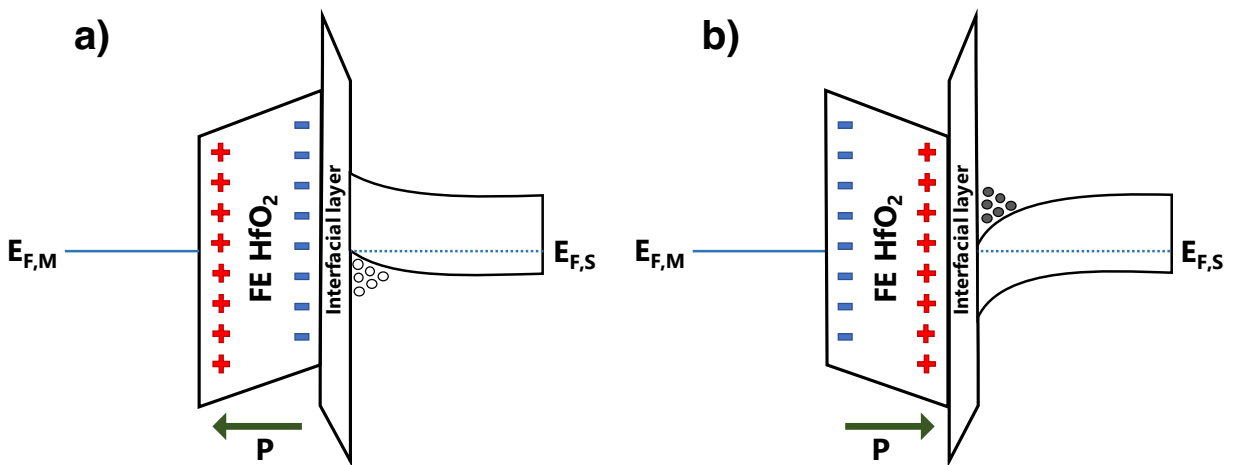
Figure 4.3: Equilibrium band diagram explanations of FeFET variable V_T (not drawn to scale). (a) After a negative gate bias, the FeFET's polarization points towards the gate, putting the device in the high V_T state. (b) After a positive gate bias, the FeFET's polarization points towards the channel, putting the device in the low V_T state.

Figure 4.3 provides additional insight into the effect of ferroelectric polarization on the resulting V_T of the FeFET. The example shown here is for an n-channel FeFET fabricated on p-type silicon. Considering a generic MFIS FeFET structure, after the application of a negative gate bias (as shown in Figure 4.3(a)), the ferroelectric polarization is pointing towards the gate. The ferroelectric dipoles are represented as "+" and "-" signs in the

ferroelectric layer. Without loss of generality, the alignment of ferroelectric dipoles in such a way drives the semiconductor into accumulation (of holes) at equilibrium. Similarly, in Figure 4.3(b), the reversal of polarization is effectively represented as the flipping of dipoles, which therefore drives the semiconductor into inversion (with electrons filling the channel) at equilibrium. Therefore, the former scenario corresponds to the *high V_T state*, as additional voltage will need to be applied to the gate to invert the channel; the latter scenario corresponds to the *low V_T state*, as the channel in this example has already been inverted.

Following the discussion above, it is clear that at the flatband condition of the FeFET (where all charges associated with every capacitor in the gate stack, from the ferroelectric capacitance to interfacial layer capacitance to semiconductor depletion capacitance, must be zero), the ferroelectric must be at one of its two possible coercive voltages, $\pm V_C$. This can also be understood by considering the PE hysteresis loop of the ferroelectric – the only way for the surface charge of the ferroelectric to be equal to zero is when the ferroelectric is switching. This therefore allows one to approximate the memory window of the FeFET as follows [81]:

$$MW = V_C^+ - V_C^- = 2 \cdot |V_C| = 2 \cdot E_C \cdot t_{FE} \approx \Delta V_{FB} \approx \Delta V_T \tag{4.3}$$

Of course, Equation 4.3 does not take into consideration the effect of nonidealities such as fixed, trapped, or interface charges. This derivation as well as the band diagram examples in Figure 4.3 above also assume that the polarization is strong enough to fully drive the MFIS FeFET from inversion to accumulation and vice versa during switching. The picture is complicated if the polarization charge is not strong enough to achieve this, which will to first order cause the actual memory window to be smaller than predicted in Equation 4.3 [81]. Equation 4.3 indicates that a larger coercive field and ferroelectric layer thickness will result in a device with a larger memory window. In actuality, when working with a given material system, it is very likely that modulation of the coercive field will be significantly more challenging than changing the film thickness.

4.2 FeFET Device Structure

Baseline FeFET Device

The baseline FeFET characterized is shown in Figure 4.4, which details the device design, the gate oxide stack, and corresponding TEM. To fabricate the device, after thinning down the device active layer on 150 mm SOI (silicon on insulator) substrates through a combination of wet oxidation and HF etching to 30 nm in thickness, the active regions are defined and the ferroelectric gate oxide is deposited via atomic layer deposition. Once the tungsten gate is defined, the source/drain regions are implanted with As$^+$ ions, thereby completing the self-aligned gate-first process.

The activation of the implanted source/drain regions and the crystallization of the ferroelectric HfO_2 layer are carried out simultaneously with one rapid thermal annealing (RTA) step at 500 °C for 30 seconds. The devices are then isolated with a thick 350 nm interlayer dielectric of SiO_2, and an additional metal deposition/etch step is utilized to make contact to the source/drain/gate regions. Additional details regarding the fabrication process flows used to realize the devices characterized here are described in Appendix A.

Based on Figure 4.4(b), the HfO_2 layer is roughly 1.3 nm thick, and the Zr-doped HfO_2 layer is 4.2 nm thick. The 15 cycles of pure hafnia were deposited in an effort to reduce charge injection from the substrate through the oxide stack without compromising the overall EOT of the stack, as it follows from conventional ferroelectric theory that reduced charge injection plus a higher-κ IL (and therefore increased total capacitance) would reduce the depolarization effect and thereby improve the retention properties of the device. The ratio of cycles of Hf to Zr is chosen as 4:1 as previously investigated in [61].
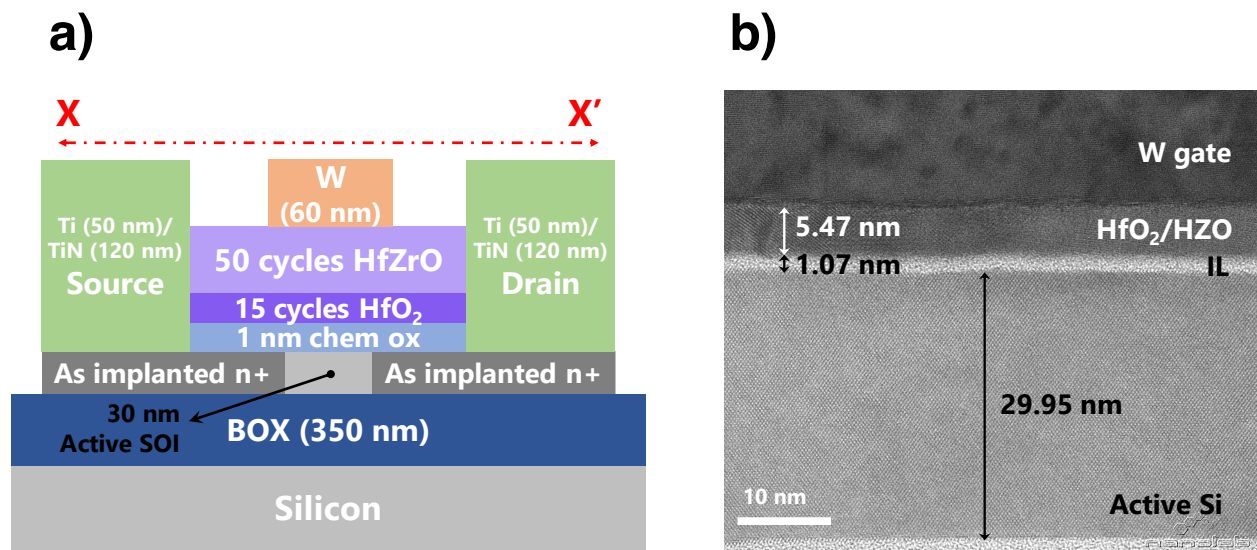
Figure 4.4: Baseline FeFET device structure characterized. (a) 2D cross-section across channel of the FeFET, showing the gate stack composition and active device layer. (b) TEM (tunneling electron microscopy) image of the gate stack of a fabricated device with W/L = 1/1 μm.

4.3 FeFET Programmability

Baseline FeFET Device

To understand projected FeFET performance, it is imperative to first characterize the programmability of representative single FeFET devices. Programmability of a device refers to assessing the memory window size achievable for a given applied gate voltage of a defined magnitude and duration. All devices discussed in this section are characterized using a Keysight B1500A Semiconductor Parameter Analyzer, in conjunction with a HV-SPGU (high voltage semiconductor pulse generator unit), which allows for flexible specifications of test pulses with user-defined widths and magnitudes.
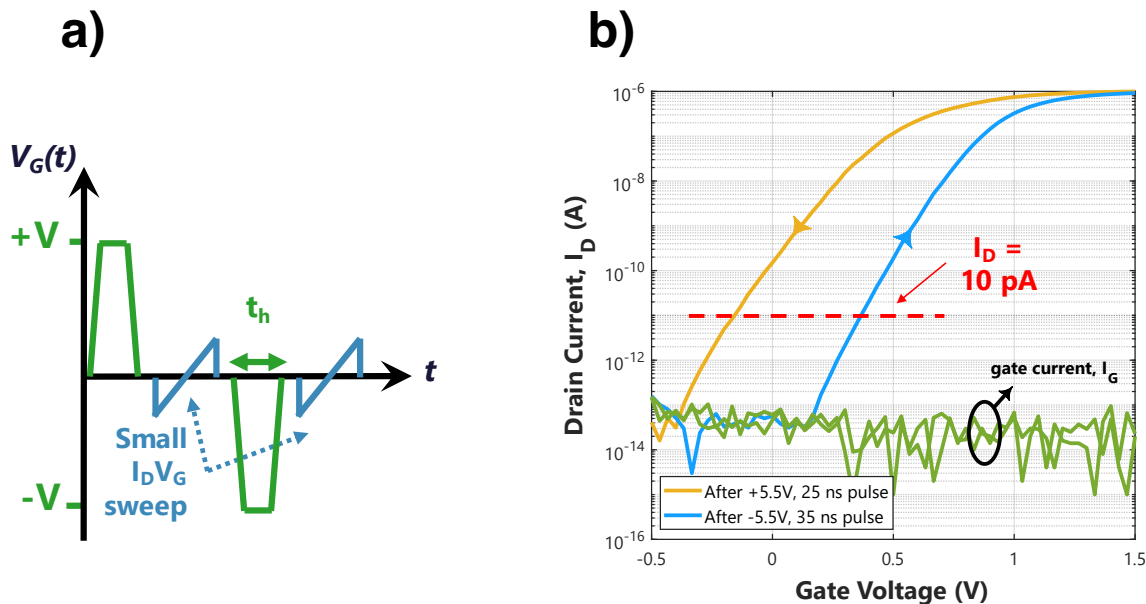
Figure 4.5: (a) The pulsing and $I_D V_G$ sweeping sequence applied to the gate of the FeFET in order to assess the total achievable memory window of the device. Both $\pm V$ and t_h can vary. (b) A typical $I_D V_G$ achieved using programming conditions of +5.5V at 25 ns, and -5.5V at 35 ns. The device characterized has W/L = 500/250 nm.

Programmability is tested following the test sequence shown in Figure 4.5(a). The device is programmed in the low V_T state by first applying a pulse of duration t_h and magnitude $+V$ at the gate, and then taking a DC $I_D V_G$ sweep over a much smaller voltage range, which can then be used to determine the V_T of the device after programming. The same procedure is then repeated to assess the second state as well, using a voltage pulse of magnitude $-V$. In the example device shown in Figure 4.5(b), the $I_D V_G$ sweep range is -0.5 to +1.5V. This small range sweep is chosen to have a negligible disturbance on state –

ideally, well below the coercive voltages of the FeFET. Figure 4.5(b) details the $I_D V_G$ hysteresis traced out by this programmability test, and the memory window here is defined as the difference in $\pm V_T$ at a constant current level of 10 pA. This hysteresis changes in size with program pulse magnitude and duration and, as it turns out, is asymmetric for the device in question.

Figure 4.6 illustrates the change in $\pm V_T$, defined at a constant current criteria of $I_D = 10$ pA, at varying programming voltages and durations, for a single FeFET memory device. The inset of Figure 4.6(a) incorporates the data from Figure 4.5(b) as an inset to illustrate typical $I_D V_G$'s after applying programming pulses of +5.5V at 25 ns, and -5.5V at 35 ns. In general, the trends shown in Figure 4.6(a) and (b) demonstrate that the magnitude of $\pm V_T$ increases with increasing voltage magnitude and/or duration. The memory window saturates once full switching of the ferroelectric layer is achieved, and charge trapping within the oxide inhibits further shifts in $\pm V_T$.
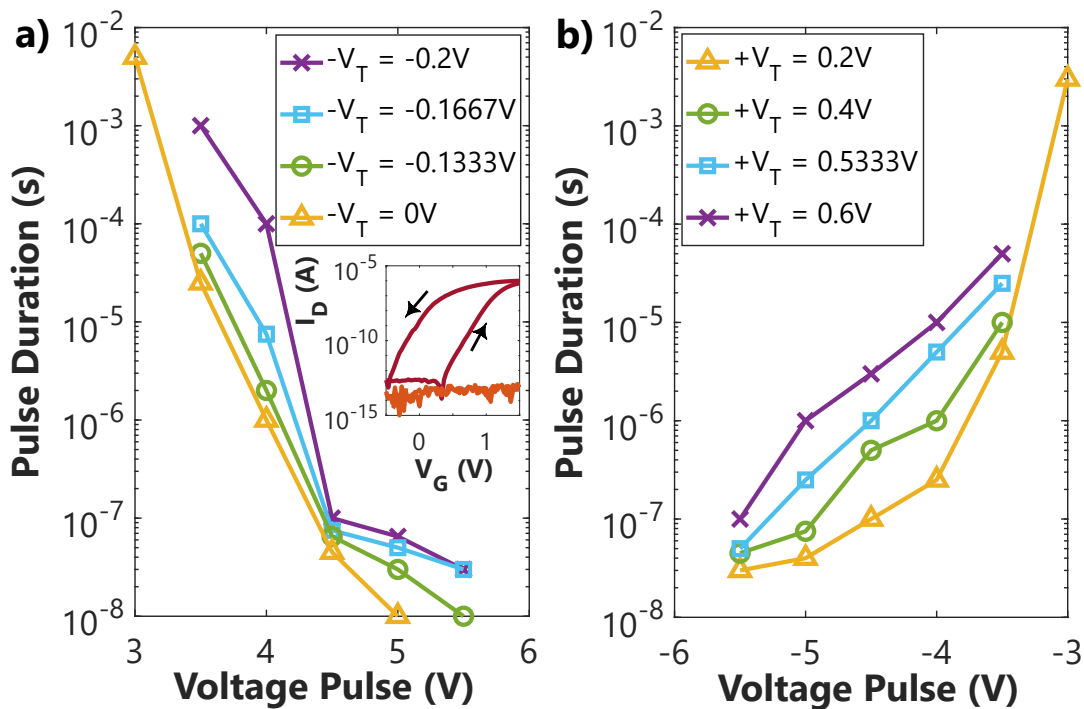
Figure 4.6: Programmability of the FeFET device. (a) Curves of constant $-V_T$ obtained at varying pulse durations/amplitudes to illustrate the voltage/time tradeoff for a single FeFET with W/L = 500/250 nm. Inset includes $I_D V_G$ characteristics. (b) Curves of constant $+V_T$ obtained at varying pulse durations/amplitudes for the single FeFET.

One observation of note in understanding the programmability results detailed in Figure 4.5 is the fact that the FeFETs are fabricated on *silicon on insulator* substrates thinned

down to just 30 nm in thickness. In comparing Figure 4.5(a) to Figure 4.5(b), as well as
the asymmetric choice of pulse durations (25 ns for the positive pulse versus 35 ns for the
negative pulse), it is clear that these devices suffer from a slight asymmetry in program-
ming conditions, with changes in the high V_T state appearing much more gradually in
comparison to changes in the low V_T state. This effect is a result of the fact that the thin 30
nm SOI body is quite hole deficient within its volume, and therefore the semiconductor
potential is closely coupled to the gate voltage, particularly when negative gate biases are
applied (which would put the device in accumulation) [7]. Therefore, a smaller percent-
age of the applied voltage is seen by the ferroelectric as well, in this scenario. With fewer
holes present in the system to aid in switching and a smaller effective voltage drop across
the ferroelectric, it is therefore required to apply additional voltage (or increase the pulse
duration) to achieve a greater ΔV_T.
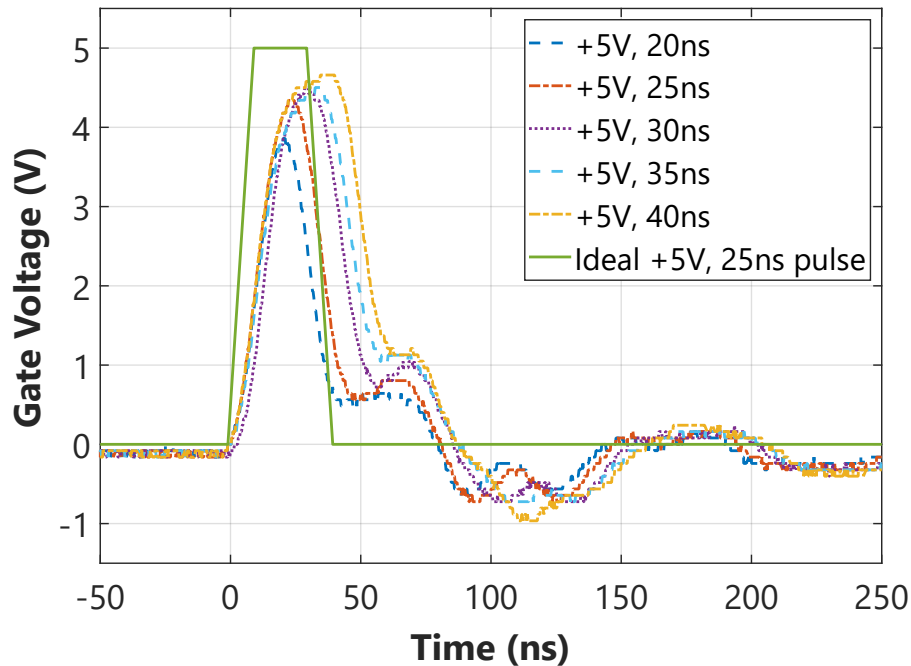
Figure 4.7: The applied ideal pulse of +5V, 25 ns with 10 ns rise/fall times plotted against
the actual voltage waveforms measured, using an oscilloscope, at the gate of a transistor
with W/L = 500/250 nm.

One additional option for improved programmability results would be to utilize the
GIDL (gate-induced drain leakage) current to supply additional holes in the system to
aid in switching, as shown in [7]. This could potentially allow for more symmetric pro-
gramming conditions to be used at the gate terminal to write and erase the FeFET. How-

ever, this strategy is achieved at the expense of also involving the drain terminal in the programming of the FeFET device.

Though Figure 4.6 illustrates the measurable program and erase speeds of the FeFET, it should be noted that at the lower limit of pulse speeds tested, significant oscillation of the gate voltage pulse was observed, as shown in Figure 4.7. Figure 4.7 plots the defined gate voltage pulse sent to the device by the SPGU, and also plots the resulting waveforms seen at the gate terminal (captured with an oscilloscope, model Agilent DSO1024A). Furthermore, the peak voltage values seen by the device were degraded as the pulse widths were reduced below 50 ns. These effects are due to the non-negligible contribution of extrinsic device impedances at shorter programming pulse durations. For this reason, long channel devices ($L_g \geq 250$ nm) were used to enable robust program/erase condition testing, as going down to very small gate lengths ($L_g \sim 50$ nm) requires the use of a much thinner SOI body, which would thereby significantly increase the associated device series resistance and requiring even longer programming pulses.

Optimized FeFET Device

Programmability results for an optimized FeFET device are detailed in this subsection. The process flow and exact process optimizations will be discussed in Chapter 6 and Appendix A of this dissertation.

The primary differences to highlight between the baseline FeFET and the optimized FeFET are 1) the difference in gate oxide thickness – the baseline FeFET uses an additional 1.5 nm of non-ferroelectric HfO_2, whereas the optimized FeFET uses only 4.5 nm of Zr-doped HfO_2; and 2) the interfacial layer itself – the baseline FeFET uses roughly 8Å of SiO_2 grown in SC1, whereas the optimized FeFET uses a nitrided, thermally grown interfacial layer.

Figure 4.8(a) shows results of a doubly swept $I_D V_G$ curve of the optimized FeFET. Nearly a 1V memory window can be achieved with just ±2.5V sweep. Compared to the published literature, this is quite a low voltage requirement. For example, the baseline devices discussed above and as reported in [125] do not demonstrate any appreciable MW at ±2.5V.

Nonetheless, the required voltage to operate the device is a strong function of switching time. Therefore, as shown in Figure 4.8(b) and (c), the measured device current at a read voltage of $V_G = \pm0.25$V is shown as a function of pulse width. Voltages in the range of $\pm2.5 - 3$V and pulse widths in the range of 100 ns - 10 μs are investigated. We define the high current/low V_T state as the ERS state, and the low current/high V_T state as the PGM state. It is shown that below a pulse duration of 1 μs, ±2.5V is not good enough to provide the current level seen in the DC hysteresis in Figure 4.8(a). As the voltage amplitude increases, the current increases, signifying switching of a larger amount of polarization, as expected. At $V_G = 3$V, the current approaches the level seen in DC hysteresis, even for a pulse width of ~ 100 ns. Similarly, for the PGM state, $V_G = -3$V brings the current level down to almost the level seen in the DC hysteresis at a pulse width of ~ 250 ns. Again, it
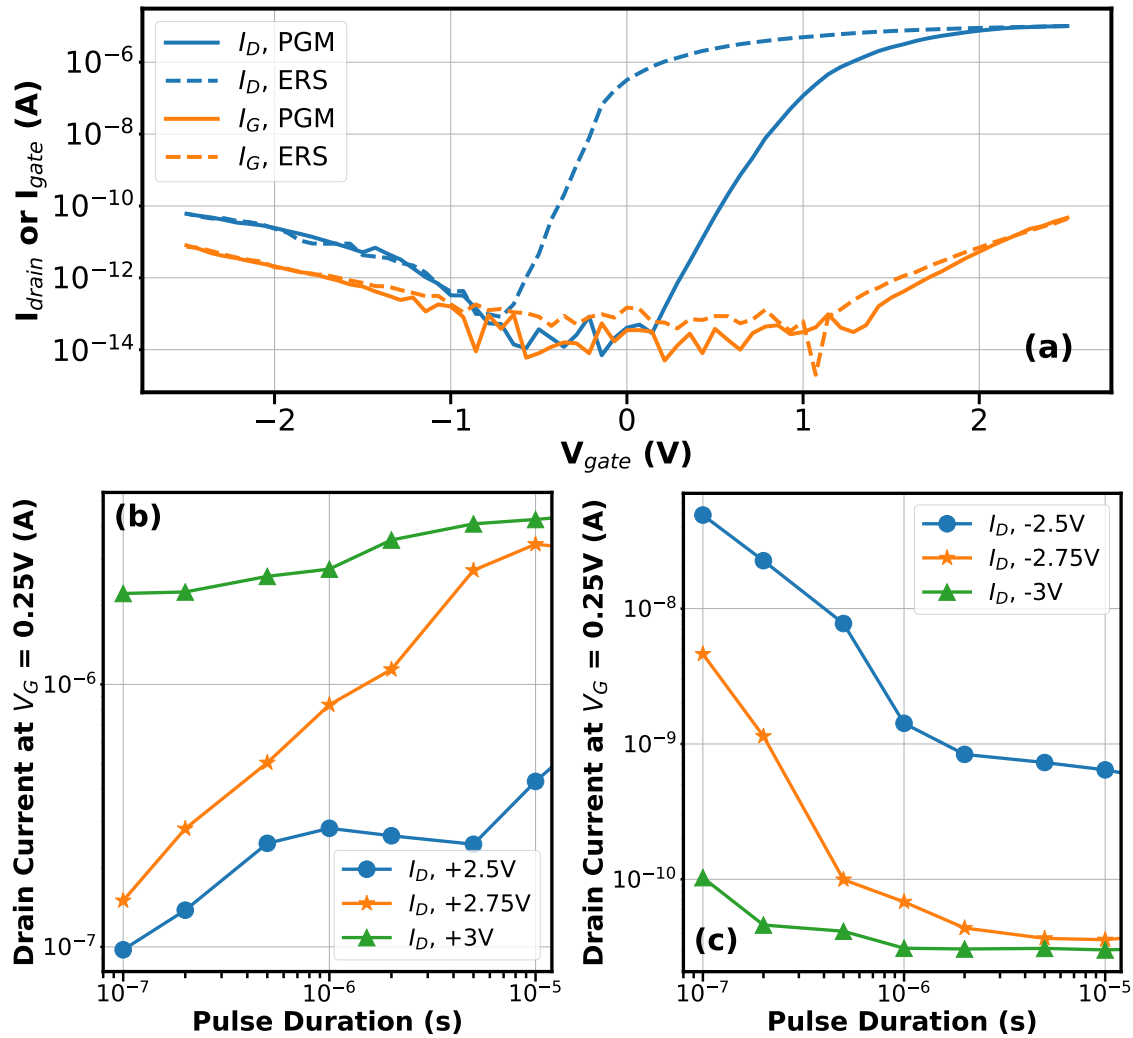
Figure 4.8: (a) $I_D V_G$ of a typical FeFET with 4.5 nm HZO on a 1.5 nm nitrided IL. The device is doubly-swept from ± 2.5V at a drain bias of $V_D = 50$ mV. (b) Typical `ERS` characteristics for the FeFET. (c) Typical `PGM` characteristics for the FeFET. Voltage magnitudes range from $\pm 2.5 - 3$V and pulse durations from 100 ns to 10 μs.

is important to note the asymmetry of the `PGM` and `ERS` characteristics, which are predicted in fully depleted SOI FeFETs as the accumulation of a thin SOI body is hindered by lack of holes [7].

Lastly, the retention behavior of the optimized FeFET is shown in Figure 4.9. The retention characteristics appear to be virtually unaffected for a testing duration of 10^4 seconds for both states at both room temperature, 25 °C, and elevated temperature at 85 °C. The superior retention performance of the FeFET can be understood by considering the property

of the ferroelectric material most closely linked to retention, which is the coercive field E_C.

In a typical FeFET, the phenomena which causes the loss of state information over time is an effect called *depolarization*, which can be thought of as an additional electric field component at equilibrium which triggers the loss of the remaining polarization in the device [81]. Therefore, the coercive field comes into the picture in considering how "close" the ferroelectric material comes to its E_C at zero applied gate bias. If the intrinsic coercive field of the material is much higher, then the retention characteristics are projected to be better as well for the FeFET. This is perhaps one of the biggest benefits *and* drawbacks of the HfO_2 FeFET – although its coercive field is typically an order of magnitude higher than that of conventional perovskite ferroelectrics [81], leading to generally superior retention characteristics, the high coercive field in many cases also forces the device to be operated at higher applied electric fields (and therefore voltages) as well. In the next chapter, which focuses exclusively on understanding FeFET reliability issues, it will become clear how the intrinsically high E_C of HfO_2 adversely impacts device performance in other ways.
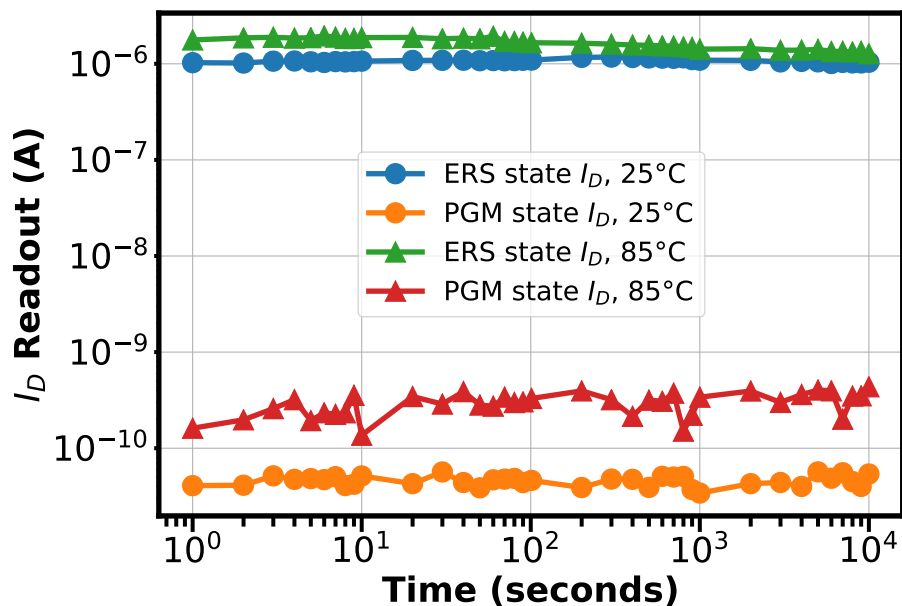
Figure 4.9: Retention testing at room temperature ($25\,°C$) and at elevated ($85\,°C$) for 10^4 seconds. Gate read voltage is chosen to the same at both testing conditions after correcting for the leftward V_T shift due to an effective substrate doping change at elevated temperature.

Chapter 5

FeFET Reliability: Building a Deeper Understanding

5.1 Introduction and Background on FeFET Reliability

In comparison to many other emerging nonvolatile memory technologies, the FeFET based on doped hafnium oxides boasts many advantages, not limited to but including its compact one-transistor cell structure, its low programming power, flexibility with CMOS processing, and so on [25]. However, at the present, the HfO_2 FeFET suffers from several reliability issues which must necessarily be resolved. Some of these issues stem from the motion and evolution of oxygen vacancies and interstitials in the imperfect HfO_2 layer, from the spread in grain sizes, as well as from its intrinsically high coercive field (which is beneficial from a retention perspective, but disadvantageous from a low-voltage operating perspective, as discussed in Chapter 4) [94, 103]. The different reliability concerns for HfO_2-based FeFETs will be addressed and explored in detail in this chapter.

Device to Device Variability

Control of device variability is an issue battled by nearly every exploratory memory technology under consideration, although the underlying physical mechanisms leading to these variability issues are distinct. In the HfO_2 based FeFET, several different factors relating to process integration, dielectric versus ferroelectric grain distribution, spread in grain sizes, the stochasticity from ferroelectric switching, defect distribution within the film, etc. all pose potential sources of variation [71, 94, 95, 104, 103]. Though there are corrective methods which can be taken at the system level in terms of optimizing write pulses (i.e, by increasing pulse widths and/or amplitudes) and other similar potential schemes, it is first and foremost more advantageous to control issues related to processing schemes or film deposition engineering in order to alleviate fundamental issues in the underlying technology.
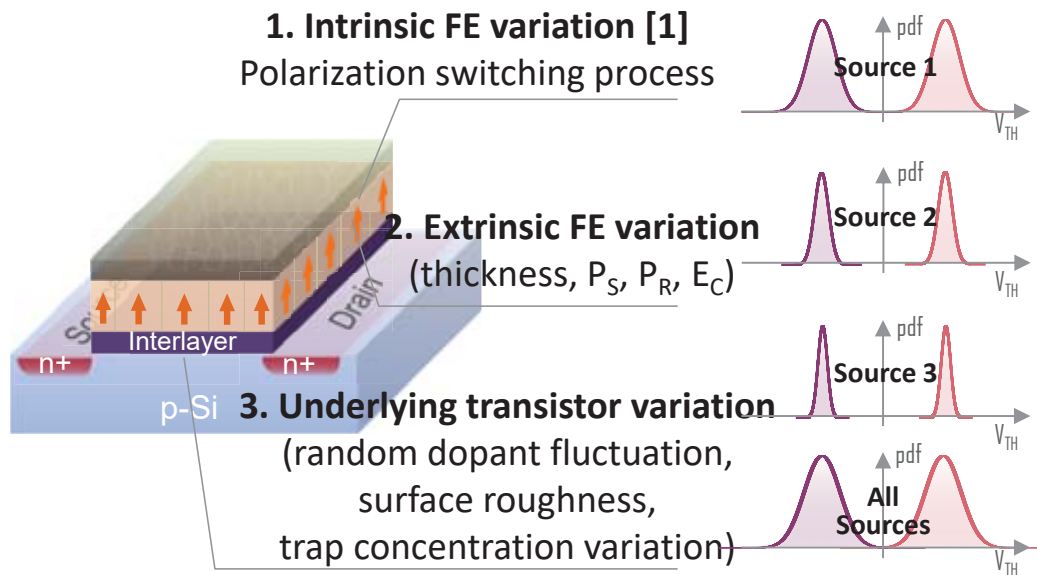
Figure 5.1: Sources of possible FeFET variability, categorized into intrinsic & extrinsic sources of variation, and underlying transistor variation (unrelated to the ferroelectric material itself). Figure adapted from [95].

