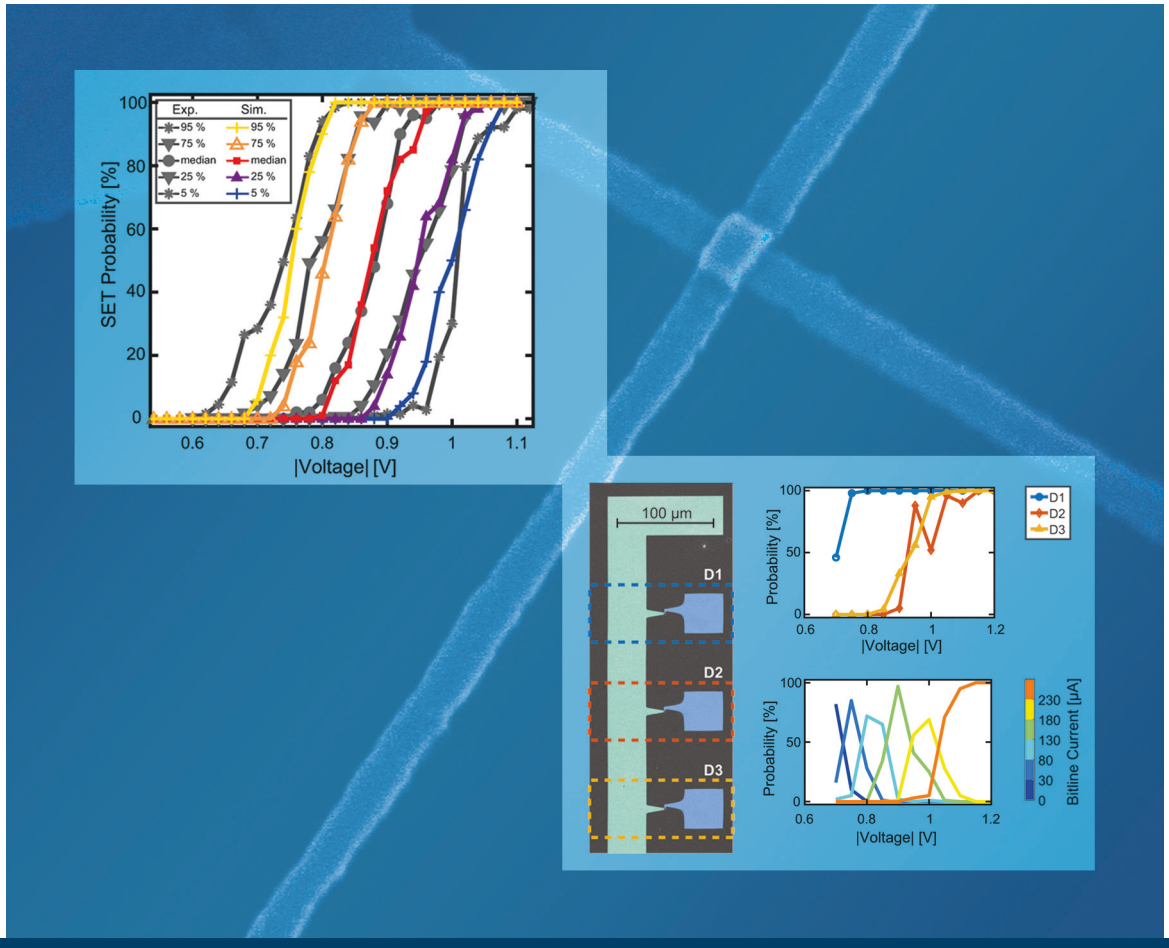


Hafnium oxide based memristive devices as functional elements of neuromorphic circuits

Felix Johannes Cüppers

Information
Band / Volume 97
ISBN 978-3-95806-702-8

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum

Schriften des Forschungszentrums Jülich
Reihe Information / Information Band / Volume 97

Forschungszentrum Jülich GmbH
Peter Grünberg Institut (PGI)
JARA-Institut Energy-efficient information technology (PGI-10)

Hafnium oxide based memristive devices as functional elements of neuromorphic circuits

Felix Johannes Cüppers

Schriften des Forschungszentrums Jülich
Reihe Information / Information Band / Volume 97

ISSN 1866-1777 ISBN 978-3-95806-702-8

Bibliografische Information der Deutschen Nationalbibliothek.
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten
sind im Internet über http://dnb.d-nb.de abrufbar.

Herausgeber und Vertrieb:	Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag 52425 Jülich Tel.: +49 2461 61-5368 Fax: +49 2461 61-6103 zb-publikation@fz-juelich.de www.fz-juelich.de/zb
Umschlaggestaltung:	Grafische Medien, Forschungszentrum Jülich GmbH
Druck:	Grafische Medien, Forschungszentrum Jülich GmbH
Copyright:	Forschungszentrum Jülich 2023

Schriften des Forschungszentrums Jülich
Reihe Information / Information, Band / Volume 97

D 82 (Diss. RWTH Aachen University, 2023)

ISSN 1866-1777
ISBN 978-3-95806-702-8

Vollständig frei verfügbar über das Publikationsportal des Forschungszentrums Jülich (JuSER)
unter www.fz-juelich.de/zb/openaccess.

Abstract

Due to the approaching limit of the computational speed of classical von-Neumann architectures, data transfer-intensive cognitive applications in future information technology demand a paradigm shift. "Beyond-von Neumann" concepts such as biologically inspired neuromorphic circuits with adjustable synaptic weights promise an energy-efficient increase in computing power. In this context, novel memristive devices such as redox-based resistive random access memories (ReRAM) are investigated intensively. They combine nonvolatility, scalability and energy efficiency. Moreover, they also allow the programming of multiple different resistive states, which further increases the memory density in addition to the compact design. Due to their mixed ionic-electronic function, they differ significantly from purely electronic systems. Important criteria for the use of memristive devices in neuromorphic circuits are the operation parameters for the two switching modes abrupt and analog switching, the stochasticity of the switching processes SET and RESET, the variability of the resistance states HRS and LRS as well as the number of programmable states. In addition to the quantification of these parameters, the physical understanding of the processes taking place is crucial in order to make predictive statements about applicability and reliability in circuits. In this context, the exchange with and further development of physical models is essential. A typical filamentary ReRAM cell operating in the bipolar valence change mechanism (VCM) is composed of one or more insulating metal oxide layers and two metal electrodes, which differ in terms of work function and chemical reactivity. A preferred choice for the metal oxide layer by the industry is HfO_2, since it is already available in semiconductor device fabrication lines. By intentionally introducing an additional sub-stoichiometric titanium oxide layer and using a chemically reactive titanium electrode and an inert platinum electrode, reproducible and stable switching behavior is obtained. In this work, the described switching modes are systematically analyzed on nanoscale $Pt/HfO_2/TiO_x/Ti/Pt$ devices based on statistical ensembles. The devices are highly comparable to industrially available options. With the aid of compact model simulations, the results are physically interpreted to obtain a comprehensive description of the devices as a foundation for usage in future "Beyond-von Neumann" concepts. The results allow an evaluation of the HfO_2-based ReRAM cells with respect to their application in novel neuromorphic circuits.

ReRAM devices of atomic layer deposition (ALD)-grown 3 nm thick HfO_2 films

were fabricated as cross-point devices with sizes ranging from 10000 nm^2 to 3600 nm^2. Functional devices of 1600 nm^2 were demonstrated as nano-plug structures. Extensive statistical characterization of the electroforming behavior and the switching stability as well as calibration of the switching properties by means of parameter variation form the basis for the differentiated analysis of the switching kinetics with rectangular voltage pulses between 100 ns and 1 s. The study of the abrupt switching kinetics in SET and RESET revealed the influence of the high-resistance state (HRS) on the switch-on process (SET) on the one hand and the influence of a series resistor on the switch-off process (RESET) on the other hand. Using the physically motivated compact model "JART v1b" developed in cooperation between IWE-II of RWTH Aachen University and PGI-7, it could be shown that, in addition to the delay time, the transition time in SET, which is difficult to access experimentally, also depends significantly on the HRS. Furthermore, if the low-resistance state (LRS) approaches the internal series resistance, a significant time delay of the RESET process is caused. By restricting HRS and LRS to a medium resistance range, the delay times can be minimized. Thus, these transition regions can be efficiently used for analog switching. Quantitative studies in this operation mode revealed that by appropriate choice of the voltage amplitudes, the behavior of the cells can be controlled to meet the requirements of neuromorphic circuits such as symmetry and programmability of intermediate conductance states. Further detailed investigations on the stochasticity of SET voltages over repeated switching operations and between different devices, performed on an extended device ensemble, allowed evaluation of parallel-connected devices as artificial synapses. The synapse was demonstrated both experimentally and in simulation using an extended version of the JART v1b model that includes device variability. The subsequent successful demonstration of synapses in a spiking neural network highlights the potential of memristive devices for neuromorphic circuits. The results illustrate that the range of applications can be further extended through a focused combination of device development and circuit design. In summary, this work shows that nanosized filamentary ReRAM devices have a high potential for use as artificial synapses in neuromorphic circuits of the future computer generation. The obtained results contribute to a deeper physical understanding of the analog and abrupt switching behavior and demonstrate the wide range of possible applications.

Kurzfassung

Aufgrund der nahenden Grenze der Rechengeschwindigkeit klassischer von Neumann-Architekturen verlangen datentransfer-intensive kognitive Anwendungen in der zukünftigen Informationstechnologie nach einem Paradigmenwechsel. „Beyond von Neumann"-Konzepte wie biologisch inspirierte neuromorphe Schaltungen mit einstellbaren synaptischen Gewichten ermöglichen eine energieeffiziente Steigerung der Rechenleistung. In diesem Kontext werden neuartige memristive Bauelemente wie redox-basierte resistive Speicher mit wahlfreiem Zugriff (ReRAM) intensiv erforscht. Sie vereinen Nicht-Flüchtigkeit, Skalierbarkeit und Energieeffizienz miteinander. Zudem erlauben sie die Programmierung vieler unterschiedlicher Widerstandszustände, was die Speicherdichte zusätzlich zur kompakten Bauform erhöht. Dabei unterscheiden sie sich aufgrund ihrer gemischt ionisch-elektronischen Funktionsgrundlage stark von rein elektronischen Systemen. Wichtige Kriterien für den Einsatz memristiver Bauelemente in neuromorphen Schaltungen sind die Operationsparameter für die zwei Schaltmodi abruptes und analoges Schalten, die Stochastizität der Schaltprozesse SET und RESET, die Variabilität der Schaltzustände HRS und LRS sowie die Anzahl der programmierbaren Zustände. Neben der Quantifizierung dieser Parameter ist das physikalische Verständnis der ablaufenden Prozesse entscheidend, um prädiktive Aussagen über Einsatz und Zuverlässigkeit in Schaltungen treffen zu können. In diesem Zusammenhang ist der Austausch mit und die Weiterentwicklung von physikalischen Modellen essenziell. Eine typische im bipolaren Valenzwechselmechanismus (VCM) schaltende filamentäre ReRAM-Zelle ist aus einer oder mehreren isolierenden Metalloxidschichten und zwei Metallelektroden aufgebaut, wobei Unterschiede hinsichtlich der Austrittsarbeit und der chemischen Reaktivität bestehen. Eine von der Industrie bevorzugte Wahl für die Metalloxidschicht ist HfO_2, da es bereits in den Herstellungslinien für Halbleiterbauelemente verfügbar ist. Durch gezielte Einführung einer zusätzlichen sub-stöchiometrischen Titanoxid-Schicht und Einsatz einer chemisch reaktiven Titanelektrode und einer inerten Platinelektrode wird reproduzierbar stabiles Schaltverhalten erreicht. In dieser Arbeit werden die beschriebenen Schaltmodi an industrienah hergestellten nanoskalierten $Pt/HfO_2/TiO_x/Ti/Pt$-Bauelementen auf Basis statistischer Ensembles systematisch analysiert. Durch Zuhilfenahme von Kompakt-Simulationsmodellen werden die Ergebnisse physikalisch interpretiert, um eine umfangreiche Beschreibung der Bauelemente zu erzielen als Grundlage für den Einsatz in zukünftigen „Beyond-von Neumann"-Konzepten. Die Ergeb-

nisse erlauben eine Bewertung der HfO_2-basierten ReRAM-Zellen hinsichtlich ihrer Anwendung in neuartigen neuromorphen Schaltungen.

ReRAM Bauelemente aus Atomlagen-Abscheidung (ALD) gewachsenen 3 nm dicken HfO_2 Schichten wurden als Kreuzpunkt-Bauteile mit Größen zwischen 10 000 nm^2 und 3600 nm^2 hergestellt. Als Nanolochstrukturen konnten funktionsfähige Bauelemente von 1600 nm^2 demonstriert werden. Umfangreiche statistische Charakterisierung des Elektroformierungsverhaltens und der Schaltstabilität sowie eine Kalibrierung der Schalteigenschaften mittels Parametervariation bilden die Grundlage für die differenzierte Analyse der Schaltkinetik mit rechteckigen Spannungspulsen zwischen 100 ns und 1 s. Die Studie zur abrupten Schaltkinetik in SET und RESET deckte einerseits den Einfluss des hochohmigen Zustands (HRS) auf den Einschaltprozess (SET) und andererseits den Einfluss eines Serienwiderstands auf den Ausschaltprozess (RESET) auf. Unter Zuhilfenahme des in Kooperation zwischen dem IWE-II der RWTH Aachen und PGI-7 entwickelten physikalisch motivierten Kompaktmodells „JART v1b“ konnte gezeigt werden, dass neben der Verzögerungszeit auch die experimentell schlecht zugängliche Übergangszeit im SET wesentlich vom HRS abhängt. Weiterhin bewirkt eine Annäherung des niederohmigen Zustands (LRS) an den internen Serienwiderstand eine deutliche Zeitverzögerung des RESET. Durch Einschränkung von HRS und LRS auf einen mittleren Widerstandsbereich können die Verzögerungszeiten minimiert werden. So kann der Übergangsbereich effizient für analoges Schalten genutzt werden. Quantitative Studien ergaben, dass durch geeignete Wahl der Spannungsamplituden das Verhalten der Zellen so gesteuert werden kann, dass die Anforderungen neuromorpher Schaltungen wie Symmetrie und Einstellbarkeit von Zwischenzuständen erfüllt werden. Die genauere Untersuchung der Stochastizität der SET Spannungen zwischen Schaltvorgängen und zwischen Bauelementen, die an einem erweiterten Bauelemente-Ensemble durchgeführt wurde erlaubte in Kombination mit einer um die Bauelemente-Variabilität erweiterten Version des JART v1b Modells die Evaluierung parallelgeschalteter Bauteile als künstliche Synapse, sowohl experimentell als auch in der Simulation. Die erfolgreiche Demonstration der Synapsen in einem gepulsten neuronalen Netzwerk unterstreicht das Potenzial memristiver Bauelemente für neuromorphe Schaltungen. Durch eine gezielte Kombination von Bauelemententwicklung und Schaltungsentwurf lässt sich das Anwendungsspektrum memristiver Zellen noch deutlich erweitern. Zusammenfassend zeigt diese Arbeit, dass nanostrukturierte filamentäre ReRAM-Bauelemente ein hohes Potenzial für den Einsatz als künstliche Synapsen in neuromorphen Schaltungen der künftigen Computergeneration haben. Die erzielten Ergebnisse tragen zu einem tieferen physikalischen Verständnis des analogen und des abrupten Schaltverhaltens der Bauelemente bei und demonstrieren die vielfältigen Einsatzmöglichkeiten.

Contents

1 Introduction

The enormous amount of data that is currently generated and will be generated in the next decades by a multitude of sensors in modern electronic devices causes the need for adequately powerful yet energy-efficient data processing equipment. However, current computing technologies struggle to deal with the complex challenges of this enormous amount of data since it is frequently unstructured, noisy or incomplete. Therefore, research in the field of information technology aims to develop new data analysis concepts. Besides other concepts such as quantum computing and hyperdimensional computing, neuromorphic computing (NC) is seen as a promising candidate to fulfill these requirements. While NC concepts can be executed on traditional hardware, their von-Neumann architecture limits the computation speed and efficiency. Novel memory technologies have gained attention in this context as they promise benefits for NC over the traditional counterparts composed of Complementary Metal-Oxide-Semiconductor (CMOS) logic gates and physically separated Dynamic Random Access Memories (DRAM). Accordingly, the applications for these novel memory devices are frequently termed "Beyond-von-Neumann" concepts. Within the group of emerging memory technologies, nonvolatile resistive switching memories have received a lot of attention both from industry and research due to their ease of fabrication, dense integration possibility and potential use of low cost materials. The essential device consists of a metal-insulator-metal (MIM) stack. In the group of resistive switching devices, several subtypes have been identified. One subtype has a significant advantage over the others and is therefore at the forefront of research. That advantage is that the materials used in the process are CMOS compatible and are in fact already established in today's production lines. The mentioned materials are HfO_2 and Ta_2O_5, which are used as CMOS high-k gate dielectric [1–3] and as capacitor oxide or, in the nitride form, as copper diffusion barrier [4]. The world's first available multi-project wafer service including oxide-based resistive switching memories is in fact based on Ti-doped HfO_2 [5]. The resistive switching in these devices is based

on redox processes within the insulating oxide film. The reactions occur in a locally constrained region of conical shape, which has lead to the commonly used term filamentary redox-based random access memory, short ReRAM. For these devices, many of the imposed criteria for augmenting existing memory technologies have been successfully demonstrated, such as device scaling[6], sub-nanosecond switching[7–10] and low power non-volatile information storage[11].

However, together with the development of NC concepts, new criteria for storage devices are emerging. When employing the devices e.g. as artificial synapses, the focus is shifted from previously important criteria to different ones which have not received attention before. Filamentary ReRAM devices based on HfO_2 and Ta_2O_5 need to be re-evaluated, and their capability to fulfill these new demands needs to be clarified. Features like resistance drift and noise or switching stochasticity, which were previously seen as parasitic and undesired, may now be the enabling factor for certain NC applications.

The present work utilizes a well studied and designed HfO_2-based ReRAM device, which is highly comparable to industrial and advanced research devices [5, 12–14]. The aim is to elucidate and evaluate the coexistence of stochastic switching between distinct resistances and the possibility of programming analog resistances in single filamentary valence change mechanism (VCM) ReRAM devices with respect to possible use cases in NC applications.

1.1 State of the art

Current state of technology of filamentary ReRAM devices for novel computing architectures is reviewed in this section. Previous assessments targeting the use as stand-alone memory or embedded memory are not considered. The reader is referred to respective papers on ReRAM devices for conventional memory published in the last 10 years as for example [11, 12, 14–29]. Rather, this section is divided into three categories that reflect the majority of use cases in novel applications: multilevel state programming, analog conductance tuning and exploitation of stochastic switching. In general, the results are shared between filamentary material systems, and are therefore valid for HfO_2-based devices, also.

1.1.1 Multilevel switching

Multilevel memory is considered extremely advantageous as it increases the information density without requiring additional space. The fact that ReRAM devices are capable of storing multiple bits of information has been known for a while, but deeper investigations have only been published recently. Sheng et al.[6] demonstrated more than 256 states (8 bits) in a single ReRAM device with access transistor when considering the full conductance range. With the target of low power operation in mind, even 8 states (3 bits) and 4 states (2 bits) were shown below 10 µS and below 1 µS, respectively. Comparable results have been achieved by Stathopoulos et al.[30], who were able to dramatically increase the number of programmable states by varying the interfacial metal oxide. They demonstrated a maximum of 92 distinguishable states (6.5 bits) in a single crossbar device while maintaining a read conductance below 50 µS. In contrast to the studies mentioned, the majority of publications report between 2 and 8 states in a single device [13, 31–40]. As seen from comprehensive review papers [41, 42], the general trend is to focus on few well defined conductance states within the capable range of the respective device, which typically limits the number of states to around 8 to 10.

1.1.2 Analog conductance tuning

However, the outstandingly high numbers provided by [6] and [30] indicate that filamentary ReRAM devices are capable of adopting more than 10 states. In fact, as many as 300 states were estimated by the authors of [42]. However, the limiting factor in this case is that such states are hardly distinguishable from each other. Therefore, it makes sense to refer to such operations as analog conductance tuning. This property has attracted a lot of attention since the vast majority of novel computing concepts benefit from nonvolatile analog weights. Therefore, a large number of publications reports analog capabilities in the devices. Most prominently, this feature is found in area-dependent VCM devices, which are still in an exploratory state. Because of the strong demand for analog tuning capability, by now, all of the commonly used materials for filamentary resistive switching have also been studied with respect to demonstrating analog conductance tuning. In particular, the most research has been done on HfO_2-based [43–54] , Ta_2O_5-based [55–59] , $SrTiO_3$-based [60] , Al_2O_3-based [61–63] and TiO_2-based [64, 65] devices. Comparison of the metrics of these devices is difficult to obtain because of the various target applications of the publications in com-

bination with non-standardized characterization methodology. Common challenges, which have also been consently reported in multiple review and research papers [66–68] include limited resistance window compared to operation as conventional memory device, lowered switching voltage and complex dynamics.

1.1.3 Stochastic switching features

In this context and throughout this work, the term stochasticity describes the non-deterministic character of occuring processes such as transitions between states. The term variability will be used for observed states or metrics such as measured resistances and voltages. When ReRAM devices were investigated for embedded memory applications, one of the most challenging aspects was the strong presence of variability and stochasticity. The extent of interlinking between the variability and the stochasticity is still under investigation. The established results for memory devices were recently reconsidered as working principle for neuromorphic applications. Especially the switching voltage stochasticity has attracted attention. Several studies focus on exploiting the stochasticity of the SET process for a multitude of applications. Dalgaty et al.[69] employed voltage signals with amplitudes in the non-deterministic regime to update ReRAM devices which are located in multiple places of a recurrent neural network. Yu et al.[70] employed the stochastic SET programming condition to operate a two-layer winner-take-all network that determines the angle orientation of an input line. By overlapping a long but weak programming signal from the input layer with a shorter signal from the output layer, the stochastic SET process is used to incrementally improve the network's accuracy through setting the correct devices, while the incorrect connections are weakened. Naous et al.[71] followed a similar two-layer network winner-take-all approach in their simulation based study, which employed a generic memristive device model. They also investigated the possibility to transfer the stochastic switching feature to the neurons, resulting in similar performance for the well-known MNIST dataset[72]. Wenger et al.[73] benchmarked their CMOS cointegrated devices with the same dataset. Their two-layer perceptron network, which is composed of software-sided neurons and hardware memristive devices, was able to achieve high recognition rates using a very limited number of neurons. The stochastic feature in the supervised learning scheme was exploited to SET a subset of devices proportional to the pixel intensity of each number in the dataset. Fewer studies exist on the resistance state variability. Dalgaty et al.[74] were able to

demonstrate that the intrinsic conductance variability of the higher conducting state can be exploited to efficiently implement Markov chain Monte Carlo sampling algorithms. In their study, they employed the normal random conductance variable for a supervised learning task and a reinforcement learning task. In both cases, accuracies that outperform software models with the same number of elements are achieved.

In summary, filamentary ReRAMs have overcome the stigma of their variable and stochastic nature. Instead of mitigation, the named studies have achieved exploitation of these previously undesired features. The current state of ReRAM technology is at a turning point. A few years ago, the driving force for research of ReRAM devices was to complement the existing memory technologies. The goal was to create a non-volatile, fast read-and-write storage device that fills the latency gap between DRAM memory and slower, permanent hard drive memory [75]. NC architectures, however, outperform conventional computers by orders of magnitude when it comes to complex tasks like image or speech recognition as well as optimization problems or prediction tasks. Therefore, the target of hardware research has shifted towards a memory unit that can store multiple bits of information or has properties that allow for efficient algorithm execution. Filamentary ReRAM devices, especially those based on HfO_2, pass a lot of challenges due to their inherent properties such as temperature constraints, integration density, cost and speed, but need more investigation with respect to their unconventional properties like stochasticity and analog conductance tuning. The working principle of ionic devices is also lacking complete and comprehensive models, especially compared to the available descriptions of electronic devices. For further improvements of ReRAM technology, it is imperative to gain more insights into the observed phenomena. This understanding will in turn result in better device design and give indications for the proper choice of device for neuromorphic applications.

1.2 Scope of this work

The present work aims to explore possible application options for filamentary nanocrossbar valence change mechanism memristive devices in novel computing concepts. For this purpose, the focus is on the switching properties of single devices, statistical analysis of variability of single and multiple devices and on unconventional switching phenomena such as stochastic switching and analog conductance tuning. The test vehicle device for this study is the well researched Pt/ 3 nm HfO_2 / 3 nm TiO_x /

10 nm Ti / Pt stack, which is one of the most frequently employed material combination both in industrial and research contexts. The stack is integrated into crossbar structures with metal line widths of 100 nm down to 60 nm. As a proof of concept of the scalability, resistive switching is also demonstrated in a nano-plug-style device of $(40\,\mathrm{nm})^2$ area, reducing the device volume down to $64 \cdot 10^3\,\mathrm{nm}^3$. The chapters of the work reflect the aim of this work:

Chapter 2 is dedicated to providing the reader with the required fundamental knowledge about the topics covered in this work. Therefore, Section 2.1 first gives an overview on the topic of emerging device technologies for novel computing architectures. Section 2.2 continues with a focus on the devices based on resistive switching. Specifically, the switching mechanism of counter-eightwise filamentary VCM memristive devices is summarized. Relevant material properties for the employed test device are given. Section 2.3 introduces the computing architectures and concepts, in which emerging devices may be utilized. Specific cases where memristive devices are employed will be discussed.

Chapter 3 summarizes the technical aspects and methods of this work. Thin film deposition techniques relevant to this work are described in Section 3.1. Section 3.2 summarizes the process for fabrication of the nano-structured crossbar devices that are mainly used in this work. Section 3.3 describes the measurement setups employed for electrical characterization of the fabricated devices.

Chapter 4 is dedicated to the phenomenon of resistive switching in the described test devices. The chapter is split into multiple sections, each reflecting a different aspect of the devices. Description of the statistics of the initial electroforming step in Section 4.1 is followed by the demonstration of device miniaturization shown in Section 4.2. Endurance characterization is presented in Section 4.3. Statistical analysis using fast triangular voltage sweeps is described in Section 4.4. Special attention is paid to the interplay of externally set parameters and resulting switching conditions. Subsequently, the switching kinetics upon rectangular voltage pulses over multiple orders of magnitude in time are studied in Section 4.5. The implications of the observed processes on the operation of the devices is investigated.

The following Chapter 5 is dedicated to the analog function of single devices. First, the analog conductance tuning by constant repeated voltage pulse trains is investigated in Section 5.1. During the analysis, a distinct conductance noise trend is observed, which is further investigated in the subsequent Section 5.2.

Three variants of Spike Timing Dependent Plasticity, which is an alternative pro-

gramming technique in the analog-like switching domain is the topic of Chapter 6.

In the following Chapter 7, the second operation mode, namely binary switching, is studied in more detail. The stochasticity and variability of a device ensemble is analyzed in detail. SET and RESET resistance switching are investigated separately and similarities and differences are highlighted in Sections 7.1 and 7.2, respectively.

Chapter 8 shows how transfers between experimental results on single devices and neuromorphic application concepts can be achieved. The results of the binary switching mode from the previous chapter are implemented into a Spiking Neural Network demonstrator application and the results are discussed.

Chapter 9 summarizes the results and findings of this thesis and provides an outlook for possible future investigation.

2 Fundamentals

This section of the work should provide the reader with necessary background information about the current state of research that relates to the topic of the present work. Initially, an overview over the currently emerging memory technologies will be given and contrasted to the previously available options. The investigated resistive switching devices are put into context with these other technologies in the next section. Subsequently, a classification within the group of resistive switching devices is given, including a more detailed definition of the working principle of devices investigated in this work. Neuromorphic applications have emerged as key aspect in application pathways for nonvolatile memories. An overview of the wide field of neuromorphic concepts is given in the last section of this chapter.

2.1 Emerging devices for neuromorphic computing

The vast majority of computing we know and use today is done in systems based on the von Neumann concept[76]. It follows the principle of physically separating the Central Processing Unit (CPU) from the Memory unit. In this architecture, inputs from the outside are passed to the CPU, which performs the appropriate computations and returns an output. While doing so, it may retrieve or store larger amounts of data in the memory, while small amounts may be stored in internal registers. The core strategy for speeding up computational tasks has not changed since the invention of the von Neumann concept. By increasing the density and hence the amount of transistors on a chip, parallelizing tasks and increasing the clock frequency, tasks are processed and the results are returned to memory[77]. However, this strategy is facing severe limitations. While the speed of computation has steadily increased through the development of faster hardware and algorithms, the transfer speed to the memory blocks is lacking behind. This issue is termed von-Neumann bottleneck. Another factor that is worsening the situation immensely is the rise of neuromorphic

computing tasks and according concepts, especially the ones that require gigantic amounts of synaptic weights such as Deep Neural Networks (DNNs). The working principle of DNNs and other networks is described in Section 2.3. Since the number of weights is so high, they are stored outside the CPU. However, the working principle of e.g. DNNs requires them to be accessed regularly. Hence, they need to be transferred back and forth often. This disadvantage of the von-Neumann concept imposes a significant energy and time penalty. New memory technologies aim to overcome this hurdle by providing fast write and read times in dense units that can be integrated on the same chip with the CPU. In special cases, the computations may even be executed in the memory units themselves. This concept is accordingly termed Computation-in-memory (CIM) and is beyond the scope of this thesis. In the past years, several of these emerging neuromorphic device technolgies have been proposed. Figure 2.1 provides an overview. The classification is done by the respective working priniciple of the memory device. Three physical mechanisms are included. Here, the following abbreviations are used:

- EDLC is Electric Double-Layer Capacitor.
- EC-doping is Electrochemical doping.
- FTJ is Ferroelectric Tunnel Junction.
- FeFET is Ferroelectric Field Effect Transistor.
- STT is Spin-Torque-Transfer.
- SOT is Spin-Orbit-Torque.
- DW motion stands for Domain Wall motion.
- PCM is Phase Change Material.
- CMOS is Complementary Metal Oxide Semiconductor.

Resistive switching elements are found in all three categories, however, redox-based devices are only found in the ion migration category. Here, the species of moving ions is specified. Although there are indications that cation movement may also play a role[78–82] the ReRAM devices investigated in this work belong to the anion based filamentary type. The distinction between non-filamentary and filamentary devices

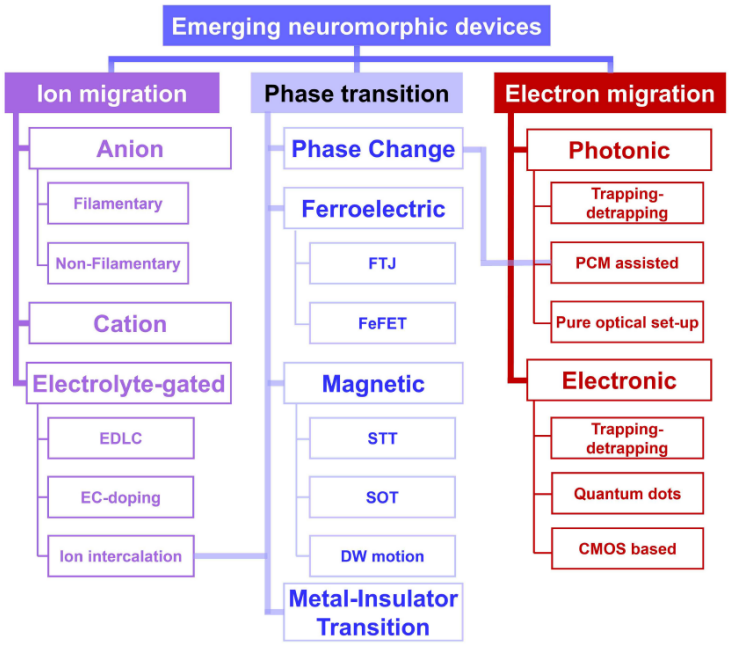


Figure 2.1: Overview of emerging neuromorphic device technologies. The classification is done by the working principle. Reprinted from [66], with the permission of AIP Publishing.

results from a measurement of the resistance states in dependence of the device area. By far the most prominent candidate of anion based filamentary devices is based on the Valence Change Mechanism (VCM).

2.2 Redox-based resistive switching devices

This section is dedicated to the phenomenon of resistance change by voltage stimulus in the case of redox-active oxide based memory cells. Basic definitions and a physical description of the working principle of filamentary valence change mechanism devices are provided in the first subsection. Ionic processes affecting the switching kinetics are the focus of the second subsection. The reader is referred to the respective textbooks for informations on other occurring physical effects and their functionalities in oxide devices[83, 84]. These may include reduction/oxidation processes on the

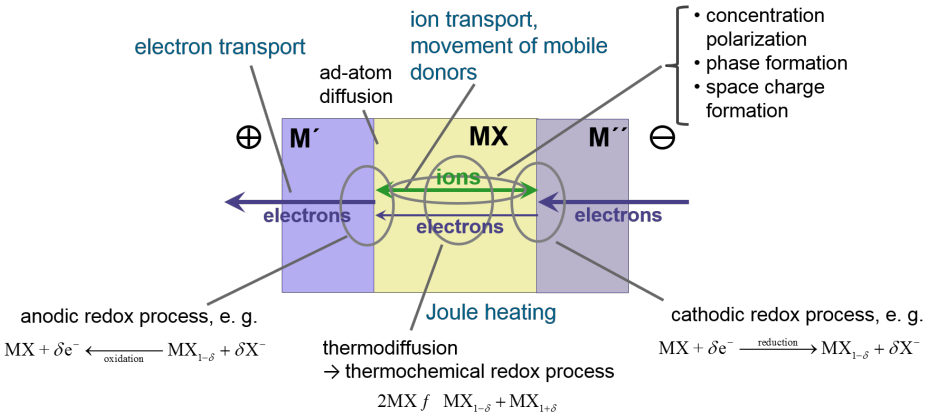


Figure 2.2: Electronic, ionic, thermal and redox processes during redox-based resistive switching in a metal (M') / metal compound isolator (MX) / metal (M") structure under bias. Reproduced with permission from [83].

nanoscale such as exchange reactions with the electrodes[85], interactions with the atmosphere [86] or so called nano-battery effects [87].

2.2.1 Physical switching model

In this section Valence Change Mechanism (VCM) devices are discussed. For this mechanism to occur, the insulator is a thin oxide layer of a mixed ionic-electronic conducting (MIEC) material. Typically, this MIEC thin film is sandwiched between two metal electrodes which differ in their oxidation enthalpy and work function. Figure 2.2, which is taken from [83] depicts the electronic and ionic processes happening on the nanoscale in a VCM device under voltage bias. Independently from the stack design, electrons are easily conducted in the metal electrodes M' and M". In contrast, the MIEC oxide layer may conduct electronic and ionic currents. The partial electronic current may result in Joule heating inside the stack. This may in turn lead to physical phenomena that are not relevant at room temperature. In this sandwich structure, reactions at the electrode interfaces may occur. Caused by the ionic currents, anodic oxidation and cathodic reduction reactions may take place at the M' and M" interfaces, respectively. This effect strongly depends on the stack design. Other reactions include oxygen species interaction with the metal electrodes, incorporation of oxygen species into the metals, phase transitions, formation of space

charge layers and concentration gradients. Noble electrode metals with low oxygen affinity make reactions of oxygen with the electrode material electrochemically unfavourable. However, incorporation of charged species is still possible. In the popular case of asymmetric electrode materials, the non-noble metal electrode can oxidize during the interface reaction of anodic oxidation. In this reaction, neutrally charged, free oxygen is created, which reacts further with the typically high oxygen affinity metals [83]. The concentration profile of oxygen vacancies, which act as donor sites in the oxide, may be manipulated by voltage bias to the stack. The oxygen affine, low work function counter electrode [88] evolves into an ohmic contact to the oxygen vacancy rich oxide and is hence often referred to as oxygen exchange layer (OEL) or as ohmic electrode (OE). In contrast, the noble, high work function electrode interface forms a Schottky-type barrier to the oxide. This interface is often referred to as Active Electrode (AE) and is expected to be the limiting electronic conductor. This Schottky-type electrode is the one where the resistance switching occurs. The switching is defined as the resistance change from a low resistance (high conductance) state, short LRS (HCS) to a high resistance (low conductance) state, short HRS (LCS) and vice versa. The HRS to LRS and the LRS to HRS transition are termed SET process and RESET process, respectively. Both processes are achieved by application of a sufficiently high voltage to the metal electrodes. In the case of VCM-type switching devices the described processes require opposite polarities, hence the mechanism is called bipolar. In contrast, devices that require a single voltage polarity to function are termed unipolar. VCM-type devices can exhibit different area scaling behavior. Material systems like $La_{0.67}Sr_{0.33}MnO_3$, $La_{0.33}Ca_{0.67}MnO_3$ and $Pr_{(1-x)}Ca_xMnO_3$ and others [89–91] show area dependent switching characteristics when sandwiched between metal electrodes. Essentially, the current through the device scales directly with the device area. These devices are often referred to as interface-type switches. Opposite to these devices, there are also resistive switching devices without area dependence. Here, the formation and rupture of a single filament during one switching cycle is understood as the underlying mechanism. Hence it is termed filamentary-type switching as the filament size shows minor to no dependence on the electrode area. In the following only these devices are discussed since they are in the main focus of this work. Figure 2.3 illustrates this abstract description for an exemplary stack of $Pt/ZrO_2/Zr$. In this specific material combination, filamentary switching is predominantly observed. Green spheres stand for mobile oxygen vacancies V_O^{**} (using the Kröger-Vink-notation [92]), while purple spheres represent Zr ions with a valence state

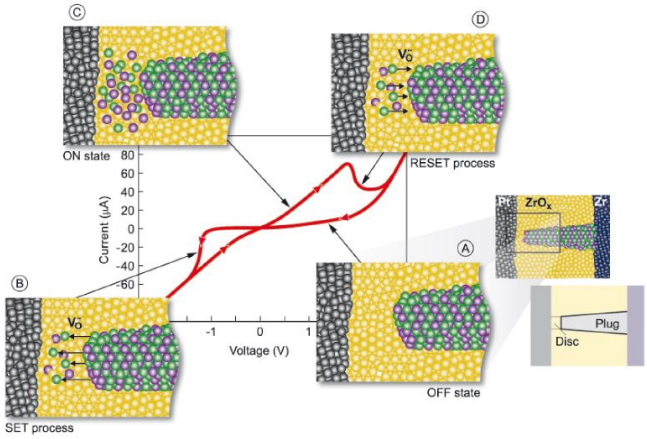


Figure 2.3: Example of a valence change memory current-voltage characteristic measurement on a $Pt/ZrO_2/Zr$ layer stack. Reproduced with permission from [83].

other than +4, i.e. +2 or +3. Yellow spheres mark Zr ions in the +4 valence state. (A) depicts the ionic distribution in the HRS, which is defined by a low concentration of oxygen vacancies close to the Pt AE. Application of sufficiently high negative voltage to the Pt electrode attracts the vacancies, which results in the SET process, see (B). At lower voltage, the LRS can be read. It is understood that a high oxygen vacancy concentration is present in close proximity to the Pt electrode, see (C). For transitioning back to the HRS, sufficiently high positive voltage is required. The result is a retraction of positively charged oxygen vacancies into a filament-shaped reservoir, see (D). Because of the conical shape and the movement between an active region and a reservoir region, the commonly used separation into 'disc' and 'plug', respectively, is introduced.

The exact mechanism of the resistance change is still under debate. Funck and Menzel recently published a comprehensive review on the conduction mechanism in ReRAM devices[93]. By studying the predominant switching properties in various oxide materials, they were able to identify two different types of oxides frequently used for ReRAM devices. The distinguishing factor was found to be the relative energy level of oxygen vacancies with respect to the conduction band edge of the oxide. For shallow defect states, it was found that the electronic conduction is limited by a Schottky barrier limited band transport, where the oxygen vacancy concentration close to the

active electrode modulates the Schottky depletion zone length, hence allows switching. In contrast, deep defect states lead to an interface limited trap assisted tunneling current. The resulting *I-V* curves exhibit different shapes: For shallow defect oxides, strong resistance nonlinearity is observed, while less pronounced to no nonlinearity is the case for deep defect states. This work focuses on HfO_2-based devices. The defect state level of V_O^{**} in HfO_2 [94, 95] as well as the observed state linearity strongly suggests that the devices used in this work fall in the category of deep defect oxides [93]. The likely conduction mechanism is hence an interface-limited electron transport over V_O^{**} defects by trap-assisted tunneling. The main limiting element however is the Schottky depletion zone at the Pt interface, which requires direct or thermally assisted tunneling into the Pt conduction band. This description is in line with the textbook model of [83]. However, the role of the interfacial TiO_x at the OEL is still under investigation.

2.2.2 Switching kinetic fundamentals

The description of the kinetics of the resistance change mechanism is important for two reasons. On the one hand, it is a relevant metric of a memory device targeted for high end applications. Filamentary VCM devices have been proven to switch in less than 1 ns [7–10]. On the other hand, the full switching kinetic curve spanning over several orders of magnitude in time may elucidate on the physical switching mechanism and different limiting reactions at different timescales [96–100]. The description of a switching kinetic curve as shown in Figure 2.4 requires certain assumptions to be made. Within the scope of this work, it is assumed that the switching process solely depends on the redistribution of doubly charged oxygen vacancies V_O^{**} within the oxide layer. Generation and recombination are not considered, but are known to play a significant role in the switching [101–103]. Further, Figure 2.4 highlights that the SET switching is extremely nonlinear with respect to the voltage-time relation. Indicated by the findings of Menzel et al. [97] as well as others [100, 104–111], the applied voltage and generated temperature create a surrounding that enables this strong nonlinearity. In summary, it was found that the ionic motion is both field and temperature accelerated. The temperature is generated through Joule heating at the active electrode interface. It may reach up to values above 1500 K temporarily. A positive feedback loop of current increase and Joule heating subsequently triggers a runaway process, which results in the typical abrupt current change observed for

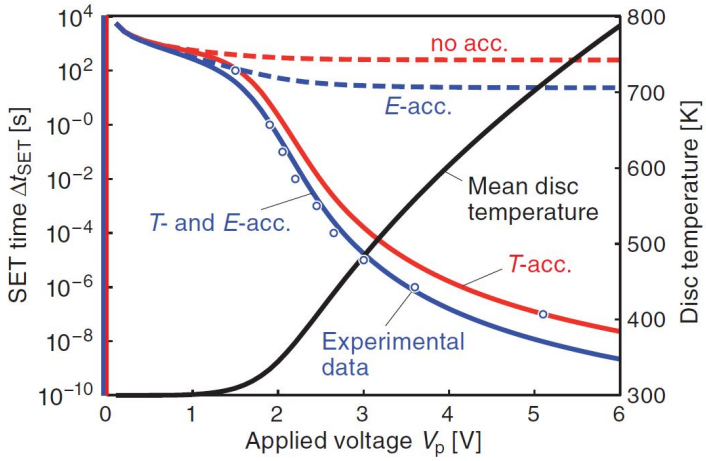


Figure 2.4: Voltage and temperature dependency on the SET transition time. By including thermal enhancement in the switching model, the strong nonlinearity in switching speed reported in VCM-type devices can be reproduced. Reproduced with permission from [97].

the SET process. Opposite to the SET switching kinetics, few studies focus on the RESET mechanism. Marchewka et al. [112] proposed a comprehensive model that describes the transient currents during RESET well. It is based on a dynamic equilibrium of oxygen vacancy drift and diffusion. While the positive polarity at the active electrode repels the doubly charged ions, a concentration gradient and a temperature gradient is formed. Due to the high overall temperature, diffusion processes are active. Subsequently, drift and diffusion forces, which are composed of chemical and thermodiffusion, counteract each other, resulting in a voltage and time dependent equilibrium. The result of this balance is the observed gradual current decrease for the RESET process. Several other studies [107, 113] report that the RESET process can also have an abrupt appearance. This is, however, caused by a configuration of the memristive element in series with a sufficiently high ohmic resistor. As the intrinsic memory resistance value approaches the series resistor value, the applied RESET voltage is divided. As a result, the observed RESET transition in a sweep experiment is shifted to higher voltage. Once the process is initiated, the resistance increases slightly, which in turn alters the voltage divider effect. Due to the nonlinear switching kinetics, a positive feedback loop of resistance increase and increasing

voltage drop starts. Accordingly, the RESET shows an abrupt reduction in current.

2.2.3 Compact model fundamentals

Modelling of the complex processes behind the resistive switching is challenging and the subject of recent investigations. One of the most advanced switching models at the time of this thesis is the fully physics-based Jülich Aachen Research Tools (JART) VCM compact model [114] in Version v1b. It will be referred to as JART VCM model in this work. As it will be employed in the following chapters, it is described in more detail here. For deeper insights about compact modelling of VCM devices, the reader is referred to more extensive publications [96, 112, 115–118] and the comprehensive PhD thesis of C. La Torre [119].

In the JART VCM model, the ReRAM device is composed of four elements. One, a Schottky diode resembles the AE/oxide interface. Second, a disc resistance R_{disc}. Third, a plug resistance R_{plug} and fourth, a series resistance R_{series}. The length of the filament l_{cell} is equal to the resistive switching oxide thickness. The disc and plug length is termed l_{disc} and l_{plug}, respectively. The filament cross-section area A is calculated by $A = \pi r_{\mathrm{fil}}^2$, where r_{fil} is the filament radius. The concentration of oxygen vacancies in the disc region and the plug region is denoted by N_{disc} and N_{plug}, respectively. The series resistance is split into two parts, which reflect electrode and line resistances on one hand and the ohmic OE/oxide interface on the other hand. Here it is important to note that the line series resistance is a function of current, since the Joule heating of metal lines with small cross-section area generates a non-negligible temperature increase.

In the model, the voltage V_{applied} is applied to the AE while the ohmic electrode is forced to ground. I is the current through the device. First, Kirchhoff's law,

$$V_{\mathrm{applied}} - [V_{\mathrm{Schottky}} + I \cdot (R_{\mathrm{disc}} + R_{\mathrm{plug}} + R_{\mathrm{series}})] = 0 \tag{2.1}$$

is solved. Although not correct in every case as recently described by Funck and Menzel [93], the conduction mechanism in the filament is approximated by band conduction with the temperature-dependent electron mobility μ_{n}. The disc and plug resistances are calculated by

$$R_{\mathrm{disc}} = \frac{l_{\mathrm{disc}}}{A \cdot z_{\mathrm{Vo}} e N_{\mathrm{disc}} \mu_{\mathrm{n}}(T)} = \frac{l_{\mathrm{disc}}}{A \cdot z_{\mathrm{Vo}} e N_{\mathrm{disc}} \mu_{\mathrm{n0}}} \exp\left(\frac{\Delta E_{\mathrm{ac}}}{k_{\mathrm{B}} T}\right) \tag{2.2}$$

and

$$R_{\mathrm{plug}} = \frac{l_{\mathrm{plug}}}{A \cdot z_{\mathrm{Vo}} e N_{\mathrm{plug}} \mu_{\mathrm{n}}(T)} = \frac{l_{\mathrm{plug}}}{A \cdot z_{\mathrm{Vo}} e N_{\mathrm{plug}} \mu_{\mathrm{n0}}} \exp\left(\frac{\Delta E_{\mathrm{ac}}}{k_{\mathrm{B}} T}\right). \tag{2.3}$$

Here, z_{Vo} is the charge state of the oxygen vacancies with respect to the situation in a perfect crystal lattice, e denotes the elementary charge, T is the temperature, k_{B} is the Boltzmann constant, and μ_{n0} is the mobility prefactor, which is temperature-independent. The mobility temperature dependence is modeled using ΔE_{ac} as activation energy. The Schottky diode current is either the thermionic emission current I_{TE} or the thermionic field emission current I_{TFE}:

$$I = \begin{cases} I_{\mathrm{TE}} & V_{\mathrm{applied}} > 0 \text{ (forward)} \\ -I_{\mathrm{TFE,reverse}} & V_{\mathrm{applied}} \leq 0 \text{ (reverse).} \end{cases} \tag{2.4}$$

Thermionic emission current is described by

$$I_{\mathrm{TE}} = A A^* T^2 \exp\left(-\frac{e\phi_{\mathrm{Bn}}}{k_{\mathrm{B}} T}\right) \left(\exp\left(\frac{eV}{k_{\mathrm{B}} T}\right) - 1\right), \tag{2.5}$$

which includes the effective Richardson constant [120]

$$A^* = \frac{4\pi e m_{\mathrm{eff}} k_{\mathrm{B}}^2}{h^3} \quad . \tag{2.6}$$

Thermionic field emission current for negative voltage is described by

$$I_{\mathrm{TFE},V<0} = A \frac{A^* T}{k_{\mathrm{B}}} \sqrt{\pi W_{00} e \left(-V + \frac{\phi_{\mathrm{Bn}}}{\cosh^2(W_{00}/k_{\mathrm{B}}T)}\right)} \exp\left(\frac{-e\phi_{\mathrm{Bn}}}{W_0}\right) \left(\exp\left(\frac{-eV}{\zeta}\right) - 1\right). \tag{2.7}$$

Here, W_{00}, W_0, and ζ are calculated by

$$W_{00} = \frac{eh}{4\pi} \sqrt{\frac{N_{\mathrm{D}}}{m_{\mathrm{eff}} \epsilon_{\mathrm{s}}}}, \tag{2.8}$$

$$W_0 = W_{00} \coth\left(\frac{W_{00}}{k_{\mathrm{B}} T}\right), \text{ and} \tag{2.9}$$

$$\zeta = \frac{W_{00}}{(W_{00}/k_{\mathrm{B}}T) - \tanh(W_{00}/k_{\mathrm{B}}T)} \ . \tag{2.10}$$

$e\phi_{\mathrm{Bn}}$, which is the effective barrier height, is computed by

$$\phi_{\mathrm{Bn}} = \phi_{\mathrm{Bn0}} - \Delta\phi = \phi_{\mathrm{Bn0}} - \sqrt{\frac{eE_{\max}}{4\pi\epsilon_{\phi_{\mathrm{B}}}}} = \phi_{\mathrm{Bn0}} - \sqrt[4]{\frac{e^3 N_{\mathrm{D}}(\phi_{\mathrm{Bn0}} - \phi_{\mathrm{n}} - V)}{8\pi^2\epsilon_{\phi_{\mathrm{B}}}^3}}\,. \tag{2.11}$$

The donor concentration goes into this equation through N_{D}. For the donor concentration in the above equations, $N_{\mathrm{D}} = z_{\mathrm{Vo}}N_{\mathrm{disc}}$ is assumed. The energy difference $e\phi_{\mathrm{n}}$, which is the energy between the Fermi level and the conduction band edge, also depends on the donor concentration. Based on the Boltzmann statistics, it is calculated by

$$\phi_{\mathrm{n}} = \frac{k_{\mathrm{B}}T}{e}\,\log\left(\frac{2(2\pi m_{\mathrm{eff}}k_{\mathrm{B}}T/h^2)^{3/2}}{z_{\mathrm{Vo}}N_{\mathrm{disc}}}\right). \tag{2.12}$$

For resistive switching to occur, the oxygen vacancy concentration must change. The ionic current I_{ion} of mobile oxygen vacancies between the plug and the disc is modeled by

$$\frac{\mathrm{d}N_{\mathrm{disc}}}{\mathrm{d}t} = -\frac{1}{z_{\mathrm{Vo}}e\,A\,l_{\mathrm{disc}}}\cdot I_{\mathrm{ion}}. \tag{2.13}$$

In the JART VCM v1b model, the plug is assumed as highly concentrated, infinite oxygen vacancy reservoir. The temperature and field dependent ionic current can be calculated by

$$\begin{aligned} I_{\mathrm{ion}} &= AJ_{\mathrm{ion,drift}} \\ &= A\left(2z_{\mathrm{Vo}}ea\nu_0 N\exp\left(-\frac{\Delta W_{\mathrm{A}}\left[\sqrt{1-\gamma^2}+\gamma\arcsin\gamma\right]}{k_{\mathrm{B}}T}\right)\sinh\left(\frac{az_{\mathrm{Vo}}eE}{2k_{\mathrm{B}}T}\right)\cdot F_{\mathrm{limit}}\right). \end{aligned} \tag{2.14}$$

Here, ν_0 is the attempt frequency, a is the hopping distance and N is the ion concentration. F_{limit} is a factor that limits the ionic current to avoid deviation from the maximum/minimum allowed concentration. The difference of forward and reverse jumps is introduced by the prefactor γ according to

$$\gamma = \frac{az_{\mathrm{Vo}}eE}{\pi\Delta W_{\mathrm{A}}}. \tag{2.15}$$

The driving concentration N in the drift current equation is chosen as the geometric

mean of both concentrations N_{disc} and N_{plug}:

$$N = \sqrt{N_{\text{disc}} \cdot N_{\text{plug}}}. \tag{2.16}$$

The electrical field for SET and RESET is calculated through

$$E = \begin{cases} \dfrac{V_{\text{disc}} + V_{\text{plug}}}{l_{\text{cell}}} & V_{\text{applied}} > 0 \text{ (RESET)} \\ \dfrac{V_{\text{disc}}}{l_{\text{disc}}} & V_{\text{applied}} < 0 \text{ (SET)}. \end{cases} \tag{2.17}$$

It is numerically ensured that the concentrations do not exceed the maximum and minimum concentrations $N_{\text{disc,max}}$ and $N_{\text{disc,min}}$.

A single temperature for the whole filament is used. It is calculated based on the dissipated power in the filament caused by the voltage drops across the plug and the disc:

$$T = (V_{\text{disc}} + V_{\text{plug}}) \cdot I \cdot R_{\text{th,eff}} + T_0. \tag{2.18}$$

T_0 is the ambient temperature. The temperature increase caused by the Joule heating is described using a singular effective thermal resistance $R_{\text{th,eff}}$.

2.3 Neuromorphic computing architectures employing memristive devices

This section introduces two important types of novel brain-inspired architectures that are thought to benefit from utilizing memristive devices, namely Fully Connected Neural Networks and Spiking Neural Networks. It is not meant as a full review, but should give the reader an impression of the possible target applications that are relevant to this work. For more detailed information, the reader is referred to recent review publications on the topic[42, 66, 121–125].

By definition of Carver Mead [126], neuromorphic architectures aim to translate the working principles found in biological systems into human-built hardware systems. The two fundamental building blocks of the human brain as it is understood today are synapses and neurons. Neurons receive, compute and generate signals from connected neurons. The connections are realized in the form of synapses, which pass the signals along and modulate them in strength and/or shape. Based on these rather loose

definitions, it is not surprising that a wide variety of brain-inspired concepts have developed over the past years. Network properties such as number of elements and interconnections, information encoding, output result interpretation and so forth may be freely chosen. In the context of this work, the memristive devices are solely investigated with respect to synapse application. Two network architectures are described in more detail in the following sections.

2.3.1 Fully Connected Neural Networks

The most frequently used and arguably the most straightforward approach to understand the concept of brain-inspired computing is a single layer, fully-connected network. A schematic example is shown in Figure 2.5. It consists of an input neuron layer and an output neuron layer. Each input neuron is connected individually to each output neuron by a synapse with a specific weight $w_{i,j}$. In the forward pass phase, also called inference, of the network learning, input information x_i is given to the input neuron layer. The signals are passed through the synapses to the output neurons, which accumulate the multiplied signals ($\sum_{j=1}^{n} w_i x_i$) and impose a nonlinear activation function on them. The obtained values from the output layer are the result of the network computation and serve as input to a problem-dependent algorithm that aims to solve the problem. The easiest example for this would be a function that determines the maximum output neuron signal and declares its assigned value the calculated solution to the problem. Other criteria are possible however. In the next stage, the backward pass, the result of the network is compared to the known label of the input, and an error value is calculated by the chosen loss function. The loss value is then propagated backwards through the network, resulting in a gradient value for

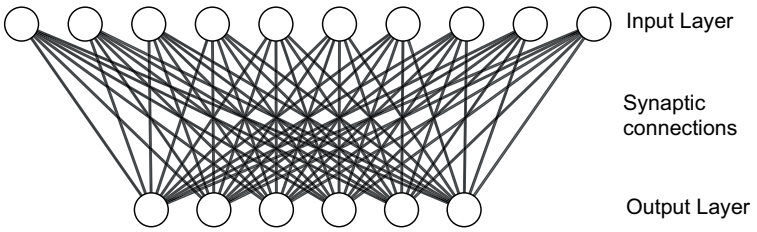


Figure 2.5: Example of a single layer fully connected neural network. Created with the web-based tool of LeNail [127].

each synapse weight. This gradient can then be turned into an update value, which is added to each of the current synapse values. The process of forward and backward pass may be iterated many times to increase the accuracy of the network. To increase the performance of such neural networks, more neuron layers can be introduced. This architecture is then called Deep Neural Network (DNN) and follows the same rules as described above, but consists of a significantly larger amount of neurons and synapses. A limited number of publications have demonstrated use of filamentary VCM devices as integral part of the structure on the side of neurons, such as [69, 128, 129]. In most cases, it is required to connect additional circuitry to the device as they do not intrinsically provide the features required for a neuron, such as volatility. However, their reversible and fast resistance transition has proven to be a beneficial property in this field. On the other side, application examples as synapses are numerous [49, 65, 68, 73, 74, 130, 131] owing to the state nonvolatility and the multi-bit capability.

2.3.2 Spiking Neural Networks

An alternative to DNNs that is motivated from the brain is the Spiking Neural Network (SNN). In this architecture, neurons are still connected to each other by synaptic elements, which modify the signals. However, the strict layer arrangement of DNNs is not necessary in SNNs. More importantly, SNNs utilize spike signals as information carrier in contrast to DNNs, where the information is encoded in clocked, rectangular signals, similar to a classical computer. Information encoding in spike form is inspired by the biological brain. Information can be stored in many ways in the spike domain. Spiking rate, spike shape, amplitude and relative timing of spikes are examples of information encoding.

Accordingly, the rules for defining SNNs are less strict compared to DNNs, and many sub-categories have evolved. Comprehensive reviews on the use of memristive devices in SNNs can be found in [132, 133]. The specifics of the utilized SNN in this work are explained in the respective Chapter 8.

3 Experimental Methods

This chapter introduces the experimental methods that are relevant in the scope of this work. Section 3.1 describes the different thin film deposition techniques that were employed to obtain the layers of the resistive switching devices. The most important layers, namely the oxides, were obtained by Atomic Layer Deposition (ALD), while the metal layers were fabricated using two different Physical Vapor Deposition (PVD) techniques, Sputtering and Electron Beam Evaporation. Next, Section 3.2 describes the steps for structuring the deposited layers into nano-crossbar devices that were solely used in the scope of this work. The main focus of this thesis is the electrical characterization of the obtained VCM devices. Therefore, Section 3.3 describes the employed measurement setups in detail.

3.1 Thin film deposition techniques

This section describes the two main thin film deposition techniques that are used in this work to fabricate the layer stack. First, the principle of oxide thin film ALD is introduced, and the details of the employed process for obtaining the hafnium oxide and titanium oxide layers are provided. The second component of the memristive device in this work are layers of tantalum, titanium and platinum metal. All metals can be deposited using sputtering and electron beam evaporation. Both methods were employed in this work. Hence, they are described and the relevant parameters are provided.

3.1.1 Atomic Layer Deposition

Principle of ALD

The principle of ALD was developed by Tuomo Suntola and Jorma Antson and patented in 1977 [134]. Figure 3.1 illustrates the process for arbitrary chemicals A and B. The basic definition of ALD is that a substrate surface, which is heated to sufficient temperature, is presented with vapor of a first chemical A (①), which can react with the surface to form a single element atomic layer (②). Subsequent removal, typically termed purge, of the vapor of chemical A and possible reaction products (③) and introduction of the second chemical B (④) leads to a formation of an atomic layer of B by reacting with the first layer of A (⑤). Element B and the possible byproducts of this reaction are removed from the reaction site (⑥), and the growth cycle is completed. By repetition of this alternating cycle, a film can be grown until the desired thickness is reached. The difference to other Chemical Vapor Deposition (CVD) techniques is this alternating pattern of injecting the precursor chemicals in contrast to simultaneous injection at the same time. Because the ALD reaction is a surface limited process [135], the growth can be controlled precisely by the number of growth cycles. Additionally, the process of surface saturation by atomic (sub-)monolayers ensures high uniformity and hence dense and thin layers. The resulting film can therefore be tuned precisely in thickness. Because the reactants are delivered to the surface as vapors, the growth is isotropic and the walls of etched holes with high aspect ratios in the sample surface can be covered homogeneously [136]. Further merits of ALD is the possibility to cover large areas, CMOS process compatible process temperatures and low impurity contents [137, 138].

