UNIVERSITE D'AIX-MARSEILLE

IM2NP

ECOLE DOCTORALE 353

CEA-Leti

Advanced Memory Lab

A Thesis submitted in fulfillment of the requirements
for the Degree of Doctor of Philosophy

Discipline: Electrical Engineering
Specialization: Micro and Nano electronics

Marios Barlas

Development and characterization of innovative nonvolatile OxRAM memory cells compatible with advanced nodes

Soutenue le JJ/MM/AAAA devant le jury :

Daniele Ielmini	Politecnico di Milano	Rapporteur
Nathalie Malbert	Université de Bordeau	Rapporteur
Jean-Michel Sallese	EPFL	Examinateur
Simon Jeannot	STMicroelectronics	Examinateur
Gérard Ghibaudo	Université de Grenoble-Alpes	Examinateur
Damien Deleruyelle	INSA Lyon	Examinateur
Laurent Grenouillet	CEA-Leti	Invité
Elisa Vianello	CEA-Leti	Invité
Jean-Michel Portal	Aix-Marseille Université	Directeur de thèse

Numéro national de thèse/suffixe local : 2017AIXM0001/001ED62

To all the people who made me who I am today

The journey of a thousand miles begins with one step.

-Lao Tzu

SOMMAIRE

Les mémoires resistives a base d ’oxydes de transition organométalliques (OTM) sont fortement étudiés comme un candidat pour des applications mémoires non volatiles. Il y a trois moteurs derrière la forte activité R&D sur les mémoires non volatiles : Tout d’abord, la demande sans cesse croissante de plus en plus de puissance de calcul pour les applications haut de gamme (HPC). C’est le moteur principal des semi-conducteurs « classiques », à savoir les mémoires CMOS, DRAM et NAND; mais à un rythme plus ralenti en raison de la mort de la loi de Moore. Deuxièmement, le passage de l’ère PC / mobile du début des années 2000 à l’ère de l’Internet des objets (IoT) conduisant à un matériel spécifique à l’application et à une éventuelle diversification des technologies possibles en fonction de l’application. L'IoT impose des exigences matérielles que les applications de mémoire classiques ne sont pas en mesure de fournir dans certaines applications, surtout si le coût est pris en compte. Troisièmement, la nécessité identifiée d'un changement perturbateur du paradigme informatique avec des domaines tels que l'intelligence artificielle (IA) et l'informatique neuromorphique, entraînant le besoin d'un nouveau matériel pour les synapses à neurones, auquel les ReRAM pourraient apporter une solution avec leur comportement de commutation analogique.

Dans le cadre de cette travail on examine d’une facon systematique

Mots clés :

Abstract

Transition Metal Oxide ReRAM is a class of non-volatile memory technologies where the switching between memory states is enabled by the reversible breakdown of the oxide by means of the creation and dissolution of a percolation path (filament). The main advantages of the technology lie in the scalability of the memory cell –mainly owed to the sub 10nm dimension of the filament, its low power consumption (< 300 pJ/ switch) and material compatibility to advanced CMOS. Nevertheless, there are two major roadblocks that have prevented so far the implementation of ReRAM in large arrays: First, the requirement for an initial breakdown happening voltages significantly higher than the operating voltage range and second, the intrinsic and extrinsic variability components arising from material interaction to its environment as well as the fundamental stochastic nature of percolative conduction.

This work, is focused on HfO_2 based ReRAM technology. In the first part, we investigate different dopants to engineer the conductive properties of HfO_2 by combining a first-principles approach and in-depth material characterization techniques. In the second part, the proposed HfSiOx alloy is integrated in the BEOL of a 130nm process and the impact of the integration of the switching zone in forming, switching, error rate evolution and data retention is investigated. In the last part, a HfO_2 based integration in the early MOL of an advanced FDSOI 300mm CMOS process is demonstrated investigating standard HfO_2 ReRAM performances and limitations.

Keywords : ReRAM, semiconductor process integration, CMOS, FDSOI, HfO_2, doping, Si, Al, defect engineering, endurance, data retention

Acknowledgements

Working in technology and device physics is certainly teamwork. It requires a great deal of patience for wafers to go through hundreds of physico-chemical steps a process that lasts for months and it's a great exercise for patience. And all that, after the initial idea has been theorized and you've tried your best to cover all bases. As a consequence, no work showing results on the level of a chip demonstrator is the work of a single person. My role in this has been to provide some guidelines and connect the dots between theory, device integration, intrinsic and extrinsic device physics. As a result, there's probably tens of people taking part in a smaller or larger degree to this journey.

First things first then, I should thank Elisa and Laurent for giving me the opportunity. I hope it's been a good 3-year long journey. In particular, I should thank them both for challenging my ideas and thoughts on the ReRAM that pushed me a step further every time. I'll thank Laurent in particular for his presence on an almost daily basis in my time in Leti be it to either encourage, question, or disapprove of my ideas and actions. To be horribly specific I particularly loved the "I x % agree with you" with x varying between 0 and 100 for every different case.

Then, there's Philippe Blaise. Through interacting with him, I got a good microscopic understanding of the material with a level of precision and exactness that I will try to translate in all of my following steps. At the same time, I should also thank boubacar Traoré for laying the first foundation in my work. His initial reflections on the HfO_2:Si system laid the groundwork for what we did next. I only think it's a bit of a pity that we didn't get the chance to work together really owed to the non-overlap of our time in Leti.

From a material aspect, I should really thank Emmanuel Nolot for providing me with a fearsome amount of measurements in XPS and XRR and also for training me on the spectral analysis techniques and letting me free to find my way around the data from that point. Moreover, all the clean room process engineers that worked in developing steps of my projects deserve their merit as well.

Getting closer to the end, I'll thank my friend and great colleague, Alessandro Grossi for being a super fun partner in all that was analyzing data on the ReRAM arrays and pinging ideas over lunch for the better part of a year! In the same basket go the friends I made

through my time in Leti like Thilo, Julia and Giuseppe that made the three years I spent there a good amount of fun. It’s truly hard to find colleagues that can also be good friends in life as well, in my opinion anyways.

An Introduction of Sorts

« La semplicità è l'ultima sofisticazione ». Simplicity is the ultimate form of sophistication -according to Leonardo da Vinci. This is maybe the most fitting place to start when it comes to ReRAMs. It is impressive that a concept so simple as breaking down a dielectric "sandwitch" could lead to so many possible applications: Over the years, the nanoionics community and designers have theorized -and even more so demonstrated on the product level- around the possible applications of ReRAMs: memory, neuromorphics, security, are just some of the applications that come to mind; there might be others that people haven't even thought of yet. Now the issue with geniuses, such as Da Vinci and genius ideas such as ReRAMs, is that usually there's a rabbit hole. It does not mean that if something is simple it is also simplistic. And this is where, albeit a great amount of research has been conducted in ReRAMs and nanoionics everyone is still yearning for that singularity that will allow the technology to soar. At the core of ReRAM lies dielectric breakdown, and breakdown is a whimsical physical process that remains elusive to this day despite all our best efforts to control it. After all, you cannot blame gravity for falling in love, as uncle Albert once said.

And so, back when I started working on ReRAMs in 2015 in the context of a master thesis, I wasn't really expecting that I was setting of in an almost four-year long journey that would essentially lead to a pursuit of restricting a percolation path. But now, I know that the journey of a thousand miles starts with a single step (Lao Tzu). And by now I'm pretty sure dear reader that you're getting tired of me quoting famous people. This work is all about pursuing something elaborate in something that is simple. We essentially tried to play billiards with atoms in a deterministic way to try and limit the possible path of dielectric breakdown that is, by nature, stochastic. In short, this research work is organized in five chapters.

In the first chapter, starting from the current semiconductor economic landscape I try to identify the current motors for ReRAM R&D are. The, I make an effort to summarize the work that I consider the most interesting so far in the field: In a bottom up approach, the types of ReRAM and different kinds of switching materials and their behavior are identified. A detailed review of the dielectric breakdown process from the microscopic level to the statistical laws governing the phenomenon. Key aspects of the physics of

the device in terms of material dependence and switching characteristics such as endurance and data retention are presented. From then on peripheral aspects of the technology such as architecture, selector devices and state of the art are briefly reviewed.

In chapter two, we move on to describe our material system of choice, HfO_2 and how doping in general, and the ion implantation process in particular, can change the system properties in a significant fashion; provided the right dopant is used. Notably the Al and Si implants are studied in detail. We combine material analysis with ab-initio calculation to link local molecular microstructure to the electronic properties of the doped material matrix and its impact on the ReRAM device.

Chapter three picks up the idea of doping and explores the impact of device geometry and confinement of the implanted volume in different device characteristics: From forming, to ReRAM operation to endurance and data retention. We exploit relatively large (for academic prototyping capacities) statistical ensembles of the devices fabricated, to understand not only the intrinsic device characteristics but also the behavior and control of extreme individuals of the population while trying to identify the eventual trade-offs and limitations of the approach. In all this, we try to combine material and device geometry to not only modulate ReRAM conduction but also control to an extent the extreme most events of the resistive memory population by dissociating the switching zone from the device environment.

Chapter four falls back into the standard HfO_2 technology, exploring a more elaborate integration strategy to essentially try and achieve similar effects: Stronger control of the extreme-most events. Even if this chapter has been fun from the technology point of view, it basically serves to demonstrate that the fundamental roadblock of ReRAM is not necessarily technology maturity, but the profound proneness to variability of not just the percolation path, but also of the materials involved in the ReRAM stack itself. As a byproduct, we also demonstrated compatibility of the ReRAM stack to GO1 (sub 1.5nm gate stack oxide) FDSOI gate-first nodes. Finally, we conclude this work by a general summary of results and some ideas for pushing the limits for the ReRAM device beyond the standard MIM capacitor and towards the front end of CMOS in chapter 5.

In a more personal aspect, this work is more my desire to build something that tries to bring something into the equation of the ReRAM and to convince myself even more than others that it works in a consistent, systematic and repeatable fashion –more to be able to sleep at night without being haunted by the phantoms of physics. There were times I had a good deal of fun during this process and others that things were on the highway to hell. Now that the deed is done, even though I didn't achieve Da Vinci's intended sophistication, I feel like I start to understand why you cannot blame gravity for falling in love.

TABLE OF CONTENTS

CHAPTER 1. FROM OXIDE BREAKDOWN TO OXRAM – A SHORT OVERVIEW

1.1 Introduction: The Memory Landscape

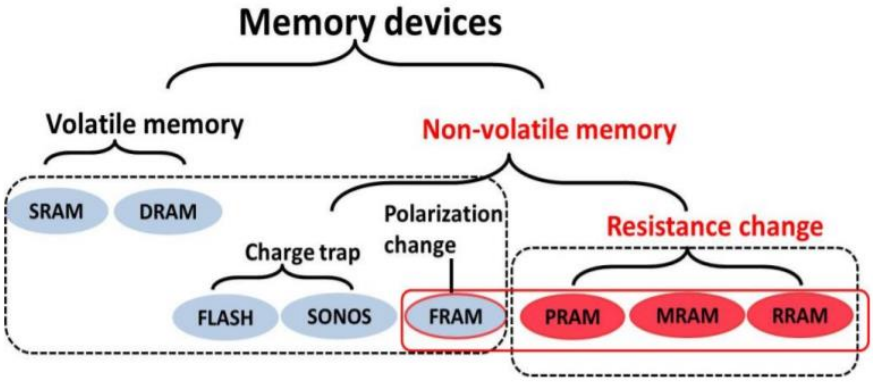


Figure 1-1 Classification of emerging NVM technologies

For the last 30 years when Moore's Law was the main drive in the field of microelectronics almost all high-speed, high-density memory applications were addressed through CMOS transistor derivatives based on a charge-trapping physical mechanism. SRAM remains the predominant choice for embedded high speed applications where non-volatility is not required. DRAM (volatility in the ~ms range) provide high memory bandwidth (> 100 GB/s) and low latency. Ever since it was invented in 1987 [1] FLASH has allowed virtually the only CMOS compatible non-volatile memory cell providing significant boosts in power reduction and density owed to Moore's law at every node giving birth to solid state storage class memory, constituting 43% of the semiconductor memory market in 2016 [2].

Nevertheless, 30 years of continuous scaling have inevitably led to approaching fundamental physical limitations. And while breakthroughs in 3D integration continue to drive DRAM and V-NAND in the sub 20nm nodes, they come at a steep price in process complexity, that has heavy impact in fabrication cost. This trend means that soon the manufacturing cost might render the use of advanced NAND prohibitive to certain classes of applications. As a result, the industry has started to explore different kinds of non-volatile memory based on completely different physical principles, known as resistive memories (Figure 1-1), where the memory state is associated to a resistance level in the cell.

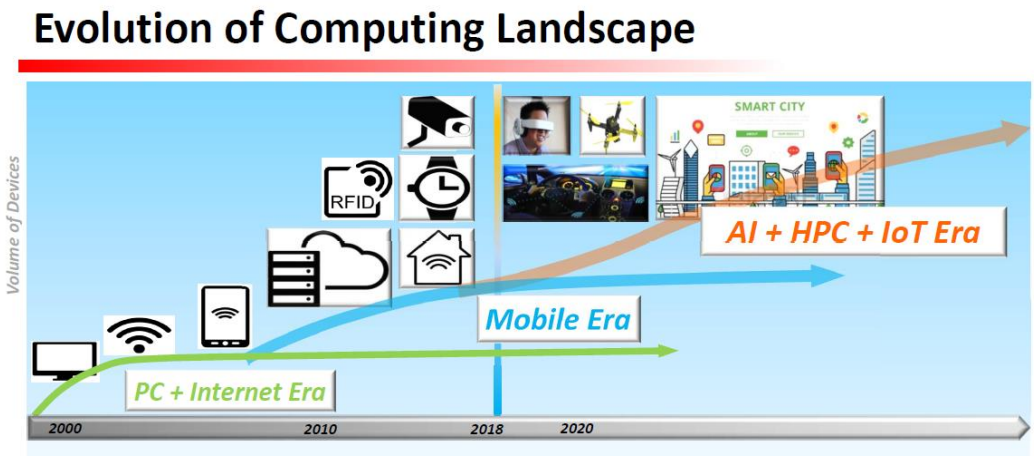


Figure 1-2 Applications driving computing roadmap [Short Course, IEDM 2018]

A number of different physical systems exhibit resistive switching behaviour. Phase-change memory (PCM) base their operation on the change of phase of the active material as a function applied voltage owed to Joule/electronic heating. Magnetic tunnel junctions (MTJ / STT-MRAM) resistance state changes with voltage are associated to the parallel and anti-parallel configuration of electronic spin between the active layers of the device [3]. The ReRAM switching mechanism is thought to be mainly owed to modulation of conductivity owed to the reversible dissolution of ion species transitioning to and from given oxidation states. Historically the latter have been differentiated between CB-RAM or Electrochemical Metallization Cell (ECM) and OxRAM or Valence change (VCM) if the conductive centres are attributed to metallic ion species or oxygen vacancies, respectively [4], [5].

All these classes are significantly different with respect to standard flash both in terms of operation principle and architecture. Contrary to charge-trapping NVMs the above *emerging* NVM technologies are two terminal devices based on Metal Insulator Metal (MIM) stacks. This comes with the distinct advantage of highly simplified integration schemes as will be discussed in further detail in the following sections. Even though NAND and DRAM are still the dominant technologies, emerging NVMs are starting to pick up momentum owed to the diversification of computing in the modern era.

Three key factors driving computing can be identified for the near and mid future: First, the ever-growing demand for more and more computational power for high-end applications (HPC). This is the main drive behind “classical” semiconductors namely

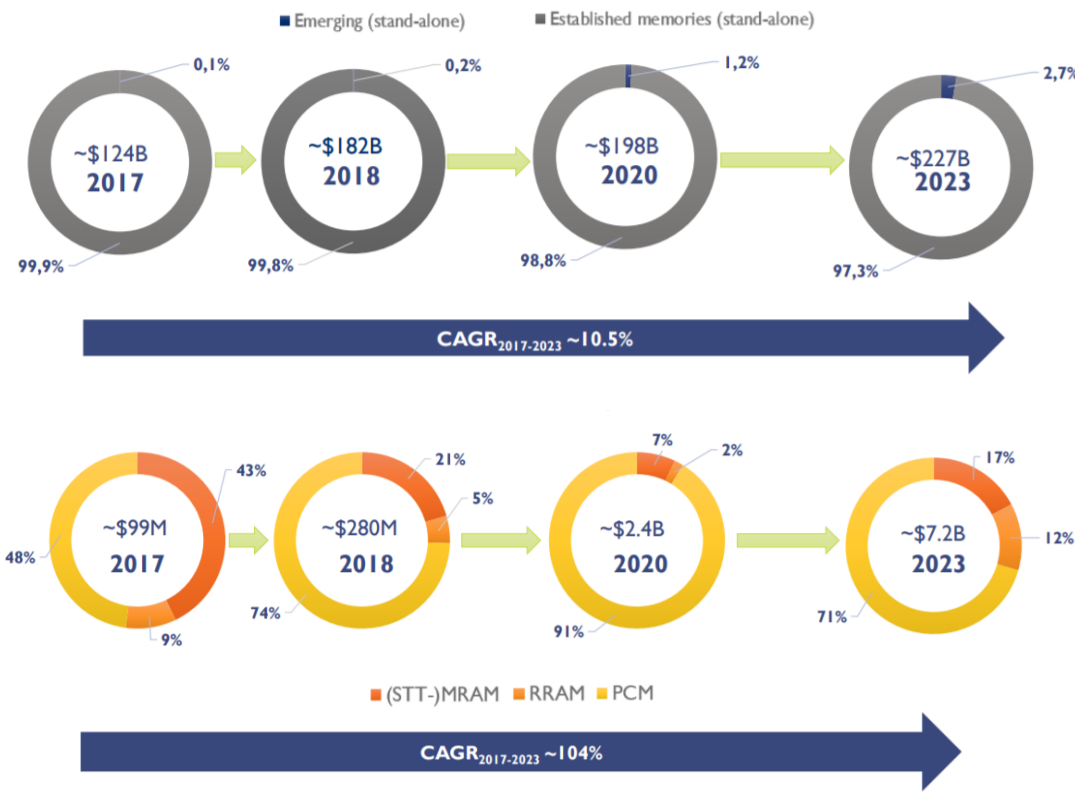


Figure 1-3 Predicted Memory Market Growth – Yole Development Group, reprinted with permission (CAGR for chiffre d' affaires growth)

CMOS, DRAM and NAND flash; albeit at a decelerating pace owed to the death of Moore's Law and the end of transistor scaling. Second, the passage from the PC/mobile era of the early 2000s to the era of the Internet of Things (IoT) (Figure 1-2) leading to application-specific hardware and an eventual diversification of the possible technologies according to the application. IoT imposes hardware requirements that classical memory applications are not able to deliver in some applications, especially if cost is taken into account. Third, the identified need for a disruptive change in the computing paradigm with fields like Artificial Intelligence (AI) and neuromorphic computing, giving rise to the need for new hardware for the neuron synapses, to which ReRAMs with their analog switching behaviour could be a solution [6]–[9].

All of these factors give rise to a growing market for new memory technologies. At present, the stand-alone memory market is projected to grow from a 124B in 2017 to ~227B in 2023. Interestingly, even though the memory sector is will still be dominated by traditional technologies, notably NAND flash and DRAM, it starts to slowly but surely give ground to resistive switching memories that, even though take less than 0.1% of the

market share nowadays, are projected to reach 2.7% in the next five years, generating a new market worth more than 7B $ by 2023 (Figure 1-3).

Out of the different competing resistive switching technologies three are leading the emerging NVM market evolution, namely PCM, STT-MRAM and ReRAM. Between the three, PCM is considered the current leader with more than 15 years of R&D, driven by major players like Intel & Micron with 3D X-Point technology in Optane chips for mainstream applications, but also by STMicroelectronics oriented more towards the microcontroller market [10], and are based on GST (Ge-Se-Te) compounds.

STT-MRAM starts to appear as a PCM contender promoted mainly by players like TSMC and Globalfoundries and promising high switching speed and almost unlimited endurance [11]. Nevertheless, its sensitivity to temperature has proven to be much stronger of a challenge than initially expected, with Samsung reporting early product-ready maturity only at the beginning of 2019 [12].

Resistive memory devices based on transition metal oxides (OxRAM) base their prospect insertion into the emerging NVM market owed to three main drivers: First, their physical principle allows for the NVM memory window to be either insensitive to scaling or even be expected to augment as the cell dimension is reduced (such is the case of non-filamentary switching mechanism) (Figure 1-4). As a result, it might be the only NVM type amongst the emerging technologies that can match density large enough to contend its more mature NAND and DRAM candidates. Second, the simplicity of the cell which promises to deliver high density at low cost. Third, their analog switching behaviour and memristive properties [13]–[16] that has recently attracted a lot of attention as a potential element to implement neuromorphic chips. The main roadblocks for ReRAM are their requirement for forming operation that increases chip design complexity for large scale implementations, their limited endurance as well as their sensitivity to intrinsic and extrinsic variability. Nevertheless, it is worth noting that, variability, albeit imposing problems for memory applications could be a benefit in disguise for neuromorphic chips. To date only Panasonic[17] and Fujitsu[18] have announced small-scale ReRAM chips in the market.

It is nowadays generally understood that the future of microelectronics will be diversified and application specific. Especially for NVM, there is no technology to solve all problems.

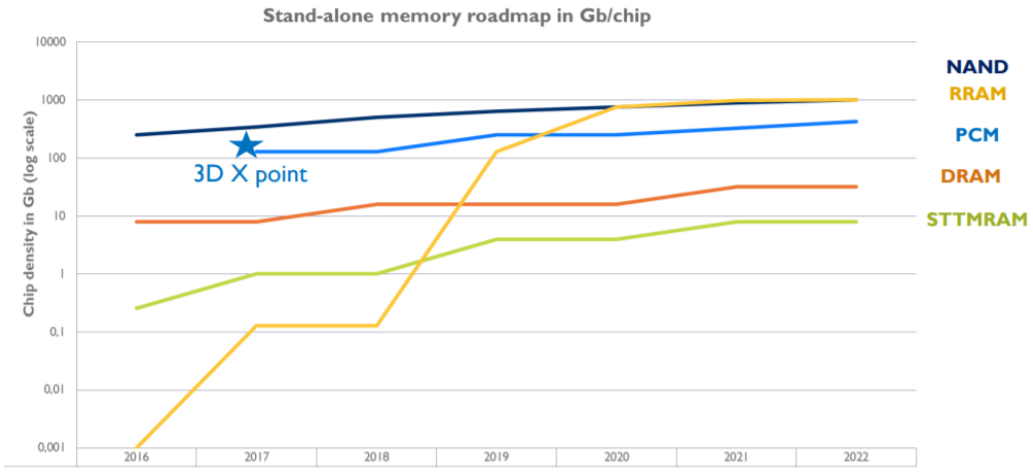


Figure 1-4 Predicted Stand-alone cell density evolution – Yole Development Group

As a result, ReRAM –which is the main focus of this work- faces the challenge and the opportunity to become the best candidate for some applications. Nevertheless, ionic manipulation, which lies at the heart of the ReRAM switching mechanism, proves a challenge bigger than the scientific community have expected as has been recently observed [19].

This work aspires to provide some solutions to key aspects towards industrialization of ReRAM. In this brief introduction, an effort was made to provide some context into the undercurrents influencing the development of the technology from an economic and application prospective. In the following section, we move to describe ReRAM in more detail. I strongly believe that, to understand ReRAM it is required to first understand the key aspects of conduction of oxides and dielectric breakdown; that after all lie at the heart of the ReRAM switching mechanism. For this reason, the next two sections detail conduction mechanisms in oxides and dielectric breakdown from a more general prospective. Then, the current theories about the physicochemical processes involved in the TMO ReRAM technology are detailed along with materials of interest and current cell and device architectures.

1.2 Some Basic notions of NVM and ReRAM

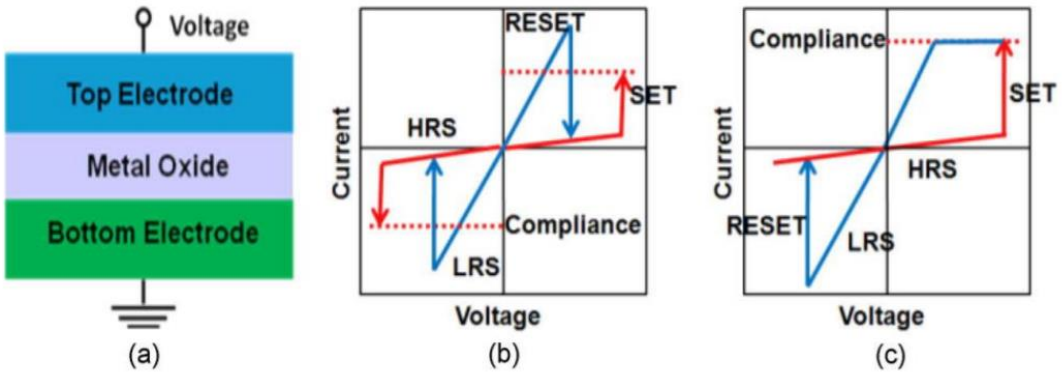


Figure 1-5 (a) MIM structure based ReRAM cell (b) Unipolar device and (c) bipolar device switching regimes respectively [4]

NVM cells can have two or more resistive states, namely a low (LRS) and a high (HRS) resistance states. Binary cells can oscillate between only two resistance levels while multi-level cells can have more than two stable resistive states, according to the programming conditions applied on the cell. In ReRAM, the as-fabricated material cell typically requires a first preliminary operation of controllable soft dielectric breakdown, commonly referred to as electroforming or forming process. Forming occurs by either applying a pulse ramp or a constant voltage pulse on one of the device electrodes until the device experiences a breakdown event.

After forming, the cell state can be reversed by applying appropriate programming conditions. The transition from HRS to LRS is referred to as the SET operation while the transition from LRS to HRS as the RESET operation. While the set voltage corresponds to the voltage at which the SET switch occurs, the reset voltage is not defined in the same straightforward manner. In most cases, the (dc) reset voltage term refers to the maximum voltage that is used during the reset phase. This discrepancy is mainly owed to the difficulty to correlate an easy-to-detect characteristic of the I-V curve to the reset operation, owed to the high variability associated to the reset phase.

ReRAM cells can exhibit either *unipolar* or *bipolar* switching. Unipolar switching occurs when the cell's resistance state can be reversed from LRS to HRS by applying a voltage of the same polarization configuration as the one used to induce the LRS state (Figure 1-5b). If the polarization in the device terminals needs to be reversed in order for the

transition from LRS to HRS to occur then the cell is characterized as bipolar (Figure 1-5c). Cell polarity is strongly dependent not just on the oxide used, but more specifically in the oxide/metal electrode combination, and will be discussed in further detail in the sections 1.3 and 1.4.

ReRAM cells can exhibit both surface-dependent and surface independent (filamentary) switching, depending on the material system and choice of electrodes. This work focuses on the TiN/HfO2/Ti system that demonstrates purely filamentary behaviour.

1.3 Materials in the ReRAM Cell and the impact in resistive switching

A very wide range of materials have been studied for resistive switching applications, including both oxides, chalcogenides as well as nitrides [20], [21]. Both binary and complex oxides such as SiO_2 [22], [23], TiO_2 [24]–[26], [27] , NiO [28] and $SrTiO_3$ [29], [30] can show resistive switching characteristics. An rather complete review of the different materials exhibiting resistive switching characteristics can be found in [31]. Nevertheless, out of the wide range of possible candidates studied over the years, the most promising in terms of memory performance have proven to be Al_2O_3[32], Ta_2O_5 and HfO_2 for which a large range of works can be found in the literature [4], [33]. The latter oxides (Ta_2O_5, HfO_2, and Al_2O_3) are particularly important for ReRAM because even more than their promising switching behavior are all widely used in CMOS technology and are thus easy to integrate in a semiconductor fabrication process.

At this point it is important to stress the sensitivity of the ReRAM switching behavior to the nature of the metallic electrodes. Both inactive and reactive metals have been used as ReRAM electrodes. For instance, the Pt/HfO_2/Pt system has been shown to exhibit unipolar switching behavior [34] while the Pt/HfO_2/TiN can be both unipolar and bipolar [35]. Using a reactive metal of column IV of the periodic table as a ReRAM electrode, such as Ti or Hf favors bipolar switching while at the same time, increasing the thickness of the scavenging layer acts to reduce the forming voltage of the cell [36], [37]. Nevertheless, this comes at the expense of impacting the data retention characteristics.

From the early stages of ReRAM, it has been understood that in order to tune ReRAM cell characteristics a systematic approach needs to be adopted in terms of material stack engineering as simple thickness scaling of either the oxide or the metallic electrode does not provide enough margin to tailor cell switching with enough liberty in order to

accommodate the needs of storage class memory. Engineering at the level of the top electrode has made it clear that the level of stoichiometry in the metal/oxide interface can add another degree of freedom in ReRAM technology design.

Bilayers or alternating nano-laminate stacks of binary oxides have been used to engineer stacks with preferential switching characteristics. The Ta_2O_5/TiO_2 system has been shown to exhibit superior endurance and resistance state uniformity characteristics [38] while HfO_2/Al_2O_3 have been shown to improve the data retention of the HfO_2 based ReRAM cell [39]–[41].

Doping TMOs to achieve improved switching characteristics has thus been heavily investigated; especially for HfO_2 [42]–[45] and Ta_2O_5 [46]. A wide variety of dopants has been proposed such as Ti, N, H, Si, Al, Mg and Ru targeting to modulate oxygen vacancy concentration and migration energies in the oxide matrix. Experimental results remain somewhat controversial however (see Al/Si doping in HfO_2 by implantation [45] or nanolaminates). Recently, Duncan et al. have proposed some theoretical guidelines for the behavior of dopants in HfO_2 [47]. Nevertheless, the creation of an alloy that can not only tailor forming and switching voltages but also allow to act in-situ in key aspects such as endurance, retention and variability is still a topic of open research in the ReRAM field.

1.4 The current understanding of the Physics of the OxRAM Cell

1.4.1 Electronic Conduction Mechanisms in oxides

From a band diagram prospective, oxides can be considered an extreme case of semiconductors. In both classes of materials, the valence and conduction bands are separated with a band gap, spanning from a few eV in classical semiconductors (Ge – 0.58 eV, Si 1.1 eV, SiC – 2.36 eV) to several eV in oxides (> 5 eV) [48].

It is important to notice that for any meaningful measurement of conduction a test structure is required, typically either an MIS (Metal Insulator Semiconductor) or a MIM (Metal Insulator Metal) contact. As material properties from one material to another differ, when different materials are brought in contact the energy bands are *bent* to satisfy thermodynamic equilibrium. The rate of band bending is dependent on the work function difference between the different materials in contact, which in turns affects the quantum mechanical barrier shape.

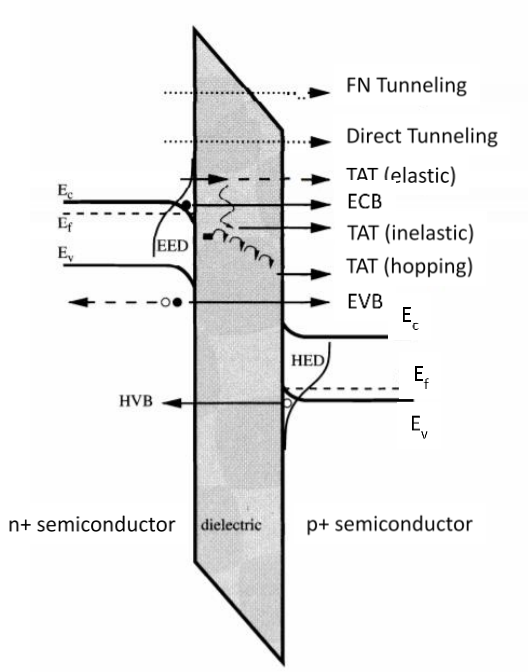


Figure 1-6 Schematic representation of an oxide as a potential barrier and the different types of electronic conduction as a function of field, temperature and presence of defects in the material and interfaces

In the absence of electrically active defects (perfect insulator) and considering the shape of the quantum-mechanical barrier alone, conduction is governed by the transmission probability of an electron through the volume of the oxide, known as *direct tunnelling*. This is owed purely to the penetration of the electron wave function into the volume of the oxide, and is therefore dependent on the barrier shape which is in turns defined by oxide band gap and oxide thickness and the rate of *band bending*.

If a semiconductor/oxide contact is considered tunnelling can be more rigorously classified in ECB (electrons from conduction band), EVB (electrons from valence band) and HVB (holes from valence band) (Figure 1-6). Such is the case of the MIS or MS (Metal semiconductor or shottky contact). In the MIM structure however, which is of interest in this work, further distinctions are not necessary owed to the absence of a band gap in the metal. A detailed review of direct tunnelling models can be found in [49].

The application of an electric field in the terminals of the test structure leads to further band bending and enhances electron transmission probability, leading to a strong increase in the observed current. At the limiting case where the barrier shape is degenerated to a triangular shape the conduction regime is known as Fowler-Nordheim (F-N). Presently, for state of the art oxides where defectivity level is very low, the complete direct tunnelling

effect becomes dominant for ultra-thin oxides (typically < 2nm) whereas F-N becomes dominant at high fields or strongly bent bands [50]. Finally, a transition can also occur by excitation of an electron in an energy range sufficient for it to surpass the energy barrier of the oxide/s or m/oxide interface and find itself in the conduction band of the metallic electrode, more widely known as thermionic or Schottky emission. Contrary to direct tunnelling schemes temperature doesn't play a role, Schottky emission is enhanced by temperature. [50].

Charges in the oxide (space charge) arising from structural defects of the oxide or other contaminants can contribute in conduction. This is a fundamental property of charges and arises naturally using dimensional analysis on the Poisson equation, more widely known as the Mott-Gurney law for space charge limited conduction [50].

The presence of charged defects (typically associated with a charge state) in the oxide volume acts to lower the effective barrier height an electron feels in the vicinity of the trap. The case where an electron can be excited into the conduction band (either by means of temperature or field) through a trap site is called Poole-Frenkel emission [50]. Another scenario is 'hopping conduction', where electrons are trapped for ultra-short times in a site forming a quasi-hydrogen state. By means of temperature of field excitations the electron is momentarily excited free from the trap only to be captured by the energetically most favourable closest neighbour. Tunnelling is therefore conducted through a series of such "hops".

The latter mechanism is probably the most complete and general way to model conduction in oxides and amorphous semiconductors, although even semi-analytical solutions are limited to strict assumptions. Even though multiple groups contributed in the development of the theory involved, a comprehensive model linking macroscopic current properties to the evolution of tunnelling probability with temperature, field and the average distance between defects (localization parameter), $J \sim P(\alpha,F,T)$ was proposed by Dr. Nevill Mott (Nobel prize 1977) [51]. Finally it is worth noting the works of Apsley & Hughes on the macroscopic current characteristics of hopping conduction as a function of defect density, temperature and field [52], [53].

The J-V relation evolution at different temperatures can give valuable information about the spatial density of defects through the values of the localization parameter and density

of states, both of which can be linked to the slope and intercepts of the linearized logJ-logV diagram. A detailed treatment of the topic can be found in [53]. Current conduction is directly related to the generation rate of defects in the oxide which leads to dielectric breakdown phenomena. In the next section a short overview of the physics of dielectric breakdown in oxides is presented.

1.4.2 A short overview of dielectric breakdown in oxides

Dielectric breakdown phenomena were observed for the first time in solid insulators by the German physicist Georg Christoph Lichtenberg in the 17th century and constitute the only macroscopically observable image of dielectric breakdown in our disposal, capturing the chaotic (yet fractal in nature [54]) behavior of dielectric breakdown pathways formed in solids; more widely known as conductive filaments within the ReRAM community.

The development of microelectronics has caused a systematic study of breakdown phenomena in oxides, in particular SiO_2 and high-k dielectrics. The origin of dielectric breakdown is understood to be linked to the generation rate of defects (breaking of molecular or atomic bonds). The increase of conductive defect density relates in turn to an increase of current density hence facilitating the breaking of bonds in a cascade process. For thick or high barrier oxides current is mainly controlled by FN tunneling while for thin oxides, direct tunneling is more critical. As the defect density increases in the oxide interface / volume current is controlled by TAT or hopping mechanism. The random evolution of defects generates a percolation path resulting in the random evolution of the conductive pathways in the oxide [55].

The physical process of BD is argued to happen in two stages: First, a random generation of defects takes place in the oxide volume known as the nucleation phase. Once a critical defect density is reached at a given site, the local field and temperature perturbations act to propagate the direction of the defect growth from one electrode to the other, driven by the electric field. Once the percolation path is formed, further augmentation of local conductivity leads to lateral wear out and enlargement of the filament.

The point where the defect density becomes critical for a permanent change to be observed in the electrical behavior of the oxide is called *breakdown event* and it is mainly associated with two experimentally measurable parameters: time-to-breakdown T_{BD} (in a constant bias/current stress regime) or breakdown voltage V_{BD} (in a varying bias/current stress

regime) both of which are shown to follow a power-law [56]. In the context of MOS technology the charge-to-breakdown (Q_{BD}) is often used, that is the total charge injected into the oxide for a measurable breakdown even to be observed.

The power law dependence of T_{BD}, Q_{BD} and V_{BD} parameters lies is the underlying reason for the weakest-link problem ([55], [57]) and the reign of Weibull statistics over DB [57].

$$F = 1 - \exp\left[-\left(\frac{x}{x_{63\%}}\right)^{\beta}\right] \xrightarrow{\text{yields}} \ln[-\ln(1-x)] = \beta \ln x - b \ln x_{63\%} \qquad (1.1)$$

The Weibull distribution probability can be projected in the dual-log space where the Weibit "ordinal" variable can be linearly related to the log of the statistical breakdown variable. The slope (shape factor) of the Weibit distribution is a particularly important parameter giving a measure of the intrinsic variability of the Weibull process whereas the intercept is generally related to the device surface. From a physical standpoint, the two parameters are often linked to device macroscopic characteristics such as oxide thickness t_{ox}, with the relations according to the models of Suñé [58] & Degraeve [59], [60]:

$$\beta = \left(\frac{a}{a_0}\right) t_{ox} \ , \qquad \eta \propto \exp\left(-\frac{1}{\beta} \ln\left(\frac{A}{a_0^2}\right)\right) \qquad (1.2)$$

where a_0 is the defect size and A, the cell surface. The slope dependence on oxide thickness is associated to the probability of presence of defects assuming defect distribution in the oxide volume is homogeneous [61] and is well supported by experimental results in CMOS technologies or in structures where reactive electrodes are not introduced [62]. Nevertheless, when applied to ReRAM the interface reactions between oxide and reactive electrode should be taken into account and can strongly affect the dependence of the parameters to their nominal "geometric" values.

Finally, it should be noted that the qualitative characteristics of breakdown events in the Weibit diagram can be used to detect different breakdown mechanisms in the same class of devices through the change in slope as a function of the region of values for the statistical variable. This is often used to differentiate between intrinsic and extrinsic events. As is usual with statistics though, care must be taken when trying to relate a

statistical property of the system to its physical parameters as the interpretation is often highly sensitive to the assumptions made.

Historically, three models have been proposed, relating microscopic physical mechanisms to the evolution of defects during BD. First, the thermochemical model proposed by McPherson, Mogul and Khamankar for SiO_2 [63] initially and later extended to high-k materials [64], [65]. In the thermochemical model the breakdown mechanism is associated to the formation of oxygen vacancies in the $Si = O$ or in general the $M = O$ bonds linked to the bond formation enthalpy as well as the oxide dielectric constant and is purely dependent on the applied electric field and a field acceleration factor coming from time dependent dielectric breakdown (TDDB) measurements. The thermochemical model is contested owed to two shortcomings: i) it does not does not accurately capture the polarity dependence of BD to gate biasing in NMOS transistors [66] neither the relation to carrier energy or current density.

This uncertainty, coupled with observation of hydrogen release during BD in SiO_2 have led to the proposal of the hydrogen release in which the breaking of the Si=H bridge leads to the augmentation of defects and eventually leads to BD [67], [68]. Nevertheless this mechanism has been only associated to SiO2:Si systems and to my knowledge, fails to explain the oxide breakdown mechanism observed in TMOs and high-k dielectrics. To the contrary, as far as high-k and specifically the HfO_2 system are concerned, recent works have associated the degradation of oxide gate current and eventual breakdown of the CMOS HKMG (high-k metal gate) oxide to the generation of oxygen vacancies in the oxide [69], [70] converging to the views of the ReRAM community for the forming and set processes in TMO ReRAM.

Even though dielectric breakdown was first studied mainly owed to CMOS applications, independent groups have reported studies in MIM structures as early as the late 1950s [71]–[74] where the term of *forming process* denoting oxide dielectric breakdown was first introduced. The observation that the forming process can be (partially) reversed [75] in certain material systems and could thus induce a non-volatile, field dependent hysteresis I-V curves could be considered as one of the earliest works in the field of ReRAM.

1.4.3 Forming process

1.4.3.1 Microscopic Description of Forming

As mentioned before, resistive switching has been investigated since the early 1960s. The image of conductive filament (CF) formation, that is a localized percolation of the material phase with increased conductivity, was first suggested by Dearnaley [76]. Generating such material phases can be induced either by fabrication or by electronic breakdown. At the heart of the process, lies the transition of a perfect or a quasi-perfect oxide to a locally defect rich material where hopping conduction is facilitated as discussed previously.

From an electrochemical standpoint conductive filament formation is seen as a series of local transitions to sub stoichiometric or a super stoichiometric phases with increases conductivity. Increased conductivity can be predicted for oxides owed to the nature of redox reactions associated with the formation of point defects of either positive or negative charge. If hypo-stoichiometric phases are thermodynamically stable, then the region of the metal organic oxide (MO_x) with $MO_{x-\delta}$ μ $\delta>0$ leads to a surplus of electrons. This can be either due to the formation of oxygen vacancies ($V_O^{\cdot\cdot}$) or to the generation of metallic cation sites ($M_i^{\cdot\cdot}$). Both cases lead to a positively-charged center with the subsequent generation of a surplus of electrons (and oxygen in gaseous form if no secondary reactions are considered) indicating a transition to an n-type semiconductor behavior. The processes are respectively described in Kröger–Vink notation in eq. (1.3) & (1.4), respectively:

$$O_O^{\times} \rightarrow V_o^{\cdot\cdot} + 2e' + \frac{1}{2}O_2(g) \tag{1.3}$$

$$M_M^{\times} + O_o^{\times} \rightarrow M_i^{\cdot\cdot} + 2e' + \frac{1}{2}O_2(g) \tag{1.4}$$

By analogy, hyper stoichiometry leads to a p-type semiconductor with surplus of holes. The point-defects that carry the negative charge in this case would be oxygen interstitials (Eq. 1.5) or cation vacancies (Eq. 1.6) both carrying negative charge states:

$$\frac{1}{2}O_2(g) \rightarrow O_i^{\cdot\cdot} + 2h^{\cdot} \tag{1.5}$$

$$\frac{1}{2}O_2(g) \rightarrow O_o^{\times} + V_M'' + 2h^{\cdot} \tag{1.6}$$

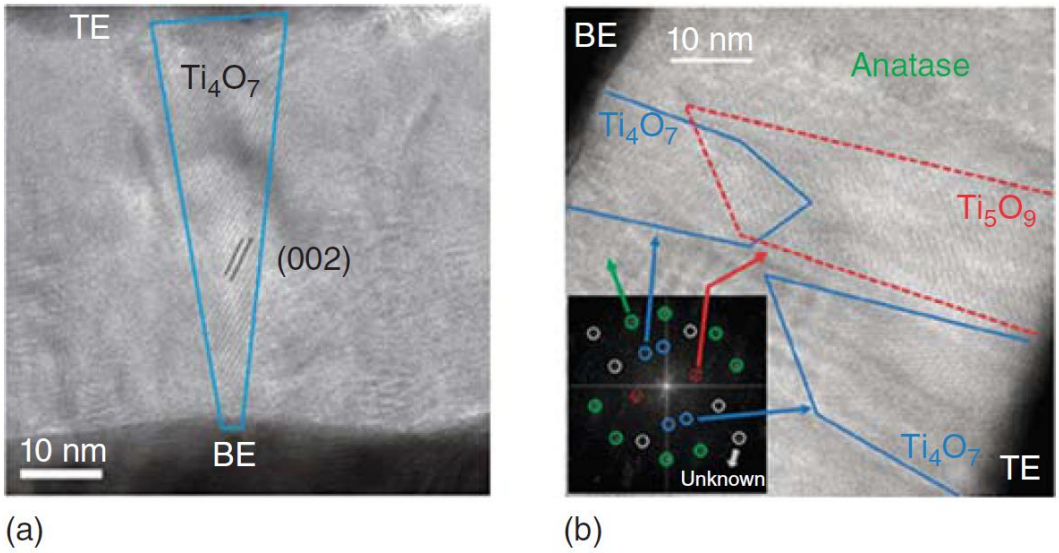


Figure 1-7 (a) Distribution gradient of sub-stoichiometric Magneli phases in TiO_2 [70] (b) Magneli-phases identification through TEM diffraction pattern analysis [71]

The model of vacancy based CF was first proposed by Dearnaley and Greene [77] and involves the migration of vacancies from anode to cathode thus generating a CF connecting the two electrodes. It implicitly thus predicts the sensitivity of such a CF to both the nature of the electrode and the partial pressure of oxygen in the environment. The migration of oxygen vacancies from a vacancy rich anode through the oxide and towards the cathode was later detected in $SrTiO_3$ [78]. Furthermore, systems such as TiO_2 exhibit well known sub-stoichiometric regions after forming called Magneli phases where the same gradient of oxygen vacancies has been observed [79], [80]. Similar sub stoichiometric phase transitions have also been reported for Ta_2O_5 [81].

It is also considered the dominant mechanism in HfO_2 based ReRAM as it explains the sensitivity of the devices to the top electrode and HfO_2 stoichiometry. Nevertheless, modified phases in HfO_2 devices have not been detected through any type of microscopic chemical analysis to date despite the strong activity in the topic. In all of the above cases, the vacancy region is considered to locally act as an n-type semiconductor. It is worth noting the case of NiO where the observed filament is attributed to Ni cation vacancies [82], [83] or the Al_2O_3:Cu system [84] as an example of a system where the CF is a cation vacancy and thus conduction is considered to be p-type.

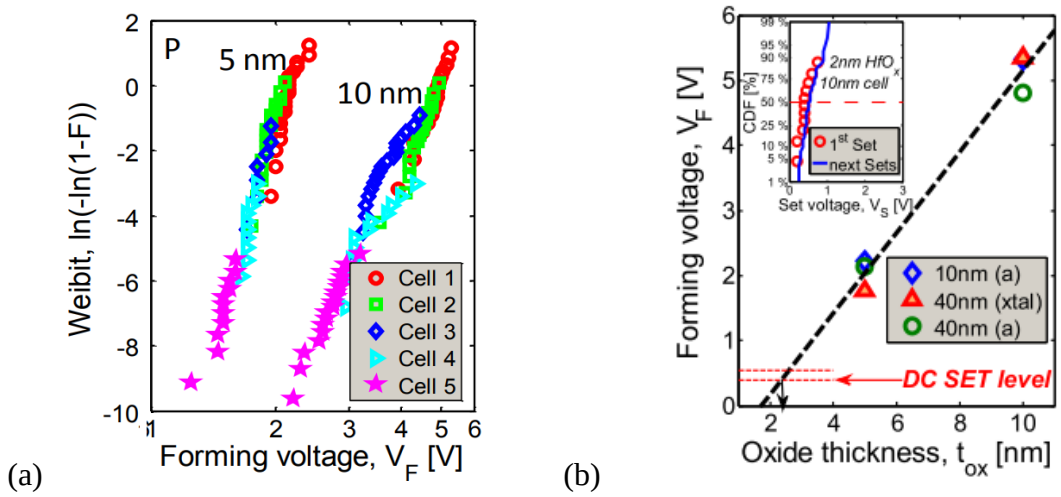


Figure 1-8 (a) Area-normalized Weibit plot of the system of HfO2(5nm)/Hf(5nm) for different oxide thicknesses and areas. (b) Impact of oxide thickness scaling o the forming voltage of HfO_2:Hf ReRAM. (Inset) Statistical range of set voltage, indicating that set is a limiting case of forming.[37]

It is interesting to note that even though clearly detectable sub-stoichiometric regions have not been detected in HfO_2 (serving as a popular system for a counter- example) it has been theorized that the forming process can be seen as a local phase transition, owed to the fact that point defects are expected to quickly relax under the influence of diffusion processes as soon as the external force is stopped [85]. Hence long-term stability of the conductive properties can be justified through a transition of phase. Nevertheless, this is to date an open topic of active research. Accepting the existence of local phases in the CF is not necessarily incompatible to the image of the percolation path as described in the previous section. Indeed, the nano-sized phases observed in TiO_2 are of the order of 5-20 nm (Figure 1-7) or smaller in the case of systems such as the HfO_2.

1.4.3.2 Impact of Oxide thickness and reactive electrode

Until now we have discussed the principle of the microscopic forming mechanism in TMOs. It is evident from the different reported results that the specificities of the filament nature, like its size, chemical composition and charge state of the conductive centers is heavily dependent not only in the material but also in the nature of the oxide/metal interfaces used. Whatever the specificities though, the macroscopic breakdown characteristics (exponential dependence to time and field, area and oxide thickness

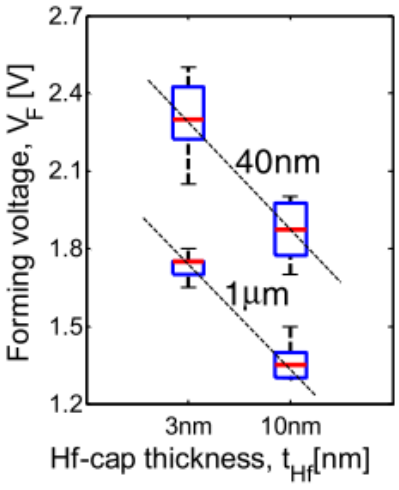


Figure 1-9 Impact of thickness scaling of the Hf reactive electrode in HfO_2:Hf ReRAM [37]

scaling) are well maintained denoting that the physics of the percolation path formation is predominant: The forming voltage scales with device area (Figure 1-8a, HfO_2:Hf system) and follows very well the Weibull distribution for dielectric breakdown while it scales proportionately to the oxide thickness (Figure 1-8b). Interestingly though the sharpening of the Weibull slope expected for the case of the thicker oxide is not observed, contrary to the respective cases of CMOS systems (see 1.4.2). This suggests that the defects are not homogeneously distributed in the oxide.

Several works have demonstrated that inserting a reactive electrode in the cathode contact facilitates the forming process. In particular Ti, Hf, Zr (column IV elements of the periodic table) have all been shown to reduce the forming voltage of the native HfO_2 sysstem or similarly Ta for the case of the Ta_2O_5 system [86]. In all cases the process is facilitated and the forming voltage reduced as compared to the native system being in contact with an inert electrode such as Pt. The reactive metals are considered to have a lower migration barrier with their native binary oxides hence facilitating the oxygen exchange reaction in the metal/oxide interface. It is worth noting that oxygen segregation gradients near the cathode/oxide interface have been observed for the TiO_2/Pt system [87].

The specificities of the role of the electrodes in the microscopic mechanisms of defect

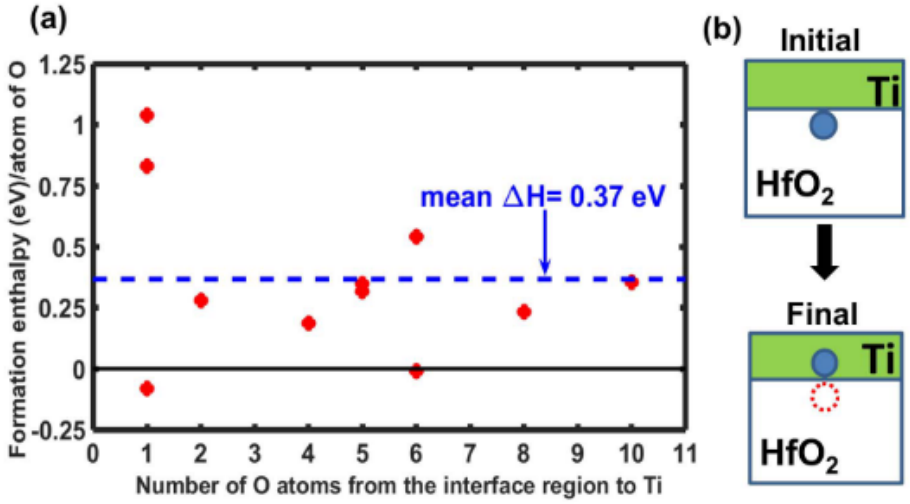


Figure 1-10 (a) Formation enthalpy of removing O from the interface region to top Ti layer. The enthalpy is divided by the number of removed O atoms for comparison (b) Pictorial model showing the process of O exchange across the interface used in studying the formation enthalpy of (a). [Traoré et al. T.ED 65(2), 2018]

generation are highly material dependent but in general it is understood that reactive metals have have a lower formation enthalpy of a sub oxide phase with an oxygen ion from the binary oxide and thus facilitate the migration of oxygen ions into the cathode electrode subsequently generating vacancies and a surplus of electrons. Traoré et. Al (amongst other groups) have studied the effect of oxygen exchange in the HfO_2:Ti interface by means of density functional theory (DFT) calculation. Figure 1-10 shows the energy cost to extract O ions from the HfO_2:Ti interface (formation enthalpy / atom O) as a function of the number of O ions migrating towards Ti calculated using DFT. The energy required to extract an oxygen ion is a function of the atoms extracted from the vicinity and the distance from the interface. In the graph, the highest energy values correspond in O atoms originating from layers a few atomic layers away from the interface while the zero or negative values correspond to O ions at the interface. Moreover the decreasing energy cost with increasing number of atoms can be interpreted as a lowering of the interface barrier as the oxide region near the interface becomes more substoichiometric. Given the highly local nature of DFT, this process can explain the initial stages of filament formation.

1.4.4 Switching Mechanism

1.4.4.1 Unipolar and Bipolar Switching

As briefly discussed in §1.3 post forming ReRAM devices can exhibit two modes of operation that being the bipolar or unipolar switching behavior depending on whether bias polarity needs to be reversed, respectively, in order for the device to successfully reset from LRS to HRS. Unipolar operation has been shown for several binary oxides amongst which TiO_2 [88][89] and HfO_2 [90][91]. Even though the microscopic mechanism can vary from material to material, the physical principles involved are attributed primarily to a competing behavior of drift and diffusion processes of the moving species. Moreover, systems such as HfO_2 can show both unipolar [34], [35] and bipolar switching [92] behavior depending on the nature of the electrodes. Finally, it is worth noting that even though filamentary switching has been reported for both unipolar and bipolar modes, there is no report of unipolar surface-dependent ReRAM system to the author's knowledge to date.

A unified semi-analytical model has been recently proposed by Yu *et al* to describe the unipolar and bipolar nature of ReRAM devices in terms of the moving species kinetics [93]. The main concept lies with modeling O^{2-} ion kinetics in terms of drift and diffusion processes. In that image the direction of the diffusion associated forces is dictated by the concentration gradient while drift forces will be mainly controlled by the field gradient. In bipolar switching drift and diffusion will act synergistically to break down the formed filaments causing ions to migrate back and either recombine with oxygen vacancies or oxidize metal precipitates. In unipolar devices however drift and diffusion will act antagonistically to each other. Lee et al. [94] have suggested that during the reset process in unipolar devices, filament dissociation is owed to a thermally activated oxidation process due to Joule self-heating, which in turn causes diffusion forces from the anode to overcome the drift forces.

Yu's model is based on the experimental observation that unipolarity is associated with noble metal interfaces whereas bipolarity with at least one reactive interface. With that in mind, it was assumed that the introduction of a reactive interface causes local oxidation in

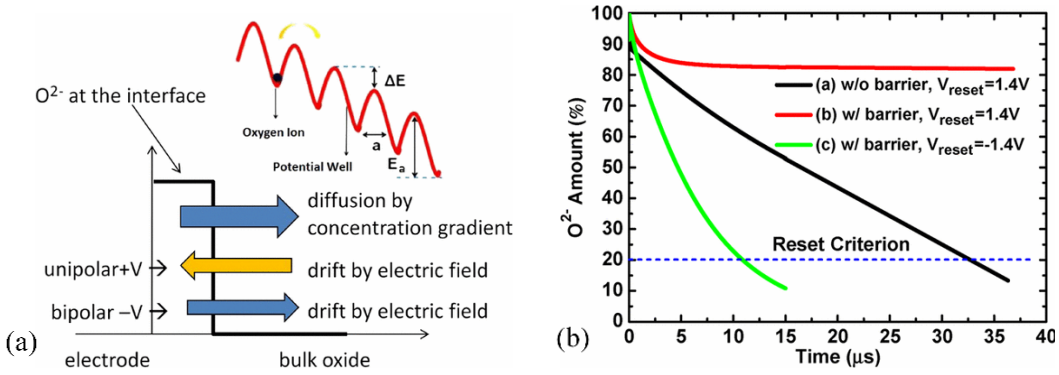


Figure 1-11 Schematic representation of driving forces excreted on O^(2-) in Yu's model. Migration is assumed to occur due to O^(2-) hopping in subsequent potential wells in the oxide characterized by discrete energy states. In bipolar stacks drift and diffusion are assumed to act synergetically whereas in unipolar mode they act competitively to each other. (ii) Simulated O^{2-} concentration under 1.4V bias (discrete pulse) for three reset cases. Case (a): Without interfacial barrier and under positive bias, unipolar reset is successful. Case (b): With interfacial barrier and under positive bias, unipolar reset is unsuccessful. Case (c): With interfacial barrier and under negative bias, bipolar reset is successful. [94]

the reactive electrode/oxide introducing a potential barrier layer. In the absence of a barrier, oxygen ion diffusion is possible [Figure 1-11 ii-(a)] and unipolar switching is possible. The introduction of such barrier acts to block diffusion from the anode if a positive bias is applied and unipolar reset probability drops [Figure 1-11 ii-(b)]. On the other hand, under negative bias O^{2-} ions can overcome the barrier making bipolar mode possible [Figure 1-11 ii-(c)].

A diffusion process it is mainly controlled by local temperature rather than local field and is thus also known as a thermally controlled mechanism (TCM) ReRAM. It has been argued that using thinner electrodes[95] or electrodes with higher effective thermal resistance [96] could improve ReRAM characteristics (both bipolar and unipolar).

1.4.4.2 Universal Switching Characteristics

Closing this section, it is worth mentioning that no matter the material system, or its switching behavior, certain symmetries common to all resistive switching devices can be identified. Notably, it has been observed that for all VCM and ECM cells whether unipolar or bipolar, the LRS associated resistance is strongly correlated to the compliance current (maximum permitted current through the cell during set transition) (Figure 1-12a). and the maximum current during reset is also correlated to the set current (Figure 1-12b).

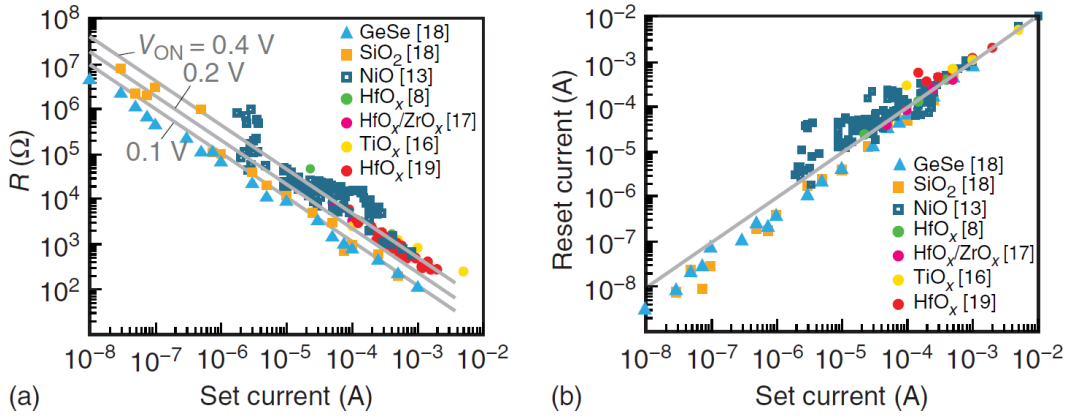


Figure 1-12 (a) LRS state resistance correlation (b) HRS resistance correlation to set compliance current. Switching Behavior from multiple material systems and switching mechanisms. [R. Waser, Nanoelectronics and Information technology, 2012]

Moreover, it has been reported that the switching time depends exponentially with applied voltage for both the forming process [97] and during switching [98] supporting the hypothesis of a gradual defect generation followed by a nucleation-propagation process or in equivalent terms the formation and dissolution process of the conductive filament.

Several models of varying detail have been proposed to describe the mechanism. From a compact model standpoint the variable gap or variable diameter models have been proposed; a comprehensive review can be found in [67]. It is also worth noting the approach of Nardi *et al*. in the microscopic modeling of forming and set/reset kinetics using Nerst-Planck equations for the oxygen vacancies [I had it somewhere, sandisk publication].

1.5 ReRAM Architectures

Integrating ReRAM cells in a functional large scale memory array imposes the use of an element in series with the MIM cell called a selector device. There are two reasons that impose the use of selectors in ReRAM memory array technology. From a fundamental physics prospective, the nigh-uncontrollable nature of the dielectric breakdown event occurring during the forming and set operations. If too many oxygen vacancies are generated during the forming/set process, it can lead to inability to reset and device failure. This can be especially critical for the forming operation where forming occurs at elevated voltages. As breakdown occurs, the resistance between the two terminals of the cell drops

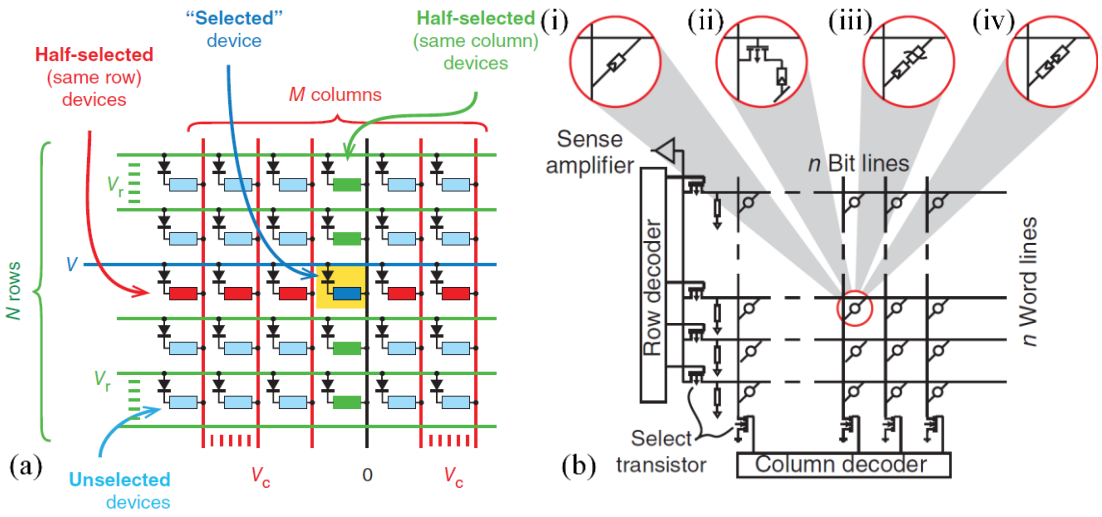


Figure 1-13 (a) Cross point memory array architecture and (b) different types of ReRAM cell architectures (i) 1R (ii) 1T1R (iii) 1D1R (iv) 1S1R [58]

acutely resulting in an exponential enhancement of breakdown more commonly referred to as current overshoots [99].

