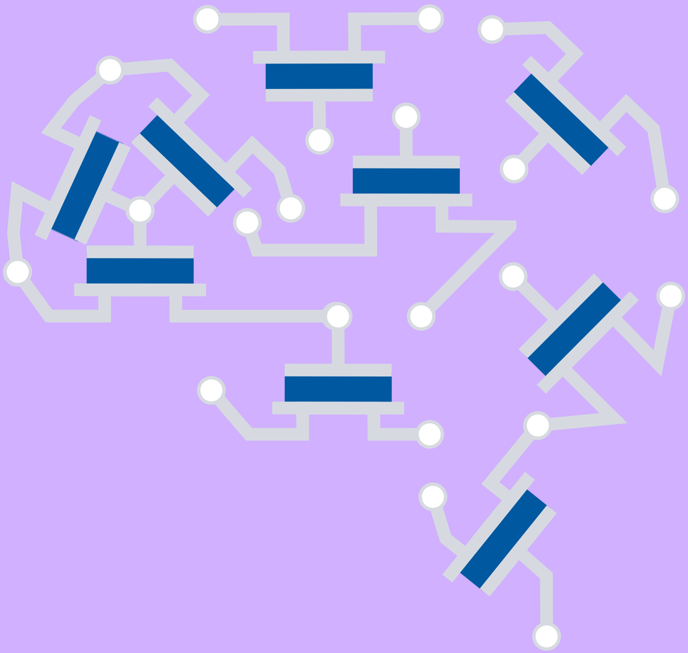
Mattia Halter

Ferroelectric memristors for neuromorphic applications: design, fabrication, and integration



Diss. ETH No. 28298

MATTIA HALTER

FERROELECTRIC MEMRISTORS FOR NEUROMORPHIC APPLICATIONS: DESIGN, FABRICATION, AND INTEGRATION

DISS. ETH NO. 28298

FERROELECTRIC MEMRISTORS FOR NEUROMORPHIC APPLICATIONS: DESIGN, FABRICATION, AND INTEGRATION

A dissertation submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

MATTIA HALTER
MSc ETH

born on 17 March 1991
citizen of Switzerland

accepted on the recommendation of

Prof. Dr. Mathieu Luisier, examiner
Prof. Dr. Beatriz Noheda, co-examiner
Prof. Dr. Bert Offrein, co-examiner
Dr. Jean Fompeyrine, co-examiner

2022

To axolotls

ABSTRACT

An artificial synaptic element consisting of a three terminal Ferroelectric Field-Effect Transistor (FeFET) with an oxide channel is presented in this thesis. Bio-inspired computing emerged as the forefront technology to harness the growing amount of data generated in an increasingly connected society. Dedicated hardware solutions are required to leverage its full potential, especially regarding power consumption and parallelism by co-locating memory and computing. A common denominator among most proposed neuromorphic computing architectures is a neural network consisting of neurons and synapses. In the analog domain, the state of a synapse is emulated by a programmable and persistent electrical conductance, for which multiple physical effects can be exploited. Among them, the ferroelectric effect promises a low power operation and high endurance due to the electrostatic nature of the polarisation switching.

In the first part of this thesis, the process development for the materials is reviewed. A Back-End-Of-Line (BEOL) compatible crystallisation of $HfZrO_4$ (HZO), a CMOS friendly and scalable material, in the metastable ferroelectric phase is demonstrated by a millisecond flash lamp anneal. Also, the effect of the electrodes and film thickness is studied. It is found that TiN and WOx electrodes both support the stabilisation of the metastable ferroelectric phase and that the ferroelectricity vanishes for very thin HZO. Furthermore, the development of a semiconducting WOx channel is presented, including the effect of the deposition method and processing conditions on its electrical properties.

In the second part of the manuscript, the developed materials are combined in a FeFET device: a simple gate-first device layout is designed and then used to establish a direct link between the ferroelectric polarisation and the channel conductance. The fine-grained domain structure of HZO is used to demonstrate a programmable and persistent multi-state conductance. Moreover, the FeFETs display a good linearity and symmetry, a low cycle-to-cycle noise, fast programming speed, and low write energy. The device area, dynamic range, endurance, and large device-to-device variability call for additional improvement.

In the last part, the process is further developed with the objective of decreasing the device area, reducing the device-to-device variability, and in-

creasing the dynamic range and endurance. An important change to allow for such improvements is the growth of WOx by atomic layer deposition instead of sputtering. In addition, a more complex design enables the integration in cross bar arrays. The result is a sub-µm size artificial synaptic element with a quasi-continuous resistance tuning and a fine-grained weight update. Moreover, the change of conductance appears over two timescales. It is found that a fast, saturating ferroelectric effect and a slow, less saturating ionic drift and diffusion process are responsible for the multitime scale behaviour. The FeFET exhibits an excellent endurance and ferroelectric retention thanks to the good interface between the ferroelectric and the oxide channel. Its reduced footprint is an important step towards dense integration. Also, it is found that as a consequence of the two physical effects leading to different timescales, the symmetry and linearity of the device deteriorate. Taking all these characteristics into account, the performance of the FeFET is assessed by simulating the classification of the MNIST dataset, resulting in an excellent accuracy of 88 % accuracy, making it well suited for neuromorphic and cognitive computing.

ZUSAMMENFASSUNG

In dieser Arbeit wird ein ferroelektrischen Feldeffekttransistor (FeFET) mit einem Oxidkanal als künstliches synaptisches Gewicht vorgestellt. Bioinspiriertes Rechnen hat sich zur führenden Technologie entwickelt, um die wachsende und in einer zunehmend vernetzten Gesellschaft erzeugte Datenmenge zu verarbeiten. Um das volle Potenzial auszuschöpfen sind jedoch spezielle Hardwarelösungen erforderlich: insbesondere im Hinblick auf den Stromverbrauch und die Parallelität sind Lösungsansätze die Speicher und Rechnungsleistung örtlich kombinieren, vielversprechend. Ein gemeinsamer Nenner der meisten vorgeschlagenen neuromorphen Paradigmen ist das neuronale Netz, welches aus Neuronen und Synapsen besteht. Der Zustand einer Synapse wird im analogen Bereich durch ein Bauteil mit einem programmierbaren und nichtflüchtigen elektrischen Leitwert nachgebildet, wofür mehrere physikalische Effekte verwendet werden können. Unter denen ist der ferroelektrische Effekt vielversprechend da die elektrostatische Natur des ferroelektrischen Effekts eine stromsparende Polarisationsumschaltung und eine hohe Lebensdauer bietet.

Im ersten Teil wird die Prozessentwicklung für die verwendeten Materialien besprochen. Eine Back-End-Of-Line (BEOL) kompatible Kristallisation von $HfZrO_4$ (HZO), einem CMOS-freundlichen und skalierbaren Material, in der metastabilen ferroelektrischen Phase, wird durch eine pulsierte Erhitzung im Millisekundenbereich demonstriert. Zusätzlich werden die Auswirkungen der Elektroden und der Schichtdicke untersucht: Sowohl TiN als auch WOx Elektroden unterstützen die Stabilisierung der metastabilen ferroelektrischen Phase und bei sehr dünnem HZO verschwindet die Ferroelektrizität. Außerdem wird die Prozessentwicklung eines halbleitenden WOx-Kanals vorgestellt, einschließlich der Auswirkungen der Abscheidungsmethode und der Prozessbedingungen auf seine elektrischen Eigenschaften.

In einem zweiten Teil werden die entwickelten Materialien in einem FeFET-Bauelement kombiniert: Ein einfacher gate-first Herstellungsprozess wird entworfen und FeFET Bauelemente realisiert, um eine direkte Verbindung zwischen der ferroelektrischen Polarisation und der Leitfähigkeit des Kanals herzustellen. Die feingliedrige Domänenstruktur von HZO wird verwendet, um eine in mehrere Zustände programmierbare und

nichtflüchtige Leitfähigkeit zu demonstrieren. Außerdem weisen die FeFETs eine gute Linearität und Symmetrie, eine geringe Zyklus-zu-Zyklus-Variation, eine schnelle Programmierung und eine niedrige Schreibenergie auf. Die Bauteilfläche, das Leitfähigkeitsfenster, die Ausdauer und die große Variation unter Bauelementen erfordern jedoch noch Verbesserungen.

Im letzten Teil wird das Verfahren weiterentwickelt, mit dem Ziel die Bauteilfläche und die Variation unter den Bauelementen zu verringern, sowie das Leitfähigkeitsfenster zu vergrössern und die Ausdauer zu erhöhen. Eine wichtige Änderung in diese Richtung ist das Wachstum von WOx durch Atomlagenabscheidung (anstelle von Sputtern). Darüber hinaus ermöglicht ein komplexeres Design die Integration von Crossbar-Arrays. Das Ergebnis ist ein künstliches synaptisches Gewicht im sub-µm-Bereich mit einer quasi kontinuierlichen Leitfähigkeit und einer feinkörnigen Gewichtsveränderung. Darüber hinaus erfolgt die Änderung der Leitfähigkeit über zwei Zeitskalen: Ein schneller, sättigender ferroelektrischer Effekt und ein langsamer, weniger sättigender ionischer Drift- und Diffusionsprozess sind für das Schalten auf mehreren Zeitskalen verantwortlich. Der FeFET hat eine ausgezeichnete Ausdauer und Zustandsbeibehaltung aufgrund der guten Qualität der Schnittstelle zwischen der ferroelektrischen Schicht und dem Oxidkanal. Sein verringerter Platzbedarf ist ein wichtiger Schritt in Richtung einer dichten Integration. Als Folge der beiden physikalischen Effekte auf unterschiedlichen Zeitskalen, nimmt die Symmetrie und Linearität des Bauelements ab. Dann, wird die Klassifizierung des MNIST-Datensatzes unter Berücksichtigung der nicht-idealer FeFET-Eigenschaften simuliert und bewertet. Das gute Ergebnis von 88 % Genauigkeit zeigt, dass er sich gut für neuromorphe und kognitive Rechnungsarchitekturen eignet.

ACKNOWLEDGEMENTS

This Ph.D study was carried out as part of a collaboration between ETH Zurich and IBM Reserach Europe - Zurich Laboratory.

First of all, I would like to express my sincere gratitude to my advisor Mathieu Luisier for the opportunity to join his "Computational Nanoelectronics" group at the Integrated Systems Laboratory (IIS) and for always being helpful and supportive for my Ph.D study and related research. I am thankful for his trust and the freedom he gave me in my work that made the external collaboration possible.

I am grateful to Bert Offrein for offering me a Ph.D position in his "Neuromorphic Devices and Systems" group at IBM Research and for his trust and support ever since. I value the informal and mindful discussions we had, the honest feedback and freedom that I experienced.

I am also grateful to Beatriz Noheda for accepting to be the co-examiner of this thesis

Jean Fompeyrine was my primary advisor for the first half of this long adventure. I am deeply thankful to him for creating a friendly, trusting, and fun working environment and for continuously having expressed his satisfaction, a great motivation to keep going. I am still impressed by his far reach in the community, and want to thank him for initiating the many collaborations. If he did not know something, he certainly knew somebody who did. I will always value his thoughtful opinion.

For the second half of my Ph.D, Laura Bégon-Lours has been an incredibly helpful and supportive advisor. I would like to express my sincere gratitude for always offering a helping hand, for the many patient proofreading and valuable discussions, for guidance and theoretical advise, and simply for being a friend. The best partner in crime for ferroelectrics.

I especially want to thank Diana Dávila Pineda, who is partially responsible for starting my career as a researcher almost 6 years ago. Ever since we have been colleagues and I always could rely on her. She taught me a great deal in the cleanroom, where I spent most of my time during my Ph.D. Her endless efforts in improving the work environment are deeply appreciated. Apart from the professional aspect, I truly value her friendship, all the laughs we have, and her everlasting positivity. Thanks for letting me practice for the driver licence with your car.

I am grateful for the many conversations I have enjoyed with Youri Popoff, my long standing office mate with a good taste for coffee. Many of my processing skills I have learned from him and he sharing his perspective has always improved my day. He is always interested in helping solving a problem and has surely done so for me many times.

I want to thank Valeria Bragaglia for all the help and discussion related to our projects, and for her trust in my processing opinions.

I am thankful to all members of the "Neuromorphic Devices and Systems" group for the friendly environment they created.

This thesis is a collaborative work. Therefore, I would like to thank the following people:

Éamon O'Connor for his pioneering work regarding the growth and crystallisation of HZO.

Ute Drechsler for her highly valuable advice in processing, for being the aorta of the cleanroom and ensuring its operation and for contributing to this work by performing CMP.

Marilyne Sousa for the direct contribution to this work by performing STEM analysis.

Steffen Reidt for keeping good care of the FIB and for the relaxing laughs amid hectic times.

Antonis Olziersky for running countless(!) e-beam exposures and for accepting my designs and requests.

Daniele Caimi for dicing many wafers into chips and for all the helpful advice I got in the cleanroom.

Richard Stutz for keeping various deposition tools operational and for the many technical insights I learned about tools. I appreciate the informal discussions we had.

Heinz Siegwart for inspiring engineering advice and together with Youri Popoff for building a new automated electrical probe station.

I am truly grateful to my parents Simone, Manfred, Paul, and Claudia for their encouragement and unfailing belief in me.

Finally, I would like to express my deepest gratitude for the one companion I love most, my partner Julia. There are no words to express how grateful I am for her support, her encouragement, her patience, and understanding.

This work was funded by Horizon 2020: ULPEC (no 732642) and BeFerroSynaptic (no 871737).

CONTENTS

1

INTRODUCTION

> *Everything needs to change – and it has to start today!*
>
> — Greta Thunberg

The amount of data created during the last thirty years ($\sim$320 ZB[1]) is about the same as what will be created during the next three years ($\sim$364 ZB projected for 2022-2024), a phenomenon accelerated by the Covid-19 pandemic [1–3]. A considerable share will come from the rapidly growing Internet-of-Things (IoT) [4], which connects the physical world and computing entities. The development of sensors and actuators connected to the internet and facilitating applications such as environmental monitoring, smart healthcare systems, or smart transport [5, 6], comes in pair with the emergence of artificial neural networks (ANNs) for data processing. The conventional von-Neumann architecture cannot sustain such evolution, because of the energy and performance bottleneck [7] coming from the massive data movement between its physically separated memory and processing units. Novel processing architectures, device technologies, and computational paradigms have therefore recently emerged. In-memory computing [8] co-locates memory and processing, but a major challenge remains in supporting operations under a wide range of effective timescales [9]. This requirement arises from the need to adapt the computation to the input time scale in real-time online applications (e.g. real-world sensory signals) and because multi timescales are inherent to spiking neuro-dynamics (e.g. neural activation decay, combination of short and long term plasticity mechanisms [9–11]). Thus, novel hardware with multi timescale characteristics at the synapses and neurons level, but also matching interfacing circuits, are required. To emulate the complex and multi timescale plasticity processes of biological synapses [9], physical effects acting at different timescales must be orchestrated in an artificial synapse [11]. Non-volatile memory devices, so called memristors, are used to emulate the synaptic functionality, to locally store network parameters [12, 13], and to serve as analog computing element [14]. Being able to tune the timescale

---

[1] $1\,\text{ZB} = 10^{21}$byte

(tunable volatility) in Spiking Neural Networks (SNN) would potentially bring the ability to mimic many of the basic processing and storage operations of the mammalian brain [15] and facilitate reservoir computing or unsupervised learning [16]. Moreover, SNNs allow to benefit from the sparse and energy efficient spike representations where the information exchange and processing are event-driven. The spiking energy is thus consumed only when and where it is needed [17].

Another application of artificial synapses are Artificial Neural Networks (ANNs). In recent years, ANNs have remarkably evolved [18] with increasing complexity and applications [19]. The training and operation of ANNs are dominated by Vector-Matrix Multiplications (VMM), which are required to calculate the propagation of the signal from the input to the output. VMMs can be performed in the analog domain using memristors in a crossbar configuration. The matrix values are mapped to the conductances of the memristors and the vector values are applied to the crossbar inputs encoded as voltages. By reading the summed current at each output, the VMM is performed fully parallel and directly in the memory [14]. The underlying memristors should ideally possess a gradual (analogue) modulation of its conductance, a dynamic range of 8 to 100, a long endurance ($>10^9$ cycles), a low power operation ($<10\,\mathrm{pJ}$), and a long retention (>10 years) [20–22].

A promising memristor concept is the Ferroelectric Field-Effect Transistor (FeFET). The FeFET concept dates back to the 1950's [23] and was investigated for a long time for binary memory applications such as FeRAM [24] or Fe-NAND [25]. These implementations were based on perovskite ferroelectric layers. The difficulty of directly integrating a ferroelectric on Si and the continuous need for scaling [26] prevented a wider commercialisation. These shortcomings have been lifted with the discovery of ferroelectricity in hafnia compounds [27] and the rigorous research [28] it has triggered. HfO_2 is integrated by Intel since 2007 [29] and is fully compatible with the Complementary Metal-Oxide-Semiconductor (CMOS) technology. The Atomic Layer Deposition (ALD) of hafnia compounds enables a precise thickness control, conformality, and 3D structures [30]. The HfO_2 - ZrO_2 solid solution has a CMOS and even Back-End-Of-Line (BEOL) compatible [31] crystallisation temperature and shows ferroelectric properties down to a few nanometres of layer thickness [32–34]. The low power, voltage-driven, and multi-state nature of FeFETs makes them viable candidates for the development of memristors, which has further led to the application of FeFETs [28, 35, 36] in neuromorphic chips. In contrast to the ferroelectric tunnel junction (FTJ), hafnia-based ferroelectric films do not have to be scaled to tunnelling dimensions where the ferroelectricity vanishes. Instead, they can be used with an optimum thickness in the gate stack. FeFETs have the advantage over two-terminal devices of separating the read and write operations [37]. This permits to write to a high impedance gate (low power), while reading from an ohmic channel [38]. FeFETs were demonstrated in the Front-End-Of-Line (FEOL) down to the 22 nm node as binary [39] and to the 28 nm node as multi-state [40] devices. Integrating FeFETs in the BEOL [38, 41, 42] leaves more area for the control electronics, while at the same time it relaxes size constraints on the FeFETs: Larger number of ferroelectric domains per device translates into a larger number of intermediate states.

In this thesis we present the development of a BEOL-compatible, back-gated junction-less FeFET that utilizes a $HfZrO_4$ (HZO) ferroelectric gate dielectric and a tungsten oxide (WO_x) thin film channel. WO_x, a transition metal oxide, was chosen as n-type channel due to its relatively mobile oxygen ions and thereby tunable conductivity as a function of the oxygen content [43, 44]. Therefore, the resistance modulation of the channel is possible on two different timescales: one relying on the the fast electronic screening of the polarisation charges and one taking advantage of the slow ionic drift and diffusion processes of oxygen. A brief introduction to the theory of brain inspired computing, memristors and ferroelectricity will be discussed in Chapter 2. In Chapter 3 we introduce semiconductor processing and material characterisation methods relevant to this work. Further, we report on the development of the HZO and WO_x layers in Chapter 4. The fabrication process to combine these materials to form a FeFET as well as its electrical characterisation are described in Chapter 5: After reporting on the first FeFET generation, we present the development, fabrication, and characterisation of a scaled and improved second FeFET generation. Finally, we simulate our FeFETs in an analog memristor crossbar array by using the MLP+NeuroSimV3 [45] framework with an excellent classification accuracy of 88 % on the MNIST [46] dataset.

2

THEORY

> *Progress is impossible without change, and those who cannot change their minds cannot change anything.*
> — George Bernard Shaw

2.1 BRAIN INSPIRED COMPUTING

The terms neuromorphic computing and Artificial Neural Networks (ANNs) are often heard in the context of Artificial Intelligence (AI). They both are brain inspired, but stand for fairly different concepts and communities. This section gives an overview of these two approaches.

2.1.1 Artificial Neural Networks

Information and communication technologies have become omnipresent in every part of our lives, ranging from smartphones to smart vacuum cleaners, healthcare to education, transport to climate-change and lately the Internet-of-Things (IoT). To cope with this huge demand, but also to exploit the opportunities offered by the availability of such data, new machine learning (ML) techniques like ANNs have emerged. ANNs are computing systems inspired by the mammalian brain in the sense that they are based on a collection of connected nodes called artificial neurons. The latter loosely model the neuron of a biological brain. Each connection between neurons can transmit a signal to another neuron and resembles the synapse found in the biological brain. An artificial neuron can receive signals from many connections. These signals are summed up before the result is passed to a non-linear function. The connections have a weight, resembling the strength of a synapse, which is adjusted during an initial learning phase. Figure 2.1 shows a schematic of a simple artificial neural network with an input layer, two hidden layers, and an output layer, also called Multi-Layer-Perceptron (MLP). The vector $\vec{x}$, which represents a data point, e.g. all pixels of an image, is presented to the network at the

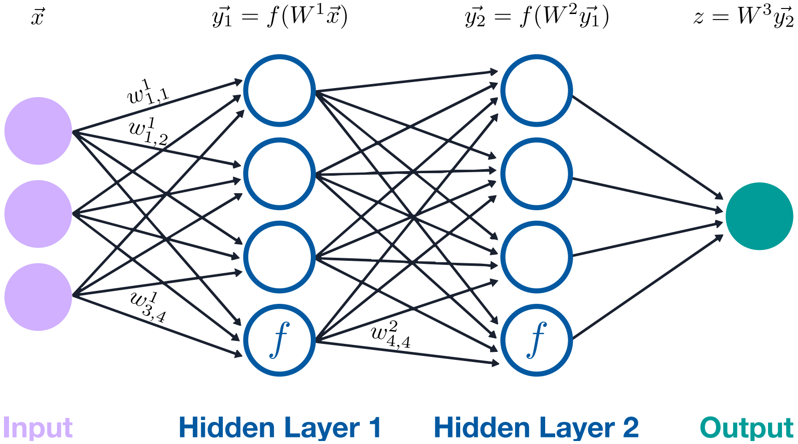


Figure 2.1: Multi layer perceptron with two hidden layers. All neurons of one layer are connected to all neurons of the next layer. The connection strength is represented as a weight $w^l_{i,j}$. Each neuron performs a non-linear transformation (f) of its input to its output.

input. The strength of each synapse is denoted as $w^l_{i,j}$, where $l = 1, 2, ..., n$ refers to the layer index, j to the index of the signal-emitting neuron, and i to the index of the receiving neuron, while n is the total number of layers, including the output (here $n = 3$). All weights between two layers can be represented as a matrix W^l, using i,j as matrix indices. A common approach to train such a network is the Back Propagation (BP) algorithm [47], which belongs to the class of supervised learning algorithms. Supervised means that for each given data point $\vec{x}$ of the training data-set, the result $\hat{z}$ must be known. For example, if an image classification data-set used for training, the class to which each image belongs must be known. In other words, all data are labelled. The BP algorithm has three phases:

1. First, the input signal $\vec{x}$ is passed through the network and ripples towards the output. When passing trough a synapse between adjacent neuron layers, the signal is multiplied by the corresponding weight $w^l_{i,j}$. At the input of a neuron, all incoming signals are summed up, the neuron then applies a non-linear function before presenting the result at its output. For the ith neuron of the first hidden layer:

$$\vec{y}^1_i = f(\sum_{j=1}^{N^0} x_j w^1_{j,i})) = f(W^1\vec{x}),$$

where N^0 is the number of neurons of the input layer. The weighted sum that each neuron performs at its input can be summarised for the entire layer by a Vector-Matrix Multiplication (VMM) $W^1\vec{x}$. This operation is repeated until the signal reaches the output. The non-linear transformation of the signal at each layer can be expressed in a more generalised way as:

$$\vec{y}^l = f(W^l\vec{y}^{l-1}), \;\; \vec{y}^0 = \vec{x}, \;\; \vec{y}^n = z.$$

At the output of Figure 2.1 we thus get:

$$z = f(W^3 f(W^2 f(W^1\vec{x}))).$$

2. In a second step, a loss function (e.g. $\mathcal{L} = [\hat{z} - z]^2 = [\hat{z} - f(W^3\vec{y}^2)]^2$) is calculated at the output of the ANN, where $\hat{z}$ is the targeted output, i.e. the label of the input. Learning is achieved by minimizing the loss function. A simple approach is gradient descent, which involves calculating the partial derivative of the loss function with respect to each weight in the network by back propagating the error. By knowing the local gradient (slope) of the loss function with respect to a particular weight ($\partial\mathcal{L}/\partial w^l_{i,j}$), a small step $\Delta w^l_{i,j}$ in the opposite direction of the slope can be taken to slightly minimise the contribution of that weight to the total loss. For the last layer this means:

   $$w^3_{i,j,new} = w^3_{i,j} + \Delta w^3_{i,j} = w^3_{i,j} - \eta\frac{\partial\mathcal{L}}{\partial w^3_{i,j}},$$

   where η is the learning rate defining the magnitude of that step. For the hidden layers that do not have direct access to the loss function, the chain rule is applied. Hence, for the hidden layer 2:

   $$\mathcal{L} = [\hat{z} - f(W^3 f(W^2\vec{y}^1)]^2,$$

   $$w^2_{i,j,new} = w^2_{i,j} + \Delta w^2_{i,j} = w^2_{i,j} - \eta\frac{\partial\mathcal{L}}{\partial w^2_{i,j}} = w^2_{i,j} - \eta\frac{\partial\mathcal{L}}{\partial\vec{y}^2}\frac{\partial\vec{y}^2}{\partial w^2_{i,j}}.$$

3. The last step is to update all the weights according to the calculated $\Delta w^l_{i,j}$.

Repeating these 3 steps for many data points ($(\vec{x},\hat{z})$) minimises the error, but does not guarantee to find the global minimum of the loss function. It should be noted that learning in the case of BP involves VMMs both in the forward propagation and error back propagation phases. The multiplication of an m-by-n matrix with a vector of dimension n requires $m \ x \ n$ multiplications and the same number of additions. This is thus considered to have a quadratic complexity. Already Richard Bellman [48] asserted 65 years ago that the high dimensionality of data is a fundamental hurdle in many science and engineering applications. The learning complexity quickly grows with the dimensions of the data. As a result, a common approach consisted of first pre-processing the data to reduce its dimensionality (feature extraction) and to facilitate its processing [49]. The therefore used ML and signal processing approaches like Gaussian Mixture Models (GMMs), Support Vector Machines (SVM), logistic regression, or MLPs with one hidden layer, are shallow architectures that typically contain at most one or two layers of non-linear feature transformations [50]. They perform well in solving simple or well-constrained problems, but are not robust against input variations. They are further limited in modelling and representational power.

Recent findings in neuroscience [51] revealed that the neocortex does not explicitly pre-process sensory signals, but lets them pass through a complex hierarchy of modules and by that learns patterns by their regularities. Consequently, to extract and memorise complex information such as human speech or visual scenes, deep architectures are required to transform the input (e.g. sound wave) to a high level representation (e.g. linguistic level). Deep neural networks (DNN) or MLPs with many hidden layers are good examples of deep architectures. The idea is not new: already in the 1970's the learning of the parameters of such networks was performed by using the BP algorithm [52]. The difficulty with DNNs is that the optimisation problem is highly non-convex with many local minima. Depending on the initial state of the network, learning can get stuck in poor local optima, an issue that is amplified with increasing depth of the network. Augmenting the width of a network improves the modelling power and creates many closely optimal configurations [50], reducing the risk of falling into poor local minima. Too many neurons (parameters) in a network can also lead to overfitting [50]. But wide (modelling power) and deep (complex data transformation) was not possible back in the 1970's.

Multiple factors contributed to the success of DNNs after all: better learning algorithms like Stochastic Gradient Descent (SGD) that often allows to jump out of local minima [50], different non-linearities, and an increased computational power allowing for wide and deep neural networks. Performing the training of such networks requires massive amounts of data

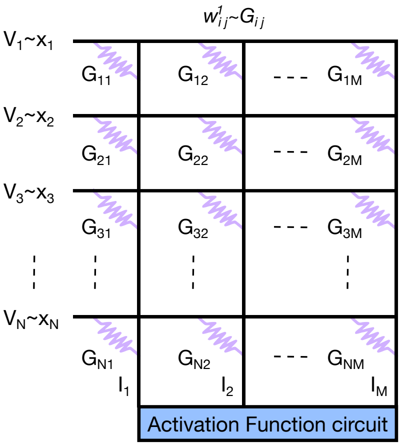


Figure 2.2: Memristor crossbar array: Each input line is connected to each output line by a memristor with electrical conductance $G_{j,i}$. The summed current I_i is sensed at each output after applying the inputs x_j as voltages V_j. The activation function is applied in the digital domain.

and computing resources which in turn requires an amount of energy that is not sustainable. The reason is that the conventional von Neuman architecture at the core of the CMOS technology is not well matched to the parallel processing nature of neural networks [7]. Constantly moving data from the memory to the CPU and back is highly inefficient. The parallelism of General-Purpose Graphical Processing Units (GPGPU) was quickly found to be better suited [53] so that this kind of hardware had since been specialised for the training of neural networks. Pipelining the BP [54] has further allowed to parallelise the learning across multiple GPUs. Even if optimised for matrix-vector multiplications, GPUs still perform the multiplication in the digital domain and need to read the data from their memory, not solving the von Neuman bottleneck.

AlphaGo Fan [55], Google's computer program that played the board game Go is based on DNNs and ran on 1920 CPUs and 280 GPUs, consuming a peak power of half a MW [56], an enormous amount of power compared to the human brain which uses about 20 W. Even if ANNs are inspired by the human brain, they are nowhere close to be as efficient. Thus, the necessity to build dedicated hardware for the training and operation of ANNs is evident. As seen before, the computational cost in ANNs mainly arises from the VMM.

A promising concept is to perform the VMM in the analog domain on a memristor crossbar array (Figure 2.2). Each input line (horizontal lines) is connected by a memristor to all output lines (vertical lines). In a memristor, the conductance is not fixed and can be changed in a non-volatile way. The matrix values are stored in the conductance of the memristor $G_{j,i}$ and the vector is applied to the inputs encoded as voltages V_j. By reading the summed current I_i at each output i, the vector-matrix multiplication, or a Multiply And Accumulate (MAC) operation is performed fully parallel and directly in the memory [14], as illustrated in Figure 2.2 and described as follows:.

$$I_i = f(\sum_{j=1}^{N^l} V_j G_{j,i}).$$

Such an approach is an example of in-memory computing [8]. The computation is performed at the same physical location as where the weights are stored. Different memristor implementations and their specifications are described in Section 2.2.

In our example we looked at a simple MLPs with two hidden layers. Other architectures of ANNs exist, often designed to perform a specialised tasks, e.g. Convolutional Neural Networks (CNNs) are typically used for image recognition or natural language processing. One thing they all have in common is the need for VMM.

2.1.2 Neuromorphic Computing

The term neuromorphic was originally introduced by Carver Mead in the late 1980 [57] for Very Large-Scale Integration (VLSI) systems implementing circuits operating in their weak-inversion or sub-threshold regime and emulating neuro-biological architectures existent in the nervous system (neurons, synapses, networks). [58] The goal of the original neuromorphic engineering approach was twofold: On one hand the understanding of neural computation by building an artificial copy of the real neural circuits. On the other hand, the development of low power computational architectures for sensory input [57]. Today, a few research groups continue on this path, mainly focusing on reconfigurable networks with biologically plausible neural dynamics [59, 60] or spike-based learning and plasticity circuits [61]. In recent years, the term has been adopted to neuromorphic computing and is used within multiple disciplines to describe a broader range of concepts.

2.1.2.1 Digital Architectures

Neuromorphic computing is used to describe mixed-signal and pure digital VLSI systems that simulate a spiking neural network (SNN) and its dynamics. Such platforms were made possible thanks to technological progress in integrated circuits. SNN is a popular and reputed model for its capacity to capture informational dynamics observed among real biological neurons and to represent and integrate several information dimensions, e.g. time, space, frequency, phase, and to deal with large volumes of data into a single model [62]. Prominent examples such as the European research project "FACETS" [63] or SpiNNaker [64] allowed for the first time to directly execute neural models from neuroscience on hardware instead of numerically simulating them as was done before. IBM's TrueNorth [65] was introduced in 2014 and was a breakthrough in the sense that it demonstrated the integration of a massive amount of silicon neurons while keeping the average consumed power for running a recurrent network in biological real-time about four orders of magnitude lower than a conventional computer: TrueNorth is made of 4096 pure digital asynchronous cores, each of them being able to simulate 256 neurons with 256×256 synaptic connections. Intel introduced the Loihi [66] chip in 2018, in many ways similar to the TrueNorth, but focusing on smaller and complex neural and synaptic features like spike-based learning mechanisms. The Loihi chip possesses 128 cores that each have 128 kB of synaptic state and another 20 kB of routing tables that can be flexibly allocated between its 1024 neurons. Furthermore, it supports compression and weight sharing, variable weight precision, delays, and tags for reinforcement learning. Plasticity rules can be assigned to synapses programmatically. This flexibility allows to implement almost arbitrary SNNs [67]. The asynchronous event driven nature (favouring the idle state) combined with sparse data, results in a low power operation. Hence, these fully digitally implemented architectures are well suited for accelerating neuroscience simulations where data is encoded based on spikes. Learning algorithms for SNN that are emulating the functionality of the brain are generally unsupervised. The data of real-world applications usually is not encoded in spikes and first needs to be converted according to some spike-based information encoding model, e.g. spike rate encoding.

2.1.2.2 *Novel Materials*

Since many years, the materials science and device physics community have been researching on materials for new types of memory. The term neuromorphic has appeared in this community after Strukov *et al.* [68] reported memristive behaviour [69] in nano-structured oxide layers. The ability to change their resistivity reminds of the biological synapse ability to adjust the connection strength between neurons and even more to emulate the synaptic weight in ANNs. This quickly led to the idea of using memristors as analog artificial synapses in ANNs and store their weight locally. Soon many more device concepts based on different physical phenomena with volatile and non-volatile characteristics were brought forward. These technologies show promising non-linear characteristics through their physics that could be used to emulate biological synapses [57]. The challenge here is to find a device architecture with optimal characteristics that can be integrated and driven by CMOS. This goes hand in hand with finding suitable learning schemes that can be applied to memristors. Additionally, memristive devices have also been shown to emulate neurons [70]. Memristor will be discussed in more details in Section 2.2

2.1.2.3 *Computational Neuroscience*

The computational models and algorithms community also uses the term neuromorphic. This comes as no surprise, software having to be co-developed with emerging neuromorphic architectures (CMOS, memristive, or mixed) [57]. The focus of this research lies on exploring spike-based learning methods that approximate the BP algorithm [71, 72] or identifying complex non-linear plasticity mechanisms that can be reproduced by memristive devices or CMOS learning circuits [12, 57, 73, 74]. Furthermore, combining computational neuroscience modelling with AI and ML training algorithms provide useful specification for the design of volatile and non-volatile memristors or new low power SNN architectures that make use of the low power computation principles of the mammalian brain. [75–77]

2.1.3 Interrelationships

After this short introduction we can try to summarise: on the one hand we have loosely brain-inspired ANNs that are trained by ML algorithms and are usually simulated in the digital domain. The digital domain allows almost infinite freedom to design ANN systems at the cost of being extremely power hungry, even when running on GPUs, a phenomenon rapidly increasing with network complexity. New dedicated hardware that is optimised for the high parallelism of ANNs is required to solve this problem.

On the other hand we have neuromorphic computing in neuroscience, mainly focusing on SNNs running on VLSI chips adopting spikes to represent, exchange, and compute data in analogy to action potentials in the brain, i.e. to simulate neuro-biological processes. A key element of such neuromorphic circuits is their non-von-Neumann architecture, e.g. consisting of multiple cores, each implementing distributed computing and memory [78]. In contrast to ANNs with ML training algorithms, information exchanges and processing are event-driven and the spiking energy is thus consumed only when and where it is needed [17].

Regardless of the specific learning algorithm or architecture, the neural network consisting of neurons and synapses is an important processing element. VMMs for instance are a common feature of spiking (e.g. SNNs) and non-spiking (e.g. ANNs) networks, explaining the significant research effort that is aimed at realizing dense, fast, and energy-efficient crossbar arrays for in-memory computing. Moreover, memristors are interesting for SNNs due to their ability to implement neuro-biological functions such as spike integration, short and long term memory, and synaptic plasticity [17] that they represent in the analog domain instead of simulating them in the digital domain. Thus, the most benefit for neuromorphic circuits is a hybrid integration where the front-end CMOS technology is combined with novel memristor devices.

For real applications, a fundamental challenge remains to train neuromorphic systems directly in the spiking domain. This would allow to benefit from the sparse and energy-efficient spike representation, and to continuously update knowledge on portable devices without the need for heavy cloud computing systems [78]. Computational neuroscience is a key ingredient to inspire neuromorphic engineering and circuits by learning how the mammalian brain performs computations at a variety of timescales, how small neurons assembled up to entire brain regions interact with peripheral sensors and actuators, or how information is encoded in spikes. It thus becomes clear that the success of novel energy-efficient neuromorphic circuits requires strong collaborations between all these different disciplines.

2.2 MEMRISTORS

2.2.1 The Memristor

In the previous section we introduced a device called memristor in the context of crossbar arrays. The memristor was first proposed by Leon Chua [79] in 1971 as the missing fundamental electrical component linking electric charges and magnetic fluxes. The memristor completes a theoretical quartet that also comprises the resistor, inductor, and capacitor. Chua defined the memristance $M(q) = d\Phi_m / dq$ as a memristor characteristic. Here, Φ_m does not represent the total magnetic flux Φ_m through an inductor (e.g. trough the surface delimited by a coil of conducting wire), but rather the flux linkage λ, which is an actual circuit quantity [80]. For inductors λ is the same as the total magnetic flux Φ_m. For memristors this is not the case as the electric field in a memristor is not as negligible as in an inductance. The flux linkage can be regarded as the integral of the voltage over time ($\lambda(t) = \int_{-\infty}^{t} V(t)\, dt$) [80]. Using the well-known charge-current relationship $q(t) = \int_{-\infty}^{t} I(t)\, dt$, we can derive a more convenient representation of the memristance:

$$\begin{gathered} M(q(t)) = \frac{d\lambda}{dq} = \frac{V(t)dt}{I(t)dt} = \frac{V(t)}{I(t)}, \\ V(t) = M(q(t))I(t). \end{gathered} \tag{2.1}$$

Equation 2.1 can be interpreted as Ohm's Law, except that $M(q(t))$ at any time $t = t_0$ depends on the entire past history of $I(t)$ [81]. The memristance is a charge-dependent resistance. If a current flows, $q(t)$ varies with time and so does $M(q(t))$. Finally, if no current flows $I(t) = 0$ then $q(t)$ stays constant and thus $M(q(t))$ is constant, resembling the memory effect [82]. As Tellini *et al.* [80] pointed out, the memristance $M(q(t)) = d\lambda(t)/dq(t)$ actually defines a proportionality relationship between the differentials of the original quantities (λ and q) it is supposed to relate. This is in contrast to the fundamental elements, resistor, capacitor, and inductance, that directly establish a relationship between V and I, q and V, and λ and I, respectively. If we were perfectly consistent, then the memristance should have been defined as:

$$
\begin{aligned}
M(q(t)) = \frac{\lambda(t)}{q(t)} &\Leftrightarrow \lambda(t) = M' q(t) \\
\lambda(t) = \int_{-\infty}^{t} V(t)\, dt &= M' \int_{-\infty}^{t} I(t)\, dt
\end{aligned} \tag{2.2}
$$

As Chua [79] pointed out himself, if M does not depend on q, M becomes equivalent to a linear and time-independent resistor, which is the case if the memristance relates λ and q (Equation 2.2). Tellini *et al.* [80] intuitively showed with circuit theory and by using Chua's matrix representation of circuit elements [83] that the memristor cannot be a 4th fundamental circuit element. It can be seen simply as a generalisation of the resistor, a time-variant resistor. Therefore, memristors, memcapacitors, and meminductances should be defined in the context of basic circuit element definitions according to Desoer [84] by including time-variant elements. Because of the mistaken association of the magnetic flux instead of the time integral of the voltage, the intrinsic resistive nature of the memristor was sometimes overlooked [80]. Di Ventra *et al.* [85] noticed that as a result of the dissipative behaviour, the memristor violates the time-reversal invariance. Later in 1976, Chua *et al.* [86] defined a more general *memristive system*:

$$
\begin{aligned}
\dot{\vec{x}} &= \vec{f}(\vec{x}, u, t), \\
\dot{\vec{y}} &= \vec{g}(\vec{x}, u, t) u,
\end{aligned}
$$

where $\vec{x}$ is the state variable of the system, u a *generic* input quantity, and y a *generic* output quantity, e.g. y is a current and u is a voltage, a voltage-controlled memristive system. This more general definition of memristive systems in principle allows to include arbitrary physical quantities. As mentioned above, memcapacitors and meminductances [69] are good examples.

There is much controversy about the term memristor and if currently proposed devices in research actually meet the targeted specifications [68, 80, 85, 87] and if the constraints should even exist [85]. For the applications of memristive systems to AI, in particular for memristive-system crossbar arrays, there is no requirement to exactly satisfy the definition (e.g. single-valued curve in the λ-q plane [80]). Thus, we will be (knowingly) using the term memristor for non-volatile resistive device implementations, regardless of the precise definition and constraints.

2.2.2 Memristors for Neuromorphic Computing

Before we introduce different memristor device concepts, we want to go over some memristor characteristics that are often used to describe them. It has to be noted here that some requirements for memristors depend on the application, the hardware they are integrated on, and the training algorithm. Learning can be achieved with binary [88, 89] or with multi-state synapses [90–92]. Nevertheless, relaxed requirements usually also mean smaller application scopes. Gokmen *et al.* [21] and Yu [20] both published desirable performance metrics.

ANALOG MULTILEVEL STATES In a computer, the weights are encoded in binary numbers, while in neuromorphic devices they are encoded as the conductance of the memristors (analog domain). Some memristors only allow for a discrete number of conductance states, while analog synaptic weights generally refer to devices with a continuous change of resistance. The synaptic plasticity characteristics observed in biological synapses show an analog-like behaviour with multilevel synaptic weight states [20]. Most neuro-inspired algorithms rely on analog synaptic weights. Moreover, small gradual changes of the conductance are required, for example in machine learning algorithms where continuously adapting the synaptic weights in small steps is needed to approach a minimum of the loss function. Although applications for binary memristors exist, it generally holds that the more analog the better. Gokmen *et al.* see 1000 and Yu >100 states as desirable, but that does not mean that less does not work [90, 92, 93]. Figure 2.3a shows a potentiation and depression response of a device with many intermediate states (light blue coloured squares).

DYNAMIC RANGE The dynamic range, or on/off ratio, is defined as the ratio between the highest and lowest conductance state of the memristor, as shown in Figure 2.3a. A too small dynamic range makes the mapping of the weights from the algorithm onto the devices difficult so that analog-to-digital converters will struggle to differentiate between the different levels. If the dynamic range is extremely large, one element in a crossbar array can dominate the current. Values between 8 to 100 [20, 21] are targeted, but depend on the application and the noise of the devices.

LINEARITY AND SYMMETRY Ideally, the potentiation and depression conductance change should be linear and symmetric with respect to the number of identical write pulses. This would allow for a linear mapping of the weights to the devices. In general, this is not the case and often the non-linearity is determined by fitting the potentiation and depression curves with an exponential function. Often the conductance changes rapidly at the beginning and then saturates towards the end of the process. The consequence is not only non-linearity, but also a non-symmetric potentiation and depression. Both are undesired as the next conductance change depends on the current state. Figure 2.3a shows an ideal linear and symmetric (purple dashed line), a non-linear and symmetric (dark blue coloured line), and a non-linear/non-symmetric potentiation (dark blue line) and depression (teal line). By choosing the appropriate write pulse scheme, a better linearity and symmetry can be obtained [20]. Figure 2.3c shows 4 common schemes. The two identical pulse schemes keep the control CMOS circuits and the memristor operation simple. The other two induce considerable overhead in the control circuitry and memristor operation as the current state must be sensed first and pulses of changing shape need to be generated [94]. This is important for online learning. For inference applications, linearity and symmetry requirements are relaxed as the target is an absolute value that is set once and small changes relative to the current state are not constantly required.

ENDURANCE The endurance of a memristor defines how many times it can be switched before breaking down. Endurance is measured by applying AC signals that switch the device back and forth until it physically fails (e.g. dielectric breakdown) or the dynamic range vanishes. A good memristor should reach $>10^9$ cycles [20].

CMOS COMPATIBILITY To reach large and efficient neuromorphic systems, all components (neurons, synapses, control circuits) should be co-integrated and scaled. Thus, a memristor technology must be compatible with CMOS processes. This usually imposes restrictions on the materials that can be used, on the process temperature, and on the device area. CMOS compatibility beyond fabrication considerations means that the available voltages are limited to $0\,\mathrm{V}$ to $5\,\mathrm{V}$, while the amount of current depends on the technology node and transistor size. Write and read signals must conform to theses boundary conditions.

ENERGY CONSUMPTION The energy required to read and write a memristor state depends on the amplitude and duration of the read/write signal, the memristors conductance, and the energy consumption of the con-

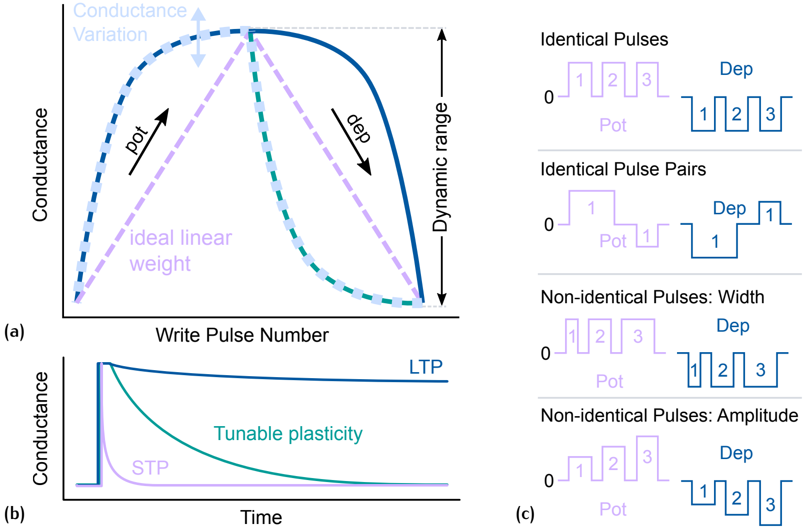


Figure 2.3: Memristor characteristics: **(a)** Different potentiation and depression shapes indicating the degree of linearity and symmetry. The dynamic range is the ratio between the highest and lowest conductance. **(b)** Different plasticity time scales. **(c)** Illustration of common pulsed potentiation and depression schemes, usually leading to a different potentiation and depression shape.

trol circuits. Ideally, low power writing is enabled by a high impedance memristor (almost no current flowing), while optimal reading is a trade-off between high impedance (small current, low power) and the capability to sense the current with an analog-to-digital converter without having to extend the integration time (slow operation). According to Yu [20], $<10\,\mathrm{fJ}/\mathrm{programming}$ pulse is the desired target.

PLASTICITY TIME-SCALES / RETENTION The retention measures how long a state is stable, i.e. how long information can be stored. In particular after the training is complete, the synapses should behave as long-term memory with data retention of >10 years. During training the state is constantly changed and this constraint can be relaxed. To mimic many of the mammalian brain synaptic plasticity dynamics, e.g. combining the ability of short-term and long-term memory, a tunable and controlled retention time is the ultimate goal. This would allow to approach brain-inspired low-power computing paradigms. Figure 2.3b shows the different retention times for Long-Term Potentiation (LTP), Short-Term Potentiation (STP), and a tunable retention time.

2.2.3 Memristor Device Implementations

In the following section we will present the working principles of the most common non-volatile memristive device concepts: Phase Change Memory (PCM) [95–100], Valence Change Memory (VCM) [101–104], Ferroelectric Tunnelling Junction (FTJ) [32, 105–110], and Ferroelectric Field Effect Transistor (FeFET) [26, 28, 36, 38, 111, 112]. Other concepts, not discussed here are Electro Chemical Metallisation (ECM) [113, 114], Electro Chemical Resistive Access Memory (ECRAM) [42, 115–118], and Magnetic Tunnelling Junction (MTJ) [119–121].

2.2.3.1 Phase Change Memory (PCM)

Being developed since the 1960s [95], PCM was demonstrated as pure digital binary memory technology and is already integrated in Intel Optane memory since 2018 [122]. Thus, PCM is the most advanced technology presented here. As the name suggests, the working principle is based on changing the phase of a material. A simple schematic of a PCM cell is illustrated in Figure 2.4a. It consists of a small local heating element that is in contact with the nanometric-volume phase change material. The phase change material usually is a compound of Ge, Sb, and Te and it can be re-

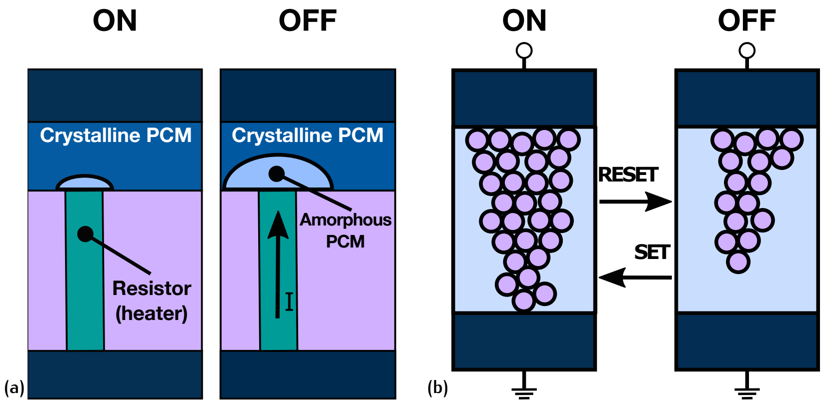


Figure 2.4: Memristor device types: **(a)** Phase Change Memory (PCM), **(b)** Valence Change Memory (VCM).

versibly switched between a high resistive, amorphous and a low resistive, crystalline state [78]. To transition from the thermodynamically unstable amorphous to the crystalline phase, PCM must be heated to a high enough temperature but below the melting point. Reversing the crystallisation is possible by heating it above the melting point and then quickly cooling it down, not giving enough time to crystallise (quenching) [99]. The required heat is generated by passing an electric current through the device. To reduce the required current, the volume of the phase change material that needs to be switched is minimised, either by reducing the volume of the layer (refined cell) or by reducing the area of one contact to the phase change layer (mushroom cell). Figure 2.4a shows a mushroom cell. Because lower write currents are achieved with the refined cell, substantial research is directed to various refined cell structures [97, 99]. State-of-the-art individual devices have demonstrated $<10\,\mu\text{A}$ crystallisation current, $\sim 25\,\text{ns}$ amorphisation current, $>10^{12}$ endurance cycles, >10 years retention, and scalability to sub-20 nm nodes [99]. The most recent advances in PCM technology can be found in recent reviews [97, 123, 124]. Continous resistance states between the crystallised and amorphous state are possible by applying fast quenching pulses of increased energy (pulse amplitude or width, always short edges for quenching). This results in a continuous amorphisation, e.g. in the mushroom cell starting from the small contact interface. Up to 3 bits of intermediate states have been demonstrated [125]. The transition from amorphous to crystalline on the other hand is very abrupt and can not be controlled to reach intermediate states [99].

The accumulative property arising from the amorphisation kinetics, the multi-state and scalability are the key advantages of this technology for neuromorphic applications. Nevertheless, there are multiple challenges still in need to be solved. The 1/f noise observed leads to a limited precision of VMMs and the conductance drift of intermediate states limits their retention. Furthermore, the accumulative behaviour is highly nonlinear and stochastic. Also, the fabrication of ultra-scaled and dense arrays is not straight forward due to effects like etch damage. [78, 126]. Projected PCM promise to mitigate the 1/f read noise problem [98, 127]. Utilizing multi-layer PCM was reported to reduce the drift [127]

2.2.3.2 *Valence Change Memory (VCM)*

The next class of potential analog artificial synapses is the VCM, a concept that was mainly developed for storage applications in the past 15 years [78]. VCM rely on oxygen-ion migration effects that lead to a valence change, hence the name valence change memory [128]. Many transition metal oxides, sandwiched between a metal with a high and one with a low workfunction, show bipolar switching without the injection of metal cations from the electrode. Instead, the transport of anions is considered essential for the switching. It is understood that oxygen-related defects, e.g. oxygen vacancies are much more mobile than transition metal cations. Thereby, the valence state of the transition metal is changed by either an enrichment or depletion of oxygen vacancies, which leads to a change in the electronic conduction [128].

We can differentiate between two categories of VCM, based on the location of the switching event in the oxide layer: Confined filamentary switching or switching at interfacial regions. The confined filamentary switching is the most studied and advanced of the two in terms of integration and scaling, and schematically illustrated in Figure 2.4b. Similar to ECM, filamentary VCM devices require an initial forming step. Before the device can be operated, a Conductive Filament (CF) needs to be formed between two electrodes, leading to the ON-state, as sketched in Figure 2.4b on the left side. Electro-forming is accomplished by applying a large field across the oxide until a pathway of oxygen vacancies is formed. This usually happens along grain boundaries or other regions with an increased defect concentration. Apart from the high electric field, it is believed that Joule heating further enhances the mobility of the vacancies until a CF is formed. The necessary voltage across the junction is called forming voltage [129]. By applying a field of opposite direction, a depletion of oxygen vacancies at one electrode leads to a discontinued CF at its tip and the OFF-state is measured, as illustrated on the right side of Figure 2.4b. By applying pulses of a suitable amplitude and duration, the CF diameter and dissolution can be controlled and multiple states in between can be reached [78]. The fact that the switching happens very locally, at the tip of one CF, makes this technology scalable.

The intrinsic stochasticity of the switching process on the other hand leads to high device variability, reduced control of multilevel operation, and significant read disturbance of the states. Another challenge is the low resistance of the ON-state (usually in the few kΩ range). Nevertheless, multiple CMOS-VCM co-integrations for inference have been shown [102–104], underlining the advance of the technology.

The other type of VCM where the switching occurs at the interface without a filament, show much less variability, controlled analog tuning, and much less read instability [130, 131]. The resistance depends on the area and can be increased simply by scaling, although scaling is still an open issue [78]. Furthermore, to reach a large resistance modulation, high electric fields are required.

A similar concept, but with 3 terminals is the VCM redox transistor (ECRAM), where the oxygen concentration of the channel is controlled by the gate [42, 117, 118]. It is a promising technology with similar benefits and challenges as the interfacial VCM, except that the read and write paths are separated. Without Joule heating, the ion movements are slow, large fields are required, and fast operations are nearly impossible.

2.2.3.3 Ferroelectric Tunnelling Junction (FTJ)

The next two device concepts both rely on the ferroelectric (FE) effect. Ferroelectricity is a spontaneous electric polarisation of a non-centrosymmetric crystalline material that can be reversed by an external electric field. Regions of opposing polarisation are called domains. Ferroelectricity is described in more details in Section 2.3.