Processes in this work

In the scope of this work, two processes are used to grow layers of HfO_2 and TiO_x. The same reaction chamber in an *Oxford Instruments* FlexAl^{TM} ALD system is utilized for both processes. Figure 3.2 (a) shows the tool, which is part of the Nanocluster tool that is located in the Helmholtz Nano Facility, Forchungszentrum Jülich GmbH. It is connected by a gate valve to a loadlock terminal, which is outfitted with an Edwards turbo molecular pump and a lamp heater. Both allow for effective removal of gas contaminants of loaded wafers by achieving a base pressure of about 10^{-9} mbar and a temperature of 200 °C, respectively. The precursor cabinet allows permanent

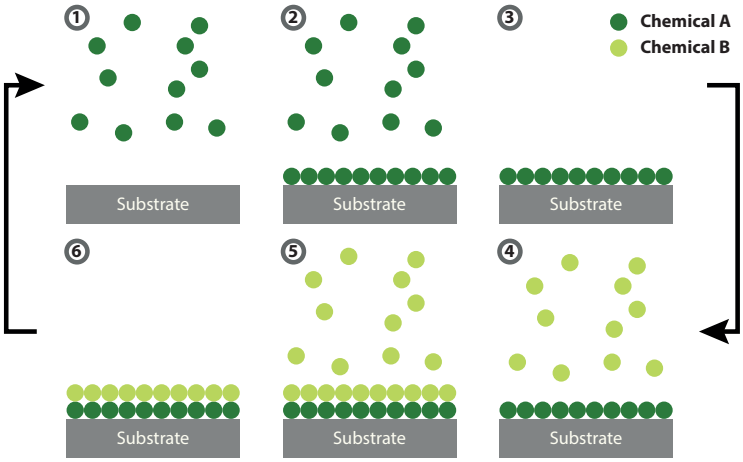


Figure 3.1: Schematic ALD principle. By alternating injection (① and ④), reaction with the surface (② and ⑤) and removal (③ and ⑥) of chemicals A and B, the layer is grown.

installation of up to six bubbled precursor containers. Three oxygen sources are available: A radio-frequency remote plasma generator can provide oxygen plasma, a water container can provide water vapor and an ozone generator. In Figure 3.2 (b) a cross-section of the reaction chamber is depicted. The precursors in this system are transported from their storage container in the precursor cabinet via a carrier gas, which is Argon. This type of ALD is called bubbler ALD since the carrier gas is injected below the liquid surface in the precursor containers. As written in Figure 3.2 (b), the reaction chamber is equipped with a remote plasma source as well as gas inlets for precursors and water vapor. A detailed description of the parameter optimization on this machine can be found in the PhD thesis of Alexander Hardtdegen [140]. The following parameters are the result of this optimization and were kept constant thoughout this work. The HfO_2 process in this work utilizes oxygen plasma as oxidizing agent. The table heater temperature is set to 300 °C. Note that this temperature is likely not reached at the substrate surface and is therefore not equal to the deposition temperature. The metal-organic precursor for this process is tetrakis(ethylmethylamido)hafnium (TEMAH, $Hf(NCH_3C_2H_5)_4$). Its structure is depicted in Figure 3.3 (a). During the growth, the precursor container is heated to 70 °C to adjust the vapor pressure of the liquid. The growth cycle for plasma assisted ALD of HfO_2 is depicted in (b). Here, the filled color regions indicate open ALD

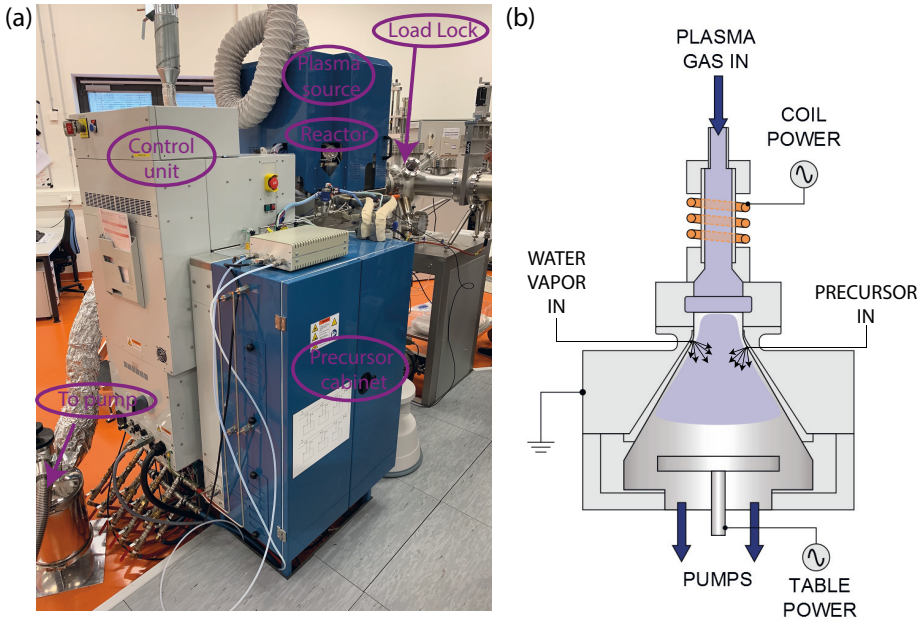
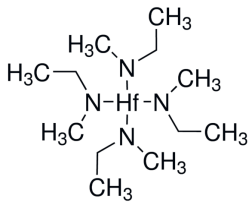


Figure 3.2: (a) *Oxford Instruments* FlexAl ALD located in the Nanocluster at Forschungszentrum Jülich GmbH. The visible parts are labeled. (b) Cross-section of the reaction chamber. Taken and adapted from [139].

valves. The second line labeled "Ar" indicates the gas line that goes directly into the chamber. The third line ("Ar purge") indicates the status of the gas line that goes through the precursor carrying line, but not through the containers. The fourth line ("Ar bubbler") is the line that goes through the precursor liquid and hence is the actual carrier gas. The process works as described for the schematic principle above. The first step is to inject the carrier Argon gas into the heated precursor container. At the same time, Argon gas is provided both through the direct line and the purge line to the chamber. During the surface reaction and the following precursor purge step, these gas flows are maintained. Afterwards, all ALD valves are closed for a duration of 3 s, allowing the exhaust pump to remove the TEMAH molecules and reaction products from the atmosphere. To ensure reliable plasma ignition, a dedicated oxygen gas flow stabilization step is performed next. During the subsequent plasma step, the surface layer is oxidized. Byproducts of this oxidation reaction as well as the remaining oxygen gas are purged from the chamber during the final step of the cycle.

(a)



TEMAH

(b)

	TEMAH injection	TEMAH purge	TEMAH pump	O_2 stabilisation	O_2 plasma	Post plasma purge
TEMAH						
Ar	100 sccm	100 sccm				100 sccm
Ar purge	250 sccm	250 sccm	20 sccm	20 sccm	250 sccm	250 sccm
Ar bubbler	250 sccm	20 sccm	20 sccm	20 sccm	20 sccm	250 sccm
O_2				60 sccm	60 sccm	
Plasma power					250 W	
Duration	1 s	15 s	3 s	3 s	3 s	6 s

Time

Figure 3.3: (a) Chemical formula of the Hf precursor tetrakis(ethylmethylamido)hafnium, short TEMAH. (b) Process flow diagram for one full plasma assisted growth cycle of HfO_2.

This process has a characteristic growth rate of about 0.1 nm cycle, which means that 30 cycles are necessary to reach the 3 nm thickness utilized in this work.

The process for the deposition of TiO_x is slightly different since it utilizes water vapor as oxidizing agent. Hence, it is called a thermal process. The precursor molecule of tetrakis(dimethylamino)titanium (TDMAT, $Ti(N(CH_3)_2)_4$) is illustrated in Figure 3.4 (a). The sample table temperature is also 300 °C. The precursor container is heated to 60 °C for the growth. The process parameters are listed in the process flow diagram in Figure 3.4 (b). During the first step, the chamber is pressurized by injecting Ar gas through the purge line. For the second step, which is the precursor injection, the ALD valve to bubble the precursor is opened and Ar carrier gas is injected into the chamber. The subsequent purge removes the reaction products and the remaining TDMAT from the chamber. The water vapor is injected by opening the according ALD valve for 20 ms, which is long enough to draw a sufficient amount of water vapor from the container. Note that the water container is not bubbled, but the vapor is drawn purely by the low pressure in the chamber. The water vapor and reaction products are purged for a duration of 20 s. This process results in a growth rate of 0.035 nm per cycle, which means that about 85 cycles are required to grow the desired thickness of 3 nm.

The described processes for HfO_2 and TiO_x are performed without removing the sample from the chamber. They are, however, seperated by a 5 minute long pump

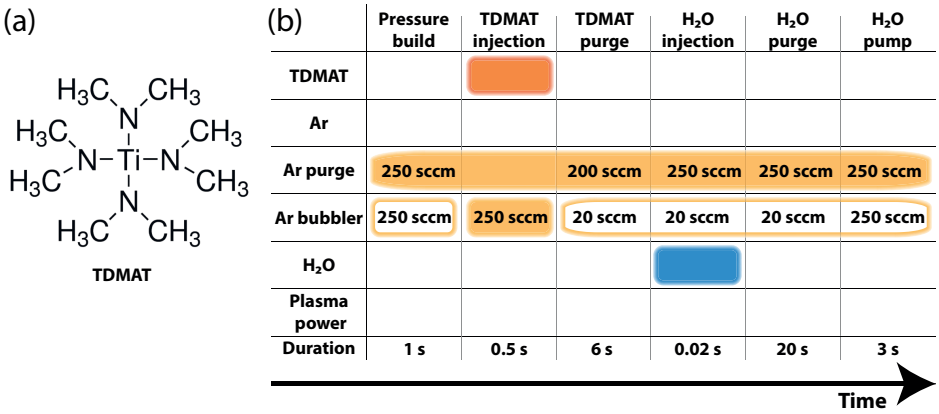


Figure 3.4: (a) Chemical formula of the Ti precursor tetrakis(dimethylamino)titanium, short TDMAT. (b) Process flow diagram for one full thermal growth cycle of TiO_x.

step, during which the chamber is evacuated. This step aims at removing possibly remaining TEMAH, oxygen gas and byproducts of the HfO_2 growth that might influence the subsequent growth of TiO_x. The resulting amorphous films are dense and coat the sample surface uniformly.

3.1.2 Physical Vapor Deposition

Physical vapor deposition (PVD) is the method of choice to deposit the metal layers in this work. In the scope of this work, tantalum, titanium and platinum films were employed in the fabrication electrodes. However, they were fabricated using different PVD methods. In the following, the two employed fabrication methods, namely sputter deposition and electron beam evaporation for bottom and top metal layers, respectively, are introduced.

Sputter deposition

Two main variants can be determined for sputter deposition, namely direct current (DC) and radio frequency (RF) sputtering. Both may be used in combination with magnetron sputtering. The substrate and the sputter target are mounted opposite to each other in a vacuum chamber. The process pressure in the chamber is typically kept between 10^{-3} and 10^{-1} mbar. Often Argon gas is used as the sputtering gas.

DC sputtering involves a constant voltage that is applied to the target while the substrate is grounded. Once the plasma is ignited, the Ar ions generated in the electric field are accelerated towards the target. The surface impacts of the ions cause a random collision sequence, which can result in ejection of near-surface target atoms [141]. Ejected atoms traverse the distance to the opposite substrate, adhere on its surface, and finally deposit there. The sputter yield, i.e. the number of ejected target atoms per incident Ar ion, depends on the surface binding energy of the target atom species, the relative masses of the ionized gas atoms and target atoms, the ion kinetic energy, and the impact incident angle.

In contrast, RF sputtering involves a capacitive coupling of the cathode, i.e. the target, to an RF generator. Because the mobility of the free electrons in the ionized gas is much higher, the cathode quickly develops a negative DC bias with respect to the anode. Subsequently, Ar ions are accelerated towards the cathode and the sputter process is initiated as described for the DC case. During one electric field half-cycle, positive charges are accumulated on the cathode. In the following half-cycle, they are neutralized by the electrons that are attracted. The net current for a full cycle is hence zero. RF sputtering is especially useful for non-conductive target materials since the static charging of the target is avoided, which would increasingly disrupt the plasma or prevent ignition altogether.

One modification to both variants is using magnetron sputtering. The deposition rate is increased by a permanent magnet mounted behind the target. It contributes a magnetic field of constant strength close to the sample surface. The moving electrons in the surface vicinity are confined to circular trajectories. Therefore, more Ar ions are generated on these trajectories. Additionally, the ionized gas hits the target surface at a more shallow angle, which increases the sputter rate. The local increase in sputter rate can create elliptical erosion marks in the target. The uniformity of the resulting film can be decreased using this method, but can be omitted by using a non-parallel sample-target constellation.

For the bottom electrode, a tantalum film serves as adhesion promoting layer between the Si/SiO_2 substrate and the platinum layer that is part of the actual VCM device. Both bottom metal layers are deposited in the same *Scienta Omicron* off-axis sputter tool, which is part of the Nanocluster in the Helmholtz Nano Facility in the Forschungszentrum Jülich GmbH. It features an extremely low base pressure of 10^{-10} mbar. The process parameters for the 5 nm thick Ta layer and the 25 nm thick Pt layer are as listed in Table 3.1:

Table 3.1: Sputter parameters for the bottom metal layer fabrication.

Parameter	Ta process	Pt process
Target manufacturer	*EVOCHEM Advanced Materials*	*EVOCHEM Advanced Materials*
Ar flow	30 sccm	30 sccm
DC power	–	40 W
RF frequency	13.56 MHz	–
RF power	100 W	–
Target tilt	51 °	51 °
Substrate rotation	12 rpm	12 rpm
Substrate heater temperature	22 °C	22 °C
Process pressure	0.003 mbar	0.0056 mbar
Deposition time	163 s	375 s

Electron Beam Evaporation

Another method to deposit metal layers that was used for the top metal layers in this work is electron beam evaporation. It involves generation of an electron beam by extracting electrons from a tungsten tip. This focused beam is redirected by a generated magnetic field to a cooled crucible which contains the desired thin film material. In contrast to sputter deposition this process requires the best possible vacuum to minimize collisions of gas molecules with the beam. The electron beam is able to locally vaporize the material in the crucible. The so derived atoms in the gas phase are distributed in the chamber and condensate on the cooler surfaces, including the sample that is mounted opposite to the crucible.

The parameters for the evaporation of titanium and platinum, i.e. the top electrode metal layers, are summarized in Table 3.2.

Table 3.2: Electron beam evaporation parameters for the top metal layer fabrication.

Parameter	Ti process	Pt process
Crucible material manufacturer	*MaTeck Material Technologie & Kristalle GmbH*	*EVOCHEM Advanced Materials*
Acceleration voltage	10 kV	10 kV
Beam current	500 mA	500 mA
Beam power reduction	9 % to 10 %	35 % to 36 %
Preheating duration	2 min	2 min
Process pressure	$5 \cdot 10^{-5}$ mbar	$5 \cdot 10^{-5}$ mbar
Deposition rate	0.5 Å	1.5 Å

3.2 Nano-crossbar fabrication

Industrial application of VCM-type devices require device sizes that are compatible with current CMOS technologies. Device footprints, i.e. the area that a device requires, should therefore be in the range of few ten nanometers. Therefore, the VCM device of this work is fabricated in nanometer sized crossbar structures. This integration process utilizes patterning by electron beam lithography (EBL), which is described in the following. A schematic view of an EBL writing system is shown in Figure 3.5. Electrons are extracted from a tungsten tip by an electric field and subsequently accelerated towards the sample. In the scope of this work, the acceleration voltage was kept constant at 100 kV, which allows for excellent contrast between illuminated and dark regions in the resist. After extraction and acceleration, the electrons are manipulated through a series of magnetic lenses and apertures to achieve a beam of defined focus and current. Both determine the required time for a region to receive the correct electron dose, which is the defining parameter for EBL resists. The resist that is used in this work is *MicroChemicals* AZ nLOF 2020. The dose for this resist was kept constant at 70 µC/cm^2 in accordance to a dose series experiment performed prior. The fabrication procedure, starting from the thermally oxidized silicon wafers, is described in the following. Sketches in Figure 3.6 illustrate the intermediate process steps and the final device geometry. First, the passivated substrates are covered by a sputtered bottom electrode layer of 5 nm Ta and 25 nm Pt (①). The *Scienta Omicron* sputter system that is described in Section 3.1.2 is used for this. The resist covering for the BE structuring is done in the HNF clean room. A (2:1) mixture of (*MicroChemicals* AZ EBR solvent : *MicroChemicals* AZ nLOF 2020) is applied on

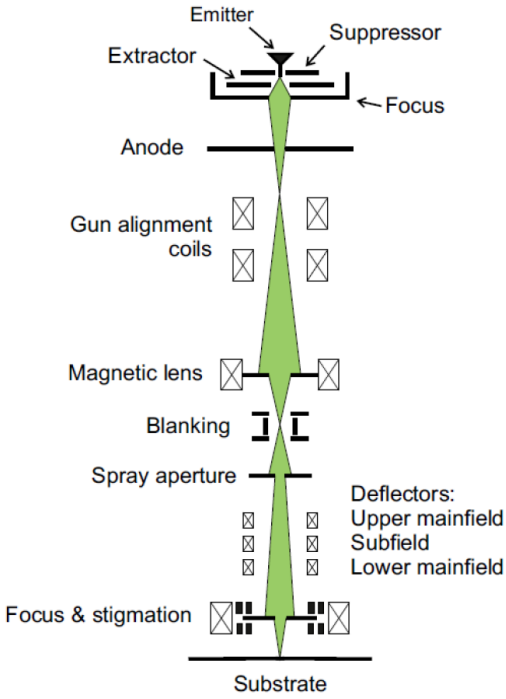


Figure 3.5: Schematic view of an electron beam lithography system. Electrons are emitted, focused and manipulated to illuminate the sample that is covered with EBL resist. Redrawn from [142].

the dried surface via spin-coating (4000 rpm for 1 minute) and soft-baking at 90 °C for 3 minutes. The dilution of the resist leads to a layer thinning down to about 220 nm in thickness. It is required to obtain narrow resist lines after development by avoiding high resist aspect ratios, which are likely to collapse. The resist is structured by EBL (②) as described above. The *Raith* EBPG 5200 EBL system is operated by Dr. Stefan Trellenkamp and Dr. Florian Lentz and is located in the HNF. After exposure to the electron beam, the activated resist requires a short post exposure bake, which is done at a temperature of 110 °C for 1 minute. In the following, the bottom electrode layers are etched via Reactive Ion Beam Etching (RIBE). This dry etching process allows anisotropic etching of the metal lines. Details of RIBE processes can be found in references [143–145]. The remaining protective resist is removed by the according solvent based on dimethyl sulfoxide (DMSO) (③). Subsequent treatment in acetone and isopropyl alcohol removes remaining residues. The structured BE have line widths

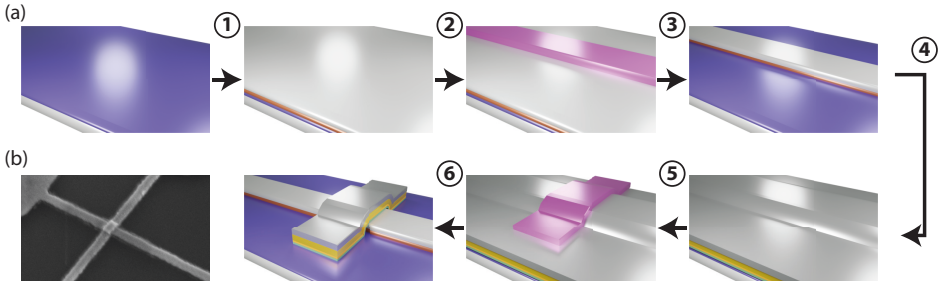


Figure 3.6: (a) Nano crossbar fabrication process. (b) SEM image of a 100 nm x 100 nm crossbar.

between 60 and 100 nm. The BEs are covered with the remaining layers for the device stack, i.e. 3 nm HfO_2, 3 nm TiO_x, 10 nm Ti and 20 nm Pt (④). The oxides are deposited via ALD as described in Section 3.1.1. The metals are obtained by electron beam evaporation, see Section 3.1.2. The top electrode and the oxides are structured using the same e-beam lithography process as for the BE. The fabricated resist lines are perpendicular to the BE lines and identical in width (⑤). Through RIBE etching and resist removal in the identical process as before, the crossbar structure is finalized (⑥). Figure 3.6 shows a Scanning Electron Microscopy image of the final structure. Note that no length scale is depicted due to the tilting of the sample, which illustrates the topography. The physical characterization of the sputtered, evaporated and ALD grown layers is documented in the respective recent works on technology [140, 145, 146]. The reader is referred to these works for details on the layer characterization. Layer control by X-Ray Reflectivity (XRR) measurement was performed on blank SiO_2 substrates. Details on XRR can be found in [147]. Figure 3.7 shows the XRR measurements for the bottom electrode in (a), the oxide double layer in (b) and the top electrode metals in (c) as black symbols. The actual layer thicknesses, which are obtained from the fits drawn in red, are written in the diagrams. Note that the layer thicknesses are not perfectly in line with the nominal values, yet the sum of the actual thicknesses is very close or identical to the sum of the nominal values. This is due to the difficulty to identify the precise interface position by XRR. All layers have been tested individually for their deposition rate. Hence, the origin of the deviations from the nominal values is the limiting fitting accuracy and the actual layer thicknesses are likely identical to the nominal values.

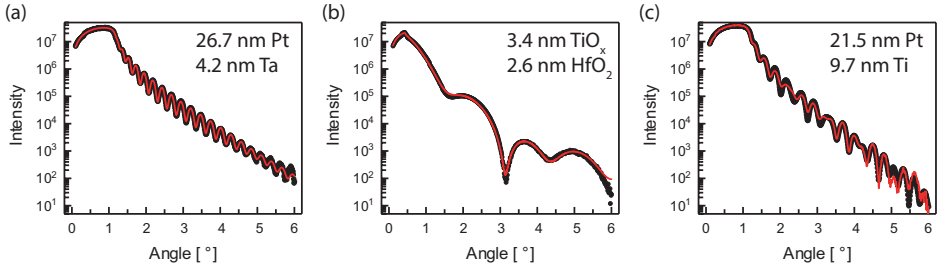


Figure 3.7: X-Ray Reflectivity measurements of the layers on blank SiO_2 substrates. (a) Bottom electrode metals. (b) ALD grown oxides. (c) Top electrode metals.

3.3 Electrical Analysis

This section focuses on the measurement setups that were used for electrical characterization of the VCM devices. It is mainly split into two categories, namely setups that are used for sweep measurements and setups that are used for application of square voltage pulses, typically on shorter timescales than the sweep measurement systems.

3.3.1 Sweep measurement setups

One of the most employed technique for electrical measurement of VCM devices is the current-voltage sweep. It consists of three consecutive voltage ramps. Starting at 0 V, the voltage is first decreased to a negative voltage, then increased to a positive voltage and decreased back to 0 V. At the same time, the current that goes through the device is recorded. In this work, two setups were employed for sweep measurements, which are introduced in the following.

Agilent B1500A setup

The first measurement setup for sweeps is an *Agilent B1500A* instrument, which is connected to a *Karl Suess Microtec PA-200* semiautomatic prober. Within the scope of this work, the measurement setup is used for electroforming and voltage sweeps on previously electroformed samples to verify normal function. The advantage of this setup is that it allows automated sequential contacting and measurement of devices on the test sample by programmable table movement. The probe station is depicted

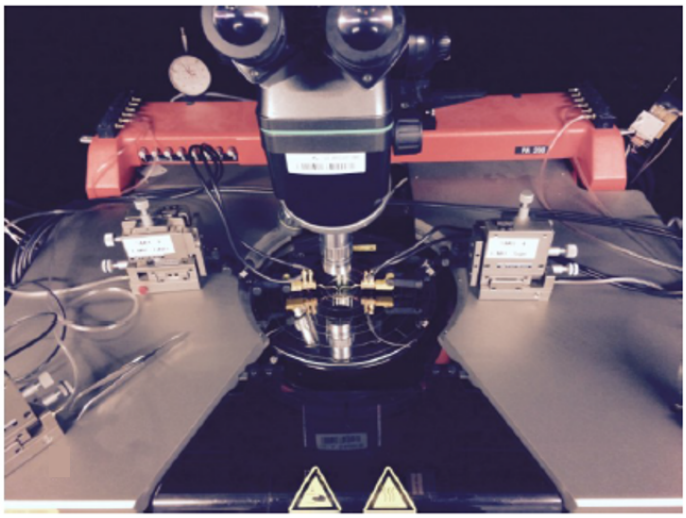


Figure 3.8: Photo of the *Karl Suess Microtec PA-200* semiautomatic prober. The four installed Source Measurement Units of the *Agilent B1500A* are connected to the visible probes.

in Figure 3.8. For specifications about the instrument, the reader is referred to the product page of the device [148].

Custom current compliance sweep setup

The second setup for measuring voltage sweeps was developed by Tyler Hennen from the IWE 2, RWTH Aachen University with the goal of reducing the current overshoot that occurs in commercial measurement hardware when the current compliance is reached through a fast transition such as the electroforming and SET process of a VCM device. The measurement setup and its significant advantages over commercially available equipment was described in detail in the paper of Hennen et al. [149] and is therefore only briefly introduced in this thesis.

The setup consists of multiple individual components, which are shown in Figure 3.9. The arbitrary waveform generator *Rigol DG5102* applies the programmed voltage waveform to the tungsten needles contacting the device under test (DUT). The signal is simultaneously recorded on channel A of the *Picoscope 6403C* oscilloscope. To achieve the set goal of reducing the current overshoot phenomenon, a

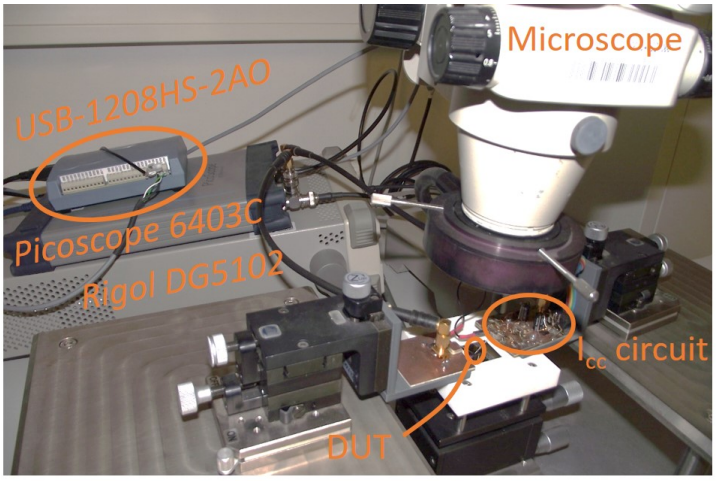


Figure 3.9: Custom built measurement setup designed by Tyler Hennen from IWE 2, RWTH Aachen University. Main components are labeled. The reader is referred to reference [149] for schematics of the current compliance circuit.

current limiting circuit is connected directly to the needle, acting as an active current compliance, see Figure 3.9. This circuit has been developed specifically for measuring ReRAMs [149]. To determine the current flow through the device, the voltage V_{out} is recorded on channel B of the *Picoscope 6403C*. The current is obtained by dividing the voltage by a known resistance. The main component of the current compliance circuit are two transistors that are supplied with the appropriate voltages by the *USB-1208HS-2AO Hub*. The adjustable range of the current compliance is determined by the resistance of $R_{\text{compliance}}$. During this work, $R_{\text{compliance}}$ was chosen to achieve tunable current compliance between 50 µA to 800 µA. The reader is referred to the paper of Hennen et al. [149], where the exact circuit for the current compliance is illustrated and explained. For this work, a few additional points apart from the circuit are to be noted:

- All measurements performed on this setup use a triangular voltage signal, i.e. are sweep measurements. Initially, 0 V is applied and sweeped to the negative polarity $V_{\text{SET, stop}}$, which is defined as the minimum applied voltage. After reaching $V_{\text{SET, stop}}$ the voltage is sweeped back to 0 V and in the positive polarity to $V_{\text{RESET, stop}}$, the maximum applied voltage, and finally reduced back to 0 V. During the sweep the slew rate is kept constant and is between 6000 V/s and

7000 V/s in this work. This means it is about three orders of magnitude higher than the *Agilent B1500A* setup.

- The *Picoscope 6403C* operates at its maximum sampling rate of 1.25 billion samples per second. Since the internal memory of 512 MSamples is limited, long voltage waveforms are split into shorter sequences. Further, the *Rigol DG5102* waveform generator outputs small discrete voltage steps when a voltage ramp is programmed. Since the change from one step to the next is typically slower than the sampling frequency of the oscilloscope, the signal is oversampled. The data is hence smoothed by calculating the moving average of the voltage and the current and down-sampled to 1000 data points per sweep.

- After an *I-V* curve is measured, a significant current offset of roughly 10 µA can be observed. This offset is subtracted during the analysis. For this, the HRS branch of the sweep is linearly fitted around the 0 V mark and the y-intercept current is subtracted from all data points. The issue is illustrated in Figure 3.10 (a). Figure 3.10 (b) shows the resulting changes in the switching curve. After the correction of the offset the HRS resistance can be calculated. All shown *I-V* curves from this setup are offset corrected. However, the static resistance value is still relatively imprecise. Instead, the differential resistance given for measurements from this setup. In first approximation, a linear *I-V* relation is assumed. It is fitted between (-0.1) V and (-0.4) V via a least squares algorithm.

- The fixed voltage resolution of channel B on the *Picoscope 6403C* oscilloscope leads to a limited current resolution. Currents below 100 nA are therefore impossible to resolve with sufficient accuracy. The device properties measured in the scope of this work are all above this setup limitation.

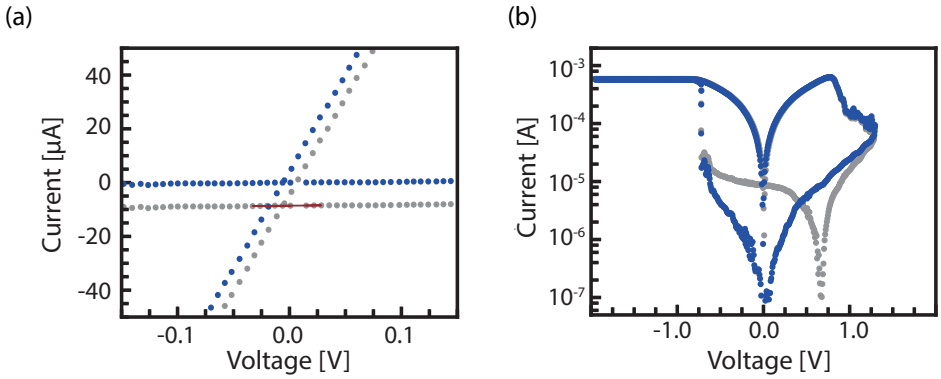


Figure 3.10: Offset subtraction procedure for the custom compliance circuit measurement setup. (a) As recorded data (grey), linear fit to the HRS (red) and resulting shifted *I-V* curve (blue). (b) The difference between the corrected curves on a logarithmic scale.

3.3.2 Pulse measurement setups

For several results of this work, square voltage pulses were employed. Square voltage pulses contain a different type of information about the device properties compared to voltage sweep measurements. Additionally, they are the likely method of operation in the later integrated devices, unless special driver circuitry is created. The operation with square voltage signals can be divided into two principles. In the first one, the transient currents are recorded and are analyzed. Read signals before and after the switching pulse are performed, but serve only as verification of the observations that occurs during the switching pulse, e.g. a lower resistance after a successful SET pulse compared to before the switching pulse. The second principle ignores the transient switching current or does not record it altogether. In this method, the analysis is carried out by studying the read currents prior and after the switching pulse. Once again, the method of choice depends on the target of the investigation, and the latter is closer to the likely mode of operation in future applications.

In the following, three setups are introduced that are employed in this work to generate square voltage pulses.

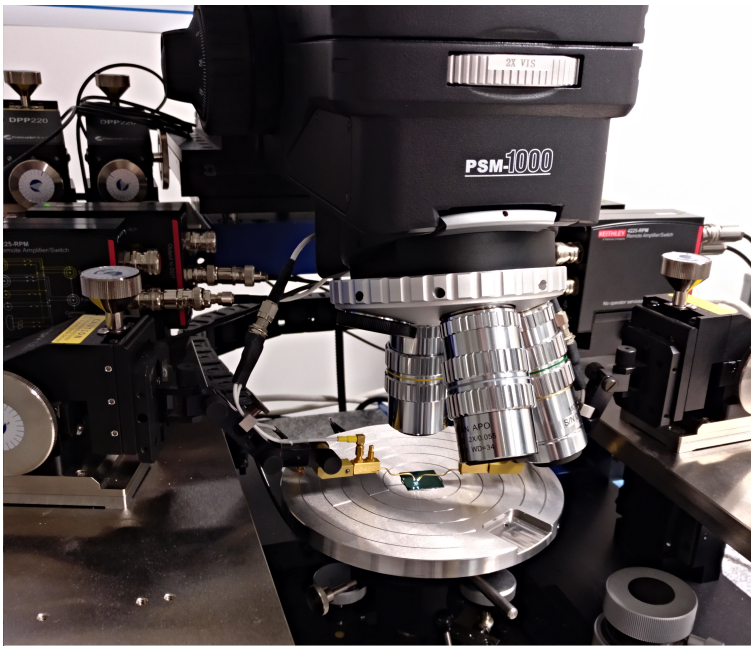


Figure 3.11: *Keithley 4200-SCS* setup with sample stage, probe needles, microscope and PMUs in the background.

Keithley 4200 SCS setup

The first introduced setup for square voltage pulses is a *Keithley 4200-SCS* equipped with *4225-Pulse Measurement Units* (PMUs) with additional *4225-RPM Remote Amplifier/Switch* modules which allow detection of currents in the nanoampere range. Due to the capability of recording the transient currents it falls into the first category of setups in this work. Figure 3.11 shows a photo of the probe station, which is a *Cascade Microtech MPS150* system equipped with a *Motic PSM-1000* microscope and two manually operated *DPP220* probe positioners. The *Keithley 4200-SCS* instrument can generate square current resolved voltage signals up to a voltage of +/- 40 V at times down to 70 ns with rising and falling flanks of 20 ns. The current range is adjustable between 100 nA and 200 mA, allowing current resolution for a wide range of device resistances. For more specifications of the device, the reader is referred to the product page [150]. The device is controlled by a custom LUA script which sends commands and receives data via the GPIB interface.

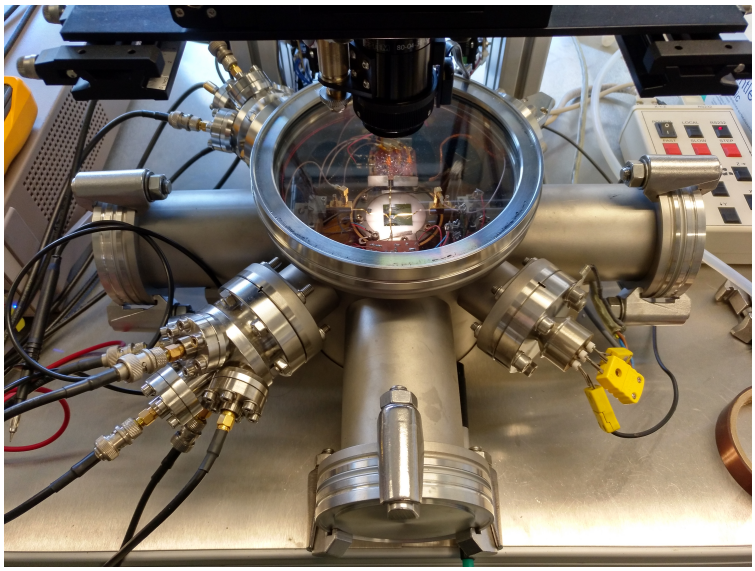


Figure 3.12: Photo of the resistance network pulse measurement setup probe station including sample stage, probe needles and cable connections to outside the chamber. Measurement instruments are located in a nearby instrument rack.

Resistance network pulse measurement setup

The second setup used for fast pulse measurements in this work consists of an *Agilent B1110A* Pulse-/ Pattern Generator and a *Tektronix TDS6804B* Digital Storage Oscilloscope. The setup was created from these components by Bernd Rösgen and Marcel Gerst. In this setup, the voltage signals are generated by the *Agilent B1110A* waveform generator. Simultaneous to being applied to the device, the voltage pulse is monitored with the *Tektronix TDS6804B* oscilloscope on Channel 1. Further, the voltage drop over a 39.2 Ωresistor connected in series to the DUT is recorded and amplified by a *Texas Instruments OPA847* high speed operation amplifier. The resulting signal is a 22 times magnified signal and is monitored on Channel 2 of the oscilloscope. The voltage is then recalculated into a current that flows through the device. A second resistor and *OPA847* amplifier results in a 422 times magnified signal, which is routed to Channel 3 on the oscilloscope.

The device resistance read-out is designed differently to other setups. A *Stanford Research Systems SR830* Lock-In Amplifier is utilized for this task. An AC voltage

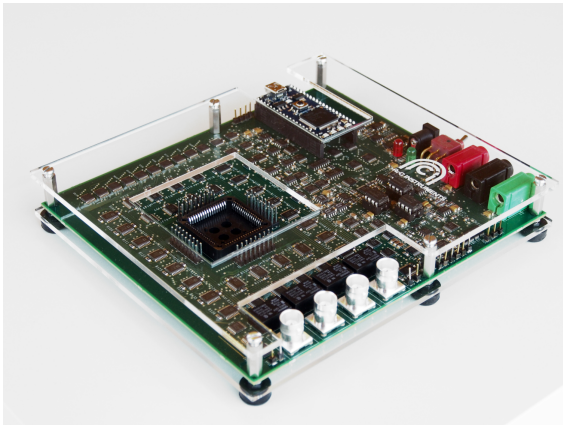


Figure 3.13: Photo of the *ArC ONE* instrument measurement unit. Three ways of connecting DUTs are possible: by connecting a single DUT to the BNC connectors, by using the pin banks to connect a probe card or by placing a bonded chip in the according socket. Photo taken from [151].

signal with a bias of 470 mV and a frequency of 1 MHz is applied to the described resistance network. This results in a sinusoidal voltage read signal of 25 mV on the DUT. Figure 3.12 shows the setup probe stage. The setup is additionally capable of low temperature measurements and atmospheric variations in the vacuum chamber. In the present work, this feature is not utilized. The instruments allow for square voltage pulse testing over a range of nine orders of magnitude in time, 10 ns to 10 s. Voltage amplitudes between 100 mV and 10 V in both polarities are possible. Due to the setup's capability of measuring the transient currents it also belongs in the first category of setups in this work.

ArC ONE measurement setup

The third electrical measurement setup for square voltage waveforms in this work is an *ArC ONE* instrument by *ArC Instruments*. Because transient currents of the switching pulses are not recorded, this device falls into the second defined category of voltage pulse measurement devices. Instead, the device characterization is carried out by applying the programmed voltage pulse and reading the resistance afterwards. An image of the measurement unit is shown in Figure 3.13. The *ArC ONE* offers multiple options to connect single or multiple DUTs. The first is to connect a single

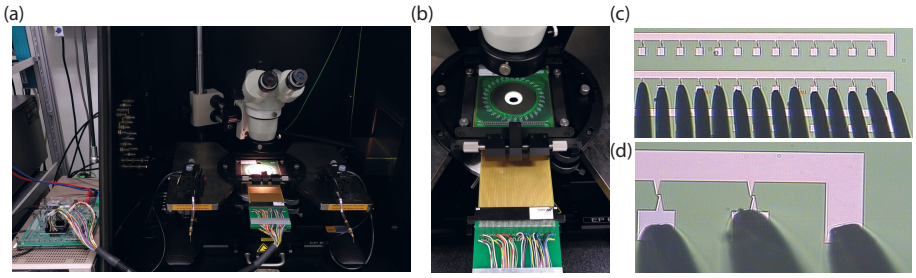


Figure 3.14: (a) Probecard configuration of the *ArC ONE* instrument. (b) Probecard in holder. (c) Microscope image of the probecard in contact with a single line array. (d) Zoom into the BE contact and the first two devices.

DUT to the BNC connectors. Multiple DUTS can be contacted either by using the pin banks via a probe card or by placing a bonded chip in the according socket on the actual measurement unit. In the scope of this work, the connection to single devices was done via the BNC connectors directly to needle probes. Contact to multiple DUTs was done by connecting the pin banks on the instrument board to a probe card. The probe card is specifically designed for the device configuration in this work and features 33 individual needles, which are routed to the respective pins via an accordingly designed cable. It was fabricated specifically for this sample geometry by *High Tech Trade* company. The described setup is illustrated in Figure 3.14 (a) and (b). Microscope images of the probecard in contact with a sample are shown in (c) and (d). The *ArC ONE* instrument comprises a single waveform generator and routing options to one or multiple outputs. Therefore, individual voltage signals on different probes are not possible. However, it allows application of voltage pulses down to 70 ns in duration and up to 12 V in amplitude. For further specifications, the reader is referred to the instrument's product page [151].

4 Resistive switching in HfO_2-based devices

This chapter describes the phenomenon of resistive switching of the devices investigated in the scope of this work. Starting from the as fabricated state, the electroforming process is described. The subsequent resistive switching is examined with regard to the operating parameters and the observed relations. Endurance characteristics are described. The focus of this chapter, however, lies in the detailed description of SET and RESET kinetics in the range from 1 s to 10 ns. It will be shown that both processes are made up of two kinetically consecutive steps with different physical origins. Their individual length is determined by the previous device state and the SET or RESET voltage amplitude applied in each case. Consequently, gradual and abrupt switching characteristics emerge. Both are possible pathways to build artificial synapses that involve analog conductance tuning. In single cells, the transition time may be used to tune the device conductance gradually. In the case of abrupt switching, the switching stochasticity with respect to cycle-to-cycle and device-to-device variation is investigated. The findings of this chapter form the foundation of the following chapters, which exploit the interrelation found in the study of the kinetic behavior.

4.1 Electroforming

As described in section 2.2, an electroforming step is typically required for obtaining resistive switching. Unless the oxide layer has been modified beforehand, e.g. by ion implantation and oxygen deficiency engineering, the required electroforming voltage is significantly higher than the SET voltage for the subsequent switching operation. Figure 4.1 a shows the electroforming with an active current compliance of 50 µA and first RESET curves without current compliance of 64 identically fabricated

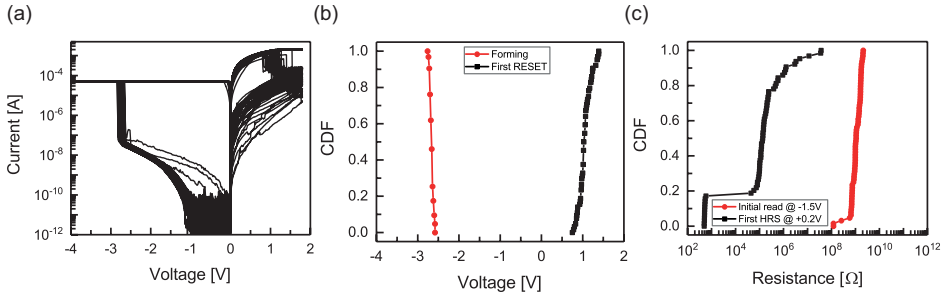


Figure 4.1: Electroforming of the investigated stack Pt / 3 nm HfO_2 / 3 nm TiO_x / Ti / Pt. (a) Current voltage curves of 64 devices of 100 nm x 100 nm size. 53 devices subsequently showed standard filamentary VCM-type switching. (b) Corresponding statistics of the electroforming voltage and the first RESET voltage. (c) Initial and first HRS resistance statistics.

100 nm x 100 nm devices. A high uniformity which is typical for ALD grown layers is observed. Figure 4.1 b demonstrates this uniformity. The electroforming voltage distribution is very sharp with a median value of -2.66 V while the extreme values in the distribution are -2.76 V and -2.60 V. The subsequent RESET process is not as defined as the initial current voltage behavior, which is attributed to the stochastic nature of the filament formation. However, the RESET voltage distribution is still narrow with a median of 1.03 V and a minimum of 0.75 V and a maximum of 1.39 V. While all devices in the ensemble showed electroforming, around 17 % are stuck in a very low resistive state below 1 kΩ, which can be seen in Figure 4.1 c. Concerning the yield of the ensemble, it is 100 % for electroforming, but only 17 % for reproducible switching. Since the dominant failure type is LRS stuck, it could be reasoned that the current overshoot obtained for the abrupt current increase during electroforming could cause this issue. The existence of the overshoot is without doubt since the RESET current is up to two orders of magnitude higher than the set current compliance of 50 µA during electroforming. This assumption would mean that the yield of the manufactured device could theoretically reach 100 % if the overshoot could be minimized.

To summarize, the nano-crossbar Pt / HfO_2 / TiO_x / Ti / Pt devices require an electroforming step prior to stable filamentray-type resistive switching. The residual leakage current of pristine devices is very low and the forming voltage distribution of various devices is quite narrow. The devices are formed into the ON-state with a negative voltage polarity applied to the Pt BE. By this, oxygen vacancies are injected into the metal oxide film from the Ti electrode. For cells of 3 nm HfO_2 and 3 nm TiO_x,

the electroforming voltage is about -2.66 V±0.10 V at a sweep rate of around 1 V/s. These uniform electroforming characteristics hence comply with silicon technology voltages. The device yield after the first RESET is sufficiently high for the studies conducted in this work.

4.2 Miniaturization of VCM devices

One of the most recognized features of VCM devices is their superior scalability. Especially applications that strive to mimic functions in the brain are expected to require many millions of devices on a limited footprint. Other neuromorphic concepts will likely benefit from dense integration in the future as well. Therefore, miniaturization of VCM devices has been a topic in research and industry from an early stage of development onward [6, 20, 152–155]. The main device size of this work is 100 nm x 100 nm as described in Section 3.2 due to the yield advantage over more scaled devices. Higher reproducibility and yield is possible in industrial production lines. However, this section highlights the potential for miniaturization of the devices used in this study. Aside from the 100 nm x 100 nm crossbar device size, 60 nm x 60 nm devices in crossbar geometry and 40 nm x 40 nm as plug devices were tested. The fabrication pathways are described in Section 3.2 and Appendix A. Figure 4.2 (a) shows the Scanning Electron Microscopy (SEM) images of a described 100 nm x 100 nm crossbar structure. In the upper panel, the contact pads for probing, labeled BE for bottom electrode, and TE for top electrode, are visible. The lower image is a zoom-in on the actual crossing point of the electrodes. The sketch in Figure 4.2 (b) illustrates the stack structure at the crosspoint. The Pt / HfO_2 / TiO_x / Ti / Pt device stack can be identified. The sketch in Figure 4.2 (c) depicts the device geometry for the plug device. Note that the dimensions are not to scale in these sketches. The Pt / HfO_2 / TiO_x / Ti / Pt device stack exists only where the SiO_2 layer is etched. Elsewhere, the stack in the overlap area is Pt / SiO_2 / HfO_2 / TiO_x / Ti / Pt, effectively preventing switching apart from the plug location. SEM imaging of the plug geometry device are challenging due to the small hole size and the top electrode roughness. Figure 4.3 shows the according electrical measurements for the three devices. Electroforming results are depicted as grey solid lines, while five subsequent switching cycles are drawn in black color. All three devices exhibit typical forming and switching properties. However, there are two major differences. The first is the current scaling in the pristine state, i.e. the electroforming curve before the rapid current increase. As expected, the current in

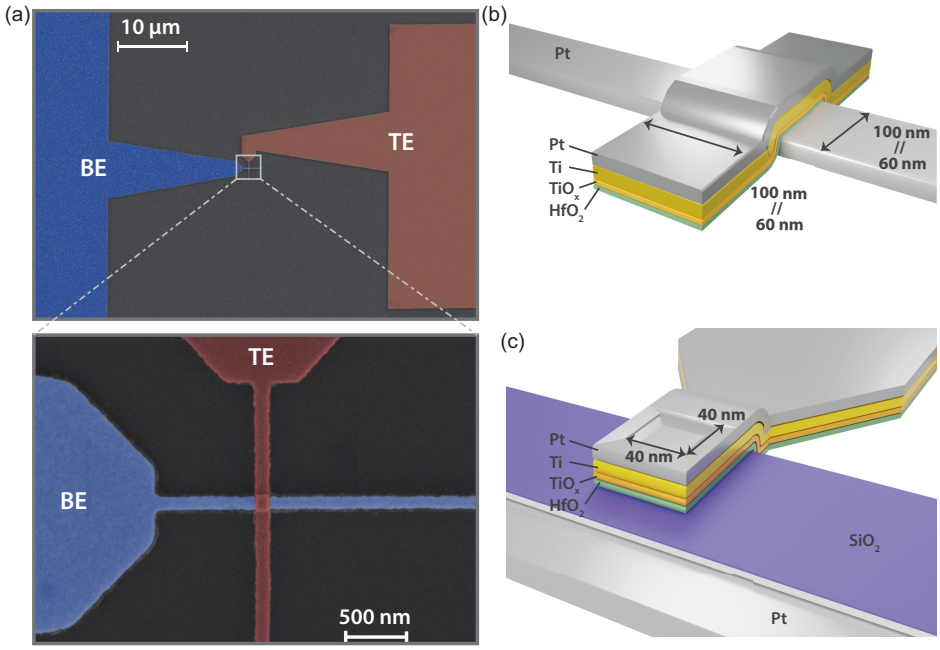


Figure 4.2: Miniaturization of the device in this work. (a) SEM pictures of the 100 nm x 100 nm device in crossbar geometry. Color added in post-processing. (b) Sketch of the device in cross-point geometry. Lines are fabricated in 60 nm and 100 nm width. (c) Sketch of the plug device with 40 nm x 40 nm etched area.

this state is decreased with decreasing device area. The second observation is, that the RESET for the 60 nm x 60 nm devices occurs at higher voltage and current compared to the 100 nm x 100 nm and 40 nm x 40 nm devices. The switching is also fairly abrupt, while it is more gradual for the other two devices. This effect can be attributed to the increased line resistance of the 60 nm wide and 1 µm long metal lines that connect the device to the contact pads. The appearance of the RESET transition is severely changed by a series element [107, 113, 156]. Avoiding the series resistance effect in scaled crossbar devices could be achieved with two strategies. The first is to use better conducting materials for the metal lines such as copper. However, this comes with its own issues as such materials are typically unwanted in proximity to silicon-based devices. The second strategy is to alter the non-lateral device geometries such as metal layer thickness t_{met} and line length l_{met}. An increase of the former and decrease

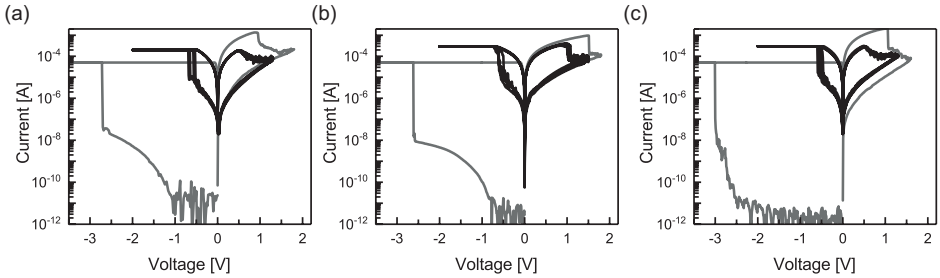


Figure 4.3: Miniaturization of the device in this work. Forming (grey solid curves) and subsequent switching cycles (black lines) (a) of the 100 nm x 100 nm and (b) of the 60 nm x 60 nm devices in crossbar geometry and (c) of the plug device with 40 nm x 40 nm area.

of the latter can reduce the metal line resistance R_{met} according to

$$R_{met} = \rho \cdot \frac{l_{met}}{b_{met} \cdot t_{met}}. \tag{4.1}$$

For scaled devices with the given materials, the specific resistivity ρ and the metal line width b_{met} are fixed. In the scope of this work, this approach was not tested due to fabrication limitations. Because the 40 nm x 40 nm device is not a crossbar device, but a plug device with a 40 µm wide bottom electrode and a 1 µm wide top electrode, the series resistance is low even compared to the 100 nm x 100 nm devices. Hence, the RESET does not appear abrupt in this device. Other than the RESET, the switching appears almost identical within the usual device variation, which is to be expected for filamentary devices where the typical assumed filament size is in the range of a few tens of nanometers in diameter.

To summarize, the successful device miniaturization of VCM devices is demonstrated. The expected device function is demonstrated when the device size is reduced down to 40 nm x 40 nm.

4.3 Endurance characteristics

Sufficient endurance is one of the key prerequisites for obtaining statistically meaningful data sets of ReRAM devices. For all device tests, the summed number of switching cycles should remain below the critical cycle number for device degradation. Ideally, a safety margin is kept, since many device tests impose significantly more stress on

the device than a controlled current voltage sweep with an active current compliance. Nonetheless, endurance testing provides a good guideline for the onset of degradation effects. For achieving high cycle numbers in endurance tests in reasonable times, the measurement procedure is typically split into two sections with the purpose of reducing the data traffic between measurement device and user PC which is often the limiting factor. The majority of the switching cycles is not recorded and therefore termed "blind cycling". In decadal spacing, measurement cycles are performed. This procedure is described in Figures 4.4 (a) and (c) for a typical rectangular pulse endurance and a sweep endurance, respectively. The first is measured with the *Keithley 4200* system, see Section 3.3.2. The latter is recorded with the current compliance circuit setup described in Section 3.3.1. In the first case, the switching itself is not recorded. Instead, the read currents after the switching pulse are measured and the resistance is calculated. In the case of a sweep endurance, a full sweep cycle is recorded and the read resistance is calculated according to the procedure described in Section 3.3.1. The endurance measurement results of the two modes are shown in Figures 4.4 (b) and (d). The rectangular pulsed endurance is shown for a single device while the sweep endurance comprises measurements of 10 devices. Therefore, Figure 4.4 (d) shows the results of each device as grey line as well as the median for HRS and LRS as black and red line, respectively. The SET process of the rectangular pulse endurance was performed at a voltage of -1.0 V for 1 µs, while the RESET was carried out at 1.45 V for 1 µs, also. Read signals were 1 ms in length and -0.2 V in amplitude. In the sweep endurance, a single blind cycle had a duration of 10 µs, during which the voltage was swept from 0 V to -2.0 V, up to 1.3 V and back to 0 V. During the SET process, a current compliance of 600 µA was active. To achieve more accurate read resistance during the measurement cycles, the total sweep duration was increased to 1 ms.

In summary, the devices show stable and reproducible switching behavior for at least one million switching cycles even for pulsed switching without active current compliance. It is known that optimized switching voltages and algorithms tailored to the device properties can yield higher endurance numbers [26, 157–159]. It is expected that such improvements are also possible for the device of this work, However, the endurance numbers presented in this section of the work demonstrate that the device is stable enough for the studies conducted in the following chapters.

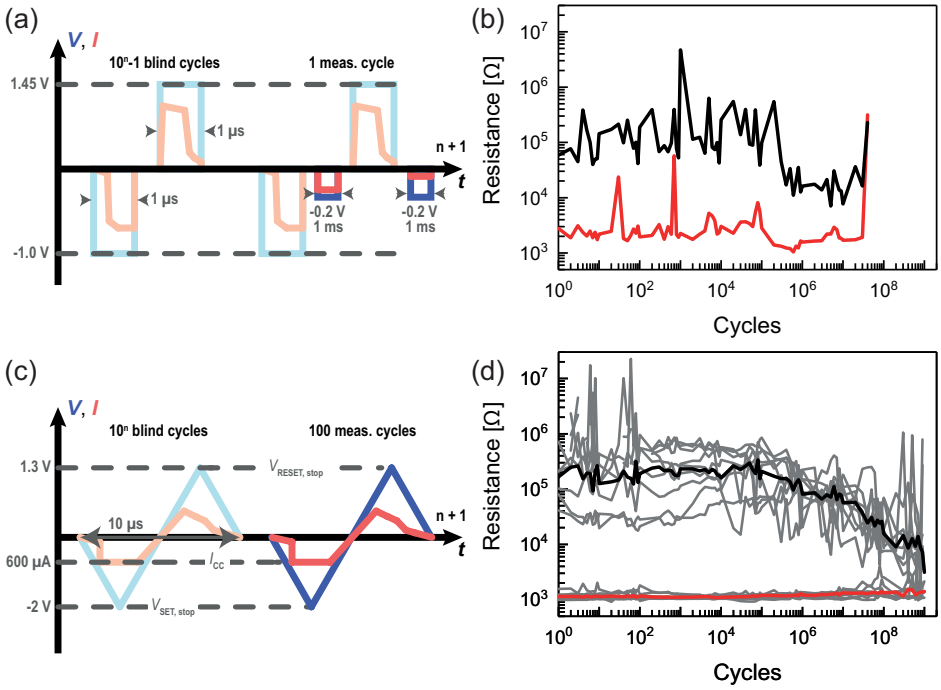


Figure 4.4: Endurance characteristics of the tested devices. (a) Endurance measurement scheme with rectangular pulses. (b) Results of endurance measured with rectangular pulses. Abrupt device failure occurs after $5 \cdot 10^7$ switching cycles. (c) Measurement scheme with triangular voltage pulses, including a current compliance. (d) According results for 10 devices (grey lines). The median for HRS and LRS is shown in black and red line, respectively. Gradual device failure occurs between 10^7 and 10^9 cycles.

4.4 Statistical analysis of resistance state tuning for the voltage sweep mode

In this section, the focus is on the controlled sweep operation of the investigated devices. By using the setup described in 3.3.1, a statistical meaningful number of data points can be collected in a reasonable amount of time. Additionally, the fast and reproducible current compliance of this setup is crucial for studying the influence of the SET and RESET operation on the resistance states. For the systematic investigation of switching dependencies, a pre-defined measurement scheme was used,

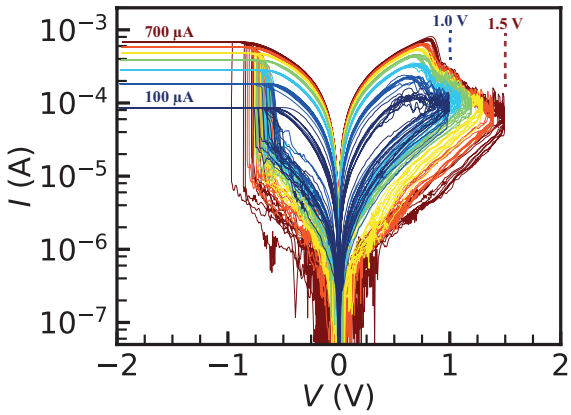


Figure 4.5: Typical *I*-*V* measurement sweeps in semilogarithmic view. Only every tenth sweep of 100 for each parameter set is shown. The external parameters [I_{CC}; $V_{RESET,\,stop}$] are on the diagonal of the matrix. From blue to red: [100 µA; 1.0 V], [200 µA; 1.0 V], [300 µA; 1.1 V], [400 µA; 1.2 V], [500 µA; 1.3 V], [600 µA; 1.4 V], [700 µA; 1.5 V].

which is hereafter referred to as the "matrix measurement". Here, the values of current compliance I_{CC} and maximum applied RESET voltage $V_{RESET,\,stop}$ are varied in discrete steps in two nested loops. The outer loop, which sets the value of I_{CC} starts at 100 µA and increases by $\Delta I_{CC} = 100\,\mu A$. In the inner loop, $V_{RESET,\,stop}$ is varied from 1.0 V to 1.5 V in steps of 0.1 V. In total, 6 x 7 = 42 different values of I_{CC} and $V_{RESET,\,stop}$ were applied. At each parameter pair [I_{CC}; $V_{RESET,\,stop}$] 100 *I*-*V* sweeps are performed from triangular voltage signals. The sweep duration is kept constant at 1 ms per sweep, and the applied voltage at -2.0 V for the SET process. Therefore, the effective sweep rate is between 6000 V/s and 7000 V/s, depending on the $V_{RESET,\,stop}$ value. A selection of sweeps from a typical matrix measurement are displayed in Figure 4.5 in a color coding. Here, 7 representative combinations of [I_{CC}; $V_{RESET,\,stop}$] are shown, referring to the main diagonal of the measured matrix. Data reliability has been confirmed on multiple devices and across samples. Reproducibility of the switching behavior at certain [I_{CC}; $V_{RESET,\,stop}$] parameters was confirmed by repetition of parameter combinations after finishing of the matrix routine. For better visibility, only every tenth out of the one hundred sweeps per [I_{CC}; $V_{RESET,\,stop}$] pair is shown.