Within the ferroelectric material itself, film concerns such as the presence of charged defects and various crystalline phases of HfO_2 (some ferroelectric, others not) directly impact the nucleation of ferroelectric domains, also affecting the electrostatics of the HfO_2 layer and affecting the internal field of the material [103]. These effects manifest in the wakeup effects clearly present in many HfO_2 material systems, which affect the starting characteristics of the ferroelectric device (therefore requiring "wasted" cycling efforts in order to unpin or evenly redistribute defects, unpin domains, or modulate the crystallinity of the film – an example is discussed in Chapter 3 in the testing of Si-doped HfO_2 capacitors).

An example demonstrating the possible variation that can be present in pristine Fe-FET devices is shown in Figure 5.2. The devices characterized, as discussed in [23], are nominally identical in dimension and geometry (with W/L = 1 μm/0.35 μm). The low V_T states, illustrated by the blue $I_D V_G$ sweeps, are obtained with +5V, 1 ms pulses, and the high V_T states are obtained with -5V, 1 ms pulses; the conditions are chosen to fully polarize the FeFET in one state or the other. Despite this carefully chosen condition, it is clear that intrinsic variation is present in both V_T states. In fact, the spread in both $\pm V_T$ is on the order of hundreds of millivolts.

From a modeling perspective, this variation in memory window hysteresis can be considered as the result of a non-uniform distribution of coercive fields within the gate oxide
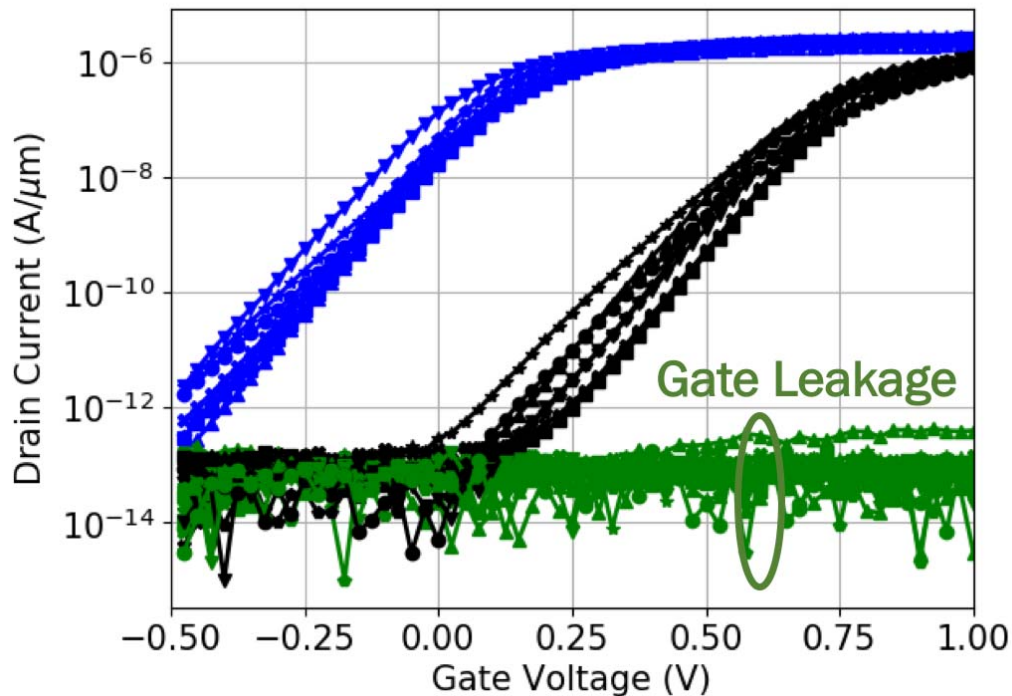
Figure 5.2: An example of an FeFET demonstrating a spread in $I_D V_G$ hysteresis across 7 different devices measured. For both states, the `PGM` and `ERS` states are chosen to fully polarize the device. Figure adapted from [23].

itself, if one chooses to model the gate oxide as a collection of ferroelectric (and additionally, non-ferroelectric) grains with varying properties such as remanent polarization and coercive field. Ultimately, the physical root cause of such a distribution is tied to the varying aforementioned material and processing factors. New engineering methods of controlling the HfO_2 film properties at sub-10 nm scales will therefore be imperative in achieving better device-to-device variation at a fundamental level.

Bulk Charge Trapping and Endurance

In addition to affecting the pristine device characteristics as described in the previous subsection, the imperfections within the HfO_2 film also heavily influence the endurance cycling characteristics and electrical performance throughout the device's lifetime. The first issue to consider is the phenomena of *charge trapping*, an effect which directly counteracts ferroelectric switching. Charge trapping is a particularly troublesome concern in HfO_2-

based FeFETs, due to the fact that HfO_2 material systems tend to possess high intrinsic defect densities [146]. Figure 5.3 details the exact mechanism for how electron charge trapping in the device gate stack counteracts the ΔV_T associated with ferroelectric switching. According to this nonideal effect, one can expect that the total memory window size achievable in an FeFET will be greatly reduced if charge trapping is a substantial concern. The direction of $I_D V_G$ hysteresis induced by trapping and detrapping of electron charges in the gate stack is the exact opposite of the direction of hysteresis of ferroelectric polarization.
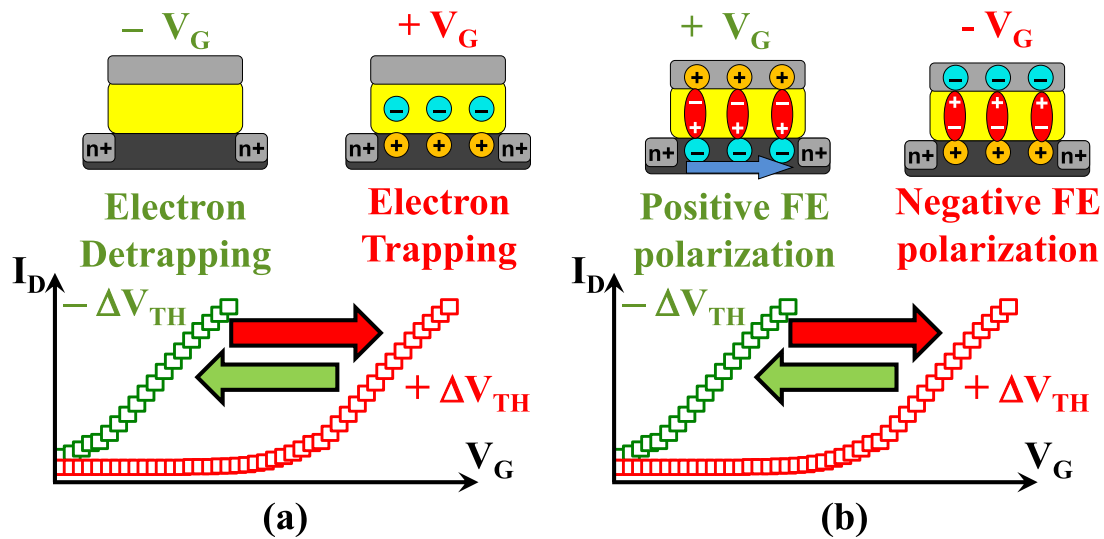
Figure 5.3: Illustration of the interplay between electron charge trapping and ferroelectric switching. (a) In a transistor, electrons can become trapped in the gate oxide upon the application of a positive gate bias, and the net effect is a right shift or $+V_T$ shift (shown in red). (b) In an FeFET, the polarization points towards the channel upon the application of a positive gate bias, and the net effect is a left shift or $-V_T$ shift (shown in green). Figure adapted from [146].

The extent of charge trapping within the gate stack has been found to be highly contingent on the voltage pulse magnitude and duration applied at the gate – in much the same way that ferroelectric polarization is influenced by the gate programming conditions. Therefore, for devices in which bulk charge trapping is a substantial issue, it is important to determine the exact PGM/ERS conditions for which the onset of charge trapping severely compensates the ferroelectric polarization. In [146], the authors first applied a negative pulse at the gate of -6V, 100 ns, in order to establish a saturated negative polarization state in the device. Then, pulses of increasing positive magnitude (+2V to +5V) and duration (100 ns to 100 μs) were applied, reestablishing the negative polarization state between

tests with the -6V, 100 ns pulse.

It was then established that charge trapping in the ferroelectric *correlates* with ferroelectric switching, due to the fact that the internal fields within the device stack generated by the polarization charge the interfacial layer/ferroelectric HfO_2 interface greatly increases the voltage drop across the thin interfacial layer. This effect essentially decreases the potential barrier seen by carriers tunneling across it, giving rise to increased charge injection and trapping in the device stack. In order to mitigate this effect, the element of time in measurement was shown to have a substantial influence, with increasing delay times from 100 ns up to 100 ms greatly improving the resulting endurance metric. The increased time allows for the detrapping of charges from within the gate stack prior to the application of a new gate pulse, which improves the positive feedback cycle of trapped charges intrigating further oxide trap generation.

Additionally, in [53], it was found that the charge trapping effect in measurement could also be suppressed by taking pulsed $I_D V_G$ measurements, rather than DC $I_D V_G$ measurements. Such a measurement reduces the amount of time the entire gate stack witnesses a given applied bias and therefore reduces the time during which charge injection can occur. The reduction of charge injection and trapping was clearly seen in the complete reversal of the hysteretic direction (from entirely charge trapping hysteresis in DC characteristics) to a nearly 1V ferroelectric memory window. This example is an extreme illustration of the effects of parasitic charge trapping in a device, showing the extent to which the memory capability of the FeFET can be entirely destroyed by use of a poor quality HfO_2 gate oxide with higher defect density than polarization charge.

Interfacial Oxide Wearout and Endurance

In addition to charge compensation due to the charge trapping in the gate stack, one final and major concern for FeFET reliability relates to the wearout of the interfacial oxide layer in the device stack. This interfacial oxide is virtually unavoidable in fabricating FeFETs on a silicon channel, and many works utilize either a thin SiO_2 interfacial layer or nitrided SiO_2 layer [2, 85, 92, 127, 148, 150].

As briefly discussed in the previous subsection, the polarization charge which gives rise to the ferroelectric properties of the HfO_2 layer actually also *increases* the magnitude of injected charge across the interfacial layer across the stack, when compared to an normal transistor designed with a non-ferroelectric HfO_2 layer. Compounded with the fact that HfO_2 intrinsically has a large coercive field, this translates to high electric field stress *and* substantial charge injection seen by the thin, low-κ interfacial SiO_2/SiON over time. Perhaps then, unsurprisingly, this leads to a discussion of the ultimate bottlenecking reliability issue seen in the HfO_2 FeFET to date: gate stack failure not due to the breakdown/wearout of the ferroelectric layer itself, but of the *interfacial layer*.

Empirically, many researchers have indeed reported troublesome endurance results on HfO_2-based FeFETs. For example, in reports such as [2, 49, 137, 148, 150], the endurance of the FeFET is very clearly limited to the range of $10^4 - 10^6$ cycles at best. To put this met-

ric in context, it is noted that the ultimate endurance limit of flash memory technology – a well-established storage solution for persistent memory, as described in the introductory Chapter 1 – is already held to an endurance requirement of $10^4 - 10^6$. As one of the projected applications of FeFETs is in the storage class memory space, which requires substantially greater accesses (in terms of both reads and writes), it is imperative to improve this number substantially to make HfO_2 amenable for such an application.

A table summarizing recent reported FeFET endurance results is shown below. For the sake of ease of comparison, the endurance of each evaluated device is taken at 25% of the starting memory window size.

Device	FE Layer	PGM/ERS	MW	Write Endurance (to 25% MW_{init})	Retention
[124] A. J. Tan et al.	4.5 nm HZO	5.5V, 25 ns/ -5.5V, 35 ns	0.75V	$\sim 10^5$	>10 years at 85 °C
[87] J. Muller et al.	10 nm Si-HfO_2	5V, 100 ns/ -5V, 100 ns	0.75V	$\sim 10^4 - 10^5$	>10 years at RT
[40] S. Dunkel et al.	t_{FE} N/A Si-HfO_2	3.8V, 10 μs/ -3.8V, 10 μs	0.41V	$\sim 10^4$	<7 days at 250 °C
[2] T. Ali et al.	10 nm Si-HfO_2	6V, 300 ns/ -6V, 300 ns	1.5V	$\sim 10^5$	>10 years at 85 °C
[28] K. T. Chen et al.	5 nm HZO	4.8V, 500 μs/ -4.8V, 500 μs	0.6V	$\sim 10^3$	>10 years at RT
[148] Yurchuk et al.	9 nm Si-HfO_2	4V, 100 ns/ -6V, 100 ns	0.7V	$\sim 10^4$	N/A
[150] Zeng et al.	10 nm HZO	4.85V, 100 ns/ -4.85V, 100 ns	1V	$\sim 10^4$	N/A

Table 5.1: Comparison of endurance metrics across several recent works in the literature on HfO_2-based FeFETs. Several other parameters relevant for assessment of memory performance, alongside endurance, are given. HZO is an abbreviation for Zr-doped HfO_2. t_{FE} indicates ferroelectric film thickness, and N/A indicates that the authors did not disclose the characteristic in question.

The key result, in considering Table 5.1, is to notice that when normalized to 25% of initial starting memory window size (to account for variation in different processing techniques, device geometries, film compositions, thicknesses, etc.), the ultimate endurance of the HfO_2 FeFET on silicon channel is extremely limited. Furthermore, the publication dates on the works in Table 5.1 span a range of 6 years, from 2014 to 2020. Clearly, the issue of limited FeFET endurance has plagued the research community since the first initial attempts to integrate ferroelectric HfO_2 into memory transistors.

The results discussed in [124] in Table 5.1 above are expanded upon in detail below. Details governing the programmability of the device as well as processing steps are found in Chapters 4 and 7. Figure 5.4 tracks both the high and low V_T states (the PGM and ERS states, respectively) as a function of fatigue cycles. The cycling condition is chosen to be 5.5V, 25 ns, and -5.5V, 35 ns. The reason for choosing slightly different pulse durations is explained in Chapter 4, and has to do with the FDSOI device structure (and lack of ability to accumulate holes and therefore bend the silicon bands when negative gate voltages are applied).
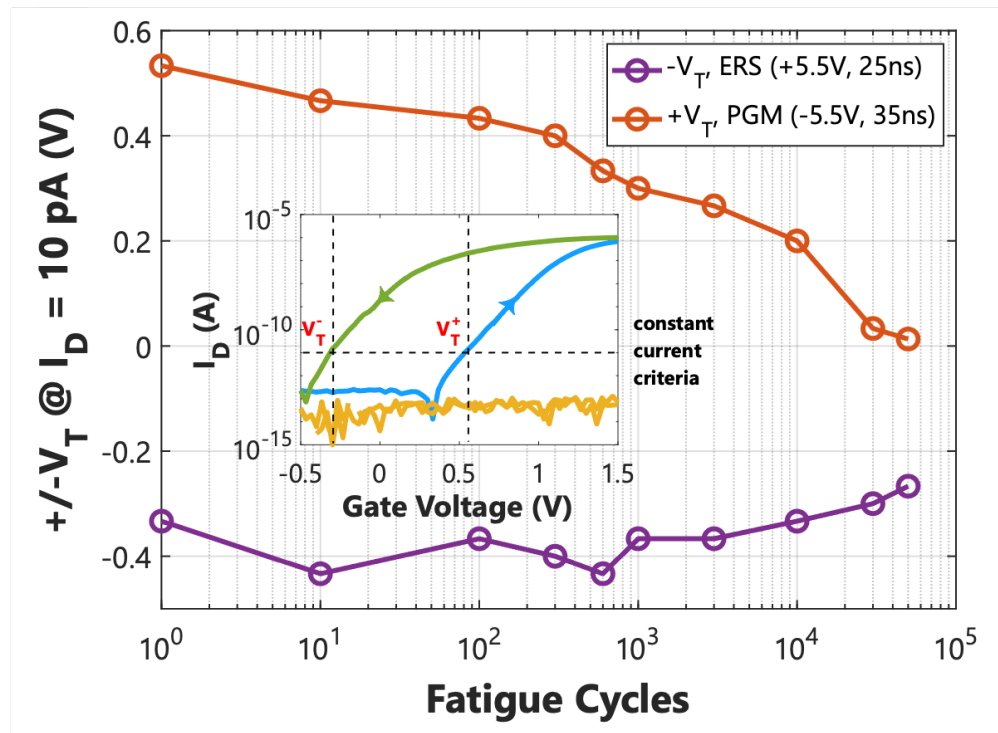
Figure 5.4: Tracking the $\pm V_T$ states of an FeFET as a function of fatigue cycles for a single device with W/L = 500/250 nm at a constant current criteria of 10 pA. Inset shows the initial starting memory window for the measured FeFET, with dashed lines indicating the constant current criteria which defines the threshold voltage.

Figure 5.4 shows that a gradual shrinking of the memory window occurs during the FeFET's lifetime. The $+V_T$ state tends to drift to the left, and the $-V_T$ state similarly tends to drift to the right. Over the duration of the cycling test, the total ΔV_T corresponding to the memory window becomes smaller, and eventually disappears once charge trapping hysteresis dominates the device characteristics.

5.2   Impact of Hot Electron Effect on Endurance

One characteristic of note in Table 5.1 is the range of film thicknesses investigated for the gate stack in ferroelectric HfO_2 transistors. Most studies on FeFETs utilize HfO_2 films which are typically between 9 - 30 nm in thickness [30, 86], as this is the range for which the ferroelectric properties can be stabilized and maximized. However, in order to reduce both the switching voltage required for the ferroelectric and degradation of the interfacial layer, scaling of the ferroelectric layer thickness is necessary. Furthermore, in order for FeFETs to be a scalable memory technology, it will be imperative to continue investigating thinner ferroelectric HfO_2 films that can enable good device performance at shorter device gate lengths. Therefore, the FeFETs characterized in this work will all use ferroelectric oxide thicknesses on the order of 4.5 nm.

One surprising result when working with such thin ferroelectric oxides is that the degradation mechanisms are more complicated than the simplistic picture of interfacial layer degradation introduced in this chapter. For ferroelectric oxides on the order of 5 nm or less, bulk charge trapping (which is typically a big concern in gate stacks with ≥ 9 nm HfO_2) is more or less negligible; rather, it is hot electron injection from the gate side which is primarily responsible for endurance degradation over time.
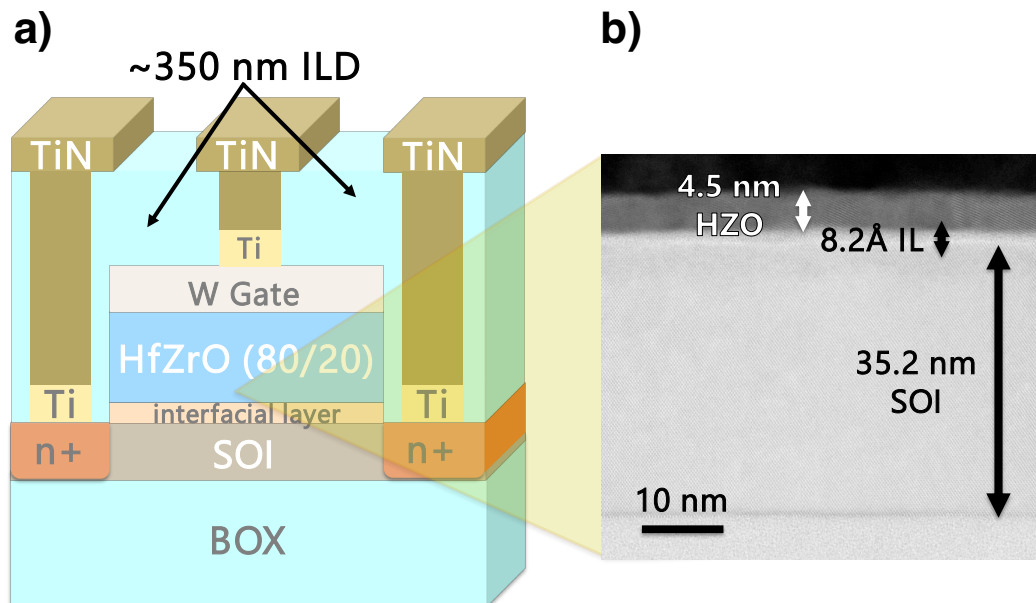
Figure 5.5: Baseline FeFET device structure characterized. (a) 2D cross-section across channel of the FeFET, showing the gate stack composition and materials used in the process flow. (b) TEM (tunneling electron microscopy) image of the gate stack of a fabricated device.

Figure 5.5(a) and (b) show a schematic and a cross-sectional TEM image, respectively, of an FeFET characterized in the investigation of additional endurance degradation mechanisms. The process flow to realize the FDSOI FeFET on SOI uses a gate-first, self-aligned process as described in [124] and in detail again in Chapter 7 of this dissertation.

FeFET Programmability as a Function of Temperature
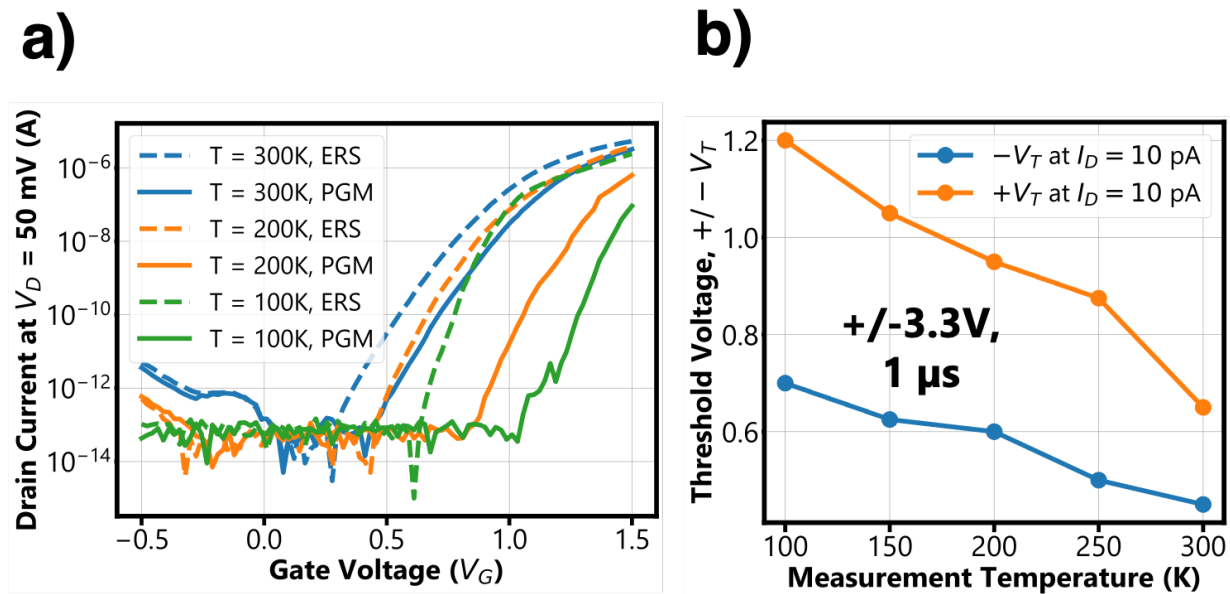
Figure 5.6: FeFET programmability as a function of measurement temperature. (a) Counterclockwise (CCW) $I_D V_G$'s obtained after applying ± 3.3V, 1 μs pulses as a function of T. (b) Despite general left shift of $\pm V_T$ with T, shift in $-V_T$ is lesser due to the increased switchability of the FE when starting with a fully inverted channel.

Figures 5.6 and 5.7 detail the programmability of the FeFET as a function of measurement temperature and programming pulse widths (t_{pw}) at a fixed gate voltage (V_G) of ± 3.3V. Notably, the memory window increases for $t_{pw} = 1$ μs as the background measurement temperature decreases from 300K to 100K, as shown in Figure 5.6(a). This effect is attributed to the dependence of ferroelectric domain wall motion on temperature; as measurement temperature increases, increased thermal fluctuation makes the depinning of domain walls less abrupt, leading to a smoother switching characteristic [39]. The absolute threshold voltages which govern the memory window characteristic are also plotted in Figure 5.6(b) at each discrete measurement temperature. Here, it is again clear that the memory window widens as temperature decreases. The left shifting of V_T to larger values with increasing temperature is primarily a consequence of effective body doping at the

various temperatures and to first order is unrelated to the properties of the ferroelectric later.

It is evident that at lower measurement temperatures, the memory becomes larger for a fixed V_G and t_{pw}, assuming the applied V_G is always greater than the coercive voltage. This trend is captured in Figure 5.7(b) for a range of different pulse widths. Figure 5.7(a) details the widening of the memory window for a fixed pulse amplitude of ±3.3V and increasing t_{pw} from 100 ns to 10 μs at room temperature.
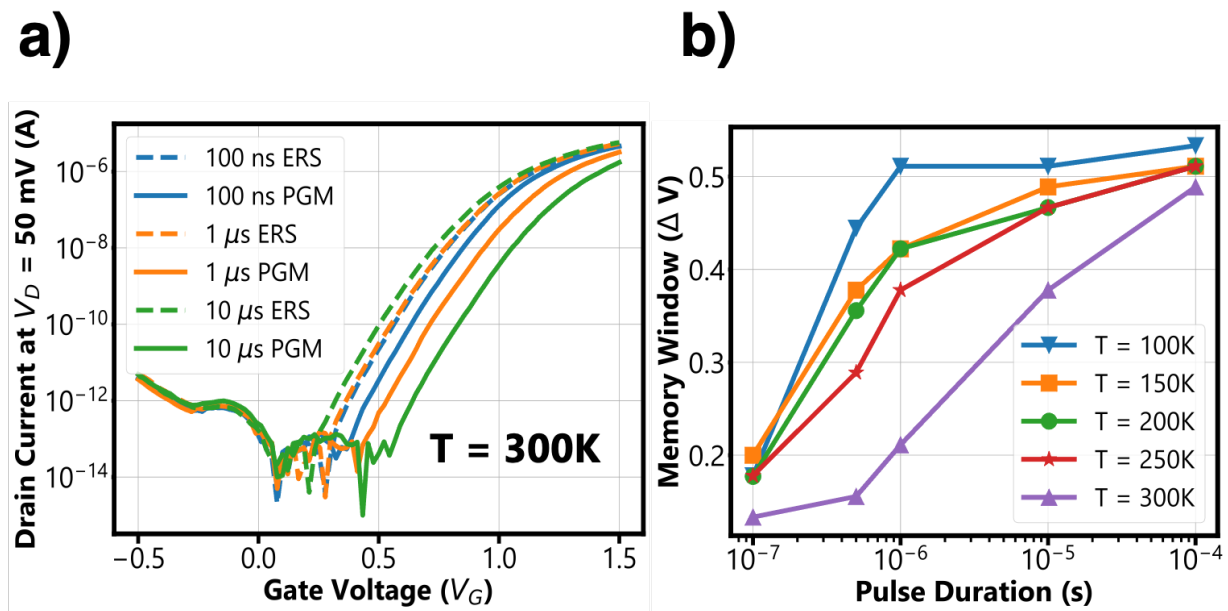
Figure 5.7: FeFET programmability as a function of measurement temperature (continued). (a) CCW $I_D V_G$'s obtained after variable width (t_{pw}) pulses, ±3.3V magnitude, at room temperature. (b) Memory window as a function of t_{pw}, ±3.3V pulses, at temperatures ranging from from 100 – 300K.

FeFET Endurance as a Function of Temperature

Based on the programmability results shown in the previous subsection, a cycling condition of ±3.3V, 1 μs is chosen to test FeFET endurance at every measurement temperature. Determination of the written FeFET state is conducted according to the procedure outlined in Figure 4.5(a) in Chapter 4, where $t_h = 1$ μs and $\pm V = \pm3.3$V.

The degradation of the memory window and its dependence on measurement temperature is captured in Figure 5.8 as well as Figure 5.9. Figure 5.8 plots the results of the endurance cycling test at every investigated measurement temperature, with the total memory window size normalized to the starting memory window size at that given
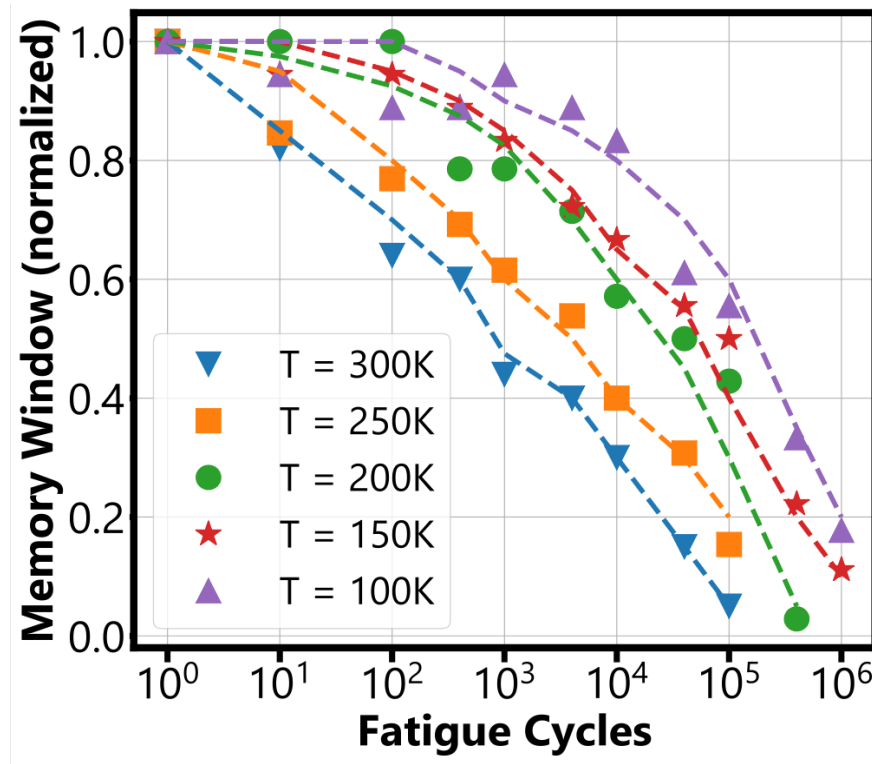
Figure 5.8: FeFET memory window, normalized to the starting memory window size for ease of comparison, as a function of fatigue cycles for measurement temperatures ranging $100 - 300K$.

temperature, for ease of comparison. It is evident that the memory window degradation characteristics show a clear dependence on measurement temperature, with the memory window persisting for a greater number of endurance cycles as the background temperature decreases. Previously suggested mechanisms for endurance degradation – namely, the electric-field driven degradation of the interfacial layer – cannot exactly explain this strong temperature-dependent result. Figure 5.8(a) therefore provides the first insight into an additional source of endurance degradation which is appears to have a particularly substantial effect in highly scaled FeFETs.

Figure 5.9(a) takes the two extreme temperatures investigated – room temperature (300K) and 100K – and tracks the shift in $\pm V_T$ until the end of the endurance test. If defect generation were primarily electric-field driven, as suggested by prior works, one would not necessarily expect the memory window to survive longer at lower measurement temperatures [102]. To understand this temperature-dependent degradation effect further, it is important to observe the *left-shifting* of both the high and low V_T states with cycling (as seen in Figure 5.9(a)). Furthermore, it is seen that this shift is exacerbated at
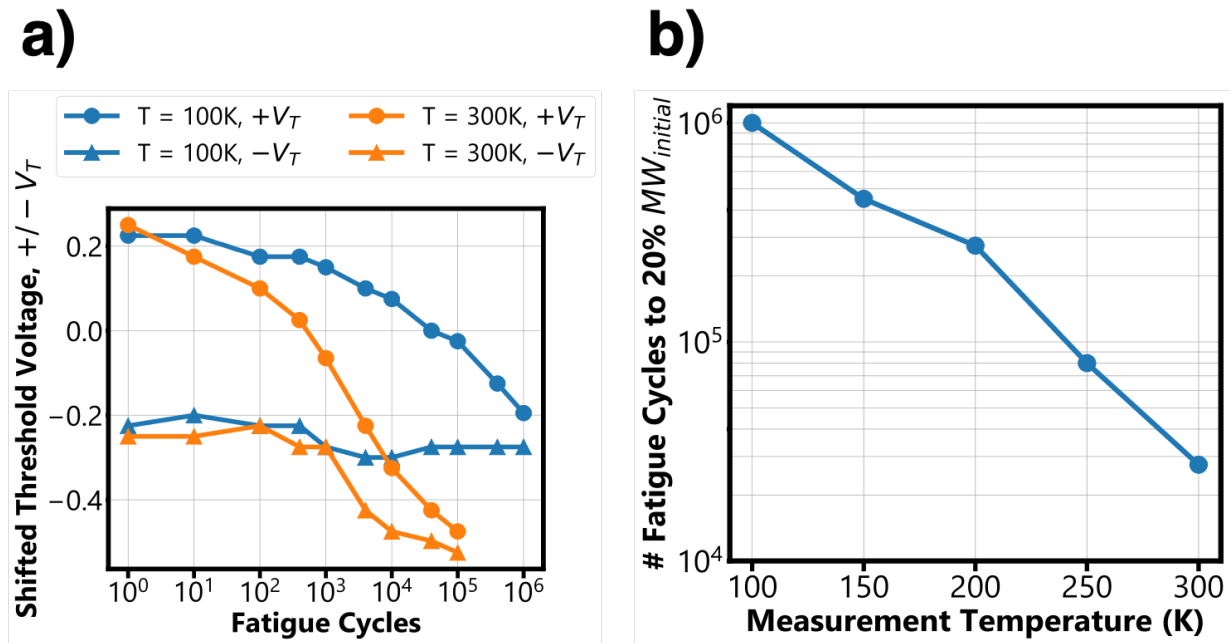
Figure 5.9: Temperature-dependent endurance characteristics of the FeFET. (a) Both $\pm V_T$ suffer from a left shift due to hole trapping, an effect which worsens with increasing measurement temperature. (b) Number of fatigue cycles to reach 20% of the initial memory window size, MW_{initial}, as function of measurement temperature.

higher measurement temperatures. The nature and direction of the shift suggests strong *hole injection* which occurs during endurance testing. It is important to note here that the p-body FDSOI is n+ doped at the source and drain, with no substrate contact to provide additional holes to the body. Therefore, the only way a substantial number of holes can be generated in this scenario (which then become trapped in the gate oxide) is through **hot electron-induced damage** at the substrate, specifically during the application of negative gate biases.