The FTJ is a two-terminal device with an ultra-thin FE film sandwiched between two asymmetric electrodes, as illustrated in Figure 2.5a. The idea is that the FE film is thin enough for a non-zero tunnelling probability for electrons. By applying an external electric field, the polarisation of the FE layer can be switched so that it points towards either one or the other electrode. To obtain a stable polarisation in either direction, the polarisation charge of the FE layer must be screened in both electrodes. If the metal electrodes have asymmetric screening abilities, the tunnelling barrier-height

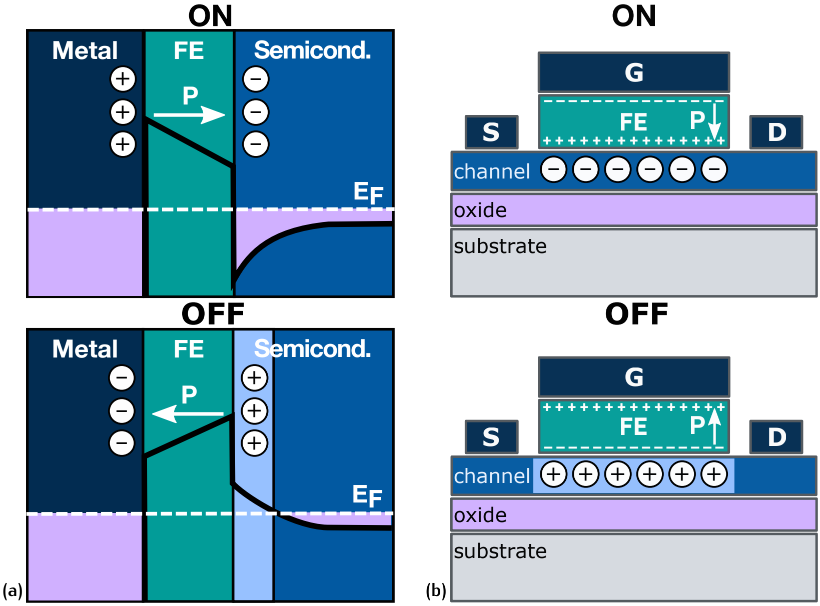


Figure 2.5: Ferroelectric memristor device types: **(a)** Ferroelectric Tunnelling Junction (FTJ) and **(b)** Ferroelectric Field Effect Transistor (FeFET).

for an electron changes with polarisation direction [132]. Apart from different electrode materials, this asymmetry can also arise from different interfaces only. For example with different terminations, with an ultra-thin dielectric layer for one of them, or because of pinned interface dipoles [133]. This effect is called Tunnelling Electro-Resistance (TER). Furthermore, if one electrode is a metal (screening width for electrons and holes in the pm range) and the other one is a doped semiconductor (screening width in the nm range) the asymmetry becomes even more notable as depicted in Figure 2.5a. While a metal can screen both positive and negative charges, semiconductors will screen one more efficiently than the other due to the difference between the majority and minority carrier concentrations. If a n-type semiconductor must screen the negative polarisation charges with positive minority carriers, the increased screening length adds to the tunnelling barrier width, leading to a change in tunnelling probability and thus electrical resistance. This effect is sometimes referred to as Giant TER.

The FTJ structure was first proposed by Esaki *et al.* [134] in 1971, but only the realisation of ultra-thin epitaxially grown FE layers [135] allowed for the demonstration of FTJs [106] in 2009. Later, Chanthbouala *et al.* showed a binary [136] and then a multi-state [137] FTJ memristor. Multiple states are possible in a FTJ because the FE layer consists of many FE domains. By an appropriate choice of the switching stimulus (amplitude or time), only a subset is switched. Despite the impressive properties, epitaxial films and in particular the required single crystal substrate and high growth temperature are not compatible with CMOS.

Since the discovery of ferroelectricity in polycrystalline HfO_2 in 2008 and the first publication in 2011 [27], the well-established and CMOS-compatible fluorite-structure material has been extensively studied, also in the context of FTJs [32, 108–110]. The main limitation with hafnia-based FE layers is that the remanent polarisation is greatly reduced for layers below 5 nm [138], which are required for tunnelling. Thus, a multilayer FTJ was recently proposed [110, 139], where the FE layer is $\sim$12 nm thick and combined with a thin $\sim$2 nm Al_2O_3 layer [140, 141]. In this case, the tunnelling takes place across the Al_2O_3 potential-well in the ON-state and across the potential-barrier combining the Al_2O_3 and FE layers in the OFF-state [142]. This circumvents the problem of diminished polarisation for thin layers, but results in very small current densities and hence larger read voltages ($2.5 \times 10^{-15}\,\mathrm{A}/\mu\mathrm{m}^2$ at 2 V for the OFF-state [110]), and larger switching voltages (6 V).

The number of ferroelectric domains scales with the area of the device, giving rise to only few domains, and thus intermediate states, in extremely-scaled devices. This is an intrinsic scaling problem of all FE-based multi-state devices. Furthermore, a trade-off has been identified between reducing the FE layer thickness to achieve scaled devices with reasonable current densities and the thereby increased self-capacitance that limits the read speed [99]. Other challenges for the hafnia-based FTJs are the stabilisation of ferroelectricity for sub-5 nm films and the observed "wake-up" effect [143]. The phase polymorphism of hafnia-based FE layers imposes a challenge as the non-FE phase has a smaller band-gap and thus, the para-electric grains act as parasitic sneak paths [144].

The switching of FTJs is field-driven and only small currents to screen the polarisation must flow, making the FE memristor a low-power memristor [145]. On individual devices, writing speeds of 20 ns to 50 ns [30, 146, 147], dynamic ranges of 7 to 16 [32, 148] (single-layer) and 10 to 100 [30, 139, 147] (multi-layer), endurances of $>10^{10}$ cycles [145], and retentions of >10 years [147] have been demonstrated for hafnia based FTJs.

2.2.3.4 Ferroelectric Field Effect Transistor (FeFET)

The FeFET is a device concept that has been known since 1957 [23] and first demonstrated in 1974 [149] as a Metal-Ferroelectric-Semiconductor (MFS) structure, which is very similar to a conventional Metal-Oxide-Semiconductor FET (MOSFET). The FeFET has three terminals, a Source (S), Drain (D), and Gate (G). It is a MOSFET where the usual high-k gate dielectric is replaced by a FE layer. Figure 2.5b illustrates a FeFET for the ON- and OFF-state. Depending on the orientation of the polarisation, an accumulation or depletion of electrons occurs in the semiconductor channel, leading to a change of resistance between S and D. It is the same electrostatic principle as for a classical FET, except that the polarisation is remanent and thus the modulation of the channels resistance as well. The state of the FeFET is sensed non-destructively between S and D. Similar to the FTJ, before the discovery of FE-HfO_2, different perovskite ferroelectrics were integrated, including $Pb[Zr_xTi_{1-x}]O_3$ (PZT), $BiTiO_3$ (BTO), and $SrBi_2Ta_2O_9$ (SBT) [150]. Because of the difficult integration of these materials on Si, large depolarisation fields and limited scaling of perovskite ferroelectrics (>100 nm thick layers), early FeFETs never saw the evolution that the flash memory underwent [36]. Nevertheless, in 2012, a 64 kbit NaND memory array was demonstrated [25].

Ferroelectricity in HfO_2 has the potential to mitigate most problems of the perovskite FeFETs: CMOS compatibility, ferroelectricity down to 5 nm thick films, wide band-gap, and good retention. [36]. Since HfO_2 is already used as high-k dielectric in the MOSFET process, hafnia-based FeFETs have seen rapid technological advances and were successfully integrated in the Front-End-Of-Line (FEOL) down to the 22 nm technology node as binary 32 Mbit arrays [39], and to the 28 nm technology node as binary 64 kbit arrays [151] and multi-state (3 states) memristors [152]. The binary characteristic and also the relatively small number of intermediate states for the ultra-scaled mutli-state FeFET is a consequence of the small number of available domains, as demonstrated by Mulaosmanovic *et al.* [152]. Their endurance is limited to $\sim 10^5$ cycles, a consequence of the thin (0.5 nm to 2 nm) Interfacial Layer (IL) grown between the FE and the Si channel. The IL serves as buffer layer to minimise inter-diffusion of elements, as enabler for high quality FE film growth, and as high quality channel interface, in contrast to a spontaneously grown native oxide [36]. Si FeFETs thus have a Metal-Ferroelectric-Insulator-Semiconductor (MFIS) gate stack. The drawback of the IL (SiO_2) is that due to its much smaller dielectric constant than HfO_2, it experiences a large voltage drop across it, thus reducing the field across the FE layer, which increases the required write voltages. Furthermore, it leads to charge trapping [153] at the IL-FE interface that screens the polarisation, reduces the influence of the polarisation on the channel, limits the readout speed, and creates a depolarisation field that destabilises the FE domains and ultimately decreases the retention [36]. Some of these issues have been addressed by stack engineering (SiON instead of SiO_2 IL [151, 154]). Moreover, the large field across the IL leads to a severe wear-out of the SiON IL and subsequent excessive charge trapping, which hinders further ferroelectric switching [155]. This is considered the reason for the limited endurance ($<10^8$ cycles [36, 156]).

The adoption of oxide-based channels is a promising alternative to mitigate some of the issues associated with Si-FeFETs. Having a semiconducting Oxide Channel (OC) makes an (oxide-)IL redundant and many of the associated concerns are reduced. Mo *et al.* [41] demonstrated FeFETs with an Indium Gallium Zinc Oxide (IGZO) channel in combination with $HfZrO_4$ (HZO) as FE layer. The fabricated device did not show any inter-diffusion or formation of an IL, maximising the field across the FE layer. The work of this thesis also belongs to the OC-FeFETs. Parts of the results were already published elsewhere [38] and can be found in Section 5. A further advantage of the OC-FeFETs is their flexible integration in the BEOL, where the size constraint can be relaxed. A much more gradual and analog be-

haviour of the resistance change with many intermediate states (18 cycles to 64 cycles) and an increased endurance ($\sim 10^7$ cycles) can be obtained [38, 157]. Furthermore, no contact implantation is needed and in contrast to Si-FeFETs, OC-FeFETs are junction-less transistors.
Some challenges still remain, in particular the scaling that intrinsically leads to less domains, imposes a trade-off between the device area and the number of intermediate states. The polycrystalline nature of hafnia-based FE layers introduces a non-uniform landscape of defects, introducing a non-uniform polarisation behaviour. This in turn gives rise to large device-to-device variability for scaled devices [36]. It is generally accepted that defects in the bulk and at the HfO_2 interface play a crucial role on the performance of prospective non-volatile memory devices. A better understanding of the observed effects and mechanisms behind it is still needed [144]. The experienced wake-up and fatigue effects, both believed to be linked to oxygen vacancy redistributions and defect generation must be tackled as well [158].

2.3 FERROELECTRICITY

Here, we will first briefly introduce the general concept of ferroelectricity and then have a closer look at hafnia based ferroelectrics. To understand ferroelectricity we must have a closer look at crystal symmetries. Let us start by defining piezoelectricity. If we apply stress to a crystal, direct piezoelectricity (see Figure2.6a) is the linear response of the charge development on the surface of the crystal to stress [159]:

$$P = \frac{Q}{A} = dX,$$

where P is the polarisation [C/m²], Q is the charge [C], A is the area [m²], d is the piezoelectric coefficient [C/N], and X is the stress [N/m²). The converse piezoelectric effect, shown in Figure 2.6b, linearly relates an applied electric field E [V/m] to the strain x on the crystal (deformation) [159]:

$$x = dE.$$

The piezoelectric coefficient d quantifies the proportionality between stress and charge and between strain and electric field. It is defined as a tensor because it depends on the direction and crystal symmetry.

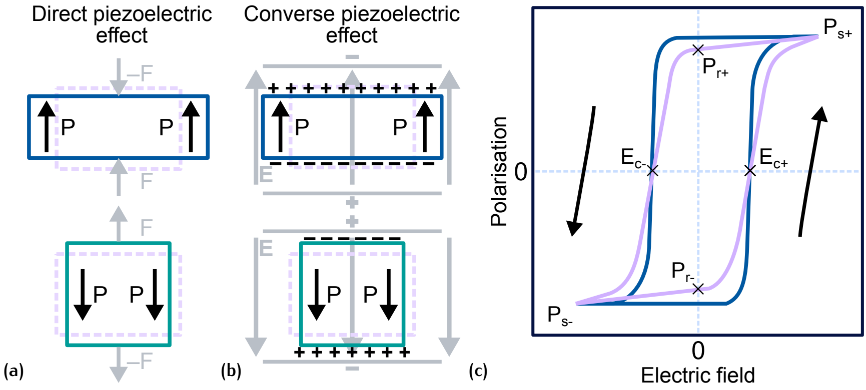


Figure 2.6: Piezoelectricity and ferroelectricity: **(a)** Direct piezoelectric effect. **(b)** Converse piezoelectric effect. **(c)** Polarisation as a function of applied electric field with the coercive field E_c, the remanent polarisation P_r, and saturating polarisation P_s.

Ferroelectricity is defined as a permanent spontaneous re-orientable polarisation that can be switched by 180° by an external electric field. Ferroelectricity is typically characterised by measuring the response of the polarisation to the electric field (dielectric displacement). The polarisation can also be looked at as a vector field that quantifies the density of dipoles in a certain volume. Per definition, the direction of the polarisation points from the negative to the positive charge. Practically, the ferroelectric polarisation can be quantified by measuring the charge that moves from one side to the other during switching. Figure 2.6b shows a dielectric layer perturbed by an electric field that aligns the dipoles and thus creates a mechanical deformation. Figure 2.6c reports a typical measurement of the polarisation with respect to the electric field. The electric field that is required to switch the polarisation from one direction to the other is called the coercive field E_c. Once switched, the polarisation remains in that direction until a large enough electric field of the opposite polarity is applied, leading to the illustrated hysteretic polarisation response. Many domains with a given dipole may coexist in a ferroelectric layer. In the ideal case, they all switch at the same E_c (blue curve). In a non-ideal case, not all dipoles have the same E_c, leading to a slope in the hysteresis (purple curve). While every dielectric material is polarisable by a strong electric field, only a ferroelectric material will have a non-zero polarisation at $E = 0$. The polarisation at zero electric field is called the remanent polarisation (P_r). The saturation polarisation

(P_s) can be higher than P_r and the difference is called the non-remanent polarisation as it is lost when the electric field is removed. This can be the case if some dipoles are not stable in one polarisation direction due to charged defects that immediately switch them back. Thus, this difference between P_r and P_s is not related to dielectric polarisation.

Piezoelectric and ferroelectric properties only exist in crystal structures lacking an inversion center, i.e. in non-centrosymmetric crystals. This means, when looking at the unit cell of a non-centrosymmetric crystal, that the charges (ions) are not distributed symmetrically along all directions, resulting in an intrinsic dipole. Usually, by slightly moving a few ions within the unit cell, the dipole switches polarity. Out of the 32 crystal point groups, only 10 have a unique polar axis in the unstrained state and thus stable electric dipoles [159]. These permanent dipoles are called spontaneous polarisation.

2.3.1 Ferroelectricity in Hafnia Compounds

Since the discovery of ferroelectricity in Si-doped hafnium oxide (Si:HfO_2) thin-films in 2011 [27], fluorite-structure binary oxides (fluorites) have attracted considerable interest. Due to the excellent compatibility with CMOS [29] and an improved scaling to 1–10 nm [160–162], integrated ferroelectrics were revived after the perovskite ferroelectrics encountered fundamental issues like scaling limits and missing integration on Si [150, 163]. Soon after the discovery of Si:HfO_2, anion (N [164]) and multiple cation (Si [27], Sr [165], La [166], Al [167], Gd [168], Y [169], Zr [161], Ge [170]) dopants that stabilise the ferroelectric phase of hafnia were investigated. Different dopants (X) display different dopant concentration windows (Hf:X ratio) to stabilise the ferroelectric phase [164, 171, 172]. $HfZrO_4$ (HZO) films have emerged as one of the most promising combinations, mainly due to the wide and more forgiving composition window (Hf:Zr ratio) [173] and due to the lowest crystallisation temperatures (500 °C [173, 174], 400 °C [108, 162]) among the aforementioned dopants. For comparison, Si-doped HfO_2, the first ferroelectric hafnia that was reported [27], has a composition window range for ferroelectricity of 4 % and crystallisation temperatures of 1000 °C are required. When using the ferroelectric HZO in device concepts such as Ferroelectric Tunnelling Junctions or Ferroelectric Field Effect Transistors for applications as artificial synapses that store analog weights, it is of advantage to use the Back-End-Of-Line (BEOL) where size constraints are relaxed. A larger area of the ferroelectric layer means

Crystal System	Point Group	Space Group	Abbreviation
Cubic	$Fm\bar{m}3$	225	c-phase
Orthorhombic	$Pbca$	61	oI-phase
Orthorhombic	$Pnam$	62	oII-phase
Tetragonal	$P4_2/nmc$	137	t-phase
Orthorhombic (polar)	$Pca2_1$	29	f-phase
Orthorhombic (polar)	$Pmn2_1$	31	f′-phase
Monoclinic	$P2_1/c$	14	m-phase
Rhombohedral (polar)	$R3m$ or $R3$	160 or 146	r-phase

Table 2.1: Space groups, point groups, space group numbers and abbreviations for the different phases of HfO_2.

more individual domains, which in turn translates into finer-grained resistance levels. To not damage the CMOS circuits, it is important to not exceed 400 °C while integrating the ferroelectric memristors. Thus, HZO is a promising candidate for integrated ferroelectrics thanks to its low crystallisation temperature and forgiving compositional window for ferroelectricity.

2.3.1.1 *Crystal Structures of Hafnia Compounds*

The stable structure of HfO_2, ZrO_2, and related materials at room temperature is the monoclinic phase ($P2_1/c$, m-phase). Other stable crystal structures exist at higher temperatures [177, 178]. They include the cubic ($Fm\bar{m}3$, c-phase) and tetragonal ($P4_2/nmc$, t-phase) symmetries, that can be stabilised at room temperature through doping [179–181] or surface enthalpy engineering (e.g. nano structuring) [182–185]. High pressure orthorhombic ($Pbca$, oI-phase) and ($Pnam$, oII-phase) structures were studied in the past because of the incompressible nature of the oII-phase suitable for ultra-hard materials [177, 178], but then proven otherwise [186, 187]. The oII-phase was found to be quenchable [188] and hence could exist at room temperature. However, all these structures are not polar as they possess an inversion center and ferroelectric properties cannot exist. The rhombohedral structure (r-phase), known to exist under stress in ZrO_2 [189, 190], was observed for strained HZO ($R3m$ or $R3$, r-phase) [191] epitaxially grown on a single crystal substrate. The r-phases of HfO_2 have a total en-

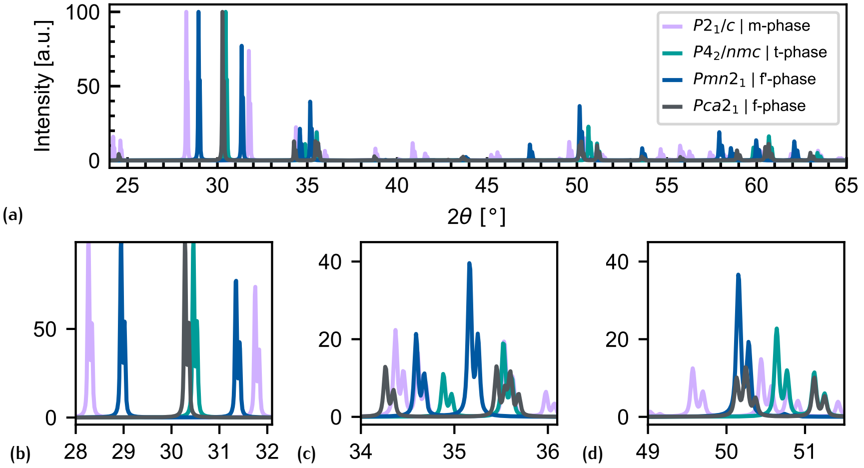


Figure 2.7: Computed X-Ray diffraction patterns for the different phases of HfO_2 depicted in Figure 2.9: **(a)** Overview. **(b-d)** Regions of interest, revealing the similarities especially between the t- and f-phase. Lattice parameters from [175]. Diffractogram created with VESTA 3 [176].

ergy that is about 158–195 meV/f.u. higher than the ground-state of the m-phase. Hence, their stabilisation is only possible through epitaxial compression by a perovskite substrate with a proper lattice spacing. Table 2.1 summarises the possible HfO_2 phases.

From the discovery of ferroelectric properties in Si:HfO_2, a non-centrosymmetric polar orthorhombic ($Pca2_1$, f-phase) structure was speculated to be responsible for the measured polarisation [27]. Due to the structural similarity to the oI-, oII-, and t-phase, XRD studies cannot clearly distinguish between these phases, especially in thin films [192]. The calculated XRD patterns in Figure 2.7 make this quite clear. The diffractograms were calculated with VESTA 3 [176] and the lattice parameters for the m-,t-, and f-phase were taken from [175], from [193] for the f′-phase. The polar f-phase was then later experimentally proven by Sang et al. [194] by using position-averaged convergent beam electron diffraction analysis. Nevertheless, the total energy of the f-phase is 49 meV/f.u. [195] higher than the bulk ground-state (m-phase) and hence, the factors leading to the formation of the f-phase are still not clear. Many factors have been proposed such as doping [196–198], oxygen vacancy incorporation [196, 199], grain size (high surface to volume ratio) [200–202], film thickness [138, 202], tensile strain [148, 201, 203], confinement by the top electrode [196, 204], and

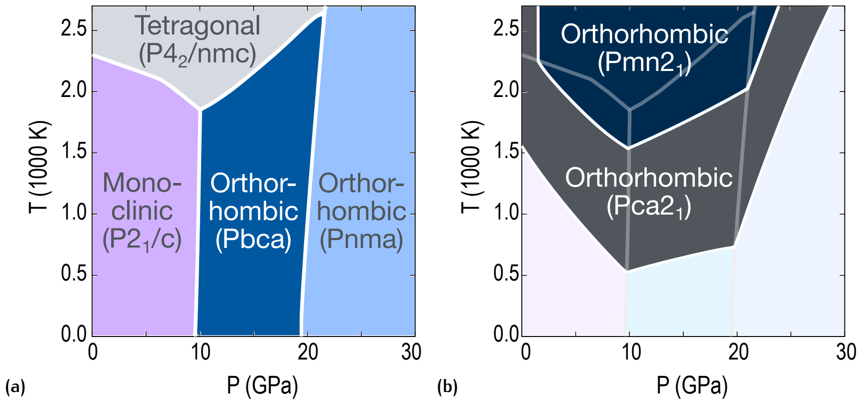


Figure 2.8: Computed phase diagram: **(a)** Equilibrium phases. **(b)** Regimes where the free energy difference between the f- or f′-phase and the equilibrium phases are small ($< k_B T/5$). Redrawn from [193].

surface energy [173, 195, 202]. Note that most of these factors are not independent. e.g.the film thickness has an influence on the grain size, which changes the surface energy and internal pressure. Although ferroelectricity has been shown in films that are much thicker than 10 nm [205], they all share crystallite grain sizes below 10 nm. In fact, to avoid larger grain sizes, that favour the m-phase when increasing the layer thickness, very thin Al_2O_3 layers can be alternatively added between each 10 nm HZO layer of thick films. The large internal pressure of small crystallites favours the lower volume c- or t-phases over the m-phase [191, 206]. Today, it is generally accepted that the small crystallite size plays a crucial role in stabilizing the f-phase [173, 202], which is postulated to be the transformation phase between the t- and m-phases [193, 196, 207].

Huan *et al.* [193] used the minima-hopping method [208] to identify low-energy phases at various pressures and temperatures (Figure 2.8a). They then singled out polar phases by applying the group theoretical symmetry reduction principles developed by Shuvalov [209]. Table 2.1 presents all low-energy phases of hafnia that are possible at various pressures and temperatures. Additionally, the rhombohedral phase is added as it was shown to exist under epitaxial strain. Two space groups, namely the $Pca2_1$ (f-phase) and $Pmn2_1$ (f′-phase) are found to be ferroelectric with a small free energy difference with regard to the equilibrium phases (Figure 2.8b). They represent a distorted version of the t-phase in its [110] and [100]

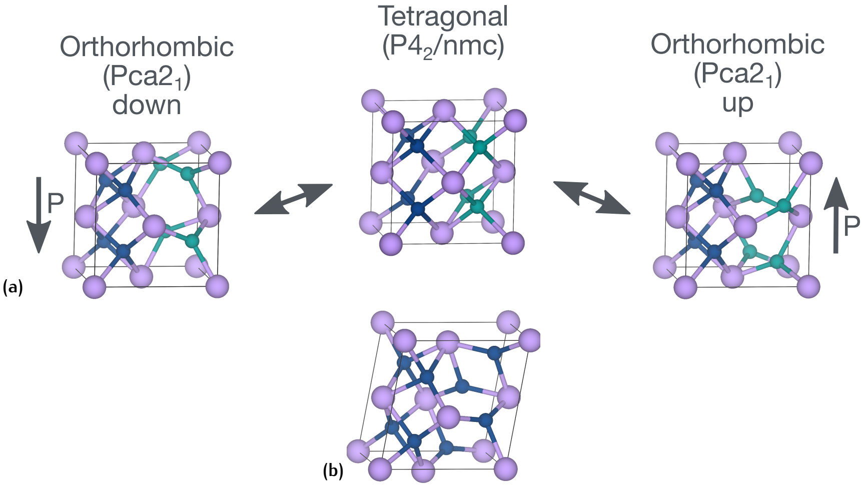


Figure 2.9: Crystal phases [176]: **(a)** Schematic of the switching from the f-phase with polarisation down through the t-phase to the f-phase with polarisation up. Lattice parameters from [175]. **(b)** m-phase crystal structure. Lattice parameters from [175].

directions, respectively. The spontaneous polarisation is switchable (180°) with small energy barriers of 40 meV for the f-phase and 8 meV for the f′-phase [193]. In both cases the switching path goes through the t-phase as schematically shown for the f-phase in Figure 2.9a [193, 210]. The oxygen atoms that have a large displacement between polarisation directions are coloured in teal, while the other oxygen (blue) and hafnium (purple) atoms only undergo minor displacement. During crystallisation it is therefore important to go through the t-phase before the f-phase can be reached. For comparison, Figure 2.9b reports the crystal structure of the m-phase.

Materlik *et al.* [175] investigated the Gibbs/Helmholtz free energies as a function of temperature, pressure, strain, and surface energy via density functional theory (DFT). They found that by temperature, pressure, and compressive film stress alone, the f-phase in HZO is not stabilised over the m-phase. ZrO_2 has a size-driven m→t-phase transformation and energy crossover for crystallites below 24 nm [211], while for HfO_2 it is below 3 nm [184]. Therefore, the size-driven transformation requires a much higher surface area to volume ratio for HfO_2 than for ZrO_2. Mixing HfO_2 with ZrO_2 to form HZO thus should ease this transition compared to pure HfO_2. Finally, by considering the surface energy in the Helmholtz equation, the f-phase has the lowest free energy of all phases for grain sizes between 8 nm and 16 nm in HZO, which is in good agreement with literature [175].

Batra *et al.* [212] calculated the surface energy for various HfO_2 phases and surface planes from first-principles using DFT. The smaller values they obtained indicate that the results by Materlik *et al.* [175] might be overestimated and their conclusion, not the model, for a stable f-phase at small crystallites might not be accurate.

Park *et al.* [202] quantitatively compared the experimental observations in their HZO films of different compositions and thicknesses to the modified surface energy model proposed by Materlik *et al.* [175]. They measured the grain size distribution as a function of composition and thickness. This data was used to compute the evolution of the f-, t-, and m-phases as a function of the composition and thickness based on the aforementioned thermodynamic model [175]. When comparing experimentally measured P_r values to the evolution of the o-phase, it appears that the overall trends are consistent, but there remains a critical mismatch. Park *et al.* pointed out that the assumption of the model that the grain-boundary and interface energies are identical to the free-surface energy is problematic for polycrystalline films [202]. Thus, they assumed a grain boundary energy to be one third of the surface energy, which still resulted in the m-phase being the stable phase, where P_r was experimentally measured to be the largest. Furthermore, the model assumes a fully amorphous layer that is transformed into a crystalline one, neglecting the observed small crystallites after ALD deposition [202]. By considering the interface energy between small crystallites and the amorphous surrounding, the f- and t- phases are energetically more favourable with respect to the m-phase. They therefore concluded that the f- and t-phase crystalline nuclei, formed during ALD deposition (~2 nm), retain their phase during post deposition anneals and act as seeds for the full crystallisation of the entire film [202]. Still, the m-phase is the stable phase across the entire temperature range involved. Later, the same group used kinetic considerations to show why the f-phase still can be stabilised [213]: The large kinetic barrier for the transition from the t- to the stable m-phase (~300 meV/f.u.) prohibits the undesirable phase transition. The transition form the t-phase to the f-phase on the other hand has a negligible energy barrier (~30 meV/f.u. [213], 40 meV/f.u. [193]) and could readily occur during cooling.

The thermodynamic model considering bulk, grain-boundary, and interface energy effects [175, 202, 212] finds the m-phase to be stable for typical crystallisation temperatures and grain sizes. However, by considering small f- and t-phase crystallite nuclei (∼2 nm) [202] resulting from the ALD deposition and the large kinetic barrier hindering the t- to m-phase transition [213], the observed stable ferroelectric o-phase in thin-films could be explained.

2.3.1.2 Wake-Up and Fatigue

Compared to conventional perovskite ferroelectrics, hafnia-based ferroelectric layers generally exhibit a pronounced "wake-up" effect. In the pristine state, the "Polarisation vs. Electric-Field" (P-E) curves are anti-ferroelectric like, with two switching events, seen in the "Current vs. Electric-Field" (I-E) measurement for each polarisation direction [214]. With an increasing number of electric field cycles, the hysteretic curve gets de-pinched, the Remanent Polarisation Window $RPW = P_{\mathrm{r}+} - P_{\mathrm{r}-}$ increases and a typical ferroelectric P-E curve is obtained. This is an opposite effect to the usually observed fatigue effect in ferroelectric films where the RPW decreases with the increasing number of field cyclings. After the wake-up, hafnia-based ferroelectrics show the typical fatigue effect [215], which manifests itself as a reduction of the RPW with increasing electric field cycling.

In the pristine state, a strong internal bias field is present [216] that is linked to an asymmetric distribution of charges at the two electrodes, polymorphic phase distribution, and charges located at domain walls [214, 216–218]. This asymmetry can be caused by the deposition process where the bottom electrode gets more oxidised (e.g. by the O_2-plasma of the ALD-HfO_2 deposition) and thus has less oxygen vacancies. Among other electrodes, this was shown for TiN [219]. During the crystallisation, the top electrode interface is oxidised (with O_2 from the HfO_2) leading to an accumulation of oxygen vacancies [217]. This, together with a nitrogen diffusion into the HfO_2, could lead to the stabilisation of the t-phase at this interface [164]. During the wake-up cycling, it is suspected that the oxygen vacancies diffuse [220], resulting in a field-induced phase change at the electrode interface [218, 221]. The reduction of this t-phase Interfacial Layer (IF) would lead to a reduced depolarisation field E_{dep}:

$$E_{dep} = \frac{P_r}{\epsilon_{FE}\,\epsilon_0}\left(1 + \frac{\epsilon_{IF}\,t_{FE}}{\epsilon_{FE}\,t_{IF}}\right)^{-1},$$

where ϵ_{FE} and ϵ_{IF} are the permittivities and t_{FE} and t_{IF} the layer thicknesses of the ferroelectric and IL respectively. The reduced E_{dep} could explain the wake-up behaviour. Some studies report a reduced wake-up effect when using higher deposition temperatures [222], cycling at enhanced temperatures [220, 223], or ALD ozone-dosage and electrode engineering [224]. However, this tends to lead to increased leakage currents and early dielectric breakdowns. Furthermore, the wake-up effect was found to depend on the accumulated pulse time, rather than on the number of cycles, thus being frequency dependent [220, 223]. This hints again to the proclaimed oxygen vacancy redistribution.

The fatigue effect that directly influences the endurance of a device was attributed to multiple possible causes. Leakage current defect spectroscopy [217] revealed that the leakage and *RPW* are not strongly influenced by unipolar stress, in contrast to bipolar stress, which results into an increase of the leakage current and consequently a reduction of the *RPW*. This was directly correlated to an increase in defects. The continuous ion displacement between the polarisation directions induces an increased defect generation, fatigue, and ultimately dielectric breakdown. Masuduzzaman *et al.* [225] proposed "hot atom" degradation as possible explanation. They could identify a cycling frequency and voltage dependent degradation, which they attributed to a local overshoot of the polarisation (from the gained energy at the domain walls to overcome the central energy barrier between polarisations) until the energy is dissipated in the surroundings. These overshoots reduce the activation energy [226] required for bond breakage. They could also show an increase in endurance by applying pulses that have a larger transition time. Hot atom degradation is difficult to separate from contributions of charge trapping at redistributed or generated defects [214]. Experiments with increasing waiting times between bipolar cycling stress showed an increased recovery of the fatigue. The authors argued that this might come from the detrapping of electrons from occupied defects that results in depinning of domains and thus recovery of the *RPW* [217]. Another explanation for the time-dependent recovery could be the recombination of interstitial oxygen ions and oxygen vacancies [227]. Recovery of the *RPW* was also observed by temperature treatments [228], but is accompanied by an early fatigue, revealing the permanent nature of the degradation, e.g. formation of a conductive filament along grain boundaries due to a high concentration of oxygen vacancies [229].

In summary, we can conclude that the wake-up effect arises from a redistribution of interface defects (oxygen vacancies), reducing the build-in field. During fatigue, oxygen vacancies are possibly created by hot ion damages or by the conversion of neutral vacancies into charged ones. A high concentration of oxygen vacancies at domain walls leads to domain wall pinning and ultimately to the formation of a conductive filament [214].

For more details on the switching kinetics of hafnia-based ferroelectrics we refer the readers to a recent review by Lee *et al.* [230].

3

METHODS

> *The use of AI for social control and oppression is already emerging, even in the face of developers' best of intentions.*
>
> — Meredith Whittaker

3.1 MATERIAL CHARACTERISATION METHODS

3.1.1 X-Ray Diffraction

For the development of thin-films, their structural characterisation is a highly relevant issue. An important characteristic is the crystallographic structure. Depending on the growth method, materials can be amorphous or crystalline. The temperature or pressure determine the different crystallographic phases that are energetically favourable in a material. In an X-Ray Diffraction (XRD) measurement, a beam of parallel X-rays with the same phase and wavelength impinges a crystallised material and scatters at the materials periodic grid of atoms (Figure 3.1a). At certain incident angles, this leads to constructive interference because of the cumulative effect of reflections in successive crystallographic planes. X-Rays are used because their wave length is comparable to the inter-atomic distances ($\sim$150 pm), and thus form an excellent probe. For constructive interferences, the incident (and reflection) angle θ, the X-ray wavelength λ, and the lattice spacing d are related by the Bragg condition [235]:

$$n\lambda = 2dsin\theta, \tag{3.1}$$

where $n = 1, 2, 3...$ is the diffraction order. In a $\theta/2\theta$ scan (Figure 3.1b), the angle of the incident beam and of the detector are varied and a reflected intensity profile, which characterises this lattice spacing, is recorded. Multiple databases exist, also open-source ones [236], where the measured profiles for many compounds can be compared to already known ones. The $\theta/2\theta$ scan is also called symmetric scan because the incident beam angle is always equal to the angle between the sample and the detector, as il-

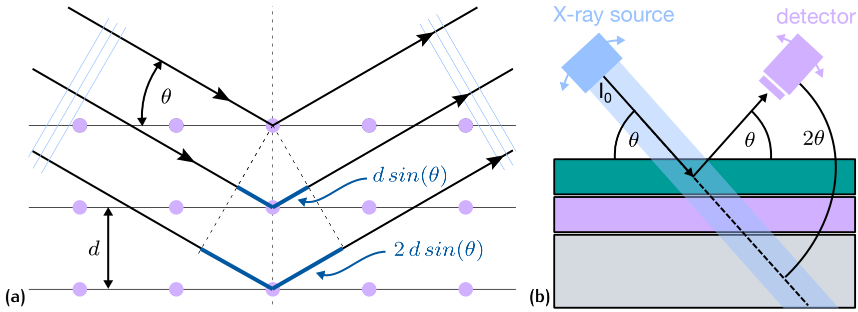


Figure 3.1: X-Ray Diffraction (XRD): **(a)** Parallel beams with the same phase and wavelength are scattered at different atoms. Constructive interference occurs when the Bragg condition is met. **(b)** Illustration of the $\theta/2\theta$ scan.

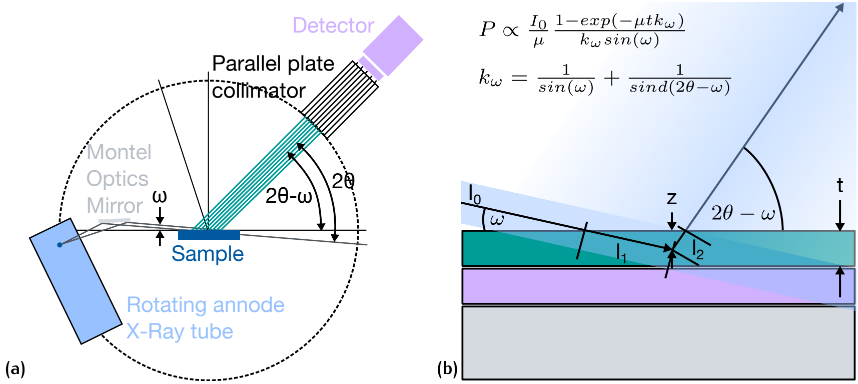


Figure 3.2: Grazing-Incidence X-Ray Diffraction (GIXRD): **(a)** Schematic of the arrangement of the sample, X-ray source, and detector and the definition of important angles. **(b)** Illustration of the sample interaction with the X-Ray beam.

Element	Attenuation coefficient μ/ρ [cm^2 g^{-1}]	Atomic mass [u]	Density ρ [g cm^{-3}]	Absorption coefficient μ [µm^{-1}]
Hf	157.1 [231]	178.49	13.3 [232]	0.2089^{*}
Zr	135.6 [231]	91.22	6.52 [232]	0.0884^{*}
O	11.63 [231]	15.99	-	-
HfO_2	$134.99^{\dagger}$	210.5	9.68 [232]	0.1306^{*}
ZrO_2	$103.41^{\dagger}$	123.22	5.68 [232]	0.0587^{*}
$Hf_{0.57}Zr_{0.43}O_2$	$125.31^{\dagger}$	172.96	$7.95^{\S}$	0.0997^{*}
Zn	58.75 [231]	65.38	7.134 [232]	0.0419^{*}
ZnO	$49.48^{\dagger}$	81.38	5.61 [233]	$0.0281^{*\ddagger}$
W	170.5 [231]	183.84	19.3 [232]	$0.3290^{*\ddagger}$
WO_3	$137.61^{\dagger}$	231.84	7.2 [232]	0.0991^{*}
Ti	202.3 [231]	47.87	4.506 [232]	0.0911^{*}
N	7.562 [231]	14.01	-	-
TiN	$158.22^{\dagger}$	61.87	5.21 [232]	0.0824^{*}
Si	64.68 [231]	28.09	2.329 [232]	0.0151^{*}
SiO_2	$36.42^{\dagger}$	60.09	2.648 [232]	0.0096^{*}

† The values were calculated with Equation 3.6.
∗ The values were calculated by $\mu = \mu/\rho \cdot \rho$.
‡ These values were included to compare with [234] for sanity check.
§ The value is a linear combination of HfO_2 and ZrO_2.

Table 3.1: Attenuation coefficient, atomic mass, density, and calculated absorption coefficients for the most common elements and compounds used in this work. The attenuation coefficients were extracted from [231] for a photon energy of 8 keV. CuK$_\alpha$ X-Rays used in this work have an energy of 8.04 keV ($\lambda_{CuK\alpha} = 154\,\mathrm{pm}$).

lustrated in Figure 3.1b. As evident from Figure 3.1a, the symmetric scan probes the lattice planes that are parallel to the surface. When dealing with very thin films, a large fraction of the intensity in a $\theta/2\theta$ scan originates from the substrate. The beam path through the thin film is too short to produce Bragg reflections with a sufficient peak-to-noise ratio. By reducing and fixing the incident angle to a very small value ($\omega < 1°$), the beam path through the thin layer of interest is maximised and the structural information contained in the intensity profile mainly stems from the thin layer. Such a scan is called Grazing-Incidence X-Ray Diffraction (GIXRD). Figure 3.2a shows a common configuration of the XRD tool for a GIXRD scan. The incident angle ω is fixed and only the detector is moved. The scattering angle between the incoming and reflected beam is still defined as 2θ. With a fixed ω, the angle between the surface and the reflected beam thus becomes $2\theta - \omega$. During a GIXRD scan, Bragg reflections arise from lattice planes that are neither parallel to the surface nor to each other [237]. Due to the fixed ω, the orientation of the probed lattice plane constantly changes with θ. For polycrystalline films with random grain orientation, GIXRD diffraction patterns are comparable to those obtained by $\theta/2\theta$ after subtraction of the substrate-contribution. However, different beam/surface angles lead to different attenuations: Figure 3.2a shows the beam path for a GIXRD scan in a layer with a thickness t. If the incident beam is scattered at an atom that is located at a depth z from the surface, $l_1 = z/sin(\omega)$ is the beam path before scattering and $l_2 = z/sin(2\theta - \omega)$ after scattering. From this we can derive the configuration factor:

$$k_\omega = \frac{l_1 + l_2}{z} = \frac{1}{sin(\omega)} + \frac{1}{sin(2\theta - \omega)}. \tag{3.2}$$

The intensity scattered at depth z scales with $exp(-\mu l_1)$ (attenuation) and with $I_0/sin(\omega)$ (beam is inclined, less incident intensity per film surface area for small ω), where μ is the absorption coefficient of the material. After being scattered, the beam is further attenuated by $exp(-\mu l_2)$ by leaving the layer. Hence, the scattered intensity at the detector from depth z $dP(z)$, is proportional to:

$$dP(z) \propto \frac{I_0}{sin(\omega)} exp(-\mu z k_\omega) dz. \tag{3.3}$$

Integrating Equation 3.3 over all depths (0 to t) yields the total scattered intensity:

$$P_t \propto \frac{I_0}{sin(\omega)} \int_0^t exp(-\mu z k_\omega)\, dz = \frac{I_0}{\mu} \frac{[1 - exp(-\mu t k_\omega)]}{k_\omega sin(\omega)}. \tag{3.4}$$

For an infinitely thick layer we obtain $P_\infty \propto I_0/\mu k_\omega sin(\omega)$. Now we can derive the absorption factor A_ω, the amount that a Bragg reflection from a thin-film sample is reduced when compared to an infinitely thick sample:

$$A_\omega = \frac{P_t}{P_\infty} = [1 - exp(-\mu t k_\omega)]. \tag{3.5}$$

The consequence is a quite constant A_ω for 2θ in the range of 20° to 60°, contrary to the case of the symmetric scan where the absorption factor $A_{\theta 2\theta} = [1 - exp(-2\mu t/sin(\theta))]$ quickly decays with increasing 2θ.
The mass attenuation coefficients μ/ρ, where ρ is the density of the materials, can be found in the NIST5632 [231] database. For compounds, a simple additive formula exists:

$$\frac{\mu}{\rho} = \sum_{i=1} w_i(\frac{\mu_i}{\rho_i}), \tag{3.6}$$

where w_i is the fraction by weight of the i^{th} atomic constituent. Table 3.1 summarises the calculated μ for the most common elements and compounds used in this work. Finally, we can insert some numbers typical for a GIXRD scan of HZO ($Hf_{0.57}Zr_{0.43}O_2$). The most interesting peak for ferroelectric HZO is located at $2\theta = \sim 30.5°$. We have to use an incident angle $\omega = 0.4°$ slightly above the total reflection angle θ_c to maximise the beam path in the thin HZO layer (usually $t = 10\,\mathrm{nm}$). From Table 3.1 we get $\mu_{HZO} = 0.0997\,\mathrm{\mu m^{-1}}$. With these values we obtain an intensity gain by using GIXRD over a symmetric scan of:

$$\frac{A_\omega}{A_\theta 2\theta} = \frac{[1 - exp(-\mu_H ZOtk_\omega)]}{[1 - exp(-2\mu_H ZOt/sin(\theta))]} = \frac{0.1348}{0.0076} = 17.85. \tag{3.7}$$

By reducing the HZO film thickness from 10 nm to 5 nm and 3 nm , the absorption factor A_ω decreases from 0.1348 to 0.0698 and 0.0425, respectively. Thus, a reduction of the intensity measured at the detector of 48 % and 68 % can be expected. Performing a symmetric $\theta/2\theta$ scan is clearly disadvantageous for such thin-films. One last point has to be noted here. A $\theta/2\theta$ scan at very low angles (0° to 10°) can be used to characterise the thickness of thin-films. Due to the reflection and refraction of the beam at boundaries of layers with different indices of refraction (n), constructive interference patterns can be observed, similar to optical ellipsometry [237]. Such scans are called X-Ray Reflectivity (XRR) measurements and are a great non-destructive option to determine layer thicknesses.

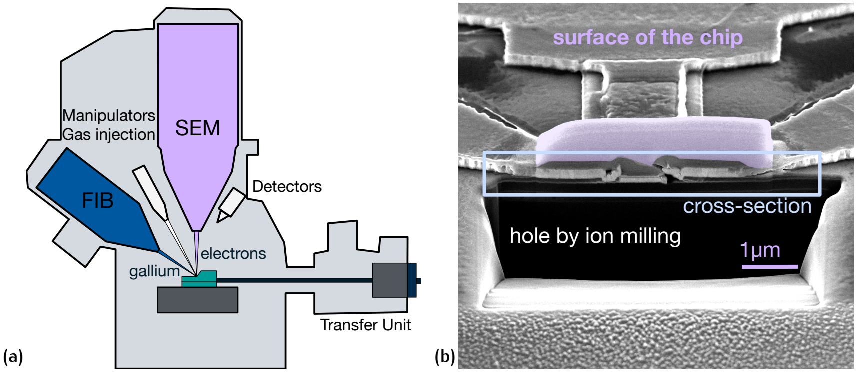


Figure 3.3: Focused Ion Beam (FIB): **(a)** Schematic of the focused ion beam system, showing the availability of an electron and ion beam for imaging and machining of samples. **(b)** Scanning Electron Microscope (SEM) image taken at an angle of 52° showing a cross-section (light blue box) that was revealed by ion-milling (black hole). To protect the structures, a Pt layer was deposited on top of the area of interest (falsely coloured in purple).

3.1.2 Focused Ion Beam

A sophisticated tool for the structural characterisation of thin-film layers is the Focused Ion Beam (FIB) system. A schematic of a typical FIB system is provided in Figure 3.3. Modern systems supplement the FIB with a Scanning Electron Microscope (SEM), building together a dual-beam (FIB-SEM) setup. Most commercially used FIB instruments use Ga ions to form the beam. Ions have much larger masses in contrast to electrons. Thus, the ion beam is used to sputter the surface, which enables precise machining of the sample. The electron source on the other hand is used to image the structures without damaging them. Furthermore, it is possible to inject gas into the chamber to perform an ion-beam or electron-beam activated deposition of materials such as platinum or carbon. This is especially useful when the surface of interest needs to be protected.

FIB analysis is of particular use when the area of investigation is not the surface, but a cross section. First, the ion- or electron-beam is used to deposit a protecting layer of Pt above the region of interest. Then, materials are removed with the ion-beam to uncover the cross section. In a third step, the sample is tilted and the cross section is imaged by the SEM.

In semiconductor processing, a visual inspection of the structures at the nanoscale can often clarify encountered inconsistencies such as failed electrical measurements, unexpected de-colouration of the layers, or delamination of layers. Furthermore, it is a great method to perform process control, e.g. assess if vias in the passivation layers were completely opened, if layer thicknesses are as expected, or if metal layers are contacting each other. While non-intrusive methods like XRD are great to characterise global film properties, the inspection with a FIB system allows to visually resolve the cross section of local structures down to feature sizes of about 10 nm.

Increasing the resolution of a cross-section to sub-nm is possible by Scanning Transmission Electron Microscopy (STEM). For such an experiment, the specimen's cross section must be thinned to a thickness of about 100 nm-200 nm to be able to transmit electrons through it. The FIB system offers the perfect prerequisite to prepare such thin specimens. A precisely movable needle can be used to extract the thin specimen from the sample and to attach it to the specimen holder used for STEM. Figure 3.3b shows an SEM image of a cross section revealed with the FIB ion gun. To protect the area of interest, a Pt layer was deposited (falsely coloured in purple). The large black area is the hole milled by the ion gun. The SEM image was then taken at an angle of 52° to look at the revealed cross section (light blue box).

3.2 SEMICONDUCTOR PROCESSING

3.2.1 Deposition: Atomic Layer Deposition

Atomic Layer Deposition (ALD) is a process to grow thin-films in a very controlled manner, at relatively low temperatures. The main advantages of ALD with respect to other deposition techniques are the thickness control at the Angstrom level, tunable film compositions, and the exceptional conformality on high-aspect ratio structures. They are all delivered from the sequential, self-saturating, gas-surface reaction control of the deposition process [238]. Other methods such as Molecular Beam Epitaxy(MBE) or Pulsed Laser Depsition (PLD) also offer great thickness and composition control, but at much higher temperatures and not in a conformal way. The most prominent examples that were made possible by ALD are the integration of a thin high-k gate dielectric with a well controlled thickness

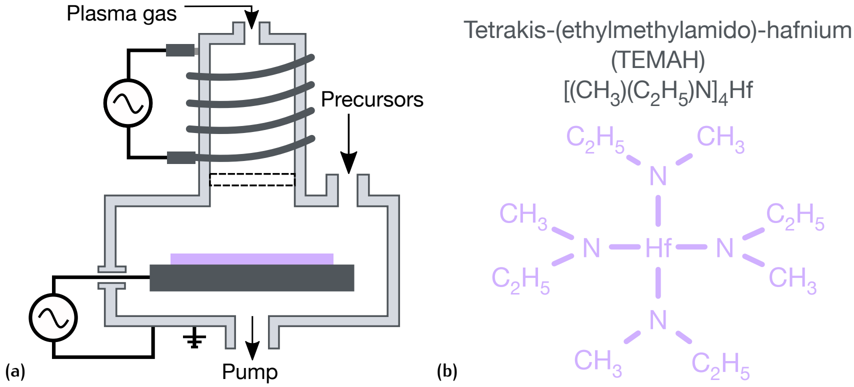


Figure 3.4: Atomic layer deposition: **(a)** Simplified schematic of the process chamber. **(b)** Formula for the Hf precursor.

by Intel [29] for the 45 nm technology node and later the realisation of 3D transistors such as Fin Field-Effect Transistors (FinFETs) [239]. The later example was made possible due to the self limiting gas surface reaction that restricts the film growth to one layer at a time and thus enables the conformality of high-aspect ratios and three dimensionally-structured materials. Many materials ranging from insulators to semiconductors and metals can be deposited by ALD. This method offers many elements to choose from to create the desired composition [238]. The main limitation of the availablity of a material by ALD is the restricted choice of reaction pathways. To stay in the self limiting growth regime, very specific reactants and counter reactants for the deposition of the desired materials are required and synthesised. Furthermore, the reactants must be volatile to be in the gas phase at moderate temperatures. Such reactants are called precursors and should be stable until they reach the sample surface.

In the following part, the growth process by ALD is explained using the example of HfO_2. The ALD process is performed in a sealed reactor as depicted in Figure 3.4a. The precursor for Hf is TEMAH[1] and is depicted in Figure 3.4b. One complete ALD deposition cycle is illustrated in Figure 3.5. In a first step, the precursor is introduced into the process chamber. The ligaments attached to the Hf element ensure its volatility at low temperature and that it reacts in a self limiting way with the surface, leaving no more than one monolayer. Excess precursors and reaction by-products remain volatile in the chamber. In a second step, these by-products and

---

[1] TEMAH: tetrakis-(ethylmethylamino)-hafnium $[(CH_3)(C_2H_5)N]_4Hf$

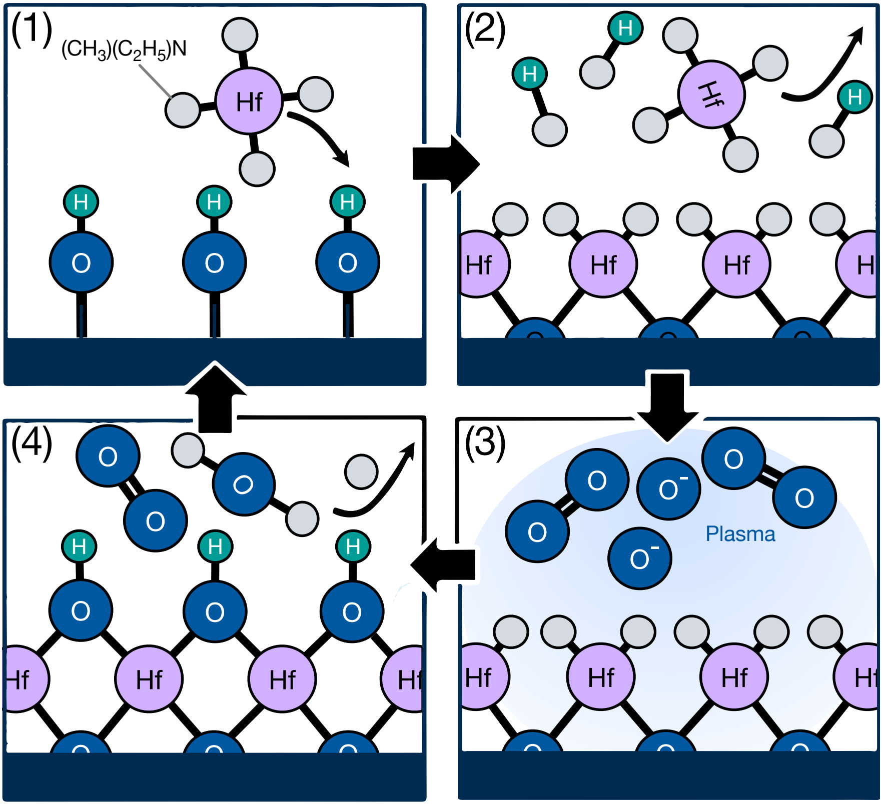


Figure 3.5: Single atomic layer deposition cycle: **(1)** The first reactant/precursor is introduced into the chamber and reacts with the surface in a self-limiting way. **(2)** By-products are purged by an inert gas. **(3)** The second reactant, here O_2-plasma, is applied to complete surface reaction. **(4)** By-products are purged by an inert gas to complete the cycle.

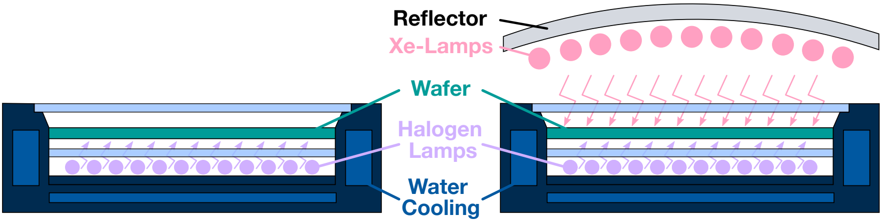


Figure 3.6: Rapid thermal annealing systems: **(left)** Widely used Rapid Thermal Annealer (RTA) with infrared heating. **(right)** Flash Lamp Annealer (FLA) which is an extension of the RTA by the ability to release short high-energy flash pulses through Xe-lamps.

excess precursors are purged with an inert carrier gas (typically N_2 or Ar, not shown here) and only the layer on the surface remains, still having ligaments attached. This is called a half reaction. The third step introduces the second precursor or reactant to complete the reaction on the surface. In our Plasma Enhanced ALD (PEALD) the second reactant is a suitable plasma. For oxides like HfO_2 an O_2-plasma is used. The oxygen radicals remove the organic ligaments and complete the reaction with the monolayer of Hf at the surface. The by-products are then purged by a inert gas and one cycle is complete. This is repeated with the same precursor until the desired thickness is reached. By alternating between different precursors, various material compositions can be grown. For $HfZrO_4$, for example, alternating cycles of TEMAH and ZrCMMM[2] precursors are used. TDMAT[3] precursors can be used to grow TiO_2. In our case TiN electrodes are grown by changing the O_2-plasma to a N_2/H_2 plasma.