Some universal switching behaviors of VCM-type ReRAMs can also be observed in this measurement.

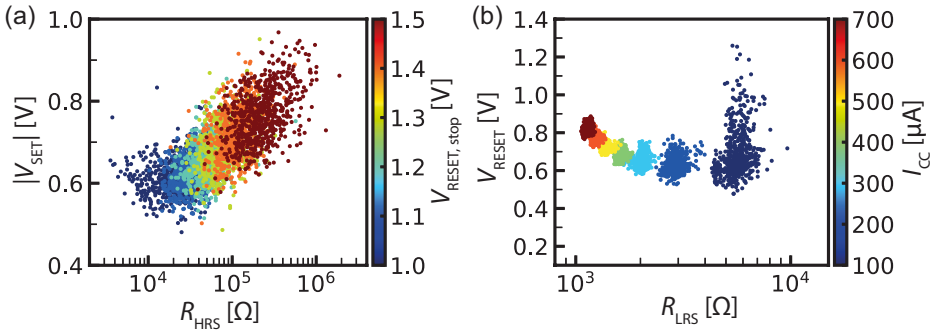


Figure 4.6: Typical sweep parameter dependencies: (a) R_{HRS} can be controlled by the applied RESET stop voltage, $V_{\mathrm{RESET,\,stop}}$, and influences the subsequent SET voltage. (b) R_{LRS} is governed by the applied current compliance, I_{CC}, during SET and influences the subsequent RESET voltage in the presence of a series resistance in the same resistance range as the LRS.

- First, the HRS read value, R_{HRS}, depends on $V_{\mathrm{RESET,\,stop}}$.

- Second, the SET voltage is increased if the previous R_{HRS} is higher [108, 110, 159, 160]. Figure 4.6 (a) illustrates this correlation. The R_{HRS} is evaluated as described in 3.3.1. In these measurements, an increase in R_{HRS} by two orders of magnitude leads to an increase of the SET voltage of about 0.3 V. The repeated cycles were essential for robustly detecting this trend, since the SET voltage variation is about 0.2 V even for fixed values of $V_{\mathrm{RESET,\,stop}}$.

- Third, the R_{LRS} is adjustable by the level of I_{CC} [6, 37, 161].

- Fourth, the RESET switching voltage V_{RESET} is almost constant under changing of I_{CC}, unless R_{LRS} approaches a similar value to the series resistance causing an increase in V_{RESET} [107, 113, 156, 159, 162]. This relation is depicted in Figure 4.6 (b), which shows an increase of V_{RESET} by about 0.2 V when R_{LRS} is reduced from about 6 kΩto roughly 1 kΩ. The series resistance of around 800 Ωis attributed to the small cross section of the lines of the nano-crossbar devices.

- Fifth, the device current at the RESET point I_{RESET} is similar to the programmed compliance current I_{CC} [161]. In Figure 4.5 it is seen that this behavior is fulfilled for I_{CC} values higher than 400 µA, while RESET overshoots are seen for smaller I_{CC} levels.

In the analysis of the matrix measurement, it is observed that the HRS level undergoes several changes as it is influenced by the external parameters. The most significant influence is posed by varying the $V_{\mathrm{RESET,\,stop}}$ as discussed above. However, an additional feature emerges when the current compliance is varied on top of that. Combining a moderately low $V_{\mathrm{RESET,\,stop}}$ and a high I_{CC} during the preceding SET operation results in a comparably low HRS. In contrast, a higher HRS is reached at the same $V_{\mathrm{RESET,\,stop}}$, when the preceding I_{CC} was lower. For high $V_{\mathrm{RESET,\,stop}}$, this observation is in fact reversed: Here, a high and a low current compliance during the preceding SET operation enables a RESET to a relatively higher and lower HRS, respectively. For intermediate voltages, a HRS is achieved, that is almost independent of the I_{CC} or the LRS prior to the RESET. The transition between the three regimes is very gradual, and does not occur spontaneously in contrast to the deep RESET behavior reported in [6]. Figures 4.7 (a), (b) and (c) depict details of this observation, showing the recorded sweeps for $V_{\mathrm{RESET,\,stop}} = 1.0\,\mathrm{V}$, $1.3\,\mathrm{V}$ and $1.5\,\mathrm{V}$, respectively. Because the inherent resistance variability overlaps with the described effect on the HRS level, it was necessary to acquire a significant number of sweeps to identify this behavior. In Figure 4.7 (a), the HRS is strongly degraded with increased I_{CC}. Figure 4.7 (b) illustrates the almost identical R_{HRS} at $V_{\mathrm{RESET,\,stop}} = 1.3\,\mathrm{V}$ for all I_{CC}. Figure 4.7 (c) shows the slightly increasing R_{HRS} trend for $V_{\mathrm{RESET,\,stop}} = 1.5\,\mathrm{V}$. This observation is statistically analyzed by the boxplot graph in Figure 4.7 (d). The box positions are slightly shifted from the actual I_{CC} positions to increase the readability. The dashed lines serve as guide to the eye and highlight the three regimes.

In addition to the representation in Figure 4.6 (a) and (b), the individual R_{HRS} data points are depicted in Figure 4.8 in an alternate form. Here, the HRS level is plotted against the $V_{\mathrm{RESET,\,stop}}$. At each voltage, the observed values are split into subgroups with respect to the preceding current compliance value, which also reflects the previous LRS level, see Figure 4.6 (b), and shown in color in Figure 4.8. Note that only every second tested I_{CC} is included in the graph. A slight voltage shift of the I_{CC} subgroups is imposed for graphical demonstration purposes. As a general statement, the overall expected behavior of resistance increase with increased $V_{\mathrm{RESET,\,stop}}$ is confirmed. However, the splitting of the current compliances at every $V_{\mathrm{RESET,\,stop}}$ reproduces the trend observed from Figure 4.7 (d). At $V_{\mathrm{RESET,\,stop}}$ of $1.0\,\mathrm{V}$, a higher current compliance leads to a relatively lower HRS level compared to the usage of a low current compliance. At intermediate voltages, the HRS tunability is nearly constant, regardless of the value of I_{CC}. In contrast, at a high $V_{\mathrm{RESET,\,stop}}$ of

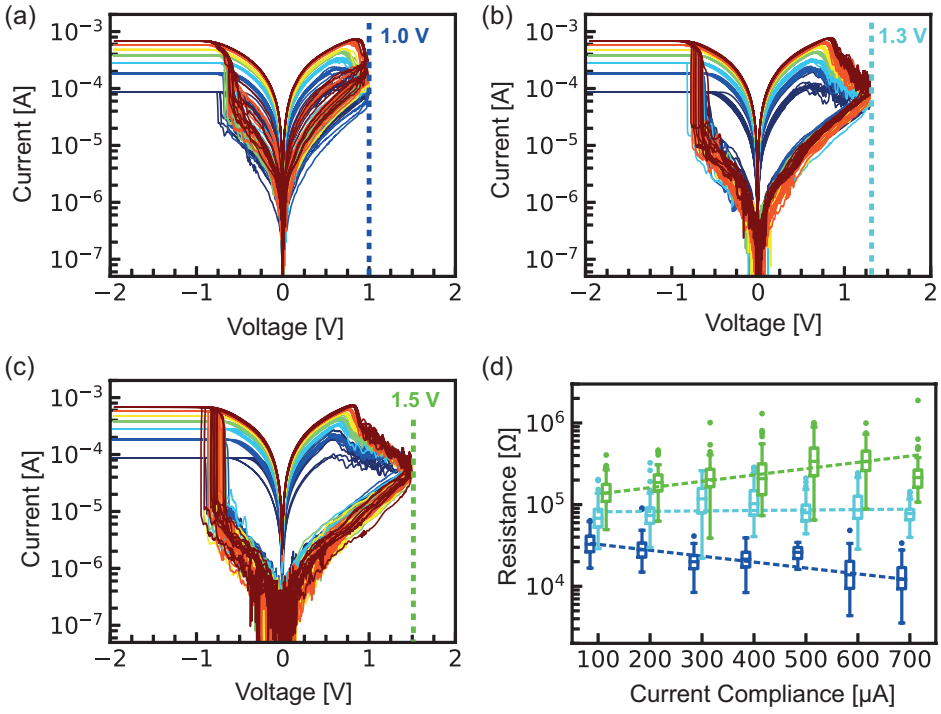


Figure 4.7: HRS current-voltage interdependence observed in the sweep experiments performed by the matrix measurement protocol. In each panel, the sweeps of all 7 I_{CC} values are shown for a single value of $V_{RESET,\,stop}$. (a) 1.0 V. (c) 1.3 V. (d) 1.5 V. (b) The reversal of order of the HRS level with increasing $V_{RESET,\,stop}$.

1.5 V, a higher I_{CC} enables the RESET operation to reach higher HRS values, while for lower I_{CC} values only relatively lower HRS levels are achieved. This phenomenon of complex HRS tunability is highlighted by the spline curves for 100 µA and 700 µA in blue and red color, respectively.

As seen in the presented figures, the device exhibits all typical features of a standard filamentary VCM-type resistive switching cell as described in several review articles and text books [6, 159–161]. Yet, the described HRS tunability phenomenon by means of a parameter cross-dependency has not been reported. Two main reasons may be responsible for this absence of representation: First, an extensive number of repeated sweeps at each parameter pair [I_{CC}; $V_{RESET,\,stop}$] is required to identify an effect that has the same order of magnitude as the inherent variability of the HRS. The effect could only be significantly demonstrated by measuring at least 100 cycles

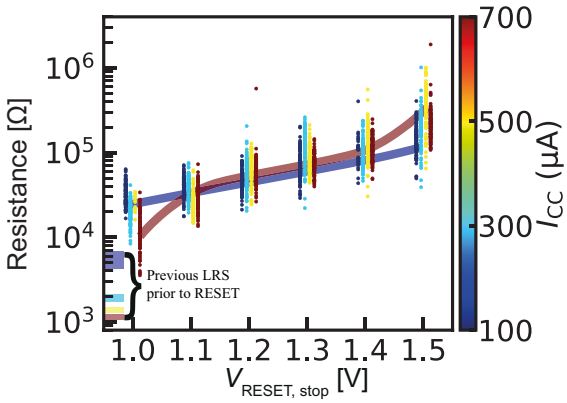


Figure 4.8: Analysis of the reachable HRS resistance values as a function of the current compliance and $V_{\mathrm{RESET,\,stop}}$ derived from the matrix measurement. At each value of $V_{\mathrm{RESET,\,stop}}$, 4 I_{CC} values are shown. A slight voltage shift is added for better readability. At low voltages, the RESET process achieves only low HRS for high current compliance. At intermediate voltages, the effect of I_{CC} on HRS is neglectable. At high $V_{\mathrm{RESET,\,stop}}$, higher HRS are achieved for high I_{CC}. Splines serve as guide to the eye.

per parameter combination. This means, that the device must endure a minimum of 5000 stable cycles before the onset of degradation. Additionally, this minimum number may even be higher in the presence of higher variability. The second reason may be the required systematic measurement protocol. Typically, when a large number of cycles is desired, the external parameters are kept constant [26, 159, 163], whereas the described phenomenon is only observable with the systematic variation as performed in the described matrix measurement.

The JART VCM model that is presented in this work in Section 2.2.3 and in other works [43, 115, 116] is not able to deliver an explanation for the shown phenomenon in the full extent. Therefore, the accurate physical interpretation is challenging due to the cross-coupling interference of electrical and thermal effects in filamentary-type VCM devices [97, 115].

Yet, the observed effect may be interpreted individually in the three voltage regimes. In the low voltage regime, a higher I_{CC} induces low HRS values at constant $V_{\mathrm{RESET,\,stop}}$. It is known that a series resistor impacts the RESET process if its resistance is in the range of R_{LRS} [107, 113] because the voltage is divided between the variable resistance and the constant series element. By using a high I_{CC}

and low amplitude for $V_{\mathrm{RESET,\,stop}}$, the switching element approaches the series resistance and the voltage divider effect becomes visible. As soon as the limiting effect of the series element is overcome, the effect vanishes due to the strong switching time nonlinearity. Hence, it is only visible for the lowest $V_{\mathrm{RESET,\,stop}}$ of 1.0 V. At intermediate $V_{\mathrm{RESET,\,stop}}$, the RESET process dynamics are strongly accelerated according to the extreme voltage-time nonlinearity inherent to VCM devices [43, 107]. The oxygen vacancies residing at the Pt / HfO_2 interface can be retracted irrespective of the previous configuration in the LRS for this voltage regime. Therefore, only very minor deviations between the I_{CC} levels are observed. The fluctuations in the voltage range of 1.2 V to 1.4 V are attributed to variability. In the voltage regime of 1.4 V and above, two effects are important. They are visible in Figure 4.7 (d) and Figure 4.6 (a). Caused by the increase in $V_{\mathrm{RESET,\,stop}}$ and the following R_{HRS} increase, the SET voltage is also increased. This is not unexpected and reported in literature[108]. Accordingly, the SET process causes a higher power dissipation, which is obtained by the product of I_{CC}, which is equal to before, and the measured SET voltage. This additional power dissipation is expected to trigger accelerated oxygen excorporation at the ohmic electrode interface, leading to a more dense oxygen vacancy configuration at the Pt / HfO_2 interface. The RESET switching dynamics for this ionic configuration then could be significantly accelerated compared to the case of low $V_{\mathrm{RESET,\,stop}}$ [43, 107]. The additionally created or moved oxygen vacancies are removed to a large extent, together with the usual amount of filament retraction. Thus, the result is a reduced remaining defect concentration and a higher R_{HRS}.

Strong programming conditions are often employed to achieve larger resistance windows. However, they are also known to degrade the endurance [164]. The hypothesis of additional creation and accelerated retraction can explain this observation. It can be assumed that the cell contains a finite number of oxygen vacancies that can be created and used for resistive switching before the device begins to degrade. If stronger programming conditions trigger enhanced generation, the reservoir of oxygen vacancies that can be utilized for switching is used up at a faster rate, hence the onset of device degradation begins earlier. As described above and expected from Section 4.3, no degradation effects were observed in the described experiment. Likely, the onset of degradation is beyond 1 million cycles even for the unfavorable measurement conditions.

In summary, this section described the phenomenon of a systematic HRS dependence both on RESET stop voltage and on the current compliance during resistive

switching. Through statistical analysis, the effect was distinguishable from the inherent resistance variability. Two effects were detected: On one hand, high I_{CC} combined with low $V_{RESET,\,stop}$ leads to R_{HRS} lowering due to the voltage divider effect. On the other hand, the increase in V_{SET}, which is caused by higher R_{HRS}, causes higher power dissipation at the same I_{CC}. Additional defects created during this process trigger an accelerated RESET process at high $V_{RESET,\,stop}$ and consequently a higher R_{HRS}. While such strong programming conditions can boost the resistance window, it can lead to faster device degradation.

4.5 Exploiting the switching kinetics of HfO_2-based ReRAM devices for SET and RESET operation

Parts of this section are taken from [43]. Hence, some figures contain the abbreviations LCS and HCS, which stand for low conductance state and high conductance state and represent the conductance counterparts to HRS and LRS, respectively. The motivation for this unit change lies in the comparably simpler calculation of the associated current at a given reading voltage, since it is a multiplication and not a division. In accordance, the following section discusses effects on conductance change frequently. HRS and LRS may be used interchangeably with LCS and HCS.

As discussed in 2.2, the dynamics of the conductance change of a typical filamentary VCM device is asymmetric in nature: The transition from HRS to LRS is quite abrupt due to a positive feedback between current increase and Joule heating[97, 100]. In contrast, the opposite RESET process is gradual in nature due to a negative thermal feedback and, eventually, the counteracting forces of drift and diffusion of oxygen vacancies approaching equilibrium concentration[112]. Based on these properties programming of multiple conductance states is achieved either by controlling the current during SET, e.g. by using a transistor in series to the ReRAM[161], or by changing the RESET voltage amplitude exploiting the gradual RESET transition[30, 33]. A layer stack modification to the typical VCM ReRAM device can change the switching properties[162, 165–167]. The introduction of an inherent conduction limiter (ICL) has a positive impact on reducing SET variability and has opened up the possibility of a gradual transition on either side, increasing single device and network performance[44, 113, 159]. In a synaptic application, such a gradual conductance

change for both SET and RESET is desired, i.e. analog conductance states should be programmable. With the background of the strong switching kinetic nonlinearity of SET and RESET, the capability of the devices to perform such analog programming should be investigated individually. Therefore, the following sections will discuss the SET switching kinetics and the RESET switching kinetics individually. The conditions that allow analog programming that are found in these sections are applied to the demonstration of analog conductance tuning. Comparison with a compact model in the final section elucidates the physical mechanism.

SET switching kinetics

In the VCM-type system $Pt/HfO_2/TiO_x/Ti/Pt$, the TiO_x layer acts as the aforementioned ICL. The reader is referred to [113] for a detailed investigation on the impact of the introduction of an artificial interface titanium oxide layer on resistive switching in voltage sweep mode. The complete stack design, including the equivalent circuit used in the compact model, is illustrated in Figure 4.9 (a). The equivalent circuit shows the different components that are combined in the device under testing (DUT). The conductance of the metal connections (G_{line}) and the titanium oxide interface layer ($G_{\mathrm{TiO_x}}$) can be combined into a conductance G_{ICL}, which represents the ICL. For simulation purposes, the HfO_2 layer is split into a plug and disc region, where the plug is conductive and represents an infinite reservoir of oxygen vacancies. In Figure 4.9 (b), two sweep measurements of bipolar resistive switching (BRS) of a single device are shown. The slew rate of the voltage ramps was set constant at 1.0 V/s. In the first case (solid black line), the initial starting LCS is low. Negative bias applied to the Pt BE leads to an abrupt increase of the current at a voltage of -0.75 V. The current through the device after the SET process is initially limited by the ICL. At a voltage below -1.3 V, a measurement system sided current compliance of 900 µA is active. The conductance is close to the maximal achievable value. At positive bias applied to the Pt BE, the device switches back to the LCS at a voltage of 1.3 V. As the voltage is stopped at 1.6 V, the device is in a low LCS again, which promotes another SET process with abrupt characteristic in the subsequent BRS cycle. The second operation mode, drawn in colored lines, each representing an individual conductance state contrasts with the first operation mode. Here, the initial LCS is more conductive. The first switching loop is taken for a voltage sequence from 0.0 V to -0.5 V to 0.0 V. The stop voltage of the SET sweep is consecutively decreased from -0.5 V to -0.85 V

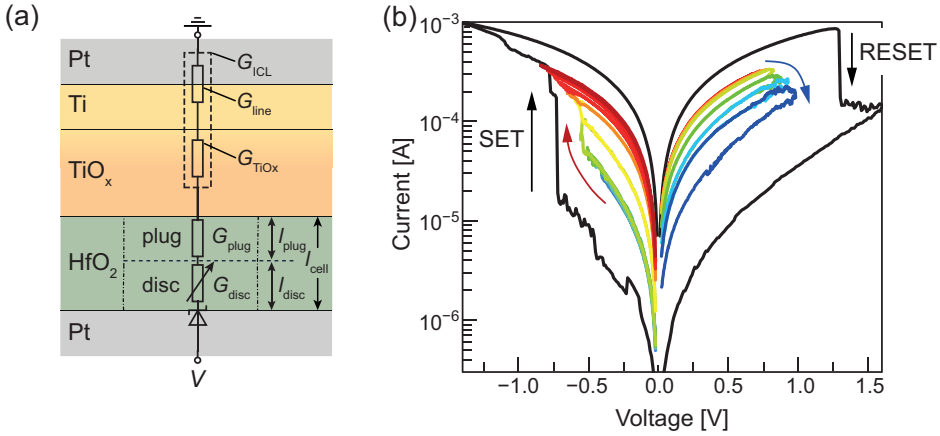


Figure 4.9: (a) Stack design of the nanosized 25 nm Pt/3 nm HfO_2/3 nm TiO_x/10 nm Ti/20 nm Pt devices and equivalent circuit including the conductance-limiting element G_{ICL} composed of contributions from the lines, G_{line}, and from the inherent TiO_x layer, G_{TiO_x}. The HfO_2-based memristive element is characterized by the conductive filament which is divided into a conductive plug, G_{plug}, and the resistive disc regime, G_{disc}; dimensions not to scale. (b) Current-voltage sweeps measured for the device in (a) visualizing the two switching operations of the BRS SET and RESET, which are the abrupt mode (black line) and gradual mode (colored lines). Reproduced with permission from [43].

in steps of 10 mV. Each negative sweep results in a new, separable conductance state, which is indicated by the different colors. In this operating mode, no measurement sided current compliance is necessary. Analogously, upon applying triangular voltage signals (0.0 V to $V_{RESET,\,stop}$ to 0.0 V) with increasing positive amplitude to the Pt BE, a gradual decrease in the conductance is achieved for $V_{RESET,\,stop}$ between 0.75 V and 1.0 V. Starting the switching hysteresis with the device in a low LCS leads to an abrupt SET and abrupt RESET characteristic, which is due to the thermoelectric coupling during the SET event and the voltage divider effect during RESET. In contrast, switching loops recorded in the defined limits of moderate LCS and HCS level enable gradual bipolar BRS for the SET and the RESET process. This shows, that the initial conductance state plays a decisive role for achieving the desired switching properties. This state-dependence of the switching mode is further investigated by analyzing the current transients for constant voltage pulses. For this measurement procedure, the device is initialized with a defined LCS. Then, a negative SET voltage pulse of variable duration and amplitude is applied. Figure 4.10 (a) shows the

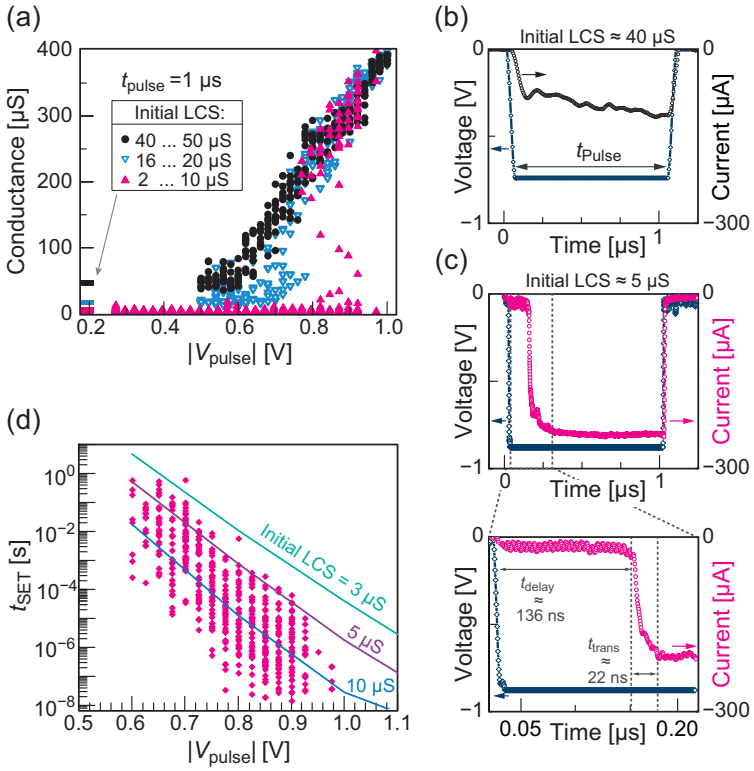


Figure 4.10: (a) Observed conductance values after 1 µs SET pulse with varying amplitude for low (closed triangles), intermediate (open triangles), and high (filled circles) initial LCS level. (b) Analog transient current response during SET obtained for a high initial LCS and a SET voltage of 0.72 V. (c) Abrupt transient current response during the SET pulse obtained for a low LCS level and a SET voltage of 0.85 V. (d) SET switching kinetics study revealing the variability of the experimental delay time at given SET voltage amplitudes. The simulation results (solid colored lines) show that the differences in the delay time can largely be explained through a variation of the initial LCS values. Reproduced with permission from [43].

read conductance states obtained by 0.2 V read signals after applying a voltage pulse with 1 µs width and variable amplitude. A clear dependence of the read conductances from the voltage pulse amplitude and from the initial LCS state (indicated by colors) is obtained. For the lowest initial LCS of about 2 µS to 10 µS (pink triangles) two regimes appear. At amplitudes below $|V_{pulse}| = 0.75$ V only minor deviations from the initial LCS are measured. At higher voltage amplitudes, a mixture of successful

and unsuccessful SET events is observed. While in general a higher amplitude leads to a higher HCS, intermediate conducting states appear as well indicating an incomplete switching event. In contrast, the highest initial LCS conductance state (closed black circles) shows an almost linear relation of conductance increase with rising $|V_{\mathrm{pulse}}|$ starting from about 0.625 V. Intermediate conductance levels, which are inaccessible from the lowest LCS, are reproducibly addressable. For intermediate initial states of 16 to 20 µS (open blue triangles), only a slight deviation from a linear behavior at voltages between 0.65 V and 0.7 V is obtained. Some events, however, are delayed in their conductance change, while others follow the linear increase in conductance. It is interesting to note that the final HCS is independent of the initial state; it is only a matter of the applied voltage amplitude as indicated by the data overlap in Figure 4.10 (a) at voltages greater than 0.7 V. This effect can be related to the voltage divider effect due to the ICL[113]. The state-dependence of the programming behavior shown in Figure 4.10 (a) is correlated to a different current response during the application of the voltage pulse. Figure 4.10 (b) shows a typical SET current transient from a high initial LCS of about 40 µS at a pulse amplitude of 0.72 V. In contrast, Figure 4.10 (c) shows the effect of a low initial LCS value, in detail 5 µS, on the transient current for a SET voltage pulse of 0.85 V and 1 µs. Typically, the current transients starting from a moderately low LCS show a slow current increase in the beginning, followed by fast current increase. This two-step SET process is related to a positive feedback between conductance increase and increasing Joule heating, which finally leads to a thermal runaway[100, 105]. The delay time of the SET process t_{delay} describes the initial low current increase and is often assumed to be equal to the SET time, t_{SET}, because it marks the onset of the fast current increase, consistent with literature[100]. However, a more precise definition of t_{SET} should include the time for the fast, abrupt current increase, which is termed transition time t_{trans} in this work, see zoom-in of Figure 4.10 (c). Therefore, t_{SET} is the sum of t_{delay} and t_{trans}. Data on the relation between t_{trans} and the pulse amplitude starting from low initial LCS is shown in Figure 4.11. Exemplarily, two effects on the SET transition behavior can be determined from Figures 4.10 (b) and (c). An increase in the SET voltage by 0.13 V leads to significant reduction of the SET transition time to the point, where the entire transition event is undergone within the pulse duration. This demonstrates a strong nonlinearity of the transition time from the applied voltage consistent with results on $SrTiO_3$-based devices reported in the literature[100]. In contrast, SET events from high initial LCS (Figure 4.10 (b)) show a different dependency. The high

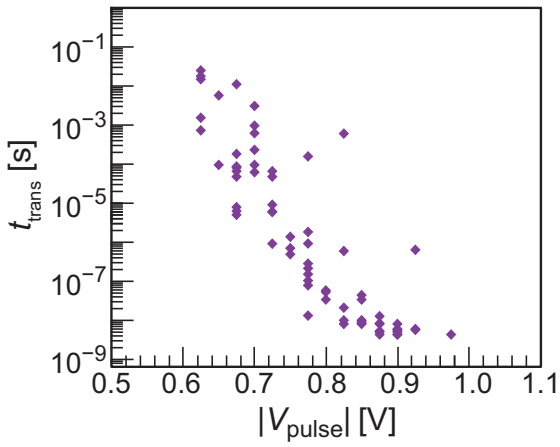


Figure 4.11: Transition time analysis of the SET kinetics measurement starting from the LCS of 2 to 10 µS. Only events where the transition involved at least 12 samples were recorded. Reproduced with permission from [43].

LCS shortens the delay time so much that it is no longer observable. The analysis of the current transients from low LCS during the SET event in 1 µs pulses lead to a better understanding of the SET switching kinetics plot given in Figure 4.10 (d), which was recorded using pulse lengths between 100 ns and 1 s. Here the variability of the SET time, which at a given voltage scatters over about four orders of magnitude, originates from different low initial states. The lower the LCS is, the longer the SET time is, consistent with previous studies [108, 109, 160] and with the simulation results presented in this study as shown below. However, the more careful analysis revealed that this variation in the experimentally derived SET time is, in detail, due to an increase in the delay time leaving the transition time almost unaffected. Further control of the SET behavior of defined VCM-type devices requires the determination of the transition time versus voltage dependence in separation from the effect of the experimental conditions on the delay time.

The study in Figure 4.10 (a) for 1 µs SET pulses was extended to cover pulse lengths from 100 ns up to 1 s, resulting in the graphs given in Figure 4.12 (a), (b) and (c) for high initial LCS of 40 - 50 µS, intermediate initial LCS of around 20 µS, and low initial LCS of around 3 µS, respectively. The median conductance averaged over ten SET pulses is given for the high and intermediate LCS in (a) and (b). Since the median value of conductance would misrepresent the reality of the mix of successful

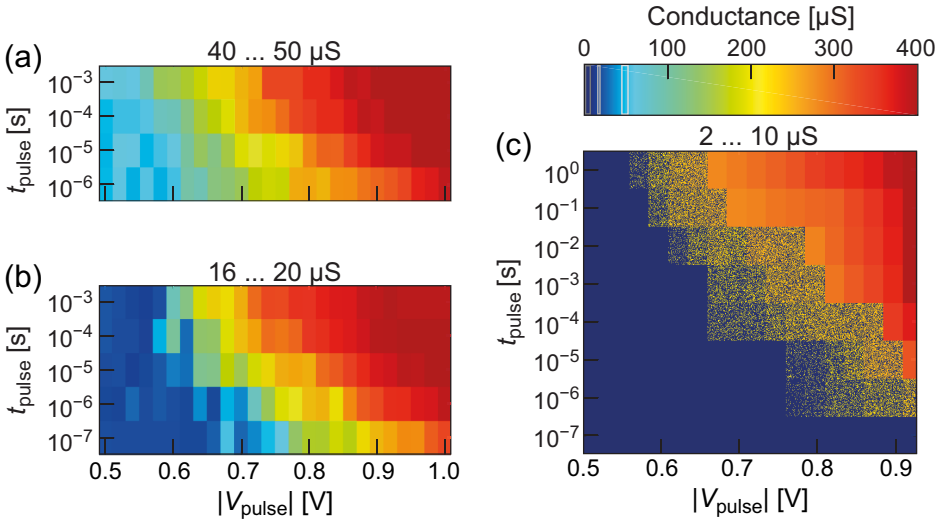


Figure 4.12: Median conductance values after SET pulses (at $V_{read} = 0.2\,V$) with varied amplitude and duration, starting from (a) high LCS and (b) intermediate LCS. (c) Conductance values after SET pulses starting from low initial LCS level. A mixture of successful and unsuccessful SET events is observed. In the HCS, the values coincide independent of previous LCS. Reproduced with permission from [43].

and unsuccessful SET events for the low LCS case, a different depiction method is chosen for Figure 4.12 (c). Combinations of pulse duration and amplitude which leave the device conductance unaffected are drawn as LCS values (dark blue color). In cases where the device undergoes a SET event or remains in the LCS, which is defined as probabilistic switching, a mixture of LCS and HCS values is drawn, with the density of HCS states representing the probability of a SET event. When the pulse voltage-time-combination always leads to a SET event, the respectively reached HCS is drawn (orange to red color). The shown diagrams can be understood in several ways. Firstly, the voltage dependence on the achieved HCS at read voltage of 0.2 V, which was already shown in Figure 4.10 (a), is evident. By application of increased voltage, a higher HCS is achieved. This holds true for all three initial states. This means, that the reached HCS level is independent of the initial LCS state. However, depending on the initial state, the voltage threshold for conductance modulation is influenced. High initial LCS levels require lower voltages for state modulation than low LCS start conditions. Yet, in the experiments presented in this section, all initial conditions lead to almost the same slope in the semi-logarithmic voltage-time-

plot, indicating identical physical processes. This observation supports the thermally activated switching model. Alternatively, the diagrams can be understood from the viewpoint of constant voltage operation. Using extended pulse durations instead of increasing the amplitude also yields a gradual transition between LCS and HCS in the case of high initial LCS conditions. At low initial LCS, longer pulse times additionally make the switching event more likely.

Apparently, the two kinds of switching characteristics observed for high/intermediate LCS and low LCS are related to the different behavior during the current transients. To access intermediate conductance values, the length of the pulse time t_{pulse} must be comparable with the transition time t_{trans}. For very low initial LCS, $t_{\mathrm{pulse}} \gg t_{\mathrm{trans}}$ holds, and thus, the pulse time for switching is a lot longer than the transition time. In consequence, the switching appears binary. Only when the switching transition happens at the end of the applied pulse, intermediate states might be accessible, but these are rare events.

RESET switching kinetics

To study the state- and voltage-dependence of the RESET process, the devices were programmed to different HCS by applying different pulses of -0.9 V up to -1.6 V for a duration of 10 µs in a preceding SET process. The respective HCS, read at 0.3 V, are plotted in Figure 4.13 (a). The obvious nonlinearity of the HCS vs. SET voltage behavior arises from the effect of the ICL element of these devices (see Figure 4.9 (a)). This interplay arises only at SET voltages below -1 V and was therefore not visible in Figure 4.10 (a). Subsequently, RESET experiments at constant voltage are performed for various HCS. Representatively, Figure 4.13 (b) shows the transient currents recorded for a RESET voltage of 1.0 V. The device shows strongly delayed RESET behavior depending on the initial HCS level. Equivalent to the analysis of the SET processs, the RESET process is experimentally defined by the RESET time t_{RESET}, which is determined when the current drops below 300 µA. In the more specific analysis, the transient RESET behavior reveals two regimes. The time addressed to the regime of low current reduction is named the delay time, $t_{\mathrm{delay,\ RESET}}$, the one related to the strong current reduction is identified as the transition time, $t_{\mathrm{trans,\ RESET}}$. The addition of both yields the RESET time t_{RESET}. In the case of low initial HCS, about 500 µS, drawn in dark blue, the RESET occurs within the first few microseconds after pulse application. In the case of high initial HCS, about 700 µS, drawn as orange

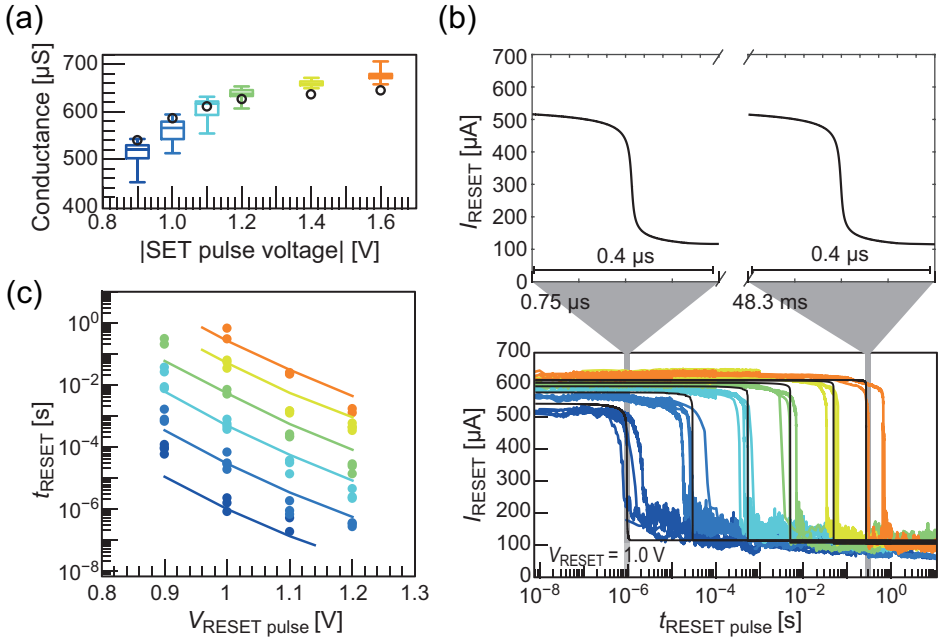


Figure 4.13: (a) HCS of the devices programmed by application of a SET pulse of 10 µs and the given voltage amplitude. The nonlinearity originates from the influence of the ICL. The black circles show the initial states of the corresponding simulations in (b). (b) RESET current transients at a constant RESET voltage of 1.0 V for different initial HCSs. High HCS values lead to pronounced delays during the RESET operation. The simulated transients (black solid lines) are able to fit the variation of the delay by assuming different initial states. The zoom of the simulated transients illustrates that the transition time is state independent. (c) RESET switching kinetics for various HCSs characterized by the SET pulse amplitude that is defined by the color code of (a). A delay of up to six orders of magnitude in switching time is observed. The simulations (solid lines) predict the voltage-time dependence well. Reproduced with permission from [43].

lines, several hundred milliseconds up to 1 second are needed for obtaining the RESET. Here, the RESET time is controlled by the delay time, which turns out highly state-dependent. In contrast, the duration of the sharp transition between HCS and LCS, i.e. the RESET transition time $t_{\mathrm{trans,\ RESET}}$, stays almost constant. Zooming into the transiton regime reveals a transition time of about 65 ns for all initial HCS. From this, the RESET process can be viewed as consisting of three distinct phases. During the first phase, the state stays almost constant and little switching occurs

because during this phase most of the applied voltage drops over the ICL. This delay phase increases significantly for higher HCS and thereby leads to the state dependent delay time. The second phase is the abrupt RESET transition. It has a constant duration at a specific voltage independent on the initial state. The third phase is the slow "phasing out" after the abrupt transition that appears due to the strong temperature decrease with increasing resistance, which slows down the switching into lower LCS states[112]. The strong dependence of the experimentally accessible RESET time was further analyzed at different voltages. A pattern of increasing RESET time with increasing HCS level, highlighted by the colored bars, is visible from Figure 4.13 (c). In the extreme case tested, the RESET is delayed by about six orders of magnitude in time while changing the voltage amplitude of the preceding SET process by 0.7 V. The simulation results, drawn as solid lines, closely match the experimental findings. The strong nonlinearity of the dependence of the switching time on the applied pulse voltage emphasizes the importance of controlling the conductance window, since unfavorable delay times arise when the conductance leaves the moderate regime. Based on the systematic experimental analyses combined with the simulations by means of the fully physical switching model (see below), a detailed description of the SET and RESET dynamics and their state-dependence is developed.

Synapses formed from stochastically and deterministically switching memristive devices

Utilizing this understanding, two alternative types of synapses can be realized with the same device. Bivalent switching between distinctive high HCS and low LCS levels is possible by pulse operation with increased voltage. In this case, the SET/RESET delay time will become significantly longer than the transition time and intermediate states are not accessible. Figure 4.14 (a) depicts a suggested pulse scheme of alternating pulse packages for LTP and LTD operation. Figure 4.14 (b) shows an exemplary extract of three alternating LTP/LTD cycles. In this specific case, the voltages were chosen high enough to obtain reproducible switching with the first two pulses. As shown in Figure 4.14 (c), the change of the normalized conductance over pulse number, which is essentially the update function of the weights, is strongly nonlinear, since the first pulse already traverses the entire dynamic range of the conductance. In the case of linear weight update, this function would follow the straight diagonal line from 0 to 1. A significant step towards a more linear weight update function is taken

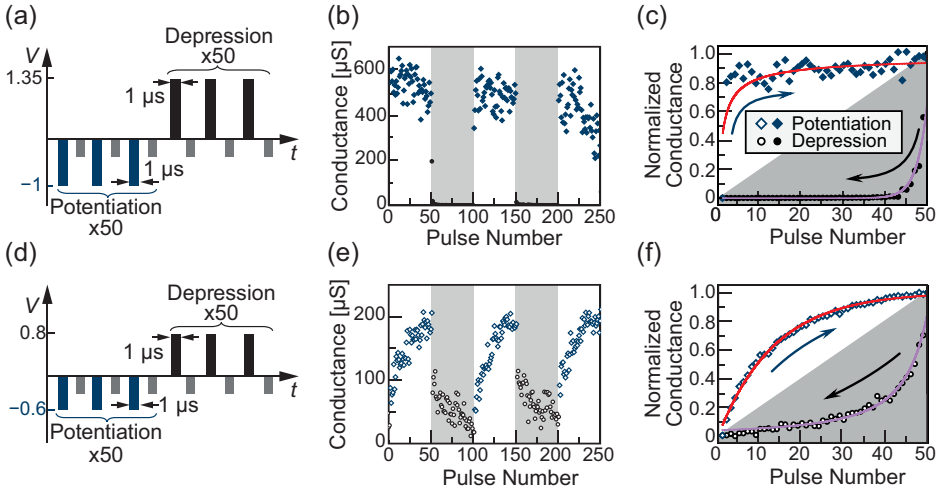


Figure 4.14: LTD and LTP operation with parameters resulting in abrupt and gradual switching behavior: (a) and (d) Operation method and parameters. (b) and (e) Resulting alternating LTP and LTD cycles. (c) and (f) Normalized conductance evolution revealing the high abruptness of the SET/RESET operation parameters in (a) and the more gradual behavior for parameters of (d). The diagonal line marks linear behavior. The fits in (c) and (f) are performed according to the model of Fusi and Abbott[168]. Reproduced with permission from [43].

by (i) choosing a pulse length that is smaller than the transition time at a specific voltage and (ii) reducing the HCS and increasing the LCS level. During operation, this is effectively achieved by reducing voltage amplitudes as given in Figure 4.14 (d). The LTP/LTD cycles depicted in Figure 4.14 (e) show conductance levels inaccessible by the method shown in (a). The normalized conductance change function over pulse repetition is therefore shaped more towards the straight diagonal line, as can be seen in Figure 4.14 (f). The major difference between the two modes seems to be the transition time from LCS to HCS. Low initial LCS levels require a high SET voltage, which, in turn, results in very short transition time (e.g. 22 ns), and is just above the resolution limit of the measurement setup. Hence, intermediate conductance steps are highly infrequent as the transition from LCS to HCS is orders of magnitude shorter than the pulse duration. Once triggered, it is likely to begin and finish within a single pulse. Additionally, a delay time of high variance is inherent. This makes the accessibility of intermediate conductance states very complicated to achieve, even if accordingly short pulse durations were available in the employed measurement setup.

In contrast to this, the transition times corresponding to lower SET/RESET voltages in Figure 4.14 (e) and (f) approach towards the range of the pulse length of 1 µs for this experiment. Furthermore, in these experiments no delay time for the SET transition was observed as the initial LCS is quite high. This makes intermediate conductance states accessible by single pulse application, since the SET process has not been finished by the end of the pulse and is subject to further changes in the subsequent pulse with the same amplitude. Comparable studies on $TiN/HfO_2/Ti/TiN$ devices have been reported by Frascaroli et al.[48]. Interestingly, the voltage regimes defining analog-type behavior in the different HfO_2-based memristive devices are quite comparable. This could be evidence of the universal applicability of the transition-time concept proposed here. From the experimental results shown, it can be concluded that analog switching is possible if the pulse length is of the order of magnitude of the transition time. Furthermore, the state-dependency of the SET/RESET transition should be eliminated. To this end, a suitable conductance window must be selected.

Compact model simulation

To validate and generalize this conclusion, a simulation study was performed using the compact model for filamentary switching based on the valence mechanism called JART VCM v1, which is part of the Juelich-Aachen Resistive Switching Tool Box (JART). The model deviates slightly from the original description[113]. The mechanism of ion conduction, which was previously modeled according to the law of Mott and Gurney[169] is now modeled according to the proposal of Genreith-Schriever[170]. This provides a more accurate description for very high electric fields. To achieve a consistent description of the SET and RESET dynamics, a polarity-dependent effective thermal resistance R_{th} is introduced. The switching parameters have been fitted within physically reasonable limits to match the experimental data. The used parameters are listed in Table 4.1 and the equivalent circuit diagram is shown in Figure 4.9 (a). The model reproduces the state- and voltage-dependence of the experimental data over many orders of magnitude in time. The solid colored lines in Figure 4.10 (e) show the simulated SET delay times for three different initial resistance values. In the simulations, a constant voltage pulse with a rise time of 1 ns is applied. The SET delay is defined at the point in time with the steepest current rise. The spread of the experimental data is well reproduced by assuming different initial resistances (according conductances of 2 µS to 10 µS) in the simulation, which

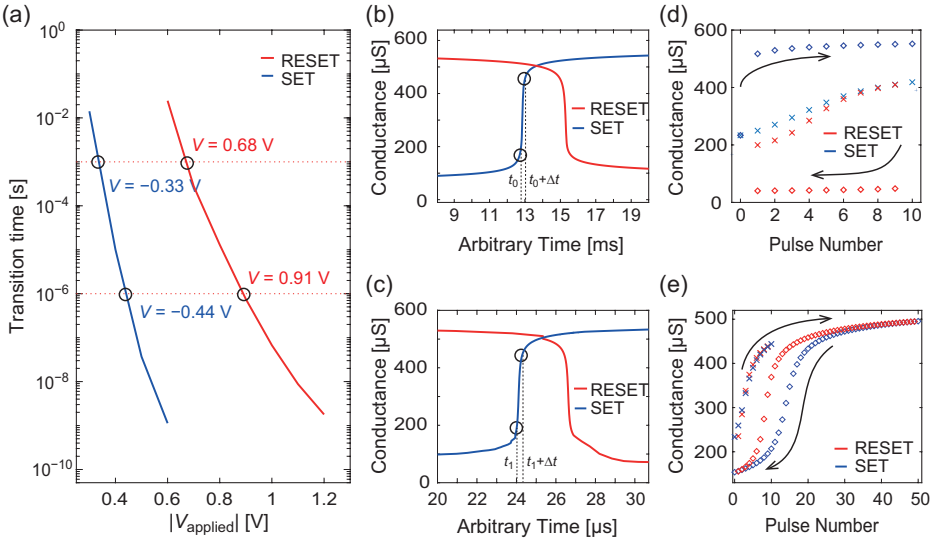


Figure 4.15: (a) Simulated transition times for SET and RESET: the marked circles show the SET and RESET voltages to achieve transition times of 1 ms and 1 µs. The corresponding SET and RESET transients for transitions times of 1 ms and 1 µs are shown in (b) and (c), respectively. The times t_0 and t_1 are arbitrary reference times at the beginning of the transition. Δt represents the time range in which linear conductance modulation is achieved. In (d), the cross symbols show the conductance evolution for consecutive SET pulses (blue) and RESET pulses (red) in the conductance regime defined by the black circles in (b). The diamonds in (d) show the conductance modulation for too high voltage amplitudes (-0.4 V/0.8 V). In (e), the cross symbols show the conductance evolution for consecutive SET and RESET pulses in the conductance regime defined by the black circles in (c). The diamonds in (e) show the simulated conductance modulation using the same pulse length and amplitudes as for the crosses, but starting from a lower initial conductance. Reproduced with permission from [43].

agrees well with the initial resistance states in the experimental data. In the simulations the SET delay time is orders of magnitudes higher than the transition time, consistent with the experimental data shown in Figure 4.10 (c) and previous simulation studies[100]. The simulated RESET current transients are shown in black in Figure 4.13 (b). The initial oxygen vacancy concentrations are chosen to match the initial experimental states at 0.2 V. In the simulations, the RESET pulse of 1.00 V is applied as a constant voltage signal with a rise time of 1.0 ns. As in the experiment, the RESET transition is strongly delayed for higher HCS. This delay results from the

Table 4.1: Simulation parameters (for the explanation of the symbols, see the work of Hardtdegen et al.[113].

Symbol	Value	Symbol	Value
l_{cell}	3 nm	A^*	$6.01 \cdot 10^5$ A/(m^{2}K^2)
l_{disc}	0.4 nm	$e\Phi_{\mathrm{Bn0}}$	0.18 eV
r_{fil}	45 nm	$e\Phi_{\mathrm{n}}$	0.1 eV
z_{VO}	2	μ_{n}	$4 \cdot 10^{-6}$ m^2/Vs
a	0.25 nm	N_{plug}	$20 \cdot 10^{26}$ m^{-3}
ν_0	$2 \cdot 10^{13}$ Hz	$N_{\mathrm{disc,max}}$	$20 \cdot 10^{26}$ m^{-3}
ΔW_{A}	1.35 eV	$N_{\mathrm{disc,min}}$	$0.008 \cdot 10^{26}$ m^{-3}
ε	$17\,\varepsilon_0$	$R_{\mathrm{th,eff,SET}}$	$15.72 \cdot 10^6$ K/W
$\varepsilon_{\Phi\mathrm{B}}$	$5.5\,\varepsilon_0$	$R_{\mathrm{th,eff,RESET}}$	$4.2444 \cdot 10^6$ K/W
T_0	293 K	$G_{\mathrm{TiO_x}}$	1538 µS
$G_{\mathrm{line}}\,(I = 0\,\mu\mathrm{A})$	1391 µS	$G_{\mathrm{line}}\,(I = 700\,\mu\mathrm{A})$	1234 µS

voltage-divider effect of the ICL. At the beginning most of the voltage drops over the ICL rather than the active switching part. As soon as the conductance decreases, the voltage drop over the device increases, and in turn, the switching speed increases. This leads to the fast RESET transition. Due to the reduced power dissipation, the conductance change slows down towards the end. This "phasing out" origins from the decrease in local temperature in combination with ionic drift and diffusion approaching equilibrium defines the behavior in this region during RESET[112]. In contrast, the plateau region at the final stages of the SET transition is due to the current limitation by the ICL. To achieve a stable analog switching, the time frame of the input signals should be in the order of the transition time. Thus, the transition time from SET and RESET pulse simulations were extracted. It turns out that the SET and the RESET transition times are state-independent (cf. zoom in Figure 4.13 (b)), which is consistent with the presented experimental findings. Figure 4.15 (a) shows the simulated SET/RESET transition time as a function of the applied voltage. Consistent with data of Ta_2O_5- and $SrTiO_3$-based filamentary VCM cells, the transition time is a highly nonlinear function of the applied voltage[100], but almost independent of the initial state. The graph also shows that the transition times required for RESET are longer than for SET. Thus, asymmetric voltage amplitudes will be required to achieve SET and RESET with comparable transition times. The desired time sequence depends on the application. For the simulation, transition times of 1 ms and 1 µs were selected. According to Figure 4.15 (a), the corresponding SET/RESET voltage pairs are (-0.33 V/0.68 V) and (-0.44 V/0.91 V)

for 1 ms and 1 µs transition time, respectively. To access intermediate values between HCS and LCS, a pulse width smaller than the transition time is required. In addition, too high HCS and too low LCS should be avoided, as this would lead to long delay times. In order to achieve optimum tunability, the conductance values must be selected within the transition regime as illustrated in Figure 4.15 (b) and (c) for the 1 ms and 1 µs case, respectively. The marked maximum and minimum conduction states allow for an almost linear tuning of the conductance with consecutive pulses of identical voltage amplitudes and duration. The length of the pulses is adjusted to achieve ten different conductance levels. To this end, the elapsed time between the beginning (see t_0 and t_1 in Figure 4.15 (b) and (c)) and the end ((t_0 resp. t_1) + Δt) of the conductance transition is divided by 10. Using the voltage amplitudes and the pulse widths determined as described before, pulse train simulations with 10 consecutive SETs followed by 10 consecutives RESETs are performed. The simulation results are shown as crosses in Figure 4.15 (d) and (e) for the 1 ms and 1 µs case, respectively. In both cases, an almost linear conductance tuning can be achieved. Furthermore, the tuning turns out to be very similar. Following the procedure described above, conductance tuning for analog memristive behavior can be achieved on every time scale. However, any deviation from this procedure unavoidably leads to undesirable results. This is demonstrated in Figure 4.15 (d), where the diamonds are the results of a tuning starting from the same initial conductance and with the same pulse widths as before, but using higher voltage amplitudes (-0.4 V/0.8 V). The transition times for these amplitudes are smaller than the chosen pulse width (see Figure 4.15 (a)). Thus, intermediate states are hardly accessible and the switching becomes binary and achieves high HCS and low LCS. The accompanying increase of stochasticity is due to the state-dependence of the delay times in these conductance regimes. The simulation results correspond well with the experimental data shown in Figure 4.14 (a)-(c). If a lower initial conductance value is chosen while the pulse width and height are the same as in the linear case, the resulting conductance follows an S-shaped modulation (diamonds in Figure 4.15 (e)). In this case, the pulse voltages and the pulse width are chosen according to the procedure described before, but higher HCS and lower LCS values are used. Thus, there is a trade-off between nonlinearity of the conductance tuning and the accessible conductance window.

Conclusion

In conclusion, it was shown that the same $Pt/HfO_2/TiO_x/Ti/Pt$ stack could be exploited for analog or binary, stochastic conductance modulation. The controlling parameters are the voltage amplitude and the pulse length. The most important issue is that a proper conductance window must be chosen. In general, SET and RESET in VCM-type memristive devices are a two-step process: A sharp conductance change within a transition time succeeds a slow conductance change described by a delay time. Both times highly depend on the applied voltage. While the delay time turns out to be highly state-dependent for a specific voltage, the transition time is relatively independent of the state of the device.

Based on these findings, two conditions for achieving analog conductance behavior in filamentary VCM cells are deduced. First, the applied pulse length must be shorter than the transition time. Otherwise, the switching will become binary. Second, the conductance window must be chosen in a way that delay times are significantly reduced below several percent of the transition time. This can be effectively done by increasing the initial LCS and by decreasing the HCS, prior to SET operation (potentiation) and to RESET operation (depression), respectively. Using the JART VCM v1 model, a generalization of this result could be made and an experimental procedure to find the optimum working condition was formulated. First, the transition time needs to be determined as a function of the applied voltages. Based on this result, the SET and RESET voltage amplitudes can be chosen according to the operation time of the targeted application. From the recorded SET and RESET transients, a proper conductance window is then defined. It is important to note that, in principle, every timescale allows for proper operation conditions. The voltages, however, may not be compatible to the application. The present study points a new direction for further research. As the transition time is identified as the most important parameter, future research should strive for the elucidation of the physical parameters influencing the transition time and the size of the addressable conductance window.

5 Analog function of VCM devices

The previous chapter elucidated on the coexistence of analog and binary switching in single memristive devices. This chapter focuses on the operation in analog mode. The possibility to program and store analog conductance in memristive devices makes them a possible candidate for different applications. Deep Neural Networks (DNNs) can be accelerated by several orders of magnitude compared to computations in the CPU at the same performance levels [68, 171]. Other applications are Computation in Memory (CIM) concepts [172–174] and as analog content addressable memory (CAM) [56, 175] with ReRAMs which can decrease the overall power consumption by an order of magnitude or more compared to SRAM architectures with the same function [56]. Both applications require initial programming and storage of analog conductance values. However, once the desired weights are programmed, the device operation is limited to reading the conductance. For networks that can learn online, the conductance modulation is of additional interest. In many cases, a rather simple but effective way of analog conductance programming, namely constant repeated voltage signals, is preferred. Depending on the specific application, different constraints such as the minimum number of programmable states, linearity of the update and symmetry between the update directions are important [67, 68]. In the following section, these parameters will be quantified for the studied device. During the investigation, a significant influence of noise is found which severely limits the number of distinguishable states. The identified noise characteristics of the programmed states, which have been reported in a similar form in other literature sources [176–179], are examined and interpreted in terms of the physical model of filamentary VCM devices.

5.1 Analog switching by constant voltage signals

As discussed for Figure 4.14, single VCM cells can be tuned in analog conductance states using a sequence of voltage pulses with constant amplitude and duration. This

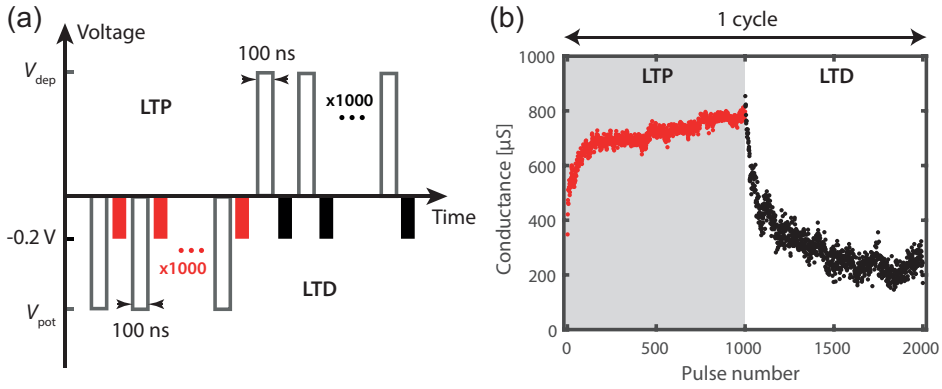


Figure 5.1: Definitions for the Long Term Potentiation (LTP) and Depression (LTD). (a) Voltage pulse sequence for LTP and LTD. (b) Increase and decrease of the conductance value during a typical LTP and LTD cycle.

feature is considered very advantageous for applications requiring frequent conductance changes in an analog manner. In this section, the analog tuning capability through constant pulses is examined in more detail. In literature, this behavior is commonly termed Long Term Potentiation (LTP) and Long Term Depression (LTD). Figure 5.1 shows the used definitions in this work. The pulsing scheme is shown in (a) and a typical LTP/LTD cycle in (b). A cycle is therefore defined as the combination of one LTP and one LTD half cycle. Each half cycle consists of 1000 programming pulses and conductance readings. The voltage for performing LTP is negative according to the bipolar SET process. Conversely, the LTD voltage is positive. LTP and LTD pulse length is always 100 ns, while the read pulses are 10 ms long. The interpulse duration is limited by the measurement setup and is on the order of milliseconds. The read amplitude is kept constant at -0.2 V.