One additional observation of note is that, unlike the results shown in the introductory section of this chapter in Figure 5.4 which shows a right shifting of the $-V_T$ state, the endurance behavior of these FeFETs is distinctly different. The device structure for both devices is nearly the same, with the primary difference being the absence of the 1.5 nm of non-ferroelectric HfO_2 in the devices characterized in this subsection. This can be confirmed by comparing the TEMs shown in Figure 4.4(b) and Figure 5.5(b), with the latter corresponding to the devices characterized in the temperature-dependent studies and the former corresponding to the device characterized in the introductory section of this chapter. With a reduction of less than 2 nm in physical thickness between the two different device splits, it is clear that this temperature-dependent degradation effect is also

intrinsically tied to the gate stack thickness.

Figures 5.10(a) and (b) support the hypothesis of hot electron-induced damage and subsequent hole injection into the gate stack as a major contributor to device degradation. Figure 5.10(a) tracks the change in the device's subthreshold swing (SS) for the $+V_T$ $I_D V_G$ sweep during the state determination test, both before and after a total of 10^5 fatigue cycles has been applied. The subthreshold swing after cycling clearly increases with measurement temperature, indicating a stronger and stronger trap generation from broken bonds near the channel. The holes released during this interfacial bond breakage therefore become trapped in the gate stack over time, leading to a greater and greater left shifting of both V_T states. Figure 5.10(b) plots the percentage increase in the subthreshold swing prior to and after cycling as a function of measurement temperature, to clearly illustrate the trend.

a)

b)
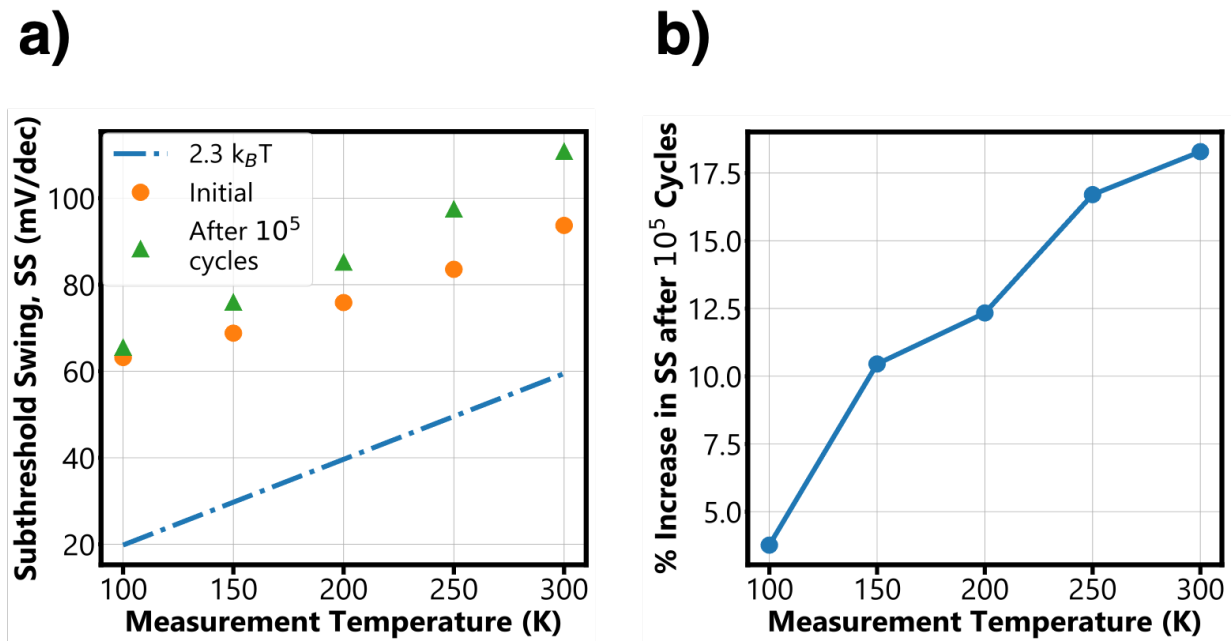
Figure 5.10: Evolution of the interface properties of the FeFET as a function of measurement temperature. (a) Subthreshold swing (SS) before and after 10^5 fatigue cycles for temperatures ranging from $100 - 300$K. (b) Percentage increase in SS with temperature, derived directly from from Figure 5.10(a).

Assessing Bulk Charge Trapping in Scaled FeFETs

The distinct disparity in the nature of endurance characteristics discussed in the previous subsection, resulting from a mere reduction in gate stack thickness of 1.5 nm, points to a

need to revisit another commonly cited source of endurance degradation introduced in the introductory section of this chapter: bulk charge trapping in high-κ oxides. It is important to understand whether this effect plays any substantial role in endurance performance for highly scaled FeFETs, as its effect has mostly been investigated in gate oxides with thicker stacks.

One proxy for understanding this effect is to investigate the evolution of the gate leakage through the device stack during the cycling tests. As seen in Figure 5.11, there appears to be a very minimal increase in gate current, or I_G, after cycling, regardless of measurement temperature. The percentage increase in gate leakage is also plotted as a function of measurement temperature in Figure 5.11(c), and it remains relatively constant. These observations suggest that bulk high-κ charge trapping plays an insignificant role in endurance degradation in these devices, as one would otherwise expect orders of magnitude increase in I_G [102].
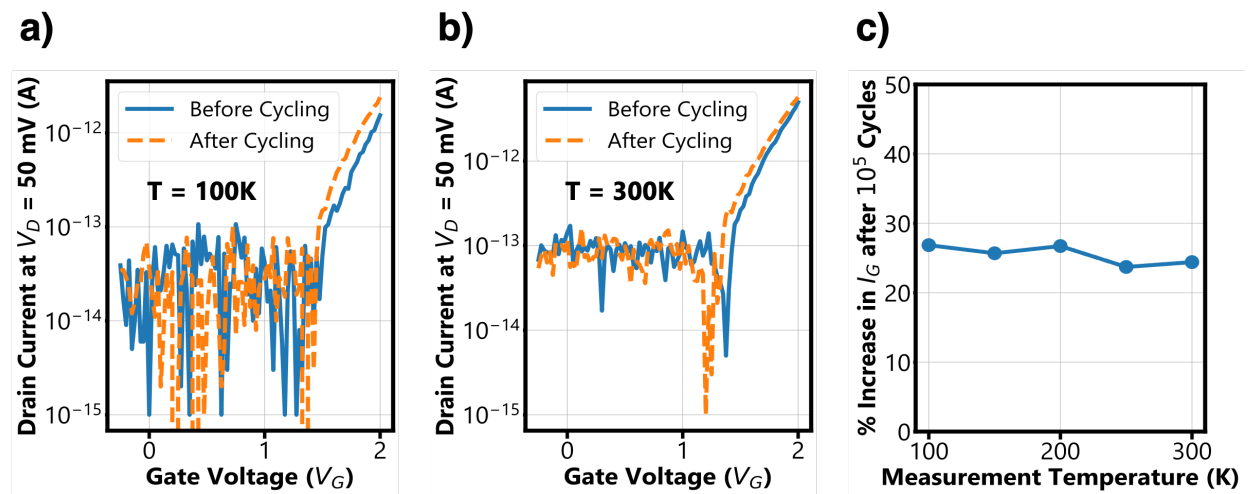
Figure 5.11: Evolution of the gate leakage in the FeFET as a function of measurement temperature. (a) $I_G V_G$'s before and after 10^5 endurance cycles for T = 100K. (a) $I_G V_G$'s before and after 10^5 endurance cycles for T = 300K. (d) Percentage increase in I_G after 10^5 cycles as a function of measurement temperature. The increase in I_G is not only negligible, but shows no real temperature dependence.

To further understand the possibility for as well as the extent of bulk charge trapping in the FeFET, one can also conduct a so-called *charge-trapping measurement*, which will directly measure the hysteresis due to charge trapping in the gate stack. The details of such a charge-trapping measurement are detailed in Figure 5.12. This transient current measurement involves first applying a state-setting pre-pulse to the gate of the FeFET in order to switch the polarization fully. Then, after allowing the drain voltage some time to ramp up and stabilize, the the drain current I_D measured during application of the sec-

ond gate pulse, which is of the same polarity as the first. This second pulse will capture the hysteresis purely from charge trapping and not ferroelectric switching. It is important also to mention that the duration of the pulse will determine the amount of charge trapping hysteresis measured, as the traps which are able to respond may have varying time constants corresponding to their location in the bandgap. A pulse duration of 500 μs is chosen accordingly to ensure that the slowest bulk traps are able to respond during the measurement. Plotting the resulting time-dependent drain current, $I_D(t)$, against the time-dependent gate voltage, $V_G(t)$, will then yield the charge trapping hysteresis.
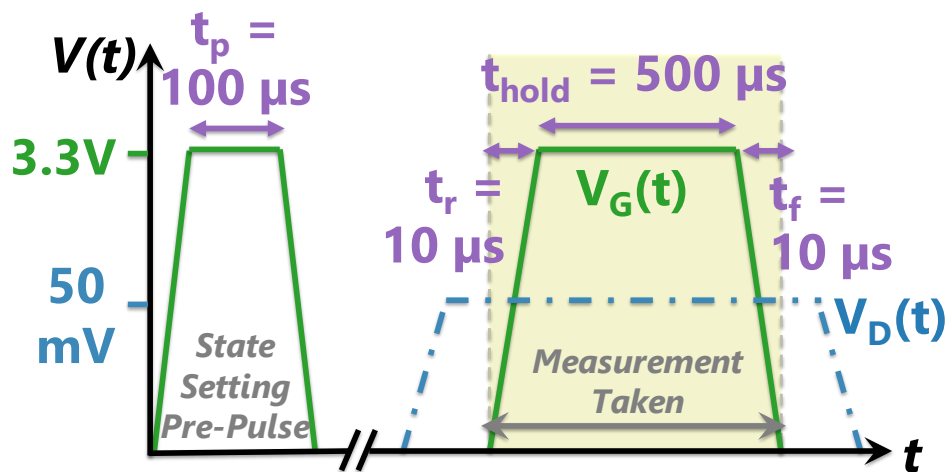
Figure 5.12: Transient current measurement and pulse scheme for capturing charge trapping within the high-κ ferroelectric oxide. No measurement is taken during the state-setting pre-pulse; the current measurement is taken only during the second pulse of the same polarity.

Curiously, the charge trapping hysteresis remains largely unchanged both before and after cycling to 10^5, as shown in Figure 5.13(a) and (b); furthermore, the magnitude of the hysteresis is virtually temperature-independent. In contrast, charge trapping measurements simulated in Figure 5.13(c) and (d) using the Ginestra^{TM} modeling platform [4] at 100K and 300K, assuming standard high-κ bulk charge trapping effects, predict increased hysteresis as background temperature decreases. This simulation result is consistent with what is observed in standard charge trapping flash memories and in FeFETs with thicker gate oxides.

Essentially, bulk trap emission, if dominant, is an effect which should increase as measurement temperature decreases, as the trapped defects cannot emit as efficiently with reduced kinetic energy in the system. This ought to lead to a larger and larger charge trapping hysteresis. Not only is there a negligible difference in the magnitude of initial
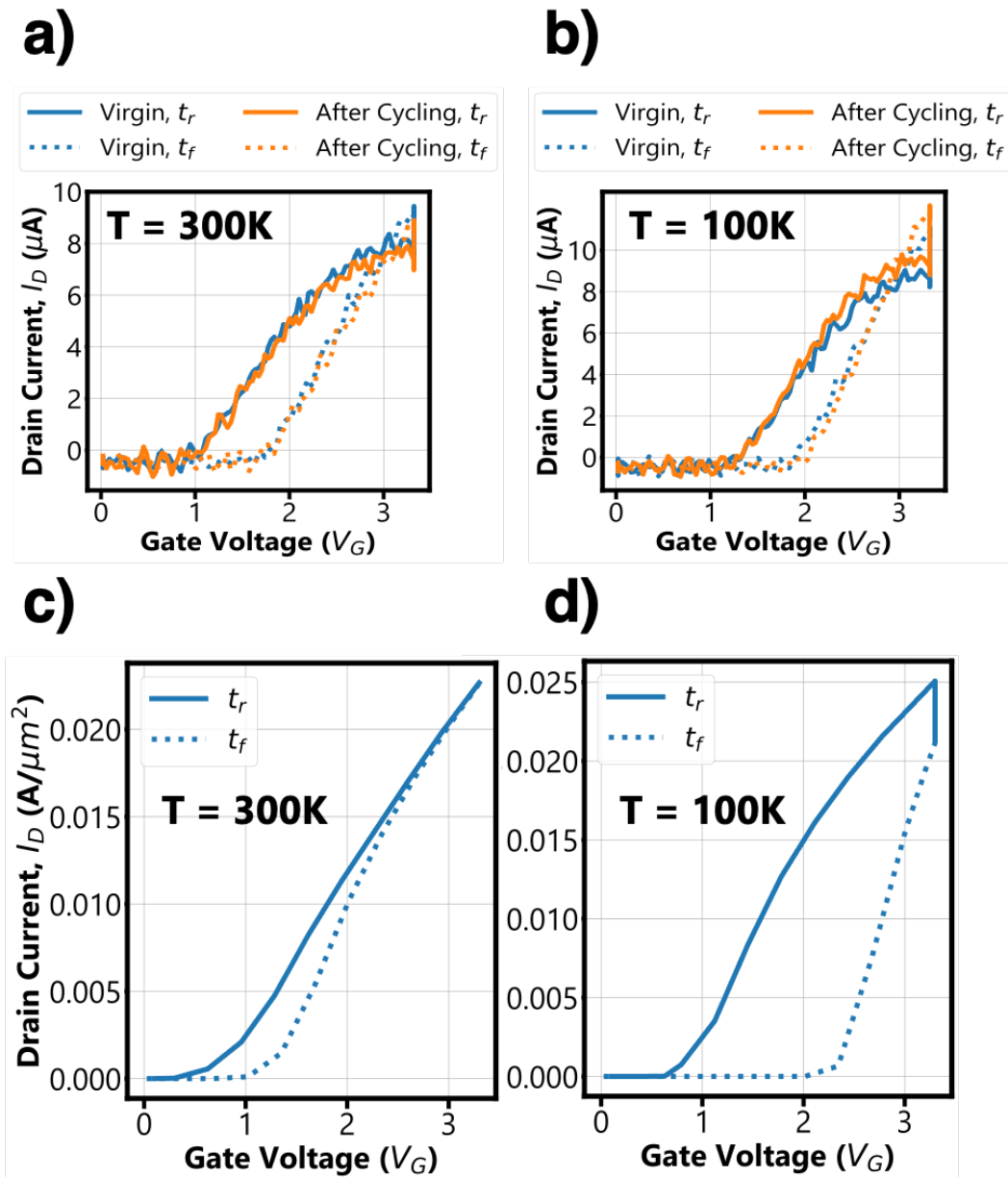
Figure 5.13: Results of the charge trapping measurement at different measurement temperatures, compared against simulation results. (a) $V_G(t)$ vs. $I_D(t)$ plotted before and after 10^5 endurance cycles, at 300K. (b) Same as (a), but for a measurement temperature of 100K. Intrinsic and post-cycling charge trapping hysteresis are comparable. (d) Simulated charge trapping at 300K. (e) Simulated charge trapping at 100K.

charge trapping hysteresis at either 100K or 300K, as evidenced in Figure 5.13(a) and (b), but the magnitude of charge trapping hysteresis also does not appear to change substantially after the endurance test.

The aforementioned observations, combined with the V_T left-shifting trends seen in Figure 5.9(a) and the subthreshold swing data in Figures 5.10(a) and (b), suggest that hot electron-induced damage is very likely the primary and dominant source of degradation in FeFETs with highly scaled gate thicknesses. This mechanism of hot electron injection and subsequent hole damage to the oxide is illustrated in the band diagram of Figure 5.14.
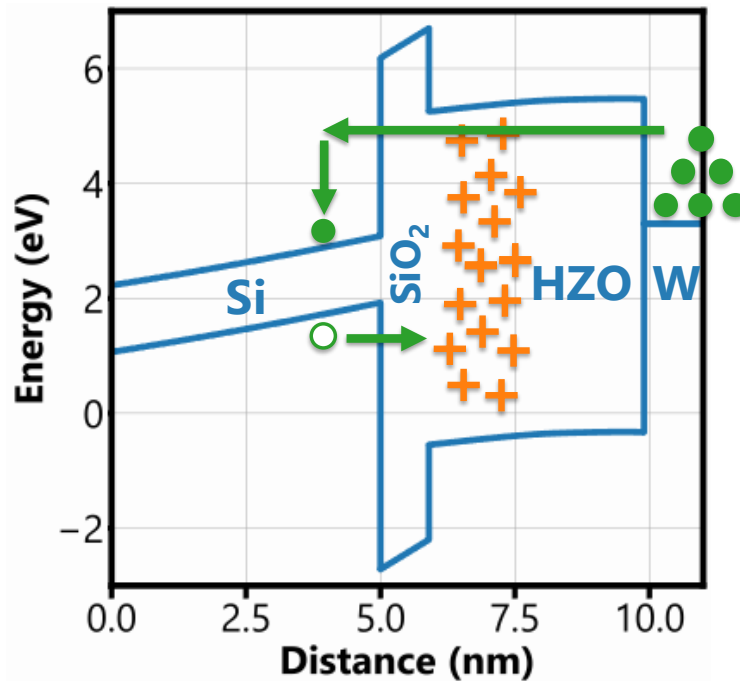
Figure 5.14: Simulated band diagram of the FeFET gate stack after endurance cycling. When $V_G = -3.3\text{V}$ is applied, positively charged holes enter the Zr-doped HfO$_2$ layer close to the channel interface, push both V_T's to the left, and eventually close the memory window.

Through rigorous temperature-dependent characterization and simulations, it is shown that the typical causes for endurance degradation for previously studied FeFETs, namely bulk charge trapping, are not a primary concern in thin ferroelectric HfO$_2$ layers. It is demonstrated that hot electron-induced damage close to the channel/interfacial layer is the main mechanism for endurance degradation in 4.5 nm HfO$_2$ FeFETs. Appropriate device design to minimize the hot electron effect may also mitigate endurance degradation in scaled FeFETs. Some of these solutions will be explored in Chapter 6 of this dissertation.

5.3   Predicting Reliability Challenges Quickly

The discussion in previous subsections has centered around identifying and exploring the various degradation mechanisms which contribute to premature failure of the HfO_2-based FeFET. However, in order to predict and assess the FeFET's electrical performance efficiently, it might be prohibitively expensive and time-consuming to run several full-fledged FeFET wafer splits, each possibly requiring multiple steps of lithography, deposition and etch.
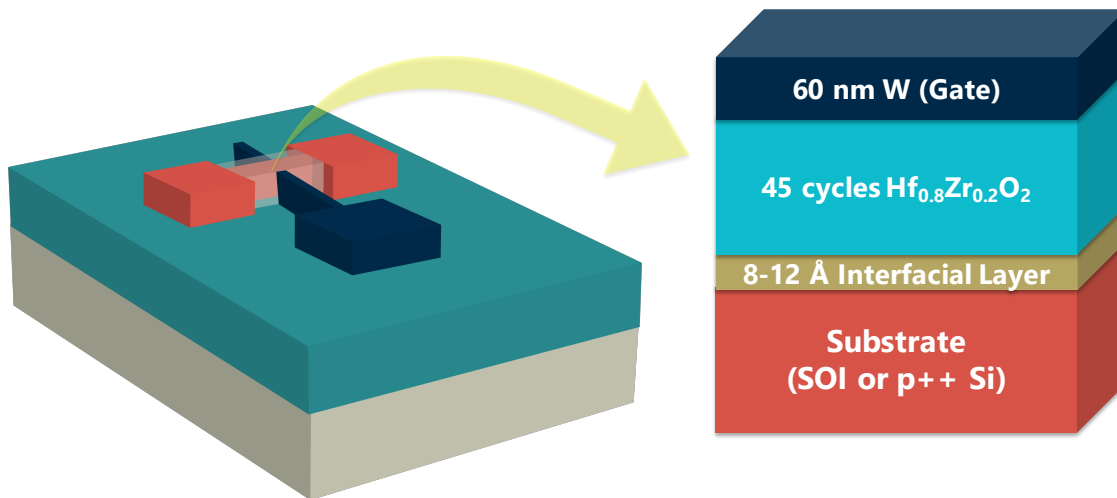
Figure 5.15: 3D schematic of a typical SOI FeFET and associated gate stack design. When performing predictive testing in the MOSCAP structure, the substrate used for testing is highly doped p-type bulk silicon rather than SOI.

The testing methodology introduced in this section provides a systematic, reliable, and rapid method to qualitatively predict the FE endurance of prospective gate stack designs *prior* to running a full FeFET fabrication process. The idea behind this testing methodology is to take the critical component which governs the performance and reliability of a MOS transistor – the gate stack – and test it in a simple MOS capacitor, or MFIS (metal-ferroelectric-insulator-semiconductor) structure first. Figure 5.15 shows cartoon schematics of the SOI FeFET tested and its corresponding gate stack design, which is then replicated in the MOSCAP testing structures.

MOSCAP Device Fabrication

MOSCAPs incorporating ferroelectric Zr-doped HfO_2 gate stacks, with a thickness of ~ 4.5 nm, are realized via a one-step lithography process on highly doped Si, and are compared against real endurance results from SOI FeFETs incorporating the same oxide designs.
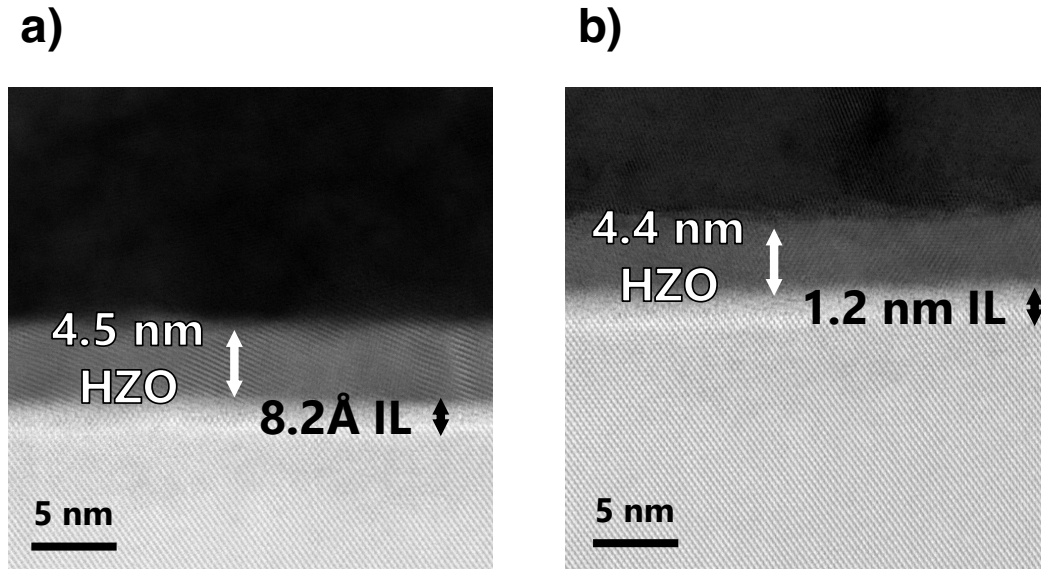
Figure 5.16: Cross-sectional TEMs of gate stacks investigated in both FeFET and MFIS structures. (a) TEM of 4.5 nm Zr-doped HfO_2 on 8.2 $\overset{\circ}{A}$ of SiO_2. (b) TEM of 4.5 nm Zr-doped HfO_2 on 8.2 $\overset{\circ}{A}$ of SiO_2 and 4 $\overset{\circ}{A}$ of Al_2O_3.

Prospective Zr-doped HfO_2 gate stacks with an ALD supercycle ratio of 4:1 Hf:Zr (as investigated in [61] are grown on highly doped, p^{++} Si ($\sim 10^{19}$ cm^{-3}) substrates. The tungsten gate electrode is patterned and etched after the ferroelectric crystalline phase anneal at 475 °C, 40 seconds. The SOI FeFET follows an identical gate deposition/definition process as the MFIS structure, with additional steps to thin down the active Si to ~ 35 nm prior to deposition of the ferroelectric oxide, device isolation via 350 nm of SiO_2 after source/drain implantation and activation, and an additional metallization step to contact the device terminals. Figure 5.15(a) and (b) show schematics of the SOI FeFET/MFIS MOSCAPs investigated, and their corresponding gate stacks.

Figures 5.16(a) and (b) showcase TEMs of the two different gate stacks investigated in this work. The stack in Figure 5.16(a) consists of 45 cycles of $Hf_{0.8}Zr_{0.2}O_2$ on $8\overset{\circ}{A}$ of self-terminated, chemical SiO_2, and the stack in Figure 5.16(b) consists of an additional 5 cycles of Al_2O_3 (and the same 45 cycles of $Hf_{0.8}Zr_{0.2}O_2$ on $8\overset{\circ}{A}$ SiO_2). In this work, the insertion of Al_2O_3 is investigated as a means to reduce leakage through the thin gate stack without heavily compromising the EOT (equivalent oxide thickness) of the gate stack.

Measurement Procedure

The methods for testing endurance evolution in the FeFET structure is detailed in Chapter 5, and the same procedure is used for evaluating FeFET endurance in this subsection as well. This subsection will focus on how to quantify and evaluate endurance in the MFIS device structure.

a)

b)
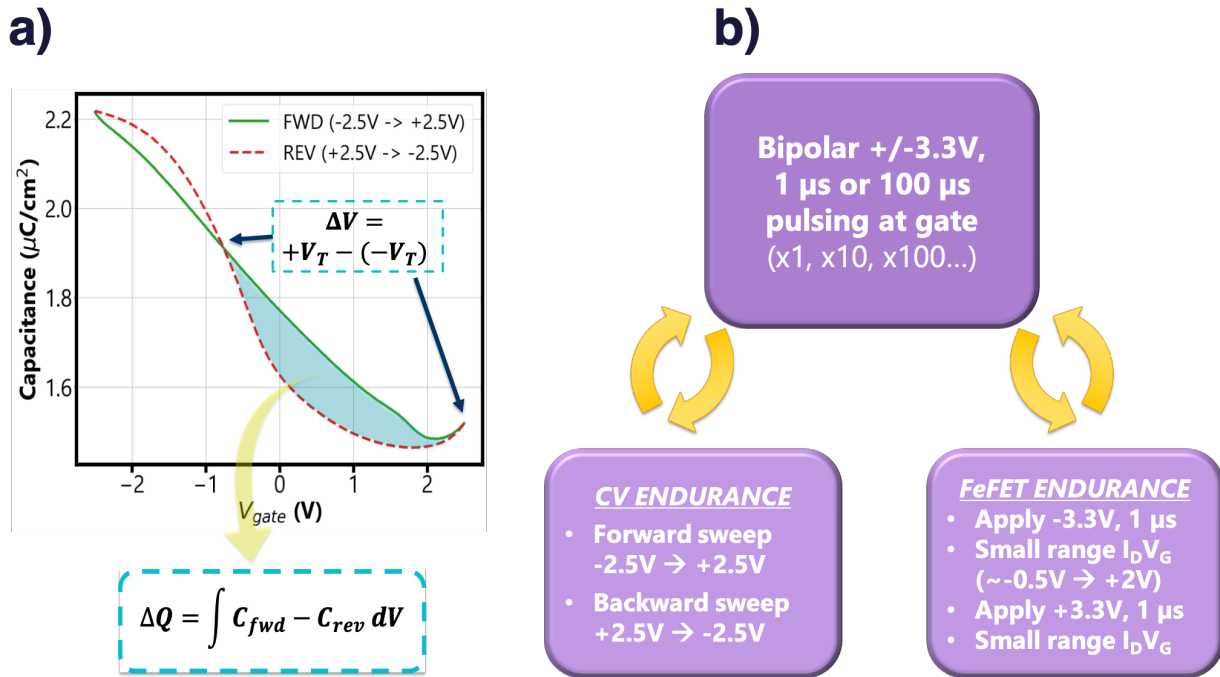
Figure 5.17: (a) Sample doubly-swept CV on as highly doped p++ substrate. Blue shaded area indicates the voltage sweep range over which the CVs are integrated to obtain the total charge, indicating the separation between both memory states. (b) Bipolar cycling and state determination for both MOSCAP and FeFET device structures.

As shown in the testing flow detailed in 5.17(b), the actual stressing sequence for both the MFIS device and FeFET device are identical – in this case, bipolar voltage pulses of ± 3.3V are applied to the gate of the device in order to fatigue the device. The state determination, likewise, is performed every decade of cycling (e.g, 10, 100, 1000, ... cycles, and so forth) until failure.

The state determination for the FeFET is performed as discussed in Chapter 5, using a small-window $I_D V_G$ sweep to minimize read disturbance. Then, to determine the memory window of the MFIS structure, following the left panel of the flowchart in 5.17(b), a bidirectional CV (capacitance-voltage) sweep is performed at 100 kHz, starting from -2.5 V to +2.5V and back to -2.5V. The sweep range is carefully chosen so as to visualize the

ferroelectric hysteresis in the MFIS structure, but also to minimize substantial gate leakage through the thin film stack at the voltage extremes.

In order to determine endurance evolution for the MFIS MOSCAP, the *integrated charge, or* ΔQ, between the $\pm V_T$ states of a doubly-swept CV (as shown in Figure 5.17(a)) is tracked after each sequence of ± 3.3V, 100 μs cycles. ΔQ is defined as:

$$\Delta Q = \int_{-V_T}^{+V_T} (C_{\text{fwd}} - C_{\text{rev}}) \cdot dV_G \tag{5.1}$$

where C_{fwd} and C_{rev} indicate the forward and backward CV sweeps. For ease of comparison to the FeFET results, the $\pm V_T$'sin the MOSCAP structure are loosely and analogously defined as the two V_G values at which the forward and backward CV sweeps intersect. These values do not correspond to the "real" $\pm V_T$ values of the MOSCAP devices, but the crossover points will be used to track how the memory window shifts and/or degrades over the course of the endurance test. Tracking the total integrated charge gives insight into the degradation behavior of the potential gate stacks

Both the shifting values of V_T and rate of decrease of ΔQ are metrics which are indicative of device fatigue, as charge-trapping through the FE oxide is well-understood as a mechanism for the closure of the memory window (MW), as discussed in Chapter 5.

Measurement Results

In comparing Figures 5.18(a) and (b) (corresponding to the 45 cycle Zr-doped HfO$_2$ and 5 cycle Al$_2$O$_3$/45 cycle Zr-doped HfO$_2$ MOSCAPs, respectively), the initial starting memory window determined by $\pm V_T$ is larger for the latter compared to the former. However, the ΔQ is smaller, evident from the reduced separation between the forward and backward sweeps. Furthermore, the insertion of Al$_2$O$_3$ reduces leakage through the gate stack (0.17 mA/cm^2 → 0.067 mA/cm^2 at +1V), which also reduces the "abruptness" of ferroelectric switching (which is typically facilitated by leakage). However, the additional thin $\sim 4\mathring{A}$ of non-ferroelectric Al$_2$O$_3$ does not reduce the predicted memory window. In fact, the initial memory window as determined by $\pm V_T$ *increases* with Al$_2$O$_3$ insertion.