3.2.2 Crystallisation: Millisecond Flash Lamp Annealing

The most common technique for the crystallisation of thin-films in the semiconductor industry is by Rapid Thermal Annealing (RTA). As the name suggests, such annealing systems have the capability to heat a sample with a fast and controlled temperature ramp of up to 100 °C/s on wafer scale, usually by near infrared lamps (Figure 3.6, left side). Often these systems have water-cooled chamber walls for a rapid cooling, which is still much slower than heating. For the crystallisation of ferroelectric HZO, we

---

[2] ZrCMMM: bis(methyl-η5-cyclopentadienyl)methoxymethylzirconium $(CH_3C_5H_4)2Zr(OCH_3)CH_3$

[3] TDMAT: Tetrakis-(dimethylamino)titanium $[(CH_3)_2N]_4Ti$

use a Flash Lamp Annealer (FLA). It consists of a xenon flash lamp array located at the top of the chamber and a conventional halogen lamp heater located under the sample holder (right part of Figure 3.6) [240]. The halogen lamp heater is used to preheat the sample, similar to a RTA. The xenon flash lamps are employed to release a flash pulse onto the sample. Prior to the release of the flash, a large capacitor is charged to support the instant release of high energy pulses. The flash duration can be set from 0.3 to 20 ms, where this duration refers to the full-width at half-maximum of the energy pulses. Energy densities of up to 110 J/cm² are possible in the FLA. The resulting temperature of the sample during a millisecond Flash Lamp Annealing (ms-FLA) thus depends on the pre-heat temperature, the flash energy, and flash duration. The pre-heat temperature is monitored by a thermocouple. Due to the fast nature of the flash pulse, there is no possibility of monitoring the temperature spikes during the pulse. The actual energy absorbed by a sample during a flash annealing process depends on the material structure of the sample. The spectral energy density of the flash lamp pulse ranges from a wavelength of about 400 nm to 1000 nm, having the largest intensity between 400 nm and 700 nm [240].

3.3 ELECTRICAL CHARACTERISATION METHODS

3.3.1 CTLM

A simple and effective way of keeping track of the oxidation state of our WO_x channels, is by monitoring their resistivity (ρ). A convenient method of measuring ρ for semiconductors is by Transmission Line Measurements (TLM). If designed correctly, not just ρ, but also the contact resistance (R_c) and transmission length (L_T) of the contact can be determined. The key idea for this method is to measure the resistance through the semiconductor for varying distances d. Therefore, multiple contacts with different spacings are necessary in the TLM model. This is depicted in Figure 3.7a. Here, W is the width of the semiconductor, Z the width of the contact, and L. The total resistance (R_T) measured between two of these contacts (A, B) can be split into multiple components, as illustrated in Figure 3.7c [241]:

$$R_T = 2R_m + 2R_c + R_{semi} \approx 2R_c + R_{semi} \tag{3.8}$$

in which R_m is the resistance of the metal contact and R_{semi} the resistance of the semiconductor. For metal contacts (in our case W), R_m is negligible

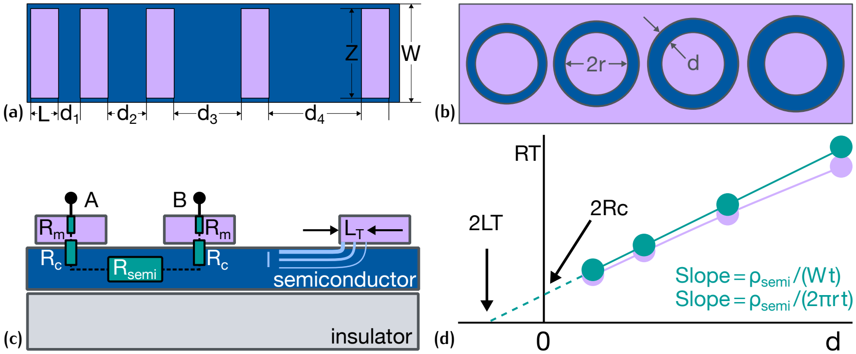


Figure 3.7: Illustration of a TLM layout and the resulting total resistance: **(b)** TLM layout with a series of square contacts. **(b)** Cirular TLM layout (CTLM). **(c)** Breakdown of the resistive components when measuring the resistance between two contacts (left) and current path from the semiconductor to the metal contact pad with the transfer length (L_T). **(d)** Typical plot of the total resistance (R_T) as a function of the contact spacing d. For CTLM, the original measurement (purple) is corrected by a factor depending on the radius (r) and d (teal).

as compared to the other two components and will be omitted for the further discussion of the TLM. A closer look at the current path from the semiconductor to the metal is depicted on the right in Figure 3.7c. Because the resistance in the metal is much lower than in the semiconductor and because the current flows through the path of least resistance, the current will transition to the metal as early as possible. Thus, the current density is largest at the edge of the contacts and drops exponentially with distance to the edge. The distance at which the current density has dropped by $1/e$ is defined as the transfer length L_T. L_T therefore defines the distance over which most of the current flows in or out of the contact. Consequently, the effective contact area is $A_{c,eff} = L_T \cdot Z$ [241]. The green data points in Figure 3.7d show a typical measurement of the resistance between two contacts (R_T) for different d. If one assumes that all contacts have the same R_c, the increasing R_T with d can be attributed to a increase of R_{semi} [241]:

$$R_{semi} = \frac{\rho_{semi}}{t} \cdot \frac{d}{W}$$

in which ρ_{semi} is the resistivity of the semiconductor, t is the thickness and W the width of the semiconductor. Thus, from the slope of the linear regression in Figure 3.7d we can extract ρ_{semi}. At zero distance (y intercept)

$2R_c$ can be extracted. The value $-2L_T$ is defined at the point where the resistance is zero. This would be the situation in which the current flows directly from one contact edge to the other.

When the semiconductor width W is wider than the contact width Z, this approximation becomes invalid because the current flows around the contacts is not accounted for. As a consequence, the semiconductor film needs to be patterned to the width of the contacts. To avoid this patterning step or the problem that $W \neq Z$, circular test structures can be used. Such structures are depicted in Figure 3.7b. They consist of an inner contact of radius r, a gap of width d, and a surrounding contact. In circular structures with $r \gg 4L_T$, $r \gg d$, and by taking the transfer length into account, R_T can be expressed as follows [241]:

$$R_T = (2R_c + R_{semi}) \cdot C = \left(2R_c + \frac{\rho_{semi}}{2\pi rt}(d + 2L_T)\right) \cdot C$$
$$C = \frac{r}{d} \ln\left(1 + \frac{d}{r}\right), \quad (3.9)$$

where C is a correction factor. It is necessary to compensate for the difference between the linear and circular TLM layouts. Otherwise, R_T would be underestimated. For practical radii up to about 200 µm and spacings of 5 µm to 50 µm, the correction factor must be applied to obtain a linear fit. The resulting linear data by applying the correction factor to the original data (purple) is depicted in Figure 3.7d. From Equation 3.9, it is evident that ρ_c can be calculated from the slope of R_T after applying the correction factor to the data [241]:

$$\rho_{semi} = \frac{\partial R_T}{\partial d} \cdot 2\pi rt.$$

So far we have used the contact resistance R_c to explain the concept of TLM. The specific contact resistivity $\rho_c = \left.\frac{\partial V}{\partial J}\right|_{V=0}$ ($[\rho_c] = \Omega\,\mathrm{cm}^2$) is a better way to describe the contact resistance as it is independent of the contact area and therefore convenient for comparing contacts of different sizes [241]. Finally, with R_c and $A_{c,eff}$, ρ_c can be calculated. In first approximation, this results in [241]:

$$\rho_c = R_c \cdot A_{c,eff} \approx R_c \cdot 2\pi r L_T.$$

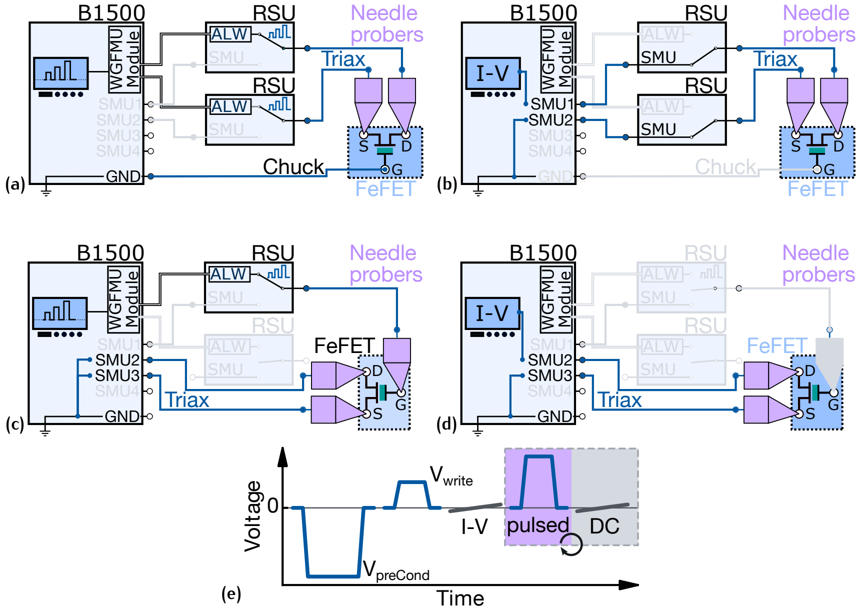


Figure 3.8: Setup for pulsed measurements: Configuration of the B1500A Semiconductor Analyser connection to the FeFET device for: **(a)** Write and **(b)** read operations for the first FeFET generation. **(c)** Write and **(d)** read operations for the second FeFET generation. **(e)**Typical measurement sequence for a R_{DS}-V_{write} plot: First the device is fully polarised in one direction (V_{preCond}), then a first state is written ((a) or (c)), and finally R_{DS} is measured ((b) or(d)).

3.3.2 Pulsed Potentiation and Depression

Here we briefly describe the measurement setup that we utilised for pulsed measurements. An Agilent B1500A Semiconductor Analyser equipped with 4 High Resolution Source Measurement Units (HRSMUs) and with a B1530A Waveform Generator/Fast Measurement Unit (WGFMU) module was used. The WGFMU module combines pulsing and Arbitrary Linear Waveform (ALW) generation capabilities with current and voltage measurement functions in a single unit. Two Remote-sense and Switch Units (RSUs) are connected by special composite cables to the WGFMU card. The waveforms created using the WGFMU's ALW voltage generation capability output through the RSU. The RSU is also the location where the actual current or voltage measurement is performed. In principle the RSU would be perfect for pulsed potentiation and depression as it can apply an ALW and at the same time sample the current. Unfortunately, the RSUs current measurement ranges from 1 µA to 10 mA. When measuring the resistance of our FeFETs ($R_{SD} \approx 100\,\mathrm{k\Omega} - 10\,\mathrm{G\Omega}$) at a small read voltage $V_{read} = 200\,\mathrm{mV}$, in order not to disturb the ferroelectric polarisation, currents in the range of 2 µA to 20 pA must be sensed. We can not increase the read voltage to augment the current because we would start to switch the polarisation (our gate is always grounded through the chuck). As an alternative HRSMUs are clearly required because they offer a measurement resolution of 1 fA. The HRSMU can be connected to the RSU through triax cables, and during operation we can switch between the ALW generation and the HRSMU, without the need for more than one needle prober per contact.

Figures 3.8a and 3.8b illustrate how a FeFET device from the first generation (Section 5.1) was contacted and how the connections were set. A state is written by applying a pulse to source (S) and drain (D) with synchronised RSUs, while the gate (G) is grounded (Figure 3.8a). The gate of the FeFET is shared among all devices on the chip and accessed through the conductive substrate by connecting the ground of the B1500 to the chuck. For the read operation (Figure 3.8b), the RSU units switch to SMU mode, where they simply pass through the HRSMU signal. The read operation is thus done with the HRSMU units by applying a $V_{read} = 200\,\mathrm{mV}$ between S and D with a dynamic integration time as required. A typical measurement sequence for a R_{DS}-V_{write} plot is performed as follows (Figure 3.8e): First the device is fully polarised in one direction (V_{preCond}), then a first state is written (Figure 3.8a), and finally R_{DS} is measured (Figure 3.8b). The writing and reading are then repeated with an increased or decreased V_{write}

until a full loop is completed. The B1500A provides its own programming environment where such sequences can be automated.

The second generation of FeFETs was connected differently. The devices deffer from the first generation by having a seperate G contact for each device which is accessible from the top. Thus, three needle probers and no chuck contact are required. Here, the write pulse is applied to the G, while S and D are grounded through a HRSMU unit. Hence, only one of the RSU units is required (Figure 3.8c). The read operation is therefore done with the HRSMU units by applying a $V_{read} = 200\,\mathrm{mV}$ between S and D with a dynamic integration time as required (Figure 3.8d). Potentiation and depression cycles are performed by switching between the write and read operations in the same way as for the first FeFET generation. When automating these write/read sequences with the B1500A software, the measurement takes quite long (up to 15 s per measurement point). Hence, we connected a computer to the B1500A by General Purpose Interface Bus (GPIB), from which we can send measurements commands from a C++ environment without using the build-in software on the B1500A. This allowed us to speed up the time required for one write/read operation to under 3 s. More details about the automated setup can be found in Appendix A.1

Working in a C++ environment on a remote computer brings another advantage for our setup. Our chuck is mounted on electrically movable x,y,z axes (Owis PS35) that can be controlled by a C++ library. This allowed us to create a single application where we can define a sequence of measurements and then perform them automatically on many devices on a chip by following a device-map containing the coordinates of all devices. More details about the automation from design to measurement including the application can also be found in Appendix A.1.

4

MATERIALS

If we amplify everything, we hear nothing.
— Jon Steward

Prior to the fabrication of ferroelectric memristors, we studied the characteristics of important material systems such as ferroelectric $Hf_xZr_{1-x}O_2$ (HZO) and semiconducting WO_x thin-films. This chapter first discusses the deposition and crystallisation of the ferroelectric HZO layer and its ferroelectric properties. In the second part, the advantages and disadvantages of different deposition techniques for the WO_x layer that is used as thin-film channel in ferroelectric field effect transistors are studied.

4.1 FERROELECTRIC HZO

In this subsection we will first look at the deposition and in a second part at the crystallisation of HZO. Especially, the latter is not trivial and multiple factors like temperature, electrodes, and layer thickness define the phase of the HZO (Section 2.3.1). Some of the results shown in this section were published in [31].

4.1.1 Deposition

One of the main advantages of fluorite ferroelectrics like HZO is their CMOS compatibility. In fact, HfO_2 and ZrO_2 are probably the two most frequently studied fluorite-structure materials for logic Field-Effect Transistors (FETs) or Dynamic Random Access Memory (DRAM) capacitors. Since the adoption of HfO_2 by Intel in 2007 [29], mature Atomic Layer Deposition (ALD) techniques have been available that provide great control over the layer thickness, composition, and 3D uniform coverage. Because of these advantages we chose ALD over other deposition methods like Physical Vapour Deposition (PVD) [242], Pulsed Laser Deposition (PLD) [191, 192, 243], or Molecular Beam Epitaxy (MBE) [244, 245]. PVD could be interesting if fast deposition rates are required, while PLD and MBE allow to grow well-oriented and epitaxial layers on specialised single crystal substrates. In our opinion, the ease of integration and composition control of ALD overweights all advantages of the alternative deposition methods when considering device fabrication and integration. A general introduction to the ALD deposition technique can be found in Section 3.2.1.

To study our HZO, we deposited TiN/HZO/TiN layer stacks on a highly conductive $n^+ Si$ substrate. We used an Oxford Instruments Plasma Enhanced Atomic Layer Deposition (PEALD) system to deposit all three layers at 300 °C without breaking the vacuum. First the $n^+ Si$ substrates were dipped in buffered hydrofluoric acid (BHF, 7:1) for ∼20 s to remove the native oxide. Then, ∼10 nm TiN was deposited using a TDMAT[1] precursor and N_2/H_2 plasma. Approximately 10 nm thick layers of HZO were grown in a process using alternating cycles of two TEMAH[2] and one ZrCMMM[3] precursor. Rutherford Back Scattering (RBS) analysis of the films indicated an actual film composition of $Hf_{0.57}Zr_{0.43}O_2$. A further 10 nm of TiN was immediately deposited in-situ in the PEALD. The layer thicknesses were later confirmed by X-Ray Reflectivity (XRR) measurements in a Bruker D8 Discover diffractometer equipped with a rotating anode generator and by cross sectional Bright Field Scanning Transmission Electron Microscopy (BF-STEM) measurements using a double spherical aberration-corrected JEOL JEM-ARM-200F microscope operated at 200 kV. The deposition rate of the TiN was found to be 0.8 Å/cycle. For two HfO_2 cycles and one ZrO_2 cycle, the smallest repeating unit of the HZO deposition, the deposition rate was 1.8 Å/cycle. After the crystallisation of the HZO, we deposited 100 nm W by PVD to form top contacts for the metal-insulator-metal (MIM) structures. Metal gate patterning and etch was performed using standard UV lithography and a fluorine-based reactive ion etch process.

4.1.2 Crystallisation

In this section we will study the crystallisation of HZO and have a closer look at important factors that influence the crystallisation, namely temperature, electrodes, and layer thickness. Because we are interested in integrating ferroelectric HZO in the Back-End-Of-Line (BEOL), our goal is to perform the crystallisation at temperatures below 400 °C, in order not to

---

[1] TDMAT: Tetrakis-(dimethylamino)titanium $[(CH_3)_2N]_4Ti$

[2] TEMAH: tetrakis-(ethylmethylamino)-hafnium $[(CH_3)(C_2H_5)N]_4Hf$

[3] ZrCMMM: bis(methyl-η5-cyclopentadienyl)methoxymethylzirconium $(CH_3C_5H_4)2Zr(OCH_3)CH_3$

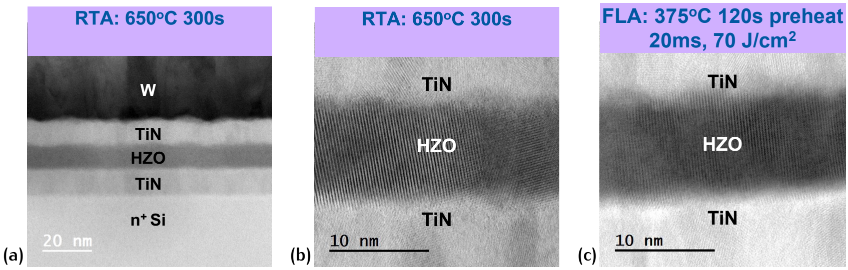


Figure 4.1: **(a)** TEM cross section of the W/TiN/HZO/TiN/n+Si MIM structure. The TEM micrographs in **(b)** and **(c)** focus on the TiN/HZO/TiN layers for the 300 s 650 °C RTA sample and the 70 J/cm² ms-FLA sample, respectively.

damage the underlying CMOS structures. For the crystallisation we used a Flash Lamp Annealer (FLA). A detailed description of the tools can be found in Section 3.2.2. Xenon flash lamps are used to release a flash pulse with a duration from 0.3 to 20 ms and energy densities of up to 110 J/cm² onto the sample surface. The resulting temperature of the sample during a millisecond Flash Lamp Annealing (ms-FLA) depends on the pre-heat temperature, the flash energy, and flash duration. To ensure that the samples are thermalised we keep them at the pre-heat temperature for 120 s before applying the flash pulse. The pre-heat temperature is monitored by a thermocouple. Using spikes is expected to benefit two fold: First, the spikes are very short and thus the temperature above 400 °C is not maintained for long, minimizing the effect on the CMOS structures. Second, we are trying to stabilise the meta-stable f-phase and hence, quenching should facilitate the stabilisation of such higher temperature phases.

4.1.2.1 *Temperature*

In a first step, we looked at the crystallisation temperature. For that we used the sample stack (TiN/HZO (10 nm)/TiN) described in Section 4.1.1. Previous studies [201, 202, 246] have shown that a film thickness around 10 nm results in grain sizes around 10 nm, which were postulated by calculations [175] and shown experimentally to maximise the fraction of the ferroelectric phase ($Pca2_1$, f-phase). Thicker films tend to crystallise in the monoclinic phase ($P2_1/c$, m-phase), while thinner films have shown an increase of the tetragonal phase ($P4_2/nmc$, t-phase). Hence, a film thickness of 10 nm is optimal to study the crystallisation temperature. We chose TiN

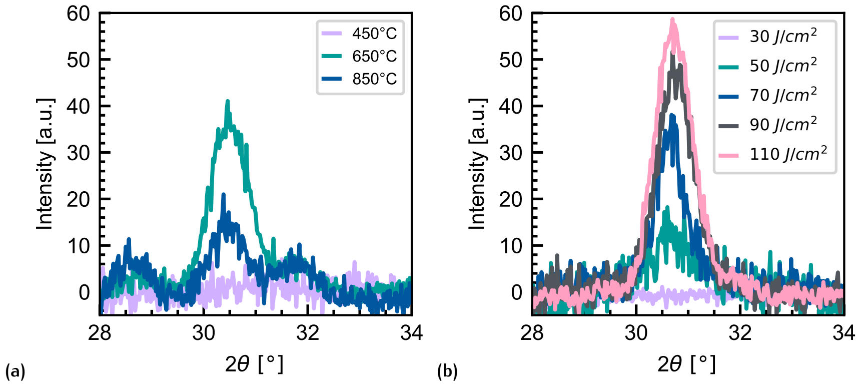


Figure 4.2: (a) GIXRD for a diffraction angle (2θ) from 28° to 34° for samples which received a regular 300 s 450 °C, 650 °C and 850 °C RTA. **(b)** GIXRD profiles for the millisecond flash lamp annealed samples with varying flash energy of 20 ms duration. In each case there was a 120 spreheat step at 375 °C.

as electrodes because they have been proved to facilitate the crystallisation of the f-phase [27]. Without a top capping electrode, the HZO films have an increased m-phase fraction. A more detailed analysis of the impact of different electrodes on the crystallisation of HZO follows in Section 4.1.2.2.

The conditions for this experiment were as follows: the pre-heat temperature was set to 375 °C, the flash pulse duration was kept constant at 20 ms, while the flash pulse energy was varied from 30 to 110 J/cm². For comparison, additional samples were annealed in the same FLA chamber for 300 s at 450 °C, 650 °C and 850 °C without a flash pulse. Figure 4.1 shows transmission electron microscopy (TEM) cross sections of the MIM devices. Figure 4.1a reports a micrograph of the full W/TiN/HZO/TiN/n+Si MIM structure. The TEM analysis confirms the thicknesses of the TiN layers (~10 nm) and of the HZO layer (~10 nm). The various layers are clearly uniform with low roughness at the interfaces. Figures 4.1b and 4.1c present more scaled micrographs focusing on the TiN/HZO/TiN stack, for the sample which received a regular 300 s/650 °C RTA with no flash anneal step, and the 70 J/cm² ms-FLA sample, respectively. The TEM images show a similar microstructure and confirm the polycrystalline nature of the TiN electrodes and the HZO layer in both cases. Therefore, it is clear that the 120 s pre-heat at 375 °C combined with 70 J/cm² ms-FLA is sufficient to fully crystallise the 10 nm HZO layer.

GIXRD results are depicted in Figure 4.2 where the range of the diffraction angle (2θ) is limited between 28° and 33°. In this region, one can easily distinguish between the non-ferroelectric monoclinic phase and the cubic, tetragonal or orthorhombic phase, only the latter one being ferroelectric. Figure 4.2a shows the reference samples that received a regular RTA treatment for 300 s. Clearly, 450 °C was not sufficient to crystallise the HZO. The sample with a regular 300 s/650 °C RTA step has a peak centered around 30.5°, which can be assigned to t- and/or f-phases in HZO. It should be noted that distinguishing them from each other is difficult using XRD owing to their similar structure. More details on the structures can be found in Section 2.3.1.1. However, this sample provides promising structural evidence for the possible existence of the orthorhombic phase responsible for ferroelectric behaviour in HZO films.

When increasing the temperature even further, the peak around 30.5° becomes smaller and the two m-phase peaks (28.5°, 31.8°) that already appeared for the 650 °C sample, grow larger. The XRD profiles for the millisecond flash lamp annealed samples with varying flash energy of 20 ms duration are shown in Figure 4.2b. In each case there was a 120 s pre-heat step at 375 °C. For the 110 J/cm², 90 J/cm² and 70 J/cm² ms-FLA samples a peak is observed in the XRD around 30.5°, with the magnitude reducing slightly for the 70 J/cm² sample. When the flash energy is reduced to 50 J/cm² this peak magnitude significantly diminishes while at 30 J/cm² it is not evident. Clearly, the XRD peaks for the ms-FLA samples are similar to that recorded in Figure 4.2a for the 300 s/650 °C RTA sample, indicating the possible presence of the f-phase.

"Polarisation vs. Electric Field" (PE) characterisation was performed on the MIM structures to investigate hysteretic behaviour of the HZO films. Figure 4.3 reports PE hysteresis loops measured at 1 kHz and at room temperature for all samples. For each sample the PE loop on a pristine device is plotted along with the PE loop obtained after electric field cycling of the same devices. As can be seen in Figure 4.3a the sample which received only 30 J/cm² and which exhibited no orthorhombic peak in the XRD analysis, displays a negligible hysteresis and linear PE characteristics, as would be expected for non-ferroelectric materials [162]. All other samples measured in a pristine state show pinched PE hysteresis loops. This is significantly improved post cycling in all cases, with the stressing enhancing both positive and negative remnant polarisation. These results are in agreement with the so-called “wake-up” effect observed previously for various ferroelectric films [143, 215, 247, 248]. In all ferroelectric samples an imprint

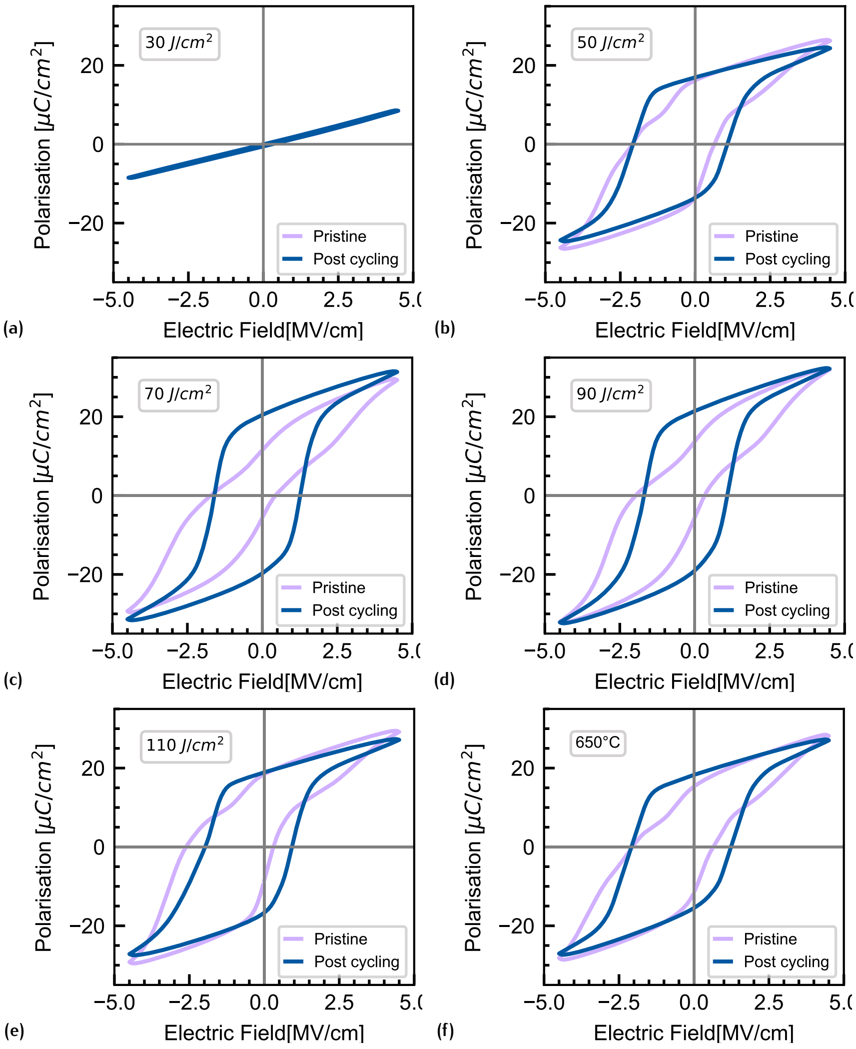


Figure 4.3: "Polarisation vs. Electric Field" (PE) characteristics measured on W/TiN/HZO/TiN/n^+Si MIM structures at room temperature and at 1 kHz. Hysteresis loops are shown for each sample, both in pristine and post electric field cycling (∼1500 cycles) conditions. **(a-e)** Millisecond flash annealed samples with varying flash pulse energies. **(f)** 300 s/650 °C RTA sample with no flash anneal.

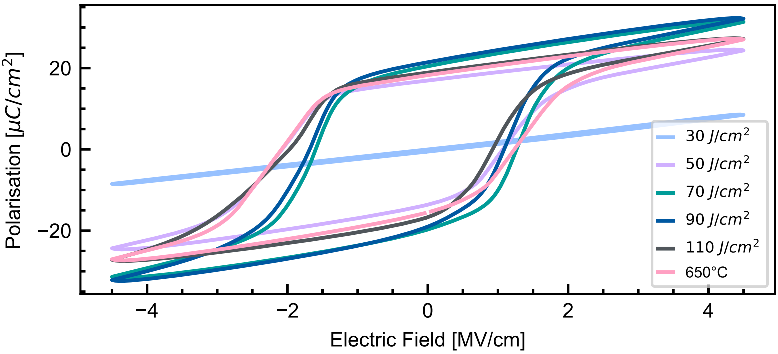


Figure 4.4: "Polarisation vs. Electric Field" (PE) characteristics measured on W/TiN/HZO/TiN/n+Si MIM structures at room temperature and at 1 kHz. Hysteresis loops (post cycling) are shown for the 300 s/650 °C RTA sample with no flash anneal and for the millisecond flash annealed samples with varying flash pulse energies. In regard to the latter there was a 120 s pre-heat step at 375 °C and the flash pulse duration was 20 ms.

can be observed. The interface screening model [249] assumes a thin layer at one interface where the spontaneous polarisation is absent and causes a charge separation between the screening and polarisation charges to be responsible for the imprint. This "passive layer" behaves as a space charge layer and hence acts as a capacitor in series. The large electric field in the passive layer induces charge transport across it, resulting in charge separation and hence a built-in potential [250, 251]. During the ALD deposition of the stack, the bottom TiN electrode gets oxidised by the first cycles of the HZO deposition. A thin $TiNO_x$ layer is formed at that interface that could lead to the above described charge separation.

For ease of comparison Figure 4.4 plots the PE hysteresis loops measured post-cycling for the various ms-FLA samples and the 300 s/650 °C RTA sample. Two of the critical parameters in evaluating the quality of a ferroelectric layer are the coercive field strength (E_c) and the remanent polarisation (P_r). In this regard the 90 J/cm² and 70 J/cm² ms-FLA samples exhibit the highest P_r values (~21 µC/cm²). E_c is ~1.1 MV/cm and the 70 J/cm² sample displays slightly lower imprint. The P_r and E_c values for the optimised ms-FLA samples in this study are comparable to those reported previously for HZO films of similar thickness [162, 166, 173, 174]. While E_c is similar for the 300 s/650 °C RTA sample the P_r values

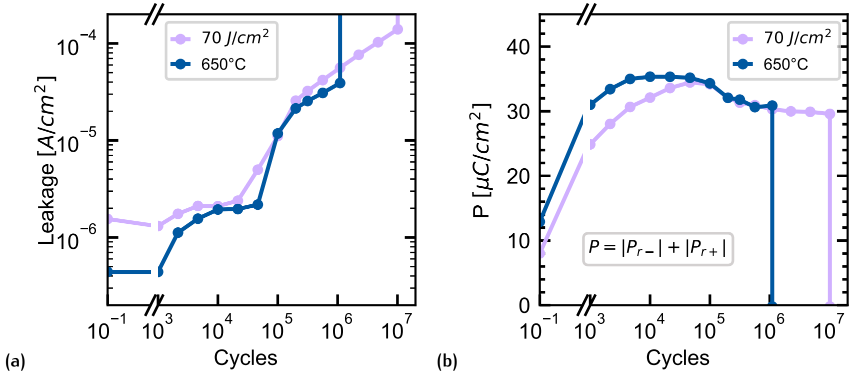


Figure 4.5: Endurance characteristics measured on the 70 J/cm² ms-FLA and 300 s/650 °C RTA sample at room temperature. Bipolar cycling stress with an AC amplitude of 3.5 V and frequency of 1 kHz up to 10^5 cycles and 100 kHz up to 10^7 cycles was applied. **(a)** DC leakage measured at ~1 MV/cm. **(b)** Polarisation extracted from PUND measurements with an amplitude of 4 V and pulse width of 1 ms.

(~18 μC/cm²) are clearly degraded as compared to the optimised ms-FLA samples.

Another clear difference between the RTA and ms-FLA annealing becomes visible in endurance measurements (Figure 4.5). A bipolar cycling with an amplitude of 3.5 V and a frequency of 1 kHz was applied to both a 70 J/cm² ms-FLA and a 300 s/650 °C sample up to 10^5 cycles. Afterwards, a frequency of 100 kHz was applied up to 10^7 cycles. Figure 4.5a shows the change in DC leakage while Figure 4.5b reports the evolution of the polarisation $P = |P_{r-}| + |P_{r+}|$ with cycling. Due to leakage currents, P was determined by Positive Up Negative Down (PUND) measurements [252]. Three distinct regimes can be identified in both samples. First, P increases due to the wake-up effect until it reaches its maximum value of 35 μC/cm² at ~46 000 cycles. Applying more cycles results in a decrease of P accompanied by an increase of the leakage current. This regime is a characteristic of the fatigue of the ferroelectric [217]. Finally, dielectric breakdown is reached at a number of cycles which determine the 'endurance' of the material. The endurance of the 70 J/cm² ms-FLA sample is one order of magnitude higher than that of the 300 s/650 °C RTA sample (Figure 4.5b). At the same time, the wake-up of the 70 J/cm² ms-FLA sample is slower, but it reaches a comparable maximum P, whereas the fatigue process appears similar in both samples. The wake-up process can be attributed to the redistribution of existing defects such as oxygen vacancies [217, 220] or phase transitions [217, 253, 254] in the device. The fatigue process on the other hand is a result of dielectric degradation due to an increasing trap density and domain pinning [217, 255]. It is explained in more details in Section 2.3.1.2. The slower wake-up phase is consistent with a delayed breakdown in ms-FLA films and can be attributed to a different microstructure obtained using the different anneal processes. The faster wake-up of the RTA sample indicates that there are less defects at the interfaces to redistribute. The relatively slow annealing compared to the ms-FLA might have allowed the diffusion of the defects already during the annealing. The improved endurance illustrates the potential of using fast annealing to stabilise the f-phase of HZO in non-volatile memory elements.

Multiple differences between the RTA (650 °C) and the 70 J/cm² ms-FLA sample have been observed: the RTA sample has a small portion of m-phase, while the ms-FLA sample does not. Due to the smaller dielectric constant, m-phase grains can form a slightly more conductive path within the f-phase. Furthermore, the negative E_c of the RTA sample is larger and the polarisation hysteresis is slightly tilted. A tilted hysteresis loop is an indication of a dielectric Interfacial Layer (IL) between the ferroelectric layer and the electrode, resulting in a depolarisation field [256]. These low quality, non-switching regions are characterised by more defects and a lower permittivity, resulting in a higher factor of degradation [217]. Only the negative E_c is substantially increased, hinting to a IL only on one side. The 50 J/cm² and 110 J/cm² ms-FLA samples show a similar polarisation hysteresis. From the GIXRD scans we know that the 50 J/cm² sample is not fully crystallised, possibly having an amorphous, non-ferroelectric layer at one electrode interface. The slight polarisation degradation of the 110 J/cm² sample compared to the 90 J/cm² one, together with a minimal positive shift of the GIXRD peak with increasing flash energy could indicate an increase of the t-phase often seen at interfaces [217, 218]. Therefore, we believe that the RTA anneal leads to a worse interface quality, which reduces the performance.

In summary ferroelectric behaviour was demonstrated for ~10 nm $Hf_{0.57}Zr_{0.43}O_2$ films using a 120 s/375 °C pre-heat combined with a 20 ms flash lamp annealing pulse. The ferroelectric characteristics achieved using the ms-FLA were comparable to that obtained using a much higher thermal budget of 300 s RTA at 650 °C, as confirmed using XRD, PE, and endurance

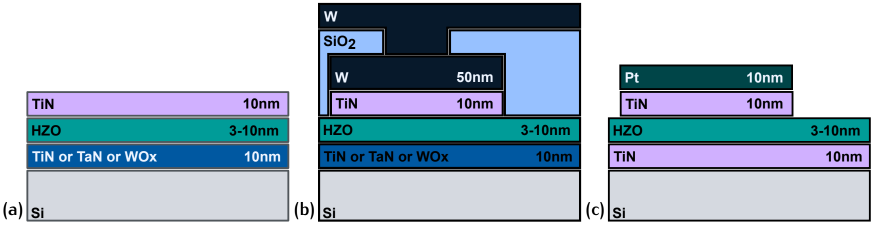


Figure 4.6: Schematic representation of structures used to analyse different bottom electrodes: **(a)** For GIXRD, **(b)** for PUND, and **(c)** for PFM.

analysis. While a similar coercive field (~1.1 MV/cm) is achieved for both the ms-FLA and RTA, superior remanent polarisation values are obtained for the optimised ms-FLA samples (~21 μC/cm²) compared to the RTA sample (18 μC/cm²) in this study. The increased endurance of the ms-FLA sample further emphasises the advantage of the millisecond flash lamp annealing technique. Given that the annealing temperature profile is one of the most critical factors for formation and stabilisation of the ferroelectric orthorhombic phase in HZO, this millisecond FLA technique offers a promising low thermal budget alternative for the crystallisation of ferroelectric HZO films.

4.1.2.2 *Electrodes*

In this section we will have a closer look at different electrodes. Confinement by the top electrode [196, 204] is a key enabler for the ferroelectric phase (f-phase) in HZO. The mechanical stress of the electrode helps to stabilise the f-phase over the monoclinic one (m-phase). For TiN a tensile stress is created that can be divided into two components, the intrinsic stress due to the film growth and the thermal stress originating from the different thermal expansion coefficients of TiN and HZO. Shiraishi *et al.* [203] have shown that larger in-plane tensile stress leads to higher remanent polarisation (P_r), while compressive stress resulted in the opposite. Others have shown that it is also possible to obtain the f-phase without the confining top electrode, but nevertheless, the resulting P_r was always smaller than the control sample with an electrode [257, 258]. Therefore, choosing the right electrode can potentially increase the polarisation due to a higher fraction of f-phase. Until now, only few studies have demonstrated slightly improved P_r with other electrodes than TiN: Mo [259], W [260, 261],

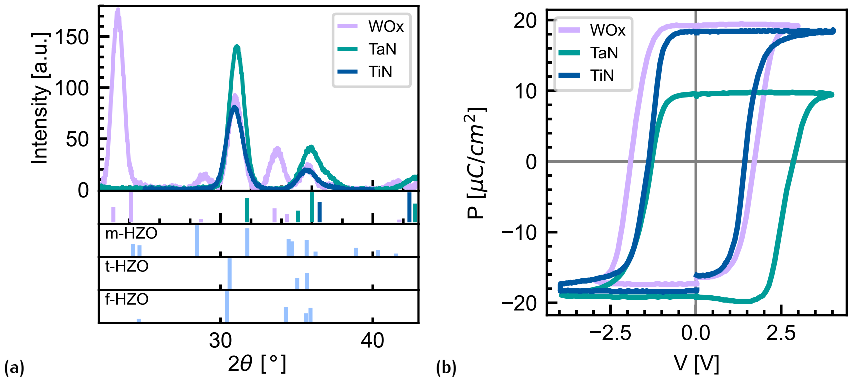


Figure 4.7: Comparison between different bottom electrodes: **(a)** GIXRD for a diffraction angle (2θ) from 22° to 43° for samples which received a ms-FLA with 70 J/cm². The lines below the measurement indicate the reported peak position for the different electrodes. **(b)** PUND measurements on capacitors with an area of 80 µm × 80 µm that were processed on the same samples. Bipolar cycling stress with an AC amplitude of 3.5 V and frequency of 100 kHz for 10^6 cycles was applied to wake-up the HZO before measuring PUND.

Ge [262], Ta [261], while most are performing less well: Ir [263], Si [262], Au [261], Pt [257, 261], and RuO_2 [264].

In the context of ferroelectric memristor fabrication, the electrodes play a crucial role, especially for the Ferroelectric Tunneling Junction (FTJ). The working principle of FTJs (Section 2.2) relies on asymmetric electrodes, in particular on electrodes (or interfaces) with different screening lengths. Thus, investigating different electrodes is interesting for FTJs and generally for the optimisation of the crystallisation in the f-phase. In addition to the standard TiN electrode that we used, we fabricated devices with TaN and WO_x bottom electrodes. The fabrication of the devices with TaN is almost identical to the TiN (Section 4.1.1), except that the bottom electrode was deposited using a Ta precursor. The WO_x sample was fabricated by first depositing 4 nm of W on the n^+ Si substrate by PVD. This thin W layer was then oxidised to WO_x in a RTA at 350 °C with 50 sccm O_2 for 6 min. Afterwards the sample was transferred to the ALD where the HZO and TiN materials were deposited. The resulting layer stack is depicted in Figure 4.6a.

Before proceeding, we characterised the root-mean-square roughness of the three electrodes by Atomic Force Microscopy (AFM): 291 pm for TiN, 371 pm for TaN, and 1250 pm for WO_x. While TiN and TaN show similar smooth values, WO_x has a four times larger roughness. The oxidation of W to WO_x is accompanied by a volume increase of about 3.4 (Section 4.2.1.1) that leads to this increased roughness.

All the samples were then crystallised in the FLA with the optimal conditions that we found in the previous Section 4.1.2.1: After pre-heating the sample to 375 °C for 120 s we applied a energy pulse of 70 J/cm². GIXRD scans were then taken and compared in Figure 4.7a. Below the GIXRD scan data, the locations of the Bragg reflections for TiN (ICSD658338 [265]), TaN (ICSD644728 [266]), WO_x (ICSD86144 [267]), m-HZO [175], t-HZO [175], and f-HZO [173] are shown. All peaks can be assigned. The absence of the m-phase for all electrodes again underlines the benefit of the ms-FLA. TaN has a peak at 31.7° that is relatively close to the f-phase of HZO and the resulting peak is a superposition of the two, leading to an increase in amplitude and a shift towards the TaN peak. Clearly, also WO_x crystallises by the ms-FLA.

Later, in Section 4.2.1.2 we show that the ms-FLA provides enough energy to reduce the WO_x to be quite conductive. Still, it is not clear if the peaks around 30.5° are mainly due to the f- or t-phase. Therefore, the samples were further processed to capacitors, as illustrated in Figure 4.6b. We then measured PUND measurements with a frequency of 5000 Hz on capacitors with an area of 80 µm × 80 µm. Prior to the PUND measurement, bipolar cycling stress with an AC amplitude of 3.5 V and frequency of 100 kHz for 10^6 cycles was applied to wake-up the HZO. All samples show a switchable polarisation (Figure 4.7b).

The TaN clearly possesses a different behaviour by having a much larger positive coercive voltage of $V_c+ = 2.43\,\mathrm{V}$ and a lower positive remanent polarisation $P_{\mathrm{r}+} = \sim 9.7\,\mu\mathrm{C/cm^2}$. The TiN and WO_x samples exhibit similar remanent polarisations with equal positive and negative values of about $|P_\mathrm{r}| = \sim 19\,\mu\mathrm{C/cm^2}$. The sample with the WO_x electrode displays slightly higher coercive fields than the TiN sample, possibly due to a partial field drop across the semiconducting WO_x. The smaller $\mathrm{P}_{\mathrm{r}+}$ as compared to $\mathrm{P}_{\mathrm{r}-}$ of TaN indicates that not all domains are stable when the polarisation points towards the TaN. Positively charged defects at the TaN interface could lead to local pinning of domains pointing away from the TaN electrode which can never contribute to $\mathrm{P}_{\mathrm{r}+}$ and would also explain the larger V_c+ [268].

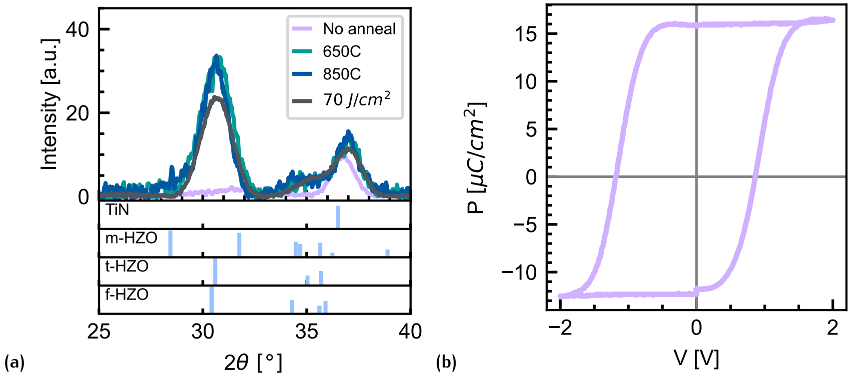


Figure 4.8: Ferroelectricity of 5 nm HZO: **(a)** GIXRD for a diffraction angle (2θ) from 25° to 40° for a sample with a 5 nm thick HZO layer which received different annealing treatments. **(b)** PUND measurements on capacitors with an area of 80 µm × 80 µm that were processed on the same samples. Bipolar cycling stress with an AC amplitude of 1.5 V and frequency of 100 kHz for 10^5 cycles was applied to wake-up the HZO before measuring PUND.

We can conclude, that in our case, TiN remains the best performing electrode, with WO_x being an alternative option with a slightly larger V_c. On the other hand, thinking of asymmetric electrodes for FTJs, WO_x is a very good candidate.

4.1.2.3 *Layer Thickness*

In this section we will investigate the influence of the HZO layer thickness on the crystallisation temperature, the resulting phase, and polarisation. According to Materlik *et al.* [175], the f-phase has the lowest free energy of all phases for grain sizes between 8 nm and 16 nm in HZO. For thin films, the grain size is comparable to the film thickness (more details can be found in Section 2.3.1.1). For smaller grain sizes, they predict the stabilisation of the t-phase over the m-and f-phase. In a first step we repeated the deposition of the TiN/HZO/TiN stack with a 5 nm thick HZO layer. We then looked at different crystallisation conditions and validated them by GIXRD, as shown in Figure 4.8a. The peak positions for TiN and the m-, t- and f-HfO_2 phase are shown below the GIXRD data.

First we applied the same optimum 70 J/cm² energy flash as for 10 nm HZO films and found a familiar peak at ~30.5°. The peak at ~36.7° comes

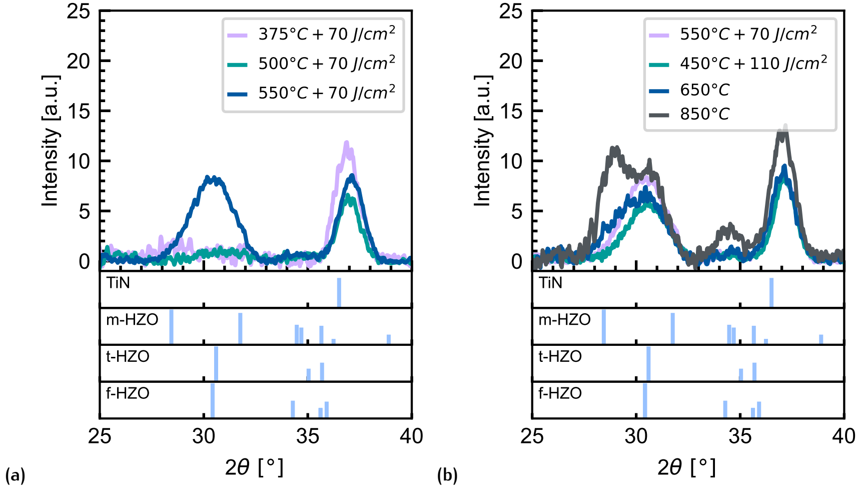


Figure 4.9: Ferroelectricity of 3 nm HZO: GIXRD for a diffraction angle (2θ) from 22° to 43° for a sample with a 3 nm thick HZO layer which received either **(a)** 70 J/cm^2 with different pre-heat temperatures or **(b)** only the pre-heat treatment for 300 s.

from the bottom TiN electrode. It is much more pronounced as compared to the 10 nm HZO sample because of the greater penetration of the X-ray beam into the bottom electrode and the weaker signal of the thinner HZO. A mathematical explanation for the intensity reduction with film thickness can be found in Section 3.1.1. The small shoulder at ∼35° comes from the (200) reflection of HZO.

For comparison, additional samples were annealed in the same FLA chamber for 300 s at 650 °C and 850 °C without a flash pulse. Again, the FLA sample shows a similar diffraction pattern as the sample that was annealed at 650 °C while having a lower thermal budged, still compatible with the BEOL. Increasing the temperature to 850 °C introduces a small amount of m-phase which reduces the polarisation. Figure 4.8b shows a PUND measurement at 100 Hz that was conducted on a 80 µm × 80 µm capacitor that was fabricated on the 70 J/cm^2 sample. Prior to the measurement, a wake-up of the HZO was achieved by bipolar cycling stress with an AC amplitude of 1.5 V and frequency of 100 kHz for 10^5 cycles. Even with a thin 5 nm HZO film, we were able to measure a positive (negative) remanent polarisation of $P_{r+} = \sim 15\,\mu\mathrm{C/cm^2}$ ($P_{r-} = \sim 12\,\mu\mathrm{C/cm^2}$).

So in a next step, we further reduced the HZO thickness to 3 nm. Again, starting with the same 70 J/cm² energy flash as for 10 nm HZO films, it became evident that such thin layers require a higher thermal budget to crystallise. From the GIXRD diffraction patterns in Figure 4.9a we can extract a required pre-heat temperature of 550 °C to crystallise the HZO while keeping the flash energy at 70 J/cm². To reduce the temperature (450 °C) we increased the flash energy to the maximum of 110 J/cm² and were still able to crystallise the HZO (Figure 4.9b), although with a slight reduction of the peak. This indicates that we are at the lower limit of the pre-heat temperature to still crystallise the HZO. Again, for comparison we repeated the crystallisation with a pre-heat temperature of 650 °C and 850 °C without the flash. In both cases, but especially at 850 °C, a considerable amount of the HZO crystallised in the m-phase.

Measuring the polarisation by PUND was not possible because the leakage current through such thin films is so large that the switching current could not be separated thereof. In principle, 3 nm is thin enough for tunnelling and the ferroelectricity could be confirmed by measuring a Tunnelling Electro-Resistance effect (TER), as expected in FTJs (Section 2.2.3.3). Therefore, we applied pulses with different amplitudes to the capacitors and measured the resistance afterwards. No TER effect was obtained. One explanation is that the peaks at ~30.5° come from the t-phase. An increase of the t-phase with decreasing film thickness was also predicted by Materlik *et al.* [175]. Symmetric electrodes is another possible reason for the absence of TER. However, the latter is almost impossible since one TiN electrode is always more oxidised than the other during growth due to the O_2 plasma from the HZO deposition. It could be that the HZO crystallised in the f-phase, but similarly to epitaxial thin-film perovskite ferroelectrics, the polarisation vanishes for films approaching a few nm thickness due to non-ferroelectric interfaces [269].

Piezoresponse Force Microscopy (PFM) is a method to measure the polarisation locally with a conducting tip in an AFM. An AC electric field is applied locally between the AFM tip and the substrate and the piezoelectric response (contraction/extension) of the ferroelectric material is measured. Both the amplitude and phase of the oscillating tip are monitored with a standard photo-diode detector and demodulated with a lock-in amplifier. The strength of the signal (amplitude) provides information on the materials piezoelectric tensor. The phase signal indicates the direction of the polarisation. To be able to sense the contraction/extension of the ferroelectric with the PFM tip through the electrodes, their thickness has to

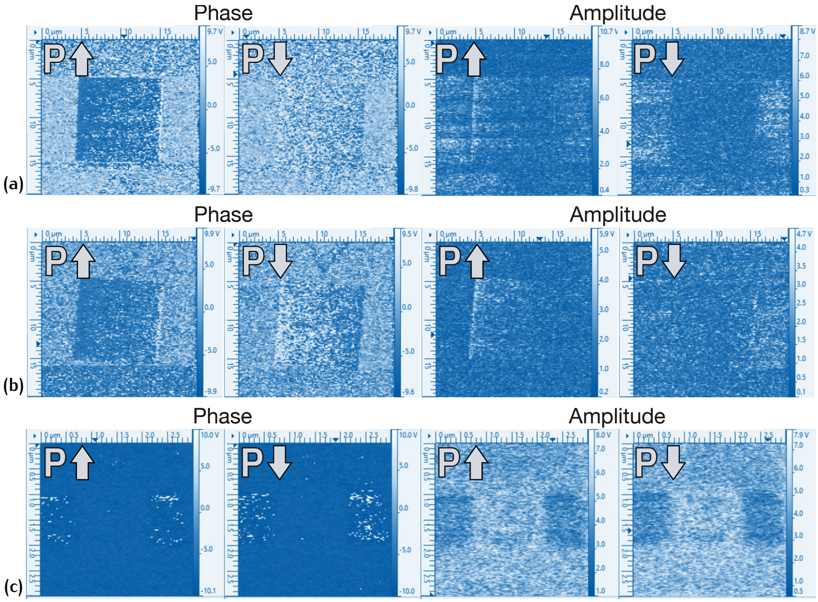


Figure 4.10: Piezoresponse Force Microscopy (PFM) on samples with HZO thicknesses of **(a)** 10 nm , **(b)** 5 nm , and **(c)** 3 nm. The PFM response was measured through a thin Pt/TiN electrode for the structure depicted in Figure 4.6c. The two rows of PFM scans on the left are showing the phase and the two on the right the amplitude for both polarisation directions. The scans show an area slightly larger than the capacitor.

be very thin. If $P_{r+} = P_{r-}$ then the measured amplitude of the contraction and extension should stay constant. The phase of the cantilevers response on the other hand will have a 180° shift depending on the polarisation direction. This is because domains with one direction respond with a contraction while domains with the other direction respond with an extension along the electric field direction that is applied by the tip.

We sent 3 samples with HZO thicknesses of 10 nm, 5 nm and 3 nm that were processed especially for this to our colleagues at CNRS Thales. Their structure is depicted in Figure 4.6c. Figure 4.10 shows the PFM scans for the various HZO thicknesses. Our polycrystalline films have grains with polarisation directions showing in all directions. The tip will only pick up the projection of the contraction/extraction along the out-of-plane direction. Hence, polycrystalline films are more difficult to measure as the signals are usually quite small and noisy. In addition the domain size ($\sim$10 nm) is smaller than the tip radius (nominally $\sim$20 nm).