Figure 5.2 illustrates three examples how the conductance of a single cell can be programmed with constant pulses. The bottom panel shows the response to -0.5 V and +0.5 V amplitude pulses for LTP and LTD, respectively. The response of the cell to these voltage pulses is small, and noise largely masks the conductance changes. A different response is observed for -0.6 V and +0.7 V, see middle panel. The conductance is noticeably increased by LTP pulses and decreased by LTD pulses. Small conductance increments throughout the continued pulses are observed. Intermediate states are repeatedly approached. In contrast, the upper panel shows the cell's response to higher voltages (-0.7 V and +0.9 V). Again, the conductance can be pro-

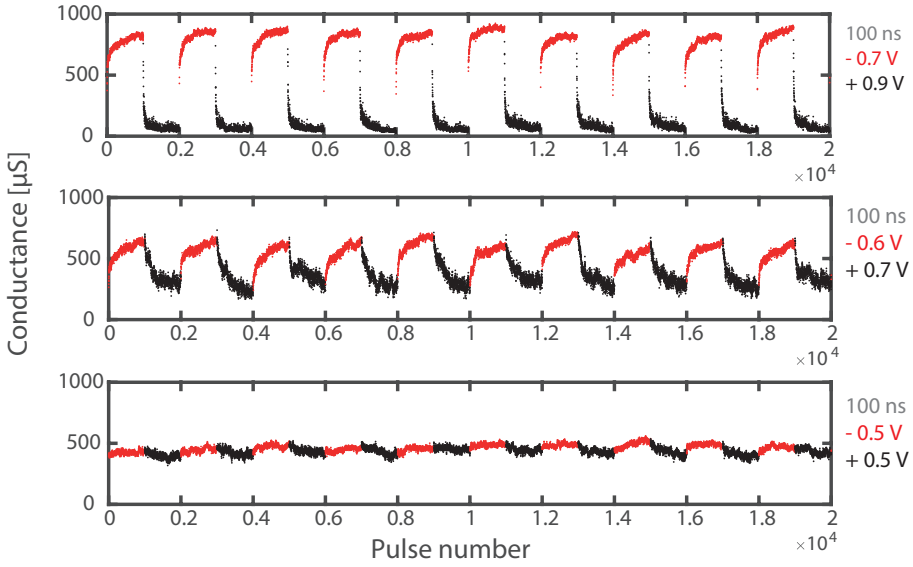


Figure 5.2: Long Term Potentiation and Depression overview. From lower to upper panel: Low amplitudes do not trigger significant conductance change while intermediate amplitudes lead to analog-type programmable conductance. Higher amplitudes cause relatively abrupt changes within the first few pulses.

grammed reproducably, but significant conductance gaps appear in the intermediate range, indicating a nonlinearity in the programming. In particular for the LTP half cycle, the first applied pulse frequently induces a change of about 300 µS or more. Considering the total conductance difference between the end of an LTP and an LTD cycle, the influence of higher absolute voltages is observable. From the lower to the upper panel, the total conductance difference ΔG increases from values around 10 µS to around 350 µS to around 700 µS. An extension of the presented experiment is conducted by varying the LTP voltage and the LTD voltage in a matrix-like fashion. The parameters (V_{pot} | V_{dep}) are varied between (-0.3 V | +0.1 V) and (-0.8 V | +1.0 V) in 50 mV steps. The relationship between the LTP and LTD amplitude is denoted as $V_{\mathrm{dep}} = |V_{\mathrm{pot}}| + \Delta V$. However, the unbalanced corners of the matrix are not tested because the cycling for these parameters is unstable, i.e. the device typically gets stuck in the corresponding state of the prevailing voltage polarity. Each combination is repeated 100 times, i.e., 100 alternating LTP and LTD half cycles are applied without intermediate write-verify.

First, the observations made in Figure 5.2 are investigated in more detail. Figure 5.3 (a) shows the mean conductance response to the first applied LTP pulse ($\Delta G_{\text{pot, first}}$) as well as the mean maximum conductance change ($\Delta G_{\text{pot, max}}$) within one pulse as function of the mean total LTP conductance window $\Delta G_{\text{pot, total}}$. The error bars are given by the standard deviation of the 100 cycles. The trend is consistent with the observation in Figure 5.2: above about 600 µS of the total conductance range, the first LTP conductance change coincides with the maximum conductance change at about 200 µS, suggesting that the switching is strongly nonlinear and that a third of the dynamic range is inaccessible. The corresponding LTP voltage is -0.7 V. At even higher voltages, the first LTP pulse induces a conductance change which corresponds to half of the total adressable range. At lower voltages, the first and the maximum conductance changes seperate and decrease, while the absolute conductance range also decreases. However, this regime allows for smaller conductance increments in the LTP sequence. Importantly, this means that near linear conductance change is possible. For LTD, which can be seen in Figure 5.3 (b), a different trend can be observed. Even for the highest total conductance ranges, which are reached by employing accordingly high voltages, the maximum conductance step is limited to around 200 µS. Across all voltages, the first conductance step is found below the maximum value. This means that a more gradual conductance transition is found for the first LTD pulses. However, the black open circles indicate that the maximum conductance step is around 200 µS for the majority of voltages. Only for small voltages does the maximum conductance step fall below 200 µS. However, the maximum conductance step is sometimes higher than the total achieved conductance range $\Delta G_{\text{dep, total}}$. This is also the case for the LTP. It is indicated by the red dashed line, which marks $\Delta G_{\text{total}} = \Delta G_{\text{max}}$. It is clear that a conductance step larger than the total bridged conductance range is unrealistic. The reason for these points can be found in the lower panel of Figure 5.2. For low applied voltages, the conductance noise from one pulse to the next can be larger than the actual spanned conductance range from the beginning of an LTP and LTD cycle. This means, that the value of the maximum conductance step ΔG_{max} at ΔG_{total} around 0 µS in Figure 5.3 is actually caused by the noise present in the measurement. Hence, the data points that are displayed below this threshold represent a severely degraded signal to noise ratio, i.e. individual noise spikes are stronger than the signal. The presence of noise in the analog switching regime significantly complicates the analysis of the underlying conductance change. For this reason, the approach of Gong et al.[51] is adopted. Here, a

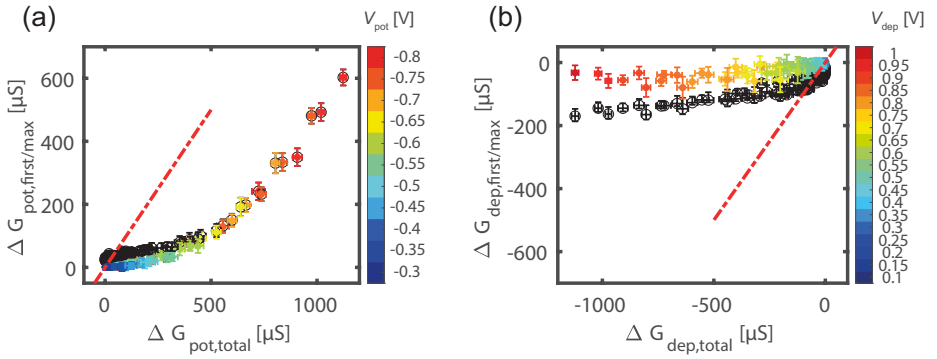


Figure 5.3: Correlation between the conductance change of the first pulse (colored solid symbols) and the maximum conductance change (black open symbols) with the total tunable conductance at the specific voltage for (a) LTP and (b) LTD. Red dashed lines represent $\Delta G_{\text{total}} = \Delta G_{\text{max}}$.

Gaussian Process Regression (GPR) algorithm is used to seperate the noise from the underlying signal. It is assumed that the measured conductance values are normally distributed around a hidden arbitrary function. Importantly, GPR is independent of the physical switching mechanism and does not require any user-made assumptions, which makes it an ideal tool to study the noisy data of filamentary resistive devices. For a detailed description and validation of the technique to study analog resistive devices, the reader is referred to the original paper [51]. Figure 5.4 shows example cycles from the data in Figure 5.2. The blue lines in (a) through (c) represent the optimized noise-free GPR fits to each half cycle. (d) to (f) show the histograms of the residual conductances of the LTP, while (g) to (i) show the LTD residuals. The coarsely sampled Gaussian distribution is evident, as shown by the blue lines.

The exemplary fits in Figures 5.4 (a) to (c) illustrate that each LTP and LTD cycle can be divided into two parts as indicated by the dashed vertical lines. The first part, labeled "SW" for switching part, is characterized by a monotonic conductance drift that depends on the polarity of the signal. For LTP, the slope is positive, while LTD has a negative slope. In all measured half cycles, the monotonic fit section ended before reaching the end of the pulse train, i.e. the slope of the GPR fit changed sign. This second region, labeled "NSW" for non-switching part, is characterized by noise around a mean conductance value, which does not change in response to the applied pulses. A similar observation was made by Brivio et al. [176]. In the following, both parts will be treated separately, starting with the switching part.

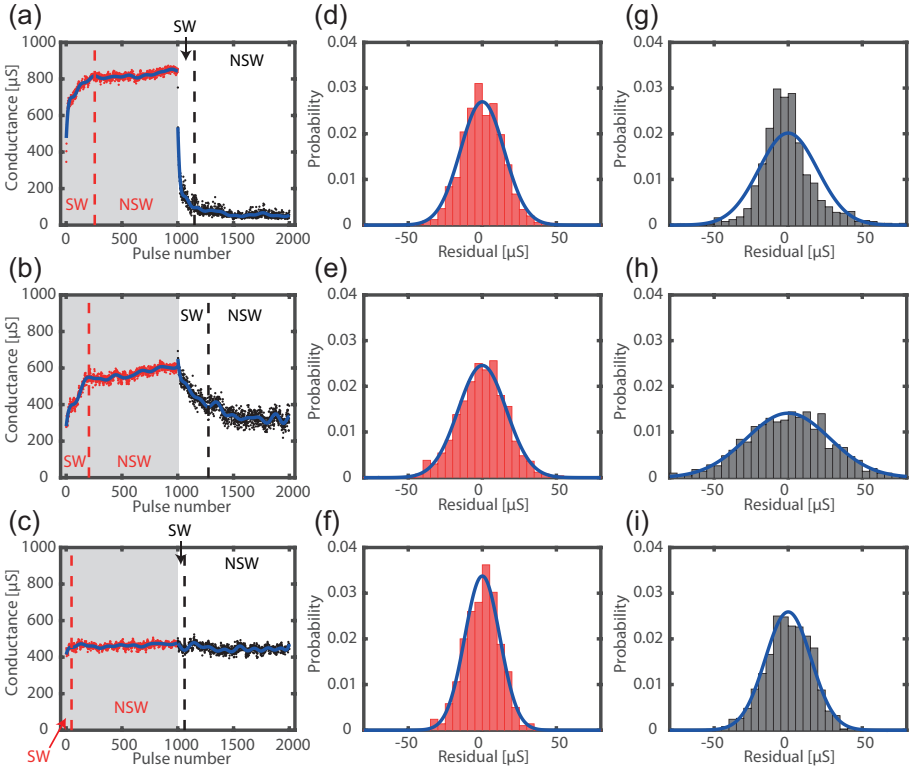


Figure 5.4: Gaussian Process Regression analysis of the data shown in Figure 5.2. (a) to (c) Example fits to LTP and LTD. Voltages are (a): (-0.7 V | +0.9 V), (b): (-0.6 V | +0.7 V) and (c): (-0.5 V | +0.5 V). Times are all 100 ns. (d) to (f) According residuals histograms of the LTP fits. (g) to (i) According residuals histograms of the LTD fits. Blue lines show the normal probability distributions as verification of the assumption of normally distributed noise around the GPR fit.

Switching part

First, the LTP/LTD half cycles are examined in terms of the number of pulses until the fit changes sign for the first time, i.e. the conductance saturates and the mean value remains constant. A high number of pulses is generally desired in analog-like devices because it means that the number of programmable conductance steps is higher. The respective term to describe this number is resolution in this work and can be seen as a theoretically ideal, completely noise-free signal. Figure 5.5 shows the LTP voltage dependence of the number of pulses until saturation is reached over the absolute conductance range traversed. The points are mean values, while the error bars represent the standard deviation of the 100 half cycles per voltage combination, i.e. the variation from . For all measured half cycles, saturation was reached well before the end of the half cycle pulse train at 1000 pulses. The influence of the voltage on the number of pulses to saturation is not clear when considering only the data points. The grey to black lines indicate the described voltage offset between LTP and LTD, which is given by $\Delta V = |V_{\mathrm{dep}}|\text{-}|V_{\mathrm{pot}}|$. Figure 5.6 shows the according graph for LTD. It is evident from the grey lines in Figures 5.5 and 5.6 that voltage combinations with negative ΔV, i.e. where $|V_{\mathrm{dep}}| < |V_{\mathrm{pot}}|$, do not produce significant conductance windows for programming. Exemplarily, this can also be seen in the lower panel of Figure 5.2. The relevant voltage combinations are therefore limited to positive ΔV, i.e. $|V_{\mathrm{dep}}| > |V_{\mathrm{pot}}|$. Although the error bars in Figures 5.5 and 5.6 make it clear that the observation of resolution is statistically sound, they make it considerably more difficult to make sense of the graphs. Therefore, in Figure 5.7 (a) and (b), the mean values of Figures 5.5 and 5.6 are presented without error bars. Only combinations where ΔV is equal to or greater than 50 mV are shown. Additionally, colored lines are added to connect the points of equal voltage and highlight the following observation. Two different trends are evident for both polarities. For the LTP data, the nearly horizontal colored lines indicate that the resolution is purely determined by the amplitude V_{pot}. For a given LTP voltage, the conductance window $\Delta G_{\mathrm{pot,\,total}}$ is determined by the voltage offset ΔV as indicated by the grey lines. The opposite is the case for the LTD data. Here, the resolution varies significantly for a given amplitude V_{dep}. Instead of resolution, the LTD voltage appears to determine the conductance window $\Delta G_{\mathrm{dep,\,total}}$. The resolution depends strongly on the offset voltage. The highest resolutions are achieved with moderate offset voltage, while strong ΔV lower the number even with the same V_{dep}.

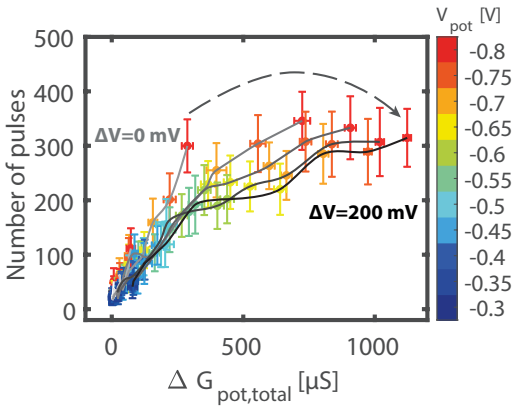


Figure 5.5: Total conductance range versus number of pulses until saturation is reached for LTP half cycles. Solid lines show the voltage offset between LTP and LTD, where $\Delta V = |V_{\mathrm{dep}}|\text{-}|V_{\mathrm{pot}}|$.

To illustrate the significance of resolution in the actual measured data and the widely varying behavior, two sets of example LTP/LTD cycles are shown in Figure 5.8. The selected set combinations are highlighted in Figure 5.7 by square symbols and star symbols. Figure 5.8 (a) shows examples for $\Delta V = 200\,\mathrm{mV}$, while (b) shows example cycles for $\Delta V = 50\,\mathrm{mV}$ combinations. The vertical lines for LTP and LTD mark the first negative gradient in the data, i.e. resolution. LTP and LTD voltages are shown in the Figure. The three LTP voltages between (a) and (b) are identical, but the LTD amplitude differs. Comparison between the LTP curves in (a) and (b) illustrates that both the resolution and the absolute final conductance are determined by the LTP amplitude and are largely independent of the offset voltage and the previous conductance state. In contrast, the LTD half cycles are significantly affected by the offset voltage. In (a), the 200 mV difference leads to low resolutions for all three example curves. The resolutions are comparatively high for the 50 mV difference in (b) and follow an increasing trend with LTD voltage. The higher resolution comes at the cost of a reduced conductance window.

This significantly different behavior for LTP and LTD raises the question which offset voltage and which combination thereof actually provides symmetrical resolutions. Since it is difficult to compare the resolution values for the data points in Figure 5.7, the numbers for LTP are plotted against the LTD numbers in Figure 5.9. The diagonal line marks the identity. The grey lines are labeled with the according ΔV, while

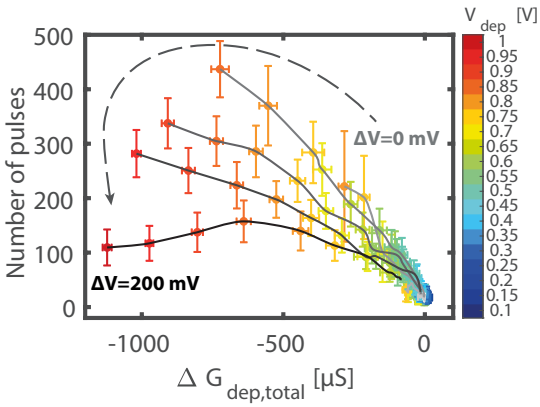


Figure 5.6: Total conductance range versus number of pulses until saturation is reached for LTD half cycles. Solid lines are voltage offsets.

the color of the circles represent the LTP voltages. Amplitudes $|V_{\mathrm{pot}}| < 0.5\,\mathrm{V}$ are not shown due to the low resolution number, the small observed conductance range, and for clarity of the plot. The lines for the offset voltages show a consistent trend. 50 mV difference between LTP and LTD amplitude favors the resolution of the LTD half cycle. However, 150 mV and 200 mV lead to increasingly lower LTD resolution. The most symmetrical combinations are found for 100 mV. The LTP resolution data points for each LTP amplitude are almost vertically aligned, while the points for LTD are spread further apart. This underlines that the LTP is largely independent of the opposite programming half cycle, while the LTD is strongly dependent on it.

Regardless of the resolution numbers, a considerable amount of nonlinearity can be seen in the switching part of the LTP and LTD half cycles, for example in Figure 5.8. To systematically study nonlinearity in the device of this work, the data are first filtered for sufficient resolution above 100. In addition, a requirement for the total conductance modulation of at least 200 µS is imposed. For both criteria pulse combinations that do not show significant change are removed. Subsequently, both the conductance range and the pulse count are normalized to ensure comparability. The median lines of the described analysis are shown in Figure 5.10. The LTD line is inverted to show the difference to the LTP. The colors are coded for the LTP voltage as indicated in the color scale on the left. LTD voltages can be calculated from the given offset. The imposed resolution and conductance criteria only result in postive offset voltages from 50 mV to 200 mV, see Figures 5.10 (a) to (d). Additionally, the

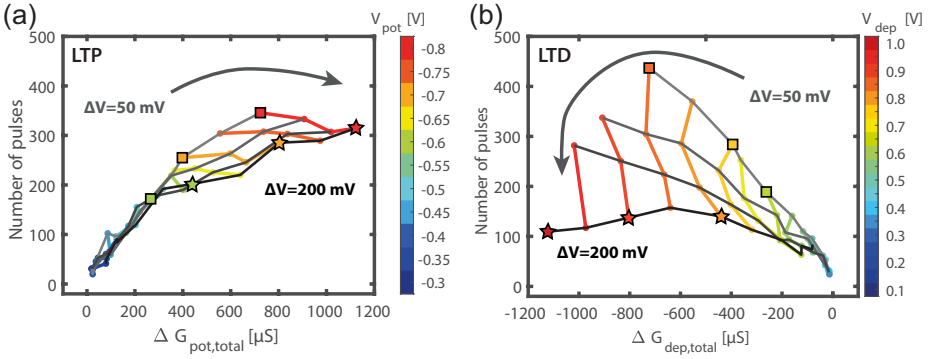


Figure 5.7: Resolution versus conductance range plot for (a) LTP and (b) LTD half cycles. Solid grey lines are voltage offsets and colored lines are same voltages.

criteria are met only for absolute LTP voltages above 0.55 V. It is obvious that all tested pulse combinations exhibit significant nonlinearity. The theoretical linear conductance response is illustrated by the gray diagonal line in the diagrams. While the LTP seems to be mainly dependent on the respective LTP voltage, the LTD curves are both dependent on the voltage, see for example Figure 5.10 (c), but also on the offset voltage. At low offset voltages, the response is close to the linear line, while higher offset voltages lead to more nonlinearity. The reasons for this behavior are very similar to the discussion for resolution. Importantly, LTD exhibits a broader range of nonlinearities that can be controlled by external parameters, whereas the LTP appears to be more deterministically nonlinear.

Another important metric for analog tuning with memristive devices is the symmetry between LTP and LTD. Therefore, the normalized data are again represented as the difference between LTP and LTD, i.e., the absolute LTD curve is substracted from the LTP curve: ΔG_{pot}-$|\Delta G_{\mathrm{dep}}|$. The results corresponding to the data in Figure 5.10 are shown in the asymmetry plots in Figures 5.11 (a) through (d). Positive values indicate that the LTP has higher nonlinearity, while the opposite is true when the values are negative. Good symmetry is indicated by values close to zero. The colors are chosen with respect to the LTP voltage as indicated by the colorbar on the left. The written offset voltages enable assignment of the used LTD amplitude. The graphs allow for easy visual differentiation of the asymmetry of the nonlinear curves. A clear trend is observed. For low offset voltages, the LTP is dominating for all voltages. Increasing the offset voltage strengthens the LTD, until the LTD is dominating

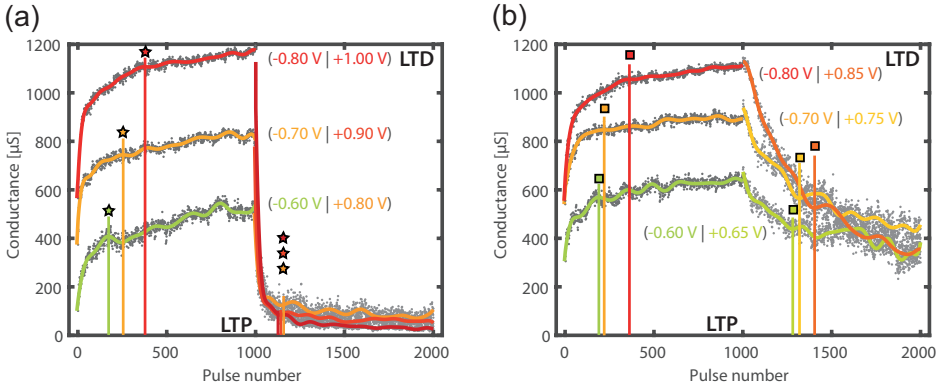


Figure 5.8: Two sets of exemplary LTP/LTD cycles for important voltage combinations. Vertical lines indicate the resolution. The offset voltage $\Delta V = |V_{\mathrm{dep}}| - |V_{\mathrm{pot}}|$ is always 200 mV in (a) and 50 mV in (b).

the asymmetry. At 200 mV offset and for high absolute LTP voltage, the LTP is dominant again because the switching is abrupt. In all cases, higher absolute voltages lead to more asymmetry. The least asymmetric combination is found for moderate LTP amplitude and moderate offset voltage. To illustrate the tradeoff, the best combination of every offset voltage is taken and analyzed further when the switching and noise parts are recombined. In particular the combinations (-0.60 V | +0.65 V), (-0.55 V | +0.65 V), (-0.55 V | +0.70 V) and (-0.55 V | +0.75 V) show values closest to zero and therefore represent the least asymmetric curves. The observations so far are summarized as follows:

- Voltage offsets below 0 mV, i.e. $|V_{\mathrm{pot}}| > |V_{\mathrm{dep}}|$, result in insignificant switching operations. LTD voltage should always be at least equal or higher than the LTP voltage in its absolute amplitude. Switching is significant and stable from an offset of 50 mV.

- The resolution of the LTP half cycle is determined only by the LTP amplitude, but not by the offset voltage or the start conductance. In contrast, the LTD half cycle resolution is significantly affected by the previous programming and the offset voltage for a given LTD amplitude changes the resolution.

- The resolution symmetry between LTP and LTD is therefore determined by matching the controllable LTD half cycle to the LTP half cycle, which is fixed

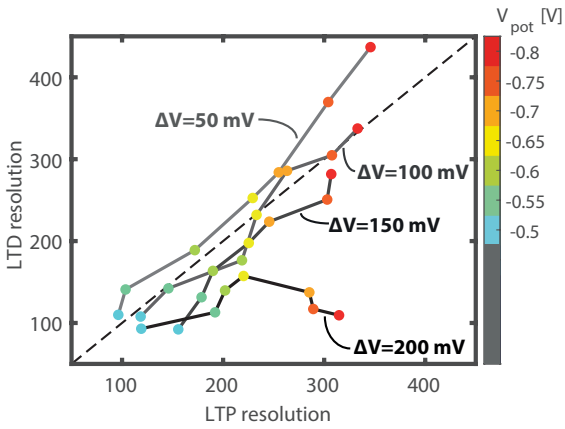


Figure 5.9: LTP versus LTD resolution for relevant voltage combinations. The dashed lines indicates identical values for LTP and LTD.

for a given LTP amplitude. A ΔV of 100 mV proved to be ideal for obtaining symmetrical resolutions for almost all tested combinations.

- Nonlinearity is present in all combinations that have stable switching operations and cannot be avoided. However, the symmetry of the normalized curves can be achieved by choosing an LTD curve that matches the opposite LTD curve as close as possible in its asymmetry. The lower nonlinearity of the lower LTP amplitudes allows a better match with an LTD half cycle.

The ideal combination of LTP and LTD for symmetry in the resolution number and symmetry in the normalized curve is therefore a moderate LTP amplitude of -0.55 V with an LTD amplitude increased by 100 mV, i.e. 0.65 V. Other combinations may have a higher conductance range, but suffer from the asymmetries described earlier.

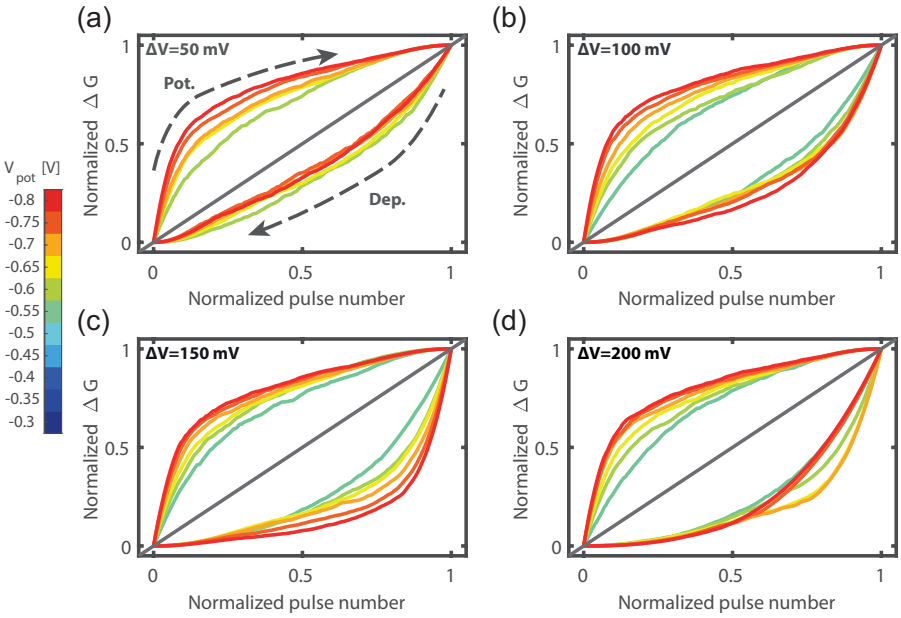


Figure 5.10: Nonlinearity plots for pulse combinations where the resolution and the conductance range are considered sufficient. The LTP curves go from 0 to 1, while the LTD lines are inverted to go from 1 to 0. Colors are according to LTP voltage, and offset voltages $\Delta V = |V_{\mathrm{dep}}|\text{-}|V_{\mathrm{pot}}|$ are (a) 50 mV,(b) 100 mV, (c) 150 mV and (d) 200 mV.

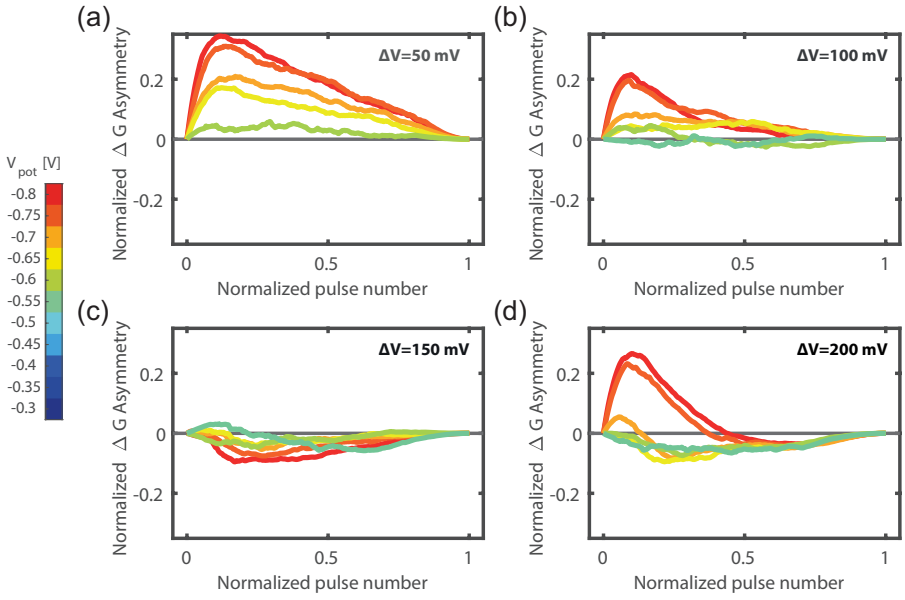


Figure 5.11: Asymmetry plots for the relevant subset of pulse combinations. Asymmetry is defined as ΔG_{pot}-$|\Delta G_{\mathrm{dep}}|$. (a) to (d) represent offset voltages.

Noise part

To further explore the occurrence of noise in this experiment, a procedure similar to that described in [176] is applied. As discussed above, the analog transition is completed after the first derivative of the fitted GPR signal changes sign for the first time. The subsequent pulses do not contribute to a change in the mean conductance, but rather cause noise around the mean value. Therefore, the mean conductance of the last pulses $\mu(G_{\text{end}})$ and the according standard deviation $\sigma(G_{\text{end}})$ will be calculated in this part of the cycle. Figure 5.12 (a) and (b) show the results for LTP and LTD, respectively. Here, the colors show the amplitude of the applied voltage pulse as before. The diagonal grey lines represent different values of signal-to-noise ratio (SNR), as defined by

$$\text{SNR}_{\text{avg}} = \frac{\mu(G_{\text{end}})}{\sigma(G_{\text{end}})}. \tag{5.1}$$

Since the mean and standard deviation are considered, the fraction describes the average case of the SNR. The diagrams shown include both the influence of the noise as a function of the average conductance (x-axis) as well as the effect of the voltage amplitude (color scale). The LTP analysis reveals that, in general, relatively low noise levels are present at the end of each cycle. The SNR_{avg} is typically above 10. Considering the results from Figure 5.3, however, voltages above -0.5 V show higher noise than the actual conductance change. Considering only voltages below -0.5 V, the value of SNR_{avg} is typically above 50 for LTP. At lower LTP voltages, the value of SNR_{avg} seems to increase further. This effect is due to the nearly constant standard deviation, while the mean conductance decreases due to the decreasing voltage. This observation is in excellent agreement with the results from Section 4.5.

A different effect is observed in the LTD case. Similar to the LTP, the different voltage levels lead to different mean final conductances, but because of the more gradual transition of the RESET process, a wider range of conductance values is covered. In particular, this means that different voltage amplitudes at the same terminal conductance can be compared and their influence on noise excitation observed. For a mean conductance of around 200 to 500 µS, voltages from 0.1 V to 0.7 V can be compared. There is a clear trend of increase in noise amplitude with increase in pulse amplitude. Another effect is the decrease in noise with decreasing mean conductance. Between 30 µS and 200 µS, the mean conductance and noise level appear to be almost proportional at a constant SNR_{avg} of around 5. As such conductances can only be achieved with voltages above 0.9 V, the influence of voltage on the noise can not be

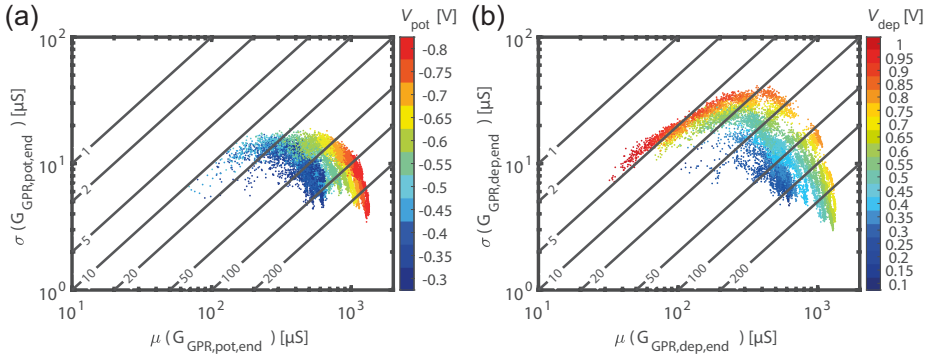


Figure 5.12: Noise analysis at the end of the LTP (a) and LTD (b) half cycles. The mean and the standard deviation for each measured cycle is shown with respect to the applied voltage amplitude. Diagonal lines mark distinct signal-to-noise ratio levels.

extracted from this experiment. Looking at the results from Figure 5.3, the noise at the end of the LTD half cycles leads to SNR_{avg} values around 5 to 20 for voltages that induce conductance switching. Compared to the observation of the LTP, the noise levels at the end of LTD half cycles are the limiting factor in terms of SNR_{avg}. In general, the noise levels are higher when LTD pulses were applied, as can also be seen in the histograms in Figure 5.4. Brivio et al.[176] came to similar results in their analysis to a certain extent. In the supplementary section of their paper, they compare the SET and RESET process in terms of the noise amplitudes. Although they do not give a quantitative value, it is clear that the dominant noise is observed during RESET cycles, i.e. LTD. However, in their analysis they conclude that the main source of noise stems from the mean resistance range of the device, while the effect of applied voltage is not considered in detail. In contrast, the results obtained in this chapter show that a higher voltage also increases the observed noise, especially in the center of the easily programmable conductance range. Since the noise-versus-conductance curve is bell-shaped (see Figures 5.12), it seems logical to prefer programming in the high conductance range, since it provides a good signal-to-noise ratio. However, this decision implies the use of high voltage amplitudes, which has been shown to have a negative impact on analog programming, see the previous section. Interestingly, the conductance range from about 50 µS to about 600 µS, which allows analog tuning, is also the range with the highest observed conductance noise. The reason for this observation is not entirely clear, but could be related to the complex nonlinear conductance change mechanism in HfO_2 based devices. This will be investigated in more

detail in Section 5.2.

Reduction of programmable states by noise

In the two preceding sections, the switching and noise component of the analog conductance modulation were analyzed individually. In this section, an attempt is made to recombine the two for the most promising parameter combinations, which were found in the switching section and determined to be (-0.60 V | +0.65 V), (-0.55 V | +0.65 V), (-0.55 V | +0.70 V) and (-0.55 V | +0.75 V). The approach chosen is a worst-case approximation for the number of programmable states in a 1 σ range. The used noise characteristic is taken from the LTD curve for two reasons: First, it is fully available for the relevant conductance range. Second, LTD pulses in the experiment caused slightly more noise than LTP pulses, indicating that the LTD curve represents the higher noise levels, which is consistent with the worst-case approximation. For the same reason, the upper envelope of the noise spectrum is chosen. Figures 5.13 (a) to (d) show the results. Here, the data points show the noise spectrum for LTD pulses with amplitude of +0.65 V, +0.65 V, +0.70 V and +0.75 V, respectively. The dashed line marks the upper limit spline used for the following analysis. To determine the number of levels that can be assumed for a given pulse combination, the window between minimum conductance and maximum conductance for each combination is filled with as many distributions as possible, keeping a distance of 1 σbetween the levels. The resulting distribution averages and sigma values are shown as colored dots in Figure 5.13 (a) through (d). The number of levels resulting from this approximation are 7, 5, 6 and 7, respectively. Figures 5.14 (a) to (d) show the corresponding normal distributions. Since a spacing of 1 $\sigma$is required, they overlap considerably. Increasing the requirement to a 2 σ spacing allowed the testing of four distributions in all cases, but better separation. Finally, the parameter combinations are compared with respect to their respective LTP and LTD resolution, taken from Figure 5.6 and 5.5, and the number of levels. Figure 5.15 shows that almost identical resolutions are found only for the combination (-0.55 V | +0.65 V). For the other parameters asymmetric resolution values occur. Note that the asymmetry of the resolution is not equal to the normalized asymmetry, which was acceptable for all four parameters. At the same time, the lowest number of separable levels is found for the combination (-0.55 V | +0.65 V), which is due to the strong presence of noise in the programmable conductance range. The most important findings of this section are summarized as follows:

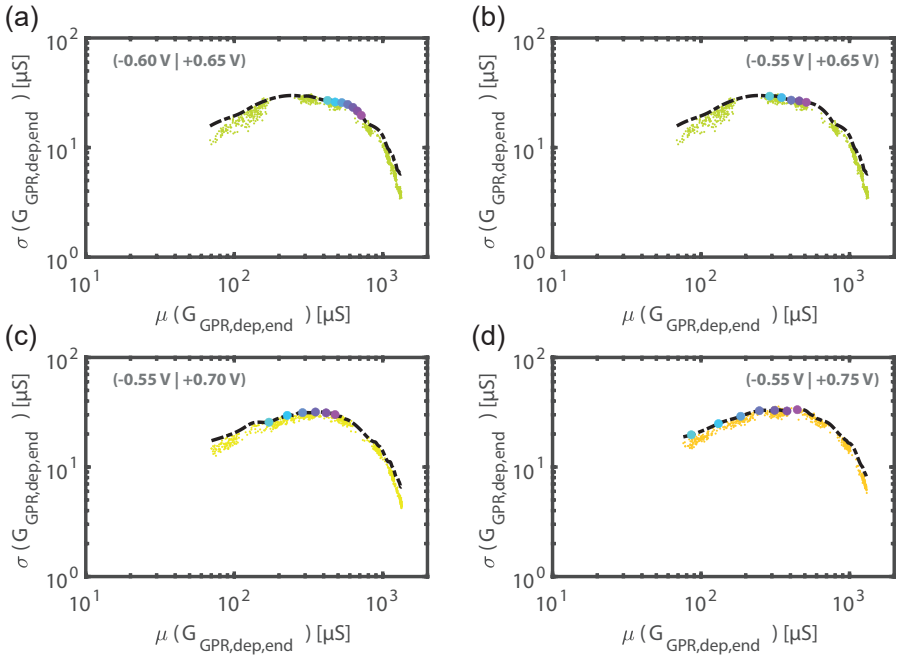


Figure 5.13: Recombining switching and noise for analog conductance modulation for the best combinations regarding conductance range and asymmetry. (a) and (b) show noise data for +0.65 V LTD pulses, (c) for +0.7 V and (d) for +0.75 V. Dashed lines indicate the upper noise boundary. Points mark seperable levels within the programmable conductance range.

- The operation parameters chosen strongly affect the response characteristic in the analog-like operation in various ways.

- In conductance modulation, switching and noise occur simultaneously. In order to understand the underlying properties, it is helpful to separate the two processes. In the context of this work, this was achieved by applying a Gaussian process regression, which assumes an undistorted form, and subsequent separation of signal and noise.

- It was found that the noise-free switching part is best understood with three quantities: resolution, nonlinearity, and asymmetry. All three quantities are influenced by the absolute amplitudes of LTP and LTD, but also their relative amplitudes. In general, LTD amplitudes should be chosen moderately (100 mV)

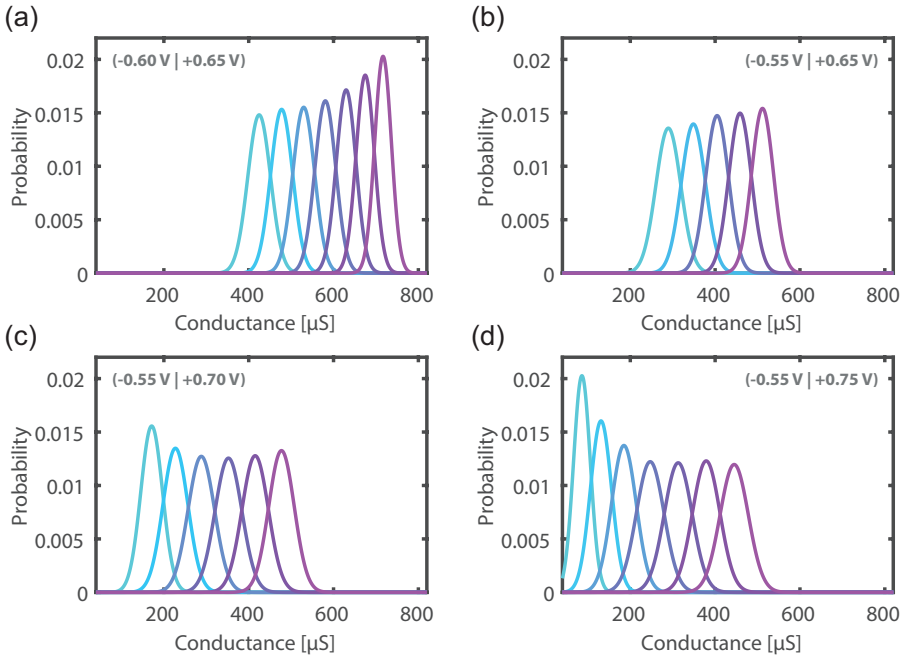


Figure 5.14: Histograms of seperable levels within the programmable conductance range for the ideal combinations from the previous analysis.

higher than the LTP to account for both the resolution and the nonlinearity of the physical SET process. At the same time, moderate LTP amplitudes allow a more symmetrical balancing. This means, that both the nonlinearity and the asymmetry mainly stem from the physics of the SET process. The best approach was found to match the LTD half cycle, which is characterized by the more controllable RESET dynamics, to match the LTP half cycle.

- During the noise analysis, it was found that maximum noise is present for devices in intermediate conductance. At the lower end as well as at the upper end, the noise decreases. On the low conductance side, the signal-to-noise ratio stabilizes to a near-constant value of about five, while the signal-to-noise ratio the high conductance side increases dramatically. Further investigation of this phenomenon is found in the following section.

- The switching part of the most promising sets of parameters in terms of programmable conductance range, resolution, and symmetry of inherent nonlinear-

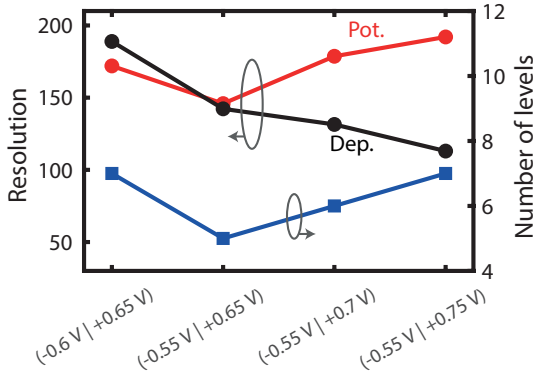


Figure 5.15: Summary of resolution and levels for the chosen parameters.

ity were recombined with the measured noise characteristics to determine the number of realistically programmable levels. In stark contrast to the theoretical resolution, which was in the proximity of 200, a 1 σ noise estimate was found to reduce this number to only 5 to 7 levels.

- The most promising switching combination is found exactly in the conductance range with the highest absolute noise values. This further limits the possibility of matching multiple states. At the same time, this is the range where the relative noise, i.e. the signal-to-noise ratio transitions from constant to higher, favorable values. This could be related to the conduction mechanism in the cells.

These observations can be understood as a consequence of the results found for the switching kinetics analysis in Section 4.5. The switching processes for LTP and LTD, i.e. SET and RESET, are fundamentally different from each other. The SET process is thermally activated and self-accelerating due to the thermal runaway phenomenon. In consequence, the LTP half cycle is mostly independent from the previous state and shows little to no dependence on the offset voltage or absolute conductance. Therefore, the absolute LTP voltage becomes the critical factor for resolution and nonlinearity. The RESET process is inherently different in two aspects. First, the gradually decreasing conductance causes a self-deceleration because the lower currents generate less heat. Second, the counteracting forces of drift and diffusion [112] further stabilize the conductance decrease. For these two reasons, the LTD half cycle

allows much better control. On one hand, the LTD voltage can be selected to adjust the resolution and the nonlinearity. On the other hand, the previously programmed configuration has a tremendous influence on the RESET dynamics, which can be seen in the strong dependence on the offset voltage between LTP and LTD.

It has been shown that the analog function of the devices is severely limited by the presence of Gaussian noise in the measurement, which significantly reduces the number of realistically separable states. It is interesting to note that the optimal switching conditions for LTP/LTD are found in a region with strong noise. This could indicate that the two phenomena are related in their physical nature. The ionic configuration in the conductance range for analog function is required to be easily disturbed by voltage pulses with small amplitudes, since stronger amplitudes trigger exponentially accelerated conductance changes. While such configurations are favorable for analog programming, they could be prone to unintended fluctuations or drifts for the same reason. The exact characteristics of the observed Gaussian noise in the experiment will be discussed in detail in the following section.

5.2 Analog state stability

The previous section has shown that the analog tunability is significantly affected by the presence of noise. In addition, an interesting relationship was found between the mean conductance and the noise standard deviation. This section aims to further explain this phenomenon in more detail. An experiment is designed to study noise in an isolated fashion over the range of programmable conductance that is relevant for applications where analog states may be employed. To do this, a single device is programmed into a defined conductance state by using a simple pulsed program-verify scheme utilizing both LTP and LTD voltages. Following the programming, the conductance is read by a signal of -0.2 V with the maximum frequency allowed by the setup, which is every 2 ms after the first read. Due to the delay of the measurement script, the first read signal is 5 ms delayed to the last program-verify read. The reading continues for about 1 s. The device function is then verified through cycling the device. This sequence is repeated 1000 times for the targeted conductance level. The same procedure is repeated for the next conductance level and so on. The programmed conductances are equally spaced in 20 µS increments from 20 µS to 800 µS, resulting in a total of 40 000 measurements of 1 s duration each. The programmed conductance has a tolerance of only ± 2 µS, which means that the conductances are well sepera-

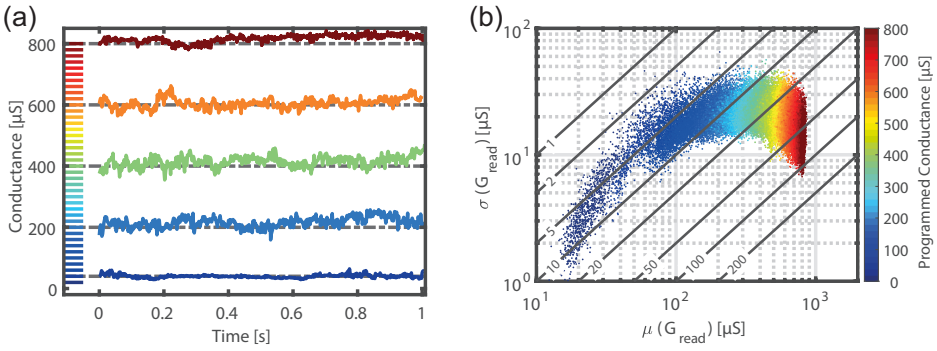


Figure 5.16: Overview of the designed read noise experiment. (a) Five exemplary read noise traces. Colors on the left indicate the programmed conductance levels. (b) Noise characteristic irrespective of time after programming and differences between repetitions.

ble after the last programming pulse. Exemplary measurements for 5 programmed conductances are shown in Figure 5.16 (a). The colors indicate the programmed conductance. The processing of the data is analogous to the noise analysis in the previous section. However, as the data consists only of read signals and no switching is performed, the full measurement time of 1 s is considered for the analysis. The mean conductance and the standard deviation are shown in Figure 5.16 (b). The previously observed bell-like shape is reproduced. The typical SNR is 5 until a conductance of roughly 100 µS is reached. Above that, the SNR increases. In contrast to the results of the previous section, the vertical spread of the conductances is relatively large. On one hand, this may be attributed to the 10x larger amount of data recorded in this experiment. On the other hand, the recorded duration is significantly longer. Consequently, the time dependence is further analyzed.

A closer look at a single measurement reveals that the noise is more pronounced in the first part of the values, directly after the end of programming. This effect is referred to as intra-trace noise in the context of this work. Figure 5.17 (a) shows as an example the first 300 ms of a trace programmed at 200 µS. The trace is divided into 100 ms intervals. Larger conductance jumps are observed in the first interval, while the second and third interval show comparatively less noise. The according conductance histograms for the three intervals in Figure 5.17 (b) emphasize this observation. The standard deviation is continously decreasing, from 20.1 µS for the first interval to 10.6 µS in the third interval. The intervals after 300 ms show a continuous

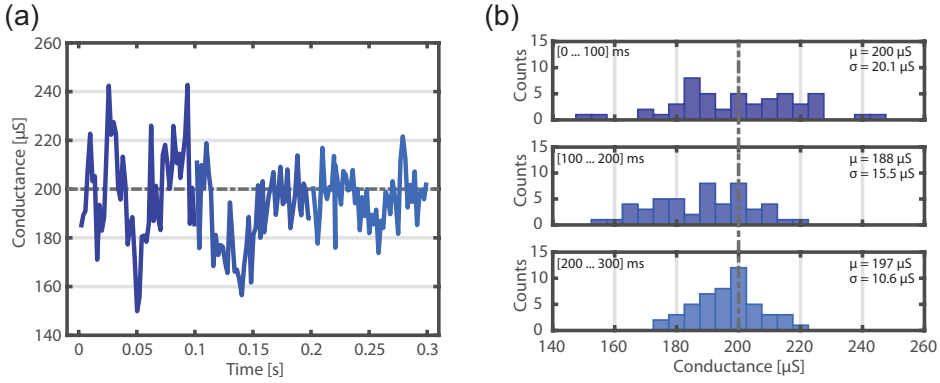


Figure 5.17: Example for intra-trace noise in the experiment. (a) First 300 ms of a trace programmed to 200 µS, divided into 100 ms intervals. (b) According conductance histograms. The standard deviation gradually decreases.

decrease, however with smaller margins. This observation is applied to the data of Figure 5.17 (b), which is divided into 10 time intervals of 100 ms each. The mean and standard deviation are shown in Figure 5.18 (a) with the data of the first interval plotted in blue and of the last interval in yellow. As expected, the majority of standard deviation values is lower for the last intervals compared to the first interval. The spread is still relatively wide in the vertical direction, but the trend is well reproduced. To evaluate the time dependence for all time intervals, the mean of the standard deviation is plotted against the mean of the mean conductance shown in Figure 5.18 (b). The reported trend of reduced noise for longer times is observed over the whole conductance range. The difference between first and second interval is most significant, while the following intervals change at a lower rate.

The vertical spread of points for one time interval in Figure 5.18 (a) remains after extraction of the intra-trace component. The difference between the individual traces in one time interval is that they are programmed individually and slightly different every time. In fact, each program-verify cycle is seperated by full switching into a high conductance and a low conductance state, resetting the conditions for every cycle. Figure 5.19 (a) shows the read signals of three consecutively recorded traces which were all programmed with the target of 200 µS. The resulting, actual conductances of 198.89, 198.57 and 200.08 µS are all within the ± 2.00 µS tolerance. The recorded conductance traces show very different trends, ranging from a conductance increase (green trace) to conductance decrease (blue). This phenomenon is termed inter-trace

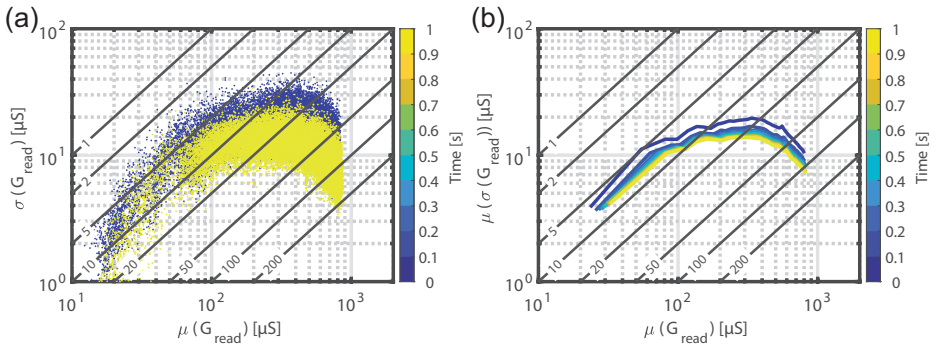


Figure 5.18: Analysis of the intra-trace noise for all programmed conductances. (a) First interval (blue) and last interval (yellow), indicating the reduction in noise with time. (b) Meaned relation for all times.

noise in this work. The magnitude of inter-trace conductance differences spans already 200 µS at the end of the measurement in this simple example. This implies that the inter-trace noise is far more severe than the intra-trace noise. Furthermore, the conductance difference between the traces appears to increase with time, whereas the intra-trace noise decreases. To capture the behavior for all coductances tested, the standard deviation of the mean conductances for the ten time intervals is plotted against the averaged conductances, as shown in Figure 5.19 (b). This analysis essentially removes the effect of the intra-trace noise, because the standard deviation in one time interval is not considered in this analysis. Instead, this shows the drifting apart of the mean conductance for the time intervals of the traces. In agreement with the observation of the example, the graph shows that the difference between the traces increases with time. Figure 5.20 shows the comparison of the two observed effects as an example for the traces programmed at 200 µS. In the first 100 ms, the noise types are nearly identical, because the traces are all programmed to the same conductance. The noise in the first few milliseconds is therefore dominated by the intra-noise phenomenon. For longer times, the splitting up of different traces increases, while the intra-trace noise decreases. Overall, the designed experiment allows several observations.

- The bell-like shape of the noise is reproduced. The SNR is also constant here at low to medium conductance values, but increases at higher conductance values.
- The spread of the noise is more severe than in the previous experiment which

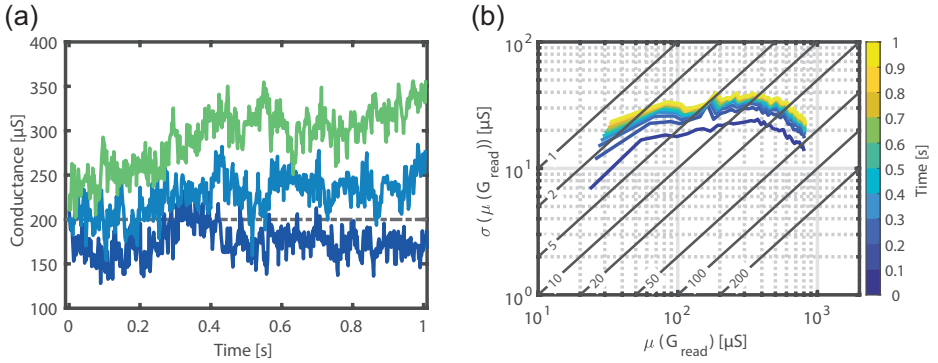


Figure 5.19: Inter-trace noise analysis. (a) Three subsequently recorded traces, programmed to 200 µS, showing increasing difference with time. (b) Inter-trace noise analysis for all times and conductances. The noise gradually increases in magnitude.

included LTP and LTD programming. Two underlying phenomena were found.

- The intra-trace noise evolves during the trace. It is characterized by a high standard deviation at the beginning of the trace and gradually decreases as the trace progresses. Intra-trace noise dominates in the first 100 ms of a trace, but is comparatively low at the end of the measurement.

- The inter-trace noise is the difference between traces taken at different times. It is characterized by a low magnitude right after conductance programming, but increases over the course of the trace. Its final magnitude is comparatively high, therefore it appears to be the dominating source of noise for read events performed at a longer times after programming.

Discussion

The difference between the time dependence of intra-trace noise and inter-trace noise suggests that overlapping mechanisms play a role when programming an analog conductance.

The observed intra-trace noise amplitude is explained first: The temporal development of the intra-trace noise amplitude can be divided into two sections. A time-dependent decay right after programming is followed by a constant noise value for long times. Typically, such decays at low applied voltage are explained by relaxation of volatile states which were charged during programming. One example of

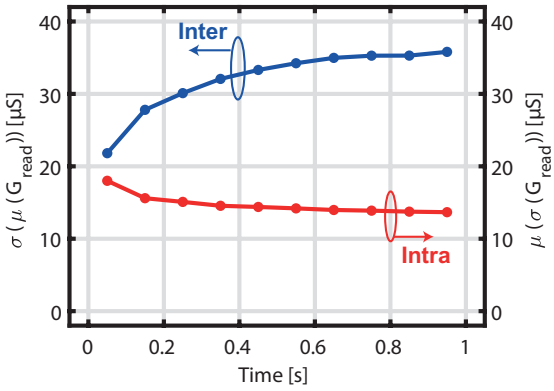


Figure 5.20: Example of intra-trace and inter-trace noise for the 200 µS programmed conductance. While intra-trace noise gradually decreases with time, the difference between traces increases and is overall stronger in magnitude.

such charged states has been reported by La Torre et al.[180], who observed a characteristic current decay for Ta_2O_5 and ZrO_2-based resistive switching devices when the subloop-feature found in the low conductance state was measured. Characteristic current decays after programming are also frequently found in area-dependent VCM devices, for example $SrTiO_3$ and TiO_2-based devices[181]. Different to those studies, the mean current does not really decrease in the presented experiment, but the noise does. The proposed explanation is, that as the charged states are emptied after the programming, the associated current jumps contribute to the measured intra-trace conductance noise. Towards the end of the measurement time, around 1 s after programming, the intra-trace noise has stabilized to a value different from zero. This indicates a time-independent source of noise.

Other studies also report on noise phenomomena in VCM devices. For example, Brivio et al.[176] have extracted the noise in a similar LTP and LTD experiment as shown in Section 5.1. The samples are HfO_2-based and very comparable to the devices of this work. The simulation and experimental data are given as function of resistance. However, upon transforming the graph to conductance data, a similar bell-shaped curve is obtained as seen in Figure 5.21 (a). The model they apply suggests that the oxygen vacancy concentration profile is slightly disturbed with each incoming voltage pulse. This means that the mean conductance averaged over many pulses is not changed, but noise, which they term Stimulated Telegraph Noise, is

induced. Another study was performed by Mao et al.[177], who measured the noise properties of their integrated Ta_2O_5-based devices. They show the measurement in conductance, and the bell shape is reproduced, see Figure 5.21 (b). However, they do not provide an explanation for the observed behavior, but fit the results to include the noise characteristic in their neural network simulations. An extensive study on the noise properties in the low conductance state was recently published by Wiefels et al.[178]. They employed ZrO_2-based resistive switches. Through a Kinetic Monte-Carlo simulation, it is found that vacancy jumps towards and away from the active Schottky interface have the most significant impact on the current fluctuations. Furthermore, the JART VCM v1b compact model is extended by a stochastic state machine to match these current fluctuations. The measurement data is presented as $\Delta R\,/\,V$ over the mean resistance. Transforming the curve into conductances yields the bell-shaped characteristic, which can be found in Figure 5.21 (c). However, towards low conductances the shape deviates from the other reports slightly. Finally, a recent study of Perez et al. [179] found that an array of HfO_2-based devices shows similar noise properties as depicted in Figure 5.21 (d). They split the experiment into LTP and LTD and found that the LTP is significantly more noisy, which was attributed to partial abrupt switching in the device ensemble. However, their LTD curve is almost identical to the other reports, most likely due to the more gradual nature of the RESET process. Both the model of Brivio et al.[176] and Wiefels et al.[178] suggest that the oxygen vacancy profile along the filament direction (vertical direction in the context of this work) is the decisive source of noise.

Neither study, however, addresses the bell-shaped curve that emerges in the conductance plots. To understand this feature, the conductivity mechanism of the device plays a crucial role, as it relates the oxygen vacancy profile to the measured currents. The recent ab initio simulation study of Funck and Menzel [93] proposes that the current conduction in typical VCM devices is in strong relation with the oxygen vacancy energy level in the oxide's band gap. For shallow defect levels, electrons tunnel through the Schottky depletion zone at the active electrode in its full length into the conduction band. For deep defect states, the model describes interface-limited electron transport and trap-assisted tunneling over the oxygen vacancy defects as the correct mechanism. In both cases, the Schottky barrier at the active electrode is current limiting, but the transport mechanism is different. As summarized by Funck and Menzel, the oxygen vacancy defect energy levels in HfO_2, ZrO_2 and Ta_2O_5 may all be considered in the category of deep defects. Furthermore, it is known that the con-

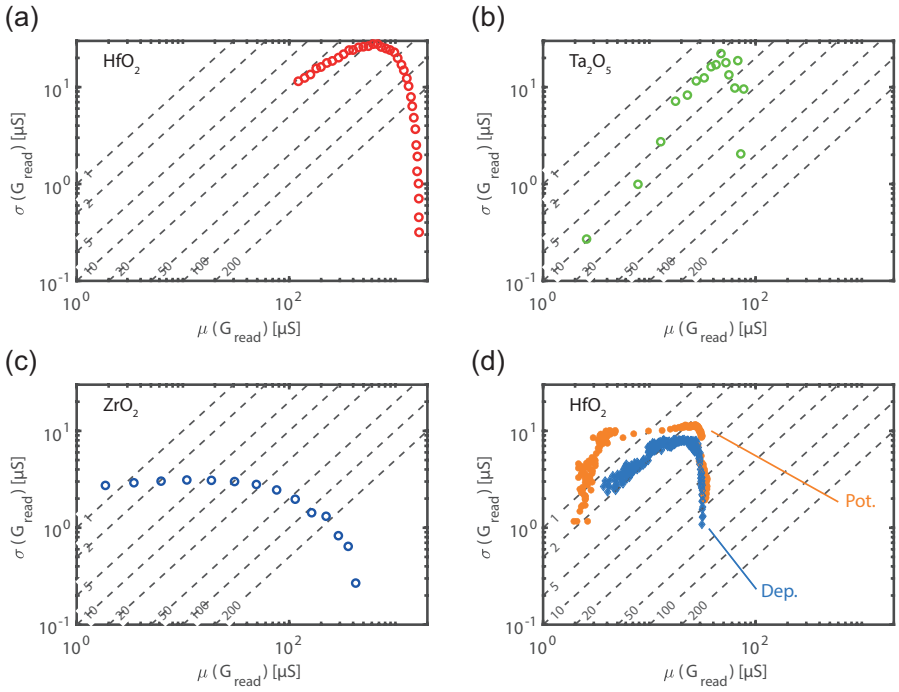


Figure 5.21: Noise characteristics reported in literature. (a) Transformed simulation data from the study of Brivio et al.[176]. (b) Experimental data from the study of Mao et al.[177]. (c) Transformed compact model simulation of Wiefels et al.[178], based on the JART VCM v1 model also used in this work. (d) Experimental data from the study of Perez et al.[179].

duction mechanism in these systems changes from an exponential *I-V* characteristic for low conductance to a linear one at higher conductance.