From a memory array design prospective, the crosstalk between memory cells of large memory arrays, quickly becomes a limiting factor for both the size and array cell density. Figure 1-13a shows a typical configuration of a cross point memory array architecture where cells are arranged by columns and rows. There are many read and programming schemes that are being the scope of this short review. In general, a cell in the array is selected by combined biasing of its row and columns to an appropriate read or write voltage. This results in not just the selected cell to be biased but all the column and row cells as well. Even though the cells corresponding to non-intersecting biases are not optimally polarized, unwanted crosstalk can emerge: First, read and write disturb, where the applied biases not only affect the state of the selected cell, but also that of nearby cells in shared or non-shared columns and rows [100]. Second, inaccurate determination of the state of the selected memory cell as a result of additional sneak-path currents coming from half selected or floating bias cells that collectively lower the readout resistance [101]–[103]. The latter becomes particularly limiting if an array sub-block is fully programmed in LRS, where the leakage current of each cell is the largest.

Hence, a control element (selector) is integrated in the memory cell in series to the device (1S1R) so as to both limit current overshoots and leakage current through the device as opposed to the 1R memory cell architecture (Figure 1-13i). The selector element can be

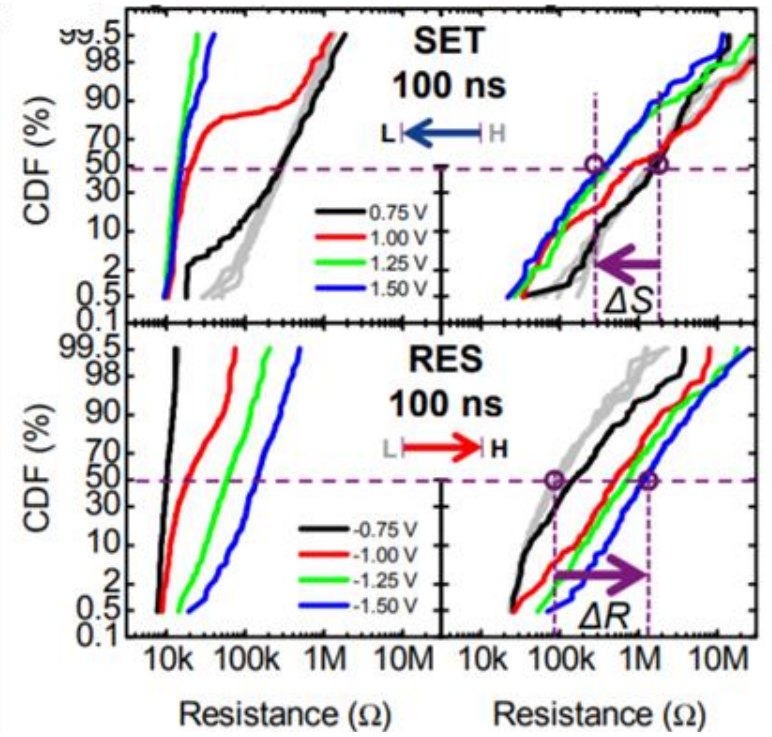


Figure 1-14 LRS and HRS device-to-device resistance cumulative distributions for different pulse heights during the SET and RESET pulsed operations.

active or passive in nature depending on if its conductive properties can be controlled independently from the memory cell by means of a bias other than the one applied on the memory cell. The active selectors to date include vertical BJTs [104], [105], or FETs [106] leading to a 1T-1R architecture approach (Figure 1-13ii). Passive selector technologies include MS Shottky diodes [107] and p-n junctions (Figure 1-13iii). Passive selectors include MIM switches whose conductive properties can be modified by means of an applied field of the same range as the memory such as oxide based nonlinear field-tunable tunnel barrier switches [108] or ovonic threshold switches (OTS) [109].

All different selector technologies come with their respective pros and cons. Notably, MOSFETs although excellent in terms of current controllability and parametrization (both by tuning transistor sizing and operating voltage) consume too much surface. On the other hand, diodes cannot be used with bipolar elements while the switching properties of passive selector elements need to be carefully engineered to fit the memory switching range. For this reason, in this work the selector of choice is the standard FET as its tunability gives enough margin to focus on the investigation of the memory cell engineering.

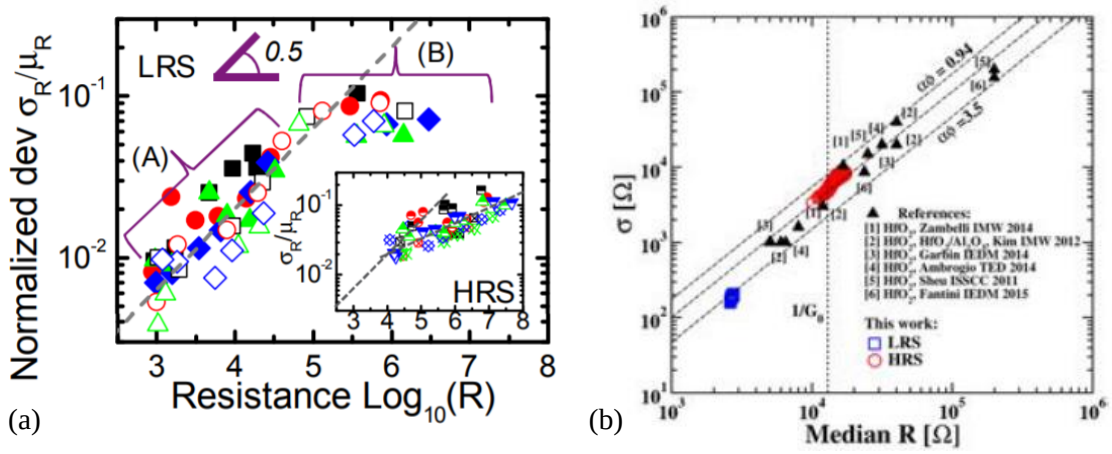


Figure 1-15 (a) Normalized relative spread of standard deviation over mean resistance for LRS and HRS as a function of the observed resistance[110] (b) Generalized sigma too median relation for both LRS and HRS from [111]

1.6 Variability of the ReRAM States

One of the major challenges for ReRAM is controlling variability of the states. Figure 1-14 shows the spread of both LRS and HRS for pulsed programming conditions; different set pulse height induces a bimodal distribution for the CDF of the LRS. The HRS is less sensitive to bimodality for varying programming conditions yet is shows a significant drift of the mean resistance with varying height of the reset voltage.

Variability in ReRAM can be traced in both intrinsic and extrinsic causes. One the one hand, the percolative nature of the filament entails that in every program-erase cycle the local inter-defect distance will be on average slightly different than the previous case. This is particularly true for HRS where the path is broken and conduction is volume-aware: After each reset, the inter-defect distance will slightly vary with respect to the previous cycle. As a result, and since conduction is controlled mainly by delocalized centers the random percolations of the initial defect configuration give rise to a large number of possible resistance states. In LRS intrinsic variability is less critical as the defect is more confined around the volume of the filament.

For filamentary ReRAM this behaviour has been modelled under the image of the quantum point contact model (QPC) attributing the totality of intrinsic variability to the different percolations of the filament near the anode [110]. In their approach, the stochastic nature of HRS and LRS is owed to the variation of number of atoms, in the "tip" of the filament to anode interface, the percolations of local distance as well as the local metal electrode

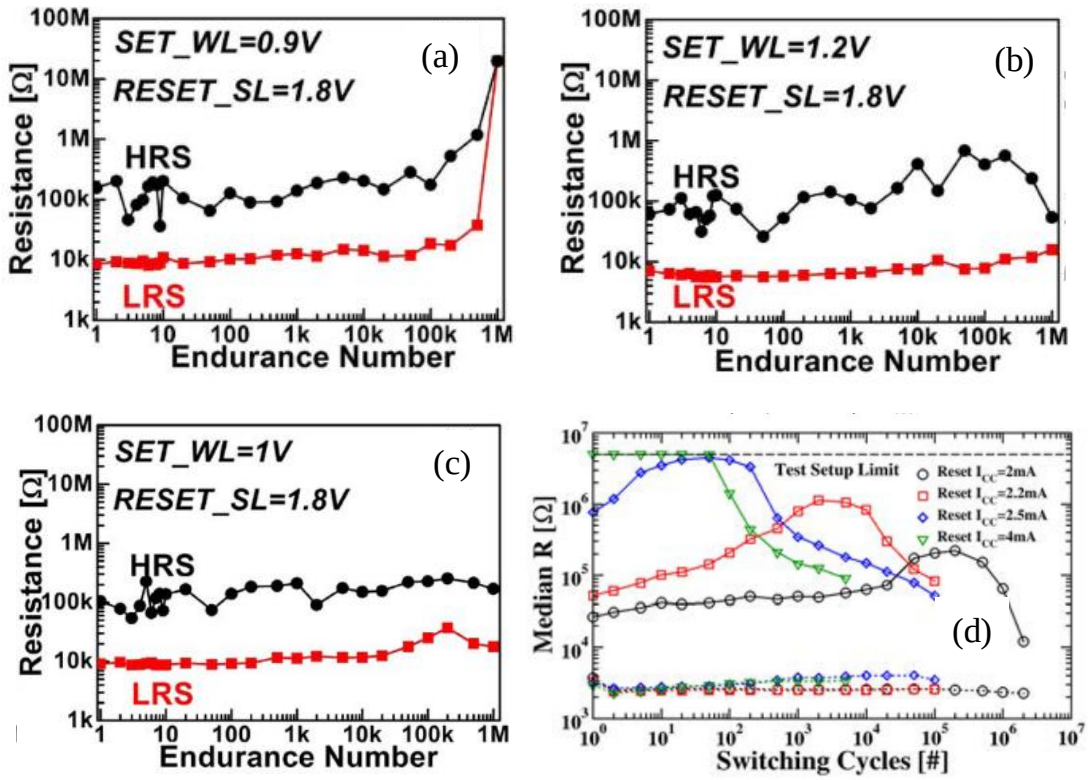


Figure 1-16 Impact of low (a) , high (b) and balanced (c) set pulse height in the endurance characteristics of HfO_2 ReRAM cells. [112] low set voltage favors failure towards the HRS, while a high condition leads to failure towards LRS (b). (d) Impact of the Reset pulse voltage modulation through change of the series resistance of the transistor in a 1T1R 4kbit array of HfO2/Ti (10/10 nm) cells [111].

morphology. It is worth noting that such approaches do not take into account bulk material defects such as local phase transitions, interstitial impurities or local crystal dislocations that could influence the shape and conductive properties of the percolation path.

On the note of intrinsic variability, a fundamental relation between the standard deviation of the cycle-to-cycle (C2C) [Figure 1-15a] [110] or device-to-device (D2D) [Figure 1-15(b)] [111] distribution of either state has been demonstrated experimentally by separate groups. In both cases, the arbitrary potential barrier fluctuation in the anode/oxide interface has been used as the percolating parameter controlling the statistics of the resistance and its spread in the image of the QPC model. It is nevertheless important to note that the 2D QPC model fails to explain bimodality in the CDF, which is not surprising considering that a bimodal distribution is a typical property of two coexisting mechanisms in the devices, notably intrinsic and extrinsic variability. Extrinsic variability is particularly hard to model as it originates from a number of factors. Device-to-

environment interaction, such as reactive electrode or oxide to sidewall interaction locally modifying vacancy diffusivity circuit-device interactions and metal line voltage fluctuations are only some of the effects that can impact the LRS and HRS statistic. Finally, LRS seems to be more prone to bimodality and thus extrinsic variability while in HRS intrinsic variability is dominant. Unfortunately, to effectively study such fine effects populations larger than the Mbit range would be required, results for which, are not available in the literature to date.

1.7 Endurance Mechanisms of TMO ReRAM

In Figure 1-16 (a)-(c) the median resistance of HRS and LRS is shown versus the accumulated programming operations in the device in pulsed operation [112]. The programming pulse width (duration) was fixed at 100ns and the pulse height during the set and reset operations was varied. The drift/diffusion effects of the active species in ReRAM and the inevitable local reconfiguration of material properties lie at the heart of the endurance / retention properties of the device. When the set voltage is too low or too high the resistance values of both states tend to drift towards those of the HRS or those of the LRS respectively. The trend is symmetrical when the reset voltage is varied for a given set condition (Figure 1-16d) or time is varied [111] [112].

This behaviour can be explained if the field controlled vacancy drift is taken into account: Each set/reset cycle will induce an ionic current of vacancies that generates the different filament percolations. If the set voltage or current compliance are higher more defects will be generated resulting in a stronger path that requires more energy to dissolve. It is interesting to note that all three parameters, notably the pulse height, pulse width and maximum current through the device, correspond to the energy transfer to generate and dissolve the percolation path (E_{cycle}=IVt). Degradation towards HRS or LRS is inevitable as in every operation it is impossible to completely counterbalance all the defect sites generated in the previous operation. Hence, statistically, the fragile optimal equilibrium is broken: When vacancies are dominant the device is stuck to LRS while when vacancies are lost from the active zone resistance leans towards HRS. The failure towards LRS was simulated by MC by Yu et al. [113].

1.8 Retention Mechanisms of TMO ReRAM

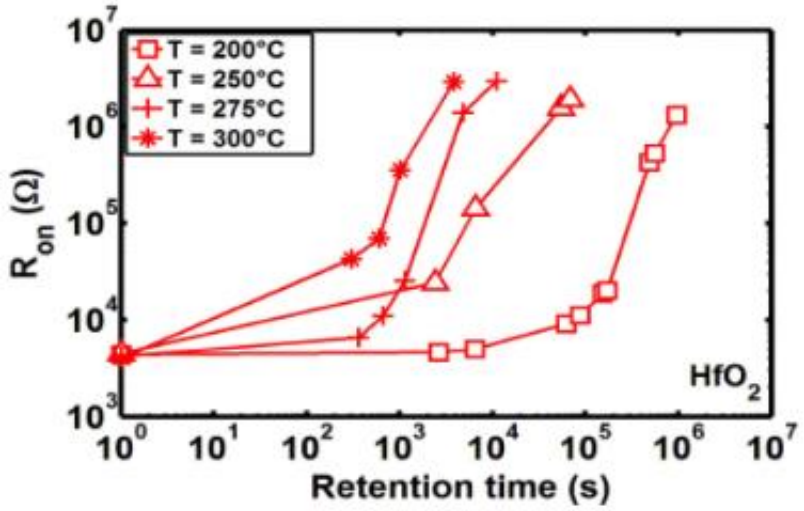


Figure 1-17 LRS media drift with time and temperature for HfO_2/Ti based ReRAM [117]

Yu's work was one of the first kinetic Monte Carlo models reproducing the statistical behaviour of ReRAM in the image of the stochastic movement of oxygen ions and vacancies allowing to make predictions for both endurance and retention failure towards LRS. And while it is true that when a small number of devices are included into the statistic (like in earlier works in ReRAM) the LRS becomes the principal origin of failure [114]–[116][117] latter works in small kbit arrays point out that early HRS failures post programming are also non negligible [118].

The drift of resistance to higher values (LRS retention failure) has a monotonic behaviour and is activated with time and temperature (Figure 1-18a) -this is true for both the median values and upper quantiles suggesting one uniform mechanism at work- and can be explained by the diffusion of oxygen vacancies in the reactive electrode. On the other hand, the HRS can drift both towards higher and lower values as is usually observed between different quantiles in the HRS CDF, depending on the time the system is given to relax and ambient temperature. Lin et al. [118] have recently proposed a mechanism to explain the non-uniform evolution of the HRS drift.

The HRS retention characteristics have been interpreted through a three stage mechanism involving two physical processes. The activation range of each stage is highly dependent

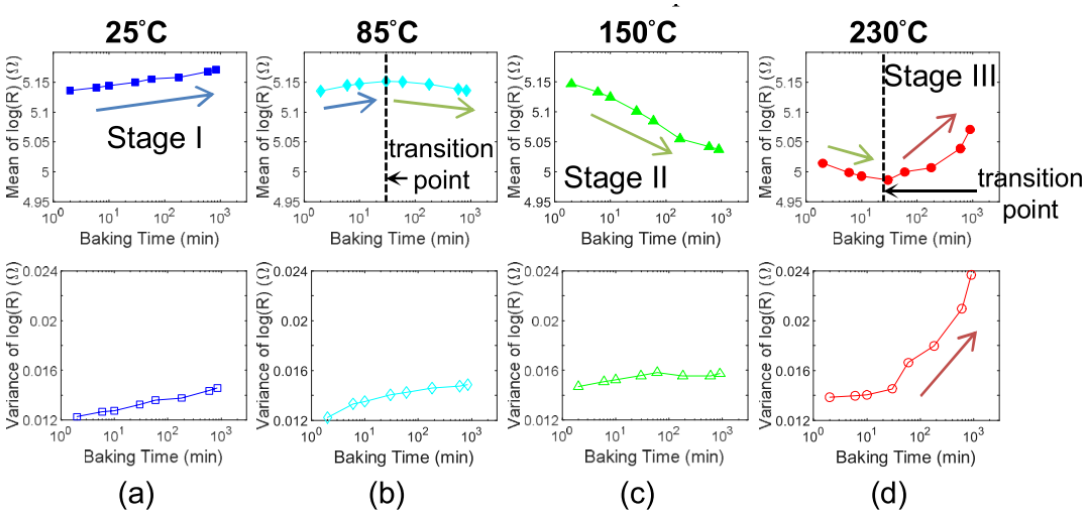


Figure 1-18 The mean and variance of the resistance distributions change with time at 25?C, 85?C, 150?C, and 230?C. First, the mean resistance increases in stage I. Then, the mean value drops in stage II. When the temperature is high enough, the stage III can be observed in a reasonable time and the mean will increase again. The variances slightly increase with time in the stages I and II but dramatically rise in the stage III. [118]

on the temperature. In Stage I oxygen vacancies migrate in the oxide bulk by slow diffusion process (Figure 1-18b) leading to an enlargement of the statistical spread of resistances and a transition towards normally distributed valued. State II is a mix of migration and ion-vacancy generation followed by migration activated by temperature. The transition point between stages I, II is shown in Figure 1-18b and denotes the temperature dependence of the regimes. The vacancy generation mechanism leads to statistically lowering of the resistances (Figure 1-18c). Stage III is predominant at high temperatures and is attributed to migration of oxygen ions into the reactive electrode. All three stages can coexist at high temperatures resulting in the different dynamics between the upper and lower quantiles of the CDF of the HRS.

The activation of different physical processes between stages I, II and stage III as argued by the authors is also supported by the change of the slope of Var(R)-log(time) (Figure 1-18 a-d, lower). While in stages I, II the device-to-device variance increases, the rate of increase with time is similar. Only after the activation of stage III does the variance also show a significant change.

1.9 The place of ReRAM in the CMOS integration

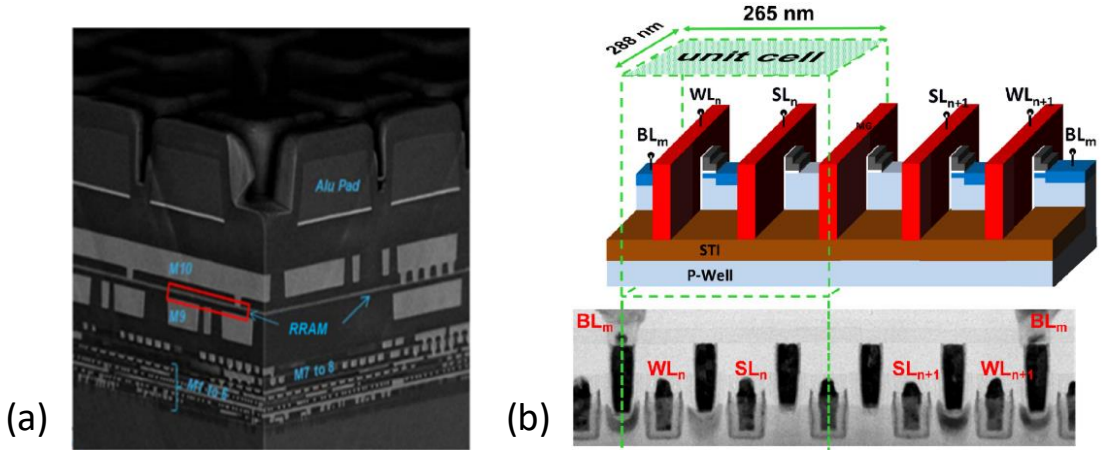


Figure 1-19 16 Kbit BEOL integrated 1T-1R HfO_2 ReRAM array in a 28nm FDSOI process from STMicroelectronics (b) Schematic representation of dummy-gate FEOL integrated HfO_2 based ReRAM in a 14nm FinFET process from TSMC [].

Until now we have tried to briefly review key aspects of the physics and material involved in ReRAM devices and provide the scope of applications that drive the industry's economic setting, motivation and constraints. And it is precisely because of the economic motives driving advanced research in the field of semiconductors that any new technology faces the challenge of being adopted by the industry. Considering the huge success of Si based technologies that are the golden norm in microelectronics, any new technology aspiring to make it out of the lab and into a fab has to satisfy three main targets: Performance, cost and compatibility to existing technologies.

If a novel technology aspires to make it to large scale implementation, it first needs to show significant gain as compared to existing ones. The less disruptive an innovation, the smaller its cost/effort has to be to be implemented in current products. Considering the undisputed reign of Si based technologies in microelectronics this is the challenge that all novel NVMs face to date, considering that none has at this moment achieved breakthroughs significant enough to justify a dedicated novel NVM fab. Moreover, as advances in selector technology have been slow, and logic functions are still done on standard CMOS, it is unlikely that fab facilities will emerge centered around exotic CMOS incompatible materials and transistors will still be required in new-class NVM related products. As a result, most new classes of NVMs (PCM, ReRAM, STT-MRAM, etc) are by default integrated in the middle to far back end of the line (BEOL) of industrial CMOS

processes typically between metals 4 and 10 where cross-contamination regulation are significantly relaxed and materials that are typically considered contaminants to Si can be safely introduced. Several companies have demonstrated test vehicles for ReRAM in that level (Figure 1-19a) [TSMC,Panasonic,Renesas,Toshiba,STMicroelectronics]

This provides the advantage of both reduced cost, as lithography is cheaper at these levels and higher flexibility for novel material introduction. However, this flexibility comes at the cost of density: With each metal line, lithography critical dimension is relaxed to satisfy power transfer and self-heating requirements. Moreover, with the most advanced CMOS nodes, the main limiter in terms of density is no longer the transistor pitch and MOL metals but more the BEOL metal pitch a trend that is expected to increase in the next node generations [119]. As a result, to improve density, the NVM element needs to approach the FEOL as much as possible. This is also true if NVMs aspire to target beyond Von-Neumann architectures for near-memory or in-memory computing.

In this environment, HfO_2 based ReRAM can actually be facing an opportunity where competitors such as STT-MRAM and PCM may still be limited to the upper BEOL metals. Owed to HfO_2 being already integrated in the gate stack of HKMG nodes, the roadblock of material constriction is lifted. Moreover, HfO_2 based ReRAM with its high endurance and promising retention as compared to other materials has a good starting point to make it to the FEOL. The feasibility of such an approach has already been demonstrated by TSMC for a 16Kbit test vehicle in a 14nm FinFET process (Figure 1-19b) where a dummy-gate fin is used as the memory cell in series to the actually control transistor, with the added advantage of not requiring any additional lithography levels (!). Finally, the 21st century AI drive is a booster for ReRAM as new computing paradigms are required, with ReRAM achieving breakthroughs in that field [120]. Nevertheless, with competitors such as STT-MRAM holding the advantage of maturity, the competition is still severe [121].

CHAPTER 2. MATERIAL ENGINEERING FOR HFO_2 BASED RERAM BY SI/AL ION IMPLANTATION

2.1 Introduction

As discussed in Chapter 1, even though multiple material systems have been investigated for RRAM applications, HfO_2, Ta_2O_5 and $Al2O_3$ have attracted most attention over the years. These oxides, and in particular HfO2 and $Al2O_3$ are native to HKMG CMOS processes and pose no roadblocks to integrate with advanced CMOS, contrary to more exotic systems such as Strontium Titanate ($SrTiO_3$, STO) [29], [30] or TiO2 [122]. From a material growth perspective, they can be synthesized using highly controllable, industrially friendly deposition methods like Atomic Layer Deposition (ALD) or Pulsed Vapor Deposition (PVD) allowing for precise thickness control, high uniformity across wafer as well as high throughput.

Integrating $Ta2O_5$ MIM in CMOS, increases fab cross-contamination adding to integration complexity and restricts the memory point in the far BEOL. Panasonic is the only company to date to have commercialized $Ta2O_5$ ReRAM in their MN101L series, performing up to 10^5 write/erase cycles [123].

As a result, industry and academia have paid significant attention to metal-organic oxides like Al_2O_3 and HfO_2. The latter, in particular is also used in the gate stack of advanced HKMG CMOS nodes and thus can be integrated almost at any level of the integration route from FEOL to BEOL. Furthermore, HfO_2 based RRAM has demonstrated promising programming endurance characteristics on unitary cells [124]–[126] in the order of >1 x 10^8 cycles for the median-to-median memory window. As a result, it is one of the most promising material candidates for ReRAM applications especially for storage class memory. Nevertheless, integrating the technology in large arrays requires several levels of optimization.

First, the forming operation constitutes a significant roadblock as regards the integration of the technology in large arrays. Typical forming voltage values for ReRAM based on 10 nm HfO_2 layers are in the order of 3.5V, which strongly augments circuit periphery [127] and impose the use for complex algorithms during forming [128]. In order to reduce the footprint of the overall 1T1R memory cell, advanced CMOS nodes should ideally be co-

integrated to the memory element. Adding to surface economy, using advanced nodes can allow to operate at lower voltages contributing in overall power reduction in the memory chip. Nevertheless, the high-voltage forming operation required for HfO_2 based RRAM cells to function as memory stresses the control transistor in the 1T1R cell. Hence, the forming step should be relaxed or ideally eliminated in the technology.

Reducing forming voltage can be achieved either by reducing the oxide thickness and modulate oxide crystallinity [37] or by reducing the stoichiometry for a given oxide (§1.4.3). The former acts to enhance the electric field for a given applied bias, accelerating the generation of electrical defects and consequently, dielectric breakdown [64]. The latter targets on pre-engineering a higher defect density in the as-fabricated oxide, by decreasing the oxygen fraction thus favoring the generation of metallic states [129], [130].

Reducing oxide thickness to achieve forming-free or quasi forming-free devices eventually leads to degradation of the HRS/LRS ratio, also known as *memory window*. Furthermore, it comes at the expense of significant degradation of the cell data retention. The use of reactive electrodes such as Ti or Hf has been used to reduce forming voltages. The presence of the reactive electrode and in particular one that consists of the column IV elements of the periodic table is indispensable in order to achieve stable resistive memory effect for HfO_2 based OxRAM. The reason lies in the high solubility of oxygen in these

a) elements, making them ideal to the metal-oxide and the reactive

b) facilitate ion exchange between electrode layer. Nevertheless their

role in the forming operation is limited to the interaction of the Oxide/Metal interface, at least as regards the forming operation [86]. As a result, in order to introduce the generation of a critical density of bulk defects, necessary for the first resistive switch operation (forming process) further engineering of the oxide is required. Alternatively, thermal anneals have been used during fabrication process to facilitate the intermixing of the metal oxide/reactive electrode interface, generating conductive states during the fabrication process of the devices. In this approach, a high temperature step is typically required, which renders the elements incompatible to the MOL/BEOL [86].

Methods to increase data retention of the cell have also been investigated. From a process integration point of view, superior thermal stability characteristics have been demonstrated

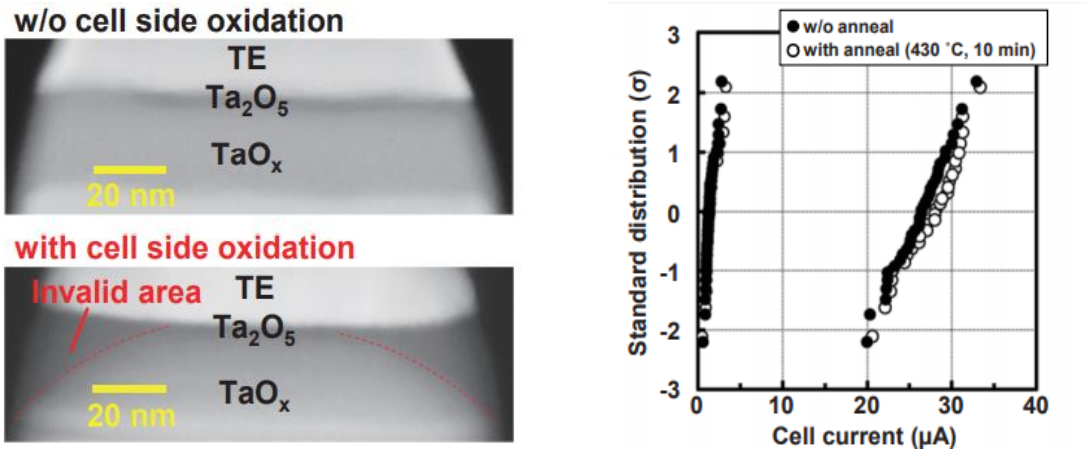


Figure 2-1 a) TEM cross-section of sidewall passivated TaO_x based fully embedded RERAM device. Active area (TaO_x) and passivated area (Ta_2O_5) present strong contrast under TEM b) HRS and LRS distributions before and after annealing (430 °C, 10 min).

in Ta_2O5 RRAM cells by localizing the switching zone away from the MIM sidewalls [131]; -where defects are generated owed to the etching process of the stack (Figure 2-1)- or by optimizing the integrated reactive electrode geometry to favor switching in a local zone [132]. Moreover, multiple groups have investigated the mechanism of alloying HfO_2 in order to engineer ternary oxides with superior thermal stability both theoretically and experimentally. Al_2O_3 has been found to have superior thermal stability, associated with its higher barrier height for the migration of oxygen vacancies, when compared with HfO_2 [133], [134], [135]. It nevertheless comes at the expense of impractically high forming voltages, even for ultra-thin layers [136]–[138]. HfO_2/Al_2O_3 bilayers or nano-laminates have demonstrated superior thermal stability of the programming states with respect to single layer HfO_2 [41], [139] while having lower forming voltage as compared to an Al_2O_3 single layer of comparable thickness and stoichiometry.

Both Si and Al doped HfO_2 has been reported to exhibit superior data retention characteristics [134]. Despite that, contradictory results have been reported for Al as regards the impact on the forming voltage. Both reduction [Stanford] [140] and increase [39], [134] of the forming voltages of Al doped HfO_2 layers have been reported by different groups using different alloying methods (namely doping by supercycling ALD, ion implantation and CVD in the previously cited works). This indicates that the physics of the doping mechanism is not yet fully understood and seems to be very sensitive to the choice of the deposition technique, process conditions and integration method.

In this chapter, the alloying mechanism of HfO_2 doped with Al and Si by means of ion implantation is studied in detail. Ion implantation is chosen for its tunability, accurate control of the dopant concentration an across-wafer uniformity as will be detailed in the following paragraph. Preliminary electrical results are reported and interpreted in combination to ab-initio calculations and material physicochemical characterization. Chapter 3 focuses on the integration of the given alloys, targeting to relax the memory cell forming constraints, without vertical scaling of either the oxide or the reactive electrode thicknesses. At the same time, the impact on data retention endurance and other aspects of reliability of large arrays is also studied in detail as a function of the doping species.

2.2 Ion Implantation in HfO_2: The case of Si and Al doping

Creating a tunable, homogeneous alloy of HfO_2 requires precise control of the dopant concentration. Out of the industrially compatible HfO_2 growth techniques available, while the alloy stoichiometry can be controlled using CVD, the technique provides the least control of both dopant concentration in terms of target accuracy as well as across-wafer variability when considering ultrathin oxides. On the contrary, ALD allows synthesizing HfO_2 a monolayer at a time with as low as 1.3% oxide thickness variation across a 300 mm wafer with short process times. The synthesis of a ternary alloy like HfSiOx or HfAlOx by CVD or PVD requires alternating deposition cycles of ultra-thin layers of binary oxides, making across-wafer uniformity both in terms of total film thickness and in terms of film properties extremely challenging for sub 10nm oxides, such as the ones typically used in ReRAM applications. Even though it is possible to synthesize both HfAlOx and HfSiOx by alternating ALD cycles of HfO_2 and AlO_2 / SiO_2 [134] with excellent uniformity, it is impossible to achieve sub-stoichiometric oxides or alloys thereof owed to the ALD window [141], [142]: ALD is a self-limiting process that leads to materials of almost perfect stoichiometry. As a result, trap rich oxides or alloys thereof cannot be synthesized using ALD alone, which only leaves thickness scaling as the only parameter available to tune RRAM forming voltage constrictions.

To circumvent this limitation, we propose to dope the oxide by means of ion implantation. In this way, the dopant introduced has the capacity to modify the stoichiometry of the material as the initial oxygen content in the oxide does not change which is expected to

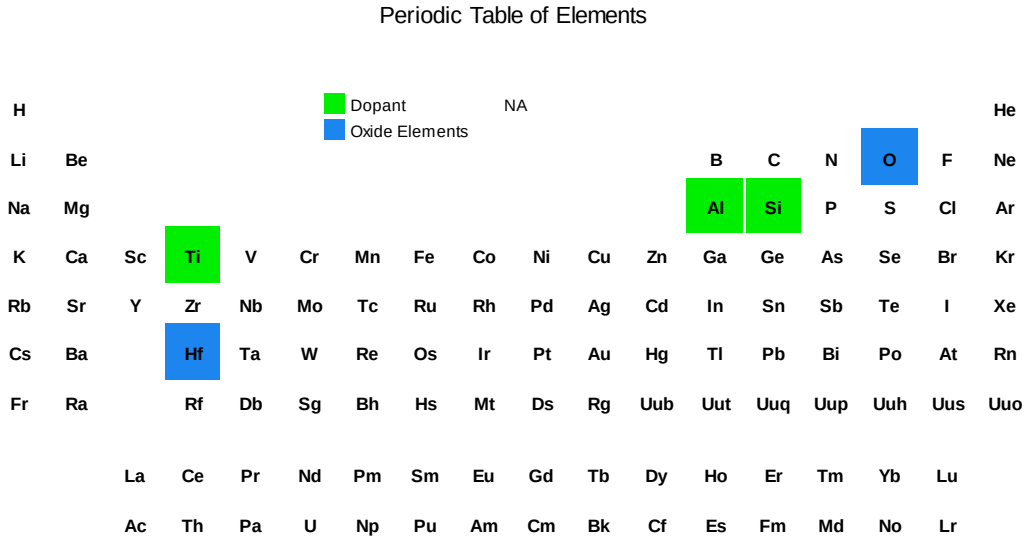


Figure 2-2 The periodic table of elements showing the elements used in the RERAMs developped in this work, notably: HfO_2 , the top electrode reactive electrode (Ti) and the oxide dopants (Al, Si).

increase the electrically active bulk defects thus enhancing bulk conductivity. This is nevertheless sensitive to the thermodynamic stability of the target alloy as well as the bonding state of the dopant inside the HfO_2 matrix. From this prospective, the choice of dopant is critical.

Nevertheless, practical limitations also apply to this approach as well. Primarily, doping by ion implantation (either direct [143], recoil [144] or plasma [145]) leads to doping variation across the depth of the implanted material, which follows the profile of the implanted species. In the case of direct implantation, it follows the Pearson distribution [146]. The lower and upper limits of the approach are both tied to the thickness of the target oxide. The thinnest oxide that can be controllably doped is limited by the lowest energy available in the implanter, for which the position of the dopant profile inside the oxide as well as its tail can be reasonably controlled. The thickest oxide is limited by the dose required to achieve the target dopant concentration of the implanted species, which leads to impractically long process duration. Finally, as ion acceleration in the ion implantation process, and therefore the projected range of the profile, depends on their atomic weight, the previously described limits are expected to vary according to the dopant of choice. As the atomic weight of Al is very close to the one of Si (Figure 2-2), the former's implantation profile is practically identical to that of Si for given implantation conditions.

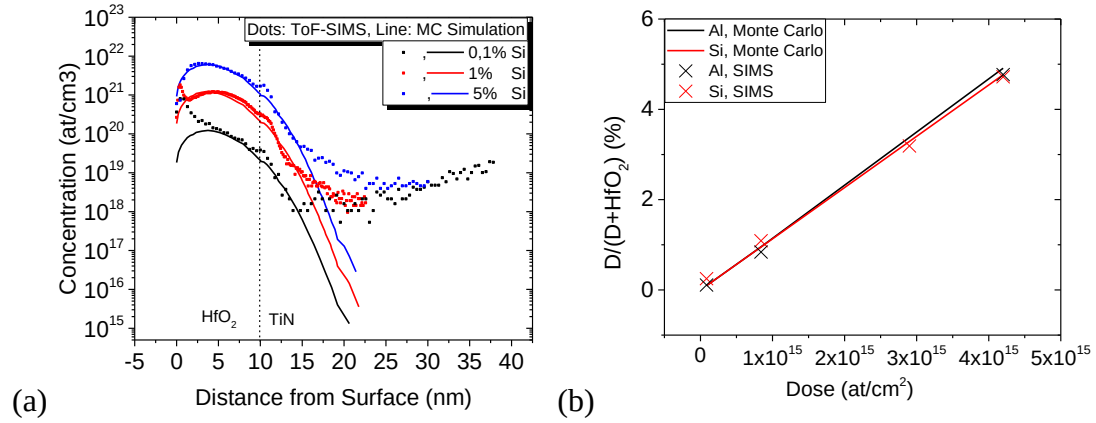


Figure 2-3 (a) Monte-Carlo simulated (continuous line) Si implant and ToF-SIMS (point) profiles for ion implant doses corresponding to (0.1-5) % concentration of the Si ion in a 10 nm HfO_2 layer. (b) Correlation of implanted ion dose to dopant concentration for values extracted by simulated (continuous line) and experimentally measured (points) profiles in the material for Al and Si dopants.

The thickness of the oxide is fixed to 10 nm, which is sufficiently thick to concentrate the dopant profiles but also reasonably thin for practical RRAM applications. The implantation energy is chosen such that the projected range of the implanted profile is in the middle of the oxide layer, expected to render the alloy properties symmetric with respect to the distance of the Metal/oxide interfaces of the top and bottom electrodes, respectively. The experimentally measured Si doping profile is in good agreement with the MC simulated one (using the Sentaurus sprocess simulator) as shown in Figure 2-3a.

$$C\,(\%) = \frac{C_D}{C_D + C_{HfO_2}} = \frac{\frac{I_{D,impl}}{t_{ox}}}{\frac{I_{D,impl}}{t_{ox}} + \frac{3\rho_{HfO_2} N_A}{(M_{Hf} + 2M_O)}} \qquad (2.1)$$

The concentration of implanted dopants per total number of atoms, can be extracted using equation (1.1) where: $I_{D,impl}$ is the numerical integral of the dopant profile across the thickness of HfO_2, t_{ox} is the oxide thickness, ρ_{HfO2} the HfO_2 mass density and M the atomic weights of hafnium and oxygen multiplied by the respective number of atoms of each element in the HfO_2 molecule (3) and N_A Avogadro's number. The dopant concentration is calculated in terms of number of atoms of dopant per *total number of atoms* in the HfDOx supercell thus giving rise to a multiplicity factor of 3 in the term corresponding to the atomic concentration of HfO_2. Figure 2-3b shows the correlation between the implanted species dose and the resulting percentagewise concentration for dopants

implanted at 4 keV. The dose to concentration relation is linear and the target values extracted by using the MC simulated implantation profiles lie in excellent agreement to the ones extracted from the Si and Al experimental profiles measured by means of TOF-SIMS for identical implantation conditions.

The dopant concentration is chosen in the range of 0.1% to 5%, after examining the possible scenaria of the bonding states of an Al or Si atom when introduced into an HfO_2 cluster, by means of ab-initio calculation. The calculations were performed in collaboration with Boubacar Traoré and Philippe Blaise of the Advanced Memory and Simulation groups respectively, and provide a theoretical toolset that gave me the necessary theoretical "mind tools" to design the material analysis and help understand the device behavior. In the following, I provide a common framework to understand the physics of the doped HfO_2 system by interpreting the theoretically calculated properties as a function of the experimentally measured material and device electrical characteristics.

2.3 Alloying of Al and Si dopants in HfO_2: Electronic Properties

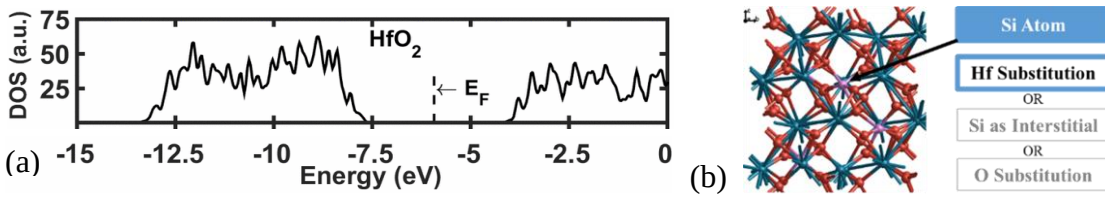


Figure 2-4 (a) Shape of the band gap of HfO_2 as calculated by DFT theory from [32] (b) Schematic representation of the HfDOx supercell. Hf atoms represented with blue, O atoms with red and dopant atoms in purple.

The typical bandgap structure of m-HfO_2 is in the order of 4.8 eV as shown in Figure 2-4a from [146]. The Fermi level lies close to the mid gap region. The underestimation of the HfO_2 band gap as compared to the typical experimental values of (5.1-6) eV [147] arises from the choice of functional used to describe the exchange-correlation energy. For these calculations the Perdew-Burke-Ernzerhof functional in the Generalized Gradient Approximation (GGA-PBE) were used in the DFT calculations [148], [149] using SIESTA [150], [151]. Double Zeta Polarized (DZP) basis set with an energy shift of 50 meV and a mesh cutoff of 300 Ry were used. Starting from a 2x2x2 (96 atom) supercell of m-HfO_2, Si/Al atoms are added while allowing for both the atomic coordinates and geometry to be

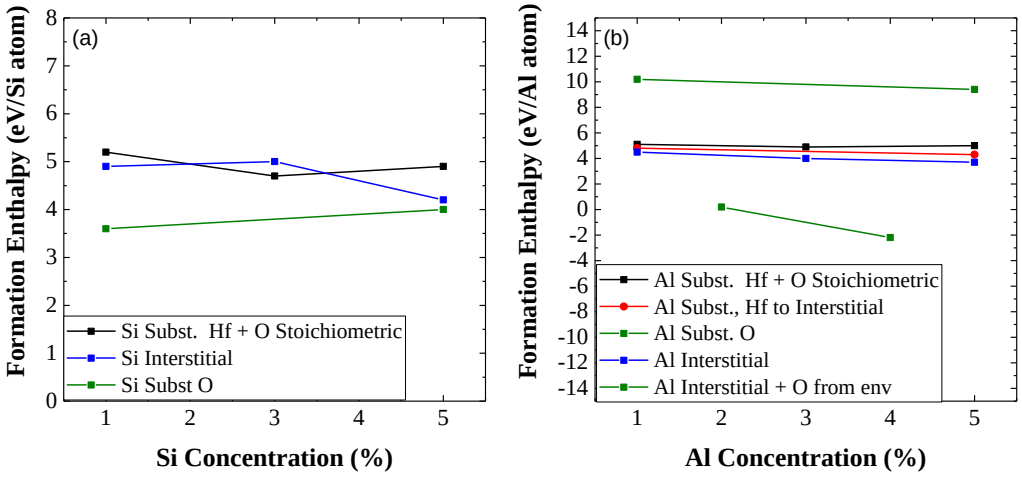


Figure 2-5 Formation Enthalpy for each of the examined scenaria of Hf or O substitution or interstitial dopant for a) Si and b) Al dopants.

relaxed. The supercell size is enough to reproduce qualitative and semi-quantitative characteristics of the alloy material down to approximately 1% dopant concentration (1 out of 96 atoms). M-HfO_2 has been used as the material phase of reference as ALD deposited HfO_2 films can be considered amorphous only *macroscopically*. The ALD as-deposited HfO_2 film typically exhibits monoclinic nano crystalline regions of the order of 1 nm (as will be discussed in the end of this chapter). Hence, in the scale of ab-initio calculations m-HfO_2 suggests the best phase of reference.

When Si or Al atoms are introduced in an HfO_2 cluster, they can *i)* substitute either O or Hf atoms, or *ii)* occupy interstitial positions in the local cell. Alloy cluster formation enthalpy, and the qualitative characteristics of the material band gap as a function of each clustering configuration were studied by calculating the density of electronic states (DOS) for each of the scenarios considered. Even though secondary reactions of dopants with the environment such as oxidation of the dopant from the ambient atmosphere have not been considered owed to the complexity of the calculation.

$$\Delta H_i^X = E_t(HfXO_2) - E_t(HfO_2) - n\cdot\mu_X \quad (2.2)$$

$$\Delta H_{Hf/O}^X = E_t(HfXO_2) - E_t(HfO_2) - n\cdot\mu_X + n\cdot\mu_{Hf/O} \quad (2.3)$$

The formation enthalpies for interstitial (ΔH_i^X) and substitutional ($\Delta H_{Hf/O}^X$) insertion of dopant in HfO_2 were calculated using Eq. 2.1 and 2.2, for n number of X dopants. The

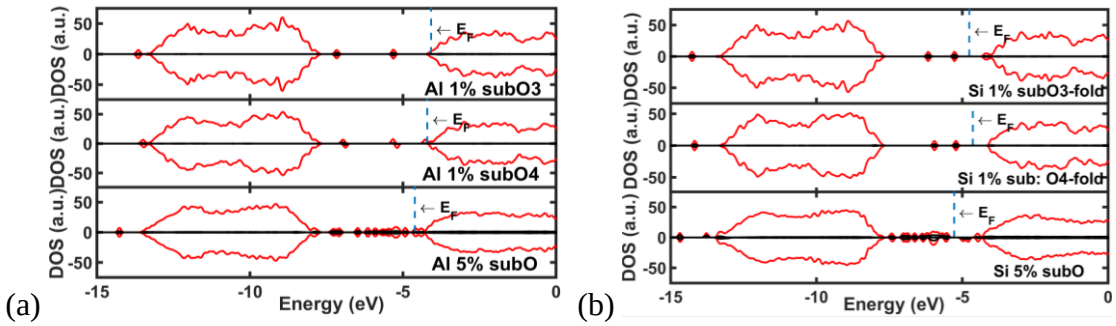


Figure 2-6 Representation of the calculated DOS of HfO_2 when oxygen substitution by the dopant is considered for a) Al and b) Si dopants as a function of the dopant concentration.

total energy E_t of the $HfXO_2$ supercell is referenced to its bulk counterpart (FCC-Al or diamond-Si); μ is the chemical potential.

In Figure 2-5 a, b we summarize the formation enthalpies of the ternary alloy for each dopant and dopant position assumption, respectively. Moreover, in the case of aluminum, additional scenaria are shown, notably that of the substitution of Hf by Al in a hypoxic environment as well as the reaction of Al interstitials with Oxygen from a nearby reservoir. Evidently substitution of Hf with or without the addition of oxygen requires the same amount of energy for both dopants and is again comparable to the energy needed to be spent for the dopant to be inserted as an interstitial. For Si, substitution of oxygen seems to be mildly favored in the model (Figure 2-5a). The only scenario that represents a spontaneous reaction is that of Al interacting with a readily available oxygen surplus inside the material (Figure 2-5b).

Oxygen substitution by Al or Si gives rise to traps inside the alloy bandgap proportionately to the dopant concentration for both dopants (Figure 2-6 a, b). The Fermi level of the doped oxide shifts towards more positive values with respect to the midgap, indicating that the resulting alloy is n-type like. A material with such configuration is expected to be more conductive with respect to pure HfO_2 with electrons being the conductive carriers as a result.

The substitution process of Hf by either Si or Al atoms can give differing band gap characteristics according to the availability of oxygen in the vicinity. If oxygen is can be supplied during the substitution of Hf a trap-free bandgap for both Si and Al (Figure 2-7 a, b) is recovered. If on the other hand, oxygen cannot be balanced as the dopant occupies Hf substitutional positions, the substituted Hf atoms occupy interstitial positions that generate conductive states inside the oxide band gap (Figure 2-7c). Considering that the

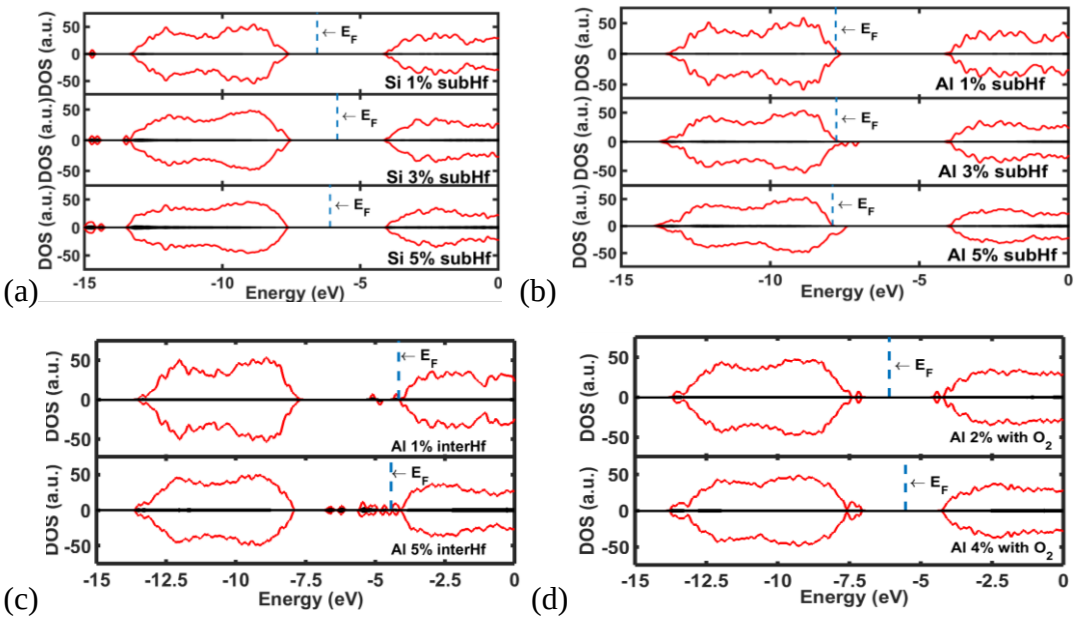


Figure 2-7 Representation of the calculated DOS of: HfO_2 when Hf substitution by the dopant is considered for Al (ab) and Si (b) dopants while alloy stoichiometry is maintained. (c),(d) DOS of Al doped HfO_2 when Al occupies interstitial positions in the HfO_2 cell (c) and when interstitial Al reacts with an external oxygen reservoir.

substitutional process formation enthalpy is almost identical for both dopants and does not depend on the lack or surplus of oxygen close to the reaction site (Figure 2-5) it is expected that Hf substitution by Si will push Hf atoms to interstitial positions with similar bandgap characteristics.

The case where the dopants occupy interstitial positions in the HfO_2 supercell is thermodynamically equivalent to the substitution of Hf previously described as it has very similar formation enthalpy (Figure 2-5). In terms of electronic properties it is expected to generate conductive states inside the band gap of HfO_2 (Figure 2-8a, b).

The case where Al is added into the system while an oxygen source is also present has been studied (Figure 2-7d). This case is of particular importance, as it constitutes the only scenario whose formation enthalpy is negative: As long as Al atoms are not tightly bound to a molecule they will immediately oxidize without the need to add extra energy. This observation suggests that dangling bonds in the HfAlOx cluster could naturally with molecular oxygen forming alumina clusters. This suggests that the most thermodynamically stable condition of the HfAlOx cluster are sub clusters of the alumina and Hafnia binary oxides. Finally, both dopants give very similar characteristics in terms of DOS if inserted in HfO_2 in interstitial positions (Figure 2-8 a, b), both leading to the appearance of conductive states inside the band gap.

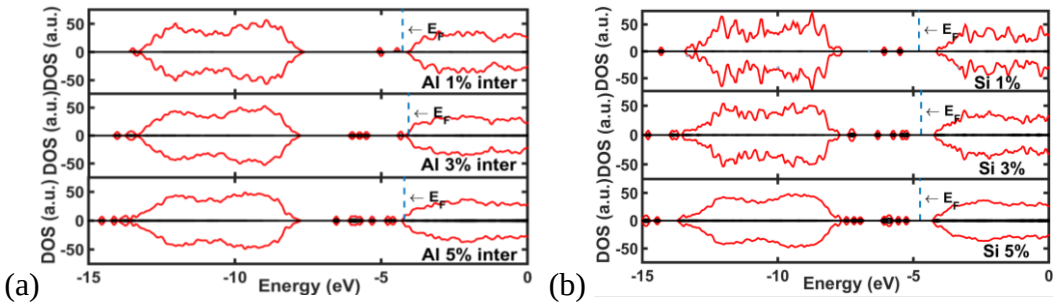


Figure 2-8 Representation of the calculated DOS of HfO_2 when the dopant occupies interstitial positions for (a) Al and (b) Si dopants.

From these calculations alone, there is no clear property that can strongly justify the exclusion of any one scenario. Nevertheless, from a thermodynamic standpoint Hf substitution by the dopant or the dopant taking an interstitial position are equi probable scenaria that produce materials with very different electronic densities of states. Particularly for the case of Hf substitution the resulting DOS will heavily depend on the stability of the generated Hf interstitials.

Therefore, a conclusive answer as regards the dominant physical processes during the alloying of Si and Al with HfO_2 molecules cannot be reached by ab-initio calculations alone. This is because i) the assumptions made are potentially incomplete ii) More than one processes can occur in sequence iii) in practical application the choice of the technique to synthesize the material can strongly impact the resulting oxide and iv) the impact of the interfaces and by extension the ambient is not considered as periodic conditions have been imposed for this study.

It is worth noting that in all cases where traps appear in the oxide band gap, the resulting DOS of the trap sites are comparable in magnitude indicating a flat DOS. This is particularly important as it can be experimentally detected in the electrical properties of the film. The trap sites appear closer to the edge of the conduction band of HfO_2 indicating that the resulting oxide is more likely to be n-type like, also supported by the shift of the Fermi level closer to the conduction band edge.

As a result, calculations alone cannot collapse realistic with nonrealistic scenario and device electrical data are indispensable to complement our understanding further.

2.4 Pristine device electrical characteristics of Si/Al implanted HfO_2 based RRAM

2.4.1 Device Fabrication

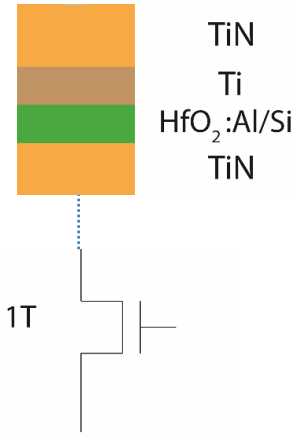


Figure 2-9 Schematic representation of a 1T-1R embedded ReRAM cell using nMOS transistor as the pass gate.

1T-1R devices have been integrated in the BEOL of a 300 mm CMOS process, employing a 65 nm CMOS technology. The OxRAM structure consists of a 30 nm thick TiN bottom electrode deposited by PVD, followed by a 10 nm HfO_2 layer deposited by water chemistry ALD at 300°C (Figure 2-9).

HfO_2 was implanted with either Si or Al at 4 keV, except for the reference wafer. Monte Carlo simulation was used to calibrate the implantation process. The profiles of implanted Si and Al are expected to have very similar characteristics as previously described in §1.2, Figure 2-3. The implantation dose is such, to target approximately (0.1, 1, 5) % overall concentrations of D/(D+ HfO_2) in HfO_2, where D stands for the implanted species atomic concentration by unit volume, namely Al and Si (Figure 2-3b)**.** The stack is completed by the deposition of a 10 nm Ti oxygen scavenging layer and 50 nm TiN top electrode, both deposited by PVD. The MIM structure is then patterned and encapsulated. Finally, standard BEOL CMOS process is used to form the electrode contacts.

2.4.2 Electrical Characteristics of the Devices in Pristine State

First we investigate the impact of the implanted species and dose on the low field conductivity of the material by extracting the pristine state resistance from quasi static I-V measurements at room temperature in the range of (0-200) mV (Figure 2-10). For the reference and Al implanted HfO_2, the pristine resistance ($R_{Pristine}$) remains pinned to values

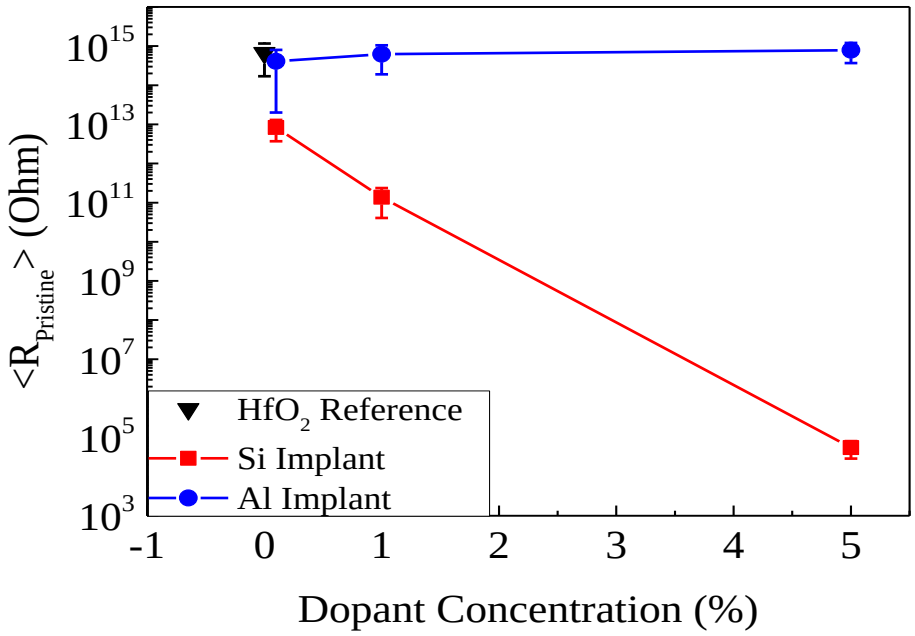


Figure 2-10 Evolution of the pristine resistance as a function of dopant concentration for the Si, Al implanted species in a HfO_2(10 nm)/Ti(10 nm) ReRAM stack.

close to 10^{15} Ohm, owed to the resolution limit of the parameter analyzer. On the contrary, $R_{Pristine}$ decreases monotonically with implanted Si fraction, spanning a 10 orders of magnitude drop from the reference sample to the highest Si dose.

The electrical signature of the pristine devices, when seen under the light of the ab initio calculations discussed in §2.3, points towards a strong increase of the density of conductive states in the Si implanted devices as compared to the reference and the devices implanted with Al. As all device variants have the same technological parameters, other than the implanted species chosen, it is safe to conclude that this enhancement of the conductivity of HfO_2 can be associated with the alloying mechanism of the implanted Si species.

Despite that, the dominant alloying mechanism for either dopant is still not entirely transparent. What can be inferred is that in the case of Al the mechanisms that explain the experimental pristine state evolution versus the doping process is either Hf substitution with the addition of oxygen from a nearby source so that the system still maintains full stoichiometry, or the oxidation of the aluminum species. For these to processes the resulting alloy is trap-free. Furthermore, the enhancement of conductivity observed in the case of Si implantation in HfO_2 can be attributed to an increased trap density on the one

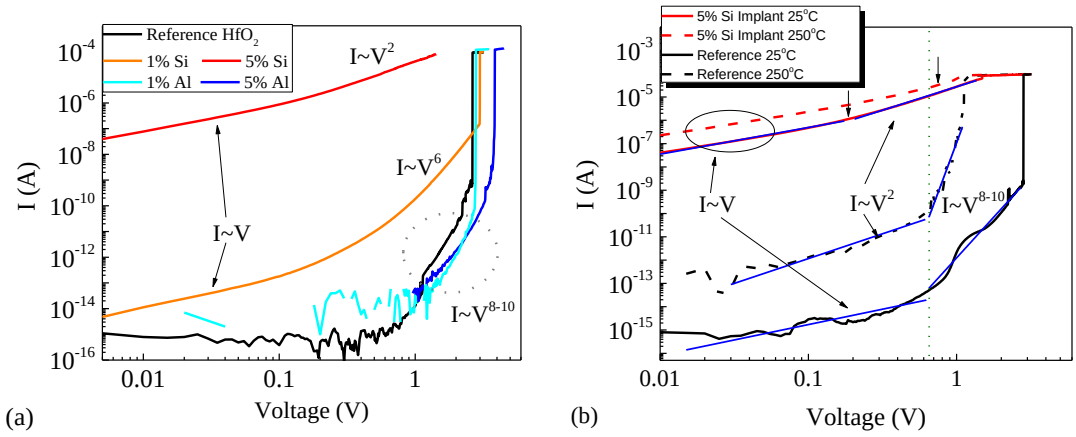


Figure 2-11 Log-Log diagrams of the quasi static I-V characteristics for (a) the evolution of the current as a function of the implanted species and doping level, compared to the reference. (b) Pre-forming conduction characteristics of the reference and 5% Silicon implanted samples at room and elevated temperatures, respectively.

hand, yet the nature and origin of such traps is still unclear. The low-field conductivity measured (in terms of pristine resistance in Figure 2-10) does not provide sufficient information on the either nature, origin or positioning of the conductive centers. It is also unclear if such modification of the conductive properties of the stack is associated to the HfO_2:Si alloying process alone or if the Ti layer plays an additional role in the generation of traps in the presence of Si, close to the HfO_2:Si / Ti interface of the ReRAM reactive electrode.

In Figure 2-11a, high-resolution quasi-static forming I-V curves are shown for the reference, Si and Al samples in log-log scale. On the one hand charge transport in the undoped and Al doped samples presents with weak, ohmic-like voltage dependency characteristics at low voltages, whereas it is essentially dominated by a strong exponential $I \sim V^{(8-10)}$, field-dependent conduction scheme, activated above ~0.7V. Increasing the Al fraction seems to shift the maximum forming voltage while the onset of the field controlled mode remains the same, suggesting a decrease in the material static dielectric constant [6]. On the other hand, introducing Si gradually increases the base current while at the same time decreases the exponential dependency on the applied voltage, as illustrated in Figure 2-11a. In the 2 extremes of the reference and 5% Si implanted samples the distinction between different conduction regimes is investigated versus the increase of temperature: For the reference, the quadratic dependency overtakes the linear at high temperatures; nevertheless, the onset of the field dependent conduction seems temperature independent.

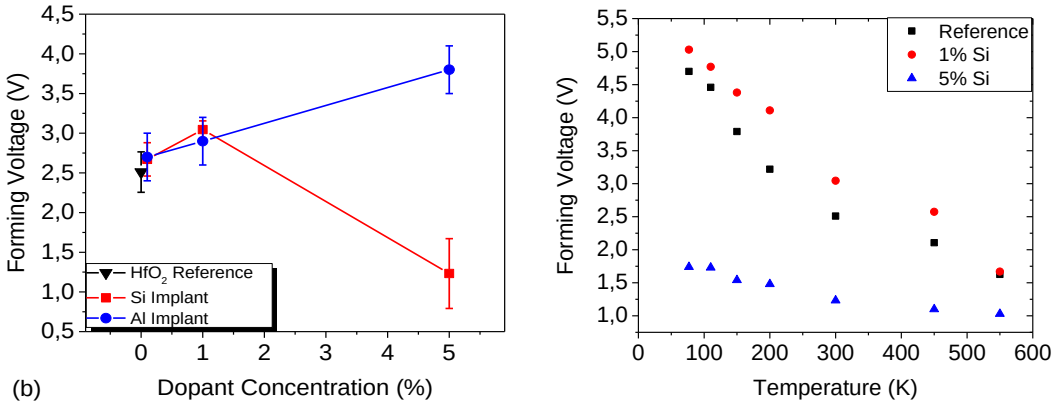


Figure 2-12 (a) Forming voltage for Si/Al doped and reference samples at room temperature. (b) Forming voltage evolution as a function of temperature for the reference and Si doped samples from 77 K to 550 K.

For the 5% Si implanted devices, the strong exponential field dependency vanishes. At the 5% limit, the pre-forming current of the Si-doped samples is ohmic until ~0.2V above which it becomes quadratic. At high temperatures, the ohmic region extends as high as ~0.8V, very close to the forming voltage (Figure 2-11b).