While scanning, the topography of the sample is measured at the same time as their contraction/extension and needs to be decoupled, which further introduces noise. The experiment was conducted as follows: First the capacitors were woken-up by an AC signal applied through the cantilever with an amplitude of 3.5 V, 3 V, and 2 V for the 10 nm, 5 nm, and 3 nm samples, respectively. Then, the ferroelectricity of the capacitor was poled in one direction by applying a field large enough to switch all possible domains. Finally, an area slightly larger than the capacitor was scanned with a smaller AC field that does not switch the domains, but is large enough to get a measurable response. The results for the 10 nm- and 5 nm-thick HZO films are quite similar and show a constant amplitude for both polarisation directions, except some small regions at the edge of the contacts. The phase clearly displays a contrast between the two polarisation directions, agreeing with the PUND findings shown above. For the 3 nm HZO film, no difference between the polarisation directions can be observed, underlining the absence of ferroelectric switching. Therefore, we conclude that our 3 nm thick film has no switchable polarisation and that the X-ray diffraction peak at $\sim$30.5° cannot be clearly assigned to the f- or t-phase.

After having looked at the influence of the temperature, electrodes, and film thickness on the ferroelectric properties of our HZO, it is evident that 10 nm HZO films have the largest polarisation ($\sim$19 $\mu\mathrm{C\,cm^{-2}}$) and can be crystallised in the f-phase with a low thermal budged of 375 °C plus a 70 $\mathrm{J/cm^2}$ energy flash in the FLA. The absence of ferroelectricity in the 3 nm HZO layers make the realisation of a FTJ difficult. In FeFETs on the other hand, the ferroelectric gate dielectric thickness can be adapted to the optimum HZO thickness for ferroelectricity.

4.2 SEMICONDUCTING WO_x

This part is about the deposition, crystallisation, and reduction of WO_x films. It is important to understand the properties of WO_x since it is the thin film channel in our FeFET devices. WO_x is a transition metal-oxide and has the ability to change its conductivity by modifying its oxidation state. WO_x has been investigated in the context of sensing [270] and electrochromic applications [271, 272] as well as conductive metal oxide [267]. Lately, it has also gained attention for memristive applications [273], mainly as valence change memory (VCM) [274–277], electrochemical metallisation (ECM) [278, 279], or oxygen diffusion devices [280–284]. We will look at two deposition methods and how to control its conductivity for the purpose of a semiconducting thin-film channel in a FeFET.

4.2.1 WOx by Physical Vapour Deposition plus Thermal Oxidation

4.2.1.1 Deposition

The initial deposition method that we used to produce WO_x films is by depositing a thin W layer by Physical Vapour Deposition (PVD) with a consecutive Rapid Thermal Oxidation (RTO). For the deposition of the W we used a Von Ardenne CS320S Clustersystem sputter tool with a 200 mm W target that is located 84 mm above the substrate. By setting the DC plasma power to 120 W and regulating the pressure to 5.9 mbar and without heating the sample we obtained a deposition rate of 0.22 nm/s. These films were then transferred to a Rapid Thermal Annealer (RTA) where they were heated to 350 °C, while applying 50 sccm O_2 to oxidise the thin W layer to WO_x (RTO). The same temperature was sufficient to simultaneously crystallise the WO_x. The oxidation/crystallisation time depends on the thickness of the W layer. To make sure that all W atoms are oxidised this time should be calculated generously. In particular, in a device concept such as the FeFET, a remaining thin film of conducting W would undermine the field effect in the channel.

For the FeFET devices we are interested in very thin films that are in the order of the screening length of the polarisation charges from the HZO.

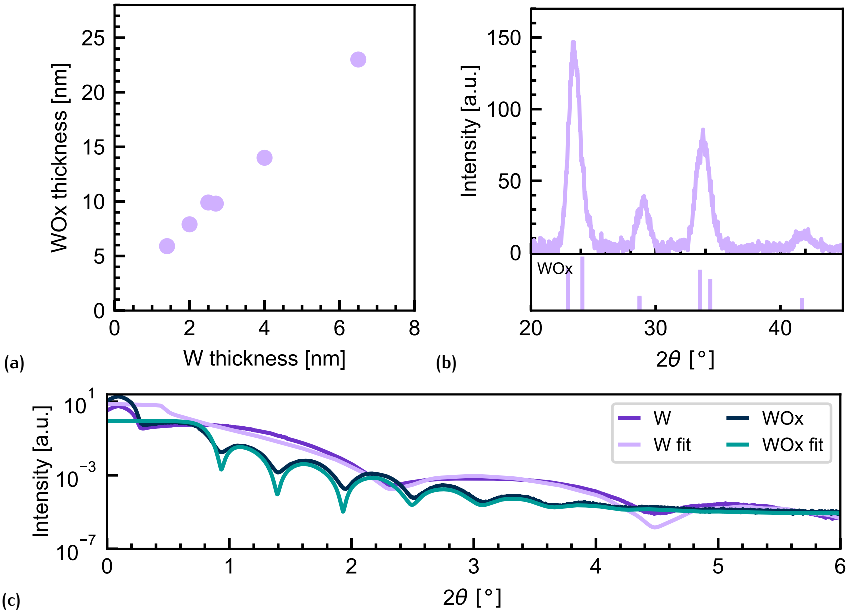


Figure 4.11: Rapid thermal oxidation of W to WO$_x$: **(a)** Relation between the thickness of the original W layer and the resulting WO$_x$ thickness. It is quite linear and translates to roughly a factor of 3.4. **(b)** GIXRD scan of a 23 nm-thick WO$_x$ layer with the corresponding peak positions for the tetragonal phase (P-42_1m) [267]. **(c)** XRR scans before and after the oxidation to determine the film thicknesses.

Oxidising W to WO_x is accompanied by a large volumetric change and hence, the W seed layer must be even thinner. Figure 4.11a shows the relation between the W and the resulting WO_x layer thickness after oxidation. The thickness of the layers were determined by fitting X-Ray Reflectivity (XRR) measurements, as can be seen in Figure 4.11c for the case of 4 nm W and 14.5 nm WO_x, resulting in a extracted thickness of 4.017 nm and 14.76 nm, respectively. To determine the phase of the WO_x layers, GIXRD scans were performed and the observed peaks (Figure 4.11b, 23 nm WO_x) were found to be in good agreement with the tetragonal phase ($P\text{-}42_1m$, ICSD86144 [267]). The relation between the original W and the resulting WO_x film thickness is fairly linear with a slope of ~3.4. For a WO_x channel thickness of less than 10 nm we thus require a W seed-layer of less than 3 nm, which by PVD would probably lead to non-continuous films. Our approach for such thin layers by PVD was to first deposit 4 nm of W, which should be thick enough to form a continuous film, and then we thinned the W layer to the desired thickness by Ar sputtering.

4.2.1.2 Reduction

After the W oxidation by RTO we obtain stoichiometric WO_3, which is electrically insulating. WO_3 has a high mobility of oxygen [267] and thus can be easily reduced ($WO_{x<3}$) by annealing in a reducing environment such as Na [267], H_2 [285], or CO [286]. When in contact with non-stoichiometric oxides, WO_x can be reduced by simply annealing in vacuum, as we will see later (Section 4.2.2.2). For the reduction experiment we prepared the following layer stack: Si/SiO_2 (20 nm)/W (2 nm to 2.5 nm). First the 2 nm- and 2.5 nm-thick W layers were oxidised for 6 min to 6 nm- and 10 nm-thick WO_x layers. After the oxidation, we performed different reduction treatments: 30 min H_2 anneal at 150 °C + 30 min vacuum anneal at 350 °C (reduction 1), 30 min vacuum anneal at 350 °C (reduction 2), or no reduction. H_2 is a good reducing gas, but might lead to hydrogen trapped in the layers, which could lead to uncontrolled reductions during further processing. After the reduction, W-structures designed for circular transmission line measurement (CTLM) were deposited on top of the WO_x by lift-off. With these structures (Figure 4.12c) we were able to measure the resistivity (ρ) of the WO_x layer. We further processed the samples to simple FeFET devices and measured again the resistivity after the processing (Figure 4.12b). The device processing included: First, the deposition of 10 nm HZO and 10 nm TiN by ALD, followed by 100 nm of W by PVD. Next, the HZO was crystallised in the FLA at 375 °C + 70 J/cm². The gate was

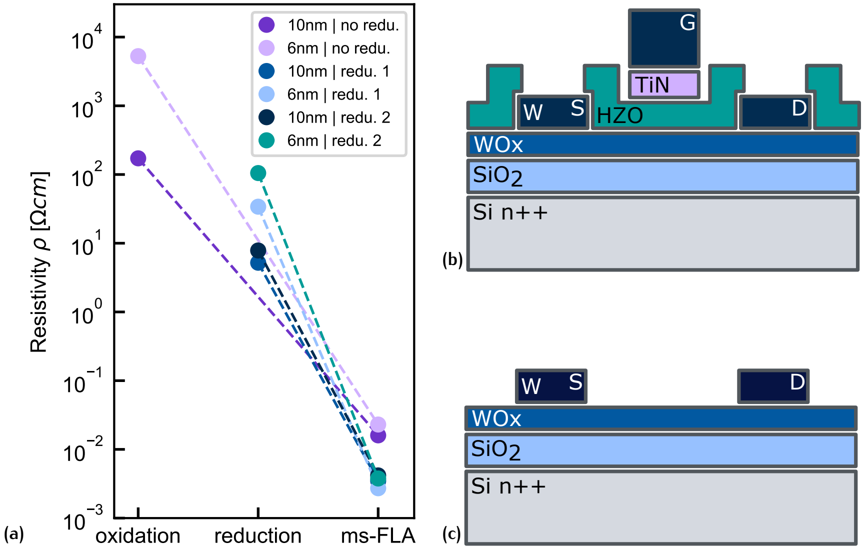


Figure 4.12: Reduction of WO$_x$: **(a)** The resistivity of 6 nm- and 10 nm-thick WO$_x$ layers after different reduction treatments and after processing the layers to simple FeFET structures (including a ms-FLA to crystallise the HZO), measured on CTLM structures. **(b)** Processed sample structure. **(c)** Sample structure after reduction and contact lift-off.

then defined by dry-etching the top TiN/W layer. Finally, the HZO was etched above the source, drain, and CTLM pads to ensure good contact for electrical measurements.

The samples with no reduction show that after the oxidation, ρ is large, as expected. Both reduction treatments result in a smaller ρ as compared to after the oxidation, but the ρ values still remain quite high, between $10\,\Omega\,\mathrm{cm}$ to $100\,\Omega\,\mathrm{cm}$. Resistivities in that range are interesting for FeFETs due to the limited free carriers and because this results in a channel sheet resistance in the tens of $\mathrm{M}\Omega/\square$. After the full process, which includes the crystallisation of the HZO by a ms-FLA, the $\mathrm{WO_x}$ on all samples was greatly reduced by 3 to 4 orders of magnitude. The samples which received no reduction treatment have the highest resistivity after the ms-FLA ($\sim 2 \times 10^{-2}\,\Omega\,\mathrm{cm}$). Such low channel sheet resistance values ($\sim 20\,\mathrm{k}\Omega/\square$) are not suitable for a FeFET memristor due to the high currents/power consumption and because the modulation of the carrier concentration by the ferroelectric polarisation is minimal. A large carrier concentration leads to a small polarisation-charge screening-length, which means that most of the channel is unaffected by polarisation switching. From this experiment we can conclude that the $\mathrm{WO_x}$ should not see the ms-FLA step of the HZO crystallisation. Hence, a new device structure was proposed where the gate is on the bottom. This enables to deposit and anneal the gate stack (TiN/HZO/W) first, before forming the channel by oxidising the W to $\mathrm{WO_x}$. A detailed description of the fabrication of such devices can be found in Section 5.1.2 and the electrical results in Section 5.1.3 [38].

4.2.2 WOx by Atomic Layer Deposition

After the characterisation of the first FeFET generation (Section 5.1.3) that utilised $\mathrm{WO_x}$ by RTO, we realised that even thinner channels ($<8\,\mathrm{nm}$) are required. Oxidising thin W layers is not ideal for this task. We need a deposition technology that is CMOS compatible and displays a good thickness control. The Atomic Layer Deposition (ALD) tool that we already used for thin HZO and TiN layers appeared as a suitable choice. Therefore, we added the W precursor BTBMW[4] to our PEALD tool.

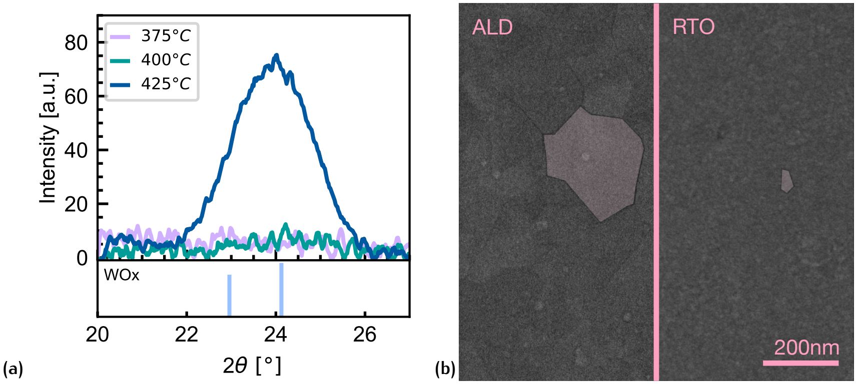


Figure 4.13: WO$_x$ grown by ALD: **(a)** Increased crystallisation temperature of 425 °C for 4 nm thick films. **(b)** Comparison of the film morphology between a 30 nm thick layer grown by RTO and ALD. Two SEM images with false-coloured grain boundaries.

4.2.2.1 Deposition

Having the ability to grow WO$_x$ by ALD opens the path to very thin channels with an excellent thickness control and uniformity across large areas. To start we implemented the deposition recipe supplied by the ALD tool company. The resulting layers were amorphous. To see if it is possible to grow crystalline WO$_x$ directly in the ALD, we varied the deposition pressure, temperature, and O$_2$-plasma power. Within the ALD tools allowed specifications we were not able to crystallise the WO$_x$ while it was growing. From now on all the WO$_x$ layers were grown with the same recipe at 375 °C with a pressure of 15 Torr and a O$_2$-plasma power of 250 W. The deposition rate was 0.46 Å/cycle. The target thickness for the channel of the FeFETs was 4 nm, which were achieved within 87 cycles. To guarantee a fully oxidised WO$_3$ layer, the samples were subsequently annealed in the RTA at 350 °C with 50 sccm O$_2$ for 6 min. Although these conditions were sufficient to crystallise 8 nm-thick WO$_x$ films, 4 nm-thick films remained amorphous. Similarly to what we observed for HZO, a higher thermal budget is required as the thickness decreases. When we increased the temperature to 425 °C the 4 nm thick WO$_x$ crystallised, as depicted in Figure 4.13a. Comparing the surface between a 30 nm thick layer grown by RTO and ALD, we observed that the film morphology differs between the two.

---

[4] BTBMW: bis(tert-butylimino)bis(dimethylamino)tungsten(VI) $[(CH_3)_3CN]_2W[N(CH_3)_2]_2$

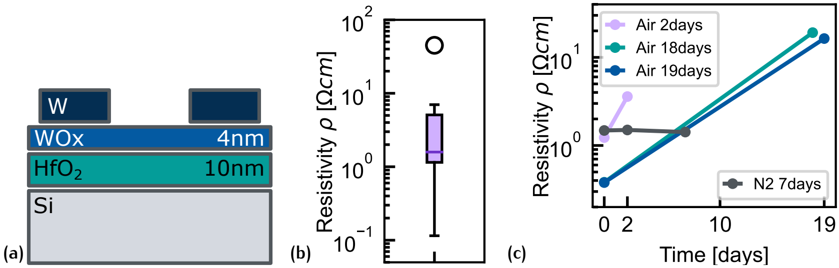


Figure 4.14: RTA reduction of WO_x grown by ALD: **(a)** Schematic of the layer stack with CTLM contacts. **(b)** Resulting resistivity spread after repeating a vacuum reduction at 350 °C for 2 min on 8 samples. The boxes extend from the lower to the upper quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data. Flier points are those past the end of the whiskers. **(c)** Drifting resistivity due to exposure to the clean room air. There is no drift when storing the samples in a N_2 desiccator.

Figure 4.13b presents two SEM images were we false-coloured two grains. While the RTO grain spans about 70 nm along its longest axis, the ALD grain reaches 280 nm. The grain size of the ALD WO_x is therefore much larger. Although the WO_x crystallises at 425 °C, keeping CMOS compatibility in mind, we decided to continue with an amorphous WO_x channel.

4.2.2.2 Reduction

In this section we investigate the reduction/oxidation of the WO_x grown by ALD caused by the environment, e.g. reduction anneals, exposure to ambient air, and contact with other oxides. We therefore fabricated samples with a similar layer structure as in back-gated FeFETs (Section 5.2.2), as depicted in Figure 4.14a. To determine the resistivity of the WO_x we deposited CTLM structures on top by a W lift-off process. First, we looked at the reduction of the WO_x layer by annealing it in vacuum for 2 min at 350 °C. This experiment was repeated 8 times. The resulting resistivity spread is reported in Figure 4.14b, with a median of 1.36 Ω cm. Clearly, there are some outliers, underlining the limited control of the reduction.

Next we want to report the influence of the ambient air on exposed WO_x layers. We realised during fabrication of the FeFETs that the resistivity of our WO_x layers changed when the samples were not immediately processed and covered by other passivation layers. A reduced WO_x layer

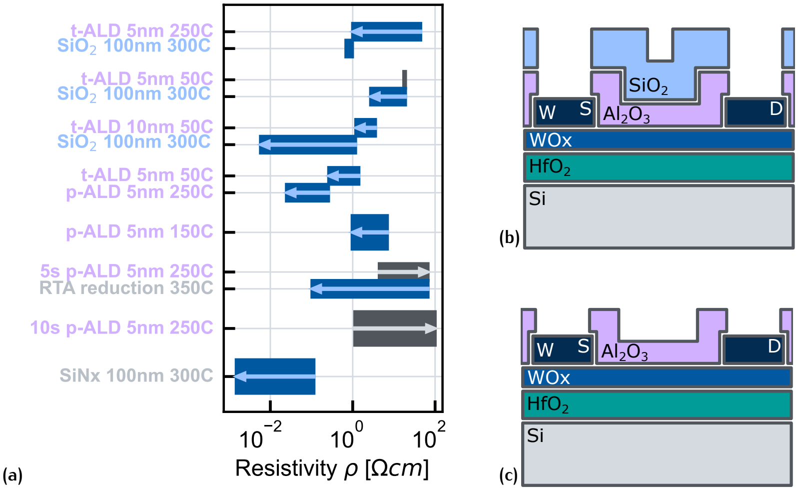


Figure 4.15: **(a)** Influence of the deposition of passivation layers on the resistivity of WO_x grown by ALD: Resistivity measurements before and after the deposition. Reducing results are displayed as blue boxes and oxidising ones in gray. If two subsequent depositions/treatments were conducted, they are grouped together. The colors of the labels correspond to the colors of the layers depicted in (b) and (c). **(b,c)** Schematic of the layer structure for two and one layer depositions, respectively.

slowly oxidises by the O_2 in the ambient air. Figure 4.14c shows the drift in resistivity that we measured on different samples. In between the measurements, the samples were stored with no special care and exposed to the clean room ambient air. When we stored a sample in a N_2 desiccator, no drift of the resistivity was observed, proving that the drift for the other samples originates from the exposure to air. Hence, it is important to store the samples in a N_2 desiccator or under vacuum until the WO_x is encapsulated by passivation layers.

These passivation layers are discussed next. It is crucial to bury the WO_x to protect it from ambient air as just seen. At the same time we came to know that the deposition of passivation and etch-stop layers such as Al_2O_3, SiN_x, or SiO_2 can have a big impact on the WO_x. Actually, our understanding is that any oxide that is in contact with WO_x and that is not fully stoichiometric tends to scavenge O_2 from the WO_x. We thus performed experiments where we measured the resistivity before and after the deposition of the aforementioned passivation layers under different conditions (temperature, plasma time) and with different tools (plasma vs. thermal ALD). Reducing results are displayed as blue boxes and oxidising ones in gray (Figure 4.15a). If two subsequent depositions/treatments were conducted, they are grouped together. The colors of the labels correspond to the colors of the grown layers depicted in Figures 4.15b and 4.15c. First we looked at the effect of growing the Al_2O_3, etch-stop layer on the resistivity of WO_x. Obviously, at elevated temperatures, independently of the ALD tool, the WO_x is reduced by the Al_2O_3 deposition. Even when lowering the temperature of the thermal ALD (t-ALD) to 50 °C, a smaller but notable reduction takes place. This effect is amplified for thicker Al_2O_3 layers. In the plasma ALD (p-ALD) an increase of the O_2-plasma time from the standard 1 s to 10 s and 15 s are the only conditions where the WO_x layer was more resistive after the deposition. In all cases, growing the SiO_2 passivation layer by Plasma Enhanced Chemical Vapour Deposition (PECVD) at 300 °C on top of the Al_2O_3 layer reduced the WO_x. Thicker Al_2O_3 layers reinforce this effect. Growing a SiN_x layer by PECVD on top of the WO_x led to a reduction as well. From all theses experiments we understand that other passivation layers that are in contact with the WO_x will scavenge Oxygen and reduce it. Low-temperature depositions or p-ALD depositons with increased plasma time can reduce or even reverse this effect. However, as soon as the sample is heated to elevated temperatures, the low temperature or long plasma films will scavenge the oxygen just as much.

For a controlled resistivity of the WO_x at the end of a device fabrication, we thus devised the following strategy: we do not reduce the WO_x after the crystallisation to keep it as resistive as possible. The Al_2O_3 etch-stop layer is deposited with increased plasma time. The SiO_2 passivation layer is deposited by PVD, a deposition technique that requires high-power RF-plasma, but no sample heating. In order to avoid overheating the sample by the RF-plasma, the deposition can be performed in multiple partial growths. At the end of the device processing, the resistivity of the WO_x can be slowly tuned towards more conductive values by gentle anneals in the RTA at around 200 °C. This resistivity tuning only works in one direction as a de novo oxidation is not possible.

5

FERROELECTRIC FIELD EFFECT TRANSISTORS

This chapter describes the design, fabrication, and characterisation of Ferroelectric Field-Effect Transistors (FeFETs). In a first step, relatively large µm-sized devices were fabricated with a rather simple and fast fabrication process, focusing on demonstrating the non-volatile impact of the ferroelectric polarisation on the channel resistance. In a second step, the devices were reduced in size to sub-µm dimensions and a more thorough fabrication process was developed to allow for three terminal crossbar arrays. The two FeFET generations are discussed in separate subsections.

5.1 FIRST FEFET GENERATION: µm-SIZED DEVICES

For the development of the thin film processing steps and for a proof of principle that the resistivity of WO_x can be modulated by the ferroelectric polarisation, devices that are as simple to process as possible were realized. This means making use of optical lithography, a fast turn-around transfer method for patterns larger than 2 µm.

5.1.1 Device Design

In this section we will describe the lithography mask set and its containing devices in more details. By using a shared gate (G) for all devices that can be accessed through the substrate we minimised the lithography steps to produce the FeFETs. Figure 5.1a shows the schematic of a FeFET's cross-section. The gate stack consists of a highly doped $n^{+}Si$ substrate, a 10 nm thick TiN layer, and a 10 nm thick HZO layer that extends over the entire chip. On top of the gate stack, thin WO_x films of different thicknesses (8 nm, 11.3 nm and 15 nm) form the channel. Local W-based source and drain contacts serve as electrodes to the channel. To electrically measure the devices, larger contact pads are required. Therefore, a passivation layer consisting of 5 nm of Al_2O_3 as an etch-stop layer and 100 nm of SiO_2 is

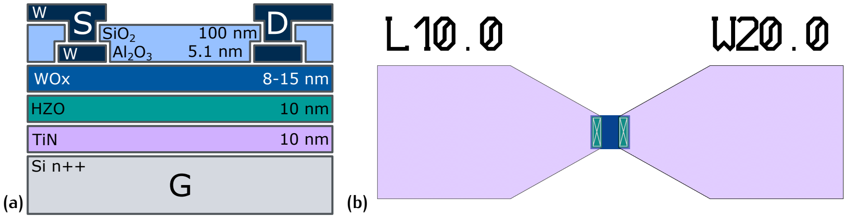


Figure 5.1: FeFET design: **(a)** Cross section of the device. **(b)** Top view of the device in a GDS layout file. The blue box shows the extend of the WO_x layer, the small teal squares on each side are the local S and D contacts (W_{ch}), the gray boxes with a cross indicate the openings through the passivation, and the large purple structures extending to the right and left are the S and D contact pads for electrical measurement.

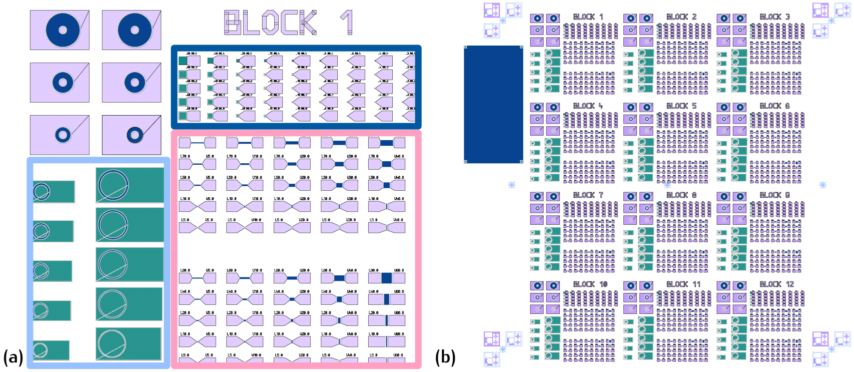


Figure 5.2: Chip design: **(a)** Overview of a block containing a span of FeFETs, capacitors, and CTLM structures. **(b)** Overview of the chip showing multiple blocks.

placed between the S/D and contact pads to prevent a direct connection to the HZO. Figure 5.1b depicts the top view of a FeFET with a channel width (W_{ch}) of 20 µm and a channel length (L_{ch}) of 10 µm. The channel length is defined as the distance between source and drain contacts while the width is defined by the structuring of the WO_x layer.

Because the optimum channel dimensions are not clear beforehand, channel geometries from long ($L_{ch} \gg W_{ch}$) to wide ($L_{ch} \ll W_{ch}$) were placed on the chip (rose box in Figure 5.2a). Both the channel length and width are varied from 5 to 100 µm. For the characterisation of the ferroelectric layer we placed a series of Metal-Semiconductor-Ferroelectric-Metal (MFSM) capacitors next to the FeFETs with a changing area from 3×3 to 60×60 µm² (dark blue box). The TiN (M) bottom contact is accessed through the substrate, the WO_x/W (SM) contact from the top, in the same manner as the S and D contacts of the FeFETs are accessed. For the characterisation of the WO_x channel we added Circular Transmission Line Measurement (CTLM) structures (light blue box). They allow to measure the resistivity of the WO_x and the contact resistance of the S and D contacts to the WO_x channel. This method is described in more detail in Section 3.3.1. Figure 5.2a shows the Graphic Design System (GDS) layout of the aforementioned devices that were placed together to form a block. The chip has a size of 2×2 cm² and hence, multiple of these blocks were placed on the chip to maximise the use of the area, as can be seen in Figure 5.2b. In addition of having multiple devices of the same dimensions for statistical analysis, such a block multiplication permits to detect spacial process non-uniformities.

In summary, we have a design with 4 optical lithography steps, namely the lift-off of the S and D contacts, the definition of the channel, the opening of the passivation, and the definition of the contact pads. Overall there are 50 devices per block and 12 blocks on the chip, for a total of 600 devices.

5.1.2 Device Fabrication

To keep the fabrication simple, we opted for a back-gated FeFET. Together with a conductive n^+ Si substrate this allows to reduce the processing steps to the minimum since the substrate acts as the shared gate (G) for all devices on the chip. Furthermore, this enables to first crystallise the HZO before the WO_x channel is formed, a necessity for these devices. We fabricated three chips with different channel thicknesses: 8 nm, 11.3 nm,

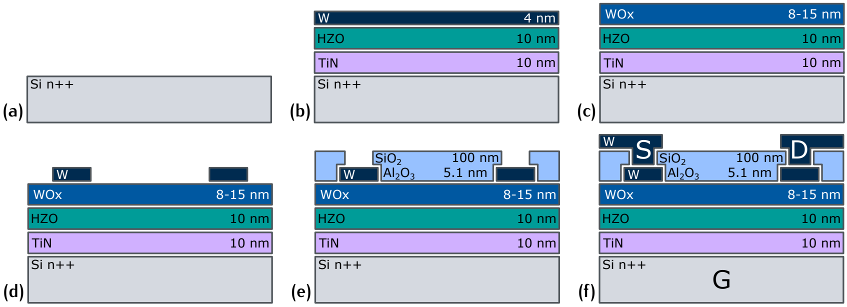


Figure 5.3: Process Flow of the µm-sized FeFETs: **(a)** Starting with a conductive substrate, **(b)** the gate layer stack consisting of 10 nm TiN, 10 nm HZO, and 10 nm W is deposited. **(c)** W is oxidised and crystallised to WO_3. **(d)** Source and drain are deposited by a lift-off process. **(e)** 5 nm of Al_2O_3 and SiO_2 are deposited and openings to the contacts are etched. **(f)** 100 nm of W is used for the contact pads.

and 15 nm. Besides the three chips with the ferroelectric gate stack, a similar chip was fabricated with a HfO_2 gate dielectric instead of the HZO layer. We will refer to this chip as "control sample" as it serves the purpose to prove that the ferroelectricity in HZO is responsible for the resistance modulation of the channel. In the following paragraphs, the process flow is described step-by-step:

1. Starting from a highly doped n++ Si substrate (Figure 5.3a), the gate stack is grown first: 10 nm of TiN were deposited using a TDMAT[1] precursor and N_2/H_2 plasma in an Oxford Instruments Plasma Enhanced Atomic Layer Deposition (PEALD) system. An approximately 10 nm thick layer of HZO was grown in a process using alternating cycles of TEMAH[2] and ZrCMMM[3] at 300 °C. For the control sample with a non-ferroelectric HfO_2 gate dielectric, approximately of 10 nm HfO_2 were grown by only using cycles of TEMAH precursors at 300 °C. The sample was then immediately transferred to a sputter chamber for the deposition of 4 nm of W by Physical Vapour Deposition (PVD). The resulting thin film layers are depicted in Figure 5.3b.

---

[1] TDMAT: Tetrakis-(dimethylamino)titanium $[(CH_3)_2N]_4Ti$

[2] TEMAH: tetrakis-(ethylmethylamino)-hafnium $[(CH_3)(C_2H_5)N]_4Hf$

[3] ZrCMMM: bis(methyl-η5-cyclopentadienyl)methoxymethylzirconium $(CH_3C_5H_4)2Zr(OCH_3)CH_3$

2. In a next step, the HZO layer must be crystallised in the ferroelectric orthorhombic phase. It is important that the HZO is covered by a capping layer (W in our case), as described in more detail in Section 2.3.1. For the crystallisation of HZO a millisecond Flash Lamp Anneal (ms-FLA) [31] with a background temperature of 375 °C was performed.

3. After crystallisation, the 4 nm thin W layer was reduced to ~2.5 nm and ~2 nm by Ar sputtering for the samples with a channel thickness of 11.3 nm and 8 nm, respectively. For the sample with a channel thickness of 15 nm the W layer was left at 4 nm. W was then crystallised and oxidised to 8 nm, 11.3 nm, and 15 nm WO_3 in a Rapid Thermal Annealer (RTA) at 350 °C for 6 min with 50 sccm O_2. Afterwards a reduction of the WO_3 to WO_x was performed in the same RTA by H_2 annealing at 150 °C and vacuum annealing at 350 °C (Figure 5.3c)

4. The source (S) and drain (D) contacts were formed by depositing W in a PVD system and a subsequent lift-off process (Figure 5.3d). The WO_x channel was then structured with an SF_6 plasma using a Reactive Ion Etcher (RIE).

5. The passivation above the FeFET structures consists of 5 nm of Al_2O_3 by thermal ALD using TMA[4] as precursor and 100 nm of SiO_2 by Plasma-Enhanced Chemical Vapor Deposition (PECVD). Vias were etched using an RIE with a CHF_3/O_2 plasma (Figure 5.3e).

6. Finally, the contact pads were realised by depositing 100 nm W by PVD and a subsequent definition in an RIE with a SF_6/O_2 plasma (Figure 5.3f).

Because of the micrometer scale of the features, optical instead of electron-beam lithography was applied, a pattern transfer method with a high throughput, ideal for the development of a process. For all four lithography steps the positive photoresist AZ1505 was applied and exposed by direct laser writing.

As visible in the Bright Field Scanning Transmission Electron Microscopy (BF-STEM), our fabrication process results in sharp interfaces between the layers and crystalline WO_x grains (Figure 5.4a to 5.4c). The Energy-Dispersive X-ray Spectroscopy (EDS) line profile shown in 5.4a

---

[4] TMA: trimethylaluminum $(CH_3)_3Al$

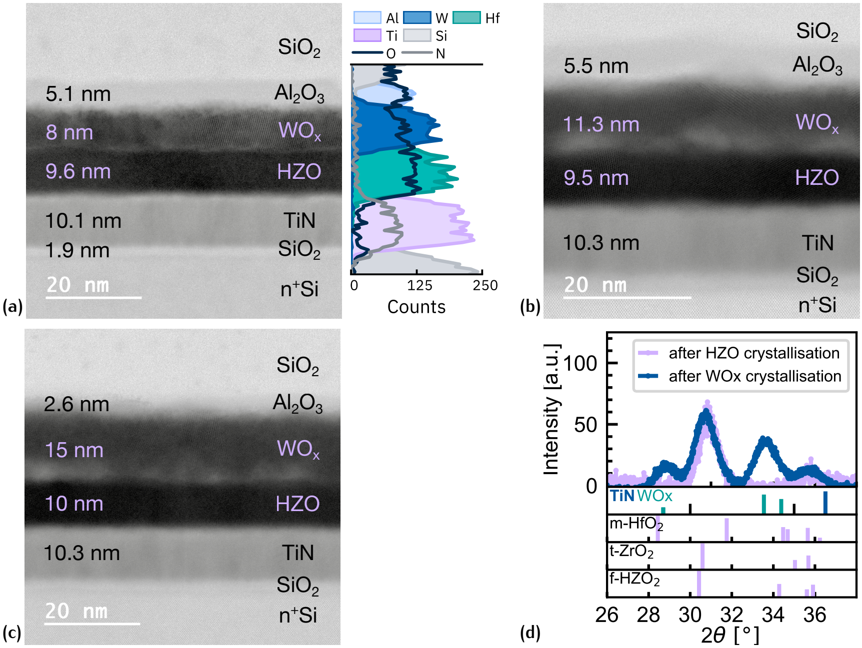


Figure 5.4: Structural data of the FeFET. **(a)** Cross-sectional BF-STEM image with energy-dispersive X-ray spectroscopy (EDS) line profile of the $SiO_2/Al_2O_3/WO_x/HZO/TiN/n^+$ Si gate region for the 8 nm sample. **(b)** Same as (a), but for the 11.3 nm sample. **(c)** Same as (a) and (b), but for the 15 nm sample. **(d)** GIXRD for a diffraction angle (2θ) from 26° to 38° showing the presence of the f- or t-phase in HZO after crystallisation and after the W layer was oxidised to WO_x. Lattice parameters for the tetragonal P-42_1m phase of WO_x were taken from (ICSD-86144) [267].

confirms the targeted elemental distributions and reveals regions of intermixing between the various layers. Grazing Incidence X-ray Diffraction (GIXRD) analysis displays the characteristic peak at 30.5° of the f- or t-phase in HZO (Figure 5.4d). The diffractogram is consistent with data from Metal-Ferroelectric-Metal (MFM) structures with the same HZO as published in Ref. [31]. No monoclinic phase (peaks at 28.2° and 31.8°) [175] is present in our samples, which is a consequence of the low temperature crystallisation technique. Following the oxidation and crystallisation of W to WO_x, GIXRD still exhibits no monoclinic HZO phase, but displays two additional peaks at 28.6° and 33.3° that can be attributed to the tetragonal $P\text{-}42_1m$ phase of WO_x (ICSD-86144) [267]. The HfO_2 peak positions were calculated from the lattice parameters of the m-, t-, and f-phase that were taken from [175].

As described above, the WO_x channel was reduced by a H_2 annealing at 150 °C and vacuum annealing at 350 °C. To reach a suitable resistivity of the channel, this step needs to be controlled. Circular Transmission Line Measurement (CTLM) structures of different sizes (described in more details in Section 3.3.1) were defined by the same lift-off step as S and D. After the reduction treatment we still measured very high resistivities and were not able to get meaningful CTLM results. The subsequent Al_2O_3 deposition at 250 °C in the thermal ALD and SiO_2 passivation in the PECVD at 300 °C is quite close to the reduction temperature. At the end of the full process, it became evident that the WO_x was reduced by the passivation deposition discussed in Section 4.2.2.2. Th resistivity of WO_x was reduced from not measurable to $3.27 \times 10^{-1}\,\Omega\,\text{cm}$ after the complete process. Both the Al_2O_3 and the SiO_2 layers like to scavenge oxygen, as described in more details in Section 4.2.2.2. It is difficult to control the exact amount of reduction by the passivation deposition. The general approach is to leave the WO_x quite resistive. A further reduction can still be achieved at the end of the full process by annealing. On the other hand, oxidising the WO_x is not possible anymore. If the WO_x turns out to be too conductive, a too high carrier concentration will diminish the electrostatic influence of polarisation on the channel resistance. It is for the same reason that a gate-first approach was chosen as the crystallisation of the HZO at 375 °C would reduce the WO_x to an almost metallic state.

Finally, the S and D can be accessed from the top of the chip by two needle probes. The G is accessible through the highly n^+ doped Si substrate and is shared between all devices on our chip. Because the TiN gate is not structured and extends over the whole sample the source and drain contacts fully overlap with the gate.

5.1.3 Device Results

In this section we will present the electrical measurements that were performed on the FeFETs and capacitors. First, we assess the quality of our ferroelectric films and then we characterise the memristive behaviour of our FeFETs. These results were published in Ref. [38].

5.1.3.1 *Ferroelectric Properties of HZO*

To get the final proof that we indeed have the ferroelectric orthorhombic phase and not the tetragonal one, we conducted "Capacitance vs. Voltage" ($C-V$) and "Polarisation vs. Voltage" ($P-V$) measurements on the Metal-Semiconductor-Ferroelectric-Metal (MSFM) capacitor structures. A ferroelectric typical butterfly-shaped hysteresis curve, with a capacitance per unit area $C_{OX} = 27\,\mathrm{fF/\mu m^2}$ (Figure 5.5a) was measured on a $60\,\mu\mathrm{m} \times 60\,\mu\mathrm{m}$ W/WO$_x$/HZO/TiN/n^+Si capacitor on the chip with the 8 nm thick WO$_x$ layer. For comparison, an equally sized TiN/HZO/TiN Metal-Ferroelectric-Metal (MFM) capacitor was measured on a different chip (Figure 5.5b). The asymmetric behaviour of the capacitance in the MSFM structure originates from the asymmetric electrodes (WO$_x$, TiN).

$P-V$ measurements (Figure 5.6a) show typical characteristics: In the pristine state, the $P-V$ curve is Anti-Ferroelectric (AFE)-like with hysteresis, especially on the negative voltage side [174]. We applied 10^5 switching cycles of triangular pulses with an amplitude of $\pm 3.8\,\mathrm{V}$ at a frequency of 100 kHz, resulting in a pinched $P-V$ curve with a positive (negative) remanent polarisation $P_{r+} = 12.4\,\mu\mathrm{C/cm^2}$ ($P_{r-} = 11.8\,\mu\mathrm{C/cm^2}$). Furthermore, a slight imprint with a positive coercive voltage of $+V_C = 0.91\,\mathrm{V}$ and a negative one of $-V_C = -1.27\,\mathrm{V}$ were observed due to the asymmetric electrodes. The same experiment was repeated on the samples with 11.3 nm- and 15 nm-thick WO$_x$, as shown in Figures 5.6b and 5.6c, respectively. No clear difference can be noticed. We conclude that the HZO layers are very similar on all 3 samples. From here on, if not otherwise stated, measurements were carried out on the 8 nm-thick WO$_x$ sample.

We continued by assessing the cycling endurance of our HZO. The total remanent polarisation ($P = |P_{r-}| + |P_{r+}|$) was determined by Positive Up Negative Down (PUND) measurements. The cycling frequency was set to 1 kHz up to 10^4 cycles, 10 kHz up to 10^5 cycles, and 100 kHz for cycles

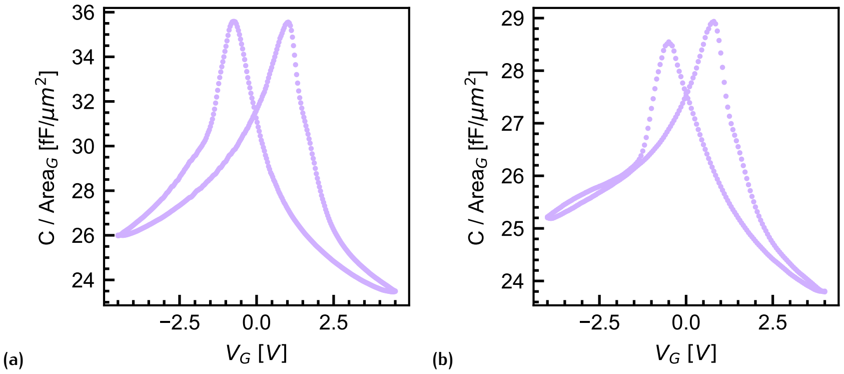


Figure 5.5: Capacitance measurements on a 60 µm × 60 µm **(a)** TiN/HZO/TiN MFM and **(b)** W/WO_x/HZO/TiN/n^+Si MSFM structure.

above 10^5. To determine the influence of the WO_x electrode we repeated the endurance measurements on a TiN/HZO/TiN MIM capacitor of the same size. The cycling endurance of our HZO is 10^8 for an MFM structure (Figure 5.7a) and 8×10^6 in the case of the MSFM configuration present in our FeFET (Figure 5.7b). The reduced endurance of the MSFM capacitor could originate from the oxidation of the W (capping layer during crystallisation) to the WO_x: oxygen from the HZO layer diffuses into the W interface to form WO_x, which increases the oxygen vacancies in the HZO. During cycling stress, a conductive path formed by oxygen vacancies is thus facilitated (Section 2.3.1.2).

5.1.3.2 Channel Properties

Having confirmed the ferroelectric nature of our HZO gate dielectric, the electrical characterisation of the WO_x channel in a FeFET device was performed next. WO_x is an n-type semiconductor. When the HZO ferroelectric polar tion points towards (outwards) the interface with WO_x, free-carriers accumulate (deplete) at the interface and screen the electric field so that the channel resistance (R_{DS}) decreases (increases): it is a junction-less transistor. Note that for both states, the polarisation is screened at the WO_x interface and that no depolarisation field destabilises it. The screening length (x_d) increases as the carrier density (N_D) decreases. To investigate the effect of the polarisation P_r and the channel thickness d_{WOx}, samples with different d_{WOx} and one with a non-ferroelectric HfO_2 gate dielectric were measured. R_{DS} was measured between source and drain after each 2 µs-long write

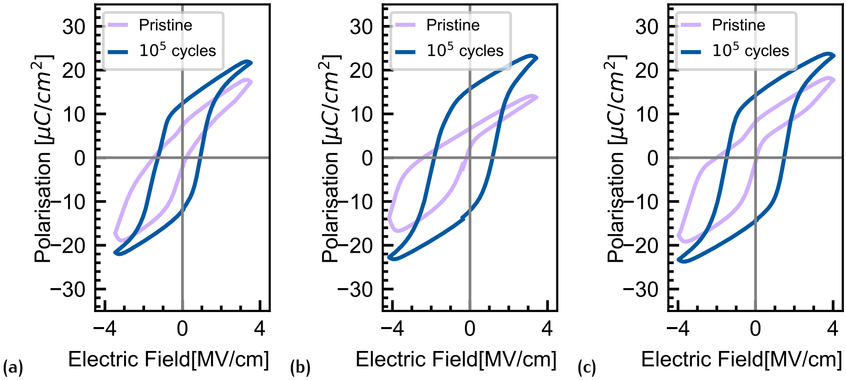


Figure 5.6: "Polarisation vs. Voltage" ($P - V$) characteristics measured on 60 µm × 60 µm W/WO$_x$/HZO/TiN/n$^+$Si MFM structures at 5 kHz in the pristine state and after 1×10^5 cycles: **(a)** $d_{\mathrm{WOx}} = 8$ nm, **(b)** $d_{\mathrm{WOx}} = 11.3$ nm, and **(c)** $d_{\mathrm{WOx}} = 15$ nm.

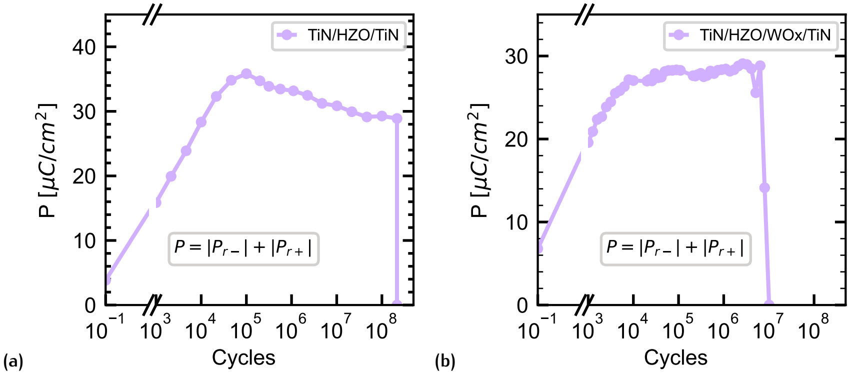


Figure 5.7: Endurance measurements of 10 nm HZO. The total remanent polarisation ($P = |P_{\mathrm{r}-}| + |P_{\mathrm{r}+}|$) was determined by Positive Up Negative Down (PUND) measurements with 1 kHz and ±3.5 V. The cycling frequency was set to 1 kHz up to 1×10^4 cycles, 10 kHz up to 1×10^5 cycles, and 100 kHz for cycles above 1×10^5: **(a)** TiN/HZO/TiN MFM configuration cycled at ±3.5 V. **(b)** W/WO$_x$/HZO/TiN MSFM configuration cycled at ±3.0 V with a −0.5 V offset.

pulse (V_{write}) applied to the gate. The measurement scheme can be seen in Figure 3.8 and is described in more details in Section 3.3.2. The channel width and length (L_{ch}, W_{ch}) of the measured devices were (20 µm, 5 µm), (5 µm, 5 µm), (5 µm, 10 µm), (10 µm, 10 µm) for the 8 nm, 10.3 nm, 15 nm, and HfO_2 sample, respectively. For ease of comparison, R_{DS} is normalised by R_{ON} (Figure 5.8a to 5.8d). A clear hysteresis in R_{DS} can be observed for devices with a ferroelectric HZO gate dielectric. To confirm that the modulation of the channel resistance originates from P_{r} and not from another effect, an identical device with a non-ferroelectric HfO_2 gate dielectric and an 8 nm thick WO_x channel was measured. R_{DS} shows no hysteresis in the non-ferroelectric HfO_2 sample (Figure 5.8c) and further proves that the hysteresis originates from the ferroelectricity in HZO. In addition to the polarisation in the HZO, the type and concentration of the free charge carriers [287, 288] as well as d_{WOx} influence the Dynamic Range (DR). The channel geometry was found to impact the DR only at high aspect ratios where the $L_{ch}/W_{ch} > 10$. This behaviour can be attributed to a drop in the write field far away from the source and drain. For a maximum reduction in the channel off-current, the screening length x_{d} should be larger than d_{WOx}. Using Poisson's equation, the relationship between x_{d} and N_D can be expressed as follows: [288–290]

$$x_{\mathrm{d}} = \frac{\epsilon_0 \epsilon_{\mathrm{WOx}}}{C_{\mathrm{HZO}}} \left[\left(1 + \frac{2C_{\mathrm{HZO}}^2 V_{\mathrm{GS}}}{qN_{\mathrm{D}}\epsilon_0\epsilon_{\mathrm{WOx}}} \right)^{1/2} - 1 \right], \tag{5.1}$$

where ϵ_0 is the vacuum permittivity, ϵ_{WOx} the permittivity of WO_x, C_{HZO} is the HZO capacitance per unit area ($C_{HZO} = 31.4\,\mathrm{fF/\mu m^2}$, Figure 5.5a), and V_{GS} is the polarisation charge-induced potential across HZO. The electron carrier concentration ($N_{\mathrm{D}} = 1.01 \times 10^{20}\,\mathrm{cm^{-3}}$) and the channel mobility ($\mu_{\mathrm{H}} = 0.19\,\mathrm{cm^2\,V^{-1}\,s}$) were determined by Hall measurements carried out on a similar sample with the same layer structure and deposition conditions. The WO_x on the similar sample has been crystallised and reduced to match the resistivity of the WO_x of the FeFET sample. The permittivity of WO_x ($\epsilon_{\mathrm{WOx}} = 189$) was calculated using the following equation of two capacitances in series:

$$\frac{1}{C_{\mathrm{WOxHZO}}} = \frac{1}{C_{\mathrm{HZO}}} + \frac{d_{\mathrm{WOx}}}{\epsilon_0 * \epsilon_{\mathrm{WOx}}}, \tag{5.2}$$

where C_{WOxHZO} is the capacitance per unit area of the W/WO_x/HZO/TiN stack, C_{HZO} the capacitance per unit area of the TiN/HZO/TiN stack,

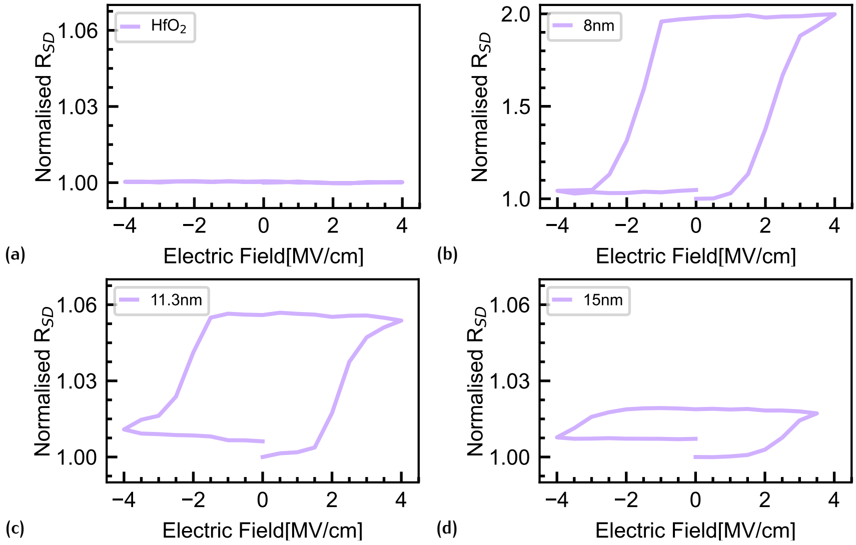


Figure 5.8: Channel resistance modulation of different channel thicknesses and gate dielectrics: **(a, b)** Comparison of simultaneously processed samples with HZO and HfO$_2$ gate dielectric. The non ferroelectric HfO$_2$ sample does not show any channel resistance hysteresis. **(b,c,d)** Influence of the channel thickness (d_{WOx}) on the DR.

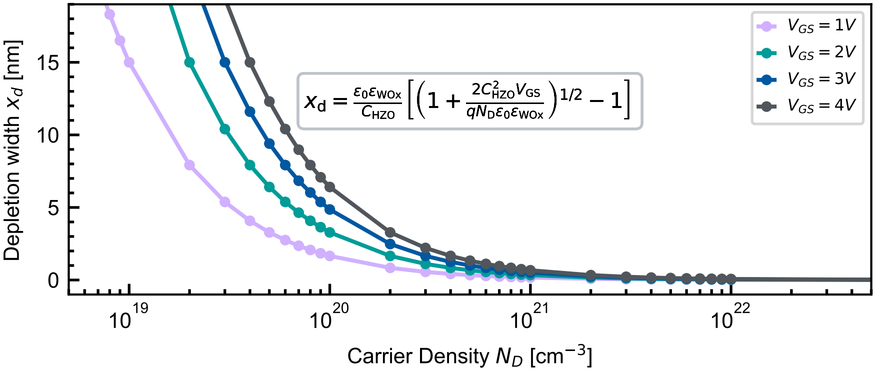


Figure 5.9: Depletion width x_d of electrons in WO$_x$ as a function of the carrier concentration N_D, calculated for different V_GS according to Equation 5.1.

$d_{\mathrm{WOx}} = 8\,\mathrm{nm}$ the thickness of the $\mathrm{WO_x}$ channel, and ϵ_0 the vacuum permittivity. From Figure 5.5b we get $C_{\mathrm{WOxHZO}} = 27.5\,\mathrm{fF/\mu m^2}$. Using Equation 5.1, a depletion width $x_\mathrm{d} = 1.7\,\mathrm{nm}$, $3.3\,\mathrm{nm}$, $4.8\,\mathrm{nm}$ and $6.4\,\mathrm{nm}$ for $V_{\mathrm{GS}} = 1\,\mathrm{V}$, $2\,\mathrm{V}$, $3\,\mathrm{V}$ and $4\,\mathrm{V}$, was calculated (Figure 5.9), respectively. For a constant polarisation, the largest effect is obtained if $d_{\mathrm{WOx}} < x_\mathrm{d} = 6.4\,\mathrm{nm}$ or $N_\mathrm{D} < 1 \times 10^{20}\,\mathrm{cm^{-3}}$. Three samples with different d_{WOx} were realised to benchmark this estimation with experimental data. The polarisation does not change between the three structures (Figure 5.6). By increasing d_{WOx} from 8 nm to 11.3 nm and 15 nm the DR decreases from ≈1.9 to ≈1.05 and ≈1.01, as shown in Figure 5.8b, 5.8c, and 5.8d, respectively. The total resistance of the channel in this junction-less FeFET can be approximated by the resistance of two channels in parallel [291]: one of thickness x_d in which the sheet carrier density and thus the resistivity is modulated upon polarisation switching and one of thickness $d_{\mathrm{WOx}} - x_\mathrm{d}$ with a constant resistivity. Those results agree well with the x_d calculated by Equation 5.1.

5.1.3.3 Memristor Properties

For neuromorphic applications multiple (analog) levels of the channel resistance, good retention properties, low device-to-device and cycle-to-cycle variability, fast conductance updates, and low power consumption are important characteristics of ideal devices. [21, 292–294] The exact requirements vary depending on the details of operation and from one implementation to the other. As an example, inference workloads would use off-line trained weights transferred to the chip to operate the network. Hence, the precision of the weights (≥3 bit) is more relaxed than in the case of a chip designed to perform on-line learning. [295] In our device structure, weights are defined through the intermediate states of the channel resistance, enabled via the multi-domain nature of the ferroelectric HZO layer. [137, 296, 297] The size of the ferroelectric domains in HZO was found to be in the order of the thickness of the ferroelectric film. [31] If the size of the channel compares to the ferroelectric domain size, discrete resistance levels are reached. [298] Our channel dimensions differ by two to three orders of magnitude, which enables analog resistance levels by switching only a subset of the domains. [137] The fraction of the switched ferroelectric domains depends on the amplitude, width, and number of applied write pulses. Different pulsing schemes on HZO have been investigated in the past. [299] For on-line learning algorithms running on crossbar arrays integrated on CMOS, potentiation and depression pulse schemes with a constant pulse amplitude and width are preferred to those with

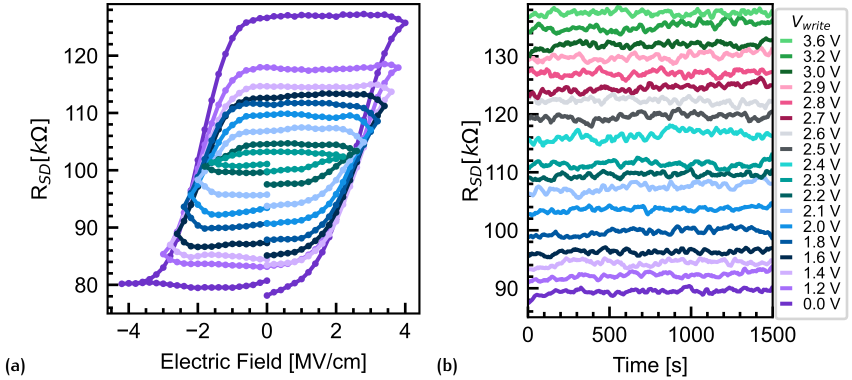


Figure 5.10: Analogue multi-level behaviour of a FeFET of 20 µm width and 5 µm length. **(a)** Channel resistance (R_{DS}) after the application of 5 µs write pulses (V_{write}) of varying amplitudes. Each data point corresponds to a resistance measurement between S and D at $V_{\mathrm{read,D}} = 200\,\mathrm{mV}$. The different curves correspond to different consecutive measurements with reducing V_{write} range. **(b)** Retention measurement for 1500 s. A $V_{\mathrm{read,D}} = 200\,\mathrm{mV}$ was uninterruptedly applied while the current was measured every 5 s to determine R_{DS}.