By combining the noise mechanism of oxygen vacancy profile perturbations and the conduction mechanism for HfO_2, a comprehensive picture for the bell-shaped curve is proposed. This is illustrated in Figure 5.22. According to the model the intra-trace noise curve in the last measurement interval is assigned to the three ion configurations at the AE interface labeled A, B, and C. At low conductances (region A), the mechanism of trap-assisted tunneling determines the noise. Since the current flow happens via the oxygen vacancies, the vertical position of the vacancy closest to the AE has a strong impact on the noise. Slight position changes of this vacancy alter both the mean conductance and the noise simultaneously. Therefore, the signal to

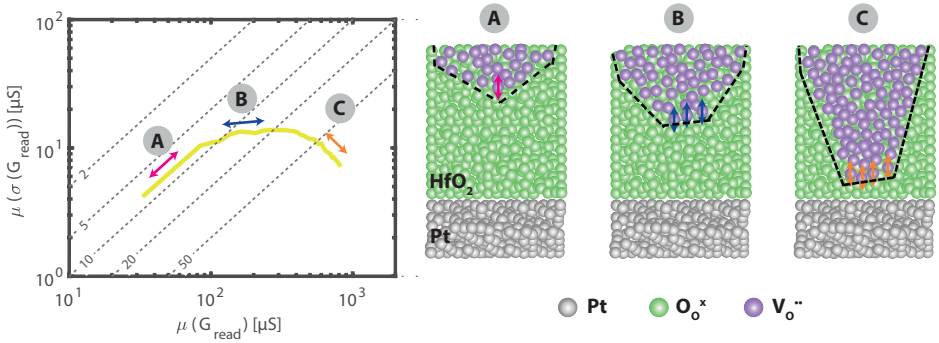


Figure 5.22: Proposed model for the intra-trace conductance noise based on the current conduction mechanism in HfO_2 devices. In region A, both conductance and noise are linked through the position of the last vacancy. In region C, many oxygen vacancies are accumulated, reducing the influence of single position perturbations. Region B is the transition between A and C.

noise ratio stays constant in this regime. For the other extreme at high conductance (region C), many oxygen vacancies are accumulated close to the Pt interface. Position changes have only a minor influence on the current, because there are many alternative defects that serve as conducting pathway for the current. The bell-shape of the measured noise hence describes the gradual transition between these two extreme scenarios (region B). Starting at low conductance, a single current conducting vacancy tip is pointed towards the AE. At the peak of the bell-shaped noise curve, its position is closer to the AE, but it is still a mostly single conductance path. Further increase activates more available paths, which lowers the influence of a single vacancy. Towards high mean conductance, many conductance paths through alternative defect positions are available. Therefore, the influence of single position perturbations is gradually minimized.

The compact model of Wiefels et al. is basically identical to the one that was previously employed to match the experimental data in Section 4.5. In a later chapter, the parameters of Table 4.1 will be slightly refined to match a stochastic switching dataset, see Chapter 8. Therefore, the parameters extracted there, see Table 8.1, are used to calculate the intra-state noise characteristics for the sample of this work. As will be explained later, the model has the capability to model cycle-to-cycle variability and device-to-device variability properties. For the shown simulation in this section, the device-to-device variability is set to zero and the variability in the simulation data

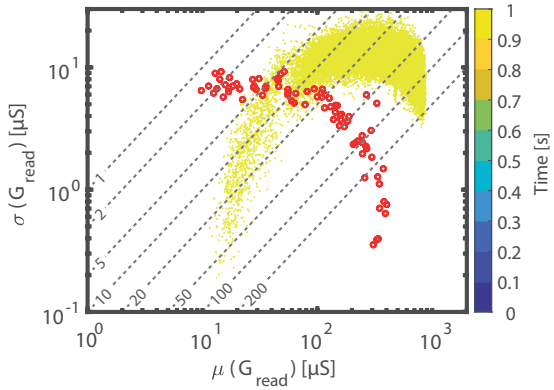


Figure 5.23: Comparison of the intra-trace noise in the last time interval (yellow dots) with the JART VCM v1 compact model simulated noise (red circles), using the state machine probabilities from [178] and the simulation parameters from Table 8.1. Device-to-device variability is set to zero.

hence is only cycle-to-cycle variability. The results are shown in Figure 5.23. Since the simulation does not include relaxation effects, but comprises the cycle-to-cycle variability, the final time interval of the data is chosen for comparison. The simulation results are shown as red circles. Each point shows the calculated mean conductance and conductance standard deviation of a 1 s long simulated read signal. While the general bell shape is reproduced as expected, a significant discrepancy between the simulation and the measurement data is evident. However, the differences can be understood as follows. The difference in the mean conductance could be removed by increasing the filament radius to allow higher current, i.e. lowering the device conductance. The experimental data originate from a device which has a relatively high conductance, which explains the difference. The comparatively small difference in noise amplitude on the y axis can be reduced by adapting the probabilities of the state machine. In the shown simulation, they were left unchanged from the ones given in the study of Wiefels et al.[178]. The compact model utilizes electronic band transport as described in Section 2.2.3. The trap-assisted tunneling character of the conduction, which was described in the proposed model above, is therefore missing. The descreasing noise with increasing average conductance is well reproduced because the difference between the trap-assisted conduction mechanism and the band conduction mechanism is small for high oxygen vacancy concentrations. However,

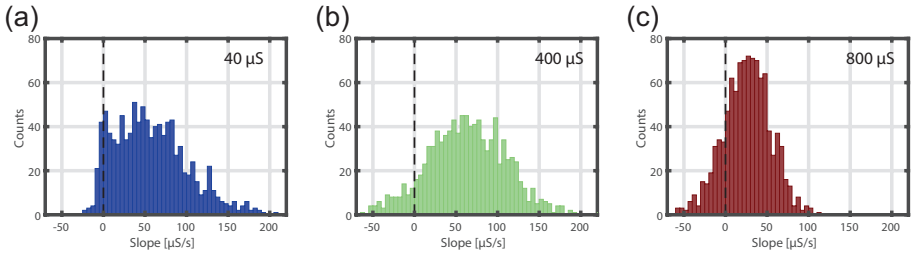


Figure 5.24: Slopes of the linear fits to each trace for programmed conductance of (a) 40 µS, (b) 400 µS and (c) 800 µS.

the simulation fails to reproduce the trend to constant signal-to-noise ratio at low conductances. This feature is inherent to the physical model of the simulation and would be hard to adjust through parameter choice only. A better fit to the data would therefore require to change the underlying physical model of the compact model, which is ongoing work at the time of this thesis.

The second feature is the inter-state noise, which is the increasing difference between traces of the same programmed conductance with time. To evaluate the underlying physical mechanism, a linear fit is applied to each trace. The conductance at $t = 0\,\mathrm{s}$ is set to the programmed conductance. The slopes for three representative programmed conductances, namely 40 µS, 400 µS and 800 µS are shown as histograms in Figure 5.24 (a) to (c), respectively. The dashed vertical line at zero slope highlights that all three distributions are above zero on average, but also extend below zero in different degrees. For the programmed conductance of 40 µS, the majority of data is in the positive regime. At 400 µS, the trend is even more pronounced. At 800 µS, a larger portion of traces shows a negative slope and the mean slope is closer towards zero. The distributions show that the inter-trace noise is not identical to a deterministic conductance drift, since there are both positive and negative slopes. However, one component that could explain the tendency to higher conductances could be the oxygen vacancy concentration gradient that is built up along the filament direction during programming. While the exact fields and temperatures in the disc region are difficult to estimate, it can be expected that diffusion of oxygen is possible, which can explain the positive conductance change component. However, the negative component is unclear and requires more experiments in the future.

To summarize, it was found that the noise of programmed analog conductance states is composed of two parts: First, a single programmed conductance shows a base-

line noise associated with oxygen vacancy position perturbations. The characteristic bell-shaped noise curve could be explained by taking into account the conduction mechanism of HfO_2-based VCM devices and the potential oxygen vacancy jumps along the filament direction. The first few hundred milliseconds of the traces show an additional decaying noise, which may be related to the relaxation of volatile states that were charged during the programming. Second, the programmed conductance states exhibit a non-deterministic conductance drift. The majority of traces drifts towards higher conductances, which may be attributed to diffusion caused by the oxygen vacancy gradient. The counteracting force of negative drift slopes is still under debate and requires more experiments.

6 Spike Timing Dependent Plasticity

In the previous chapter the conductance programming was described using rectangular wave voltage signals with constant amplitude. This mode of operation is mainly used for learning in deep neural networks or for computation in memory concepts. In brain-inspired Spiking Neural Networks (SNNs), an alternative approach for programming the conductance, the so called Spike Timing Dependent Plasticity (STDP) is often used. The current understanding of the biological system is that the synaptic strength should be modulated when two spike emitting nodes send signals, which may coincide at a synaptic connection [182]. In biology, these voltage signals are generated by neurons, and the connecting elements are synapses. The electronic device analogy are electronic neurons and electronic synapses. Electronic neurons can be constructed from transistor circuits or from novel, volatile devices, which are the topic of active research[183–188]. Nonvolatile memristive devices are better suited as electronic synapses. The two main pathways for synapses in SNNs have been identified as Spike Rate Dependent Plasticity (SRDP) and STDP. For SRDP, the device conductance is modulated when a train of identical voltage signals with a certain frequency reaches the synapse. The modulation should be dependent on the frequency of the incoming signals. The study of Nishi et al.[111] has shown that the switching dynamics of VCM devices, operated in the binary mode, does not change when a single pulse is split into shorter pulses, i.e. the frequency of the applied pulse(s) does not change the device response. In Section 5.1, the analog operation mode was investigated. The findings suggest that the analog operation follows the principles of the binary mode. Hence, it is expected that VCM devices in general behave frequency-independent. For this reason, SRDP is not included in this study. For STDP, the device conductance should be modulated when the pre-synaptic neuron and post-synaptic neuron fire with a certain timing relative to each other. Because

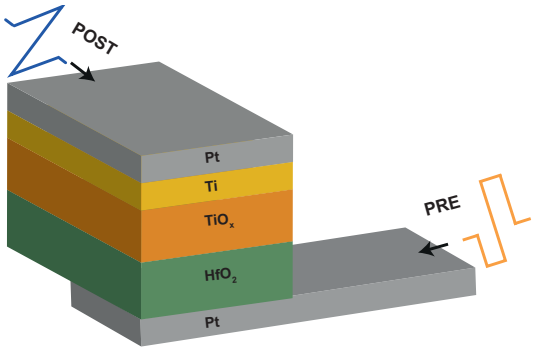


Figure 6.1: Concept of STDP in VCM crosspoint devices. Signals are sent to the device from the presynaptic side, here defined as bottom electrode, and the postsynaptic side, here top electrode. Signals coincide depending on their timing.

VCM devices are inherently different from their biological counterparts, the present study does not attempt to mimick their signal amplitudes and shapes. However, it will be shown that the filamentary VCM device dynamics allows for implementation of some of the basic principles of STDP. The principle of STPD in VCM crosspoint devices is shown in Figure 6.1. By definition of this work, the voltage signal coming from the presynaptic side is applied to the bottom electrode, while the signal from the postsynaptic side is applied to the top electrode.

In the following, three combinations of presynaptic and postsynaptic waveforms will be discussed. The first combination is composed of a short rectangular signal and a longer triangle voltage waveform. It is designed to translate the well understood relationship between voltage amplitude and conductance modulation into a STDP scheme. The second combination is composed of two identical triangle waveforms. While this selection of waveforms is easier to implement in a circuit, the results show that the control of the STDP response is more challenging. The third combination is composed of two double triangle waveforms. This combination is the most complex. It demonstrates that a single overlapping event can induce more than one switching operation. The final STDP response is determined by the sequence of positive and negative voltage polarity signals and the voltage balancing.

Square + Triangle Waveforms

Figure 6.2 shows the first applied STDP scheme. It is strongly asymmetric in pulse geometry chosen, see Figure 6.2 (a). The upper panel shows the presynaptic signal,

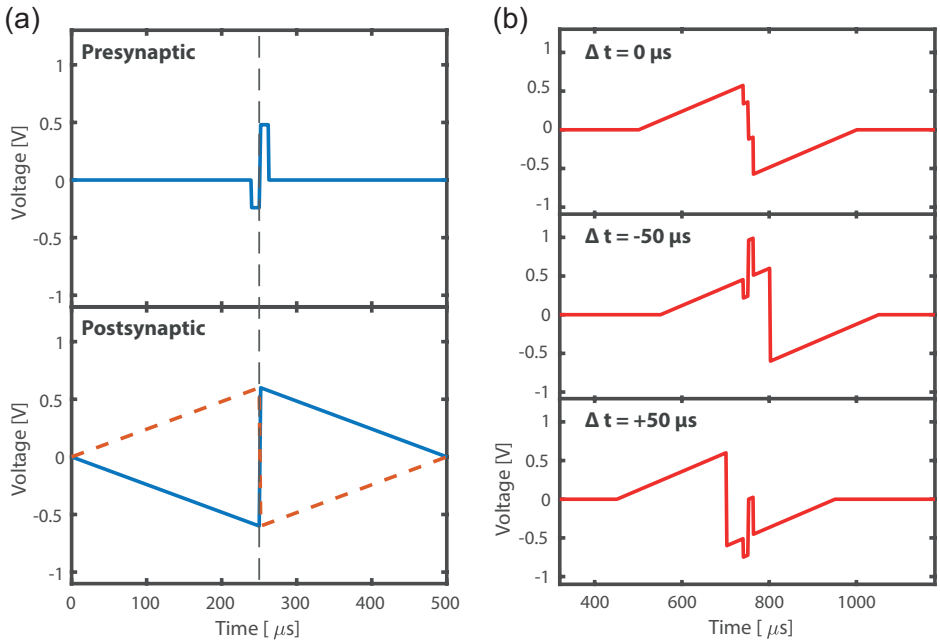


Figure 6.2: First tested STDP waveform combination. (a) shows the presynaptic (upper panel) and postsynaptic (lower panel) waveforms. The postsynaptic waveform is shown in its original form (solid line) and inverted (dashed line), which is in accordance to the active electrode referencing of this work. (b) shows three distinct examples of overlap between the waveforms.

which is applied to the BE of the device. It consists of two 10 µs long rectangular pulses with opposite polarity. The negative amplitude is intentionally smaller to compensate for the asymmetry of the device response. The reason behind this is explained in Section 5.1. The postsynaptic signal, shown in the lower panel, is chosen as two triangular voltage ramps with symmetrical amplitude in negative and positive polarity. The waveform duration is much longer than the presynaptic one at 250 µs per ramp. The solid line in the graph shows the voltage waveform on the device top electrode. For convenience, the inverted voltage is also shown as dashed line. The vertical dashed line in both panels shows the reference time for the following overlaps. In both waveforms, the voltage changes from the maximum (minimum) to the opposite polarity minimum (maximum) amplitude within 2 µs. Figure 6.2 (b) shows three exemplary cases of time-dependent overlap of the voltage signals. The displayed voltage is the summed voltage on the BE, while the TE would be on ground in this

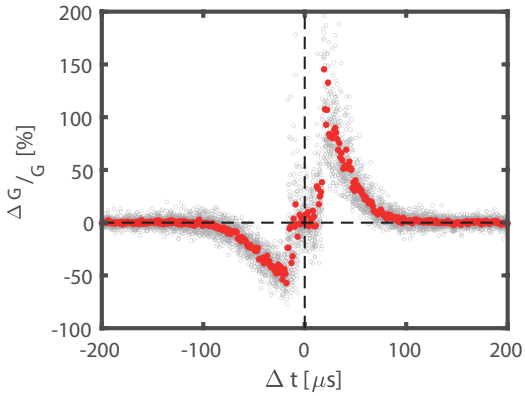


Figure 6.3: Synaptic response to the waveforms shown in Figure 6.2 at various Δt.

case. This follows the convention in this work. In the upper panel, the special case of $\Delta t = 0\,\mu s$ is illustrated. The addition of presynaptic and inverted postsynaptic signal leads to a reduction in both positive and negative absolute amplitude compared to the individual signals as they cancel each other out. At Δt = -50 µs, i.e. when the presynaptic signal reaches the device slightly before the postsynaptic signal, the waveform sum leads to an enhancement of the positive amplitude. This case is shown in the middle panel. The total negative polarity on the device is essentially given by the inverted postsynaptic signal. The lower panel shows the opposite case, when the presynaptic signal arrives after the postsynaptic one. Here, the positive polarity is shaped by the inverted postsynaptic waveform, while the negative side is affected by the addition of presynaptic and inverted postsynaptic waveforms. The experiment for this waveform was carried out in the following way: Δt was varied between -500 µs and +500 µs in steps of 1 µs. Importantly, the testing was started at $\Delta t = 0$ and continued to higher absolute values of Δt symmetrically ($\Delta t = 0\,\mu s$, +1 µs, -1 µs, ...). This is crucial in avoiding unwanted conductance runaway to potentially stuck states. The sequence was repeated 15 times to eliminate the effect of outliers. The device was cycled 5 times between repetitions to ensure normal functionality. Before each repetion, it was programmed to an initial conductance of 200 µS, which was identified to be in the conductance range that allows analog conductance modulation, see Section 5.1. However, the conductance was free to change in response to the applied voltages, meaning that the G in $\Delta G / G$ is not constant and most likely not 200 µS. This is important to note when $\Delta G / G$ goes to high values. Figure 6.3 shows the

resulting synaptic response to the given waveforms at various Δt. Shown is the relative conductance change ΔG compared to the conductance G before the waveform is applied to the device. Grey open circles show every measurement, while red circles represent the mean value of the 15 repetitions. The x axis is shortened to the relevant range between -200 µs and +200 µs. For higher timing differences, the conductance remained constant apart from noise, i.e. $\Delta G / G = 0$. This shows that the application of the individual waveforms did not elicit a synaptic response. The shape of the synaptic response is easily understood from the examplary waveforms illustrated in Figure 6.2 (b). Starting at negative Δt and going towards zero, the signal overlap enhances the positive polarity more and more, which induces increasing LTD of the device conductance. As Δt approaches zero, the overlap of the waveforms first leads to a reduction in pulse length of the strengthened positive polarity. In fact, this can be seen by the few points which lie between the minimum $\Delta G / G$ and the data points around zero. In the region around $\Delta t = 0$, the overlap cancels each other out as shown in the upper panel of Figure 6.2 (b). For increasing Δt, the above described process repeats for the negative polarity, leading to LTP of the device conductance. The presynaptic voltages are chosen as (-0.24 V | +0.48 V), and the symmetric voltage maximum/minimum of the postsynaptic waveform is (-0.60 V | +0.60 V). The extreme voltage combinations are therefore (-0.84 V | +1.08 V), which is higher than the typical voltages found for LTP and LTD using constant signals, see Section 5.1. The data in Figure 6.3 already shows signs of strong programming conditions close to the described voltages, which are reached at Δt = -10 µs and Δt = +10 µs. However, the symmetrical testing procedure described above allows the conductance to stay within a programmable window, and the ramped shape of the postsynaptic waveform quickly reduces the programming amplitudes to typical levels. To study the influence of the voltage levels, the described experiment is repeated using the same waveform shape. The amplitudes are changed systematically, see Figure 6.4 (a). The results of Figure 6.3 are shown in yellow color, meaning that both higher and lower amplitudes were tested. The lower panel in (a) only shows the inverted postsynaptic waveform. The mean synaptic responses for the various amplitudes is shown in Figure 6.4 (b), upper panel. The overall shape of the response curve largely remains identical to before. However, for stronger voltage amplitudes than the above described case of (-0.24 V | +0.48 V) | (-0.60 V | +0.60 V), the LTP response increases dramatically, indicating strong programming conditions. The lower panel of Figure 6.4 shows the maximum and minimum applied voltage for the tested Δt range. Both panels to-

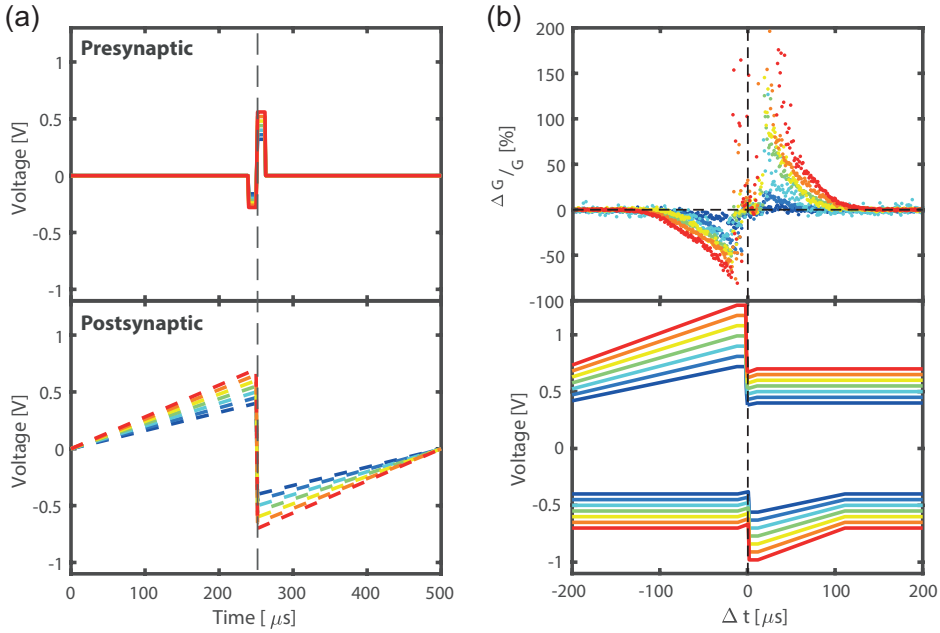


Figure 6.4: Amplitude scalability demonstration. (a) Presynaptic (upper panel) and inverted postsynaptic (lower panel) waveforms. (b) According synaptic responses (upper panel) and maximum and minimum amplitudes of the waveforms (lower panel).

gether demonstrate that the given waveform combination mainly utilizes the voltage control to program the device. Overall, this clearly illustrates the voltage scalability of this STDP waveform combination. Regarding the time response, the postsynaptic waveform could be made even longer, i.e. the ramps having a lower slope. It is easily imagineable that the synaptic response would be stretched to longer timescales. This highlights the versatility of this waveform combination.

This STDP programming example utilized a waveform scheme which essentially exploits the timing of the two pulses to generate a single, nearly rectangular voltage pulse. It essentially translated the time dependence into a voltage dependence. As discussed in Section 4.5, it is relatively easy to program roughly targeted conductances with a single pulse if the starting conductance is in a suitable range. Due to the strong nonlinearity of the switching kinetics, however, true time encoding is far more challenging.

Triangle + Triangle Waveforms

Another possibility for waveforms in STDP presents itself in using two triangular waveforms, where one is slightly asymmetric to compensate for switching asymmetry. To make the waveforms comparable, the postsynaptic signal of the previous example is kept, see Figure 6.5 (a) lower panel. Note the original signal as solid line and the inverted signal, which is with reference to the active electrode, as dashed line. The presynaptic waveform is a similar triangular shape, see upper panel. However, the amplitude in negative polarity is lower, which is the attempt to compensate the switching asymmetry. In the shown waveforms, the minimum and maximum voltages of the presynaptic signal are (-0.2 V | +0.5 V) and for the postsynaptic signal (-0.5 V | +0.5 V). Figure 6.5 (b) upper panel illustrates the resulting waveform for $\Delta t = 0\,\mu s$. The signals cancel each other out for the most part. Only the difference in slope is evident by the stronger postsynaptic amplitude. The middle and lower panel of the Figure show the cases for Δt = -100 µs and Δt = +100 µs, respectively. The resulting waveform is once again a nearly rectangular pulse with higher amplitude, surrounded by lower amplitude voltage signals where the waveforms are of different polarity. Because of the presynaptic waveform asymmetry, negative Δt lead to a rectangular voltage shape, while positive Δt lead to a distorted shape. This distortion should be beneficial in shortening the time that strong amplitude is applied. The shown waveforms are tested identically to the first waveform combination. The resulting synaptic response is shown in Figure 6.6. Important to note, especially compared to the first waveform combination in Figure 6.3, is the difference in x and y axis length. The values of $\Delta G / G$ exceed the mark of 400 % frequently for Δt around +100 µs. At the same time, the response starts and ends around -200 µs and +200 µs, while the first waveform combination was limited to the Δt range between -100 µs and +100 µs. This confirms that the individual length of the waveforms plays a significant role in the extent of the synaptic response curve. Going from negative Δt towards zero, the curve changes from the expected zero response to increasingly negative values. The Δt range between -200 µs and around 0 µs looks similar to the response curve of the first waveform combination, see Figure 6.3. The main difference is the stronger response, which leads to flattening towards $\Delta G / G$ = -100 %, which is of course the minimum possible value. As expected from the waveforms, the response around zero is zero. However, the following LTP repsonse is very strong, but similar in shape to the first waveform combination. To investigate if this waveform combination also

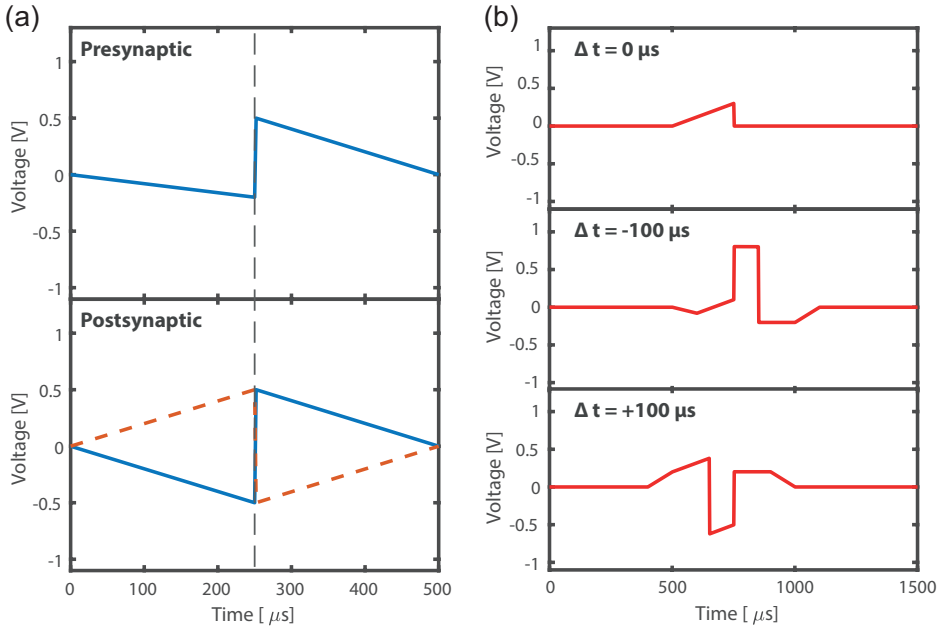


Figure 6.5: Second tested STDP waveform combination. (a) shows the presynaptic (upper panel) and postsynaptic (lower panel) waveforms. The postsynaptic waveform is shown in its original form (solid line) and inverted (dashed line), which is in accordance to the active electrode referencing of this work. (b) shows three distinct examples of overlap between the waveforms.

allows for voltage scaling, the experiment is repeated with more voltages. The data of Figure 6.6 is shown as yellow curves in Figure 6.7. The tested presynaptic and postsynaptic voltage scaled combinations are shown in (a), while (b) shows the respective synaptic reponses (upper panel) and the maximum and minimum voltage amplitudes of the tested waveforms. The synaptic responses curves show that a slight increase in voltage (orange and red color) strengthens both LTD and LTP significantly, with $\Delta G / G$ values approaching -100 % for Δt around -100 µs and reaching extreme values up to +5000 % around +100 µs. Of course, such values exceed the window of analog conductance modulation significantly and point towards bistable switching. Upon lowering the voltage scaling (green to blue curves), the synaptic response on the other hand quickly diminishes, likely because the voltages are just not sufficient enough, see lower panel in Figure 6.7 (b). This result is unexpected as the given waveforms should not introduce such strong responses for moderate voltages. However, one explanation

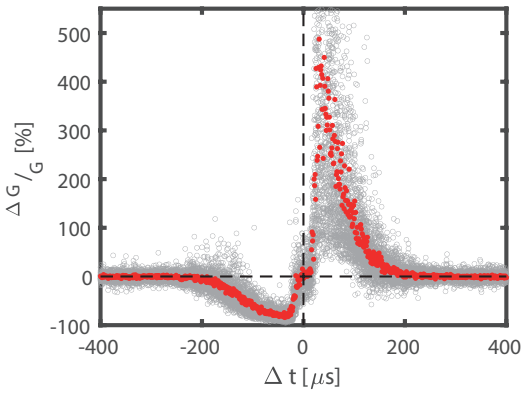


Figure 6.6: Synaptic response to the waveforms shown in Figure 6.5.

may be found in the way this waveform combination is expected to interact with the device. Because the signals are on the same timescale in duration, the overlap of pre- and postsynaptic signal not only encodes a maximum or minimum voltage like in the first example, but additionally manipulates the duration of the "strong" amplitude. This entanglement of voltage and duration is illustrated in Figure 6.8. Here, five exemplary negative Δt are shown. By increasing Δt from -10 µs to -100 µs, both the peak positive voltage is reduced and the pulse duration is increased. While the maximum positive voltage is shown the lower panel of Figure 6.7 (b), the duration information has not been considered so far. However, as discussed before, the filamentary VCM devices in this work exhibit very strong voltage-time nonlinearity, meaning that the time encoding in the shown voltage scheme does not affect the devices significantly because the time only marginally changes compared to the exponential dependence of VCM devices on time. The shown STDP waveforms may therefore be a better candidate for resistive devices with a less pronounced voltage-time nonlinearity.

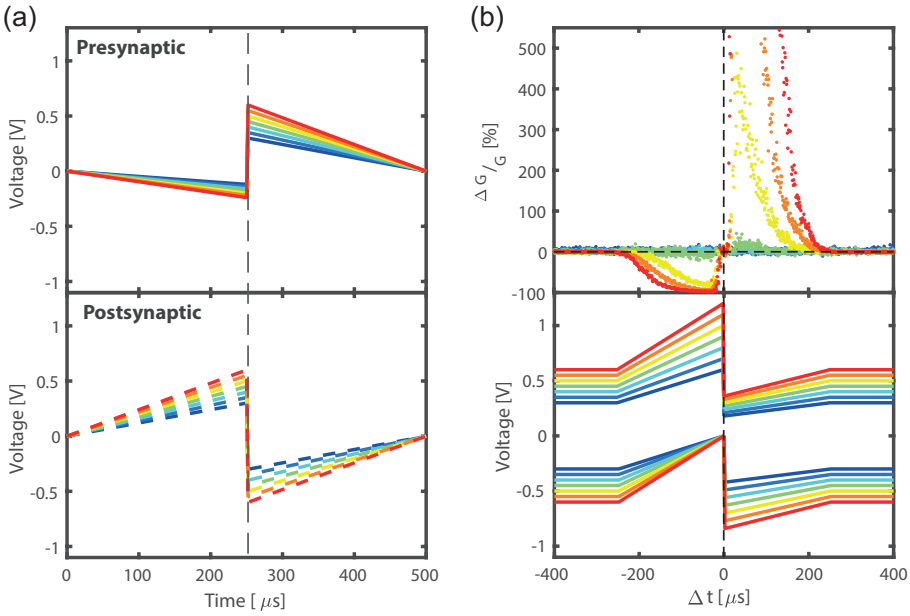


Figure 6.7: Amplitude scalability demonstration for the second waveform combination. (a) Presynaptic (upper panel) and inverted postsynaptic (lower panel) waveforms. (b) According synaptic responses (upper panel) and maximum and minimum amplitudes of the waveforms (lower panel).

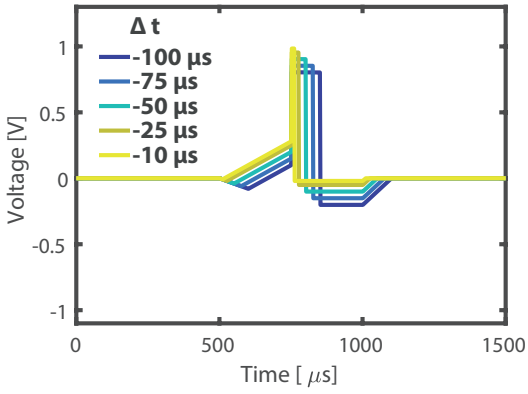


Figure 6.8: Demonstration how voltage amplitude and duration are entangled in the second STDP waveform combination.

Double triangle + double triangle Waveforms

In the previous two examples, the conductance modulation was achieved by overlapping two voltage waveforms which resulted in one dominating voltage pulse and some surrounding voltage signals with smaller amplitude. This section will demonstrate an alternative approach. Instead of a single voltage pulse that is sufficiently high to modulate the conductance, multiple sufficiently high voltage signals are generated. The deciding factor for the conductance change therefore is not a singular additive voltage amplitude and polarity, but instead the balance and sequence of the consecutive signals. For this purpose, the waveforms shown in Figure 6.9 (a) are employed. The upper panel shows the presynaptic signal, while the lower panel illustrates the postsynaptic signal and the inverted polarity to conform with the active electrode definition in this work. The waveforms are termed "double triangular" and consist of both negative and positive polarity sloped voltages. The reference time is at 300 µs, where the signals have their maximum, as indicated by the vertical dashed line. The amplitudes of the presynaptic waveform is slightly lowered compared to the postsynaptic one. Figure 6.9 (b) shows three cases for different Δt. In the upper panel, the case of $\Delta t = 0\,\mu s$ shows that once again the waveforms are chosen so that the positive polarity is generally favored to compensate the switching asymmetry in the VCM devices. The middle and lower panel illustrate the cases of Δt = -100 µs and $\Delta t = +100\,\mu s$, respectively. as described, the timing of these overlaps leads to significant amplitudes in both polarities. However, there is a distinct difference between the two cases: At Δt = -100 µs, a relatively strong positive voltage peak is followed by a negative one and a low amplitude positive one. The finishing, significant pulse of the sequence is of negative polarity. In contrast, at $\Delta t = +100\,\mu s$, the leading peak is an insignificant positive one, then a comparably strong negative one followed by a positive peak. The last significant pulse of the sequence is therefore of positive polarity. The discussed waveforms are reflected in the synaptic response curve shown in Figure 6.10. At Δt = -100 µs, the device is strongly potentiated, while $\Delta t = +100\,\mu s$ triggers LTD. At $\Delta t = 0\,\mu s$, the conductance remains unchanged, similar to the previous two examples. Different to the previous waveforms is the larger flat region around $\Delta t = 0\,\mu s$. Another difference to the previous examples is that the synaptic response curve shows both LTD and LTP for both signs of Δt, which results in a total of 4 peaks in the response curve. The reason for this is the different dynamic response of LTP and LTD. The strength of these peaks can be modulated by scaling the voltages

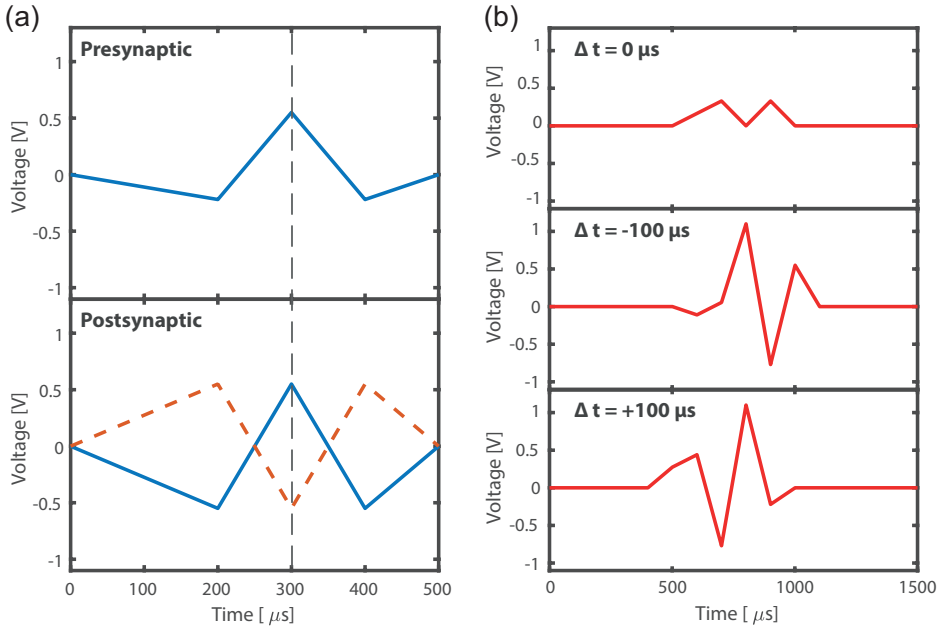


Figure 6.9: Third tested STDP waveform combination. (a) shows the presynaptic (upper panel) and postsynaptic (lower panel) waveforms. The postsynaptic waveform is shown in its original form (solid line) and inverted (dashed line), which is in accordance to the active electrode referencing of this work. (b) shows three distinct examples of overlap between the waveforms.

of the input signals. The tested waveforms are shown in Figure 6.11 (a). The data of Figure 6.10 is shown in yellow color. Slightly increasing the voltages strengthens all four peaks in the response curve significantly, see Figure 6.11 (b) upper panel. By lowering the voltages, the four peaks are reduced in magnitude. Because of the accelerated dynamics of the SET process, this lowering is most visible for the peak at Δt = -100 µs. The maximum and minimum voltage amplitudes of the waveforms in the lower panel of Figure 6.11 (b) are not able to explain this effect since both minimum and maximum voltages peak at Δt = -100 µs. In fact, they peak a second time at Δt = +100 µs with the same value. However, the resulting synaptic response is an LTD. This points to the fact this waveform combination induces bipolar signals, where the balancing between LTP and LTD as well as the finishing pulse in the voltage sequence are the deciding factor.

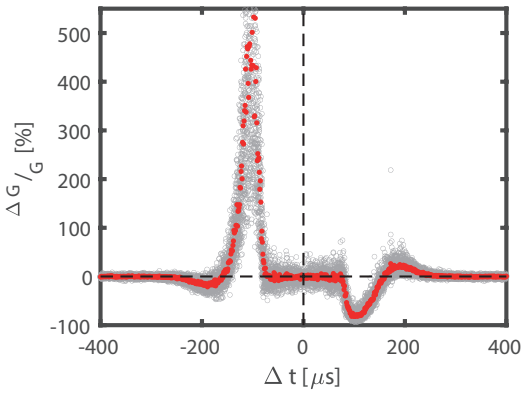


Figure 6.10: Synaptic response to the waveforms shown in Figure 6.9.

Summary of STDP operation

The presented three operation examples illustrated the versatility of the studied HfO_2-based filamentary VCM devices as synaptic devices. The tested waveform combinations, which are specifically designed for operation with the tested device, were successful in achieving plastic behavior. The first combination of a short rectangular bipolar pulse and a slow bipolar triangular pulse exploited the well-understood switching dynamic behavior of the SET and RESET process to generate ideal voltage waveforms for analog conductance modulation. By using the vastly different timescales of pre- and postsynaptic waveforms, the relative timing of the pulses translates the strong voltage-time nonlinearity of the switching kinetics, see Section 4.5, into an operation principle that operates on a linear timescale. In the chosen example, the presynaptic waveform was on the 10 µs time scale, while the postsynaptic was in the 100 µs time scale. With voltage modifications that are easily derived from the switching kinetic curves, these time scales may be changed down towards the Nanosecond range or towards the Millisecond range, which is closer to the biological counterpart. A more linear response may be facilitated by manipulating the postsynaptic waveform. For example, a more linear synaptic response could be realized by changing the voltage ramps to a slightly convex shape. Another example would be the expansion of the flat region around the diagram origin by decreasing the slope between the maximum and minimum amplitude of the postsynaptic waveform. The second example demonstrated why signals on the same timescale are inherently more difficult to adjust to

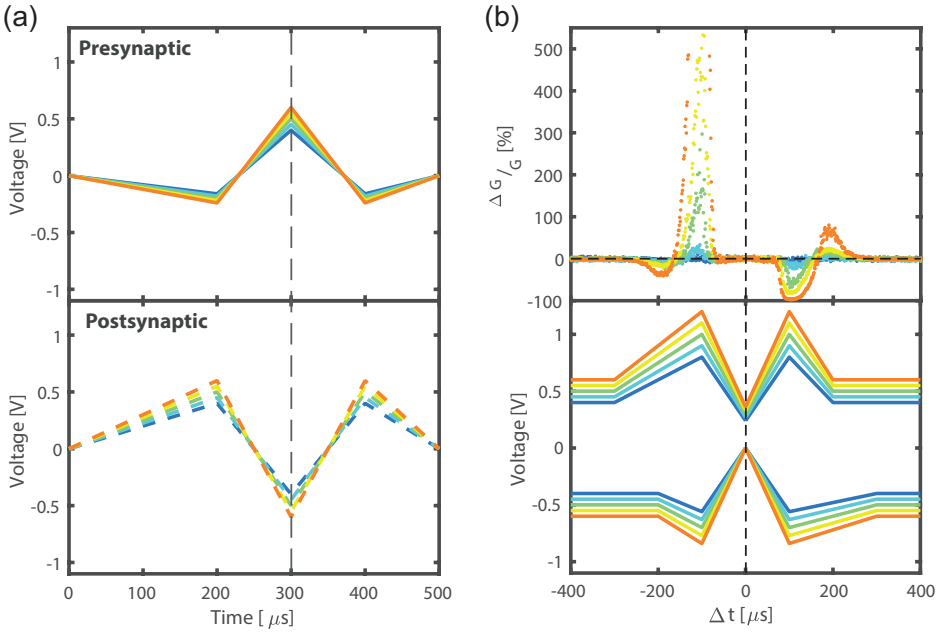


Figure 6.11: Amplitude scalability demonstration for the third waveform combination. (a) Presynaptic (upper panel) and inverted postsynaptic (lower panel) waveforms. The vertical dashed line marks the reference time. (b) According synaptic responses (upper panel) and maximum and minimum amplitudes of the waveforms (lower panel).

filamentary VCM devices. While synaptic response was observed and the response timescale followed the input waveforms, precise tuning proved difficult. The operation with such waveforms is likely better suited for devices with less pronounced voltage-time nonlinearity. In the context of ReRAM devices, area-switching materials have demonstrated such behavior and may therefore be a good candidate for this. The third example illustrated another operation principle, where the coincidence of pre- and postsynaptic waveforms generates a sequence of voltage peaks which individually are sufficiently strong to induce conductance modulation. Here, the timing sequence and the relative balancing dictates the device response. This approach proofed to be the most complex, but also most versatile STDP type as multiple peaks emerged, whereas before the correlation between sign of the the timing difference and conductance modulation direction were in an injective relationship to each other. Another point to mention is that the appearance of extreme responses is not necessarily undesired. For example, the complete turning on or off of a synaptic connection in a

special event can be desired in some applications. This of course strongly depends on the application.

In summary, STDP presents an alternative approach to program filamentary VCM devices. The presented cases underline that the devices in this work are a potential candidate for application as electronic synapses in Spiking Neural Networks. The operation should however be carefully chosen to the device physics. If properly executed, the device response can be tuned accurately.

7 SET and RESET switching variability aspects

Section 4.5 discussed the switching process as a two-step process composed of a delay part and a transition part. However, the discussion about the variability phenomenon was deliberately left out to bring across the message of the underlying switching dynamics. In order to address aspects of switching variability in the studied HfO_2/TiO_x devices, a statistical approach is presented in this section.

7.1 Switching variability of the SET process

As seen in Figure 4.10 and 4.12, the SET switching time of a single device varies between 3 and 5 orders of magnitude for a given voltage amplitude. Vice versa, for a given pulse duration, the required voltage can vary by about 300 mV, see Figure 4.10 (a). In fact, the procedure for recording the SET switching kinetics reflects this inhomogenous response to identical signals: Typically, the variation of pulse duration and pulse amplitude is performed through a nested loop. Within the loop, each pair of duration and amplitude is repeated multiple times to obtain a significant number of measureable switching times. As this is the most straight forward approach, it is followed by many publications in literature [7–9, 38, 96, 97, 100, 104, 105, 107–112, 160, 189–198]. In consequence to this procedure, a large number of switching attempts are discarded as they are either unsuccessful in switching the device or offer poor resolution in the timescale of the switching event. The first case occurs most frequently when the voltage is too low or the duration too short, while the latter case occurs more often for high voltages at long durations. In both situations, the variability of the switching event plays a big role. For the same voltage pulse to the same device in a comparable HRS, the current response can significantly differ. An example of this behavior is shown in Figure 7.1. Here, four example current transients

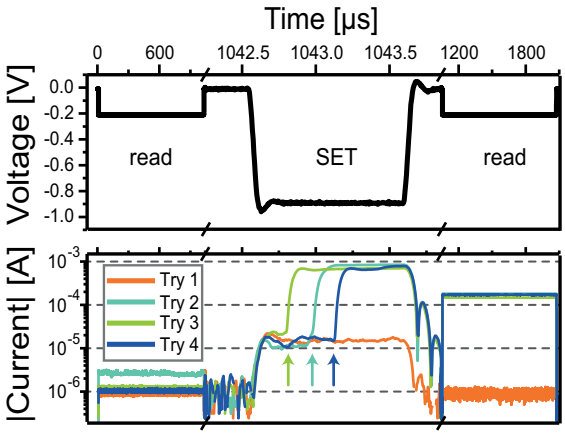


Figure 7.1: Upper panel: Typical SET voltage stress sequence. A read pulse for verification of the HRS is followed by the voltage stress pulse. The resulting resistance state is detected by another read pulse. Lower panel: Recorded possible current responses are shown in logarithmic scale. The SET transitions are indicated by arrows. Reproduced with permission from [116].

are shown. The shown SET attempts were recorded on the same device in direct succession, with only a RESET to the HRS before the next SET attempt. Before the SET pulse is applied, a read signal of -0.2 V is applied to the device, confirming the device being in HRS, which was in the range of 200 kΩ to 350 kΩ. The following SET voltage pulse with a duration of 1 µs and an amplitude of -0.88 V causes a variable current response of the device. In the case of Try 1, the current remains at a low level for the pulse duration, exhibiting no significant increase. The subsequent read signal confirms that the device has not undergone a significant change in resistance as the current is still low. In comparison, Tries 2, 3 and 4 show an abrupt increase in absolute current during the voltage stress. As typical for filamentary VCM type devices, this transition does not occur at a deterministic time but is highly variable ($t_{\text{delay, Try 4}} > t_{\text{delay, Try 2}} > t_{\text{delay, Try 3}}$). Furthermore, the current at the end of the pulse is not identical for the tries with an abrupt increase but also suffers from variability. The subsequent read signal current level reflects the current at the end of the SET pulse, where $|I_{\text{end, Try 2}}| > |I_{\text{end, Try 4}}| > |I_{\text{end, Try 3}}|$. However, the unsuccessful SET attempts which are typically discarded when studying the switching kinetics comprise another interesting feature, that is the switching probability. In Figure 7.2 (a), the switching kinetic data points of the previous section are drawn in colors with

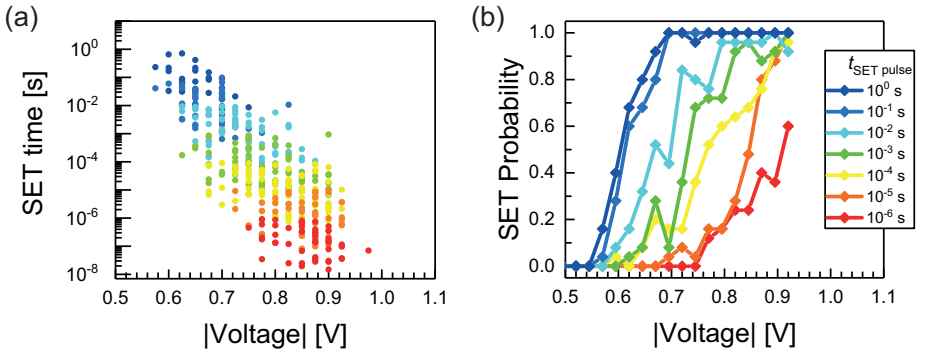


Figure 7.2: (a) SET kinetics measurement and (b) according SET probability traces as a function of the pulse amplitude.

respect to the pulse duration they were recorded with, see the according legend in Figure 7.2. Each pulse with a given voltage amplitude and pulse length was repeated 25 times, hence a SET probability can be calculated. In Section 4.5, Figure 4.12 comprises these values as pixel densities. In Figure 7.2 (b), the resulting values are plotted with respect to the pulse amplitude. Figure 7.3 shows the same values with respect to the pulse duration. Note that each probability is calculated from only 25 trials, making this dataset relatively susceptible to statistical error. Nevertheless, the trends of the switching kinetics are clearly reproduced. The demonstrated derivation from the switching kinetics into probabilities reflects the switching variability for a single device through multiple cycles. However, this cycle-to-cycle variability can differ from device to device, similar to the other properties such as sweep characteristic and resistance distribution. To confirm this assumption, the following experiment is conducted: A total of 15 devices are contacted individually and subjected to various voltage pulses using the measurement system described in Section 3.3.2. The voltage stress duration was chosen as 1 µs to be as close to industrial relevance as possible while maintaining sharp pulse geometry. The voltage range is chosen from -0.6 V to -1.1 V with -20 mV increments. In comparison to the SET kinetics measurement, the voltages are relatively high. However, at short timescales, some of the variable devices may require higher voltages than observed in the SET kinetics. Furthermore, the devices are less likely to degrade because of the short pulse duration. In the experiment, each SET attempt is repeated 50 times. This number is a compromise between measurement speed and statistical significance. Between the attempts, the

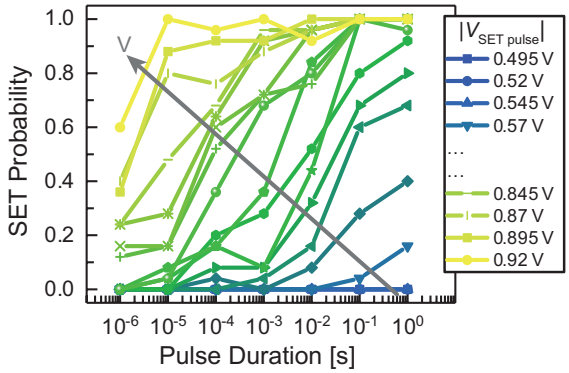


Figure 7.3: SET probability derived from the SET kinetics experiment as a function of pulse duration.

device was subjected to one and a half switching cycles of SET (-2.0 V; 100 µA compliance), RESET (1.3 V; no current compliance) and SET (-2.0 V; 100 µA compliance) by sweeping the voltage. Afterwards, the device is prepared for the SET attempt by precisely programming it into a resistance value between 200 kΩ and 350 kΩ through a voltage adjusting read-verify scheme. SET pulses that lead to a read resistance lower than 20 kΩ are counted as successful events. At this point, it is convenient to introduce the term "SET probability trace", which describes the SET probability data points with respect to voltage. In general, each individual trace is expected to begin at zero probability for low voltage. In a finite range of intermediate amplitudes, the trace should increase from 0 to 1 in a monotonous fashion as the SET event should occur more frequently the higher the voltage is. At high voltages, the trace is expected to stay at 1 since every SET attempt should be successful. However, deviations from this ideal behavior are expected because of the limited sample number that the probability is calculated from. Examplarily, this can be seen in the SET probability traces shown in Figure 7.2 (b) for the 100 ms trace, which has a very ideal shape, and the neighbouring trace of 10 ms pulse width, which has several kinks, likely owing to the low sampling number of SET attempts. In Figure 7.4 (a), the individual SET probability traces of each of the 15 devices is shown for pulse durations of 1 µs. As expected, the variation between devices is significant. While the SET probability trace of some devices begins at a voltage of 0.65 V, others start at 0.90 V or at even higher amplitude. The voltage range of intermediate probability is around 150 mV to 200 mV for all measured devices. This cycle-to-cycle voltage range is therefore smaller than the

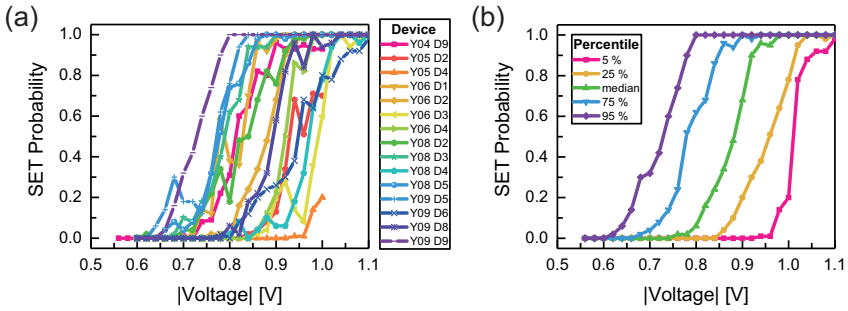


Figure 7.4: SET probability traces of 15 devices for 1 µs pulse length. (a) Individual devices. (b) Statistical analysis.

device-to-device spread, which can be estimated at around 300 to 350 mV. For better readability, the same data is shown in Figure 7.4 (b), where the 5, 25, 50, 75 and 95 percentiles at each tested voltage are shown. Since the tested number of devices was 15, the 5 % and 95 % lines are identical to the envelope of the ensemble. This dataset will be analyzed further in the next chapters as it provides the statistical input for the simulations. As described in the measurement procedure, the HRS before attempting the SET was programmed through a read-verify pulse scheme. The target resistance was 200 kΩ to 350 kΩ. In Figure 7.5 (a), the measured resistance just before the pulse is applied is shown as a device resolved stacked bar diagram. The double log scale is chosen for two reasons. On the x axis, it is chosen to accommodate for the lognormal resistance distribution. On the y axis, it highlights the uniformity of the distribution while correctly representing the small distribution tails. As indicated by the sharp distribution, the HRS remained roughly in the programmed window. Further, only small deviations between devices are noticeable. However, the programmed resistance is not perfectly matched with the defined resistance target window at the distribution edges. The tails of the distribution are slightly too low and too high. This can also be seen in the CDF plot in Figure 7.5 (b). This effect has been studied before [178, 199–202] and is understood as the effect of oxygen vacancy diffusion towards and away from the insulating gap. This random movement effectively smears out the programmed resistance distribution. Since the programmed state was reached just before the SET is attempted and each attempt is followed by two full switching cycles, the broadening should not impact the measurement results in this experiment. Figure 7.5 (c) highlights the high uniformity between devices when they are programmed

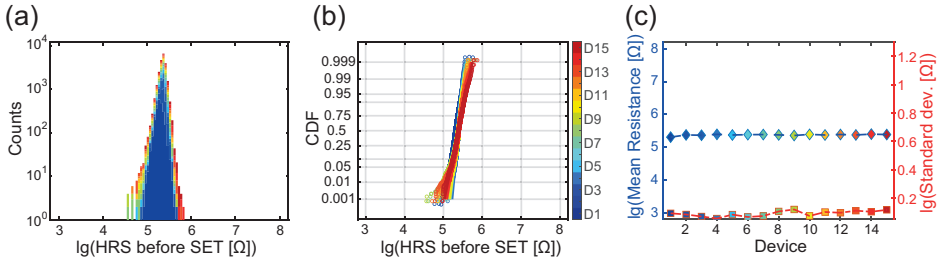


Figure 7.5: (a) Programmed HRS of 15 devices before 1 µs SET pulse is applied. (b) CDF analysis. (c) Mean and standard deviation of log-distributed resistances.

by a read-verify scheme. Here, diamond symbols demonstrate the mean of the decadal logarithmic resistance distribution, while squares denote the standard deviation of the same. In Figure 7.6 (a) and (b), the LRS values resulting from successful SET events are shown. Since the resistance in this case is not programmed by the previously employed read-verify scheme the distributions are significantly wider spread. Important to note is, however, that the various LRS resistances have been reached with a range of different voltages. The effects of exceedingly higher voltages on the LRS resistance have been shown in Section 4.5. Due to the series resistance, the LRS is further decreased with higher voltage. Nonetheless, the device-to-device spread is significantly more noticeable compared to the programmed HRS. While some devices reach just below the SET threshold resistance of 20 kΩ, others frequently reach below 2 kΩ. The difference between devices is further underlined in the CDF plot in Figure 7.6 (b). Black circles indicate the previous HRS of all devices, see Figure 7.5 (b). There is no correlation between the shape or position of the SET probability trace and resulting LRS distribution, i.e. the resulting LRS can be high or low independent from the onset of the switching trace. The phenomenon of SET probability of VCM type resistive switches consists of multiple parts. For rectangular voltage signals, it can be explained by the SET kinetics in large parts. By considering both successful and unsuccessful switching events at a given combination of voltage amplitude and pulse length, SET probabilities can be defined. The probability is a product of cycle-to-cycle variability and shows a typical voltage window of about 150 mV where the switching is not zero, but also non-deterministic. The SET probability of an ensemble of individual devices is significantly affected by d2d variability. The extremes of the rather small device ensemble already reached a difference of around 350 mV. The voltage range of the d2d variability is therefore larger than the c2c voltage range, i.e. the voltage difference of

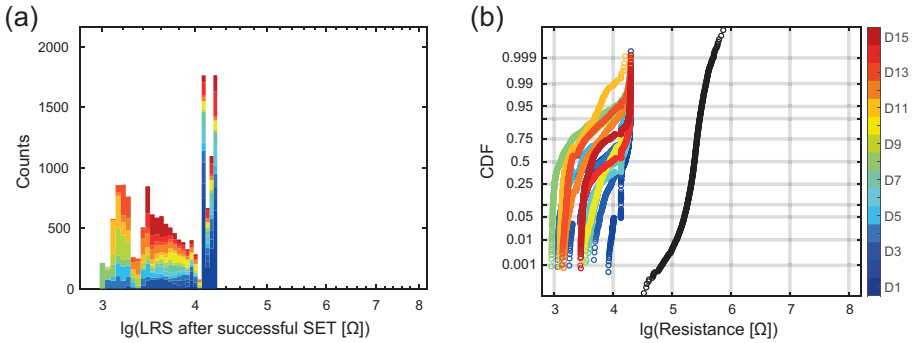


Figure 7.6: (a) Resulting LRS of 15 devices after an 1 µs SET pulse is applied. (b) CDF analysis.

two different devices is larger than the non-deterministic switching voltage window of a single device. The statistical analysis of this observation will be employed in the next section for an extension of the JART VCM v1 compact model and a comparison to the respective simulation. With respect to the resistance distribution, it is apparent that by employing a read-verify scheme, the HRS is precisely controllable apart from minor current drift events. However, physical boundaries such as oxygen diffusion impose a limit on the distribution sharpness, as was also discussed for the observation of noise in analog conductance states, see Section 5.2. In contrast to the HRS, LRS distributions after successful SET events proved to be strongly heterogeneous between devices, ranging over one order of magnitude in resistance. This property should be considered in an application where a blind SET pulse is applied to the device in contrast to a programming via current compliance control or read-verify scheme.