This type of field and temperature dependence presents similar qualitative traits to a space charge limited conduction (SCLC) scheme like variable range hopping (VRH) or phonon assisted tunneling [53], [152]. For the reference and Al implanted samples the low low-field current values suggest an almost defect-free band gap as significant current is only measured at high fields, and can be interpreted by either Hf substitution by Al where stoichiometry is maintained or Al oxidation according to ab-initio results. Temperature possibly acts to excite more traps, with limited effect owed to a low density of states [152]. On the other hand, the increase of the base current and gradual linearization of the I-V current transfer characteristics in the Si implanted devices point towards an increase in the density of bulk defects participating in charge transport. The latter consistent with any of the atomistic configurations where Hf interstitial are produced, oxygen is substituted by the dopant or the dopant occupies interstitial positions.

Furthermore, considering the implanted Si profile lies in the center of the HfO_2, we can directly tie the implantation of Si with the creation of bulk defects that form a flat density of electronic states inside the band gap of the material. Nevertheless, either interstitial Si or Si substituting Hf knocking it to interstitial positions or substituting oxygen, all give rise to traps inside the bandgap. In the following, we proceed to demonstrate the impact of

doping on the forming process of RRAM devices and continue in section 1.5 with physicochemical and structural material characterization that can provide more detailed insight into the physics of the material.

2.4.3 Dependence of the RRAM forming voltage on dopant concentration and temperature

The forming voltage (V_f) dependence on the dopant concentration is shown in Figure 2-12a for both Si and Al implanted devices. While Al implanted samples exhibit a monotonic increase of V_f with implanted dose as compared to the no-implant reference, the V_f of the Si implanted samples exhibits a "bell-like" shape, after which an acute drop is observed at the 5% Si content. The latter trend is observed for a wide range of temperature in Si, as illustrated in Figure 2-12b where the median forming voltage dependence with temperature is shown, for the case of Si implanted devices and for the undoped reference HfO_2. Interestingly, the thermal dependence of V_f is stronger than the HfO_2 no-implant reference when V_f is higher (0.1%-1% Si case), and weaker when V_f is lower (5% Si-case). For the latter case, where V_f is close to 1 V in the (77-550) K range, an almost forming-free device is formed, as V_f tends to reach values comparable to the set voltage. The behavior of the forming voltage can be explained either by means of the modification of the dielectric constant using the image provided by MacPherson's thermochemical model [64] or by means of defect generation as a result of oxygen ion migration facilitated by the existing oxygen vacancies engineered [153], and will be discussed in more detail in Chapter 3.

Considering the monotonic evolution of the pristine resistance reduction with increasing Si concentration, the higher median values of forming voltage observed in the 1% Si implanted HfO_2 RRAM devices, as compared to the no-implant reference seem contradictory at first look. As the corresponding values for a given temperature of the reference and 1% Si implant devices are generally close in absolute value, this anomalous evolution can be the effect of a wafer-to-wafer variability more significant than the device-to-device variability across wafer for a given split. Hence, it is not captured by the error bars, which represent the standard deviation (Figure 2-12a) of measurements coming from samples from the same wafer. It also suggests that the traps induced in this region of Si implanted doses does not create significant enough modifications to the HfO_2 switching matrix.

2.5 Material Properties of the HfO_2:Si and HfO_2:Al systems alloyed by ion implantation

2.5.1 Physico-chemical signatures of the alloying mechanism

While the electrical characterization of the fabricated RRAM can partially link the electronic properties of the material to the experimentally measured conductivity and provide an explanation on the behaviour of the forming voltage, it does cannot provide further insight on the alloying mechanism itself. This is because i) traps arise inside the band gap for both oxygen substitution and interstitial dopant for both Al and Si. ii) Hf substitution can provide either a trap full or a trap free band gap, depending on the assumptions on the availability of the nearby oxygen reservoir Figure 2-6 & Figure 2-7). iii) The case of oxygen substitution is highly improbable both because of its distinctively high characteristic formation enthalpy (Figure 2-5) and because it gives rise to traps inside the HfO_2 band gap for both dopants which comes in direct contradiction to the experimentally measured low field conductivity of Al (Figure 2-10).

It is for this reason that more than one characteristic signatures need to be combined, not only from the electrical domain but also from the structural and physio-chemical ones. The expected alloy mass density constitutes a particularly useful quantity as experimental values can be extracted by analysing XRR spectra and compared to expected mass density values calculated by ab-initio methods. XPS analysis is used to provide information on the nature of bonds between atoms in the alloy. Finally, the stability of the dopant in temperature is examined by means of ToF-SIMS. By combining the previous techniques to electrical data, the doping mechanism and underlying physics of HfO_2 can be better understood. In the following, we combine the various signatures mentioned above in order to understand the behaviour of the material and the related dopant when introduced into HfO_2 by means of ion implantation as well as provide a self-consistent explanation on the impact of the doping mechanism on the electrical characteristics observed in each case.

2.5.2 *Sample Preparation*

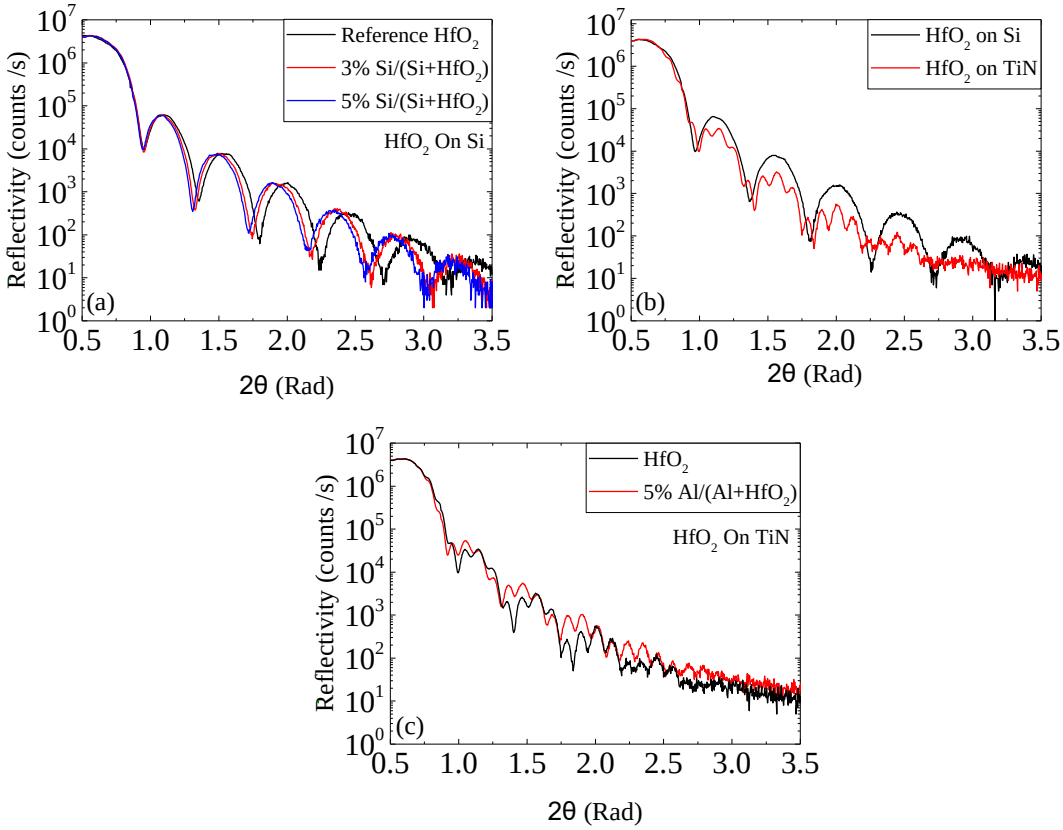


Figure 2-13 Impact of (a) Si implantation (b) the addition of a thin TiN (30 nm) substrate and (c) Al implantation in the presence of a TiN substrate, on the XRR Reflectivity spectra of a thin HfO_2 layer.

As the typical ReRAM MIM capacitors utilize TiN electrodes, we create two sample families: Type I, where HfO_2 is deposited directly on the Si bulk and type II where HfO_2 is deposited on a 30 nm thick PVD deposited TiN layer. The thickness of HfO_2 is fixed at 10 nm, deposited by water chemistry ALD at 300 ^oC on 300 mm Si [100] wafers. The native Si oxide was removed by means of RCA cleaning. Type I samples serve to simplify the deconvolution of XRR spectra while type II samples act to investigate the impact of the presence of the TiN in the stability of the alloyed material. For the case of Al doping only samples on TiN were fabricated. Following the HfO_2 deposition Si and Al have been implanted at room temperature with Si or Al at 4 KeV with doses such to achieve 1%, 3% and 5% concentrations (Figure 2-3b). Both Dopant:HfO_2 on Si and on TiN were studied.

Next, the samples are annealed at 400 ^oC, 600 ^oC and 800 ^oC for 2 min under He atmosphere using RTA. This minimizes the temperature ramp-up and ramp-down times during the annealing process as well as provides with an inert ambient, ensuring minimal

to no secondary reactions such as in-situ oxidation inside the RTA chamber. Hence, the annealing process is not expected to introduce parasitic oxygen sources. These annealing steps target to study material response in temperatures comparable to the BEOL and FEOL process steps, respectively.

2.6 Impact of Si, Al implantation on the mass density of HfO_2

X-Ray reflectometry (XRR) is a non-destructive structural characterization technique where an X-Ray beam is directed towards the material stack, under varying incident angles θ. The reflected beam at 2θ, is then collected providing a direct measurement of the stack reflectivity at a given angle. By fitting the reflectivity curve to a model for the material stack, information on the mass density, layer thickness and interfaces can be acquired. The technique can be used in all monocrystalline, polycrystalline and amorphous samples. In this study a Ka-Cu X-Ray source was used and the scan angle was varied from 0 to 3.5 rad. Owed to the high penetration depth of the x-ray beam (~ 1 μm) the reflectivity measurement is sensitive to the sum of the layers in the stack of interest (§1.5.2).

The characteristic reflectivity spectrum of HfO_2 implanted with Si is shown in Figure 2-13a. For angles smaller than 0.8 rad the incident x-ray beam experiences total reflection and the incident x-rays do not penetrate into the material. At higher angles, as the beam penetrates into the material part of it is absorbed from the top layer, resulting in a drop of reflectivity. If the stack comprises of more than one layers, the difference in the electronic density of states and fill factor of each material modulates the x-ray refractive index and gives rise to oscillations in the spectrum known as Kiessing fringes [154], [155]. The thicker the film the smaller the period of such oscillations. Moreover, the amplitude of the oscillations as well as the critical angle, provide information on the mass density of the material; under the assumption of a given layer thickness and given density of electronic states for each layer. Finally, as the number of layers of the stack increases, diffraction is enhanced owed to the increasing number of interfaces. This gives rise to a more complex spectrum, as is the case for the type II samples with HfO_2/TiN bilayer on Si bulk (Figure 2-13b). Doping HfO_2 with Si or Al acts to shift the position of the secondary diffraction peaks towards lower angles (Figure 2-13 a & c).

To extract meaningful information for the XRR reflectivity spectrum, a model for each case is constructed, defining the target structure nominal thickness and expected density

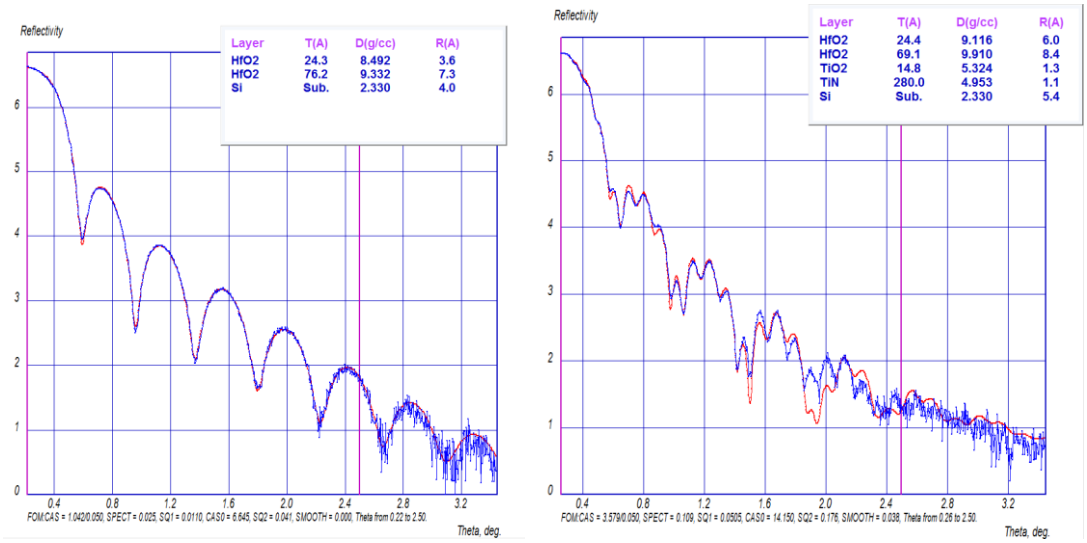


Figure 2-14 Model fits solving the structure of Si doped HfO_2 (a) on Si and (b) on TiN. Additional layers need to be imposed to reach convergence.

for each layer. The measured spectra are analysed by using the JVXrrs fitting software. A generic algorythm is used in a recursive fashion to fit each spectrum with an appropriate model corresponding to each stack, leading the refinement of the fitted layer thickness, mass density and interface roughness values. In the following we focus more in the extraction of density as i) it can be directly correlated to ab-initio ii) no significant variation of the target thickness for each layer is expected owed to the precision of ALD deposition for HfO_2 and PVD deposition for TiN iii) Si/HfO_2 and TiN/HfO_2 interface roughness is considered of secondary importance owed to the nature of both ALD and PVD.

In the case of doping by ion implantation, the material properties are expected to vary to a certain extent, as a result of the shape of the profile of the implanted species. Indeed, when fitting the experimental data to a model, convergence cannot be reached if only one HfO_2 layer is assumed for the doped samples. As the material matrix is modified both the shape of DOS and fill factors wi d_{av} =9.13 g/cc Si or Al doping. To acco d_{av} =8.74 g/cc plit the HfO_2 layer in 2 dummy layers to facilitate the convergence of the generic algorithm, essentially accounting for local variations in the material mass density. The final value of mass density of the physical HfDOx (where D stands for either Al, Si) is then taken as the weighted average of the mass densities of each dummy layer, where the dummy layer fitted thickness is used as the weight.

In Figure 2-14 the fitted spectra of HfSiOx on Si and on TiN (Figure 2-14 a, b respectively)

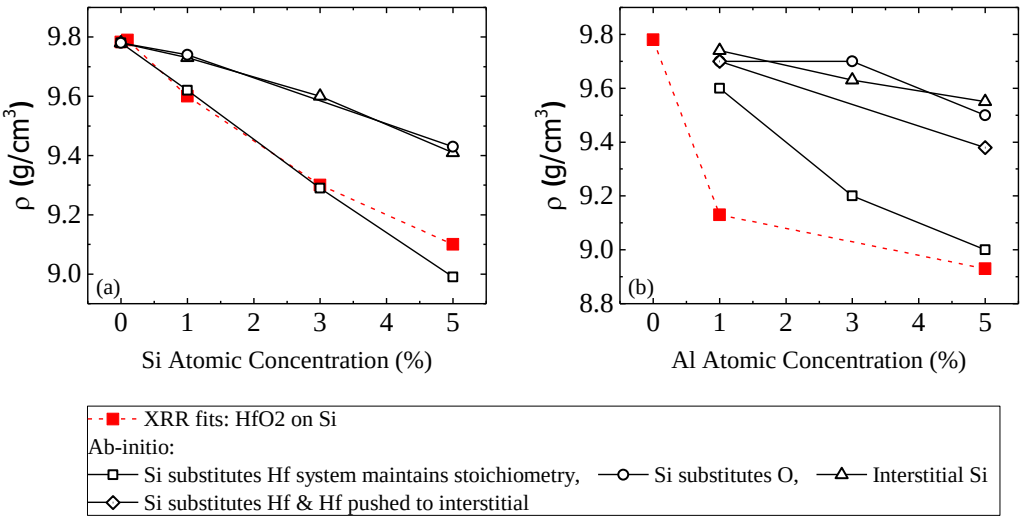


Figure 2-15 Evolution of mass density values (a) Si implanted and (b) Al implanted HfO_2 as a function of the dopant concentration. Experimental results (in red) are correlated to ab-initio calculated values (in black) for each of the assumed scenaria.

with 5% Si concentration in HfO_2 are shown. We immediately notice that the extracted mass density of the doped HfO_2 on Si (Figure 2-14a) is higher as compared to the corresponding measure when the TiN substate is present (Figure 1.14b) for a given doping condition. At the same time, the spectum of the on TiN doped HfO_2 shows more complex Kiessing fringes owed to the addition of the TiN layer (thicker wrt HfO_2) as well as the addition of the TiN/HfO_2 interface which enhances diffraction.

Finally, similarly to the doped layer for the samples with a TiN layer below HfO_2, it is imperative to also model the native TiON that occurs owed to the natural exposure of TiN to the ambient post-depostion. These additional contributions to the parameter extraction model add complexity and reduce the speed of convergence, thus making it more challenging to completely deconvolute each layer's properties. As a result, as the mass densities of TiON & TiN layers are far smaller than those of HfO_2 and as the HfO_2/TiN interface is gradually mixed owed to the implantation process the fitted values for the doped HfO_2 on TiN converge to lower values wrt the their on-Si counterparts. For this reason, in the following the ab-initio calculated mass density values are compared to the on-Si doped HfO_2 to minimize the contributions of the HfO_2/TiN interface for the Si dopant which is of primary interest for our application.

In Figure 2-15a the evolution of the ab-initio calculated mass density is shown for each of the scenaria of the configuration of the Si atom inside the HfO_2 matrix (§ 1.3). The

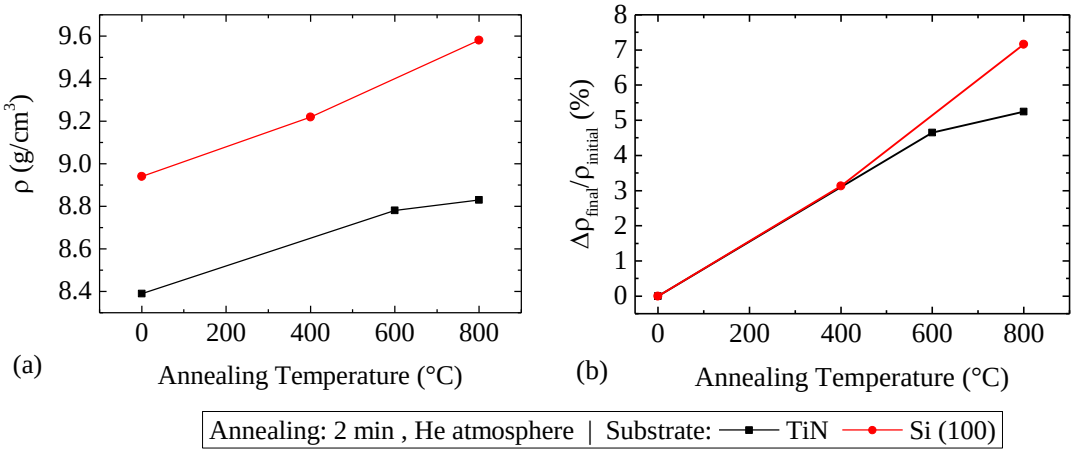


Figure 2-16 Evolution of mass density as a function of annealing temperature for HfO_2 implanted with 5% Si for Si and TiN substrates. (a) Absolute (b) percentage value evolution.

substitution of Hf atom by Si scenario (where stoichiometry is maintained by a source of nearby oxygen) produces mass density values in good agreement with the experimentally extracted values. Similar conclusions can be reached for the case of the Al doped samples.

Nevertheless, Hf substitution with undisturbed stoichiometry cannot provide a common explanation to the opposing trends of the preforming conductivity of the devices (§ 1.4). Moreover, the modification of electronic properties of the doped material could interfere with the model accuracy, as the change in the electronic DOS with each dopant cannot be taken into account in full. It is for this reason that we further investigate the material properties using x-ray photoelectron spectroscopy (XPS).

It must be pointed out that in the ab-initio scenario calculation, the case where Hf is substituted by Si (and the cell is left to fully relax the mesh later) while stoichiometry is maintained is *extreme*. That is because in reality, the algorithm replaces Hf atoms with Si atoms directly. Hence, Hf atoms are directly removed from the system, proportional to the Si concentration, creating a significant change in the total volume of the local cell. The fact that this is the only scenario that leads to values close to the experimentally observed ones, implies a significant amount of disorder in the actual material. In actuality the material density can recover with annealing temperature (Figure 2-16) signifying that even though Hf atoms can be knocked out of their initial positions there is no significant loss of matter in the material matrix (as can be the case of backscattering in recoil implantation

processes). This can lead to significant implications in the material electronic structure and electrical characteristics as will be shown later.

2.7 Binding state of Al and Si in the HfO_2 matrix as a function of the dopant concentration

2.7.1 XPS Measurement Principle

The bind state of Si in the HfO_2 matrix has been studied in the as-implanted material by means of x-ray photo spectroscopy [156]. As x-ray irradiates the sample, photoelectrons from the core states of the atoms in the material are extracted from the atoms near the surface with characteristic kinetic energy that depends on the atom and the nature of the atomic bonds of the material. They are then collected by a Cylindrical Mirror analyser (CMA) and converted into electrical signal. The binding energy of each emitted core electron can be calculated by means of (Eq. (2.4)) where KE the kinetic energy (measured), hν the photon energy from the x-ray source (controlled), φ the spectrometer work function (constant, characteristic of the tool) and BE the binding energy (unknown).

$$KE = h\nu - \varphi - BE \tag{2.4}$$

Even though the penetration depth of the x-rays is in the order of 1μm the meaningful information that photoelectrons can provide is limited to photoelectrons that come from the surface of the sample. That is because photoelectrons emitted deep inside the bulk experience multiple scattering events modifying their kinetic energy and consequently the extracted binding energy. Such photoelectrons eventually contribute to signal background noise. In this study, XPS was performed on wafer level in a 300 mm tool using an Al-Ka x-ray source. The acquired spectra were analyzed using the Thermo Avantage XPS analysis software.

It is important to note that even though XPS is considered a surface analysis technique, for the case of ultra-thin layers such as the 10 nm HfO_2 studied in this work, the active zone where the XPS photo peaks provide meaningful information is in the range of (3-5) nm from the surface. As a result, the peaks carry significant qualitative information about the

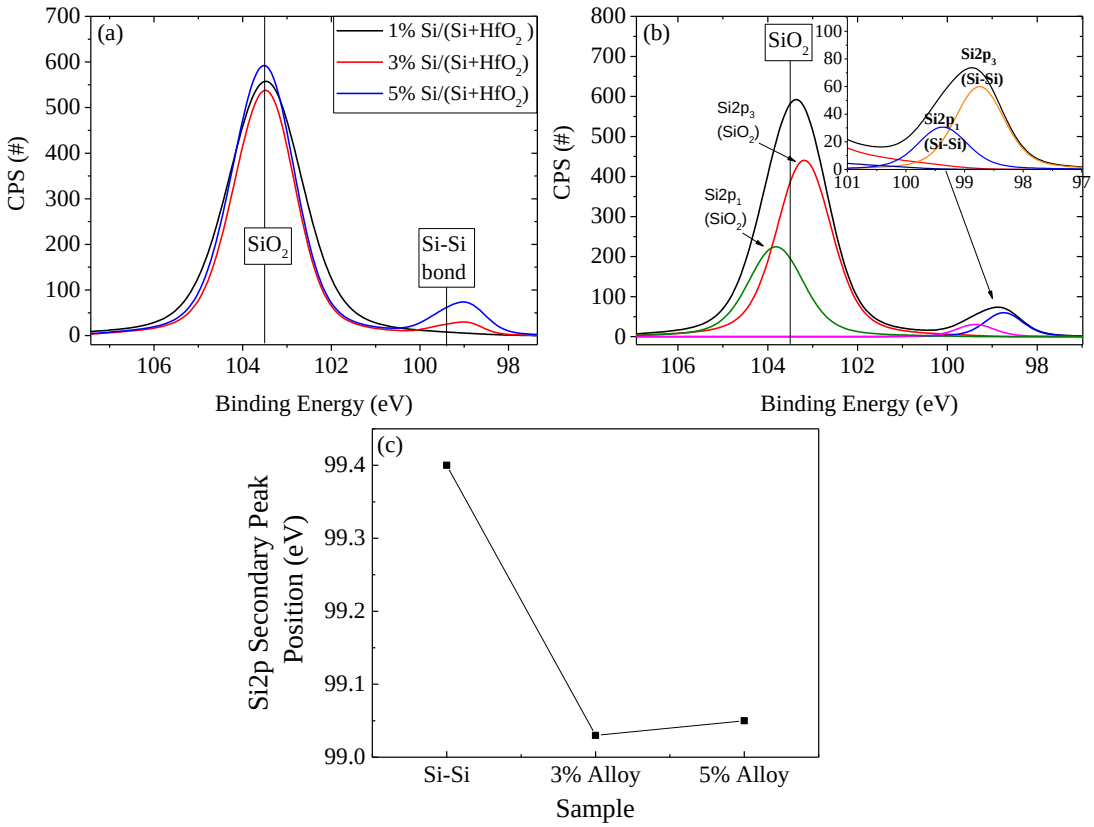


Figure 2-17 (a) Si2p spectrum of Si implanted HfO_2 as a function of the dopant concentration. (b) Deconvolution of the photopeaks of a 5% Si implanted HfO_2 alloy showing the presence of SiO_2 and Si-Si bonds (b) Shift of the position of the Si-Si bonds in the implanted alloy as a function of the Si concentration as compared to the characteristic energy of the Si-Si bond.

presence of bonds for more than 30% of the volume of the HfO_2 film. For all the spectra presented below, the background signal has been removed during peak analysis.

2.7.2 Binding State of implanted Al, Si in HfO_2 as a function of the dopant concentration.

Figure 2-17a shows the Si2p spectrum as a function of the implanted dose for the Si doped samples. As Si is implanted into the HfO_2 matrix, we notice the appearance of a double signature: A stable SiO_2 peak at 103.5 eV as well as a secondary one in the range of the Si-Si bond (99.3 eV), the amplitude of which is directly proportional to the implanted dose. The appearance of the secondary photopeak suggests that Si is not fully oxidized when implanted into the HfO_2 matrix but it rather creates potentially electrically active bonds. This peak is related to the Si_2p3 & Si2p1 orbitals of the Si-Si bond (Figure 2-17b and inset). Moreover, its position is directly impacted by the amount of Si incorporated into the HfO_2 matrix, shifting by approximately 0.4 eV between the reference position of the

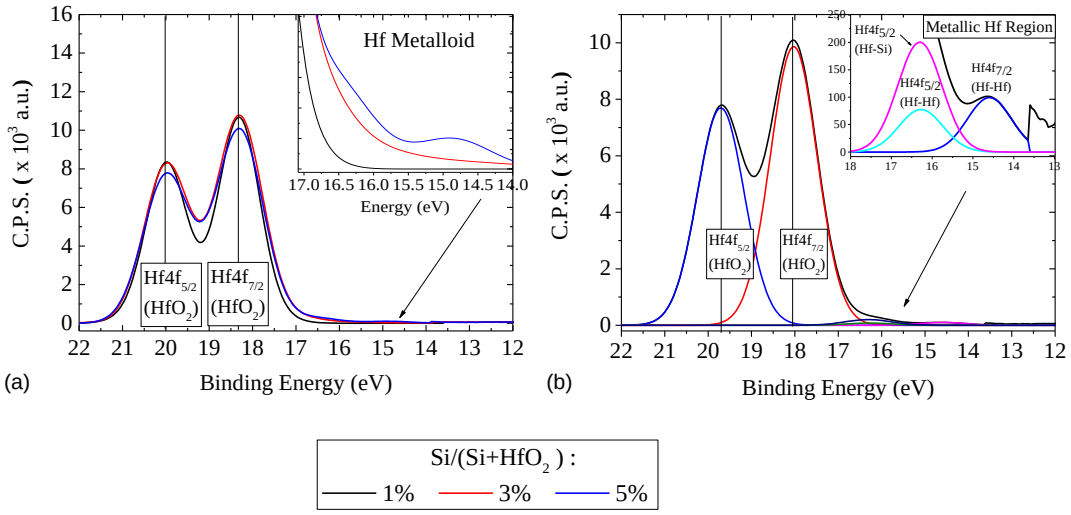


Figure 2-18 (a) Evolution of the Hf4f spectrum of Si implanted HfO_2 (b) Photopeak deconvolution of 5% Si implanted HfO_2 showing the presence of Hf-Hf & Hf-Si bonds on top of the oxidized state of Hf.

Si-Si element towards lower energies (Figure 2-17c). Considering that the characteristic binding energies of the hafnium sub oxides are in the range of (35-12) eV while oxygen lies in the range 200 eV, the previously observed shift in Figure 2-17c suggests that Si is partially bound to Hf atoms. This is in agreement with the previously observed trends through XRR analysis.

Examining the Hf4f spectrum (Figure 2-18) we immediately notice a similar behaviour. Even though the Hf4f spectrum is dominated by the oxidized states, ($Hf4f_{5/2}$ and $Hf4f_{7/2}$ for HfO_2 at 20 eV and 18.3 eV respectively), a closer observation reveals the appearance of metallic Hf states in the spectrum, in the region of 18.3 eV to 14.3 eV whose magnitude is tied to the dose of implanted Si (Figure 2-18a). Furthermore, deconvolution of the measured Hf4f signal reveals Hf-Si bonds was well as Hf-Hf bonds (Figure 2-18b & inset). This suggests that the implanted Si atoms are partially bound to both Si and Hf, in line with the assumption of the creation of an HfSiOx suboxide. Interestingly, the appearance of the Hf-Hf metallic peaks in the range of 16.5 eV to 14.3 eV suggests that the material conductivity is potentially enhanced, and is related to the Si incorporation process. This trend is in agreement with the experimentally measured behaviour of the pre-forming current in the Si implanted HfO_2 ReRAM devices presented in §2.4.

Analysing the Al implanted samples, the trend observed differs significantly: Figure 2-19a shows the Al2p spectrum of Al implanted HfO_2 as a function of the dopant incorporation

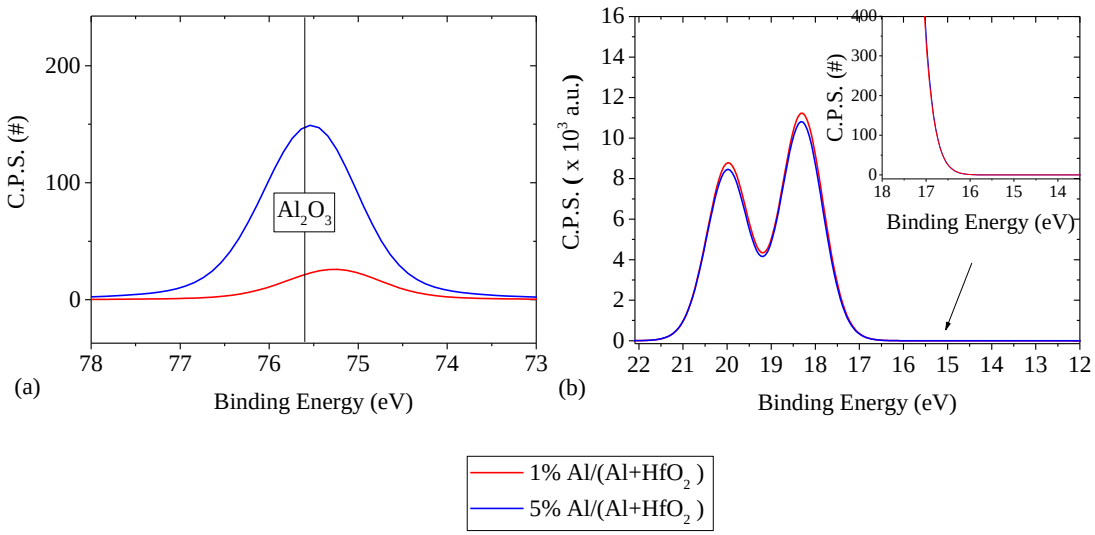


Figure 2-19 (a) Al2p spectrum showing only the presence of Al_2O_3 peaks (b) Evolution of the Hf4f spectrum as a function of Al implanted species concentration. No Hf-Hf or Hf-Al bonds are detected.

in the oxide. From this, only the peak corresponding to the aluminium oxide close to 75.6 eV is detected, with surface proportional to the Al fraction implanted. Hence, Al is in a perfectly oxidized state (to the limit of experimental detection accuracy). At the same time, no metallic peaks are detected in the Hf4f spectrum of these samples (Figure 2-19b). The absence of metallic states is also supported by the low pre-forming current of Al implanted devices as already discussed in §2.4.

The analysis so far, combined with the ab-initio calculations previously discussed in this chapter shows that effectively none of the scenario assumed can, self-consistently, explain all experimental results. One the one hand, Hf substitution by Si seems to explain the mass density evolution, but cannot correctly predict the shape of the bandgap and existence of metallic states detected by XPS. On the other hand, the case of Al implantation seems to favour the Hf substitution scenario (considering density profile evolution alone), but no Al-Hf bonds could be detected in the XPS spectra of the Al implanted HfO_2 films. As the Al is found to be in a completely oxidized state, the oxidation of Al seems to be favoured, which in turns is not expected to strongly impact density on the basis of the ab-initio results. As the implantation occurs under identical conditions, one could expect that the impact should be similar. Moreover, it has been theorized that Al could also augment conductivity when introduced in HfO_2 [40]. These results seem contradictory at a first glance.

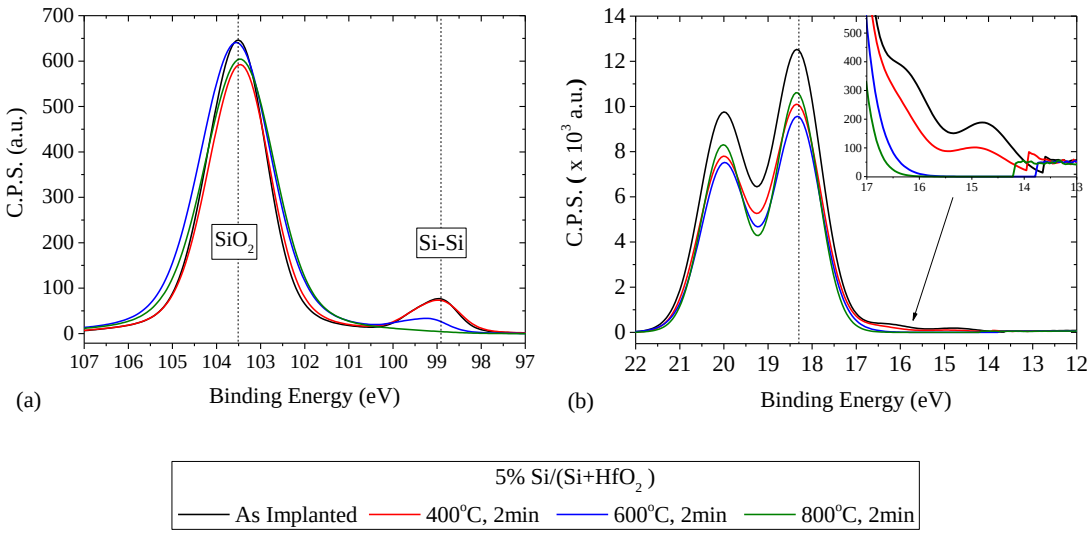


Figure 2-20 Vanishing of the Si-Si bond with increasing annealing temperature in the Si2p spectrum of Si implanted HfO_2 (a) with parallel disappearance of the metallic part of the Hf4f spectrum in the corresponding samples (b).

This apparent discrepancy of the binding states of the two dopants can be understood by studying the thermal response of the Si dopant in HfO_2. In Figure 2-20a & b, the Si2p and Hf4f spectra of a 5% Si/(Si+HfO_2) implanted samples are presented respectively. Each curve corresponds to a different sample, processed under identical conditions yet annealed at a different temperature for 2 min under He ambient. The binding state of Si is found to depend the thermal budget applied to the film post implantation. The Si-Si bonds detected in the Si2p spectrum start to lessen in magnitude until they completely vanish after annealing at 800°C where only the peak corresponding to SiO_2 remains

At the same time, the Hf-Si and Hf-Hf metallic states of the Hf4f spectrum show similar behaviour. After annealing at 800 °C only HfO_2 and SiO_2 peaks can be detected in the matrix. This suggests a potential thermally activated oxidation process of Si and/or re-organization of the local structure of the material. The film annealed at high temperatures, is expected to be less conductive as compared to its non-annealed counterpart owed to the absence of metallic states. Nevertheless, further investigation is required about the charge state of the Hf-Hf and/or Hf-Si bonds.

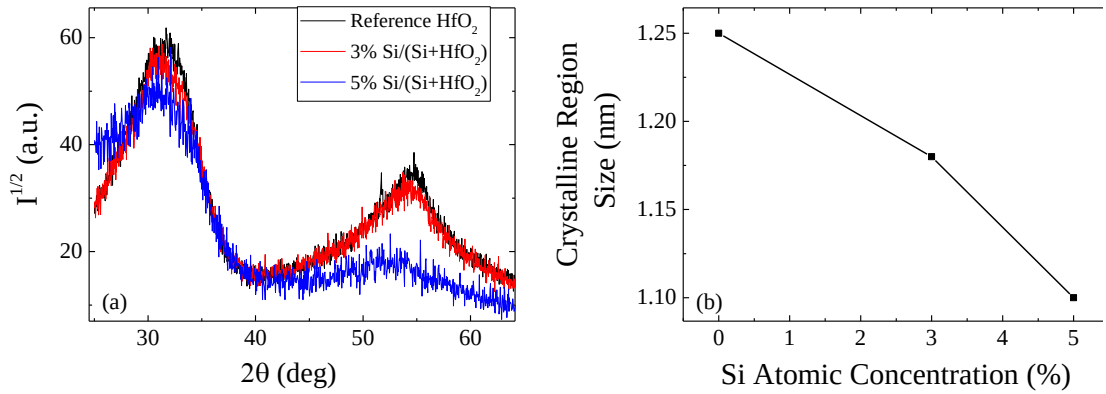


Figure 2-21 (a) Evolution of XRD diffraction spectra of HfO_2 with Si implant concentration and (b) mean crystallite size extracted from (a). XRD was performed on RTA annealed (400 ^oC, 2 min, He atmosphere) HfO_2 and Si implanted HfO_2 layers deposited on TiN.

2.7.3 .ALD – HfO_2 and impact of Si implantation on Crystallinity

The Si ion implantation process does not significantly impact the initial crystallinity of ALD deposited HfO_2 as the average crystalline region size, determined by XRD spectroscopy is found to be below 1.3 nm (Figure 2-21a, b). As a result, one can conclude that the observed progressive vanishing of Hf-Si, Hf-Hf and Si-Si peaks in the XPS spectra of the annealed samples is linked to local reorganisation of the material structure for the temperature range of intererst for BEOL ReRAM integration (~400°C).

2.7.4 Si implant thermal stability in HfO_2

The dopant profile stability was studied by TOF-SIMS analysis for samples annealed between 400°C and 800°C for 2 min in He atmosphere using RTA annealing. The Si profile is found to be stable below 600°C. Diffusion signatures are found only for the 800°C annealed samples. At that limit, the Si diffusion profile in the HfO_2 layer is found to be dependent on the nature of the substrate.

Si:HfO_2 on Si samples (Figure 2-22a) samples present with strong diffusion of Si with the depth concentration profile becoming homogeneous. The profile concentration tails (7-9 nm) of the annealed sample near the HfO_2/Si interface increase, suggesting Si dopant self-diffusion inside the Si crystal; further detail is lost owed to the convolution with the background concertation of the crystal. Interestingly, the peak exhibited in the first 3 nm close to the Air/HfO_2 interface points towards an out-diffusion of Si in the ambient.

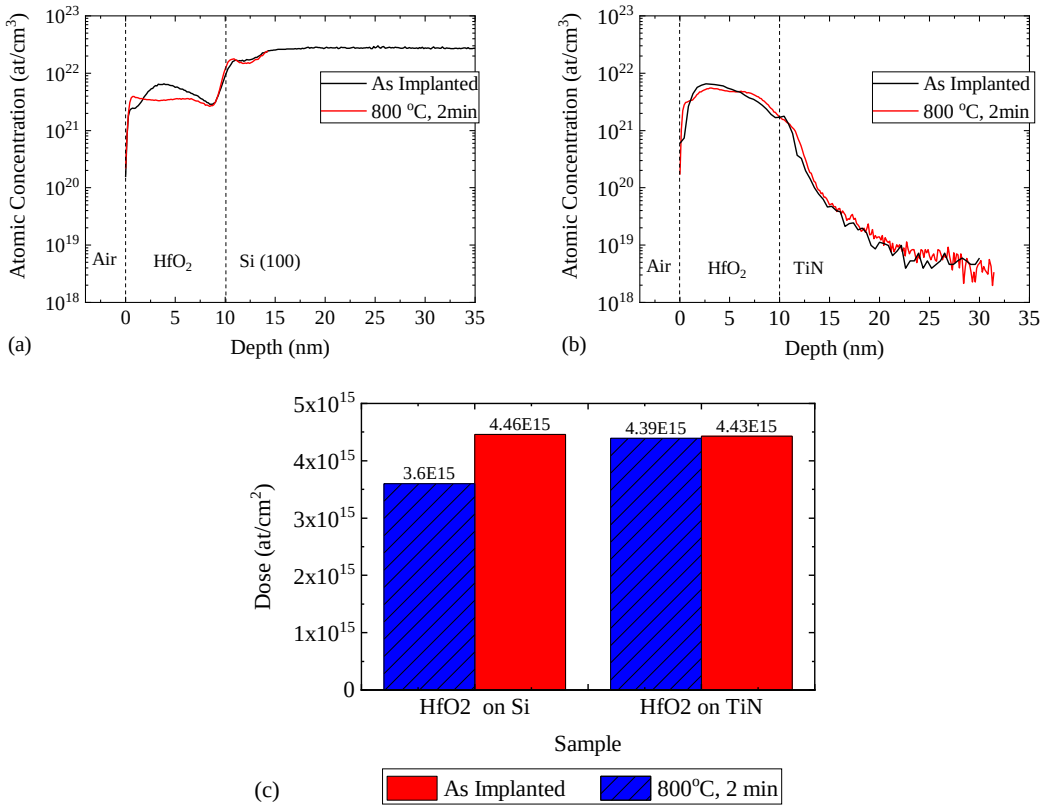


Figure 2-22 Evolution of Si implant in HfO_2 with annealing temperature. (a) on Si (100) substrate (b) on TiN. (c) Total dose measured experimentally from SIMS. Significant dose loss is only observed for when the Si implanted HfO_2 matrix is deposited on Si (100) crystal.

The Si:HfO_2 on TiN layer diffusion process varies significantly. In this case, the profile of the Si dopant in HfO_2 is hardly impacted by temperature (Figure 2-22b) and only minor profile re-distribution is observed. Significant dose loss is observed only when Si doped HfO_2 is deposited on Si (Figure 2-22c) in the order of 21% wrt the as-implanted reference. This dose loss is owed to diffusion both towards the Si bulk and towards the ambient.

Considering the evolution of the two concentration profiles it is straightforward to understand that at high temperature process steps, Si can diffuse in HfO_2 if a concentration-controlled diffusion process is dominant and the segregation coefficient in the interfaces is not constrained; this seems to be the case for HfO_2 on Si (Figure 2-22a). The presence of TiN acts as a blocker for the diffusion of Si in the HfO_2/TiN interface, and as a result, a segregation gradient cannot be locally created to activate the diffusion process towards TiN. Nevertheless, considering that no dose loss is registered in that case, the out diffusion towards the ambient in the air/HfO_2 interface seems to still be blocked.

There are two possible scenario explaining this behaviour: i) a phase transition of the HfO_2 when deposited on TiN induced by either mechanical stress of the TiN layer (in the order of -2 to -6 GPa, see Chapter 4) or in-situ annealing of the oxide during ion implantation that induces a phase-change in HfO_2 and modifies the diffusivity of Si in the oxide. Nevertheless, the implantation conditions have been optimized to minimize the impact of self-heating and the thermal conductivity of both conductive TiN and Si are high, so self-annealing during ion implantation is considered less probable.

ii) the development of a built-in electric field with direction towards the TiN layer that acts antagonistically to the concentration controlled diffusion. This component will tend to push Si towards the TiN bulk, but owed to the extremely low diffusivity of Si in TiN, the diffusion is blocked leading to a net of almost zero. The presence of a built in-field can be justified if a significant net charge in the oxide is generated by the introduction of Si in HfO_2. In the presence of a charge in the oxide, the TiN metallic gate will tend to mirror the charge, creating local dipoles. The Air/oxide interface however cannot create dipoles as efficiently, hence creating an electric field with direction towards the TiN layer. This scenario requires high levels of net charge in the oxide supported by further findings (§2.9).

2.8 A microscopic understanding of the alloying mechanism in HfO_2

The experimental data presented in the previous sections can be explained by more than one atomistic configurations as already discussed. In the case of Si implantation, the electrical results point towards the emergence of conductive states inside the gap. Such states could arise by either Hf or Si occupying interstitial positions. The Hf substitution scenario by the dopant is generally supported by the strong reduction of mass density with dopant concentration, as predicted by ab-initio and supported by XRR results. The interstitial atom being either Hf or Si is imposed by the rise of electrical conductivity. Moreover, XPS confirms the existence of Hf-Hf metallic states, indicative of Hf interstitials as well as Hf-Si & Si-Si states shifted towards the Hf binding energy spectrum. From this, we can infer that the dominant process for Si is substitution of the Hf atom, resulting in it occupying interstitial positions. The possibility that a non zero fraction of Si atoms occupying interstitial positions thus contributed to conductivity cannot be excluded.

Explaining the Al doping process during ion implantation is less straightforward at first glance. On the one hand, the strong mass density variation with doping seems to support the idea of Hf substitution where oxygen is added in parallel to maintain stoichiometry. Nevertheless, this raises the question why Si gives raise to Hf interstitials while Al doping does not. Moreover, no Hf-Al or Al-Al bonds can be traced in the XPS spectrum of Al implanted samples, but rather Al found in an oxidized state (Figure 2-19). This suggests that the oxidation of Al incomplete bonds in HfO_2 is spontaneous, in line with the negative formation enthalpy calculated by ab-initio.

The results for both Si and Al implantation can be understood if seen under the scope of a twostep process: i) Hf substitution by the dopant followed by ii) dopant native oxide formation. In both cases, the dopant introduced by ion implantation substitutes Hf atoms in the HfO_2 cell, disrupting local symmetry, knocking Hf atoms to interstitial positions and creating a trap-rich complex alloy. The energy required for the Hf substitution process is provided by the dopant atoms as they are being accelerated inside the ion beam of the implantation chamber. In this image, Hf-Hf and Hf-D (where D stands for Al, Si) are expected. As Hf is much heavier than either Si or Al, it being knocked to quasi interstitial positions with the dopant taking its place can naturally distort the supercell, strongly enough to modify the molecular volume hence reducing mass density.

As the film is exposed to the ambient, the incomplete dopant-Hf bonds are prone to oxidation or phase transitions as the structure will seek to restructure itself to an energetically more favorable configuration. The only difference between Al and Si is that, Al oxidation is an exothermic reaction while Si oxidation is endothermic. Otherwise put, Al oxidation is more thermodynamically favorable than the incomplete Hf-Al bond. This is supported by multiple signatures both theoretical and experimental:

Firstly, from ab-initio calculation (Figure 2-5) it is observed that as soon as Al is inserted into the HfO_2 matrix and is left free to interact, the Al_2O3 cluster reaction has negative formation enthalpy if an oxygen source is provided and the bandgap of the resulting material is free from traps. In this fabrication approach, oxygen is inescapably provided by the ambient. Moreover, this trail of thought is consistent with the XPS results shown before. The strong variation of XRR density can also be explained if seen under the scope of the formation of local HfO_2 and Al_2O_3 nanoclusters: As the mass density of Al_2O_3 is

typically lower than that of HfO_2, the resulting alloy composed by nanoclusters of the two binary oxides has lower density on average, proportionately to the presence of Al_2O_3.

Secondly, even though Si implanted HfO_2 presents with conductive states, which we associate to Hf interstitials with the parallel formation of HfSiOx alloy, these bonds seem to be stable up to 400-500^{o}C approximately. For higher applied thermal budgets all Hf-Hf, Hf-Si and Si-Si bonds vanish pointing towards a local reorganization of the material to a more stable state as previously discussed in §2.7.2. The key point being that the end state after the application of a high thermal budget gives purely oxidized states for Si, in analogy with the final state of the Al implanted samples. The conductive states detected in the HfSiOx alloy can be attributed to the higher thermodynamic stability of the disordered state of HfSiOx as compared to that of HfAlOx. SiO_2 and HfO_2 nanoclustering still occurs, in analogy to the case of its Al implanted counterpart but at a higher thermal budget.

Exploiting this effect can favor resistive switching in an implanted zone as we will show in the following chapters. More importantly, by understanding the defect generation during this mechanism, valuable insight into the conductivity of HfO_2 can be gained. We identify Si as a dopant that leads to a thermodynamically stable, trap-rich alloy that can be exploited to reduce the forming voltages of ReRAM devices. It is possible that other dopants can prove promising candidates for defect engineering in HfO_2 as long as their native oxide formation process when placed inside HfO_2 is energetically less favorable than the formation of the HfDOx. In addition to that, Si and HfO_2 are standard materials in HKMG CMOS and can be effortlessly integrated in any process. Other dopants with similar properties can be found at the same column of the periodic table of elements with Ge being a promising candidate owed to its native HKMG CMOS compatibility.

Finally, the impact of the introduction of either dopants in the total charge state of the oxide has not been studied yet. Although it is not possible to directly associate the generation of a bulk charge owed to the introduction of Si or Al in the HfO_2 matrix from the XPS spectra alone, there are two signatures that support this hypothesis. Specifically for the case of Si (presenting the stronger technological interest): First, the evolution of the current-voltage characteristics as a function of the implant can be described reasonably well a variable range hopping which is essentially, a charge-limited conduction scheme. The fact that conduction is facilitated, supports the generation of a positive charge. Secondly, the dependence of the diffusion profile of Si in HfO_2 to the substrate strongly

suggests that the mechanism of diffusion has a non-negligible electrostatic component that is inherently sensitive to the boundary conditions introduced by a dielectric or dielectric/metal interface.

In the following, the presence of charges in the oxide as a function of the Si implant are investigated using optical and electrical techniques.

2.9 Charge Generation in HfO_2 as a function of Si incorporation

2.9.1 Oxide charging dynamics

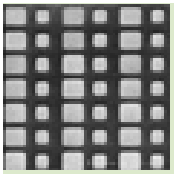
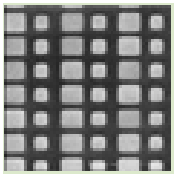
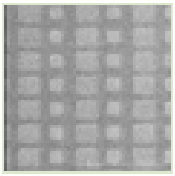
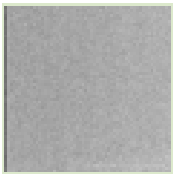
Si Conc. (%)	REF.	1	3	5
SEM Image Contrast				

Figure 2-23 Change of contrast as a function of implanted Si atoms in HfO_2 under a 300 V CD-SEM top-down direct observation. In the pattern shown above, the observed layer is HfO_2 deposited on patterned TiN 20x20 μm^2 M4 conductive TiN plugs (white). (Black) SiO_2 post-metal dielectric. The transition to white for the Si implanted HfO_2 implies a change in electronic properties towards more metallic behavior (electron trapping in the oxide)

We have so far described the alloying process of either Si or Al by ion implantation in HfO_2 presenting the different possible atomic condonations and various experimental signatures supporting the substitution of the Hf atom by the Si or Al dopant in the HfO_2 matrix. From an electrical point of view, we have demonstrated that the Si implant process leads to an increase in the electrical conductivity of the material, a decrease of the ReRAM element forming voltage, as well as decrease in the thermal dependence of the forming process. From the behavior of the conductive modes of the pristine ReRAM pristine state, we link this behavior to the generation of volume defects assumed either the Hf-Hf or Hf-Si states observed in the XPS spectra. The appearance of such states would be expected to behave as bulk traps and therefore, be associated with the generation of a net charge.

The first experimental signature of charge unbalancing of the implanted layer can be observed right after the implantation process. In Figure 2-23 a top down SEM observation is presented for an as-implanted HfO_2 layer deposited on TiN during integration of

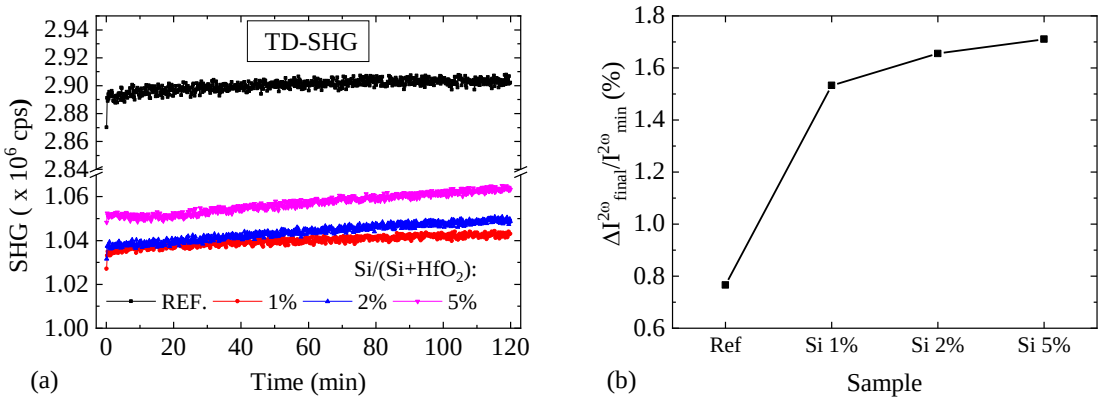


Figure 2-24 (a) Second harmonic generation signal dynamic (SHG)

ReRAM devices on a 200 mm process (Chapter 3). The layer contrast changes as a function of the implanted Si fraction signifying a change in the electronic properties of the layer. Constidering that the chromatic contrast of HfO_2 gradually approaches that of the conductive TiN layer below, we can expect the oxide conductivity to increase. Nevertheless, even though it appears a net charge is being generated in the layer, this first signature does is not suffice to understand if the charge is coming from the surface or the volume of the layer.

The charging dynamics of the implanted HfO_2 has been investigated using optical second harmonic generation response (SHG), a useful optical technique to study buried interfaces and volume defects [157]–[159]. In this technique a laser beam (λ=0.78 μm) is shot to the sample upon which part of it is reflected (HfO_2/Air interface) and part penetrates into the material. The transmitted beam travels into the HfO_2 bulk until it reaches the HfO_2/Si(100) interface. Owed to the broken inversion symmetry at that interface, second harmonics are generated. The phase, modulus and time response of the second harmonic of the electric field ($I^{2\omega}(t)$) can provide information on the quality of the interface and the amount of bulk defects in the oxide.

In Figure 2-24a the $I^{2\omega}(t)$ response from Si implanted HfO_2 on Si(100) is studied. Si is used as a substrate to maximize the second harmonic response. The measured DC component of the SH intensity decreases with increasing Si content in HfO_2 until 2% and then starts to increase again at 5%. The time-response of the SH signal is stronger with increasing Si content. This result can be understood if i) the charging of the Si:HfO_2/Si(100) interface and the bulk defects into the Si:HfO_2 layer in relation to the SHG signal response are taken into account:

The time dependence of the SHG signal is typically called EFISH (electric field-induced second harmonic) in the literature and mainly originates from two processes:

A. Electron-hole pair creation from the absorption of the fundamental radiation (1.59 eV), which is subsequently separated from the pre-existing electric field close to the broken symmetry interface, from the SCR (space charge region) in the Si and/or the interface traps at the interface and/or the oxide-trapped charge. This charge separation screens the initial electric field, reducing the SH intensity.
B. Multiphoton electron injection to various trap sites at the interface or in the bulk of the oxide, which increases the SH intensity. The order of the photo process depends on the band offset between Si valence band and oxide conduction band. For SiO_2 and Al_2O_3 it is typically a 3-photon injection process.

The main difference between the SHG signal levels of the REF. sample and the implanted ones is that the Si implanted oxide matrices absorb the SH radiation travelling away from the interface (and towards the air). Hence, the signal level is decreased. With continuous excitation, and as the SCR is uniform (uniformly doped wafers), the difference in the pre-existing electric field at the interface will be given mostly by interface defects/oxide trapped charge. That is why in the implanted samples, the initial SHG signal (close to t=0), associated with the pre-existing electric field, *decreases* for increasing implanted dose (stronger pre-existing field, process A). This is supported from the fact that the time-dependency is reduced until it reaches a minimum, due to process A (screening of pre-existing field of comparable magnitude to the one induced by process A). Then it starts to increase due to process B. For the 1% and 2% samples, the pre-existing field is lower and process B is dominant, increasing SHG with time (no screening of the fields and reducing of initial SHG is observed, Figure 2-24b).

Even though the bulk of the SHG signal originates from the HfO_2/Si interface, its strong attenuation for all samples denotes that process A is dominant at all times with process B becoming more significant (but still of secondary importance) for samples with higher Si concentration. This implies that oxide volume charge is for all cases more important than interface charges (since field screening is always more significant that multi-photon injection here). Even though the SHG signal provides useful information about the behavior of defects, a quantitative of the amount of charge is not possible owed to the lack

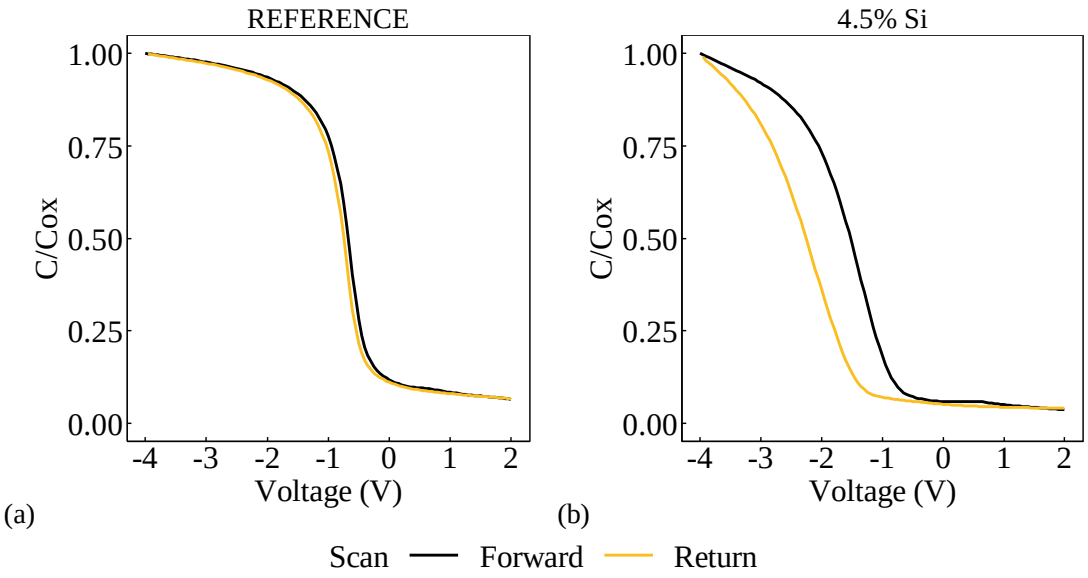


Figure 2-25 Normalized closed- C-V characteristics of p-Si(100) 50x50 μm^2 MOS capacitors at 10 KHz with (a) HfO_2 and (b) Si implanted HfO_2 with 4.5% Si content measured at 300 K.

of analytical models that accurately relate charges in the interfaces/bulk to the signal response, at present.

2.9.2 Estimation of oxide-trapped charge as a function of Si incorporation

The most suitable system to characterize the presence of charges in an oxide is still the Metal Oxide Semiconductor (MOS) capacitor where the oxide is deposited on a Si substrate. The MOS system has been used as a test vehicle for oxide electrical characterization, with a specific target to extract the level of charge in the oxide volume. In an MOS system, three major species of charge can be identified: i) D_{it}: interface charge states in the Si/Oxide interface ii) border traps: slow traps in the Si/oxide interface and inside the volume of the oxide iii) oxide trapped charge in the oxide volume and iv) fixed charge in the Si/oxide or oxide/Metal interfaces.

In Figure 2-25 the high-frequency CV characteristics of TiN/HfO_2/p-Si with $N_A \sim 8.54\ 10^{15}$ ($\rho \sim 1\ \Omega \cdot cm$), ALD deposited HfO_2 with $t_{ox} = 10$ nm for the no-implant reference (Figure 2-25a) and 4.5% Si implanted HfO_2 (Figure 2-25b) are shown. Both samples have been annealed for 1s at 1050°C (spike, flash lamp annealing) to limit leakage currents. From the qualitative characteristics of the CV curves of the two samples it can be understood that:

i) The leakage current through the oxide is stronger in the implanted sample as compared to the reference owed to the enhanced roll-off effect observed in both accumulation and depletion regions [160], [161].

ii) The amount of positive charge trapped in the volume of the oxide and/or near the HfO_2/Si interface is higher in the implanted sample as its CV to values that are more negative.

iii) The amount of slow-responding border traps either in the HfO_2/Si interface or inside the oxide volume is increased with implantation, as the surface enclosed by the closed-loop CV curve is increased.

To quantify the impact of Si implantation in HfO_2 in terms of charge generation we study the flat band shift and CV closed-loop surface evolution (total charge in the border traps) of the MOS structures as a function of ion implantation. Samples implanted with Si varying from 3% to 4.5% and annealed at 750°C for 30s or 1050°C for 1 s have been studied. The flat-band voltage of the MOS system is given by [162]:

$$V_{fb} = \varphi_{MS} - \frac{Q_f}{C_{ox}} - \frac{Q_{it}(\varphi_S)}{C_{ox}} - \frac{Q_{ot}}{C_{ox}} - \frac{Q_M}{C_{ox}} \tag{2.5}$$

An accurate determination of the fixed surface charge would require minimizing all other contributions. Considering that the implantation process can impact both the charge in the Si/HfO_2 interface and oxide volume charge, supported both by SEM observation SHG and XPS we consider that the flat band shift is primarily impacted by charges in i) the volume of the oxide and/or near the Si/HfO_2 interface and ii) border traps.

The presence of mobile charges in the oxide can only by measured by closed-loop capacitance CV at elevated temperatures (~200°C)[162] [163] and is hence excluded. Dit quantification has also been ignored in this analysis, as it impacts mostly the frequency response characteristics of the system is specific to the Si/HfO_2 interface, and any conclusions made by studying the Dit evolution cannot be transferred to the MIM system. By considering the initial charge levels in the reference oxide to be much lower as compared to the implanted samples, we can estimate the amount of oxide-trapped charge using the approximation [162]:

$$\mathrm{Q_{ot}} = -\mathrm{C_{ox}} \cdot \Delta \mathrm{V_{FB}} \quad , \quad \Delta \mathrm{V_{FB}^{I-R}} = \mathrm{V_{FB}(C_{Si})} - \mathrm{V_{FB}^{REF}(Q_{ot} = 0)} \tag{2.6}$$

To account for the uncertainty of the TiN-Si work function φ_{MS} and the initial amount of charge in the reference HfO_2, the relative flat band shift between the reference and the implanted samples is calculated and used as in Eq. (2.4) to approximately calculate the oxide trapped charge increase in the Si implanted samples.