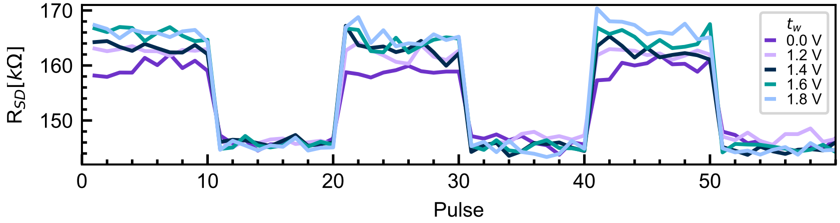


Figure 5.11: Stability of an intermediate state after repeating a writing pulse and reading sequence ten times on a FeFET with a 10 µm wide and 40 µm long channel. A writing pulse of 3 V and width t_w from 0.5 µs to 5 µs was applied, then the channel resistance R_{DS} was measured at $V_{read} = 200\,\mathrm{mV}$. The same sequence was repeated ten times, then with a writing voltage of −3 V.

varying amplitude. Nevertheless, for the proof of concept the multi-state nature of a 20 µm wide and 5 µm long FeFET was investigated by applying voltage pulses of varying amplitudes (V_{write}), while keeping a fixed pulse duration of 5 µs (Figure 5.10a). A more detailed description of the writing and reading procedure can be found in Section 3.3.2. This pulse scheme results in the best linearity in potentiation and depression. [299] By sweeping V_{write} from $-4\,\text{V}$ to $4\,\text{V}$, R_{DS} shows a hysteretic cycle from 80 kΩ to 125 kΩ with various intermediate states ($DR \approx 1.55$). By reducing the range of V_{write} numerous R_{DS} sub-loops can be accessed, as indicated in Figure 5.10a. The asymmetry in the hysteresis loop is due to the imprint in the ferroelectric layer. The robustness of the intermediate state upon application of the same pulse up to ten times without reset was measured for different pulse widths and pulse amplitudes. For pulses in the range 0.5 µs to 5 µs, no cumulative effect was observed (see for example in Figure 5.11 with $V_G = 3\,\text{V}$). Furthermore, the retention properties have been studied, as demonstrated in Figure 5.10b. First, an intermediate state was written by a 5 µs pulse. Then, a source-to-drain voltage $V_{\text{DS}} = 200\,\text{mV}$ was continuously applied for 1500 s, while R_{DS} was measured every 5 s. Between each measured intermediate state the FeFET was reset to its low resistive state (R_{ON}) by setting $V_{\text{write}} = -4\,\text{V}$ during 1 ms. The FeFET possesses stable retention properties for 18 differentiable channel resistances (>4 bit) for the full 1500 s. The stable retention measurement hints to an absence of depolarisation or other screening mechanisms. This is in agreement with a partially depleted channel, which is still able to screen the polarisation charges.

For on-chip learning, artificial synapses require a finer mesh of intermediate levels. In addition, symmetric and linear potentiation and depression are desirable. With respect to symmetry the field-driven ferroelectric switching is advantageous over competing technologies that often show abrupt or unidirectional switching. [294, 299] The requirement of low variability is relaxed as the training occurs on a specific hardware and thus incorporates the variability in its solution. [295] To investigate the linearity and symmetry of the potentiation and depression, multiple write pulses of increasing and decreasing amplitudes were applied. For the depression V_{write} was increased from 0 V to 3.5 V and for the potentiation decreased from 0 V to $-3\,\text{V}$ in steps of 100 mV (Figure 5.12a). It has to be noted there, that potentiation and depression are used to describe the conductance. Hence, the potentiation means a decrease in resistance and the depression means an increase in resistance. The duration of the write pulses was

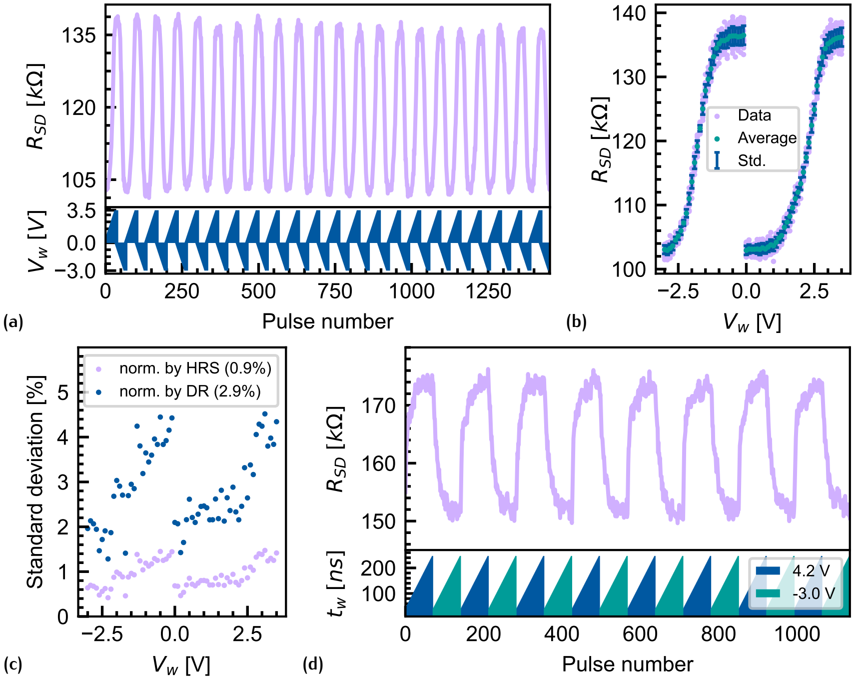


Figure 5.12: Potentiation and depression of a 20 µm wide and 5 µm long FeFET. **(a)** Top panel: Multiple potentiation and depression cycles of the channel resistance (R_{DS}) with varying pulse amplitude (V_{write}) at a constant pulse width (t_{write}). Bottom panel: Corresponding write pulse sequence . After each pulse R_{DS} was measured. **(b)** Absolute cycle-to-cycle variation of R_{DS} including the data itself, its average and standard deviation error bars. **(c)** Standard deviation of the R_{DS} cycle-to-cycle variations normalised by the R_{ON} and by the DR. **(d)** Multiple potentiation and depression cycles of R_{DS} with increasing t_{write} from 40 ns to 250 ns at a constant V_{write}.

kept constant at 10 µs. When averaging over the 22 measured cycles (Figure 5.12b), multiple states with small standard deviation can be observed. Normalising the cycle-to-cycle standard deviation by R_{ON} and DR reveals an average value of 0.9 % and 2.9 %, respectively (Figure 5.12c). The number and overlap of states are defined by the potentiation and depression step size. The latter could be reduced further to increase the resolution.

When fitting the potentiation range from 1 V to 3.1 V and depression range from −0.9 V to −3.0 V of the same 22 cycles by linear regression (Figure 5.13a), an adjusted residual-square value of 0.952 is obtained. The residuals normalised by the R_{DS} window as a function of the pulse number is depicted in Figure 5.13b. Chen *et al.* [45] proposed a method where the potentiation and depression is fitted by an exponential function to extract non-linearity factors (α_p/α_d). Zero means linear, while positive and negative values with an absolute greater than zero represent a non-linear change. The same authors published a simulator for the classification of the MNIST dataset that accepts real device characteristics and non-idealities such as the non-linearity. When our data is fitted according to [45], we obtain good non-linearity factors of $\alpha_p = 0.38$ $\alpha_d = 1.0$ (Figure 5.13c) for the potentiation and depression, respectively. Note however that, it is difficult to fit the s-like shaped potentiation and depression of ferroelectric-based memristors with an exponential function.

For a more detailed analysis of the symmetry, Gaussian Process Regression (GPR) was used to predict a noise free signal (Figure 5.13d) [300]. Gong *et al.* define the Signal to Noise Ratio (SNR) as follows:

$$SNR = \frac{\Delta R_{SD}}{r},$$

where ΔR_{SD} is the change in resistance and r the residuals from the noise-free signal fit for each pulse. In our case, the noise is the cycle-to-cycle variation. Plotting ΔR_{SD} (Figure 5.13e) as a function of the pulse number reveals a diminishing ΔR_{SD} towards the extremes, a consequence of the pinching ferroelectric polarisation. The noisier signals (Figure 5.13f) towards the extremes is the result of small ΔR_{SD}, while r does not get smaller. The symmetry factor (SF) was then calculated using the following equation: [300]

$$SF = |\frac{\Delta R_{SD+} - \Delta R_{SD-}}{\Delta R_{SD+} + \Delta R_{SD-}}|,$$

where ΔR_{SD+} is the depression and ΔR_{SD-} is the potentiation change in resistance at a certain resistance level. By this definition, SF can take values between 0 and 1 where 0 is the perfect symmetry. The less linear the range of the data becomes, the larger is SF (Figure 5.13g). The average across the full resistance range is $SF = 0.203$ while the most linear part in the center reaches a very good symmetry factor of $SF = 0.08$.

Short programming pulses are advantageous as fast writing and low-power consumption are important for neuromorphic applications. By varying the pulse width from 40 ns to 250 ns with a fixed amplitude (Figure 5.12d), the shortest applied pulse of 40 ns already changes the resistance and demonstrates the very fast writing capabilities of the FeFET. It is expected that even shorter pulses could successfully program the memristor. [137] In our device, little energy is consumed while writing a state. When applying $V_{\text{write}} = 3.5\,\text{V}$ a gate current of $I_{\text{gate}} = 3.02 \times 10^{-8}\,\text{A}$ is measured. Applying a write pulse duration of $t_{\text{write}} = 200\,\text{ns}$ results in $E = \frac{V_{\text{write}} \cdot I_{\text{gate}} \cdot t_{\text{write}}}{W_{ch} \cdot L_{ch}} = 2.1 \times 10^{-17}\,\text{J}\,\mu\text{m}^{-2}$, where L_{ch} is the length and L_{ch} the width of the gate. This very low write energy is promising for the implementation of such FeFETs in crossbar arrays. It should be noted here that when scaled up, additional sources of power consumption such as the charging/discharging of the metal lines will be introduced.

5.1.3.4 Conclusion and Possible Improvements

We proposed a device concept based on the ferroelectric field effect into a thin WO_x channel using HZO gate dielectrics that can be used as a synaptic element in hardware-supported neural networks. By utilising a junctionless transistor design, no high temperature source and drain activation is required. The fabrication process is compatible with an integration in the Back-End-Of-Line of CMOS technology. It relies on earth-abundant materials, which makes FeFETs attractive and flexible for large-scale integration. By comparing HZO- and HfO_2-based devices and carefully analysing capacitor and transistor data, we unambiguously show that the channel resistance is directly coupled to the polarisation of the HZO layer and that it can be programmed in a non-volatile manner. Multilevel states programmed over more than 4 bit-depth with a stable retention over 1500 s and an almost symmetric potentiation and depression are demonstrated, together with a low programming energy. The property of the WO_x layer and the geometry of the device can be arranged so that a well-suited resistance range is obtained, favourable to build large-scale crossbar arrays.

Table 5.1 summarises the characteristics of our first FeFET generation. From this we can identify some characteristics that can potentially be improved by adapting the design and processing. The large device-to-device variability is believed to come from the non-uniform WO_x, which in turn

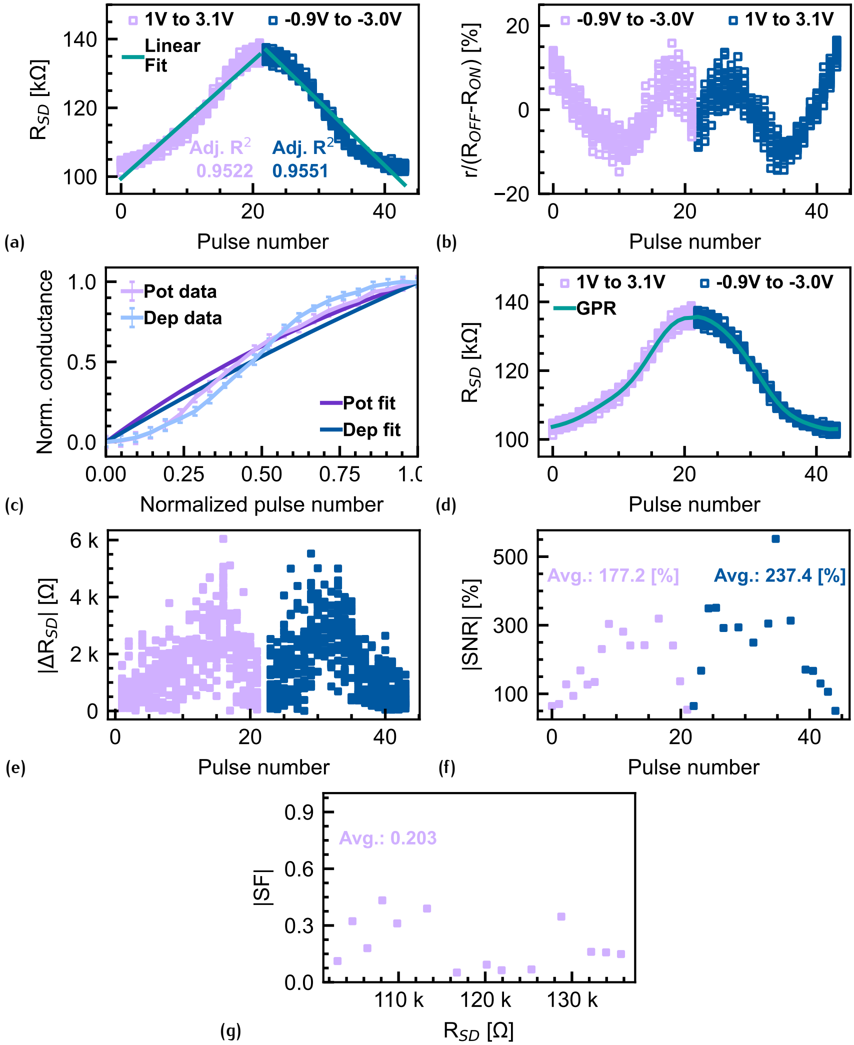


Figure 5.13: Extraction of the linearity and symmetry metrics of the same device as reported in Figure 5.12a to 5.12c. Data from multiple cycles with 22 depression (purple, 1 to 3.1 V) and 22 potentiation pulses (blue, −0.9 to −3 V) are reported. **(a)** Linear regression fit (teal). **(b)** Absolute residuals r normalised by the channel resistance window. **(c)** Exponential fit for linearity parameter extraction [45]. **(d)** GPR predicted noise free signal (teal). **(e)** Absolute change of R_{DS} after each potentiation and depression pulse. **(f)** Absolute SNR. **(g)** Symmetry factor (SF).

	1 gen. FeFET	Target
Dynamic range	1.55 to 1.98	8 to 100 [20]
R_{ON}	$>100\,\mathrm{k\Omega}$	$>10\,\mathrm{M\Omega}$
Number of states	$22^{\ddagger}$, $18^{\S}$	>100 [20]
Programming time	50 ns to 5 µs	ns
Programming voltage	$3.1\,\mathrm{V}/-3\,\mathrm{V}$	$<5\,\mathrm{V}$
Linearity (α_d/α_p)	0.38/1.0	0 [45]
Linear regression	0.952/0.955	1/1
SF [0 to 1] (average)	0.203	0 [300]
Symmetry ($\lvert\alpha_p - \alpha_d\rvert$)	0.62	0 [299]
Write energy	0.525 fJ	<10 fJ [20]
Area	$\geq 5 \times 5\,\mu\mathrm{m}^2$	$<10 \times 10\,\mathrm{nm}^2$
Device-to-device var.	Huge (not shown)	0
Cycle-to-cycle var.	$\sim 2.9\,\%^{\dagger}$, $\sim 0.9\,\%^{*}$	0
Endurance	$8 \times 10^{6\#}$	$>10^9$ [20]

† Normalised by the resistance window ($R_{\mathrm{ON}} - R_{\mathrm{ON}}$.)
§ Differentiable states (>4 bits)
∗ Normalised by R_{ON}.
‡ V_{write} step size dependent.
# MFSM capacitor (not FeFET)

Table 5.1: Performance of the first FeFET generation and corresponding targets.

is a consequence from the oxidation of an extremely thin W layer that is probably already non-uniform to start with. Changing the WO_x growth method could help decrease the device-to-device variability. Moreover, the device sizes are in the µm-range because we used optical lithography for fast process development and turn-around times. Replacing optical with e-beam lithography will enable sub-µm devices. Last, controlling the channel thickness and the carrier concentration of WO_x might allow to increase the dynamic range.

5.2 SECOND FEFET GENERATION: SUB-µm-SIZED DEVICES

After the successful demonstration of a multi-state FeFET in the µm-range, we adapted the design and fabrication process for a second FeFET generation. The main differences are: We used e-beam lithography to fabricate sub-µm FeFETs. Thinner and conformal WO_x layers were deposited by ALD (see Section 4.2.2). Every device has its own gate contact which can be accessed from the top (no shared gate through the substrate), which enables crossbar array configurations.

5.2.1 Device Design

In this section we describe the design of the second FeFET generation and give an overview of the chip-layout. For the same reason as in the first generation, we use a bottom gate approach to allow for a crystallisation of the HZO before the WO_x is deposited. Figure 5.14a shows a schematic of a cross section along the channel length, Figure 5.14b along the channel width. On the bottom we have two metal lines (M1, M2) that are used to contact the 10 nm thick TiN Gate (G) contact. M1 is only required for FeFETs in a crossbar array. Above the gate contact, 10 nm of HZO covers the entire chip. Then, the 4 nm-thick WO_x channel is located above the gate stack directly in contact with the HZO. The Source (S) and Drain (D) contacts on both sides of the channel are made of Pt/W. In our design, G fully overlaps with S and D to avoid large topographic steps beneath the channel. Two metal lines (M3, M4) are used to route the S and D contacts to large contact pads for electrical measurements. M4 is required for FeFETs in a crossbar array where the S and D lines are crossing. SiO_2 is utilised as a passivation layer between metal lines and to encapsulate the WO_x channel.

Figure 5.14c shows the GDS layout of a FeFET, indicating the two cross sections displayed in Figure 5.14a and 5.14b. The channel length (L_{ch}) is defined as the distance between the S and D contacts, while the channel width (W_{ch}) is determined by the structuring of the WO_x layer. The blue box shows the extend of the WO_x layer, the small teal squares at each side are the local S and D contacts (Pt/W), the boxes with a cross indicate the openings through the passivation, and the large dark structures extending to the left, right, and top are the S, D, and G contact pads for electrical mea-

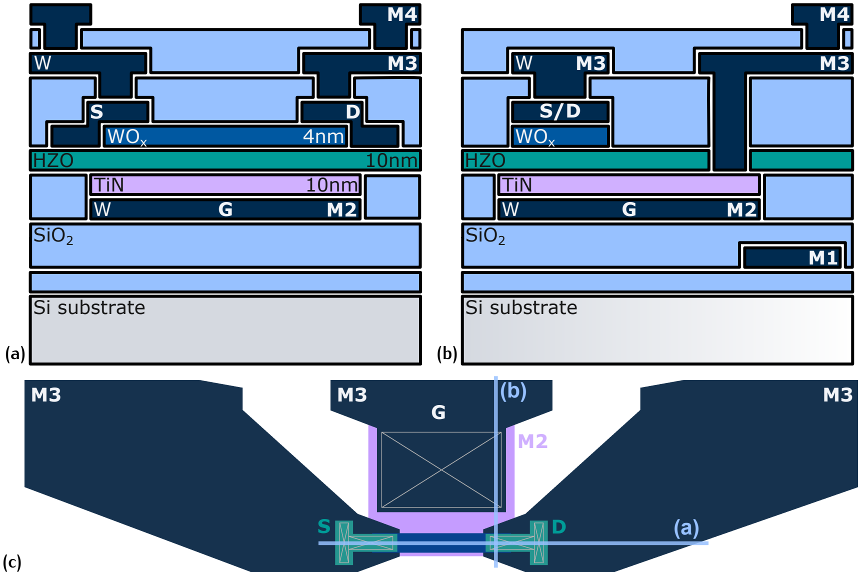


Figure 5.14: Second FeFET generation design: **(a)** Schematic of the cross section along the channel length and **(b)** along the channel width. **(c)** Top view of a FeFET where the two cross sections shown in (a) and (b) are indicated by blue lines.

surements, respectively. The purple square is the M2 gate layer extending beneath the FeFET.

FeFETs with long ($L_{ch} \gg W_{ch}$) to wide ($L_{ch} \ll W_{ch}$) channel geometries were placed in each block (purple box in Figure 5.15a). Both the channel length and width vary from 300 nm to 2 μm. For each geometry, 5 identical devices were created in each block. Apart from single devices, we designed three terminal (3T) crossbar arrays with diagonal gate lines (rose box). Crossbars ranging from 1×1 to 10×10 and with different device geometries were added to the design. Figure 5.14c illustrates a 3×3 3T-crossbar array, where the sources are connected by light blue horizontal lines (M4), the drains by dark blue vertical lines (M3), and the gates by purple diagonal lines (M2). Since not all diagonals connect the same number of devices and because there are more diagonals (5) than horizontal (3) or vertical (3) lines, always two of the diagonals are connected by a gray line (M1). In the case of the 3×3 array, diagonals with two devices are connected to diagonals with one device, resulting in 3 independent diagonal lines. As evidenced in Figure 5.14a, the vertical, horizontal, and diagonal lines are connected to large extended contacts that allow to measure the entire array with three needle probes and a movable stage. More details can be found in Appendix A.2. For the characterisation of the ferroelectric layer we placed a series of Metal-Semiconductor-Ferroelectric-Metal (MFSM) capacitors next to the FeFETs with a changing area from 3 μm × 3 μm to 60 μm × 60 μm (dark blue box). For the characterisation of the WO_x channel we added Circular Transmission Line Measurement (CTLM) structures (light blue box). Figure 5.15a shows the GDS design of the aforementioned devices that were grouped together to form a block. The chip has a size of 2 cm × 2 cm and hence, 6 of these blocks were placed on the chip to maximise the space usage, as can be seen in Figure 5.15b. Additionally to the blocks, we placed long metal lines for the characterisation of the W resistivity on the far left and Hall structures for the WO_x were positioned on the far right.

In summary, we have a design with 10 e-beam lithography steps to realise 240 FeFETs and 15 3T-crossbar arrays per block. This results in 1440 devices and 90 3T-crossbar arrays on the full chip. Without the extra metal lines required for the crossbar arrays, 6 e-beam lithography steps would be sufficient.

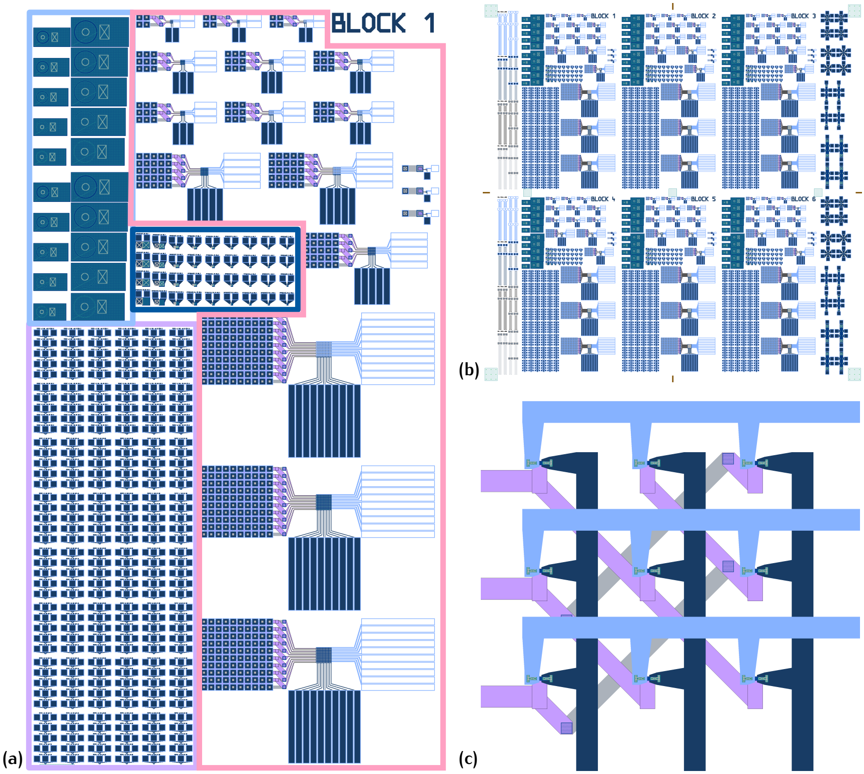


Figure 5.15: Chip design: **(a)** Overview of a block containing FeFETs, FeFET-arrays, capacitors, and CTLM structures. **(b)** Overview of the chip showing multiple blocks. Long metal lines to characterise the W resistivity were placed on the left of the 6 blocks. Hall structures for the WO_x were placed on their right. **(c)** Close-up of a 3 × 3 cross-bar array with diagonal gate lines (purple), horizontal source lines (light blue), and vertical drain lines (dark blue).

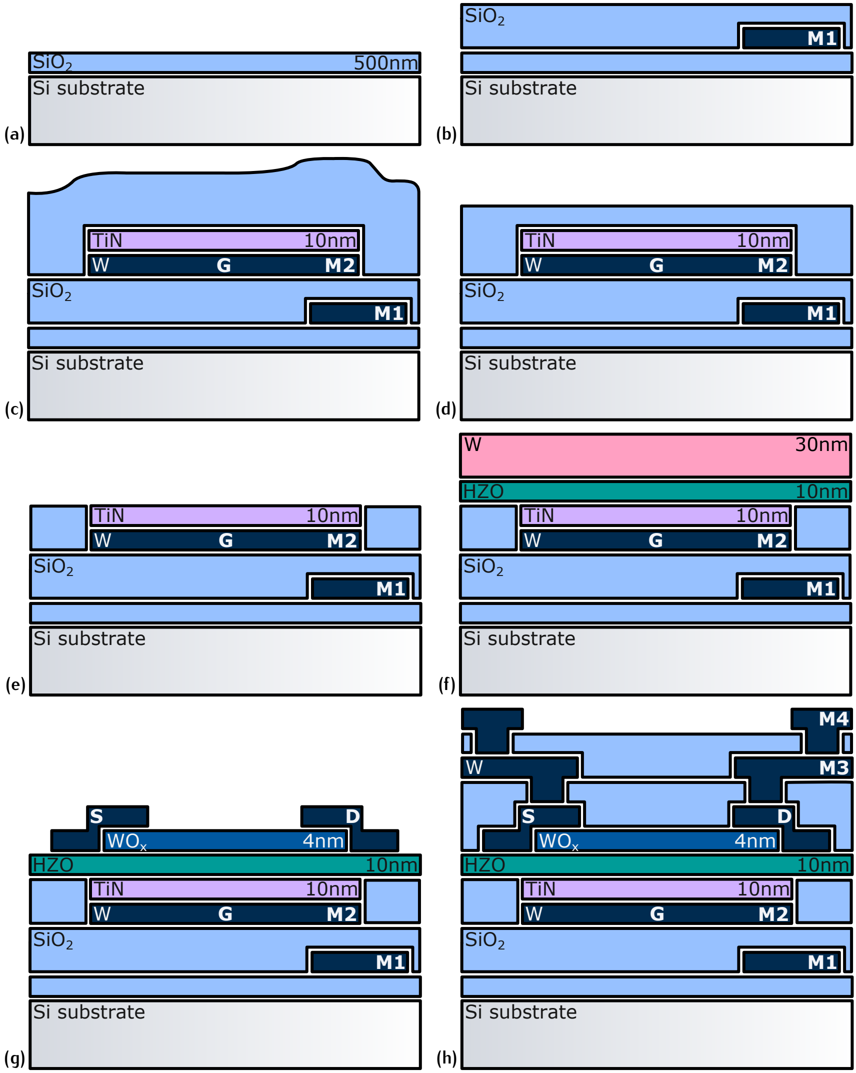


Figure 5.16: Process flow of the µm-sized FeFETs: **(a)** Starting with an oxidised Si substrate, **(b)** the contact M1 (100 nm W) and a passivation layer (100 nm SiO_2) are deposited. **(c)** The contact M2 (100 nm W/10 nm TiN) and a second passivation layer (600 nm SiO_2) are then deposited. **(d)** After CMP, the topography is removed. **(e)** Precise stopping on the TiN by RIE. **(f)** 10 nm of HZO and 30 nm of W are grown. **(g)** The sacrificial W layer is removed, WO_x is grown and structured, and S,D contacts are deposited by a lift-off process. **(h)** Final structure.

5.2.2 Device Fabrication

The starting point for the fabrication of the sub-µm FeFETs is a 4-inch thermally oxidised Si substrate (500 nm SiO_2). We used a positive (AR-P 672) or negative resist (AR-N 7520) for the e-beam lithography steps. AR-P was used if most of the wafer or chip had to be protected (exposed areas dissolve, e.g. etching small vias in the passivation) and AR-N was used when most of the wafer or chip needed to be etched (the exposed areas remain, e.g. metal lines). As e-beam lithography is a sequential process where every structure is exposed one after another, the exposure time needs to be optimised by using the right resist. The following paragraphs explain the full fabrication process:

1. **Markers and M1:** E-beam lithography markers were fabricated in the same step as the metal level M1 by first growing 100 nm of W by Physical Vapour Deposition (PVD). After exposing the AR-N resist, the W was structured with an SF_6 plasma using a Reactive Ion Etcher (RIE). Above M1 we deposited 5 nm of Al_2O_3 at 300 °C by Plasma Enhanced Atomic Layer Deposition (PEALD) using TMA[5] as a precursor (etch-stop layer) and 100 nm of SiO_2 by Plasma-Enhanced Chemical Vapor Deposition (PECVD), as depicted in Figure 5.16b. Then, AR-P was used to etch vias to access M1 by applying an RIE with a CHF_3/O_2 plasma. The Al_2O_3 etch-stop layer was removed by a wet etch in an undiluted AZ-726-MIF developer (2.38 % TMAH, etch rate: $\sim 2\,\mathrm{nm\,min^{-1}}$).

2. **M2:** 100 nm of W by PVD and 10 nm of TiN by PEALD were deposited using a TDMAT[6] precursor and N_2/H_2 plasma). Using AR-N resist, the W and TiN were structured with an SF_6 plasma in an RIE.

3. **CMP:** To avoid large topographies from M1 and M2, 5 nm of Al_2O_3 (etch-stop layer) and 600 nm of SiO_2 were deposited on top of M2 and then removed by Chemical Mechanical Polishing (CMP) on the entire wafer until only a thin layer of SiO_2 was left above M2. Figure 5.16c shows a schematic before and Figure 5.16d after the CMP.

4. **Dicing:** At this point, the 4-inch wafer was diced into 12 2 cm × 2 cm chips. The following steps were carried out at the chip level.

---

[5] TMA: trimethylaluminum $(CH_3)_3Al$

[6] TDMAT: tetrakis-(dimethylamino)titanium $[(CH_3)_2N]_4Ti$

5. **Gate:** The remaining SiO_2 above M2 was precisely etched on the entire chip until the surface height of the SiO_2 layer was equal to the surface height of the TiN as depicted in Figure 5.16e. A slight overetch was performed as the remaining SiO_2 on the TiN surface would be detrimental to the device performance. To determine the remaining height of the SiO_2, cross sections from a Focused Ion Beam (FIB) were used.

6. **HZO:** Next, the gate stack was grown in the PEALD: an approximately 10 nm-thick layer of HZO was created by using alternating cycles of TEMAH[7], and ZrCMMM[8] at 300 °C. It was immediately capped by a $\sim$30 nm sacrificial W layer by PVD. W is only used for the crystallisation of HZO and is preferred over TiN as capping layer due to its easier removal. The resulting stack is depicted in Figure 5.16f. For the crystallisation of HZO, a millisecond Flash Lamp Anneal (ms-FLA) [31] with a background temperature of 375 °C was performed.

7. **WOx:** The W capping layer was removed by a 2 min wet chemical etch in H_2O_2 at 50 °C. Immediately after, 4 nm of WO_x was deposited using a BTBMW[9] precursor and an oxygen plasma at 375 °C in a PEALD system. To fully oxidise the WO_x, it was annealed in a rapid thermal annealer (RTA) at 350 °C with 50 sccm O_2 for 6 min. After the structuring of the WO_x by using AR-N and an SF_6 plasma in the RIE, S and D contacts (Pt (20 nm)/W (20 nm)) were deposited on top of the WO_x channel by a lift-off process, as illustrated in Figure 5.16g.

8. **Device capping:** On top of the device we grew 5 nm of Al_2O_3 at 250 °C in the PEALD with an extra long plasma time of 5 s to avoid a significant reduction of the WO_x by the Al_2O_3. We then deposited 100 nm of SiO_2 by PVD to keep the process temperature low and avoid a reduction of the WO_x.

9. **Device access:** AR-P was used to etch vias in the SiO_2 and access S and D through an RIE. The Al_2O_3 etch-stop layer was removed by an undiluted AZ-726-MIF developer. Then, another AR-P layer was required to etch through the HZO and access the G. This was done in an Inductive Coupled Plasma RIE (ICP-RIE) system by using a CF_4 plasma.

---

[7] TEMAH: tetrakis-(ethylmethylamino)-hafnium $[(CH_3)(C_2H_5)N]_4Hf$

[8] ZrCMMM: bis(methyl-η5-cyclopentadienyl)methoxymethylzirconium $(CH_3C_5H_4)2Zr(OCH_3)CH_3$

[9] BTBMW: bis(tert-butylimino)bis(dimethylamino)tungsten(VI) $[(CH_3)_3CN]_2W[N(CH_3)_2]_2$

10. **M3 and M4:** The last two metal levels, M3 and M4, were each realised by growing 100 nm of W by PVD. In between, 5 nm of Al_2O_3 (etch-stop layer, PEALD) and 200 nm of SiO_2 were grown by PVD and vias between M3 and M4 were etched in the RIE. The final cross section is depicted in Figure 5.16h.

In the next paragraphs we describe the structural characterisation we performed during and after the processing:

CMP: At first, we did not include CMP in our process. With 10 e-beam lithography steps, we realised that the topography introduced by the first few layers became a problem, especially for M3 and M4 that displayed delamination close to the vias. Figure 5.17a and 5.17b show Scanning Electron Microscope (SEM) images of the S, D, and G area of a FeFET on a chip without CMP and one with CMP, respectively. The purple boxes indicate areas with large topography originating from M1 and M2. The delamination was probably also facilitated by the stress present in thin W layers. Figure 5.17c shows a cross-sectional SEM image that was taken after exposing the device cross-section with a Focused Ion Beam (FIB) system. After CMP and RIE etching, the SiO_2 passivation is on the same height as the M2 contact, demonstrating that the topography was successfully removed
.

HZO AND WO_x: After the deposition of the WO_x layer by PEALD and the subsequent oxidation in the RTA, we performed Grazing Incident X-Ray Diffraction (GIXRD) analysis (Figure 5.17d). HZO peak positions were calculated from the lattice parameters of the m- [175], t- [175], and f-phase [173]. The data for the tetragonal $P\text{-}42_1m$ phase of WO_x was taken from ICSD-86144 [267]. No monoclinic HZO phase was detected, owing to the low temperature ms-FLA. The peaks around $\sim$30.5° and $\sim$35.8° confirm that the HZO crystallised in the f- or t-phase, as expected. In contrast to the first FeFET generation with thicker WO_x, the 4 nm ALD WO_x did not crystallise under the same conditions. For comparison, we added the diffractogram of a thicker crystallised WO_x layer. As already discussed in Section 4.2.2, thin films require an increased thermal budged for crystallisation. For 4 nm of WO_x it was found to be 425 °C. To remain BEOL-compatible, we left the WO_x amorphous. No reduction treatment was applied to the WO_x as the reduction by the encapsulating passivation layers was found to be sufficient (Section 4.2.2.2).

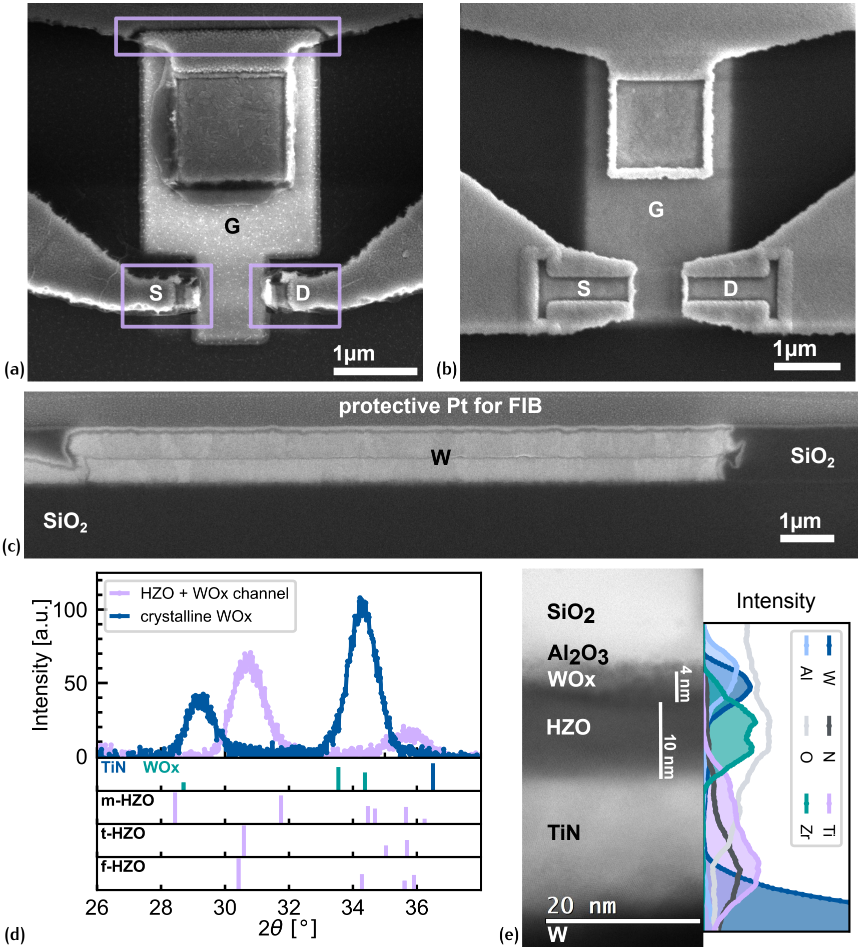


Figure 5.17: Structural characterisation of the sub-μm-sized FeFETs: **(a)** SEM image showing the S, D, and G area with a delamination of M3. The purple boxes indicate areas with large topography. **(b)** Similar area as in (a), but on a chip where we performed CMP. **(c)** Cross-sectional SEM image taken in the FIB system, proving the removal of the topography from M1 and M2. **(d)** GIXRD for a diffraction angle (2θ) from 26° to 38° showing the presence of the f- or t-phase in HZO after ms-FLA crystallisation and deposition of the amorphous WO_x channel (purple). Crystalline WO_x for comparison (blue). **(e)** Cross-sectional BF-STEM image with EDS line profiles of the $SiO_2/Al_2O_3/WO_x$/HZO/TiN/Si gate region.

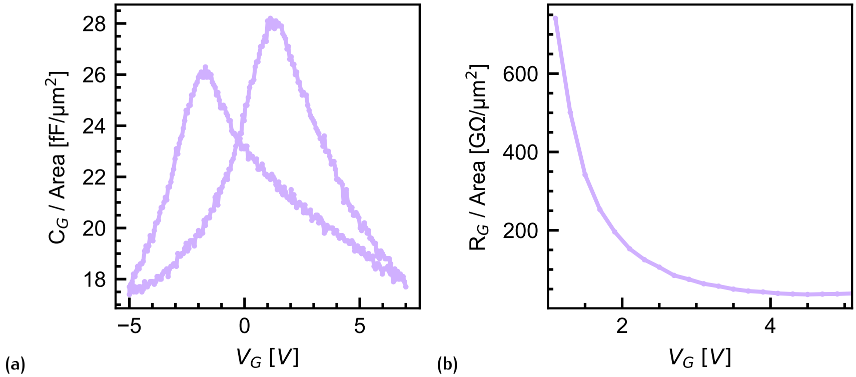


Figure 5.18: **(a)** Gate capacitance (C_G) and **(b)** gate resistance (R_G) of FeFETs normalised by the gate area (W_{ch} of 2 µm and varying L_{ch} from 300 nm to 2 µm) and averaged over 30 devices as a function of the gate voltage (V_G). Due to the overlap of S, D, and G, the gate area is slightly larger than the channel area ($L_{ch} \times W_{ch}$).

INTERMIXING Bright-Field Scanning Transmission Electron Microscopy (BF-STEM) analysis displays the expected layer thicknesses and an amorphous WO_x is confirmed (Figure 5.17e). The rough topography originating from the columnar growth of the 200 nm W below the TiN gate (Figure 5.17c) results in a projection effect (overlapping of measured elements) at layer boundaries, as can be seen in the Energy-Dispersive X-ray Spectroscopy (EDS) line profile in Figure 5.17e. Although the projection effect creates some uncertainty in the interpretation of the EDS data, two observations can be made: First, at the HZO/TiN interface a TiON layer was formed, possibly during the HZO growth by ALD. Second, Al diffusion into the WO_x is measured. It is more pronounced than the artificial mixing by the projection effect. This is an interesting feature as Al is expected to promote the reduction of WO_x [301].

5.2.3 Device Results

5.2.3.1 *Gate Stack Characterisation*

GATE CAPACITANCE AND RESISTANCE First, the gate capacitance was measured across 30 FeFETs of different sizes (long and wide channels) and normalised by the gate area: A typical butterfly shaped capacitance

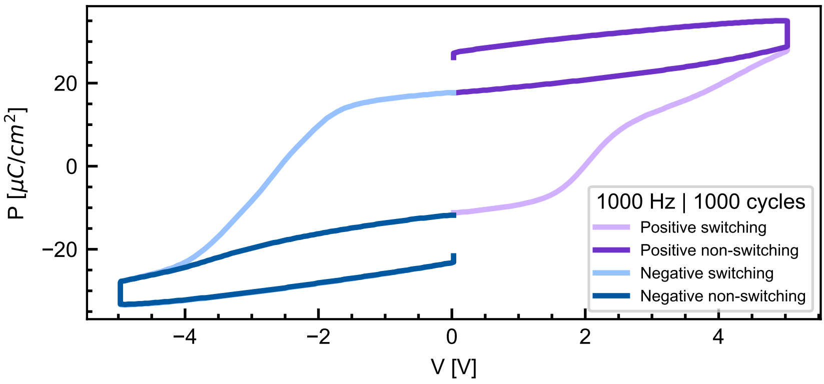


Figure 5.19: PUND measurements of a MFSM capacitor with a 40 µm × 40 µm area.

dependence on the gate voltage (V_G) was observed with $C_G = 24\,\mathrm{fF}/\mu\mathrm{m}^2$ at $V_G = 0\,\mathrm{V}$ (Figure 5.18a). This value is in good agreement with what we measured on the first FeFET generation. The gate resistance (R_G) was measured on the same 30 FeFETs by applying a V_G from 1 V to 5 V, while grounding source and drain (Figure 5.18b). At 1 V the gate resistance was 740 GΩ/µm², while at 5 V it decreased to 39 GΩ/µm². Both values are high impedance and allow for low power writing. In addition, decreasing the device size further reduces the power dissipation.

GATE POLARISATION The ferroelectric properties of the HZO layer were analysed by measuring the Positive-Up Negative-Down (PUND) characteristics of a 40 µm × 40 µm TiN/HZO/WO_x /W (MFSM) capacitor on the same sample as the FeFETs (Figure 5.19), resulting in a positive (negative) remanent polarisation $P_{r+} = 17.7\,\mu\mathrm{C/cm}^2$ ($P_{r-} = 11.2\,\mu\mathrm{C/cm}^2$) and a positive (negative) coercive field $V_{c+} = 2\,\mathrm{V}$ ($V_{c-} = -2.57\,\mathrm{V}$). The asymmetric coercive field is due to asymmetric electrode work-functions (WO_3: 6.8 eV [302], $W_{18}O_{49}$: 6.4 eV [302], TiN: 4.55 eV [303]) that result in a build-in field. The reduced negative remanent polarisation ($P_{r-} = 11.2\,\mu\mathrm{C/cm}^2$) as compared to the positive one ($P_{r+} = 17.7\,\mu\mathrm{C/cm}^2$) is an indication of partially switched domains due to incomplete screening by the depleted WO_x layer and the thereby resulting depolarisation field [304] across HZO. FeFETs with low hole-density channels show partial switching because of incomplete screening (charge balance requirements [305, 306]).

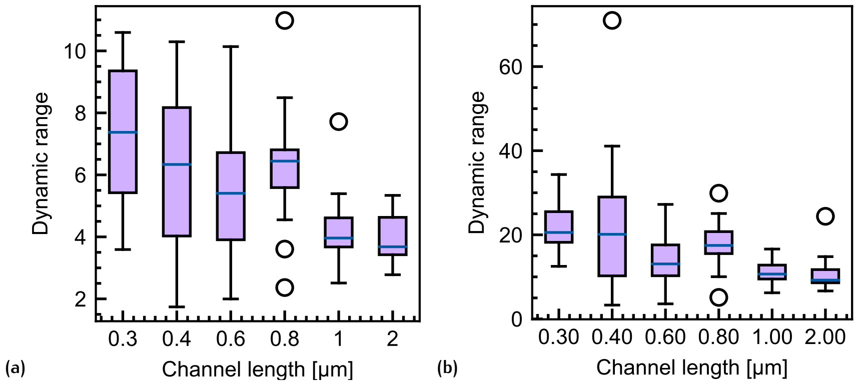


Figure 5.20: Channel length effect: Dynamic range of 120 devices with $W_{ch} = 600\,\text{nm}$ and a L_{ch} varying between 300 nm and 2 µm. Write pulses with V_w from $-6\,\text{V}$ to 6 V and t_w of **(a)** 500 µs and **(b)** 500 ms were applied. The boxes extend from the lower to the upper quartile values of the data, with a line at the median. The whiskers extending from the box show the data range. Flier points are those past the end of the whiskers.

5.2.3.2 *Switching Locations and Time Scales*

SWITCHING LOCATIONS Potentiation (depression) cycles on 120 devices with a constant channel width $W_{ch} = 600\,\text{nm}$ and a channel length varying from $L_{ch} = 300\,\text{nm}$ to 2 µm were measured by applying write pulses with an amplitude (V_w) up to 6 V (-6 V). The same measurement was repeated for a pulse width (t_w) of 500 µs and 500 ms. Three cycles with a step size $Vstep = 200\,\text{mV}$ were measured on each device to extract R_{SD} when the device is in its High Resistive State (HRS) and Low Resistive State (LRS). The Dynamic Range ($DR = HRS/LRS$) as a function of L_{ch} is presented as box plots in Figure 5.20a for 500 µs and in Figure 5.20b for 500 ms. Clearly the DR increases for shorter channels, independently from t_w. The above mentioned incomplete screening of the FeFETs with a low channel hole-density could be responsible for a less effective source-drain resistance (R_{SD}) modulation in the centre of long channels. The availability of holes decreases with distance from the contacts and thus, less negative charges can be screened (reduced P_{r-}) in the centre of the device. If this is the case, we should see a channel resistance (R_{Ch}) which does not scale with the geometry of the channel.

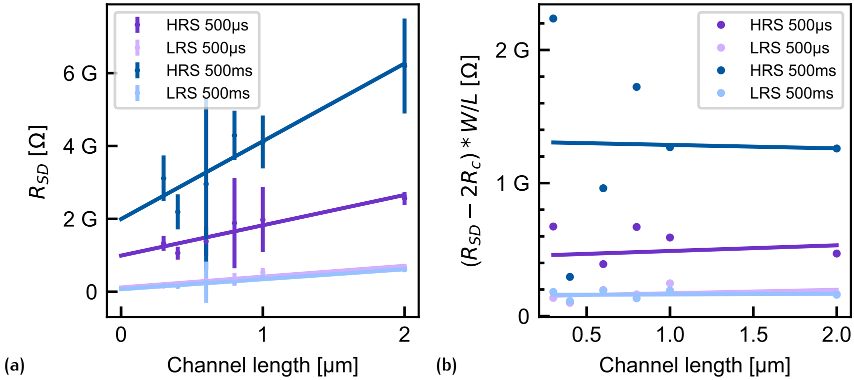


Figure 5.21: **(a)** LRS and HRS as a function of the FeFET channel length. The contact resistance can be extracted at the y-intercept. The error bars denote the standard deviation. **(b)** Sheet resistance of the *LRS* and *HRS* obtained by subtracting the $2R_c$ from R_{SD} and then normalising the result by the channel dimensions (W_{ch}/L_{ch}).

t_w	$2R_c$ (LRS)	$2R_c$ (HRS)	DR_{R_c}	ρ_{Ch} (LRS)	ρ_{Ch} (HRS)	$DR_{R_{Ch}}$
500 μs	118 MΩ	1000 MΩ	8.4	71 Ω cm	200 Ω cm	2.8
500 ms	69 MΩ	2000 MΩ	29	65 Ω cm	512 Ω cm	7.4
Factor ms/μs	0.58	2	3.45	0.92	2.56	2.64

Table 5.2: Contact resistance (R_c) and channel resistivity (ρ_{Ch}) extracted from Figure 5.21a.

We can now define the total source-to-drain resistance as:

$$R_{SD} = R_{Ch} + 2R_c$$
$$R_{Ch} = \frac{\rho_{Ch}}{t} \cdot \frac{L_{ch}}{W_{ch}},$$

where R_c is the contact resistance, and ρ_{Ch} the resistivity of the WO_x channel. Plotting the *LRS* and *HRS* as a function of the channel length (Figure 5.21a) is similar to Transmission Line Measurements (TLM) and allows to extract the contact resistance ($2R_c$) at the y-intercept and ρ_{Ch} from the slope of the linear regression. The extracted values for $t_w = 500\,\mu s$ and $t_w = 500\,ms$ are summarised in Table 5.2. Multiple observations can be made:

- First, the channel resistance R_{Ch} and the contact resistance R_c are modulated to a different extend by the write pulses. While R_c displays a modulation of $DR_{R_c} = 8.4$, the dynamic range of the channel $DR_{R_{Ch}} = 2.8$ is three times lower. Figure 5.21b shows the sheet resistance R_{sh} of the *LRS* and *HRS* for $t_w = 500\,\mu s$ and $t_w = 500\,ms$ as a function of the channel length:

  $$R_{sh} = (R_{SD} - 2R_c)\frac{W_{ch}}{L_{ch}} = R_{Ch}\frac{W_{ch}}{L_{ch}} = \frac{\rho_{Ch}}{t}.$$

  The fact that the linear regression has almost no slope shows that R_{Ch} scales with the geometry of the channel and that there is very little dependence of the $DR_{R_{Ch}}$ on the channel length. Consequently, an incomplete screening of the polarisation in the centre of long channels seems unlikely. Instead, the total *DR* in Figures 5.20a and 5.20b that includes both contributions (DR_{R_c} and $DR_{R_{Ch}}$) decreases for longer channels as the contribution of $DR_{R_{Ch}}$ increases. The same trends were observed when increasing t_w from 500 µs to 500 ms.

- Second, an overall increase of the total *DR* by a factor of ∼2.8 was observed by increasing t_w to 500 ms. This is consistent with the scenario of an oxygen migration, which would be stronger near the contacts were the field is larger (DR_{R_c} increases more than $DR_{R_{Ch}}$). This is a first indication of a successful exploitation of oxygen migrations in addition to the polarisation switching to modulate R_{SD}.

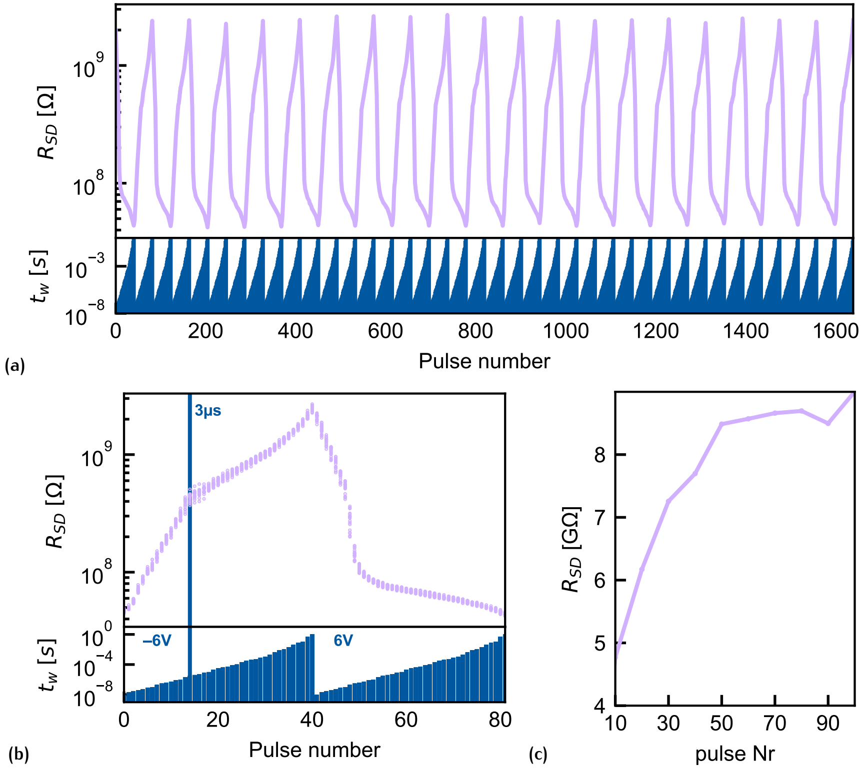


Figure 5.22: Potentiation and depression cycles with a constant pulse amplitude scheme on a FeFET with $L_{ch} = 400\,\text{nm}$ and $W_{ch} = 1\,\mu\text{m}$ by modulating t_w from 10 ns to 1 s while keeping V_w constant at 6 V for the depression and −6 V for the potentiation. **(a)** Channel resistance as a function of the pulse number for 20 consecutive cycles. On the bottom, the corresponding duration t_w of each pulse is provided. **(b)** Superposition of the 20 cycles to visualise the cycle-to-cycle variability and the potentiation and depression shape. **(c)** Cumulative increase of R_{SD} by repeatedly applying a square pulse with $V_w = 6\,\text{V}$ and $t_w = 100\,\text{ms}$.