7.2 Variability of the RESET process

In Figure 4.13, the RESET process kinetics were studied. It was found that the RESET transition time is mostly independent of the initial conductance, but the delay time is strongly elongated for low LRS. Due to the small number of measurement points, the resulting HRS states were not analyzed further. In this section, the HRS levels resulting from a range of RESET voltages are studied. In Figure 7.7, the results of a similar experiment to the previous section for the SET process are shown.

The device was cycled one and a half times in total. Starting with a SET process (-2.0 V; 100 µA compliance), the device is switched to LRS. The following RESET sweep (1.3 V; no current compliance) results in an HRS of more than 100 kΩ. The HRS is confirmed by a read sweep. Afterwards, the same read-verify scheme as for the SET experiment is employed, however the target is a resistance between 3 kΩ and 5 kΩ. Once the target resistance is reached, a RESET pulse with a duration of 1 µs and variable amplitude is applied. In this experiment, the RESET voltage was varied between 0.5 V and 1.5 V with a step size of 100 mV. Each voltage was tested for 500 times to ensure sufficient resolution in the distribution tails. Since the LRS is not as susceptible to influence from noise and random ionic motion as the HRS or analog states, see Section 5.2 and [178, 202–204], the programmed resistance remains almost constant until the pulse is applied. As can be seen in Figure 7.7 (a) and (b), the resistance distribution measured after the RESET pulse does not increase significantly for voltages lower than 0.9 V. At higher amplitudes, the resistance after the RESET pulse increases significantly. The distribution for each voltage amplitude can be fit with a lognormal distribution, see the straight lines in Figure 7.7 (b). For increasing pulse amplitude, the mean resistance as well as the width of the distribution increases. The respective values from the lognormal fits, i.e. the mean of the decadal logarithm and the standard deviation of the same are shown in Figure 7.7 (c) as diamond and square symbols, respectively. The mean increases from the programmed LRS at [$\log(3\,\mathrm{k\Omega}) \approx 3.48$ ... $\log(5\,\mathrm{k\Omega}) \approx 3.70$] to around 100 kΩ at 1.0 V and 1 MΩ at 1.4 V. The standard deviation increases up to 1.0 V and stays close to 0.45 until 1.4 V. The highest resistances are observed at 1.5 V, after which the experiment is stopped to prevent damage to the device. However, the standard deviation at 1.5 V increases strongly, which can also be seen by the flat slope in the CDF plot. In contrast to the SET experiment, the measured resistance ensembles do not show splitting into two distributions as would be expected from a probabilistic process. This suggests that the RESET process is more controllable and exhibits less or no stochasticity compared to the SET process. From an application point of view, in many cases it is desireable to program a relatively high HRS with narrow distribution to ensure easy discrimination of from LRS and to avoid inaccurate or false read outs. From the presented data, voltages of 1.3 V and 1.4 V seem to result in high HRS values while ensuring comparably low variability. Therefore, the experiment described above is extended by additional 13 individual cells, which were subjected to the same routine described before for amplitudes of 1.3 V and 1.4 V. The results are shown in Figure 7.8.

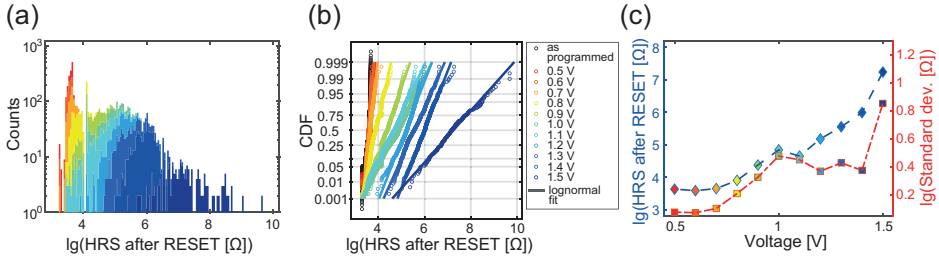


Figure 7.7: RESET voltage dependence for 1 µs pulses. (a) Resulting HRS in dependence of applied amplitude. (b) CDF analysis. Lines are lognormal fits. (c) Mean and standard deviation of the decadal logarithm of the resistance.

(a), (b) and (c) correspond to 1.3 V, while (d), (e) and (f) represent the results of 1.4 V pulse amplitude. Device 1 is the same data as in Figure 7.7. The same trend of mean resistance increase by higher voltage as for the single device is noticeable. At the same time, the standard deviation increases only slightly. Also important is the device-to-device variability for both the mean and the standard deviation. Some devices seem to have generally a higher HRS capability than others at the same voltage. Accordingly, some devices appear to be more variable than others. All devices show resistance distributions that are well described by a single lognormal fit, i.e. no splitting is observed. This proves once again that the RESET process is not stochastic in nature.

The observed difference between SET and RESET behavior can be understood with the physical model for filamentary switching. On the one hand, the initially low currents in HRS limit the potential to SET at a fixed time for a given voltage. Instead, the beginning of the thermal runaway phenomenon is stochastic and not predictable before the SET pulse is applied. The switching therefore appears probabilistic. The reason for this could be that many ion configurations result in the same resistance. However, one configuration may be more volatile than another when voltage is applied, resulting in the very different switching times at a given voltage. On the other hand, the current during RESET from the LRS is high during the pulse. If the LRS resistance is far enough away from the series resistance, the delay time is neglectable. Because the current in the LRS is high, the device is Joule heated and ionic motion is possible soon after the voltage is applied. The gradual nature of the RESET is due to the shifting equilibrium of drift and diffusion as described in detail by Marchewka et al.[112]. Hence, the observed distribution is uniform. The RESET process is therefore not stochastic in nature.

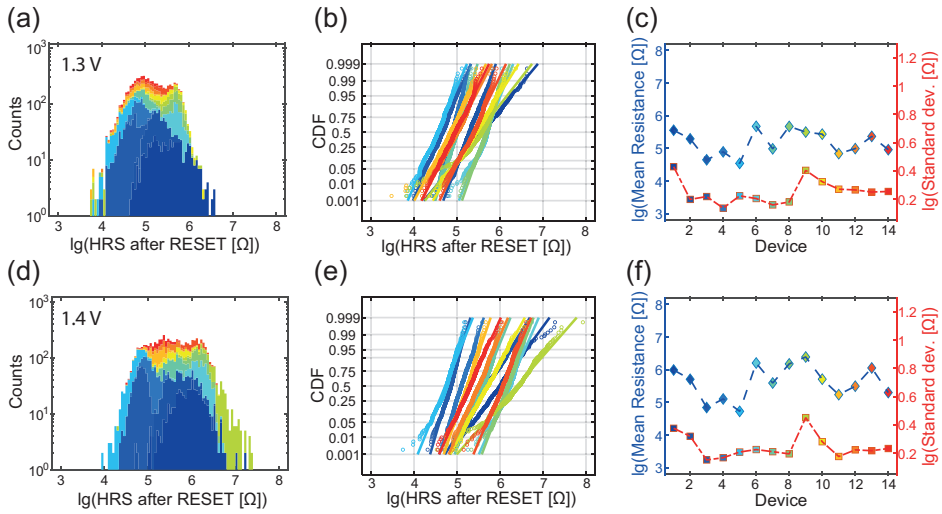


Figure 7.8: Device-to-device variability for 1 µs pulses of 1.3 V ((a), (b) and (c)) and 1.4 V ((d), (e) and (f)) amplitude. (a) and (d) Histograms of the resulting resistance. (b) and (e) CDF analysis. Lines are lognormal fits. (c) and (f) Mean and standard deviation of the decadal logarithm of the resistance.

8 Application of resistive switching features for neuromorphic hardware

Parts of this section are taken from [116]. Memristive devices are seen as promising candidate for replacing conventional CMOS and DRAM architectures where non-volatility, energy-efficiency and scalability is important. One such field of applications is neuromorphic computing. Memristive devices exhibit a multitude of unconventional properties that would hinder their use in classical computing architectures. Not all properties are found in every material and every material stack, but a few examples include time and temperature dependent resistance decay, resistance read noise and stimulated read noise, sub-loop switching features and existence of switching variability. In the emerging field of neuromorphic computing, which itself is unconventional compared to previous computing paradigms, such features may offer powerful enhancement strategies. In the following, an application pathway that exploits the SET switching stochasticity aspect of filamentary VCM devices is discussed. An in-depth experimental analysis of d2d and c2c variability of the binary mode is presented. By matching the experimentally observed variabilities to the JART VCM compact model, the concept of parallel operation of devices as a synapse is evaluated through experiments and simulation. The combined variability of these devices is exploited for a stochastic online learning network. It is demonstrated that stochastic switching features can be employed for a pattern classification task.

8.1 Background

Several approaches have been exploited for synapse emulation utilizing the stochastic nature of the resistance change in ReRAM devices. From the literature, two main pathways are identified, these are the single-device and the compound-device architectures. On the side of single device architectures, extensive studies exist on the variability phenomenon both for d2d and c2c aspects [54, 71, 73, 205–207]. The idea to incorporate multiple resistive devices into a synapse has been introduced earlier [208–214]. All variants followed the strategy to compensate for the short-comings of single devices by forming compound synapses. Singha et al. concluded that a synapse construction based on multiple parallel devices does not yield a significant advantage over single analog switches, but can be chosen as an alternative pathway when other tradeoffs emerge. However, while extracting the typical switching voltage stochasticity from c2c, their work doesn't consider the aspect of d2d variability, which should yield a significantly different result for a single, analog device synapse compared to the multi-device approach [211]. Boybat et al. employed parallel PCM devices as analog synapses for three different neural network tasks. They conducted a very detailed study of variability between individual devices and the switching stochasticity over multiple cycles. However, the main goal of their work was to stabilize analog conductance changes in the synapse's update under application of repeated current pulses [215]. Common for previous works published on the subject of stochastic switching of ReRAM devices in neural networks is the utilization of behavioral models. These lead to voltage-dependent switching probability models such as the Poisson distribution [71, 205, 207, 208, 213], sigmoidal distribution [73], Gaussian distribution [70] and lognormal distribution [210, 216] and even linear dependence [211]. By definition, these models only capture the minimal required behavioral aspects and possess little to no predictive character for any setup modification. However, with the aim to correlate the single ReRAM device behavior with the neuromorphic circuit behavior, it is imperative to use more detailed compact models that are able to capture the full dynamical spectrum of the employed device type.

8.2 Extension of variability in the JART model

The equivalent circuit diagram of the JART VCM compact model is shown in Figure 8.1 (a). The parameters for the deterministic model are listed in Table 8.1. In a

previous work [115], d2d and c2c variability were implemented in the JART VCM v1b model by pulling a random set of parameters from a truncated Gaussian distribution (seed parameters) for d2d variability and then changing these parameters, throughout the simulation around this seed parameter to produce c2c variability. The truncation of the Gaussian distribution determines the maximum deviation of the parameters around its mean value and was the same for the initialization as well as for the variation throughout the simulation. The variability parameters were chosen as the minimum and maximum oxygen vacancy concentration in the disc $N_{\text{disc, min}}$ and $N_{\text{disc, max}}$, as well as the radius of the switching filament r_{fil} and the length of the disc region l_{disc}. The choice of parameters was motivated based on the experimental findings of [217], where it was shown that the filament can form at different positions in the cell leading to variability in LRS and HRS.

Compact modeling aims to provide a tool for the design of circuits as well as bridging the gap between device-level technology and circuit design. Therefore, if a certain application is considered, the compact model must adequately model the device behavior under the conditions of the specific experiment. For this section, the relevant experiment is the measurement of the SET probabilities at different applied voltages starting from a specific range of HRS. Due to the large cycling variability

Table 8.1: Simulation parameters (for the explanation of the symbols, see[113], [115] and [116].

Symbol	Value	Symbol	Value
l_{cell}	3 nm	A^*	$6.01 \cdot 10^5\,\text{A}/(\text{m}^2\text{K}^2)$
l_{disc}	0.25 nm	$e\Phi_{\text{Bn0}}$	0.18 eV
r_{fil}	30 nm	$e\Phi_{\text{n}}$	0.1 eV
z_{VO}	2	μ_{n}	$4 \cdot 10^{-6}\,\text{m}^2/\text{Vs}$
a	0.25 nm	N_{plug}	$20 \cdot 10^{26}\,\text{m}^{-3}$
ν_0	$2 \cdot 10^{11}$ Hz	$N_{\text{disc,max}}$	$0.25 \cdot 10^{26}\,\text{m}^{-3}$
ΔW_{A}	1.6 eV	$N_{\text{disc,min}}$	$0.2 \cdot 10^{23}\,\text{m}^{-3}$
ε	$17\,\varepsilon_0$	$R_{\text{th,eff,SET}}$	$4 \cdot 10^7\,\text{K/W}$
$\varepsilon_{\Phi\text{B}}$	$5.5\,\varepsilon_0$	$R_{\text{th,eff,RESET}}$	$14 \cdot 10^6\,\text{K/W}$
T_0	293 K	R_{series}	$1300\,\Omega$

and the large spread between different devices observed in the experiment described in Section 7, it was decided to change the way variability is introduced into the model. Instead of the previous version, the ranges for d2d and c2c variability are split. This is schematically depicted in Figure 8.1 (b). In the new version, a cell

is initialized by drawing the seed variability parameters from a truncated Gaussian distribution. The c2c variability is achieved by changing these parameters during the simulation. However, the range in which these parameters can change during cycling is limited independently from the range of the d2d variability through a fixed percentage around the seed value. This enables tuning d2d and c2c variability independently to fit the model parameters to the measurements. The new implementation, therefore, uses three different parameters to modify the different variability of the model. The relative standard variation which determines the width of the truncated Gaussian distribution is used to initialize the devices. This quantity influences mostly the d2d variability since it determines whether the parameters will be initialized closer or further away from the median value on average. The c2c variability is controlled by two parameters, namely the c2c percentage and the maximum step size. The c2c percentage determines the range around the drawn set of seed parameters for each device in which the parameters can change through repeated switching. On the other hand, the maximum step size determines the maximum amount by which the variability parameters may change between two successive switching cycles. The influence of these different parameters can be seen in Figure 8.1 (c) through (f). Figures 8.1 (c), (d) and (e) show the effects of different amounts of c2c variability on the SET probabilities while (f), (g) and (h) show the effects of different amounts of d2d variability. It can be observed that increasing the c2c variability makes the behavior of single devices more stochastic. This implies that increasing voltages do not always result in an increase of SET probability but might also decrease it. Increasing the d2d variability spreads the SET probability curves across a larger voltage range. The parameters related to the variability are given in Table 8.2. They are kept constant in this work showing the high degree of consistency between model and experiment.

Table 8.2: Simulation parameters (for the explanation of the symbols, see the work of Bengel et al.[115].

Symbol	Min /Median / Max	Symbol	Value
$N_{\text{min, var}}$ [10^{23} m^{-3}]	0.1 / 0.2 / 0.3	relative standard deviation	1
$N_{\text{max, var}}$ [10^{26} m^{-3}]	0.05 / 0.25 / 20	c2c percentage	15 %
r_{var} [nm]	25 / 30 / 35	maximum stepsize	10 %
l_{var} [nm]	0.175 / 0.25 / 0.35	-	-

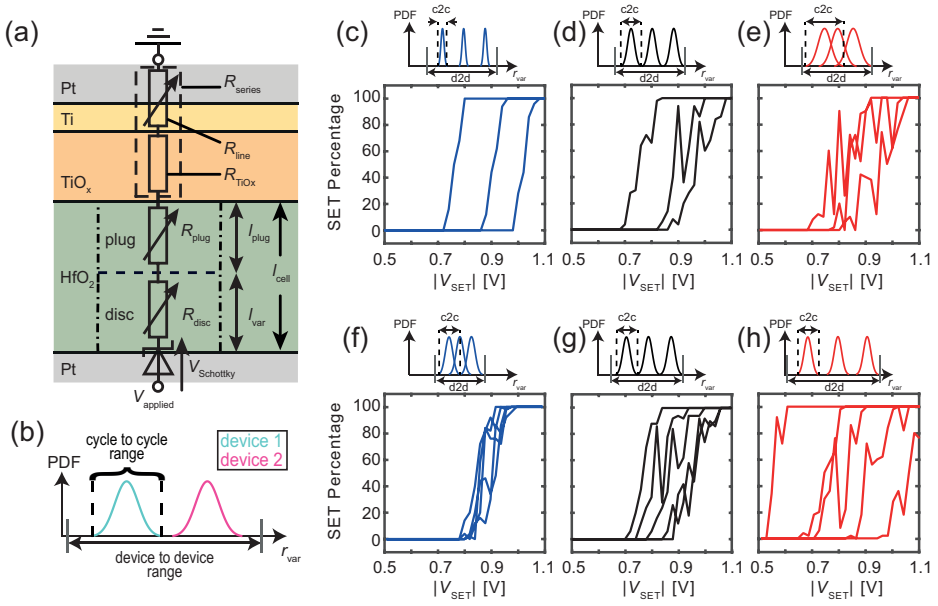


Figure 8.1: (a) Equivalent circuit diagram of the JART VCM v1b model for the HfO_2/TiO_x based memristive devices. (b) Schematics of the modification made to the variability model exemplary for the variability model parameter r_{var}. For each device, a seed parameter is drawn from the d2d range. Around this seed parameter, the cell's variability parameters can change by the smaller c2c range. (c) to (e) Effect of different amounts of the c2c variability on the SET probability behavior for three cells. For (c), the c2c percentage was 5 %, for (d) it was 15 %, and for (e) it was 25 %. (f) to (h) Effect of different amounts of d2d variability. This was achieved by decreasing the d2d range and the variation coefficient for (f). For (g) the values from Table 8.2 were chosen and for (h) the d2d range was increased. The changes are conceptually shown by the small PDF diagrams given on top of diagram (c) to (h). Redrawn with permission from [116].

8.3 Experimental results and compact model simulations for single devices

As described in Chapter 7, 15 individually contacted devices were tested experimentally for their SET probability traces. In this specific context, a SET probability trace is the probability-voltage relation that shows the required voltage range for a device to traverse from zero percent switching probability to 100 % switching probability. By definition, the SET is counted as successful if the resistance is lower than 20 kΩafter the SET attempt. If the resistance is still above 20 kΩ, the trial is counted as unsuccessful. The SET probability is then given as the fraction of successful events over the total number of trials. In Chapter 7, voltage stresses of 1 µs duration were employed. The voltage range from -0.6 V to -1.1 V chosen with increments of -20 mV ensures that the entire trace is recorded with a sufficient resolution in voltage, where non-deterministic switching occurs, i.e. where the SET probability lies between 0 % and 100 %. At each voltage step, 50 trials are performed. This amount is a compromise between measurement speed and statistical significance. Hence, each given probability value has to be seen with an inaccuracy of at least 2 %. Each attempt is accompanied by a forced SET, a RESET and another SET using a sweep signal. The resistance state prior to voltage stressing is then accurately programmed to be in the range of 200 kΩand 350 kΩbefore the SET trial, which corresponds to read currents between -0.57 µA and -1 µA. The HRS read currents immediately before pulse application are shown as the red histograms in Figure 8.2. The distribution lies well within the defined limits described above. Minor deviations at the lower and upper boundary are noticeable. This behavior, which was described in Section 5.2, has been studied before [178, 218] and can be explained by ionic noise that is typically present in filamentary VCM devices. Figure 8.3 depicts the measured SET traces from Chapter 7 in grey lines and symbols and the simulated traces of the described experiment in different colors. For better readability of the comparison, the gathered device traces are analyzed statistically in the following manner: The median trace value, as well as 5 %, 25 %, 75 % and 90 % percentiles at every tested voltage, are given. In this context, the term “edge cases” refers to the SET probability traces at the lower and upper voltage extreme. For the experimental dataset, the 5 % and 95 % lines essentially reflect the edge cases because of the limited device count. Therefore, the relative uncertainty in these percentiles is quite large. For the simulation dataset, the actual edge cases may

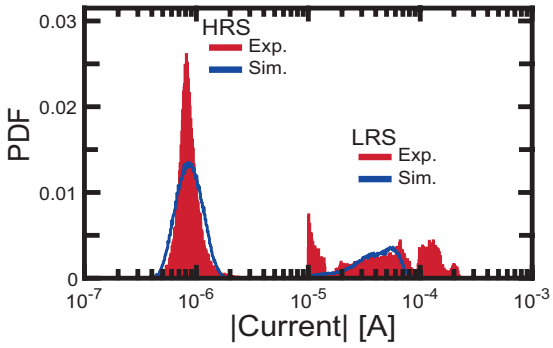


Figure 8.2: Comparison of the read currents for HRS and LRS for simulation (blue lines) and experiment (red bars) by their respective probability density functions (PDF). Redrawn with permission from [116].

be located at slightly lower or higher voltages than the 5 % and 95 % lines, respectively. Here, the relative uncertainty is reduced because of the higher device number. While the agreement between simulation and experiment of the median and the 25 % and 75 % cases is nearly flawless, the compact model exhibits a slight mismatch for the 5 % and 95 % lines. However, this deviation is minor. The percentile lines are, however, an important characteristic for comparison to the simulation dataset. Overall, the experimental data shows the expected behavior of combined c2c and d2d variability. Following the median trace highlights the c2c aspect. It traverses from zero SET probability at low voltages to deterministic switching, i.e. a SET probability of 100 %, at high voltages. The regime of non-deterministic switching has a width of around 160 mV. For the voltages in this range, the c2c variability leads to a mixture of successful and unsuccessful events. The percentile marks are indicative of the d2d spread of the devices. While the median trace of the experimental datasets shows the beginning of the non-deterministic regime at 0.80 V and the end at 0.96 V, the 25 % trace is shifted to lower voltages of 0.70 V and 0.86 V, respectively. The opposite trend is observed for the 75 % trace which shows a range between 0.86 V and 1.04 V. The range between these two traces, the interquartile range, is therefore almost perfectly constant at 160 mV for all voltages. The same observation can be made with the 5 % and 95 % traces. Here, a range of 280 mV is covered. These numbers are important measures for comparison to the simulation, but also comparison to other devices and device types. In cases where the SET event was successful, i.e. the resistance was

below 20 kΩ, the resistance is noted. The red histogram at the higher current level in Figure 8.2 summarizes the read currents of the measured low resistance states. A significant spread from 10 µA up to 300 µA is visible. This spread further signifies the presence of variability in the devices. Three important points must be mentioned in this context: First, the displayed histogram is gathered from 15 different devices. Closer analysis reveals, that each device itself has a less significant spread. Hence, the d2d aspect of this measurement has to be taken into account. On top of that is the second point: The shown data summarizes both the d2d and the c2c variability of the devices. The third point is that in this measurement, low resistive states that result from SET pulses with varying amplitude are shown. The results of Section 4.5 have demonstrated the impact of stronger voltages on the resistance state after the SET process. Therefore, voltage stresses that are barely enough to switch the device will lead to higher resistances than voltage stresses that switch the device with high certainty. Because of these three aspects, the spread of the low resistive state is not unexpected.

The simulation for this experiment is carried out with the model parameters described above. A total of 250 device seed parameters are drawn in the described way. Each drawn device is tested in the same way as described above for the experimental devices. Hence, the distributions of the HRS prior to the voltage stress and the LRS in cases where switching took place can be compared. Clearly visible in Figure 8.2 is the nearly perfect agreement of the HRS distributions before the SET attempt. Comparing LRS values is a bit more complicated. As visible from the comparison of distributions in Figure 8.2, the simulation lacks parts of the distribution both at the lower current end and the higher current end. This difference is caused by two different phenomena. The lower current end difference is caused by the imperfect description of the switching transition time by the model. As discussed in Section 4.5, the SET transition can be abruptly interrupted if the switching pulse is ended, but the device has not fully undergone the SET process. In the present model, the transition speed is higher than the experimental one, thus causing mainly full switching events. Therefore, the lower current end is not simulated as often as in the experiment. The higher current end difference is due to the experimentally observed effect, that a stronger voltage leads to a higher read current, even after the switching event is completed. The employed simulation model cannot fully describe this relation, since in these simulations the defined maximum oxygen vacancy concentration is reached when the SET event has taken place. In total, the switching model still describes the

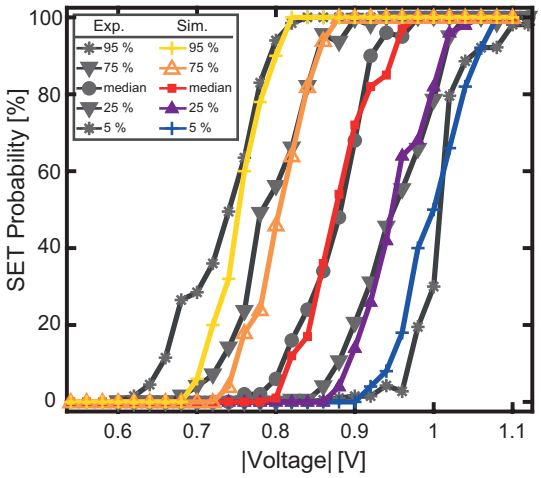


Figure 8.3: Statistics of the d2d and c2c SET switching variability for 1 µs voltage stresses for experiment and simulation. Redrawn with permission from [116].

experimental data very accurately. The missing effects may be addressed in a future work. As stated above, a much higher number of devices is simulated compared to the experimentally measured dataset. For the evaluation, the same approach as for the experimental dataset is chosen, and the median at each voltage as well as the 5 %, 25 %, 75 % and 95 % are calculated. The results are compared to the experimental results in Figure 8.3. The good agreement for the median SET probability trace is visible. Moreover, all percentile traces are also well met, with only very minor deviations at the 5 % and 95 % traces. In summary, the simulation of SET events on the timescale of 1 µs reveals a high degree of agreement between measurement and simulation. All statistical values, including current levels and switching voltage, are well met. Slight deviations are caused by the inherent variability of both measurement and simulation on one hand and by minor effects not accounted for in the model on the other hand.

To verify our simulation model beyond the correct c2c and d2d description on the 1 µs time scale, the predictive capability for a different experiment is tested. A crucial prerequisite for a model is to correctly display the switching dynamics over multiple orders of magnitude in switching time [43, 97, 100, 115]. For this purpose, a single device was experimentally tested and the results were compared to the com-

(a)

Parameter	$N_{min,\,var}$ [10^{23} m^{-3}]	$N_{init,\,var}$ [10^{23} m^{-3}]	$N_{max,\,var}$ [10^{23} m^{-3}]	r_{var} [nm]	l_{var} [nm]
Seed Value	0.25	7.10	0.40	34.8	0.28

(b)

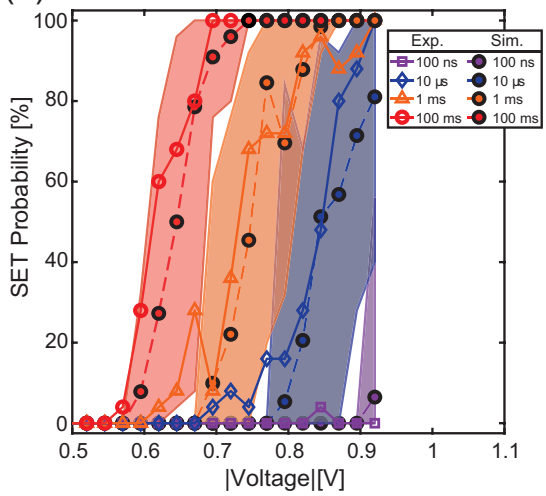


Figure 8.4: (a) Seed parameters of the simulated device. (b) SET probability as a function of the pulse amplitude for different pulse widths. An experimental and a simulation device, which are both close to the median on the 1 µs timescale, are compared. Both voltage window and voltage onset in dependence of pulse duration are correctly described, verifying the predictive character of the compact model. The envelope areas quantify the maximum possible voltage shifts resulting from insufficient repetitions at a given voltage. Redrawn with permission from [116].

pact model simulation. The tested device, which was previously shown in Section 7 in Figure 7.2, follows the median trace of the 1 µs experiment closely and is therefore considered as a good example. The device is switched by voltage pulses that vary over multiple orders of magnitude in pulse duration and voltage stress. The device preparation procedure for SET probability testing is identical to before. However, in this experiment, the device is stressed with a voltage pulse of variable duration (100 ms down to 100 ns) and amplitude (-400 mV to -925 mV in -25 mV decrements). The test procedure is repeated 25 times at each combination of pulse width and pulse amplitude. Figure 8.4 (b) displays the experimental SET probabilities at each combination as a solid line. Noticeably, the onset of the SET probability curve shifts to lower voltages when increasing the pulse duration. This reflects the typical SET

switching kinetics. However, the voltage range, where the probability is neither 0 % nor 100 %, remains roughly constant at around 160 mV, with some deviations caused by the limited number of tries. To reproduce this behavior over multiple orders of magnitude in switching time, a simulation identical to the described experiment is conducted by choosing a single device seed parameter that follows the simulated median trace of the previously described 1 µs experiment. The seed parameters are listed in Figure 8.4 (a). Note that the c2c percentage and the maximum stepsize are 15 % and 10 % as indicated in Table 8.2. However, for this median-like simulation device, the pulse duration dependent SET probability traces are recorded not only once, but 50 times. Figure 8.4 (b) contains the results of this simulation. Here, the dashed lines represent the SET probability at each combination of pulse duration and voltage stress, calculated from the 1250 attempts at this combination. At the same time, the colored areas outline the range that is covered by each subset of 25 tries per combination. This method was chosen to highlight the fact, that a SET probability trace for the same device can show a shift of the voltages when recorded twice. The true SET probability trace is revealed by repeating the same pulse conditions a significant amount of times. In our experiment, 25 attempts already yielded stable results, but with each additional trial, the accuracy of the SET probability trace can be increased. In Figure 8.4, only every second pulse duration is shown for better readability. However, the not shown pulse durations follow the same trend of onset voltage and show the same width of the traces. By comparing the experimental results to the simulation, it is directly visible that the median simulated curves closely follow the experimental ones apart from some minor deviations likely caused by the limited number of trials in the experiment. The key characteristic, namely the voltage range of non-deterministic switching, is met reasonably well. The good agreement between measurement and simulation validates the JART VCM v1b switching model for this kind of experiment. The predictive character, namely that the compact model can describe the switching dynamics precisely on the 1 µs scale as well as for multiple orders of magnitude in pulse time, has been demonstrated. Therefore, the model is employed in the following sections for creating a multi-device compound synapse and a pattern classification network.

8.4 Theoretical and experimental considerations for a multi-device synapse

In this section, the implementation of the proposed parallel synapse model from a theoretical standpoint is described and the operation is experimentally demonstrated. Several devices are connected in parallel and are biased with identical voltage stresses of 1 µs duration. For the proposed probabilistic update in the following SNN application, it is desired to apply a voltage pulse of a given amplitude and get a corresponding probabilistic bitline current response as the outcome. Therefore, different current response levels have to be accessible with some probability. By increasing (reducing) the voltage stress, the probability of getting a higher (lower) current response should increase. In the following, the effect of ReRAM device variability on the collective synapse behavior will be discussed by using our compact model to simulate different exemplary cases.

8.4.1 Theoretical considerations

For the proposed synapse implementation, it is required for the synapse to be able to adopt intermediate current levels between the two extreme cases of all devices in HRS and all devices in LRS. If all devices were to show identical, deterministic behavior, no intermediate current levels can be achieved, and the accumulated synapse current will either be low or high for low and high voltage signals, respectively. This undesirable or even worst-case scenario is illustrated in Figure 8.5 (a) and (b), where (a) shows the SET probability traces of 3 devices with identical seed parameters and no c2c variability, while (b) shows the probability for normalized bitline current levels at each voltage. As can be expected the synapse will only show two achievable current levels since the LRS and HRS values are the same for all three devices which switch in a deterministic fashion at a specific voltage. The addition of d2d variability leads to a significant change in the synapse behavior. Figure 8.5 (c) and (d) show three devices without c2c variability but with significant d2d variability. Different voltage onsets yield (n+1) separable bitline current levels, with n being the number of devices per synapse. Each of the levels has exactly 100 % probability in a distinct voltage interval as the devices still show deterministic switching at a voltage specific to each device. The simulation was performed by modifying the seed parameters to achieve a device switching at a low, a moderately high and a high voltage. The device switching at a

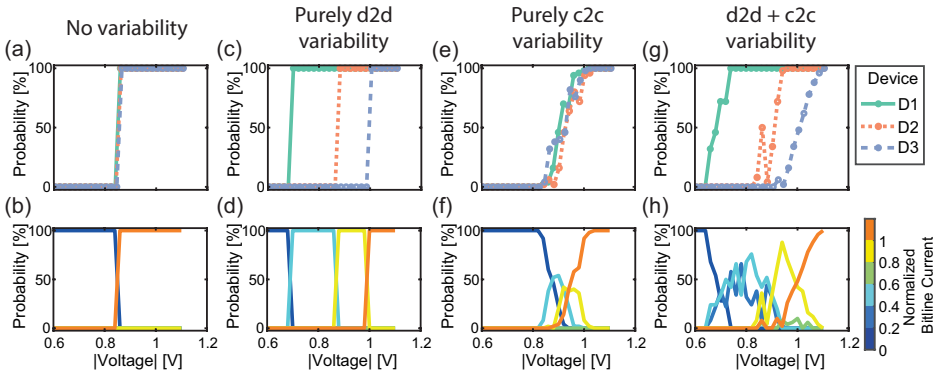


Figure 8.5: Simulation of different synapse behavior to showcase the relation between SET probability traces of individual devices (upper row) and probabilities for the achieved bitline currents (lower row) by adapting the compact model. (a) and (b) represent three identical deterministically behaving devices. (c) and (d) show three different but deterministic devices. (e) and (f) show three identical but probabilistically behaving devices. (g) and (h) show three different and probabilistic devices. Redrawn with permission from [116].

low (high) voltage is realized by choosing a small (large) filament radius (r_{var}), a small (large) disc length (l_{var}) and a high (low) initial oxygen vacancy concentration in the disc ($N_{\mathrm{disc,\,init}}$). Adding c2c variability to situation (a) in Figure 8.5 instead of d2d variability leads to the behavior observed in Figure 8.5 (e) and (f). The devices shown here have the same parameter seed. Throughout the simulation, their parameters were varied to achieve c2c variability. At each voltage 50 tries were performed for each device. It should be noted that this limited number of repetitions at each voltage is the reason for the three different SET probability traces. Increasing the number of repetitions will make the traces comparable and even identical for an infinite number of repetitions. Since only a limited number of cycles is simulated, the SET probability traces are different. In Figure 8.5 (f), the resulting bitline current range probabilities form a voltage window. In this voltage window, the intermediate bitline current ranges can be addressed with a single voltage pulse. However, the voltage window has a width of only around 200 mV although the total range of applied voltages is 600 mV. The previous cases are, however, purely theoretical cases. Filamentary switching VCM type devices exhibit significant d2d and c2c variability as a consequence of the underlying physical mechanism as well as tolerances in device fabrication. This makes it virtually impossible to eliminate variability. By combining d2d and c2c

variability, we arrive at the most realistic case shown in Figure 8.5 (g) and (h). Compared to the previous case, the voltage window is significantly wider (around 400 mV), which is caused by the early onset voltage of device D1 and the late onset of D3. In this specific case, it would be sufficient to employ voltage levels with a spacing of around 100 mV to address the intermediate bitline current ranges. Another feature of real devices that can be observed here is that the number of accessible levels is larger than (n+1). This can be attributed to the variability in the LRS state of the various devices as shown in Figure 8.2. It should be noted here that similar to Figure 8.5 (e) and (f) resulting bitline current probabilities in Figure 8.5 (h) are not an unambiguous consequence of the SET probability traces in Figure 8.5 (g) due to the limited number of tries at each voltage level. Each trial of all devices at a specific voltage can be viewed as a discrete stochastic process. In summary, to realize an analog synapse with binary switching devices, the interplay between d2d and c2c variability is very important. Deterministic devices limit the number of levels to two while adding d2d leads to an increase of the number of levels in a deterministic fashion (the levels can be programmed with 100 % probability). Adding c2c variability leads to probabilistic devices and a probabilistic update. The voltage window to achieve this, however, might be limited. Differing probabilistic devices widen this window, but are not necessarily required for the proper synapse functionality if their individual c2c variability covers a sufficiently wide voltage window. For an increasing number of devices per synapse, the range of conductances will shift towards higher values. However, if we scale this range by the number of devices in each synapse, we can see that the resulting fraction stays constant. Thus, the bounds to the synaptic efficacy stay constant except for a shift towards higher conductance levels, which might lead to a higher power consumption. On the other side, the analog tunability improves as more intermediate levels become accessible. Lastly, this improvement in the synapse behavior will increase the area that each synapse occupies. As d2d and c2c cannot be eliminated in real devices the observations need to be tested by measurements.

8.4.2 Experimental demonstration of the multi-device synapse

To verify the proposed explanation of the synapse behavior in dependence of the included variability, two exemplary cases of experimentally realized synapses are shown. For this, the voltage stress is applied to three devices simultaneously. In order to do so, a probe card arrangement as described in the experimental section 3 is employed.

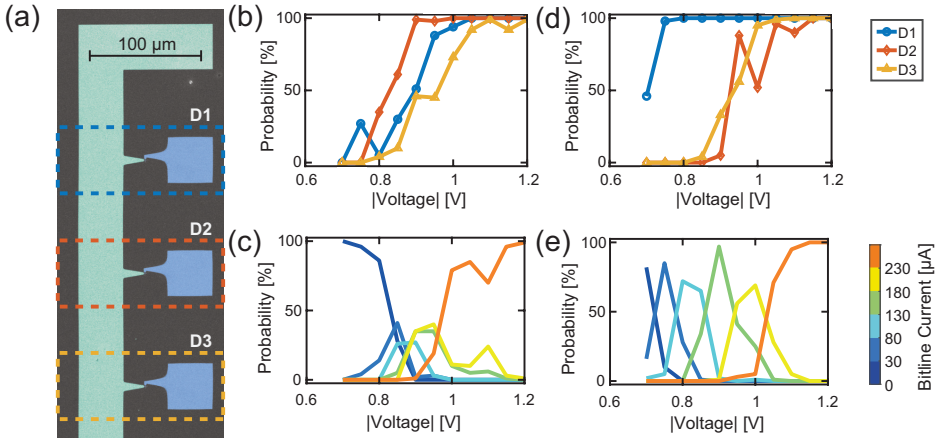


Figure 8.6: Experimental demonstration of favorable synapse characteristics. (a) SEM picture of a section of the device structure contacted in a probe card arrangement. Blue: Top electrodes. Green: Bitline electrode (color added in postprocessing). (b) SET probability traces of three fairly similar devices. (c) Resulting bitline current probabilities, showing the narrow voltage range for tunability. (e) SET probability traces of three significantly different devices. (f) Resulting separated bitline current ranges. Redrawn with permission from [116].

From the 32 x 1 line array, a subset of three devices is contacted to see the direct dependence of the device SET probability traces on the synapse characteristic. The structure is shown in the scanning electron microscopy (SEM) picture in Figure 8.6 (a). The addressed cases reflect the cases discussed in Figure 8.5 (g) and (h). In this experiment, the three devices in parallel were repeatedly stressed with voltages from -0.7 V to -1.2 V in steps of -50 mV. Each voltage was tested 100 times. The initialization was carried out for each device individually and followed the same procedure as described for Figure 8.3 (g). After each voltage stress, the summed bitline current is recorded for a read voltage of -0.2 V. On top of that, each device is read individually with a read voltage of -0.2 V. First, three devices with very similar SET probability traces are contacted, see Figure 8.6 (b). The resulting synapse behavior is depicted in Figure 8.6 (c). As expected, the bitline current ranges are only addressable in a very limited voltage range of about 0.2 V. This means, that intermediate synapse currents require very precise voltage stresses to the devices. This combination of devices reflects the case of identical probabilistic devices as described in Figure 8.5 (e) and (f). In contrast, the second subset of devices contacted shows a significant voltage

margin between the individual SET probability traces, see Figure 8.6 (d). At 0.70 V, the highest probability is observed for the lowest current range of 0 µA to 30 µA, since only infrequently switching events occur and the devices remain in the HRS. A lower probability is evident for the second current range of 30 µA to 80 µA. By increasing the voltage to 0.75 V, the probability of reaching this second level is increased. However, the first level or the third level may occur with a low probability. This trend continues in a very regular pattern at 100 mV intervals until currents of 230 µA or more are observed. Here, the currents do not increase further with voltage because the synapse's dynamic range is reached. Therefore, at 1.20 V, the probability for reaching the last level, i.e. currents above 230 µA, is 100 %. The dynamic voltage window of this synapse, therefore, lies in the voltage range from 0.70 V to 1.05 V, which results in a voltage window of 0.35 V. This behavior is favorable over the case of nearly identical devices, since a higher voltage spacing for the levels can be utilized, which in turn reduces the challenges for integration. In summary, the proposed synapse structure fulfills the imposed requirements. The synapse's tunability window is mainly influenced by the switching probability behavior of the individual devices it is composed of. Two possible ways for enlarging this window can be derived: First, the presence of d2d variability can improve the synapse behavior. However, only having d2d variability can still lead to unwanted synapse behavior if nearly identical devices happen to appear in a given synapse. A better approach is to introduce more c2c variability in each device while reducing the d2d variability to a minimum. By this method, the voltage window remains large enough for several voltage levels with moderate spacing. At the same time, the minimized d2d variability ensures conformality of the synapse characteristic.

8.4.3 Performance criteria for multi-device synapses

For three different synapse sizes, namely three, eight and twelve devices per synapse, 50 synapses are simulated. At each voltage stress, 50 SET tries are conducted. For comparing the arising effects when scaling up the synapses, three new parameters are introduced as indicated by the sketch in Figure 8.7 (a):

1. The difference between the first and the last SET probability trace with respect to the pulse voltage. For this, the area between the two extreme traces is calculated by the trapezoidal numerical integration of the difference of the first and last trace, see the shaded area in Figure 8.7 (a).

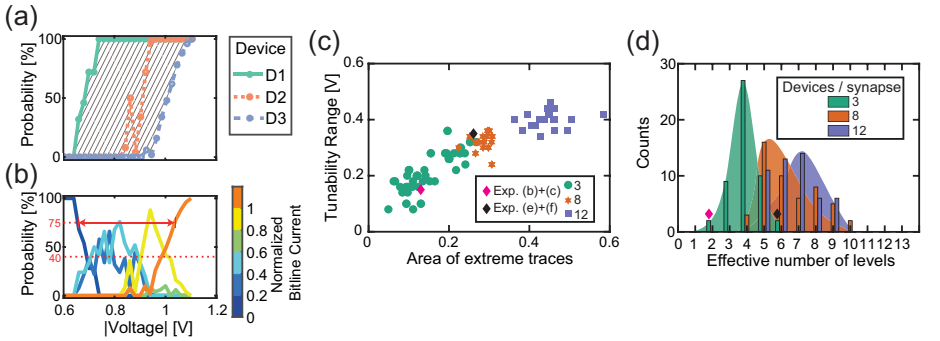


Figure 8.7: (a) and (b) Sketch of the introduced analysis parameters based on the individual SET probability traces and the bitline current probabilities. The grey shaded area in (a) shows the first parameter, while the solid red line in (b) shows the second and the dashed red line in (b) shows the third. (c) Dependence between tunability range and extreme trace area for the simulations comprising 50 instances of three, eight and twelve devices each. (d) Statistics of the effective number of addressable intermediate bitline current ranges for each synapse size. Redrawn with permission from [116].

2. The synapse tunability voltage range. It is calculated by the difference in voltage of the first current range at a probability of 75 % and the voltage of the last current range at a probability of 75 %. This value describes the range of voltages in which the intermediate current ranges (between 'all devices HRS' and 'all devices LRS') can be stochastically addressed. It should be adequately large for the chosen voltage levels of the application. The value is marked with an arrow in Figure 8.7 (b).

3. The effective number of realistically addressable levels. For this, the number of levels that reach a probability of 40 % or more over the whole voltage range is counted. It is highlighted by the horizontal dashed line in Figure 8.7 (b).

Figure 8.7 (c) displays the relation between the SET probability difference and the synapse tunability voltage range for all synapse sizes and each individually initialized synapse. The simulations with three devices per synapse are indicated by green circles. The achievable voltage window for a synapse constructed from three devices ranges from quite low values to the desired larger ranges. For three devices per synapse, an increasing relation of the tunability range with the area enveloped by the highest and lowest SET trace is obtained. The exact underlying relation is however masked

by the significant variability, which stems from the combination of variable devices. The simulation results are controlled by the experimental data. For this purpose, the desired data points are determined from measurements shown in Figure 8.6 (b) and (c) and in Figure 8.6 (d) and (e), respectively, and are added to the graph in Figure 8.7 (c) as diamond-shaped symbols. The data points from the experiment lie within the range of the simulated multi-device synapses for the case of three devices. This match clearly demonstrates the accuracy of the developed compact model. A further enlarged voltage tunability range of a multi-device synapse can be achieved by increasing the number of devices per synapse. Figure 8.7 (c) also displays the simulation results for eight and twelve devices per synapse. For the purpose of comparing these synapses, the bitline current ranges were normalized with the assumption that each device contributes 38 µA (corresponds to 5.2 kΩ) to the overall bitline current. Furthermore, instead of splitting the resulting currents into six levels as for the three devices per synapse, the currents are grouped in eleven and 15 levels for eight and twelve devices per synapse, respectively. For the eight and twelve devices per synapse structures, the tunability window is above 0.2 V, and levels at around 0.4 V. The synapse comprising twelve devices shows an even higher SET probability difference, but no significant increase in the tunability window is evident. Most important for synapses comprising eight and twelve devices is the absence of synapses with an undesirable low voltage tunability window. This can be attributed to the low chance of drawing eight or twelve nearly identical devices for the synapse, respectively. By increasing the synapse size, it becomes increasingly likely to draw devices from the full d2d range, hence making sure that sufficient variability is present in the synapse. For the three devices, there is a chance of getting three highly similar devices with their limited c2c voltage range, thus causing low voltage window synapses as e.g. shown experimentally in Figure 8.6 (b) and (c). Figure 8.7 (d) shows the statistical analysis for the 50 simulated synapses regarding the effective number of addressable intermediate bitline current levels for each synapse. As expected, the number of levels with a probability of 40 % or more increases with synapse size, allowing for more accurate tuning of the larger synapses. The experimentally determined data points are plotted as diamond symbols. Again, the data points derived for the extreme cases lie at the edges of the simulated range. It is therefore expected that increasing synapse size will yield higher network performances, especially if overlapping patterns are shown since such require improved synapse tunability.

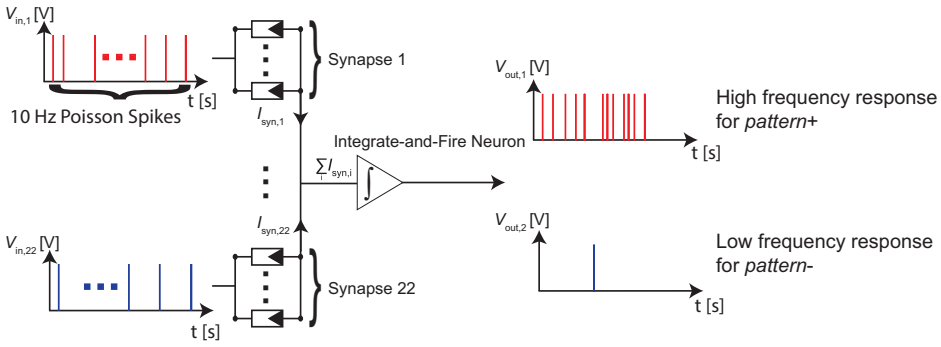


Figure 8.8: The neural network consists of 22 synapses connected to an integrate-and-fire type output neuron. The synapses consist of multiple VCM ReRAM devices connected in parallel. The studied range is between 1 and 24 cells per synapse. It is trained to react to a *pattern+* (which is presented to 11 of its synapses) with a high number of spikes and to a *pattern-* (which is also presented to 11 of its synapses) with a low number of spikes. Redrawn with permission from [116].

8.5 Spiking Neural Network setup

To highlight the power of exploiting device stochasticity, the device model is employed in a binary classification problem with overlapping features. To systematically assess the classification accuracy, the complexity of the problem is raised through increasing the overlap, i.e. the mutual information between the patterns. The parameters of the variability model are drawn to initialize the required number of devices, which depends on the number of ReRAM per synapse. The devices are usually initialized with a resistive state that roughly lies between the LRS and HRS.

8.5.1 Network setup and general learning procedure

The investigated neural network consists of 22 synapses connected to an integrate-and-fire type output neuron. The synapses each consist of one or multiple ReRAM cells connected in parallel. The studied range is between 1 and 24 cells per synapse. The network structure is shown in Figure 8.8. The patterns are synthesized to have control over the complexity and to study the network accuracy as a function of the problem complexity. To generate the patterns half of the synapses are picked randomly and stimulated with a Poisson distributed spike train. The rest of the synapses are not stimulated. This is called *pattern+*. To generate *pattern-*, an overlapping pa-

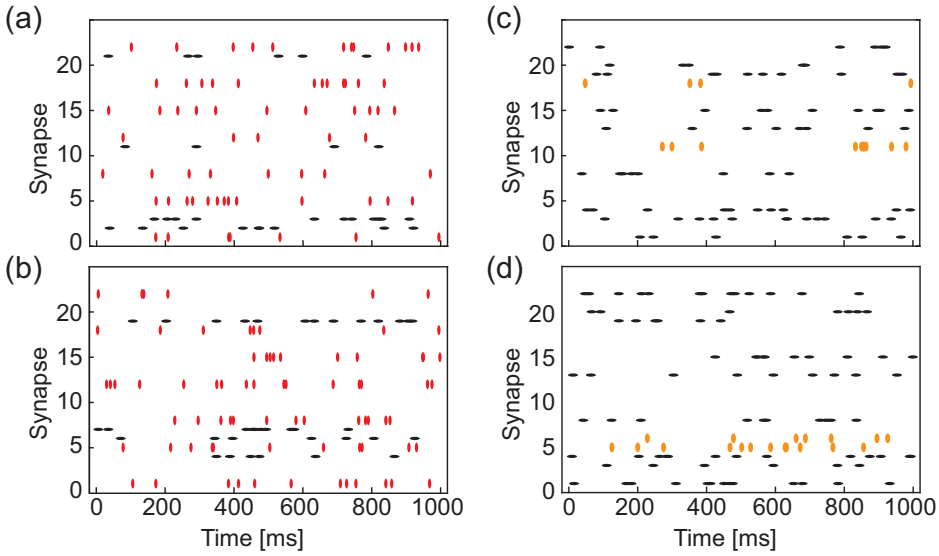


Figure 8.9: Exemplary raster plots showing the Poisson distributed spike trains which are applied to the different synapses for 1 s. (a) and (b) show an exemplary *pattern+* and *pattern-*, respectively, with an overlap of seven between the patterns. The overlapping synapses (1, 5, 8, 12, 15, 18, 22) are marked as red vertical symbols. (c) and (d) show two noisy *pattern+* examples. Based on the ideal pattern (not shown) two synapses were flipped (11 and 18 in (c) and 5 and 6 in (d)). The flipped synapses are marked as orange vertical symbols. Redrawn with permission from [116].

rameter M is defined, which is the number of common features or amount of mutual information. M number of synapses from *pattern+* are then selected randomly as the common feature of *pattern-*. The rest of the features of *pattern-* is chosen randomly from the remaining synapses that are not included in *pattern+*. The synapses making up *pattern-* are also stimulated via a Poisson distributed spike train. As an additional level of complexity, noise is introduced in the patterns by randomly flipping a specified number of features in *pattern+* and *pattern-*. Seven noisy test patterns are generated for *pattern+* and seven for *pattern-*. These training patterns will be used during the inference phase for the evaluation of the network's performance. A few exemplary patterns are shown as raster plots in Figure 8.9. Figure 8.9 (a) and (b) show one exemplary *pattern+* and *pattern-*, respectively, with an overlap of seven. The seven overlapping patterns are marked as red vertical symbols, while the red horizontal lines represent the non-overlapping patterns. Figure 8.9 (c) and (d) show two *pattern+* in

which two synapses have been flipped each. The flipped patterns are marked as orange vertical symbols, while the non-flipped patterns are marked as black horizontal symbols. To train the network, a technologically plausible training algorithm is utilized, namely a stochastic Delta rule algorithm, which is the simplest form of gradient descent for single-layer networks [37]. The Delta rule can be formulated as

$$\Delta w_i = \lambda \cdot (\hat{y} - y) \cdot x_i, \tag{8.1}$$

where Δw_i denotes the amount the weights that have to change in the network, λ is the learning rate which can be used to scale the amount of weight change per update, $\hat{y}$ is the target, y is the neuron output activity and x_i selects the synapses to which the current pattern is applied. For the binary classification problem, the target for *pattern+* and *pattern-* are the maximum and minimum firing rate of the neuron $FRMAX$ and 0, respectively. Therefore, the update rule becomes

$$\Delta w_i = \lambda \cdot \frac{FRMAX \cdot label - FR}{FRMAX} \cdot x_i, \tag{8.2}$$

where *label* is 1 when *pattern+* is applied and 0 if *pattern-* is applied. FR is the firing rate of the neuron in response to the applied pattern. This training rule is formalized in Algorithm 1. To assess the untrained accuracy all training patterns are presented to the network. At this stage, the accuracy of the network is 50 % in most cases, which is the accuracy of guessing randomly. This first step is done to show that the network is trained during the next steps and starts from a bad accuracy. The simulations are performed in the following fashion. The synapses are simulated using Cadence Spectre, and the current $\sum I_{syn}$ (see Figure 8.8) accumulated at the output node is saved. This output current is then fed into an integrate-and-fire neuron model realized in MATLAB which samples the current at 1 ms intervals. The current is summed up until the neuron threshold I_{TH} is reached and the integrated current is reset to zero. Each instant of time at which the neuron threshold is reached is counted as a spike of the neuron. The total number of spikes produced for 1 s is then compared with the decision threshold $FRMAX/2$. If the number of spikes is larger the pattern is interpreted as *pattern+* and if it is smaller it is interpreted as *pattern-*. After the initial evaluation, the network is repeatedly trained and tested for 10 epochs. During the training, a noisy training pattern is applied to the specified synapses and the number of spikes FR is counted. This number is then compared with the target number for the current pattern which is either $FRMAX$ for *pattern+* or 0

Algorithm 1 Delta Rule implementation with stochastic synapses

$w_\mathrm{i} = rand()$;
while Test Accuracy $< 100\%$ **or** # epochs < 11 **do**
    apply labeled train pattern
    calculate FR
    $\Delta w_\mathrm{i} = \lambda * \frac{FRMAX * label - FR}{FRMAX} * x_\mathrm{i}$
    if label=0 **then**
        if epoch=1 **then**
            $V_{\mathrm{train,\ i}} = V_{\mathrm{RESET,\ nominal}} * x_\mathrm{i}$
        else
            $V_{\mathrm{train,\ i}} = \mathrm{round}(\ V_{\mathrm{RESET,\ nominal}} * |\Delta w_\mathrm{i}|)$
        end if
    else
        if epoch=1 **then**
            $V_{\mathrm{train,\ i}} = V_{\mathrm{SET,\ nominal}}\ x_\mathrm{i}$
        else
            $V_{\mathrm{train,\ i}} = \mathrm{round}(\ V_{\mathrm{SET,\ nominal}}\ |\Delta w_\mathrm{i}|)$
        end if
    end if
    apply $V_{\mathrm{train,\ i}}$
end while

for *pattern-*. The difference between the real and wanted number of spikes is scaled by *FRMAX*, multiplied with the learning rate λ which is set to a value of 1.2 in our case, and multiplied with a vector that corresponds to the assignment of synapses of the current training pattern. This relationship is described by Equation 8.2. It represents the desired weight change for all the synapses that just received a pattern. In this way, the following programming pulse will be weaker if the neuron's response is close to the ideal response (0 or *FRMAX*) so as not to disturb the achieved weights too much, and stronger if the neuron's response to the current pattern is far away from the ideal response. After the calculation of the distance between the ideal and the actual neuron response, the programming pulses are applied to the synapses that also received the training signals. The nominal SET (-0.8 V) and RESET (1.3 V) voltages are scaled by Δw_i and rounded to the closest 100 mV increment. This scaling of the voltages modulates the switching probability of the ReRAM cells in the synapse. As the probability of switching is increased if the network's error is higher and decreased if it is lower, this represents a technologically plausible stop learning mechanism. Additionally, this can be seen as a form of randomized or stochastic rounding [219] similar to the one implemented in [213]. In previous works, stochastic rounding has

been found to improve neural networks, enabling to reduce the bit size of the weights [220] or enabling to reduce the input bit size [221] while keeping the accuracy constant [221]. The scaled programming voltages are applied for 1 µs, leading to a SET/RESET of the ReRAM cells of the respective synapses. This training procedure is performed for three *pattern+* and three *pattern-* in a random succession for each epoch. After these six training rounds, the network's performance is tested on seven test patterns for *pattern+* and *pattern-*. This procedure is repeated until the accuracy on the test pattern set reaches 100 % or until 10 epochs are reached.

8.5.2 Hardware aware network optimization

The network also contains hyperparameters. Hyperparameters are these parameters used to control the learning process. Unlike the weights, the hyperparameters are not derived through training but have to be specified prior to the training. For the utilized network, the hyperparameters are defined by the learning rate λ, the neuron threshold current I_{TH}, the maximum achievable spike rate $FRMAX$, which also determines the spike decision threshold, and the SET and RESET voltages for the synapses. While these parameters have a significant impact on the network performance, they are typically not trained but rather preset to a constant value. Therefore, their choice has to be motivated in a different way, which usually involves trial and error. In this work, however, a plausible procedure to tune two of the hyperparameters, namely I_{TH} and $FRMAX$ is described, before the training of the neural network starts. This alleviates the problem of finding optimal hyperparameters through guessing, since two parameters less have to be optimized manually. The devices exhibit significant d2d and c2c variability as discussed in the previous sections. This makes finding global parameters for the neuron threshold I_{TH} and the decision threshold ($FRMAX/2$), which ensure good network convergence, challenging. In any case, a global combination of I_{TH} and $FRMAX$ for all neurons would be a compromise and would degrade the performance by having this global constraint. A novel approach is to mitigate this issue at a local scale by self-adapting the values at the individual neuron level. Before training the network, the neurons are first strongly excited by turning all synapses on through use of a stronger SET pulse (-1.3 V for 1 µs) to achieve a global SET probability of 100 %. Then, the output current that is achieved for several different exemplary *pattern+* training signals, applied to 11 random synapses, is sampled. All synapses are excited because of the external noise, i.e. flipped bits, which are introduced during the actual

network operation. By varying I_{TH}, the maximum possible number of spikes that can be generated by this output neuron can be sampled. In a second step, all synapses are turned off by applying a RESET pulse (+1.3 V for 1 µs) and the output current is sampled for several different exemplary *pattern-* training patterns. By varying I_{TH} in this state, the minimum possible number of spikes that can be generated by this output neuron can be determined. By averaging over the responses, the value of I_{TH} that results in the largest difference between the neurons response to *pattern+* and *pattern-* is found. Using this I_{TH}, the decision threshold $FRMAX/2$ is defined as the difference between the weakest response of the neuron to the different *pattern+* signals and the strongest response of the neuron to the different *pattern-* signals. It is found that the optimum I_{TH} is increased with the number of ReRAM cells per synapse as a higher current will pass through the parallel devices. The decision threshold $FRMAX/2$ is found more or less independent of the number of devices per synapse. The values are distributed between 20 Hz and 30 Hz. The hyperparameter tuning algorithm is formalized in Algorithm 2. In summary, the proposed algorithm maximizes the distance

Algorithm 2 Hyperparameter tuning algorithm

$w_{\mathrm{i}} = rand()$;
apply V_{SET} = (-1.3 V ‖ 1 µs)
for j = index test patterns **do**
    $I_{+,\,\mathrm{j}} = \sum_{n=1}^{22} I_{\mathrm{syn,\,n,\,j}}$
end for
for k = index I_{TH} **do**
    $FR_{+,\,\mathrm{k}}(I_{\mathrm{TH,\,k}})$ = mean(number of spikes, j)
end for
apply V_{RESET}=(1.3 V ‖ 1 µs)
for j = index test patterns **do**
    $I_{-,\,\mathrm{j}}$= $\sum_{n=1}^{22} I_{\mathrm{syn,\,n,\,j}}$
end for
for k=index I_{TH} **do**
    $FR_{-,\,\mathrm{k}}(I_{\mathrm{TH,\,k}})$ = mean(number of spikes, j)
end for
$I_{\mathrm{TH,\,k}} = \max(FR_{+,\,\mathrm{k}} - FR_{-,\,\mathrm{k}})$

between the neurons response for *pattern+* and *pattern-*. An additional advantage is that it enables adaptation to failed devices and even would enable retraining of the network. This might be useful if after a certain time some of the devices start to fail. In that case, the described procedure can be repeated and adapted hyperparameters can be found that consider the failed devices. Once these hyperparameters have been

algorithmically optimized, the training begins.