The flat band voltage shift is found to be linearly dependent to the Si implant, irrespective of the annealing temperature (Figure 2-26). Expressing the relative flat-band shift between the reference and implanted samples in terms of Eq. (2.5), and taking into account the only volume and interface charge terms we have:

$$\Delta V_{FB}^{I-R} = -\frac{\alpha}{C_{ox}^{Si}} C_{Si} + \frac{Q_{REF}}{C_{ox}^{REF}} \ , Q_{Si} \sim \alpha C_{Si} \tag{2.7}$$

Where C_{si}[%] is the concentration of Si in HfO_2. The term α [V^{-1}] can be interpreted as the rate of flat band shift induced by the generation of charge in the layer with doping. Eq. (2.4) differs to (2.6) only in that it introduces the correction term Q_{REF}/C_{ox}^{REF} to account for the presence of initial charge in the reference sample. Finally the total oxide trapped charge in the oxide for each Si implanted sample can be calculated by (the charge in the reference layer is also positive):

$$Q_{ot} = -\Delta V_{FB}^{I-R} \cdot C_{ox}^{Si} - \frac{Q_{REF}}{C_{ox}^{REF}} \tag{2.8}$$

The presence of hysteresis in the CV curves denotes the presence of a different type of charge other than oxide trapped or fixed charge. The hysteretic behavior in the room temperature CV curve implies the generation of a charge population responsive sensitive to the AC component of the applied bias.

The total charge in the border traps is calculated by numerical integration of the CV closed loop curves according to:

$$Q_{bt} = \int_{forward} C(V)dV - \int_{return} C(V)dV \tag{2.9}$$

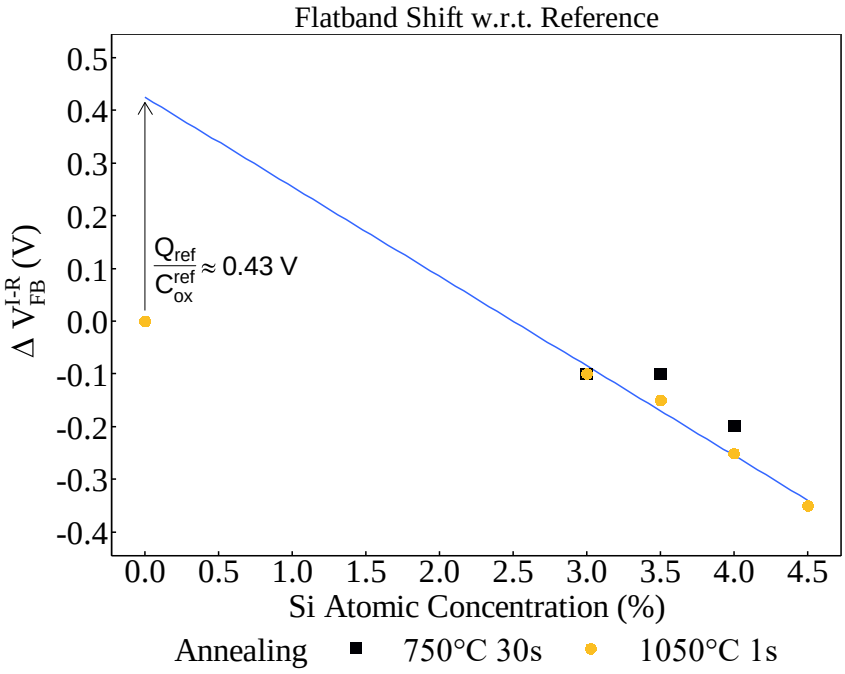


Figure 2-26 Extracted flat band shift of Si implanted HfO_2 MOS capacitors with respect to the no-implant reference sample.

Moreover, the charge inducing the permanent shift in the flat band voltage is closer to the Si/HfO_2 interface. On the other hand, the hysteresis of the CV gives more a measure of the complete charging and discharging of the defects inside the volume of the oxide, and is thus believed to allow probing of species closer to the TiN metallic electrode that are normally not detectable in the flat band shift. The calculated oxide trapped charges and border traps are shown in Figure 2-27 (a), (b) and the total extracted charge in Figure 2-27 (c). The volume density of defects calculated from the total charge estimate is shown in Figure 2-28.

It is important to note that the projected density of defects in the implanted samples *increases* by at least one order of magnitude as compared to the no-implant reference from 4 x 10^{18} $eV^{-1}cm^{-3}$ to 1.5 x 10^{19} $eV^{-1}cm^{-3}$ (for the conversion it has been considered that each volume defect contributes 1 electron) an increase of 270% (!).

The absolute values of the volume charge extracted with the methodology presented above, are sensitive to the initial estimation of charge in the reference oxide. Especially the charge density profile would require dedicated characterization techniques such as the etch-off or the photo I-V methods[162]. The values reported above are nevertheless consistent with reported values in the literature for low-temperature HKMG technologies

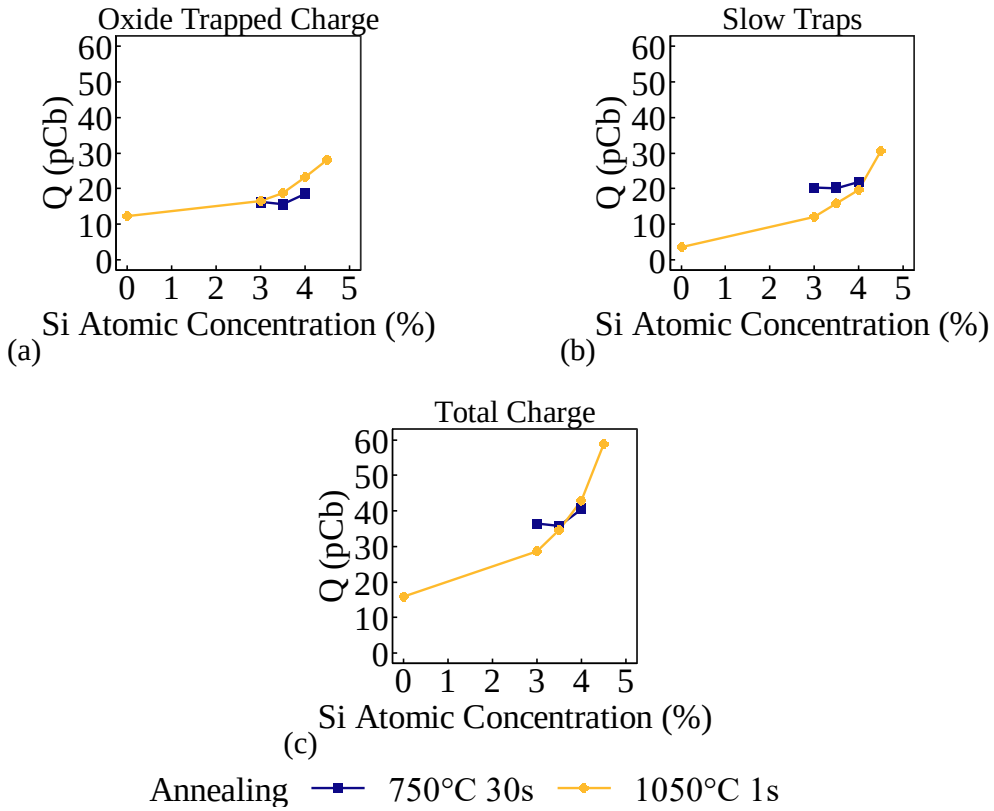


Figure 2-27 (a) Oxide trapped charge calculated by flat band shift (b) Total charge responding to the slow-field variation contribution in CV hysteritic behaviour. (c) Total charge estimate from summing components of (a) and (b).

in the order of 1.5 x 10^{19} $eV^{-1}cm^{-3}$ [164] , and are reported to vary from 7.5 x 10^{18} $eV^{-1}cm^{-3}$ (state of the art) reaching as high as 1 x 10^{20} $eV^{-1}cm^{-3}$ for SiGe systems [165].

Considering the charge compensation from the TiN gate in the oxide/Metal interface, the defects closer to the gate cannot be probed by MOS techniques. As a result, the amount of charge measured by such techniques might be underestimated with regard to the total charge trapped in the oxide. Even though the charges near the gate are of little importance for CMOS technology, the can be critical for the conduction and operation of ReRAM devices. The previously presented XPS & XRR analysis supported by the electrical data of ReRAM & MOS devices and the SEM observations (Figure 2-23) suggest that the primary charge generation mechanism is closely tied to the alloying process during ion implantation and a non-negligible amount of defects is created inside the bulk of the oxide. Hence, the previously extracted defect density values are considered reasonably representative for both systems.

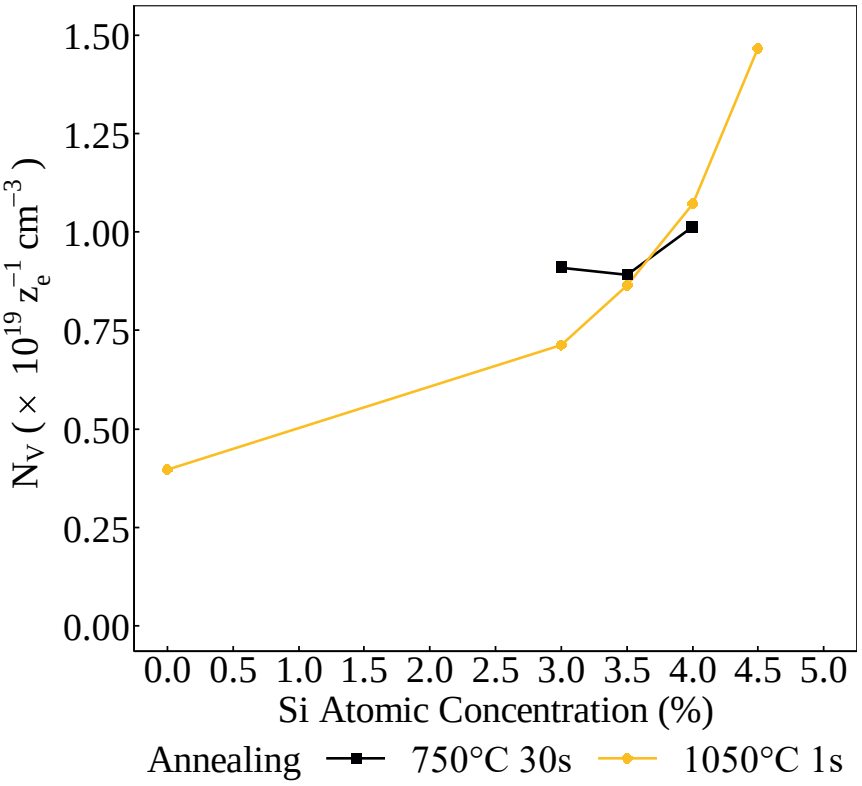


Figure 2-28 Estimated evolution of volume density of defects with Si ion implantation in HfO_2. z_e stands for the charge state of the bulk defect.

2.10 Discussion: On the origin of conductive states in HfO_2 and associated doping strategy for ReRAM applications

Even though many groups have investigated doping as an option to improve properties of HfO_2 ReRAM, a clear roadmap as per the dopant properties needed in order to favor a specific aspect of ReRAM features is yet needed. There is a plethora of experimental data, sometimes contradictory to each other. This is owed to the fact that i) the understanding of the alloying process itself is not complete especially if more than one dopant species are considered and ii) the multi-parametric nature of the problem. The end material observed is often impacted by experimental parameters such as choice of the alloying process or the choice of integration into the device, which largely impacts the material and finally leads to adding more than one physical mechanisms that are hard to decouple.

Moreover, there is no consensus to date as per the nature of conductive states in HfO_2. While it is experimentally observed that sub-stoichiometric HfOx will be more conductive, it is not clear if it is owed to a narrowing of the bandgap, emergence of conductive states inside the bandgap or band conduction.

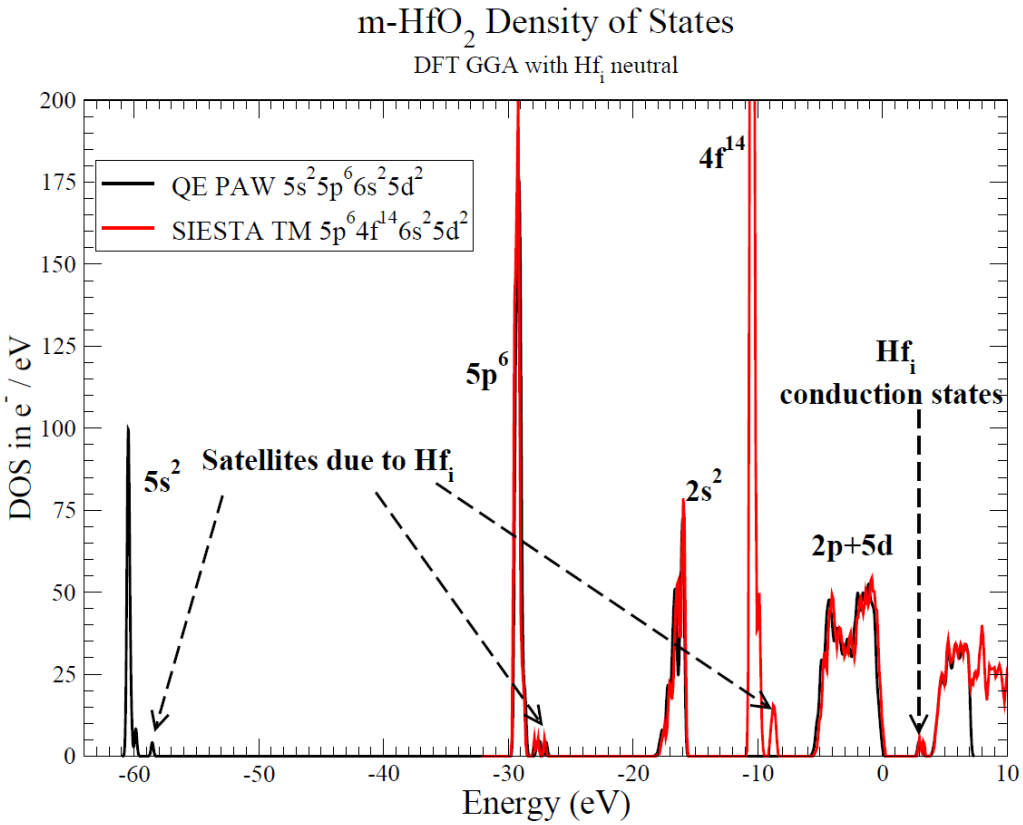


Figure 2-29 DOS for pure m-HfO_2 (black) and m-HfO_2 with excess Hf atoms in interstitial positions calculated by DFT. Satellite peaks appear for orbitals with Hf_i states.

In Figure 2-29 the density of states of pure m-HfO_2 as calculated based on the Generalized Gradient Approximation (GGA) is shown for the fully ordered cell (black) and a cell where Hf atoms are knocked-out in interstitial positions (red). Satellite peaks for nearly all orbitals are traced.

The case of the $4f^{14}$ orbital and the associated satellite peaks are of particular interest. The emergence of the satellite peaks in the DOS as a consequence of Hf_i proves that there should be a detectable difference on the spectral energy shifts of the characteristic bonds as a function of the environment. The Hf4f metallic bond of Hf-Hf can then directly be tied to interacting Hf_i. The calculation is representative of the physical mechanism albeit the corrections that need to be applied for a closer quantitative description (notably for the potential perturbations caused by the excitation of electrons in the core). Moreover, the electrical results show increase of the material conductivity as well as the increase of total charge in the oxide.

Hence, two conclusions can be drawn: First, the introduction of Si leads to Si_{Hf} while it

still partially interacts with O. The process could be seen as a cationic Frenkel defect creation of the form $\emptyset \rightleftharpoons V'_{Hf} + Hf_i$ generating a positive charge. The augmentation of conductance in tandem to the appearance of charged Hf4f Hf-Hf could be understood from the atomistic standpoint: The creation of a state in the conduction band means removing an electron from the full $4f^{14}$ shell towards an empty level. The introduction of Si binding with O to an extent being the trigger of charge unbalancing; the gist being that, material conductivity will vary in proportion to the Hf_i population. Understanding of this mechanism can not only help to systematically understand how to engineer HfO_2 for ReRAM applications but also provides a clear image of the origin of electron carrier centers forming the filamentary conduction path in HfO_2 based ReRAM.

CHAPTER 3. SI IMPLANTED HFO_2 RERAM: THE IMPACT OF LOCAL DEFECT ENGINEERING IN RERAM

3.1 The concept of switching zone localization

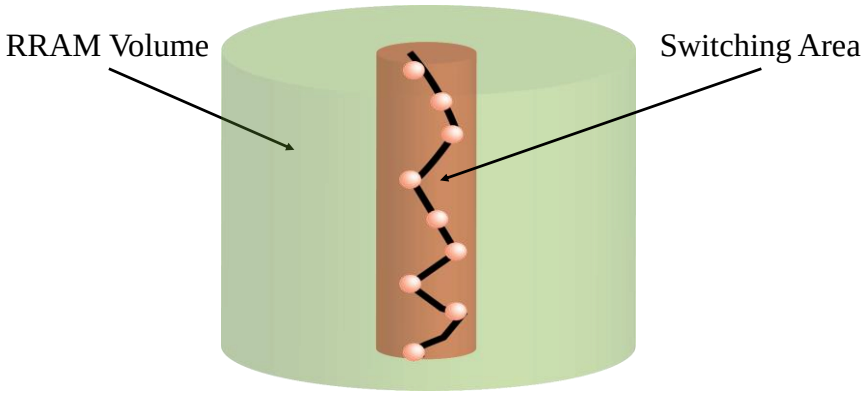


Figure 3-1 Schematic representation of a conductive zone engineered by Si ion implantation.

As already discussed in the previous chapter, filament localization is considered to act beneficially in the data retention of ReRAM devices (§2.1). Deterministically constricting the switching zone, laterally into the defined cell is nevertheless a non-trivial problem, considering the stochastic nature of filamentary ReRAM devices. To date, the approaches presented consist in limiting the regions of reactivity between the ion exchange layer and the metal oxide by either modifying the sidewall of the cell to make it inert (PANASONIC) or limiting the reactive surface of the ion exchange electrode; therefore, restricting the flow of ions (RENESAS).

On the other hand, forming voltage reduction has been achieved by favoring the creation of defect states, either near the oxide/reactive electrode interface or inside the bulk of the oxide. It is rather straightforward to understand that a trap rich region inside the volume of an oxide is electrically more favorable than a volume with less defects. As such, the conductive path should preferentially be created in trap-rich regions where the forming voltage is expected to be reduced.

In the previous chapter, we studied in depth the process of engineering defects in HfO_2 by means of Si ion implantation and the material modifications that are induced as a consequence thereof. In the following, we proceed to demonstrate a new concept for a HfO_2 based ReRAM cell employing a simple integration flow, that aims to localize the

switching area as well as tune forming voltages. After detailing the integration flow of the proposed cell, we compare its electrical characteristics to non-localized switching ReRAM cells and identify the improvements and eventual tradeoffs of the approach.

3.2 Integration of a Localized Switching ReRAM Cell in the BEOL

3.2.1 Description of the test vehicle

Considering the thermodynamic stability of the HfO_2:Si alloying mechanism where the engineered defects start vanishing above the 600°C threshold, the identified alloy can be ideally integrated in CMOS gate-last, MOL and BEOL integration flows. In the following, we demonstrate a BEOL integration flow using 1T-1R memory cells integrated in 4 Kbit array topologiy; the control transistor is based on a 130 nm CMOS node and the test vehicle is demonstrated in a 200 mm technology. This serves to demonstrate the functionality of the approach, inducing short and cost effective fabrication cycles.

The MIM capacitors acting as the ReRAM memory cells are integrated between metal 4 (M4) and metal 5 (M5). The 130 nm CMOS circuitry is pre-processed on the baseline wafers and therefore we avoid a detailed description of its integration flow. The design of the cell is such that the bottom electrode is contacted to its environment through M4.

3.2.2 Integration of a ReRAM cell in the BEOL

Two different types of cells are integrated and compared in terms of electrical performances. The first cell variant consists of an OxRAM cell with uniform doping in the lateral direction of the HfO_2 layer, hereafter referred to as Blanket Si Implant (BSI) device. In the second variant, the HfO_2 layer is doped locally, in a lithographically defined area in the center of the MIM cell and will be hereafter referred to as Locally Si Implant (LSI) device. The 100 nm PVD deposited TiN MIM bottom electrode contact is first patterned (Figure 3-2a). The bottom electrode TiN layer is then planarized using a chemical mechanical polishing (CMP) step. Following the patterning of the TiN bottom electrode, 10 nm of HfO_2 are deposited by means of ALD (Figure 3-2b), whereupon the integration splits in the two variants. In the first approach (BSI), the oxide is implanted directly with the dopant, namely Si. In the second approach (LSI), an additional e-beam lithographic level serves to constrict the implanted area in a narrow region in the center of the MIM capacitor (Figure 3-2c, d). The resist is then stripped. Following that, 10 nm of Ti and 100

(a) (b)

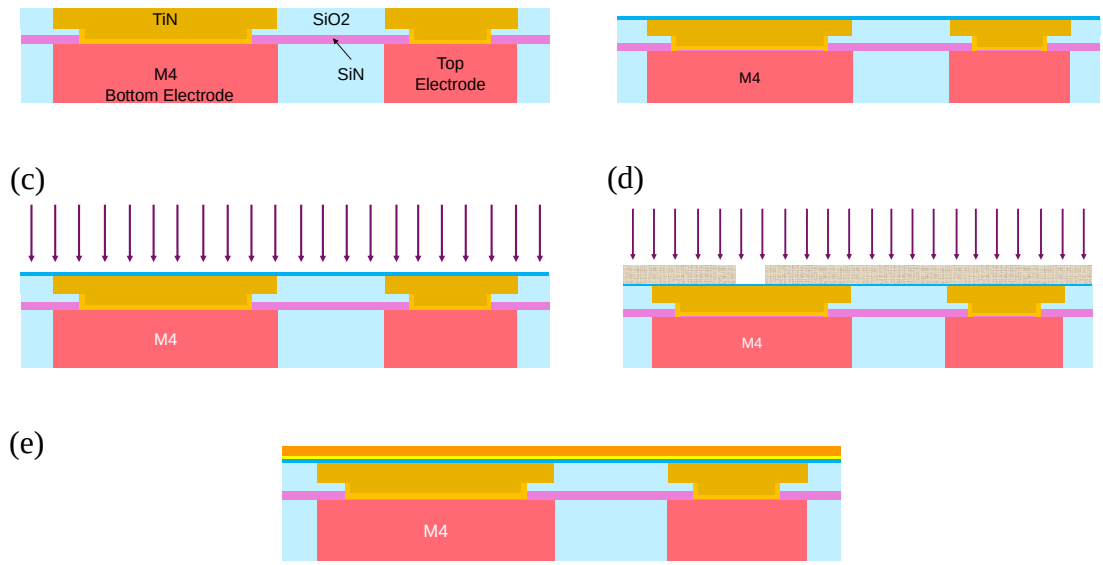


Figure 3-2 Integration flow of the MIM memory block: (a) Patterning of the TiN bottom electrode on M4, (b) Deposition of HfO_2 (c) BSI device: Full sheet ion implantation. (d) LSI device: Patterning and implantation through a local aperture directly on the oxide (e) deposition of the reactive and contact electrode metals (Ti/TiN).

nm of TiN, serving as the reactive and contact electrodes are respectively deposited by PVD at 350 ^{o}C (Figure 3.2e).

The two device types follow the same integration flow hereafter: a DUV lithography process is used to define cylindrical MIM capacitors of identical sizes between the two integration routes (Figure 3-3a). The MIM size ranges from 600 nm down to 300 nm diameter. The stack is then etched using a dry RIE plasma chemistry in two steps: First, BCl_3 chemistry is used to etch the Ti/TiN layer and then Ar/Cl_2 chemistry is used for the etching of the HfO_2 layer. Even though BCl_3/Cl_2 chemistry is also suitable to etch HfO_2 [] Ar/Cl_2 is preferred in this approach as it allows precise control of the HfO_2 etch rate.

Next, a 30 nm SiN layer is deposited followed by standard PMD (SiO_2) as shown in Figure 3-3c. SiN acts to insulate the active region of the memory from the oxygen rich SiO_2 environment that can degrade ReRAM performances. The VIA4 (M4,M5) contact then is then defined by a dedicated lithographic step, followed by an SiN/SiO_2 layer etching step and an O_2 dry + wet resist strip and cleaning steps, respectively (Figure 3-3 d). Similar lithographic and etching steps are used to expose the TiN top electrode of the MIM cell (Figure 3-3 de). More mature versions of this flow can exploit highly selective SiN/SiO_2

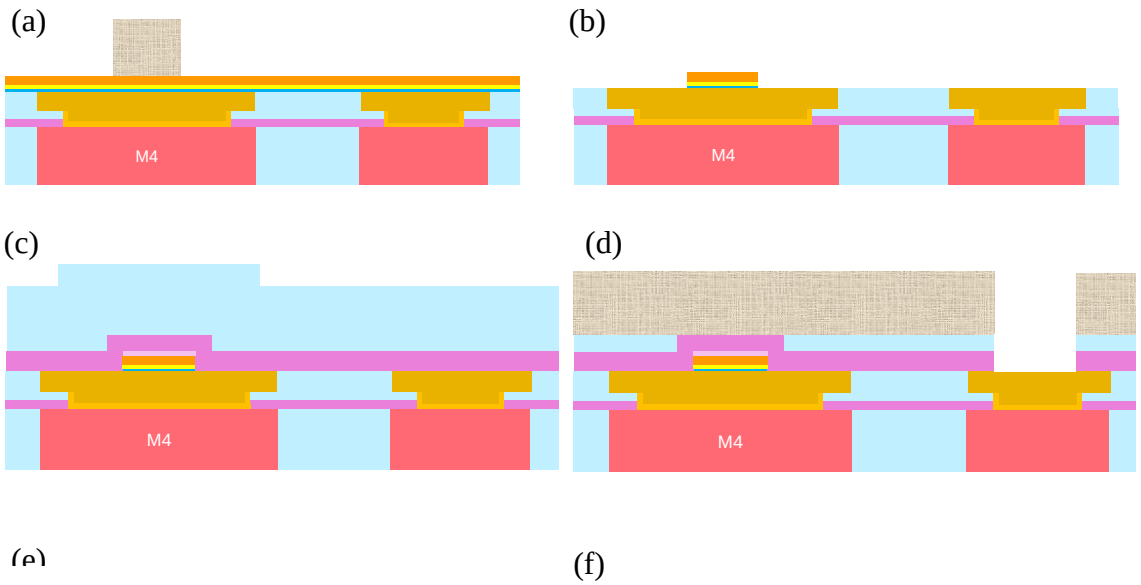


Figure 3-3 Integration flow of the MIM memory block: (a) Patterning of the TiN bottom electrode on M4, (b) Deposition of HfO_2 (c) BSI device: Full sheet ion implantation. (d) LSI device: Patterning and implantation through a local aperture directly on the oxide (e) deposition of the reactive and contact electrode metals (Ti/TiN).

etching chemistries to reduce the number of masks required. Finally, M5 is deposited and patterned and etch defining the interconnect (Figure 3-3 f).

The integrated memory cell right after contact etching of the top electrode is shown in Figure 3-4. This integration approach, although simple and flexible, presents several challenges in terms of process accuracy that we will briefly discuss in the following.

3.2.2.1 Bottom Electrode Surface Roughness

In this 200 mm process, CMP is used for the planarization of the bottom electrode in the beginning of the memory process block. The CMP on TiN is specifically optimized to minimize surface roughness (< 1.3 nm) and decrease topology effects across the memory cell (Figure 3-5a). This allows to homogenize the bottom electrode electrical behavior

3.2.2.2 Etching of the SiN passivation layer

In this integration approach, the HfO_2 is etched defining a cylinder-shaped MIM capacitor. The etching of the HfO_2 layer can, compromise the electrical integrity of scaled ReRAM devices: Removal of the HfO_2 layer makes the separation between the two metal layers

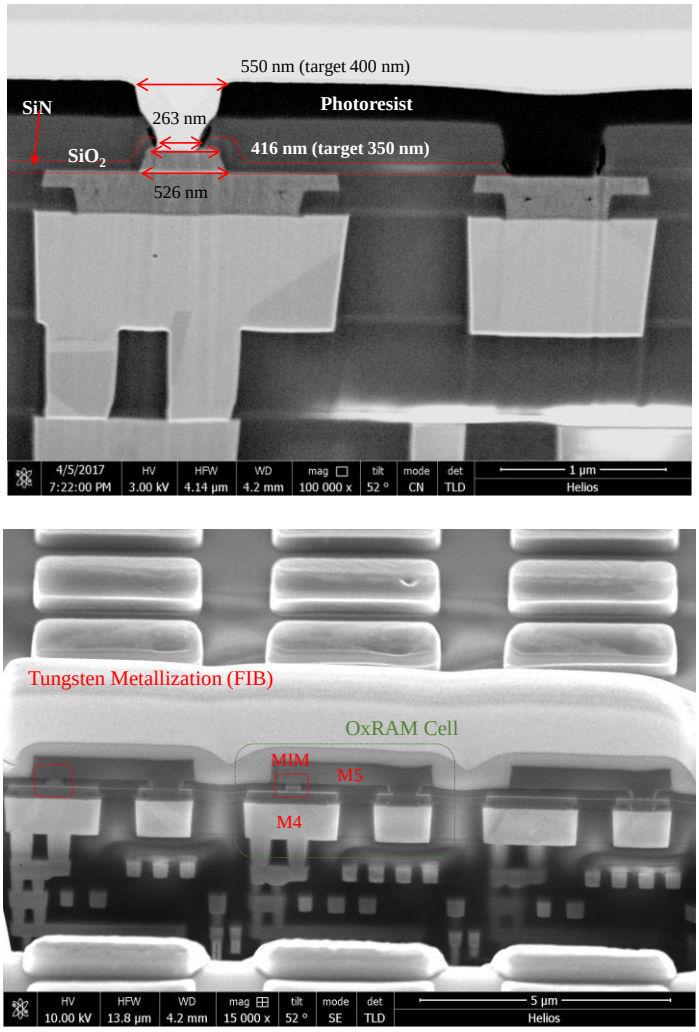


Figure 3-4 (Top) FIB SEM image of a 400 nm MIM ReRAM element integrated in M4, (structure on the left) post the top contact via (structure on the right) opening etching step. (Bottom) FIB profile of 400 nm MIM cell integrated inside a 4 Kbit ReRAM array

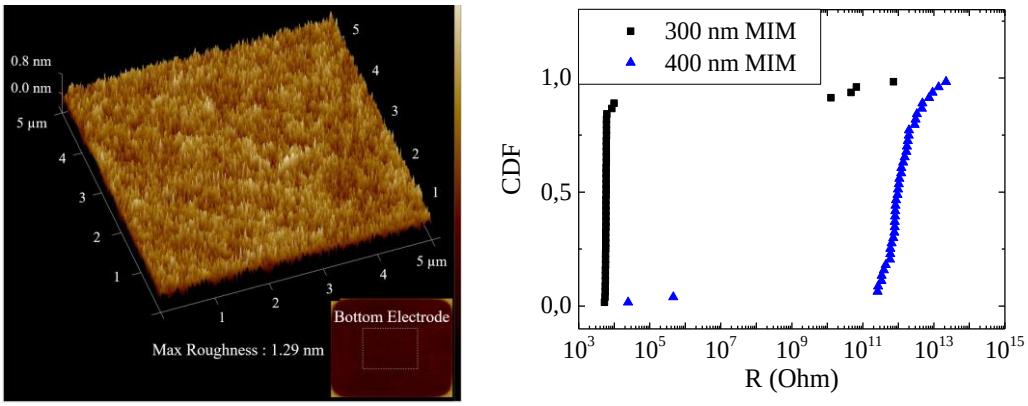


Figure 3-5 (a) topography profiling center-to-edge post CMP after TiN bottom electrode planarization (b) Diameter dependent shorts of devices below (black) and above (blue) the via critical dimension for reference HfO_2 (center-to-edge devices).

above and below the MIM structures with sizes smaller than the VIA4 critical dimension

(400 nm, Figure 3-4) dependent on the SiN layer.

.The center-to-edge etching variability can lead to complete removal of SiN layer during the top contact VIA4 opening step. As a consequence, the top electrode metal deposition contacts directly the bottom electrode TiN layer through the MIM sidewall if the MIM size is smaller than the critical size of the via. This leads to electrical shorts, as shown in Figure 3-5b. Reduction of the SiN etch rate so that the SiN sidewall recess can be efficiently controlled is critical in order to guarantee the cell electrical integrity.

3.3 Impact of local Si implantation on Pristine Devices and the Forming Operation

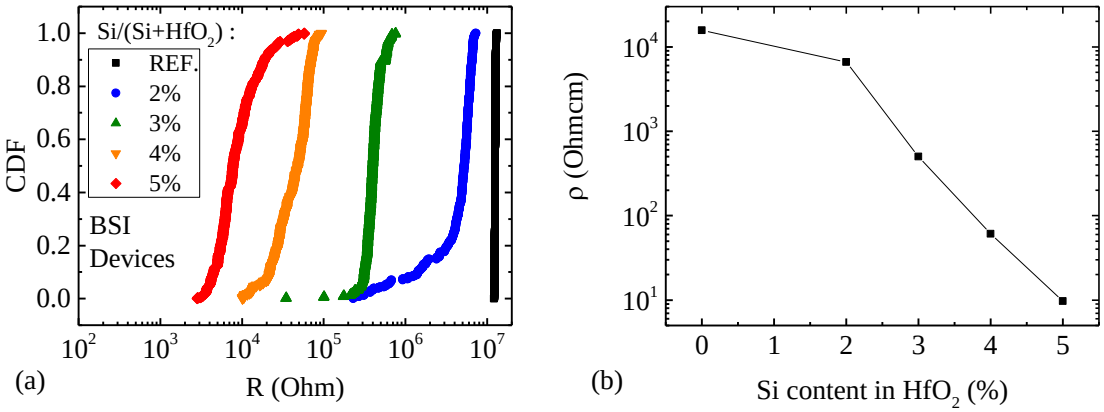


Figure 3-6 (a) Pristine resistance statistical spread of 4 kbit arrays implanted with Si using the BSI approach. (b) Median pristine resistivity evolution of OxRAM devices in a 4 kbit array with Si fraction for BSI and LSI devices, assuming Ohmic conduction in the low field approximation.

3.3.1 Pristine Resistance evolution with Si implant concentration

The impact of Si doping is measured in 4 Kbit arrays to achieve significant statistics. The cumulative probability distributions of the measured populations as a function of the Si content in the HfO_2 layer is shown in Figure 3-6a, extracted by reading the arrays using a 100 ns rectangular pulse at 100 mV. As in the previously shown case Si acts to enhance the conductivity of the layer, as already discussed in Chapter 2. The pre-forming resistivity extracted as a function of the Si content shown in Figure 3-6b.The good uniformity of the distributions acts to confirm the reproducibility of the physics described before. Deviation of the tails of the distributions are attributed to extrinsic process variation sources.

The median value of the pristine resistance of LSI devices stays significantly higher than that of the corresponding BSI devices (Figure 3-7a). Moreover, the pristine resistance ratio

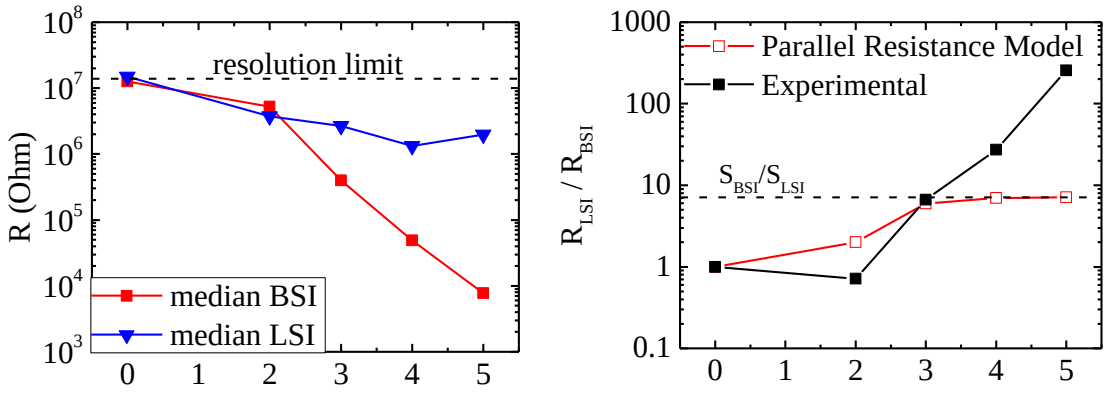


Figure 3-7 (a) Evolution of the median pristine resistance measured on 4 Kbit arrays for BSI and LSI devices (b) (Black) LSI/BSI pristine resistance ratio as measured in (a) and as calculated using Eq. (3.1).

of LSI to BSI devices does not follow the nominal expected surface ratio of the lithographically defined BSI (400 nm) to LSI (150 nm) zones of ~7.1. Below the 2% threshold the measurement is limited by the parameter analyzer current resolution. From the 3% point onwards the resistance drops drastically reaching a resistance ratio of 256 (Figure 3-7b).

To understand how low field conduction evolves in the LSI devices we consider the following Ohmic model: The total measured resistance is considered to comprise fo two parallel components, that of the implanted zone and that of the non implanted zone, $R_M^{LSI} = R_{Impl}||R_{no-Impl}$. Considering the cylindrical geometry of the MIM capacitor, the measured resistance for LSI devices can be expressed analytically in the ohmic approximation by Eq. (3.1) for low electric field:

$$R_M^{LSI} = \rho_{impl}\frac{4}{\pi}\frac{t_{ox}}{d_{LSI}^2 + \frac{\rho_{impl}}{\rho_{HfO_2}}(d_{MIM}^2 - d_{LSI}^2)} \qquad (3.1)$$

Where ρ_{impl} stands for the resistivity of the implanted material for a given concentration of Si in HfO_2, ρ_{HfO_2} the resistivity of the no-implant reference, t_{ox} the thickness of the oxide, d_{MIM} the MIM capacitor DUV defined diameter, and d_{LSI} the lithographically defined diameter of the implanted zone.

MC simulations of the implant suggest that the lateral straggle of the LSI zone is less than 10% of the total radius (Figure 3-8) while diffusion experiments of Si in HfO_2 (§2.9) and temperature dependence of the detected bonds from XPS (§2.7.1) suggest that the material structure post the BEOL thermal budget (T_{max}=400°C) is not impacted significantly.

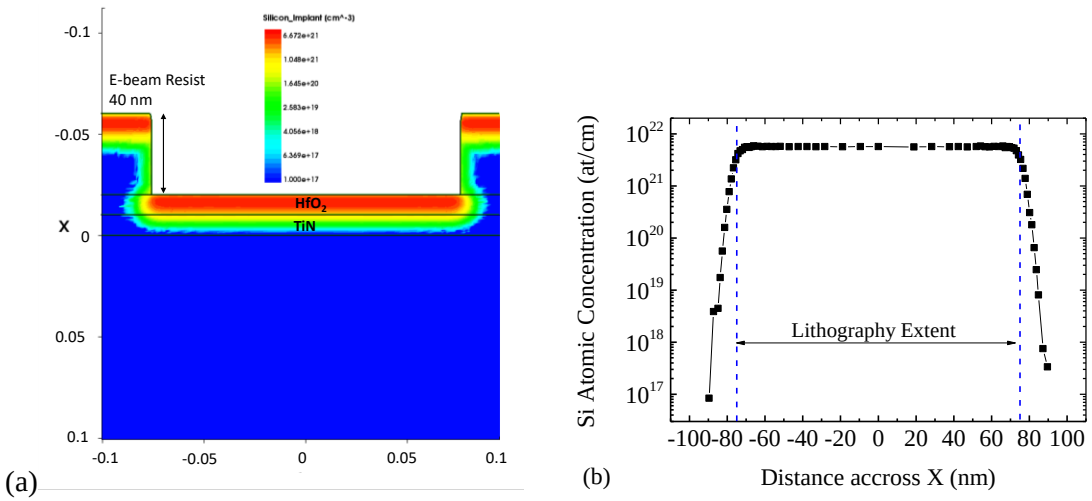


Figure 3-8 Lateral profile of the Si implanted species in a 150 nm implanted zone. (a) 2D Surface (b) Atomic concentration of Si across the radius of the device for an implantation step corresponding to 5% Si concentration. The calculated lateral straggle is in the order of (4-7) nm for the 5% LSI device.

Hence, the resistivity in the implanted zone of LSI devices is considered identical to that of BSI devices for a given Si content (Figure 3-6b).

The calculated LSI pristine resistance over the experimentally measured BSI pristine resistance ratio R_{LSI}/R_{BSI} is shown in Figure 3-7b. At low Si concentration (<2%), the LSI resistivity is very low and the ratio is close to 1 suggesting that the whole DUV volume participates in current conduction. Above the 3% threshold the LSI zone overtakes conduction reaching asymptotically the surface ratio of 7.1. Nevertheless, the expected and calculated resistance ratios differ by more than one order of magnitude. This points towards an unexpected conclusion: The effective surface of LSI devices is significantly lower, with $d_{LSI}^{eff} \sim 25$ nm if only the surface ratio approximation is considered.

3.3.2 Forming voltage evolution with Si implant concentration

The forming process has been studied for both types of devices on the array level to achieve significant statistics (Figure 3-9a). To form the devices, ramped voltage stress is used. A series of consecutive 100 ns long pulses with increasing voltage amplitude is applied on each of the cells consecutively inside the array from zero to 4 V with a 100 mV step. The current compliance through the cell is limited to 250 µA, by means of the transistor pass gate. In order to determine the forming voltage of the devices, the normal quantile plot of the resistance values of the array cells is constructed for every pulse

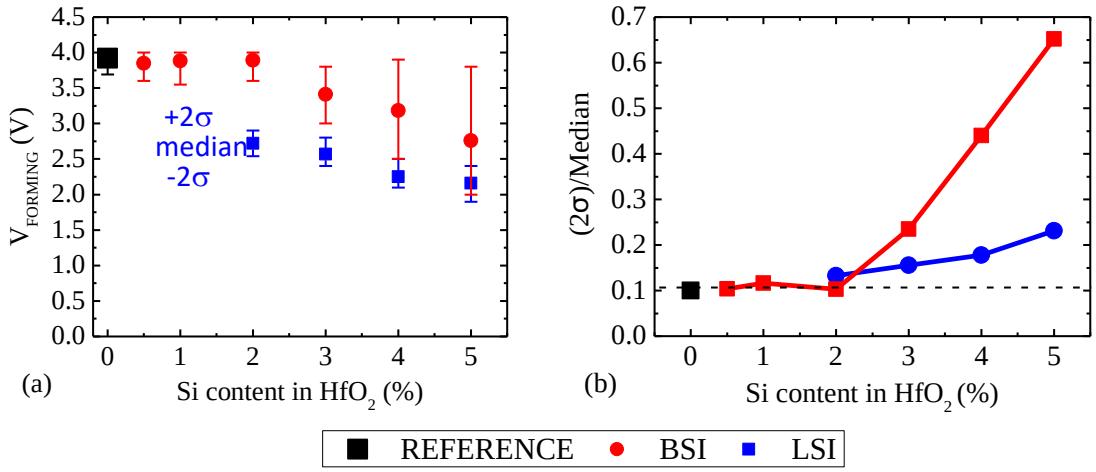


Figure 3-9 (a) Median forming voltage evolution and 2σ dispersion of OxRAM devices and (b) $\Delta V_F(2\sigma,-2\sigma)/V_{Median}$ ratio evolution with Si fraction in a 4 kbit array for BSI and LSI devices in pulsed mode (100 ns pulse, voltage ramp). While forming voltage is more significantly reduced in LSI devices its relative variability is lower as compared to BSI.

applied in the sequence and the resistance values at the median and 2σ are compared versus a target value expected from a formed cell. If the said resistance value is lower or equal to the target value[1] in the median, then the voltage of the pulse is considered as the forming voltage of the devices. The statistical spread of the given iteration is calculated in terms of σ and the 2σ is used to quantify the statistical variability of the forming process.

[2]Comparing the behavior of LSI and BSI devices as regards the forming operation it can be seen that the forming voltage decreases with increasing doping concentration as compared to the no-implant reference in both cases. This supports the device proof-of concept that the formed zone will occur in the implanted, trap-rich region. Unexpectedly, LSI devices present with forming voltage values, systematically lower than the ones of BSI devices with the same level of doping even though the LSI surface (either lithographic or effective) is expected to be significantly lower than the one of the BSI variants. Furthermore, even though forming voltage variability tends to increase with doping; it is controlled more effectively in the LSI approach (Figure 3-9b).

Considering the anomalous evolution of the pristine resistance values, this behavior of the forming voltage between BSI and LSI devices is not straightforward to interpret.

[1] The typical value of LRS for formed devices with a compliance of 250 µA during the SET operation is in the range of 3.4 KOhm (see Figure 3-26a). Here the threshold used is 5 KOhm.

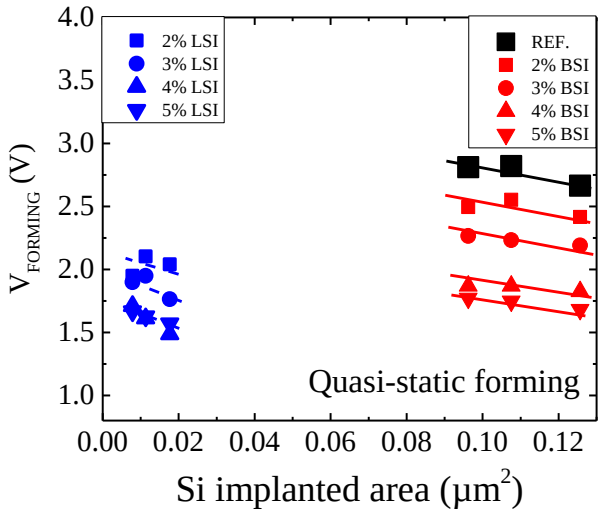


Figure 3-10 Quasi-static forming voltage evolution with surface for reference, BSI and LSI devices. MIM diameter varies in the 800 nm to 400 nm range while LSI zone in the 250 nm to 150 nm range.

Nevertheless, it is confirmed independently from quasi-static IV measurements on devices of multiple sizes and implanted surfaces (Figure 3-10).

For BSI devices the DUV defined MIM capacitor surface, whereas for LSI devices the e-beam defined, implanted region are considered. In both cases, forming voltage decreases with increasing doping concentration and weakly increases with decreasing total active surface. Again, in both cases LSI devices have forming voltages consistently lower than the ones measured in BSI devices.

3.3.3 Impact of Si doping on the diffusivity of O vacancies

To this date the specific details of the forming are still being contested. Even though the image of defect generation by means of oxygen vacancy migration towards the reactive electrode, creating a preferential percolation path (filament) is generally accepted, the specificities of the vacancy migration mechanism are not fully understood yet. The process has been described either by using the Nernst-Plank equations approach [Ielmini, Sandisk] or by stochastic MC algorithms []. In the semi-analytical approach of the Nernst-Plank equations, the oxygen vacancy diffusivity (D) with temperature and field is of major importance as it is critical in determining the rate of generation / migration of oxygen vacancies from the HfO_2 bulk towards the reactive electrode.

At the same time, ab-initio calculations have pointed out that the presence of Hf-Hf inter states (such as the ones found in the Si implanted material) can facilitate the migration of

oxygen vacancies [153]. In the following, we make an effort to measure the change in diffusivity of the (assumed) oxygen species as a function of Si doping using the temperature dependence of resistance at low field: The pristine resistance values of 400 nm 1R MIM capacitors have been extracted under different temperature conditions, from 25°C to 200ºC by measuring the current between zero and 0.4 V with 5mV step. The current response for a given stress point is measured across a characteristic measurement time. To resolve low currents the measurement time is fixed in 10 PLC which corresponds to 200 ms / stress point, a time considered sufficient to observe field-induced oxygen migration in HfO_2 [166].

The resistance value and diffusivity can be linked by means of the Einstein relation and fundamental relations as shown in equations:

$$R = \frac{1}{\mu e n}\frac{L}{S}\ (a), \qquad D = D_0 e^{-\frac{E_{act}}{k_B T}}\ (b), \qquad D = \frac{\mu k_B T}{q_e}\ (c) \tag{3.2}$$

The diffusivity using thermodynamical arguments can be expressed as [167]:

$$D(T) = \frac{k_B T}{h}\frac{n d^2}{2a}\exp\left(\frac{\Delta S_{vib}}{k_B}\right)\exp\left(-\frac{\Delta U_{vib}}{k_B T}\right)\exp\left(-\frac{\Delta E}{k_B T}\right) \tag{3.3}$$

From which $D_0(T)$ is defined as:

$$D_0(T) = \frac{k_B T}{h}\frac{n d^2}{2\alpha}\exp\left(\frac{\Delta S_{vib}}{k_B}\right)\exp\left(-\frac{\Delta U_{vib}}{k_B T}\right) \tag{3.4}$$

Combining the set of equations (2.4)-(3.6) yields:

$$R = \left(\frac{t_{ox} k_B T}{S n q_e^2 D_0(T)}\right) e^{\frac{E_{act}}{k_B T}} = \frac{L k_B T}{S n q_e^2}\frac{h}{k_B T}\frac{2a}{n d} e^{-\frac{\Delta S_{vib}}{k_B}} e^{\frac{\Delta U_{vib}}{k_B T}} e^{\frac{\Delta E}{k_B T}} \tag{3.5}$$

$$R = R_0 e^{\frac{E_{act}}{k_B T}},$$

$$R_0 \equiv \frac{L h}{S n q_e^2}\frac{2a}{n d}\exp\left(-\frac{\Delta S_{vib}}{k_B}\right), \tag{3.6}$$

$$E_{act} = \Delta U_{vib} + \Delta E$$

In (3.4) R is the measured resistance, t_{ox} the oxide thickness (length of the assumed resistor in the Ohmic approximation), S the resistor surface, E_{act} the diffusivity activation energy (assuming oxygen species diffusion/migration), μ the ion mobility, q_e the electron charge, d T the temperature. In this approximation we consider the electron density in the oxide n and the vibrational enthalpy ΔS_{vib} temperature independent in the first order, as the enthalpy will depend on the system's microstates; characteristic of the material. The activation energy E_{act} is defined as the sum of vibrational internal energy ΔU_{vib} and the external energy provided into the system ΔE to account for the temperature dependence of the D_0 pre-exponential factor.

Figure 3-11a shows the temperature dependence of the median low field pristine resistance of 50 BSI devices. The experimental resistance medians follow well eq. (3.6) reasonably well. Doping acts to significantly decrease the diffusivity activation energy (Figure 3-11b). This quantity can be interpreted as the oxygen species migration barrier from one site to the other inside the oxide. Lowering the activation energy, leads to a strong increase of diffusivity while its temperature dependence is reduced. This is consistent with the strong dependence of the forming voltage of undoped HfO_2 ReRAM cells with temperature, and its independence of the latter in strongly Si doped samples (Chapter 2, Fig. 2.12).

$$\mathbf{J} = \mathbf{J_{diff}} + \mathbf{J_{drift}} = -\left[D\nabla c + \frac{Dzq_e}{k_B T} c(\nabla\varphi)\right] \tag{3.7}$$

According to the oxygen vacancy migration model, the quasi-particle being the charge carrier is governed by a concentration (c) dependent diffusion term and a field (φ) dependent drift term (Eq.(3.8)). Hence a given concentration or field gradient will result in a diffusion of the charge carrying quasi-particle. If oxygen vacancy migration is considered with a fundamental charge state zq_e the quasi-particle will move in a random fashion even if the field is directional oved to perturbations of the local potential landscape. The image is complementary to the variable range hopping mechanism ().

Probing the diffusivity of vacancies in this fashion can be understood as follows: When bias is applied for a given time the initially existing trap sites in the oxide will be filled with electrons, causing vacancies to drift towards the reactive electrode. Nevertheless, for ions to be displaced higher energy is required, hence no permanent redistribution or creation/breaking of bonds takes place. Augmenting the concentration of Si creates more trap sites as already discussed in chapter 2. For this measurement's case, the decreasing activation energy can be interpreted as the vacancies sensing a total field stronger than the externally applied one, suggesting the creation of a built-in

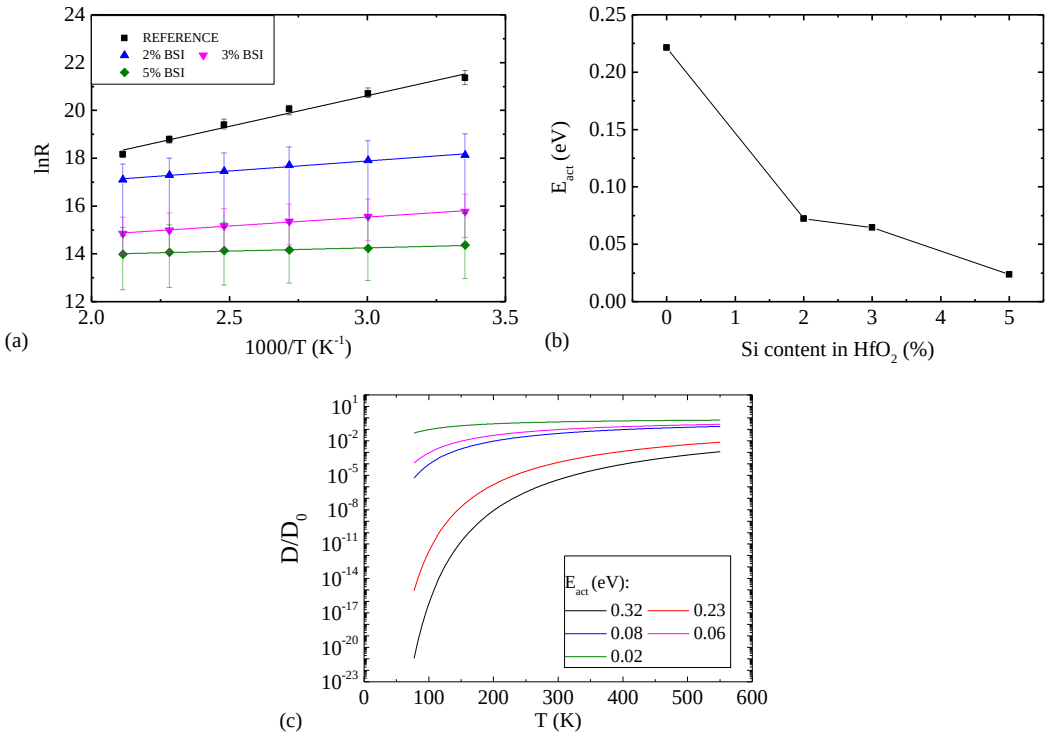


Figure 3-11 (a) Measured resistance as a function of temperature for different Si concentration as measured in BSI devices. (b) Extracted activation energy of diffusivity and associated resistance pre-exponential factor. (c) Evolution of the diffusivity exponential term as a function of temperature for different ranges of activation energy.

field in the oxide. Such field can indeed be created by the existence of a charge distribution in the volume of the oxide, supported by the experiments presented in Chapter 2.

Hence the facilitation of the migration of a vacancy can be considered equivalent to the local potential being lowered as the population of charge trap sites available for an electron to hop in and out of is increasing, and their mean distance is decreasing as a consequence. From that standpoint, forming can naturally be seen as a breakdown event accelerated by the drift of defects. The more defects the lower the external field needed to form as breakdown is a cascade event. In this context the lowering of he forming voltage in Si implanted devices is a direct consequence of the increase of diffusivity as observed in BSI devices (Figure 3-9, Figure 3-10) or full-sheet Si implanted HfO_2 (Ch2., Figure 2.12). Nevertheless, this does not suffice to explain why the forming voltage of LSI devices is lower than that of BSI while both their nominal surface is smaller and pristine resistance higher as previously demonstrated.

Using the diffusivity model above, and considering that $V_{bd} \propto 1/J \propto 1/D$ a further reduction of the forming voltage would require either an enhancement of the electric field in LSI devices, or localization of the motion of vacancies. The first will result in augmentation of the ionic current

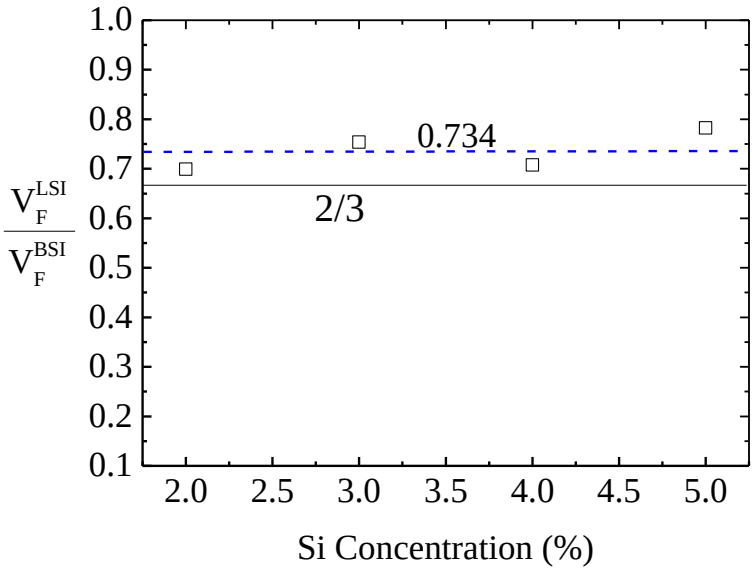


Figure 3-12 Evolution of the ratio between median forming voltages of LSI devices over median forming voltage of BSI devices as calculated from the experimental data of Figure 3-9.

according to Eq. (3.7) while the second in an increase of diffusivity owed to the *localization of the Brownian motion* of the quasi particle. The concept of diffusivity scaling with the dimensionality of the motion is consequence of diffusion being a Markov process[2]. Continuing on the previous argumentation $V_{bd} \propto 1/J \propto 1/D_{bd} \propto 1/J \propto \alpha$: One expects that as the vacancy potion is restricted in a lower dimensional system the resulting breakdown voltage would be reduced equivalently, the ratio of breakdown voltages between two otherwise identical systems of different dimensionality would be:

$$\frac{V_{bd,2}}{V_{bd,1}} = \frac{D_1}{D_2} = \frac{\alpha_2}{\alpha_1} \tag{3.8}$$

Where V_{bd} stands for the breakdown voltage, D is the diffusivity and α the dimensionality factor of the motion namely 1, 4, or 6 for a 1D, 2D or 3D random walk.

In Figure 3-12 we present experimental data of the evolution of the ratio of the median forming voltage of LSI over that of BSI devices from Figure 3-9. We find that the ratio is, within the accuracy of the experimental error independent on the doping level and close to 2/3. This supports the assumption that the filament growth during the forming process follows a random walk process in 3D space for BSI devices and in the 2D space for LSI

[2] See Diffusivity in 1D and Diffusivity as a Markov Process and dimensionality scaling

devices. The fact that the ratio does not appear to be dependent on the Si doping concentration means that the effect is primarily geometrical in origin.

This hypothesis, albeit being aligned with experimental data does not justify the origin of localization but only requires it. Nevertheless, as detailed in the previous chapter, evidence suggests that the charge inside the volume of Si implanted HfO_2 is increasing in proportion to the Si implanted species. If the electrostatic behavior of the charges is taken into account, the charge gradient between the implanted and non-implanted volume can result in a discontinuity of the electric field in the transition between the two regions. This can locally enhance the electric field and localize the vacancy motion in that interface as will be detailed in the following section.

3.4 Electrostatics of ReRAM Devices with cylindrical geometry

In Chapter 2 the defect generation mechanism in Si implanted HfO_2 was analyzed in detail. The increase of conductivity traced in the pristine devices is associated with the presence of Hf-Hf states that appear in proportion to the Si implant dose. Furthermore, the concentration of Hf-Hf states in 5% Si implanted oxide was found to be in the order of0.32% of the total bonds detected. As previously discussed, evidence suggests that defect generation occurs in the full volume of the oxide and not only in the interfaces.

As the defect generation mechanism is associated to the Si implant, the implanted volume charge density is modeled as:

$$N_{V_1}(r)\left[\frac{Cb}{m^3}\right] = 0.0032 \cdot zq_e C_{Si^{impl}}(r) \qquad (3.9)$$

Where z the $Hf_i^{\bullet}$ assumed to be 1 for simplicity, q_e the elementary electron charge and C the lateral profile of Si at depth $t_{ox}/2$ (peak of implant profile) implanted species the length of the implanted region as calculated by Monte Carlo simulation in Sentaurus TCAD (Figure 3-8). Any charge in the no-implant region is considered to be sufficiently smaller than the one in the implanted region so that it can be ignored in the first order analysis.

The 2D Gauss law is solved numerically using COMSOL finite element analysis package, for a cylindrical geometry of regions I, II and III extending for a thickness of t_{ox} (10nm) in zero bias conditions:

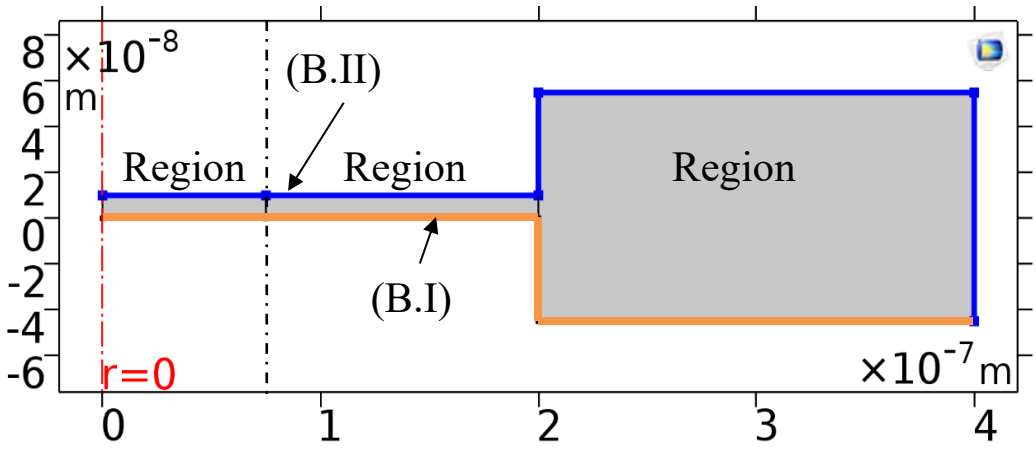


Figure 3-13 Simulated dielectric regions of a MAD200 ReRAM cell (I): Implanted Region, with charge defect density $\rho_v \neq 0$ (II) Non-implanted region. (III) SiO_2 passivation layer.

i) Region I: An inner cylindrical region with dielectric constant ε_{r1} with a given volume charge density $\rho_{V_1}(r)$ with 75 nm radius, representing the LSI region.

ii) Region II: An outer cylindrical region with dielectric constant ε_{r2} with a given volume charge density ρ_{V_2}. Such that $N_{V_1} \gg N_{V_2}$ and ρ_N is considered essentially zero representing the HfO_2 surrounding the LSI region. Region II extends up to 200 nm.

iii) Region III: An encapsulating SiO_2 environment in contact with region II and the metallic electrodes with $N_{V_3} = 0$. The oxide thickness is 100 nm in accordance to the ReRAM cell characteristics.

The potential variation of the dielectric constant of the implanted region owed to material alloying is approximated by linearly interpolating between the limiting values of SiO_2 and HfO_2 respectively, in proportion to the atomic concentration of Si P (%).

$$\varepsilon_{HfSiO_x} = P \cdot \varepsilon_{SiO_2} + (1 - P) \cdot \varepsilon_{HfO_2} \qquad (3.10)$$

Dirichlet boundary conditions are used for the top and bottom electrode nodes. Boundary condition (B.I) represents the iso-potential surface in contact with TiN and Cu interconnect carrying the top electrode voltage, while boundary condition (B.II) represented the bottom electrode set to ground.

$$V = \begin{cases} \text{const} & \text{(B.I)} \\ 0 & \text{(B.II)} \end{cases} \qquad (3.11)$$

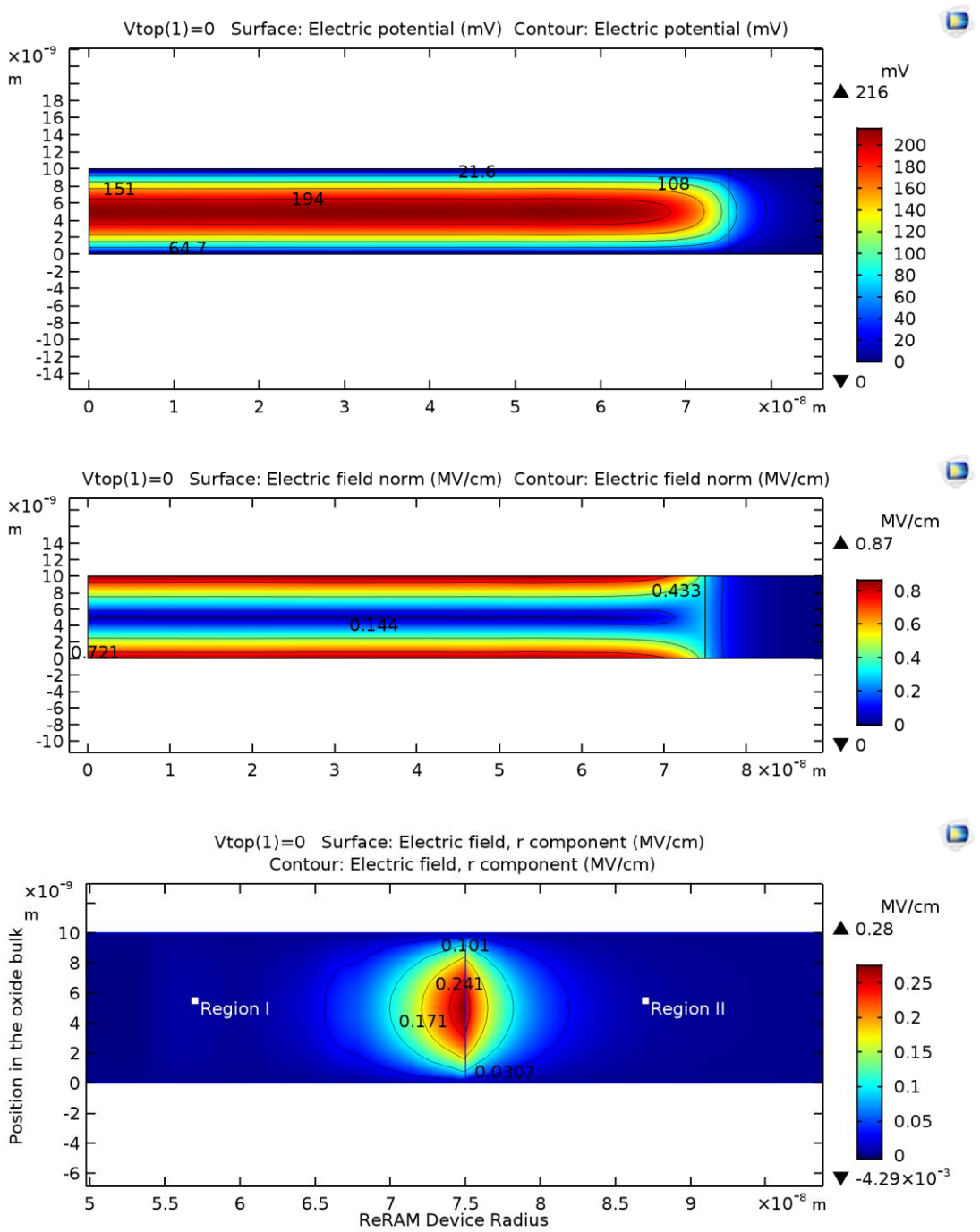


Figure 3-14 2D Surface plot of (a) the built-in potential (b) the total electric field norm and (c) the radial component of the electric field. Here, $\varepsilon_{r1} = 19.2$ *(5% Si HfSiOx),* $\varepsilon_{r2} = 20$ *(*HfO_2*).*

First, the zero-bias case is studied to understand the behavior of the built-in field. The space charge introduced gives rise to a built-in potential in the order of 216 mV (Figure 3-14a) inside region I. The total electric field norm is shown in Figure 3-14 b and is found to be in the order of ~0.87 MV/cm; while the radial component of the field is shown in

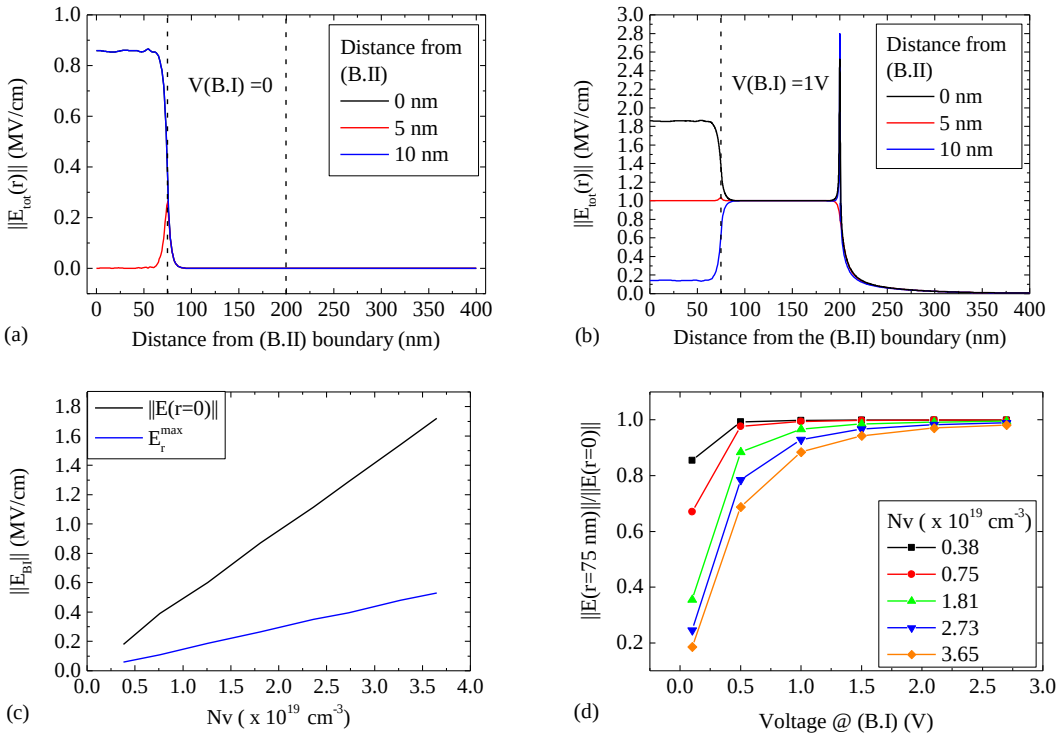


Figure 3-15 Electric Field norm close to the grounded boundary (B.II), in the middle of the layer ($z=t_{ox}/2$) and near the (B.I) boundary without (a) and under the influence of external bias (b). (c) Projected evolution of electric field norm and radial components as a function of the charge density in region I. (d) Peak-to-base ratio of the norm of the electric field in the middle of the layer for different amounts of charge

Figure 3-14c. The latter is of particular interest as its magnitude and distribution is only dependent on the amplitude of the space charge distribution (total charge in the LSI volume), while its spatial extent depends on the sharpness of the charge density profile and the dielectric constant mismatch between the two regions. In this model, the outer HfO_2 charge density is assumed to be negligible, supported by the calculated space charge evolution during Si ion implantation (Ch2. Fig 2.2xx , increase of charge in the order of 275%).

The norm of the electric field is studied for (a) zero-bias conditions and (b) 1V applied bias on the (B.I) boundary, representing the HfO_2/Ti top electrode. (B.II) is always on ground. At zero-bias conditions (Figure 3-15a) the assumed charge density results in a significant built-in field, with the magnitude of the radial component being 30% of the total magnitude in the center of the device. In this case, the norm of the field is maximized at the metallic interfaces while it drops to zero at $t_{ox}/2$ (*Figure 3-14*b). The symmetry of the norm in the top and bottom interfaces arises from pure geometrical considerations for

the structure and the distribution of charge. As the LSI volume is populated by a net charge, local dipoles will be created. Near the top and bottom interfaces, in contact with the metallic Ti and TiN surfaces, respectively, an image charge will be created to compensate for the dipoles in the side of the oxide. As a result the field in the z direction of the oxide will be screened while near the interfaces will be maximized, pointing "outwards" from the bulk of the oxide and symmetrically with respect to $t_{ox}/2$.

In the presence of an externally applied electric field breaks the symmetry between the two interfaces: The field across the top interface where the external bias is applied is canceled out owed to the presence of the built-in field ($\mathbf{E_{BI}} \cdot \mathbf{E_{ext}} < 0$). In the bottom interface,

$\mathbf{E_{BI}} \cdot \mathbf{E_{ext}} > 0$ and the fields act synergistically so the total field is increased. For high enough external bias the field induced the surface charge induced by biasing eventually dominates the dipole-induced built-in surface density and the external field dominates. The convolution between the external-bias dependent vertical component and the bias-independent radial component eventually decreases the importance of the field discontinuity in the LSI/non-implanted HfO_2 interface for high fields.

The projected dependence of the total built-in field in the center of the device (r=0) and that of the peak of the radial component vary linearly with the total amount of charge (Figure 3-15c). For levels of charge ($N_v < 1 \times 10^{19}$ cm^{-3}) such as the ones found in the reference or devices implanted with Si below 2% both the vertical and radial components are negligibly small (< 0.1 MV/cm). The base/peak total field ratio is shown in (Figure 3-15d) for the range of biases typical of that used in the forming process of Si implanted devices. For the charge initially projected into the oxide ($N_v < 1.81 \times 10^{19}$ cm^{-3}) the base to peak ratio becomes unity in the range of 1.5V. For ($N_v \sim 3.7 \times 10^{19}$ cm^{-3}) the voltage where the ratio approaches unity is augmented to ~2.5V.

In this framework, the reduction of forming voltage observed in Si implanted devices can be understood as follows: The built-in field created owed to the space charge density increases diffusivity of vacancies this reducing the forming voltage. This can be easily understood if in Eq. 3.7 the potential term is expressed as $\varphi = \varphi_{ext} + \varphi_{BI}(C_{Si})$. Hence, if oxygen vacancies (V_O^-) are considered to be diffusing under the influence of a built-in field generated by positive immobile charge centers (Hf_i^+), they will experience an additional drift force with direction towards the high potential interface during the forming process;

further supported by the measured decrease of the diffusivity activation energy (Figure 3-11). The mechanism is in principle similar to diffusion enhancement by charged vacancies []. In this image, the reduction of forming voltage in BSI devices can be understood.

The further lowering of the forming voltage reported in LSI devices, cannot be explained by diffusivity enhancement owed to the vacancy species *alone*. Nevertheless, the enhancement of the electric field in the boundary between the charge-no charge region can introduce an additional confinement of the drifting/diffusing vacancies in the thin cylindrical shell between the implanted and non-implanted regions. Considering the spatial distribution of the field norm, the vacancy generation and motion can be understood conceptually in the following image: In the presence of high amounts of charge and under the influence of an external electric field, the electric field norm will be maximum in the grounded electrode, smaller towards the biased electrode while it decreases the when moving away from the two interfaces becoming zero in the center of the oxide owed to the symmetry of the structure. At the same time, the radial component of the electric field is maximized in the boundary between the implant/no-implant regions and in particular inside the bulk of the oxide away from the two interfaces, independently on the applied potential.

A vacancy is expected to be generated in the high-electric field region close to the bottom electrode. Under the influence of electrostatic force, vacancies will migrate towards the top electrode. As they move away from the bottom electrode / oxide interface the vertical component of the electrostatic force will decrease near $t_{ox}/2$ and motion will be dominated by the concentration gradient of the vacancies and the radial component of the electrostatic force. Owed to the presence of E_r the generated vacancies will still be dragged near the electric field discontinuity region, in the cylindrical shell interface between the implant/no-implant zones. This is potentially facilitated from the fact that ionic motion is not inhibited in the z direction owed to the uniformity of the Ti reactive electrode in the region. As they approach the top electrode / oxide interface they are injected into the Ti layer owed to the re-enhancement of the electric field. Under this scope, ionic motion is expected to occur

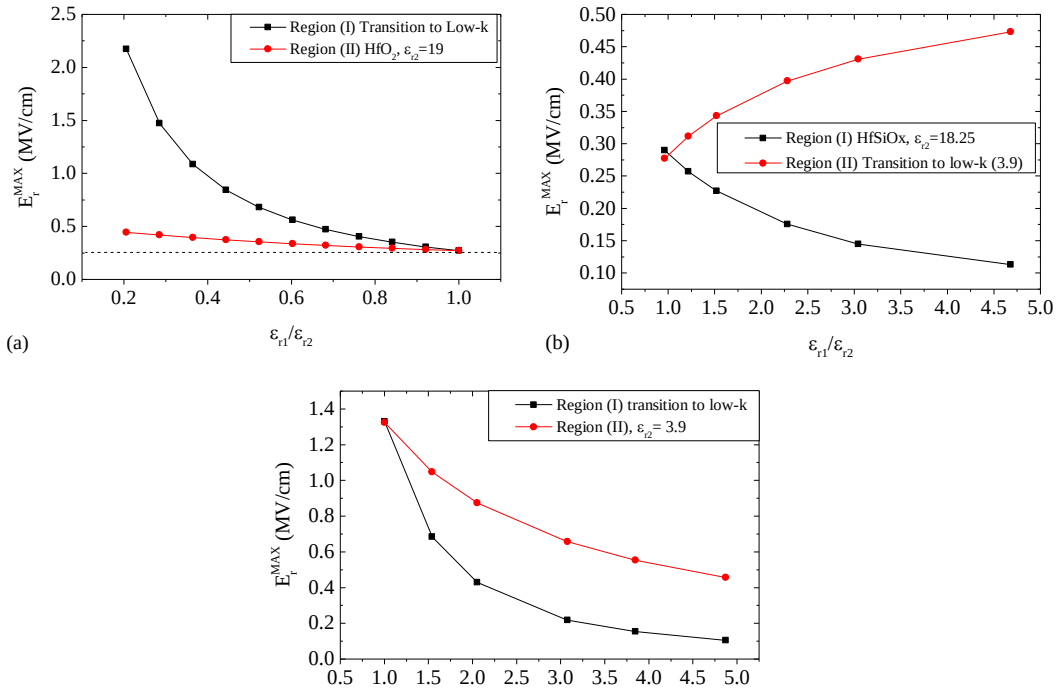


Figure 3-16 Dependence with dielectric constant mismatch

in the 2D surface on the outer region of the implanted volume. Considering the vertical field component fluctuates with the position in the oxide depth, the motion could be considered quasi-random within the 2D surface and hence, the diffusivity of the vacancy is further increased owed to the dimensionality scaling of the random walk. In this image, a further reduction of the diffusivity is expected. As this enhancement is related to the properties of space, the enhancement factor is expected to be independent on the doping concentration, as long as 2D confinement holds, supporting by the independence of the V_F^{LSI}/V_F^{BSI} ratio to the concentration of Si in HfO_2 (Figure 3-12).

The impact of the fringing field E_r between the charge/no-charge region is of particular importance as it scales with charge and dielectric constant mismatch, hence it will also affect BSI devices. In particular owed to the substitution process of Si in the HfO_2 matrix the dielectric constant is expected to drop. Moreover, the BSI device charged region is the whole DUV defined cylinder, with the edge being in direct contact with the SiN spacer. Owed to the dielectric mismatch, the fringing radial field will my more significant inside the region with the lowest dielectric constant.

The magnitude of E_r is dependent on the dielectric constant mismatch for a given charge density. Different scenarii have been examined for the typical values of E_r when the

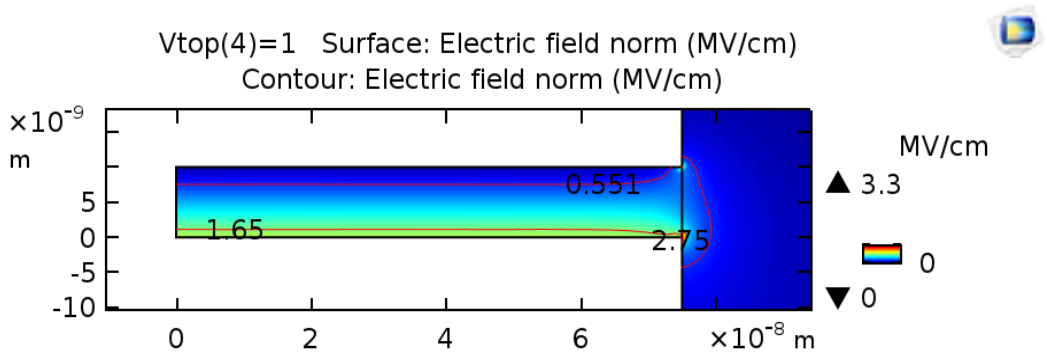


Figure 3-17 Electric field norm for a BSI device of the same dimension as the LSI zone, in contact with a SiO_2 low-k.

charged region I transitions to low-k values owed to alloying (Figure 3-16a) in contact with a low-k region; when region II retains high dielectric constant but while the environment transitions to low-k (Figure 3-16b); or when region I transitions to a low-k while being in contact with another low-k region. Case (a) is representative of the LSI device where switching is confined inside the doped HfO_2 where cases (b) and (c) are more representative of the BSI device, examining the cases where the dielectric constant of the doped HfO_2 varies lightly or strongly with doping. In both cases, the fringing field is more significant in the environment region for high dielectric mismatch, the worst case being when the doped region dielectric constant is close to low-k values.

In Figure 3-15 a representative image of the electrostatics of a BSI device under 1V bias is shown. The electric field is maximum close to the bottom interface (ground) while a significant spiking in the corners of the two electrodes appears owed to charge, dielectric constant mismatch and geometry field focusing (corner effect). Considering that the oxide is defect rich for high doping, it is expected that the edge region will participate in the percolative conduction. Nevertheless, the region material properties in the real BSI device are strongly impacted by the different oxide in contact, the lithography and etching process creating line edge roughness and the decreased quality of the Ti layer on top. This can explain the increasing variability of forming in BSI with Si doping as previously reported Figure 3-9b). The variability of forming is still increasing with Si doping, but at a much lower rate. This could be attributed to the increase of the lateral straggle of the doping profile as $\sim \sqrt{D}$, D being the implant dose.

It is unknown if the fringing field in BSI causes localization, yet data do not support this scenario. A possible reason that the localization effect of the fringing field is lost could be

strong scattering because of the increased surface roughness along the BSI/environment edge. Finally, this image is also expected to strongly effect the state variability of LSI and BSI devices versus those of the reference for a given programming condition.

As a final note, it should be noted that a careful study of the electric field distribution can provide an understanding of the pure HfO_2 device as well. In the absence of volume charge and for a given applied bias, the electric field is homogenous and constant throughout the volume of the MIM capacitor. This suggests that, in a first order defect generation is equi-probable at any point inside the volume. Nevertheless, the fringing field in the edge is still present to a certain extent owed to geometrical corner effects and dielectric constant mismatch. Hence, even though the major population of devices are formed within a small range (as is the case of the reference) if the statistical ensemble is large enough, some extreme events where the edge has participated into the formation of the percolation path can still be found, giving rise to devices forming behavior strongly different from the median. The behavior and importance of the fringing field can potentially explain reliability effects in ReRAM devices and will be briefly discussed in relation to experimental data in the following sections.

3.5 Impact of LSI on the DC SET and RESET voltages of HfO_2 ReRAM devices

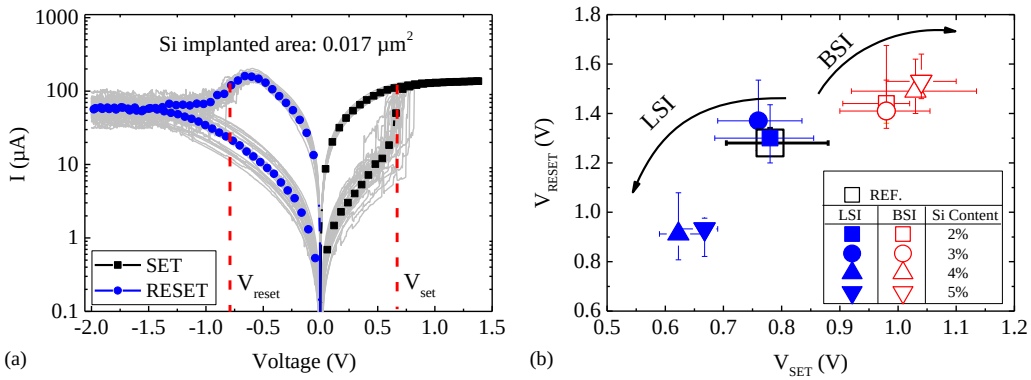


Figure 3-18 (a) Across-wafer device-to-device quasi-static IV curves after 20 SET/RESET sweeps. Median SET (black) median RESET (blue). (b) Correlation between the median quasi-static SET & RESET voltages for BSI and LSI devices extracted from (a).

Figure 3.13a shows the 20th DC SET/RESET sweep of 30, 5% LSI devices. The highlighted values correspond to the median current device to device. The DC SET & RESET voltages are determined experimentally, defined as the voltage at which the

current gradient is maximized as per Eq. 3.12 and is determined by numerical differentiation over the experimental IV data.

$$V_{SET,RESET} = V\left[\frac{dI_{SET,RESET}}{dV}\Big|_{max}\right] \quad (3.12)$$

The extracted SET and RESET voltage median values (30 devices across wafer per data point) and corresponding error represented by the 10% to 90% normal quantiles are shown in correlation of one to the other for every variant tested (Figure 3.13b). The impact of the localization of the switching zone has a direct implication on the switching voltages of BSI and LSI devices. While both set and reset voltages of BSI devices tend to increase as compared to the no-implant reference, the corresponding ones of LSI devices strongly decrease. Another potential reason is that as the Ti region on top of the BSI sidewalls has reduced reactivity owed to passivation from the environment. If the percolation path is formed near the edge, then the passivated Ti volume in the vicinity is less efficient in storing oxygen vacancies. Hence, a higher field is needed to perform the transition as observed in the figure above.

In the following, an in-depth study of the implications of localization of the filament is presented with respect to programming states, reliability, device endurance and data retention.

3.6 Impact of filament localization on the statistical properties of LRS and HRS

Table 3-1 Summary of forming and programming conditions used on 4 Kbit arrays.

		FORMING		PROGRAMMING			
		FC.1	FC.2	PC.1			PC.2
		T_{BD} @ V_{SL}		*SET* | *LSI*	SET | *REF*	RESET	SET
Parameters	I_{cc} (mA)	0.03	0.03	0.24		3.0	0.09
	V_{BL} (V)	3	4	1.7	2.0	0	1.7
	V_{SL} (V)	0	0	0		2.5	0
	t_{total}^{stress} (µs)	5	25	-		-	-
	t_{pulse} (ns)	100					

The evolution of SET and RESET voltages in BSI and LSI devices insinuates that LSI devices can operate at lower voltages as compared to undoped HfO_2 or BSI devices. It has been observed that while the current compliance is the predominant parameter that defines

the intrinsic level of LRS, the top electrode bias plays an important role in controlling the statistical spread of the distribution of LRS states of large memory populations []. In other words, current compliance fixes the median value of the CDF, associated with the true intrinsic behavior of the ReRAM population while the top electrode voltage acts to control extrinsic variability which appears in the tails of the CDF. As a result, the stronger the extrinsic variability of a chip, the higher the top electrode voltage required to control the tails of the CDF.

To examine the impact of local silicon implantation on the control of extrinsic variability of ReRAM devices, the following experiment is designed: 4 Kbit ReRAM arrays are formed, using a low compliance current TDDB approach where the total stress time and applied top electrode voltage are optimized between doped devices and the no-implant reference as in table 3-1 and then programmed 100 consecutive times under identical programming conditions. Programming using PC.1 (Table I, PC.1) allows a stable median-to-median window on the array level for all device variants and is used for comparison. The state of the LRS and HRS state of the matrix is assessed at the 100^{th} operation by evaluating the level of overlap between the statistical distributions of LRS and HRS on the 4 kbit array level. The resistance values of the array addresses are mapped on the lognormal distribution space, which consists the best approximation to describe the statistics of both LRS and HRS. The complete statistics is in essence multimodal and more than one distributions are required to fully describe either state including the CDF tails. Figure 3-19 (a)-(d) shows the lognormal distribution probit diagram. The memory window (MW), defined as the worst-case HRS/LRS resistance ratio at different extremes (Eq. (3.13)), provides a good measure of the extrinsic variability is the matrix memory window when assessed on the array level at a given programming cycle and was analyzed as a function of the Si content for both LSI and BSI devices.

$$\mathrm{MW_i} = \frac{\mathrm{R_{-i}^{HRS}}}{\mathrm{R_i^{LRS}}}\ , \mathrm{i} = 0, \sigma, 2\sigma, 3\sigma \qquad (3.13)$$

LSI devices present drastic reduction of the LRS distribution upper tail and partial reduction of the HRS distribution lower tail with Si increasing, that allows complete separation between LRS and HRS at 3σ (Figure 3-19 a-c) for the 5% Si LI devices. This is in line with the lower set/reset voltages measured on LSI unitary cells (Figure 3-18). On the contrary, BSI devices at 5% Si content present significant overlap between HRS and

LRS (Figure 3-19d). This can be attributed to the increasing role of the fringing field in the switching of BSI devices at high volume charge, supported by the discussion in §3.5. The impact of doping and switching zone localization is clearly visible in the array statistics: While the median MW stays mostly unaffected in BI devices until 3% Si, the 3σ MW steadily increases with increased Si content. Nevertheless, they all degrade above 3% (Figure 3-20a): the HRS is limited by the degradation of the pristine state of the device (Figure 3-6) as the material becomes metallic. On the contrary, the MW of LSI devices always increases even at 3σ with increasing Si while the median-to-median MW is stable (Figure 3-20a). The trend is more pronounced at 3σ with the 5% Si LSI devices exhibiting an improvement of 1 order of magnitude as compared to the non-implanted reference. The lower LRS variability can be attributed to the localization of the defects in the switching

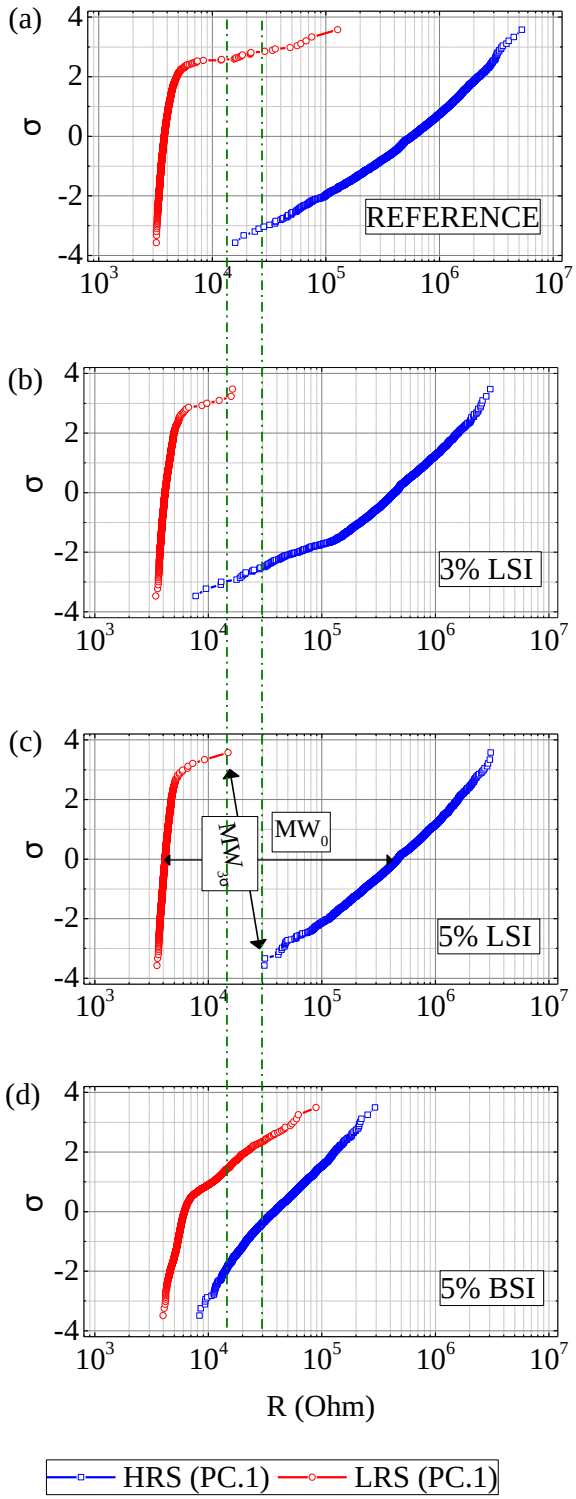


Figure 3-19 (a)-(d) LRS & HRS state distributions evolutions in 4 kbit arrays varying BSI/LSI Si fraction as compared to standard HfO_2. Array state assessed post forming and 100 cycles at (PC.1). MW (Memory Window) defined as the resistance ratio of HRS/LRS.

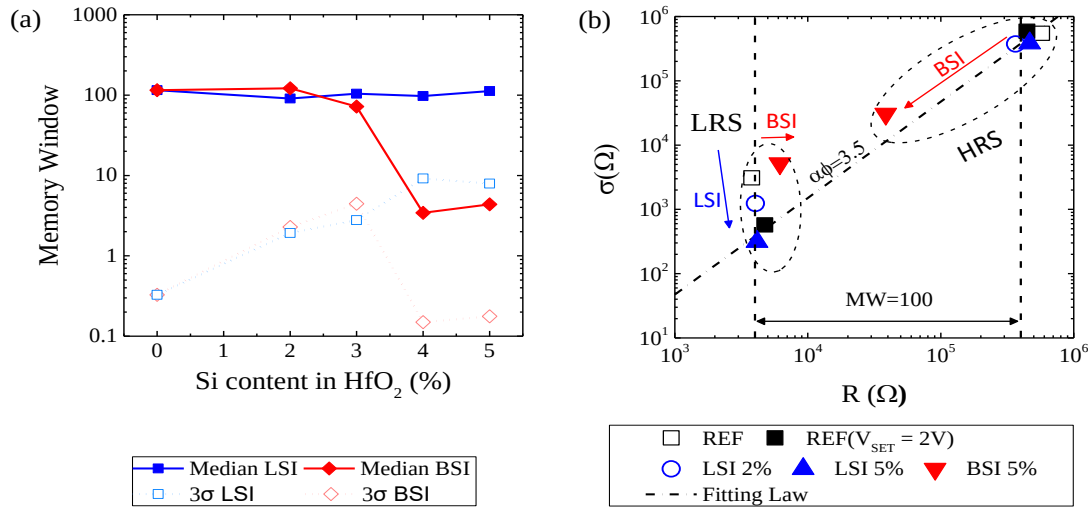


Figure 3-20 (a)-(d) (a) MW median, and 3σ evolution as a function of Si fraction for LSI and BSI in a 4 kbit array. (b) σ/median evolution of HRS & LRS using PC.1 in BSI and LSI devices.

area with increasing Si content in the LI devices and provides direct evidence of the reduction of the extrinsic component of device-to-device variability.

Figure 3-20b shows the relationship between median resistance and the standard deviation of LRS and HRS, confirming that the increase of Si doping in LSI devices allows to reduce LRS variability without impact on the variability of the HRS. To obtain a comparable LRS variability for the non-implanted reference, a higher set voltage is required. The relationship between median resistance and standard deviation follows the σ/R law (Eq. (3.15)) where G_0 the fundamental quantum conductance unit and αΦ are characteristics of the tunnel barrier; with $\alpha\Phi = 3.5$ corresponding to a non-degraded working condition [111]:

$$\sigma_R = \sqrt{G_0 \exp(-\alpha\Phi)}\, R^{3/2} \qquad (3.14)$$

To verify the most significant contribution in the variability of the devices in question the device-to-device and cycle-to-cycle variability of LRS and HRS were analyzed. 4 kbit arrays were cycled for 4k cycles, allowing a 1-1 correlation between the device population and the number of cycles. PC.1 with a SET voltage of 1.7 V was used for LSI devices while 2.0 V SET voltage for the reference stack (Table 3-1). To identify upper and lower

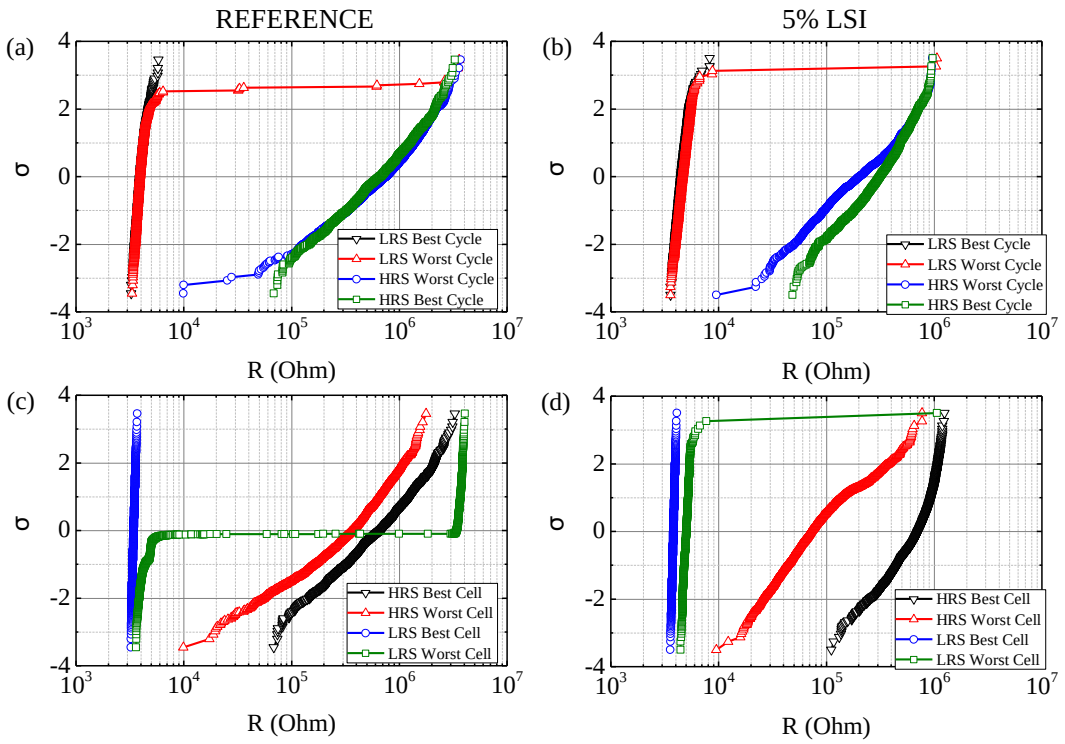


Figure 3-21 Cell-to-cell and (b) cycle-to-cycle variability of reference and 5% LSI 4 Kbit array over 4000 Cycles using PC.1 after annealing at 165°C for 1000 min.

boundaries the best (smallest distribution tail) and worst (largest distribution tail) cell-to-cell (Figure 3-21 a, b) and cycle-to-cycle (Figure 3-21 c, d) distributions measured during the test, are reported for the no-implant reference and the 5% LSI devices. In order to determine the best and worst LRS and HRS distributions the following method was used. The LRS evolution of each address for the 4096 cycles, and that of all the addresses in the matrix for each cycle (LRS & HRS) were grouped separately. Then, the arithmetic sum of each group is calculated serving as a figure of merit to determine the group of states with the best and worst cases. The best case LRS is considered as the one whose states exhibit the minimum sum while the best HRS state is considered as the one that maximizes the sum of resistances. The opposite is true for the worst-case evaluation. This way the best/worst LRS & HRS groups are evaluated.

When considering the cycle-to-cycle behavior of the cells in the tail of the LRS cell-to-cell distribution (Figure 3-21 c, d), it can be observed that even in the worst-case, the failure events are less than 1% for LSI (failures to set lie beyond the 2σ region). On the other hand, the erratic programming attempts of the worst cell for the no-implant HfO_2 fails for approximately 50% of the cycles in LRS. This suggests a distinct advantage of

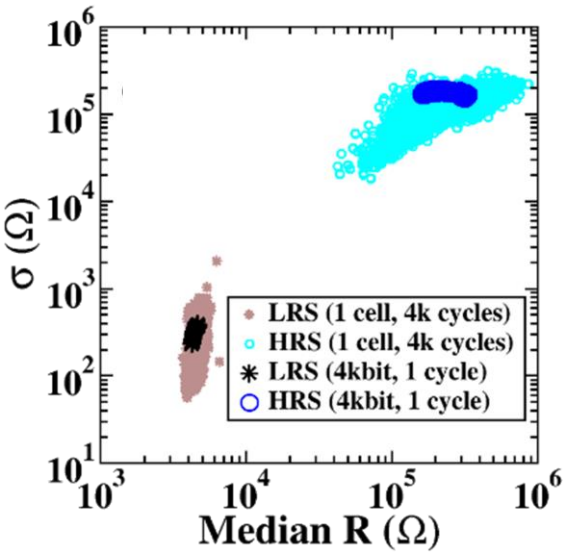


Figure 3-22 σ/R_{Median} dispersion of the 4 kbit array of 5% LSI devices over 4000 cycles.

the LSI approach towards reducing erratic cells in large OxRAM arrays at lower operating voltages.

The fact that HRS variability is not impacted can be easily understood if thought under the scope of the intrinsic and extrinsic variability components. Evidently, LRS variability is dominated by the extrinsic component and is sensitive to process maturity, circuit design and array topology etc. On the other hand, HRS variability is governed by the intrinsic characteristics of the variability of the filament dissolution. The characteristics dimension of this process is in the deep atomic scale and depends on the local potential of the oxygen atoms close to the bottom electrode of the ReRAM cell and can potentially be linked to bottom electrode granularity, local density variations of the oxide or bulk defects which are hard to control from a process point of view.

In Figure 3-23 the σ/median R relationship evolution extracted from the 4 Kbit LRS and HRS distributions is reported, showing that the variability for both cell-to-cell and cycle-to-cycle can be reliably controlled even for the erratic cells in the tails of the distributions.

3.7 Memory Array programming optimization strategy

3.7.1 Introduction

Up to this moment, the forming and switching properties of Si implanted HfO_2/Ti based ReRAM arrays have been examined (§3.6). The programming conditions used, correspond to ones that are identical or very close to the ones that are optimal for the no-implant reference operation in terms of LRS/HRS state spread (σ/median ratio) and long-term cycling stability. Such optimization has been studied extensively [111] for this material stuck and identical technological parameters on the array level and is used as reference (PC.1 of Table 3-1 or Condition 0 of Table 3-2).

For such high current compliance current conditions, the LSI samples studied show a significant reduction of LRS variability both device-to-device and cycle-to-cycle as was shown in §3.5. From the analysis so far, the conditions used on the array level are kept constant and closer to the optimal ones for the reference stack in order to compare the impact of the technology and local doping. These operating conditions do not however necessarily guarantee the optimal array statistical properties nor the long-term reliability of the LSI memory arrays.

The question of operating condition optimization is a multi-parameter problem and can also depend on the application. Mainly, programming conditions can be optimized for separation of the LRS and HRS states, data retention or maximum endurance. The most common tradeoff in terms of programming lies in the separation of the memory states and data retention versus the maximum number of cycles the device can sustain. Higher programming currents during SET lead to a more compact and stable LRS state at the expense of device lifetime. Lower programming currents lead to a degradation of the distance between states but allow to increase the maximum number of cycles in the devices [, , , ,]. Furthermore, overdriving reset voltage parameters also leads to accelerated ageing of the ReRAM cell [].

Finally, when large device populations are considered, such as in the case of memory chips, the optimal parameters for chip operation can differ significantly with respect to the ones extracted using a small ensemble (<100). This can be attributed primarily to the augmentation of the statistical ensemble allowing for a larger range of extreme events to appear. Chip topological effects are not statistically significant on a global scale and are

not discussed in detail for the sake of brevity (Moran's geo-statistics gives p-values > 0.05, not shown here).

3.7.2 Programming Conditions optimization on LSI devices: Qualitative Characteristics

In a first order, optimal programming conditions are such that allow for a complete separation between LRS and HRS. Secondly, they should also stabilize the memory states in the long-term, as the device is cycled. Finally, in the ideal scenario the state should be retained for as long as possible. Optimally, under a given set of conditions LRS and HRS overlap should be minimized. Furthermore, it is important to note that given the memristive [] nature of ReRAM, SET and RESET conditions need to be optimized in tandem as will be shown in the following. The full range of SET and RESET condition couplets explored is shown in Table 3-2. P.C. 0 is the reference optimal condition for the standard HfO_2 stack in this technology. P.C. 1, 2 explore the impact of reducing the maximum current during the SET operation, 3 the impact of top electrode voltage under low current compliance SET while 4-8 the impact of the bottom electrode voltage during the RESET operation.

Table 3-2 DoE list of conditions exploring the SET/RESET programming range of LSI 5% samples on the array level.

Programming Condition (P.C.)	SET			RESET				t_{pulse} (ns)
	V_{BL} (V)	V_{SL} (V)	I_{cc} (µA)	V_{BL} (V)	V_{SL} (V)	I_{cc} *(mA)*	V_{WL} *(V)*	
0	2.0	0	240	0	2.5		3.0	
1	1.7	0	240	0	2.5		3.0	
2	1.7	0	130	0	2.5		3.0	
3	2.0	0	130	0	2.5		3.0	
4	1.7	0	240	0	2.5	> 1	2.5	100
5	1.7	0	130	0	2.5		2.5	
6	2.0	0	130	0	2.5		2.5	
7	2.0	0	240	0	2.5		2.5	
8	1.7	0	240	0	2.5		2.0	

Specifically during the RESET phase, the critical parameter determining the efficacy of the reset pulse is the word line voltage, which directly affects the source-drain series resistance of the selection transistor hence directly controlling the bottom electrode voltage of the floating bottom electrode of the ReRAM element. P.Cs. 1 through 8 are applied on 512 bits, each, on a 4 Kbit array evenly distributed across the memory. The cells of each of the 8 groups are then programmed across 10^5 cycles (SET/RESET operations) and the statistical behavior of each population is studied. This population size allows to observe tail bits up to ±2σ especially in the LRS while eliminating the die-to-die variability component of extrinsic variability as well as maintain sustainable testing times. The behavior of the devices is traced right after the forming operation.

In Figure 3-23 the behavior of each population is summarized. In each cycle, the normal probit is calculated for each state and the evolution of the median resistance, as well as the resistance values within the range of ±σ and ±2σ of each state is shown as a function of the programming cycles. (P.C. 1) imposes current compliance of 240 µA. Such condition quickly leads to a strong compacting of the LRS distribution across the range of cycles tested in accordance with the study presented in the previous sections. Nevertheless, the "hump" appearing in the HRS is a direct quantitative signature of accelerated ageing and is expected to quickly lead to breakdowns. Reducing the current compliance to 110 µA as well as the programming voltage of the top electrode (P.C. 2) leads to poor control of the LRS upper tail bits. Interestingly, a small increase of the top electrode voltage efficiently controls the LRS tail even at a low current compliance (P.C. 3).

Interestingly, a poor / incomplete reset leads to a poor set state as can be seen by the failure to control the LRS tail bits of P.Cs. 4, 5 under the same reset condition, irrespective of the current compliance and requires to further augment the top electrode voltage to compact the LRS distribution anew (P.C. 6, 7). The limiting case of insufficient reset voltage pulse is shown in 8 where the medians merge to the LRS state.

The current compliance directly impacts the median value and is as a result primarily impacted by the physics of the SET mechanism, effectively directly controlling the local filament temperature that facilitates ion migration. On the other hand, the top electrode voltage has direct impact on the statistical limits. This signifies that the set voltage values required on the array level are higher than the intrinsic set requirements of the cell imposed by the physics.

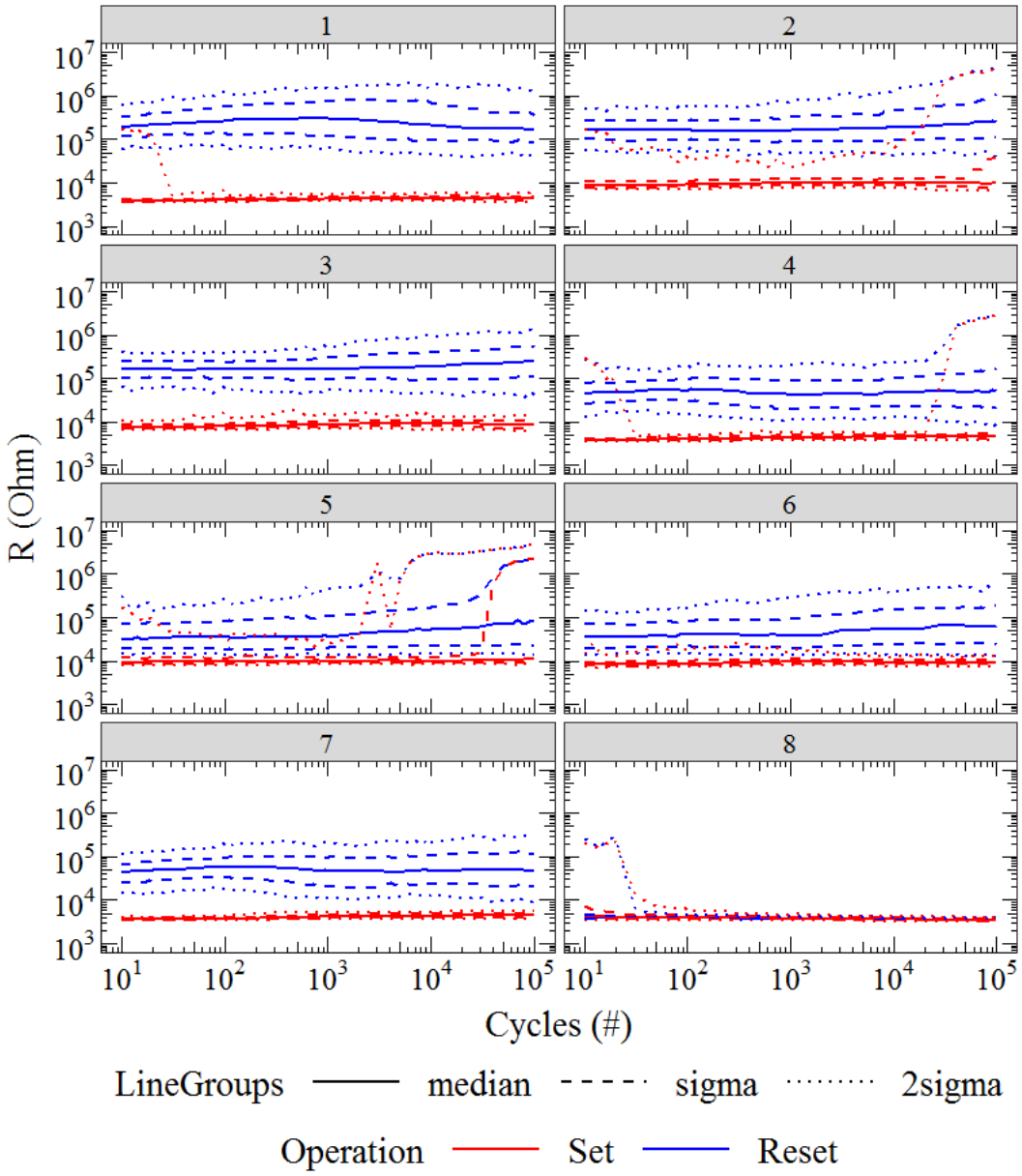


Figure 3-23 1-8: Impact of SET/RESET condition couplets of Table 3-2 on the median HRS and LRS states of 512 devices / group and within ±σ and ±2σ confidence intervals.

The nominal applied set voltage is likely attenuated by the RC parasitics of the array interconnect and varies wrt the one that is applied on the array address.

3.7.3 Programming Conditions optimization on LSI devices: Quantitative Characteristics

From the graphical analysis presented below, conditions 1, 3 and 7 lead to a stable memory window with more than 95% confidence assuming the log-normal distribution is followed. Nevertheless, this approach is sensitive to the extent that the population follows the pre assumed distribution. As the interest falls in the control of erratic tail bits, the higher the confidence range required, the less reliable the graphical approach is as tail bits usually do not follow the statistical behavior of the main population […]. As a result, an approach other than statistical ranking is required in order to rigorously assess the behavior of erratic bits. For this reason, we choose to define the Bit Error Rate (BER) at a given cycle and condition device-to-device as the maximum between the total sum[3] of LRS (R_{LRS}) cells found above or HRS cells (R_{HRS}) found below a resistance threshold R_{th}, over the total number of bits N_{bits}.

$$\mathrm{BER} = \frac{1}{\mathrm{N_{bits}}} \max\left\{ \sum_{\mathrm{BL,WL}} [\mathrm{R_{LRS}} > \mathrm{R_{th}}], \sum_{\mathrm{BL,WL}} [\mathrm{R_{HRS}} < \mathrm{R_{th}}] \right\} \qquad (3.15)$$

The notion of the assessment against a threshold comes from the principle of operation the electronics of a memory chip. As the memory resistance is essentially an analog value it needs to be translated into a digital 0 or 1 (considering 1 bit per cell) through an ADC by comparing the output to a reference threshold value.

The total BER is considered to be accurately approximated by the maximum contribution between the LRS and HRS states at a given assessment attempt. Hence, the optimal condition should be one that minimizes the BER. Nevertheless, this also implies that a given BER is sensitive to the choice of threshold. As a result, the R_{th} is optimized such that it minimizes the BER across the total number of cycles programmed, and is summarized in condition (3.16) below.

$$\mathrm{R_{th}^{opt}}: \mathrm{BER}\left(\mathrm{R_{th}^{opt}}\right) = \min \mathrm{BER}(\mathrm{R_{th}}) \; \forall \; \mathrm{R_{th}} \; \in \Omega \, , \Omega = \left[\mathrm{R_{min}^{LRS}} \, , \mathrm{R_{max}^{HRS}}\right] \qquad (3.16)$$

[3] The conditional sum has been expressed using the Iverson Notation.

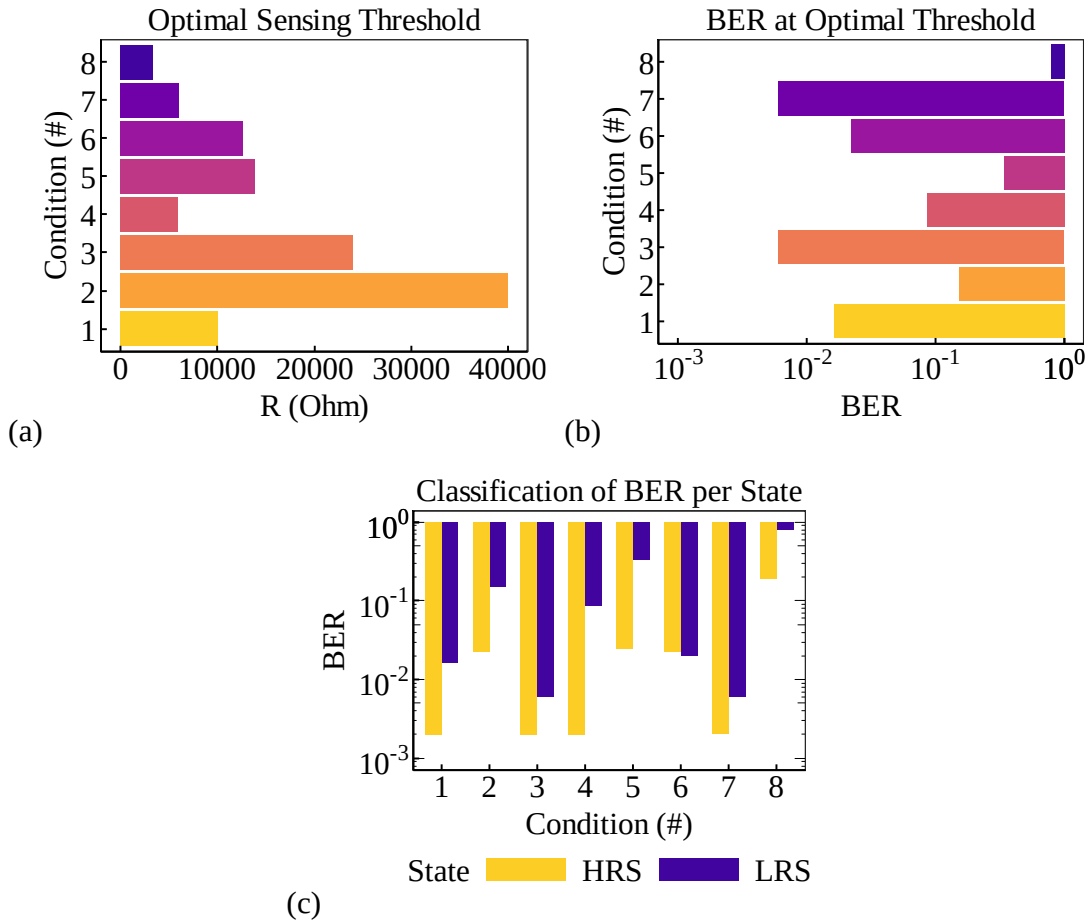


Figure 3-24 (a) Optimal figures of merit calculated for each programming condition mimimizing Eq.{(3.15),(3.16)} (a) Sense threshold (b) Total Bit Error Rate (BER), (c) Bit error rate of each of LRS and HRS per programming condition.

Equations (2.4) and condition (3.16) constitute a constrained optimization problem that can be uniquely determined by a simple numerical algorithm.

In Figure 3-24a, the optimal calculated threshold is extracted for each programming condition after 10^5 cycles. Programming conditions with strong programming currents result in an optimal threshold of very low resistances. Interestingly, even though reducing compliance current during set increases the resistance of the LRS state, a low BER can still be obtained, with the difference being that the assessment threshold is increased from 6.09 KOhm to 25.96 KOhm as can be seen by directly comparing conditions 3 and 7. This suggests that even though erratic tails bits of the LRS are not entirely controllable at low current compliance conditions (~ 110 μA), in the sense that a bimodal distribution can still emerge in the LRS, they are not necessarily prohibitive for ReRAM operation as long as reset failures do not become significant. Moreover, using a softer reset condition results in

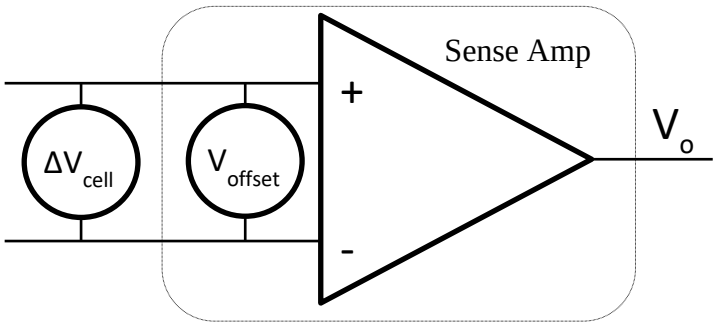


Figure 3-25 Schematic representation of a differential sense amplifier. Cell input is abstracted as a virtual voltage source ΔV_{cell} while the internal op amp noise is abstracted as V_{offset}

a lower BER for the high compliance current case (Conditions 1 &7). This can be attributed to accelerated ageing of the devices resulting in reset failures.

If errors are separately classified to their source of either the HRS or LRS state it is evident that the most significant contribution to the BER comes from the degradation of the low resistive state. Nevertheless, overdriving reset also acts to degenerate the lower tails of HRS. Equivalently this can be also seen as a gradual two-way degeneration of the memory window as a result of oxide ageing. Even though conditions 3 and 7 give similar metrics in terms of BER (Figure 3-24b), a low current compliance (condition 3) set presents two distinct advantages: i) reduces power requirements during set by 54% and ii) allows to decelerate the ageing process of the devices. Finally, if the BER contribution of each state is seen separately it is immediately seen that the strongest contribution to errors comes from set failures during the SET operation (Figure 3-24c).

The optimization of the sense threshold on which the assessment of the BER has been based should be stressed. In the previous study, the increased variability of the memory window, even after 1 x 10^5 cycles can be mitigated by shifting the assessment threshold to higher values as long as the HRS tail is not strongly overlapping with LRS. The fact that an optimal value for sensing appear in the range of 25 KOhm instead of 8 KOhm has very important advantages on the design of the sense circuit periphery.

For memory chips, sensing is based on differential operation amplifiers using the voltage drop across the cell terminals as the differential input. The differential read is the translated into an output voltage that is propagated into the ADC. Even in the most optimized op amps the minimal differential voltage resolution lies in the (2-4) mV range. The limit rises

from the internal noise of the op amp ultimately. If a voltage drop across the memory cell ΔV_{cell} is lower than the voltage noise floor V_{offset} the signal can no longer be resolved (An abstracted representation is given in Figure 3-25). The noise limit is ultimately limited by internal design and technology parameters of the sense amplifier such as transistor mismatch, transconductance noise etc.

As a consequence, the resistance threshold is directly related to the critical output voltage of the sense amp. As a result, a larger threshold translates directly to a more relaxed resolution requirement between LRS and HRS, as a larger resistance will cause a more pronounced voltage drop that allows to rise well above the sense amp noise margin. Furthermore, as the amplifier gain can be relaxed owed to the stronger threshold signal, a smaller W/L ratio can also be used leading to a reduction of the sense periphery surface. The tradeoff being of course a loss of speed in the time resolution, but at this range of values the augmentation of the signal integration time requirements is expected to be negligible.

3.8 LSI Reliability: Endurance and Error Analysis

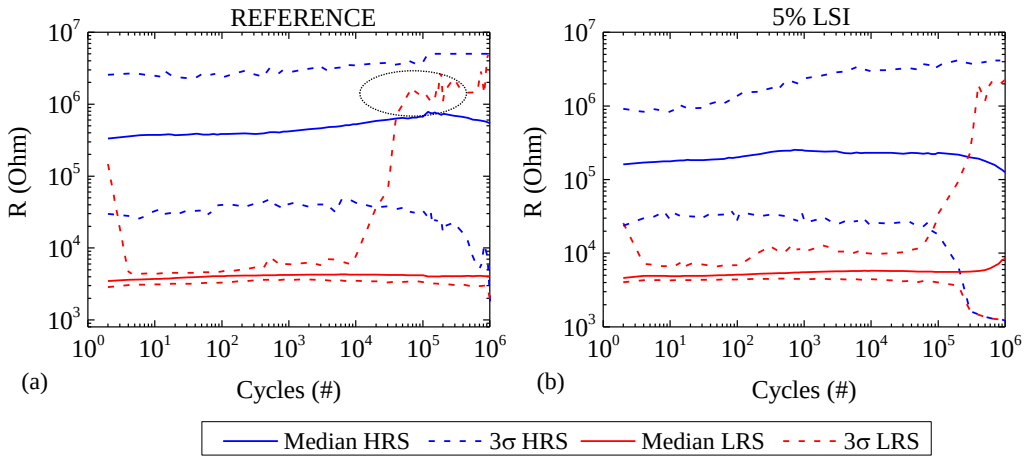


Figure 3-26 Median and 3σ evolution of LRS and HRS states for a 4 Kbit matrix for (a) reference HfO_2 using (P.C. 0) (b) 5% LSI using (P.C. 3).

Using the previous methodology the ageing of 4 Kbit arrays for the limiting cases of the reference and 5% LSI samples is studied. This allows to expand the resolution in the range of 3σ or 2.5 x 10^{-4} for the BER. Devices are cycled right after being formed using suitable forming conditions for each split. Cycling is done at the optimum condition for each of the

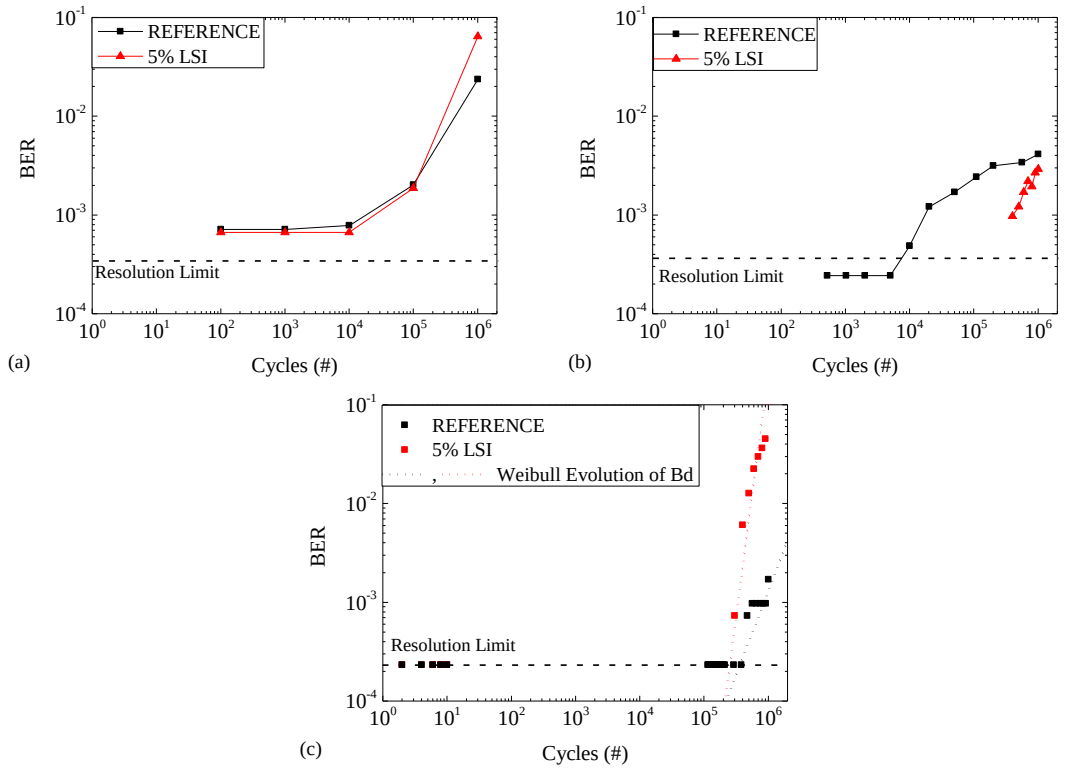


Figure 3-27 (a) Total corrigible BER (b) BER of permanent LRS-write failures (c) oxide breakdown related errors for a 4 Kbit array of reference HfO_2 and 5% LSI devices.

variants, notably (P.C. 0) for undoped HfO_2 and (P.C. 3) for 5% LSI of Table 3-2 as previously determined. In Figure 3-26 we present the behavior of the median resistance as well as the resistance at $\pm 3\sigma$ of LRS and HRS for the arrays tested across 1 x 10^6 cycles.

In both cases, an early-lifetime material wakeup phenomenon is observed as can be seen by the overlap of the upper tail of LRS and lower tail of HRS states at 3σ which quickly disappears after the first 4-7 SET/RESET cycles. The reference sample presents with a compact LRS state owed to the higher programming current (250 µA). Nevertheless erratic LRS bits start to appear as early as 10.000 cycles. On the other hand, similar behavior is exhibited in the LRS of LSI devices only after 1 x 10^5 cycles. On the other hand, the lower tail of the HRS of the reference devices crosses the 25 KOhm assessment threshold past the 1.8 x 10^5 cycles, whereas in LSI devices this limit is reduced to ~ 1 x 10^5 cycles. Moreover, observing the median behavior of HRS and LRS the reference devices experience less ageing overall as compared to the LSI devices. There is as a result, a tradeoff between the gains in the control of the LRS erratic bits versus the accelerated ageing leading to irreversible oxide breakdown in the LSI technology as compared to the reference stack.

If the tail bit population in the LRS of each technology is studied, it can be shown that early-lifetime LRS failures (transition from HRS to LRS) can be classified in three populations according to their behavior with respect to the programming attempts: i) random corrigible errors ii) permanent LRS -write-failures and iii) oxide breakdown errors. Out of the three, only type (i) errors can be corrected by reprogramming the cell or using an ECC approach []. Such errors originate from random temporal and/or spatial fluctuations of the programming pulses across the metal lines of the array or by the intrinsically stochastic nature of the filamentary mechanism. Both LRS-write-failures and oxide breakdown are catastrophic and lead to dead cells in the array.

This classification results from observing the capacity of cells considered erroneous (either found above the sense threshold value after a SET operation or below the sense threshold after a reset operation) to respond to a reprogram algorithm [patent submitted by Alessandro, cite]. Cells that are irresponsive to a given number of reprogramming attempts are classified as either LRS-write-failures if found above the threshold or as oxide breakdown errors if stuck below the assessment threshold.