SWITCHING TIME SCALES The influence of t_w on the DR was further analysed by performing constant amplitude potentiation ($V_w = 6\,\text{V}$) and depression ($V_w = -6\,\text{V}$) cycles with changing t_w from 10 ns to 1 s (Figure 5.22a). By exploiting such wide pulses, a dynamic range of almost 60 was reached (average of 20 cycles), a considerable increase as compared to the state-of-the-art oxide-channel FeFETs [38, 41, 157]. Figure 5.22b is a superposition of the 20 cycles. It confirms the strong dependence of R_{SD} on t_w: there are two different regimes for the potentiation and depression. For short pulses ($t_w < 3\,\mu\text{s}$, $V_w = -6\,\text{V}$) a steep depression is observed, followed by a less steep change ($t_w > 3\,\mu\text{s}$, $V_w = -6\,\text{V}$) that does not saturate up to the maximum t_w of 1 s. In the long pulse regime, cumulative switching is observed: repeating the same pulse (e.g. $t_w = 100\,\text{ms}$, $V_w = -6\,\text{V}$) 100 times increased the channel resistance after the first 10 cycles still by a factor 2 (Figure 5.22c). The large change by the first 10 cycles is not shown ($R_{SD} = {\sim}60\,\text{M}\Omega$ at 0 cycles). The absence of a saturation and a change of slope indicates that the resistance modulation for $t_w > 3\,\mu\text{s}$ is based on an additional and slower physical process with respect to the previous regime ($t_w < 3\,\mu\text{s}$): the first regime ($t_w < 3\,\mu\text{s}$) is attributed to the ferroelectric switching as it is known to occur at very fast time scales [307]. In the second regime ($t_w > 3\,\mu\text{s}$), the energy of each pulse was potentially large enough to additionally enable oxygen migration [281] between HZO and WO_x. The oxidation or reduction of the WO_x channel in turn adds to the modulation of R_{SD}.

SWITCHING EFFECTS The resistance modulation of the two regimes can be better understood by examining their underlying conduction mechanisms. Several electrode-limited and bulk-limited conduction mechanisms depend on temperature in different ways [308]. Temperature-dependent $I_D{-}V_{DS}$ measurements of the channel were performed (20 °C to 60 °C) after programming the device in the LRS ($V_w = 6\,\text{V}$) and in the HRS ($V_w = -6\,\text{V}$). Write pulses of $t_w = 500\,\mu\text{s}$ (where the ferroelectric effect is saturated) and longer write pulse trains of $90 \times t_w = 100\,\text{ms}$ (to enhance the oxygen migration) were applied. No difference in the conduction mechanism between the two time scales was observed. Both the LRS and HRS show a linear $I_D{-}V_{DS}$ characteristic and are best fitted with the Ohmic conduction model [309]. An exact description of the fitting and extraction of the parameters can be found in Appendix A.3.

Comparing the dynamic range as a function of temperature for the fast and slow time scale (Figure 5.23a and 5.23b) on the other hand displays

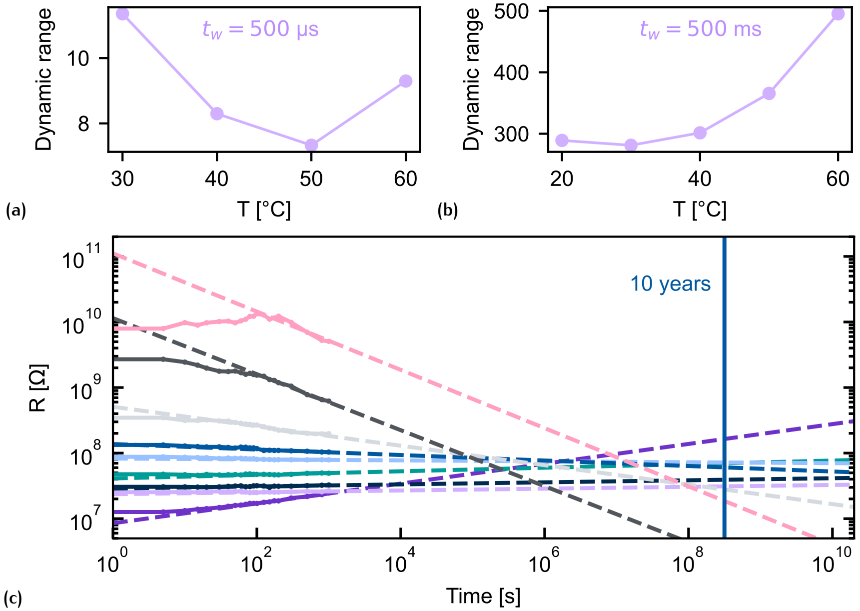


Figure 5.23: Temperature dependent measurements: **(a)** for $t_w = 500\,\mu\text{s}$ and **(b)** for longer write pulse trains of $90 \times t_w = 100\,\text{ms}$. **(c)** Retention measurements at elevated temperature (85 °C) for long set pulses $t_w = 500\,\text{ms}$. The solid lines are acquired experimental data and the dashed lines are linear fits in the log-log scale to predict the evolution of the states.

a clear difference. For the fast ferroelectric dominated time scale, the dynamic range decreases with increasing temperature. This effect could originate from a phase transition from the orthorhombic ferroelectric phase of HZO to its tetragonal, anti-ferroelectric phase at elevated temperatures, as observed in ZrO_2 [173] and Si:HfO_2 [310]. For the slower time scale, a large increase of the dynamic range with temperature is observed. This is consistent with oxidation and reduction processes that are facilitated at elevated temperatures by the higher mobility of the oxygen ions. This further demonstrates the dual effect that modulates our channel resistance.

Retention measurements at elevated temperatures (85 °C) and with long pulses ($t_w = 500\,\mu s$) display a resistance change of almost 3 orders of magnitude (Figure 5.23c). The larger the resistance change obtained within this retention study, the faster (hours to minutes) it drifted back to resistance values measured without heating and with shorter pulses. Temperature-dependent current measurements revealed that oxygen exchange between the WO_x and HZO layers is a valid assumption. The same interpretation holds for the retention results at 85 °C: The higher mobility of oxygen ions at elevated temperatures allows to reach higher resistance values while at the same time increasing the relaxation rate of the resistance through oxygen diffusion. This relaxation observed at elevated temperatures also highlights the trade-off between the dynamic range and retention. While the modulation by the ferroelectric is stable and well-suited for long term memory, oxygen gradients introduced by longer pulses eventually relax back to a steady-state through ion diffusion. The more consecutive pulses are applied, the larger the resistance change and the longer the relaxation back to a stable long-term state, are a characteristic that can be used to implement the spontaneous decay of the conductance for short-term plasticity. We also believe that voltage-based STDP models [76, 311, 312] could be well-suited for such artificial synapses.

5.2.3.3 *Characteristics for ANNs*

POTENTIATION AND DEPRESSION VARIABILITY In contrast to SNNs, where a tunable plasticity is desired, the efficient operation of ANNs on analog memristive crossbar arrays requires artificial synapses that behave as long-term memory with long data retention and display low variability. Based on the better performance of short-channel devices, as can be seen in Figure 5.20, a more thorough investigation of the potentiation and depression characteristics for in-memory computing with analog crossbar arrays was conducted on a FeFET with $L_{ch} = 300\,\text{nm}$. Figure 5.24a shows

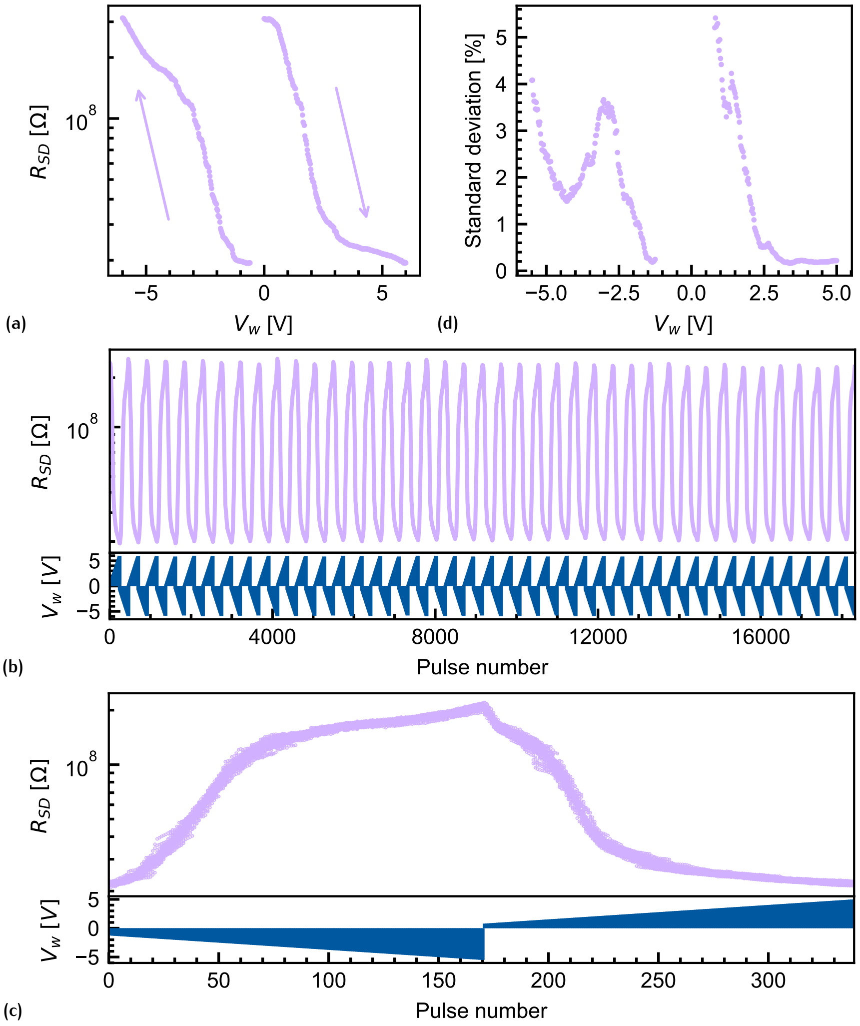


Figure 5.24: Potentiation and depression of a FeFET with $L_{ch} = 300\,\text{nm}$ and $W_{ch} = 2\,\mu\text{m}$: **(a)** One potentiation (0 V to 6 V) and depression (−0.6 V to −6 V) cycle with a constant $t_w = 500\,\mu\text{s}$ showing quasi continuous channel resistance states (dep: 241 states, pot: 217). **(b)** 40 sub-range potentiation (0.6 V to 5 V, 171 levels) and depression (−1.25 V to −5.5 V, 169 levels) cycles. R_{SD} as a function of the pulse number (top) and the corresponding V_w (bottom) are reported. **(c)** Superposition of all 40 cycles from (c) to visualise the cycle-to-cycle variability. **(d)** Standard deviation of R_{SD} normalised by the resistance window $(HRS - LRS)$.

a single potentiation and depression cycle of the R_{SD}. Potentiation from the HRS to the LRS was performed by increasing V_w from 0 V to 6 V in 25 mV steps and the depression back to the HRS by decreasing V_w from −0.6 V to −6 V in 25 mV steps. The pulse duration t_w was kept constant at 500 µs. The 241 (217) steps for the potentiation (depression) exhibit a quasi-continuous resistance range with a monotonic decrease (increase) of the resistance and a dynamic range of 16. The cycle-to-cycle variability was analysed by performing 40 sub-range cycles of depression (0.6 V to 5 V) and potentiation (−1.25 V to −5.5 V) that are shown in Figure 5.24b. The corresponding V_w of each pulse is given at the bottom. Figure 5.24c is a superposition of all 40 cycles to help visualise the Cycle-to-Cycle Variation ($CtCV$). By reducing the V_w range, the dynamic range decreases to 10.4 on average. Figure 5.24d is a visualisation of the $CtCV$ (standard deviation) as a percentage of the channel resistance range ($HRS - LRS$). The $CtCV$ does not exceed 6 % and is around 1.9 % on average.

To study the device-to-device variations, 20 identical FeFETs were measured and the normalised standard deviation of the HRS and LRS was calculated. In both cases it is 39 %. At the same time, the dynamic range has a smaller normalised standard deviation of 28 %, reflecting the fact that for most devices the entire resistance window moved up and down: if the LRS was smaller than average, the HRS was also smaller than average for 70 % of the measured devices. The device-to-device variability can be explained by process variations across the sample. The polycrystalline nature of HZO results in different ferroelectric properties from device to device [36]. Also, the WO_x is very sensitive to the oxygen content and it usually reduced at elevated temperatures (>250 °C) by other oxides or nitrides interfacing it, such as SiO_2, Al_2O_3, or SiN. Hence, local temperature differences during processing can lead to different local reduction states of the WO_x.

LINEARITY AND SYMMETRY When fitting the potentiation range from −1.25 V to −5.5 V and depression range from 0.6 V to 5.0 V of the same 22 cycles by linear regression (Figure 5.25a), an adjusted residual-square value of 0.958 and 0.671 is obtained for the depression and potentiation, respectively. Especially the potentiation is not linear and thus poorly fitted. For a more detailed analysis of the symmetry and the Signal-to-Noise Ratio (SNR), Gaussian Process Regression (GPR) was used to predict a noise-free signal (Figure 5.25b), as proposed by Gong *et al.* [300]. Different memristor technologies require different fitting formulas, making it difficult to com-

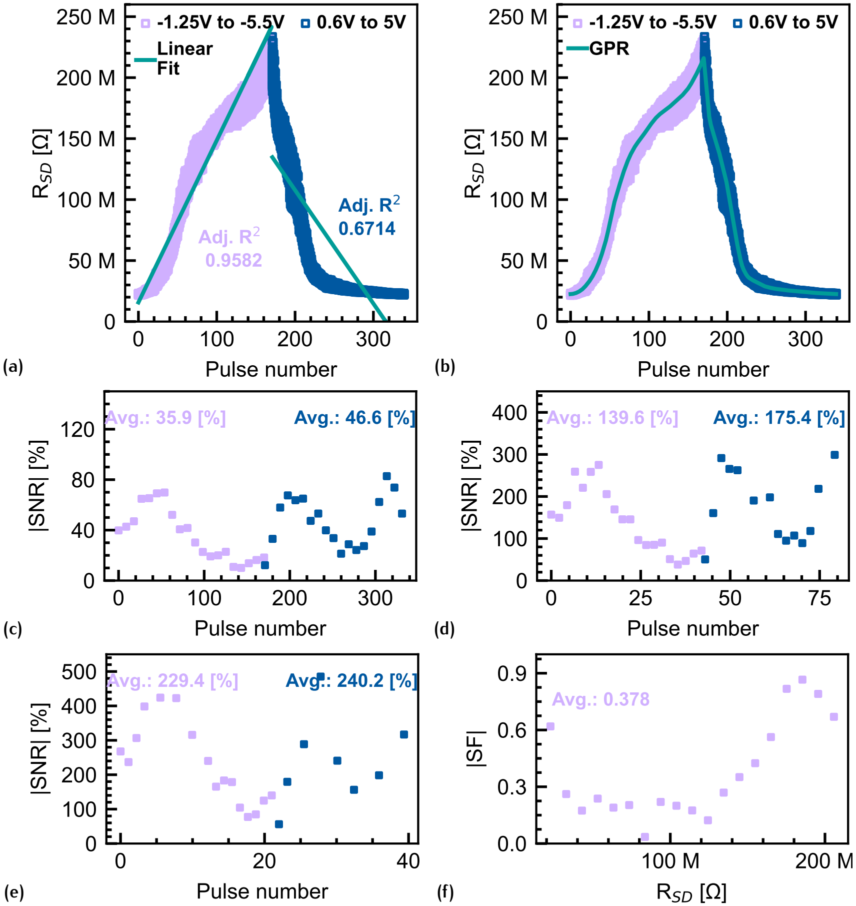


Figure 5.25: Extraction of the linearity and symmetry metrics. FeFET data from multiple cycles with 22 depression pulses (purple, 1 V to 3.1 V) and 22 potentiation pulses (blue, −0.9 V to −3 V) are shown. **(a)** Linear regression fit (teal). **(b)** GPR-predicted noise-free signal (teal). **(c, d, e)** Absolute SNR for V_{step} of 25 mV, 100 mV, and 200 mV, respectively. **(f)** Symmetry factor (SF) extracted according to Equation 5.2.3.3.

pare key performance parameters on a common ground. GPR addresses this issue by not assuming any specific functional form such as linear or exponential. The SNR is defined as follows:

$$SNR = \frac{\Delta R_{SD}}{r}, \tag{5.3}$$

where ΔR_{SD} is the change in resistance and r the residuals from the noise-free signal fit for each pulse. In our case, the noise arises from the CtCV and not from a stochastic switching. From Equation 5.3 we see that the SNR depends on ΔR_{SD} and thus on V_{step}. Figures 5.13f, 5.13e, and 5.13g show the SNR for V_{step} of 25 mV (170 steps), 100 mV (42 steps), and 200 mV (21 steps), respectively. We can change R_{SD} in very small steps by choosing a small V_{step}. Smaller ΔR_{SD} divided by r result in a smaller SNR. If we increase the step size, we get less intermediate states and the SNR increases. For the first FeFET generation, we used $V_{step} = 200\,\text{mV}$, thus the same is used here for comparison. While the first FeFET generation has a SNR of 177.2 %(dep)/237.4 %(pot), the second generation is characterised by a slightly increased SNR of 229 %(dep)/240 %(pot).

The symmetry factor (SF) was then calculated using the following equation: [300]

$$SF = \left|\frac{\Delta R_{SD+} - \Delta R_{SD-}}{\Delta R_{SD+} + \Delta R_{SD-}}\right|,$$

where ΔR_{SD+} is the depression and ΔR_{SD-} is the potentiation change in resistance at a certain resistance interval. By this definition, SF can take values between 0 and 1, where 0 is the perfect symmetry. Compared to the first generation, SF increased, especially in the parts close to the LRS and HRS that we attributed to oxygen migrations (Figure 5.13d). The average across the full resistance range is $SF = 0.378$, which is not a good performance metric with respect to symmetry. Introducing multiple resistance modulation mechanisms thus has a negative effect on the symmetry.

ENDURANCE The endurance of a FeFET with $L_{ch} = 800\,\text{nm}$ and $W_{ch} = 600\,\text{nm}$ is shown in Figure 5.26a. Cycling pulses of ±3 V and ±4 V with a frequency of 100 kHz were applied to the gate, while S and D were grounded. The HRS and LRS were measured by performing 3 cycles of potentiation and depression. When switching at ±3 V we observed a small continuous decay of the dynamic range window to about 70 % of its initial value after 10^{10} cycles, but no failure could be identified. To the best of our knowledge this is the best endurance reported on hafnia-based FeFETs. Especially compared with Si-based channel FeFETs this is a major

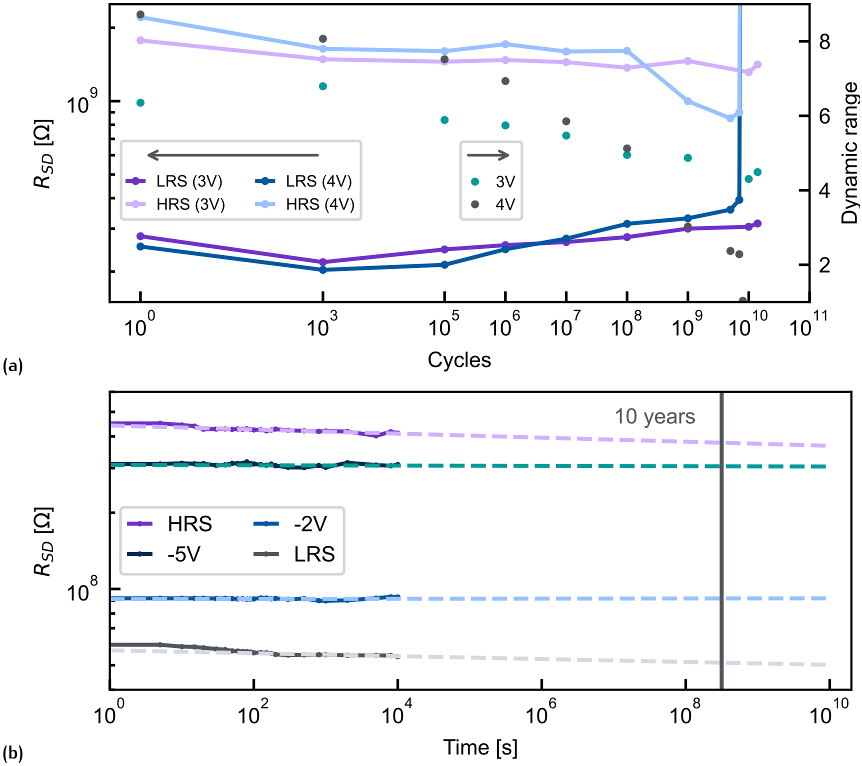


Figure 5.26: Endurance and retention: **(a)** Endurance of a FeFET with $L_{ch} = 800\,\text{nm}$ and $W_{ch} = 600\,\text{nm}$. Triangular pulses with a frequency of 100 kHz were applied up to 10^{10} cycles. The amplitude of the pulses was ±3 V and ±4 V. The evolution of the *HRS* and *LRS* (left axis) and the corresponding dynamic range (right axis) are shown. **(b)** Retention measurements at room temperature for a FeFET with $L_{ch} = 300\,\text{nm}$ and $W_{ch} = 2\,\mu\text{m}$ showing a good retention of >10 years for the four programmed states. Only the *HRS* has a small drift. The solid lines are the experimental data and the dashed lines are linear extrapolations in the log-log scale.

improvement over the typically reported 10^6 cycles [36], although lately up to 10^8 cycles were reported [313]. The increased endurance of oxide channels was attributed to the absence of an interfacial layer (with a small dielectric constant) between the channel and the ferroelectric layer as it is the case for Si channels [36, 153]. Furthermore, the deposition of the channel by atomic layer deposition produces clean interfaces with a reduced number of defects due to its conformal nature. Increasing the cycling voltage to ±4 V (on another device with the same dimensions) accelerated the fatigue and the device failed after 8×10^9 cycles. The failure took the form of an open device, which could be the result of local heating around the contacts due to a sudden high current caused by a HZO breakdown or of some other mechanisms degrading the contacts. Cycling at even higher fields was not tested as online learning happens in small changes and not by constantly switching between the extreme states.

RETENTION The longer the programming pulses, the larger the FeFET's resistance range is. At the same time this means that the crossbar arrays must be operated at slower frequencies. For inference applications where the state of the artificial synapse is not constantly changed, the large dynamic range made available by long pulses can be exploited. While the update frequency, write pulse width, and amplitude become less important, a long retention is key for inference applications. Therefore, retention measurements were conducted as follows: first a state was set by applying a write pulse of $t_w = 500\,\mu s$. Then the channel resistance was monitored ($V_{SD} = 200\,mV$) every few minutes up to 10^4 s. This was repeated for the *LRS*, two intermediate states, and the *HRS* (Figure 5.26b). By fitting a linear regression in the Log-Log scale, no drift is observed for the intermediate states. Both, the *LRS* and *HRS* display a small drift towards lower values. Extrapolating the fit to 10 years yields a change of about 11 % and 15 %, respectively. This excellent retention time confirms the advantage of using metal-oxide thin films over Si as channels as there is no back-switching of the ferroelectric domains due to charge trapping at the oxide interlayer formed between Si and the ferroelectric [153]. This low drift opens the path to inference and memory applications for the FeFET devices presented here.

ENERGY With a channel resistance between 20 MΩ and 2 GΩ (depending on the geometry) and a read voltage of 100 mV, between 5 pW and 500 pW are dissipated during a read operation ($P_{read} = V_{SD}^2 / R_{SD}$). The write operation of a single device ($t_w = 500\,\mu s$) to the high gate impedance ($39\,G\Omega/\mu m^2$ at 5 V, Figure 5.18b) has a lower energy consumption, between 1.7 fJ and 1.2 pJ for V_w of 1 V and 5 V, respectively ($E_{write} = t_w V_w^2 / R_G$).

5.2.3.4 MNIST Simulation

The performance of our FeFETs as artificial synapses in a crossbar array was investigated by using the *MLP+NeuroSimV3.0* [45] framework. The online learning accuracy of a pseudo crossbar array of 400 input, 250 hidden, and 10 output neurons trained on the MNIST [46] database was simulated by using the aforementioned values. The Non-Linearity (NL) parameters were extracted by fitting the potentiation and depression curves according to [45] (Figure 5.27a). They are 2.32 and -4.63 for the potentiation and depression respectively. Ferroelectric-based artificial synapses often exhibit a S-like potentiation and depression and hence the non-linearity factor extraction by fitting an exponential function, as required by [45], does not fully capture its characteristics. As conductance Range Variation (CRV) parameter in the simulator, the same value is usually applied to all devices. It thereby models the CRV as an increase or decrease of the dynamic range. Hence, the *MLP+NeuroSimV3.0* code was slightly adapted (Appendix A.4) to apply a random CRV to every device of the network, which takes into account our device-to-device variability. The spread of the HRS, LRS, and DR around the average (solid lines labeled "No Var") are shown in Figures 5.27b and 5.27c. Figure 5.27d reports the learning accuracy on the MNIST database. By only considering the NL parameters and the Finite Number of States ($FNoS$) as a non-ideality, an excellent performance of 92 % recognition accuracy is achieved. By further introducing the $CtCV$ to the simulation, the performance remaines as high as to 89 % and with the HRS, LRS, and DR spread included, 88 % is reached. We therefore conclude, that a device-to-device variability of 39 % can be accommodated by the network and has only minor impact on the classification accuracy.

5.2.3.5 Conclusion and Possible Improvements

We demonstrated a scaled (sub-µm) BEOL-compatible FeFET artificial synapse with an amorphous $WO_{x<3}$ channel. The device concept was engineered to leverage two controllable resistance modulation mechanisms activated on two different write pulse time scales: a fast ferroelectric field effect ($t_w < 3\,\mu s$) and an oxidation/reduction of the channel by oxygen

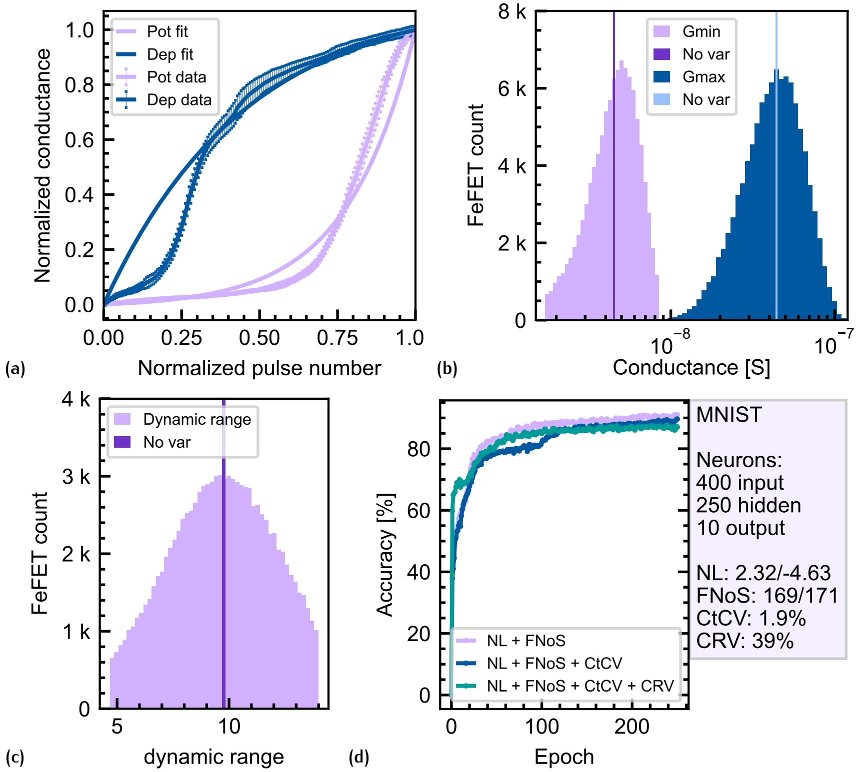


Figure 5.27: Online learning performed with *MLP+NeuroSimV3.0* [45] **(a)** Exponential fit of the potentiation and depression curves to extract the corresponding non-linearity parameters. Same data as in Figure 5.24c. **(b,c)** Histogram of the *HRS*, *LRS*, and *DR* after taking the device-to-device variability into account (see Appendix A.4). **(d)** MNIST classification performance of our FeFETs with different degrees of non-idealities included: non-linearity factors and finite number of steps (purple), + cycle to cycle variation (blue), and + conductance range variation (teal).

movement at a slower time scale ($t_w > 3\,\mu s$). Key enablers were the control of the channel oxidation state and of its thickness down to 4 nm, as well as having the channel in direct contact with the ferroelectric HZO gate without the formation of a spurious interlayer. The dual nature of the resistance modulation mechanisms was derived from the write-time-dependent dynamics and from the two different potentiation and depression slopes. The temperature-dependent current measurements showed opposite dynamic range trends for the two time scales, confirming that the resistance changes originate from two different mechanisms. Moreover, the temperature-dependent retention measurements highlighted the role of the oxygen drift across the layers in the slow regime ($t_w > 3\,\mu s$).

With this extra knob to extend the dynamic range, our scaled FeFETs have a dynamic *HRS/LRS* ratio at room temperature that is 30 times larger than the first FeFET generation. The plasticity of our synapse shows a different response over multiple timescales, making our FeFETs interesting candidates for neuromorphic engineering. The Ohmic nature and a resistance in the MΩ regime of our WO_x channel are welcomed features for precise and low-power readout operations. The extremely fine-grained, quasi-continuous monotonic resistance changes with more than 200 steps between the *LRS* and *HRS*, together with an excellent cycle-to-cycle variability led to a MNIST classification accuracy of 88 % with the *MLP+NeuroSimV3.0* framework. The endurance was extended to unprecedented $>10^{10}$ cycles and an excellent retention of >10 years was obtained with only little dynamic range loss. Therefore, our FeFET technology is not only promising for online learning, but also for in-memory computing and inference applications. Table 5.3 summarises the memristor characteristics of the second FeFET generation. For comparison the values form the first generation are also included.

We reduced the device-to-device variability by growing WO_x by ALD instead of rapid thermal oxidation. While this parameter decreases as expected, there is still room for improvement. We believe that the easily reducible WO_x could impose an intrinsic problem: on theone hand it allows to modulate the channel resistance even after fabrication, but on the other hand it could also create local resistivity gradients during processing, leading to large device-to-device variations. In a first step, a reduction of the overall processing temperature could mitigate this issue. In particular, the WO_x deposition and oxidation parameters should be further optimised for lower temperatures. Furthermore, eliminating all contacts to water would minimise the presence of interstitial hydrogen atoms and potentially decrease the device-to-device variability.

By reducing the channel thickness and increasing its resistivity we were able to considerably increase the dynamic range. Moreover, we found that oxygen migrations can be used to further modulate the channel resistance. A consequence of this is an increased write pulse voltage and time, which in turn results in more power dissipation. A reduction of the required voltage could be achieved by optimising the HZO thickness in the gate stack. A thinner layer would decrease the voltage needed for the same electric field. The trade off here is between the write amplitude, gate-leakage, and polarisation.

For an improved "third" FeFET generation, we propose to try to mitigate some of the shortcomings described above. Building on the finding of multiple timescales and physical locations where the R_{SD} modulation occur, we envision adding a 4th electrode with a non-ferroelectric gate on top of the WO_x channel to further oxidise or reduce the channel without affecting the ferroelectric state. This additional knob might lead to a volatile resistance component that can tune the switching dynamics similarly to [314]. Oxygen movements induced by this 4th electrode are expected to diffuse back quickly without the stabilising field of the ferroelectric layer. Having such an additional volatile component would permit to implement a wider range of neuromorphic engineering paradigms.

	1st FeFET gen.	2nd FeFET gen.	Target
Dynamic range	1.55 to 1.98	10 to 60$^{\$}$	8 to 100 [20]
R_{ON}	>100 kΩ	10 MΩ to 500 MΩ	>10 MΩ
Number of states	22‡, 18§	170‡	>100 [20]
Programming time	50 ns to 5 μs	10 ns to 1 s	ns
Programming voltage	3.1 V/−3 V	<6 V	<5 V
Linearity (α_d/α_p)	0.38/1.0	2.32/−4.63	0 [45]
Linear regression	0.952/0.955	0.958/0.671	1/1
SF [0 to 1] (average)	0.203	0.378	0 [300]
Symmetry ($\lvert\alpha_p - \alpha_d\rvert$)	0.62	6.85	0 [299]
Write energy	0.525 fJ	<1.2 pJ (6 V)	<10 fJ [20]
Area	≥5 × 5 μm²	<1 × 1 μm²	<10 × 10 nm²
Device-to-device var.	>100 %	39 %†	0
Cycle-to-cycle var.	∼2.9 %†, ∼0.9 %*	<6 %, avg. 1.9 %†	0
Endurance	8 × 10$^{6\#}$	>10^{10}	>10^{9} [20]
Individual gate	No	Yes	Yes

† Normalised by the resistance window ($R_{\mathrm{ON}} - R_{\mathrm{ON}}$.)
§ Differentiable states (>4 bits)
∗ Normalised by R_{ON}.
‡ V_{write} step size dependent.
# MFSM capacitor (not FeFET).
$ Depends on pulse width.

Table 5.3: Performance of the first and second FeFET generation, and corresponding targets.

6

SUMMARY AND OUTLOOK

> *Algorithmic bias, like human bias, results in unfairness. However, algorithms, like viruses, can spread bias on a massive scale, at a rapid pace.*
>
> — Joy Buolamwini

This chapter provides a summary of the thesis and gives an outlook on future work.

6.1 SUMMARY

In this thesis a Ferroelectric Field-Effect Transistor (FeFET) technology for artificial synaptic weights was developed. The effort focused first on the material optimisation of the oxide channel and the hafnia-based ferroelectric gate dielectric. Both elements were then combined to fabricate two FeFET generations with the goal to emulate the complex and multi timescale plasticity of biological synapses. Apart from pursuing ideal memristor characteristics such as gradual analog conductance changes, linearity, ohmic conduction, long endurance, small cycle-to-cycle variabilities, and low power operations, the objective was also to develop a Complementary Metal-Oxide-Semiconductor (CMOS)-compatible and Back-End-Of-Line (BEOL) friendly process. This imposed some restrictions on materials, processing steps, and maximum temperature budget. Therefore, a HfO_2 and ZrO_2 composite mixture (HZO) was selected as ferroelectric layer because both are well-established materials in the CMOS industry and because their combination has one of the lowest crystallisation temperatures of all ferroelectric hafnia composites. WOx was chosen as channel material due to its relatively mobile oxygen-ions and thereby tunable conductivity as a function of the oxygen content. Furthermore, oxide channels are expected to mitigate problems related to a presence of poor semiconductor-oxide interfaces, as encountered with Si-channels.

After presenting all tools and processes that were used in this thesis, the development of a BEOL-compatible crystallisation of HZO to its ferroelectric phase was demonstrated in Chapter 4 by applying a millisecond flash lamp anneal (ms-FLA) at 375 °C. Apart from the low temperature crystallisation, the samples showed no monoclinic phase fraction, an increased remanent polarisation ($P_r = \sim 21\,\mu C/cm^2$) and a better endurance of 10^7 cycles, as compared to samples that were annealed with a standard Rapid Thermal Annealer (RTA) at 650 °C. The reduced number of defects at the interface is believed to be at the origin of the performance enhancement. Looking at different electrode materials, it was found that a WOx electrode gives an almost equally high P_r as TiN. The Thickness-dependent studies revealed that the thermal budged required for crystallisation increases for decreasing HZO thicknesses. Layers as thin as 3 nm failed to display ferroelectricity, thus providing another advantage to FeFETs, where the ferroelectric field-effect can be fully exploited over Ferroelectric Tunnelling Junctions (FTJ) where polarisation is traded-off against thickness. The WOx layers were deposited with two different techniques: Rapid Thermal Oxidation (RTO) and Atomic Layer Deposition. It was found that WOx is very sensitive to oxygen reductions and that extra precaution has to be taken during processing to avoid reaching carrier concentrations that diminish the effect of the polarisation on the channel conductance.

Two FeFET generations were then designed, fabricated, and characterised in Chapter 5. In a first step, relatively large $\geq 5 \times 5\ \mu m^2$ gate-first FeFETs were used to demonstrate the direct link between ferroelectric polarisation and channel conductance. The fine-grained domain structure of HZO enabled a programmable and persistent multi-state synapse with more than 4 bits of distinguishable conductance states. Furthermore, these FeFETs displayed a good linearity and symmetry, a low cycle-to-cycle noise (~3.7 %), fast programming speeds(50 ns to 5 µs), and low write energy (0.53 fJ). The device area ($\geq 5 \times 5\ \mu m^2$), dynamic range (2), endurance (8×10^6), and large device-to-device variability required on the other hand improvement.

Finally, a scaled device design was employed to allow for three-terminal crossbar arrays and to mitigate some of the shortcomings plaguing the first FeFET generation. Replacing the RTO-grown WOx of the first FeFETs by a much more controlled ALD process led to thinner and more uniform channels, which in turn increased the electrostatic effect of the polarisation on its conductivity. The result was a sub-µm-sized artificial synapse with a quasi-continuous resistance tuning by a factor of 60 and a fine-grained weight update of more than 200 conductance values. The modulation of the channel displayed a strong time dependence with two mechanisms activated on two different write pulse time scales. This dual nature was highlighted based on temperature-dependent current measurements that showed opposite dynamic range trends for the two timescales, confirming that the observed resistance changes originate from different mechanisms: a fast saturating ferroelectric effect and a slow, less saturating ionic drift and diffusion process. The scaled FeFETs have an excellent endurance of more than $>10^{10}$ cycles and a ferroelectric retention of more than >10 years, demonstrating the benefit of a metal-oxide channel over a silicon one. The footprint of the fabricated devices was successfully reduced to $<1 \times 1$ µm², an important step towards dense integration. Furthermore, it was found that the two physical effects at different timescales result in a deterioration of the symmetry and linearity, as compared to the first generation. A comparison of the performance metrics from the two device generations was presented in Table 5.3. Taking all characteristics into account, the performance of the second FeFET generation was assessed by simulating the classification of the MNIST dataset, resulting in a good accuracy of 88 %, making our technology well-suited for neuromorphic and cognitive computing approaches.

6.2 REMAINING CHALLENGES

Despite the promising results, there are still challenges that need to be solved for FeFETs. The intrinsic problem of few domains in ultra-scaled devices makes it difficult to densely integrate them while maintaining a fine grained landscape of intermediate states. The choice of integration in the BEOL as reported in this thesis solves this problem at the cost of application scope because of a less dense integration. This trade-off is intrinsic to the ferroelectric effect.

Further, the programming voltage and pulsing scheme used in this work are not optimal for the operation by CMOS circuits. The programming voltage can be reduced by reducing the HZO thickness but will be a trade-off with polarisation for layers with a thickness below 10 nm. Here, there is still room for optimisation. The increasing pulse amplitude or width programming scheme require more complex CMOS circuits that can create arbitrary waveforms and require a read operation before each write operation. While for inference applications this is less of a problem, for online learning this leads to an increased energy dissipation and reduces the parallelism due to the reading of each state. Thus, cumulative switching behaviour with identical pulses is preferred but no straight forward. Ferroelectric single-crystal $BaTiO_3$ is well described by the nucleation-limited model where cumulative switching is observed [315]. For polycrystalline HZO, the grains are small and probably switched as a whole, lacking the nucleation-limited character. Controlling grain size and orientation in a more textured HZO could lead to a more cumulative switching behaviour for identical pulses. On the other hand, Mulaosmanovic *et al.* [316] demonstrate cumulative switching in HZO by using trains of pulses and by selecting the appropriate pulse amplitude and length. Trains of pulses without a read operation in between each write pulse is comparable to the increasing pulse-width write-scheme. It can not fully be considered cumulative switching as observed in Ref. [315]. A better understanding of the switching kinetics in HZO will further help to progress towards accumulative switching. Careful defect engineering and a combination with other effects like oxygen migration could be another potential path in this direction.

In our FeFETs we activate oxygen migration for an increased dynamic range. Oxygen migration requires large electric fields and long stimuli, the consequence is an increased energy consumption. Increasing the oxygen mobility by optimising the oxide channel and type of passivation material [317] could be a possible mitigation.

Finally, we can compare our FeFET characteristics to other state of the art device technologies introduced in Chapter 2. Both technologies, filamentary RRAM and PCM are advantageous with respect to scalability and operation voltage. In contrast to the direct link between number of domains and ferroelectric gate area, filaments in RRAM devices and the phase change mechanism in PCM do not directly depend on the device area. Furthermore, both of them can usually be operated with voltages smaller than $\pm 2.5\,\mathrm{V}$ [99, 318]. Also, cumulative switching is observed for both device technologies. Although being advantageous over the aforementioned characteristics where the FeFETs still need to be improved, they have disadvantages that are superior in ferroelectric memristors and especially in FeFETs. Both PCM and RRAM show stochastic switching with increased noise and less stable states. In particular the very fine-grained stable intermediate levels of the FeFETs are difficult to achieve. The large current density needed to change the phase in PCM can be problematic for transistors and can quickly deteriorate interfaces. Furthermore, drift of intermediate states due to structural relaxation in the melt-quenched amorphous state and a wide distribution of the crystallisation voltage are an issue in PCM. Looking at the fabrication process, achieving a dense, void-free, reliable phase-change material with the desired stoichiometry and resistivity post integration, remains difficult [99]. In RRAM devices, a filament needs to be formed before operation which usually requires a higher voltage and the process needs to be well controlled. Often the resulting device resistance are too small (tens of kΩ) for large arrays as they draw too much current. Also, the limited endurance when fully switching from LRS to HRS is a problem in RRAM devices, but can be improved by reducing the switching window [319].

Therefore, if ultra-scaling is not required, the FeFET technology displays many advantages, especially with retention, number of intermediate states, write energy and simple integration over other memristor technologies like PCM and RRAM. Reducing the write voltage and enabling accumulative switching with identical pulses are the remaining engineering challenges to make FeFETs a superior memristor technology.

6.3 OUTLOOK

The results presented in this thesis demonstrate the advantages of oxide channel FeFETs integrated in the BEOL. Three developments can be envisioned as future steps beyond this thesis. First, the processing should be further optimised, especially with respect to the reduction of WOx, hopefully mitigating the device-to-device variability. Second, this device technology should be integrated in three-terminal crossbar arrays with selectors for the gate to demonstrate learning on a real array instead of in simulations. This should be accomplished for small arrays within the BeFerroSynaptic EU H2020 project. Third, to enable more brain inspired and low power algorithms in Spiking Neural Networks (SNN) to be transferred to the analog domain, it is important to further engineer the FeFETs to support a tunable plasticity. This could be achieved by introducing a 4th electrode with a non-ferroelectric gate on top of the WOx channel for a volatile channel resistance component. Having such an additional volatile component would permit to implement a combination of short and long term learning rules and ultimately a wider range of neuromorphic engineering paradigms.

A

APPENDIX

A.1 AUTOMATION: DESIGN-TO-MEASUREMENT

This section is intended to demonstrate the degree of automation that we implemented, especially for the design and later the characterisation of Fe-FETs. We processed many devices (>1000) on a single chip, which means that measuring each of them by hand is not an option. Therefore, we build an electrical probe station with motorised x, y, and z axes. These axes can be remotely controlled by a computer and thus allow to move the chip below the needle-probes automatically to measure many devices sequentially.

A.1.1 GDS Design

Designing a GDS can be done programmatically by python libraries like IPKISS. Generating a GDS by executing a python script has the advantage of a high flexibility when it comes to producing a spread of device-sizes or simply changing any parameter in retrospect. Each type of device can be defined as a class that is parametrised and a subsequent "for loop" can place many variations of a device on the GDS. At the same time, coordinates can be stored directly when a device is placed to create a device-map that can be used for the automated probe station. The following code shows a simplified MIM capacitor device class and how it is placed multiple times on a GDS design. The chip designs presented in Section 5.1.1 and 5.2.1 were created in the same manner:

```python
class MIM(Structure):
    a_mesa = PositiveNumberProperty(required=True)  # size of the square capacitor
    gap = PositiveNumberProperty(default=5.0)  # micro metres

    def define_elements(self, elems):
        # define important coordinates
        contact2Coord = Coord2(self.a_mesa / 2.0, self.a_mesa / 2.0)
        contact1Coord = contact2Coord + Coord2((self.gap + self.a_mesa), 0)
        centerOfDevice = (contact1Coord + contact2Coord) * 0.5

        # create bottom metal
        elems += Rectangle(layer=bottomMetalLayer,
                           center=centerOfDevice,
                           box_size=(2 * self.a_mesa + self.gap + Ccl, self.a_mesa + Ccl))
        # create via to botom metal
        elems += Rectangle(layer=via1Layer, center=contact2Coord,
                           box_size=(self.a_mesa - Ccl, self.a_mesa - Ccl))

        # create top metal
        elems += Rectangle(layer=topMetalLayer, center=contact1Coord,
                           box_size=(self.a_mesa, self.a_mesa))
        # create via to top metal
        via2 = Rectangle(layer=via2Layer, center=contact1Coord,
                           box_size=(self.a_mesa - Ccl, self.a_mesa - Ccl))
        via2.contactParam = {"isContactStr": True, "name": "MIM_%i" % self.a_mesa,
                                  "type": "MIM_Capacitor",
                                  "length": self.a_mesa, "width": self.a_mesa,
                                  "shape": "square",
                                  "isArray": "0", "xPosA": "-1", "yPosA": "-1",
                                  "terminals": "2"}
        elems += via2
        return elems

#Place structures on GDS ---------------------------------------------------------
my_layout = Structure(name="myLayout") # create structure to hold entire GDS design

capacitor_size = [60.0, 20.0, 5.0] # define capacitor sizes
n_same_devices = 2 # repeat same device multiple times
dX = 180 #distance between devices in x
dY = 120 #distance between devices in y
curX = 0.0
curY = 0.0

for x in range(0, n_same_devices):
    curX = 0
    nX = 0
    for a_m in capacitor_size:

        tempMIM = MIM(a_mesa=a_m, gap=10) # create instance of a MIM capacitor
        my_layout += SRef(tempMIM, position=(curX, curY)) # add instance to the GDS layout
        curX += dX
    curY += dY


createXML(my_layout.elements, "MIM.xml", my_layout) # create device-coordinate-map
my_layout.write_gdsii("MIM.gds", unit=1E-6, grid=1E-9) # save layout to GDS file
```

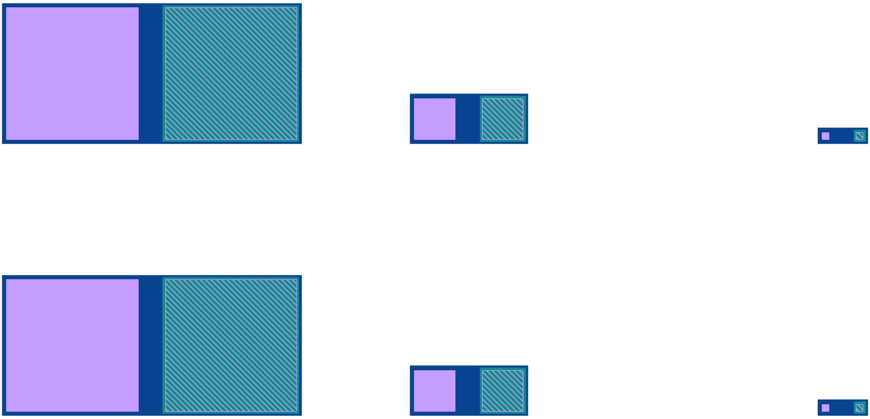


Figure A.1: GDS example created with python and the IPKISS library.

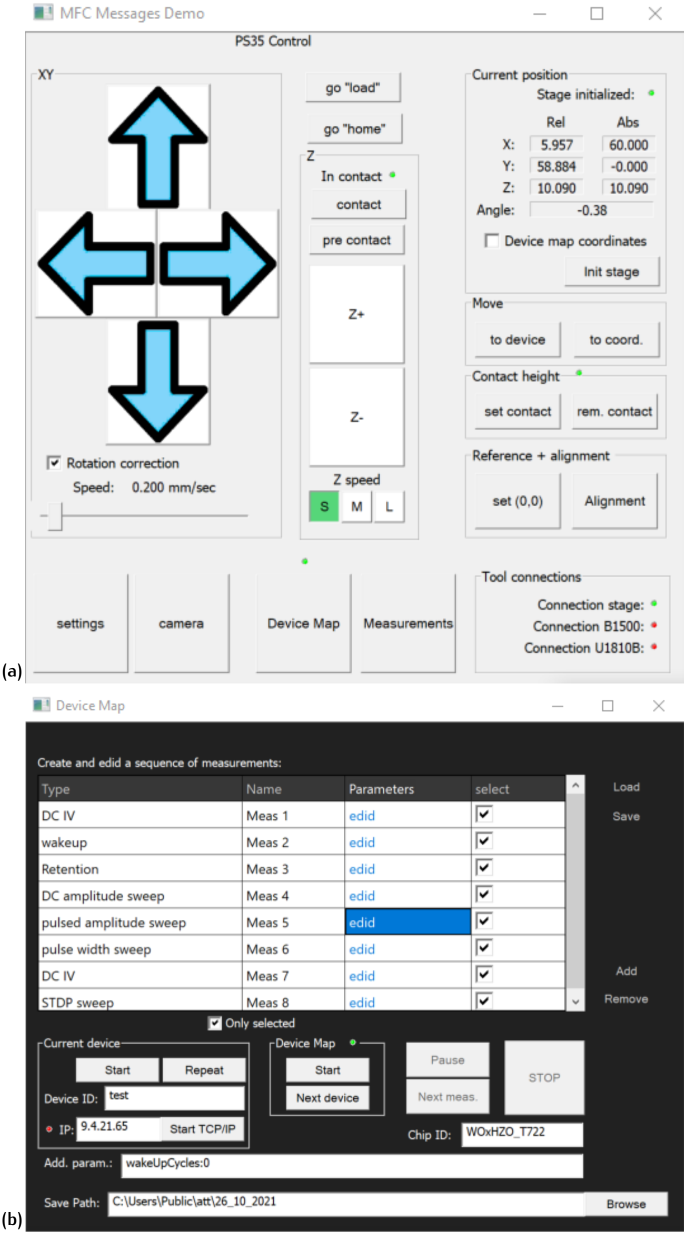
L26-L31 define the custom `contactParam` parameter of the `via2` structure. It contains information that will be used to create a device-map for the automated probe station. Having defined a parametrised device class, we can now place it many times on the GDS with different dimensions.
L34-L52: With a simple "for-loop", the capacitor structures are placed on the GDS. With the custom function `createXML()` (definition not shown) on L55, all structures with a `contactParam` are found and their information and final position on the GDS is saved to a xml file. Then the design is written to the GDS file, which is shown in Figure A.1.

A.1.2 Automated Probe Station

In Section 3.3.2 we described our measurement setup. The Agilent B1500A Semiconductor analyser is connected by GPIB to the same computer as the x, y, z stages. Both of them can be controlled using `C++` libraries provided by the manufacturers. We have therefore created a measurement software with a simple GUI interface (MFC) that has the following functionalities:

1. Stage control: (Figure A.2a)
    - Move the chuck in x, y, z direction.
    - Set contact height.
    - Align to sample rotation.
    - Set the reference position for the device-map.
2. Measurements: (Figure A.2b)
    - Create a list of measurements to perform on each device.
    - Change parameters of each measurement.
    - Save and load measurement lists.
3. Device-map: (Figure A.3a and A.3b)
    - Load a device-map xlm file.
    - Display the device-map.
    - Apply a filter to select only a subset of devices.
    - Save and load filters.

We now can process a chip and then by loading the device-map, automatically get statistics from many devices, including by fast pulsed measurements. For example potentiation and depression cycles on 120 devices were measured and reported in Figure 5.20.



	Rel	Abs
X:	5.957	60.000
Y:	58.884	-0.000
Z:	10.090	10.090
Angle:	-0.38	

☐ Device map coordinates

☑ Rotation correction

Speed: 0.200 mm/sec

Create and edid a sequence of measurements:

Type	Name	Parameters	select
DC IV	Meas 1	edid	☑
wakeup	Meas 2	edid	☑
Retention	Meas 3	edid	☑
DC amplitude sweep	Meas 4	edid	☑
pulsed amplitude sweep	Meas 5	edid	☑
pulse width sweep	Meas 6	edid	☑
DC IV	Meas 7	edid	☑
STDP sweep	Meas 8	edid	☑

☑ Only selected

Current device

Device ID: test

● IP: 9.4.21.65

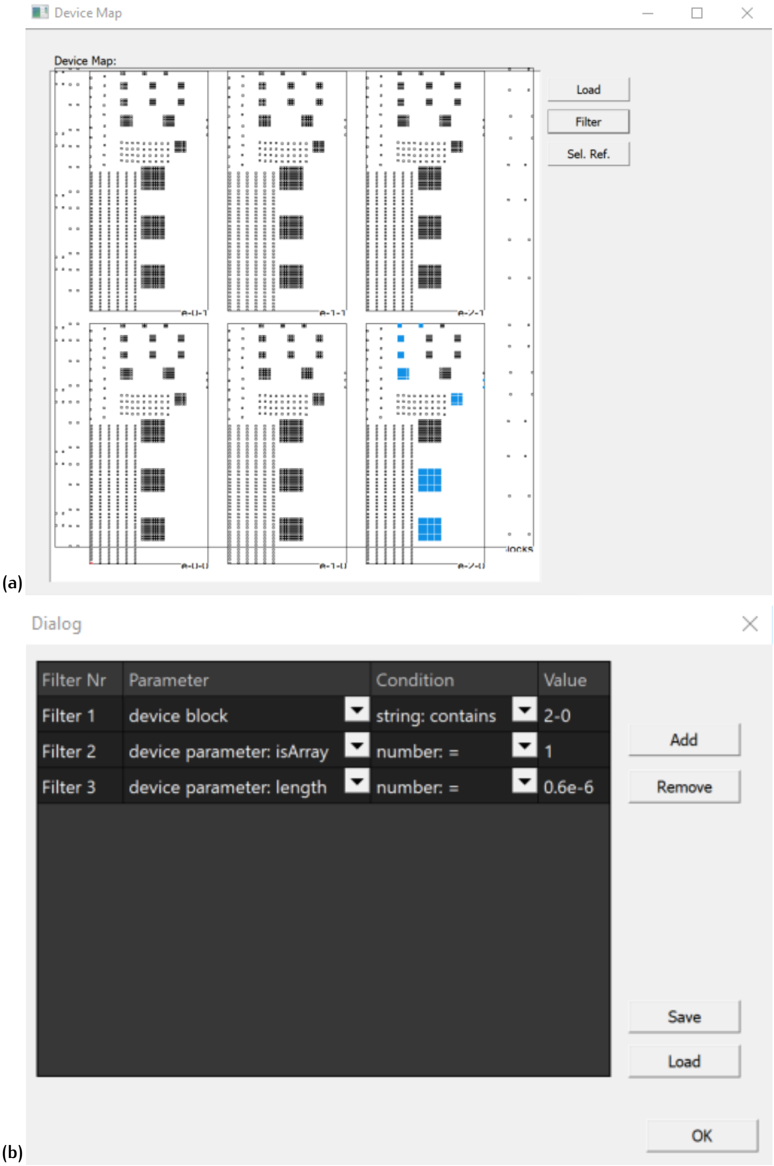
Device Map ●

Chip ID: WOxHZO_T722

Add. param.: wakeUpCycles:0

Save Path: C:\Users\Public\att\26_10_2021

Figure A.2: GUI: **(a)** Stage control window. **(b)** Measurement window with a list of measurements.



Filter Nr	Parameter	Condition	Value
Filter 1	device block	string: contains	2-0
Filter 2	device parameter: isArray	number: =	1
Filter 3	device parameter: length	number: =	0.6e-6

Figure A.3: GUI: **(a)** Device-map window. **(b)** Filter options. Here, only arrays containing devices with $L_{ch} = 0.6\,\mu\text{m}$ on block 2-0 are selected (blue).

A.2 3-TERMINAL CROSSBAR ARRAYS

We designed 3 terminal (3T) crossbar arrays with two goals in mind: First, automated probing of each device in the array by 3 needle probes (no wire bonding). Second, improved layout to make a passive crossbar operation possible. In crossbar arrays, the issue of cross-talk within elements of the matrix is crucial. In particular, when controlling the gate of an element (i, j), the design of the circuit has to be such that voltage drops on other gates do not result in the modification of the value of other cells. In a 3T crossbar array, the Source (S) and Drain (D) of a FeFET connect each horizontal (input) and vertical (output) in the same way as 2T crossbar arrays are connected. In that way, an applied voltage at the input is multiplied by the channel conductance and can be read at the output as a summed current.

To program our FeFETs we ground S and D and apply a write signal to the gate (G). We think it is important to ground S and D simultaneously for an uniform switching of the gate ferroelectric. If the gate line is chosen horizontally or vertically, all devices of that specific gate line are written because all of them have either S or D grounded. This is why we came up with a design were the gate lines are diagonals, as illustrated in Figure A.4a. There are more diagonals $(2n-1)$ than S or D lines in a $n \times n$ array. Furthermore, not all diagonals connect the same amount of devices. Only the center diagonal $(k = 1)$ connects the n devices like the S and D lines. Therefore, we can electrically connect two diagonals with less than n devices (e.g. $k = 2$ and $h = 3$) so that at the end, each gate line coming from the left in Figure A.4a (G1 to G5) is coupled to a total of n devices.

Our goal is to measure each device in the array by three needle probes. Therefore, a special design with elongated contact pads (S_i and D_j) for the S and D lines was applied, as can be seen in Figure A.5a. In an automated probe station, this allows to keep the spacing between needles constant and to move the sample bellow the needles to measure each cross point. For the gate contact we had to apply the same diagonal idea (Figure A.4b) to ensure that each gate contact pad ($W_{i,j}$) is connected to the device $A_{i,j}$ in the crossbar (and $n-1$ other devices). In Figure A.4b the connection example for G3 is shown. The advantage of this diagonal and elongated contacts scheme is displayed in Figure A.5.Three needles from an electrical prober can be kept in the same spacing to measure all $A_{i,j}$ devices in the crossbar. The diagonal layout ensures that the potential drop between $W_{i,j}$ and (S_i,D_j) is only seen by the device $A_{i,j}$. Only $A_{i,j}$ is connected to all

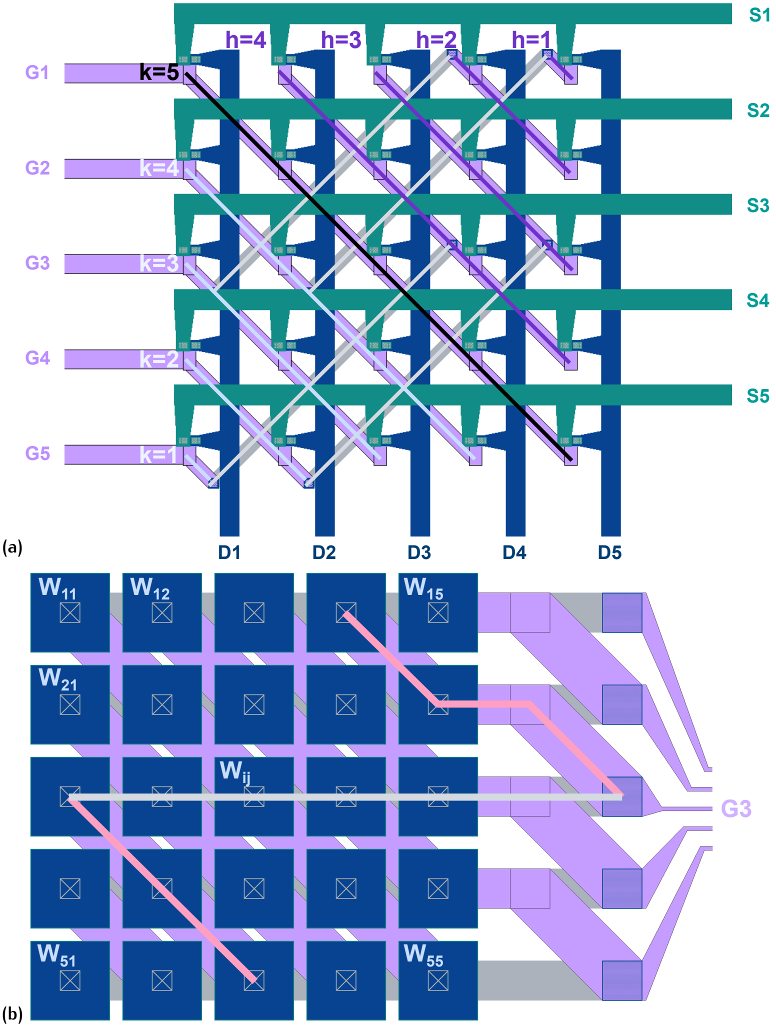


Figure A.4: Passive 3T crossbar array with diagonal gates: **(a)** Schematic of the crossbar showing how two diagonal gate lines are connected together. **(b)** Same principle is applied to the gate contact pads.

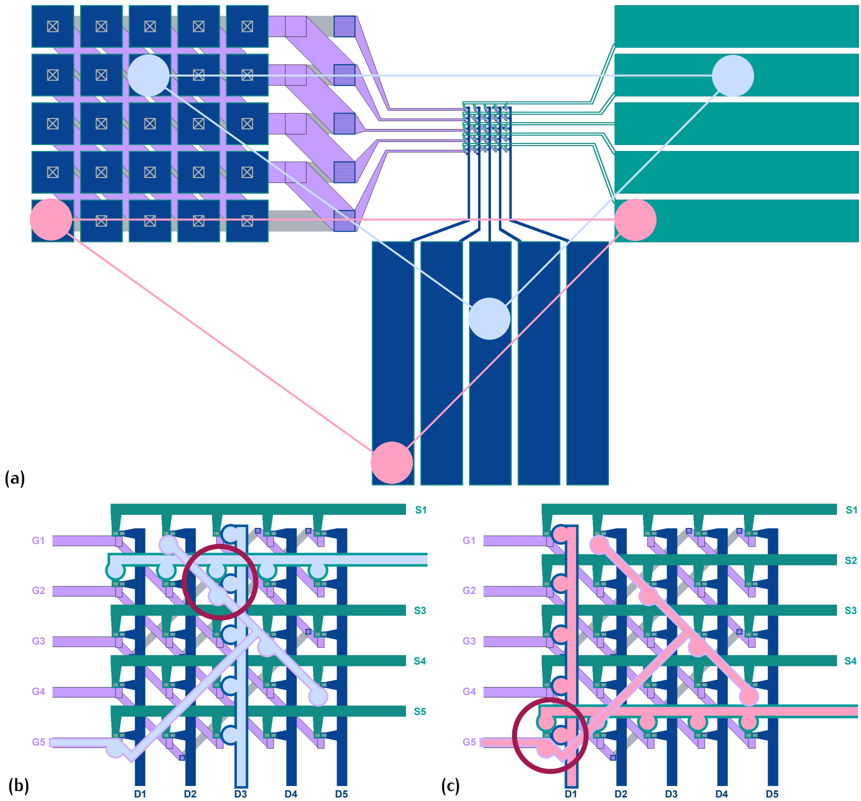


Figure A.5: Diagonal 3T crossbar array: **(a)** Illustration of different measurement positions for the 3 needles by moving the chip below. **(b)** Signal paths for the two positions of (a). Only one device sees all three signals.

3 needles, all other devices are connected to at most 1 needle, making it impossible to see a potential drop.

It turns out that this is not true and the problem is that the channel resistance in our case ($R_{SD} = 10\,\mathrm{M\Omega}$ to $10\,\mathrm{G\Omega}$ at $V_{SD} = 200mV$) usually is much lower than the gate resistance ($R_G > 740\,\mathrm{G\Omega}$ at $V_{write} = 1\,\mathrm{V}$). If a S_i/D_j pair is grounded, all other S_i and D_j are only separated by R_{SD} and thus practically also grounded. Therefore, a passive operation without write disturbance is not possible. Introducing a selector to each gate removes this problem. At the same time, with a selector, the diagonal gate has no advantage over a gate line that is perpendicular to the selector line.

Nevertheless, the elongated pads still allowed to automatically measure every device in an array, which is useful for statistics. Figure A.6 shows the dynamic range that was measured on a 5×5 and 10×10 array of FeFETs with $L_{ch} = 600\,\text{nm}$ and $W_{ch} = 1\,\mu\text{m}$. Because this is a passive crossbar, the measured current between a S_i and D_j is the current of the cross point and an additional current flowing through the sneak-paths. Thus, the measured resistance is lower than expected. The measured Dynamic Range (DR_{meas}) is much smaller than expected due to the sneak current that is not modulated like the current of the cross point. We can define the resistance of the cross point that we want to measure as R_{Sel}. Let us assume a very simple picture where all resistances except the cross point have the same value (R_{notSel}). We can write the measured resistance between S_i and D_j as:

$$R_{meas}(R_{sel}, R_{noSel}, n, m) = \left(\frac{1}{R_{sel}} + \frac{(n-1)(m-1)}{(n+m-1)R_{notSel}} \right)^{-1},$$

where n and m are the array dimensions. DR_{meas} can then be calculated:

$$DR_{meas} = \frac{R_{meas}(HRS, R_{noSel}, n, n)}{R_{meas}(LRS, R_{noSel}, n, n)},$$

where HRS is the High Resistive State and LRS is the Low Resistive State. Figure A.7 plots the evolution of DR_{meas} with increasing array sizes for a DR of 10 and 100. A lower bound would be if all resistances are in the LRS (largest sneak current) and the upper bound if all are in the HRS (smallest sneak-path). The larger the DR is, the bigger the impact of the devices in the LRS on the sneak-path current is.

The problem of sneak-paths can be avoided by sinking all other lines to ground or by applying a signal to all inputs and reading all outputs simultaneously to perform a matrix-vector-multiplication. This requires to connect all array lines (e.g. by wire-bonding) and a setup that can handle so many connections.

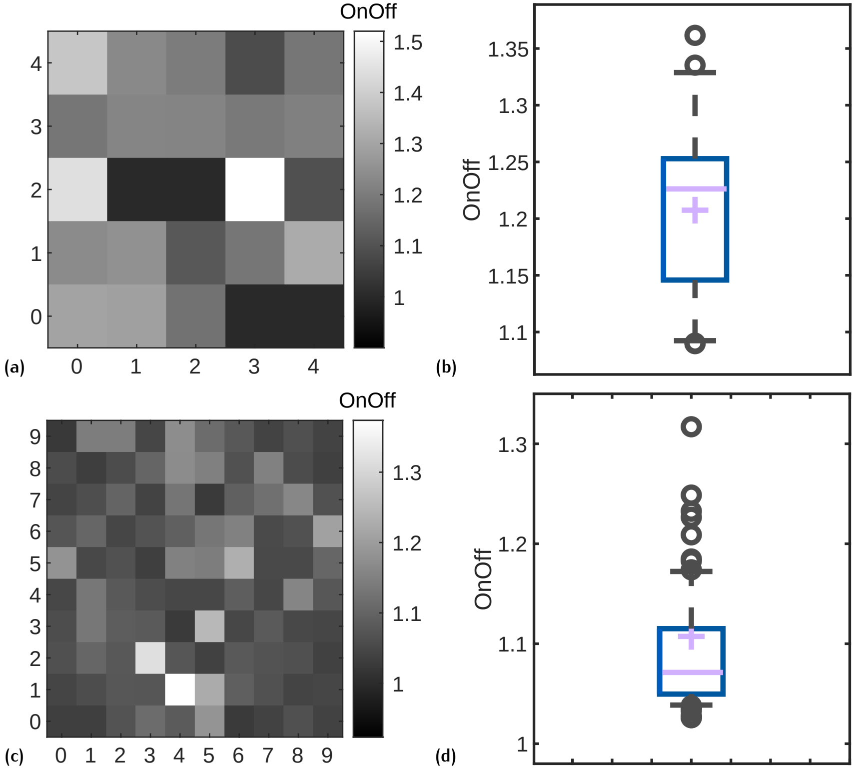


Figure A.6: Automated measurement of entire arrays: **(a, c)** Measured dynamic range on a 5 × 5 and 10 × 10 array. **(b, d)** Corresponding box plots. The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data. Flier points are those past the end of the whiskers.

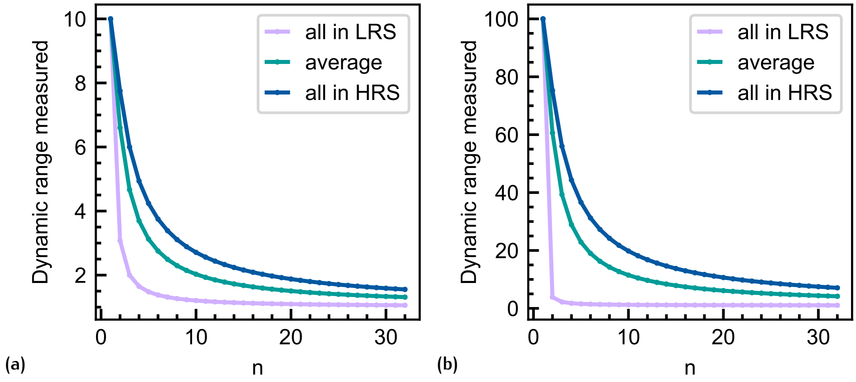


Figure A.7: Dynamic range (*DR*) in a passive crossbar array: Reduction of the measured dynamic range with increasing array dimensions ($n \times n$). **(a)** $DR = 100$ and **(b)** $DR = 10$.

A.3 TEMPERATURE-DEPENDENT CURRENT MEASUREMENTS

Temperature-dependent current measurements of the channel were conducted for two reasons: on the one hand to characterise the WO_x channel conduction mechanism. On the other hand, to further prove that two different effects are contributing to the resistance modulation of the channel at different timescales. The experiment was conducted as follows: at each temperature, the device was first set to its *HRS*, then the channel current (I_D) was measured by applying an $I-V$ sweep from $V_{SD} = -200\,\text{mV}$ to $V_{SD} = 200\,\text{mV}$. Then the same was repeated for the *LRS*. After the $I-V$ measurements were conducted, the temperature was increased. To stabilise the temperature, we waited 10 min before starting with the next $I-V$ sweeps. The following equation describes the ohmic conduction:

$$J_{SD} = \sigma E = \mu q N_C exp\left[\frac{-(E_C - E_F)}{kT}\right]\frac{V_{SD}}{L_{ch}} \tag{A.1}$$

$$\Leftrightarrow$$

$$Log\left(J_{SD}\right) = Log\left(\frac{\mu q N_C}{L_{Ch}}\right) + \frac{-(E_C - E_F)}{k}\frac{1}{T} + Log\left(V_{SD}\right), \tag{A.2}$$

where J_{SD} is the current density, σ the electrical conductivity, μ the electron mobility, q the electronic charge, N_C the carrier concentration, $(E_C - E_F)$

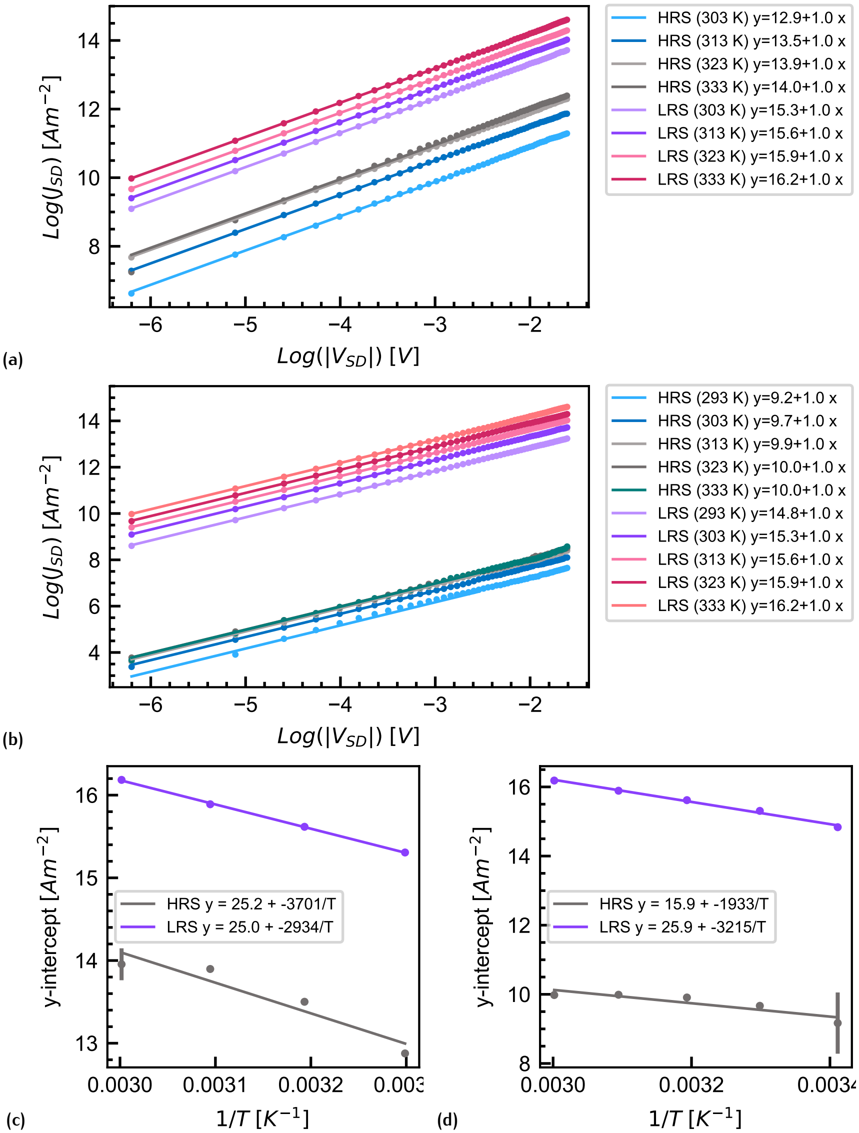


Figure A.8: Analysis of the I_D temperature dependence: **(a)** $Log(J_{DS})-Log(|V|)$ at different temperatures for the *HRS* and *LRS*. The *HRS* was set with $t_w = 500\,\mu\text{s}$ and $V_w = -6\,\text{V}$. **(c)** Arrhenius plot with the y-intercept from (a) as a function of $1/T$. **(b),(d)** Same as (a),(b) but the *HRS* was set by 90 pulses of $t_w = 100\,\text{ms}$ and $V_w = -6\,\text{V}$, trying to saturate the *HRS*.

	LRS $(E_C - E_F)$	*LRS* (μN_C)	*HRS* $(E_C - E_F)$	*HRS* (μN_C)	DR at 60 °C
$t_w = 500\,\mu s$			0.32 eV	1.65×10^{21}	9
$t_w = 90 \cdot 100\,ms$	0.28 eV	3.17×10^{21}	0.17 eV	1.55×10^{17}	500

Table A.1: Contact resistance (R_c) and channel resistivity (ρ_{Ch}) extracted from Figure A.8.

the energy difference between the conduction band and the Fermi level, k the Boltzmann constant, T the absolute temperature, V_{SD} the source-drain voltage, and L the length of the channel. Figure A.8 shows temperature-dependent current measurements of the channel. The channel current density (J_{DS}) as a $Log(J_{DS})-Log(|V|)$ plot for the *HRS* and *LRS* at different temperatures is displayed in sub-plot a. Here, the *HRS* was set by a short pulse of $t_w = 500\,\mu s$ and $V_w = -6\,V$ to stay in a regime where the ferroelectric effect is dominant. All currents were well fitted by a line with slope $= 1$, a characteristic of ohmic conduction. Figure A.8b reports the Arrhenius plot of the y-intercepts from the linear regression of Figure A.8a as a function of $1/T$. According to Equation A.2 we can extract $(E_C - E_F)$ from the slope (slope $= -(E_C - E_F)/k$). This then allows to use Equation A.1 to determine the μN_C product, under the assumption that μN_C is temperature-independent [309].

The same experiment was repeated where the *HRS* was set by 90 pulses of $t_w = 100\,ms$ and $V_w = -6\,V$, representing a much longer timescale (Figure A.8c). As can be seen in Figure 5.22c, a complete saturation of the *HRS* is usually not reached. Again, the current can be fitted with a linear regression with slope $= 1$. The corresponding Arrhenius plot (Figure A.8d) can be used to extract $E_C - E_F$ and μN_C. Since the *HRS* does not saturate, it is difficult to argue that at each temperature the same state is reached and hence also explains why the conduction in Figure A.8d does not linearly depend on $1/T$. In other words, using the same pulses to set the *HRS* at different temperatures will result in a different oxidation state of the WO_x for each temperature. The extracted parameters are summarised in Table A.1 where the values for the *HRS* should be treated with caution as explained above.

Finally, Figure 5.23a shows a dynamic range that increases with temperature, exactly what we expect if the resistance modulation is dominated by oxygen migration: the mobility of oxygen increases with temperature, which leads to an enhanced oxidation and reduction of the channel. In summary, this experiment clearly indicates the ohmic nature of our WO_x channel at read voltages ($V_{SD} = \pm 200\,\mathrm{mV}$) for both the LRS and HRS. Especially, by looking at the dynamic range dependence on temperature we can further prove that two different mechanisms at two different timescales modulate the channel resistance.

A.4 NEUROSIM MODIFICATIONS

In the original *MLP+NeuroSimV3.0* [45] code, the conductance variations from device-to-device are equally applied to all weights in the network. This shortcoming is due to the random generators seed, which is the current time in seconds (original `L3: std::time(0)`). Hence, the same output is produced during an entire second. The initialisation of the devices usually happens faster than a second and thus all devices get the same variation or at most two different variations. Since we want to encode a realistic device-to-device conductance variations, the code was slightly adapted (modified L2: `rd()` instead of `std::time(0)`) to generate a random variation for each device:

L8: Calculate the dynamic range `OnOff`.

L9-L10: Calculate the conductance variation for `minConductance` and `OnOff`.

L13-L14: Gaussian distribution functions with `minConductanceVar` and `OnOffVar` around 0.

L17-L20: Define min and max variations measured on samples.

L25-L30: Draw a random sample from the Gaussian distribution function of the conductance variation.

L31: Add the random conductance variation to `minConductance`.

L35-L40: Draw a random sample from the Gaussian distribution function of the dynamic range variation.

L41: Add the random dynamic range variation to the dynamic range `OnOff`.

L43: Multiply `minConductance` with the random dynamic range to get `maxConductance`.

This way we can account for the variation in conductance and dynamic range. If a random value from far into the Gaussian tail is picked, which is higher or lower than the maximum or minimum variation measured on the devices, the process is repeated and a new variation is picked until it falls within the boundaries. The resulting distributions are shown in Figure 5.27

A.4.1 Original Code

```
//It's OK not to use the external gen, since here the device-to-device vairation is a one-time deal
std::mt19937 localGen;
localGen.seed(std::time(0));

/* Conductance range variation */
conductanceRangeVar = false; //Consider variation of conductance range or not
maxConductanceVar = 0; //Sigma of maxConductance variation (S)
minConductanceVar = 0; //Sigma of minConductance variation (S)
gaussian_dist_maxConductance = new std::normal_distribution<double>(0, maxConductanceVar);
gaussian_dist_minConductance = new std::normal_distribution<double>(0, minConductanceVar);

if (conductanceRangeVar) {
  maxConductance += (*gaussian_dist_maxConductance)(localGen);
  minConductance += (*gaussian_dist_minConductance)(localGen);

  // Conductance variation check
  if (minConductance >= maxConductance || maxConductance < 0 || minConductance < 0 )
  {
    puts("[Error] Conductance variation check not passed. The variation may be too large.");
    exit(-1);
  }
  // Use the code below instead for re-choosing the variation if the check is not passed
  //do {
  //  maxConductance = avgMaxConductance + (*gaussian_dist_maxConductance)(localGen);
  //  minConductance = avgMinConductance + (*gaussian_dist_minConductance)(localGen);
  //} while (minConductance >= maxConductance || maxConductance < 0 || minConductance < 0);
}
```

A.4.2 Modified Code

```
std::random_device rd{};
std::mt19937 localGen(rd()); //Change to rd() to create a random seed for every device

/* Conductance range variation *///=Variation of HRS and LRS from device-to-device
conductanceRangeVar = true; //Consider variation of conductance range or not
if (conductanceRangeVar)
{
  OnOff = maxConductance / minConductance;
  minConductanceVar = 0.39 * minConductance; //Sigma of minConductance variation (S)
  OnOffVar = 0.29 * OnOff; //Sigma of OnOff variation

  //create gaussian distribution function around 0
  gaussian_dist_minCond = new std::normal_distribution<double>(0, minConductanceVar);
  gaussian_dist_OnOffVar = new std::normal_distribution<double>(0, OnOffVar);

  //max and min values in all devices measured in % of the mean
  double minCond_clip_max = 1.88;
  double minCond_clip_min = 0.38;
  double OnOffVar_clip_max = 1.43;
  double OnOffVar_clip_min = 0.48;

  /*If random sample from gaussian distribution is more/less than the max/min device
   measured, repeat if this is not done, sometimes a random sample is so far in the
   tail of the gaussian, that it makes the resulting conductance negative*/
  double minCondChange = 0;
  do{
    minCondChange = (*gaussian_dist_minCond)(localGen);

  }while((minConductance + minCondChange) <= (minConductance * minCondVar_clip_min)
      || (minConductance + minCondChange) >= (minConductance * minCondVar_clip_max));
  minConductance += minCondChange;

  /*Multiply minConductance by a OnOff value from a gaussian distribution around the
  mean. Repeat if OnOff is too far in the tale*/
  double OnOffChange = 0;
  do{
    OnOffChange = (*gaussian_dist_OnOffVar)(localGen);

  }while((OnOff + OnOffChange) <= (OnOff * OnOffVar_clip_min)
      || (OnOff + OnOffChange) >= (OnOff * OnOffVar_clip_max));
  OnOff += OnOffChange;

  maxConductance = minConductance * OnOff;
  printf("conductance=%.4e / %.4e / %f\n", minConductance,maxConductance,OnOff);

  // Conductance variation check
  if (minConductance >= maxConductance || maxConductance < 0 || minConductance < 0 )
  {
    puts("[Error] Conductance variation check not passed. The variation may be too large.");
    exit(-1);
  }
}
```

BIBLIOGRAPHY

1. Völske, M., Bevendorff, J., Kiesel, J., Stein, B., Fröbe, M., Hagen, M. & Potthast, M. *Web Archive Analytics* in *INFORMATIK 2020* (eds Reussner, R. H., Koziolek, A. & Heinrich, R.) (Gesellschaft für Informatik, Bonn, 2021), 61.
2. Umair, M., Cheema, M. A., Cheema, O., Li, H. & Lu, H. Impact of COVID-19 on IoT Adoption in Healthcare, Smart Homes, Smart Buildings, Smart Cities, Transportation and Industrial IoT. *Sensors* **21**, 3838 (2021).
3. Reinsel, D., Rydning, J. & Gantz, J. *Worldwide Global DataSphere Forecast, 2021-2025: The World Keeps Creating More Data — Now, What Do We Do With It All?* 2021.
4. Abdo, G. & Mahale, V. *Worldwide Managed Edge Services Forecast, 2021-2025* 2021.
5. Krishnamurthi, R., Kumar, A., Gopinathan, D., Nayyar, A. & Qureshi, B. An Overview of IoT Sensor Data Processing, Fusion, and Analysis Techniques. *Sensors* **20**, 6076 (2020).
6. Al-Sarawi, S., Anbar, M., Abdullah, R. & Al Hawari, A. B. *Internet of Things Market Analysis Forecasts, 2020-2030* in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)* (IEEE, 2020), 449.
7. Indiveri, G. & Liu, S.-C. Memory and Information Processing in Neuromorphic Systems. *Proceedings of the IEEE* **103**, 1379 (2015).
8. Yang, R. In-Memory Computing With Ferroelectrics. *Nature Electronics* **3**, 237 (2020).
9. Feldman, D. E. Synaptic Mechanisms for Plasticity in Neocortex. *Annual Review of Neuroscience* **32**, 33 (2009).
10. Kasabov, N., Dhoble, K., Nuntalid, N. & Indiveri, G. Dynamic Evolving Spiking Neural Networks for on-Line Spatio- And Spectro-Temporal Pattern Recognition. *Neural Networks* **41**, 188 (2013).
11. Beilliard, Y. & Alibart, F. Multi-Terminal Memristive Devices Enabling Tunable Synaptic Plasticity in Neuromorphic Hardware: A Mini-Review. *Frontiers in Nanotechnology* **3**, 1 (2021).
12. Payvand, M., Nair, M. V., Müller, L. K. & Indiveri, G. A Neuromorphic Systems Approach to in-Memory Computing With Non-Ideal Memristive Devices: From Mitigation to Exploitation. *Faraday Discussions* **213**, 487 (2019).
13. Indiveri, G., Linares-Barranco, B., Legenstein, R., Deligeorgis, G. & Prodromakis, T. Integration of Nanoscale Memristor Synapses in Neuromorphic Computing Architectures. *Nanotechnology* **24**, 384010 (2013).
14. Kub, F., Moon, K., Mack, I. & Long, F. Programmable Analog Vector-Matrix Multipliers. *IEEE Journal of Solid-State Circuits* **25**, 207 (1990).
15. Youngblood, N., Ríos, C., Gemo, E., Feldmann, J., Cheng, Z., Baldycheva, A., Pernice, W. H., Wright, C. D. & Bhaskaran, H. Tunable Volatility of Ge 2 Sb 2 Te 5 in Integrated Photonics. *Advanced Functional Materials* **29**, 1807571 (2019).
16. Diehl, P. U. & Cook, M. Unsupervised Learning of Digit Recognition Using Spike-Timing-Dependent Plasticity. *Frontiers in Computational Neuroscience* **9**, 1 (2015).
17. Ielmini, D., Wang, Z. & Liu, Y. Brain-Inspired Computing via Memory Device Physics. *APL Materials* **9** (2021).
18. Chakraborty, I., Jaiswal, A., Saha, A. K., Gupta, S. K. & Roy, K. Pathways to Efficient Neuromorphic Computing With Non-Volatile Memory Technologies. *Applied Physics Reviews* **7**, 021308 (2020).
19. Brown, T. B. *et al. Language Models Are Few-Shot Learners* 1877 (2020).

20. Yu, S. Neuro-Inspired Computing With Emerging Nonvolatile Memorys. *Proceedings of the IEEE* **106**, 260 (2018).

21. Gokmen, T. & Vlasov, Y. Acceleration of Deep Neural Network Training With Resistive Cross-Point Devices: Design Considerations. *Frontiers in Neuroscience* **10**, 1 (2016).

22. Xiao, T. P., Bennett, C. H., Feinberg, B., Agarwal, S. & Marinella, M. J. Analog Architectures for Neural Network Acceleration Based on Non-Volatile Memory. *Applied Physics Reviews* **7**, 031301 (2020).

23. Ross, I. M. *Semiconductive Translating Device* 1957.

24. Yeh, C.-P., Lisker, M., Kalkofen, B. & Burte, E. P. Fabrication and Investigation of Three-Dimensional Ferroelectric Capacitors for the Application of FeRAM. *AIP Advances* **6**, 035128 (2016).

25. Zhang, X., Takahashi, M., Takeuchi, K. & Sakai, S. 64 Kbit Ferroelectric-Gate-Transistor-Integrated NAND Flash Memory With 7.5 v Program and Long Data Retention. *Japanese Journal of Applied Physics* **51**, 04DD01 (2012).

26. McAdams, H. *et al.* A 64-Mb Embedded FRAM Utilizing a 130-Nm 5LM Cu/FSG Logic Process. *IEEE Journal of Solid-State Circuits* **39**, 667 (2004).

27. Böscke, T. S., Müller, J., Bräuhaus, D., Schröder, U. & Böttger, U. Ferroelectricity in Hafnium Oxide Thin Films. *Applied Physics Letters* **99**, 102903 (2011).

28. Yu, S., Hur, J., Luo, Y.-C., Shim, W., Choe, G. & Wang, P. Ferroelectric HfO_2 -Based Synaptic Devices: Recent Trends and Prospects. *Semiconductor Science and Technology* **36**, 104001 (2021).

29. Mistry, K. *et al. A 45nm Logic Technology With High-K+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-Free Packaging* in *2007 IEEE International Electron Devices Meeting* (IEEE, 2007), 247.

30. Luo, Q. *et al.* A Highly CMOS Compatible Hafnia-Based Ferroelectric Diode. *Nature Communications* **11**, 1391 (2020).

31. O'Connor, É., **Halter, M.**, Eltes, F., Sousa, M., Kellock, A., Abel, S. & Fompeyrine, J. Stabilization of Ferroelectric Hf X Zr 1-x O 2 Films Using a Millisecond Flash Lamp Annealing Technique. *APL Materials* **6**, 121103 (2018).

32. Begon-Lours, L., **Halter, M.**, Popoff, Y., Yu, Z., Falcone, D. F., Davila, D., Bragaglia, V., Porta, A. L., Jubin, D., Fompeyrine, J. & Offrein, B. J. Analog Resistive Switching in BEOL, Ferroelectric Synaptic Weights. *IEEE Journal of the Electron Devices Society* **9**, 1275 (2021).

33. Bégon-Lours, L., **Halter, M.**, Popoff, Y. & Offrein, B. J. Ferroelectric, Analog Resistive Switching in Back-End-of-Line Compatible TiN/$HfZrO_4$ /TiO X Junctions. *physica status solidi (RRL) - Rapid Research Letters* **15**, 2000524 (2021).

34. Liang, Y.-K., Wu, J.-S., Teng, C.-Y., Ko, H.-L., Luc, Q.-H., Su, C.-J., Chang, E.-Y. & Lin, C.-H. Demonstration of Highly Robust 5 Nm $Hf_{0.5}Zr_{0.5}O_2$ Ultra-Thin Ferroelectric Capacitor by Improving Interface Quality. *IEEE Electron Device Letters* **42**, 1299 (2021).

35. Han, H., Yu, H., Wei, H., Gong, J. & Xu, W. Recent Progress in Three-Terminal Artificial Synapses: From Device to System. *Small* **15**, 1900695 (2019).

36. Mulaosmanovic, H., Breyer, E. T., Dünkel, S., Beyer, S., Mikolajick, T. & Slesazeck, S. Ferroelectric Field-Effect Transistors Based on HfO_2 : A Review. *Nanotechnology* **32**, 502002 (2021).

37. Breyer, E. T., Mulaosmanovic, H., Mikolajick, T. & Slesazeck, S. Perspective on Ferroelectric, Hafnium Oxide Based Transistors for Digital Beyond Von-Neumann Computing. *Applied Physics Letters* **118**, 050501 (2021).

38. **Halter, M.**, Bégon-Lours, L., Bragaglia, V., Sousa, M., Offrein, B. J., Abel, S., Luisier, M. & Fompeyrine, J. Back-End, CMOS-Compatible Ferroelectric Field-Effect Transistor for Synaptic Weights. *ACS Applied Materials & Interfaces* **12**, 17725 (2020).

39. Dunkel, S. *et al. A FeFET Based Super-Low-Power Ultra-Fast Embedded NVM Technology for 22nm FDSOI and Beyond* in *2017 IEEE International Electron Devices Meeting (IEDM)* **1** (IEEE, 2017), 19.7.1.

40. Mulaosmanovic, H., Muller, F., Breyer, E. T., Dunkel, S., Trentzsch, M., Beyer, S., Mikolajick, T. & Slesazeck, S. *HfO_2 -Based Ferroelectric FETs: Performance of Single Devices and Mini-Arrays* in *2020 International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA)* **1** (IEEE, 2020), 146.

41. Mo, F., Tagawa, Y., Jin, C., Ahn, M., Saraya, T., Hiramoto, T. & Kobayashi, M. *Experimental Demonstration of Ferroelectric HfO_2 FET With Ultrathin-Body IGZO for High-Density and Low-Power Memory Application* in *2019 Symposium on VLSI Technology* (IEEE, 2019), T42.

42. Kim, S. *et al. Metal-Oxide Based, CMOS-compatible ECRAM for Deep Learning Accelerator* in *2019 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2019), 35.7.1.

43. CHARLTON, M. G. Hydrogen Reduction of Tungsten Trioxide. *Nature* **169**, 109 (1952).

44. Ingham, B., Hendy, S. C., Chong, S. V. & Tallon, J. L. Density-Functional Studies of Tungsten Trioxide, Tungsten Bronzes, and Related Systems. *Physical Review B* **72**, 075109 (2005).

45. Chen, P.-Y., Peng, X. & Yu, S. *NeuroSim+: An Integrated Device-to-Algorithm Framework for Benchmarking Synaptic Devices and Array Architectures* in *2017 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2017), 6.1.1.

46. Li Deng. The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Processing Magazine* **29**, 141 (2012).

47. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* 800 (MIT Press, 2016).

48. Bellman, R. *Dynamic Programming* 1st ed. (Princeton University Press, Princeton, NJ, USA, 1957).

49. Arel, I., Rose, D. C. & Karnowski, T. P. Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier]. *IEEE Computational Intelligence Magazine* **5**, 13 (2010).

50. Ahmad, J., Farman, H. & Jan, Z. in *SpringerBriefs in Computer Science* 31 (2019).

51. Lee, T. S. & Mumford, D. Hierarchical Bayesian Inference in the Visual Cortex. *Journal of the Optical Society of America A* **20**, 1434 (2003).

52. Linnainmaa, S. Taylor Expansion of the Accumulated Rounding Error. *BIT* **16**, 146 (1976).

53. Oh, K.-S. & Jung, K. GPU Implementation of Neural Networks. *Pattern Recognition* **37**, 1311 (2004).

54. Chen, X., Eversole, A., Li, G., Yu, D. & Seide, F. *Pipelined Back-Propagation for Context-Dependent Deep Neural Networks* in *Interspeech* (ISCA, 2012).

55. Silver, D. *et al.* Mastering the Game of Go With Deep Neural Networks and Tree Search. *Nature* **529**, 484 (2016).

56. Wang, Z. *et al.* Reinforcement Learning With Analogue Memristor Arrays. *Nature Electronics* **2**, 115 (2019).

57. Indiveri, G. Introducing 'Neuromorphic Computing and Engineering'. *Neuromorphic Computing and Engineering* **1**, 010401 (2021).

58. Mead, C. Neuromorphic Electronic Systems. *Proceedings of the IEEE* **78**, 1629 (1990).

59. Moradi, S., Qiao, N., Stefanini, F. & Indiveri, G. A Scalable Multicore Architecture With Heterogeneous Memory Structures for Dynamic Neuromorphic Asynchronous Processors (DYNAPs). *IEEE Transactions on Biomedical Circuits and Systems* **12**, 106 (2018).

60. Qiao, N., Mostafa, H., Corradi, F., Osswald, M., Stefanini, F., Sumislawska, D. & Indiveri, G. A Reconfigurable on-Line Learning Spiking Neuromorphic Processor Comprising 256 Neurons and 128K Synapses. *Frontiers in Neuroscience* **9**, 1 (2015).

61. Qiao, N., Bartolozzi, C. & Indiveri, G. An Ultralow Leakage Synaptic Scaling Homeostatic Plasticity Circuit With Configurable Time Scales Up to 100 Ks. *IEEE Transactions on Biomedical Circuits and Systems* **11**, 1271 (2017).

62. Lobo, J. L., Del Ser, J., Bifet, A. & Kasabov, N. Spiking Neural Networks and Online Learning: An Overview and Perspectives. *Neural Networks* **121**, 88 (2020).

63. Schemmel, J., Briiderle, D., Griibl, A., Hock, M., Meier, K. & Millner, S. *A Wafer-Scale Neuromorphic Hardware System for Large-Scale Neural Modeling* in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* (IEEE, 2010), 1947.

64. Furber, S. B., Galluppi, F., Temple, S. & Plana, L. A. The SpiNNaker Project. *Proceedings of the IEEE* **102**, 652 (2014).

65. Merolla, P. A. *et al.* A Million Spiking-Neuron Integrated Circuit With a Scalable Communication Network and Interface. *Science* **345**, 668 (2014).

66. Davies, M. *et al.* Loihi: A Neuromorphic Manycore Processor With on-Chip Learning. *IEEE Micro* **38**, 82 (2018).

67. Davies, M., Wild, A., Orchard, G., Sandamirskaya, Y., Guerra, G. A., Joshi, P., Plank, P. & Risbud, S. R. Advancing Neuromorphic Computing With Loihi: A Survey of Results and Outlook. *Proceedings of the IEEE* **109**, 911 (2021).

68. Strukov, D. B., Snider, G. S., Stewart, D. R. & Williams, R. S. The Missing Memristor Found. *Nature* **453**, 80 (2008).

69. Di Ventra, M., Pershin, Y. V. & Chua, L. O. Circuit Elements With Memory: Memristors, Memcapacitors, and Meminductors. *Proceedings of the IEEE* **97**, 1717 (2009).

70. Mulaosmanovic, H., Chicca, E., Bertele, M., Mikolajick, T. & Slesazeck, S. Mimicking Biological Neurons With a Nanoscale Ferroelectric Transistor. *Nanoscale* **10**, 21755 (2018).

71. Neftci, E. O., Mostafa, H. & Zenke, F. Surrogate Gradient Learning in Spiking Neural Networks: Bringing the Power of Gradient-Based Optimization to Spiking Neural Networks. *IEEE Signal Processing Magazine* **36**, 51 (2019).

72. Bellec, G., Scherr, F., Subramoney, A., Hajek, E., Salaj, D., Legenstein, R. & Maass, W. A Solution to the Learning Dilemma for Recurrent Networks of Spiking Neurons. *Nature Communications* **11**, 1 (2020).

73. Chicca, E. & Indiveri, G. A Recipe for Creating Ideal Hybrid Memristive-Cmos Neuromorphic Processing Systems. *Applied Physics Letters* **116**, 120501 (2020).

74. Brivio, S., Conti, D., Nair, M. V., Frascaroli, J., Covi, E., Ricciardi, C., Indiveri, G. & Spiga, S. Extended Memory Lifetime in Spiking Neural Networks Employing Memristive Synapses With Nonlinear Conductance Dynamics. *Nanotechnology* **30**, 015102 (2019).

75. Lillicrap, T. P. & Santoro, A. Backpropagation Through Time and the Brain. *Current Opinion in Neurobiology* **55**, 82 (2019).

76. Diederich, N., Bartsch, T., Kohlstedt, H. & Ziegler, M. A Memristive Plasticity Model of Voltage-Based STDP Suitable for Recurrent Bidirectional Neural Networks in the Hippocampus. *Scientific Reports* **8**, 9367 (2018).

77. Berdan, R., Vasilaki, E., Khiat, A., Indiveri, G., Serb, A. & Prodromakis, T. Emulating Short-Term Synaptic Dynamics With Memristive Devices. *Scientific Reports* **6**, 18639 (2016).

78. Christensen, D. V. *et al.* 2022 Roadmap on Neuromorphic Computing and Engineering (2021).

79. Chua, L. Memristor-the Missing Circuit Element. *IEEE Transactions on Circuit Theory* **18**, 507 (1971).

80. Tellini, B., Bologna, M., Chandía, K. J. & Macucci, M. Revisiting the Memristor Concept Within Basic Circuit Theory. *International Journal of Circuit Theory and Applications* **49**, 3488 (2021).

81. Chua, L. Resistance Switching Memories Are Memristors. *Applied Physics A* **102**, 765 (2011).

82. Chua, L. in *Memristors and Memristive Systems* 17 (Springer New York, New York, NY, 2014).

83. Chua, L. O. The Fourth Element. *Proceedings of the IEEE* **100**, 1920 (2012).

84. Desoer, C. A. *Basic Circuit Theory* (Tata McGraw-Hill Education, 1969).

85. Di Ventra, M. & Pershin, Y. V. On the Physical Properties of Memristive, Memcapacitive and Meminductive Systems. *Nanotechnology* **24**, 255201 (2013).

86. Chua, L. & Sung Mo Kang. Memristive Devices and Systems. *Proceedings of the IEEE* **64**, 209 (1976).

87. Wang, F. Z., Li, L., Shi, L., Wu, H. & Chua, L. O. Φ Memristor: Real Memristor Found. *Journal of Applied Physics* **125** (2019).

88. Yu, S., Li, Z., Chen, P.-Y., Wu, H., Gao, B., Wang, D., Wu, W. & Qian, H. *Binary Neural Network With 16 Mb RRAM Macro Chip for Classification and Online Training* in *2016 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2016), 16.2.1.

89. Yu, H., Ni, L. & Huang, H. in *Studies in Computational Intelligence* 275 (2017).

90. Cai, F., Correll, J. M., Lee, S. H., Lim, Y., Bothra, V., Zhang, Z., Flynn, M. P. & Lu, W. D. A Fully Integrated Reprogrammable Memristor-CMOS System for Efficient Multiply-accumulate Operations. *Nature Electronics* **2**, 290 (2019).

91. Yao, P., Wu, H., Gao, B., Eryilmaz, S. B., Huang, X., Zhang, W., Zhang, Q., Deng, N., Shi, L., Wong, H. S. & Qian, H. Face Classification Using Electronic Synapses. *Nature Communications* **8**, 1 (2017).

92. Hu, M. *et al.* Memristor-Based Analog Computation and Neural Network Classification With a Dot Product Engine. *Advanced Materials* **30**, 1705914 (2018).

93. Prezioso, M., Merrikh-Bayat, F., Hoskins, B. D., Adam, G. C., Likharev, K. K. & Strukov, D. B. Training and Operation of an Integrated Neuromorphic Network Based on Metal-Oxide Memristors. *Nature* **521**, 61 (2015).

94. Chen, P.-Y., Lin, B., Wang, I.-T., Hou, T.-H., Ye, J., Vrudhula, S., Seo, J.-s., Cao, Y. & Yu, S. *Mitigating Effects of Non-Ideal Synaptic Device Characteristics for on-Chip Learning* in *2015 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)* (IEEE, 2015), 194.

95. Ovshinsky, S. R. Reversible Electrical Switching Phenomena in Disordered Structures. *Physical Review Letters* **21**, 1450 (1968).

96. Neale, R. & Aseltine, J. The Application of Amorphous Materials to Computer Memories. *IEEE Transactions on Electron Devices* **20**, 195 (1973).

97. Burr, G. W. *et al.* Phase Change Memory Technology. *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena* **28**, 223 (2010).

98. Koelmans, W. W., Sebastian, A., Jonnalagadda, V. P., Krebs, D., Dellmann, L. & Eleftheriou, E. Projected Phase-Change Memory Devices. *Nature Communications* **6**, 8181 (2015).

99. Le Gallo, M. & Sebastian, A. An Overview of Phase-Change Memory Device Physics. *Journal of Physics D: Applied Physics* **53**, 213002 (2020).

100. Khaddam-Aljameh, R. *et al. HERMES Core - A 14nm CMOS and PCM-based in-Memory Compute Core Using an Array of 300ps/LSB Linearized CCO-based ADCs and Local Digital Processing* in *2021 Symposium on VLSI Circuits* (IEEE, 2021), 1.

101. Covi, E., Brivio, S., Serb, A., Prodromakis, T., Fanciulli, M. & Spiga, S. Analog Memristive Synapse in Spiking Networks Implementing Unsupervised Learning. *Frontiers in Neuroscience* **10**, 1 (2016).

102. Covi, E., George, R., Frascaroli, J., Brivio, S., Mayr, C., Mostafa, H., Indiveri, G. & Spiga, S. Spike-Driven Threshold-Based Learning With Memristive Synapses and Neuromorphic Silicon Neurons. *Journal of Physics D: Applied Physics* **51**, 344003 (2018).

103. Mochida, R., Kouno, K., Hayata, Y., Nakayama, M., Ono, T., Suwa, H., Yasuhara, R., Katayama, K., Mikawa, T. & Gohou, Y. *A 4M Synapses Integrated Analog ReRAM Based 66.5 TOPS/W Neural-Network Processor With Cell Current Controlled Writing and Flexible Network Architecture* in *2018 IEEE Symposium on VLSI Technology* **2018-June** (IEEE, 2018), 175.

104. Yao, P., Wu, H., Gao, B., Tang, J., Zhang, Q., Zhang, W., Yang, J. J. & Qian, H. Fully Hardware-Implemented Memristor Convolutional Neural Network. *Nature* **577**, 641 (2020).

105. Tsymbal, E. Y. & Kohlstedt, H. Tunneling Across a Ferroelectric. *Science* **313**, 181 (2006).

106. Garcia, V., Fusil, S., Bouzehouane, K., Enouz-Vedrenne, S., Mathur, N. D., Barthélémy, A. & Bibes, M. Giant Tunnel Electroresistance for Non-Destructive Readout of Ferroelectric States. *Nature* **460**, 81 (2009).

107. Gruverman, A., Wu, D., Lu, H., Wang, Y., Jang, H. W., Folkman, C. M., Zhuravlev, M. Y., Felker, D., Rzchowski, M., Eom, C. B. & Tsymbal, E. Y. Tunneling Electroresistance Effect in Ferroelectric Tunnel Junctions at the Nanoscale. *Nano Letters* **9**, 3539 (2009).

108. Chouprik, A., Chernikova, A., Markeev, A., Mikheev, V., Negrov, D., Spiridonov, M., Zarubin, S. & Zenkevich, A. Electron Transport Across Ultrathin Ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ Films on Si. *Microelectronic Engineering* **178**, 250 (2017).

109. Berdan, R., Marukame, T., Ota, K., Yamaguchi, M., Saitoh, M., Fujii, S., Deguchi, J. & Nishi, Y. Low-Power Linear Computation Using Nonlinear Ferroelectric Tunnel Junction Memristors. *Nature Electronics* **3**, 259 (2020).

110. Max, B., Hoffmann, M., Mulaosmanovic, H., Slesazeck, S. & Mikolajick, T. Hafnia-Based Double-Layer Ferroelectric Tunnel Junctions as Artificial Synapses for Neuromorphic Computing. *ACS Applied Electronic Materials* **2**, 4023 (2020).

111. Moll, J. & Tarui, Y. A New Solid State Memory Resistor. *IEEE Transactions on Electron Devices* **10**, 338 (1963).

112. Boscke, T. S., Muller, J., Brauhaus, D., Schroder, U. & Bottger, U. *Ferroelectricity in Hafnium Oxide: CMOS Compatible Ferroelectric Field Effect Transistors* in *2011 International Electron Devices Meeting* (IEEE, 2011), 24.5.1.

113. Kund, M., Beitel, G., Pinnow, C.-U., Rohr, T., Schumann, J., Symanczyk, R., Ufert, K.-d. & Muller, G. *Conductive Bridging RAM (CBRAM): An Emerging Non-Volatile Memory Technology Scalable to Sub 20nm* in *IEEE InternationalElectron Devices Meeting, 2005. IEDM Technical Digest.* **00** (IEEE, 2005), 754.

114. Cha, J.-H., Yang, S. Y., Oh, J., Choi, S., Park, S., Jang, B. C., Ahn, W. & Choi, S.-Y. Conductive-Bridging Random-Access Memories for Emerging Neuromorphic Computing. *Nanoscale* **12**, 14339 (2020).

115. Fuller, E. J., Gabaly, F. E., Léonard, F., Agarwal, S., Plimpton, S. J., Jacobs-Gedrim, R. B., James, C. D., Marinella, M. J. & Talin, A. A. Li-Ion Synaptic Transistor for Low Power Analog Computing. *Advanced Materials* **29**, 1604310 (2017).

116. Van de Burgt, Y., Lubberman, E., Fuller, E. J., Keene, S. T., Faria, G. C., Agarwal, S., Marinella, M. J., Alec Talin, A. & Salleo, A. A Non-Volatile Organic Electrochemical Device as a Low-Voltage Artificial Synapse for Neuromorphic Computing. *Nature Materials* **16**, 414 (2017).

117. Tang, J. *et al. ECRAM as Scalable Synaptic Cell for High-Speed, Low-Power Neuromorphic Computing* in *2018 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2018), 13.1.1.

118. Li, Y., Fuller, E. J., Sugar, J. D., Yoo, S., Ashby, D. S., Bennett, C. H., Horton, R. D., Bartsch, M. S., Marinella, M. J., Lu, W. D. & Talin, A. A. Filament-Free Bulk Resistive Memory Enables Deterministic Analogue Switching. *Advanced Materials* **32**, 2003984 (2020).

119. Lequeux, S., Sampaio, J., Cros, V., Yakushiji, K., Fukushima, A., Matsumoto, R., Kubota, H., Yuasa, S. & Grollier, J. A Magnetic Synapse: Multilevel Spin-Torque Memristor With Perpendicular Anisotropy. *Scientific Reports* **6**, 31510 (2016).

120. Ostwal, V., Zand, R., DeMara, R. & Appenzeller, J. A Novel Compound Synapse Using Probabilistic Spin-Orbit-Torque Switching for MTJ-Based Deep Neural Networks. *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits* **5**, 182 (2019).

121. Mansueto, M., Chavent, A., Auffret, S., Joumard, I., Vila, L., Sousa, R. C., Buda-Prejbeanu, L. D., Prejbeanu, I. L. & Dieny, B. Spintronic Memristors for Neuromorphic Circuits Based on the Angular Variation of Tunnel Magnetoresistance. *Nanoscale* **13**, 11488 (2021).

122. Tallis, B. *Intel Announces Optane Memory M15: 3D XPoint on M.2 PCIe 3.0 X4* 2019.

123. Fong, S. W., Neumann, C. M. & Wong, H.-S. P. Phase-Change Memory—Towards a Storage-Class Memory. *IEEE Transactions on Electron Devices* **64**, 4374 (2017).

124. Zheng, Q., Wang, Y. & Zhu, J. Nanoscale Phase-Change Materials and Devices. *Journal of Physics D: Applied Physics* **50**, 243002 (2017).

125. Stanisavljevic, M., Pozidis, H., Athmanathan, A., Papandreou, N., Mittelholzer, T. & Eleftheriou, E. *Demonstration of Reliable Triple-Level-Cell (TLC) Phase-Change Memory* in *2016 IEEE 8th International Memory Workshop (IMW)* (IEEE, 2016), 1.

126. Sky, M. B., Sosa, N., Masuda, T., Kim, W., Kim, S., Ray, A., Bruce, R., Gonsalves, J., Zhu, Y., Suu, K. & Lam, C. Crystalline-as-Deposited ALD Phase Change Material Confined PCM Cell for High Density Storage Class Memory. *Technical Digest - International Electron Devices Meeting, IEDM* **2016-Febru**, 3.6.1 (2015).

127. Ding, K., Wang, J., Zhou, Y., Tian, H., Lu, L., Mazzarello, R., Jia, C., Zhang, W., Rao, F. & Ma, E. Phase-Change Heterostructure Enables Ultralow Noise and Drift for Memory Operation. *Science* **366**, 210 (2019).

128. Waser, R., Dittmann, R., Staikov, C. & Szot, K. Redox-Based Resistive Switching Memories Nanoionic Mechanisms, Prospects, and Challenges. *Advanced Materials* **21**, 2632 (2009).

129. Joshua Yang, J., Miao, F., Pickett, M. D., Ohlberg, D. A., Stewart, D. R., Lau, C. N. & Williams, R. S. The Mechanism of Electroforming of Metal Oxide Memristive Switches. *Nanotechnology* **20** (2009).

130. Moon, K., Fumarola, A., Sidler, S., Jang, J., Narayanan, P., Shelby, R. M., Burr, G. W. & Hwang, H. Bidirectional Non-Filamentary RRAM as an Analog Neuromorphic Synapse, Part I: Al/Mo/Pr 0.7 Ca 0.3 MnO 3 Material Improvements and Device Measurements. *IEEE Journal of the Electron Devices Society* **6**, 146 (2018).

131. Govoreanu, B. *et al. A-Vmco: A Novel Forming-Free, Self-Rectifying, Analog Memory Cell With Low-Current Operation, Nonfilamentary Switching and Excellent Variability* in *2015 Symposium on VLSI Technology (VLSI Technology)* **59** (IEEE, 2015), T132.

132. Zhuravlev, M. Y., Sabirianov, R. F., Jaswal, S. S. & Tsymbal, E. Y. Giant Electroresistance in Ferroelectric Tunnel Junctions. *Physical Review Letters* **94**, 246802 (2005).

133. Garcia, V. Ferroelectric Tunnel Junctions (2020).

134. Esaki, a. L., Laibowitz, R. B. & Stiles, P. J. Polar Switch. *IBM Tech. Discl. Bull* **13**, 114 (1971).

135. Gajek, M., Bibes, M., Fusil, S., Bouzehouane, K., Fontcuberta, J., Barthélémy, A. & Fert, A. Tunnel Junctions With Multiferroic Barriers. *Nature Materials* **6**, 296 (2007).

136. Chanthbouala, A. *et al.* Solid-State Memories Based on Ferroelectric Tunnel Junctions. *Nature Nanotechnology* **7**, 101 (2011).

137. Chanthbouala, A. *et al.* A Ferroelectric Memristor. *Nature Materials* **11**, 860 (2012).

138. Kim, S. J., Mohan, J., Lee, J., Lee, J. S., Lucero, A. T., Young, C. D., Colombo, L., Summerfelt, S. R., San, T. & Kim, J. Effect of Film Thickness on the Ferroelectric and Dielectric Properties of Low-Temperature (400 °C) $Hf_{0.5}Zr_{0.5}O_2$ Films. *Applied Physics Letters* **112**, 172902 (2018).

139. Ryu, H., Wu, H., Rao, F. & Zhu, W. Ferroelectric Tunneling Junctions Based on Aluminum Oxide/ Zirconium-Doped Hafnium Oxide for Neuromorphic Computing. *Scientific Reports* **9**, 20383 (2019).

140. Max, B., Hoffmann, M., Slesazeck, S. & Mikolajick, T. *Ferroelectric Tunnel Junctions Based on Ferroelectric-Dielectric $Hf_{0.5}Zr_{0.5}O_2/Al_2O_3$ Capacitor Stacks* in *2018 48th European Solid-State Device Research Conference (ESSDERC)* **2018-Septe** (IEEE, 2018), 142.

141. Max, B., Pešić, M., Slesazeck, S. & Mikolajick, T. Interplay Between Ferroelectric and Resistive Switching in Doped Crystalline HfO_2. *Journal of Applied Physics* **123**, 134102 (2018).

142. Luo, Y.-C., Hur, J., Wang, P., Khan, A. I. & Yu, S. *Modeling Multi-States in Ferroelectric Tunnel Junction* in *2020 Device Research Conference (DRC)* **2020-June** (IEEE, 2020), 1.

143. Zhou, D., Xu, J., Li, Q., Guan, Y., Cao, F., Dong, X., Müller, J., Schenk, T. & Schröder, U. Wake-Up Effects in Si-Doped Hafnium Oxide Ferroelectric Thin Films. *Applied Physics Letters* **103**, 192904 (2013).

144. Chouprik, A., Negrov, D., Tsymbal, E. Y. & Zenkevich, A. Defects in Ferroelectric HfO_2. *Nanoscale* **13**, 11635 (2021).

145. Begon-Lours, L., **Halter**, **M.**, Pineda, D. D., Bragaglia, V., Popoff, Y., la Porta, A., Jubin, D., Fompeyrine, J. & Offrein, B. J. *A Back-End-of-Line Compatible, Ferroelectric Analog Non-Volatile Memory* in *2021 IEEE International Memory Workshop (IMW)* **4** (IEEE, 2021), 1.

146. Begon-Lours, L., **Halter**, **M.**, Popoff, Y., Yu, Z., Falcone, D. F. & Offrein, B. J. *High-Conductance, Ohmic-Like $HfZrO_4$ Ferroelectric Memristor* in *ESSCIRC 2021 - IEEE 47th European Solid State Circuits Conference (ESSCIRC)* (IEEE, 2021), 87.

147. Wu, T.-Y. *et al. Sub-nA Low-Current HZO Ferroelectric Tunnel Junction for High-Performance and Accurate Deep Learning Acceleration* in *2019 IEEE International Electron Devices Meeting (IEDM)* **2019-Decem** (IEEE, 2019), 6.3.1.

148. Goh, Y., Hwang, J., Lee, Y., Kim, M. & Jeon, S. Ultra-Thin $Hf_{0.5}Zr_{0.5}O_2$ Thin-Film-Based Ferroelectric Tunnel Junction via Stress Induced Crystallization. *Applied Physics Letters* **117**, 242901 (2020).

149. Shu-Yau Wu. A New Ferroelectric Memory Device, Metal-Ferroelectric-Semiconductor Transistor. *IEEE Transactions on Electron Devices* **21**, 499 (1974).

150. Sakai, S. & Ilangovan, R. Metal-Ferroelectric-Insulator-Semiconductor Memory FET With Long Retention and High Endurance. *IEEE Electron Device Letters* **25**, 369 (2004).

151. Trentzsch, M. *et al. A 28nm HKMG Super Low Power Embedded NVM Technology Based on Ferroelectric FETs* in *2016 IEEE International Electron Devices Meeting (IEDM)* **63** (IEEE, 2016), 11.5.1.

152. Mulaosmanovic, H. *et al. Evidence of Single Domain Switching in Hafnium Oxide Based FeFETs: Enabler for Multi-Level FeFET Memory Cells* in *2015 IEEE International Electron Devices Meeting (IEDM)* **2016-Febru** (IEEE, 2015), 26.8.1.

153. Yurchuk, E. *et al.* Origin of the Endurance Degradation in the Novel HfO_2-based 1T Ferroelectric Non-Volatile Memories. *IEEE International Reliability Physics Symposium Proceedings* **2**, 1 (2014).

154. Mulaosmanovic, H., Breyer, E. T., Mikolajick, T. & Slesazeck, S. Ferroelectric FETs With 20-Nm-Thick HfO_2 Layer for Large Memory Window and High Performance. *IEEE Transactions on Electron Devices* **66**, 3828 (2019).

155. Yurchuk, E., Muller, J., Muller, S., Paul, J., Pesic, M., Van Bentum, R., Schroeder, U. & Mikolajick, T. Charge-Trapping Phenomena in HfO_2-Based FeFET-Type Nonvolatile Memories. *IEEE Transactions on Electron Devices* **63**, 3501 (2016).

156. Bae, J. H., Kwon, D., Jeon, N., Cheema, S., Tan, A. J., Hu, C. & Salahuddin, S. Highly Scaled, High Endurance, Ω-Gate, Nanowire Ferroelectric FET Memory Transistors. *IEEE Electron Device Letters* **41**, 1637 (2020).

157. Kim, M.-K. & Lee, J.-S. Ferroelectric Analog Synaptic Transistors. *Nano Letters* **19**, 2044 (2019).

158. Pešić, M., Künneth, C., Hoffmann, M., Mulaosmanovic, H., Müller, S., Breyer, E. T., Schroeder, U., Kersch, A., Mikolajick, T. & Slesazeck, S. A Computational Study of Hafnia-Based Ferroelectric Memories: From Ab Initio via Physical Modeling to Circuit Models of Ferroelectric Device. *Journal of Computational Electronics* **16**, 1236 (2017).

159. Ihlefeld, J. F. in *Ferroelectricity in Doped Hafnium Oxide: Materials, Properties and Devices* 1 (Elsevier, 2019).

160. Cheema, S. S. *et al.* Enhanced Ferroelectricity in Ultrathin Films Grown Directly on Silicon. *Nature* **580**, 478 (2020).

161. Müller, J., Böscke, T. S., Bräuhaus, D., Schröder, U., Böttger, U., Sundqvist, J., Kücher, P., Mikolajick, T. & Frey, L. Ferroelectric Zr 0.5 Hf 0.5 O 2 Thin Films for Nonvolatile Memory Applications. *Applied Physics Letters* **99**, 112901 (2011).

162. Hyuk Park, M., Joon Kim, H., Jin Kim, Y., Lee, W., Moon, T. & Seong Hwang, C. Evolution of Phases and Ferroelectric Properties of Thin $Hf_{0.5}Zr_{0.5}O_2$ Films According to the Thickness and Annealing Temperature. *Applied Physics Letters* **102**, 242905 (2013).

163. Takahashi, M. & Sakai, S. Area-Scalable 109-Cycle-High-Endurance FeFET of Strontium Bismuth Tantalate Using a Dummy-Gate Process. *Nanomaterials* **11**, 101 (2021).

164. Xu, L., Nishimura, T., Shibayama, S., Yajima, T., Migita, S. & Toriumi, A. Ferroelectric Phase Stabilization of HfO_2 by Nitrogen Doping. *Applied Physics Express* **9**, 091501 (2016).

165. Schenk, T., Mueller, S., Schroeder, U., Materlik, R., Kersch, A., Popovici, M., Adelmann, C., Van Elshocht, S. & Mikolajick, T. *Strontium Doped Hafnium Oxide Thin Films: Wide Process Window for Ferroelectric Memories* in *2013 Proceedings of the European Solid-State Device Research Conference (ESSDERC)* (IEEE, 2013), 260.

166. Muller, J. *et al. Ferroelectric Hafnium Oxide: A CMOS-compatible and Highly Scalable Approach to Future Ferroelectric Memories* in *2013 IEEE International Electron Devices Meeting* (IEEE, 2013), 10.8.1.

167. Mueller, S., Mueller, J., Singh, A., Riedel, S., Sundqvist, J., Schroeder, U. & Mikolajick, T. Incipient Ferroelectricity in Al-Doped HfO_2 Thin Films. *Advanced Functional Materials* **22**, 2412 (2012).

168. Mueller, S., Adelmann, C., Singh, A., Van Elshocht, S., Schroeder, U. & Mikolajick, T. Ferroelectricity in Gd-Doped HfO_2 Thin Films. *ECS Journal of Solid State Science and Technology* **1**, N123 (2012).

169. Olsen, T., Schröder, U., Müller, S., Krause, A., Martin, D., Singh, A., Müller, J., Geidel, M. & Mikolajick, T. Co-Sputtering Yttrium Into Hafnium Oxide Thin Films to Produce Ferroelectric Properties. *Applied Physics Letters* **101**, 082905 (2012).

170. Xu, L., Shibayama, S., Izukashi, K., Nishimura, T., Yajima, T., Migita, S. & Toriumi, A. *General Relationship for Cation and Anion Doping Effects on Ferroelectric HfO<inf>2</inf> Formation* in *2016 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2016), 25.2.1.

171. Park, M. H., Schenk, T., Fancher, C. M., Grimley, E. D., Zhou, C., Richter, C., LeBeau, J. M., Jones, J. L., Mikolajick, T. & Schroeder, U. A Comprehensive Study on the Structural Evolution of HfO_2 Thin Films Doped With Various Dopants. *Journal of Materials Chemistry C* **5**, 4677 (2017).

172. Xu, L., Nishimura, T., Shibayama, S., Yajima, T., Migita, S. & Toriumi, A. Kinetic Pathway of the Ferroelectric Phase Formation in Doped HfO_2 Films. *Journal of Applied Physics* **122**, 124104 (2017).

173. Müller, J., Böscke, T. S., Schröder, U., Mueller, S., Bräuhaus, D., Böttger, U., Frey, L. & Mikolajick, T. Ferroelectricity in Simple Binary ZrO_2 and HfO_2. *Nano Letters* **12**, 4318 (2012).

174. Kim, H. J., Park, M. H., Kim, Y. J., Lee, Y. H., Moon, T., Kim, K. D., Hyun, S. D. & Hwang, C. S. A Study on the Wake-Up Effect of Ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ Films by Pulse-Switching Measurement. *Nanoscale* **8**, 1383 (2016).

175. Materlik, R., Künneth, C. & Kersch, A. The Origin of Ferroelectricity in $Hf_{1-x}Zr_xO_2$: A Computational Investigation and a Surface Energy Model. *Journal of Applied Physics* **117**, 134109 (2015).

176. Momma, K. & Izumi, F. VESTA 3 for Three-Dimensional Visualization of Crystal, Volumetric and Morphology Data. *Journal of Applied Crystallography* **44**, 1272 (2011).

177. Ohtaka, O., Fukui, H., Kunisada, T., Fujisawa, T., Funakoshi, K., Utsumi, W., Irifune, T., Kuroda, K. & Kikegawa, T. Phase Relations and Equations of State of ZrO_2 Under High Temperature and High Pressure. *Physical Review B* **63**, 174108 (2001).

178. Ohtaka, O., Fukui, H., Kunisada, T., Fujisawa, T., Funakoshi, K., Utsumi, W., Irifune, T., Kuroda, K. & Kikegawa, T. Phase Relations and Volume Changes of Hafnia Under High Pressure and High Temperature. *Journal of the American Ceramic Society* **84**, 1369 (2004).

179. Kisi, E. H. & Howard, C. Crystal Structures of Zirconia Phases and Their Inter-Relation. *Key Engineering Materials* **153-154**, 1 (1998).

180. Tomida, K., Kita, K. & Toriumi, A. Dielectric Constant Enhancement Due to Si Incorporation Into HfO_2. *Applied Physics Letters* **89**, 142902 (2006).

181. Lee, C.-K., Cho, E., Lee, H.-S., Hwang, C. S. & Han, S. First-Principles Study on Doping and Phase Stability of HfO_2. *Physical Review B* **78**, 012102 (2008).

182. Garvie, R. C. The Occurrence of Metastable Tetragonal Zirconia as a Crystallite Size Effect. *The Journal of Physical Chemistry* **69**, 1238 (1965).

183. Navrotsky, A. Thermochemical Insights Into Refractory Ceramic Materials Based on Oxides With Large Tetravalent Cations. *Journal of Materials Chemistry* **15**, 1883 (2005).

184. Shandalov, M. & McIntyre, P. C. Size-Dependent Polymorphism in HfO_2 Nanotubes and Nanoscale Thin Films. *Journal of Applied Physics* **106**, 084322 (2009).

185. Sharma, G., Ushakov, S. V. & Navrotsky, A. Size Driven Thermodynamic Crossovers in Phase Stability in Zirconia and Hafnia. *Journal of the American Ceramic Society* **101**, 31 (2018).

186. Al-Khatatbeh, Y., Lee, K. K. M. & Kiefer, B. Phase Relations and Hardness Trends of ZrO_2 Phases at High Pressure. *Physical Review B* **81**, 214102 (2010).

187. Al-Khatatbeh, Y., Lee, K. K. M. & Kiefer, B. Phase Diagram Up to 105 GPa and Mechanical Strength of HfO_2. *Physical Review B* **82**, 144106 (2010).

188. Haines, J., Léger, J. M. & Atouf, A. Crystal Structure and Equation of State of Cotunnite-Type Zirconia. *Journal of the American Ceramic Society* **78**, 445 (1995).

189. Sakuma, T., Yoshizawa, Y.-I. & Suto, H. The Rhombohedral Phase Produced in Partially-Stabilized Zirconia. *Journal of Materials Science Letters* **4**, 29 (1985).

190. Burke, D. P. & Rainforth, W. M. Intermediate Rhombohedral (R-ZrO_2) Phase Formation at the Surface of Sintered Y-TZP's. *Journal of Materials Science Letters* **16**, 883 (1997).

191. Wei, Y. *et al.* A Rhombohedral Ferroelectric Phase in Epitaxially Strained $Hf_{0.5}Zr_{0.5}O_2$ Thin Films. *Nature Materials* **17**, 1095 (2018).

192. Shimizu, T., Katayama, K., Kiguchi, T., Akama, A., Konno, T. J., Sakata, O. & Funakubo, H. The Demonstration of Significant Ferroelectricity in Epitaxial Y-Doped HfO_2 Film. *Scientific Reports* **6**, 32931 (2016).

193. Huan, T. D., Sharma, V., Rossetti, G. A. & Ramprasad, R. Pathways Towards Ferroelectricity in Hafnia. *Physical Review B* **90**, 064111 (2014).

194. Sang, X., Grimley, E. D., Schenk, T., Schroeder, U. & LeBeau, J. M. On the Structural Origins of Ferroelectricity in HfO_2 Thin Films. *Applied Physics Letters* **106**, 162905 (2015).

195. Materlik, R., Künneth, C. & Kersch, A. The Origin of Ferroelectricity in Hf 1-x Zr X O 2 : A Computational Investigation and a Surface Energy Model. *Journal of Applied Physics* **117**, 134109 (2015).

196. Hoffmann, M. *et al.* Stabilizing the Ferroelectric Phase in Doped Hafnium Oxide. *Journal of Applied Physics* **118**, 072006 (2015).

197. Xu, L., Nishimura, T., Shibayama, S., Yajima, T., Migita, S. & Toriumi, A. Kinetic Pathway of the Ferroelectric Phase Formation in Doped HfO_2 Films. *Journal of Applied Physics* **122**, 124104 (2017).

198. Künneth, C., Materlik, R., Falkowski, M. & Kersch, A. Impact of Four-Valent Doping on the Crystallographic Phase Formation for Ferroelectric HfO_2 From First-Principles: Implications for Ferroelectric Memory and Energy-Related Applications. *ACS Applied Nano Materials* **1**, 254 (2018).

199. Lin, Y.-J., Teng, C.-Y., Chang, S.-J., Liao, Y.-F., Hu, C., Su, C.-J. & Tseng, Y.-C. Role of Electrode-Induced Oxygen Vacancies in Regulating Polarization Wake-Up in Ferroelectric Capacitors. *Applied Surface Science* **528**, 147014 (2020).

200. Kim, H. J., Park, M. H., Kim, Y. J., Lee, Y. H., Jeon, W., Gwon, T., Moon, T., Kim, K. D. & Hwang, C. S. Grain Size Engineering for Ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ Films by an Insertion of Al_2O_3 Interlayer. *Applied Physics Letters* **105**, 192903 (2014).

201. Hyuk Park, M., Joon Kim, H., Jin Kim, Y., Moon, T. & Seong Hwang, C. The Effects of Crystallographic Orientation and Strain of Thin $Hf_{0.5}Zr_{0.5}O_2$ Film on Its Ferroelectricity. *Applied Physics Letters* **104**, 072901 (2014).

202. Park, M. H., Lee, Y. H., Kim, H. J., Schenk, T., Lee, W., Kim, K. D., Fengler, F. P. G., Mikolajick, T., Schroeder, U. & Hwang, C. S. Surface and Grain Boundary Energy as the Key Enabler of Ferroelectricity in Nanoscale Hafnia-Zirconia: A Comparison of Model and Experiment. *Nanoscale* **9**, 9973 (2017).

203. Shiraishi, T., Katayama, K., Yokouchi, T., Shimizu, T., Oikawa, T., Sakata, O., Uchida, H., Imai, Y., Kiguchi, T., Konno, T. J. & Funakubo, H. Impact of Mechanical Stress on Ferroelectricity in (Hf 0.5 Zr 0.5)O 2 Thin Films. *Applied Physics Letters* **108**, 262904 (2016).

204. Kim, S. J. *et al.* Large Ferroelectric Polarization of TiN/$Hf_{0.5}Zr_{0.5}O_2$ /TiN Capacitors Due to Stress-Induced Crystallization at Low Thermal Budget. *Applied Physics Letters* **111**, 242901 (2017).

205. Riedel, S., Polakowski, P. & Müller, J. A Thermally Robust and Thickness Independent Ferroelectric Phase in Laminated Hafnium Zirconium Oxide. *AIP Advances* **6**, 095123 (2016).

206. Lu, C.-H., Raitano, J. M., Khalid, S., Zhang, L. & Chan, S.-W. Cubic Phase Stabilization in Nanoparticles of Hafnia-Zirconia Oxides: Particle-Size and Annealing Environment Effects. *Journal of Applied Physics* **103**, 124303 (2008).

207. Kisi, E. H., Howard, C. J. & Hill, R. J. Crystal Structure of Orthorhombic Zirconia in Partially Stabilized Zirconia. *Journal of the American Ceramic Society* **72**, 1757 (1989).

208. Goedecker, S. Minima Hopping: An Efficient Search Method for the Global Minimum of the Potential Energy Surface of Complex Molecular Systems. *The Journal of Chemical Physics* **120**, 9911 (2004).

209. Shuvalov, L. A. Symmetry Aspects of Ferroelectricity. *J. Phys. Soc. Japan* **28**, 38 (1970).

210. Barabash, S. V., Pramanik, D., Zhai, Y., Magyari-Kope, B. & Nishi, Y. Ferroelectric Switching Pathways and Energetics in (Hf,Zr)O 2. *ECS Transactions* **75**, 107 (2017).

211. Yashima, M. & Tsunekawa, S. Structures and the Oxygen Deficiency of Tetragonal and Monoclinic Zirconium Oxide Nanoparticles. *Acta Crystallographica Section B Structural Science* **62**, 161 (2006).

212. Batra, R., Tran, H. D. & Ramprasad, R. Stabilization of Metastable Phases in Hafnia Owing to Surface Energy Effects. *Applied Physics Letters* **108**, 172902 (2016).

213. Park, M. H., Lee, Y. H., Kim, H. J., Kim, Y. J., Moon, T., Kim, K. D., Hyun, S. D., Mikolajick, T., Schroeder, U. & Hwang, C. S. Understanding the Formation of the Metastable Ferroelectric Phase in Hafnia-Zirconia Solid Solution Thin Films. *Nanoscale* **10**, 716 (2018).

214. Fengler, F., Park, M. H., Schenk, T., Pešić, M. & Schroeder, U. in *Ferroelectricity in Doped Hafnium Oxide: Materials, Properties and Devices* 381 (Elsevier, 2019).

215. Schenk, T., Schroeder, U., Pešić, M., Popovici, M., Pershin, Y. V. & Mikolajick, T. Electric Field Cycling Behavior of Ferroelectric Hafnium Oxide. *ACS Applied Materials & Interfaces* **6**, 19744 (2014).

216. Schenk, T., Hoffmann, M., Ocker, J., Pešić, M., Mikolajick, T. & Schroeder, U. Complex Internal Bias Fields in Ferroelectric Hafnium Oxide. *ACS Applied Materials and Interfaces* **7**, 20224 (2015).

217. Pešić, M., Fengler, F. P. G., Larcher, L., Padovani, A., Schenk, T., Grimley, E. D., Sang, X., LeBeau, J. M., Slesazeck, S., Schroeder, U. & Mikolajick, T. Physical Mechanisms Behind the Field-Cycling Behavior of HfO_2 -Based Ferroelectric Capacitors. *Advanced Functional Materials* **26**, 4601 (2016).

218. Grimley, E. D., Schenk, T., Sang, X., Pešić, M., Schroeder, U., Mikolajick, T. & LeBeau, J. M. Structural Changes Underlying Field-Cycling Phenomena in Ferroelectric HfO_2 Thin Films. *Advanced Electronic Materials* **2** (2016).

219. Matveyev, Y., Negrov, D., Chernikova, A., Lebedinskii, Y., Kirtaev, R., Zarubin, S., Suvorova, E., Gloskovskii, A. & Zenkevich, A. Effect of Polarization Reversal in Ferroelectric TiN/$Hf_{0.5}Zr_{0.5}O_2$ /TiN Devices on Electronic Conditions at Interfaces Studied in Operando by Hard X-Ray Photoemission Spectroscopy. *ACS Applied Materials & Interfaces* **9**, 43370 (2017).

220. Starschich, S., Menzel, S. & Böttger, U. Evidence for Oxygen Vacancies Movement During Wake-Up in Ferroelectric Hafnium Oxide. *Applied Physics Letters* **108** (2016).

221. Batra, R., Huan, T. D., Jones, J. L., Rossetti, G. & Ramprasad, R. Factors Favoring Ferroelectricity in Hafnia: A First-Principles Computational Study. *Journal of Physical Chemistry C* **121**, 4139 (2017).

222. Weeks, S. L., Pal, A., Narasimhan, V. K., Littau, K. A. & Chiang, T. Engineering of Ferroelectric HfO_2-ZrO_2 Nanolaminates. *ACS Applied Materials and Interfaces* **9**, 13440 (2017).

223. Fengler, F. P. G., Pesic, M., Starschich, S., Schneller, T., Bottger, U., Schenk, T., Park, M. H., Mikolajick, T. & Schroeder, U. *Comparison of Hafnia and PZT Based Ferroelectrics for Future Non-Volatile FRAM Applications* in *2016 46th European Solid-State Device Research Conference (ESSDERC)* (IEEE, 2016), 369.

224. Kashir, A., Kim, H., Oh, S. & Hwang, H. Large Remnant Polarization in a Wake-Up Free $Hf_{0.5}Zr_{0.5}O_2$ Ferroelectric Film Through Bulk and Interface Engineering. *ACS Applied Electronic Materials* **3**, 629 (2021).

225. Masuduzzaman, M. & Alam, M. A. *Hot Atom Damage (HAD) Limited TDDB Lifetime of Ferroelectric Memories* in *2013 IEEE International Electron Devices Meeting* (IEEE, 2013), 21.4.1.

226. McPherson, J. W. & Mogul, H. C. Underlying Physics of the Thermochemical E Model in Describing Low-Field Time-Dependent Dielectric Breakdown in SiO2 Thin Films. *Journal of Applied Physics* **84**, 1513 (1998).

227. Mannequin, C., Gonon, P., Vallée, C., Latu-Romain, L., Bsiesy, A., Grampeix, H., Salaün, A. & Jousseaume, V. Stress-Induced Leakage Current and Trap Generation in HfO_2 Thin Films. *Journal of Applied Physics* **112** (2012).

228. Huang, F. *et al.* Fatigue Mechanism of Yttrium-Doped Hafnium Oxide Ferroelectric Thin Films Fabricated by Pulsed Laser Deposition. *Physical Chemistry Chemical Physics* **19**, 3486 (2017).

229. Lanza, M., Bersuker, G., Porti, M., Miranda, E., Nafría, M. & Aymerich, X. Resistive Switching in Hafnium Dioxide Layers: Local Phenomenon at Grain Boundaries. *Applied Physics Letters* **101**, 193502 (2012).

230. Lee, D. H. *et al.* Domains and Domain Dynamics in Fluorite-Structured Ferroelectrics. *Applied Physics Reviews* **8**, 021312 (2021).

231. Hubbell, J. & Seltzer, S. *Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients (Version 1.4)* 2004.

232. Rumble, J. R. in *CRC Handbook of Chemistry and Physics* 102nd ed. Chap. 4 (CRC Press/Taylor & Francis, Boca Raton, FL, 2021).

233. Berger, L. I. in *CRC Handbook of Chemistry and Physics* 102nd ed. Chap. 12 (CRC Press/Taylor & Francis, Boca Raton, FL, 2021).

234. Birkholz, M. Modelling of Diffraction From Fibre Texture Gradients in Thin Polycrystalline Films. *Journal of Applied Crystallography* **40**, 735 (2007).

235. Stephens, P. W. in *International Tables for Crystallography* (eds Gilmore, C. J., Kaduk, J. A. & Schenk, H.) 252 (International Union of Crystallography, Chester, England, 2019).

236. Gražulis, S., Daškevič, A., Merkys, A., Chateigner, D., Lutterotti, L., Quirós, M., Serebryanaya, N. R., Moeck, P., Downs, R. T. & Le Bail, A. Crystallography Open Database (COD): An Open-Access Collection of Crystal Structures and Platform for World-Wide Collaboration. *Nucleic Acids Research* **40**, D420 (2012).

237. Birkholz, M. in *International table of crystalography* 581 (2019).

238. Johnson, R. W., Hultqvist, A. & Bent, S. F. A Brief Review of Atomic Layer Deposition: From Fundamentals to Applications. *Materials Today* **17**, 236 (2014).

239. Auth, C. *et al.* A 22nm High Performance and Low-Power CMOS Technology Featuring Fully-Depleted Tri-Gate Transistors, Self-Aligned Contacts and High Density MIM Capacitors. *Digest of Technical Papers - Symposium on VLSI Technology* **980**, 131 (2012).

240. Skorupa, W., Yankov, R., Voelskow, M., Anwand, W., Panknin, D., McMahon, R., Smith, M., Gebel, T., Rebohle, L., Fendler, R. & Hentsch, W. *Advanced Thermal Processing of Semiconductor Materials in the MSEC-Range* in *2005 13th International Conference on Advanced Thermal Processing of Semiconductors* (IEEE, 2005), 53.

241. Schroder, D. K. *Semiconductor Material and Device Characterization* 3rd ed., 1 (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2005).

242. Fan, Z., Xiao, J., Wang, J., Zhang, L., Deng, J., Liu, Z., Dong, Z., Wang, J. & Chen, J. Ferroelectricity and Ferroelectric Resistive Switching in Sputtered $Hf_{0.5}Zr_{0.5}O_2$ Thin Films. *Applied Physics Letters* **108**, 232905 (2016).

243. Nukala, P., Antoja-Lleonart, J., Wei, Y., Yedra, L., Dkhil, B. & Noheda, B. Direct Epitaxial Growth of Polar (1 - X)HfO_2 -(X)ZrO_2 Ultrathin Films on Silicon. *ACS Applied Electronic Materials* **1**, 2585 (2019).

244. Zacharaki, C., Tsipas, P., Chaitoglou, S., Fragkos, S., Axiotis, M., Lagoyiannis, A., Negrea, R., Pintilie, L. & Dimoulas, A. Very Large Remanent Polarization in Ferroelectric Hf 1-X Zr X O 2 Grown on Ge Substrates by Plasma Assisted Atomic Oxygen Deposition. *Applied Physics Letters* **114**, 112901 (2019).

245. Zacharaki, C., Tsipas, P., Chaitoglou, S., Bégon-Lours, L., **Halter, M.** & Dimoulas, A. Reliability Aspects of Ferroelectric TiN/$Hf_{0.5}Zr_{0.5}O_2$ /Ge Capacitors Grown by Plasma Assisted Atomic Oxygen Deposition. *Applied Physics Letters* **117**, 212905 (2020).

246. Park, M. H., Kim, H. J., Kim, Y. J., Lee, W., Moon, T., Kim, K. D. & Hwang, C. S. Study on the Degradation Mechanism of the Ferroelectric Properties of Thin $Hf_{0.5}Zr_{0.5}O_2$ Films on TiN and Ir Electrodes. *Applied Physics Letters* **105**, 072902 (2014).

247. Schroeder, U., Yurchuk, E., Müller, J., Martin, D., Schenk, T., Polakowski, P., Adelmann, C., Popovici, M. I., Kalinin, S. V. & Mikolajick, T. Impact of Different Dopants on the Switching Properties of Ferroelectric Hafniumoxide. *Japanese Journal of Applied Physics* **53**, 08LE02 (2014).

248. Carl, K. & Hardtl, K. H. Electrical After-Effects in Pb(Ti, Zr)O3 Ceramics. *Ferroelectrics* **17**, 473 (1977).

249. Grossmann, M., Lohse, O., Bolten, D., Boettger, U., Schneller, T. & Waser, R. The Interface Screening Model as Origin of Imprint in PbZrxTi1-xO3 Thin Films. I. Dopant, Illumination, and Bias Dependence. *Journal of Applied Physics* **92**, 2680 (2002).

250. Le Rhun, G., Bouregba, R. & Poullain, G. Polarization Loop Deformations of an Oxygen Deficient Pb(Zr_0.25,Ti_0.75)O_3 Ferroelectric Thin Film. *Journal of Applied Physics* **96**, 5712 (2004).

251. Koval, V., Viola, G. & Tan, Y. in *Ferroelectric Materials - Synthesis and Characterization* tourism, 13 (InTech, 2015).

252. Rabe, K. M., Dawber, M., Lichtensteiger, C., Ahn, C. H. & Triscone, J.-M. in *Physics of Ferroelectrics* 1 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007).

253. Park, M. H., Kim, H. J., Kim, Y. J., Lee, Y. H., Moon, T., Kim, K. D., Hyun, S. D. & Hwang, C. S. Study on the Size Effect in Hf$_0$.5Zr$_0$.5O$_2$ Films Thinner Than 8 Nm Before and After Wake-Up Field Cycling. *Applied Physics Letters* **107**, 192907 (2015).

254. Lomenzo, P. D., Takmeel, Q., Zhou, C., Fancher, C. M., Lambers, E., Rudawski, N. G., Jones, J. L., Moghaddam, S. & Nishida, T. TaN Interface Properties and Electric Field Cycling Effects on Ferroelectric Si-Doped HfO$_2$ Thin Films. *Journal of Applied Physics* **117**, 134105 (2015).

255. Pesic, M., Fengler, F. P. G., Slesazeck, S., Schroeder, U., Mikolajick, T., Larcher, L. & Padovani, A. *Root Cause of Degradation in Novel HfO$_2$-based Ferroelectric Memories* in *2016 IEEE International Reliability Physics Symposium (IRPS)* **2016-Septe** (IEEE, 2016), MY–3–1–MY–3.

256. Damjanovic, D. in *The Science of Hysteresis* 337 (Elsevier, 2006).

257. Chernikova, A., Kozodaev, M., Markeev, A., Matveev, Y., Negrov, D. & Orlov, O. Confinement-Free Annealing Induced Ferroelectricity in Hf$_0$.5Zr$_0$.5O$_2$ Thin Films. *Microelectronic Engineering* **147**, 15 (2015).

258. Mueller, S., Mueller, J., Singh, A., Riedel, S., Sundqvist, J., Schroeder, U. & Mikolajick, T. Incipient Ferroelectricity in Al-Doped HfO$_2$ Thin Films. *Advanced Functional Materials* **22**, 2412 (2012).

259. Chen, K. T., Liao, C. Y., Lo, C., Chen, H. Y., Siang, G. Y., Liu, S., Chang, S. C., Liao, M. H., Chang, S. T. & Lee, M. H. Improvement on Ferroelectricity and Endurance of Ultra-Thin HfZrO$_2$ Capacitor With Molybdenum Capping Electrode. *2019 Electron Devices Technology and Manufacturing Conference, EDTM 2019*, 62 (2019).

260. Karbasian, G., dos Reis, R., Yadav, A. K., Tan, A. J., Hu, C. & Salahuddin, S. Stabilization of Ferroelectric Phase in Tungsten Capped Hf 0.8 Zr 0.2 O 2. *Applied Physics Letters* **111**, 022907 (2017).

261. Cao, R., Wang, Y., Zhao, S., Yang, Y., Zhao, X., Wang, W., Zhang, X., Lv, H., Liu, Q. & Liu, M. Effects of Capping Electrode on Ferroelectric Properties of Hf$_0$.5Zr$_0$.5O$_2$ Thin Films. *IEEE Electron Device Letters* **39**, 1207 (2018).

262. Goh, Y. & Jeon, S. The Effect of the Bottom Electrode on Ferroelectric Tunnel Junctions Based on CMOS-compatible HfO$_2$. *Nanotechnology* **29**, 335201 (2018).

263. Shimizu, T., Yokouchi, T., Shiraishi, T., Oikawa, T., Krishnan, P. S. S. R. & Funakubo, H. Study on the Effect of Heat Treatment Conditions on Metalorganic-Chemical-Vapor-Deposited Ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ Thin Film on Ir Electrode. *Japanese Journal of Applied Physics* **53**, 09PA04 (2014).

264. Park, M. H., Kim, H. J., Kim, Y. J., Jeon, W., Moon, T. & Hwang, C. S. Ferroelectric Properties and Switching Endurance of $Hf_{0.5}Zr_{0.5}O_2$ Films on TiN Bottom and TiN or RuO2 Top Electrodes. *Physica Status Solidi - Rapid Research Letters* **8**, 532 (2014).

265. Aigner, K., Lengauer, W., Rafaja, D. & Ettmayer, P. Lattice Parameters and Thermal Expansion of Ti(C_xN_{1-x}), Zr(C_xN_{1-x}), Hf(C_xN_{1-x}) and TiN_{1-x} From 298 to 1473 K as Investigated by High-Temperature X-Ray Diffraction. *Journal of Alloys and Compounds* **215**, 121 (1994).

266. Keller, H.-L. Darstellung Und Kristallstruktur Von Hoch-Tl3PbBr5. *Journal of the Less Common Metals* **78**, 281 (1981).

267. Aird, A., Domeneghetti, M. C., Mazzi, F., Tazzoli, V. & Salje, E. K. H. Sheet Superconductivity in : Crystal Structure of the Tetragonal Matrix. *Journal of Physics: Condensed Matter* **10**, L569 (1998).

268. Lee, H., Kim, T. H., Patzner, J. J., Lu, H., Lee, J. W., Zhou, H., Chang, W., Mahanthappa, M. K., Tsymbal, E. Y., Gruverman, A. & Eom, C. B. Imprint Control of BaTiO3 Thin Films via Chemically Induced Surface Polarization Pinning. *Nano Letters* **16**, 2400 (2016).

269. Ihlefeld, J. F., Harris, D. T., Keech, R., Jones, J. L., Maria, J.-P. & Trolier-McKinstry, S. Scaling Effects in Perovskite Ferroelectrics: Fundamental Limits and Process-Structure-Property Relations. *Journal of the American Ceramic Society* **99**, 2537 (2016).

270. Williams, D. E., Aliwell, S. R., Pratt, K. F. E., Caruana, D. J., Jones, R. L., Cox, R. A., Hansford, G. M. & Halsall, J. Modelling the Response of a Tungsten Oxide Semiconductor as a Gas Sensor for the Measurement of Ozone. *Measurement Science and Technology* **13**, 314 (2002).

271. Lee, W. J., Fang, Y. K., Ho, J.-J., Hsieh, W. T., Ting, S. F., Huang, D. & Ho, F. C. Effects of Surface Porosity on Tungsten Trioxide(WO3) Films' Electrochromic Performance. *Journal of Electronic Materials* **29**, 183 (2000).

272. Guo, J., Guo, X., Sun, H., Xie, Y., Diao, X., Wang, M., Zeng, X. & Zhang, Z.-B. Unprecedented Electrochromic Stability of a-WO_{3-x} Thin Films Achieved by Using a Hybrid-Cationic Electrolyte. *ACS Applied Materials & Interfaces* **13**, 11067 (2021).

273. Qu, B., Younis, A. & Chu, D. Recent Progress in Tungsten Oxides Based Memristors and Their Neuromorphological Applications. *Electronic Materials Letters* **12**, 715 (2016).

274. Wheeler, D., Alvarado-Rodriguez, I., Elliott, K., Kally, J., Hermiz, J., Hunt, H., Hussain, T. & Srinivasa, N. *Fabrication and Characterization of Tungsten-Oxide-Based Memristors for Neuromorphic Circuits* in *2014 14th International Workshop on Cellular Nanoscale Networks and their Applications (CNNA)* (IEEE, 2014), 1.

275. Qian, K., Cai, G., Nguyen, V. C., Chen, T. & Lee, P. S. Direct Observation of Conducting Filaments in Tungsten Oxide Based Transparent Resistive Switching Memory. *ACS Applied Materials and Interfaces* **8**, 27885 (2016).

276. Yang, R., Terabe, K., Liu, G., Tsuruoka, T., Hasegawa, T., Gimzewski, J. K. & Aono, M. On-Demand Nanodevice With Electrical and Neuromorphic Multifunction Realized by Local Ion Migration. *ACS Nano* **6**, 9515 (2012).

277. Hong, D. S., Chen, Y. S., Li, Y., Yang, H. W., Wei, L. L., Shen, B. G. & Sun, J. R. Evolution of Conduction Channel and Its Effect on Resistance Switching for Au-Wo3-X-Au Devices. *Scientific Reports* **4**, 1 (2014).

278. Qu, B., Du, H., Wan, T., Lin, X., Younis, A. & Chu, D. Synaptic Plasticity and Learning Behavior in Transparent Tungsten Oxide-Based Memristors. *Materials & Design* **129**, 173 (2017).

279. Liang, L. *et al.* Vacancy Associates-Rich Ultrathin Nanosheets for High Performance and Flexible Nonvolatile Memory Device. *Journal of the American Chemical Society* **137**, 3102 (2015).

280. Chang, T., Sheridan, P. & Lu, W. *Modeling and Implementation of Oxide Memristors for Neuromorphic Applications* in *2012 13th International Workshop on Cellular Nanoscale Networks and their Applications* (IEEE, 2012), 1.

281. He, X., Yin, Y., Guo, J., Yuan, H., Peng, Y., Zhou, Y., Zhao, D., Hai, K., Zhou, W. & Tang, D. Memristive Properties of Hexagonal WO3 Nanowires Induced by Oxygen Vacancy Migration. *Nanoscale Research Letters* **8**, 50 (2013).

282. Tsai, T.-l., Lin, Y.-h. & Tseng, T.-y. Resistive Switching Characteristics of WO_3/ZrO_2 Structure With Forming-Free, Self-Compliance, and Submicroampere Current Operation. *IEEE Electron Device Letters* **36**, 675 (2015).

283. Lee, C., Choi, W., Kwak, M., Kim, S. & Hwang, H. *Excellent Synapse Characteristics of 50 Nm Vertical Transistor With WO_x Channel for High Density Neuromorphic System* in *2021 Symposium on VLSI Technology* (2021), 1.

284. Won, S., Lee, S. Y., Park, J. & Seo, H. Forming-Less and Non-Volatile Resistive Switching in by Oxygen Vacancy Control at Interfaces. *Scientific Reports* **7**, 1 (2017).

285. Lassner, E. & Schubert, W.-D. in *Tungsten* 85 (Springer US, Boston, MA, 1999).

286. Salleh, F., Tengku Saharuddin, T. S., Samsuri, A., Othaman, R., Mohamed Hisham, M. W. & Yarmo, M. A. Reduction Behaviour of WO3 to W Under Carbon Monoxide Atmosphere. *Materials Science Forum* **840**, 305 (2016).

287. Brotherton, S. D. in *Introduction to Thin Film Transistors* 9 (Springer International Publishing, Heidelberg, 2013).

288. Bang, S., Lee, S., Park, J., Park, S., Jeong, W. & Jeon, H. Investigation of the Effects of Interface Carrier Concentration on ZnO Thin Film Transistors Fabricated by Atomic Layer Deposition. *Journal of Physics D: Applied Physics* **42**, 235102 (2009).

289. Nakata, M., Tsuji, H., Sato, H., Nakajima, Y., Fujisaki, Y., Takei, T., Yamamoto, T. & Fujikake, H. Influence of Oxide Semiconductor Thickness on Thin-Film Transistor Characteristics. *Japanese Journal of Applied Physics* **52**, 03BB04 (2013).

290. Goetzberger, A. & Nicollian, E. H. Transient Voltage Breakdown Due to Avalanche in Mis Capacitors. *Applied Physics Letters* **9**, 444 (1966).

291. Bégon-Lours, L. *et al.* Factors Limiting Ferroelectric Field-Effect Doping in Complex Oxide Heterostructures. *Physical Review Materials* **2**, 084405 (2018).

292. Kim, S., Gokmen, T., Lee, H.-M. & Haensch, W. E. *Analog CMOS-based Resistive Processing Unit for Deep Neural Network Training* in *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)* **2017-Augus** (IEEE, 2017), 422.

293. Gokmen, T., Onen, O. M. & Haensch, W. Training Deep Convolutional Neural Networks With Resistive Cross-Point Devices, 1 (2017).

294. Yu, S., Chen, P. Y., Cao, Y., Xia, L., Wang, Y. & Wu, H. Scaling-Up Resistive Synaptic Arrays for Neuro-Inspired Architecture: Challenges and Prospect. *Technical Digest - International Electron Devices Meeting, IEDM* **2016-Febru**, 17.3.1 (2015).

295. Obradovic, B., Rakshit, T., Hatcher, R., Kittl, J., Sengupta, R., Hong, J. G. & Rodder, M. S. A Multi-Bit Neuromorphic Weight Cell Using Ferroelectric FETs, Suitable for SoC Integration. *IEEE Journal of the Electron Devices Society* **6**, 438 (2018).

296. Mulaosmanovic, H., Ocker, J., Muller, S., Noack, M., Muller, J., Polakowski, P., Mikolajick, T. & Slesazeck, S. Novel Ferroelectric FET Based Synapse for Neuromorphic Systems. *Digest of Technical Papers - Symposium on VLSI Technology*, T176 (2017).

297. Oh, S., Kim, T., Kwak, M., Song, J., Woo, J., Jeon, S., Yoo, I. K. & Hwang, H. HfZrO$_x$ -Based Ferroelectric Synapse Device With 32 Levels of Conductance States for Neuromorphic Applications. *IEEE Electron Device Letters* **38**, 732 (2017).

298. Mulaosmanovic, H., Ocker, J., Müller, S., Schroeder, U., Müller, J., Polakowski, P., Flachowsky, S., van Bentum, R., Mikolajick, T. & Slesazeck, S. Switching Kinetics in Nanoscale Hafnium Oxide Based Ferroelectric Field-Effect Transistors. *ACS Applied Materials & Interfaces* **9**, 3792 (2017).

299. Jerry, M., Chen, P.-Y., Zhang, J., Sharma, P., Ni, K., Yu, S. & Datta, S. *Ferroelectric FET Analog Synapse for Acceleration of Deep Neural Network Training* in *2017 IEEE International Electron Devices Meeting (IEDM)* **6** (IEEE, 2017), 6.2.1.

300. Gong, N., Idé, T., Kim, S., Boybat, I., Sebastian, A., Narayanan, V. & Ando, T. Signal and Noise Extraction From Analog Memory Elements for Neuromorphic Computing. *Nature Communications* **9**, 2102 (2018).

301. Haubner, R., Schubert, W. D., Lassner, E. & Lux, B. Influence of Aluminum on the Reduction of Tungsten Oxide to Tungsten Powder. *Int. J. Refract. Hard Met.* **6**, 161 (1987).

302. Greiner, M. T. & Lu, Z.-H. Thin-Film Metal Oxides in Organic Semiconductor Devices: Their Electronic Structures, Work Functions and Interfaces. *NPG Asia Materials* **5**, e55 (2013).

303. Vitale, S. A., Kedzierski, J., Healey, P., Wyatt, P. W. & Keast, C. L. Work-Function-Tuned TiN Metal Gate FDSOI Transistors for Subthreshold Operation. *IEEE Transactions on Electron Devices* **58**, 419 (2011).

304. Mehta, R. R., Silverman, B. D. & Jacobs, J. T. Depolarization Fields in Thin Ferroelectric Films. *Journal of Applied Physics* **44**, 3379 (1973).

305. Si, M. & Ye, P. D. The Critical Role of Charge Balance on the Memory Characteristics of Ferroelectric Field-Effect Transistors. *IEEE Transactions on Electron Devices*, 1 (2021).

306. Si, M., Lin, Z., Noh, J., Li, J., Chung, W. & Ye, P. D. The Impact of Channel Semiconductor on the Memory Characteristics of Ferroelectric Field-Effect Transistors. *IEEE Journal of the Electron Devices Society* **8**, 846 (2020).

307. Li, J., Nagaraj, B., Liang, H., Cao, W., Lee, C. H. & Ramesh, R. Ultrafast Polarization Switching in Thin-Film Ferroelectrics. *Applied Physics Letters* **84**, 1174 (2004).

308. Chiu, F.-C. A Review on Conduction Mechanisms in Dielectric Films. *Advances in Materials Science and Engineering* **2014**, 1 (2014).

309. Sze, S. & Ng, K. K. *Physics of Semiconductor Devices* 739 (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2006).

310. Böscke, T. S., Teichert, S., Bräuhaus, D., Müller, J., Schröder, U., Böttger, U. & Mikolajick, T. Phase Transitions in Ferroelectric Silicon Doped Hafnium Oxide. *Applied Physics Letters* **99**, 112904 (2011).

311. Clopath, C., Büsing, L., Vasilaki, E. & Gerstner, W. Connectivity Reflects Coding: A Model of Voltage-Based STDP With Homeostasis. *Nature Neuroscience* **13**, 344 (2010).

312. Ziegler, M., Riggert, C., Hansen, M., Bartsch, T. & Kohlstedt, H. Memristive Hebbian Plasticity Model: Device Requirements for the Emulation of Hebbian Plasticity Based on Memristive Devices. *IEEE Transactions on Biomedical Circuits and Systems* **9**, 197 (2015).

313. Kim, M.-K., Kim, I.-J. & Lee, J.-S. CMOS-compatible Ferroelectric NAND Flash Memory for High-Density, Low-Power, and High-Speed Three-Dimensional Memory. *Science Advances* **7**, 1 (2021).

314. He, C. *et al.* Artificial Synapse Based on Van Der Waals Heterostructures With Tunable Synaptic Functions for Neuromorphic Computing. *ACS Applied Materials and Interfaces* **12**, 11945 (2020).

315. Boyn, S. *et al.* Learning Through Ferroelectric Domain Dynamics in Solid-State Synapses. *Nature Communications* **8**, 14736 (2017).

316. Mulaosmanovic, H., Dunkel, S., Trentzsch, M., Beyer, S., Breyer, E. T., Mikolajick, T. & Slesazeck, S. Investigation of Accumulative Switching in Ferroelectric FETs: Enabling Universal Modeling of the Switching Behavior. *IEEE Transactions on Electron Devices* **67**, 5804 (2020).

317. Ruan, D. B., Liu, P. T., Chiu, Y. C., Kuo, P. Y., Yu, M. C., Gan, K. J., Chien, T. C. & Sze, S. M. Mobility Enhancement for High Stability Tungsten-Doped Indium-Zinc Oxide Thin Film Transistors With a Channel Passivation Layer. *RSC Advances* **8**, 6925 (2018).

318. Shen, Z., Zhao, C., Qi, Y., Xu, W., Liu, Y., Mitrovic, I. Z., Yang, L. & Zhao, C. Advances of RRAM Devices: Resistive Switching Mechanisms, Materials and Bionic Synaptic Application. *Nanomaterials* **10**, 1 (2020).

319. Swaidan, Z., Kanj, R., El Hajj, J., Saad, E. & Kurdahi, F. RRAM Endurance and Retention: Challenges, Opportunities and Implications on Reliable Design. *2019 26th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2019*, 402 (2019).

CURRICULUM VITAE

PERSONAL DATA

Name	Mattia Halter
Date of Birth	March 17, 1991
Place of Birth	Bettingen BS, Switzerland
Citizen of	Switzerland

EDUCATION

December 2017 – March 2022	PhD candidate at the Chair of Computational Nanoelectronics, Department of Information Technology and Electrical Engineering (D-ITET), ETH Zurich, Zurich, Switzerland
September 2015 – April 2017	Graduate studies in Micro and Nanosystems, Department of Mechanical and Process Engineering, ETH Zurich, Zurich, Switzerland *Final degree:* MSc ETH
September 2011 – August 2014	Undergraduate studies in Information Technology and Electrical Engineering, ETH Zurich, Zurich, Switzerland *Final degree:* BSc ETH
August 2006 – June 2011	Kantonsschule Chur, Chur, Switzerland *Final degree:* Matura

EMPLOYMENT

September 2014 – February 2015	Internship: Thermal Sensing and Energy Harvesting *greenTEG AG,* Zurich, Switzerland

PUBLICATIONS

Articles in peer-reviewed journals:

J1. O'Connor, É., **Halter, M.**, Eltes, F., Sousa, M., Kellock, A., Abel, S. & Fompeyrine, J. Stabilization of ferroelectric $Hf_xZr_{1-x}O_2$ films using a millisecond flash lamp annealing technique. *APL Materials* **6**, 121103 (2018).

J2. **Halter, M.**, Bégon-Lours, L., Bragaglia, V., Sousa, M., Offrein, B. J., Abel, S., Luisier, M. & Fompeyrine, J. Back-End, CMOS-Compatible Ferroelectric Field-Effect Transistor for Synaptic Weights. *ACS Appl. Mater. Interfaces* **12**, 17725 (2020).

J3. Zacharaki, C., Tsipas, P., Chaitoglou, S., Bégon-Lours, L., **Halter, M.** & Dimoulas, A. Reliability aspects of ferroelectric $TiN/Hf_{0.5}Zr_{0.5}O_2/Ge$ capacitors grown by plasma assisted atomic oxygen deposition. *Appl. Phys. Lett.* **117**, 212905 (2020).

J4. Begon-Lours, L., **Halter, M.**, Popoff, Y., Yu, Z., Falcone, D., Davila, D., Bragaglia, V., La Porta, A., Jubin, D., Fompeyrine, J. & Offrein, B. J. Analog Resistive Switching in BEOL, Ferroelectric Synaptic Weights. *IEEE J. Electron Devices Soc.*, 1 (2021).

J5. Bégon-Lours, L., **Halter, M.**, Popoff, Y. & Offrein, B. J. Ferroelectric, Analog Resistive Switching in Back-End-of-Line Compatible $TiN/HfZrO_4/TiO_x$ Junctions. *Phys. Status Solidi RRL* **15**, 2000524 (2021).

Conference contributions:

C1. Begon-Lours, L., **Halter, M.**, Pineda, D. D., Bragaglia, V., Popoff, Y., la Porta, A., Jubin, D., Fompeyrine, J. & Offrein, B. J. *A Back-End-Of-Line Compatible, Ferroelectric Analog Non-Volatile Memory* in. 2021 IEEE International Memory Workshop (IMW) (IEEE, Dresden, Germany, 2021), 1.

C2. Begon-Lours, L., **Halter, M.**, Pineda, D. D., Popoff, Y., Bragaglia, V., Porta, A. L., Jubin, D., Fompeyrine, J. & Offrein, B. J. *A BEOL Compatible, 2-Terminals, Ferroelectric Analog Non-Volatile Memory* in. 2021 5th IEEE Electron Devices Technology & Manufacturing Conference (EDTM) (IEEE, Chengdu, China, 2021), 1.

C3. Begon-Lours, L., **Halter, M.**, Popoff, Y., Yu, Z., Falcone, D. F. & Offrein, B. J. *High-Conductance, Ohmic-like $HfZrO_4$ Ferroelectric Memristor* in. ESSCIRC 2021 - IEEE 47th European Solid State Circuits Conference (ESSCIRC) (IEEE, Grenoble, France, 2021), 87.

Patents:

P1. Bragaglia, V., Bégon-Lours, L., La Porta, A., Fompeyrine, J. & **Halter**, **M.** *pat.* P202003687US01 (2021).

Unpublished conference presentations:

U1. **Halter**, **M.** *Milisecond Flash Lamp Annealing for the stabilization of ferroelectric* $Hf_xZr_{1\text{-}x}O_2$ 2nd MEM-Q Workshop. Rethymno, Crete, Greece, 2018.

U2. **Halter**, **M.** *Milisecond Flash Lamp Annealing for the ferroelectric phase stabilization in* $Hf_xZr_{1\text{-}x}O_2$ 2019 MRS Spring Meeting and Exhibit. Phoenix, Arizona, 2019.

U3. **Halter**, **M.** *XRD and PFM evidence for the stabilization of ultra-thin ferroelectric HfxZr1-xO2 by millisecond flash lamp annealing* Poster. Poster. F2Cp2 joint conference (ISAF 2019). Lausanne, Switzerland, 2019.

NOTES