8.6 Spiking Neural Network results

First, the achievable accuracy of the neural network under increasing overlap between *pattern+* and *pattern-* and how the accuracy is improved by increasing the number of devices per synapse is investigated. The studied ranges for the overlap were $M = 4$ to 10 while the studied range of devices per synapse was 1, 4, 8, 12 and 24 devices. Figure 8.10 (a) shows the simulation results of the neural network's accuracy as a function of the overlap between the patterns and the number of devices per synapse. If the accuracy reached 100 % after a certain training epoch, the training was stopped and this accuracy is taken as the final value of this run. Otherwise, the accuracy after 10 epochs was used. Figure 8.10 (b) to (f) show the evolution of the accuracy over the training epochs for all 10 runs (thin grey lines) as well as the mean curve (thick red line) for an overlap M between the patterns of nine. Figure 8.10 (b) shows the results if each of the 22 synapses consists of only one ReRAM device, and (c) to (f) for 4, 8, 12 and 24 devices per synapse, respectively. From this figure, multiple effects can be observed. Due to the different sources of variability that already exist in the initialization phase of the network (d2d, Poisson inputs, etc.), ten runs are performed for each combination of overlap between the patterns and number of devices per synapse. From Figure 8.10 (a) it can be observed that the overlap between the two patterns influences the network's performance. Generally it is observed that the network reliably reaches an accuracy of 100 % for overlaps smaller than 5, independent of the number of devices per synapse. For larger overlaps, the average accuracy is degraded. However, this effect is stronger for networks where only a small number of devices are used. This shows that increasing the number of devices per synapse is a way to improve the performance of the neural network if the classification problem becomes more difficult. This can also be observed in Figure 8.10 (b) to (f). While networks with only one or four devices per synapse struggle to reach an accuracy of 100 % during training, perfect accuracy can be achieved after only one training epoch for networks with more devices. While some runs also achieve high accuracies after a few training rounds for the small networks, other runs struggle as their accuracy is stuck at a low value or oscillates over the epochs. A closer look at the training is depicted in Figure 8.11. Here, the conductances of the different categories of synapses for the runs in 8.10 (b), (d) and (f) as well as the number of spikes that were gener-

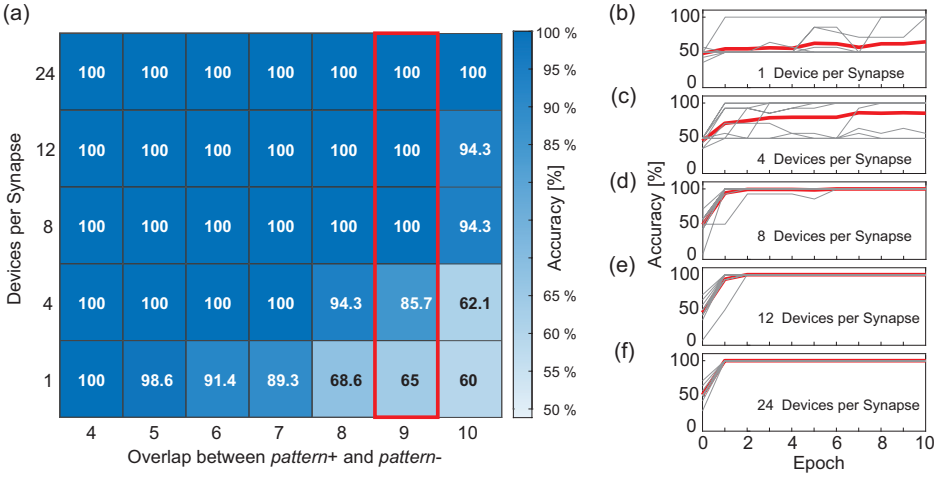


Figure 8.10: (a) Simulation results of the neural network's accuracy as a function of the overlap between the patterns and the number of devices per synapse. (b) to (f) show the accuracy over the training epochs for all ten runs (thin grey lines) as well as the mean curve (thick red line) for an overlap between the patterns M of nine. (b) shows the results if each of the 22 synapses consists of only one ReRAM device, (c) to (f) accrodingly for 4, 8, 12 and 24 devices per synapse. Redrawn with permission from [116].

ated for positive and negative patterns are shown. Figure 8.11 (a), (b) and (c) show the conductances normalized to the number of one, eight and 24 devices per synapse, respectively, of the synapses receiving *pattern+* (orange line and diamonds), *pattern-* (blue line and circles) and both patterns (black line and triangles) over the training epochs. Again, the overlap between the patterns was nine. The solid and dashed lines show the mean values while the different symbols show the values of the actual synapses. Figure 8.11 (e), (d) and (f) show the corresponding number of spikes FR that are generated if *pattern+* (orange) or *pattern-* (blue) is presented to the neuron. The lines again show the mean values while the symbols show the number of spikes generated for the unique patterns. Similarly to Figure 8.10, it is observed that a higher number of devices improves the network's performance. While Figure 8.11 (a) and (d) (one device per synapse) show that the training is not complete after ten epochs, the training already finishes after the sixth epoch for (b) and (e) (eight devices per synapse) or after the first training epoch for (c) and (f) (24 devices per synapse). Looking closer at Figure 8.11 (a) shows the reason why the training is not successful.

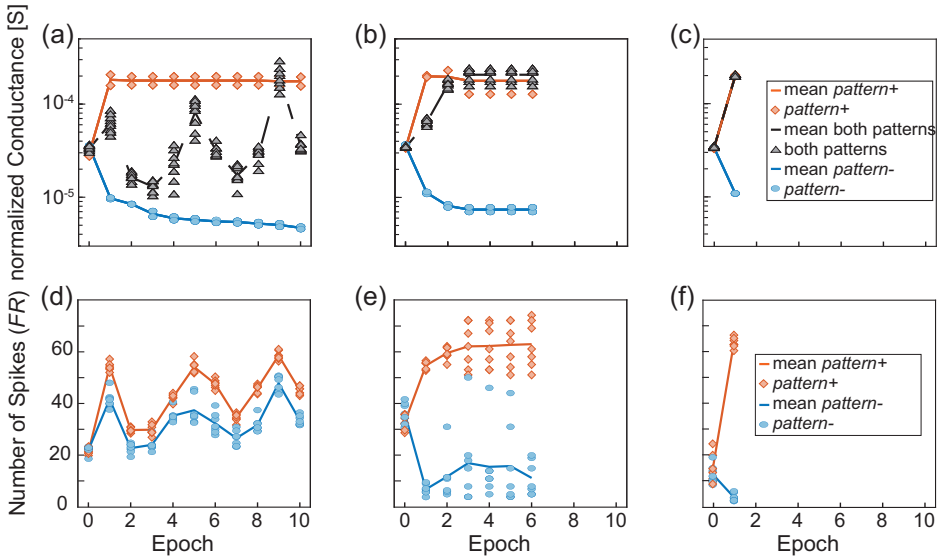


Figure 8.11: (a), (b) and (c) show the conductances normalized to the number of devices per synapse of the synapses receiving *pattern+* (orange line and diamonds), *pattern-* (blue line and circles) and both patterns (black line and triangles) over the training epochs. The overlap is nine in all cases and the number of devices per synapse was one in (a) and (d), eight in (b) and (e) and 24 in (c) and (f). The solid and dashed lines show the mean values while the different symbols show the values of the actual synapses. (d), (e) and (f) show the corresponding numbers of spikes *FR* that are generated if *pattern+* (orange) or *pattern-* (blue) is presented to the neuron. The lines again show the mean values while the symbols show the number of spikes generated for the unique patterns. Redrawn with permission from [116].

While the synapses receiving only *pattern+* (orange) or only *pattern-* (blue) are programmed to distinct values that stay constant throughout the training, the synapses receiving both patterns (black) are not programmed to a stable conductance level and change throughout the training, oscillating between the conductance boundaries. Since the overlap was nine in this example, the group of synapses receiving exclusively *pattern+* or *pattern-* each only consists of two elements while the black group consists of nine elements. The consequences of this can be seen in Figure 8.11 (d) which shows the number of spikes generated in this case for *pattern+* in orange color and *pattern-* in blue color. The distance between the neuron's responses to *pattern+* and *pattern-* is small and both are subject to abrupt changes. This prevents a converging of the delta rule algorithm as the training voltages are not scaled down. A contrast to this

can be seen in Figure 8.11 (b) and (e) which are achieved for an increased number of devices per synapse of eight. In this case, the synapses receiving both patterns quickly reach a stable conductance range. As can be expected the neuron responds to this increase (decrease) in current with a significantly higher (lower) number of spikes for *pattern+* (*pattern-*). Lastly, Figure 8.11 (c) and (f) show an even improved picture. The training is already finished after the first training epoch as all synapses achieve a stable conductance value. It can be observed in Figure 8.11 (a), (b) and (c) that the neural network only reaches 100 % accuracy when the synapses receiving both patterns (black) are completely excited. The explanation for this finding is that fully excited and fully depressed synapses are representative of more stable device states in the sense that they require higher voltages to be switched. It has been shown that a higher HRS requires higher SET voltages to set the device and that a smaller LRS requires higher RESET voltages to reset it. Therefore, if a synapse is found in a fully excited or depressed state, it will require higher absolute voltages to switch it to the opposite state as if the synapse was only partially excited or depressed.

In summary, the presented results show that increasing the number of devices per synapse greatly increases the performance of the network as it allows for a more gradual tuning of the weights and helps with reaching the stop learning condition, i.e. 100 % accuracy. Finally the network's performance when the inputs are noisy is investigated. The number of flipped bits represents an external noise source. For the case of N flipped bits, N random synapse assignments are changed for every training and test pattern. This makes the classification problem significantly more difficult since it causes the wrong synapses being trained. Here, a range of zero flipped bits up to two flipped bits is tested. For even higher numbers of flipped bits, the accuracy is heavily degraded and no network architecture is able to reliably achieve accuracies much higher than random guessing. An overview of the results for one and two flipped bits can be seen in Figure 8.12 (a) and (b). As expected, the accuracy worsens when some of the input bits are flipped in each training and test run. Another important feature to be observed in Figure 8.12 (c), (d) and (e) is that the unique runs showcased by the orange diamonds and the blue circles show a significantly larger spread if the number of flips is increased, resembling the significant rise in pattern to pattern variability. The different test patterns are fixed before the training starts and they are not changed over the epochs. While some patterns produce more easily distinguishable spike numbers (close to zero for *pattern-* or about 60 for *pattern+*), other patterns provide not such a clear spike response. The number of patterns

producing an unclear response increases with the number of flipped input bits. For zero flips all patterns provide a clear spike response, for one flip there is one pattern that falls out of line and for two flips most of the patterns provide an unclear spike response. This degradation can also be seen in the median spike response for *pattern+*, which is close to 60 for zero flips, around 50 for one flip and only at 40 for two flips. The dependence of the accuracy on the number of devices per synapse and the overlap between the patterns is however more complicated than before. One trend which can be observed is that for smaller overlaps the smaller networks usually perform better than their larger counterparts. The proposed explanation for this is that when the network becomes larger it stops training the weights after the first few epochs as the error-adjusted SET and RESET voltages become too small to significantly adjust the weights. Figure 8.13 (a) and (c) show this exemplarily for the case of two flips, an overlap of 4 and the networks containing one (a) or 24 (c) devices per synapse. The normalized conductances of the synapses in Figure 8.13 (a) are much more shallow than their counterparts in Figure 8.13 (c), which enables the network to reach a better final result. The synapses in Figure 8.13 (c) show very little change after around the second epoch which means that the network has stopped training at this point. As seen in Figure 8.11 larger networks generally lead to a faster convergence, as they are able to find weight values for high accuracies quicker. In the presence of flipped bits in the inputs, this behavior will still hold. However, as some of the synapses are trained in the opposite direction, this initial stable solution does not yield 100 % accuracy. Their smaller counterparts take longer to find a stable solution as the weights are easier disturbed. This gives the smaller synapses a certain robustness against the incidence of flipped bits. An increase in overlap stresses the point that larger networks perform generally better for one flip, see Figure 8.12 (a). For two flips the same statement is true, see Figure 8.12 (b). Multiple effects determine the final accuracy of a network with a given size and overlap that can be achieved. On the one hand, smaller networks have an advantage if the overlap between the patterns is small or medium as their less stable weights can be tuned even if the error becomes smaller, see Figure 8.13 (a) and (c). However, larger networks perform better for one flipped input and larger overlaps, which can be explained in the same way as for the zero flip cases. The comparison of the smallest and largest considered network for an overlap of nine and two flipped input bits in Figure 8.13 (b) and (d) shows why larger networks can find better solutions for high overlaps than smaller networks. While the unique synapses receiving both patterns (grey triangles) for the small network are mostly

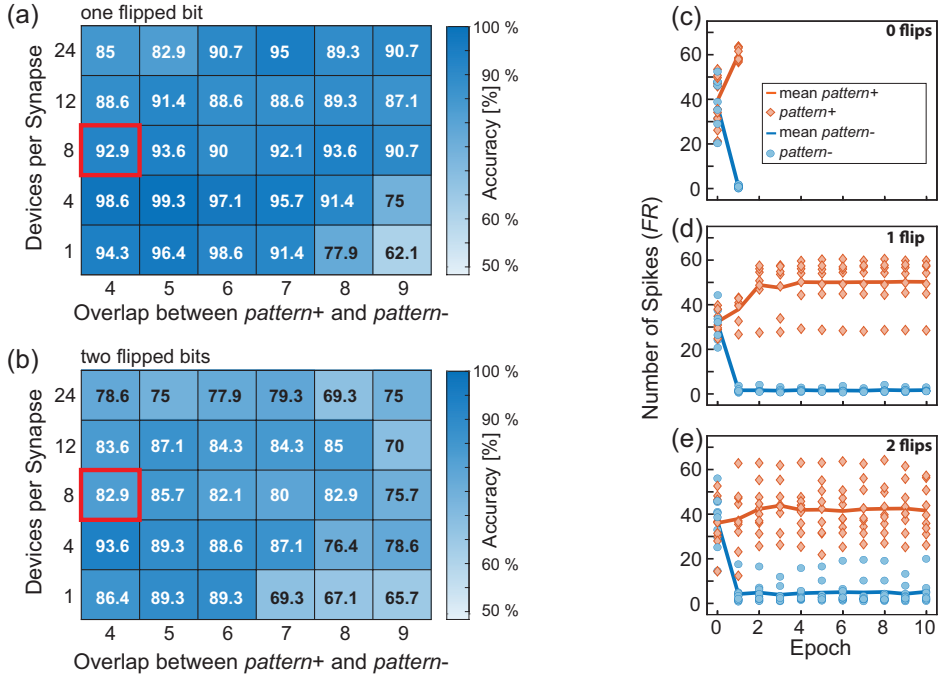


Figure 8.12: (a) and (b) show the accuracies achieved for (a) one or (b) two flipped bits, averaged over ten runs and evaluated after the same criterion that was used in Figure 8.10. (c), (d) and (e) show the number of spikes generated by *pattern+* (orange) and *pattern-* (blue). The solid lines show the median while the diamonds or circles show the responses to the unique *pattern+* or *pattern-*, respectively. (c) shows the results for zero flips, (d) for one flip and (e) for two flips. Redrawn with permission from [116].

programmed to less stable medium conductance states as depicted in Figure 8.13 (b), the larger network can program them to a more stable high conducting state as shown in Figure 8.13 (d).

Overall, the accuracy is reduced if the inputs are noisy. As can be expected, this effect is stronger if more inputs are noisy. As the drop in accuracy is closely related to the utilized training rule, it seems reasonable to investigate how to increase noise resilience. As seen from the device analysis, this noise resilience will have to take into account the specialties of ReRAM programming. One idea towards this might be to use a non-linear scaling of the voltage in the delta rule algorithm. If - for the sake of argument - it is assumed that the first spike response was at a maximum distance from

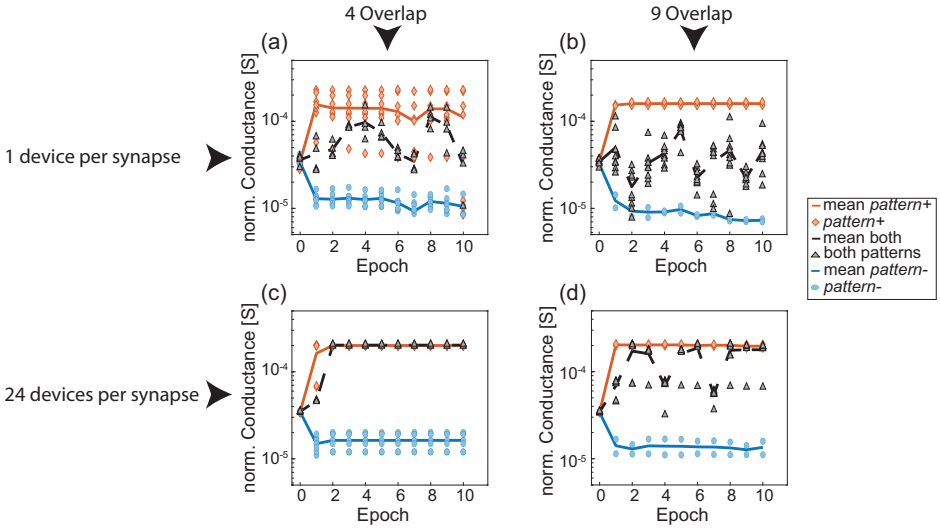


Figure 8.13: (a) to (d) show the conductances normalized to the number of devices per synapse of the synapses receiving *pattern+* (orange line and diamonds), *pattern-* (blue line and circles) and both patterns (black line and grey triangles) over the training epochs. The solid and dashed lines show the mean values while the different symbols show the values of the actual synapses. Redrawn with permission from [116].

the goal spike response, this would result in Δw_i of 1 as described in Equation 8.2, giving the nominal programming SET voltage or RESET voltage in the training step. If the next spike response to an applied pattern was now at half the distance from the goal spike response, the resulting Δw_i would be 0.5 corresponding to half the nominal SET or RESET voltage to be applied afterwards. This dividing of the training voltage in half is however problematic when applied to the tested devices. As can be seen in Figure 8.3, the range of voltages that lead to a 50 % SET probability is around 300 mV, considering also the edge cases. This range is not significantly changed if one goes to other SET probabilities. If now the nominal SET voltage was -1 V, to ensure a 100 % probability, half of that would be -500 mV, which would result in no switching at all. For half the algorithmic response the actual response is more or less set to zero. Of course, this assumed case does not occur frequently, and typically the presented voltage scaling approach works. Also the SET probabilities are affected by the HRS with smaller HRS states leading to an increased probability for a constant voltage. Still, there is room for improvement if adapted scaling of the training voltages

is performed.

8.7 Discussion and summary

The following points summarize the main findings of this study: First of all, it has become obvious that for a correct device description it is necessary to study multiple devices on multiple timescales and with high repetition numbers. Findings based on single device experiments or limited cycle numbers are to be considered with extreme caution, as inaccurate c2c and d2d variability assumptions can change the results on higher levels of integration architectures significantly. Primarily, this can be seen in the shown examples in Figure 8.5 and in Figure 8.1. Various combinations of c2c and d2d variability lead to different synapse behavior. If not modelled with respect to a minimum statistical range of devices, wrong conclusions on the synapse behavior might be drawn, leading to suboptimal operation, which will increase the mismatch between simulation and experimental investigation of a network. Second, it is crucial to evaluate the agreement between experimental results and the employed simulation tool in detail as shown in this work. Several aspects need to be addressed accurately:

- Resistance distributions of HRS and LRS
- SET voltage onset and distribution
- Switching dynamics
- Device-to-device spread

Ultimately, simulation tools like the proposed JART VCM compact model are unavoidable for testing novel neuromorphic concepts for their feasibility in real-world situations. In this context, accurate device compact models may be seen as an important step on the way towards large-scale neuromorphic applications, just like transistor models are at the foundation of current processing units. Third, the concept of parallel devices for a single synapse is investigated. The bottom-up derivation of favorable synapse behavior in Section 8.4.1 concluded that for the desired application it is very important to have a certain amount of variability in total, which may be composed of both c2c and d2d variability to different extents. This minimum variability in the synapse composition translates into favorable tunability of the synaptic weight. It was shown that additional devices in the synapse compound enhance this tunability factor, hence enhancing the synapse performance. This effect lead to an increased

number of realistically addressable levels, a larger operation voltage window and more distinct levels per synapse. Three according parameters to assess the quality of a compound synapse were introduced and verified through experiment and simulation. An interesting outlook for the future presents itself. In more advanced integration routes the amount of d2d will most likely be reduced, while the c2c amount will remain as it is a consequence of the physical nature of the VCM-type resistance change. Therefore, the addressed case with small d2d variability may arise. However, the presented approach should be resilient towards this development, as the requirement lies within the interplay of c2c and d2d variability. However, adjustments regarding the operation voltages may become necessary because the voltage window is significantly reduced under these circumstances, requiring a voltage spacing in the tens of millivolts. Fourth, the proposed synapse structure was implemented into an exemplary neural network which was trained using a technologically plausible algorithm making use of the concept of stochastic rounding. By developing an optimized hyperparameter tuning scheme for the devices, the network was able to converge to 100 % accuracy for easy tasks. As expected from the previous discussion, higher complexity problems, i.e. higher overlap between the patterns, required additional devices per synapse to maintain high accuracy. Here, a top-down view on the network training stage revealed that a higher device count per synapse leads to more resilience against perturbations in the form of pattern overlap. However, this stability proved to have a weakness when additionally considering input noise, i.e. flipped bits, in the training stage. As the final synapse weight was reached after a single epoch for large synapses, noise in the form of flipped bits lead to degraded accuracies. In contrast, fewer devices per synapse required multiple epochs for reaching the minimum error, therefore averaging over multiple flipped bit events. Hence, for low overlaps, a lower device count surpassed the performance of higher device numbers per synapse, while high overlap tasks were better solved by higher device count synapses. One mitigation strategy of this unexpected result may present itself in a more conservative voltage scaling approach, which begins at a lower voltage and employs smaller voltage increments. By this technique, the prolonged learning stage allows averaging over multiple flipped bit patterns and therefore adds noise robustness to the network. The need for adjustments like the developed hyperparameter tuning algorithm and the device-aware network operation emphasizes the importance of algorithms that are tailored to the physical substrates.

As memristive devices have become widely used in neuromorphic applications in

recent years, the concept of using multiple devices per synapse has been applied to different realizations of memristive devices. Examples for experimental realizations of this concept were done for Electro-Chemical Metallization (ECM) cells [208] and Phase Change Mechanism (PCM) cells [215], as well as other VCM systems. In addition, since the primary requirement for employing the concept is switching voltage variability, it can be considered applicable to other VCM systems such as Ta_2O_5 [111], TiO_2 [222] or $SrTiO_3$ [223]. However, in many systems, the possibility of analog switching has been demonstrated. Further studies are required for a parallel configuration of such analog type switches since the concept proposed in this chapter is based on digital switches with two distinguishable states. However, the diverse resistance switching phenomena observed in these systems will require careful design of the synapse operation algorithms. For instance, higher resistance variabilities will reduce the realistic number of addressable synapse current levels, while a tighter switching voltage distribution may reduce the voltage window where conductance tunability is possible. The parameters derived in Section 8.4 are able to capture these device-related characteristics and offer comparable quantities for different devices and device types. On the network level, mainly theoretical results have been obtained so far due to the difficulty of large scale integration possibilities. Singha et al. [211] used simulations to investigate this synapse concept for showing Spike Timing Dependent Plasticity (STDP) behavior. Their findings showcase that increasing the number of parallel devices in the synapse brings the synapse closer to the optimal analog case. However, in their study, they did not consider resistance variability nor d2d variability. At the current state of memristive device research, these two issues have not been resolved, but may be reduced in the future. The modelling in this work therefore represents a more realistic picture of the current state of the art. Even including the described artefacts, it was possible to achieve promising results, suggesting that the concept can compensate for some of the perceived device shortcomings. Also, they did not go to the network level to investigate the performance of a neural network based on their synapses. Bill and Legenstein [209] proved the feasibility of the proposed synapse concept in a STDP update rule from a theoretical point of view and with simulations of idealized bistable devices. Their study came to the similar conclusion, that the network classification error can be reduced by increasing the synapse resolution, i.e. increasing the number of devices per synapse M. However, in their abstract model, they did not consider conductance variability in the states, leading to the assumption that each synapse can assume up to M+1 discrete conductance. A

more realistic case is shown in this chapter, where the actual addressable number of states per synapse is lower than M+1, caused by the conductance variability. However, the overall trend of performance gain is maintained, which is in line with their study. Furthermore, the study predicts a strong resilience of parallel device synapses against device non-uniformity, which we can confirm from our study. Overall, the results obtained from previous literature studies and this work agree that the proposed concept of multiple devices per synapse is a promising approach, presenting a feasible alternative to single analog devices as synapse elements. However, by introducing multiple devices per synapse, new challenges arise due to the device characteristics, which were shown to have direct impact on synapse levels and tunability window. Moreover, multiple devices per synapse results in an increased area footprint and peripheral CMOS circuitries. Solutions such as the demonstrated hyperparameter algorithm will be required to access the full potential of this promising approach.

9 Conclusion and Outlook

The goal of this work was to investigate the two switching modes, namely abrupt and analog switching, in established filamentary VCM-type HfO_2 based ReRAM devices for the application as synaptic elements in neuromorphic circuits. The test vehicle of this work was the industrially highly relevant $Pt/HfO_2/TiO_x/Ti/Pt$ stack. Utilizing ALD grown oxides, the integration into nano-crossbar devices resulted in reliable resistive switching, allowing for in-depth characterization of the physical processes that cause the two switching modes. Electrical measurements by voltage sweeps and voltage pulses were conducted to gain deeper understanding of the operation parameters for both modes. Importantly, the results were verified on device ensembles and through multiple iterations, which yielded statistically relevant datasets. Further, the device analysis was conducted in close exchange with simulations from the JART VCM v1b compact model, allowing for physical interpretation of the observed relations and further development of the model.

The most important results of this work are summarized as follows:

(1) Reliability of the initial electroforming step, device scalability and endurance was investigated on device ensembles. The high quality and uniformity of the ALD films resulted in dense and pinhole-free layers that showed highly reproducible electroforming at CMOS compatible voltage levels, underlining the industrial relevance of the investigated system. By adapting a nanoplug structure, device miniaturization down to 40 nm x 40 nm size was demonstrated. The switching properties were maintained, which emphasizes the possibility for extremely dense integration in future applications. Endurance measurements of multiple devices showed that more than 1 million switching cycles is possible without further optimization, which is sufficient especially for neuromorphic circuits, where high switching numbers are less critical than in embedded memory applications.

(2) The SET and RESET transient current analysis revealed that a two-step process is underlying in both switching events. For the SET, a delay time that is dependent

on the voltage amplitude and the previous HRS programming is followed by a voltage dependent transition time. In the HRS resistance range where abrupt switching is prevalent, the delay time is significantly longer than the transition time. Accordingly, analog programming via the transition time is inaccessible. The physical origin was identified as the delayed thermal runaway process, which is triggered after the highly variable delay time has passed. The RESET delay time is highly dependent on peripheral circuit elements such as series resistances. A delay time increase of up to six orders of magnitude is observed when the voltage divider effect is active, i.e. when the switching element resistance approaches the series resistance. The following transition time is voltage-dependent and shorter than the delay time. Importantly, its duration is independent from the previous delay time. From these results, it was possible to identify the origin of analog programming capability in the devices as the effective utilization of the transition time. In accordance, the use of short voltage pulses with moderate amplitude is imperative for analog programming. At the same time, the low LRS range and the high HRS range should be avoided to avoid the appearance of delay times.

(3) A quantification of the analog properties was presented to identify tailored operation parameters for a desired programming response. For this purpose, metrics for describing the noise-free LTP and LTD curves were introduced: resolution, linearity and asymmetry. Many neuromorphic applications benefit from completely linear as well as symmetric LTP and LTD behavior. The experimental reality is that no combination of voltage amplitudes yielded completely linear response. Instead, it was found that the LTP process exhibits a purely amplitude-dependent behavior, which is largely independent of the previous LTD programming. This is true both for resolution and nonlinearity. In contrast, the LTD resolution and nonlinearity is impacted both by the amplitude as well as the previous programming, making this process the more controllable of the two. The consequence of this relation is that symmetry of the individually nonlinear processes and symmetry in resolution is obtained when the LTD operation parameters are chosen in a way that matches the less adjustable LTP behavior. Hence, a range of viable operation conditions is systematically found and the nonlinearity, asymmetry and resolution can be chosen according to the application specific requirements.

(4) The noise feature extracted from the LTP and LTD analog programming was investigated in more detail. It was found that the most suitable conductance range for analog programming is the range that also has the highest absolute conductance

noise. A constant signal-to-noise ratio at low conductance transitions into an improved signal-to-noise ratio at high conductance, resulting in the typical bell-like shape for the plot of the standard deviation versus the mean value of conductance. Frequent literature reports of similar noise characteristics suggest that this behavior is an inherent property of filamentary VCM-type devices. Recent advancements of the understanding of electronic conduction in the cells suggest that the physical explanation can be found in the importance of single oxygen vacancy position perturbations due to the complex relation of ionic configuration in trap-assisted tunneling-based devices. Although the ionic configuration possibilities are nearly infinite, the presented considerations lead to the conclusion that the realistically distinguishable number of conductance levels in the presented devices is limited to around eight. Material modifications or noise reduction strategies could improve this number.

(5) The SET process stochasticity was identified as promising property for implementing an analog synapse by utilizing devices switched in the binary operation mode. Through statistical analysis, a quantification of cycle-to-cycle and device-to-device variability was obtained. By arrangement of parallel devices an artificial synapse was constructed and the functionality was experimentally demonstrated. The key insight was that the mentioned variability components complement each other in the synapse structure. At least cycle-to-cycle variability is required for the concept to work. Through simulations with an extended version of the JART VCM v1b model, the observations were confirmed and the artificial synapse unit was utilized in a spiking neural network for a pattern recognition task. Variation of the device per synapse count and the problem complexity illustrated the viability of the proposed synapse concept.

The present work reports important findings for the future adoption of the filamentary VCM device technology in neuromorphic circuits. However, there are a couple of open questions which should be addressed in upcoming studies:

(1) The demonstrated noise extraction resulted in a characteristic bell-shape curve, which is reported in literature, too. However, an adequate model that captures this property in the full extent is still lacking. Nevertheless, such features have proven to be highly beneficial in some new applications, e.g. Bayesian neural networks. Modeling and calibration could therefore pave the way for further development of such concepts.

(2) The demonstrations of applicability for synapse functionality in this work are limited to supervised learning tasks. However, there are many concepts that employ unsupervised learning algorithms. Applicability of filamentary devices in such

networks should be tested in the future.
(3) Through lower conductance programming, currents are reduced and the device is operated more energy-efficiently. By on-chip co-integration with commercially available transistors, this goal can be achieved. However, it needs to be determined how much of the reported properties translate to such co-integrated devices and should therefore be tested accordingly.

List of Publications

Peer Reviewed Journals

F. Cüppers, K. Hirai and H. Funakubo, "On the switching dynamics of epitaxial ferroelectric CeO_2-HfO_2 thin film capacitors", *Nano Convergence*, vol. 9, no. 56, 2022.

K. Schnieders, C. Funck, **F. Cüppers**, S. Aussen, T. Kempen, A. Sarantopoulos, R. Dittmann, S. Menzel, V. Rana, S. Hoffmann-Eifert, and S. Wiefels , "Effect of electron conduction on the read noise characteristics in ReRAM devices", *Applied Physics Letters: Materials*, vol. 10, no. 10, 101114, 2022.

M. von Witzleben, S. Wiefels, A. Kindsmüller, P. Stasner, F. Berg, **F. Cüppers**, S. Hoffmann-Eifert, R. Waser, S. Menzel and U. Böttger , "Intrinsic RESET speed limit of valence change memories", *ACS Applied Electronic Materials*, vol. 3, no. 12, pp. 5563-5572, 2021.

C. Bengel*, **F. Cüppers***, M. Payvand, R. Dittmann, R. Waser, S. Hoffmann-Eifert and S. Menzel, "Utilizing the Switching Stochasticity of HfO_2/TiO_x-Based ReRAM Devices and the Concept of Multiple Devices for the Classification of Overlapping and Noisy Patterns", *Frontiers in Neuroscience*, vol. 15, p. 621, 2021.

C. Bengel, A. Siemon, **F. Cüppers**, S. Hoffmann-Eifert, A. Hardtdegen, M. von Witzleben, L. Helllmich, R. Waser and S. Menzel, "Variability-Aware Modeling of Filamentary Oxide based Bipolar Resistive Switching Cells Using SPICE Level Compact Models", *Transactions on Circuits and Systems I*, vol. 67, no. 12, pp. 4618-4630, 2020.

M. Lübben, **F. Cüppers**, J. Mohr, M. von Witzleben, U. Breuer, R. Waser, C. Neumann and I. Valov, "Design of defect-chemical properties and device performance in memristive systems", *Science Advances*, vol. 6, no. 19, pp. eaaz9079/1-10, 2020.

F. Cüppers, S. Menzel, C. Bengel, A. Hardtdegen, M. von Witzleben, U. Böttger,

R. Waser and S. Hoffmann-Eifert, "Exploiting the switching dynamics of HfO_2-based ReRAM devices for reliable analog memristive behavior", *Applied Physics Letters: Materials*, vol. 7, no. 9, pp. 091105/1-9, 2019.

A. Hardtdegen, C. La Torre, **F. Cüppers**, S. Menzel, R. Waser, S. Hoffmann-Eifert,"Improved Switching Stability and the Effect of an Internal Series Resistor in HfO_2/TiO_x Bilayer ReRAM Cells", *IEEE Transactions on Electron Devices*, vol. 65, no. 8, pp. 3229-3236, 2018.

H. Zhang, S. Yoo, S. Menzel, C. Funck, **F. Cüppers**, D. J. Wouters, C. S. Hwang, R. Waser and S. Hoffmann-Eifert, "Understanding the Coexistence of Two Bipolar Resistive Switching Modes with Opposite Polarity in $Pt/TiO_2/Ti/Pt$ Nanosized ReRAM Devices", *ACS Applied Materials and Interfaces*, vol. 10, no. 35, pp. 29766-29778, 2018.

* Authors contributed equally

Conference Proceedings

S. Aussen, **F. Cüppers**, R. Waser, S. Hoffmann-Eifert, "Direct Comparison of the SET Kinetics of a TiO_x/Al_2O_3-based Memristive Cell in Filamentary- and Area-Mode", *2022 IEEE International Conference on Nanotechnology*, 2022.

M. E. Galicia, S. Menzel, F. Merchant, M. Müller, H. Chen, Q. Zhao, **F. Cüppers**, A. R. Jalil, Q. Shu, P. Schüffelgen, G. Mussler, C. Funck, C. Lanius, S. Wiefels, M. von Witzleben, C. Bengel, N. Kopperberg, T. Ziegler, R. Walied, A. Krüger, L. Pöhls, R. Dittmann, S. Hoffmann-Eifert, V. Rana, D. Grützmacher, M. Wuttig, D. Wouters, A. Vescan, T. Gemmeke, J. Knoch, M. Lemme, R. Leupers and R. Waser, "NEUROTEC I: Neuro-inspired Artificial Intelligence Technologies for the Electronics of the Future", *DATE 2022 Multi-Partner Projects*, 2022.

S. Menzel, S. Wiefels, C. Bengel, **F. Cüppers**, J. Mohr, S. Hoffmann-Eifert, D. Wouters, "Reliability Aspects of Memristive Devices for Computation-in-Memory Applications", *17th International Workshop on Cellular Nanoscale Networks and their Applications (CNNA)*, 2021.

S. A. Chekol, **F. Cüppers**, R. Waser and S. Hoffmann-Eifert, "An $Ag/HfO_2/Pt$ Threshold Switching Device with an Ultra-Low Leakage ($< 10\,fA$), High On/OffRatio

($> 10^{11}$), and Low Threshold Voltage ($< 0.2\,\mathrm{V}$) for Energy-Efficient Neuromorphic Computing", *IEEE International Memory Workshop (IMW)*, 2021.

A. Hardtdegen*, **F. Cüppers***, M. von Witzleben, U. Boettger, S. Menzel, R. Waser and S. Hoffmann-Eifert, "Characterization of HfO_2/TiO_x ReRAM Cells in Pulse Operation Mode", *2018 IEEE International Conference on Nanotechnology*, 2018.

* Authors contributed equally

Conference Talks and Posters

F. Cüppers, C. Bengel, M. Payvand, R. Waser, S. Menzel and S. Hoffmann-Eifert, "Stochastically Switching HfO_2/TiO_x ReRAM Devices for Spiking Neural Network Synapses", *MRS Fall Meeting*, Boston, USA, 29 Nov - 08 Dec, 2021.

F. Cüppers, S. Menzel, C. Bengel, A. Hardtdegen, M. von Witzleben, U. Böttger, R. Waser and S. Hoffmann-Eifert, "Bipolar analogue memristive function for neuromorphic computing enabled by stack design of HfO_2-based ReRAM devices", *The International Conference on Memristive Materials, Devices & Systems (MEMRISYS)*, Dresden, Germany, 08-11 July, 2019.

F. Cüppers, A. Hardtdegen, S. Menzel, D. J. Wouters, R. Waser and S. Hoffmann-Eifert, "HfO_2/TiO_2 Memristive Devices as Synapses in Future Artificial Neural Networks", *Cognitive Computing 2018 - Merging Concepts with Hardware*, Hannover, Germany, 18-20 December, 2018.

A. Hardtdegen*, **F. Cüppers***, M. von Witzleben, S. Menzel, U. Böttger, R. Waser and S. Hoffmann-Eifert, "Gradual SET and Gradual RESET Behaviour for HfO_2/TiO_x Bilayer ReRAM Cells", *Faraday Discussion on New memory paradigms: memristive phenomena and neuromorphic applications*, Aachen, Germany, 15-17 October, 2018.

* Authors contributed equally

Bibliography

[1] J. H. Choi, Y. Mao, and J. P. Chang. "Development of hafnium based high-k materials-A review". In: *Mater. Sci. Eng. R-Rep.* 72.6 (2011), pp. 97–136 (cit. on p. 1).

[2] H. Harris, K. Choi, N. Mehta, A. Chandolu, N. Biswas, G. Kipshidze, S. Nikishin, S. Gangopadhyay, and H. Temkin. "HfO_2 gate dielectric with 0.5 nm equivalent oxide thickness". In: *Appl. Phys. Lett.* 81.6 (2002), pp. 1065–1067 (cit. on p. 1).

[3] G. D. Wilk, R. M. Wallace, and J. M. Anthony. "High-κ gate dielectrics: Current status and materials properties considerations". In: *Journal of Applied Physics* 89.10 (2001), pp. 5243–5275 (cit. on p. 1).

[4] C. Chaneliere, J. Autran, R. Devine, and B. Balland. "Tantalum pentoxide (Ta_2O_5) thin films for advanced dielectric applications". In: *Materials Science & Engineering R-Reports* 22.6 (1998), pp. 269–322 (cit. on p. 1).

[5] LETI. *LETI and CMP announce world's first multi-project wafer service with integrated silicon OxRAM*. Tech. rep. 2018. URL: `https://mycmp.fr/wp-content/uploads/2021/02/leti_nr_cmp_pr_mad200_aug2-18.pdf` (cit. on pp. 1, 2).

[6] X. Sheng, C. E. Graves, S. Kumar, X. Li, B. Buchanan, L. Zheng, S. Lam, C. Li, and J. P. Strachan. "Low-Conductance and Multilevel CMOS-Integrated Nanoscale Oxide Memristors". In: *Advanced Electronic Materials* (2019), p. 1800876 (cit. on pp. 2, 3, 45, 51–53).

[7] M. von Witzleben, T. Hennen, A. Kindsmüller, S. Menzel, R. Waser, and U. Böttger. "Study of the SET switching event of VCM-based memories on a picosecond timescale". In: *J. Appl. Phys.* 127.20 (2020), p. 204501 (cit. on pp. 2, 15, 121).

[8] M. von Witzleben, S. Walfort, R. Waser, S. Menzel, and U. Böttger. "Determining the electrical charging speed limit of ReRAM devices". In: *IEEE J. Electron Devices Soc.* 9 (2021), pp. 667–678 (cit. on pp. 2, 15, 121).

[9] M. von Witzleben, S. Wiefels, A. Kindsmüller, P. Stasner, F. Berg, F. Cüppers, S. Hoffmann-Eifert, R. Waser, S. Menzel, and U. Böttger. "Intrinsic RESET speed limit of valence change memories". In: *ACS Appl. Electron. Mater.* 3.12 (2021), pp. 5563–5572 (cit. on pp. 2, 15, 121).

[10] A. C. Torrezan, J. P. Strachan, G. Medeiros-Ribeiro, and R. S. Williams. "Sub-nanosecond switching of a tantalum oxide memristor". In: *Nanotechnology* 22 (2011), p. 485203 (cit. on pp. 2, 15).

[11] B. Govoreanu, A. Redolfi, L. Zhang, C. Adelmann, M. Popovici, S. Clima, H. Hody, V. Paraschiv, I.P. Radu, A. Franquet, J.-C. Liu, J. Swerts, O. Richard, H. Bender, L. Altimime, and M. Jurczak. "Vacancy-Modulated Conductive Oxide Resistive RAM (VMCO-RRAM):An Area-Scalable Switching Current, Self-Compliant, Highly Nonlinear and Wide On/Off-Window Resistive Switching Cell". In: *Electron Devices Meeting (IEDM), 2013 IEEE International* 13 (2013), pp. 256–259 (cit. on p. 2).

[12] L. Grenouillet et al. "16kbit 1T1R OxRAM arrays embedded in 28nm FDSOI technology demonstrating low BER, high endurance, and compatibility with core logic transistors". In: *2021 IEEE International Memory Workshop (IMW)*. 2021 IEEE International Memory Workshop (IMW), 2021, pp. 1–4 (cit. on p. 2).

[13] V. Milo, C. Zambelli, P. Olivo, E. Perez, M. K. Mahadevaiah, O. G. Ossorio, Ch. Wenger, and D. Ielmini. "Multilevel HfO_2-based RRAM devices for low-power neuromorphic networks". In: *APL Mater.* 7.8 (2019), pp. 81120/1–10 (cit. on pp. 2, 3).

[14] J. Hazra, M. Liehr, K. Beckmann, M. Abedin, S. Rafq, and N. Cady. "Optimization of Switching Metrics for CMOS Integrated HfO_2 based RRAM Devices on 300 mm Wafer Platform". In: *2021 IEEE International Memory Workshop (IMW)*. 2021 IEEE International Memory Workshop (IMW), 2021, pp. 1–4 (cit. on p. 2).

[15] Y. Bai, Y. Zhang, H. Wu, and H. Qian. “High-density WO_x-based RRAM with a W-doped AlO_x insertion layer”. In: *5th IEEE International Memory Workshop (IMW), Monterey, CA*. 2013 5th IEEE International Memory Workshop (IMW), 2013, pp. 120–123 (cit. on p. 2).

[16] S. Balatti, S. Ambrogio, D. Ielmini, and D. C. Gilmer. “Variability and failure of set process in HfO_2 RRAM”. In: *2013 5th IEEE International Memory Workshop (IMW 2013)* (2013), pp. 38–41 (cit. on p. 2).

[17] T. Cabout, E. Vianello, E. Jalaguier, H. Grampeix, G. Molas, P. Blaise, O. Cueto, M. Guillermet, J. F. Nodin, L. Perniola, S. Blonkowski, S. Jeannot, S. Denorme, S. Candelier, M. Bocquet, and C. Muller. “Effect of SET temperature on data retention performances of HfO_2-based RRAM cells”. In: *2014 IEEE 6th International Memory Workshop*. 2014 IEEE 6th International Memory Workshop, 2014 (cit. on p. 2).

[18] C. Y. Chen, L. Goux, A. Fantini, A. Redolfi, G. Groeseneken, and M. Jurczak. “Doped Gd-O Based RRAM for Embedded Application”. In: *2016 IEEE 8th International Memory Workshop (IMW)*. 2016 IEEE 8th International Memory Workshop (IMW), 2016, pp. 1–4 (cit. on p. 2).

[19] Z. Chen, H. Wu, B. Gao, D. Wu, N. Deng, H. Qian, Z. Lu, B. Haukness, M. Kellam, and G. Bronner. “Performance Improvements by SL-Current Limiter and Novel Programming Methods on 16MB RRAM Chip”. In: *2017 IEEE International Memory Workshop (IMW)*. 2017 IEEE International Memory Workshop (IMW), 2017, pp. 1–4 (cit. on p. 2).

[20] B. Govoreanu et al. “10x10 nm^2 Hf/HfO_x Crossbar Resistive RAM with Excellent Performance, Reliability and Low-Energy Operation”. In: *2011 IEEE International Electron Devices Meeting - IEDM ’11*. IEDM Tech. Dig., 2011, pp. 31.6.1–31.6.4 (cit. on pp. 2, 45).

[21] A. Grossi, E. Nowak, C. Zambelli, C. Pellissier, S. Bernasconi, G. Cibrario, K. El Hajjam, R. Crochemore, J. Nodin, P. Olivo, and L. Perniola. “Fundamental variability limits of filament-based RRAM”. In: *2016 IEEE International Electron Devices Meeting (IEDM)*. 2016 IEEE International Electron Devices Meeting (IEDM), 2016, pp. 4.7.1–4.7.4 (cit. on p. 2).

[22] A. Grossi, C. Zambelli, P. Olivo, E. Nowak, G. Molas, J. F. Nodin, and L. Perniola. “Cell-to-Cell Fundamental Variability Limits Investigation in OxRRAM Arrays”. In: *IEEE Electron Device Lett.* 39.1 (2018), pp. 27–30 (cit. on p. 2).

[23] T. Kempen, R. Waser, and V. Rana. “50x Endurance Improvement in TaO_x RRAM by Extrinsic Doping”. In: *2021 IEEE International Memory Workshop (IMW)* (2021), pp. 1–4 (cit. on p. 2).

[24] Sheyang Ning, Tomoko Ogura Iwasaki, and Ken Takeuchi. “Write stress reduction in 50nm Al_xO_y ReRAM improves endurance 1.4× and write time, energy by 17%”. In: *2013 5th IEEE International Memory Workshop*. 2013 5th IEEE International Memory Workshop, 2013, pp. 56–59 (cit. on p. 2).

[25] A. Padovani, L. Larcher, P. Padovani, C. Cagli, and B. De Salvo. “Understanding the role of the Ti metal electrode on the forming of HfO_2-based RRAMs”. In: *2012 4th IEEE International Memory Workshop (IMW)* (2012), 4 pp.–4 pp. (Cit. on p. 2).

[26] G. Sassine, D. Alfaro Robayo, C. Nail, J.-F. Nodin, J. Coignus, G. Molas, and E. Nowak. “Optimizing Programming Energy for Improved RRAM Reliability for High Endurance Applications”. In: *2018 IEEE International Memory Workshop (IMW)*. 2018 IEEE International Memory Workshop (IMW), 2018, pp. 1–4 (cit. on pp. 2, 48, 54).

[27] S. Subhechha, R. Degraeve, P. Roussel, L. Goux, K. De Meyer, J. Van Houdt, and G. S. Kar. “Experimental Determination of the Driving Force for Switching in TiN/a-Si/TiO_x/TiN RRAM Devices”. In: *2019 IEEE 11th International Memory Workshop (IMW)*. 2019 IEEE 11th International Memory Workshop (IMW), 2019, pp. 1–4 (cit. on p. 2).

[28] R. Yasuhara, T. Ninomiya, S. Muraoka, Z. Wei, K. Katayama, and T. Takagi. “Consideration of conductive filament for realization of low-current and highly-reliable TaO_x ReRAM”. In: *2013 5th IEEE International Memory Workshop, IMW 2013*. 2013 5th IEEE International Memory Workshop, IMW 2013, 2013, pp. 34–37 (cit. on p. 2).

[29] C. Zambelli, A. Grossi, P. Olivo, D. Walczyk, J. Dabrowski, B. Tillack, T. Schroeder, R. Kraemer, V. Stikanov, and C. Walczyk. “Electrical characterization of read window in ReRAM arrays under different SET/RESET cycling

conditions". In: *2014 IEEE 6th International Memory Workshop (IMW)*. 2014 IEEE 6th International Memory Workshop (IMW), 2014, pp. 1–4 (cit. on p. 2).

[30] S. Stathopoulos, A. Khiat, M. Trapatseli: S. Cortese, A. Serb, I. Valov, and T.Prodromakis. "Multibit memory operation of metal-oxide bi-layer memristors". In: *Sci Rep* 7 (2017), p. 17532 (cit. on pp. 3, 56).

[31] P. Bousoulas, I. Giannopoulos, P. Asenov, I. Karageorgiou, and D. Tsoukalas. "Investigating the origins of high multilevel resistive switching in forming free Ti/TiO_{2-x} -based memory devices through experiments and simulations". In: *J. Appl. Phys.* 121.9 (2017), pp. 94501/1–9 (cit. on p. 3).

[32] W. Chen, W. Lu, B. Long, Y. Li, D. Gilmer, G. Bersuker, S. Bhunia, and R. Jha. "Switching characteristics of W/Zr/HfO_2/TiN ReRAM devices for multi-level cell non-volatile memory applications". In: *Semicond. Sci. Tech.* 30 (2015), p. 075002 (cit. on p. 3).

[33] W. Kim, S. Menzel, D. J. Wouters, R. Waser, and V. Rana. "3-Bit Multi Level Switching by Deep Reset Phenomenon in Pt/W/TaO_x/Pt-ReRAM Devices". In: *IEEE Electron Device Lett.* 37.5 (2016), pp. 564–567 (cit. on pp. 3, 56).

[34] A. Prakash, J. Park, J. Song, J. Woo, E. Cha, and H. Hwang. "Demonstration of Low Power 3-bit Multilevel Cell Characteristics in a TaO_x-Based RRAM by Stack Engineering". In: *IEEE Electron Device Lett.* 36 (2015), pp. 32–34 (cit. on p. 3).

[35] A. Prakash, D. Deleruyelle, J. Song, M. Bocquet, and H. Hwang. "Resistance controllability and variability improvement in a TaO_x-based resistive memory for multilevel storage application". In: *Appl. Phys. Lett.* 106.23 (2015), pp. 233104/1–4 (cit. on p. 3).

[36] M. Terai, Y. Sakotsubo, S. Kotsuji, and H. Hada. "Resistance Controllability of Ta_2O_5/TiO_2 Stack ReRAM for Low-Voltage and Multilevel Operation". In: *IEEE Electron Device Lett.* 31.3 (2010), pp. 204–206 (cit. on p. 3).

[37] M. Payvand, Y. Demirag, T. Dalgaty, E. Vianello, and G. Indiveri. "Analog Weight Updates with Compliance Current Modulation of Binary ReRAMs for On-Chip Learning". In: *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. 2020, pp. 1–5 (cit. on pp. 3, 51, 151).

[38] U. Böttger, M. von Witzleben, V. Havel, K. Fleck, V. Rana, R. Waser, and S. Menzel. “Picosecond multilevel resistive switching in tantalum oxide thin films”. In: *Sci. Rep.* 10.1 (2020), p. 16391 (cit. on pp. 3, 121).

[39] E. Perez, M. K. Mahadevaiah, E. P. Quesada, and C. Wenger. “Variability and Energy Consumption Tradeoffs in Multilevel Programming of RRAM Arrays”. In: *IEEE Trans. Electron Devices* 68 (2021), pp. 2693–2698 (cit. on p. 3).

[40] S. Poblador, M. Gonzalez, and F. Campabadal. “Investigation of the multilevel capability of TiN/Ti/HfO_2/W resistive switching devices by sweep and pulse programming”. In: *Microelectronic Engineering* 187-188 (2018), pp. 148–153 (cit. on p. 3).

[41] Furqan Zahoor and Tun Zainal Azni Zulkifli andFarooq Ahmad Khanday. “Resistive Random Access Memory (RRAM): an Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (mlc) Storage, Modeling, and Applications”. In: Nanoscale Research Letters, 2020 (cit. on p. 3).

[42] C. Sung, H. Hwang, and I. K. Yoo. “Perspective: A review on memristive hardware for neuromorphic computation”. In: *J. Appl. Phys.* 124.15 (2018), pp. 151903/1–13 (cit. on pp. 3, 20).

[43] F. Cüppers, S. Menzel, C. Bengel, A. Hardtdegen, M. von Witzleben, U. Böttger, R. Waser, and S. Hoffmann-Eifert. “Exploiting the switching dynamics of HfO_2-based ReRAM devices for reliable analog memristive behavior”. In: *APL Materials* 7.9 (2019), pp. 091105/1–9 (cit. on pp. 3, 54–56, 58, 59, 61, 62, 64, 66, 68, 139).

[44] J. Woo, K. Moon, J. Song, S. Lee, M. Kwak, J. Park, and H. Hwang. “Improved Synaptic Behavior Under Identical Pulses Using AlO_x/HfO_2 Bilayer RRAM Array for Neuromorphic Systems”. In: *IEEE Electron Device Lett.* 37.8 (2016), pp. 994–997 (cit. on pp. 3, 56).

[45] N. Bai, B. Tian, G. Mao, K. Xue, T. Wang, J. Yuan, X. Liu, Z. Li, S. Guo, Z. Zhou, N. Liu, H. Lu, X. Tang, H. Sun, and X. Miao. “Homo-layer hafnia-based memristor with large analog switching window”. In: *Appl. Phys. Lett.* 118.4 (2021), p. 043502 (cit. on p. 3).

[46] C. Giovinazzo, J. Sandrini, E. Shahrabi, O. T. Celik, Y. Leblebici, and C. Ricciardi. "Analog Control of Retainable Resistance Multistates in HfO_2 Resistive-Switching Random Access Memories (ReRAMs)". In: *Appl. Electron. Mater.* 1 (2019), pp. 900–909 (cit. on p. 3).

[47] L. Zhao, H. Chen, S. Wu, Z. Jiang, S. Yu, T. Hou, H. P. Wong, and Y. Nishi. "Multi-level control of conductive nano-filament evolution in HfO_2 ReRAM by pulse-train operations". In: *Nanoscale* 6 (2014), pp. 5698–5702 (cit. on p. 3).

[48] J. Frascaroli, S. Brivio, E. Covi, and S. Spiga. "Evidence of soft bound behaviour in analogue memristive devices for neuromorphic computing". In: *Sci Rep* 8 (2018), pp. 7178/1–12 (cit. on pp. 3, 67).

[49] E. Covi, S. Brivio, A. Serb, T. Prodromakis, M. Fanciulli, and S. Spiga. "Analog Memristive Synapse in Spiking Networks Implementing Unsupervised Learning". In: *Frontiers in Neuroscience* 10 (2016), pp. 6–13 (cit. on pp. 3, 22).

[50] E. Covi, S. Brivio, M. Fanciulli, and S. Spiga. "Synaptic potentiation and depression in Al:HfO_2-based memristor". In: *Microelectron. Eng.* 147 (2015), pp. 41–44 (cit. on p. 3).

[51] N. Gong, T. Idé, S. Kim, I. Boybat, A. Sebastian, V. Narayanan, and T. Ando. "Signal and noise extraction from analog memory elements for neuromorphic computing". In: *Nature Communications* 9 (2018), p. 2102 (cit. on pp. 3, 76, 77).

[52] W. Wu, H. Wu, B. Gao, P. Yao, X. Zhang, X. Peng, S. Yu, and H. Qian. "A Methodology to Improve Linearity of Analog RRAM for Neuromorphic Computing". In: *2018 IEEE SYMPOSIUM ON VLSI TECHNOLOGY*. IEEE. IEEE, 2018, pp. 103–104 (cit. on p. 3).

[53] B. Long, Y. Li, and R. Jha. "Switching Characteristics of Ru/HfO_2/TiO_{2-x}/Ru RRAM Devices for Digital and Analog Nonvolatile Memory Applications". In: *IEEE Electron Device Lett.* 33.5 (2012), pp. 706–708 (cit. on p. 3).

[54] S. Yu, B. Gao, Z. Fang, H. Yu, J.F. Kang, and H.-S. P. Wong. "A Low Energy Oxide-Based Electronic Synaptic Device for Neuromorphic Visual Systems with Tolerance to Device Variation". In: *Adv. Mater.* 25.12 (2013), pp. 1774–1779 (cit. on pp. 3, 132).

[55] Zongwei Wang, Minghui Yin, Teng Zhang, Yimao Cai, Yangyuan Wang, Yuchao Yang, and Ru Huang. “Engineering incremental resistive switching in TaO_x based memristors for brain-inspired computing”. In: *Nanoscale* 8 (2016), pp. 14015–14022 (cit. on p. 3).

[56] C. Li, C. E. Graves, X. Sheng, D. Miller, M. Foltin, G. Pedretti, and J. P. Strachan. “Analog content-addressable memories with memristors”. In: *Nat Commun* 11 (2020), p. 1638 (cit. on pp. 3, 73).

[57] W. J. Chen, C. H. Cheng, P. E. Lin, Y. T. Tseng, T. C. Chang, and J. S. Chen. “Analog Resistive Switching and Synaptic Functions in WO_x/TaO_x Bilayer through Redox-Induced Trap-Controlled Conduction”. In: *ACS Appl. Electron. Mater.* 1.11 (2019), pp. 2422–2430 (cit. on p. 3).

[58] M. Hu, C. E. Graves, C. Li, Y. Li, N. Ge, E. Montgomery, N. Davila, H. Jiang, R. S. Williams, J. J. Yang, Q. Xia, and J. P. Strachan. “Memristor-Based Analog Computation and Neural Network Classification with a Dot Product Engine”. In: *Adv. Mater.* 30.9 (2018), pp. 1705914/1–10 (cit. on p. 3).

[59] E. J. Merced-Grafals, N. Davila, N. Ge, R. S. Williams, and J. P. Strachan. “Repeatable, accurate, and high speed multi-level programming of memristor 1T1R arrays for power efficient analog computing applications”. In: *Nanotechnology* 27.36 (2016), pp. 365202/1– (cit. on p. 3).

[60] T. Wan, B. Qu, H. Du, X. Lin, Q. Lin, D. W. Wang, C. Cazorla, S. Li, S. Liu, and D. Chu. “Digital to analog resistive switching transition induced by graphene buffer layer in strontium titanate based devices”. In: *J. Colloid Interface Sci.* 512 (2018), pp. 767–774 (cit. on p. 3).

[61] R. Wang, T. Shi, X. Zhang, W. Wang, J. Wei, J. Lu, X. Zhao, Z. Wu, R. Cao, S. Long, Q. Liu, and M. Liu. “Bipolar Analog Memristors as Artificial Synapses for Neuromorphic Computing”. In: *Materials* 11.11 (2018), pp. 2102/1–14 (cit. on p. 3).