In Figure 3-27a, the total BER for erratic yet corrigible erroneous bits is presented for the two splits examined, each cycled in its corresponding optimal conditions. Interestingly, such errors behave in a similar fashion in both technologies. However, LRS-write-failures start to appear in HfO_2 as early as past 10.000 cycles. In the LSI devices, this type of error appears after 1×10^5 cycles (Figure 3-26b). The tradeoff to this improvement is that, oxide breakdown is accelerated in LSI devices starting at $3x10^5$ cycles (Figure 3-27c). It seems that in both cases ReRAM operation is hindered by either type of non-correctible errors in the end, either of the 1st or 3rd type. Non-correctible errors are particularly detrimental to memory arrays, especially NAND type arrays that are the standard for high-density storage class memory, an incorrigibly erroneous cell will deactivate the entire word line it belongs to.

3.9 Data Retention assessment on large populations

3.9.1 Introduction

As in every memory technology, one of the most important metrics to be identified is the ability to retain the program data for a given amount of time. Depending on the application, different constrictions apply for assessing data retention. In standard FLASH technologies

several models have been proposed to model the destabilization of the state of memory cells given different mechanisms [148]. Nevertheless, measuring the actual failure time of the memory states is not possible owed to practical limitations: At ambient or typical operation temperatures (< 85°C) such failure times for either LRS or HRS will be in the order of months or years. Hence, an accelerated testing procedure is required to shorten testing times while allowing extrapolating the failure time at lower temperature with a fair amount of accuracy.

When modeling data retention, once the physics of the failure mechanism is determined, the statistical failures of drifting bits in time can be modeled by associating the mechanism to a given statistical distribution. Annealing the samples to higher temperatures is then used to accelerate drifting bits and shorten testing times. Nevertheless, in the case of ReRAM technology, as the physical mechanisms dominating ion movement and redistribution in time have not been completely understood to date, a complete modeling of the distribution, involving the tail bits has not yet been demonstrated even though several models have been proposed. Such approaches, involve extracting an activation energy for the drift of bits, after observing the time evolution of the CDF in LRS and HRS. Such activation energies, given their statistical origins should not be seen as strictly representative of the physics of the device as they encompass both intrinsic and extrinsic mechanisms; their values can greatly vary according to the extraction criterion used, as will be shown in the following.

In this paragraph, we start by demonstrating the qualitative drift of bits with time and temperature, and make first comparisons by direct bit count. Then, an effort is made to extrapolate time-to-fail values from thermally accelerated tests done on HfO_2 and LSI devices for both the LRS and HRS.

3.9.2 Design of Experiment for Data Retention

4 kbit arrays are set in LRS and in HRS using programming conditions (PC 0, 3 Table 3-2) for HfO_2 and (PC 3, Table 3-2) for 3% and 5% LSI samples. The choice of samples is such that allows to explore the corners of and intermediate point of the LSI technology in comparison to standard HfO_2. The two different programming conditions in HfO_2 target to understand the impact of the maximum programming current in the HfO_2 no-implant reference. LSI devices are assessed using (PC.3) (low current compliance) since this

optimizes both their lifetime and error rates. The corresponding optimal condition for the reference stack is (PC.0).

In the absence of a program-and-verify capability in the test equipment, we are forced to subtract erratically programmed cells from the array both for LRS and HRS during data treatment. For this reason, the choice of the sense threshold as this is chosen is particularly important. In the following we examine two different cases for the technologies tested: Error assessment at the optimal threshold for error optimization as determined in the previous section set in 26 KOhm. For every dataset, erratic cells found above the threshold in LRS or below the threshold in HRS right after programming (time zero) are removed from the statistics. The time evolution of the remaining population is then analysed. In all cases, less than 1.5% of the cells of the 4 kbit are excluded in the worst case leading to yields in the order of (98-99) %.

Post programming the devices are subjected to thermal annealing for one day. The state of the devices is measured after 10 min, 30 min, 120 min and 1360 min of accumulated thermal stress. The statistical evolution of the states is traced for each case.

3.9.3 Data Retention at High Current Compliance: Qualitative Characteristics

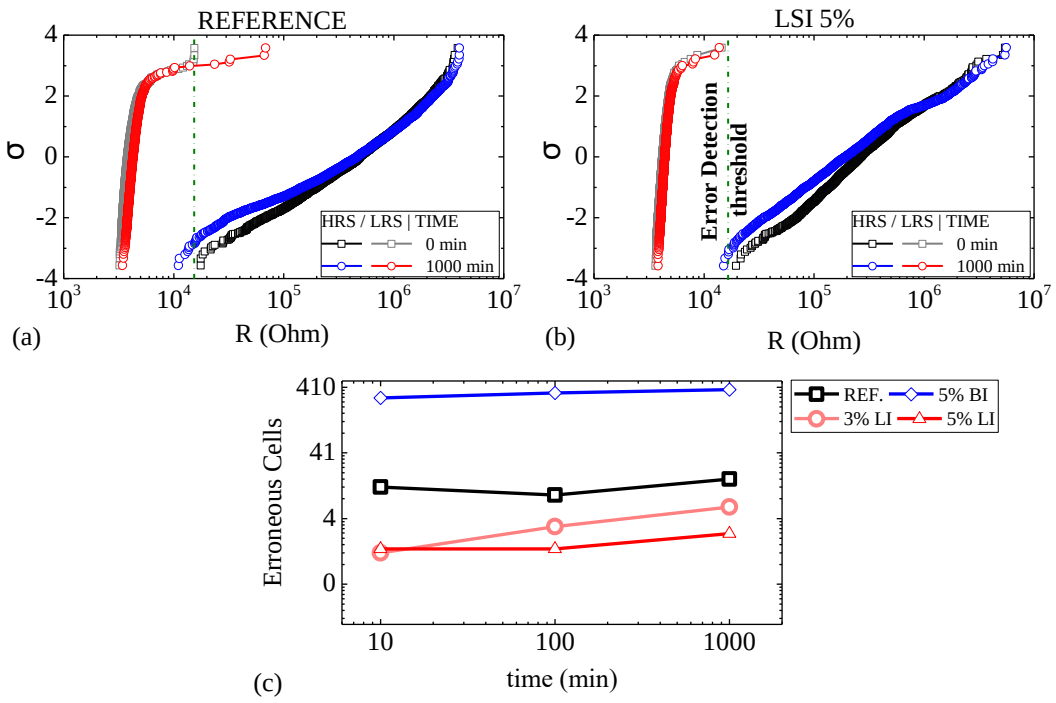


Figure 3-28 LRS and HRS state stability for (a) reference and (b) 5% LSI 4 Kbit array programmed using (P.C. 0) for 1000 min stress at 165°C. (b) BER evolution from early lifetime to 1000 min at 165°C.

In the first comparison between implanted devices and the no-implant reference, high current compliance is used (PC 0, Table 3-2). In this first assessment the evaluation of the state stability is done by direct count of erroneous cells versus the most extreme device in LRS upper tail (Figure 3-28 a, b) after subjecting the devices in 1000 min of stress at 165ºC in total. In this image, the LSI LRS is more stable with respect to the reference HfO_2 as it minimizes the number of erroneous bits (reflected in Figure 3-28c). Interestingly, the detected failures in states between BSI, LSI and no-implant reference devices appear to be detectable in the early stages of accelerated stress. Nevertheless, this assessment is not suffice to project the lifetime of the states. Moreover, in both cases HRS bits drift towards LRS.

To understand the behaviour of data retention, especially between different technology flavours it is imperative to i) understand the impact of the programming condition and ii) evaluate erroneous populations at a certain threshold, at least as far as the assessed "error bits" are concerned. The choice of threshold is also important since from the analysis of

§3.7.3 it is evident that in a realistic memory array the choice of threshold will strongly impact the assessment of erroneous bits.

3.9.4 Impact of the programming current and Si implantation on the BER of 4 kbit arrays

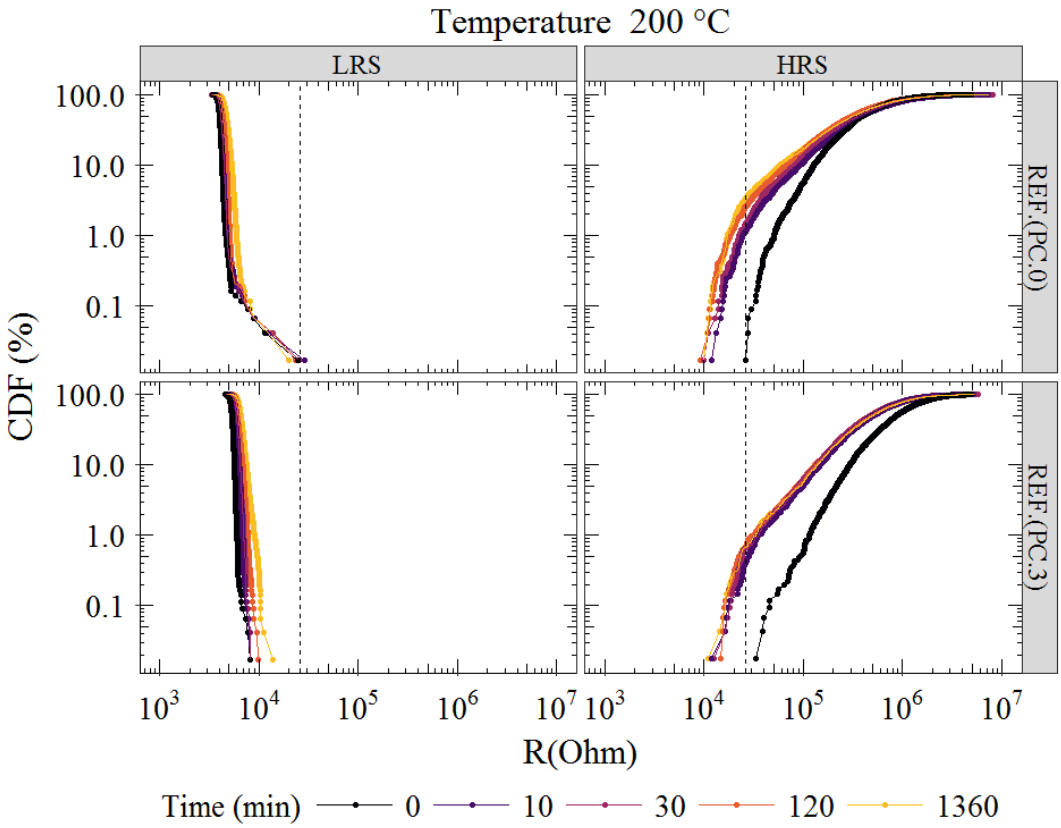


Figure 3-29 CDF (HRS) and CCDF (LRS) evolution with thermal stress (200°C) for an HfO_2 matrix programmed using high (PC.0) and low (PC.3) current compliance.

The evolution of the 4 Kbit array distributions is shown in Figure 3-29 for both HRS & LRS for a total of 24h of thermal stress at 200°C. First, the impact of the programming current on the faulty bits evolution is studied. Reference HfO_2 4 Kbit arrays have been programmed to the LRS and HRS respectively, and the temperature-accelerated time evolution of the BER was assessed for each condition (Table 3-2). Interestingly,

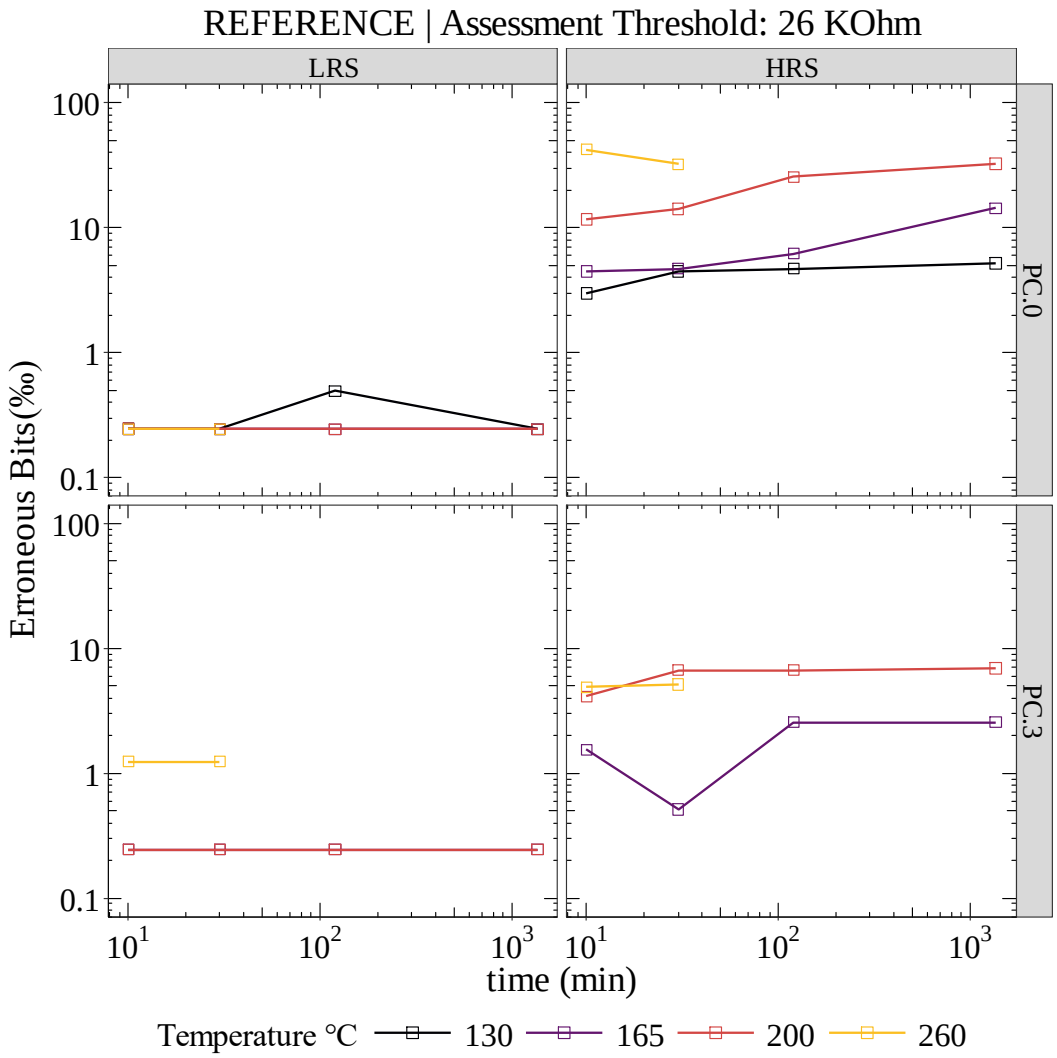


Figure 3-30 Evolution in time of the counted erroneous bits at LRS and HRS for high programming current (~250%A, PC.0) and low programming current (~ 130 µA, PC.3). The reset conditions are maintained identical in the two cases for reference HfO_2 samples.

programming at a higher current (PC.0) during the SET operation although stabilizes the LRS tail, even at 260°C (compatible to solder-reflow processes) seems to increase failures in the HRS by more than one order of magnitude.

This behaviour has previously been reported on data retention measurements of ReRAM devices on unitary cells [Paper Gab, Marinella Gabri] and has important implications in memory operation: As HfO_2 based cells require strong programming currents (> 150µA) on the array level, suggests that the HRS drift can have more significant contributions in the erroneous bits of the array as compared to the ones attributed to the LRS drift.

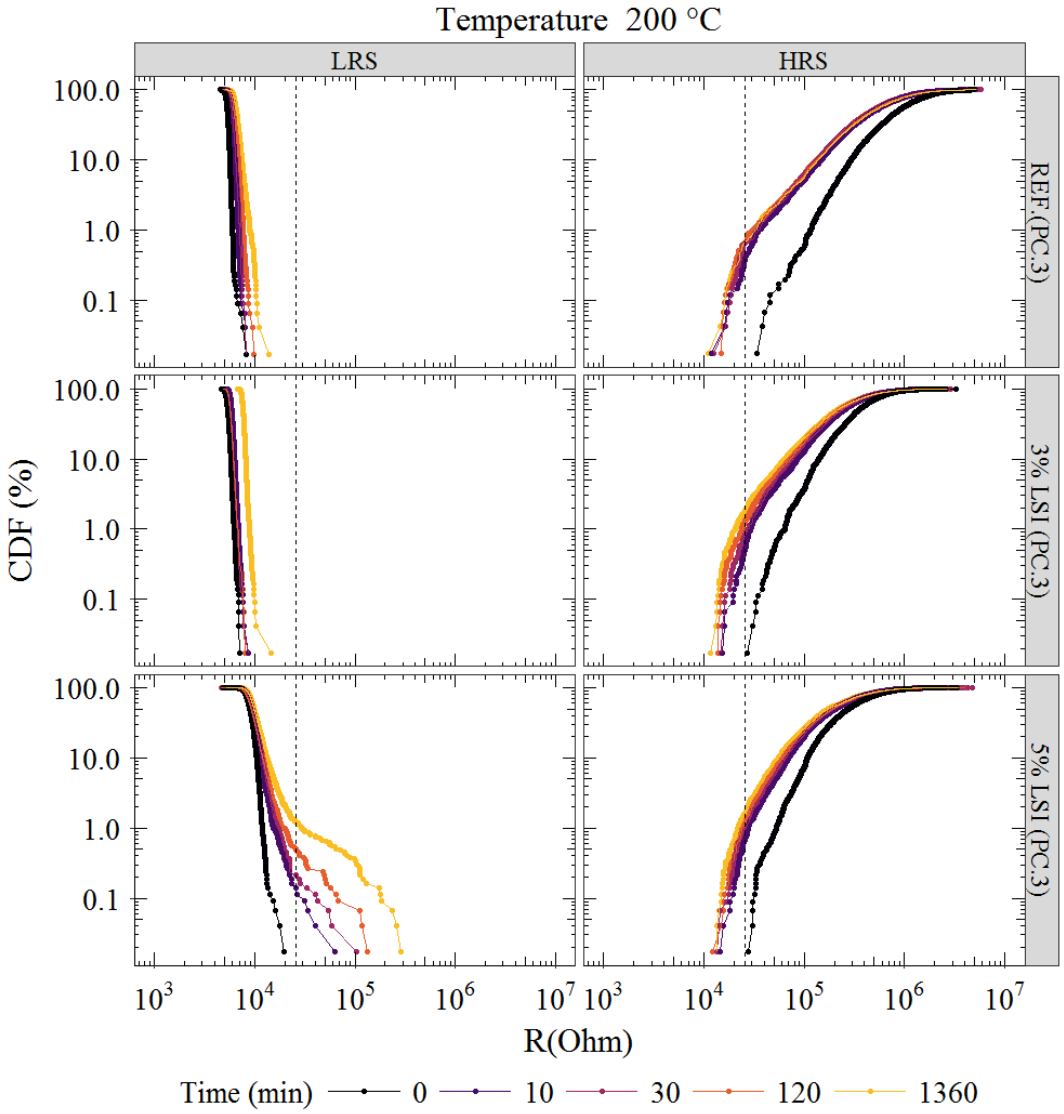


Figure 3-31 Comparison of reference, 3% and 5% evolution of states with time at 200°C thermal stress

Secondly, we examine the data retention of the different technology variants at 140 μA current compliance (PC.3) determined to be the optimum for LSI devices. Similarly, the CDF of the LSI devices is shown, in comparison to the reference in Figure 3-31. Generally, less than 5% of devices drift significantly across 24h thermal stress. In this case, the 5% LSI devices used in this study perform worse at the tail as compared to their 3% counterparts. In all cases, device medians are drifting extremely slowly in time as will be shown in the following.

As in the previous case, the drift of the CDF tails for each state and device type is first assesed by direct count of the erroneous bits (Figure 3-32). For all temperatures

considered, LRS extreme failures are insignificant. In all cases considered less than 10 ‰ failures can be assessed while < 25 ‰ failures in HRS can be counted even after 30 min of stress at 260°C. Moreover, it is interesting to note the weak evolution of the erroneous bit count with time, suggesting that the erratic bits determined here are early-fails and count be related to extrinsic effects. This type of assessment is of course sensitive to the choice of threshold and its distance to the most erratic bit in the array after programming.

Complete understanding of the failure mechanisms would require the capacity to model the CDF and its evolution with time and temperature. This is nevertheless not practical as the distributions are bimodal and the distribution of the tail is difficult to determine. To reliably model the CDF tails, a larger failure population would be required.

Weibull analysis, where the weibit failures evolution with time is analyzed has been suggested [Perez, WODIM 2017]. This method is sensitive to i) the choice of threshold to determine the “failure bits” and ii) fitting is sensitive to the failure percentiles determined. In our case, considering the extremely low numner of failures observed in the systems studied, the method leads to poor quality of the fits.

As a result, in order to obtain more quantitative results, we make an effort to quantify the dynamic evolution of the resistances in time for the different temperature conditions. In Figure 3-33 the evolution of the resistance of the CDF in the median and 1% tail is presented for the two programmed states and all temperatures tested. First, it is immediately noticed that the median resistance evolution is extremely weak, with the bulk of the shift happening in under 10 min of thermally accelerated stress. The CDF tails are relatively easier to drift as compared to the medians; it is worth noting that this occurs for less than 120 devices at the worst case for the technology variants examined on the 4 Kbit array level. The weak evolution of the resistance, with drifts smaller than 800 Ohm over 24h at these percentiles makes it challenging to accurately determine an evolution law. Nevertheless, in the following an effort is made to determine a potential failure semi-empeirical relation in order to calculate expected failure times.

In the following, we choose to model the drift using a power relation of the resistance with time (Eq. (3.17)). This semi-empirical relation, although difficult to dirrectly be correlated with physics has been observed in sereval physical systems in hard matter as well as

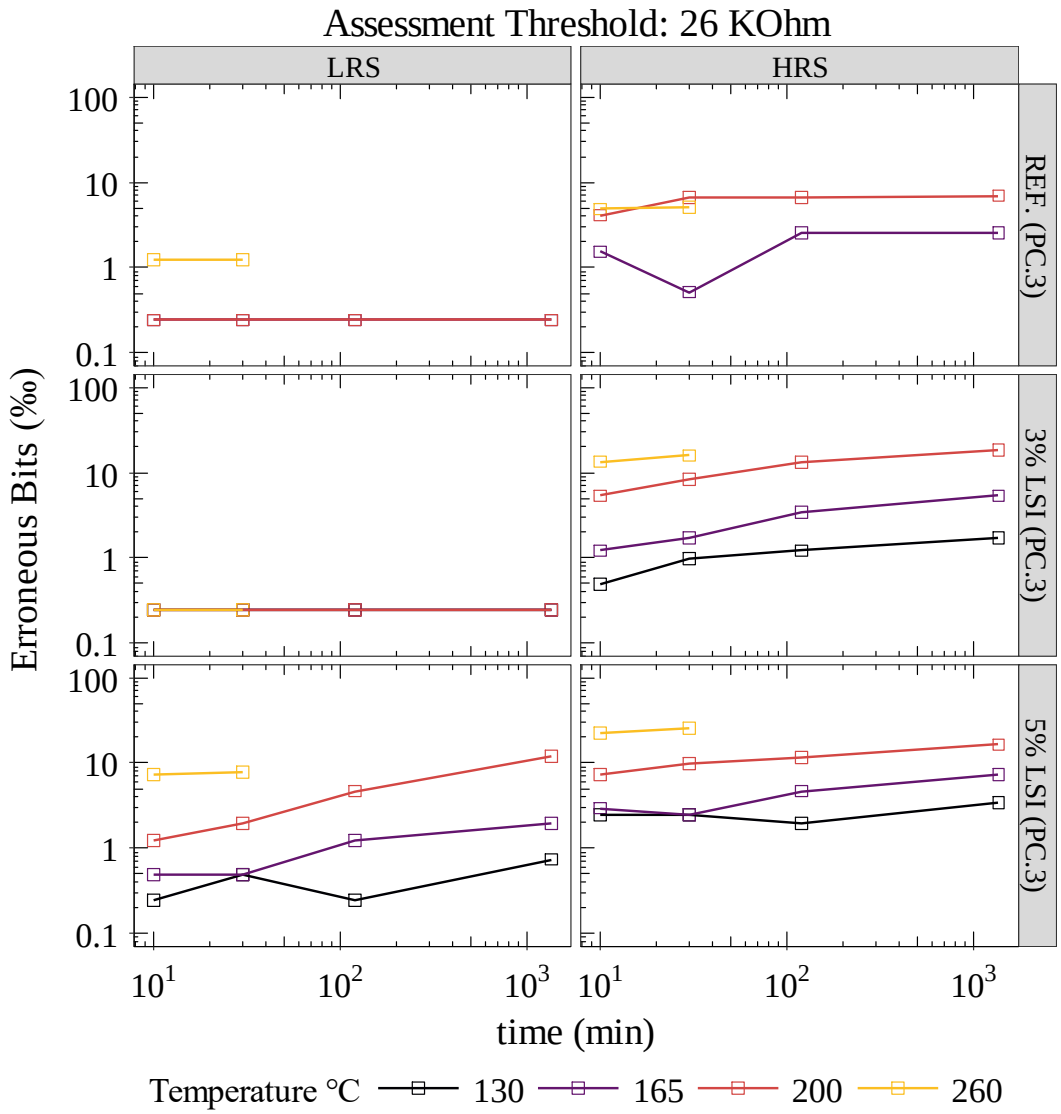


Figure 3-32 Erroneous bit count evolution with time and temperature for the reference, 3% and 5% LSI device variants at (PC.3)

biological samples [] and can be correlated to spontaneous emission processes such as continuous-time random walk [].

In our particular case, if an oxygen ion occupying a site is subject to random energy fluctuations owed to temperature over time, it could similarly exhibit random-walk like behaviour. This process can explain the strong initial relaxation of the resistance with subsequent weak perturbations in the long-term.

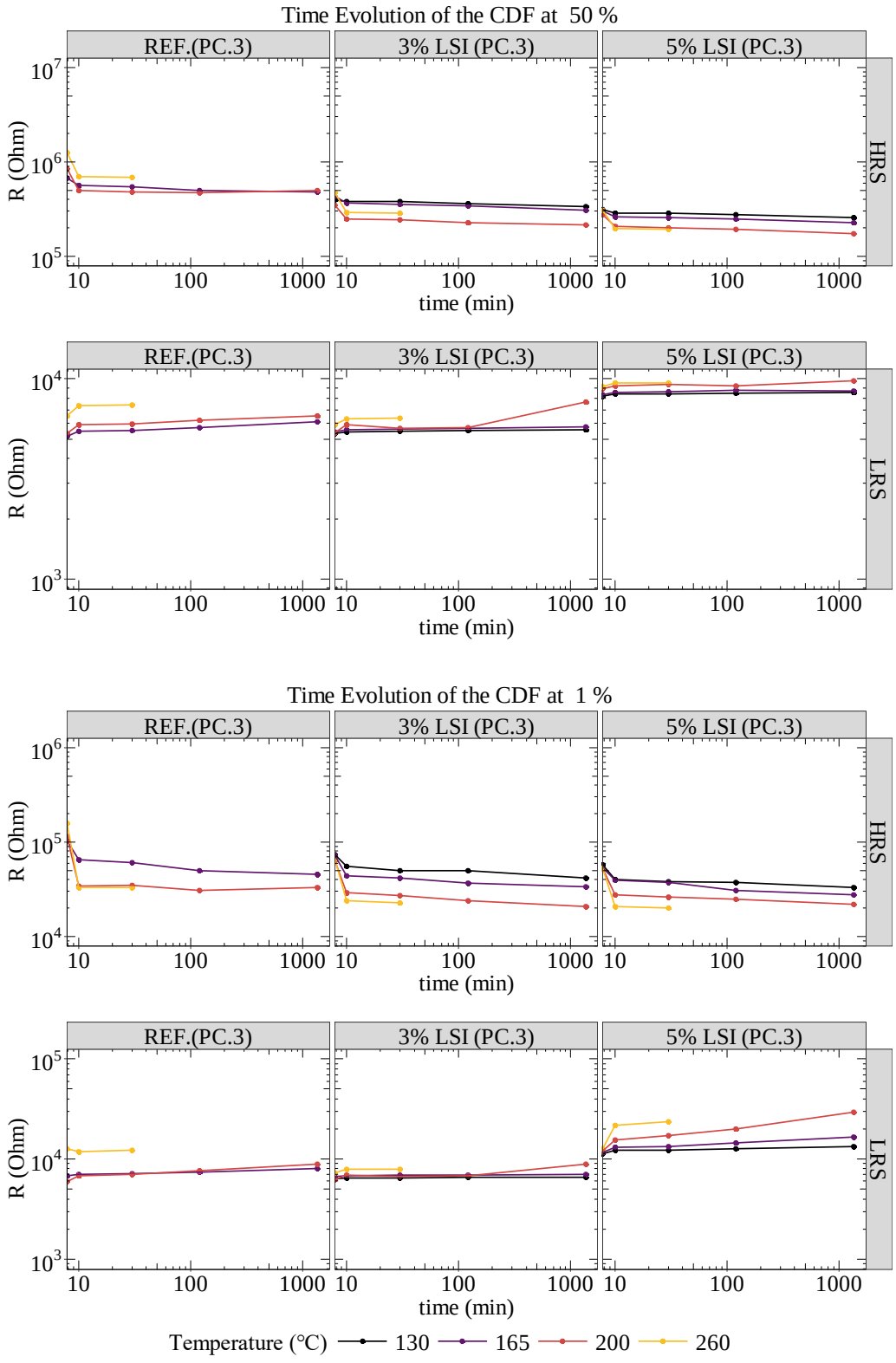


Figure 3-33 Time evolution of the median LRS and HRS resistance values (Top) and those at 1% tail of the HRS CDF or LRS CCDF, respectively (Bottom).

3.9.5 Empirical model for resistance drift

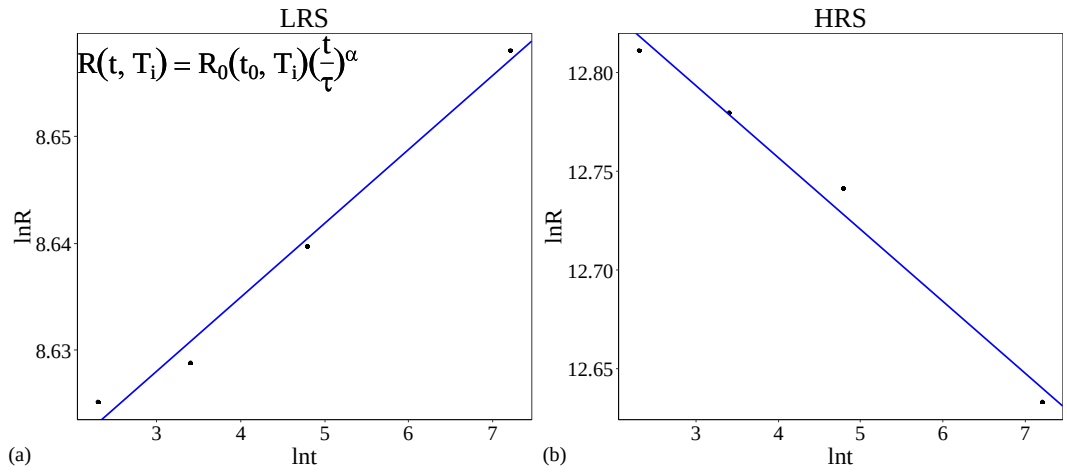


Figure 3-34 Power Law evolution of the resistance drift of the 1% percentile of the CCDF and CDF respectively, of a 4 Kbit array over 24h in 200°C for (a) LRS and (b) HRS.

$$R(t, T_i) = R(t_0, T_i) \cdot \left(\frac{t}{\tau}\right)^{\alpha} \tag{3.17}$$

$$\ln R = \alpha \ln t + (\ln R_0 - a \ln \tau_0) \tag{3.18}$$

In Figure 3-34 the evolution of the median resistance with time for both the LRS and HRS at 200°C over 24h is shown in a logarithmic scale. The proposed relation (Eq. 3.10) accurately follows the experimental data. The log R versus log time relation allows to determine the acceleration factor α of the resistance drift and the characteristic time τ. For the most part of the cases presented, fit quality is in the range of 0.95-0.99 for R^2 of the linear model fitted to the linearized equation (3.18). Using the experimental value of resistance R_0 after programming, the exponential factor α and characteristic time τ can be determined for each case as the model fitting parameters. Given these parameters the time evolution at a given temperature can be modeled solving (3.17) for time.

As in every assessment, a failure criterion needs to be defined. The time to fail is defined as the moment in time, when the resistance value is double its initial value for LRS and half its initial value for HRS, respectively (Eq. **Error! Reference source not found.**). T his definition, allows to observe the spontaneous thermal drift of the resistance without being sensitive to a subjective assessment threshold.

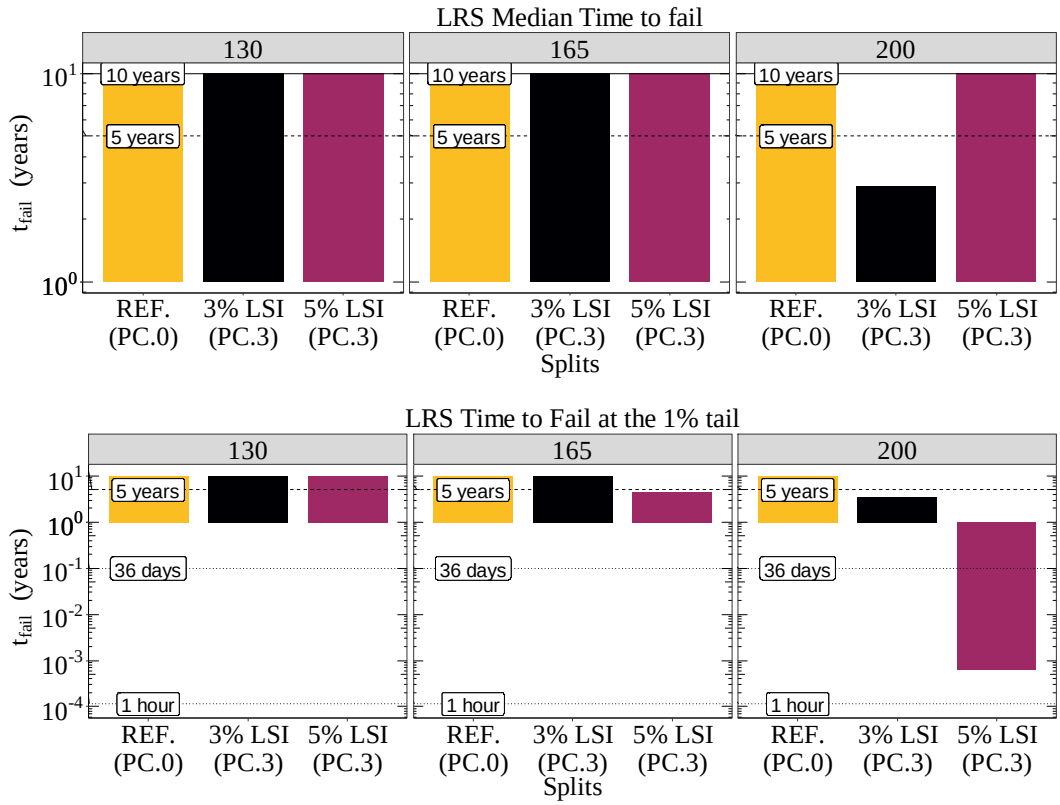


Figure 3-35 Calculated time to fail for a given temperature for 3 different samples of interest in the LRS at 130°C, 165°C and 200°C respectively. (Top) Median time to fail (bottom) time to fail of the 1% LRS tail.

$$t_{fail} = \tau\left[\frac{1}{\alpha}\ln\left(\frac{R_{Fail}}{R_0}\right)\right] \text{ where } \frac{R_{Fail}}{R_0} = \begin{cases} 2 \text{ for LRS} \\ \frac{1}{2} \text{ for HRS} \end{cases} \qquad (3.19)$$

The calculated failure times for the low resistive state at the experiment temperatures are shown in Figure 3-35. Considering the extremely weak resistance drift, the calculated time to fail is at times is unphysically long. For comparison purposes, such times have been scaled down to a maximum of 10 years. The continuous straight line denotes 10 years and the dotted line 5 years respectively. Overall, it is immediately seen that the LRS under operating programming conditions is extremely stable both in the median as well as in the tail. Only at 200°C is there any noticeable degradation for the resistance. The 5% LSI sample LRS tail seems to destabilize somewhat past the 165°C limit. HfO_2 LRS, even though programmed at high current and showing more failures in the bit count (Figure 3-32) is predicted to be more stable with time at higher temperatures. This suggests a competitive mechanism activating with temperature in Si implanted samples making itself especially apparent when LSI samples are programmed with lower currents.

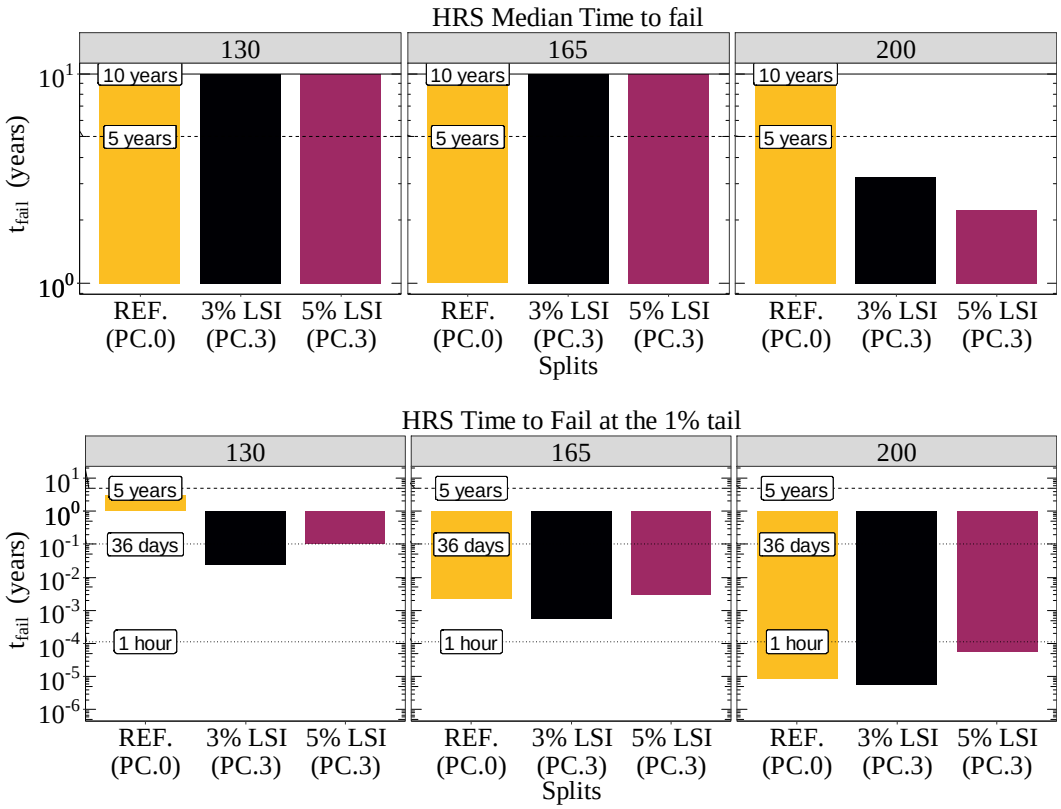


Figure 3-36 Calculated time to fail for a given temperature for 3 different samples of interest in the LRS at 130, 165 and 200C respectively. (Top) Median time to fail (Bottom) time to fail of the 1% HRS tail.

The corresponding image of the high resistive state is shown in Figure 3-36. It can be immediately seen that, even though the median HRS is very stable until the 200°C milestone LSI samples present with stronger drift as compared to the HfO_2 reference. More importantly, all samples show a very strong destabilization of the HRS tail with increasing temperate. According to the calculation, the HRS tail is projected to stabilize within the first 1-3 years at 130°C and within a matter of 40 min a 200°C irrespective of the sample. The generally independent the failure time of the HRS to split suggests that the origin of the HRS failure is not only dependent on the oxide modification by the implanted Si, but secondary parameters might also play significant role and will be discussed later in this chapter. It is hence straightforward to see that the primary component of data retention loss in the memory technologies tested for their respective optimal programming conditions is not the short-term drift of the LRS but rather a long-term drift of the HRS with LRS retention loss starting to become significant only in the limit of high temperatures of the 5% LSI sample.

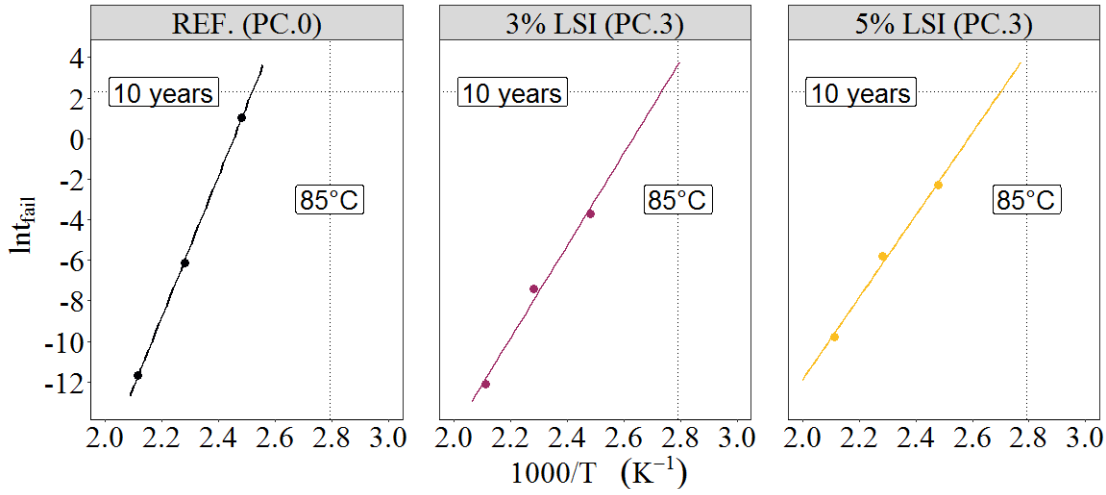


Figure 3-37 Calculated failure time of the power-law model for different temperatures at the 1% HRS CDF tail.

In Figure 3-37 the dependency of each of the HRS tail component bits is shown as a function of inverse absolute temperature. The calculated failure times are found to follow well an Arrhenius process (Eq. (3.20)).

$$t_{Fail} = t_{F,0} e^{-\frac{E_{act}}{k_B T}} \qquad (3.20)$$

By extrapolating each Arrhenius model to lower temperatures an estimate for the bit failure time at operating temperatures can be determined. Here, the projection is done to the consumer electronics standard of 85°C. Pure HfO_2 HRS tail bits are definitely more stable than LSI doped devices. Si doping is found to somewhat decrease the activation energy of the HRS tail from 3.05 eV in the no-implant reference programmed at high current to approximately 1.8 eV on the doped samples. Even though the comparison is made for different set programming conditions, the lower activation energy of the HRS found in the LSI samples programmed using low current condition denotes that the Si alloying with HfO_2 is comparatively less thermally stable that its no implant reference. Nevertheless, all three technology variants exceed the consumer electronics limit seemingly guaranteeing more than 10 years of data retention at 85°C in all three cases with LSI devices not reaching the automotive / space standard of 165°C. The extrapolated values for the three technology variants studied are summarized in Table 3-3.

Table 3-3 Summary of data retention calculated activation energy for the HRS tail at 1% of the CDF projected data retention time at the consumer electronics standard and maximum applicable temperature.

Split	Origin of Data Retention loss	E_{act} (eV)	T_{max} for 10 year data retention
REF. (PC.0)	HRS Tail	3.05 ± 0.08	125^{o}C
LSI 3% (PC.3)	HRS Tail	1.8 ± 0.2	90^{o}C
LSI 5% (PC.3)	HRS Tail	1.7 ± 0.1	90^{o}C

3.9.6 Solder Reflow Tolerance

The devices presented in this work have been subjected in high temperature annealing for a total of 30 min at 260^{o}C using high current (PC.0) and low current (PC.3) programming conditions. Figure 3-38 summarizes the results presented before (Figure 3-30 & Figure 3-32) for the total stress time at solder-reflow compatible conditions.

Following the previously observed trend, the LRS of all samples is the most stable of the two states showing less than 3‰ erroneous bits for the no-implant reference. 3% LSI presents with excellent LRS stability with errors below the detectable threshold of 0.3‰ in the 4 Kbit while 5% LSI is the least stable out of the technology variants. At the same time, the high resistive state is still the major contributor of erroneous bits with more than 5 times the erroneous bits identified as compared to the LRS. This is in part due to the assessment threshold being at the 26 KOhm range, a value inevitably closer to the extrememost erratic bits of the HRS rather than those of the LRS. Nevertheless, given the previous analysis it is clear that the root cause of erratic bits is primarily the HRS lower tail.

At the same time, it is inevitable that evaluation should be done at a realistic assessment threshold, that being the optimal for the chip operation. All memories tested can be successfully reprogrammed past the this solder-reflow like annealing step.

3.9.7 Discussion

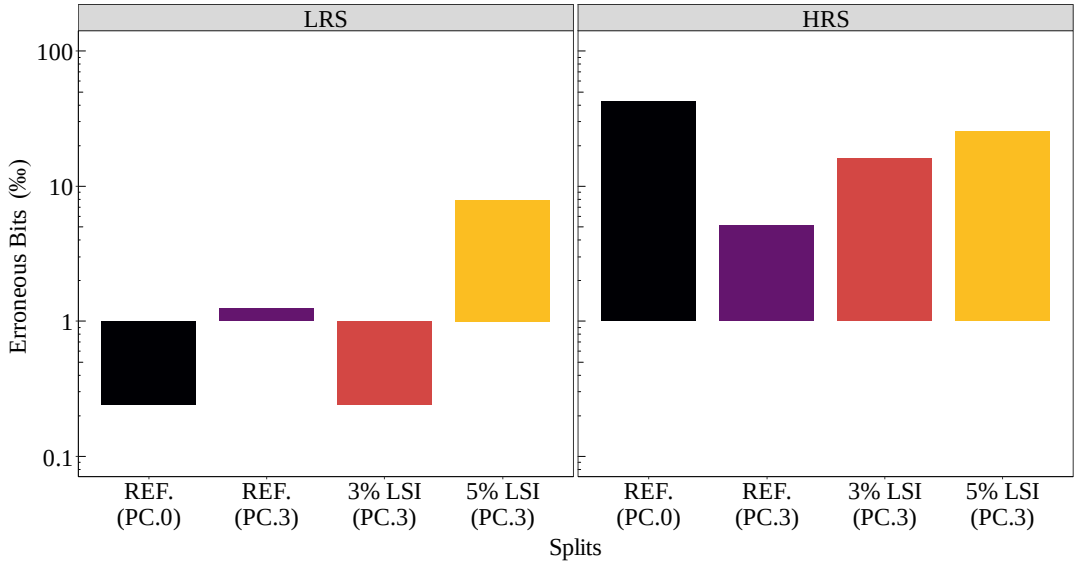


Figure 3-38 Count of Erroneous cells for the technology variants tested after 30°C at 260°C thermal stress.

It is worth stressing the tradeoff between LRS and HRS stability as a function of the programming current during the SET operation, which is again recovered here (Figure 3.34 comparison between REF. PC.0 & PC.3).

Observing the behaviour of the HRS error bits it is straightforward to see that gradually, the low-current programmed LSI HfO_2 devices present similar behaviour to the high-current programmed no-implant reference. This can be an interesting trade-off that can be attributed to the effect of defect engineering and the gradual modification of the diffusivity of the material as a result of material modification on the one hard (LSI devices) and programming current on the other (high-current vs low current programming).

This could be better understood if seen under the effect of diffusivity: As Si is incorporated in HfO_2 the material matrix is modified so that electrically active defects are created one the one hand, while the migration of oxygen ions towards the reactive electrode is facilitated on the other (reflected by the increase of the observed diffusivity, Figure 3-11). Considering this symmetry, it could be theorized that high compliance current programming acts to increase the ion flux towards the reactive electrode, but in turns, resulting in a more sub stoichiometric oxide with higher diffusivity. As a larger fraction of oxygen migrates and is thus oxidized inside the volume of the reactive electrode the LRS is gradually oxidized. On the contrary, the very same mechanism leads to degrading

the HRS retention: As the redistribution of oxygen is facilitated oxygen atoms can migrate to neighbouring available sites thus reshaping the distribution of carrier centres and disturbing the percolation path. This can facilitate the drift of the HRS as experimentally observed.

Closing this section, a final remark needs to be made on the proposed empirical model. Even though the proposed relation seems to accurately fit the experimental data, a complete derivation of the relation could not be treated. Even though several works from different fields propose the emergence of power law dependencies for random walk processes (thought to be the case for the migration of ions during the data retention experiment resistance drift) certain aspects of the model cannot be directly justified in a strict physical basis. The origin of the linear dependence of the fitting exponent α (Figure 3.35) even though previously reported for similar problems [citation] to fit well experimental data, is not yet fully understood.

From an experimental point of view, the capacity of the proposed relation to accurately predict the extrapolated failure time at a temperature outside the one of the annealing experiment is required. This originates from the fact that the experimental data spans only 24 hours. To improve accuracy and validate model capacity, more points in time, and a larger time scope spanning a few weeks of thermal stress are required. Nevertheless, the time required for such experiment is practically unattainable without dedicated equipment. In this work, the proposed law serves to make quantitative comparisons between technology variants exposed to identical experimental conditions in a relatively short time.

3.10 Conclusions

In this chapter, the alloying of Si with HfO_2 resulting in bulk defect creation in the oxide is exploited in two different integration approaches to demonstrate working ReRAM in the BEOL of a 130nm CMOS process. Two different integration schemes are studied for the implant: Full implant on the total volume of HfO_2 and localized implant approach inside a controlled volume.

While in both cases Si implantation serves to reduce the forming voltage of the devices (§3.3) the effect is stronger in the locally implanted devices. The effect is attributed to a local field enhancement in the implant/no-implant junction that potentially induces field-

confinement of the ion motion increasing oxygen vacancy diffusivity (§3.4). Local implantation also serves to reduce set/reset transition points (§3.5).

From a performance prospective, blanket implantation shows high variability, attributed to a strong interaction with the memory environment. On the contrary, locally implanted devices show a highly compact LRS. Error analysis shows that local implantation allows nearly error-free operations up to ~2E5 cycles (§3.7). On the other hand, the facilitation of vacancy migration induced by ion implantation impacts the maximum cyclability of the array as devices show accelerated ageing effects wrt the reference HfO_2 stack (§3.8).

Data retention studies have shown that for the optimal operating condition the HRS of the cell in the worst-case becomes the limiting factor for ReRAM state loss over time (§3.9). The HRS tail dynamics is modeled using a power law for the time-to-fail allowing to project the state lifetime as a function of the content of Si in HfO_2 for reference HfO_2 and locally implanted devices. All devices studied qualify data retention of 10yrs at 90°C. Nevertheless, the temperature limit is as high as 125°C for the reference stack, signifying higher vacancy diffusivity in the implanted stack.

The results to date imply that further tuning of the dose and implantation conditions of the stack can result in minimizing the impact in data retention and endurance while allowing for a significant reduction of the forming requirements even in thick oxide stacks.

CHAPTER 4. RERAM COINTEGRATION WITH GATE-FIRST FDSOI ON 300 MM TECHNOLOGY

4.1 Introduction

Up until now, we have studied and tried to address the issues of forming and reducing programming power in HfO_2 based ReRAM technology. We have demonstrated a systematic approach to relax forming constraints, reduce programming current and programming errors while partially trading off in high-temperature data retention and maximum device cyclability. Indeed, the distinctive advantage of ReRAM as compared to other emerging non-volatile memory technologies lies in its potential for low-power / low-voltage operation (set/reset currents < 250 µA, set/reset voltage < 2.5V) as well as in the compatibility of the related materials to the front end of the HKMG technology.

Passive selector technology can only suppress leakage current of devices in LRS; indeed even the state of the art passive selectors seem to require external current control elements during the set operation [108]. Hence, the 1T-1R bit cell architecture based on CMOS transistors is still the most reliable and design-friendly solution for ReRAM integration in large arrays. In order to maintain high-density, ReRAM needs to be co-integrated with advanced CMOS nodes. At the same time, the strong affinity of HfO_2 based ReRAM to CMOS technology makes it a promising candidate for low-cost high-density memory closely co-integrated to the CMOS level. Approaching the memory cell to the selector, can reduce interconnect delay, liberate chip space on the upper metal levels and ultimately allow to physically implement near-memory and memory-everywhere paradigms.

Here, we present an industrially compatible integration strategy, where the ReRAM cells lie below M1 and are scaled down to an 110 nm (post etching+spacer). In the following we develop and optimize a ReRAM fabrication flow involving, a bilayer-hard mask strategy, coupled with an optimized ReRAM sidewall etching chemistry. Functional 1T-1R ReRAM cells co-integrated with gate-first FDSOI transistors were demonstrated. The approach has the capacity to reduce the number of masks needed while demonstrating working, 1T-1R cells with reduced total footprint.

4.2 Co-integration of HfO_2 based ReRAM in FDSOI gate-first process flow

4.2.1 Integration concept and process flow

In this implementation, the ReRAM element is integrated below M1 in a 300 mm FDSOI silicon nanowire-like gate-first process. The gate length is fixed at 50 nm and can be trimmed down to 50 nm while the width at 400 nm. The transistor fabrication process is performed in CEA-Leti and follows a standard nMOS gate-first SNW flow. Nevertheless, owed to the lard width required to achieve current levels sufficient for ReRAM operation, the large W/L ratio the device geometry is closer to a standard FDSOI geometry from the strict electrostatics point of view. Post the source/drain silicidation step, a contact etch-stop layer (CESL), and a passivation post metal dielectric (PMD) are deposited. A chemical mechanical polishing step (CMP) is used to planarize the surface.

The ReRAM bottom electrode via contact is etched after a standard lithographic step, after which follow a Ti/TiN liner and W filling steps. After the W CMP planarization step, the memory layers are deposited consecutively, notably a 10 nm TiN bottom electrode, 5 nm HfO_2 and varying thicknesses of 5 nm and 10 nm of Ti reactive electrode. The top contact electrode is 30 nm thick. The TiN is deposited by PVD at 350^{o}C. Both Ti and TiN are deposited within the same PVD chamber and the wafer is kept under inert atmosphere conditions during intermediate steps of the process. Different PVD process variants have been evaluated as will be detailed in the following sections. Finally, a 30 nm SiN and 20 nm SiO_2 bilayer is deposited serving as a hard mask. Contrary to the 200 mm integrated cells discussed in Chapter 3, the scaled devices in this chapter require to reduce the top and bottom TiN electrode thickness to maintain a device height to critical dimension aspect ratio that can be sustained by the process.

On top of the bilayer mask a spin-on-carbon (SoC) and photoresist (including anti-reflective coatings) are deposited. I-line stepper lithography is used to pattern the MIM ReRAM element with diameters varying from 1 µm down to ~110 nm. Post lithography, the photoresist and SoC layers are etched (stop on SiO_2), transferring the pattern to the double hard-mask. Residues are then stripped using dry oxygen plasma and a dedicated

FEOL	Gate First Trigate on FDSOI
	CESL
	Passivation Oxide + CMP
ReRAM Bottom Contact	Drain contact Litho & Etch
	W Fill
Memory	OxRAM + SiN/SiO_2 HM
	OxRAM Litho + Etch
	SiN Spacer Dep. + Trimming
	PMD
ReRAM Top Contact	Contact Litho
	Contact Etch: Selective stop
	W Fill
BEOL	**M1, M2 & Alu PADS**

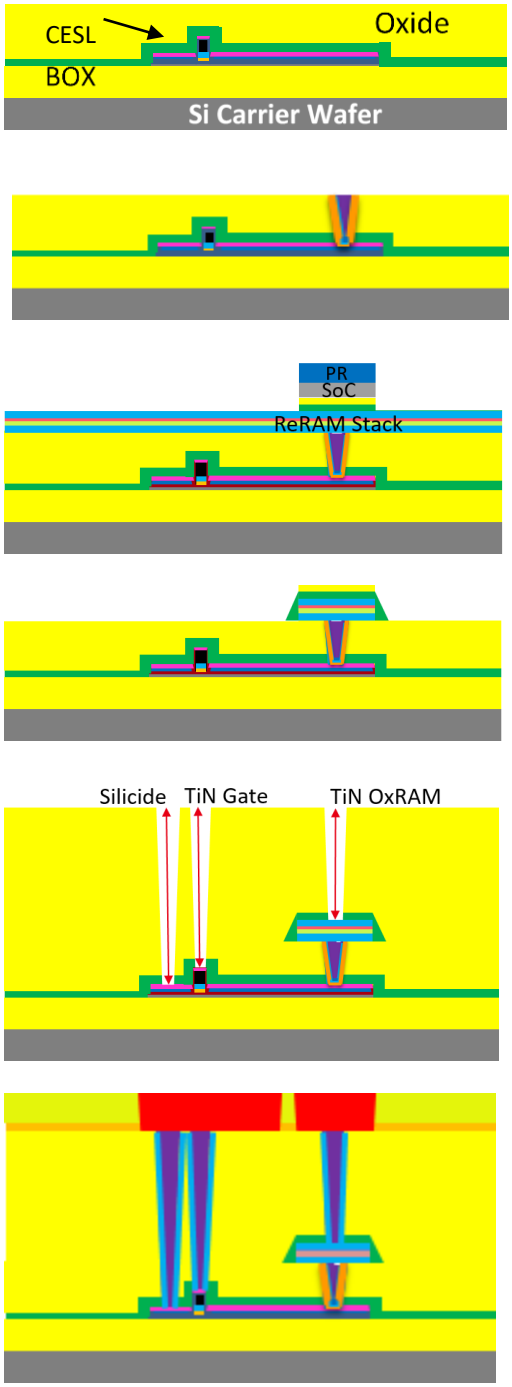


Figure 4-1 (a) FEOL block: Transistor fabrication up to PMD (oxide) and CMP. (b) Transistor contact opening blck (c) ReRAM memory module.

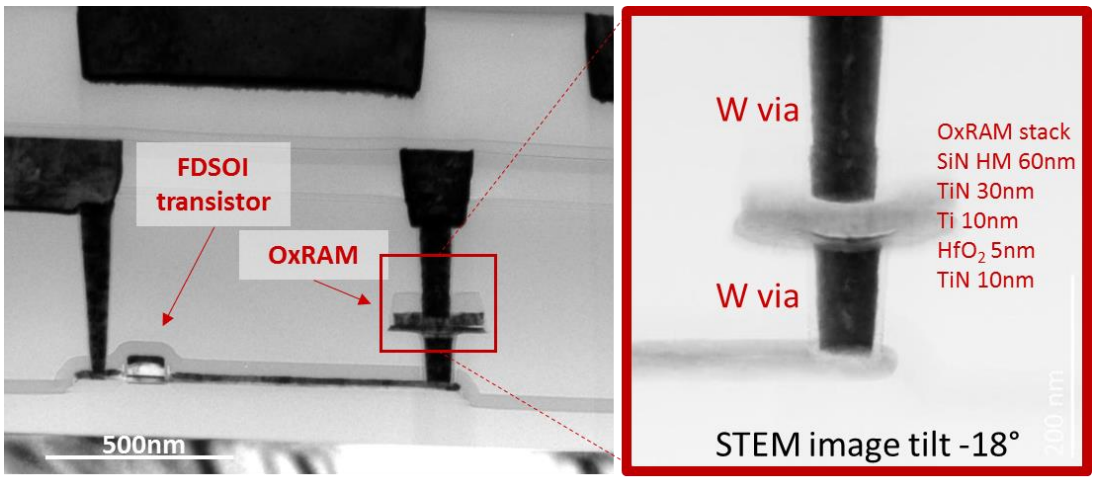


Figure 4-2 HR-TEM of a 1T-1R memory cell with a 200 nm MIM capacitor integrated with a 60 nm gate length trigate on FDSOI post the M1 interconnect block.

wet chemistry. The hard mask is then used to etch the TiN/Ti/HfO_2/TiN stack stopping on the SiO_2 layer.

The memory process module ends with deposition and etching of the SiN spacers that serve to isolate the memory element from the SiO_2 environment and protect both Ti and HfO_2 from the impact of humidity that is known to be detrimental to ReRAM cells [168]. The SiO_2/SiN hard mask allows to trim the nitride spacers without impacting the thickness of the SiN hard mask, owed to the selectivity of SiN chemistry with respect to SiO_2. This allows to minimize SiN overetch on the TiN top electrode. The module ends by deposition of PMD and CMP.

The memory element is contacted to metal interconnect using a high aspect ratio etching step. In particular, all 1T-1R contact nodes, source, gate and MIM top surface are opened during the same etching step. In order to contact the different nodes, the PMD and CESL need to be etched on the transistor side while the PMD and double hard-mask on the MIM side. The via critical dimension is in the order of 60 nm while the aspect ratio of the depth to CD profile varies for all three contact nodes and approaches 9.1 at the worst case. The etch-stop surfaces are different for each of the contacts: Silicide for the source side and gate of the transistor and TiN for the top electrode of the 1R MIM element, respectively. This etching step makes it possible to economize additional lithography steps that increase complexity and cost. This single-step multi-level etching strategy is made possible by precise control of the hard-mask bilayer thickness in combination with a highly selective RIE chemistry.

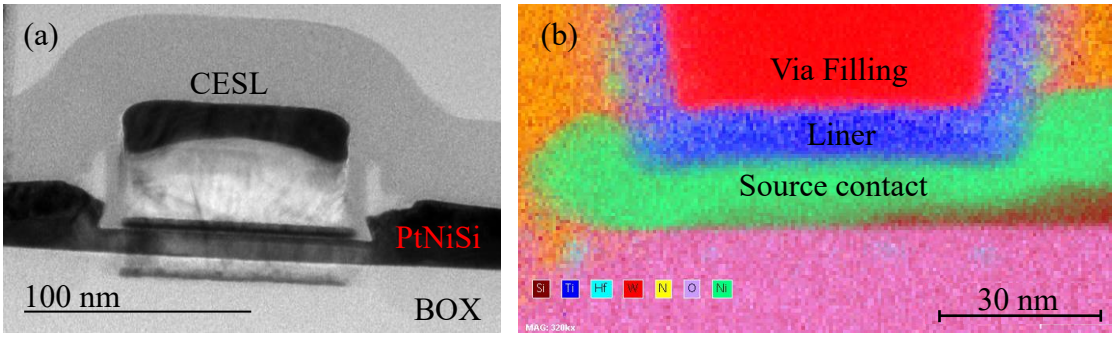


Figure 4-3 (a) HR-TEM of a Trigate transistor (b) Trigate source contact post ReRAM top electrode contact etching and Via filling process steps. No structural defects can be identified.

The bilayer hard mask serves double role. First, it protects the memory stack sidewalls from oxygen plasma exposure, which risks to impact the memory element operation, especially for the strongly scaled devices. Direct application of the SoC and photoresist trilayer, on the TiN top electrode followed by etching of the stack would expose the stack sidewalls to the oxygen plasma required to clear out organic PR and SoC residues. On the contrary, SiN and SiO_2 do not require to be stripped and can remain as part of the stack owed to material affinity. Moreover, tuning the thickness of the SiN/SiO_2 bilayer acts to directly tune the etch rate during contact opening step making the previously described, single-step contact opening etch process possible. The structure after the M1 deposition and patterning step can be seen in Figure 4-2. The total 1T1R footprint is ~ 0.2 μm^2.

4.2.2 FEOL and MOL technology characteristics

In Figure 4-3 (a) an HR-TEM image of the transistor node (t_{ox} = 1.4nm) is shown while Figure 4-3 (b) shows the transistor source contact after the top electrode etching step and contact filling process. No morphological defects can be traced in either case for the FEOL and MOL parts. The TiN liner in the via acts as blocker for W out diffusion in the PMD.

The fabricated transistors present with uniform behaviour across-wafer delivering on-currents according to the projected sizing (Figure 4-4a). Impact ionization effects start appearing in the current transfer characteristics of the nMOS above the 2V range on the drain. Impact ionization and its cycle-to-cycle or device-to-device variability can be of varying importance for the operation of resistive memory. First, the LRS state variability can be directly impacted as the LRS state is directly related to the compliance current during the set process. Reset is not expected to be strongly impacted as the maximum

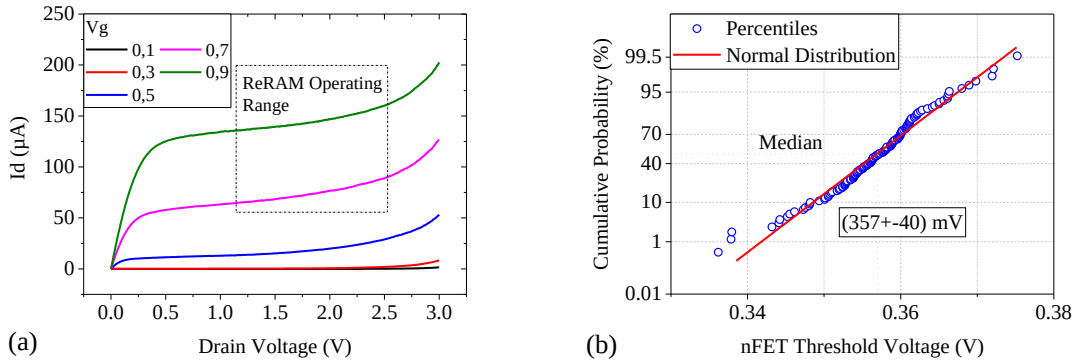


Figure 4-4 (a) I_d-V_g current transfer characteristics for the transistor node (b) Distribution of Threshold voltage

current through the transistor for a given reset condition is usually higher than the current needed to reset the memory element. Minor variations of the source-drain series resistance as a result are considered to have no effect on the HRS variability of the OxRAM.

Nevertheless, the exponential loss of control of the transistor gate at high fields on the drain side severely restricts the maximum voltage applicable on the top electrode node, which is critical for the forming operation of the ReRAM element. Firstly, to reduce the element degradation low-current forming operations are ideally required, as previously discussed [chapter 3]. Secondly, high forming voltages can even risk degradation of the transistor element, given that thick oxide forming operations require total stress times in the order of 50 µs. Even cost, during the forming event, the acute voltage drop between the top and bottom electrodes of the MIM stack (connected to the transistor drain) will most certainly result in strong spikes during the forming event, risking to degrade both the 1R yield, stress the transistor as well as impact the long-term reliability characteristics of the 1T1R cell (endurance, retention)

Considering the strong scalability of the ReRAM element (in principle < 20 nm) it is evident that the limit on scaling down the 1T1R comes from the transistor. Hence, transitioning to more advanced nodes is not only required in order to save power, but even more importantly to gain in density. As a result, impact ionization is a major blocker for the successful integration of ReRAM in advanced nodes as the loss of control of the gate vertical field to the source-drain lateral field becomes more significant with scaling. Even though some margin for transistor engineering is still possible, it comes with certain performance trade-offs and cannot completely mitigate the gradually increasing restrictions on the maximum applicable source / drain voltages.

Using a thicker gate oxide increases the minimum achievable gate length for a given technology, owed to both attenuation of the gate vertical field but also due to technological limitations. The use of thicker gate spacers could push the onset of impact ionization to higher source/drain voltage values. However, this will decrease the on current capacity of the transistor as it directly increases the metallurgical gate length of the device. Finally, using more gradual junctions (LDD engineering) could increase the long-term reliability of the transistor, but is not expected to strongly mitigate the onset of impact ionization.

Considering that the typical operating range for HfO_2 based technologies lies in the range of 1.3-2.5V the importance of transistor reliability and the compatibility of the two technologies can become a major roadblock. In the light of these limitations, cell engineering allowing to reduce the operation constraints of ReRAM in terms of both forming, set and reset voltages as well as operating current is not only required to improve technology performances but is also important for future memory scalability and co-integration to advanced CMOS.

In the following, we focus on integration of HfO_2 ReRAM; for the reasons described above the oxide thickness is limited to 5 nm while 5 and 10 nm Ti are used for the reactive electrode.

4.2.3 *MOL ReRAM integration: Impact of Structural Integrity of the Memory cell in electrical characteristics*

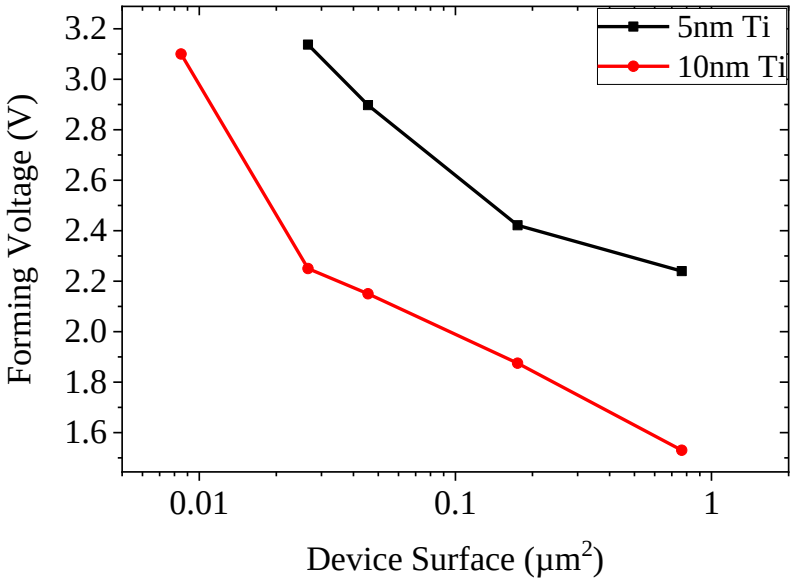


Figure 4-5 Evolution of forming voltage with size and thickness of the Ti layer, with 5 nm HfO_2.

The median forming voltage of the fabricated devices as extracted by quasi-static IV forming is shown in Figure 4-5 as a function of the MIM device critical dimension for 40 measured devices. As expected, the forming voltage decreases with increasing Ti thickness and increases with decreasing surface.

Nevertheless, the forming voltages scale weakly with surface down to 300 nm C.D. devices whereupon a stronger dependence of the forming event on the surface is observed. One universal law fails to describe this behavior, contrary to data found in the literature for similar structures. The acute change of the curve's slope implies that either a secondary effect becomes dominant or the effective surface of the devices is smaller than the one defined by lithography.

Statistical analysis was performed on the forming events of the on-wafer tested devices. As forming consists of a breakdown event, Weibull statistics is expected to be followed. Analysis shows that the Weibull slope increases with increasing device size, with the trend becoming less pronounced as the thickness of Ti decreases (Figure 4-6). In this case it can be directly observed that even though the forming voltage of the 5nm Ti samples is higher than the one of those with 10nm Ti, its slope has higher values, suggesting the forming variability is lower. The dependence of the Weibull slope on device area implies that the

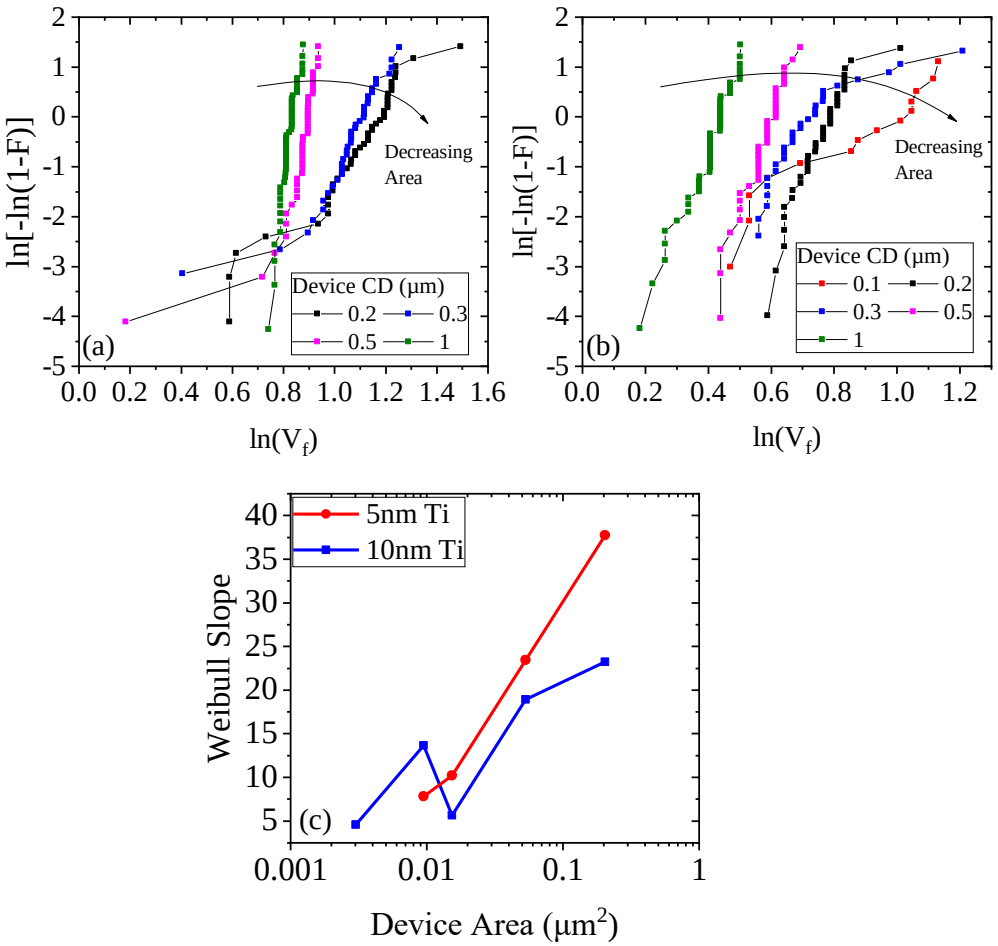


Figure 4-6 Weibull plots with device size and Ti thickness for 5 nm (a) and 10 nm (b) Ti scavenging layer. HfO_2 thickness is fixed at 5nm. (c) Evolution of slope of distributions shown in (a), (b) as a function of device size.

Ti layer, responsible for oxygen scavenging and -as a result- generation of defects at the Ti/HfO_2 interface is affected by device lateral dimension; in contradiction to the standard case where it is area insensitive [4, 5].

This size dependency is direct evidence of the impact of the memory sidewall, both in terms of the oxide and the reactive layer thicknesses to the breakdown characteristics of the defect generation in ReRAM and will be discussed later in further detail.

Quasi-static and pulsed operation reaching up to 10^8 cycles of the memory cell is shown in Figure 4-7 a,b respectively for large structures. However, both the memory states show significant device-to-device and cycly-to-cycle variability. This further supports the previous hypothesis of the degradation of the Ti/HfO_2 interface properties, originating

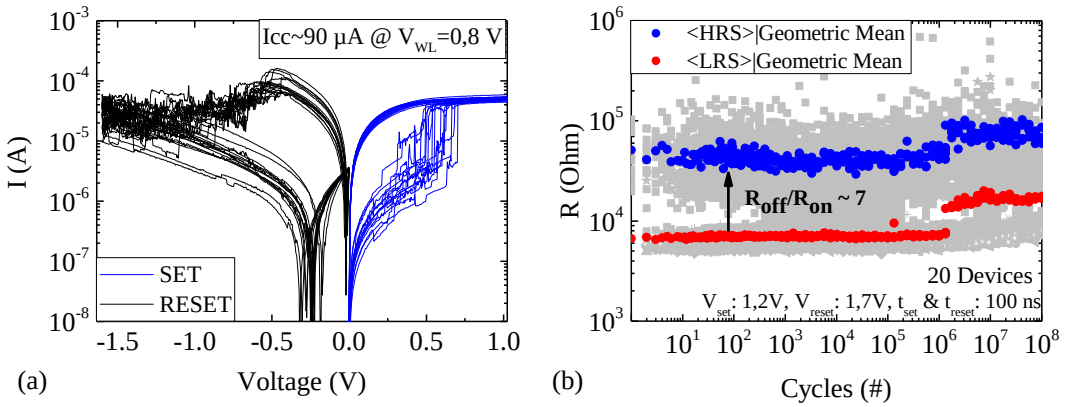


Figure 4-7 (a) Quasi-Static IV characteristics of a 1 µm device. (b) Pulse-cycling IV characteristics of a 1 µm. 5nm HfO_2, 5 nm TiN stack.

from the OxRAM sidewalls, especially as the lower dimension elements are non-functional.

The "reverse hump" app earing during the return loop of the static IV reset curve in Figure 4-7 a is a non-linear current feedback arising from the floating node potential between the ReRAM element bottom electrode and the source of the transistor and can be understood as follows: as the ReRAM element is reset, its resistance passes at very high values in the order of hundrends of Kohms. This causes the potential drop across the ReRAM node to fall severely after reset. Hence, the effective V_{ds} on the transistor will decrease (Vs = 0), and for low drain voltages the transistor structure is forced in cutoff owed to the very low values of the floating potential. At the same time, the OFF current of this transistor technology is in the order of ~1 pA, much smaller than the leakage current through the memory. As voltage is continuously applied, the bulk defects in the volume of the ReRAM are filled. At the same time, the pathway on the side of the transistor is blocked resulting in a current feedback back to the reading circuit. The floating node will be at non-zero bias for as long as charge remaining in the ReRAM and current feedback to the sense circuit is just a consequence of discharging in the ReRAM and potentially the metal line.

4.2.4 Morphological Characteristics of the memory node

TEM & EELS analysis reveals area-dependent cracking/delamination in the vicinity of the top TiN layer (Figure 4-8 a,b), suggesting the presence of significant plane stress. The presence of the crack can expose the Ti layer to contaminants such as oxygen (Figure 4-8 a,b) or nitrogen. Elemental detection analysis in TEM for large (1µm) (Figure 4-8c) and

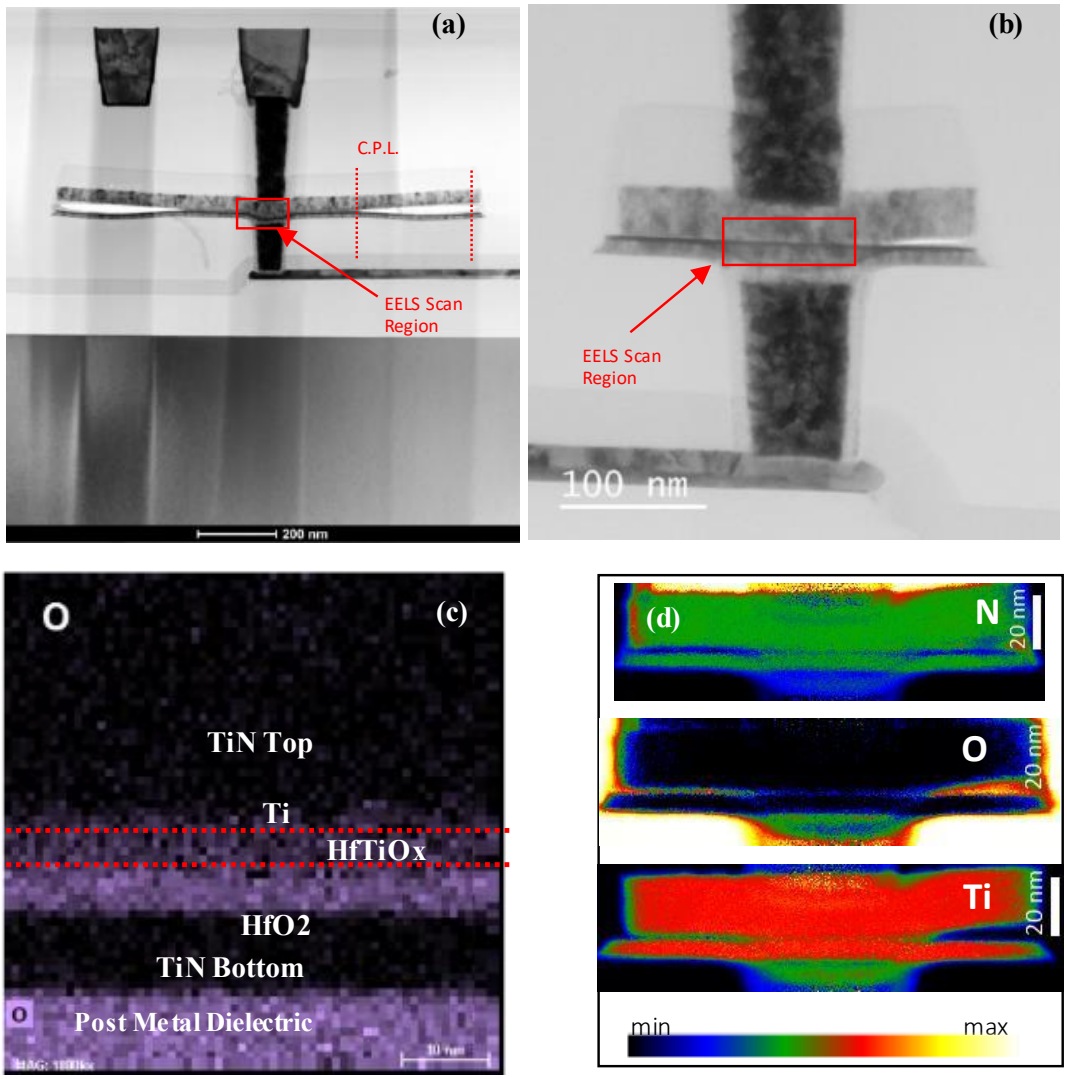


Figure 4-8 TEM Image of (a) 500nm and (b) 100nm radius device. (c), (d) STEM EELS in the TiN(30nm)/Ti(10nm)/HfO2 (5nm) stack corresponding to (a), (b). Traces of O can be seen in the TiN/Ti interface of the scaled device. 5nm HfO_2, 5 nm TiN

scaled (0.2 µm) devices (Figure 4-8d) reveals that the Ti/TiN bilayer quality is oxygen rich close to the crack region while the nitrogen concentration inside the Ti layer is abnormaly high. The homogeneity of a chromatic tone in the Ti/TiN layer regions in the Ti, N, and O spectra of (Figure 4-8d) implies that the Ti layer has been potentially nitridized. This can occur if delamination has taken place during the etching / stripping steps when the 3D structure is "released" or during the subsequent SiN spacer deposition step.

In the presence of either O or N excess, the reactive electrode (Ti) is expected to be strongly passivated leading to lower reliability and yield. The strong dependence of

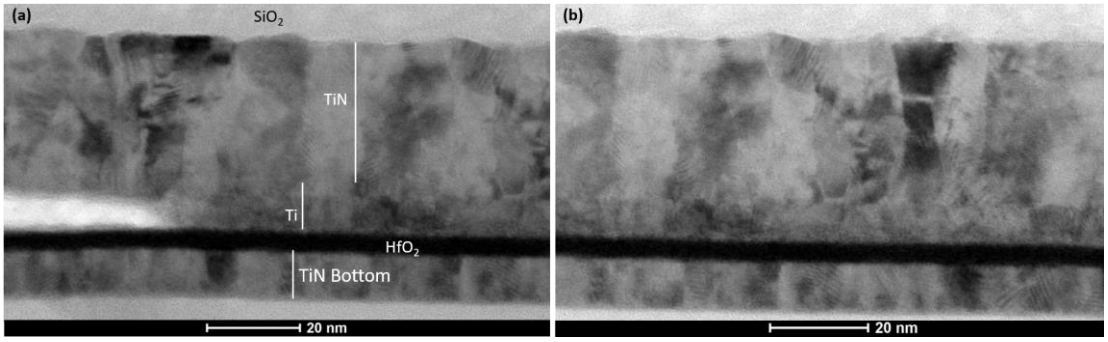


Figure 4-9 HR-TEM of a (a) Crack on the edge (b) Interface in the center of a 1 μm device. 5nm HfO_2, 5 nm TiN stack.

Weibull parameters to device area and Ti thickness can be attributed, first to the strong change of effective surface owed to crack propagation followed by a strong modifycation of the local material structure owed to segregation of either contaminants or Nitrogen and Oxygen towards the Ti/HfO_2 interface introduced by the stripping and SiN spacer deposition process steps (Figure 4-8 c,d).

4.2.5 Some considerations for interface mechanical failure

Closer TEM observation reveals that the crack propagates close to the Ti/TiN interface (traces of Ti can be seen on the HfO_2 upper surface, Figure 4-9a). Nevertheless, origin of interface contact failure is hard to determine from TEM analysis alone.

Contact zone modeling (CZM) [169], [170] can reproduce the defect behavior as a function of the TiN/Ti interface adhesion properties [170], [171] and projected stress level in TiN. The mechanichal properties of HfO_2, Ti and TiN are unknown for ultra-thin films in the deep nano scale dimensions. In performing this simulation one of the major challenges lies in the complete absence of bibliographical data to serve as reference and measure of comparison. Here, the Young's moduli for the bulk materials are considered. Furthermore, the failure is assumed in the Ti/TiN interface owed to its high mechanical mismatch parameters [172]. Normal and shear stress components are assumed to be of equal importance (normal/shear ratio =1). The Benzeggagh-Kenane Criterion for mode mixity is assumed to be 2.3 so that to Mode I separation is dominant [173], as experimentally observed. Wafer level bow measurements suggest he presence of high compressive stress in TiN in the order of -6GPa (as measured for fullsheet deposited TiN) The interface

adhesion parameters are fitted such that the calculated displacement and crack propagation length observed in TEM are closely reproduced. For the fitting procedure, the 1um structure is used to determine the interface parameters; then the simulation is repeated for multiple sizes and the result compared to the experimental values found in the 200nm MIM ReRAM. The interface parameters resulting from the model fit to the two experimental points are summarized in Table 4-1.

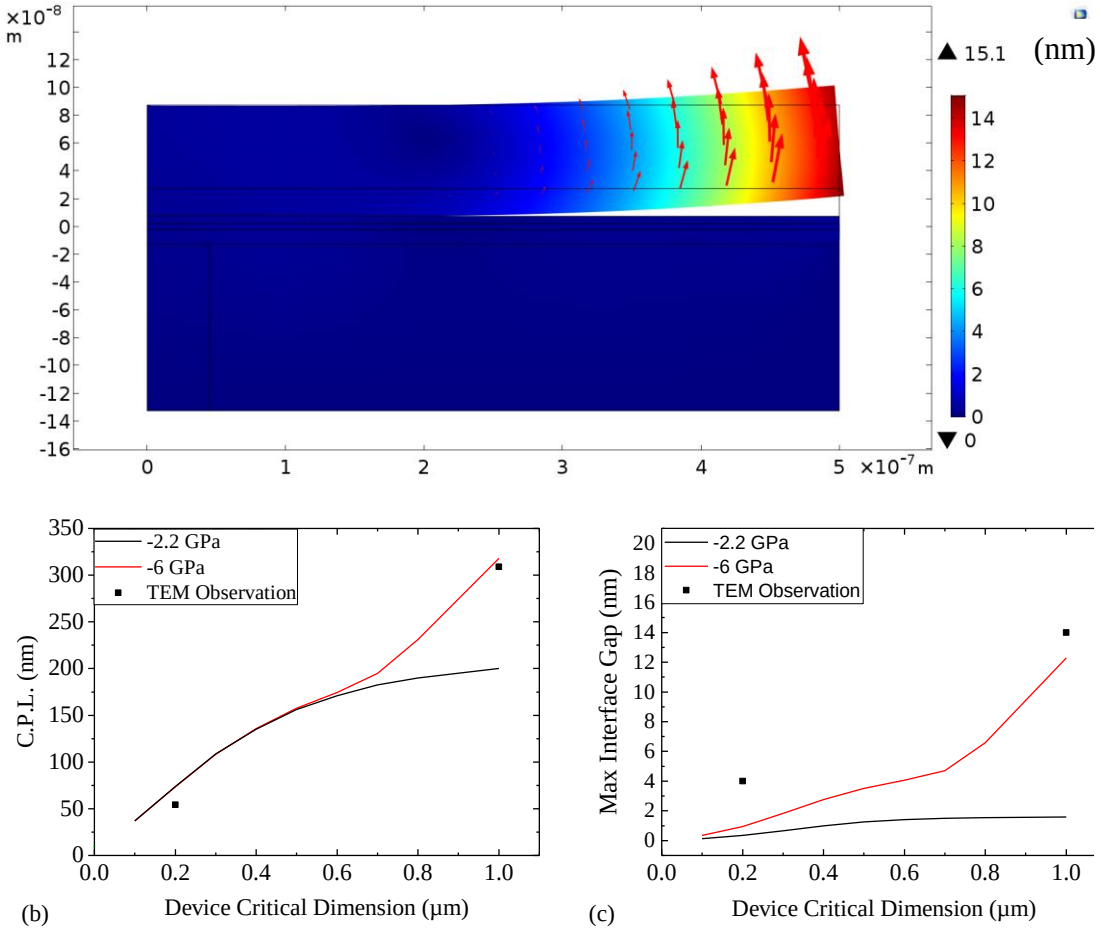


Figure 4-10 (a) 2D Simulation of contact loss in a TiN/Ti/HfO₂/TiN structure with layer thickness identical to the fabricated devices. Initially compressive plane stress is assumed in the TiN leading to the observed delamination. (b) Crack propagation length and (c) Maximum separation of layers for different simulated levels of compressive stress.

Table 4-1 Extracted mechanichal property characteristics for the Ti/HfO₂ interface

η	G_{ic} (J/m^2)	G_{iic} (J/m^2)	σ_{ic} (Pa)	τ_{iic} (Pa)
2,3	6	24	960e 6	3e9

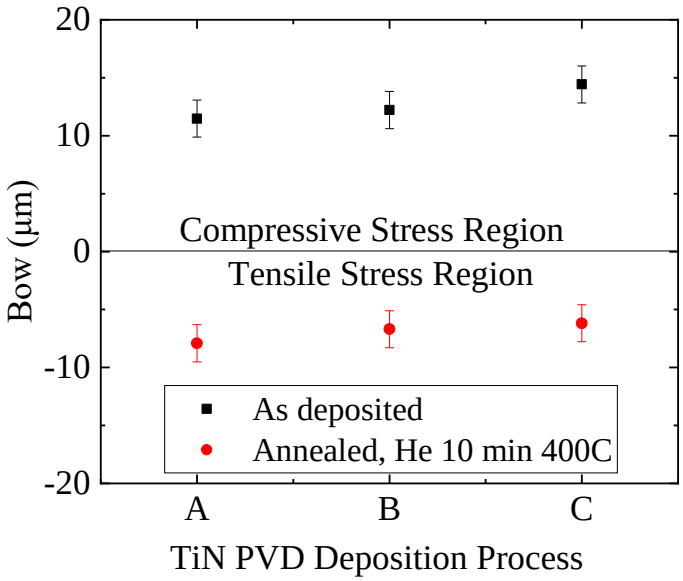


Figure 4-11 Bow measurements of full-sheet deposited TiN/Ti/HfO_2 layer before (black) and after (red) annealing in He atm for 10 min at 400^{O}C

Nevertheless, strict quantification is particularly challenging owed to: i) the highly non-linear nature of the problem and ii) the almost complete absence of bibliographical sources to provide some reference for the properties of the material interfaces implicated here. Finally, it should be noted that the mechanical properties of the films can largely vary according to synthesis conditions.

Contact Zone Modeling indicates that that the presence of significant mechanical stress in the TiN layer can induce the observed crack and delamination effect. Furthermore, the morphological deformation confirms that the stress is compressive in nature. The origin of mechanical failure can be traced to the presence of nano-sized cracks in the edge of the MIM stack originating from the dry etching and/or the following wet chemistry that serves to remove etching residues. Simulation shows that even though the maximum gap between the delaminated interfaces can be reduced by reducing the initial stress in the TiN layers, the crack propagation length (C.P.L.) cannot be mitigated unless the adhesion properties of the surface are enhanced or the sidewall defects mitigated.

Changing the material and interface properties of the HfO_2/Ti/TiN stack is a multi-parameter problem, especially taking into account the constrictions of i) maintaining the electrical characteristics of the metal layers and ii) maintaining process uniformity across the 300 mm wafer. Stress is expected to be relaxed if the nitrogen content is increased in

the PVD chamber at the expense or increased resistivity. Temperature cannot be significantly increased owed to BEOL limits. TiN thicker than 50 nm could render the layers rigid enough to prevent delamination, but cannot guarantee maintaining adhesion in the interface. To understand the behavior of stress in the ReRAM stack, measurements were performed on HfO_2/Ti/TiN (5nm/5nm/10nm) stacks deposited full-sheet on 300 mm wafers for 3 different PVD processes: Processes A, B are performed in an DC chamber while process C in an RF chamber. In all cases, comparable electrical results are targeted for different deposition conditions of the Ti and TiN films. The stress in the trilayer after the deposition is found to be comparable for all 3 cases as shown from the comparable positive bow values (indicative of compressive stress) in Figure 4-11 implying that the level of stress in the stack is not primarily determined by the deposition conditions but from the nature of interface mismatch. The observed bow measurement inversion requires further investigation and the effect has not been taken into account in the CZM simulation shown above.

4.2.6 Stress-robust Scaled ReRAM process: Morphological Characteristics

From the same on-wafer study, it is observed that the mechanical stress in the tri-layer changes becomes tensile from compressive after annealing at 350°C for 10 min, a thermal budget comparable with that of the SiN spacer deposition step that follows after the etching of the MIM stack. This could be linked to the mixing of the Ti/TiN or Ti/HfO2 interface but the specificities of the mechanism require further study. It is possible that mechanical stress is an inherent property of the MIM stack based on Ti, TiN and HfO_2. Hence, instead of targeting to synthesize a stress-free stack, an alternate route is chosen, that of minimizing the sensitivity of the MIM sidewall. To achieve that, a short nitridation pulse is performed right after the etching of the MIM stack. This serves to form an ultra-thin TiN layer in the Ti sidewall in the order of 3-5 nm. The nitrated Ti surface is les reactive

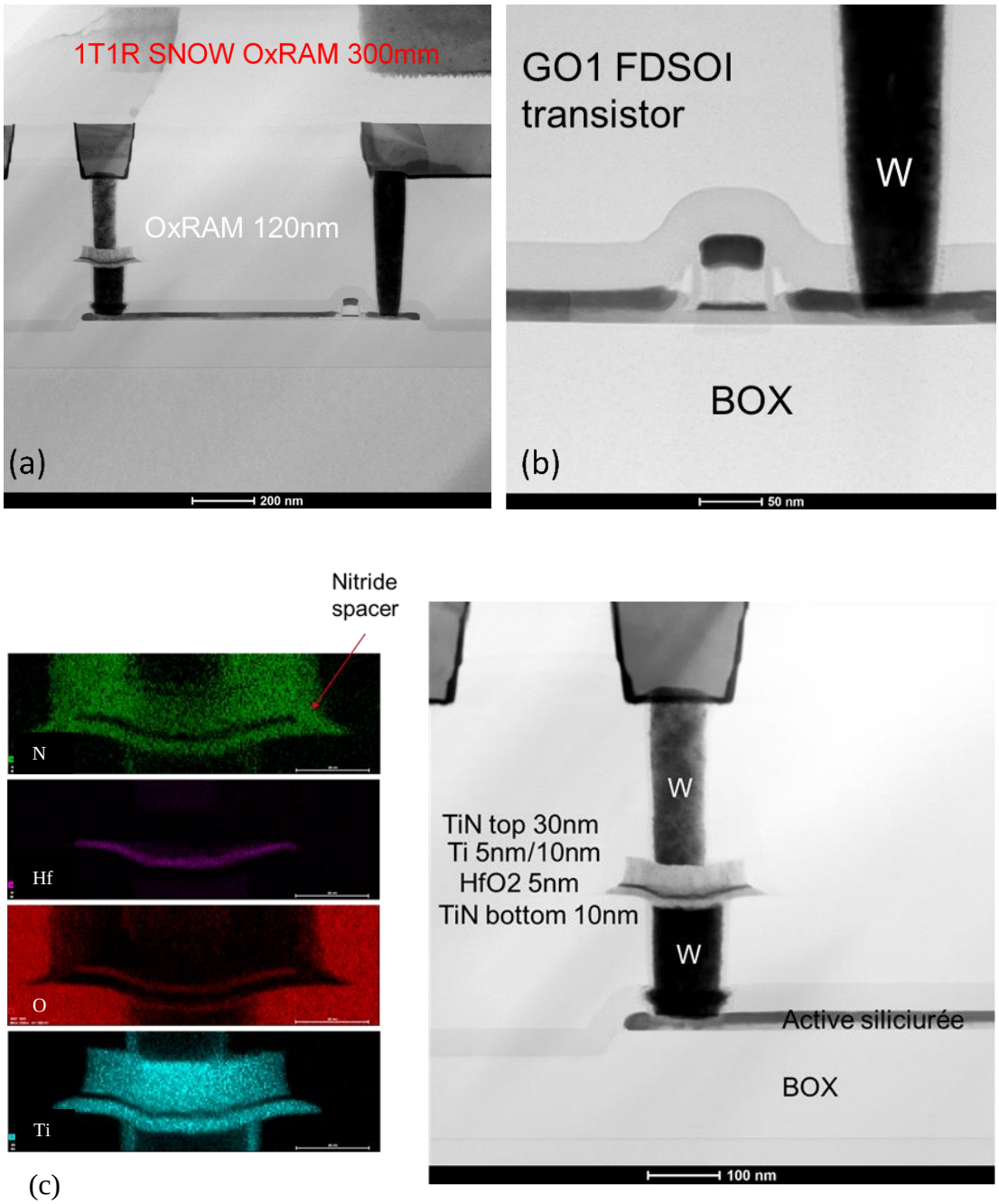


Figure 4-12 (a) 1T1R ReRAM cell (b) focus in the FDSOI node (c) Focus on the 100 nm ReRAM cell with subsequent EDX elemental analysis. 5nm HfO_2, 5 nm TiN stack.

hence it is not as strongly impacted by the mask stripping chemistry that follows. As a result, even if line edge roughness cannot be entirely mitigated, structural defects do not propagate into the structure and mechanical integrity is maintained (Figure 4-12).

4.2.7 Impact of scaling and Ti thickness in ReRAM devices integrated on FDSOI technology

4.2.7.1 Characteristics of the Forming Process with surface scaling

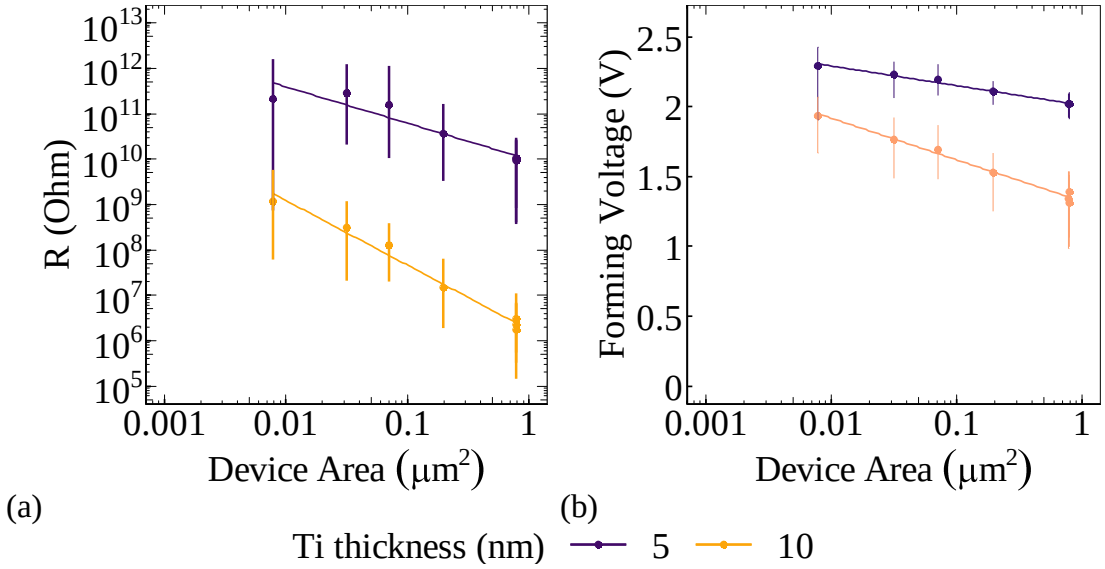


Figure 4-13 (a) Evolution of pristine resistance and (b) forming voltage as a function of device area for different thickness of Ti scavening layer and 5nm HfO_2 layer.

The structurally defect-less MIM structures exhibit very different behavior as compared to the previous integration. In Figure 4-13a, b the pristine resistance and forming voltage of 50 measured devices per size is shown. Much weaker dependence of both pristine resistance and forming voltage with device area can be identified.

Interestingly, even though the forming voltage of a 10 nm Ti MIM stack is lower than that of the corresponding stack with 5 nm Ti, the surface dependence of the forming voltage is *sharper* for the thicker Ti layer. This can be understood if the interaction of the Ti layer edge to its environment is taken into account: The thicker Ti layer can interact more strongly to the environment. To maintain structural integrity, the Ti sidewall has been exposed to a controllable nitridation process; as a consequence, the thicker Ti is expected to be more strongly impacted by the formation of TiN in the sidewall, especially as the surface of the structures is reduced.

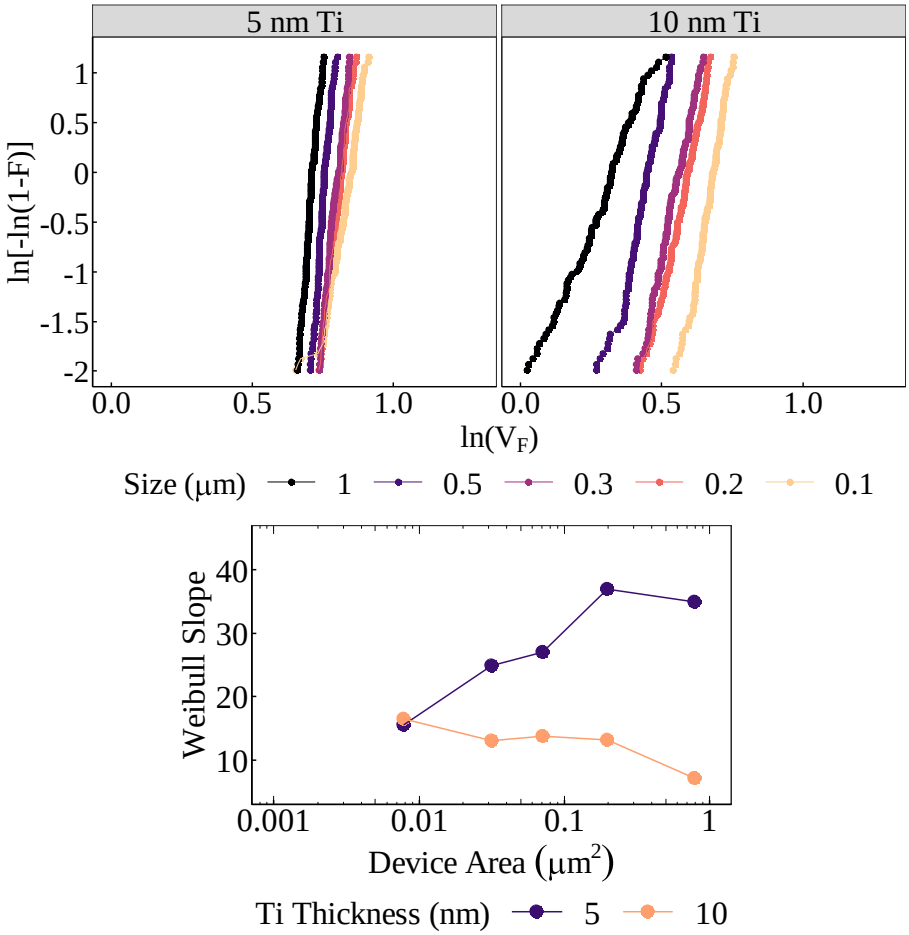


Figure 4-14 (Top) Weibit plots with surface of the forming voltage statistic for 50 devices each of 5nm HfO_2, with 5nm and 10 nm Ti layers. (Bottom) Weibull slope (shape factor) evolution with surface for the 5 and 10 nm Ti layer samples.

Studying the variability evolution in terms of scaling on a Weibull scale (Figure 4-14 a-c) one recovers the expected behavior of Figure 4-6c. Here the extreme most forming events have been excluded (outliers). Even though a thicker Ti layer is more sensitive to downscaling in terms of the behavior of the absolute value of the forming voltage, the variability of forming around the expected value is reduced with increasing Ti thickness. More specifically, comparing Figure 4-6c & Figure 4-14c one can see that between the two process flavors, and under identical deposition parameters the statistical behavior between the two stacks is different.

Notably, in the first flavor (Figure 4-6c) the Weibull slope is particularly sensitive to area scaling for both 5 and 10 nm Ti. The first process flavor values on large devices (1μm C.D.) rise as high as 40 but then quickly degrade to less than 5 in the nanometric level and

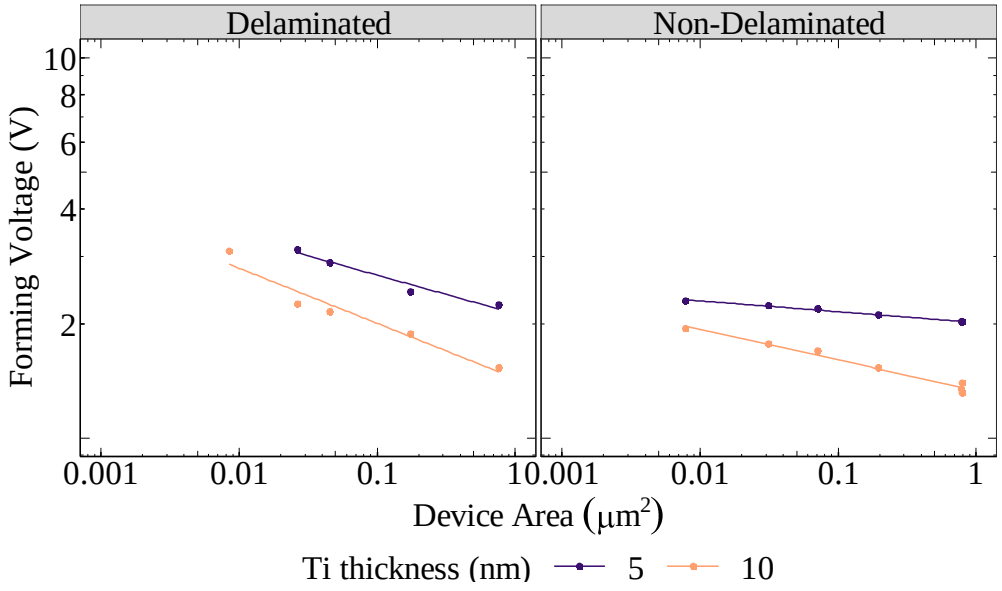


Figure 4-15 Forming voltage (at 63% quantile) evolution for the delaminated devices and non delaminated 5nm HfO_2, devices corresponding to an integration lacking or including Ti scavenging layer sidewall nitridation step (see text).

can be related to the lateral oxidation of the Ti layer owed to the mechanical failure of the stack as previously described. On the other hand, in the second flavor, even though a similar trend is observed on the 5 nm samples the 10 nm are virtually unaffected by scaling. Nevertheless, the latter shows lower Weibull slope values for all sizes, indicting stronger forming variability for the 10 nm stack.

The structural integrity of the device has a direct impact on the surface scaling law of the forming event. From Weibull statistics and under the assumption that the failure rate of the weakest link problem will follow a power law (Eq. (4.1)) in analogy to the trend with charge-to-breakdown:

$$\lambda(V) = \xi V^{a} \tag{4.1}$$

the logarithm of the forming voltage will scale linearly with the logarithm of the surface:

$$\ln\left(V_{F|63\%}\right) = -\frac{1}{\beta} ln\left(\frac{A}{a_0^2}\right) + ln\left(\frac{1}{\xi^{1/\alpha}}\right), \text{with } \beta = \frac{a}{a_0} t_{ox} \tag{4.2}$$

Here β is the slope of the Weibit plot versus breakdown voltage, α_0 the mean distance between bulk defects in the oxide, t_{ox} the oxide thickness and ξ an appropriate constant.

Fitting the model to both delaminated and non-delaminated samples, allows to determine the degree of compliance to the theoretical case. The model is fitted directly versus MIM surface as any uncertainty will be incorporated to the linear regression intercept (4.2). The slope allows to determine the α/α_0 ratio for t_{ox} = 5nm. The linear regression model is fit using maximum likelihood estimation method. The results are presented in Table 4-2.

The quality of the fit is shown to be lower in the delaminated samples, indicating that the effective surface is not the nominal, as is expected owed to the crack propagation. Both frequentist (R^2 and p-value) and Bayesian (AIC for Akaike information Criterion and BIC for Bayesian Information Criterion) figures of merit are used to evaluate fit quality. In all cases the p-value statistic is statistically significant ($p<0.05$) denoting that the percolation formation process is still dominant in all cases. Even though the R^2 values do not seem to vary more than 5% overall, the AIC & BIC of non-delaminated devices are approximately 3-4 times smaller as compared to those of the delaminated devices: This signifies that *in relative terms* the nominal surface scaling is followed significantly better in non-delaminated devices. Considering that the statistical ensemble in both cases is the same, this observation is not a size effect but could be attributed to the non-correspondence of active surface to nominal surface of the device.

Table 4-2 Extracted parameters for Weibull breakdown statistics and quality of fit for every sample tested

Ti.th (nm)	**Morphology**	**β**	**α_0/α (Å)**	**R^2**	**p**	**logLikelihood**	**AIC**	**BIC**
5	Delam.	9.86	5.07	0.955	2.3E-02	8.5	-11.1	-12.9
10	Delam.	6.9	7.25	0.945	5.5E-03	7.5	-8.9	-10.1
5	Non-Delam.	35	1.43	0.985	9.2E-06	25.9	-45.9	-46.1
10	Non-Delam.	12.3	4.05	0.978	2.6E-05	17.2	-28.4	-28.5

Under the assumption of the invariance of the Weibull slope to the surface scaling (See ch. 1), Eq. (4.2) allows to calculate the expected value of the shape factor (slope) of the Weibit plot as well as the ratio of α/α_0 for every sample for t_{ox}= 5nm. Further parameter extraction is tricky owed to the inter-dependent nature of the α, α_0 and ξ. Nonetheless, under the assumption of α=1 the mean defect distance can be extrapolated for each sample variant, from the Weibull slope. Both α=1 and α=0.5 have been assumed in the literature

(see [174] and references therein). In practicality, the first assumption implies a linear response of the weakest-link failure rate λ to the applied field as compared to the square root dependence for the other scenario. Within the linear approximation the mean defect radius α_0 is calculated within the range of ~5-7 Å for the delaminated devices and ~1.4-4 Å for the non-delaminated ones.

The former samples show higher defect radius distance while the latter lower, consistent with the smaller breakdown voltages between the two. The 5 nm Ti samples show a slightly smaller defect radius with respect to the ones with 10 nm Ti. This could be explained if the HfO_2/Ti interface extension is taken into account. Nevertheless, the calculation of such an interface from the simple analytical model, as well as the parameter ξ is sensitive to the data the assumption for the acceleration factor α, the overall shape of the Weibit distribution etc. Hence, in my opinion this approach is useful to give practical understanding for the trend of breakdown, and potentially provide some useful comparisons in an order-of-magnitude fashion for the defect radius that could be used to compare with more accurate percolation path modeling.

Finally, it is worth to mention that the values of β calculated via the surface scaling law correspond to values that are within the range of fitted values by directly fitting the Weibit plots (Figure 4-6c, Figure 4-14c). Nonetheless, for the defective devices the surface-scaling fit approach gives Weibull shape factor values within the range of 7-9, more comparable to values of the smaller devices (Figure 4-6c) for both Ti thicknesses, while for structurally robust devices the values calculated are within the range of larger devices for either thickness. This could also be an argument in support of the impact of scaling of the Ti layer, both lateral and vertical in its overall chemical composition and electronic properties.

4.2.7.2 Cycling Characteristics

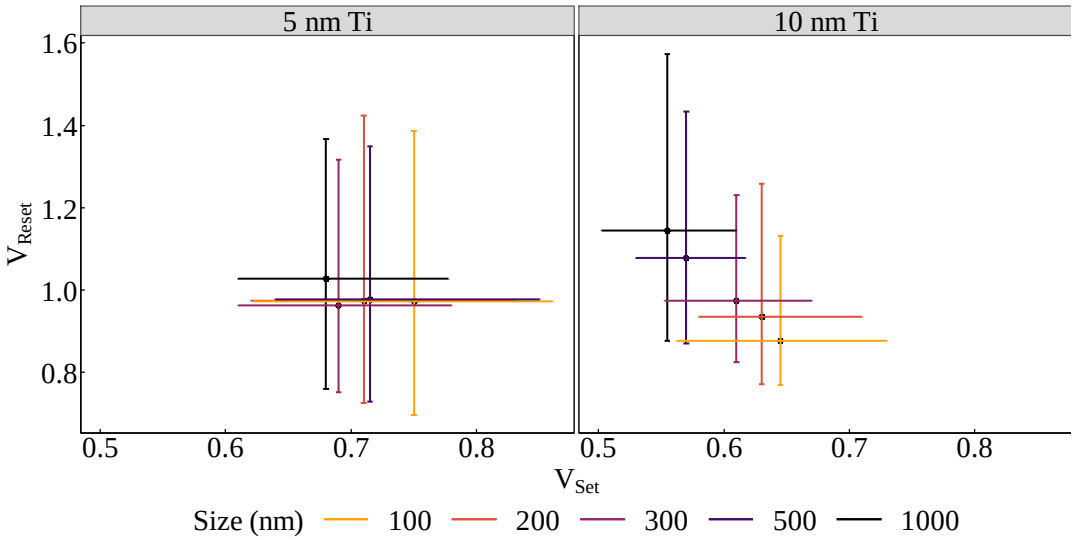


Figure 4-16 Median set and reset voltage (as defined in the text) for the 5 nm Ti (left) and 10 nm Ti (right) samples.

Post forming the devices are cycled using both DC and pulsed programming regimes. The set and reset voltage medians are plotted in relation to one another for all the device variants and all available sizes (Figure 4-16). In this experiment, the DC sweep is between -2V and 0V during the reset operation and between 0V and 2V for the set operation. Both voltages in this experiment are defined as the voltage where the current gradient over the swept voltage is maximized, in analogy to Figure 3-18. This is expected to better capture the critical juncture of filament formation and dissolution during set and reset respectively. The voltage device-to-device statistic is extracted for all devices on-wafer at the same programming cycle. The error bars correspond to the 10% and 90% quantiles of a 50 device statistic.

For both device variants [(5,10)nm Ti] the set voltage is found to decrease with increasing surface. This can be intuitively understood if the conduction properties of the high resistive state are taken into account. At HRS the percolation path is dissolved, hence conduction is volume dependent (filament dissolution). As a result, decreasing the MIM surface is expected to increase the required field to achieve of breakdown or the formation of a new percolation path. The reset voltage is weakly dependent on the device area in the case of the 5 nm Ti variants, while it is strongly dependent on area for the thicker (10 nm) Ti

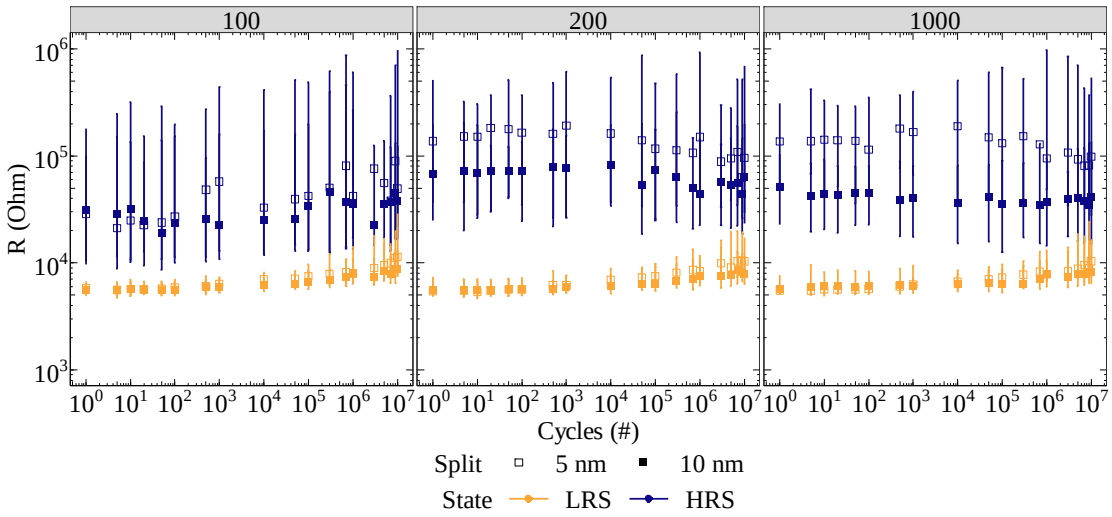


Figure 4-17 Endurance cycling of devices with critical dimension of 0.1 and 1 µm respectively for 5nm HfO_2 and the two device variants of 5 and 10 nm Ti scavenger. SET conditions: 1.3V Top electrode, I_{cc}=120 µA ; RESET: 1.5V Bottom Electrode, I_{cc} = 400 µA.

devices. Moreover the two quantities are found to be negatively correlated: Pearson correlation cor(v_{set},v_{reset}) is found to be -0.42 and -0.99 for the 5nm and 10nm Ti variants, respectively.

In Figure 4-17 the endurance characteristics of the previously studied across-wafer devices population is shown for the limiting cases of 1 µm and 100 nm device sizes. In a first order, the devices can be successfully cycled up to 10^7 times, while maintained median-to-median separation. The error bars represent the 10% to 90% normal quantiles of the studied population. As a primary observation it can be seen that the thinner Ti layer leads to a more efficient reset (higher HRS) for a given SET condition.

The device-to-device variance of the population tested is shown in Figure 4-18 for the devices with critical dimension of 100nm, 200nm and 1000nm. The increase of variability previously discussed in the statistics of the forming process is found to impact the cycling characteristics of the devices as well. The HRS of the devices with 10 nm Ti shows smaller dispersion than those with 5 nm for a given dimension. Although the HRS values of the 5nm variant are smaller than those of the 10 nm one in absolute values, this is not merely a scale effect as it appears in all sizes. More interestingly, the variability of LRS is impacted by the thickness of the Ti layer, in particular for large device sizes. Otherwise the cycle-to-cycle evolution of variance seems comparable in all cases.

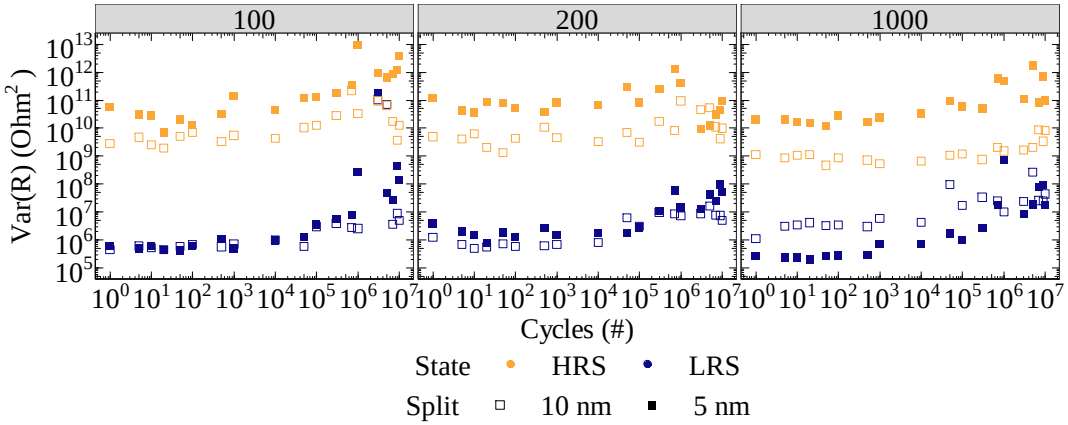


Figure 4-18 Evolution of the variance HRS,LRS for the on-wafer tested devices (50) as a function of device ageing for different surfaces, and Ti splits. The HfO_2 thickness is 5nm.

The dependence of the variance of each state to the thickness of Ti can be considered a direct impact of the sidewall interacting with the environment and the process variability associated to the sidewall surface. As HRS is mainly dependent of volume conduction, the effect is not size-dependent and thus appears in all device sizes. Only the variance of the LRS state is significantly impacted (even though the absolute value of the LRS is not as can be observed from Figure 4-17) pointing towards the increase of bimodality of the distribution of the statistical ensemble in the state. This is also supported by the sensitivity of the variance in both Ti thickness and cell size, denoting that the peripheral Ti sidewall surface plays a role. This property is not linked to intrinsic ReRAM effects but could be attributed to extrinsic factors, namely variability of the Ti passivation properties (nitrated Ti) as well as the variability of process steps link dry etching, stripping and / or space deposition. It does serve however to show that even under an optimized process and integration scheme, a standard ReRAM cell geometry is very sensitive to extrinsic effects.

The observed behavior is quantified in Figure 4-19 in a more designer-friendly figure-of-merit: The Memory Window (MW) for after 100 (early wake-up) and 10^7 cycles is traced, using a similar method as the one described by Eq. 3.13. Considering the significantly smaller statistical ensemble studied in this case as compared to the 4kbit array devices shown in Chapter 3, the MW is traced between the lower 25% limit of the CDF corresponding to the HRS and the upper 75% limit of the CDF corresponding to the LRS, describing the tail behavior and the median-to-median.

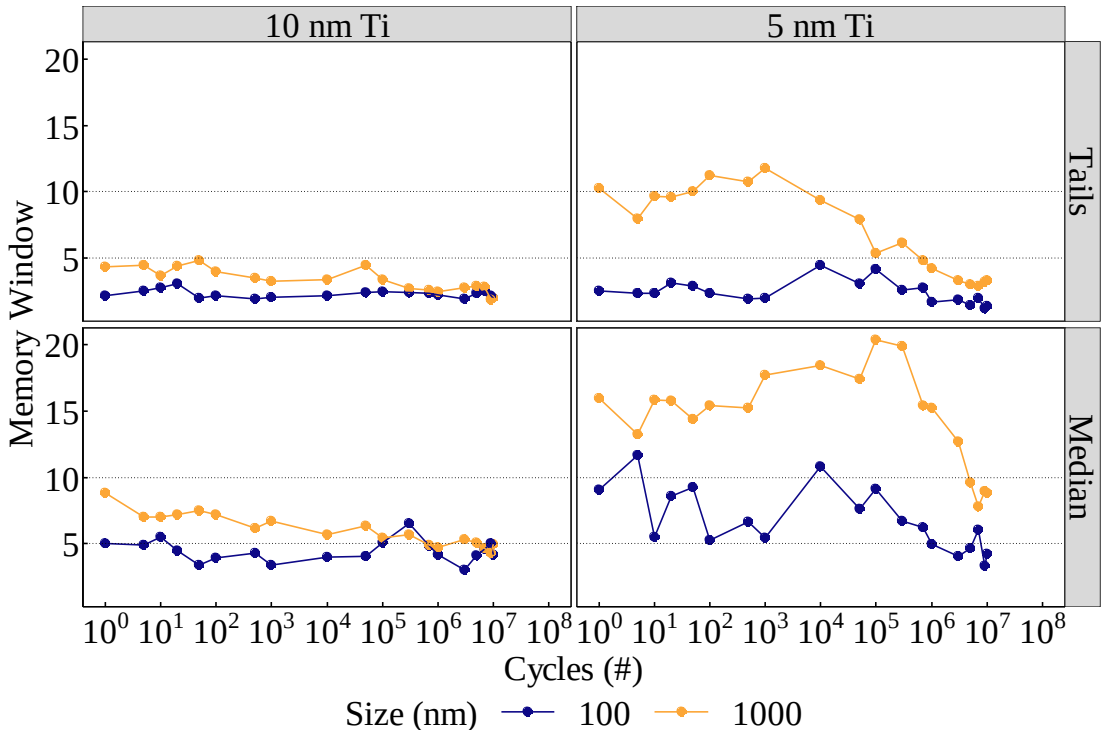


Figure 4-19 Memory window for median-to-median and tails [HRS(25%)/LRS(75%)] as defined in chapter 3.

The memory window calculated in this approach is found to be bigger for the 5nm Ti structures as compared to the 10 nm ones both for the median and tail memory windows under the same programming conditions. The trend is more important when larger structures are considered. Finally, the same calculation after cycling denotes a noticeable degradation of the devices.

The impact of ageing of the memory window metric is shown in Figure 4-20. It can immediately be noticed that for the given programing conditions the 10 nm Ti thickness acts unfavorable for the memory window irrespective of the cell size or cycling. This could be owed to a failure to properly reset the cell as oxygen migration in the thicker Ti layer is expected to be more facilitated. Hence, a stronger reset is expected. Nevertheless, a stronger reset cannot be easily achieved owed to the aggressive scaling of the transistor (nFET) which leads to the source-drain voltage-transfer characteristics to quickly saturate. Thus, operation at lower current would be the only solution at the expense of more erratic bits appearing in the LRS upper tail.

For the 5 nm Ti devices an optimum memory window can be achieved ranging from 17 in the median of larger devices and degrading to values as low as 3-4 in the MW_{tail} of 100

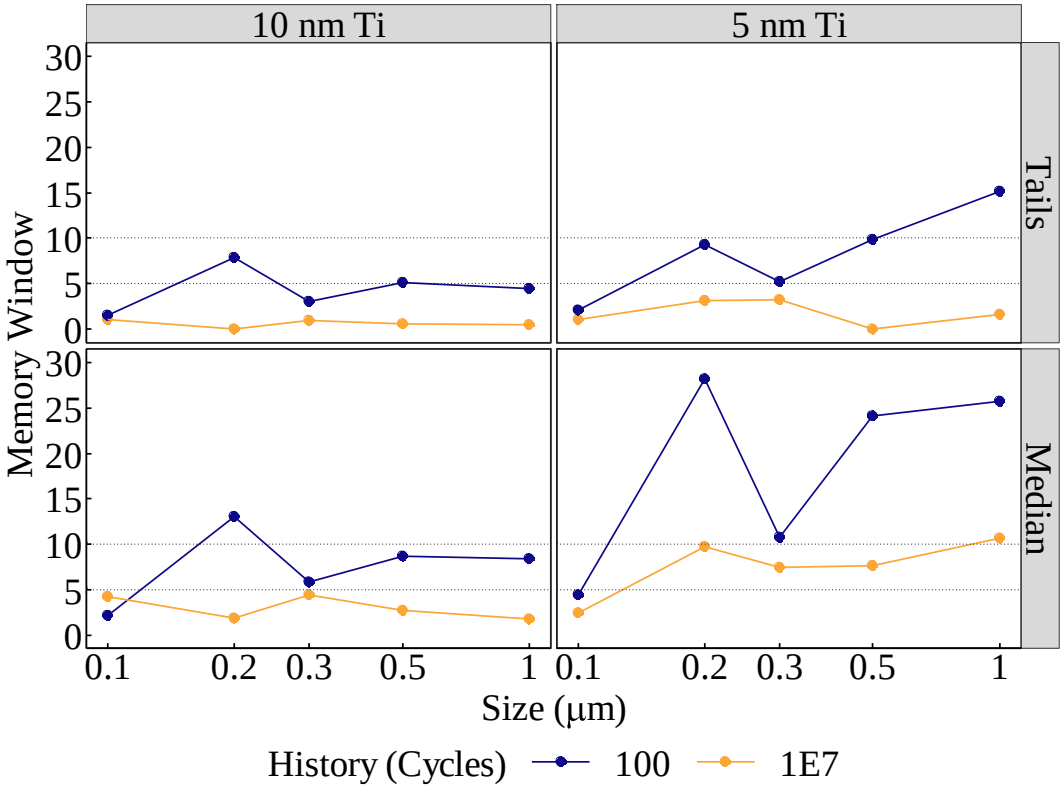


Figure 4-20 Evolution of the Memory window (MW) of 5nm HfO_2 based ReRAM devices as a function of cycling for 10nm and 5nm Ti layer, median-to-median and at the tail MW (as defined in the text) for fresh cells.

nm devices. The given stack seems to be optimally matched to the given nFET voltage transfer capacity. The intrinsic cell degradation according to memory window median occurs after 10^6 cycles while according to the 25% HRS – 75% LRS tail MW at as early as 10^5 cycles. Even though large devices start with higher memory windows, they still degrade at those limits, indicating an intrinsic cell failure arising either from material ageing or electrostatic modifications owed to the defect density, while smaller devices present with weaker window characteristics throughout their lifetime. Nevertheless, they do not show any visible degradation if only he beginning and end of the life time of the devices is taken into account. This could be related to a volume effect: The larger the device the more net charge is generated with cycling through incomplete dissolution of the randomly shaped percolation path at every cycle (proportional to oxide volume and number of programming attempts).

4.3 Conclusions

HfO_2 based ReRAM devices have been integrated with FDSOI nFETs in a 300 mm route. Two variants have been demonstrated based on the Ti layer thickness and covering multiple cell sizes. The process route has experienced multiple evolutions of process maturity, namely as regards the structural integrity of the cell. It is observed that PVD deposited TiN exhibits significant levels of compressive stress in the as-deposited layer which can be catalytic to the propagation of nano scaled cracks from the sidewall towards the center of the cell leading to interface separation and delamination. In structurally non-integer devices, the forming voltage behavior as well as cycling variability is heavily impacted from the propagation of sidewall defects, despite the percolative nature of the conduction in ReRAM. This shows the sensitivity of the device in the environment characteristics.

Passivation of the Ti sidewall leads to structurally robust devices without the requirement to relax the stress. Bow-evolution measurements before and after thermal annealing show a stress reversal in the TiN/Ti/HfO_2 system for different TiN deposition conditions, which could indicate that mechanical stress is possibly an intrinsic property of the HfO2/Ti ReRAM system, possibly associated to interface mixing; as has been previously theorized for ReRAM devices [175].

In terms of electrical operation, significant tradeoffs are observed in terms of the Ti-thickness related variants. Even though a thicker Ti layer reduces the forming operation requirements, it leads to a stronger size-dependency of the forming voltage and increased forming variability. Moreover, in the long-term, ReRAM cells with thick Ti layers are harder to reset and show degraded memory window as well as higher variance for all sizes in HRS and for larger devices in LRS. The latter impact of thick Ti layer in the variability of states has also been reported by separate groups recently [176]. This makes them hard to integrate with advanced nFETs owed to the HfO_2 cell bipolarity (imposing a reset polarization through the nFET source node) one the one side, and the nFETs' intrinsic inability to transfer a "clear" 1 on the other irrespective of cell size. Moreover, for any practical applications this will require very strong error correction inferring high surface toll for the ECC circuit on the product level. As a result, simple thickness optimization of the reactive electrode is probably not sufficient to bring ReRAM to application level maturity.

4.4 Future Prospects

In chapter 3 we demonstrated a large-scale BEOL integration route in a 200 mm technology, using it as a test vehicle to benchmark standard HfO_2 based ReRAM with the LSI implantation process. In this chapter, we show the different steps to transfer a similar process in a 300 mm test vehicle. Despite the difference in technology node and the inherent advantages in process control a 300 mm process presents against a 200 mm process, as well as the effort made for process engineering on the standard cell level, no particular amelioration on the electrical performance metrics was observed between pure HfO_2 devices in both generations. It can even be argued that the merging of ReRAM technology to advanced CMOS, imposes strict limitations in the current levels and capability of transferring voltage owed to the scaling of the CMOS node in the 1T1R structure – especially if the memory array is optimized for density. Hence, Si implanted HfO_2 devices do present with given advantages in the freedom; of the approach to engineer both cell material and cell geometry at the same time owed to the flexibility of the implantation process.

Nevertheless, the approach as developed in the previous chapters necessitates an additional lithography step to define the locally implanted area adding to cost. In the integration described in this chapter, the memory element is fully aligned to the contact via. This can provide an interesting flexibility in terms of defining the implantation process. Namely, the integration can be modified in several ways such that the implant does not impose additional lithographic steps:

In one implementation, the deposition process of Ti and TiN can be skipped until the opening of the top contact via. Via opening will then expose the HfO_2 layer below allowing the implant surface to be fixed by the via itself. Following that, Ti/TiN can be deposited inside the via with varying thicknesses. This is by no means blocking as for 65 nm MOL/BEOL processes and beyond the standard W via filling process involves the deposition of a 5 nm Ti adhesion layer followed by 5-10 nm TiN liner before W filling takes place. As such, the materials required for the ReRAM element operation are a natural part of the BEOL integration flow and pose no major roadblocks. In another implementation the full stack of TiN/HfO_2/Ti/TiN can be deposited and patterned, and the implant can be introduced again right after the via opening step. In this approach the implant is done using the TiN/Ti top electrode bilayer as a hard mask. This approach can

be interesting to help tune the implant profile so that only the Si implant tail penetrates the HfO_2.

In both cases certain pros and cons in both concepts can be found: On the one side, the reduction of lithography steps poses a strong argument. On the other hand, the top via misalignment margin starts to become important for devices whose critical dimension approaches that of the via. In this case, for fully defined MIM capacitors the impact of the proximity of the implanted zone to the MIM sidewall on the electrical behavior of the devices needs to be evaluated in large populations to verify the minimum limits where the confinement effect can be sustained. A solution to this could be to completely avoid etching the HfO_2 layer, at the expense of further constrictions in terms of cross-contamination inside the 300 mm integration route. Both approaches, could not be fully explored within the time-limits of this thesis.

CHAPTER 5. CONCLUSIONS AND PERSPECTIVES

In the previous chapters the HfO_2 ReRAM system has been studied in detail. Combining first principle calculations to physical characterization we have proposed a mechanism for the impact of Si and Al doping in the material nanostructure and its electronic properties in Chapter 2. The identified increase of conductivity with Si doping was then exploited to propose a new integration approach on the device level exploiting local ion implantation. This has led in lowering of forming while achieving superior bit error rates and lowering the extrinsic components of variability. Nevertheless, it comes with a small tradeoff in maximum endurance and ultimately maximum operating temperature and shorter data retention of the LRS as is described in detail chapter 3.

From a device perspective, Si implantation is an interesting variant for integration, notably to relax the high forming constrictions while allowing to reduce extrinsic variability that is particularly hard to control in the HfO2/Ti system. Potentially, its most interesting aspect is that it can allow to find a good tradeoff between high endurance, data retention and forming constraints. While it is true that data retention is lower for the highly doped LSI devices as shown in Chapter 3, it is worth noting that still the doped devices with 10nm HfO_2 and local Si doping are still much better in terms of thermal stability than their 5nm HfO_2 counterparts fabricated with the same process in an identical test vehicle [Cecile, G. Molas 5 nm MAD]. This means that the reduction of the energy barrier induced by doping can be partially mitigated by the enlargement of the overall diffusion path of the oxygen vacancies – a direct consequence of the use of a thicker oxide in the device stack. Such topic merits detailed optimization and study and goes beyond the targets and constraints imposed by a PhD thesis.

Chapter 4 deals with a full 300mm integration where the HfO_2 ReRAM where the compatibility of the stack with an advanced node is demonstrated and the effort was focused in approaching the FEOL in a way that can be tolerated by the constraints of a CMOS fab. Maybe the most interesting aspect of this part of the work is in my opinion the statistical evidence that the Ti layer thickness; which is one of the two main knobs to tune ReRAM characteristics seems to be impacted by ReRAM environment: In chapter four we have demonstrated the sensitivity of forming variability with the Ti thickness layer as well as the acceleration of the memory ageing process with increasing Ti layer thickness. This

implies a non-negligible interaction of This was supported by the statistical variability of forming as well as the sensitivity of the variance of the LRS and HRS states to the thicker Ti layers and larger device sizes. This stresses even more the need for the development of a dedicated ReRAM liner material capable to remain inert across the ReRAM fabrication process and lifetime. Even more so, it shows that a simple thickness tuning cannot suffice to significantly improve the technology if more than one parameters are considered: A thicker Ti layer will result in lower forming voltages but makes the device more sensitive to higher variability possibly induced by interaction with the material and is expected to reduce overall endurance and data retention. On the other hand, replacing the SiN liner is a non-trivial problem, for lack of materials that are both insulating, can block oxygen and are compatible with CMOS integration.

With respect to the work presented in this thesis there is certainly a large margin for further work. In particular, from a material and device prospective even though material reconstruction through thermal annealing was demonstrated, for all Si implanted films used in this work, only a single implant dose sate has been studied. Varying the implant doserate, implantation tilt and twist angles and observing the charge generation / film stoichiometry and device pre-forming I-V characteristics can allow to understand in a quantitative fashion the defect generation mechanism. This is crucial to allow for process and device simulation in order to predictively tailor the oxide conduction characteristics. Of course this is a non-trivial matter. The challenge in introducing the appropriate alloying and/or defect generation model in TCAD, which would allow in a comprehensive device simulation, lies in the multiple phases coexisting in the HfO_2 film as the XRD results presented in chapter 2 indicate.

From an experimental standpoint, there are several unexplored parameters: First, surface scaling of the locally implanted devices and the eventual limits have yet to be fully explored. For that matter, it is important to not only explore the device sensitivity to surface scaling of the implanted area, but also the impact of scaling both implanted zone area wrt the total MIM stack area. The A_{LSI}/A_{MIM} ratio can be particularly important for identifying the scaling rules of the technology; both in terms of suppression of the extrinsic effects as well as in terms of potential intrinsic scaling limits. Secondly, scaling the thickness of the Si implanted HfO_2 upwards should be explored in trying to find an optimal tradeoff between forming characteristics, endurance and data retention.

Moreover, as the results of HfO2/Ti devices indicate, the Ti layer might not the best candidate for the HfO_2 ReRAM stack owed to the difficulty to control extrinsic effects. At the same time, while the locally Si implanted devices are more robust in terms of error evolution, the eventual tradeoff of the tail endurance post 10^5 cycles and temperature sensitivity are inhibiting factors for some applications. As it has been already shown [124], replacing the Ti reactive electrode with Hf can improve endurance and retention characteristics and could also be beneficial for the case of HfO_2:Si/Hf stacks.

In all solutions studied so far, the ReRAM element is always in the form of a MIM capacitor in the BEOL or in the early MOL. Nevertheless, as recent works have demonstrated, it is achievable to use the gate of 14nm finFETs as the switching layer, even though it is unknown if further modifications in the process are required (such as optimization of the HfO_2 synthesis process or addition of an oxygen scavenging layer). As part of the current PhD work, an effort was made to evaluate the potential of the FDSOI 28nm node (FD28) as a resistive switching element [P-3]. It has to be noted that there are major differences between the two aforementioned nodes. In particular, the gate oxide of an FDSOI 28nm device is a HfSiON ISSG (in situ Stream Generation) oxide while the finFET process is probably a sub-nm ISSG SiO_2 followed by an ALD deposited HfO_2. In this case study, the FD28 devices evaluated have a highly doped channel targeting to enhance the local electric field in the pn junction region as well as enhance the contact size between gate and the pn junctions. This is expected to increase the gate-to-junction current.

Applying bias between the gate and drain/source of the channel doped devices allows to successfully break down the oxide (Figure 5-1) while by reversing the polarity the effect can be reversed leading to a forming followed by a RESET/SET like behavior. Unfortunately, it is particularly hard to sustain resistive switching for more than a few cycles. It is unknown at this stage if the formation of the percolation path occurs in the HfSiON layer or in the region where the spacer overlaps with the junction. Even though the SiN spacers of the FD28 technology are more than 20nm thick and comprise of a silicon nitride with few electrical defects, there is not enough data to exclude the scenario of a partial spacer breakdown. This acts to show that at least for a FEOL implementation in a node generation such as the FD28.

Starting from the fact that the highly doped substrate can facilitate the gate-to junction conduction we propose a 2T2R complimentary memory cell integrated in the FEOL. On

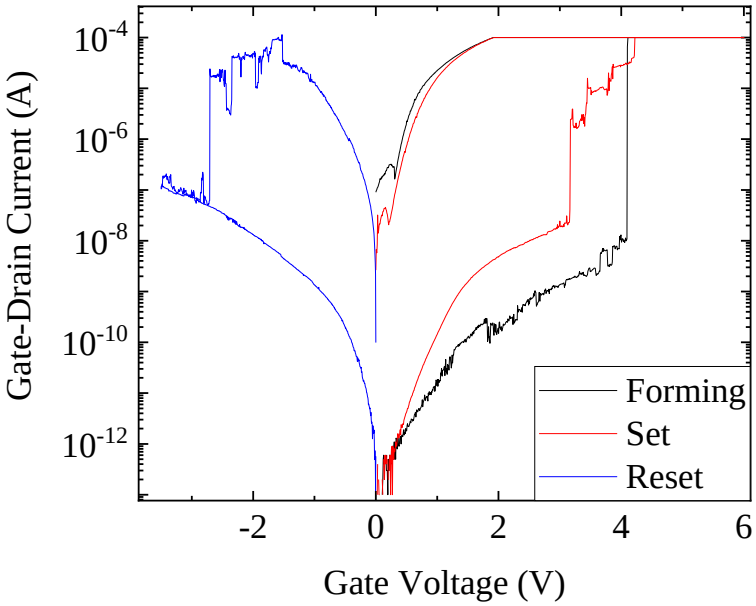


Figure 5-1 Initial breakdown (forming), breakdown recovery (Reset) and recurring breakdown (set) of an FD28 gate stack comprising of HfSiON. During the forming and set operations the source node is left floating, the drain node is set on ground while positive voltage is applied on the gate. The polarity of the gate and drain nodes are reversed for reset to happen.

the one hand the 2T2R architecture is known to allow the reduction of variability in SRAMs and similar arguments can be applied for ReRAM 2T2R architectures as long as the two elements are programmed in LRS and HRS respectively [Portal, paper]; which is expected to be beneficial for the overall state variability as the two varying ReRAM devices are expected to partially cancel each other out.

On the other hand, the cell could ideally be integrated using the same mask as the one defining the transistor gate; provided that a suitable tradeoff in the gate oxide characteristics can be identified. The proposed structure circuit is shown in Figure 5-2. We propose to exploit the dead space of the STI region between the two transistor active zones adding a dummy polysilicon gate stack as shown in Figure 5-2b. In this configuration the select line (SL) can be merged on the dummy poly-Si acting as the gate top electrode the transistor gates can act as the word line (WL) and the junctions as the complimentary bit lines. A proposed mode of operation for the different phases for forming, set and reset is shown in Table 5-1.

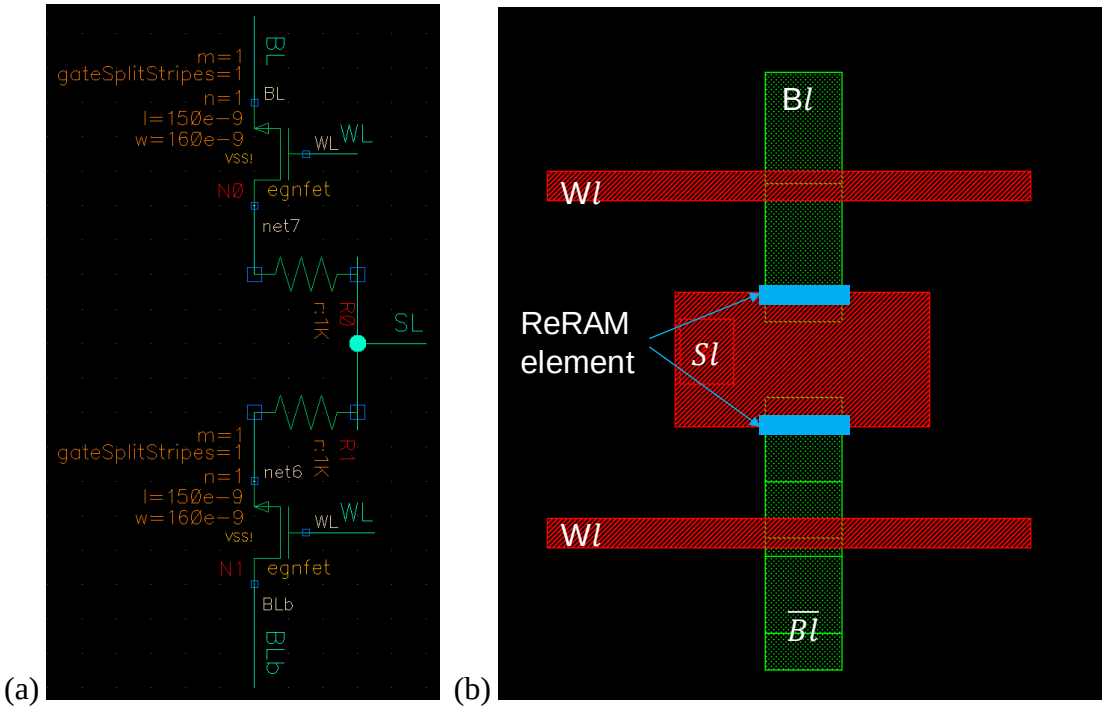


Figure 5-2 (a) Circuit schematic of a 2T-2R structure (b) 2T-2R proposed structure layout using the FD28 design rules. The ReRAM element is integrated using POLY layer of the gate stack patterned on the STI between two nodes.

This type of layout can then be arranged in different matrix topologies according to the constraints of density and area in the design kit. In one vertical SL matrix configuration the 2T2R cell size reaches down to 0.18um^2 or 0.09 um^2 if the cells are operated in a 1T1R configuration (Figure 5-3). It is interesting to note that the limitations of cell density do not come from the FEOL transistor part but mainly from the relaxed pitch of M1 to M3. Here minimal dimension structures capable of providing 120 – 200 µA of current have been used as the complimentary BL device is expected to me more robust in terms of state variability readout. In fact, high variability between the two cells could be preferred This particular matrix design trades off the capacity for parallel operation between the cells in Wl_n and $Wln_{\pm1}$ to maximize cell density while at the same time maintaining the overall design aspect ratio as close to one as possible. Moreover, it is worth noting that this design variant is fully compliant to the complete set of critical DRCs. That is to say that if specific ReRAM friendly design rules were to be implemented cell density and design flexibility could further increase. This is often the case in memory design where SRAM design rules are complete different set to digital and analog design rules.

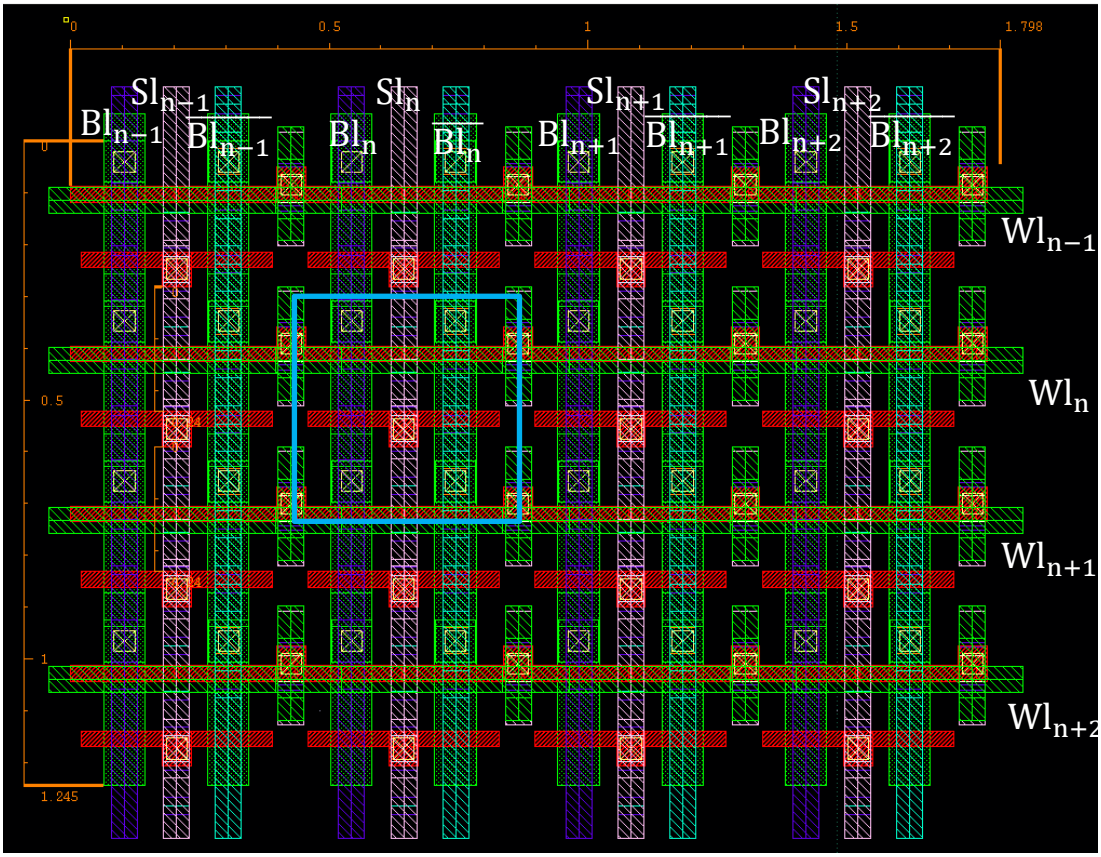


Figure 5-3 Example vertical SL arrangement of a memory array configuration exploiting dummy poly fingers on STI for the ReRAM element. Cell and array layout have been optimized for density using minimal thin oxide (GO1) FD28 transistor sizes.

Even though such designs have been integrated in exploratory tapeouts it was not possible for the printed circuit to be evaluated within the time constraints of this work. Moreover, for the proposed technology to function in a reliable and stable fashion, it is possible that a specific re-engineering of the gate stack is required that is no longer limited in the field enhancement in the Silicon crystal, but also acts to find a good trade-off between transistor reliability and reversible oxide switching. It is evident that such ReRAM-friendly FEOL integration approaches will partially compromise leakage current through the gate oxide and possibly the overall gate stack reliability metrics (DIBL, GIDL). As a result, they are inherently incompatible with multi-project fabrication cycles and would require a dedicated memory route inside the fab as well as dedicated memory products. Although the possibilities can be very alluring, the hard truth is that a true FEOL integration of ReRAM is still a highly challenging endeavor.

Table 5-1 Proposed operating configurations for a 2T2R cell in a vertical SL NAND memory array configuration. A two phase forming is proposed to overcome the stochasticity of the parallel cell configuration during this operation.

	FORMING		**RESET**	**SET**	**READ**
	Phase 1	Phase 2			
SL	gnd		gnd	Vset	gnd
WL	Vcomp		> Vcomp	Vcomp	Vread
BL	Vform	gnd	gnd/Vreset	gnd	Vdd/2 / gnd
$\overline{\mathbf{BL}}$	gnd	Vform	Vreset/gnd	gnd	Vdd/2 / gnd

On the other hand, it is also possible that the HfSiON layer used in the flavor of FD28 that was used for evaluation here is inherently unfavorable for repeatable resistive switching. Nevertheless, as we move to more advanced process flavours or more aggressive technology nodes this layer is replaced by an 11A thick ISSG SiO_2 followed by ALD deposited HfO_2. This latter approach could be more favorable for resistive switching if the Si/HfO_2 interface trap levels are increased, to facilitate the oxide breakdown. That said, it is evident that a FEOL integrated device is a highly challenging problem and there is still a lot of room for investigation. Nevertheless, the high cost of R&D in sub 20nm nodes makes research in the topic prohibitive.

As is often the case with new technologies, ReRAMs are slowly exiting the trough of disillusionment of the Gartner Hype cycle. I strongly hope that, as has been the case of the CMOS, this promising technology will slowly, but surely find it path through systematic and persistent cycles of research and development.

BIBLIOGRAPHY

[1] "Scaling Survival Guide in the More-than-Moore Era," in *International Electron Devices Meeting (IEDM)*, 2018.

[2] F. Masuoka, M. Momodomi, Y. Iwata, and R. Shirota, "New ultra high density EPROM and flash EEPROM with NAND structure cell," in *1987 International Electron Devices Meeting*, 1987, pp. 552–555.

[3] T. Kawahara, K. Ito, R. Takemura, and H. Ohno, "Spin-transfer torque RAM technology: Review and prospect," *Microelectron. Reliab.*, vol. 52, no. 4, pp. 613–627, 2012.

[4] B. H. P. Wong *et al.*, "Metal – Oxide RRAM," *Proc. IEEE*, vol. 100, no. 6, pp. 1951–1970, 2012.

[5] J. S. Meena, S. M. Sze, U. Chand, and T.-Y. Tseng, "Overview of emerging nonvolatile memory technologies.," *Nanoscale Res. Lett.*, vol. 9, no. 1, p. 526, 2014.

[6] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, and H.-S. P. Wong, "A Low Energy Oxide-Based Electronic Synaptic Device for Neuromorphic Visual Systems with Tolerance to Device Variation," *Adv. Mater.*, vol. 25, no. 12, pp. 1774–1779, Mar. 2013.

[7] E. Covi, S. Brivio, A. Serb, T. Prodromakis, M. Fanciulli, and S. Spiga, "HfO2-based memristors for neuromorphic applications," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2016, pp. 393–396.

[8] Z. Wang, S. Ambrogio, S. Balatti, and D. Ielmini, "A 2-transistor/1-resistor artificial synapse capable of communication and stochastic learning in neuromorphic systems ," *Frontiers in Neuroscience* , vol. 8. p. 438, 2015.

[9] D. Ielmini, S. Ambrogio, V. Milo, S. Balatti, and Z.-. Wang, "Neuromorphic computing with hybrid memristive/CMOS synapses for real-time learning," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2016, pp. 1386–1389.

[10] F. Arnaud *et al.*, “Truly Innovative 28nm FDSOI Technology for Automotive Micro-Controller Applications embedding 16MB Phase,” in *International Electron Devices Meeting (IEDM)*, 2018, pp. 424–427.

[11] D. Apalkov *et al.*, “Spin-transfer Torque Magnetic Random Access Memory (STT-MRAM),” *J. Emerg. Technol. Comput. Syst.*, vol. 9, no. 2, p. 13:1--13:35, 2013.

[12] EETimes, “Samsung Says It’s Shipping 28-nm Embedded MRAM,” 2019. [Online]. Available: https://www.eetimes.com/document.asp?doc_id=1334410#.

[13] L. Chua, “Memristor - The missing circuit element,” *IEEE Trans. Circuit Theory*, vol. 18, no. 5, pp. 507–519, 1971.

[14] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, “Nanoscale memristor device as synapse in neuromorphic systems,” *Nano Lett.*, vol. 10, no. 4, pp. 1297–1301, 2010.

[15] L. Chua, “Resistance switching memories are memristors,” *Appl. Phys. A Mater. Sci. Process.*, vol. 102, no. 4, pp. 765–783, 2011.

[16] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, “The missing memristor found,” *Nature*, vol. 453, no. 7191, pp. 80–83, 2008.

[17] PANASONIC, “ReRAM embedded Super low-power consumption MCU MN101L.” [Online]. Available: https://industrial.panasonic.com/ww/products/semiconductors/microcomputers/mn101l.

[18] Fujitsu, “4 Mbit ReRAM ‘MB85AS4MT.’”

[19] M. Lepedus, “What happened to ReRAM,” *Semiengineering*, 2017. [Online]. Available: https://semiengineering.com/what-happened-to-reram/.

[20] Z. Zhang *et al.*, “All-Metal-Nitride RRAM Devices,” *IEEE Electron Device Lett.*, vol. 36, no. 1, pp. 29–31, 2015.

[21] H.-D. Kim *et al.*, “Large resistive-switching phenomena observed in Ag/Si3N4/Al memory cells,” *Semicond. Sci. Technol.*, vol. 25, no. 6, 2010.

[22] B. J. Choi, A. B. K. Chen, X. Yang, and I.-W. Chen, "Purely Electronic Switching with High Uniformity, Resistance Tunability, and Good Retention in Pt-Dispersed SiO2 Thin Films for ReRAM," *Adv. Mater.*, vol. 23, no. 33, pp. 3847–3852, Sep. 2011.

[23] F.-T. Chin *et al.*, "Switching characteristics in Cu:SiO2 by chemical soak methods for resistive random access memory (ReRAM)," *Solid. State. Electron.*, vol. 103, pp. 190–194, 2015.

[24] H. Schroeder and D. S. Jeong, "Resistive switching in a Pt/TiO2/Pt thin film stack – a candidate for a non-volatile ReRAM," *Microelectron. Eng.*, vol. 84, no. 9, pp. 1982–1985, 2007.

[25] F. Lentz, B. Roesgen, V. Rana, D. J. Wouters, and R. Waser, "Current Compliance-Dependent Nonlinearity inTiO2 ReRAM," *IEEE Electron Device Lett.*, vol. 34, no. 8, pp. 996–998, 2013.

[26] M. Zangeneh and A. Joshi, "Design and Optimization of Nonvolatile Multibit 1T1R Resistive RAM," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 22, no. 8, pp. 1815–1828, 2014.

[27] B. J. Choi *et al.*, "Resistive switching mechanism of TiO 2 thin films grown by atomic-layer deposition," *J. Appl. Phys.*, vol. 98, no. 3, 2005.

[28] U. Russo, D. Ielmini, C. Cagli, and A. L. Lacaita, "Filament Conduction and Reset Mechanism in NiO-Based Resistive-Switching Memory (RRAM) Devices," *IEEE Trans. Electron Devices*, vol. 56, no. 2, pp. 186–192, 2009.

[29] K. Szot, R. Dittmann, W. Speier, and R. Waser, "Nanoscale resistive switching in SrTiO3 thin films," *Phys. Status Solidi - Rapid Res. Lett.*, vol. 1, no. 2, pp. 86–88, 2007.

[30] Y. Watanabe *et al.*, "Current-driven insulator-conductor transition and nonvolatile memory in chromium-doped SrTiO3 single crystals," *Appl. Phys. Lett.*, vol. 78, no. 23, pp. 3738–3740, 2001.

[31] D. S. Jeong *et al.*, "Emerging memories: resistive switching mechanisms and

current status," *Reports Prog. Phys.*, vol. 75, no. 7, p. 076502, 2012.

[32] Y. Wu, B. Lee, and H.-. P. Wong, "Al2O3-Based RRAM Using Atomic Layer Deposition (ALD) With -1μA RESET Current," *IEEE Electron Device Lett.*, vol. 31, no. 12, pp. 1449–1451, 2010.

[33] M. Azzaz *et al.*, "Endurance/Retention Trade Off in HfOx and TaOx Based RRAM," in *2016 IEEE 8th International Memory Workshop (IMW)*, 2016, pp. 1–4.

[34] S. Long *et al.*, "Cycle-to-Cycle Intrinsic RESET Statistics in HfO2-Based Unipolar RRAM Devices," *IEEE Electron Device Lett.*, vol. 34, no. 5, pp. 623–625, 2013.

[35] L. Goux *et al.*, "On the Gradual Unipolar and Bipolar Resistive Switching of TiN\HfO2\Pt Memory Systems," *Electrochem. Solid-State Lett.*, vol. 13, no. 6, p. G54, 2010.

[36] C. Cagli *et al.*, "Experimental and theoretical study of electrode effects in HfO2 based RRAM," in *2011 International Electron Devices Meeting*, 2011, p. 28.7.1-28.7.4.

[37] B. Govoreanu *et al.*, "10×10nm 2 Hf/HfO x crossbar resistive RAM with excellent performance, reliability and low-energy operation," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2011, pp. 729–732.

[38] Y. Kim *et al.*, "Bi-layered RRAM with unlimited endurance and extremely uniform switching," in *2011 Symposium on VLSI Technology - Digest of Technical Papers*, 2011, pp. 52–53.

[39] M. Alayan, E. Vianello, B. De Salvo, L. Perniola, A. Padovani, and L. Larcher, "Correlated Effects on Forming and Retention of Al Doping in HfO2-Based RRAM," *IEEE Des. Test*, vol. 34, no. 3, pp. 23–30, 2017.

[40] Y. Song, B. Magyari-Köpe, Y. Lin, and Y. Nishi, "Improving the High Resistance State Retention Degradation of Al-doped HfOx Based on Ab Initio Simulations," in *2015 IEEE 7th International Memory Workshop, IMW 2015*, 2017, pp. 0–3.

[41] X. Huang *et al.*, "HfO_2/Al_2O_3 multilayer for RRAM arrays: a technique to

improve tail-bit retention," *Nanotechnology*, vol. 27, no. 39, p. 395201, 2016.

[42] D. Duncan, B. Magyari-Köpe, and Y. Nishi, "Hydrogen doping in HfO2 resistance change random access memory," *Appl. Phys. Lett.*, vol. 108, no. 4, p. 43501, 2016.

[43] B. M. Long, S. Mandal, J. Livecchi, and R. Jha, "Effects of Mg-Doping on HfO2-Based ReRAM Device Switching Characteristics," *IEEE Electron Device Lett.*, vol. 34, no. 10, pp. 1247–1249, 2013.

[44] H. X. and Q. L. and Y. L. and H. L. and M. W. and X. L. and H. S. and X. Y. and S. L. and S. L. and M. Liu, "Nitrogen-induced improvement of resistive switching uniformity in a HfO 2 -based RRAM device," *Semicond. Sci. Technol.*, vol. 27, no. 12, p. 125008, 2012.

[45] H. Xie *et al.*, "Resistive switching properties of HfO2-based ReRAM with implanted Si/Al ions," *AIP Conf. Proc.*, vol. 1496, no. 1, pp. 26–29, Nov. 2012.

[46] H. Jiang and D. A. Stewart, "Using Dopants to Tune Oxygen Vacancy Formation in Transition Metal Oxide Resistive Memory," *ACS Appl. Mater. Interfaces*, vol. 9, no. 19, pp. 16296–16304, May 2017.

[47] D. Duncan, B. Magyari-Köpe, and Y. Nishi, "Properties of Dopants in HfOx for Improving the Performance of Nonvolatile Memory," *Phys. Rev. Appl.*, vol. 7, no. 3, pp. 1–10, 2017.

[48] M. Wu, Y. I. Alivov, and H. Morkoç, "High-κ dielectrics and advanced channel concepts for Si MOSFET," *J. Mater. Sci. Mater. Electron.*, vol. 19, no. 10, pp. 915–951, May 2008.

[49] S. Selberherr and A. Gehring, "Tunneling Models for Semiconductor Device Simulation," in *Handbook of Theoretical and Computational Nanotechnology*, vol. 10, 2006, pp. 469–543.

[50] F. Chiu, "A Review on Conduction Mechanisms in Dielectric Films," *Adv. Mater. Sci. Eng.*, vol. 2014, 2014.

[51] N. F. Mott, "Conduction in non-crystalline materials," *Philos. Mag. A J. Theor. Exp. Appl. Phys.*, vol. 19, no. 160, pp. 835–852, Apr. 1969.

[52] N. Apsley and H. P. Hughes, “Temperature-and field-dependence of hopping conduction in disordered systems, I,” *Philos. Mag. A J. Theor. Exp. Appl. Phys.*, vol. 30, no. 5, pp. 963–972, Nov. 1974.

[53] N. Apsley and H. P. Hughes, “Temperature- and field-dependence of hopping conduction in disordered systems, II,” *Philos. Mag.*, vol. 31, no. 6, pp. 1327–1339, 1975.

[54] L. Niemeyer, L. Pietronero, and H. J. Wiesmann, “Fractal Dimension of Dielectric Breakdown,” *Phys. Rev. Lett.*, vol. 52, no. 12, pp. 1033–1036, 1984.

[55] J. H. Stathis, “Percolation models for gate oxide breakdown,” *J. Appl. Phys.*, vol. 86, no. 10, p. 5757, 1999.

[56] B. P. Linder, S. Lombardo, J. H. Stathis, A. Vayshenker, and D. J. Frank, “Voltage dependence of hard breakdown growth and the reliability implication in thin dielectrics,” *IEEE Electron Device Lett.*, vol. 23, no. 11, pp. 661–663, 2002.

[57] E. J. Gumbel, *Statistics of Extremes*. 1958.

[58] J. Suñé, I. Placencia, N. Barniol, E. Farrés, F. Martín, and X. Aymerich, “On the breakdown statistics of very thin SiO2 films,” *Thin Solid Films*, vol. 185, no. 2, pp. 347–362, 1990.

[59] R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, and H. E. Maes, “A consistent model for the thickness dependence of intrinsic breakdown in ultra-thin oxides,” in *Proceedings of International Electron Devices Meeting*, 1995, pp. 863–866.

[60] R. Degraeve *et al.*, “New insights in the relation between electron trap generation and the statistical properties of oxide breakdown,” *IEEE Trans. Electron Devices*, vol. 45, no. 4, pp. 904–911, 1998.

[61] E. Y. Wu, J. Suñé, and S. Member, “On Voltage Acceleration Models of Time to Breakdown — Part I : Experimental and Analysis Methodologies,” vol. 56, no. 7, pp. 1433–1441, 2009.

[62] E. Y. Wu, J. Suñé, and S. Member, “On Voltage Acceleration Models of Time to Breakdown — Part II : Experimental Results and Voltage Dependence of Weibull

Slope in the FN Regime," vol. 56, no. 7, pp. 1442–1450, 2009.

[63] J. McPherson and H. Mogul, "Underlying physics of the thermochemical model in describing low-field time- dependent dielectric breakdown in thin films," *J. Appl. Phys.*, vol. 84, 1998.

[64] J. McPherson, J. Y. Kim, A. Shanware, and H. Mogul, "Thermochemical description of dielectric breakdown in high dielectric constant materials," *Appl. Phys. Lett.*, vol. 82, no. 13, pp. 2121–2123, 2003.

[65] J. W. Mcpherson, "Microelectronics Reliability Time dependent dielectric breakdown physics – Models revisited," *Microelectron. Reliab.*, vol. 52, no. 9–10, pp. 1753–1760, 2012.

[66] C. Prasad *et al.*, "Dielectric breakdown in a 45 nm high-k/metal gate process technology," in *2008 IEEE International Reliability Physics Symposium*, 2008, pp. 667–668.

[67] D. Ielmini and R. Waser, *Resistive Switching: From fundamentals to Nanoionic Redox Processes to Memristive Device Applications*. Wiley-Interscience, 2018.

[68] J. H. Stathis, "Physical and Predictive Models of Ultrathin Oxide Reliability in CMOS Devices and Circuits," *IEEE Trans. Device Mater. Reliab.*, vol. 1, no. 1, pp. 43–59, 2001.

[69] N. Raghavan *et al.*, "Feasibility of SILC Recovery in sub-10-Å EOT Advanced Metal Gate – High- κ Stacks," vol. 34, no. 8, pp. 1053–1055, 2013.

[70] G. Ribes *et al.*, "Review on High- k Dielectrics Reliability Issues," *IEEE Trans. Device Mater. Reliab.*, vol. 5, no. 1, pp. 5–19, 2005.

[71] R. Sutherland, J. A. Williamson, and R. Collins, "Electroforming and conduction in thin amorphous sulphide films," *J. Phys. D. Appl. Phys.*, vol. 5, 1972.

[72] H. Pagnia and S. Sotnik, "Bistable Switching in Electroformed Metal-Insulator-Metal Devices," *Phys. Status Solidi*, vol. 11, 1988.

[73] D. V Morgan and M. J. Howes, "Electroforrning and Switching in Copper Oxide

Films," *Phys. Status Solidi*, vol. 191, pp. 191–195, 1974.

[74] G. Dearnaley, A. M. Stoneham, and D. V. Morgan, "Electrical phenomena in amorphous oxide films," *Reports Prog. Phys.*, vol. 33, no. 3, pp. 1129–1191, 1970.

[75] D. P. Oxley, "ELECTROFORMING , SWITCHING AND MEMORY EFFECTS IN OXIDE THIN FILMS," *Electrocompon. Sci. Technol.*, vol. 3, no. C, pp. 217–224, 1977.

[76] G. Dearnaley, D. V. Morgan, and A. M. Stoneham, "A MODEL FOR FILAMENT GROWTH AND SWITCHING IN AMORPHOUS OXIDE FILMS," *J. Non. Cryst. Solids*, vol. 4, pp. 593–612, 1970.

[77] P. D. Greene, E. L. Bush, and I. R. Rawlings, "Proceedings of the Symposium on Deposited Thin Film Dielectric Materials," in *Proceedings of the Symposium on Deposited Thin Film Dielectric Materials*, 1968, p. 167.

[78] M. Janousch, G. I. Meijer, U. Staub, B. Delley, S. F. Karg, and B. P. Andreasson, "Role of Oxygen Vacancies in Cr-Doped SrTiO3 for Resistance-Change Memory," *Adv. Mater.*, vol. 19, no. 17, pp. 2232–2235, Sep. 2007.

[79] D. Kwon *et al.*, "Atomic structure of conducting nanofilaments in TiO2 resistive switching memory.," *Nat. Nanotechnol.*, vol. 5, no. 2, pp. 148–153, 2010.

[80] G. Hwan Kim *et al.*, "Improved endurance of resistive switching TiO2 thin film by hourglass shaped Magnéli filaments," *Appl. Phys. Lett.*, vol. 98, no. 26, p. 262901, Jun. 2011.

[81] F. Miao *et al.*, "Anatomy of a Nanoscale Conduction Channel Reveals the Mechanism of a High-Performance Memristor," *Adv. Mater.*, vol. 23, no. 47, pp. 5633–5640, Dec. 2011.

[82] S. P. Mitoff, "Electrical Conductivity and Thermodynamic Equilibrium in Nickel Oxide," *J. Chem. Phys.*, vol. 35, no. 3, pp. 882–889, Sep. 1961.

[83] T. Karakasidis and M. Meyer, "Grain-boundary diffusion of cation vacancies in nickel oxide:mA molecular-dynamics study," *Phys. Rev. B*, vol. 55, no. 20, pp. 13853–13864, May 1997.

[84] A. Belmonte *et al.*, "Analysis of the Excellent Memory Disturb Characteristics of a Hourglass-Shaped Filament in Al_2O_3/Cu-Based CBRAM Devices," *IEEE Trans. Electron Devices*, vol. 62, no. 6, pp. 2007–2013, 2015.

[85] D. S. Jeong *et al.*, "Emerging memories: resistive switching mechanisms and current status," *Reports Prog. Phys.*, vol. 75, no. 7, p. 76502, 2012.

[86] W. Kim, D. J. Wouters, S. Menzel, C. Rodenbucher, R. Waser, and V. Rana, "Lowering forming voltage and forming-free behavior of Ta2O5 ReRAM devices," in *European Solid-State Device Research Conference*, 2016, vol. 2016–Octob, pp. 164–167.

[87] D. S. Jeong, H. Schroeder, U. Breuer, and R. Waser, "Characteristic electroforming behavior in Pt/TiO2/Pt resistive switching cells depending on atmosphere," *J. Appl. Phys.*, vol. 104, no. 12, p. 123716, Dec. 2008.

[88] C. Rohde, B. J. Choi, D. S. Jeong, S. Choi, J.-S. Zhao, and C. S. Hwang, "Identification of a determining parameter for resistive switching of TiO2 thin films," *Appl. Phys. Lett.*, vol. 86, no. 26, p. 262907, Jun. 2005.

[89] B. J. Choi *et al.*, "Resistive switching mechanism of TiO2 thin films grown by atomic-layer deposition," *J. Appl. Phys.*, vol. 98, no. 3, p. 33715, Aug. 2005.

[90] H. Y. Lee *et al.*, "Electrical evidence of unstable anodic interface in Ru/HfOx/TiN unipolar resistive memory," *Appl. Phys. Lett.*, vol. 92, no. 14, p. 142911, Apr. 2008.

[91] Y. Yin Chen *et al.*, "Insights into Ni-filament formation in unipolar-switching Ni/HfO2/TiN resistive random access memory device," *Appl. Phys. Lett.*, vol. 100, no. 11, p. 113513, Mar. 2012.

[92] S. Poblador, M. C. Acero, M. B. González, and F. Campabadal, "Resistive switching with bipolar characteristics in TiN/Ti/HfO2/W devices," in *2017 Spanish Conference on Electron Devices (CDE)*, 2017, pp. 1–4.

[93] S. Yu and H.-S. P. Wong, "A Phenomenological Model for the Reset Mechanism of Metal Oxide RRAM," *IEEE Electron Device Lett.*, vol. 31, no. 12, pp. 1455–1457, Dec. 2010.

[94] M. Lee *et al.*, “Electrical manipulation of nano-filaments in metal oxides for resistance based memory,” *Nano Lett.*, vol. 9, no. 4, pp. 1477–1481, 2009.

[95] S. H. Chang *et al.*, “Effects of heat dissipation on unipolar resistance switching in Pt/NiO/Pt capacitors,” *Appl. Phys. Lett.*, vol. 92, no. 18, p. 183507, May 2008.

[96] S. W. Ryu, Y. B. Ahn, H. J. Kim, and Y. Nishi, “Ti-electrode effects of NiO based resistive switching memory with Ni insertion layer,” *Appl. Phys. Lett.*, vol. 100, no. 13, p. 133502, Mar. 2012.

[97] G.-H. Buh, I. Hwang, and B. H. Park, “Time-dependent electroforming in NiO resistive switching devices,” *Appl. Phys. Lett.*, vol. 95, no. 14, p. 142101, Oct. 2009.

[98] S. Menzel, M. Waters, A. Marchewka, U. Böttger, R. Dittmann, and R. Waser, “Origin of the Ultra-nonlinear Switching Kinetics in Oxide-Based Resistive Switches,” *Adv. Funct. Mater.*, vol. 21, no. 23, pp. 4487–4492, Dec. 2011.

[99] D. C. Gilmer *et al.*, “Effects of RRAM Stack Configuration on Forming Voltage and Current Overshoot,” in *2011 3rd IEEE International Memory Workshop (IMW)*, 2011, pp. 1–4.

[100] F. Longos *et al.*, “No Title,” in *International Memory Woroshop*, 2013.

[101] J. Liang and H.-. P. Wong, “Cross-Point Memory Array Without Cell Selectors—Device Characteristics and Data Storage Pattern Dependencies,” *IEEE Trans. Electron Devices*, vol. 57, no. 10, pp. 2531–2538, 2010.

[102] A. Flocke, T. G. Noll, C. Kugeler, C. Nauenheim, and R. Waser, “A Fundamental Analysis of Nano-Crossbars with Non-Linear Switching Materials and its Impact on TiO2 as a Resistive Layer,” in *2008 8th IEEE Conference on Nanotechnology*, 2008, pp. 319–322.

[103] J. Liang, S. Yeh, S. S. Wong, and H.-S. P. Wong, “Effect of Wordline/Bitline Scaling on the Performance, Energy Consumption, and Reliability of Cross-Point Memory Array,” *J. Emerg. Technol. Comput. Syst.*, vol. 9, no. 1, p. 9:1--9:14, 2013.

[104] C.-H. Wang *et al.*, “Three-dimensional 4F2 ReRAM cell with CMOS logic compatible process,” in *2010 International Electron Devices Meeting*, 2010, p.

29.6.1-29.6.4.

[105] G. Servalli, "A 45nm generation Phase Change Memory technology," in *2009 IEEE International Electron Devices Meeting (IEDM)*, 2009, pp. 1–4.

[106] R. Fackenthal *et al.*, "19.7 A 16Gb ReRAM with 200MB/s write and 1GB/s read in 27nm technology," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 338–339.

[107] G. Tallarida *et al.*, "Low Temperature Rectifying Junctions for Crossbar Non-Volatile Memory Devices," in *2009 IEEE International Memory Workshop*, 2009, pp. 1–3.

[108] S. G. Kim *et al.*, "Breakthrough of selector technology for cross-point 25-nm ReRAM," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, p. 2.1.1-2.1.4.

[109] Y. Koo and H. Hwang, "Zn(1-x)Te(x) Ovonic Threshold Switching Device Performance and its Correlation to Material Parameters," *Sci. Rep.*, vol. 8, no. 1, p. 11822, Aug. 2018.

[110] A. Fantini *et al.*, "Intrinsic switching variability in HfO2 RRAM," in *2013 5th IEEE International Memory Workshop*, 2013, pp. 30–33.

[111] A. Grossi *et al.*, "Fundamental variability limits of filament-based RRAM," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2017, p. 4.7.1-4.7.4.

[112] Y. Y. Chen *et al.*, "Balancing SET/RESET Pulse for > 1E10 Endurance in HfO2/Hf 1T1R Bipolar RRAM," *IEEE Trans. Electron Devices*, vol. 59, no. 12, pp. 3243–3249, 2012.

[113] S. Yu, X. Guan, and H.-. P. Wong, "Understanding metal oxide RRAM current overshoot and reliability using Kinetic Monte Carlo simulation," in *2012 International Electron Devices Meeting*, 2012, p. 26.1.1-26.1.4.

[114] Y. Y. Chen *et al.*, "Postcycling LRS Retention Analysis in HfO_2/Hf RRAM 1T1R Device," *IEEE Electron Device Lett.*, vol. 34, no. 5, pp. 626–628, 2013.

[115] Y. Y. Chen *et al.*, “Endurance/Retention Trade-off HfO2/Metal Cap 1T1R Bipolar RRAM,” *IEEE Trans. Electron Devices*, vol. 60, no. 3, pp. 1114–1121, 2013.

[116] M. Alayan, E. Vianello, B. de Salvo, L. Perniola, A. Padovani, and L. Larcher, “Correlated Effects on Forming and Retention of Al Doping in HfO_2-Based RRAM,” *IEEE Des. Test*, vol. 34, no. 3, pp. 23–30, 2017.

[117] B. Traore *et al.*, “On the Origin of Low-Resistance State Retention Failure in HfO2-Based RRAM and Impact of Doping/Alloying,” *IEEE Trans. Electron Devices*, vol. 62, no. 12, pp. 4029–4036, 2015.

[118] Y. Lin *et al.*, “A comprehensive study of 3-stage high resistance state retention behavior for TMO ReRAMs from single cells to a large array,” in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, p. 2.5.1-2.5.4.

[119] D. C. Edelstein, “20 Years of Cu BEOL in manufacturing, and its future prospects,” in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, p. 14.1.1-14.1.4.

[120] T. Wu *et al.*, “A 43pJ/Cycle Non-Volatile Microcontroller with 4.7µs Shutdown/Wake-up Integrating 2.3-bit/Cell Resistive RAM and Resilience Techniques,” in *ISSCC*, 2019.

[121] “Intel Says FinFET-Based Embedded MRAM is Production-Ready.” [Online]. Available: https://www.eetimes.com/document.asp?doc_id=1334343.

[122] C. Rohde, B. J. Choi, D. S. Jeong, S. Choi, J. S. Zhao, and C. S. Hwang, “Identification of a determining parameter for resistive switching of TiO2 thin films,” *Appl. Phys. Lett.*, vol. 86, no. 26, pp. 1–3, 2005.

[123] PANASONIC, “MN101L Reference Manual,” 2013. [Online]. Available: https://industrial.panasonic.com/ww/products/semiconductors/microcomputers/mn101l.

[124] Y. Y. Chen *et al.*, “Endurance/retention trade-off on HfO2 metal cap 1T1R bipolar RRAM,” *IEEE Trans. Electron Devices*, vol. 60, no. 3, pp. 1114–1121, 2013.

[125] Y. Y. Chen *et al.*, “Balancing SET/RESET Pulse for >1010 Endurance in HfO 2

1T1R Bipolar RRAM," *IEEE Trans. Electron Devices*, vol. 59, no. 12, pp. 3243–3249, 2012.

[126] Y. Y. Chen *et al.*, "Understanding of the endurance failure in scaled HfO2-based 1T1R RRAM through vacancy mobility degradation," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2012, pp. 10–13.

[127] L. Perniola *et al.*, "Universal Signatures from Non-Universal Memories: Clues for the Future...," in *2016 IEEE 8th International Memory Workshop, IMW 2016*, 2016.

[128] A. Grossi *et al.*, "Relationship among Current Fluctuations during Forming, Cell-To-Cell Variability and Reliability in RRAM Arrays," *2015 IEEE 7th Int. Mem. Work. IMW 2015*, 2015.

[129] S. U. Sharath *et al.*, "Towards forming-free resistive switching in oxygen engineered HfO2-x," *Appl. Phys. Lett.*, vol. 104, no. 063502, 2014.

[130] G. Molas *et al.*, "Controlling oxygen vacancies in doped oxide based CBRAM for improved memory performances," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2014, pp. 137–139.

[131] Y. Hayakawa *et al.*, "Highly reliable TaO x ReRAM with centralized filament for 28-nm embedded application T14 T15," in *2015 Symposium on VLSI Circuits (VLSI)*, 2015, vol. 724, no. 2011, pp. 14–15.

[132] M. Ueki *et al.*, "Low-Power Embedded ReRAM Technology for IoT Applications," in *Symposium on VLSI Technology*, 2015, pp. 108–109.

[133] B. Traoré *et al.*, "HfO 2 -Based RRAM : Electrode Effects , Ti / HfO 2," *IEEE Trans. Electron Devices*, vol. 63, no. 1, pp. 360–368, 2016.

[134] Y. Y. Chen *et al.*, "Tailoring Switching and Endurance/Retention Reliability Characteristics of HfO2/Hf RRAM with Ti, Al, Si Dopants," in *VLSI Technology (VLSIT), 2014 Symposium on*, 2014, no. 59, pp. 2013–2014.

[135] B. Traoré *et al.*, "Microscopic understanding of the low resistance state retention in HfO<inf>2</inf> and HfAlO based RRAM," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2015, vol. 2015–Febru, no. February, p. 21.5.1-

21.5.4.

[136] W. A. Hubbard *et al.*, “Nanofilament Formation and Regeneration during Cu/Al2O3Resistive Memory Switching,” *Nano Lett.*, vol. 15, no. 6, pp. 3983–3987, 2015.

[137] Y. Wu, L. Byoungil, and H. P. Wong, “Al2O3-Based RRAM Using Atomic Layer Deposition (ALD) With 1- RESET Current,” *IEEE Electron Device Lett.*, vol. 31, no. 12, pp. 1449–1451, 2010.

[138] L. Gao, Y. Li, Q. Li, Z. Song, and F. Ma, “Enhanced resistive switching characteristics in Al $_2$ O $_3$ memory devices by embedded Ag nanoparticles,” *Nanotechnology*, vol. 28, no. 21, p. 215201, 2017.

[139] E. Perez, M. K. Mahadevaiah, C. Zambelli, P. Olivo, and C. Wenger, “Data retention investigation in HfO 2 -based RRAM arrays by using accelerated tests,” in *Workshop on Dielectrics in Microelectronics (Wodim)*, 2018, vol. 2, pp. 4–5.

[140] H. Xie *et al.*, “Resistive switching properties of HfO2-based ReRAM with implanted Si/Al ions,” in *AIP Conference Proceedings*, 2012, vol. 1496, pp. 26–29.

[141] M. Leskelä and M. Ritala, “Atomic layer deposition (ALD): From precursors to thin film structures,” *Thin Solid Films*, vol. 409, no. 1, pp. 138–146, 2002.

[142] R. W. Johnson, A. Hultqvist, and S. F. Bent, “A brief review of atomic layer deposition: From fundamentals to applications,” *Mater. Today*, vol. 17, no. 5, pp. 236–246, 2014.

[143] J. F. Ziegler, “Ion Implantation - Science and Technology,” in *Science And Technology*, 2nd ed., J.F. Ziegler, Ed. Elsevier Ltd, 1988.

[144] R. S. Nelson, “The theory of recoil implantation,” *Radiat. Eff. Defects Solids*, vol. 2, no. 3, pp. 47–50, 2006.

[145] J. Pelletier and a Anders, “Plasma-blased ion implantation and deposition: A review of physics, technology, and applications,” *Ieee Trans. Plasma Sci.*, vol. 33, no. 6, pp. 1944–1959, 2005.

[146] A. F. Tasch, H. Shin, and C. Park, “An Improved Approach to Accurately Model Shallow B and BF[sub 2] Implants in Silicon,” *J. Electrochem. Soc.*, vol. 136, no. 3, p. 810, 1989.

[147] M. C. Cheynet, S. Pokrant, F. D. Tichelaar, and J.-L. Rouvière, “Crystal structure and band gap determination of HfO2 thin films,” *J. Appl. Phys.*, vol. 101, no. 5, p. 054101, 2007.

[148] J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized gradient approximation made simple,” *Phys. Rev. Lett.*, vol. 77, no. 18, pp. 3865–3868, 1996.

[149] P. Hohenberg and W. Kohn, “Inhomogeneous electron gas,” *Phys. Rev. B*, vol. 136, no. 3B, pp. 865–871, 1964.

[150] P. Ordejon, E. Artacho, and J. M. Soler, “Selfconsistent order-N density-functional calculations for very large systems,” vol. 53, no. 16, pp. 441–444, 1996.

[151] J. M. Soler *et al.*, “The SIESTA method for *ab initio* order- *N* materials simulation,” *J. Phys. Condens. Matter*, vol. 14, no. 11, pp. 2745–2779, 2002.

[152] D. R. Islamov, V. A. Gritsenko, C. H. Cheng, and A. Chin, “Percolation conductivity in hafnium sub-oxides,” *Appl. Phys. Lett.*, vol. 105, no. 26, pp. 2014–2017, 2014.

[153] S. Clima *et al.*, “First-principles simulation of oxygen diffusion in HfOx: Role in the resistive switching mechanism,” *Appl. Phys. Lett.*, vol. 100, no. 13, 2012.

[154] L. G. Parratt, “Surface studies of solids by total reflection of x-rays,” *Phys. Rev.*, vol. 95, no. 2, pp. 359–369, 1954.

[155] I. W. Hamley and J. S. Pedersen, “Analysis of neutron and X-ray reflectivity data. I. Theory,” *J. Appl. Crystallogr.*, vol. 27, pp. 29–35, 1994.

[156] D. Briggs and M. P. Seah, *Practical Surface Analysis, Auger and X-ray Photoelectron Spectroscopy*. 1996.

[157] V. Fomenko, E. P. Gusev, and E. Borguet, “Optical second harmonic generation studies of ultrathin high-k dielectric stacks,” *J. Appl. Phys.*, vol. 97, no. 8, pp. 1–8,

2005.

[158] W. Wang *et al.*, “Coupled electron-hole dynamics at the Si/SiO2interface,” *Phys. Rev. Lett.*, vol. 81, no. 19, pp. 4224–4227, 1998.

[159] J. G. Mihaychuk, J. Bloch, Y. Liu, and H. M. van Driel, “Time-dependent second-harmonic generation from the Si-SiO(2) interface induced by charge transfer.,” *Opt. Lett.*, vol. 20, no. 20, pp. 2063–5, 1995.

[160] C. Scharrer and Y. Zhao, “High-frequency capacitance measurements for monitoring EOT of thin gate films,” *Measurement+Control*, vol. 39, no. 1, pp. 24–26, 2006.

[161] C. H. Choi, Y. Wu, J. S. Goo, Z. Yu, and R. W. Dutton, “Capacitance reconstruction from measured C-V in high leakage, nitride/oxide MOS,” *IEEE Trans. Electron Devices*, vol. 47, no. 10, pp. 1843–1850, 2000.

[162] D. K. Schroder, *Semiconductor Material and Device Characterization*, 3rd Ed. Wiley-Interscience, 2006.

[163] E. H. Nicollian and J. R. Brews, *MOS (metal oxide semiconductor) physics and technology*, Second Ed. New York, Chichester, Brisbane, Toronto, Singapore: Wiley-Interscience, 1982.

[164] A. Tsiara *et al.*, “Performance and Reliability of a Fully Integrated 3D Sequential Technology,” *Symp. VLSI Technol.*, pp. 75–76, 2018.

[165] M. von Haartman and M. Östling, *Low-Frequency Noise In Advanced Mos Devices*. Springer, 2007.

[166] S. Zafar, H. Jagannathan, L. F. Edge, and D. Gupta, “Measurement of oxygen diffusion in nanometer scale HfO2 gate dielectric films,” *Appl. Phys. Lett.*, vol. 98, no. 15, p. 152903, 2011.

[167] H. Yildirim, A. Kara, S. Durukanoglu, and T. S. Rahman, “Calculated pre-exponential factors and energetics for adatom hopping on terraces and steps of Cu(1 0 0) and Cu(1 1 0),” *Surf. Sci.*, vol. 600, no. 2, pp. 484–492, 2006.

[168] T. Bertaud *et al.*, "Resistive switching of Ti/HfO2-based memory devices: impact of the atmosphere and the oxygen partial pressure," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 41, p. 12018, 2012.

[169] P. P. Camanho, C. G. Davila, and M. F. de Moura, "Numerical Simulation of Mixed-Mode Progressive Delamination in Composite Materials," *J. Compos. Mater.*, vol. 37, no. 16, pp. 1415–1438, Aug. 2003.

[170] J. R. REEDER and J. H. REWS, "Mixed-mode bending method for delamination testing," *AIAA J.*, vol. 28, no. 7, pp. 1270–1276, 1990.

[171] J. W. Hutchinson and Z. Suo, "Mixed Mode Cracking in Layered Materials," vol. 29, J. W. Hutchinson and T. Y. B. T.-A. in A. M. Wu, Eds. Elsevier, 1991, pp. 63–191.

[172] M. Barlas *et al.*, "OxRAM integration above FDSOI transistor drain: Integration approach and process impact on electrical characteristics," in *Solid State Devices and Materials*, 2017.

[173] A. T. Zehnder, *Fracture Mechanics in Lecture Notes in Applied and Computational Mechanics (Vol. 62)*. 2012.

[174] J. Sune, "New physics-based analytic approach to the thin-oxide breakdown statistics," *IEEE Electron Device Lett.*, vol. 22, no. 6, pp. 296–298, 2001.

[175] S. Ambrogio, S. Balatti, S. Choi, and D. Ielmini, "Impact of the Mechanical Stress on Switching Characteristics of Electrochemical Resistive Memory," *Adv. Mater.*, vol. 26, no. 23, pp. 3885–3892, Jun. 2014.

[176] S. Z. Rahaman *et al.*, "The Role of Ti Buffer Layer Thickness on the Resistive Switching Properties of Hafnium Oxide-Based Resistive Switching Memories," *Langmuir*, vol. 33, no. 19, pp. 4654–4665, May 2017.

LIST OF TABLES

LIST OF FIGURES

LIST OF DISCLOSURES

Publications:

Impact of Si/Al implantation on the forming voltage and pre-forming conduction modes in HfO2 based OxRAM cells *ESSDERC/ESSCIRC 2016* M. Barlas et al.

OxRAM integration above FDSOI transistor drain: Integration approach and process impact on electrical characteristics *SSDM 2017* M. Barlas et al.

Improvement of HfO2 based RRAM array performances by local Si Implantation *International Electron Devices Meeting* 2017 M. Barlas, et. al

A. Grossi, E. Vianello, M. Aly, M. Barlas, L. Grenouillet, J. Coignus, E. Beigne, T. Wu, B. Le, M. Wootters, C. Zambelli, E. Nowak and S. Mitra, "Resistive RAM Endurance: Array-level Characterization and Correction Techniques Targeting Deep Learning Applications," *IEEE Trans. Electron Devices*, 2019.

Patents:

[P-1] Selector device for a memory cell

M. Barlas, L. Grenouillet, E. Vianello, P. Blaise, B. Sklenard
Accepted: 09.09.2016, Granted: 21.09.2017, ID No: FR 16 58444, Extension to US ongoing

[P-2] Non-volatile oxide-based resistive memory cell and its fabrication process

L. Grenouillet, M. Barlas, P. Blaise, B. Sklenard, E. Vianello
Accepted: 09.09.2016, Granted: 21.09.2017, ID No: FR 16 58446, Extension to US ongoing

[P-3] Non-volatile memory cell for high density integration

J.M. Portal, M. Barlas, L. Grenouillet, E. Vianello
Accepted: 12.05.2017, ID No: FR 1754181, EU patent granted Extension to US pending

[P-4] OXRAM memory cell with reduced variability and its fabrication process

M. Barlas, P. Blaise, L. Grenouillet, B. Traoré
Accepted: 12.05.2017, ID No: FR 17 61627, EU patent granted Extension to US pending