However, when referring to Figure 5.18(c) and (d), it is apparent that the stack with Al$_2$O$_3$ experiences a stronger degradation in overall memory window size (nearly a 1V shift in the $-V_T$ state as compared to the sample without Al$_2$O$_3$, which shifts ~ 350 mV), and a correspondingly stronger decrease in overall ΔQ as well. These observations are extracted from Figures 5.18(a)-(d) and summarized in Figure 5.19(a) and (b) as total normalized memory window (MW) vs. cycles, and normalized ΔQ vs. cycles, respectively, for both gate stacks. By the end of 10^6 fatigue cycles, ΔQ drops to 23% for the sample with Al$_2$O$_3$ as compared to 53% of its starting value for the sample without Al$_2$O$_3$. This observation is important to note, as many FeFET researchers often engineer the device gate stacks to increase the starting memory window size to be as large as possible. Whereas cycling endurance of the device is often a function of the starting memory window size,
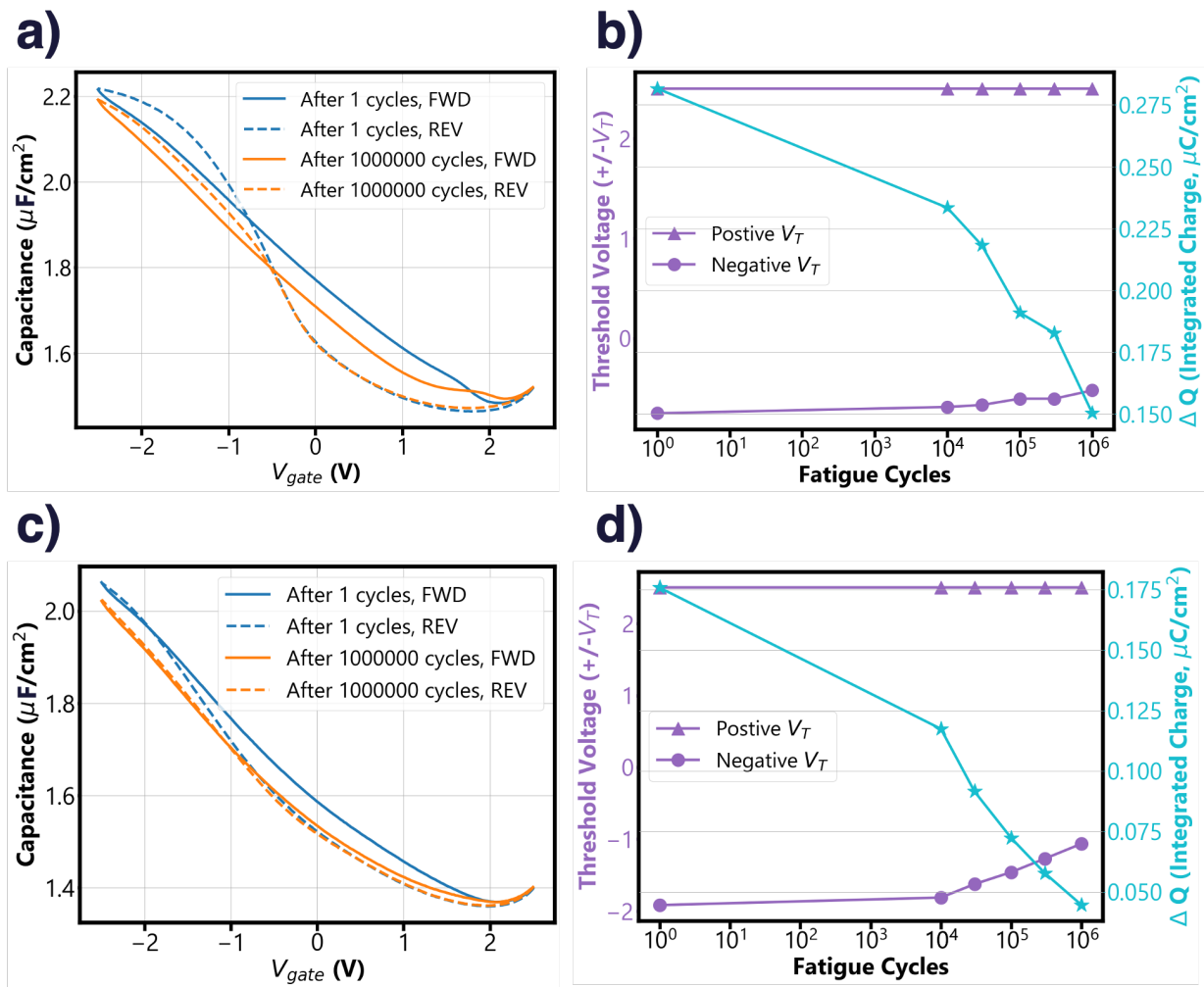
Figure 5.18: CV and ΔQ evolution for the gate stacks investigated. (a) Double-swept CV's before and after endurance cycling for the 45 cycle Zr-doped HfO$_2$ MOSCAP. (b) Total ΔQ and $\pm V_T$ shifts as a function of fatigue cycles. (c) Same as (a), but for the 5 cycle Al$_2$O$_3$/45 cycle Zr-doped HfO$_2$ stack. (d) Same as (b), again for the 5 cycle Al$_2$O$_3$/45 cycle Zr-doped HfO$_2$ stack.

increased charge trapping in the gate stack can ultimately offset the benefit achieved in engineering a larger memory window.

Figure 5.20 quantitatively correlates the MOSCAP endurance to FeFET performance. Figure 5.20(a) and (b) illustrate the shift in both $\pm V_T$ states for the SOI FeFETs ($W/L = 1/0.3\mu$m) with 45 cycles Zr-doped HfO$_2$, with and without Al$_2$O$_3$, respectively. Evident from the $I_D V_G$'s taken periodically during the first thousands of cycles is the fact that al-
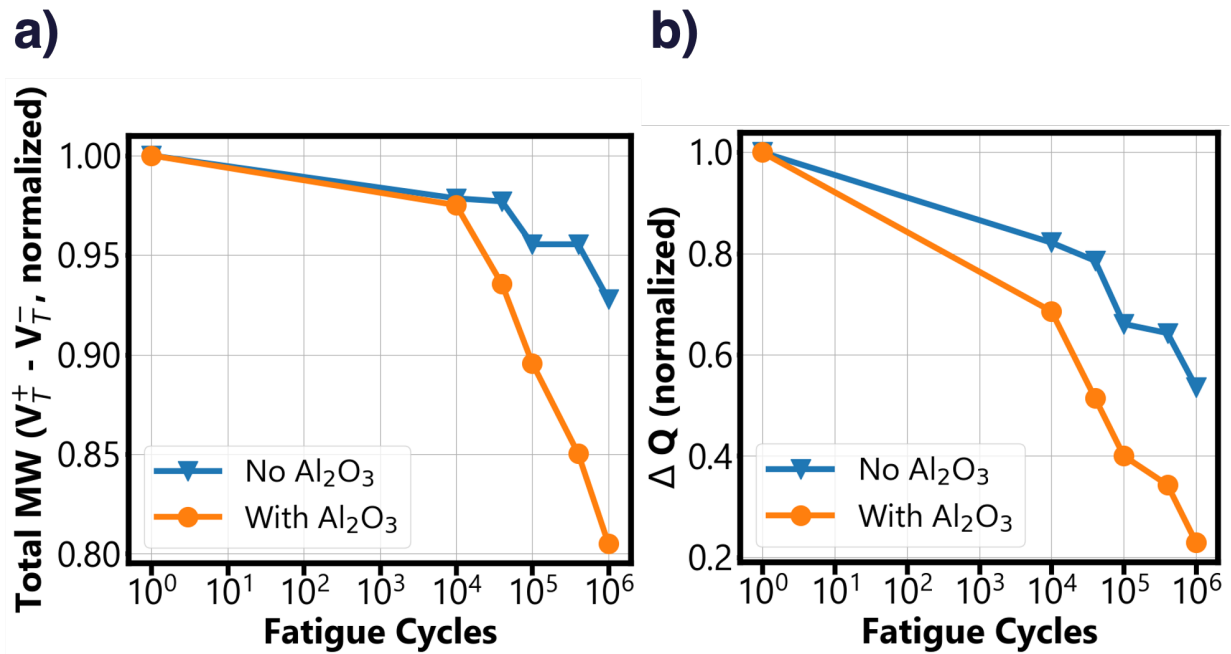
Figure 5.19: Evaluation of the endurance evolution for the two gate stacks investigated. (a) Comparison of the normalized memory window between MOSCAP with and without Al_2O_3. (b) Comparison of the normalized ΔQ between both MOSCAPs.

though the device with Al_2O_3 starts with a larger memory window, it experiences a faster rate of degradation than the device without. This result aligns with the prediction from the MOSCAP endurance results in Figures 5.19(a) and (b), and is summarized in Figure 5.20(c). Finally, we note that substrate doping/geometry differences affect the final voltage distribution from channel to electrode and may account for quantitative differences.

In summary, a rapid and conclusive testing method has been proposed to qualitatively predict the endurance characteristics of SOI FeFETs fabricated using ferroelectric HfO_2. This methodology can be extended to test other design aspects for endurance in addition to oxide design (pre/post-annealing conditions, gate electrode materials, etc.).
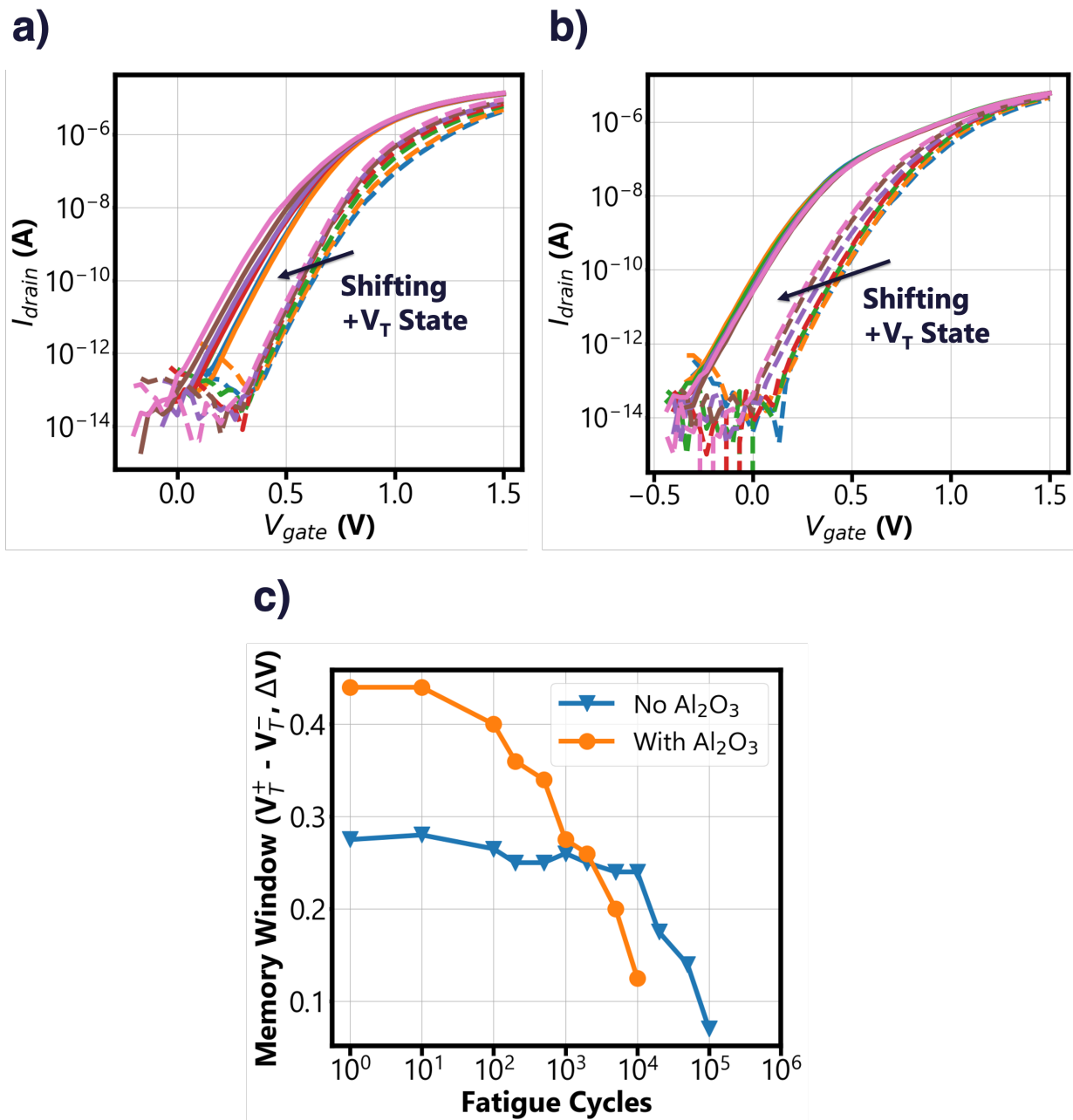
Figure 5.20: FeFET endurance evaluation for the two gate stacks investigated. (a) Example of shifted $I_D V_G$'s for the FeFET made with 45 cycles Zr-doped HfO_2 during endurance testing. (b) Same as (a), but for the 5 cycle Al_2O_3/45 cycle Zr-doped HfO_2 stack. (c) Absolute memory as a function of fatigue cycles for both FeFETs.

Chapter 6

Processing Techniques for the HfO$_2$ FeFET

The primary focus of this chapter is to underline the importance of process integration for the HfO$_2$-based FeFET, and then to identify the cause-and-effect relationships between process design and resulting device behavior. As with any other nascent memory technology, it is imperative to, upon identifying key reliability and performance concerns, address them appropriately through device design engineering.

Details regarding the processing techniques and fabrication flows used to realize all FeFET structures are provided in Appendix A.

6.1   Interfacial Oxide Design for Interface State Improvement

In this section, the nature of the interface states induced during the integration of ferroelectric Zr-doped HfO$_2$ on silicon is examined closely. Metal-ferroelectric-insulator-silicon (MFIS) capacitors, with a thin layer of Zr-doped HfO$_2$ grown by atomic layer deposition as the ferroelectric and various interfacial oxide layers as the insulator, are investigated. Since a high-temperature post-anneal is necessary to induce the formation of the ferroelectric phase in this oxide stack, as discussed in Chapter 2, the integrity of the oxide/silicon interface must be preserved after high-temperature processing. It is shown here that a SiO$_x$N$_y$ nitrided interlayer provides an improved midgap interface state density among all interfacial oxides investigated. Furthermore, the density of the interface states (D_{it}) is quantified using the AC conductance technique, and a model for the interface trap distribution across the silicon bandgap is proposed in order to explain and verify the experimental measurements.

Experimental Design

One of the potential challenges barring successful integration of ferroelectric HfO$_2$ on silicon is the poor interface between the silicon substrate and the ferroelectric material. The atomic layer deposition process (ALD) can be tuned to produce a high quality HfO$_2$-based gate oxide, but the subsequent stabilization of the ferroelectric phase involves a moderately high temperature anneal (commonly 400 °C or higher), which may perturb the interface between oxide and silicon. The effects of this so-called phase anneal on film crystallinity are well-documented [86, 98], but the effects on device interfacial properties are yet to be properly understood.

By utilizing the AC conductance technique proposed by Nicollian and Goetzberger [96], the midgap trap density (D_{it}) is measured directly. Furthermore, it is shown that a nitrided interfacial layer between the Zr-doped HfO$_2$ layer and the silicon substrate minimizes degradation of the interface during the ferroelectric phase anneal, yielding a device with a peak midgap trap density in the mid-10^{11} cm^{-2} eV^{-1} range, which is amenable for modern day device technologies.

Three different interfacial oxides on lightly doped (10^{15} cm^{-3}) p-type silicon substrates were investigated: chemical oxide, thermal SiO$_2$, and thermal nitrided SiO$_2$. The chemical oxide was grown by placing a cleaned substrate in a heated solution of 2 NH$_4$OH:5 H$_2$O$_2$:200 H$_2$O, producing several monolayers of oxide. Thermal SiO$_2$ was grown by placing a silicon substrate in a rapid thermal annealing (RTA) chamber at 900 °C in an O$_2$ ambient for 20 seconds. Nitrided SiO$_2$ was grown by placing a silicon substrate in a RTA chamber at 900 °C in O$_2$ for 10 seconds followed by N$_2$O for 5 seconds in situ. Next, alternating monolayers of HfO$_2$ and ZrO$_2$ were grown by ALD to achieve an overall film composition of Hf$_{0.5}$Zr$_{0.5}$O$_2$. 100 total cycles were deposited to yield a film thickness of roughly 8.9 nm as determined by X-ray reflectometry (XRR). 60 nm of TiN was then sputtered onto the stacks, which were subsequently annealed in an RTA tool at 500 °C for 30 seconds in an N$_2$ ambient to crystallize the Zr-doped HfO$_2$ film.

The mixed crystalline phases in the resulting polycrystalline films were characterized with grazing incidence X-ray diffraction (GIXRD). Electrical measurements were taken using a Keysight B1500A Semiconductor Parameter Analyzer.

Preliminary Device Characterization and Analysis

For structural characterization, grazing-incidence X-ray diffraction was performed on the annealed Zr-doped HfO$_2$ films. As depicted in Figure 6.1(a), the GIXRD scan shows a set of diffraction peaks which suggests the presence of orthorhombic and/or tetragonal phases of hafnia [98]. The absence of diffraction peaks corresponding to the monoclinic phases of HfO$_2$ or ZrO$_2$ [112] or any other impurity phases of Hf$_x$Zr$_{1-x}$O$_2$ [140] can also be seen in the same scan.

To complement the XRD results, ferroelectric properties have been further verified through PE (polarization vs. electric field) and CV (capacitance vs. voltage) data shown
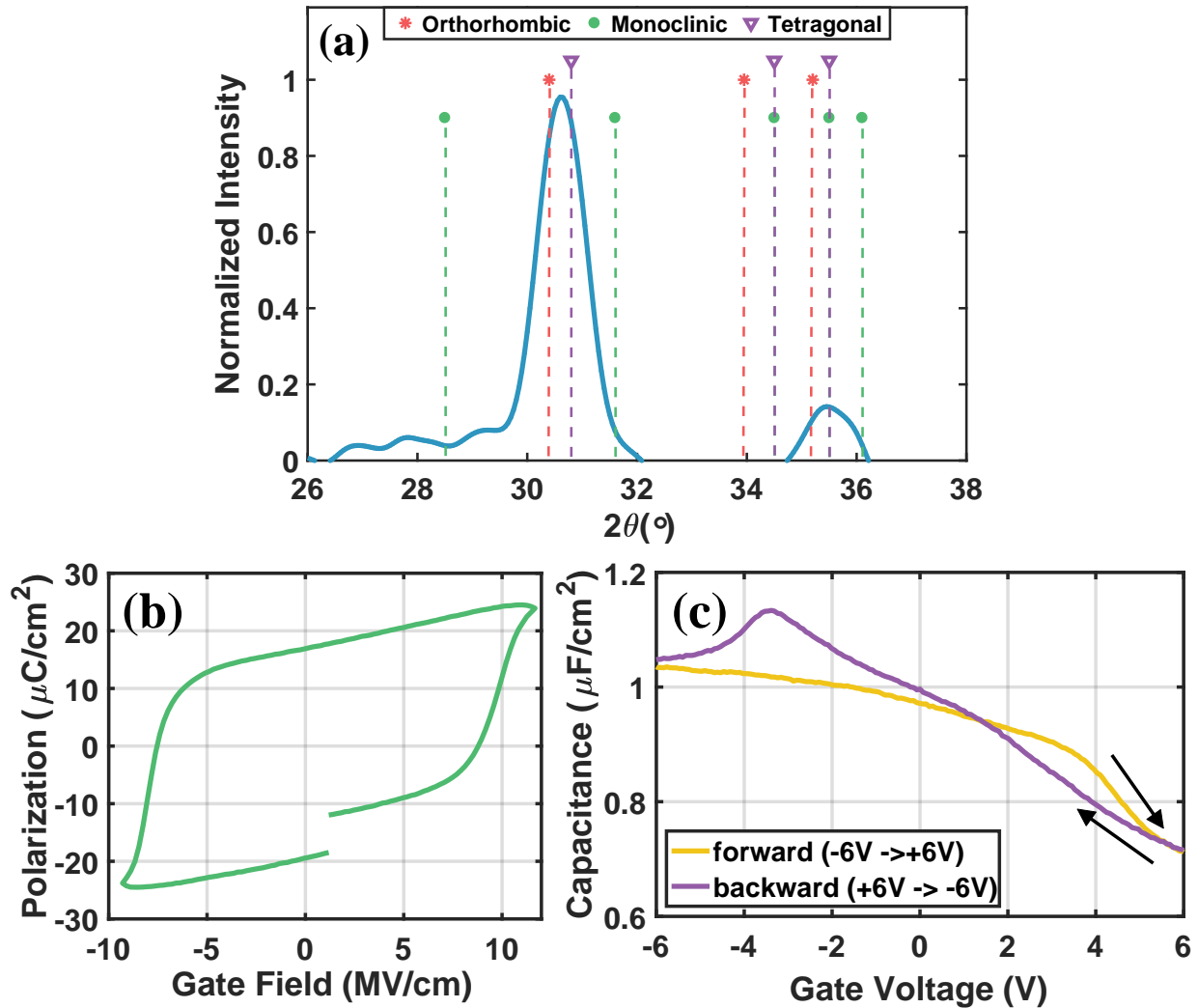
Figure 6.1: Proof of ferroelectricity in annealed HZO films. (a) A GIXRD diffractogram, taken with an incident angle of $0.35°$, suggests orthorhombic/tetragonal crystalline phases in a 100 cycle ALD $Hf_{0.5}Zr_{0.5}O_2$. (b) PE (polarization vs. electric field) measurement on a 100 cycle ALD $Hf_{0.5}Zr_{0.5}O_2$ film. (c) CV measurement on a 100 cycle ALD $Hf_{0.5}Zr_{0.5}O_2$ film.

in Figures 6.1(b) and 6.1(c), respectively. The representative PE and CV data shown here are taken on device stacks nearly identical to the ones discussed in the experimental section, except a degenerately doped (10^{19} cm^{-3}) Si substrate is used as a highly conductive back electrode. The remanent polarization value is 14.73 μC/cm^2 over a sweep range of $\pm$ 10 MV/cm, as seen in Fig. 6.1(b).

After the 500 °C ferroelectric phase anneal, it is observed that stretch-out occurs in the
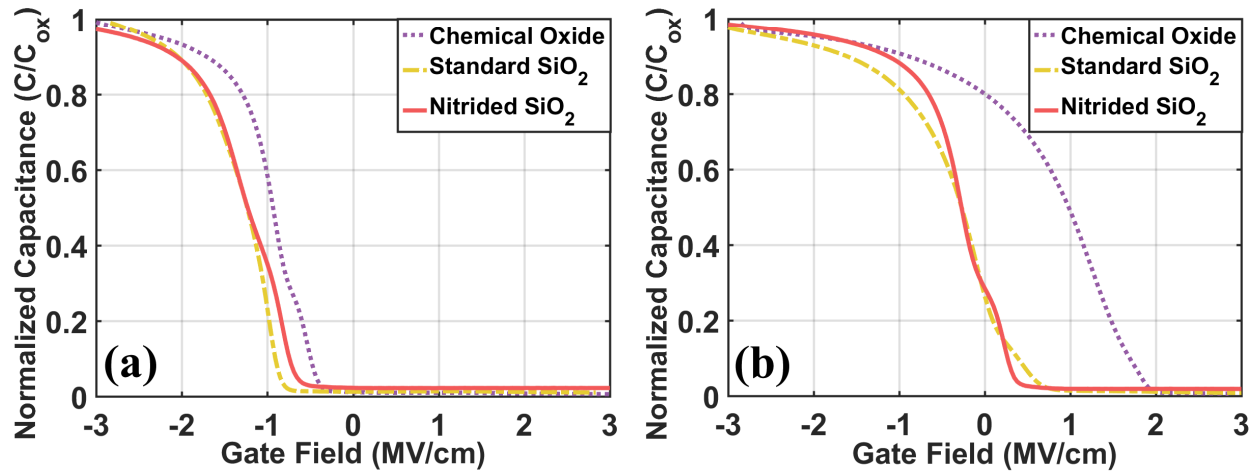
Figure 6.2: Normalized CV curves for samples with chemical oxide (dotted purple), standard SiO₂ (dash-dotted yellow), and nitrided SiO₂ (solid red). (a) Representative CV curves before the ferroelectric phase anneal. (b) Representative CV curves after annealing.

CV characteristics of the capacitors with chemical oxide or with standard SiO₂ as the interfacial layer, as seen in Figure 6.2. Furthermore, in comparing Figures 6.2(a) and 6.2(b), the stretch-out suggests that the midgap D_{it} in particular increases more dramatically for the sample with SiO₂ or chemical oxide as an interfacial layer. This phenomenon suggests that a strong degradation of the interface of the oxide/Si stack has occurred as a result of the anneal, particularly because the Si/SiO₂ system is known to exhibit an intrinsically low interface trap density [20, 109]. The capacitor with nitrided SiO₂ demonstrates the least stretch-out, suggesting that this device structure possesses the lowest overall D_{it} after annealing.

The CV curves in Figure 6.2 are plotted with capacitance values normalized to C_{ox} and applied voltage normalized to film thickness (represented as electric field) to account for differences in equivalent oxide thickness. For comparison, a representative CV prior to annealing is shown in the inset. The voltage sweep direction for each device is from -3V to $+3$V. The cause of the strong flatband voltage shift for the sample with chemical oxide in comparison to the samples with thermally grown interfacial oxides may be due to differing nucleation and growth mechanisms of the ALD Zr-doped HfO₂ film on surfaces terminated via different synthesis techniques.

It is therefore hypothesized that there is substantial diffusion of hafnium and zirconium atoms during the ferroelectric phase anneal to the underlying silicon substrate, disrupting the interfacial bonds or forming defective silicates. This observation is in light of numerous trap neutralization anneals developed for Si/SiO₂ at temperatures comparable to the ferroelectric phase anneal discussed here, whereby the interface trap density consis-

tently decreases [37, 109]. We suspect that the nitrided interface provides a strong barrier against Hf/Zr diffusion during high-temperature post-processing steps. This hypothesis is motivated by the well-known incorporation of nitrided oxides in highly scaled gate dielectrics in order to prevent boron diffusion to the channel silicon from the p+ polysilicon gate [29, 41, 51]. To establish this hypothesis, however, a detailed SIMS study will be necessary.

AC Conductance Method to Extract D_{it}

The Zr-doped HfO$_2$ capacitor with nitrided SiO$_2$ is thus chosen for further study to understand the interface trapping characteristics. The dynamic filling and emptying of trap states located at the interface between silicon and nitrided SiO$_2$ is characterized using the frequency-dependent AC conductance method [96]. As trap states often occupy particular energy levels within the bandgap, they can be characterized by changing the surface potential of the semiconductor through an applied gate voltage. When the silicon substrate is depleted, fast traps located at the interface modulate the measured capacitance and conductance values as a function of frequency [5]. Thus, it is possible to obtain information about the trap density and time response from CV and GV (conductance-voltage) measurements on MOSCAPs.
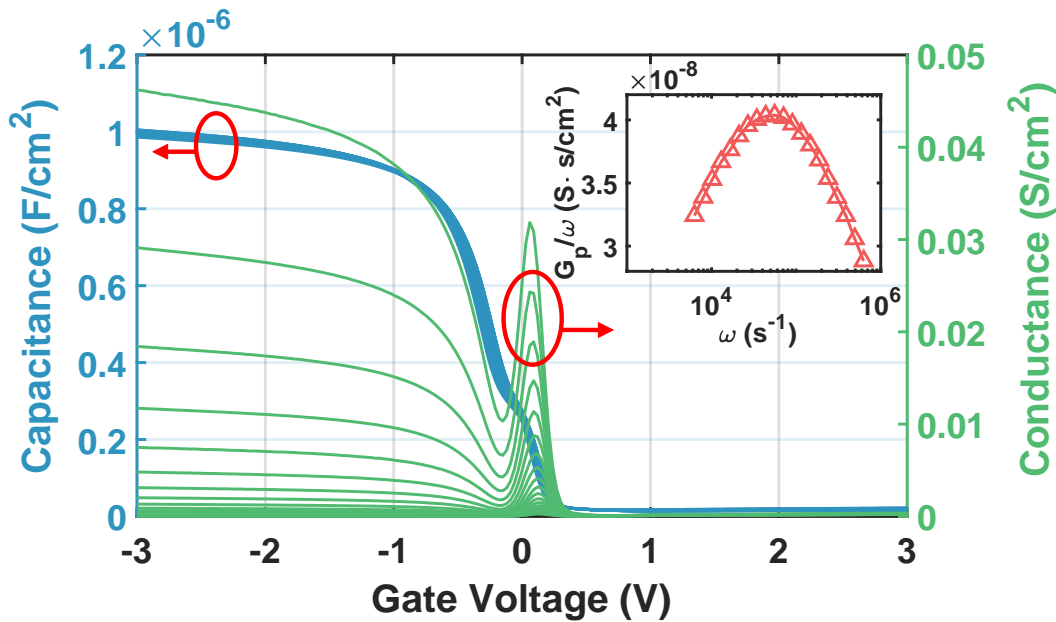
Figure 6.3: Direct measurement of D_{it} on a MOSCAP with nitrided SiO$_2$. CV/GV curves, from 1 – 100 kHz, indicate minimal frequency dispersion. Inset shows a normalized conductance vs. frequency curve for D_{it} extraction.

Peaks in the GV curves, shown in Figure 6.3, indicate a response due to interface trapping. The peak midgap conductance values, seen near an applied voltage of 0V, increases as the measurement frequency increases from $1 - 100$ kHz. Moreover, by reinterpreting the measured capacitance and conductance data as a function of frequency, the midgap interface trap density can be extracted [5, 96]. A representative normalized conductance curve is plotted versus frequency (G_P/ω vs. ω) in the inset of Figure 6.3.

The normalized conductance here is calculated as:

$$\frac{G_P(\omega)}{\omega} = \frac{\omega C_{ox} G_m}{G_m^2 + \omega^2 (C_{ox} - C_m)^2} \tag{6.1}$$

where C_m and G_m are the measured capacitance and conductance data, respectively.
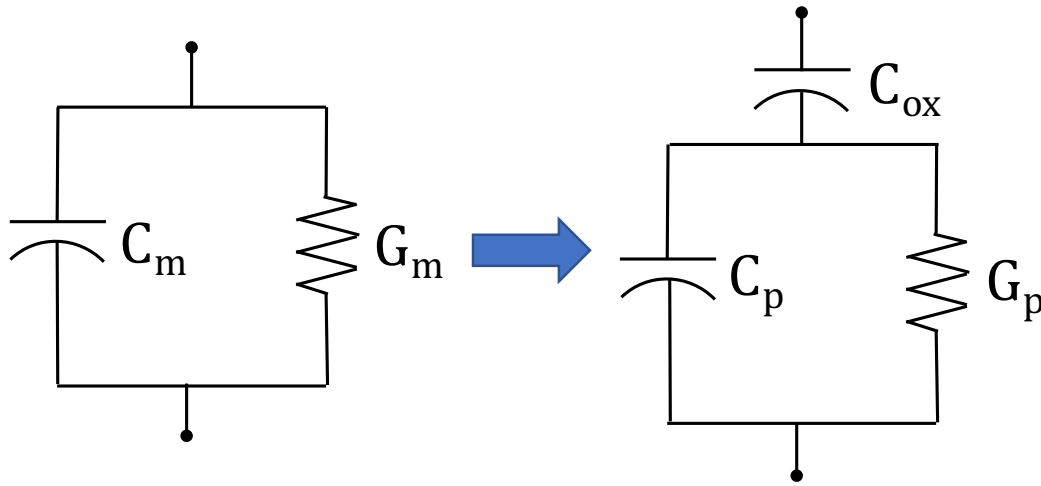
Figure 6.4: Setting the CV meter to C_p/G_p measurement mode assumes the equivalent device circuit model to the parallel capacitor/resistor model as shown on the left. Information about the capacitance associated with trapping responses at different surface potentials can be extracted using the right model.

Utilizing the equivalent circuit model for a MOSCAP as discussed in [96] and illustrated in Figure 6.4 above, D_{it} can be extracted at the peak G_P/ω value using the following relation:

$$D_{it} = \frac{2.5 \cdot G_P(\omega_{\max})/\omega}{q} \tag{6.2}$$

where q is electronic charge, and $\omega_{\max}$ is the frequency at which the maximum conductance value occurs. Using this relationship, we extract the maximum D_{it} in depletion for the capacitor with nitrided SiO$_2$ after the phase anneal as roughly 6×10^{11} cm^{-2} eV^{-1}. In comparison, the maximum midgap D_{it} of the chemical oxide or SiO$_2$ interlayer is 2.9×10^{12} cm^{-2} eV^{-1} and 2.3×10^{12} cm^{-2} eV^{-1}, respectively.

A CV Model to Describe the Influence of Interface States

Finally, a CV model to capture the influence of interface trapping is presented. The "ideal" flatband voltage of the MOSCAP, sans the effect of interface traps, is extracted from a single high frequency CV (one taken at 1 MHz). Then, a flatband capacitance is calculated based on the Debye length, and a flatband voltage (V_{FB}) is also assigned. In the device analyzed, V_{FB} is determined to be -0.201V. The oxide capacitance, C_{ox}, is defined as the combined capacitance of the interfacial oxide and the Zr-doped HfO$_2$ film, and is determined experimentally as the measured capacitance in strong accumulation.

A model CV based on standard MOSCAP theory is then constructed with and without the effect of interface traps. To the first order, the presence of D_{it} manifests itself as an additional C_{it} in the circuit model proposed by [96]. One can calculate an approximate distribution of D_{it} across a portion of the silicon bandgap, roughly from the valence band to midgap:

$$C_p = \left[\frac{1}{C_{meas}} - \frac{1}{C_{ox}} \right]^{-1} \tag{6.3}$$

$$C_{it} = C_p - C_s \tag{6.4}$$

where C_p captures the capacitance contribution from silicon depletion and interface trapping, and C_{it} subtracts off C_s, the expected depletion capacitance. C_s is determined, from standard CV calculations, as the differential change in the semiconductor charge as the surface potential is stepped.

Figure 6.5(a) contains fitted CV's, both with and without an assumed D_{it} distribution, based on experimental data taken at 1 kHz and 1 MHz. Figure 6.5(b) shows D_{it} versus the occupational energy of the traps in the silicon bandgap for the D_{it} data extracted using the AC conductance method and the fitted D_{it} used in the 1 kHz CV model. The depletion-regime D_{it} from the fitted distribution agrees well with the D_{it} extraction.

Lastly, we note a small peak in the D_{it} distribution around 200 mV from the valence band edge. This manifests in the 1 kHz CV in Figure 6.5(a) as a characteristic "shoulder" in depletion. The cause of this peak are traps which have a specific origin and location relative to the interface, which could be due to P_b centers [106, 128], but also could result from residual Hf/Zr atoms reaching the silicon interface during the phase anneal.

In conclusion, the nature of the interface states of the ferroelectric HfO$_2$/Si material system has been investigated in detail. Various interfacial oxides, crucial for obtaining a good interface between the Zr-doped HfO$_2$ and silicon, were analyzed for integrity after the phase anneal. It is shown, using the AC conductance method, that a nitrided SiO$_2$ interfacial layer provides the lowest midgap trap density and furthermore can withstand the high-temperature phase anneal by acting as a barrier against dopant diffusion. Finally, a CV model and experimental CV data were used to calculate the D_{it} distribution, which has a narrow energy peak close to the valence band. Subsequent studies will focus on optimizing the nitrided SiO$_2$ for an improved trap distribution.
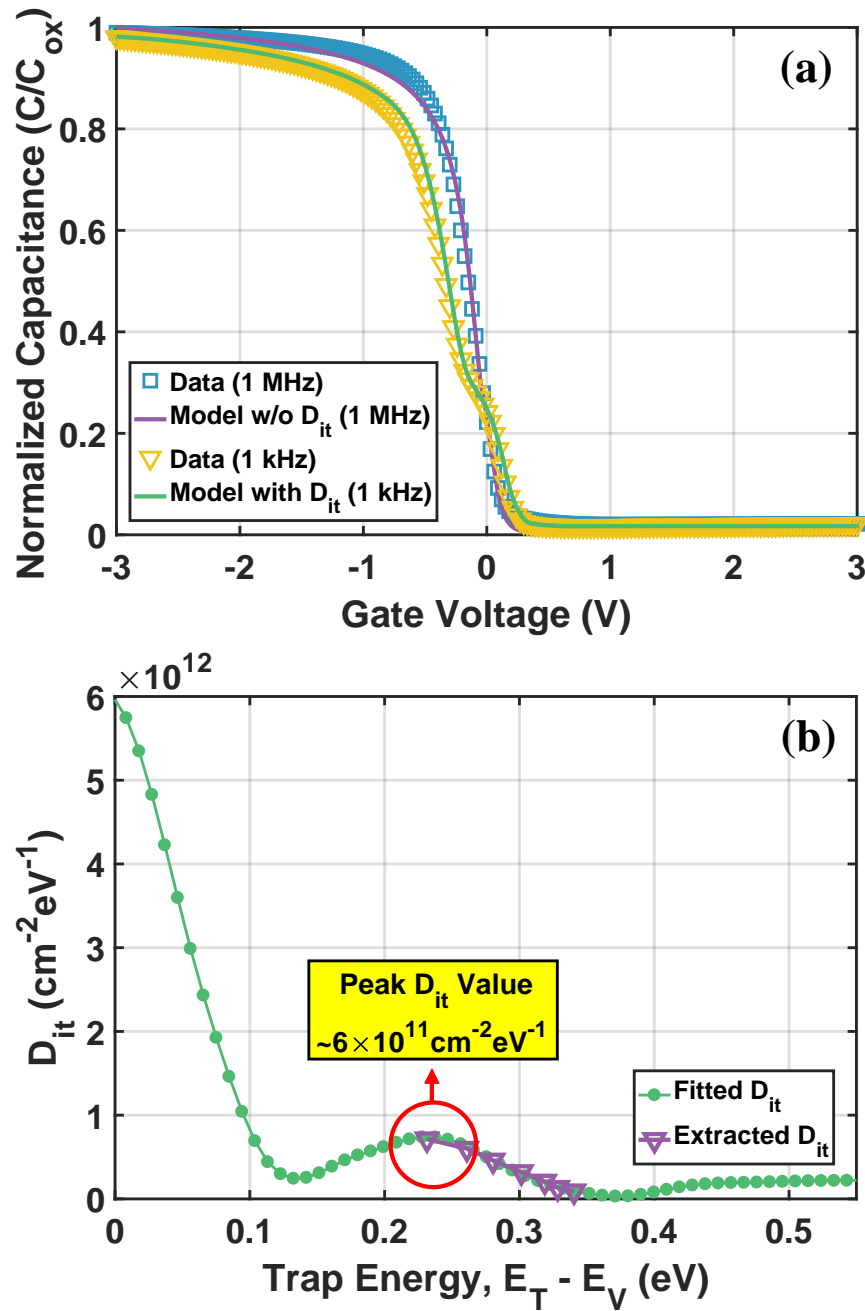
Figure 6.5: A model CV and D_{it} distribution can be fitted from the experimental V_{FB}, C_{ox}, and N_a. (a) Calculated and measured CV's at 1 kHz and 1 MHz. (b) Comparison of fitted D_{it} for CV model and measured D_{it} with AC conductance.

6.2 Record High Endurance Enabled by Direct Nitridation of Interfacial Layer

The discussion in the preceding section centers around the investigation of a nitrided SiO$_2$ interfacial layer from an interface state improvement standpoint. This solution was proposed originally in an effort to reduce trapping at the channel/oxide interface from the onset of device operation, as it was observed that the high-temperature ferroelectric phase anneal caused substantial Hf/Zr atom diffusion into the underlying substrate. Undoubtedly, as discussed in Chapter 5, the biggest reliability challenges arise from hot carriers releasing energy into the silicon lattice and continuing to disrupt interfacial bonds over the device's lifetime. Therefore, at this stage, device engineering efforts to reduce the kinetic energy of injected charges into the silicon substrate to improve overall FeFET endurance and reliability are probably most imperative.

In the devices reported in this section, a high-κ interfacial layer is combined together with a thin ferroelectric Zr-doped HfO$_2$ film (~ 4.5 nm). This choice is motivated by the previous observation by many authors (such as [22], and as discussed in Chapter 5) that the endurance cycling is strongly limited by breakdown of the interfacial layer. In fact, in a metal-ferroelectric-metal capacitor configuration, cycling endurance metrics exceeding 10^{10} are routinely observed [101]. In a recent report, Sharma et. al. circumvented the IL breakdown problem by fabricating a bottom-gate, channel last transistor, where an oxide semiconductor channel was grown directly on the FE material, thus achieving an endurance exceeding 10^{12} cycles [119]. Similarly, Kim et. al. recently reported on the fabrication of vertical 3D NAND ferroelectric thin film transistors utilizing indium zinc oxide as the semiconductor channel, showing a cycling endurance of up to 10^8 [63]. Nonetheless, when crystalline silicon is used as the channel material, as is required for high performance memory, formation of an interfacial layer is inevitable, and therefore endurance still remains a critical challenge to be addressed.

In the context of interfacial layer breakdown, it is known that 'time-to-breakdown' has an exponential relationship to the applied electric field in the interfacial layer [117]. In other words, a mild decrease in the electric field could still lead to a substantial increase in the 'time-to-breakdown', and therefore could slow the generation of traps that eventually counteract the ferroelectric hysteresis. For the same charge density, a high-κ interfacial layer reduces the electric field by the ratio of its permittivity to that of SiO$_2$. Our choice of high-κ IL is thermal nitridation of chemically formed silicon oxide. This technique provides a simple way to achieve an interfacial layer with a permittivity ~ 8. Thermally grown silicon nitride also has a comparable breakdown field to SiO$_2$ [57]. A thin 4.5 nm of ferroelectric Zr-doped HfO$_2$ is chosen to suppress the effects of bulk charge trapping, as discussed in Chapter 5, and also demonstrates the thickness scalability of ferroelectric hafnium oxides. It is shown that this combination substantially improves the device performance. In a DC $I_D V_G$ sweep, almost a 1V memory window can be achieved with just ± 2.5V. More importantly, with bipolar stress pulsing at ± 3V, 250 ns, the device endurance

exceeds 10^{10} cycles on silicon.

Experimental Procedure and EOT Comparison

The structure of the FeFET device characterized in this work is shown in Figure 6.6(a) and TEMs to compare its gate stack incorporating a nitrided IL against that of a baseline device with an SiO$_2$ IL (characterized in Chapter 5) are shown in Figure 6.6(c) and Figure 6.6(b). The process flow to realize the FeFET is described in Appendix A, with the interfacial layer formation step involving thermal nitridation of the SOI substrate at 850 °C in NH$_3$ ambient rather than a chemical growth of SiO$_2$. The device layer is p-type doped with boron to approximately 4.5×10^{14}/cm^3. The ferroelectric oxide thickness of both the control SiO$_2$ FeFET and the FeFET with a nitrided IL are the same (roughly 4.5 nm after 45 cycles of deposition). The IL thicknesses of the SiO$_2$ IL and nitrided IL are $\sim 8 \overset{\circ}{A}$ and ~ 1.5 nm, respectively.
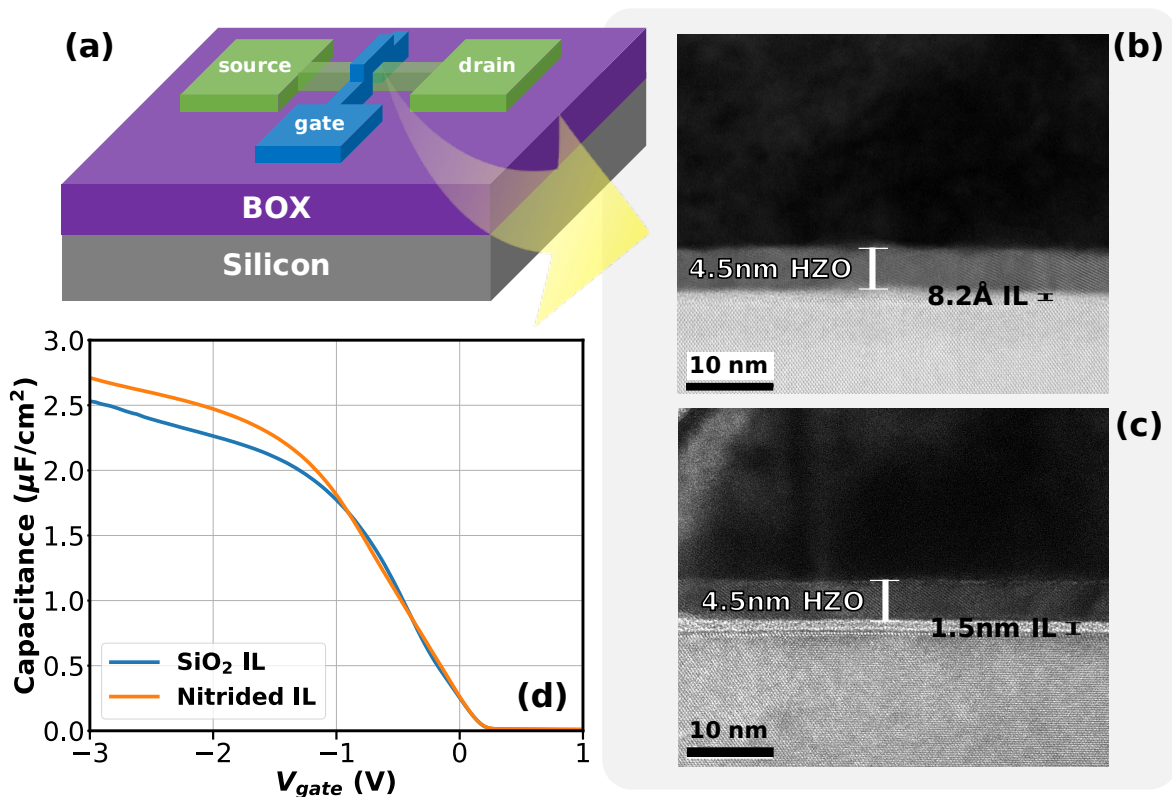
Figure 6.6: (a) Schematic of a SOI, gate-first FeFET device. (b) TEM of 4.5 nm Zr-doped HfO$_2$ gate stack with 8.2$\overset{\circ}{A}$ SiO$_2$ IL. (c) TEM of 4.5 nm Zr-doped HfO$_2$ gate stack with 1.5 nm SiN$_x$IL. (d) CV comparison of MOS capacitors with gate stacks identical to SiO$_2$ and SiN$_x$ IL FeFETs, taken at 100 kHz.

Figure 6.6(d) compares the CV's of MOS capacitors with identical gate stack designs as the baseline FeFET (with SiO$_2$ interfacial layer) and the FeFET with a nitrided interfacial layer. Though the physical thickness of the gate stack of the latter is larger, its capacitance is also larger. Using Synopsys TCAD, the estimate net EOT of the nitrided sample is determined to be roughly 1 $\overset{\circ}{A}$ smaller than the baseline sample, based upon the accumulation capacitance values. Therefore, one can make an estimation for the effective κ of the interfacial layer as follows:

$$\kappa_{\text{NIL}} = \kappa_{\text{SiO}_2} \times \frac{t_{\text{NIL}}}{t_{\text{baseline}} - \delta\text{EOT}_{\text{net}}} = 3.9 \times \frac{15}{7.5} = 7.8 \tag{6.5}$$

where κ_{NIL} and κ_{SiO_2} indicate the κ values of the nitrided interfacial layer and SiO$_2$ interfacial layer respectively; t_{NIL} and t_{baseline} indicate the physical thicknesses of the nitrided interfacial layer and SiO$_2$ interfacial layer respectively; and $\delta\text{EOT}_{\text{net}}$ is the simulated EOT difference between the two interfacial layers. This calculation indicates that the interfacial layer is nearly entirely Si$_3$N$_4$ by volume. Therefore, it is reasonable to expect a reduction in the electric field in the interfacial layer by two times, which will ultimately result in a substantial increase in the time to breakdown.

Endurance Results and Analysis

The programmability of the optimized FeFET device with the nitrided interfacial layer is discussed in detail in Section 4.3 of Chapter 4. The biggest advantage that the redesigned interfacial oxide offers, in comparison to the baseline SiO$_2$ interfacial layer devices, is the reduction in total gate voltage and associated gate pulse duration in order to achieve an even more robust memory window and separation between the two binary memory states. Based on the results of the programmability test, a cycling condition utilizing a pulse width of 250 ns and a gate voltage of $V_G = \pm 3$V is chosen to obtain reasonably fast and symmetric operation.

In many studies, the endurance is quantified by measuring $\pm V_T$ after a certain number of bipolar stress pulses. The $\pm V_T$ determination requires one to perform sweeps over a small voltage range, which typically takes ~ 1 second to complete. On the other hand, the importance of fast reading has recently been discussed (e.g. [119]). It is known from charge pumping experiments (as discussed in [102] as well as in Chapter 5 of this dissertation) that beyond several μs, charge trapping/de-trapping starts to manifest. These effects could in principle be quite complex, and could arise from the interplay between traps with varying time constants. Therefore, while slow sweeps to determine $\pm V_T$ could provide important insights into trap assisted phenomena, they are also expected to *artificially* affect the actual currents that will be observed in a real application where the device is read quickly.

Due to these considerations, a fast reading protocol is adopted in order to determine the state written to the device. The complete endurance testing protocol is detailed in Figure 6.7(a). During the stressing phase of the endurance test, bipolar voltage pulses of ± 3V,
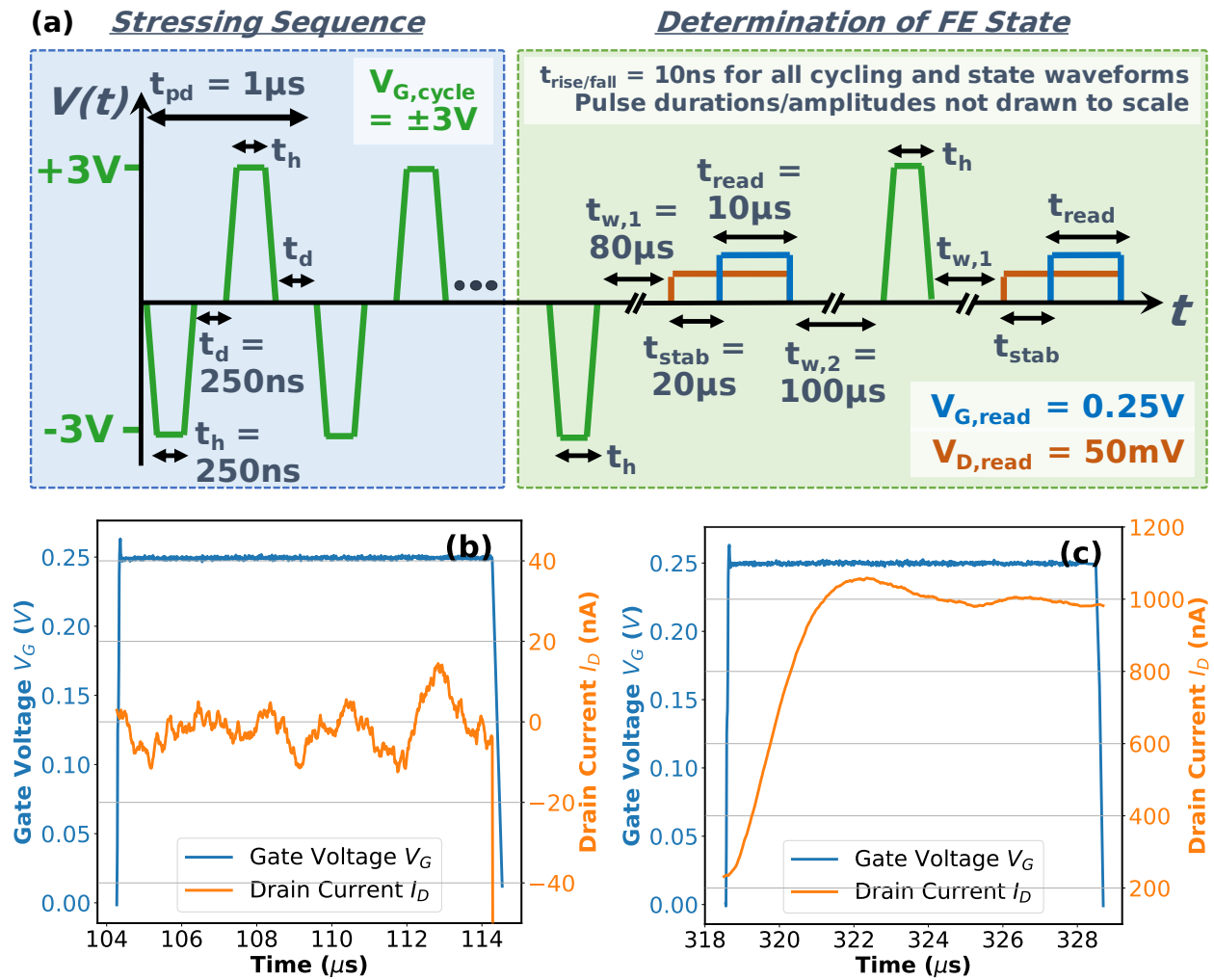
Figure 6.7: (a) Endurance stressing sequence, shown in the blue panel, and subsequent ferroelectric state determination waveforms, shown in the green panel, used to characterize the FeFETs in this work. (b) Transient current readout waveform corresponding to the high V_T state (and low readout current). (c) Transient current readout waveform corresponding to the low V_T state (and high readout current).

250 ns are applied at the gate of the FeFET, with a 250 ns delay between sequential pulses to achieve a stressing period of 1 μs total in duration. Periodically, a state determination test is conducted to evaluate the margin between the PGM and ERS states. For this, a 10 μs read pulse is applied at the gate of the FeFET (after either the ERS or PGM pulse), after ramping and stabilizing the drain voltage to 50 mV (see Figure 6.7(a), right panel, with a green background). The averaged current value during this 10 μs reading period is then

determined to be the read current. Figures 6.7(b) and (c) show the readout current during the 10 μs reading period for the high V_T (PGM) and low V_T (ERS) states, respectively. In both cases, the current saturates well within the READ duration of 10 μs.
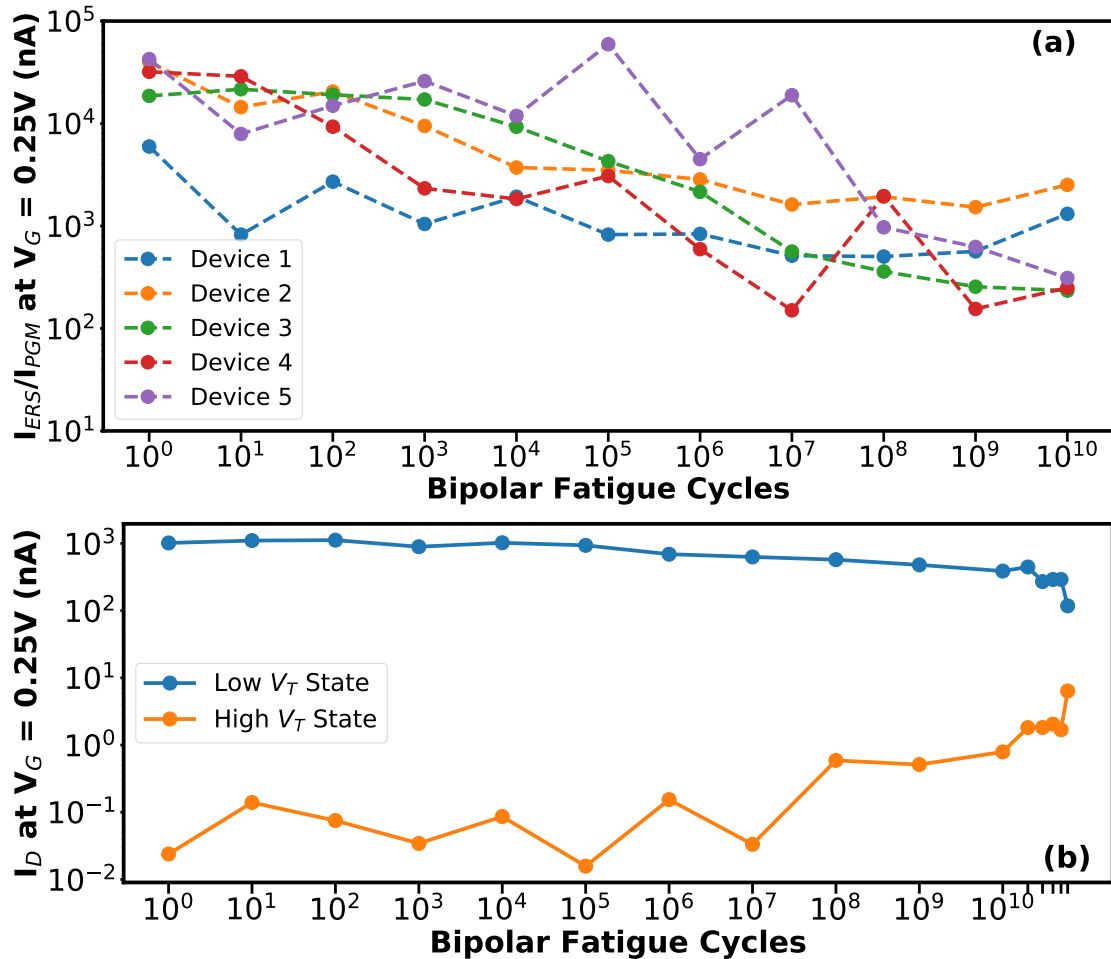
Figure 6.8: (a) Endurance characteristics of multiple FeFET devices cycled from 1 to 10^{10} fatigue cycles. Results are reported as the ratio of the low V_T I_D readout to the high V_T I_D readout at every decade of cycling. (b) A device cycled until $\sim$1 order of magnitude of current separation remains after 6×10^{10} cycles.

Figure 6.8(a) shows the results of endurance testing across different devices, proving that the devices with a nitrided IL can be reliably cycled to 10^{10}. The data is plotted as I_{ERS}/I_{PGM} vs. fatigue cycles for ease of comparison. The exact current levels are shown for a measured device in Figure 6.8(b).

Firstly, it is noted that the high and low current levels are very similar to those measured from a DC sweep. This indicates that the device is switched properly with the
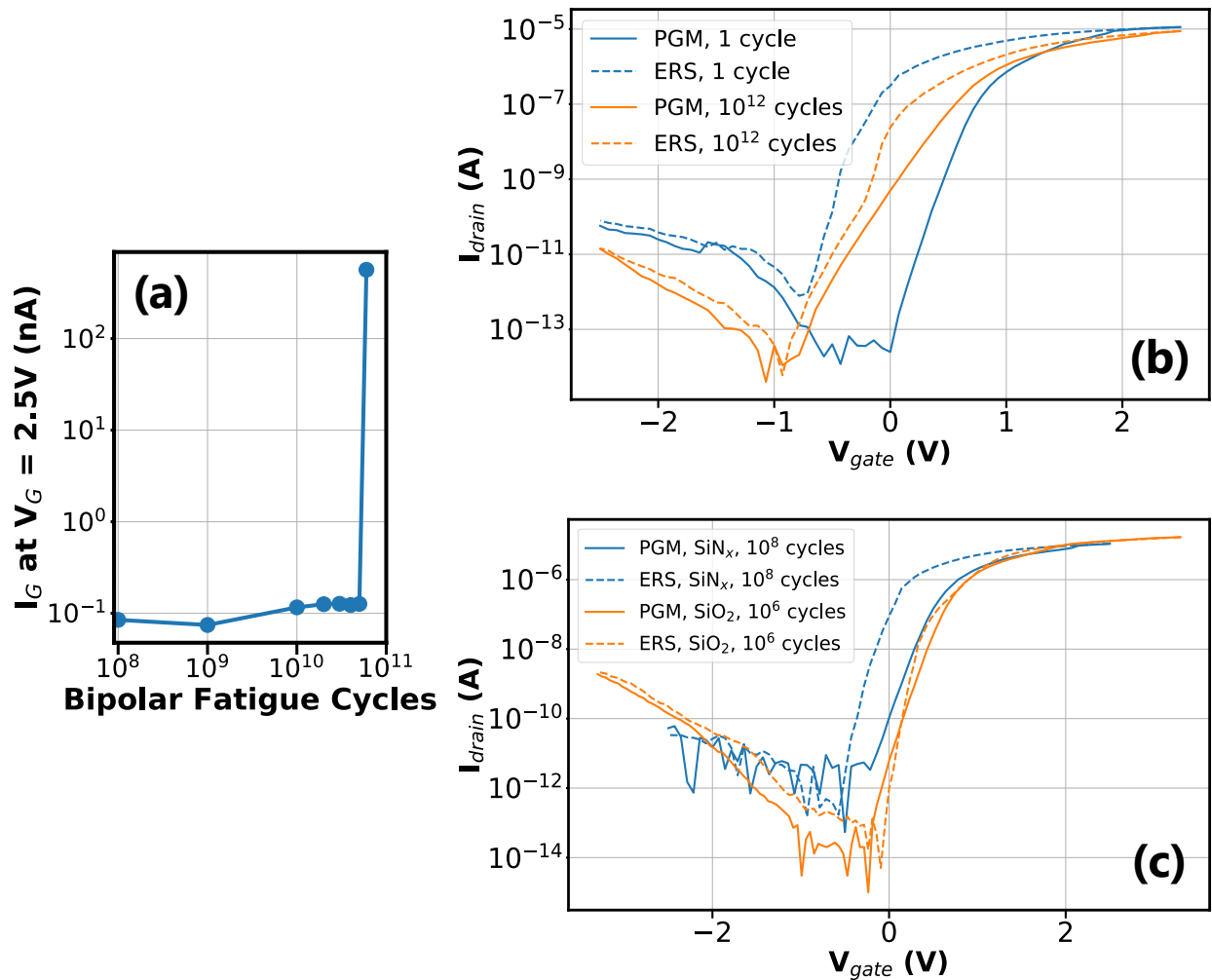
Figure 6.9: (a) I_G at $V_G = 2.5$V as a function of cycling for the same device in Figure 6.8(b), showing a strong correlation between oxide wearout and loss of memory window. (b) $I_D V_G$'s of an exemplary device cycled to 10^{12} cycles, showing some remaining ferroelectric hysteresis at the end of the endurance test. (c) Residual ferroelectric hysteresis after 10^6 cycles on the baseline FeFET vs. after 10^8 cycles in the optimized SiN$_x$ FeFET.

PGM/ERS pulses. Interestingly, the device does not show any rapid degradation after $10^4 - 10^6$ cycles, as reported in most studies. Rather, the high current level shows a slow degradation. The envelope of the low current similarly shows a slow degradation (increase); yet the separation of the current levels retains a margin of 10^3 until 6×10^{10} cycles. Beyond that point, a sudden breakdown is observed. Notably, this sudden breakdown is correlated with the gate leakage through the device shooting up several orders of magnitude, as shown in Figure 6.9(a), indicating that the gate oxide itself breaks down close to 10^{11}

cycles.

On the same wafer, there are devices that do not experience such a catastrophic break-down of the ferroelectric oxide layer. For example, Figure 6.9(b) shows DC $I_D V_G$ sweep from an exemplary device cycled to 10^{12}. The anti-clockwise ferroelectric hysteresis is still clearly visible. This result contrasts with previous studies, where it has been shown that total reversal of handedness of hysteresis happens after just 10^4-10^5 cycles. This also shows that the ferroelectric itself can remain quite robust even after one trillion cycles.

Lastly, Figure 6.8(c) shows a comparison of the baseline devices reported in Chapters 5 with SiO$_2$ interfacial layer, to the optimized FeFET with SiN$_x$ interfacial layer. By 10^6 cycles, it is clear that charge trapping within the baseline device has completely counteracted the ferroelectric hysteresis and reversed the direction of hysteresis as a result. However, by 10^8 fatigue cycles for the optimized SiN$_x$ FeFET, the ferroelectric hysteresis remains strong and a memory window of nearly ~ 0.5V can be seen in the DC $I_D V_G$ sweep. This result points to a substantially improved robustness of the interfacial layer against field cycling, which is ultimately what extends the endurance lifetime of the FeFET.

In conclusion, this section has demonstrated the successful characterization of a FeFET memory device with an engineered high-κ IL that shows larger than 10^{10} endurance cycles at a relatively small PGM/ERS voltage of $V_G = \pm 3$V and pulse width of 250 ns. Endurance measured on multiple devices shows repeatably robust behavior to over 10^{10} cycles. Total gate oxide breakdown has been identified as the main limiting factor, as opposed to defect (interface and bulk) induced clockwise hysteresis reported by many previous studies. Further optimization could reliably increase the endurance over 10^{12} cycles, as evidenced from Figure 6.9(b). One potential drawback to consider of using such a high-κ IL could be a reduction in mobility. However, somewhat reduced mobility could still be tolerable as these devices are not expected to compete with the gate-delay of logic devices.

Chapter 7

Ferroelectric Content Addressable Memory for In-Memory Computing

7.1   A New Era of Computing and Memory Challenges

The core components of any computing technology involve both memory and logic units, which are necessary for the storage of information as well as processing of the aforementioned data, respectively. Unfortunately, as CMOS technology continues to mature and advance, the gap in performance between memory and logic has not improved accordingly [17, 68]. General-purpose computing, which forms the basis of virtually all modern-day computing, revolves around *von Neumann* architecture. This architecture, first described in a paper by John von Neumann in 1945, assumes that each computation involves looking up data from memory, shuttling said data to the central processing unit (CPU) for computation, and then sending it back to memory to be re-stored [8, 131].

This data movement to and from memory and CPU exactly illustrates the so-called *"von Neumann bottleneck"*, which refers to the prohibitive restriction in bandwidth that is witnessed when there is heavy transferring of data to and from memory and CPU, especially for tasks which are computationally heavy. Within the framework of von Neumann architectures, various caching schemes have been proposed and developed in an effort to improve the latency issues by bringing computation closer to memory. However, even so, data movement to and from memory remains outstanding as the most energy-intensive component of computation – not the actual compute operations in question [131]. For example, if one considers the 45 nm CMOS technology node, the energy associated with multiplying two numbers is orders of magnitude smaller than the energy associated with fetching those same two numbers from memory [118].

Perhaps most striking is the fact that the latency issues associated with the von Neumann architecture has become a key performance issue for modern-day workloads related to artificial intelligence and machine learning applications [93, 118]. Clearly, novel solutions to tackle the separation of computation and memory units by physically bringing

them closer together on chip, are becoming more and more attractive. Some researchers
have worked on the development of specialized ASICs (application-specific integrated cir-
cuits) to overcome the von Neumann bottleneck [10, 79]. In recent years, 3D monolithic
approaches to achieve fine-grained connectedness between memory and logic units have
even been proposed [121].

Perhaps at the ultimate limit of bringing memory and logic closer together is the ap-
proach of *in-memory computing*, where a single unit can both perform a computational task
and store the data.
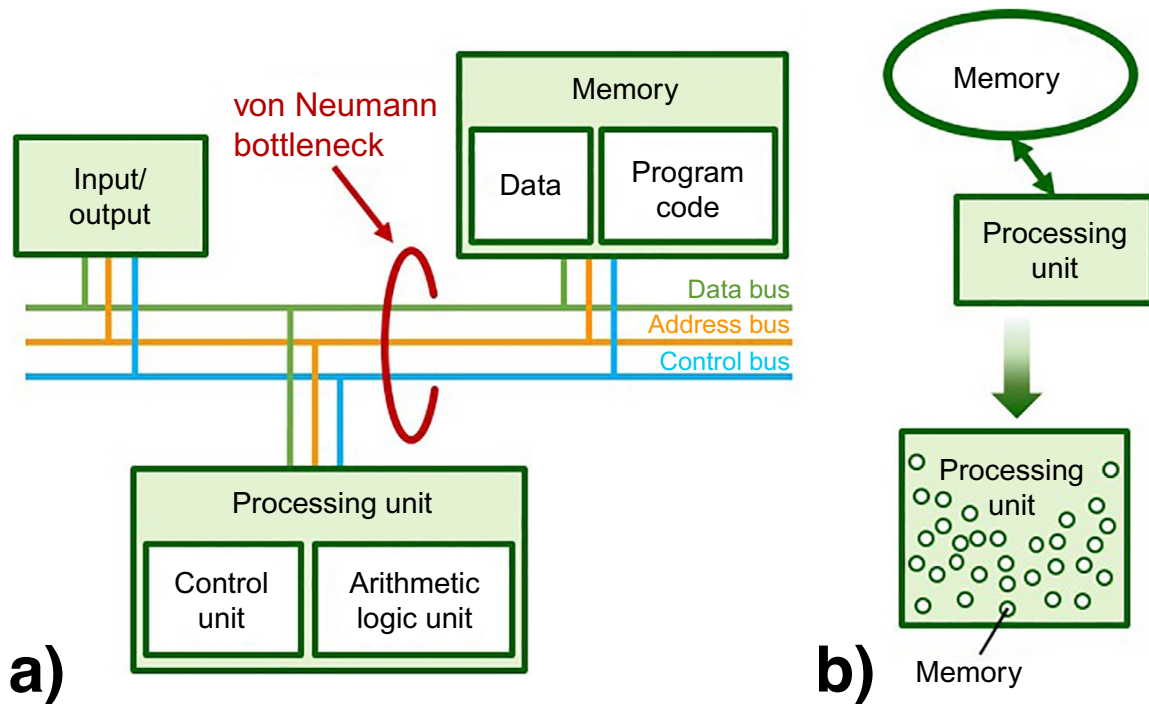
Figure 7.1: A comparison of traditional computer architecture vs. logic-in-memory. (a) A
depiction of von Neumann architecture. Data and address buses shuttle data from mem-
ory to CPU and back for computation. (b) In-memory computing schemes bring memory
into the logic processing unit itself. Figure adapted from [17].

A schematic comparison of traditional von Neumann architecture against in-memory
computing is shown in Figure 7.1. The hope, therefore, is not only to alleviate the costs
associated with increased latency and energy consumption in data movement, but also
to reduce the time complexity to execute certain computational tasks. A dense array of
memory devices with the ability to also perform computations is expected to dramatically
improve parallelism [118]. It is therefore expected that there will be a shift towards spe-
cialized, non-von Neumann, data-centric approaches for machine learning/deep learning

applications – but complete abandonment of the general purpose computer is unlikely to happen in the foreseeable future [11].

7.2 Content Addressable Memories: Overview and Applications
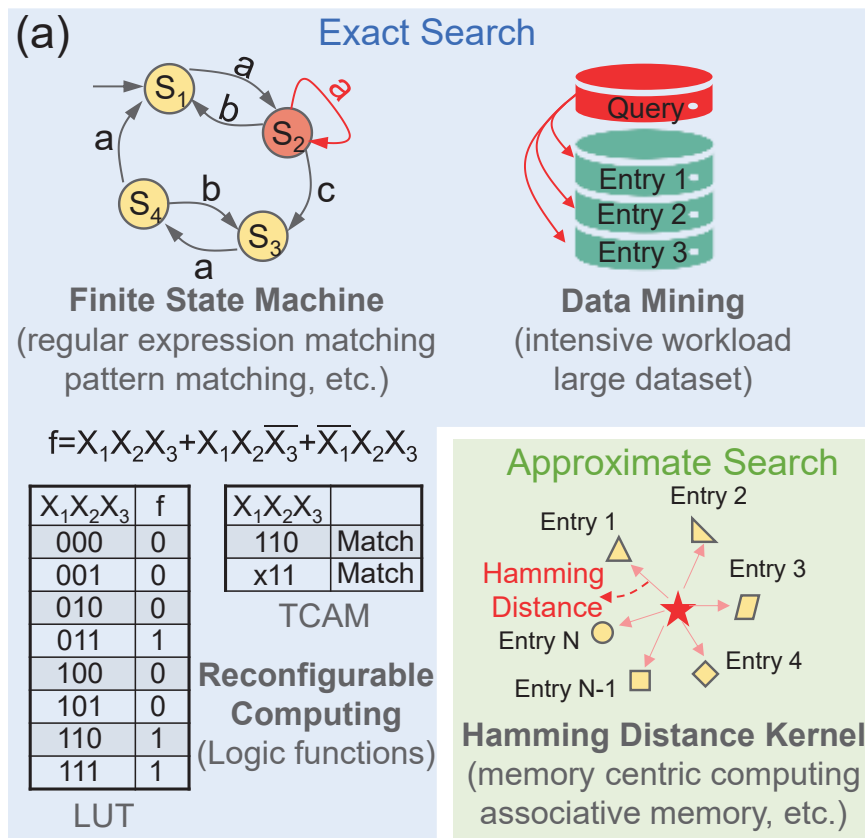
Figure 7.2: Applications of ternary content addressable memories. Highlighted in blue are applications which involve exact search queries and are more traditional applications of TCAMs. Highlighted in green is an approximate search application for memory-centric computing. Figure from [70].

Ternary content addressable memories (TCAMs) are specialized memory solutions that allow for fast parallel lookup/search of data, but typically at the expense of power consumption. Memory cells are searched by *content*, rather than by address, to determine the location and existence of a partial or exact match. For this reason, a content addressable memory can conceptually be thought of as the inverse of random access memory. Tradi-

tionally used for high-speed data processing for IP filters, network routers, etc., TCAMs are recently being explored for neural network acceleration and pattern-matching in data-intensive applications [55]. Figure 7.2 lists a few common applications of TCAMs, ranging from exact search matching to approximate search applications. Exact match applications include pattern/regular expression matching in finite state machines, data mining, and reconfigurable computing. Approximate matching applications typically involve approximating the Hamming distance between input query and stored data.

Naturally, research interest in TCAMs has gravitated towards the development of emerging memories to improve upon existing CAMs, which typically have low area efficiencies and utilize many transistors [16, 26, 56, 139]. Traditional CMOS implementations of a TCAM cell typically involves 2 SRAM cells, each of which requires 8 transistors, for a total of 16. From this point of view, utilizing emerging memories to implement TCAMs is beneficial due to their nonvolatility and potential to realize very dense solutions.

Of these solutions based on emerging memories, ferroelectric FET solutions stand out as they provide nonvolatility, reconfigurability, and the ability to achieve high density due to their scalability and ease of integration with current CMOS processes [6, 138]. Furthermore, they have been shown to successfully implement logic functions using a small number of transistors [17]. When compared against magnetic tunnel junction (MTJ), resistive RAM (ReRAM), and even SRAM-based TCAMs, FeFET-based TCAMs show superior array-level performance characteristics. A 4T-2FeFET cell design, for example, occupies 58% of the equivalent 16T SRAM-based TCAM design, and a 64-row array boasts an energy-delay product which is 7.5X and 149X better than similar MTJ- and ReRAM-based designs, respectively [141].

7.3   The Ferroelectric Content Addressable Memory (FeCAM)

In utilizing FeFET devices, the intrinsic CAM cell itself can be made to be very compact, reducing the total transistor down from 16 to just 2, as pictured in Figure 7.3.

Within a single FeCAM cell, a bit and its complement (STR and $\overline{\text{STR}}$, respectively, as labeled in Figure 7.3(a)) are stored by applying voltage pulses of opposite polarity to the gates of each respective transistor, which sets each FeFET's threshold voltage to $+V_T$ or $-V_T$. Applying -5V corresponds to storing a "0", and $+5$V to a "1". During the search phase, the select lines for each FeCAM cell are probed with the corresponding data to be matched (with $V_D = 0$V representing a bit "0", and $V_D = 50$mV representing a bit "1"). In the case of a mismatch, neither transistor in the cell will conduct high current through the match line, as one FeFET will witness a $V_D = 0$V applied drain bias while the other will be programmed in the "0" or low-current state (and thus will not contribute substantial current to the total match line readout current). In the case of a match, either the STR or $\overline{\text{STR}}$ FeFET will conduct high current to the match line upon select line activation.

a)

b)
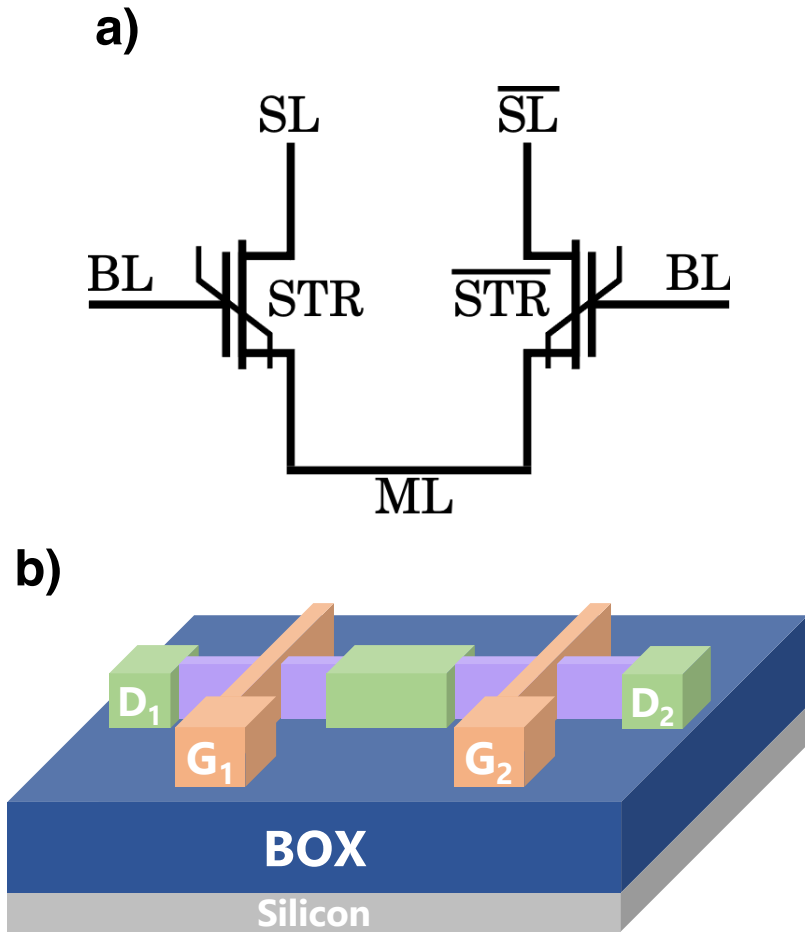
Figure 7.3: A FeCAM cell implemented with two FeFETs. (a) Circuit representation of a single CAM cell, which consists of two FeFETs with their sources shorted together. (b) Tilted 3D view of FeFET CAM cell schematic on thin SOI substrate.

SL (V_D)	STR (V_{PGM})	ML (I_D)
0V	-5V	10.9 nA
0V	+5V	3.21 pA
50mV	-5V	2.79 pA
50mV	+5V	13.1 nA

Table 7.1: The operation of the CAM cell can be described by the XNOR Boolean equation and associated truth table. Highlighted in yellow are the high-current or "match" states.

In this manner, the FeCAM cell implements the XNOR operation, as outlined in Table 7.1 and described in Equation 7.1. Table 7.1 also contains experimental data of match line current readout for the cases of a match or a mismatch. Note that the same cell can be utilized as a ternary FeCAM by writing a "1" or high-current state to both transistors to force a match when a "don't care" bit is required.

$$ML = XNOR(SL, STR) \tag{7.1}$$

Next, retention testing of the FeFET CAM cell is carried out at 85 °C, as shown in Figure 7.4. In this particular test, the match line current (I_{ML}) from a CAM cell is measured as function of time for all possible states (a miss, a match, or a forced match in the case of a "don't care" query). After 24 hours, the I_{on}/I_{off} ratio degrades roughly 30%, and there remains two orders of magnitude of current level separation after 10 years. Retention characteristics for the FeCAM cell are similarly as robust as the retention characteristics seen in a standalone FeFET, as discussed previously in Figure 4.9 of Chapter 4.
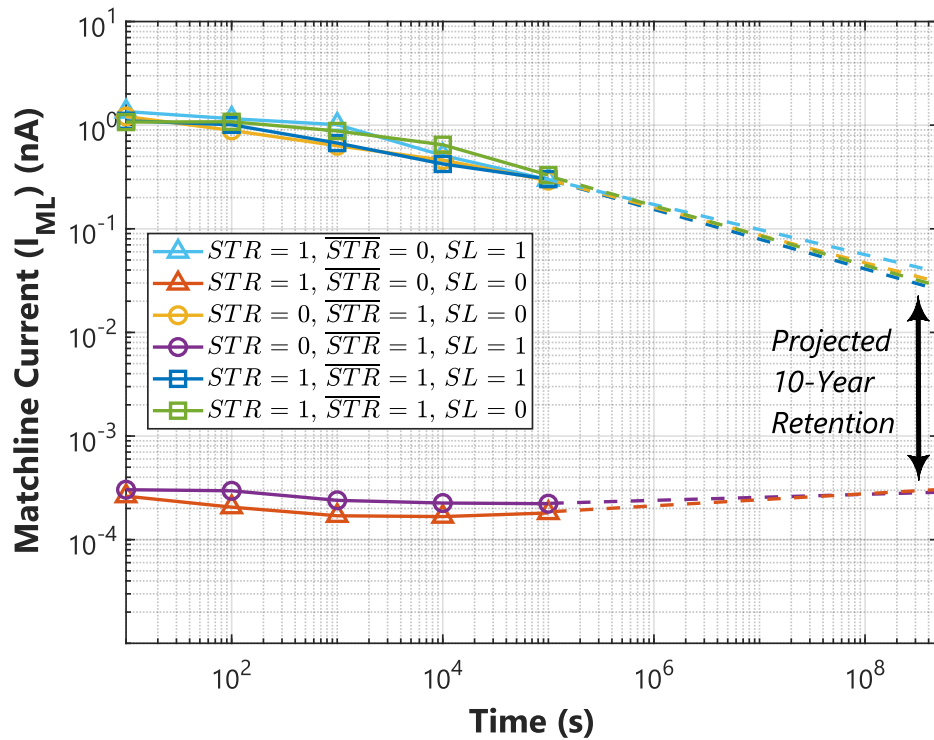
Figure 7.4: Retention testing at 85 °C for a CAM cell with device dimensions W/L = 1/1 μm. PGM/ERS conditions are -5.5V, 35 ns and $+5.5$V, 25 ns respectively.

SL	STR	$\overline{\text{STR}}$	ML
0	0	1	H
0	1	0	L
1	0	1	L
1	1	0	H
0	1	1	H
1	1	1	H

Table 7.2: Truth table corresponding to the 6 possible states of a ternary FeCAM cell: 2 match states, 2 miss states, and 2 "forced" match states. The states in which the match line current goes high (the 4 match states) are highlighted in yellow.

Table 7.2 delineates the expected match line current in the retention testing for each of the three possible states (4 forced/exact match states, and 2 miss states). H indicates that the current readout level should be high, whereas L indicates that the current readout level should be low. This truth table supports the empirical retention results reported in Figure 7.4.

7.4 FeCAM Circuits and Architecture

Experimental Demonstration of 2-Bit CAM Word

A proof-of-concept demonstration of a 2-bit FeCAM word is shown in this subsection, with an additional emphasis on a discussion of device variability with respect to projected array performance.

The truth table corresponding to the operation of the 2-bit CAM word (comprised of 4 total FeFETs) has been verified. However, although the truth table has been verified, there is noticeable variation in the current level readouts corresponding to the various states, which is a direct result of the fact that there is intrinsic device-to-device variation across the FeFET wafer. For most applications, so long as these current levels can be separated into distinct ranges, some device variability may not be an issue; however, on a much larger scale, detailed studies will need to be conducted in order to first quantify the extent of variation, and secondly, address how it impacts the performance for various applications (e.g., single clock cycle lookup/pattern-matching to data in memory).

Figure 7.5 shows the on-wafer circuit layout for the 2-bit FeFET word comprised of 2 FeCAM cells, as well as the associated circuit schematic. Table 7.3 exhaustively lists the match line current readouts for each of the possible written and probed states of the 2-bit FeCAM word at a fixed gate read voltage of $V_G = 0.3$V. The four possible combinations to test correspond to **1)** STR1 = 0, STR2 = 0 with rows highlighted in purple; **2)** STR1 = 1, STR2 = 0 with rows highlighted in blue; **3)** STR1 = 1, STR2 = 1 with rows highlighted in green;
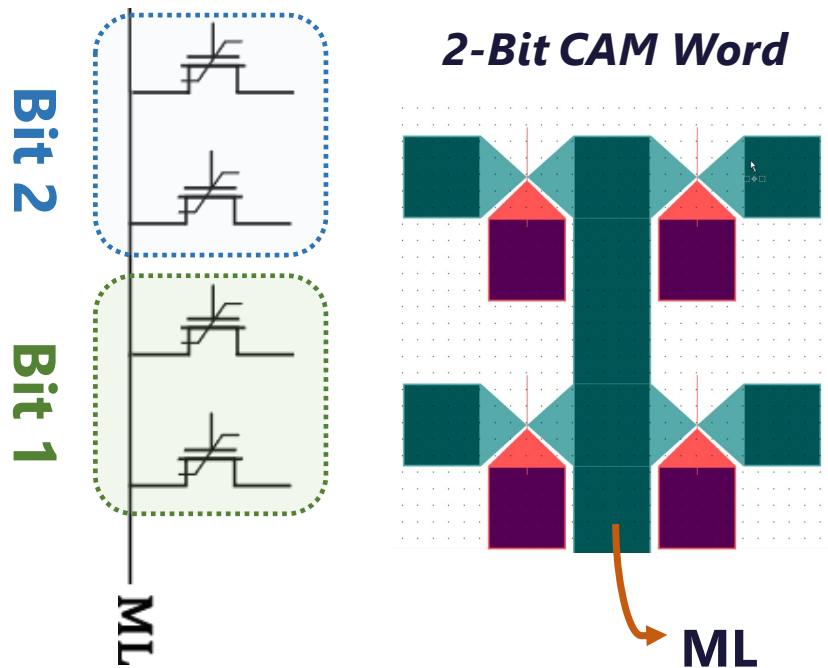
Figure 7.5: A simple 2-bit FeCAM word, both in circuit representation (left) and on-chip layout. The two bit cells share a common match line so that their current contributions can sum together.

and **4)** STR1 = 0, STR2 = 1 with rows highlighted in pink. The select lines corresponding to each individual FeCAM bit must also be probed with combinations of bit "0"s and "1"s (corresponding to $V_D = 0$V and $V_D = 50$mV, respectively).

The yellow highlighted match line current entries in Table 7.3 indicate the cases of a total exact match. Conversely, red highlighted entries indicate a total exact miss. For a perfect match, the match line current reads roughly ~ 1 μA, and for an exact miss the match line current is on the order of $\sim 0.1 - 1$ nA. Clearly, in order to also be able to distinguish partial matches/misses, there has to be enough of a margin between the on/off states for every FeFET in the word, and the less variation present, the better the projected search performance.

One final observation about the match line readout current values for the 2-bit Fe-CAM word concerns the "partial match" states achievable (for instance, when a match is achieved in one bit cell, but not the other, in the 2-bit case discussed in this subsection). In this scenario, it is preferable if the match line current readout is proportional to the total number of matched bits, if one were to extend the word length to an arbitrary n bits. And

generally, it is true that the match line current readout is proportional to the "closeness" of the match. However, the task of interpreting the degree of partial matching becomes much more challenging when intrinsic device-to-device variation exists among the FeFETs which compose each of the FeCAM cells. Partial match current readouts, as seen in Table 7.3, can range from a few hundreds of nanoamps to over 1 microamp, depending on the FeFET which is or is not conducting. This degree of uncertainty will only worsen as the number of bits in each FeCAM word increases. Future studies on FeCAMs must therefore focus on tightly controlling the threshold voltage and state variation to improve the margin of determination between the various miss, partial match, and exact match conditions.

STR1	$\overline{STR1}$	STR2	$\overline{STR2}$	SL1	$\overline{SL1}$	SL2	$\overline{SL2}$	ML
0	1	0	1	50 mV	0V	50 mV	0V	1.8 nA
0	1	0	1	0V	50 mV	50 mV	0V	759 nA
0	1	0	1	50 mV	0V	0V	50 mV	100 nA
0	1	0	1	0V	50 mV	0V	50 mV	903 nA
1	0	0	1	0V	50 mV	50 mV	0V	107 pA
1	0	0	1	0V	50 mV	0V	50 mV	155 nA
1	0	0	1	50 mV	0V	50 mV	0V	812 nA
1	0	0	1	0V	50 mV	50 mV	0V	1 μA
1	0	1	0	0V	50 mV	0V	50 mV	250 pA
1	0	1	0	0V	50 mV	50 mV	0V	165 nA
1	0	1	0	50 mV	0V	0V	50 mV	1.32 μA
1	0	1	0	50 mV	0V	50 mV	0V	1.45 μA
0	1	1	0	50 mV	0V	0V	50 mV	3 nA
0	1	1	0	50 mV	0V	50 mV	0V	293 nA
0	1	1	0	0V	50 mV	0V	50 mV	596 nA
0	1	1	0	0V	50 mV	50 mV	0V	900 nA

Table 7.3: The match line (ML) current readout for all possible probing search line bits for a 2-bit CAM word, corresponding to each of the possible stored states in both bits STR1 and STR2. Red highlighted entries indicate an exact miss, whereas yellow highlighted entries indicate an exact match.

Suggested Full FeCAM Architecture

A full example FeCAM array architecture is shown in Figure 7.6. Notably, such an array could be constructed in a 3D fashion following a back end of the line (BEOL) process with a polysilicon channel, for example. n total FeCAM cells cascaded together to share a match line form a word, and a string of m words forms a complete FeCAM array, thus enabling single clock cycle determination of a match, which enables Hamming distance calculations

for cognitive computing schemes (e.g, hyperdimensional computing [60]). Additionally, the FeCAM cell allows parallel search in hardware, thus speeding up deep neural network/convolutional neural network implementation. Further, an FeCAM crossbar array permits distributed weight updates and multiply-accumulate operations to be performed in-memory, thereby drastically reducing energy and time associated with shuttling weight data to the CPU.
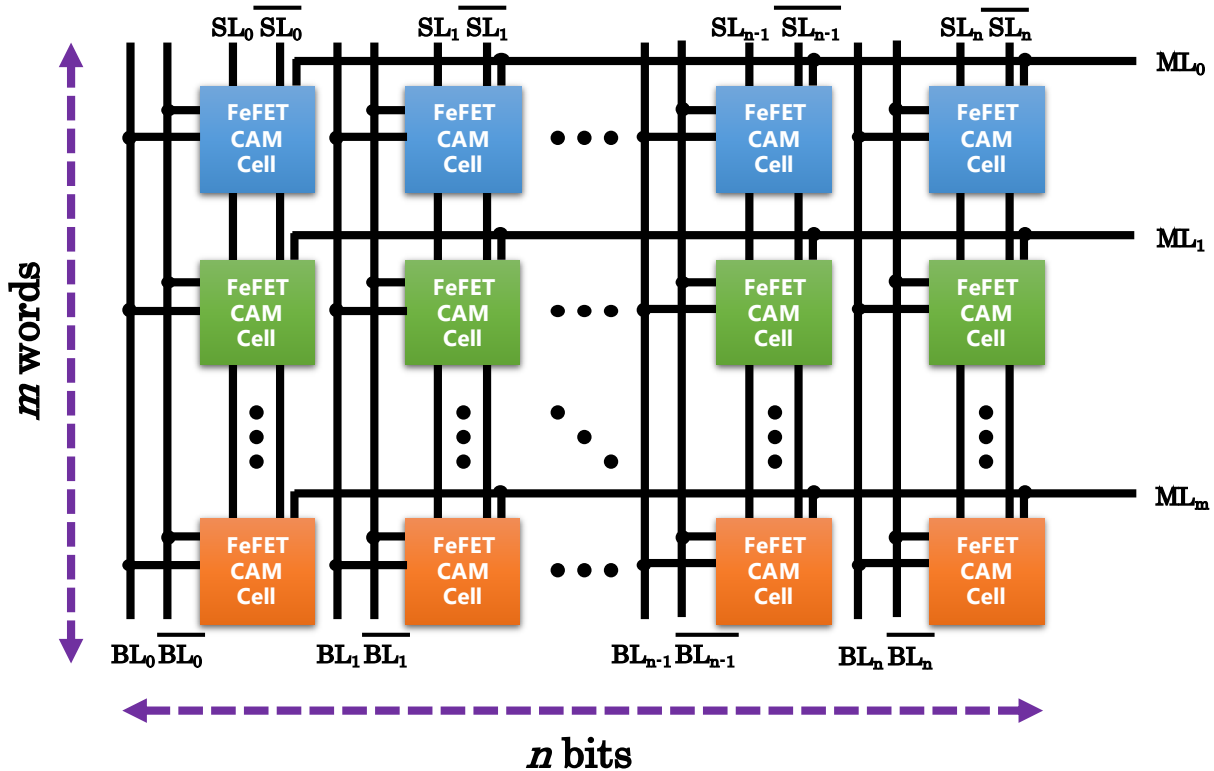
Figure 7.6: Proposed CAM cell array of dimension $m \times n$ for parallel searching. Peripheral circuitry (*not shown*) senses and compares current levels on the match lines to determine closest match to input data.

Chapter 8

Conclusions, Outlook & Future Directions

The fortuitous discovery of ferroelectricity in hafnium oxide has ushered in a new era of materials and device research. The focus areas are broad in scope, involving developing better and thinner CMOS-compatible ferroelectric films to realize all kinds of electronic components from memories – as discussed in this dissertation – to negative capacitance transistors [65, 66, 67, 134, 152], to electrostatic supercapacitors [54, 64, 100, 104], to resonators and high-precision actuators [47, 48], and beyond. Emerging memory applications are the centerpiece of this dissertation as there has been an explosion in data creation and usage in today's digital age, spurring modern challenges in the processing of data-intensive workloads. This work underscores the importance of continued materials development, of building a deeper understanding of the reliability issues facing ferroelectric HfO_2 memories, and of investigating solutions in the processing and integration space to resolve some of the technological challenges. As such, a pathway forward for continued development on CMOS-compatible, nonvolatile ferroelectric memories is established.

8.1 Summary of Work

In **Chapter 2**, two distinct hafnium oxide material systems were explored for their ferroelectric properties and projected performance: Zr-doped HfO_2 and Si-doped HfO_2. For both material systems, the entire compositional range required to elucidate ferroelectric, antiferroelectric, and dielectric responses was investigated. For the Zr-doped HfO_2 system, it was confirmed experimentally that both zirconia and hafnia form *solid solutions* across the entire compositional range of 0% HfO_2 to 100% HfO_2. In contrast, in the Si-doped HfO_2 system, only a narrow range of mol% Si doping is required to obtain ferroelectric/antiferroelectric properties. Electrical testing via PV (polarization vs. voltage) and CV (capacitance vs. voltage) loops revealed the dielectric nature of the films grown by atomic layer deposition, and were confirmed through secondary materials characteriza-

tion methods such as grazing incidence x-ray diffractometry to determine film crystallinity and x-ray spectroscopy to determine exact film composition. Other important distinctions between the two material systems include the higher overall thermal budget tolerated by the Si-doped HfO_2 system over the Zr-doped HfO_2 system ($\sim 800 - 1000$ °C for the former, versus $\sim 400 - 600$ °C for the latter), and the potential for finer granularity in dopant control in the Zr-doped HfO_2 system in comparison to Si-doped HfO_2.

In **Chapter 3**, the Si-doped capacitors characterized in **Chapter 2** are chosen for further study as potential DRAM capacitor replacements (FeRAM). As modern DRAM solutions already require trench capacitors incorporating highly engineered high-κ dielectrics, a natural solution to boost their performance is to introduce the element of *nonvolatility* by replacing the dielectric with a ferroelectric. Si-doped HfO_2 capacitors with film thicknesses ranging from 4 to 8 nm are tested for properties such as endurance, retention, and imprint. The Si-doped HfO_2 films were found to suffer from wakeup effects, needing an initial 10^4 bipolar voltage pulses to unpin and cycle defects residing at electrode interfaces. For an equivalent field stress of ± 2.5 MV/cm, the thinner ferroelectric films were found to survive for the largest number of endurance cycles, exceeding 10^9 cycles, whereas the 8 nm Si-doped HfO_2 films experienced hard oxide breakdown at 10^6 cycles. In considering the effects of the unintentional oxides which exist at the HfO_2/electrode interfaces, it is therefore suggested, since their thicknesses do not depend on the thickness of the ferroelectric HfO_2 layer, that the interfacial layers effectively screen out more of the applied voltage over time in capacitors with thinner HfO_2 layers. Elevated temperature retention testing reveals *opposite state retention* as the limiting factor in HfO_2 memory capacitors, with imprinting of the opposite state reducing the overall switchable polarization associated with the state in question.

In **Chapter 4**, the intrinsic performance of the baseline ferroelectric transistor (FeFET) is discussed and evaluated. "Baseline" refers to the usage of a chemical oxide grown in a solution of ammonium hydroxide, hydrogen peroxide, and water to achieve a thin interfacial SiO_2. The baseline FeFET is shown to have reasonable programmability characteristics, requiring programming voltages in the range of $\pm 3 - 6$ V, and pulse durations on the order of tens of nanoseconds to milliseconds. A greater spread in possible $+V_T$ states is seen in the programmability test, suggesting that the thin, hole-deficient ~ 30 nm FDSOI body can only accumulate a limited number of holes upon the application of a negative gate bias. Furthermore, it is pointed out that severe oscillations in the gate voltage are observed at the lower limit of programming speeds tested, likely due to a combination of device series resistance from the thin SOI body as well as other external impedances. Therefore, the FDSOI FeFET is not an optimal candidate to use in testing the lower limits of program/erase speeds. Lastly, the programmability and retention results from an optimized FeFET are discussed, with the process details divulged in **Chapter 6**. Interfacial oxide engineering is shown to substantially reduce the voltage required to program the FeFET, thus reducing the overall electrical stress witnessed by the device over its lifetime.

Chapter 5 focuses primarily on building a robust understanding of the key underlying reliability issues facing highly scaled FeFETs. Device-to-device variability, charge trap-

ping, and interface oxide wearout are discussed as major reliability concerns facing the FeFET. Worst yet, as illustrated in Table 5.1, many researchers working in the space of ferroelectric HfO_2 memories report ultimate device endurance numbers on the order of $\sim 10^4 - 10^6$ cycles, a troubling commonality highlighting perhaps a key integration or materials shortcoming. Much of the prior literature establishes high-κ bulk charge trapping as a major culprit in closing the FeFET memory window over time, but it is shown that for FeFETs with very thin oxides (on the order of 5 nm or less), the dominating factor turns out to be hole-trapping from hot electrons injected from gate to substrate during negative applied gate biases. In fact, through a combination of electrical measurement and simulations using the Applied MDLx Ginestra^{TM} platform [4], charge trapping was also determined to be negligible in FeFETs with very thin gate oxides. Lastly, a testing methodology on p^{++} substrates to predict the endurance properties of new FeFET gate stack designs is proposed. The efficacy of this testing methodology is confirmed through comparisons with endurance data from actual FeFETs.

In **Chapter 6**, processing techniques to improve the projected performance of FeFETs are investigated. To first order, it is important to design the interfacial oxide of the device to be robust against the high-temperature crystallization step which is required to induce ferroelectricity in the HfO_2 gate stack. During this anneal, Zr/Hf atoms are known to diffuse through thin and/or porous interfacial oxides, disrupting the existing interface bonds and causing early-stage reliability concerns. It is shown that uses a nitrided SiO_2 is most beneficial in ensuring that the interface trap density, or D_{it}, remains low after the anneal. This observation is confirmed quantitatively through the use of the *AC conductance method* for quantifying D_{it}, as well as through a CV model which self-consistently models the device stack including the quantity, nature, and time constants of the probed interface states. Next, it is shown that direct nitridation to form an interfacial layer of SiN_x leads to the most robust endurance characteristics reported in the literature to date of greater than 10^{10} cycles on silicon. The orders of magnitude improvement in the endurance metric arises from the usage of an interfacial layer with $\kappa \approx 7.8$, in comparison to $\kappa = 3.9$ in the case of SiO_2. Reducing the program/erase voltage requirements on the FeFET directly correlates to an increase in device "time-to-breakdown", thereby substantially improving cycling endurance.

In **Chapter 7**, potential use cases for FeFETs in specialized memory applications are discussed. Aside from acting a potential storage class memory candidate, as discussed in **Chapter 1**, the FeFET is also amenable for certain in-memory computing solutions. It is shown that two FeFETs with their drains tied together can act as a two-transistor implementation of a content addressable memory cell (FeCAM cell); in contrast, traditional CMOS implementations of a CAM cell typically require two SRAM cells, which are composed of at least 8 transistors each. The XNOR function (equivalently, the search function) is successfully demonstrated in the 2-FeFET CAM cell, and retention characteristics are also presented. To illustrate the detrimental impact of device-to-device variability, a 2-bit FeCAM word is also characterized, demonstrating the influence of state variability on the readout current levels (and, correspondingly, the current thresholds/margins which must

be established to distinguish between a miss, partial match, and full match). Lastly, a full FeCAM architecture comprised of m rows of n-bit words is suggested to enable single-clock-cycle lookup of data stored in memory.

8.2 Future Directions

Based on the findings in this dissertation, several suggestions for future work and directions are given as follows.

Towards Improved Cycling Robustness

For the HfO$_2$ FeFET, overcoming the critical bottlenecking factor of limited device endurance will prove to be one of the biggest challenges towards commercialization. This work has taken several steps in identifying the physical root causes of limited endurance and demonstrating the viability of interfacial oxide engineering. Future efforts to tackle device endurance should investigate other potential high-κ oxides for the interfacial layer, or perhaps at the extreme, investigate techniques to scavenge out the interfacial layer entirely (which may come at the cost of severely reduced channel mobility).

Another viable option is gate workfunction engineering. As discussed in **Chapter 5**, the high kinetic energy of electrons injected from the gate to substrate during negative biases is what fundamentally causes substrate bond breakage and subsequent hole-trapping in the ferroelectric. To counteract this issue, it is possible to use a high workfunction gate metal – for example, platinum. On the flip side, it will be important to choose a gate metal which does not run towards the other extreme of instigating electron-trapping with positive applied biases.

Though many of the high endurance strategies discussed do not fundamentally involve the ferroelectric HfO$_2$ layer itself, this will become an inevitability once the integration-induced reliability concerns are resolved. Results reported in **Chapters 3 and 6** suggest that the ferroelectric properties of the HfO$_2$ layer remain intact after $10^9 - 10^{10}$ endurnce cycles; however, by this point in the device's lifetime, severe fatiguing effects become apparent. In order to potentially extend the viability of the FeFET to last-level cache applications, it will be imperative to extend the cycling metric to 10^{12} cycles and beyond, which will undoubtedly require careful engineering of the doped HfO$_2$ itself.

In-Memory/Brain-Inspired Computing Applications

Building upon the discussion from **Chapter 7**, some additional in-memory computing and cognitive computing ideas can be explored if one considers operating an FeFET as a multi-state/multi-bit memory, rather than strictly as a binary memory element. To first order, the FeCAM array proposed in Figure 7.6 can be made even more compact, if for example a tight control of 4 possible V_T states within each FeFET can be achieved (thus leading

to a 2-bit memory implemented with just a single device). From a scalability perspective, successful V_T engineering to achieve multiple states would position the FeFET as an even more competitive emerging technology among the other two-terminal charge-based memory candidates which require an additional selector element.

Secondly, the polycrystalline nature of ferroelectric HfO_2 bestows upon it the reputation of being an "imperfect" ferroelectric in the materials research community, in stark contrast against the traditional single-crystalline ferroelectrics studied prior. However, from the perspective of designing neuromorphic hardware – for example, implementing spike-timing dependent plasticity or designing hardware neural networks – the ability to produce distinct analog responses when subjected to different input conditions is advantageous in emulating synaptic behavior. Whereas a "perfect" ferroelectric more abruptly flips its polarization once a voltage greater than the coercive voltage is applied, doped HfO_2 typically shows gradual switching characteristics. Hence, the property which leads to an intrinsic spread in V_T may be advantageous for implementing other analog-based computing hardware.

Bibliography

[1] "A century of ferroelectricity". In: *Nature Materials* 19.2 (2020), pp. 129–129. DOI: 10.1038/s41563-020-0611-1.

[2] T. Ali et al. "High Endurance Ferroelectric Hafnium Oxide-Based FeFET Memory Without Retention Penalty". In: *IEEE Transactions on Electron Devices* 65.9 (2018), pp. 3769–3774. DOI: 10.1109/ted.2018.2856818.

[3] D. Apalkov, B. Dieny, and J. M. Slaughter. "Magnetoresistive Random Access Memory". In: *Proceedings of the IEEE* 104.10 (2016), pp. 1796–1830. DOI: 10.1109/JPROC.2016.2590142.

[4] *Applied MDLx^{TM} Ginestra^{TM} Simulation Software*. URL: https://www.appliedmaterials.com/products/applied-mdlx-ginestra-simulation-software.

[5] J. M. Atkin et al. "Charge trapping at the low-k dielectric-silicon interface probed by the conductance and capacitance techniques". In: *Applied Physics Letters* 93.12 (2008), p. 122902. DOI: 10.1063/1.2990648.

[6] A. Aziz et al. "Computing with ferroelectric FETs: Devices, models, systems, and applications". In: *2018 Design, Automation Test in Europe Conference Exhibition (DATE)*. Mar. 2018, pp. 1289–1298. DOI: 10.23919/DATE.2018.8342213.

[7] J. -H. Bae et al. "Highly Scaled, High Endurance, Ω-Gate, Nanowire Ferroelectric FET Memory Transistors". In: *IEEE Electron Device Letters* 41.11 (2020), pp. 1637–1640. DOI: 10.1109/LED.2020.3028339.

[8] Brian Bailey. *Von Neumann Is Struggling*. Jan. 2021. URL: https://semiengineering.com/von-neumann-is-struggling/.

[9] Writam Banerjee et al. "Design of CMOS Compatible, High-Speed, Highly-Stable Complementary Switching with Multilevel Operation in 3D Vertically Stacked Novel HfO2/Al2O3/TiOx (HAT) RRAM". In: *Advanced Electronic Materials* 4.2 (2018), p. 1700561. DOI: 10.1002/aelm.201700561.

[10] B. V. Benjamin et al. "Neurogrid: A Mixed-Analog-Digital Multichip System for Large-Scale Neural Simulations". In: *Proceedings of the IEEE* 102.5 (May 2014), pp. 699–716. DOI: 10.1109/JPROC.2014.2313565.

[11] "Beyond von Neumann". In: *Nature Nanotechnology* 15.7 (2020), pp. 507–507. DOI: 10.1038/s41565-020-0738-x.

[12] Sabpreet Bhatti et al. "Spintronics based random access memory: a review". In: *Materials Today* 20.9 (2017), pp. 530–548. ISSN: 1369-7021. DOI: https://doi.org/10.1016/j.mattod.2017.07.007. URL: https://www.sciencedirect.com/science/article/pii/S1369702117304285.

[13] Mark T. Bohr et al. *The High-k Solution*. Oct. 2007. URL: https://spectrum.ieee.org/semiconductors/design/the-highk-solution.

[14] T. S. Böscke et al. "Ferroelectricity in hafnium oxide thin films". In: *Applied Physics Letters* 99.10 (2011), p. 102903. DOI: 10.1063/1.3634052. eprint: https://doi.org/10.1063/1.3634052.

[15] T. S. Böscke et al. "Increasing permittivity in HfZrO thin films by surface manipulation". In: *Applied Physics Letters* 95.5 (2009), p. 052904. DOI: 10.1063/1.3195623. eprint: https://doi.org/10.1063/1.3195623.

[16] E. T. Breyer et al. "Demonstration of versatile nonvolatile logic gates in 28nm HKMG FeFET technology". In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. May 2018, pp. 1–5. DOI: 10.1109/ISCAS.2018.8351408.

[17] Evelyn T. Breyer and Stefan Slesazeck. "Chapter 10.6 - Ferroelectric Devices for Logic in Memory". In: *Ferroelectricity in Doped Hafnium Oxide: Materials, Properties and Devices*. Ed. by Uwe Schroeder, Cheol Seong Hwang, and Hiroshi Funakubo. Woodhead Publishing Series in Electronic and Optical Materials. Woodhead Publishing, 2019, pp. 495–513. ISBN: 978-0-08-102430-0. DOI: https://doi.org/10.1016/B978-0-08-102430-0.00024-3. URL: https://www.sciencedirect.com/science/article/pii/B9780081024300000243.

[18] Geoffrey W. Burr et al. "Phase change memory technology". In: *Journal of Vacuum Science & Technology B* 28.2 (2010), pp. 223–262. DOI: 10.1116/1.3301579. eprint: https://doi.org/10.1116/1.3301579. URL: https://doi.org/10.1116/1.3301579.

[19] Rongrong Cao et al. "Improvement of Endurance in HZO-Based Ferroelectric Capacitor Using Ru Electrode". In: *IEEE Electron Device Letters* 40.11 (2019), pp. 1744–1747. DOI: 10.1109/led.2019.2944960.

[20] R. Castagné and A. Vapaille. "Description of the SiO2/Si interface properties by means of very low frequency MOS capacitance measurements". In: *Surface Science* 28.1 (1971), pp. 157–193. ISSN: 0039-6028. DOI: 10.1016/0039-6028(71)90092-6.

[21] Premi Chandra and Peter B. Littlewood. "A Landau Primer for Ferroelectrics". In: *Topics in Applied Physics Physics of Ferroelectrics* (), pp. 69–116. DOI: 10.1007/978-3-540-34591-6_3.

[22] K. Chatterjee et al. "Self-Aligned, Gate Last, FDSOI, Ferroelectric Gate Memory Device With 5.5-nm Hf0.8Zr0.2O2, High Endurance and Breakdown Recovery". In: *IEEE Electron Device Letters* 38.10 (2017), pp. 1379–1382. DOI: 10.1109/LED.2017.2748992.

[23] Korok Chatterjee et al. "Challenges to Partial Switching of Hf0.8Zr0.2O2 Gated Ferroelectric FET for Multilevel/Analog or Low-Voltage Memory Operation". In: *IEEE Electron Device Letters* 40.9 (2019), pp. 1423–1426. DOI: 10.1109/led.2019.2931430.

[24] Suraj S. Cheema et al. "Enhanced ferroelectricity in ultrathin films grown directly on silicon". In: *Nature* 580.7804 (2020), pp. 478–482. DOI: 10.1038/s41586-020-2208-x.

[25] An Chen. "A review of emerging non-volatile memory (NVM) technologies and applications". In: *Solid-State Electronics* 125 (2016), pp. 25–38. ISSN: 0038-1101. DOI: https://doi.org/10.1016/j.sse.2016.07.006.

[26] B. Chen et al. "Ge-Based Asymmetric RRAM Enable $8F^2$ Content Addressable Memory". In: *IEEE Electron Device Letters* 39.9 (Sept. 2018), pp. 1294–1297. ISSN: 0741-3106. DOI: 10.1109/LED.2018.2856537.

[27] Hong-Yu Chen et al. "Resistive random access memory (RRAM) technology: From material, device, selector, 3D integration to bottom-up fabrication". In: *Journal of Electroceramics* 39.1-4 (2017), pp. 21–38. DOI: 10.1007/s10832-017-0095-9.

[28] K. -. Chen et al. "Non-Volatile Ferroelectric FETs Using 5-nm Hf0.5Zr0.5O2 With High Data Retention and Read Endurance for 1T Memory Applications". In: *IEEE Electron Device Letters* 40.3 (2019), pp. 399–402. DOI: 10.1109/LED.2019.2896231.

[29] Shou Mian Chen and Suk-Yin Flora Ip. "Nitrogen in ultra-thin gate oxides: its profile and functions". In: *Solid-State Electronics* 43.11 (1999), pp. 1997–2003. ISSN: 0038-1101. DOI: 10.1016/S0038-1101(99)00166-5.

[30] Se-Na Choi, Seung-Eon Moon, and Sung-Min Yoon. "Film thickness-dependent ferroelectric polarization switching dynamics of undoped HfO2 thin films prepared by atomic layer deposition". In: *Ceramics International* 45.17, Part B (2019), pp. 22642–22648. ISSN: 0272-8842. DOI: https://doi.org/10.1016/j.ceramint.2019.07.297.

[31] T. Coughlin. "A Timeline for Flash Memory History [The Art of Storage]". In: *IEEE Consumer Electronics Magazine* 6.1 (2017), pp. 126–133. DOI: 10.1109/MCE.2016.2614739.

[32] L. E. Cross and R. E. Newnham. "History of ferroelectrics". In: *Ceramics and civilization III: High-technology ceramics–Past, present and future* (1987), pp. 289–305.

[33] Jiyan Dai. "Introduction to Ferroelectrics". In: *Ferroic Materials for Smart Systems*. John Wiley & Sons, Ltd, 2020. Chap. 2, pp. 15–46. ISBN: 9783527815388. DOI: https://doi.org/10.1002/9783527815388.ch2.

[34] *Data, Data Everywhere – Storage on the Brink?* Jan. 2020. URL: https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/building-an-enterprise-data-strategy.pdf.

[35] Scott Deboer. "Memory Technology: The Core to Enable Future Computing Systems". In: *2018 IEEE Symposium on VLSI Technology* (2018). DOI: 10.1109/vlsit.2018.8510707.

[36] M. T. Do et al. "Interface degradation and field screening mechanism behind bipolar-cycling fatigue in ferroelectric capacitors". In: *APL Materials* 9.2 (2021), p. 021113. DOI: 10.1063/5.0038719. eprint: https://doi.org/10.1063/5.0038719.

[37] L. Do Thanh and P. Balk. "Elimination and Generation of Si - SiO2 Interface Traps by Low Temperature Hydrogen Annealing". In: *Journal of The Electrochemical Society* 135.7 (1988), pp. 1797–1801. DOI: 10.1149/1.2096133. eprint: http://jes.ecsdl.org/content/135/7/1797.full.pdf+html.

[38] *DRAM Scaling Challenges Grow*. Nov. 2019. URL: https://semiengineering.com/dram-scaling-challenges-grow/.

[39] Stephen Ducharme et al. "Intrinsic Ferroelectric Coercive Field". In: *Phys. Rev. Lett.* 84 (1 Jan. 2000), pp. 175–178. DOI: 10.1103/PhysRevLett.84.175. URL: https://link.aps.org/doi/10.1103/PhysRevLett.84.175.

[40] S. Dunkel et al. "A FeFET based super-low-power ultra-fast embedded NVM technology for 22nm FDSOI and beyond". In: *2017 IEEE International Electron Devices Meeting (IEDM)* (2017). DOI: 10.1109/iedm.2017.8268425.

[41] K. A. Ellis and R. A. Buhrman. "Nitrous oxide (N2O) processing for silicon oxynitride gate dielectrics". In: *IBM Journal of Research and Development* 43.3 (May 1999), pp. 287–300. DOI: 10.1147/rd.433.0287.

[42] Timon Evenblij and Gouri Sankar Kar. *The history and future of DRAM architectures in different application domains – an analysis.* June 2020. URL: https://www.imec-int.com/en/imec-magazine/imec-magazine-june-2020/the-history-and-future-of-dram-architectures-in-different-application-domains-an-analysis.

[43] Zhen Fan, Jingsheng Chen, and John Wang. "Ferroelectric HfO2-based materials for next-generation ferroelectric memories". In: *Journal of Advanced Dielectrics* 06.02 (2016), p. 1630003. DOI: 10.1142/s2010135x16300036.

[44] S. W. Fong, C. M. Neumann, and H. -. P. Wong. "Phase-Change Memory—Towards a Storage-Class Memory". In: *IEEE Transactions on Electron Devices* 64.11 (2017), pp. 4374–4385. DOI: 10.1109/TED.2017.2746342.

[45] R. F. Freitas and W. W. Wilcke. "Storage-class memory: The next storage system technology". In: *IBM Journal of Research and Development* 52.4.5 (2008), pp. 439–447. DOI: 10.1147/rd.524.0439.

[46] Manuel Le Gallo and Abu Sebastian. "An overview of phase-change memory device physics". In: *Journal of Physics D: Applied Physics* 53.21 (2020), p. 213002. DOI: `10.1088/1361-6463/ab7794`.

[47] M. Ghatge et al. "A 30-nm thick integrated hafnium zirconium oxide nano-electromechanical membrane resonator". In: *Applied Physics Letters* 116.4 (2020), p. 043501. DOI: `10.1063/1.5134856`. eprint: `https://doi.org/10.1063/1.5134856`. URL: `https://doi.org/10.1063/1.5134856`.

[48] Mayur Ghatge et al. "An ultrathin integrated nanoelectromechanical transducer based on hafnium zirconium oxide". In: *Nature Electronics* 2.11 (2019), pp. 506–512. DOI: `10.1038/s41928-019-0305-3`.

[49] Nanbo Gong and Tso-Ping Ma. "A Study of Endurance Issues in HfO2-Based Ferroelectric Field Effect Transistors: Charge Trapping and Trap Generation". In: *IEEE Electron Device Letters* 39.1 (2018), pp. 15–18. DOI: `10.1109/led.2017.2776263`.

[50] Shashaank Gupta. "1 - Introduction to ferroelectrics and related materials". In: *Ferroelectric Materials for Energy Harvesting and Storage*. Ed. by Deepam Maurya, Abhijit Pramanick, and Dwight Viehland. Woodhead Publishing Series in Electronic and Optical Materials. Woodhead Publishing, 2021, pp. 1–41. ISBN: 978-0-08-102802-5. DOI: `https://doi.org/10.1016/B978-0-08-102802-5.00001-7`. URL: `https://www.sciencedirect.com/science/article/pii/B9780081028025000017`.

[51] E. P. Gusev et al. "Growth and characterization of ultrathin nitrided silicon oxide films". In: *IBM Journal of Research and Development* 43.3 (May 1999), pp. 265–286. DOI: `10.1147/rd.433.0265`.

[52] F. T. Hady et al. "Platform Storage Performance With 3D XPoint Technology". In: *Proceedings of the IEEE* 105.9 (2017), pp. 1822–1833. DOI: `10.1109/JPROC.2017.2731776`.

[53] Y. Higashi et al. "Impact of Charge trapping on Imprint and its Recovery in HfO2 based FeFET". In: *2019 IEEE International Electron Devices Meeting (IEDM)* (2019). DOI: `10.1109/iedm19573.2019.8993472`.

[54] Michael Hoffmann et al. "Negative Capacitance for Electrostatic Supercapacitors". In: *Advanced Energy Materials* 9.40 (2019), p. 1901154. DOI: `https://doi.org/10.1002/aenm.201901154`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.201901154`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/aenm.201901154`.

[55] M. Imani et al. "Efficient neural network acceleration on GPGPU using content addressable memory". In: *Design, Automation Test in Europe Conference Exhibition (DATE), 2017*. Mar. 2017, pp. 1026–1031. DOI: `10.23919/DATE.2017.7927141`.

[56] M. Imani et al. "Resistive CAM Acceleration for Tunable Approximate Computing". In: *IEEE Transactions on Emerging Topics in Computing* 7.2 (Apr. 2019), pp. 271–280. ISSN: 2168-6750. DOI: `10.1109/TETC.2016.2642057`.

[57] Takashi Ito et al. "Thermally grown silicon nitride films for high-performance MNS devices". In: *Applied Physics Letters* 32.5 (1978), pp. 330–331.

[58] Kiyoo Itoh. "The History of DRAM Circuit Designs – At the Forefront of DRAM Development –". In: *IEEE Solid-State Circuits Newsletter* 13.1 (2008), pp. 27–31. DOI: 10.1109/n-ssc.2008.4785688.

[59] D. Kahng and S. M. Sze. "A floating gate and its application to memory devices". In: *The Bell System Technical Journal* 46.6 (1967), pp. 1288–1295. DOI: 10.1002/j.1538-7305.1967.tb01738.x.

[60] Pentti Kanerva. "Hyperdimensional Computing: An Introduction to Computing in Distributed Representation with High-Dimensional Random Vectors". In: *Cognitive Computation* 1.2 (June 2009), pp. 139–159. DOI: 10.1007/s12559-009-9009-8.

[61] G. Karbasian et al. "Stabilization of ferroelectric phase in tungsten capped Hf0.8Zr0.2O2". In: *Applied Physics Letters* 111 (2017), p. 022907.

[62] Han Joon Kim et al. "A study on the wake-up effect of ferroelectric Hf0.5Zr0.5O2films by pulse-switching measurement". In: *Nanoscale* 8.3 (2016), pp. 1383–1389. DOI: 10.1039/c5nr05339k.

[63] Min-Kyu Kim, Ik-Jyae Kim, and Jang-Sik Lee. "CMOS-compatible ferroelectric NAND flash memory for high-density, low-power, and high-speed three-dimensional memory". In: *Science Advances* 7.3 (2021). DOI: 10.1126/sciadv.abe1341. eprint: https://advances.sciencemag.org/content/7/3/eabe1341.full.pdf.

[64] Si Joon Kim et al. "Stress-Induced Crystallization of Thin Hf1–XZrXO2 Films: The Origin of Enhanced Energy Density with Minimized Energy Loss for Lead-Free Electrostatic Energy Storage Applications". In: *ACS Applied Materials & Interfaces* 11.5 (2019), pp. 5208–5214. DOI: 10.1021/acsami.8b17211. eprint: https://doi.org/10.1021/acsami.8b17211.

[65] D. Kwon et al. "Improved Subthreshold Swing and Short Channel Effect in FDSOI n-Channel Negative Capacitance Field Effect Transistors". In: *IEEE Electron Device Letters* 39.2 (2018), pp. 300–303. DOI: 10.1109/LED.2017.2787063.

[66] D. Kwon et al. "Near Threshold Capacitance Matching in a Negative Capacitance FET With 1 nm Effective Oxide Thickness Gate Stack". In: *IEEE Electron Device Letters* 41.1 (2020), pp. 179–182. DOI: 10.1109/LED.2019.2951705.

[67] D. Kwon et al. "Negative Capacitance FET With 1.8-nm-Thick Zr-Doped HfO2 Oxide". In: *IEEE Electron Device Letters* 40.6 (2019), pp. 993–996. DOI: 10.1109/LED.2019.2912413.

[68] Ann Franchesca Laguna et al. "Ferroelectric FET Based In-Memory Computing for Few-Shot Learning". In: *Proceedings of the 2019 on Great Lakes Symposium on VLSI*. GLSVLSI '19. Tysons Corner, VA, USA: Association for Computing Machinery, 2019, pp. 373–378. ISBN: 9781450362528. DOI: 10.1145/3299874.3319450.

[69] Tae Yoon Lee et al. "Ferroelectric Polarization-Switching Dynamics and Wake-Up Effect in Si-Doped HfO2". In: *ACS Applied Materials & Interfaces* 11.3 (2018), pp. 3142–3149. DOI: 10.1021/acsami.8b11681.

[70] Chao Li et al. "A Scalable Design of Multi-Bit Ferroelectric Content Addressable Memory for Data-Centric Computing". In: *2020 IEEE International Electron Devices Meeting (IEDM)* (2020). DOI: 10.1109/iedm13553.2020.9372119.

[71] You-Sheng Liu and Pin Su. "Variability Analysis for Ferroelectric FET Nonvolatile Memories Considering Random Ferroelectric-Dielectric Phase Distribution". In: *IEEE Electron Device Letters* 41.3 (2020), pp. 369–372. DOI: 10.1109/led.2020.2967423.

[72] Patrick D. Lomenzo et al. "Ferroelectric phenomena in Si-doped HfO2 thin films with TiN and Ir electrodes". In: *Journal of Vacuum Science & Technology B* 32.3 (2014), p. 03D123. DOI: 10.1116/1.4873323. eprint: https://doi.org/10.1116/1.4873323.

[73] Patrick D. Lomenzo et al. "TaN interface properties and electric field cycling effects on ferroelectric Si-doped HfO2 thin films". In: *Journal of Applied Physics* 117.13 (2015), p. 134105. DOI: 10.1063/1.4916715. eprint: https://doi.org/10.1063/1.4916715.

[74] J. W. Lu et al. "Spintronics technology: past, present and future". In: *International Materials Reviews* 61.7 (2016), pp. 456–472. DOI: 10.1080/09506608.2016.1204097. eprint: https://doi.org/10.1080/09506608.2016.1204097. URL: https://doi.org/10.1080/09506608.2016.1204097.

[75] Arne Luker. *A Short History of Ferroelectricity*. Feb. 2011. URL: http://groups.ist.utl.pt/rschwarz/rschwarzgroup_files/Ferroelectrics_files/A%5C%20Short%5C%20History%5C%20of%5C%20Ferroelectricity.pdf.

[76] Lane W. Martin and Andrew M. Rappe. "Thin-film ferroelectric materials and their applications". In: *Nature Reviews Materials* 2.2 (2016). DOI: 10.1038/natrevmats.2016.87.

[77] Sally A. McKee. "Reflections on the Memory Wall". In: *Proceedings of the 1st Conference on Computing Frontiers*. CF '04. Ischia, Italy: Association for Computing Machinery, 2004, p. 162. ISBN: 1581137419. DOI: 10.1145/977091.977115. URL: https://doi.org/10.1145/977091.977115.

[78] Jagan Meena et al. "Overview of emerging nonvolatile memory technologies". In: *Nanoscale Research Letters* 9.1 (Sept. 2014), p. 526. DOI: 10.1186/1556-276x-9-526.

[79] Paul A. Merolla et al. "A million spiking-neuron integrated circuit with a scalable communication network and interface". In: *Science* 345.6197 (2014), pp. 668–673. DOI: 10.1126/science.1254642.

[80] T. Mikolajick et al. "Next generation ferroelectric materials for semiconductor process integration and their applications". In: *Journal of Applied Physics* 129.10 (2021), p. 100901. DOI: 10.1063/5.0037617. eprint: https://doi.org/10.1063/5.0037617.

[81] Johannes Mueller, Stefan Slesazeck, and Thomas Mikolajick. "Chapter 10.4 - Ferroelectric Field Effect Transistor". In: *Ferroelectricity in Doped Hafnium Oxide: Materials, Properties and Devices*. Ed. by Uwe Schroeder, Cheol Seong Hwang, and Hiroshi Funakubo. Woodhead Publishing Series in Electronic and Optical Materials. Woodhead Publishing, 2019, pp. 451–471. ISBN: 978-0-08-102430-0. DOI: https://doi.org/10.1016/B978-0-08-102430-0.00022-X.

[82] S. Mueller et al. "Ferroelectricity in Gd-Doped HfO2 Thin Films". In: *ECS Journal of Solid State Science and Technology* 1.6 (2012), N123–N126. DOI: 10.1149/2.002301jss.

[83] S. Mueller et al. "Next-generation ferroelectric memories based on FE-HfO2". In: *2015 Joint IEEE International Symposium on the Applications of Ferroelectric (ISAF), International Symposium on Integrated Functionalities (ISIF), and Piezoelectric Force Microscopy Workshop (PFM)* (2015). DOI: 10.1109/isaf.2015.7172714.

[84] S. Mueller et al. "Reliability Characteristics of Ferroelectric Si:HfO$_2$ Thin Films for Memory Applications". In: *IEEE Transactions on Device and Materials Reliability* 13.1 (2013), pp. 93–97. DOI: 10.1109/TDMR.2012.2216269.

[85] J. Muller et al. "High endurance strategies for hafnium oxide based ferroelectric field effect transistor". In: *2016 16th Non-Volatile Memory Technology Symposium (NVMTS)* (2016). DOI: 10.1109/nvmts.2016.7781517.

[86] J. Müller et al. "Ferroelectric Hafnium Oxide Based Materials and Devices: Assessment of Current Status and Future Prospects". In: *ECS Journal of Solid State Science and Technology* 4.5 (2015). DOI: 10.1149/2.0081505jss.

[87] J. Müller et al. "Ferroelectric hafnium oxide: A CMOS-compatible and highly scalable approach to future ferroelectric memories". In: *2013 IEEE International Electron Devices Meeting* (2013). DOI: 10.1109/iedm.2013.6724605.

[88] J. Müller et al. "Ferroelectricity in yttrium-doped hafnium oxide". In: *Journal of Applied Physics* 110.11 (2011), p. 114113. DOI: 10.1063/1.3667205.

[89] Johannes Müller et al. "Ferroelectric Hafnium Oxide A Game Changer to FRAM?" In: *2014 14th Annual Non-Volatile Memory Technology Symposium (NVMTS)* (2014). DOI: 10.1109/nvmts.2014.7060838.

[90] Johannes Müller et al. "Ferroelectricity in Simple Binary ZrO2 and HfO2". In: *Nano Letters* 12.8 (2012), pp. 4318–4323. DOI: 10.1021/nl302049k.

[91] S. Müller et al. "Incipient Ferroelectricity in Al-Doped HfO2 Thin Films". In: *Advanced Functional Materials* 22 (2012), pp. 2412–2417.

[92] Kai Ni et al. "Critical Role of Interlayer in Hf0.5Zr0.5O2 Ferroelectric FET Non-volatile Memory Performance". In: *IEEE Transactions on Electron Devices* 65.6 (2018), pp. 2461–2469. DOI: 10.1109/ted.2018.2829122.

[93] Kai Ni et al. "Ferroelectric ternary content-addressable memory for one-shot learning". In: *Nature Electronics* 3.2 (2020), pp. 130–130. DOI: 10.1038/s41928-020-0374-3.

[94] Kai Ni et al. "Fundamental Understanding and Control of Device-to-Device Variation in Deeply Scaled Ferroelectric FETs". In: *2019 Symposium on VLSI Technology* (2019). DOI: 10.23919/vlsit.2019.8776497.

[95] Kai Ni et al. "Impact of Extrinsic Variation Sources on the Device-to-Device Variation in Ferroelectric FET". In: *2020 IEEE International Reliability Physics Symposium (IRPS)* (2020). DOI: 10.1109/irps45951.2020.9128323.

[96] E. H. Nicollian and A. Goetzberger. "The Si-SiO2 Interface — Electrical Properties as Determined by the Metal-Insulator-Silicon Conductance Technique". In: *Bell System Technical Journal* 46.6 (1967), pp. 1055–1133. ISSN: 1538-7305. DOI: 10.1002/j.1538-7305.1967.tb01727.x.

[97] Min Hyuk Park et al. "Chapter 8 - Impact of Electrodes on the Ferroelectric Properties". In: *Ferroelectricity in Doped Hafnium Oxide: Materials, Properties and Devices*. Ed. by Uwe Schroeder, Cheol Seong Hwang, and Hiroshi Funakubo. Woodhead Publishing Series in Electronic and Optical Materials. Woodhead Publishing, 2019, pp. 341–364. ISBN: 978-0-08-102430-0. DOI: https://doi.org/10.1016/B978-0-08-102430-0.00008-5.

[98] Min Hyuk Park et al. "Evolution of phases and ferroelectric properties of thin Hf0.5Zr0.5O2 films according to the thickness and annealing temperature". In: *Applied Physics Letters* 102.24 (2013), p. 242905. DOI: 10.1063/1.4811483.

[99] Min Hyuk Park et al. "Study on the degradation mechanism of the ferroelectric properties of thin Hf0.5Zr0.5O2 films on TiN and Ir electrodes". In: *Applied Physics Letters* 105.7 (2014), p. 072902. DOI: 10.1063/1.4893376. eprint: https://doi.org/10.1063/1.4893376.

[100] Min Hyuk Park et al. "Thin HfxZr1-xO2 Films: A New Lead-Free System for Electrostatic Supercapacitors with Large Energy Storage Density and Robust Thermal Stability". In: *Advanced Energy Materials* 4.16 (2014), p. 1400610. DOI: https://doi.org/10.1002/aenm.201400610. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.201400610.

[101] M. Pešić et al. "Comparative Study of Reliability of Ferroelectric and Anti-Ferroelectric Memories". In: *IEEE Transactions on Device and Materials Reliability* 18.2 (2018), pp. 154–162. DOI: 10.1109/TDMR.2018.2829112.

[102] M. Pešić et al. "Deconvoluting charge trapping and nucleation interplay in Fe-FETs: Kinetics and Reliability". In: *2018 IEEE International Electron Devices Meeting (IEDM)*. 2018, pp. 25.1.1–25.1.4. DOI: 10.1109/IEDM.2018.8614492.

[103] Milan Pešić et al. "Multiscale Modeling of Ferroelectric Memories: Insights into Performances and Reliability". In: *2018 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)* (2018). DOI: 10.1109/sispad.2018.8551722.

[104] Milan Pešić et al. "Nonvolatile Random Access Memory and Energy Storage Based on Antiferroelectric Like Hysteresis in ZrO2". In: *Advanced Functional Materials* 26.41 (2016), pp. 7486–7494. DOI: https://doi.org/10.1002/adfm.201603182. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/adfm.201603182.

[105] Milan Pešić et al. "Physical Mechanisms behind the Field-Cycling Behavior of HfO2-Based Ferroelectric Capacitors". In: *Advanced Functional Materials* 26.25 (2016), pp. 4601–4612. DOI: 10.1002/adfm.201600590.

[106] E. H. Poindexter et al. "Electronic traps and Pb centers at the Si/SiO2 interface: Band-gap energy distribution". In: *Journal of Applied Physics* 56.10 (1984), pp. 2844–2849. DOI: 10.1063/1.333819.

[107] Patrick Polakowski and Johannes Müller. "Ferroelectricity in undoped hafnium oxide". In: *Applied Physics Letters* 106.23 (2015), p. 232905. DOI: 10.1063/1.4922272. eprint: https://doi.org/10.1063/1.4922272.

[108] Simone Raoux et al. "Phase change materials and phase change memory". In: *MRS Bulletin* 39.8 (2014), pp. 703–710. DOI: 10.1557/mrs.2014.139.

[109] Michael L. Reed and James D. Plummer. "Chemistry of Si-SiO2 interface trap annealing". In: *Journal of Applied Physics* 63.12 (1988), pp. 5776–5793. DOI: 10.1063/1.340317.

[110] David Reinsel, John Gantz, and John Rydning. *The Digitization of the World From Edge to Core*. Nov. 2018. URL: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf.

[111] Sebastian E. Reyes-Lillo, Kevin F. Garrity, and Karin M. Rabe. "Antiferroelectricity in thin-film ZrO_2 from first principles". In: *Phys. Rev. B* 90 (14 Oct. 2014), p. 140103. DOI: 10.1103/PhysRevB.90.140103.

[112] Robert Ruh and Peter W. R. Corfield. "Crystal Structure of Monoclinic Hafnia and Comparison with Monoclinic Zirconia". In: *Journal of the American Ceramic Society* 53.3 (1970), pp. 126–129. ISSN: 1551-2916. DOI: 10.1111/j.1151-2916.1970.tb12052.x.

[113] Hojoon Ryu et al. "Exploring New Metal Electrodes for Ferroelectric Aluminum-Doped Hafnium Oxide". In: *IEEE Transactions on Electron Devices* 66.5 (2019), pp. 2359–2364. DOI: 10.1109/ted.2019.2907070.

[114] Leonid Savtchenko et al. *Method of writing to scalable magnetoresistance random access memory element*. Apr. 2003.

[115] T Schenk et al. "Memory technology—a primer for material scientists". In: *Reports on Progress in Physics* 83.8 (2020), p. 086501. DOI: 10.1088/1361-6633/ab8f86.

[116] Uwe Schroeder et al. "Lanthanum-Doped Hafnium Oxide: A Robust Ferroelectric Material". In: *Inorganic Chemistry* 57.5 (2018), pp. 2752–2765. DOI: 10.1021/acs.inorgchem.7b03149.

[117] Klaus F Schuegraf and Chenming Hu. "Hole injection SiO/sub 2/breakdown model for very low voltage lifetime extrapolation". In: *IEEE transactions on electron devices* 41.5 (1994), pp. 761–767.

[118] Abu Sebastian et al. "Memory devices and applications for in-memory computing". In: *Nature Nanotechnology* 15.7 (2020), pp. 529–544. DOI: 10.1038/s41565-020-0655-z.

[119] A. A. Sharma et al. "High Speed Memory Operation in Channel-Last, Back-gated Ferroelectric Transistors". In: *2020 IEEE International Electron Devices Meeting (IEDM)*. 2020, pp. 18.5.1–18.5.4. DOI: 10.1109/IEDM13553.2020.9371940.

[120] Shigeru Shiratake. "Scaling and Performance Challenges of Future DRAM". In: *2020 IEEE International Memory Workshop (IMW)* (2020). DOI: 10.1109/imw48823.2020.9108122.

[121] Max M. Shulaker et al. "Three-dimensional integration of nanotechnologies for computing and data storage on a single chip". In: *Nature* 547.7661 (2017), pp. 74–78. DOI: 10.1038/nature22994.

[122] Stefan Slesazeck, Uwe Schroeder, and Thomas Mikolajick. "Embedding hafnium oxide based FeFETs in the memory landscape". In: *2018 International Conference on IC Design & Technology (ICICDT)* (2018). DOI: 10.1109/icicdt.2018.8399771.

[123] Viktor Sverdlov et al. "CMOS-compatible spintronic devices". In: *2015 30th Symposium on Microelectronics Technology and Devices (SBMicro)* (2015). DOI: 10.1109/sbmicro.2015.7298103.

[124] A. J. Tan et al. "Experimental Demonstration of a Ferroelectric HfO2-Based Content Addressable Memory Cell". In: *IEEE Electron Device Letters* 41.2 (2020), pp. 240–243. DOI: 10.1109/LED.2019.2963300.

[125] A. J. Tan et al. "Hot Electrons as the Dominant Source of Degradation for Sub-5nm HZO FeFETs". In: *2020 IEEE Symposium on VLSI Technology*. 2020, pp. 1–2. DOI: 10.1109/VLSITechnology18217.2020.9265067.

[126] Kazuyuki Tomida, Koji Kita, and Akira Toriumi. "Dielectric constant enhancement due to Si incorporation into HfO2". In: *Applied Physics Letters* 89.14 (2006), p. 142902. DOI: 10.1063/1.2355471. eprint: https://doi.org/10.1063/1.2355471.

[127] K. Toprasertpong, M. Takenaka, and S. Takagi. "Direct Observation of Interface Charge Behaviors in FeFET by Quasi-Static Split C-V and Hall Techniques: Revealing FeFET Operation". In: *2019 IEEE International Electron Devices Meeting (IEDM)* (2019). DOI: `10.1109/iedm19573.2019.8993664`.

[128] L. Truong et al. "Admittance spectroscopy of traps at the interfaces of (100)Si with Al2O3, ZrO2, and HfO2". In: *Microelectronics Reliability* 45.5 (2005). 13th Workshop on Dielectrics in Microelectronics, pp. 823–826. DOI: `10.1016/j.microrel.2004.11.039`.

[129] J. Valasek. "Piezo-Electric Activity of Rochelle Salt under Various Conditions". In: *Phys. Rev.* 19 (5 May 1922), pp. 478–491. DOI: `10.1103/PhysRev.19.478`. URL: `https://link.aps.org/doi/10.1103/PhysRev.19.478`.

[130] J. Valasek. "Piezo-Electric and Allied Phenomena in Rochelle Salt". In: *Phys. Rev.* 17 (4 Apr. 1921), pp. 475–481. DOI: `10.1103/PhysRev.17.475`. URL: `https://link.aps.org/doi/10.1103/PhysRev.17.475`.

[131] *Von Neumann Architecture*. Jan. 2021. URL: `https://semiengineering.com/knowledge_centers/compute-architectures/von-neumann-architecture/`.

[132] J. Wang, H. P. Li, and R. Stevens. "Hafnia and hafnia-toughened ceramics". In: *Journal of Materials Science* 27.20 (1992), pp. 5397–5430. DOI: `10.1007/bf00541601`.

[133] H.-S. Philip Wong and Sayeef Salahuddin. "Memory leads the way to better computing". In: *Nature Nanotechnology* 10.3 (2015), pp. 191–194. DOI: `10.1038/nnano.2015.29`.

[134] J. C. Wong and S. Salahuddin. "Negative Capacitance Transistors". In: *Proceedings of the IEEE* 107.1 (2019), pp. 49–62. DOI: `10.1109/JPROC.2018.2884518`.

[135] *World's smallest DRAM cell for low-power memory in future mobile devices*. Sept. 2019. URL: `https://www.ibm.com/blogs/research/2019/09/worlds-smallest-dram/`.

[136] T. F. Wu et al. "Hyperdimensional Computing Exploiting Carbon Nanotube FETs, Resistive RAM, and Their Monolithic 3D Integration". In: *IEEE Journal of Solid-State Circuits* 53.11 (2018), pp. 3183–3196. DOI: `10.1109/JSSC.2018.2870560`.

[137] Wenwu Xiao et al. "Memory Window and Endurance Improvement of Hf0.5Zr0.5O2-Based FeFETs with ZrO2 Seed Layers Characterized by Fast Voltage Pulse Measurements". In: *Nanoscale Research Letters* 14.1 (2019). DOI: `10.1186/s11671-019-3063-2`.

[138] Xunzhao Yin et al. "Exploiting ferroelectric FETs for low-power non-volatile logic-in-memory circuits". In: *2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. Nov. 2016, pp. 1–8. DOI: `10.1145/2966986.2967037`.

[139] Rui Yang et al. "Ternary content-addressable memory with MoS2 transistors for massively parallel data search". In: *Nature Electronics* 2.3 (2019), pp. 108–114. ISSN: 2520-1131. DOI: 10.1038/s41928-019-0220-7. URL: https://doi.org/10.1038/s41928-019-0220-7.

[140] Masatomo Yashima and Shin Tsunekawa. "Structures and the oxygen deficiency of tetragonal and monoclinic zirconium oxide nanoparticles". In: *Acta Crystallographica Section B* 62.1 (Feb. 2006), pp. 161–164. DOI: 10.1107/S0108768105030570.

[141] Xunzhao Yin, Michael niemier, and X. Sharon Hu. "Design and Benchmarking of Ferroelectric FET Based TCAM". In: *Proceedings of the Conference on Design, Automation & Test in Europe*. DATE '17. Lausanne, Switzerland: European Design and Automation Association, 2017, pp. 1448–1453. URL: http://dl.acm.org/citation.cfm?id=3130379.3130719.

[142] Alex Yoon. *Tech Brief: ABCs of New Memory*. Apr. 2018. URL: https://blog.lamresearch.com/tech-brief-abcs-of-new-memory/.

[143] Alex Yoon. *Tech Brief: Much Ado About Memory*. Jan. 2018. URL: https://blog.lamresearch.com/tech-brief-much-ado-about-memory/.

[144] S. Yu and P. Chen. "Emerging Memory Technologies: Recent Trends and Prospects". In: *IEEE Solid-State Circuits Magazine* 8.2 (2016), pp. 43–56. DOI: 10.1109/MSSC.2016.2546199.

[145] Yongsik Yu. *Tech Brief: Memory "Grows Up" with 3D NAND*. Apr. 2016. URL: https://blog.lamresearch.com/tech-brief-memory-grows-up-with-3d-nand/.

[146] Ekaterina Yurchuk et al. "Charge-Trapping Phenomena in HfO2-Based FeFET-Type Nonvolatile Memories". In: *IEEE Transactions on Electron Devices* 63.9 (2016), pp. 3501–3507. DOI: 10.1109/ted.2016.2588439.

[147] Ekaterina Yurchuk et al. "Impact of layer thickness on the ferroelectric behaviour of silicon doped hafnium oxide thin films". In: *Thin Solid Films* 533 (2013), pp. 88–92. DOI: 10.1016/j.tsf.2012.11.125.

[148] Ekaterina Yurchuk et al. "Origin of the endurance degradation in the novel HfO2-based 1T ferroelectric non-volatile memories". In: *2014 IEEE International Reliability Physics Symposium* (2014). DOI: 10.1109/irps.2014.6860603.

[149] Furqan Zahoor, Tun Zainal Azni Zulkifli, and Farooq Ahmad Khanday. "Resistive Random Access Memory (RRAM): an Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (mlc) Storage, Modeling, and Applications". In: *Nanoscale Research Letters* 15.1 (2020). DOI: 10.1186/s11671-020-03299-9.

[150] Binjian Zeng et al. "Program/Erase Cycling Degradation Mechanism of HfO2-Based FeFET Memory Devices". In: *IEEE Electron Device Letters* 40.5 (2019), pp. 710–713. DOI: 10.1109/led.2019.2908084.

[151] Dayu Zhou et al. "Wake-up effects in Si-doped hafnium oxide ferroelectric thin films". In: *Applied Physics Letters* 103.19 (2013), p. 192904. DOI: `10.1063/1.4829064`. eprint: `https://doi.org/10.1063/1.4829064`.

[152] H. Zhou et al. "Negative Capacitance, n-Channel, Si FinFETs: Bi-directional Sub-60 mV/dec, Negative DIBL, Negative Differential Resistance and Improved Short Channel Effect". In: *2018 IEEE Symposium on VLSI Technology*. 2018, pp. 53–54. DOI: `10.1109/VLSIT.2018.8510691`.

Appendix A

Basic Silicon on Insulator FeFET Fabrication Flow

The baseline process flow to realize the SOI FeFET is detailed in this appendix. All FeFETs fabricated and characterized in this dissertation are fabricated according to this process flow. The various process steps are most flexible for accommodating new gate oxide designs. The vast majority of the process steps are realized at the University of California, Berkeley Marvell Nanofabrication Laboratory.

Deposition	Etching	Cleaning	Annealing/ Oxidation	Lithography
mrc944	lam6	msink6	rtp1	asml300
oxford2	lam7	msink7	rtp8	picotrack1
cambridge	lam8	msink8	tystar1	picotrack2
				axcelis
				matrix
				technics-c

Table A.1: Minimum suggested list of tools in the Berkeley Nanolab required to fabricate silicon on insulator FeFETs. Backup tools are also provided in the proceeding subsections as necessary.

Cartoon schematics of the FeFET structure are provided in Figures 4.4(a), Figure 5.15, Figure 5.5(a), and Figure 6.6(a). The reader is strongly recommended to consult process and equipment staff at the Berkeley Nanolab for additional assistance and for the most current understanding of tool conditions.

Active Device Definition

1. **Starting wafer information**

   - *Wafer type:* 150 mm (6″), SOI (silicon on insulator)
   - *Device layer doping:* p-type, Boron, $14 - 22$ Ω·cm
   - *Wafer orientation*: (1-0-0)$\pm0.5°$
   - *Material thicknesses*: 725 ± 25 μm handle; 0.3 μm$\pm0.5\%$ SiO_2 buried oxide; $100\pm$ 20 nm device top layer

   Prior to fabrication, it is strongly recommended to use a carbon tip scribe to label the wafer on the frontside, preferably along the edge of the flat, away from regions which will be patterned. Do **NOT** scribe the backside of the wafers as some tools require vacuum to stabilize wafers during processing and/or use helium cooling.

2. **Pre-furnace Cleaning**

   Tool required: msink6

   a) Piranha Clean

      - chemical: piranha (H_2SO_4, H_2O_2)
      - temperature: 120 °C
      - duration: 10 minutes

   b) HF (hydrofluoric acid) dip

      - chemical: 1:10 HF:H_2O
      - temperature: room temperature
      - duration: 1 minute

3. **Device Layer Thinning**

   Tool required: tystar1 (alternative: tystar2)

   - recipe: 1DRYOXA
   - gas ambient: O_2 gas
   - temperature: 1050 °C
   - duration: 1 hour, 45 minutes

   This recipe is calibrated to oxidize roughly 65-70 nm of silicon, depending on tool conditions.

4. **Removal of Oxide**

   Tool required: msink6

- chemical: 1:10 HF:H_2O
- temperature: room temperature
- duration: $\sim$ 17 minutes

5. **Photoresist Coating**

   Tool required: picotrack1 (alternative: svgcoat6)

   - recipe: T1_UV210-0.6_0.87μm (keep all default settings)
   - resist type: UV210, thickness ~ 0.87 μm

   Run at least one dummy wafer prior to running real wafers to ensure no photoresist streaking occurs.

6. **PM Alignment Mark Patterning**

   Tool required: asml300

   - job file: SANGWAN $\rightarrow$ swkim
   - recticle: COMBI 45440204D176
   - field: PM layer
   - energy: 18 mJ

7. **Photoresist Developing**

   Tool required: picotrack2 (alternative: svgdev6)

   - recipe: T2_PEB130C90s_MF26A45s (keep all default settings)
   - developer type: MF26A

   For thicker layers, it is advised to run the development recipe twice to ensure photoresist is thoroughly removed.

8. **Photoresist UV Hard Baking**

   Tool required: axcelis

   - recipe: U (keep all default settings)
   - temperature: 130 °C

9. **PM Mark Etching**

   a) Active silicon etch

   Tools required: lam8 (alternative: sts2)

   - recipe: 8001_POLY_ME

- gas flows: 150 sccm HBr, 50 sccm Cl_2
- TCP RF: 300 W
- bias RF: 150 W
- etch time: $\sim$ 20 seconds

b) BOX (buried oxide) etch

Tools required: lam6 (alternative: centura-mxp, sts-oxide)

- recipe: 6001_OXIDE_ME
- gas flows: 150 sccm Ar, 25 sccm CHF_3, 25 sccm CF_4
- power: 500 W
- etch time: $\sim$ 30 seconds

It is suggested to tune the silicon/oxide etches to ensure that the PM marks are etched to a depth of at least $1200\mathring{A}$ in the substrate for adequate contrast, so that the subsequent asml300 autoalignment is successful.

10. **Photoresist Removal**

Tool required: matrix (alternative: technics-c)

- power: 400 W
- temperature: 250 °C
- time: 2 minutes, 30 seconds

11. **Photoresist Coating**

Tool required: picotrack1 (alternative: svgcoat6)

- recipe: T1_UV210-0.3_0.43μm (keep all default settings)
- resist type: UV210, thickness ~ 0.4 μm

Run at least one dummy wafer prior to running real wafers to ensure no photoresist streaking occurs.

12. **Active Device Area Patterning**

Tool required: asml300

- job file: SANGWAN $\rightarrow$ swkim
- reticle: AVA-EXT519
- field: layer #1
- energy: 16 mJ

13. **Photoresist Developing**

Tool required: picotrack2 (alternative: svgdev6)

- recipe: T2_PEB130C90s_MF26A45s (keep all default settings)
- developer type: MF26A

14. **Photoresist UV Hard Baking**

    Tool required: axcelis

    - recipe: U (keep all default settings)
    - temperature: 130 °C

15. **Active Area Etching**

    Tools required: lam8 (alternative: sts2)

    - recipe: 8001_POLY_ME
    - gas flows: 150 sccm HBr, 50 sccm Cl_2
    - TCP RF: 300 W
    - bias RF: 150 W
    - etch time: $\sim$ 14 seconds

    Etch recipe should be calibrated to etch about 30-34 nm of silicon.

16. **Photoresist Removal**

    Tool required: matrix (alternative: technics-c)

    - power: 400 W
    - temperature: 250 °C
    - time: 2 minutes, 30 seconds

17. **Post-PR Removal Cleaning**

    Tool required: msink8

    a) Piranha Clean
       - chemical: piranha (H_2SO_4, H_2O_2)
       - temperature: 120 °C
       - duration: 10 minutes

    b) HF (hydrofluoric acid) dip
       - chemical: 1:25 HF:H_2O
       - temperature: room temperature
       - duration: 10 seconds

    As the buried oxide is now exposed at this step, it is important not to allow the HF clean to undercut the defined silicon channel substantially.

Gate Stack Deposition / Gate Definition

18. **Pre-furnace Cleaning**

    Tool required: msink6

    a) Piranha Clean

    - chemical: piranha (H_2SO_4, H_2O_2)
    - temperature: 120 °C
    - duration: 10 minutes

    b) HF (hydrofluoric acid) dip

    - chemical: 1:10 HF:H_2O
    - temperature: room temperature
    - duration: 10 seconds

    As the buried oxide is now exposed at this step, it is important not to allow the HF clean to undercut the defined silicon channel substantially.

19. **OPTIONAL: Chemical SiO$_2$ Regrowth**

    Tool required: msink7

    - chemical: SC1 (1:1:10 of H_2O_2:NH_4OH:H_2O)
    - temperature: 80 °C
    - duration: 10 minutes

    The concentration of SC1, temperature, and duration are chosen to achieve $\sim 8\overset{\circ}{A}$ of SiO$_2$. The reader is encouraged to test different concentrations, temperatures and durations to tailor chemical oxide thickness as desired. If working with a thermally grown IL, this step should be skipped.

20. **OPTIONAL: Nitrided IL Growth**

    Tool required: rtp8

    - gas ambient: Ar/NH_3 (9:1 ratio)
    - temperature: 850 °C
    - duration: 1 second

    This process step realized a high-κ IL with $\kappa \approx 7.8$ and results as reported in Section 6.2 of Chapter 6. If working with a chemically grown IL, this step should be skipped.

21. **Atomic Layer Deposition of Ferroelectric HfO$_2$**

    Tool required: cambridge

- gas ambient: Ar
- oxidant: H_2O
- precursors: TDMAHf, TDMAZr
- precursor temperatures: 75 °C
- cycles: 45, ratio of 4:1 Hf:Zr
- process temperature: 250 °C

Note: a lab group-owned, process specific ALD tool (Ultratech/Cambridge Fiji G2) was used for the growth of ferroelectric HfO_2. The process can be easily emulated using the cambridge ALD tool in the Marvell Nanolab using the conditions above.

22. **Sputtering of Gate Metal**

Tool required: mrc944

- metal: tungsten (W)
- power: 0.9 kW
- process pressure: 8 mTorr
- chamber pressure: 8×10^{-7} Torr
- passes: 9 (achieves $\sim$ 60 nm)

23. **Photoresist Coating**

Tool required: picotrack1 (alternative: svgcoat6)

- recipe: T1_UV210-0.3_0.36μm (keep all default settings)
- resist type: UV210, thickness ~ 0.3 μm

Run at least one dummy wafer prior to running real wafers to ensure no photoresist streaking occurs.

24. **Gate Area Patterning**

Tool required: asml300

- job file: SANGWAN $\rightarrow$ swkim
- recticle: AVA-EXT519
- field: layer #2
- energy: 18 mJ

25. **Photoresist Developing**

Tool required: picotrack2 (alternative: svgdev6)

- recipe: T2_PEB130C90s_MF26A45s (keep all default settings)
- developer type: MF26A

26. **Photoresist UV Hard Baking**

    Tool required: axcelis

    - recipe: U (keep all default settings)
    - temperature: 130 °C

27. **Gate Area Etching**

    Tool required: lam7 (alternative: matrix-etch, centura-met)

    - recipe: Ava_W_SF6
    - gas flows: 50 sccm SF_6, 10 sccm O_2
    - TCP RF: 125 W
    - bias RF: 60 W
    - etch time: $\sim$ 24 seconds

    Etch recipe should be calibrated to etch about 60 nm of tungsten.

28. **Source / Drain Implantation**

    Implantation is completed by INNOViON Corporation, based in San Jose, California.

    - doping species: $As(75)^+$
    - total dose: 1×10^{15}
    - energy: 20 keV
    - tilt: 7 °

29. **Photoresist Removal**

    Tool required: matrix (alternative: technics-c)

    - power: 400 W
    - temperature: 250 °C
    - time: 2 minutes, 30 seconds

30. **Ferroelectric Crystallization Anneal / Dopant Anneal**

    Tool required: rtp1 (alternatives: rtp3, rtp4)

    - gas ambient: N_2
    - gas flow: 35-40 sccm
    - temperature: 475°C
    - duration: 40 seconds

Back End of Line Passivation

31. **PECVD SiO$_2$ Isolation**

    Tool required: oxford2 (alternative: oxfordpecvd4)

    - recipe: ava_200c.rec
    - gas flows: 800 sccm N$_2$O, 100 sccm 10%SiH$_4$/Ar
    - forward power: 25 W
    - temperature: 200 °C
    - duration: 5 minutes

    Run at least one dummy wafer prior to running real wafers to check for showerhead patterns. Recipe duration should be calibrated to give ~ 300 nm of SiO$_2$.

32. **Photoresist Coating**

    Tool required: picotrack1 (alternative: svgcoat6)

    - recipe: T1_UV210-0.6_0.87μm (keep all default settings)
    - resist type: UV210, thickness ~ 0.87 μm

    Run at least one dummy wafer prior to running real wafers to ensure no photoresist streaking occurs.

33. **Contact Hole Patterning**

    Tool required: asml300

    - job file: SANGWAN $\rightarrow$ swkim
    - recticle: AVA-EXT519
    - field: layer #3
    - energy: 28 mJ

34. **Photoresist Developing**

    Tool required: picotrack2 (alternative: svgdev6)

    - recipe: T2_PEB130C90s_MF26A45s (keep all default settings)
    - developer type: MF26A

    For thicker layers, it is advised to run the development recipe twice to ensure photoresist is thoroughly removed.

35. **Photoresist UV Hard Baking**

    Tool required: axcelis

- recipe: U (keep all default settings)
- temperature: 130 °C

36. **SiO$_2$ Contact Hole Etch**

    Tools required: lam6 (alternative: centura-mxp, sts-oxide)

    - recipe: 6001_OXIDE_ME
    - gas flows: 150 sccm Ar, 25 sccm CHF$_3$, 25 sccm CF$_4$
    - power: 500 W
    - etch time: $\sim$ 65 seconds

    Etch recipe should be calibrated to slightly overetch 300 nm of SiO$_2$.

37. **Photoresist Removal**

    Tool required: matrix (alternative: technics-c)

    - power: 400 W
    - temperature: 250 °C
    - time: 2 minutes, 30 seconds

38. **Oxide Breakthrough Etch**

    Tools required: lam8 (alternative: sts2)

    - recipe: 8003_POLY_OB
    - gas flows: 100 sccm CF$_4$
    - TCP RF: 200 W
    - bias RF: 40 W
    - etch time: $\sim$ 7 seconds

    <u>*WARNING:*</u> This breakthrough etch step is critical to remove any oxide left in the contact hole trenches to ensure a better contact and should not be skipped. Furthermore, the wait time between this step and the contact metallization step should be minimized ($<$ 10 minutes).

Contact Metallization

39. **Sputtering of Contact Metal Stack**

    Tool required: mrc944

    a) Titanium sputtering

- metal: titanium (Ti)
- power: 2 kW
- process pressure: 8 mTorr
- chamber pressure: 8×10^{-7} Torr
- passes: 8 (achieves $\sim$ 60 nm)

b) Titanium nitride sputtering

- metal: titanium nitride (Ti/TiN)
- power: 2 kW
- process pressure: 8 mTorr
- chamber pressure: 8×10^{-7} Torr
- gas partial pressure: 25% N_2
- passes: 20 (achieves $\sim$ 120 nm)

Set up a recipe to perform the Ti/TiN sputtering in sequence.

40. **Photoresist Coating**

Tool required: picotrack1 (alternative: svgcoat6)

- recipe: T1_UV210-0.6_0.87μm (keep all default settings)
- resist type: UV210, thickness ~ 0.87 μm

Run at least one dummy wafer prior to running real wafers to ensure no photoresist streaking occurs.

41. **Contact Metal Patterning**

Tool required: asml300

- job file: SANGWAN $\rightarrow$ swkim
- recticle: AVA-EXT519
- field: layer #4
- energy: 28 mJ

42. **Photoresist Developing**

Tool required: picotrack2 (alternative: svgdev6)

- recipe: T2_PEB130C90s_MF26A45s (keep all default settings)
- developer type: MF26A

43. **Photoresist UV Hard Baking**

Tool required: axcelis

- recipe: U (keep all default settings)
- temperature: 130 °C

44. **Contact Metal Etching**

   Tool required: lam7 (alternative: matrix-etch, centura-met)

   - recipe: 7001_AL_ME
   - gas flows: 90 sccm Cl_2, 45 sccm BCl_3
   - TCP RF: 800 W
   - bias RF: 100 W
   - etch time: $\sim$ 45 seconds

   Etch recipe should be calibrated to etch through the stack of $\sim$ 60 nm Ti/$\sim$ 120 nm TiN. It is suggested to monitor the endpoint signal during this etch step, as a clear endpoint on both channels A and B should be visible when the etch has broken through each layer of metal.

45. **Photoresist Removal**

   Tool required: matrix (alternative: technics-c)

   - power: 400 W
   - temperature: 250 °C
   - time: 2 minutes, 30 seconds

   Figure A.1 is a schematic cartoon demonstrating the expected 2D cross-sections at various steps during the overall process flow.
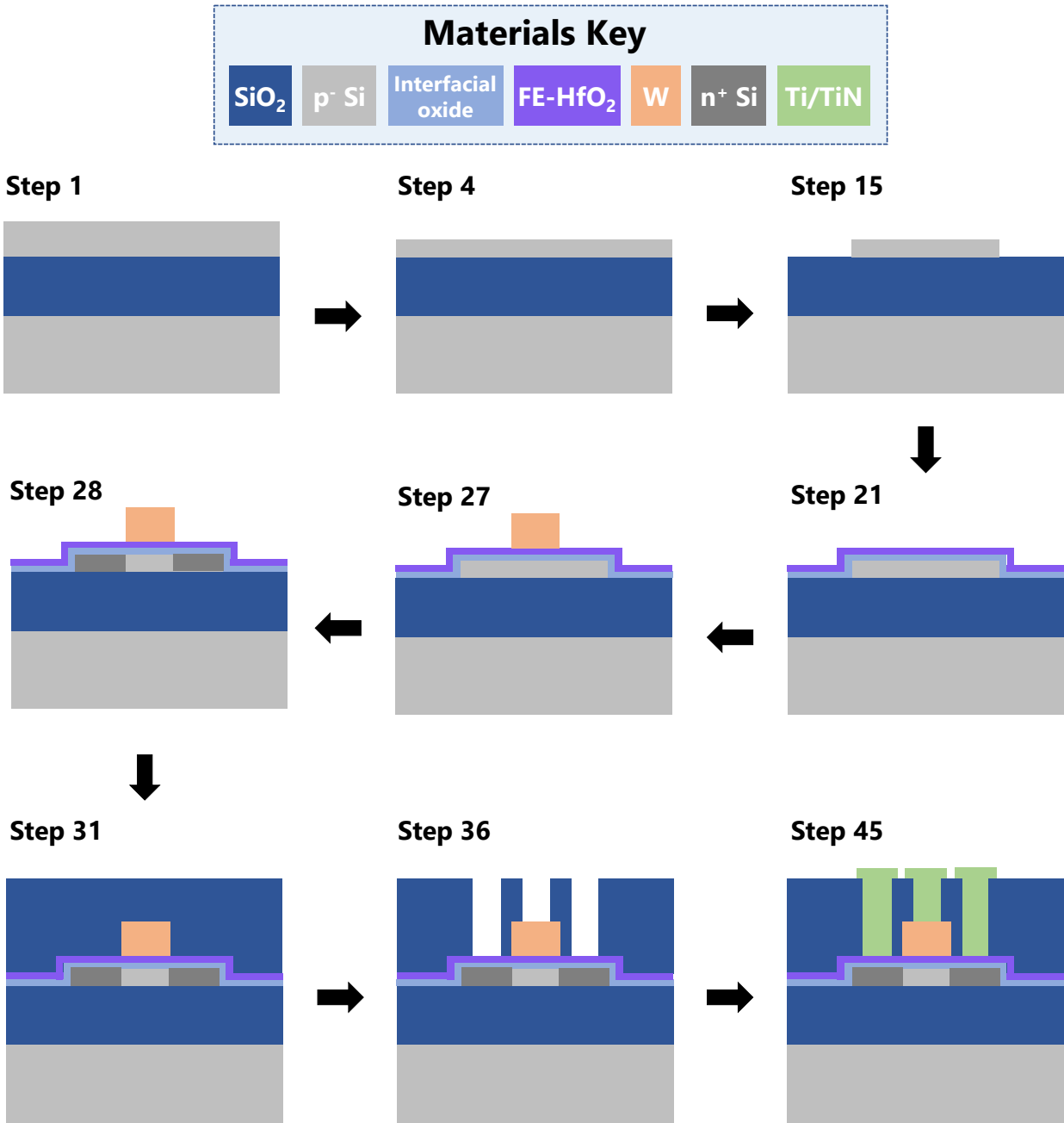
Figure A.1: 2D cross-sectional cartoons of the FeFET process flow at various steps, as described and numbered in this Appendix.

Additional Cautionary Process Notes

The process steps 1 - 45 above provide reasonable starting guidelines to realize similar Fe-FET structures as characterized in this dissertation. Additional cautionary considerations are provided in this subsection to help diagnose common processing challenges encountered at various steps in the process flow, based on personal experience.

Chemical Freshness at Wafer Cleaning Sinks

At various points in the process flow, it is mandatory to clean the wafers at one of three dedicated sinks in the Marvell Nanolab: msink6, msink7, and msink8. The hydrofluoric acid and piranha baths at msink6 and msink8 are *general purpose baths*, used by all members of the Nanolab for MOS clean and non-MOS clean process, respectively. The baths are periodically replenished/drained/supplied with additional chemicals by the staff.

During the wafer cleaning steps, it is important to monitor the temperature of the piranha bath immediately after adding hydrogen peroxide to activate it. A freshly replenished piranha bath will quickly spike in temperature from 120 °C to around 130 °C due to the highly exothermic reaction from chemicals mixing. However, a bath with old chemical will respond sluggishly, and this is a fair indicator that it is time to replenish the bath in order to achieve maximal cleaning efficiency of the organic compounds on the wafer.

It is also important to monitor the wafer after the HF dipping step. A dip in an old HF bath may leave the backside or frontside of a standard silicon wafer still slightly hydrophilic (the surface is hydrophobic when acid/water minimally "beads" up on the wafer surface; when it is hydrophilic, continuous regions of moisture will still be present on the surface). The wafer surface must be totally free of native oxide prior to gate stack deposition, and therefore be completely hydrophobic.

Active/Gate Lithography Exposures

The asml300 lithography tool typically guarantees good resolution of feature sizes down to 0.25 μm. However, on occasion, one may find that particularly for the active area or gate finger lithography steps, the smallest feature sizes may not resolve correctly, leading to features which are not perfectly straight (and in worst-case scenarios, features with high length to width aspect ratios appearing visibly bent under a microscope) after development. To avoid this, it is recommended to run a focus-exposure matrix prior to the real photolithography run to assess the exact best focus and energy conditions for exposure at that moment in time.

Wait Times between Deposition Steps

Timing between steps is critical during the various depositions involved to realize the gate stack of the FeFET. It is firstly crucial to minimize the wait time after interfacial layer growth (either using SC1 for SiO_2, or thermal nitridation for SiN_x) and prior to ALD of the

ferroelectric HfO_2 layer (< 20 minutes). If transferring the wafer out of the cleanroom and to a separate facility for oxide growth, then it is strongly recommended to use a vacuum desiccator container to carry the wafer during the transfer process.

Similarly, it is advised to minimize the wait time after ALD of the ferroelectric HfO_2 layer and prior to gate metal deposition (< 20 minutes). Exposure to ambient conditions will cause undue contamination of the various oxide/oxide/metal interfaces, adversely impacting device performance.

ILD / Isolation SiO_2 Deposition Temperature

If following the process flow above to realize gate-first FeFETs, it is important to note that the deposition temperature for the PECVD ILD SiO_2 oxide in Step 31 is set to 200 °C. Most PECVD oxide deposition recipes run at 350 - 400 °C; however, if the ferroelectric oxide used is Zr-doped HfO_2, the film will not be able to withstand prolonged exposure at such a high temperature. It was confirmed that a PECVD oxide recipe run at temperatures in the ranges of 350 - 400 °C will cause additional unwanted crystallization of the Zr-doped HfO_2 layer, additional charge trapping through the ferroelectric layer itself (which counteracts ferroelectric hysteresis), and degraded subthreshold swing in fabricated devices.

ILD Contact Hole Etching

Etching through $\sim 250 - 300$ nm of the SiO_2 ILD is a necessary step in making contact to the source/drain/gate regions of the device. This is also a critical processing step which should be undertaken carefully. If the etch stops short of making contact to the source/drain silicon regions or the metal gate, then it will be impossible to probe the device afterwards (and this mistake is extremely difficult to correct retroactively). If the etch goes through both the active silicon layer and the buried oxide SiO_2, therefore making contact with the silicon handle substrate, then it is possible that the only current that can be measured is the substrate current.

To avoid over or underetching, it is recommended to run blanket SiO_2 wafers in the etch tool (lam6, centura-mxp) prior to etching the real device wafer to first season the chamber and also calibrate the blanket film etch rate. Then, it is imperative to also run *patterned* dummy SiO_2 wafers which are patterned with the same mask layer used to pattern the contact holes. In most cases, the etch rate of the patterned SiO_2 wafers will differ from the blanket by a non-negligible amount, and will typically take *longer* to etch due to the comparatively larger volume of photoresist versus oxide being etched simultaneously. It is therefore critical to use the endpoint detection option on the etch tool to time the etch steps, and to use a film thickness measurement system (such as the nanospec or nanoduv) as a secondary checkpoint to endpoint detection to ensure the ILD oxide has been etched through and that the etch stops on the silicon active layer.