[62] X. Li, H. Wu, B. Gao, W. Wu, D. Wu, N. Deng, J. Cai, and H. Qian. “Electrode-induced digital-to-analog resistive switching in TaO_x-based RRAM devices”. In: *Nanotechnology* 27.30 (2016), pp. 305201/1–6 (cit. on p. 3).

[63] M. Prezioso, F. Merrikh Bayat, B. Hoskins, K. Likharev, and D. Strukov. “Self-Adaptive Spike-Time-Dependent Plasticity of Metal-Oxide Memristors”. In: *Sci. Rep.* 6 (2016), p. 21331 (cit. on p. 3).

[64] G. C. Adam, B. D. Hoskins, M. Prezioso, F. Merrikh-Bayat, B. Chakrabarti, and D. B. Strukov. “3-D Memristor Crossbars for Analog and Neuromorphic Computing Applications”. In: *IEEE Trans. Electron Devices* 64.1 (2017), pp. 312–318 (cit. on p. 3).

[65] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov. “Training and operation of an integrated neuromorphic network based on metal-oxide memristors”. In: *Nature* 521.7550 (2015), pp. 61–64 (cit. on pp. 3, 22).

[66] J. Zhu, T. Zhang, Y. Yang, and R. Huang. “A comprehensive review on emerging artificial neuromorphic devices”. In: *Applied Physics Reviews* 7.1 (2020), pp. 11312/1–107 (cit. on pp. 4, 11, 20).

[67] S. Brivio, D. R. B. Ly, E. Vianello, and S. Spiga. “Non-linear Memristive Synaptic Dynamics for Efficient Unsupervised Learning in Spiking Neural Networks”. In: *Front. Neurosci.* 15 (2021), p. 580909 (cit. on pp. 4, 73).

[68] T. Gokmen and Y. Vlasov. “Acceleration of Deep Neural Network Training with Resistive Cross-Point Devices: Design Considerations”. In: *Front. Neurosci.* 10 (2016), pp. 333/1–13 (cit. on pp. 4, 22, 73).

[69] T. Dalgaty, M. Payvand, F. Moro, D. R. B. Ly, F. Pebay-Peyroula, J. Casas, G. Indiveri, and E. Vianello. “Hybrid neuromorphic circuits exploiting non-conventional properties of RRAM for massively parallel local plasticity mechanisms”. In: *APL Mater.* 7.8 (2019), pp. 81125/1–12 (cit. on pp. 4, 22).

[70] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, and H. Wong. “Stochastic learning in oxide binary synaptic device for neuromorphic computing”. In: *Frontiers in Neuroscience* 7 (2013), pp. 186/1– (cit. on pp. 4, 132).

[71] R. Naous, M. AlShedivat, E. Neftci, G. Cauwenberghs, and K. N. Salama. “Memristor-based neural networks: Synaptic versus neuronal stochasticity”. In: *AIP Adv.* 6.11 (2016), pp. 111304/1–7 (cit. on pp. 4, 132).

[72] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. “Gradient-based learning applied to document recognition”. In: *Proc. IEEE* 86.11 (1998), pp. 2278–2324 (cit. on p. 4).

[73] C. Wenger, F. Zahari, M. K. Mahadevaiah, E. Perez, I. Beckers, H. Kohlstedt, and M. Ziegler. “Inherent Stochastic Learning in CMOS-Integrated HfO_2 Arrays for Neuromorphic Computing”. In: *IEEE Electron Device Lett.* 40.4 (2019), pp. 639–642 (cit. on pp. 4, 22, 132).

[74] T. Dalgaty, N. Castellani, C. Turck, K. E. Harabi, D. Querlioz, and E. Vianello. “In situ learning using intrinsic memristor variability via Markov chain Monte Carlo sampling”. In: *Nat. Electron.* 4.2 (2021), pp. 151–161 (cit. on pp. 4, 22).

[75] R. Waser, R. Bruchhaus, and S. Menzel. “Redox-based Resistive Switching Memories”. In: *Nanoelectronics and Information Technology (3rd edition).* Ed. by R. Waser. Wiley-VCH, 2012, pp. 683–710 (cit. on p. 5).

[76] Noam Nisan and Shimon Schocken. “The Elements of Computing Systems: Building a Modern Computer from First Principles”. In: The MIT Press, 2005. Chap. 5 (cit. on p. 9).

[77] S. Salahuddin, K. Ni, and S. Datta. “The era of hyper-scaling in electronics”. In: *Nat. Electron.* 1.8 (2018), pp. 442–450 (cit. on p. 9).

[78] T. Heisig, J. Kler, H. Du, C. Baeumer, F. Hensling, M. Glöß, M. Moors, A. Locatelli, T. O. Mentes, and F. Genuzio. “Antiphase Boundaries Constitute Fast Cation Diffusion Paths in $SrTiO_3$ Memristive Devices”. In: *AFM* (2020), p. 2004118 (cit. on p. 10).

[79] U. N. Gries, M. Kessel, F. V. E. Hensling, R. Dittmann, M. Martin, and R. A. De Souza. “Behavior of cation vacancies in single-crystal and in thin-film $SrTiO_3$: The importance of strontium vacancies and their defect associates”. In: *Physical Review Materials* 4.12 (2020), pp. 123404/1– (cit. on p. 10).

[80] M. P. Mueller, K. Pingen, A. Hardtdegen, S. Aussen, A. Kindsmueller, S. Hoffmann-Eifert, and R. A. De Souza. “Cation diffusion in polycrystalline thin films of monoclinic HfO_2 deposited by atomic layer deposition”. In: *APL Mater.* 8.8 (2020), pp. 81104/1–8 (cit. on p. 10).

[81] U. N. Gries, H. Schraknepper, K. Skaja, F. Gunkel, S. Hoffmann-Eifert, R. Waser, and R. A. De Souza. “A SIMS study of cation and anion diffusion in tantalum oxide”. In: *Phys. Chem. Chem. Phys.* 20.2 (2018), pp. 989–996 (cit. on p. 10).

[82] Y. Ma, D. Li, A. A. Herzing, D. A. Cullen, B. T. Sneed, K. L. More, N. T. Nuhfer, J. A. Bain, and M. Skowronski. "Formation of the Conducting Filament in TaO_x-Resistive Switching Devices by Thermal-Gradient-Induced Cation Accumulation". In: *ACS Appl. Mater. Interfaces* 10.27 (2018), 23187–23197 (cit. on p. 10).

[83] R. Waser (Ed.) *Nanoelectronics and Information Technology*. Ed. by R. Waser (Ed.) 3rd. Wiley-VCH, 2012 (cit. on pp. 11–15).

[84] D. Ielmini and R. Waser. *Resistive switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications*. Wiley-VCH, 2016 (cit. on p. 11).

[85] D. Cooper, C. Baeumer, N. Bernier, A. Marchewka, C. La Torre, R. E. Dunin-Borkowski, S. Menzel, R. Waser, and R. Dittmann. "Anomalous Resistance Hysteresis in Oxide ReRAM: Oxygen Evolution and Reincorporation Revealed by in situ TEM". In: *Adv. Mater.* 29.23 (2017), p. 1700212 (cit. on p. 12).

[86] D. S. Jeong, H. Schroeder, U. Breuer, and R. Waser. "Characteristic electroforming behavior in $Pt/TiO_2/Pt$ resistive switching cells depending on atmosphere". In: *J. Appl. Phys.* 104.12 (2008), pp. 123716/1–8 (cit. on p. 12).

[87] I. Valov, E. Linn, S. Tappertzhofen, S. Schmelzer, J. v. d. Hurk, F. Lentz, and R. Waser. "Nanobatteries in redox-based resistive switches require extension of memristor theory". In: *Nature Communications* 4 (2013), p. 1771 (cit. on p. 12).

[88] J. J. Yang, J. P. Strachan, F. Miao, M. Zhang, M. D. Pickett, W. Yi, D. A. A. Ohlberg, G. Medeiros-Ribeiro, and R. S. Williams. "Metal/TiO_2 interfaces for memristive switches". In: *Appl. Phys. A - Mater. Sci. Process.* 102.4 (2011), pp. 785–789 (cit. on p. 13).

[89] A. Sawa. "Resistive switching in transition metal oxides". In: *Mater. Today* 11.6 (2008), pp. 28–36 (cit. on p. 13).

[90] B. Arndt, F. Borgatti, F. Offi, M. Phillips, P. Parreira, T. Meiners, S. Menzel, K. Skaja, G. Panaccione, D. A. MacLaren, R. Waser, and R. Dittmann. "Spectroscopic Indications of Tunnel Barrier Charging as the Switching Mechanism in Memristive Devices". In: *Advanced Functional Materials* (2017), 1702282–n/a (cit. on p. 13).

[91] A. Sawa and R. Meyer. “Interface Type Switching”. In: *Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications*. Ed. by D. Ielmini; R. Waser. Wiley, 2016 (cit. on p. 13).

[92] F.A. Kröger and H.A. Vink. “Relations between the Concentrations of Imperfections in Crystalline Solids”. In: *Solid State Physics* (1956), 307–435 (cit. on p. 13).

[93] C. Funck and S. Menzel. “Comprehensive Model of Electron Conduction in Oxide-Based Memristive Devices”. In: *ACS Appl. Electron.* 3 (2021), pp. 3674–3692 (cit. on pp. 14, 15, 17, 99).

[94] K. Xiong, J. Robertson, M. C. Gibson, and S. J. Clark. “Defect energy levels in HfO_2 high-dielectric-constant gate oxide”. In: *Appl. Phys. Lett.* 87 (2005), pp. – (cit. on p. 15).

[95] K. Xiong and J. Robertson. “Point defects in HfO_2 high K gate oxide”. In: *Microelectron. Eng.* 80 (2005), pp. 408–411 (cit. on p. 15).

[96] S. Siegel, C. Baeumer, A. Gutsche, M. von Witzleben, R: Waser, S. Menzel, and R. Dittmann. “Trade-Off Between Data Retention and Switching Speed in Resistive Switching ReRAM Devices”. In: *Adv. Electron. Mater.* 7.1 (2020), pp. 2000815/1– (cit. on pp. 15, 17, 121).

[97] S. Menzel, M. Waters, A. Marchewka, U. Böttger, R. Dittmann, and R. Waser. “Origin of the Ultra-nonlinear Switching Kinetics in Oxide-Based Resistive Switches”. In: *Adv. Funct. Mater.* 21.23 (2011), pp. 4487–4492 (cit. on pp. 15, 16, 54, 56, 121, 139).

[98] M. Lübben, F. Cüppers, J. Mohr, M. von Witzleben, U. Breuer, R. Waser, C. Neumann, and I. Valov. “Design of defect-chemical properties and device performance in memristive systems”. In: *Sci. Adv.* 6.19 (2020), eaaz9079/1–10 (cit. on p. 15).

[99] M. Lübben, S. Menzel, S. G. Park, M. Yang, R.Waser, and I. Valov. “SET kinetics of electrochemical metallization cells - Influence of counter electrodes in SiO_2/Ag based systems”. In: *Nanotechnology* 28.13 (2017), pp. 135205/1–6 (cit. on p. 15).

[100] K. Fleck, C. La Torre, N. Aslam, S. Hoffmann-Eifert, U. Böttger, and S. Menzel. “Uniting Gradual and Abrupt SET Processes in Resistive Switching Oxides”. In: *Phys. Rev. Applied* 6.6 (2016), p. 064015 (cit. on pp. 15, 56, 60, 68, 69, 121, 139).

[101] W. Kim, S. Menzel, D. J. Wouters, Y. Guo, J. Robertson, B. Rösgen, R. Waser, and V. Rana. “Impact of oxygen exchange reaction at the ohmic interface in Ta_2O_5-based ReRAM devices”. In: *Nanoscale* 8.41 (2016), pp. 17774–17781 (cit. on p. 15).

[102] S. Bradley, A. Shluger, and G. Bersuker. “Electron-Injection-Assisted Generation of Oxygen Vacancies in Monoclinic HfO_2”. In: *Physical Review Applied* 4.6 (2015), p. 064008 (cit. on p. 15).

[103] R. Oettking, S. Kupke, E. Nadimi, R. Leitsmann, F. Lazarevic, P. Plaenitz, G. Roll, S. Slesazeck, M. Trentzsch, and T. Mikolajick. “Defect generation and activation processes in HfO_2 thin films: Contributions to stress-induced leakage currents”. In: *Phys. Status Solidi A-Appl. Mat.* 212.3 (2015), pp. 547–553 (cit. on p. 15).

[104] K. Fleck, U. Böttger, R. Waser, and S. Menzel. “Interrelation of Sweep and Pulse Analysis of the SET Process in $SrTiO_3$ Resistive Switching Memories”. In: *IEEE Electron Device Lett.* 35.9 (2014), pp. 924 –926 (cit. on pp. 15, 121).

[105] M. von Witzleben, K. Fleck, C. Funck, B. Baumkötter, M. Zuric, A. Idt, T. Breuer, R. Waser, U. Böttger, and S. Menzel. “Investigation of the Impact of High Temperatures on the Switching Kinetics of Redox-based Resistive Switching Cells using a Highspeed Nanoheater”. In: *Adv. Electron. Mat.* 3.12 (2017), p. 1700294 (cit. on pp. 15, 60, 121).

[106] V. Havel, K. Fleck, B. Rösgen, V. Rana, S. Menzel, U. Böttger, and R. Waser. “Ultrafast Switching in Ta_2O_5-based Resistive Memories”. In: *Silicon Nanoelectronics Worshop SNW 2016, Honolulu, HI, USA, 12-13 June 2016*. Silicon Nanoelectronics Worshop SNW 2016, Hawaii, 2016, pp. 82–83 (cit. on p. 15).

[107] M. B. Gonzalez, M. Maestro-Izquierdo, F. Jiménez-Molinos, J. B. Roldán, and F. Campabadal. “Current transient response and role of the internal resistance in HfO_x-based memristors”. In: *Appl. Phys. Lett.* 117.26 (2020), p. 262902 (cit. on pp. 15, 16, 46, 51, 54, 55, 121).

[108] T. Diokh, E. Le-Roux, S.Jeannot, M. Gros-Jean, P. Candelier, J. F. Nodin, V. Jousseaume, L. Perniola, H. Grampeix, T. Cabout, E. Jalaguier, M. Guillermet, and B. De Salvo. “Investigation of the impact of the oxide thickness and RESET conditions on disturb in HfO_2-RRAM integrated in a 65nm CMOS technology”. In: *2013 IEEE International Reliability Physics Symposium (IRPS)* (2013), 5E.4.1–4 (cit. on pp. 15, 51, 55, 61, 121).

[109] Y. Nishi, S. Menzel, K. Fleck, U. Böttger, and R. Waser. “Origin of the SET Kinetics of the Resistive Switching in Tantalum Oxide Thin Films”. In: *IEEE Electron Device Lett.* 35.2 (2013), pp. 259–261 (cit. on pp. 15, 61, 121).

[110] Y. Nishi, K. Fleck, U. Böttger, R. Waser, and S. Menzel. “Effect of RESET Voltage on Distribution of SET Switching Time of Bipolar Resistive Switching in a Tantalum Oxide Thin Film”. In: *IEEE Trans. Electron Devices* 62.5 (2015), pp. 1561–1567 (cit. on pp. 15, 51, 121).

[111] Y. Nishi, U. Böttger, R. Waser, and S. Menzel. “Crossover From Deterministic to Stochastic Nature of Resistive-Switching Statistics in a Tantalum Oxide Thin Film”. In: *IEEE Trans. Electron Devices* (2018) (cit. on pp. 15, 105, 121, 164).

[112] A. Marchewka, B. Roesgen, K. Skaja, H. Du, C. L. Jia, J. Mayer, V. Rana, R. Waser, and S. Menzel. “Nanoionic Resistive Switching Memories: On the Physical Nature of the Dynamic Reset Process”. In: *Adv. Electron. Mater.* 2.1 (2016), pp. 1500233/1–13 (cit. on pp. 16, 17, 56, 65, 69, 92, 121, 129).

[113] A. Hardtdegen, C. La Torre, F. Cüppers, S. Menzel, R. Waser, and S. Hoffmann-Eifert. “Improved Switching Stability and the Effect of an Internal Series Resistor in HfO_2/TiO_x Bilayer ReRAM Cells”. In: *IEEE Trans. Electron Devices* 65.8 (2018), pp. 3229–3236 (cit. on pp. 16, 46, 51, 54, 56, 57, 60, 67, 69, 133).

[114] JART. *Juelich Aachen Resistive Switching Tools (JART)*. Tech. rep. 2019 (cit. on p. 17).

[115] C. Bengel, A. Siemon, F. Cüppers, S. Hoffmann-Eifert, A. Hardtdegen, M. von Witzleben, L. Hellmich, R. Waser, and S. Menzel. “Variability-Aware Modeling of Filamentary Oxide based Bipolar Resistive Switching Cells Using SPICE Level Compact Models”. In: *TCAS 1* 67.12 (2020), pp. 4618–4630 (cit. on pp. 17, 54, 133, 134, 139).

[116] C. Bengel, F. Cüppers, M. Payvand, R. Dittmann, R. Waser, S. Hoffmann-Eifert, and S. Menzel. "Utilizing the Switching Stochasticity of HfO_2/TiO_x-Based ReRAM Devices and the Concept of Multiple Devices for the Classification of Overlapping and Noisy Patterns". In: *Frontiers in Neuroscience* 15 (2021), p. 621 (cit. on pp. 17, 54, 122, 131, 133, 135, 137, 139, 140, 143, 145, 147, 149, 150, 156, 157, 160, 161).

[117] C. La Torre, A. F. Zurhelle, T. Breuer, R. Waser, and S. Menzel. "Compact Modeling of Complementary Switching in Oxide-Based ReRAM Devices". In: *IEEE Trans. Electron Devices* 66.3 (2019), pp. 1268–1275 (cit. on p. 17).

[118] A. Siemon, S. Menzel, R. Waser, and E. Linn. "A Complementary Resistive Switch-based Crossbar Array Adder". In: *IEEE J. Emerging Sel. Top. Circuits Syst.* 5.1 (2015), pp. 64 –74 (cit. on p. 17).

[119] Camilla La Torre. "Physics-Based Compact Modeling of Valence-Change-Based Resistive Switching Devices". PhD thesis. 2019 (cit. on p. 17).

[120] S. M. Sze and Kwok K. Ng. *Physics of Semiconductor Devices*. 3rd ed. Wiley, 2007 (cit. on p. 18).

[121] S. Choi, J. Yang, and G. Wang. "Emerging Memristive Artificial Synapses and Neurons for Energy-Efficient Neuromorphic Computing". In: *Adv. Mater.* 32.51 (2020), p. 2004659 (cit. on p. 20).

[122] Y. Li, Z. Wang, R. Midya, Q. Xia, and J. J. Yang. "Review of memristor devices in neuromorphic computing: materials sciences and device challenges". In: *J. Phys. D Appl. Phys.* 51 (2018), p. 503002 (cit. on p. 20).

[123] O. Krestinskaya, A. P. James, and L. O. Chua. "Neuromemristive Circuits for Edge Computing: A Review". In: *IEEE Trans. Neural Netw. Learn. Syst.* (2019), pp. 1–20 (cit. on p. 20).

[124] R. Dittmann and J. P. Strachan. "Redox-based memristive devices for new computing paradigm". In: *APL Materials* 7.11 (2019), pp. 110903/1–10 (cit. on p. 20).

[125] K. Roy, A. Jaiswal, and P. Panda. "Towards spike-based machine intelligence with neuromorphic computing". In: *Nature* 575.7784 (2019), pp. 607–617 (cit. on p. 20).

[126] C. Mead. "Neuromorphic electronic systems". In: *Proceedings of the IEEE* 78 (1990), pp. 1629–1636 (cit. on p. 20).

[127] Alexander LeNail. *Publication-ready NN-architecture schematics.* URL: `http://alexlenail.me/NN-SVG/index.html` (visited on 11/23/2022) (cit. on p. 21).

[128] J. Woo, D. Lee, Y. Koo, and H. Hwang. "Dual functionality of threshold and multilevel resistive switching characteristics in nanoscale HfO_2-based RRAM devices for artificial neuron and synapse elements". In: *Microelectron. Eng.* 182 (2017), pp. 42–45 (cit. on p. 22).

[129] A. Mehonic and A. J. Kenyon. "Emulating the Electrical Activity of the Neuron Using a Silicon Oxide RRAM Cell". In: *Front. Neurosci.* 10 (2016), pp. 57/1–10 (cit. on p. 22).

[130] C. Li, J. Ignowski, X. Sheng, R. Wessel, B. Jaffe, J. Ingemi, C. Graves, and J. P. Strachan. "CMOS-integrated nanoscale memristive crossbars for CNN and optimization acceleration". In: *2020 IEEE International Memory Workshop (IMW)*. 2020 IEEE International Memory Workshop (IMW), 2020, pp. 1–4 (cit. on p. 22).

[131] F. Cai, S. Kumar, T. Van Vaerenbergh, X. Sheng, R. Liu, C. Li, Z. Liu, M. Foltin, S. Yu, Q. Xia, J. J. Yang, R. Beausoleil, W. D. Lu, and J. P. Strachan. "Power-efficient combinatorial optimization using intrinsic noise in memristor Hopfield neural networks". In: *Nat. Electron.* 3.7 (2020), pp. 409–418 (cit. on p. 22).

[132] A. Mehonic, A. Sebastian, B. Rajendran, O. Simeone, E. Vasilaki, and A. J. Kenyon. "Memristors—From In-Memory Computing, Deep Learning Acceleration, and Spiking Neural Networks to the Future of Neuromorphic and Bio-Inspired Computing". In: *Advanced Intelligent Systems* n/a (2020), p. 2000085 (cit. on p. 22).

[133] S. Spiga, A. Sebastian, D. Querlioz, and B. Rajendran. *Memristive Device for Brain-Inspired Computing: From Materials, Devices, and Circuits to Applications - Computational Memory, Deep Learning and Spiking Neural Networks.* Ed. by B. Rajendran S. Spiga; A. Sebastian; D. Querlioz; 1st ed. 9780081027820. Woodheat Publishing, 2020 (cit. on p. 22).

[134] T Suntola and J Antson. "Finnish Patent No. 52395, 1974". In: *US Patent* 4058430 (1977) (cit. on p. 24).

[135] S. M. George. “Atomic Layer Deposition: An Overview”. In: *Chem. Rev.* 110.1 (2010), pp. 111–131 (cit. on p. 24).

[136] Markus Bosund, Emma M. Salmi, and Risto Peltonen. *Atomic layer deposition into ultra-high aspect ratio structures with a stop-flow ALD reactor*. Tech. rep. 2016 (cit. on p. 24).

[137] B. Abendroth, T. Moebus, S. Rentrop, R. Strohmeyer, M. Vinnichenko, T. Weling, H. Stöcker, and D. C. Meyer. “Atomic layer deposition of TiO_2 from tetrakis(dimethylamino) titanium and H_2O”. In: *Thin Solid Films* 545 (2013), pp. 176–182 (cit. on p. 24).

[138] H. Zhang, N. Aslam, M. Reiners, R. Waser, and S. Hoffmann-Eifert. “Atomic Layer Deposition of TiO_x/Al_2O_3 Bilayer Structures for Resistive Switching Memory Applications”. In: *Chemical Vapor Deposition* 20.7-9 (2014), pp. 282–290 (cit. on p. 24).

[139] Oxford Instruments. *Atomic Layer Deposition (ALD)*. URL: `https://plasma.oxinst.com/technology/atomic-layer-deposition` (visited on 11/23/2022) (cit. on p. 26).

[140] Alexander Tim Hardtdegen. “Engineering of HfO_2-based gradual resistive switching devices obtained from atomic layer deposited oxide bilayers”. PhD thesis. 2022 (cit. on pp. 25, 33).

[141] John E Mahan and André Vantomme. “A simplified collisional model of sputtering in the linear cascade regime”. In: *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* 15.4 (1997), pp. 1976–1989 (cit. on p. 29).

[142] Christian Lenser. “PhD thesis: Investigation of Resistive Switching in Fe-doped $SrTiO_3$ by Advanced Spectroscopy”. PhD thesis. 2013 (cit. on p. 32).

[143] Florian Lentz. “Integration of Redox-Based Resistive Switching Memory Devices”. PhD thesis. 2014 (cit. on p. 32).

[144] Nabeel Aslam. “Resistive switching memory devices from atomic layer deposited binary and ternary oxide thin films”. PhD thesis. 2017 (cit. on p. 32).

[145] Wonjoo Kim. “Investigation of switching mechanism in Ta_2O_5 -based ReRAM devices”. PhD thesis. 2017 (cit. on pp. 32, 33).

[146] Ivonne Bente. “Characterization of the TiO_x Interface in ReRAM Cells of Atomic Layer Grown Metal Oxide on Pure Titanium Metal”. In: *Master Thesis* (2020) (cit. on p. 33).

[147] L. G. Parratt. “Surface Studies of Solids by Total Reflection of X-Rays”. In: *Physical Review* 95.2 (1954), pp. 359–369 (cit. on p. 33).

[148] Keysight. *B1500A Semiconductor Device Parameter Analyzer*. URL: https://www.keysight.com/de/de/products/parameter-device-analyzers-curve-tracer/precision-current-voltage-analyzers/b1500a-semiconductor-device-parameter-analyzer.html (visited on 11/23/2022) (cit. on p. 35).

[149] T. Hennen, E. Wichmann, A. Elias, J. Lille, O. Mosendz, R. Waser, D.J. Wouters, and D. Bedau. “Current-limiting amplifier for high speed measurement of resistive switching data”. In: *Rev. Sci. Instrum.* 92 (2021), p. 054701 (cit. on pp. 35, 36).

[150] Tektronix. *Keithley 4200A-SCS Parameter Analyzer*. URL: https://www.tek.com/en/products/keithley/4200a-scs-parameter-analyzer (visited on 11/23/2022) (cit. on p. 39).

[151] ArC Instruments. *ArC ONE*. URL: https://www.arc-instruments.co.uk/products/arc-one/ (visited on 11/23/2022) (cit. on pp. 41, 42).

[152] J. Sandrini, L. Grenouillet, V. Meli, N. Castellani, I. Hammad, S. Bernasconi, F. Aussenac, S. Van Duijn, G. Audoit, M. Barlas, J. F. Nodin, O. Billoint, G. Molas, R. Fournel, E. Nowak, F. Gaillard, and C. Cagli. “OxRAM for embedded solutions on advanced node: scaling perspectives considering statistical reliability and design constraints”. In: *2019 IEEE International Electron Devices Meeting (IEDM)*. 2019 IEEE International Electron Devices Meeting (IEDM), 2019, pp. 30.5.1–30.5.4 (cit. on p. 45).

[153] R. Fackenthal, M. Kitagawa, W. Otsuka, K. Prall, D. Mills, K. Tsutsui, J. Javanifard, K. Tedrow, T. Tsushima, Y. Shibahara, and G. Hush. “A 16Gb ReRAM with 200MB/s write and 1GB/s read in 27nm technology,” in: *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*. 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014, pp. 338–339 (cit. on p. 45).

[154] A. Kawahara, K. Kawai, Y. Ikeda, Y. Katoh, R. Azuma, Y. Yoshimoto, K. Tanabe, Z. Wei, T. Ninomiya, . K. Katayama, R. Yasuhara, S. Muraoka, A. Himeno, N. Yoshikawa, H. Murase, K. Shimakawa, T. Takagi, T. Mikawa, and K. Aono. "Filament scaling forming technique and level-verify-write scheme with endurance over 10^7 cycles in ReRAM". In: *Int. Solid-State Circuit Conference (ISSCC),* (2012), pp. 220–221, (cit. on p. 45).

[155] M. Ueki, K. Takeuchi, T. Yamamoto, A. Tanabe, N. Ikarashi, M. Saitoh, T. Nagumo, H. Sunamura, M. Narihiro, K. Uejima, K. Masuzaki, N. Furutake, S. Saito, Y. Yabe, A. Mitsuiki, K. Takeda, T. Hase, and Y. Hayashi. "Low-Power Embedded ReRAM Technology for IoT Applications". In: *29th Symposium on VLSI Circuits, Kyoto, JAPAN*. 2015 Symposium On VLSI Circuits, 2015 (cit. on p. 45).

[156] A. Hardtdegen, C. La Torre, H. Zhang, C. Funck, S. Menzel, R. Waser, and S. Hoffmann-Eifert. "Internal Cell Resistance as the Origin of Abrupt Reset Behavior in HfO_2-based Devices determined from Current Compliance Series". In: *2016 IEEE 8th International Memory Workshop (IMW), Paris, France*. 2016 IEEE 8th International Memory Workshop (IMW), Paris, France, 2016, pp. 1–4 (cit. on pp. 46, 51).

[157] S. Wiefels, M. von Witzleben, M. Hüttemann, U. Böttger, R. Waser, and S. Menzel. "Impact of the Ohmic Electrode on the Endurance of Oxide Based Resistive Switching Memory". In: *IEEE Trans. Electron Devices* 68.3 (2021), pp. 1024 –1030 (cit. on p. 48).

[158] E. Perez, O. G. Ossorio, S. Duenas, H. Castan, H. Garcia, and C. Wenger. "Programming Pulse Width Assessment for Reliable and Low-Energy Endurance Performance in Al:HfO_2-Based RRAM Arrays". In: *ELECTRONICS* 9 (2020) (cit. on p. 48).

[159] K. M. Kim, J. J. Yang, J. P. Strachan, E. M. Grafals, N. Ge, N. D. Melendez, Z. Li, and R. S. Williams. "Voltage divider effect for the improvement of variability and endurance of TaO_x memristor". In: *Sci Rep* 6 (2016), pp. 20085/1–6 (cit. on pp. 48, 51, 53, 54, 56).

[160] C. La Torre, K. Fleck, S. Starschich, E. Linn, R. Waser, and S. Menzel. "Dependence of the SET switching variability on the initial state in HfO_x-based ReRAM". In: *Phys. Status Solidi A* 213.2 (2016), pp. 316–319 (cit. on pp. 51, 53, 61, 121).

[161] D. Ielmini and S. Menzel. “Universal Switching Behavior”. In: ed. by D. Ielmini; R. Waser. 1st ed. Wiley-VCH, 2016. Chap. 11, pp. 317–340 (cit. on pp. 51, 53, 56).

[162] K. Cico, P. Jancovic, J. Derer, V. Smatko, A. Rosova, M. Blaho, B. Hudec, D. Gregusova, and K. Fröhlich. “Resistive switching in nonplanar HfO_2-based structures with variable series resistance”. In: *J. Vac. Sci. Technol. B* 33.1 (2015), 1A108/1–5 (cit. on pp. 51, 56).

[163] C. Chen, L. Goux, A. Fantini, S. Clima, R. Degraeve, A. Redolfi, Y. Chen, G. Groeseneken, and M. Jurczak. “Endurance degradation mechanisms in $TiN/Ta_2O_5/Ta$ resistive random-access memory cells”. In: *Appl. Phys. Lett.* 106.5 (2015), pp. 053501/1–3 (cit. on p. 54).

[164] P. Huang, B. Chen, Y. Wang, F. Zhang, L. Shen, R. Liu, L. Zeng, G. Du, X. Zhang, B. Gao, J. Kang, X. Liu, X. Wang, B. Weng, Y. Tang, G. Lo, and D. Kwong. “Analytic model of endurance degradation and its practical applications for operation scheme optimization in metal oxide based RRAM”. In: *Electron Devices Meeting (IEDM), 2013 IEEE International*. Electron Devices Meeting (IEDM), 2013 IEEE International, 2013, pp. 22.5.1–22.5.4 (cit. on p. 55).

[165] W. Wu, H. Wu, B. Gao, N. Deng, S. Yu, and H. Qian. “Improving Analog Switching in HfO_x-Based Resistive Memory With a Thermal Enhanced Layer”. In: *IEEE Electron Device Lett.* 38.8 (2017), pp. 1019–1022 (cit. on p. 56).

[166] T. Stecconi, R. Guido, L. Berchialla, A. La Porta, J. Weiss, Y. Popoff, M. Halter, M. Sousa, F. Horst, D. Davila, U. Drechsler, R. Dittmann, B. J. Offrein, and V. Bragaglia. “Filamentary TaO_x/HfO_2 ReRAM Devices for Neural Networks Training with Analog In-Memory Computing”. In: *Adv. Electron. Mater.* 8.10 (2022), pp. 2200448/1–14 (cit. on p. 56).

[167] S. Park, B. Spetzler, T. Ivanov, and M. Ziegler. “Multilayer redox-based $HfO_x/Al_2O_3/TiO_2$ memristive structures for neuromorphic computing”. In: *Scientific Reports* 12 (2022), p. 18266 (cit. on p. 56).

[168] S. Fusi and L. F. Abbott. “Limits on the memory storage capacity of bounded synapses”. In: *Nat. Neurosci.* 10.4 (2007), pp. 485–493 (cit. on p. 66).

[169] N. F. Mott and R. W. Gurney. *Electronic processes in ionic crystals*. Oxford Univ. Press, London, U.K., 1948 (cit. on p. 67).

[170] A. R. Genreith-Schriever and R. A. De Souza. “Field-enhanced ion transport in solids: Reexamination with molecular dynamics simulations”. In: *Phys. Rev. B: Condens. Matter* 94.22 (2016), p. 224304 (cit. on p. 67).

[171] S. Agarwal, S. J. Plimpton, D. R. Hughart, A. H. Hsia, I. Richter, J. A. Cox, C. D. James, and M. J. Marinella. “Resistive Memory Device Requirements for a Neural Algorithm Accelerator”. In: *International Joint Conference on Neural Networks (IJCNN), Vancouver, CANADA*. 2016 International Joint Conference On Neural Networks (ijcnn), 2016, pp. 929–938 (cit. on p. 73).

[172] M. Le Gallo, A. Sebastian, R. Mathis, M. Manica, H. Giefers, T. Tuma, C. Bekas, A. Curioni, and E. Eleftheriou. “Mixed-precision in-memory computing”. In: *Nat. Electron.* 1 (2018), pp. 246–253 (cit. on p. 73).

[173] P. M. Sheridan, F. Cai, Ch. Du, W. Ma, Z. Zhang, and W. D. Lu. “Sparse coding with memristor networks”. In: *Nat. Nanotechnol.* 12 (2017), pp. 784–789 (cit. on p. 73).

[174] A. Sebastian, T. Tuma, N. Papandreou, M. Le Gallo, L. Kull, T. Parnell, and E. Eleftheriou. “Temporal correlation detection usingcomputational phase-change memory”. In: *Nature Communications* 8 (2017), pp. 1–10 (cit. on p. 73).

[175] G. Pedretti, S. Serebryakov, J. P. Strachan, and C. E. Graves. “A general tree-based machine learning accelerator with memristive analog CAM”. In: *2022 IEEE International Symposium on Circuits and Systems (ISCAS)* (2022), pp. 220–224 (cit. on p. 73).

[176] S. Brivio, J. Frascaroli, E. Covi, and S. Spiga. “Stimulated Ionic Telegraph Noise in Filamentary Memristive Devices”. In: *Sci. Rep.* 9 (2019), p. 6310 (cit. on pp. 73, 77, 87, 88, 98–100).

[177] R. Mao, B. Wen, Y. Zhao, A. Kazemi, A. F. Laguna, M. Neimier, X.S. Hu, X. Sheng, C.E. Graves, J.P. Strachan, and C. Li. “Experimentally realized memristive memory augmented neural network”. In: *https://arxiv.org/abs/2204.07429* (2022) (cit. on pp. 73, 99, 100).

[178] S. Wiefels, C. Bengel, N. Kopperberg, K. Zhang, R. Waser, and S. Menzel. “HRS Instability in Oxide based Bipolar Resistive Switching Cells”. In: *IEEE Trans. Electron Devices* 67.10 (2020), pp. 4208–4215 (cit. on pp. 73, 99, 100, 102, 125, 128, 136).

[179] E. Perez, M. K. Mahadevaiah, E. P. Quesada, and C. Wenger. “In-depth characterization of switching dynamics in amorphous HfO_2 memristive arrays for the implementation of synaptic updating rules”. In: *Jpn. J. Appl. Phys.* 61 (2022), SM1007 (cit. on pp. 73, 99, 100).

[180] C. La Torre, A. Kindsmueller, D. J. Wouters, C. E. Graves, G. A. Gibson, J. P. Strachan, R. S. Williams, R. Waser, and S. Menzel. “Volatile HRS asymmetry and subloops in resistive switching oxides”. In: *Nanoscale* 9 (2017), pp. 14414–14422 (cit. on p. 98).

[181] K. Schnieders, C. Funck, F. Cüppers, S. Aussen, T. Kempen, A. Sarantopoulos, R. Dittmann, S. Menzel, V. Rana, S. Hoffmann-Eifert, and S. Wiefels. “Effect of electron conduction on the read noise characteristics in ReRAM devices”. In: *Appl. Phys. Lett.* 10.10 (2022), p. 101114 (cit. on p. 98).

[182] D. O. Hebb. *The Organization of Behavior*. New York: Wiley & Sons, 1949 (cit. on p. 105).

[183] S. A. Chekol, F. Cüppers, R. Waser, and S. Hoffmann-Eifert. “An Ag/HfO_2/Pt Threshold Switching Device with an Ultra-Low Leakage (< 10 fA), High On/OffRatio ($> 10^{11}$), and Low Threshold Voltage (< 0.2 V) for Energy-Efficient Neuromorphic Computing”. In: *2021 IEEE International Memory Workshop (IMW)*. IEEE International Memory Workshop (IMW), 2021 (cit. on p. 105).

[184] *Integrate and Fire Neurons Based on Diffusive Memristors for Spiking Neural Networks*. International Conference on Neuromorphic Systems (ICONS), July 27th-29th, 2021, 2021 (cit. on p. 105).

[185] S. A. Chekol, S. Menzel, R. W. Ahmad, R. Waser, and S. Hoffmann-Eifert. “Effect of the Threshold Kinetics on the Filament Relaxation Behavior of Ag-Based Diffusive Memristors”. In: *Adv. Funct. Mater.* 32.15 (2022), pp. 2111242/1–11 (cit. on p. 105).

[186] E. Covi, R. George, J. Frascaroli, S. Brivio, C. Mayr, H. Mostafa, G. Indiveri, and S. Spiga. “Spike-driven threshold-based learning with memristive synapses and neuromorphic silicon neurons”. In: *J. Phys. D Appl. Phys.* 51.34 (2018), p. 344003 (cit. on p. 105).

[187] R. Midya, Z. Wang, S. Asapu, S. Joshi, Y. Li, Y. Zhuo, W. Song, H. Jiang, N. Upadhay, M. Rao, P. Lin, C. Li, Q. Xia, and J. J. Yang. "Artificial Neural Network (ANN) to Spiking Neural Network (SNN) Converters Based on Diffusive Memristors". In: *Adv. Electron. Mater.* 5 (2019), p. 1900060 (cit. on p. 105).

[188] Ye Zhuo, Rivu Midya, Wenhao Song, Zhongrui Wang, Shiva Asapu, Mingyi Rao, Peng Lin, Hao Jiang, Qiangfei Xia, R. Stanley Williams, and J. Joshua Yang. "A Dynamical Compact Model of Diffusive and Drift Memristors for Neuromorphic Computing". In: *Adv. Electron. Mater.* (2021) (cit. on p. 105).

[189] Moritz von Witzleben. "Switching kinetics of valence change memory devices on a sub-100 ps timescale". PhD thesis. 2021 (cit. on p. 121).

[190] S. Koveshnikov, K. Matthews, K. Min, D. Gilmer, M. Sung, S. Deora, H. Li, S. Gausepohl, P. Kirsch, and R. Jammy. "Real-time study of switching kinetics in integrated 1T/ HfO_x 1R RRAM: Intrinsic tunability of set/reset voltage and trade-off with switching time". In: *Technical Digest - International Electron Devices Meeting, IEDM.* Technical Digest - International Electron Devices Meeting, IEDM, 2012, pp. 20.4.1–20.4.3 (cit. on p. 121).

[191] D. Ielmini, F. Nardi, and S. Balatti. "Evidence for Voltage-Driven Set/Reset Processes". In: *IEEE Trans. Electron Devices* 59 (2012), pp. 2049–2055 (cit. on p. 121).

[192] S. Yu, Y. Wu, and H. Wong. "Investigating the switching dynamics and multilevel capability of bipolar metal oxide resistive switching memory". In: *Appl. Phys. Lett.* 98.10 (2011), pp. 103514/1–3 (cit. on p. 121).

[193] M. G. Cao, Y. S. Chen, J. R. Sun, D. S. Shang, L. F. Liu, J. F. Kang, and B. G. Shen. "Nonlinear dependence of set time on pulse voltage caused by thermal accelerated breakdown in the $Ti/HfO_2/Pt$ resistive switching devices". In: *Appl. Phys. Lett.* 101.20 (2012), p. 203502 (cit. on p. 121).

[194] F. Alibart, L. Gao, B. D. Hoskins, and D. B. Strukov. "High precision tuning of state for memristive devices by adaptable variation-tolerant algorithm". In: *Nanotechnology* 23.7 (2012), pp. 75201/1–7 (cit. on p. 121).

[195] G. Cao, X. Yan, J. Wang, Z. Zhou, J. Lou, and K. Wang. "Realization of fast switching speed and electronic synapse in $Ta/TaO_x/AlN/Pt$ bipolar resistive memory". In: *AIP Adv.* 10.5 (2020), p. 055312 (cit. on p. 121).

[196] C.Y. Chen, L. Goux, A. Fantini, R. Degraeve, A. Redolfi, G. Groeseneken, and M. Jurczak. “Stack optimization of oxide-based RRAM for fast write speed (<1 µs) at low operating current (<10 µA)”. In: *J.SSE* (2016) (cit. on p. 121).

[197] Peng Huang, Yijiao Wang, Haitong Li, Bin Gao, Bing Chen, Feifei Zhang, Lang Zeng, Gang Du, Jinfeng Kang, and Xiaoyan Liu. “Analysis of the voltage-time dilemma of metal oxide-based RRAM and solution exploration of high speed and low voltage ac switching”. In: *IEEE Trans. Nanotechnol.* 13.6 (2014), pp. 1127–1132 (cit. on p. 121).

[198] W.-C. Luo, J.-C. Liu, Y.-C. Lin, C.-L. Lo, J.-J. Huang, K.-L. Lin, and T.-H. Hou. “Statistical model and rapid prediction of RRAM SET speed-disturb dilemma”. In: *IEEE Trans. Electron Devices* 60.11 (2013), pp. 3760–3766 (cit. on p. 121).

[199] S. Wiefels, U. Böttger, S. Menzel, D. J. Wouters, and R. Waser. “Empirical Tunneling Model Describing the Retention of 2.5 Mb HfO_2 based ReRAM”. In: *International Symposium On VLSI Technology, Systems and Application (VLSI-TSA)*. IEEE, 2020, pp. 37–38 (cit. on p. 125).

[200] N. Kopperberg, S. Wiefels, S. Liberda, R. Waser, and S. Menzel. “A Consistent Model for Short-Term Instability and Long-Term Retention in Filamentary Oxide-Based Memristive Devices”. In: *ACS Appl. Mater. Interfaces* 13.48 (2021), pp. 58066–58075 (cit. on p. 125).

[201] X. Huang, H. Wu, B. Gao, D. C. Sekar, L. Dai, M. Kellam, G. Bronner, N. Deng, and H. Qian. “HfO_2/Al_2O_3 multilayer for RRAM arrays: a technique to improve tail-bit retention”. In: *Nanotechnology* 27.39 (2016), pp. 395201/1–6 (cit. on p. 125).

[202] F.M. Puglisi, L. Larcher, A. Padovani, and P. Pavan. “A Complete Statistical Investigation of RTN in HfO_2-Based RRAM in High Resistive State”. In: *IEEE Trans. Electron Devices* 62.8 (2015), pp. 2606–2613 (cit. on pp. 125, 128).

[203] F.M. Puglisi, L. Larcher, P. Pavan, A. Padovani, and G. Bersuker. “Instability of HfO_2 RRAM devices: comparing RTN and cycling variability”. In: *2014 IEEE International Reliability Physics Symposium (IRPS)*. 2014 IEEE International Reliability Physics Symposium (IRPS), 2014, MY.5.1–MY.5.5 (cit. on p. 128).

[204] D. Veksler, G. Bersuker, B. Chakrabarti, E. Vogel, S. Deora, K. Matthews, D. C. Gilmer, H. F. Li, S. Gausepohl, and P. D. Kirsch. “Methodology for the statistical evaluation of the effect of random telegraph noise (RTN) on RRAM characteristics”. In: *IEDM* (2012), pp. 219–222 (cit. on p. 128).

[205] S. H. Jo, K. H. Kim, and W. Lu. “Programmable Resistance Switching in Nanoscale Two-Terminal Devices”. In: *Nano Lett.* 9.1 (2009), pp. 496–500 (cit. on p. 132).

[206] M. Suri, D. Querlioz, O. Bichler, G. Palma, E. Vianello, D. Vuillaume, C. Gamrat, and B. DeSalvo. “Bio-Inspired Stochastic Computing Using Binary CBRAM Synapses”. In: *IEEE Trans. Electron Devices* 60.7 (2013), pp. 2402–2409 (cit. on p. 132).

[207] F. Zahari, E. Pérez, M. K. Mahadevaiah, H. Kohlstedt, C. Wenger, and M. Ziegler. “Analogue pattern recognition with stochastic switching binary CMOS-integrated memristive devices”. In: *Scientific Reports* 10.1 (2020), p. 14450 (cit. on p. 132).

[208] S. Gaba, P. Sheridan, J. Zhou, S. Choi, and W. Lu. “Stochastic memristive devices for computing and neuromorphic applications”. In: *Nanoscale* 5.13 (2013), pp. 5872–5878 (cit. on pp. 132, 164).

[209] J. Bill and R. Legenstein. “A compound memristive synapse model for statistical learning through STDP in spiking neural networks”. In: *Frontiers in Neuroscience* 8 (2014), pp. 412/1– (cit. on pp. 132, 164).

[210] M. Hu, Y. Wang, Q. Qiu, Y. Chen, and H. Li. “The Stochastic Modeling of TiO_2 Memristor and Its Usage in Neuromorphic System Design”. In: *19th Asia and South Pacific Design Automation Conference (ASP-DAC), Suntec, SINGAPORE*. 2014 19th Asia and South Pacific Design Automation Conference (asp-Dac), 2014, pp. 831–836 (cit. on p. 132).

[211] A. Singha, B. Muralidharan, and B. Rajendran. “Analog memristive time dependent learning using discrete nanoscale RRAM devices”. In: *2014 International Joint Conference on Neural Networks (IJCNN)*. 2014, pp. 2248–2255 (cit. on pp. 132, 164).

[212] I. Boybat, C. Giovinazzo, E. Shahrabi, I. Krawczuk, J. Giannopoulos, C. Piveteau, M. Le Gallo, C. Ricciardi, A. Sebastian, E. Eleftheriou, and Y. Leblebici. "Multi-ReRAM synapses for artificial neural network training". In: *IEEE International Symposium on Circuits and Systems (IEEE ISCAS), Sapporo, JAPAN*. 2019 IEEE International Symposium On Circuits and Systems (ISCAS), 2019 (cit. on p. 132).

[213] Melika Payvand, Lorenz K. Muller, and Giacomo Indiveri. "Event-based circuits for controlling stochastic learning with memristive devices in neuromorphic architectures". In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS)* (2018) (cit. on pp. 132, 152).

[214] M. Payvand, M. V. Nair, L. K. Mueller, and G. Indiveri. "A neuromorphic systems approach to in-memory computing with non-ideal memristive devices: from mitigation to exploitation". In: *Faraday Discuss.* 213 (2019), pp. 487–510 (cit. on p. 132).

[215] I. Boybat, M. Le Gallo, S. R. Nandakumar, T. Moraitis, T. Parnell, T. Tuma, B. Rajendran, Y. Leblebici, A. Sebastian, and E. Eleftheriou. "Neuromorphic computing with multi-memristive synapses". In: *Nature Communications* 9 (2018), p. 2514 (cit. on pp. 132, 164).

[216] G. Medeiros-Ribeiro, F. Perner, R. Carter, H. Abdalla, M. D. Pickett, and R. S. Williams. "Lognormal switching times for titanium dioxide bipolar memristors: origin and resolution". In: *Nanotechnology* 22.9 (2011), pp. 95702/1–5 (cit. on p. 132).

[217] C. Baeumer, R. Valenta, C. Schmitz, A. Locatelli, T. O. Mentes, S. P. Rogers, A. Sala, N. Raab, S. Nemsak, M. Shim, C. M. Schneider, S. Menzel, R. Waser, and R. Dittmann. "Subfilamentary Networks Cause Cycle-to-Cycle Variability in Memristive Devices". In: *ACS Nano* 11.7 (2017), pp. 6921–6929 (cit. on p. 133).

[218] A. Fantini, G. Gorine, R. Degraeve, L. Goux, C. Chen, A. Redolfi, S. Clima, A. Cabrini, G. Torelli, and M. Jurczak. "Intrinsic program instability in HfO_2 RRAM and consequences on program algorithms". In: *2015 IEEE International Electron Devices Meeting (IEDM)*. 2015 International Electron Devices Meeting (IEDM), 2015, pp. 7.5.1–7.5.4 (cit. on p. 136).

[219] Lorenz K. Müller and Giacomo Indiveri. “Rounding Methods for Neural Networks with Low Resolution Synaptic Weights”. In: *CoRR* abs/1504.05767 (2015). arXiv: 1504.05767. URL: http://arxiv.org/abs/1504.05767 (cit. on p. 152).

[220] S. Gupta, A. Agrawal, K. Gopalakrishnan, and P. Narayanan. “Deep Learning with Limited Numerical Precision”. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. 2015, 1737–1746 (cit. on p. 153).

[221] T. Gokmen, M. J. Rasch, and W. Haensch. “Training LSTM Networks With Resistive Cross-Point Devices”. In: *Frontiers in Neuroscience* 12 (2018), p. 745 (cit. on p. 153).

[222] J. H. Yoon, J. H. Han, J. S. Jung, W. Jeon, G. H. Kim, S. J. Song, J. Y. Seok, K. J. Yoon, M. H. Lee, and C. S. Hwang. “Highly Improved Uniformity in the Resistive Switching Parameters of TiO_2 Thin Films by Inserting Ru Nanodots”. In: *Adv. Mater.* 25.14 (2013), pp. 1987–1992 (cit. on p. 164).

[223] J. L. Rieck, F. V. E. Hensling, and R. Dittmann. “Trade-off between variability and retention of memristive epitaxial $SrTiO_3$ devices”. In: *APL Mater.* 9.2 (2021), pp. 21110/1–7 (cit. on p. 164).

List of Figures

List of Tables

Appendix

Fabrication of 40 nm x 40 nm nanoplug devices

The following recipe for nanoplug devices was developed by Solomon Chekol of the PGI-7 at Forschungszentrum Jülich.

Starting substrate: 20 cm x 20 cm Si/430 nm SiO_2. Fabrication steps:

1. Bottom electrode metal deposition: Sputtering of 5 nm Ta and 25 nm Pt;
2. Cleaning: 3 minutes of ultrasonic bath in Acetone at power level 3 followed by 3 minutes of ultrasonic bath in isopropyl alcohol at power level 3, drying with N_2 gas;
3. Bottom electrode lithography:
   (a) Substrate dehydration at 120 °C for 3 min;
   (b) Spincoating of (1:2) diluted AZ nLof 2020 resist at 4000 rpm for 45 s;
   (c) Resist baking at 90 °C for 3 min;
   (d) Electron beam exposure at 100 kV using mask `"True_Planar_Device_L1"`;
   (e) Post-exposure baking at 110 °C for 3 min;
   (f) Development in AZ 726 MIF developer for 1 min, stopped in deionized water, drying with N_2 gas.
4. RIBE etching using Argon process gas.
5. Resist removal by DMSO at 80 °C for minimum of 3 hours, followed by 3 min ultrasonic bath at power level 5. Swabbing in Acetone and optional plasma ashing.
6. Deposition of the insulation layer of 20 nm SiO_2 by plasma enhanced ALD.

7. Nano hole lithography:
   (a) Spincoating of AllResist CSAR 6200.04 resist at 4000 rpm for 45 s;
   (b) Resist baking at 150 °C for 1 min;
   (c) Electron beam exposure at 100 kV using mask `"True_Planar_Device_L2"`;
   (d) Development in AR 600-55 developer for 1 min, stopped in isopropyl alcohol, drying with N_2 gas.
8. RIBE etching using CF_4 process gas.
9. Resist removal by AR 600-71 for minimum of 30 min, followed by acetone, isopropyl alcohol and deionized water bath.
10. Deposition of the switching layers of 3 nm HfO_2 by plasma enhanced ALD and 3 nm TiO_x by thermal ALD.
11. Top electrode lithography:
   (a) Spincoating of AllResist CSAR 6200.04 resist at 4000 rpm for 45 s;
   (b) Resist baking at 150 °C for 1 min;
   (c) Electron beam exposure at 100 kV using mask `"True_Planar_Device_L3"`;
   (d) Development in AR 600-55 developer for 1 min, stopped in isopropyl alcohol, drying with N_2 gas.
12. Deposition of the top electrode metal layers of 10 nm Ti and 20 nm Pt by sputtering.
13. Lift-off by AR 600-71 over night, followed by acetone, isopropyl alcohol and deionized water bath.
14. Contact pad opening lithography:
   (a) Substrate dehydration at 120 °C for 3 min;
   (b) Spincoating of (1:2) diluted AZ nLof 2020 resist at 4000 rpm for 45 s;
   (c) Resist baking at 90 °C for 3 min;
   (d) Electron beam exposure at 100 kV using mask `"True_Planar_Device_L4"`;
   (e) Post-exposure baking at 110 °C for 3 min;

(f) Development in AZ 726 MIF developer for 1 min, stopped in deionized water, drying with N_2 gas.

15. RIBE etching using CF_4 process gas.

16. Resist removal by DMSO at 80 °C for minimum of 3 hours, followed by 3 min ultrasonic bath at power level 5. Swabbing in Acetone.

Danksagungen

An dieser Stelle möchte ich mich bei allen Menschen bedanken, die mich vor und während der Erstellung dieser Dissertation begleitet und unterstützt haben. Besonders bedanken möchte ich mich bei...

- ... Prof. Dr. Rainer Waser, Leiter des PGI-7 und Doktorvater, der mir die spannende Forschung an ReRAM-Zellen ermöglicht und diese Arbeit betreut hat.
- ... Prof. Dr. Joachim Knoch als Zweitprüfer dieser Arbeit.
- ... Dr. Susanne Hoffmann-Eifert, Gruppenleiterin und Betreuerin dieser Doktorarbeit. Ohne Ihre wissenschaftliche Expertise und Erfahrung wären viele Ergebnisse dieser Arbeit nicht in der jetzigen Form zustande gekommen. Auch für Ihre unermüdlichen Ermunterungen Ergebnisse zu hinterfragen, ihren steten Einsatz zur Verbesserung des Verständnisses und den Einsatz Ihrer scharfen Beobachtungsgabe möchte ich mich bedanken.
- ... meinen Kollegen und Freunden, allen voran Dr. Alexander Hardtdegen, der mir sowohl währenddessen als auch über seine Zeit am PGI-7 hinaus unentwegt mit Rat und Tat zur Seite stand. Außerdem möchte ich mich bei Stephan Außen, Solomon Chekol, Zhaodong Wang, Hassan Sultani, Oliver Solfronk und Dr. Hehe Zhang für die allzeit gute Atmosphäre in der Gruppe bedanken. Bei Dr. Alexander Gutsche, Niclas Schmidt, Dr. Felix Hensling, Dr. René Ebeling und Dr. Jaqueline Börgers möchte ich mich für die Momente bedanken, bei denen man Arbeit nicht von Spaß unterscheiden konnte.
- ... bei den Studenten, deren Praktika und Abschlussarbeiten einen Beitrag zu dieser Arbeit geleistet haben, insbesondere bei Nils Quiring, Ivonne Bente, Jaime Raichs-Lopez, Tim Thesing und Markus Bauckhage.
- ... bei Christopher Bengel und Dr. Stephan Menzel für Ihre andauernde Kooperation beim Vergleich der Messdaten mit Modellen, ohne den viele Erkenntnisse dieser Arbeit verschlossen geblieben wären.
- ... bei René Borowski, Grigory Potemkin, Benjamin Bennemann, Clemens Wiedenhöft, Marcel Gerst, Malte Deckers und Jochen Friedrich für Ihre Unterstützung bei den diversen technischen Herausforderungen dieser Arbeit.

- ... bei meinen Freunden in Aachen sowie in der Heimat und der ganzen Welt für Ihre Unterstützung außerhalb der Arbeit.
- ... bei meiner Partnerin Camilla Winkler für Ihre Unterstützung, Gehör und aufmunternden Worte in schwierigen Zeiten.
- ... bei meinen Großeltern, Eltern und Brüdern, ohne deren anhaltende Unterstützung, Ermunterung und Bestätigung diese Arbeit unerdenklich gewesen wäre. Vielen Dank!

Schlussendlich möchte ich mich bei der gesamten Belegschaft des PGI-7, PGI-10 und IWE-2 für die gute Arbeitsatmosphäre und die Ermöglichung der wissenschaftlichen Entfaltung danken.

Band / Volume 83
Hybrid hydrogels promote physiological functionality of long-term cultured primary neuronal cells in vitro
C. Meeßen (2022), x, 120 pp
ISBN: 978-3-95806-643-4

Band / Volume 84
Surface states and Fermi-level pinning on non-polar binary and ternary (Al,Ga)N surfaces
L. Freter (2022), 137 pp
ISBN: 978-3-95806-644-1

Band / Volume 85
Dynamical and statistical structure of spatially organized neuronal networks
M. Layer (2022), xiii, 167 pp
ISBN: 978-3-95806-651-9

Band / Volume 86
Persistent firing and oscillations in the septo-hippocampal systemand their relation to locomotion
K. Korvasová (2022), 111 pp
ISBN: 978-3-95806-654-0

Band / Volume 87
Sol-Gel-Synthese, Tintenstrahldruck und Blitzlampentemperung von Tantaloxid-Dünnschichten zur pH-Messung
C. D. Beale (2022), xlix, 339 pp
ISBN: 978-3-95806-656-4

Band / Volume 88
Diversity of chiral magnetic solitons
V. Kuchkin (2022), xiv, 155 pp
ISBN: 978-3-95806-665-6

Band / Volume 89
Controlling the electrical properties of oxide heterointerfaces through their interface chemistry
M.-A. Rose (2022), vi, 162 pp
ISBN: 978-3-95806-667-0

Band / Volume 90
Modeling and Suppressing Unwanted Parasitic Interactions in Superconducting Circuits
X. Xu (2022), 123, XVIII pp
ISBN: 978-3-95806-671-7

Band / Volume 91
Activating molecular magnetism by controlled on-surface coordination
I. Cojocariu (2022), xi, 169 pp
ISBN: 978-3-95806-674-8

Band / Volume 92
Computational study of structural and optical properties of two-dimensional transition-metal dichalcogenides with implanted defects
S. H. Rost (2023), xviii, 198 pp
ISBN: 978-3-95806-682-3

Band / Volume 93
DC and RF characterization of bulk CMOS and FD-SOI devices at cryogenic temperatures with respect to quantum computing applications
A. Artanov (2023), xv, 80, xvii-liii pp
ISBN: 978-3-95806-687-8

Band / Volume 94
HAXPES study of interface and bulk chemistry of ferroelectric HfO_2 capacitors
T. Szyjka (2023), viii, 120 pp
ISBN: 978-3-95806-692-2

Band / Volume 95
A brain inspired sequence learning algorithm and foundations of a memristive hardware implementation
Y. Bouhadjar (2023), xii, 149 pp
ISBN: 978-3-95806-693-9

Band / Volume 96
Characterization and modeling of primate cortical anatomy and activity
A. Morales-Gregorio (2023), ca. 260 pp.
ISBN: 978-3-95806-698-4

Band / Volume 97
Hafnium oxide based memristive devices as functional elements of neuromorphic circuits
F. J. Cüppers (2023), vi, ii, 214 pp
ISBN: 978-3-95806-702-8

Information
Band / Volume 97
ISBN 978-3-95806-702-8

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum