

Filamentary TaO_x/HfO₂ ReRAM Devices for Neural Networks Training with Analog In-Memory Computing

Tommaso Stecconi,* Roberto Guido, Luca Berchialla, Antonio La Porta, Jonas Weiss, Youri Popoff, Mattia Halter, Marilyne Sousa, Folkert Horst, Diana Dávila, Ute Drechsler, Regina Dittmann, Bert Jan Offrein, and Valeria Bragaglia

The in-memory computing paradigm aims at overcoming the intrinsic inefficiencies of Von-Neumann computers by reducing the data-transport per arithmetic operation. Crossbar arrays of multilevel memristive devices enable efficient calculations of matrix-vector-multiplications, an operation extensively called on in artificial intelligence (AI) tasks. Resistive random-access memories (ReRAMs) are promising candidate devices for such applications. However, they generally exhibit large stochasticity and device-to-device variability. The integration of a sub-stoichiometric metal-oxide within the ReRAM stack can improve the resistive switching graduality and stochasticity. To this purpose, a conductive TaO_x layer is developed and stacked on HfO₂ between TiN electrodes, to create a complementary metal-oxide-semiconductor-compatible ReRAM structure. This device shows accumulative conductance updates in both directions, as required for training neural networks. Moreover, by reducing the TaO_x thickness and by increasing its resistivity, the device resistive states increase, as required for reduced power consumption. An electric field-driven TaO_x oxidation/reduction is responsible for the ReRAM switching. To demonstrate the potential of the optimized TaO_x/HfO₂ devices, the training of a fully-connected neural network on the Modified National Institute of Standards and Technology database dataset is simulated and benchmarked against a full precision digital implementation.

(AI) exploits large datasets to learn how to solve a broad variety of tasks such as text/image recognition, data clustering, equity pricing, and many others^[3]. However, computers based on the Von-Neumann architecture are inefficient in the execution of modern AI workloads, and the required power becomes unsustainable due to the growing volume of data involved.

The bottleneck for performance derives from the *memory wall*. The bandwidth of processors has grown faster than the memory one, with the difference between them diverging exponentially.^[4] The energetic inefficiency originates to a large extent from energy-hungry accessing of off-chip memory.^[5] With the end of Moore's law approaching^[6] and the rise of AI, the quest for alternative computing paradigms has intensified.

In-memory computing has emerged as a promising hardware architecture for data-centric applications. It consists of memory cells interconnected in a specific design to locally execute logic or arithmetical operations.^[7] The crossbar array is an example of an in-memory computing

architecture that raised a strong interest in the electron devices community. In this architecture, the memory elements are positioned at the crosspoint between bitlines and wordlines,^[8] as depicted in **Figure 1a**. When a vector of read voltages (*V*) is applied on the array elements, the currents (*I*) establishing at

1. Introduction

The volume of data created by modern society grows rapidly, driven by the increase of digitalization across a wide range of business and consumer applications.^[1,2] Artificial Intelligence

T. Stecconi, R. Guido, L. Berchialla, A. La Porta, J. Weiss, Y. Popoff, M. Halter, M. Sousa, F. Horst, D. Dávila, U. Drechsler, B. J. Offrein, V. Bragaglia
 IBM Research GmbH-Zurich Research Laboratory
 Rüschlikon CH-8803, Switzerland
 E-mail: tec@zurich.ibm.com

Y. Popoff, M. Halter
 Integrated Systems Laboratory
 ETH Zurich
 Zurich CH-8092, Switzerland
 R. Dittmann
 Peter Grünberg Institute
 Forschungszentrum Jülich GmbH
 52425 Jülich, Germany
 R. Dittmann
 JARA-FIT
 RWTH Aachen University
 52056 Aachen, Germany

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aelm.202200448>.

© 2022 The Authors. Advanced Electronic Materials published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aelm.202200448

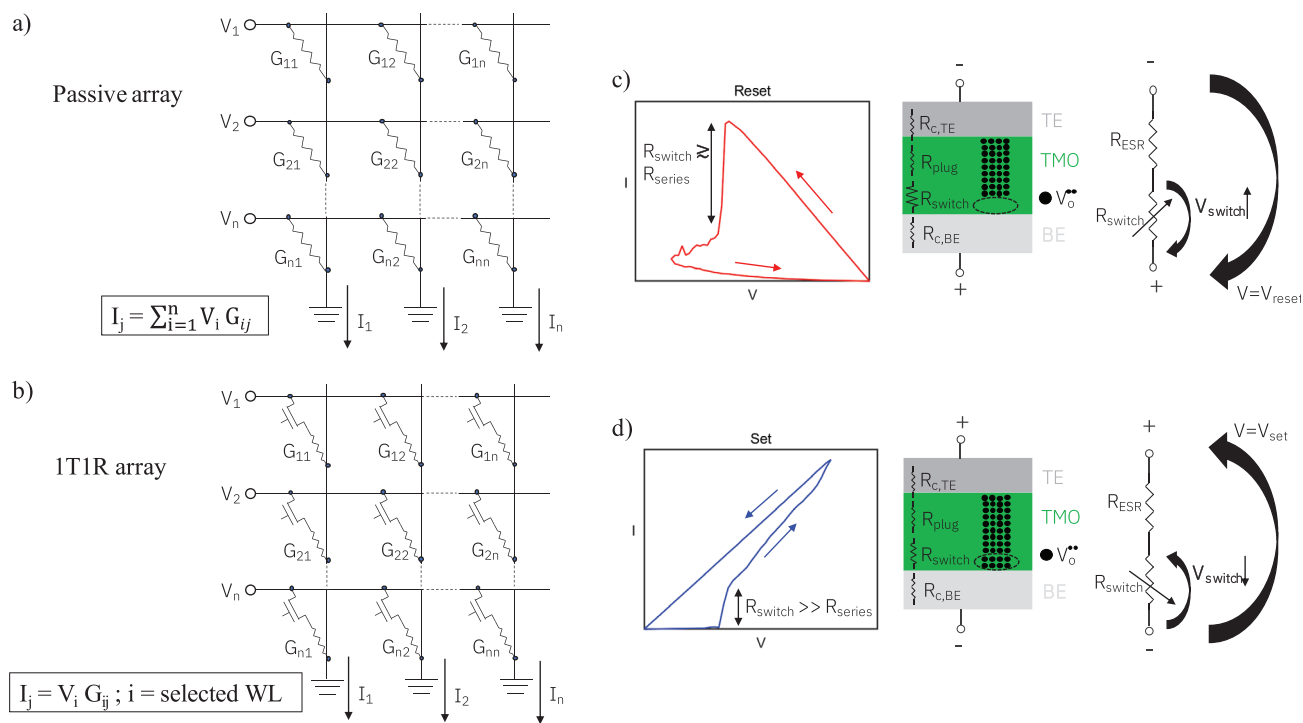


Figure 1. a,b) Circuit schematics of a passive and of a 1T1R crossbar array, respectively. c) *Reset* process, involving the depletion of V_o from a filament region next to a TMO/electrode interface. At $V = V_{\text{reset}}$, the voltage divider on the depleted region increases as long as its equivalent resistance R_{switch} dynamically increases, resulting in an abrupt *reset*. d) *Set* process, determined by the growth of the V_o density in the formerly depleted filament region. The device resistance drops abruptly, until the voltage divider on the ESR becomes relevant.

the bitlines are proportional to the array elements' conductances (G), resulting in a matrix-vector multiplication (MVM), as dictated by Ohm's and Kirchoff's law.^[9,10]

In-memory computing with crossbar arrays shifts the computing paradigm, from a digital implementation, conventionally given by blocks of digital adders,^[11] to the analog domain. This can be particularly beneficial for the inference and training of artificial neural networks (ANNs), as the synaptic operations can be directly mapped into the crossbar array functionality. Since the algorithms developed for inference and training of ANNs extensively use MVM operations,^[3] the performance and efficiency advantage of analog signal processing in crossbar arrays is based on a more favorable scaling of the computing effort with increasing number of neurons involved, and correspondingly reduced memory access.

First the demonstrators of learning tasks were realized with passive arrays^[12] (Figure 1a). Nevertheless, energy efficient and accurate programming of the single cells is limited in passive arrays by the current “sneakpaths”, especially for large arrays. By integrating a series transistor, defining 1-transistor 1-resistor (1T1R) array cells (Figure 1b), ANNs with larger arrays can be implemented to perform more complex character/image recognition tasks.^[13,14]

Crosspoint devices must provide a tunable G for ANNs training. *Memristors* are electrical elements recently developed,^[15] providing pinched current-voltage (I - V) hysteretic characteristics.^[16,17] This property reflects into a tunable G upon electrical stimuli, where the addressed G remains over time, representing a multilevel non-volatile memory.

Among a wide range of available *memristor* technologies, the two-terminal phase change memory (PCM) and resistive random-access memory (ReRAM) have raised vivid interest. Both can be integrated in dense arrays, are scalable, support fast write/read operations, multi-level cell (MLC) programming, low-power operations, have high-endurance, and are CMOS compatible.^[14,18–20] PCMs are characterized by G updates with accumulative behavior only in the direction of lower resistances,^[21] making them well-suited for ANNs inference applications. For the training of neural networks, where, instead, granular monotonic G updates are required in both directions,^[22] ReRAM-based synapses are potentially more suited.^[13]

1.1. From the Metal–Insulator–Metal to the Oxide Bilayer ReRAM

The conventional ReRAM structure is a metal–insulator–metal (MIM) stack, typically a transition metal oxide (TMO) encapsulated between an inert electrode (IE)^[23] and a reactive one (RE). In this technology, the application of a sufficiently large electric field generates anion (O^{2-}) migration, forming a conductive filament of oxygen vacancies (V_o in the Kröger–Vink notation^[24]).^[25] This process is known as the *forming*. The subsequent resistive switching operations are determined by the modulation of the V_o density profile in the TMO.^[25]

The selection of the electrodes and TMO materials is of fundamental importance for these processes as they determine the energy of formation of the oxygen vacancies.^[26] This directly

impacts the electrical performances of the device, such as the forming voltage, the on/off ratio and the writing reliability.^[27]

Moreover, the electrodes' oxygen affinity, the electrodes' work functions (WFs) and the voltage polarity of the forming process can influence the switching polarity.^[28]

In filamentary MIM devices, the switching process in the direction of higher resistances, the *reset*, is generally a gradual process, where the device resistance increases upon increasing the applied voltages.^[29] Vice versa, the switching process in the direction of lower resistances, the *set*, is a self-accelerated^[29] process, where the device resistance drops abruptly. The asymmetry between the *set* and the *reset* does not fit the requirements indicated for the training of memristor-based synapses.^[22] However, in the next lines we explain how to improve the switching symmetry, by exploiting dedicated resistive regimes where the self-acceleration of the *set* is relaxed and the onset of the *reset* is accelerated.

We can assume the device to be represented by an equivalent circuit of two resistances in series, as depicted in Figure 1c,d. One resistance represents the switching region (R_{switch}) of the device; the other one represents the passive equivalent series resistor (ESR) of the device, made of the electrodes' contacts and a portion of the filament remaining unchanged upon switching, named filament "plug".^[30]

The acceleration of the *reset* occurs when R_{switch} and ESR are comparable. In this case, while R_{switch} increases, the voltage divider between R_{switch} and ESR changes, with the voltage drop on R_{switch} increasing and the voltage drop on ESR decreasing. This positive-feedback process ends when the voltage drop on ESR becomes negligible compared to the one on R_{switch} . This phenomenon, described in Figure 1c, determines a sharp *reset* onset.

On the other hand, during the *set*, R_{switch} decreases abruptly approaching the ESR value. When ESR is no more negligible compared to R_{switch} , the decrease of R_{switch} consequently reflects into an increase of the voltage drop on ESR and a decrease of the voltage drop on R_{switch} . This phenomenon, described in Figure 1d, determines a gradual continuation of the *set*, after the initial abrupt onset.

The integration of a sub-stoichiometric oxide within the MIM-ReRAM stack, to create an oxide *bilayer*, can increase the ESR value, improving the symmetry between *set* and *reset*. In Hardtdegen et al.,^[31] a HfO₂-MIM ReRAM is compared to a TiO_x/HfO₂ *bilayer* ReRAM. The *forming* generates a conductive filament across the whole *bilayer*. The portion of the filament in the TiO_x gives an additional contribution to the ESR, compared to the other device. The increase of the ESR value enhances the mechanisms described in Figure 1c,d, leading to a more gradual *set* and a more abrupt *reset*.^[31] By operating at intermediate resistive states, where the *set* acceleration is attenuated and the *reset* remains a gradual process, the *bilayer* shows bidirectional gradual resistive switching.^[32] However, there is a trade-off against the available resistive window.

In Woo et al.,^[33] the improved switching symmetry and graduality observed in a AlO_x/HfO₂ *bilayer* ReRAM are attributed to the different mobility of the V_o in the two metal-oxides, which allow to electrically modulate the width of the conductive filament in both directions, avoiding the formation of a highly resistive filament gap.

The presence of an additional metal-oxide in the device stack, making the ReRAM an oxide *bilayer*, can also improve

the stability of the resistive states,^[31] the endurance,^[34] the reliability,^[27,35] and the switching stochasticity,^[27] by acting at the origin of the electronic-ionic mechanisms responsible for resistive switching.^[36]

A potential disadvantage introduced by the oxide *bilayer* concept is the increase of the *forming* voltages, determined by the series of two insulating materials. To reduce this effect, while keeping the improved symmetry introduced by the *bilayer* concept, a conductive metal-oxide (CMO) can be stacked onto the insulating TMO. With this solution, the voltage applied during the *forming* mainly drops on the TMO. On the other hand, if the resistivity of the CMO is higher than the resistivity of a metallic electrode, the device ESR increases with a positive impact toward the suppression of the *set* self-acceleration, as modeled in Zhao et al.^[37]

In Wu et al.,^[38] the graduality of the *set* in the CMO/TMO *bilayer* is explained by a different effect. The heat confinement determined by the low thermal conductivity of the CMO leads to the formation of multiple weak filaments in the TMO. Each weak filament partially contributes to the total switching of the device, resulting in a gradual process.

Differently from previous works where gradual bidirectional switching is observed in cells of HfO₂-based bilayer devices with fixed stack materials (AlO_x/HfO₂ in Woo et al.,^[33] TaO_x/HfO₂ in Yao et al.,^[13] and Kim et al.,^[39] TiO_x/HfO₂ in Cuppers et al.^[32]), here we present a solution to further optimize the analog response of the device, toward an increased number of multilevel states and reduced power consumption. Through a material engineering work, we studied the impact of the electrodes' work functions and of the conductive-metal-oxide' thickness and resistivity on the electrical response of the device. This investigation led us to understand the origin of the anomalous analog set property and to optimize the device' accumulative pulse response. We also verified the independence of the resistive states from the scaling of the cell area, to strengthen the hypothesis of filamentary-based switching.

By reactive sputtering, it is possible to deposit various sub-stoichiometric TaO_x films,^[40,41] corresponding to different resistivities. Such layers can be exploited to experimentally determine how different CMO resistivities impact the device switching properties.

First, we will describe the TaO_x materials development and structural characterization. Then, we present and discuss the electrical characteristics of a device made of the TaO_x/HfO₂ *bilayer*, sandwiched between TiN electrodes, to create a CMOS-compatible ReRAM. A Ti/HfO₂/TiN *baseline* device was fabricated to benchmark the upgraded switching properties exhibited by the oxide *bilayer*.^[39] To investigate the impact of the electrodes' WFs on the device switching properties, we fabricated a new oxide *bilayer* with a Pt bottom electrode (BE), instead of TiN. Guided by the observations collected from a step-by-step electrical characterization (*forming*, first *set*, first *reset*), we propose a microscopical interpretation of the resistive switching mechanisms, valid for both the oxide *bilayers* with a Pt BE and with a TiN BE.

To improve the device power consumption, reduce weight update noise and improve potentiation/depression symmetry, we optimized the TaO_x material by tuning its thickness and resistivity. Finally, we extracted the analog device model best

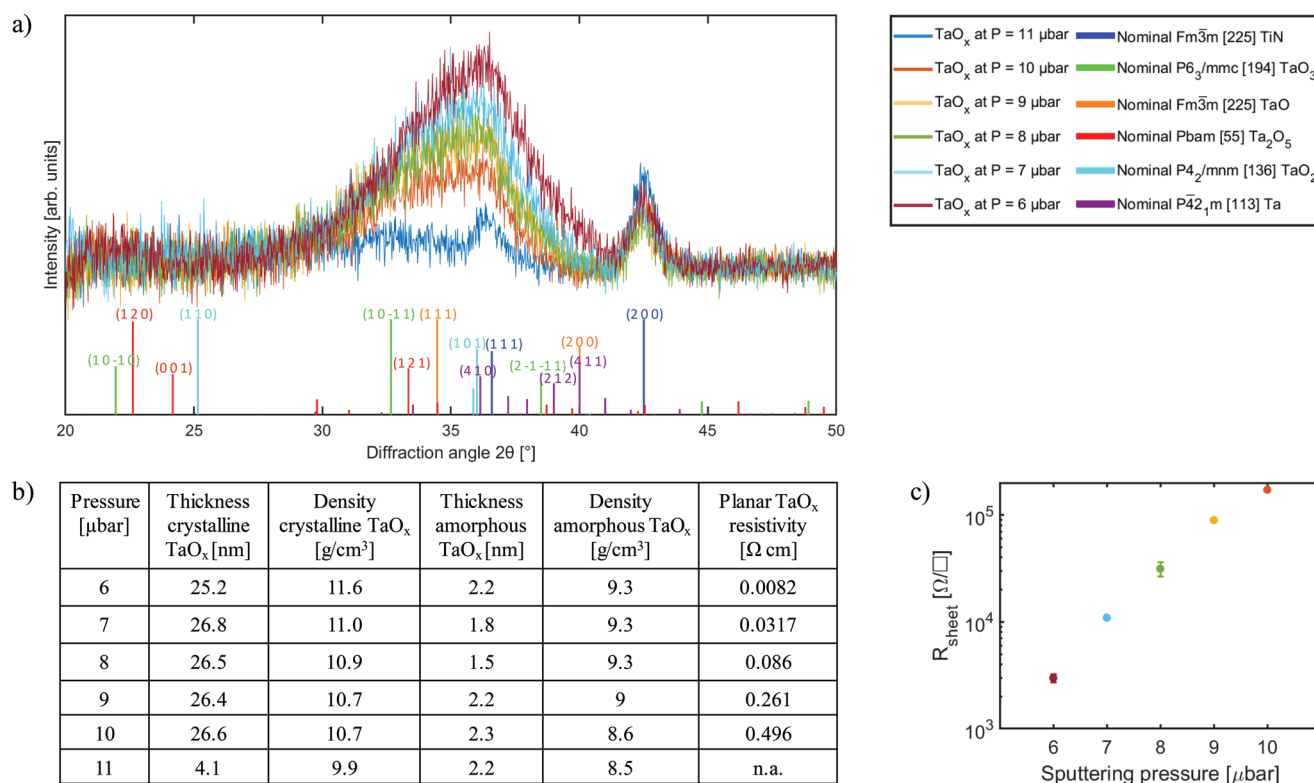


Figure 2. a) The GIXRD profiles of the TaO_x/HfO₂/TiN samples, sputtering TaO_x at increasing chamber pressures, from 6 to 11 μbar. The nominal peak positions of the various TaO_x stoichiometries, Ta and TiN, are reported from the reference database,^[45] with the correspondent space group. b) Table summarizing TaO_x films thicknesses, densities, and resistivities, respectively, determined by XRR and CTLM. c) Measured sheet resistance values plotted as a function of the sputtering chamber pressure.

fitting the optimized TaO_x/HfO₂ ReRAM pulse response.^[42] Using this model, we simulated the training of a fully-connected neural network of analog devices on the modified National Institute of Standards and Technology database (MNIST) dataset.^[43]

2. Material Development

To find the sputtering conditions for depositing a conductive TaO_x, we realized several TaO_x films for structural characterization. All these films were deposited onto HfO₂/TiN substrates, to enforce the same growth conditions as in the ReRAM devices. The sputtering chamber conditions, such as power, duration, and (O₂, Ar) mixture stream were kept fixed in all the depositions (see “Experimental Section” for details). The chamber pressure was varied from 6–11 μbar, in steps of 1 μbar between samples.

Grazing incidence X-ray diffraction (GIXRD) characterization provides information about the structure and the stoichiometry of the deposited suboxides. The GIXRD profiles of the TaO_x/HfO₂/TiN stacks, obtained by scanning at an incidence angle of 0.65°, are shown in Figure 2a. The peak at 42.5° is characteristic for the (200) TiN.^[44] The broad TaO_x peaks are shaped by the superposition of multiple contributions: in the figure we indicated only the crystallographic peaks accounting for more than 30% of the intensity. Their deconvolution (reported in Figure S1, Supporting Information) reveals the coexistence of

multiple stoichiometries of TaO_x and the Ta metal, indicating that the grown material does not have a unique composition, but rather is a complex phase.

With increasing sputtering pressures, a clear change in the broad TaO_x peak shape is identified, due to the change of the relative intensity of the contributing peaks: the oxidized components increase, at the expenses of the pure Ta one. These findings indicate that more oxygen is incorporated in the Ta lattice when sputtering at higher pressures.^[40]

We extracted the film thicknesses and densities from X-ray reflectivity (XRR) measurements (see Figure S3, Supporting Information, for details) and summarized their values in Figure 2b. The deposition rate is constant at ≈0.15 nm s⁻¹ for all the depositions, except for a chamber pressure of 11 μbar, where it falls to ≈0.03 nm s⁻¹. The fit of the XRR scans revealed the formation of a ≈2 nm thick TaO_x interface layer at the boundary with the HfO₂, with lower density than the bulk TaO_x. For increasing chamber pressures, the bulk TaO_x density gradually decreases from 11.6 g cm⁻³ at 6 μbar, to 10.7 g cm⁻³ at 10 μbar, and 9.9 g cm⁻³ at 11 μbar.

The characterization of the films’ resistivity was performed by circular transmission line measurements (CTLM). The measured sheet resistances (Ω □⁻¹) are reported as function of the chamber pressure in Figure 2c. The materials resistive window spans over more than one order of magnitude, from a few kΩ □⁻¹ up to values larger than 100 kΩ □⁻¹.

We converted the sheet resistances into resistivities (ρ) and report the values in Figure 2b. For the film deposited at 11 μbar

it was not possible to quantify the resistivity, because the measured sheet resistance exceeded the tool maximum measurable limit of $\approx 10 \text{ G}\Omega \square^{-1}$.

The characterization of the material stack was complemented by transmission electron microscopy (TEM) analysis of a pristine TiN/ TaO_x/ HfO₂/ TiN ReRAM with a $\approx 20 \text{ nm}$ thick TaO_x film deposited at $9 \mu\text{bar}$.

Figure 3a shows the scanning-TEM (STEM) image of a device in bright field (BF) mode. The contrast is sufficient to identify a thin, dense, amorphous layer at the TaO_x/HfO₂ interface. Its thickness is estimated to be $\approx 3 \text{ nm}$. Also, at the HfO₂ bottom interface with TiN, a thin oxidized TiO_x film grows, where a thickness of $\approx 1.5 \text{ nm}$ can be recognized. The oxidation of the TiN surface is attributed to the deposition of HfO₂, which takes place in an O₂ plasma atmosphere at 290°C .^[46]

Figure 3b shows the energy-dispersive spectroscopy (EDS) analysis performed by scanning from the bottom to the top TiN to study the composition across the full stack. This analysis confirms the presence of a more oxidized TaO_x phase below the crystalline TaO_x bulk, as the Ta counts in the material decrease, while the O₂ counts keep constant. The pillar growth of the TiN is believed to be responsible for the projection of the TiN roughness to the morphology of the rest of the stack, causing an overlap of the Hf and Ta counts at the TaO_x/HfO₂ interface, and of the Hf and Ti counts at the HfO₂/TiN interface. Moreover, interdiffusion can not be excluded.

Please note that during the scanning TEM (STEM) acquisitions we could not acquire sufficiently high-resolution images to classify the crystallinity properties of the bulk TaO_x. This limitation was caused by the small size of the grains identified ($< 6 \text{ nm}$, as reported in **Figure S4**, Supporting Information).

3. Electrical characterization

3.1. DC *I*–*V* Characteristics of the Oxide Bilayer ReRAM

In this section we present the DC *I*–*V* characteristics of a *bilayer* TiN/TaO_x/HfO₂/TiN ReRAM device. The thicknesses of the

TaO_x, HfO₂, and TiN electrodes are ≈ 30 , ≈ 6 , and $\approx 20 \text{ nm}$, respectively. The resistivity of the TaO_x layer is $\rho_{\text{TaO}_x} \approx 0.25 \Omega \text{ cm}$.

In parallel, we show the DC *I*–*V* characteristics of a *baseline* MIM stack, with the RE/TMO/IE structure, for comparison. Its structure is Ti/HfO₂/TiN, with a $\approx 10 \text{ nm}$ thick Ti layer.

In **Figure 4a,b** we compare their DC *I*–*V* forming characteristics, obtained by applying a positive voltage sweep. They show substantial differences: at 2 V , a leakage current of 100 pA is measured in the *bilayer* device, while in the *baseline* structure it is as large as $\approx 800 \text{ nA}$, close to the critical values required for the onset of the *forming* ($I_{\text{forming}} \approx 3 \mu\text{A}$). The *forming* voltage in the *baseline* device is $\approx 2.5 \text{ V}$, while in the *bilayer* configuration it can be as high as $\approx 5.5 \text{ V}$.

The higher pristine resistance of the *bilayer* structure can be explained by two main reasons. First, we know that even a pure Ta electrode would scavenge less O₂ from the HfO₂ layer, compared to a Ti one, due to their different oxygen affinities.^[27] Moreover, it is likely that our conductive TaO_x has a poor scavenging effect on the HfO₂ layer,^[47] since it was already partially oxidized as deposited by the O₂ plasma environment.

After the *forming*, the application of a negative voltage sweep generates opposite effects on the two ReRAM devices, as reported in **Figure 4c,d**: the *baseline* configuration shows a *reset* transition, the *bilayer* a *set* one.

The subsequent application of a positive voltage sweep resets the *bilayer* device back to higher resistances, while the same polarity sets the *baseline* device to lower resistances.

Another difference between them is about the onsets of the resistive switching mechanisms: while in the *baseline* most of the device conductance change occurs abruptly as soon as the physical conditions for switching are reached, in the *bilayer* it is a rather gradual process sustained by the increase of the applied voltages.

We repeated the same bipolar DC write/erase cycle five times on both samples. The corresponding *I*–*V* characteristics are displayed in **Figure 4e,f**. The *baseline* shows a certain variability of the V_{set} , which is attributed to the *reset* process stochasticity.^[48] This is accompanied by a large variability of the programmed high resistive states (HRS). Contrarily, the *bilayer*

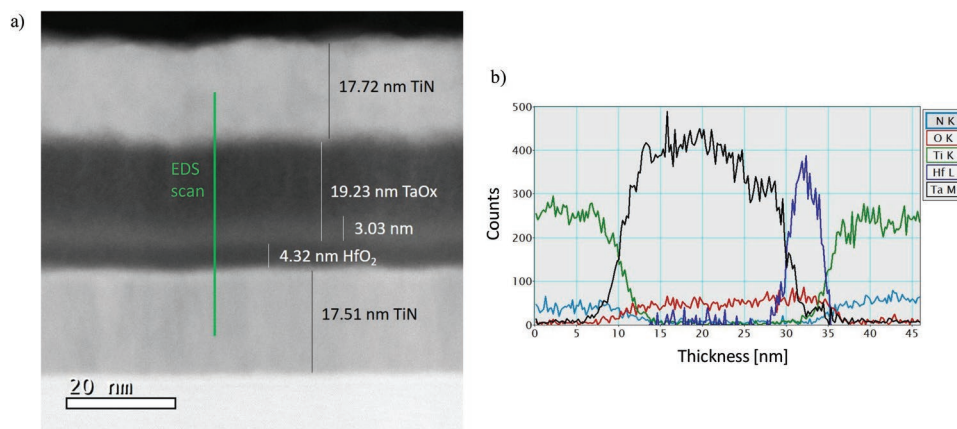


Figure 3. a) The bright field (BF) STEM image of the TiN/TaO_x/HfO₂/TiN ReRAM stack, revealing the onset of an amorphous TaO_x phase between the crystalline TaO_x bulk and the HfO₂ layer. The green vertical line corresponds to the region scanned for EDS chemical characterization, which is displayed in (b).

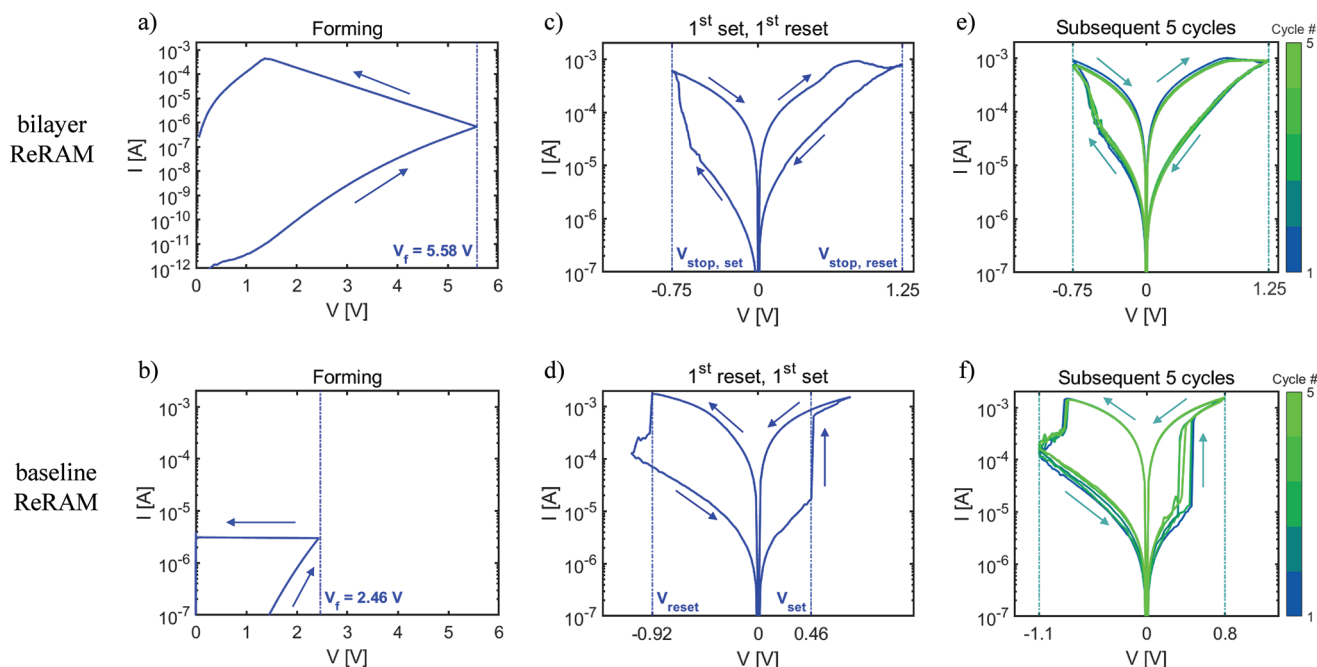


Figure 4. Comparison between the DC I - V characteristics of the *bilayer* ReRAM (first row), and the *baseline* ReRAM (second row). a,b) The *forming* curves. c,d) The first *set* and *reset*. e,f) The five programming cycles following the first *set* and *reset*.

current curves almost overlap during cycling, resulting in an improved stability of the HRS.

Despite having improved the resistive switching graduality and stochasticity, the *bilayer* structure shows higher pristine resistances compared to the *baseline* stack, leading to a large increase of the V_{forming} values. This effect hinders the device compatibility with the most advanced CMOS technologies. To address this point, future works will focus on the engineering of a sub-stoichiometric HfO_x layer to reduce the V_{forming} ,^[49] without compromising the analog and low-stochastic switching properties.

3.2. Interpretation of the Resistive Switching Mechanisms in the Oxide Bilayer ReRAM

In the filamentary ReRAM devices, the polarity of the resistive switching mechanisms depends, among other factors, on the WFs of the electrodes.^[28,50] Their role was investigated in the MIM structures made of an n-type TMO, encapsulated between a high and a low WF electrode.^[51,52] The high WF electrode/TMO interface can be modeled as a Schottky barrier, while the low WF electrode/TMO interface is a low-impedance Ohmic contact. In such a structure, after *forming* a conductive filament of V_{O} in the TMO, these defects can be either repelled from the TMO region next to the Schottky barrier, or accumulated there, depending on the applied voltage polarity.^[28] The local depletion of V_{O} increases the device resistance, while their accumulation decreases it.^[28]

The WF asymmetry can also affect the on/off switching ratio^[50] and the reliability of the bipolar switching,^[53,54] with both properties degrading for low WF asymmetries.

To verify whether the WF asymmetry plays a role in the reverse switching polarity of our *baseline* and *bilayer* structures, we investigate this aspect in more detail. In the *baseline* device, there is a clear asymmetry between the Ti electrode ($\text{WF}_{\text{Ti}} \approx 4.33 \text{ eV}$ ^[55]) and the TiN electrode ($\text{WF}_{\text{TiN}} \approx 4.7 \pm 0.9 \text{ eV}$ ^[56]). For the *bilayer* device, we did not find in literature a referential WF value for our conductive TaO_x layer. We found references for the Ta, whose WF can range from 4 to 4.8 eV,^[55] and for the Ta_2O_5 ($\text{WF}_{\text{Ta}_2\text{O}_5} \approx 4.05 \text{ eV}$ ^[57]). However, with such references we cannot claim with certainty the direction of the WF asymmetry in the *bilayer* stack, which would rather require dedicated measurements.

We decided to test our hypothesis with an alternative approach. We fabricated a *bilayer* device with a Pt BE, because Pt has the largest WF among metals ($\text{WF}_{\text{Pt}} \approx 5.65 \text{ eV}$ ^[58]) and therefore mandatorily forces a WF asymmetry in the same direction as the one of the *baseline*. However, as displayed in **Figure 5**, the *bilayer* stack with the Pt BE shows the same switching polarity as the one of the TiN-BE *bilayer* device. We conclude that the reverse switching polarity between the *bilayer* and the *baseline* ReRAMs cannot be attributed to an inversion of the WF asymmetry.

As a next step, we will interpret the switching mechanisms based on the exhibited DC I - V characteristics. In **Figure 5**, the *forming*, *set*, and *reset* operations are shown. For each process, we provide a tentative description of the microscopic mechanisms involved. As an example, **Figure 5a** shows the I - V characteristics of the *forming* process and **Figure 5b** below depicts it at the nanoscale. The applied *forming* voltage generates a conductive filament of V_{O} across the HfO_2 layer, by driving a migration of O^{2-} anions toward the interface with the TaO_x .^[25] The abrupt increase of the current upon forming the filament

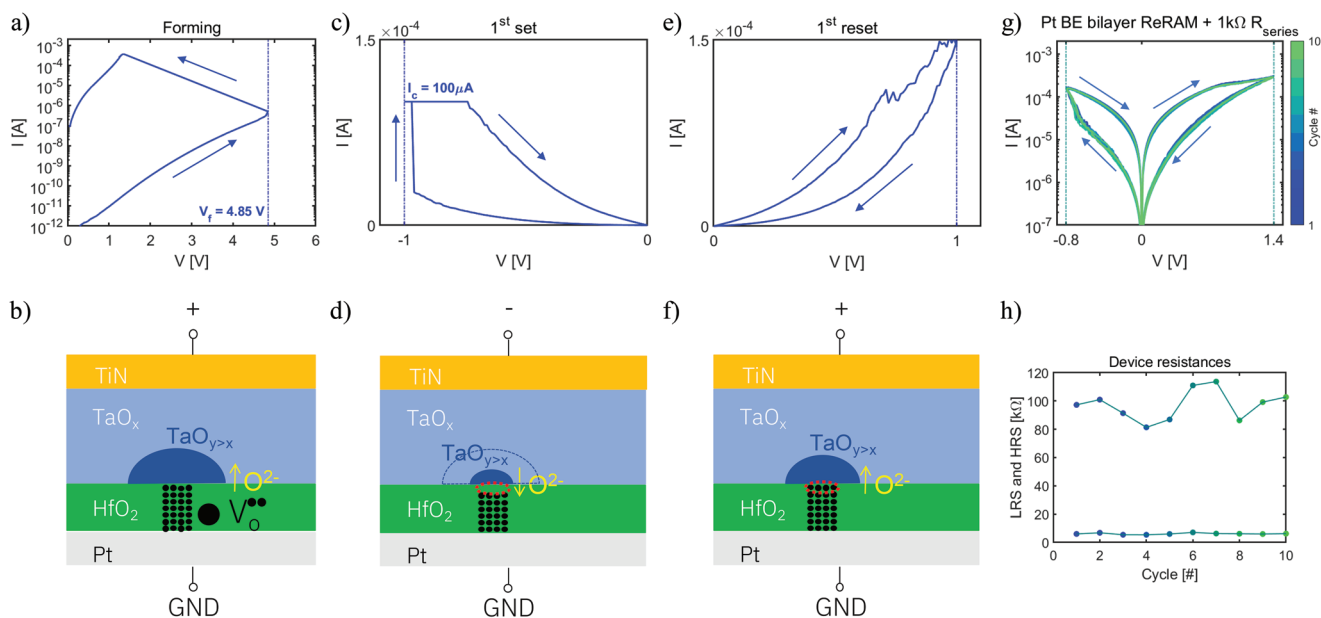


Figure 5. DC I - V characteristics of the TiN/TaO_x/HfO₂/Pt bilayer ReRAM, that is, the a) *forming*, c) first *set*, and e) first *reset*. Below, in b,d,f) the microscopical interpretation of these processes. g) The DC bipolar switching of the same bilayer stack, when a 1 kΩ external series resistor is introduced. h) Resistive states extracted from (g).

induces a local high temperature increase, which, combined to Coulomb repulsion, make the O²⁻ diffuse radially.^[25] A more oxidized region (TaO_{y>x}) is obtained, dominating the device resistance after *forming*. This is the post-forming resistance (R_{pf}), correspondent to the device HRS.

Now the device is formed and can be *set* to a low resistive state (LRS) upon the application of a negative voltage sweep, as shown in Figure 5c. We imposed a compliance of 100 μA to avoid that excessive currents could lead to very large resistance drops. In Figure 5d we depict the redox processes causing the decrease of the device resistance. Activated by the negative voltage applied, the O²⁻ formerly diffused in the TaO_y region during the *forming* process migrate back to the HfO₂ layer. Consequently, the volume of the O-rich TaO_y region shrinks, contributing to a decrease of the device resistance, while the V_o in the HfO₂ filament recombine with the incoming O²⁻, contributing to an increase of the device resistance. Since we measure a drop of the total device resistance, among these two counteracting processes the reduction of the TaO_y dominates over the oxidation of the filament tip.

Subsequently, as shown in Figure 5e, upon the application of a positive voltage sweep, the device can *reset* to a HRS. In Figure 5f we describe the correspondent redox processes, where the O²⁻ migrating from the filament gap toward the TaO_x layer gradually oxidize it, while the filament gap gradually reduces, upon increasing voltage amplitudes. In this case, since we measure an increase of the device resistance, the dominant resistive switching process is given by the local oxidation of the TaO_x.

Under these hypotheses, we conclude that after the *forming* process, the current bottleneck in the device is the TaO_x layer, both in the LRS and in the HRS, and not the filament in the HfO₂.

The main difference observed between the Pt-BE and the TiN-BE bilayer devices is in the *set* mechanism, which is abrupt

in the first case and gradual in the second. To explain this difference, we will first compare the voltage amplitudes required in the two samples to activate the *set* process and then we will discuss the role of the local electric field (F) and temperature (T) at the filament/TaO_x interface.

The bilayer stack with the Pt BE shows higher *set* voltages and higher R_{pf} values too compared to the TiN-BE one ($R_{pf,Pt-BE} \approx 305$ kΩ and $R_{pf,TiN-BE} \approx 99$ kΩ, from Figures 5a and 4a; $|V_{set,Pt-BE}| \approx 0.96$ V and $|V_{set,TiN-BE}| \approx 0.50$ V, from Figures 5c and 4c). The requirement of higher voltages to activate the *set* mechanism when the devices are programmed in higher HRSs was already observed in MIM filamentary ReRAMs^[29] and in bilayer structures too.^[32] For the latter, it was also shown that the combination of higher HRSs and onset voltages makes the *set* process sharper.^[32] In the next lines, we tentatively describe the physical origin for such phenomenon, based on the switching model derived for our bilayer structures.

While the *set* mechanism evolves (Figure 5d), the O-rich TaO_y region shrinks, but keeps the dominant contribution to the device resistance. Therefore, we can expect that most of the applied $|V_{set}|$ will keep being distributed on the shrinking TaO_y, hence leading to a local F enhancement. In parallel, the current increase leads to a T growth in the same TaO_y region, due to Joule heating. The (F , T) enhancement promotes further O²⁻ migration, which ultimately installs a positive (F , T) feedback loop, resulting in an abrupt *set* process.^[29]

A gradual *set* transition can be restored by relaxing the (F , T) enhancement in the TaO_y region with the introduction of an external series resistor. Its effect is to reduce the power dissipated in the device during the circuit transients that follow the onset of the *set* process,^[59] preventing the thermal runaway responsible for the abrupt switching.

The non-linearity of the HRS makes a 1 kΩ series resistor sufficient for this purpose, without compromising the resistive

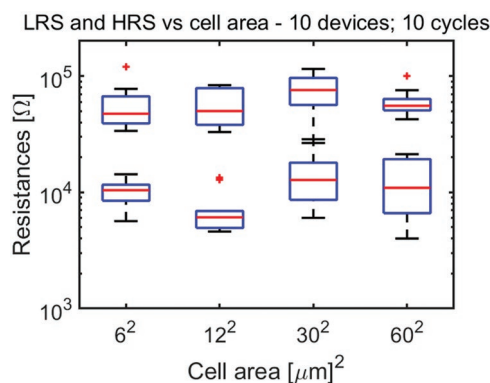


Figure 6. The boxplots of the Pt BE *bilayer* ReRAM resistive states, collected from ten devices per each of the selected device areas. Every device was tested under identical electrical conditions ($V_{\text{stop, set}} = -1.2$ V and $V_{\text{stop, reset}} = +1.6$ V).

switching window, as displayed in Figure 5g,h, where an on/off ratio of ≈ 1 order of magnitude is achieved by applying a maximum $V_{\text{stop}} = +1.4$ V.

Differently from the *filamentary-type* devices, which generally show abrupt *set* mechanisms, the *interface-type* devices are characterized by a bidirectional gradual resistive switching,^[60] typically showing area-dependent resistive states.^[60] Therefore, to confirm the hypothesis of filamentary switching for our *bilayer* ReRAMs, we studied the area-dependency of their resistive states, including $(6 \mu\text{m})^2$, $(12 \mu\text{m})^2$, $(30 \mu\text{m})^2$, and $(60 \mu\text{m})^2$ devices in the statistics. For each area, we applied 10 bipolar voltage sweeps up to the stop values -1.2 V and $+1.6$ V.

The median LRS and HRS were grouped in boxplot graphs and plotted against the device area, as shown in Figure 6. We observe that neither the HRS nor the LRS scale with the area. The fluctuations between the different cell areas can be attributed to the device-to-device variability and to the relatively small size of the collected statistics.

These results support the hypothesis that the switching mechanisms are filamentary based, even if they become gradual in both directions when a $1 \text{ k}\Omega$ series resistor is introduced.

3.3. Impact of the TaO_x Resistivity on the Switching Properties of the Oxide Bilayer ReRAM

To explore the role of the TaO_x resistivity on the switching properties of the *bilayer* devices, we compare two ReRAM types fabricated using TaO_x layers with resistivities of $\rho_{\text{TaO}_x} \approx 0.01 \Omega \text{ cm}$ and $\rho_{\text{TaO}_x} \approx 0.25 \Omega \text{ cm}$. In both cases, the TaO_x film is $\approx 30 \text{ nm}$ thick and the BE material is TiN.

In Figure 7a,b we compare their post-forming DC I - V characteristics. Both devices similarly show low-stochastic, gradual bidirectional switching transitions, making the current curves almost overlapping over 20 subsequent cycles. Moreover, the switching processes require low voltages (< 1 V) to be activated. The main difference between the two types of ReRAM is that only the device with the more resistive TaO_x can *reset* to HRS values $> 10 \text{ k}\Omega$.

To validate this result from a statistical point of view, we measured 10 devices per each TaO_x type, using the same stop voltages. In Figure 7c,d we compare the cumulative distributions

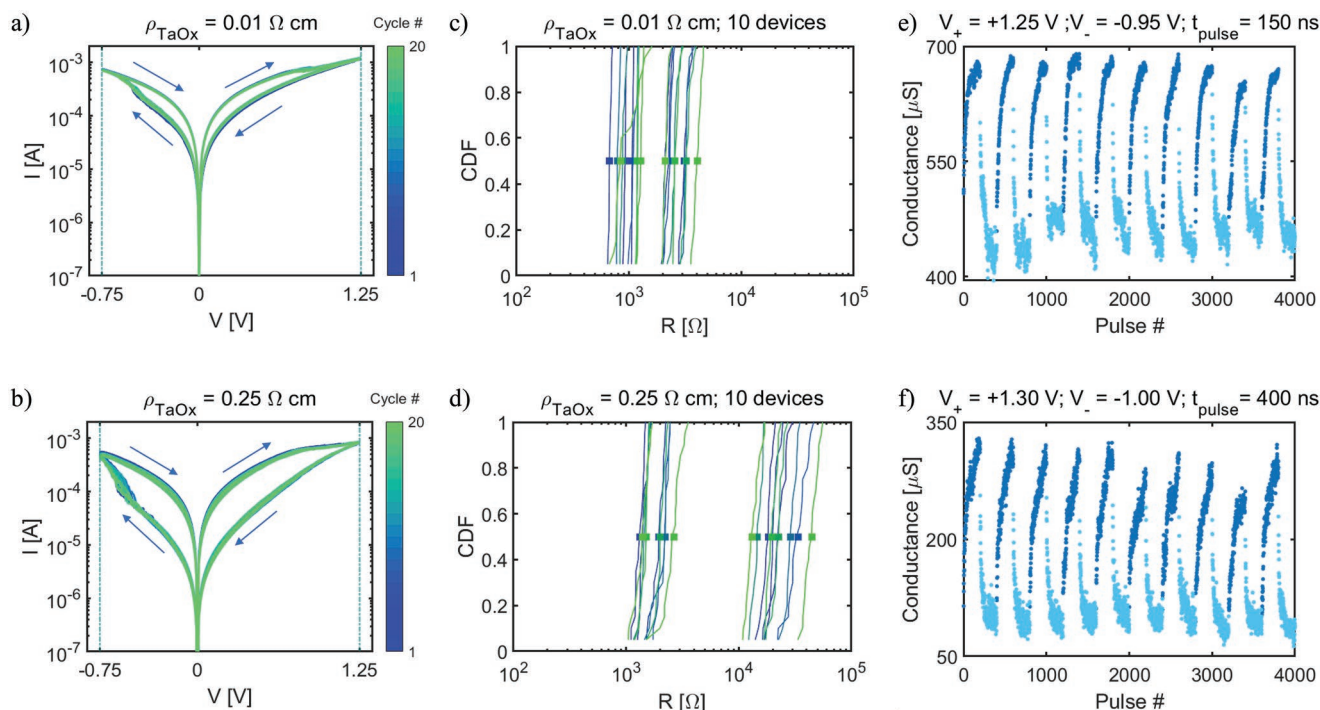


Figure 7. Electrical characterization of *bilayer* devices with different TaO_x resistivities. In the first row: $\rho_{\text{TaO}_x} \approx 0.01 \Omega \text{ cm}$, in the second one: $\rho_{\text{TaO}_x} \approx 0.25 \Omega \text{ cm}$. a,b) DC characteristics. c,d) Cumulative distributions of the resistive states. e,f) Pulsed characterization, performed by alternating 200 identical potentiation and 200 identical depression voltage pulses.

of their resistive states. Each device is represented by a pair of lines of the same color, one for the LRS and the other one for the HRS. The statistics confirms that the more resistive TaO_x improves the HRS values beyond 10 kΩ and the on/off ratio too, which increases up to ≈ 1 order of magnitude. However, the resistive states increase comes at the cost of a broader device-to-device variability (cumulative distributions more distanced), as well as a larger cycling variability (cumulative distributions more oblique).

The gradual resistive switching properties are enhanced at high bandwidth. In Figure 7e,f we compare the ReRAMs with different TaO_x types by applying short programming pulses. In both cases the programming sequence consists in alternating 200 identical *set* and *reset* pulses. This sequence is iterated ten times to check the stability of the potentiation/depression characteristics. The duration and the amplitudes of the voltage pulses are selected to maximize the number of conductive states (NoS), which improves when *G* is updated in small steps, but covers a large window too.

For the ReRAM with the low-resistive TaO_x, a pulse duration of 150 ns and pulse amplitudes of +1.25 V and −0.95 V were used for depression and potentiation, respectively. Instead, 400 ns pulses of amplitudes +1.30 V and −1.00 V were used for the ReRAM with the more resistive TaO_x.

The methodology used for determining the NoS is detailed in the Paragraph S6, Supporting Information. The device with the low-resistive TaO_x layer has 13 and 9.5 average NoS for potentiation and depression, respectively. The other device has an average NoS of 9.5 and 11.2.

3.4. Impact of the TaO_x Thickness on the Switching Properties of the Oxide Bilayer ReRAM

In previous studies, it was shown that in MIM ReRAMs with a RE/TMO/IE structure, the thickness of the RE affects the O₂ scavenging reactions at the interface with the TMO, impacting the device switching properties.^[27,47,61,62] For RE thicknesses between 1 and 5 nm, the O₂ scavenging process is enhanced by greater thicknesses, causing a decrease of the *forming* voltages.^[61,62] Instead, for REs thicker than 5 nm, the *forming* voltages stop decreasing,^[61] due to the saturation of the O₂ scavenging reactions.^[61]

In this study, we fabricated three variations of TiN-BE *bilayer* devices, where we changed the TaO_x thickness to understand its impact on the device switching properties. The thickness variations are ≈ 20, ≈ 50, and ≈ 60 nm of the highly resistive TaO_x material ($\rho_{\text{TaO}_x} \approx 0.25 \Omega \text{ cm}$).

Figure 8d shows that for increasing TaO_x thicknesses, the *forming* voltages do not decrease. Therefore, we conclude that the O₂ scavenging processes at the TaO_x/HfO₂ interface have similarly saturated in all the examined samples.

In Figure 8a–c, we compare their post-forming DC *I*–*V* characteristics. The 20 nm thick TaO_x device exhibits gradual bidirectional resistive switching transitions, similarly to the one with a 30 nm thick TaO_x layer, shown in Figure 7b. Vice versa, the device with the 50 nm thick TaO_x displays an abrupt *set* transition, and the one with the 60 nm thick TaO_x shows multiple abrupt *set* transitions. Moreover, we can highlight that the HRS values tend to decrease for increasing TaO_x thicknesses.

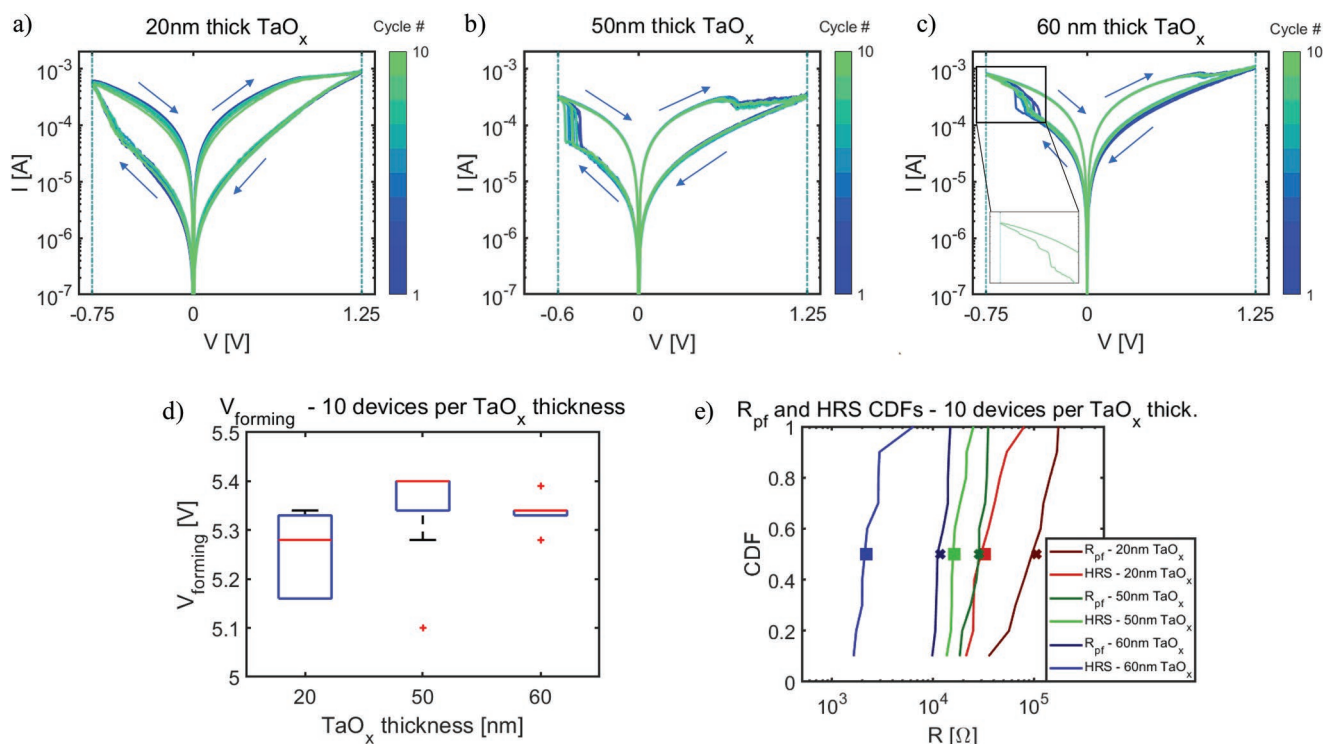


Figure 8. The DC *I*–*V* bipolar switching characteristics of the *bilayer* ReRAMs with TiN BE and TaO_x thicknesses of a) 20 nm, b) 50 nm, and c) 60 nm, respectively. d) The comparison between the *forming* voltages. e) The comparison between the *R*_{pf} and the HRS, reported as cumulative distributions.

To understand the origin of this latter trend, we first correlate the distributions of the HRSs to the R_{pf} values, plotting them all against the TaO_x thickness, as displayed in Figure 8e. We clearly observe that also the R_{pf} values decrease for increasing TaO_x thicknesses, indicating that the trend originates from the *forming* process itself.

To understand why thicker TaO_x layers result into lower R_{pf} values, we include in the discussion some additional considerations concerning the thermal properties of the TaO_x layer and their impact on the *forming* process.

The thermal coefficient of a conductive TaO_x film depends on its resistivity.^[63] Since our TaO_x material is characterized by the resistivity $\rho_{\text{TaO}_x} \approx 0.25 \, \Omega \, \text{cm}$, we can assume a low thermal coefficient $\lesssim 1 \, \text{W} \, \text{mK}^{-1}$.^[63] For this reason, in our devices, the TaO_x acts as heat confinement layer.^[38,64]

The impact of this effect on the *forming* process becomes clear if we resume the nanoscale model depicted in Figure 5a. When the filament begins to form, the heat confinement effect induces high T at the TaO_x /filament interface, promoting the ion mobility. This favors the O^{2-} migration from the HfO_2 to the TaO_x layer, enhancing the generation of conductive V_\bullet in the HfO_2 layer.

The high conductivity of the TaO_x layer makes the voltage drop insufficient to sustain the vertical drift of the O^{2-} in the TaO_x layer too. Instead, the high T generated upon forming the filament favor the lateral diffusion of the O^{2-} in the TaO_x layer.^[41,65]

Thicker thicknesses of the TaO_x layer further enhance the heat confinement effect, due to the increasing distance between the hot TaO_x /filament interface and the thermally conductive TiN-TE. With higher T , the thermal diffusion is promoted, extending the highly oxidized $\text{TaO}_{y>x}$ region depicted in Figure 5b more laterally than vertically.

Since the device resistance is measured vertically, the formation of a more conductive filament in the HfO_2 and the enhanced lateral diffusion of the O^{2-} in the TaO_x layer concur toward a reduction of the R_{pf} values.

The enhanced heat confinement effect can also explain the origin of the sharper *set* mechanisms observed in the *bilayer* devices with thicker TaO_x layers. The *set* process was modeled in Figure 5c as a redox reaction, where the reduction of the O-rich $\text{TaO}_{y>x}$ region dominates the device resistance change, over the oxidation of the filament tip. During this process, the evolution of the T at the TaO_x /filament interface is critical for the switching graduality, as already discussed.

We conclude that the abrupt *set* transitions observed in the devices with 50 and 60 nm thick TaO_x layers are driven by an enhanced heat confinement effect, overcoming the typical switching graduality observed in the devices with thinner TaO_x layers.

A particular case is the ReRAM with the 60 nm thick TaO_x , characterized by multi-step abrupt *set* transitions, as displayed in the magnified corner of Figure 8c. This phenomenon could be attributed to the formation of multiple filaments, as explained by Wu et al.^[38]

3.5. The Oxide Bilayer ReRAM with the Optimal TaO_x Material– Pulsed Characterization

The comparison between ReRAMs with different TaO_x resistivities and thicknesses indicated that thinner and more resistive TaO_x layers improve the device resistances and the gradual bidirectional switching properties, without degrading the *forming* voltages, nor the *set* and *reset* stop voltages.

In particular, the devices with a 20 nm thick TaO_x layer and $\rho_{\text{TaO}_x} \approx 0.25 \, \Omega \, \text{cm}$ show high R_{pf} values of $\approx 100 \, \text{k}\Omega$. However, we observed in Figure 8e that the DC operations degrade the HRS values by a factor $\approx 3\times$ compared to the departing R_{pf} values, due to the strong electrical stress induced by the first DC *set* voltage sweep.

To prevent it, we rather apply sub- μs programming pulses directly after the *forming* process is over. Figure 9a shows the potentiation/depression curves obtained from the application of 200 pulses of amplitudes $V_- = -0.95 \, \text{V}$ and $V_+ = +1.25 \, \text{V}$ per each direction and duration 250 ns. The sequence was iterated three times to check the states' stability over cycling.

While the potentiation trace is almost analog, the depression shows an initial abrupt jump, where the device resets close to the departing high resistances.

To get more balanced potentiation/depression characteristics, we reduced the amplitude of the *reset* pulses down to $V_+ = +1.10 \, \text{V}$ and we shifted the operational regime of the device toward higher conductances. In Figure 9b we show ten subsequent potentiation/depression curves obtained using the modified pulse scheme with more balanced amplitudes for *set* and *reset*. The G trace spans between 60 and $240 \, \mu\text{S}$, defining a $G_{\text{max}}/G_{\text{min}}$ ratio ≈ 4 , and it nicely mimics a biological analog synapse, as displayed in Figure 9c.

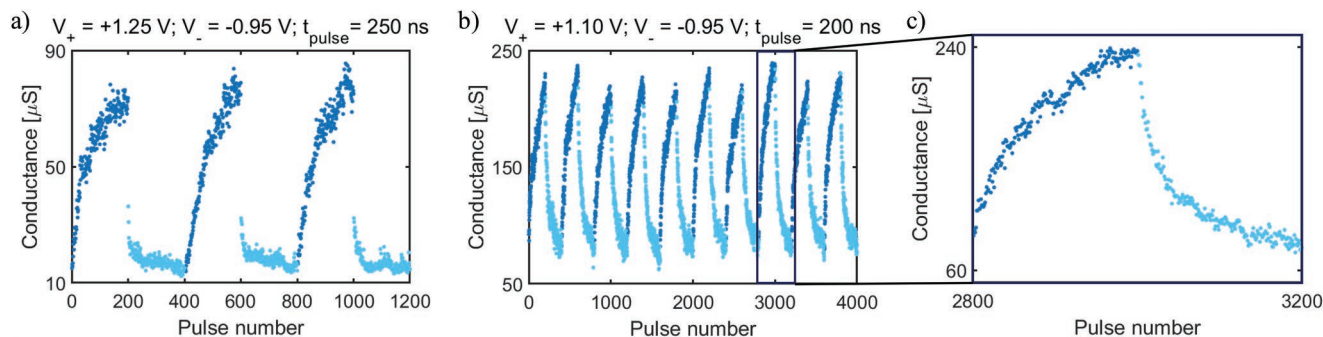


Figure 9. Pulsed characterization of the *bilayer* ReRAMs with TiN BE, 20 nm thick TaO_x , and $\rho_{\text{TaO}_x} \approx 0.25 \, \Omega \, \text{cm}$. a,b) Potentiation/depression characteristics, obtained in two different conductive regimes and using different pulse shapes. c) Magnified potentiation and depression trace.

We can estimate the relative NoS as detailed in y Paragraph S6, Supporting Information. We get an average NoS of 14.5 and 21.3 for potentiation and depression, respectively.

4. Training Simulation of a Fully Connected Network of Analog Devices

In the long term, our goal is to exploit filamentary memristors to compute ANNs and other AI workloads. To validate our *bilayer* devices, we use the publicly accessible “IBM aihwkit”,^[42] which allows to simulate our devices in real world applications. This simulation framework provides different device models, of which the “PowStepDevice” is the most suitable for the oxide *bilayer*, and which can be extracted from the measured potentiation/depression traces of Figure 9b. It provides a pulse-to-conductance response following a power-law. The model follows a series of update rules as shown in the following equations:

$$\omega_{ij} = \frac{w_{\max} - w_{ij}}{w_{\max} - w_{\min}} \quad (1)$$

$$\Delta w_{\text{up}} = dw_{\min} (1 + \beta) (\omega_{ij})^{\gamma(1+\beta_{\text{pow}})} \quad (2)$$

$$\Delta w_{\text{dn}} = -dw_{\min} (1 - \beta) (1 - \omega_{ij})^{\gamma(1-\beta_{\text{pow}})} \quad (3)$$

where w_{ij} is the weight to be updated, w_{\min} (w_{\max}) the lower (upper) bounds, dw_{\min} the linear update parameter, β its skew between the up/down directions, γ the exponential update parameter and β_{pow} its skew between the up/down directions.

The individual measurement traces were averaged and then normalized to $(-1, +1)$. We employed a *lmfit*-based algorithm^[66] to automatically fit our measurements to the aihwkit “PowStep-Device” model. The fit is depicted in Figure 10a.

As a test-vehicle for our real world validation, we selected the character recognition task on the MNIST dataset.^[67] As an architecture, we selected a fully-connected four layer network with a 784 node input layer, two hidden layers with 200 and 64 nodes, respectively, and the output-layer with ten nodes. We trained the same architecture with the analog device model and, as a benchmark, using 32-bit floating point arithmetic.

The simulation keeps into account the measured cycle-to-cycle variability, but not the device-to-device variability yet.

Figure 10b shows a comparison between the training convergences of the analog and of the digital model, with a cross-entropy objective function.^[68] The digital model converges faster to lower losses and shows a better accuracy (96%) compared to the one of the analog model (82%). This $\approx 15\%$ drop in accuracy is not unexpected, since the analog device model itself already contains non-idealities in the weight update symmetry, linearity and minimum step size.^[22]

To fully exploit the analog devices' potential, it will be crucial to modify the standard training algorithms, based on stochastic gradient descent (SGD), taking into account the specific non-ideal properties of our analog devices. As an example, the *Tiki-Taka* algorithm^[69] can achieve excellent accuracies even with non-symmetric devices.^[69] Training simulations based on this algorithm will be presented in future works.

5. Conclusion

In this work we designed, fabricated, and optimized a CMOS-compatible electrical synapse for in-memory computing applications.

The integration of a conductive TaO_x within the ReRAM stack, which creates a $\text{TaO}_x/\text{HfO}_2$ *bilayer*, improves the resistive switching stochasticity compared to a baseline ReRAM Ti/HfO_2 .

The new devices, despite having a filamentary-based physics, show anomalous but excellent gradual bidirectional resistive switching properties. Moreover, they can be programmed with fast (< 200 ns) and low voltage (< 1.5 V) pulses to cover an analog conductive window with a maximum-to-minimum ratio > 4 .

We found that thinner and more resistive TaO_x layers improve the device power consumption during its operations and the number of multilevel states.

To explain the peculiar resistive switching properties of our devices, in particular the reversed polarity compared to the *baseline* ReRAMs, we propose a microscopical interpretation of the *forming*, *set* and *reset* operations. We account for a localized O^{2-} exchange at the interface between the HfO_2 and the TaO_x . The redox reactions in the TaO_x layer are responsible for the device resistive switching.

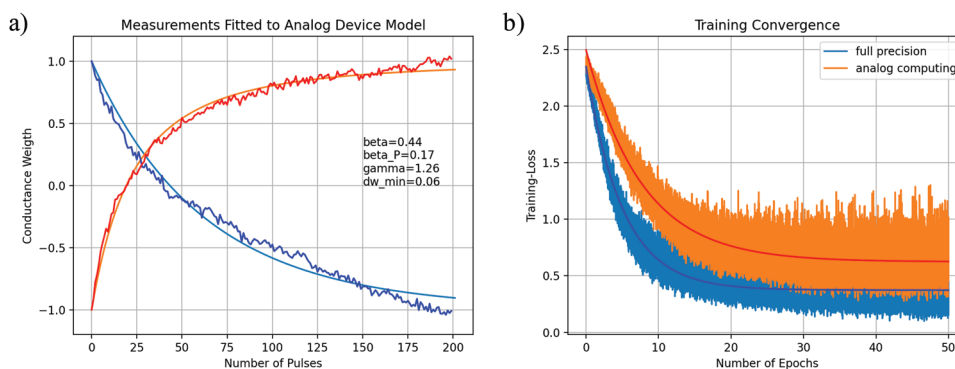


Figure 10. a) “PowStepDevice” fit of the ReRAM potentiation/depression characteristics, using *lmfit*. b) Training convergence of the MNIST dataset using the analog device model and the 32-bit floating point arithmetic. Their exponential decays are fitted with a red and a blue line, respectively.

To validate the performances in real world applications, we simulated the handwritten character recognition task with the MNIST dataset, using the IBM analog hardware acceleration kit (aihwkit). We found an accuracy drop compared to the benchmark model of only $\approx 15\%$, using SGD and backpropagation. More disruptive training algorithms such as *Tiki-Taka*^[69] are expected to further compensate for the asymmetric properties of our devices and yield a performance comparable to the benchmark simulation.

Future work will focus on product-level integration of the *bilayer* ReRAM into the CMOS stack to better control the *forming* process. This solution is expected to reduce the current overshoots during resistive switching,^[70] leading to higher resistance values and hence further reduced power consumptions.

Nonetheless, our TaO_x/HfO₂-based ReRAM already represents an attractive and flexible memory element for large-scale integration, considering its excellent granular switching properties and CMOS compatibility.

6. Experimental Section

Fabrication: The presented ReRAM devices were vertical, sharing a common bottom contact of n-doped Si. The 20 nm thick Pt BE was deposited by evaporation at room temperature; the 20 nm thick TiN BE was deposited by plasma-enhanced atomic layer deposition (PE-ALD) at 300 °C using a tetrakis-(dimethylamino)titanium (TDMAT) precursor and (N₂, H₂) plasma. The 6 nm thick HfO₂ layer was deposited by PE-ALD at 290 °C using a tetrakis-(ethylmethylamino)hafnium (TEMAH) precursor and O₂ plasma. In the *bilayer* devices, the controlled sputtering parameters for the deposition of the TaO_x films were: DC power = 50 W; (O₂, Ar) stream ratio = (1 sccm, 40 sccm); deposition time = 200 s to grow a ≈ 30 nm thick film. The TaO_x thickness variations were realized by linearly changing the sputtering duration.

STEM and EDS were used to check the ≈ 20 nm thickness, XRR for the ≈ 30 nm one, and SEM for the ≈ 50 nm and ≈ 60 nm ones (see Figures S8 and S9, Supporting Information).

In the baseline, the 10 nm thick Ti capping of the HfO₂ was DC sputtered. The 20 nm thick TiN TE was deposited by RF sputtering of a TiN target in a mixed (Ar, N₂) plasma. A 50 nm W layer was sputtered on top. The sputtering of the W/TiN/TaO_x (W/TiN/Ti in the baseline) proceeded without vacuum breaking between the deposition of the different layers, to avoid uncontrolled oxidation at the reactive interfaces. The patterning of the device geometry was performed through inductively coupled plasma etching of the W/TiN/TaO_x (W/TiN/Ti in the baseline) stack, using a mixed CHF₃ and SF₆ plasma, which stopped at the HfO₂ layer. A passivation layer of 100 nm thick SiN_x was grown by plasma-enhanced chemical vapor deposition (PECVD). The via to access the device TE was etched with a mixed CHF₃ and O₂ plasma by reactive-ion etching (RIE). 100 nm of W were sputtered and then RIE etched to define the device pads.

The described process flow avoids any lift-off steps, which could not be performed in foundries.

Structural Characterization: The X-ray characterization, which included the GIXRD and XRR scans, was performed using a Bruker D8 Discover diffractometer, equipped with a rotating anode generator.

The lamella preparation for the TEM analysis was conducted with the FEI Helios NanoLab 450S focused ion beam. The TEM analysis was carried out with a double spherical aberration corrected JEOL ARM200F operated at 200 kV. The EDS scans were performed using a liquid-nitrogen-free silicon drift detector.

Electrical Characterization: The CTLM characterization of the TaO_x layers and the DC characterization of the *bilayer* ReRAM with Pt BE were

performed with a B1500 parameter analyzer. The DC characterization of the baseline and the *bilayer* with TiN BE was performed with an Agilent 4155C parameter analyzer.

To simulate DC conditions, the quasi-static measurements were performed by dividing the applied voltage sweeps into 100 equal steps, lasting 100 ms each. The read of the device resistive state was done by performing a voltage sweep up to +250 mV and extracting the resistance at +200 mV.

External series resistors were introduced by mounting an SMD resistor in series to the probe tip biasing the device pads. All the *bilayer* devices were formed with a 10 k Ω external series resistor, to reduce the power dissipated through them during the transients following the resistive switching and limit the steady state current.

The pulsed characterization was performed using a NI PXIe-5451 arbitrary waveform generator to source the generated pulses to the device TE, and an oscilloscope NI PXIe-5164 to read the current signal flowing through the device BE. The pulsed read scheme consisted of alternating a positive and a negative pulse with amplitudes of ± 200 mV and a duration of 10 μ s, to cancel out any eventual measurement offset.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors acknowledge the Binnig and Rohrer Nanotechnology Center (BRNC) at IBM Research Europe - Zürich. The authors would like to thank Malte Rasch and Tayfun Gokmen for technical discussions. The authors thank Linda Rudin for proofreading the manuscript. The authors acknowledge the insightful discussions with Jean Fompeyrine during the initial phase of the project. This work was funded by the European Union within the H2020 “MANIC” (grant ID: 861153), “MeM-Scales” (grant ID: 871371), “NEoteRIC” (grant ID: 871330), “NEBULA” (grant ID: 871658) and “PlasmoniAC” (grant ID: 871391) projects.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

analog memory, artificial synapses, HfO₂, resistive random-access memory, TaO_x

Received: May 18, 2022
Published online: July 10, 2022

[1] R. L. Villars, C. W. Olofson, M. Eastwood, *Big data: what it is and why you should care*, White Paper, IDC, MA, USA 2011.

- [2] S. Kaisler, F. Armour, J. A. Espinosa, W. Money, in *Proc. of the Annual Hawaii Int. Conf. on System Sciences*, IEEE, Piscataway, NJ **2013**, pp. 995–1004.
- [3] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, Cambridge, MA **2016**.
- [4] W. A. Wulf, S. A. McKee, *ACM SIGARCH Comput. Archit. News* **1995**, 23, 20.
- [5] M. Horowitz, in *Digest of Technical Papers—IEEE Int. Solid-State Circuits Conf.*, Vol. 57, IEEE, Piscataway, NJ **2014**.
- [6] T. N. Theis, H. S. Philip Wong, *Comput. Sci. Eng.* **2017**, 19, 2.
- [7] A. Sebastian, M. Le Gallo, R. Khaddam-Aljameh, E. Eleftheriou, *Nat. Nanotechnol.* **2020**, 15, 529.
- [8] Q. Xia, J. J. Yang, *Nat. Mater.* **2019**, 18, 309.
- [9] D. Ielmini, H.-S. P. Wong, *Nat. Electron.* **2018**, 1, 333.
- [10] M. A. Zidan, J. P. Strachan, W. D. Lu, *Nat. Electron.* **2018**, 1, 22.
- [11] J. M. Rabaey, *Digital Integrated Circuits—A Design Perspective*, 2nd ed., Pearson, London **2003**.
- [12] M. Prezioso, F. Merrih-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, D. B. Strukov, *Nature* **2015**, 521, 61.
- [13] P. Yao, H. Wu, B. Gao, S. B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H.-S. P. Wong, H. Qian, *Nat. Commun.* **2017**, 8, 15199.
- [14] P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, H. Qian, *Nature* **2020**, 577, 641.
- [15] A. Beck, J. G. Bednorz, C. Gerber, C. Rossel, D. Widmer, *Appl. Phys. Lett.* **2000**, 77, 139.
- [16] L. O. Chua, *IEEE Trans. Circuit Theory* **1971**, 18, 507.
- [17] D. B. Strukov, G. S. Snider, D. R. Stewart, R. S. Williams, *Nature* **2008**, 453, 80.
- [18] B. Govoreanu, G. S. Kar, Y. Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini, I. P. Radu, L. Goux, S. Clima, R. Degraeve, N. Jossart, O. Richard, T. Vandeweyer, K. Seo, P. Hendrickx, G. Pourtois, H. Bender, L. Altimime, D. J. Wouters, J. A. Kittl, M. Jurczak, in *Tech. Dig. - Int. Electron Devices Meet., IEDM*, IEEE, Piscataway, NJ **2011**, pp. 31.6.1–31.6.4.
- [19] Y. Wu, B. Lee, H. S. Philip Wong, in *Proc. 2010 Int. Symp. VLSI Technol., Syst., Appl. (VLSI-TSA)*, IEEE, Piscataway, NJ **2010**, pp. 136–137.
- [20] G. W. Burr, M. J. BrightSky, A. Sebastian, H. Y. Cheng, J. Y. Wu, S. Kim, N. E. Sosa, N. Papandreou, H. L. Lung, H. Pozidis, E. Eleftheriou, C. H. Lam, *IEEE J. Emerging Sel. Top. Circuits Syst.* **2016**, 6, 146.
- [21] A. Sebastian, M. L. Gallo, G. W. Burr, S. Kim, M. BrightSky, E. Eleftheriou, *J. Appl. Phys.* **2018**, 124, 111101.
- [22] T. Gokmen, Y. Vlasov, *Front. Neurosci.* **2016**, 10, 00333.
- [23] M. Lanza, H.-S. P. Wong, E. Pop, D. Ielmini, D. Strukov, B. C. Regan, L. Larcher, M. A. Villena, J. J. Yang, L. Goux, A. Belmonte, Y. Yang, F. M. Puglisi, J. Kang, B. Magyari-Köpe, E. Yalon, A. Kenyon, M. Buckwell, A. Mehonic, A. Shluger, H. Li, T.-H. Hou, B. Hudec, D. Akinwande, R. Ge, S. Ambrogio, J. B. Roldan, E. Miranda, J. Suñe, K. L. Pey, et al., *Adv. Electron. Mater.* **2019**, 5, 1800143.
- [24] F. A. Kröger, H. J. Vink, *J. Phys. Chem. Solids* **1958**, 5, 208.
- [25] A. Padovani, L. Larcher, O. Pirrotta, L. Vandelli, G. Bersuker, *IEEE Trans. Electron Devices* **2015**, 62, 1998.
- [26] Y. Guo, J. Robertson, *Appl. Phys. Lett.* **2014**, 105, 223516.
- [27] W. Kim, S. Menzel, D. J. Wouters, Y. Guo, J. Robertson, B. Roesgen, R. Waser, V. Rana, *Nanoscale* **2016**, 8, 17774.
- [28] R. Waser, R. Dittmann, G. Staikov, K. Szot, *Adv. Mater.* **2009**, 21, 2632.
- [29] S. Larentis, F. Nardi, S. Balatti, D. C. Gilmer, D. Ielmini, *IEEE Trans. Electron Devices* **2012**, 59, 2468.
- [30] S. Menzel, M. Waters, A. Marchewka, U. Böttger, R. Dittmann, R. Waser, *Adv. Funct. Mater.* **2011**, 21, 4487.
- [31] A. Hardtdegen, C. La Torre, F. Cuppers, S. Menzel, R. Waser, S. Hoffmann-Eifert, *IEEE Trans. Electron Devices* **2018**, 65, 3229.
- [32] F. Cüppers, S. Menzel, C. Bengel, A. Hardtdegen, M. von Witzleben, U. Böttger, R. Waser, S. Hoffmann-Eifert, *APL Mater.* **2019**, 7, 091105.
- [33] J. Woo, K. Moon, J. Song, S. Lee, M. Kwak, J. Park, H. Hwang, *IEEE Electron Device Lett.* **2016**, 37, 994.
- [34] C. Y. Huang, C. Y. Huang, T. L. Tsai, C. A. Lin, T. Y. Tseng, *Appl. Phys. Lett.* **2014**, 104, 062901.
- [35] D. Y. Cho, M. Luebben, S. Wiefels, K. S. Lee, I. Valov, *ACS Appl. Mater. Interfaces* **2017**, 9, 19287.
- [36] I. Valov, *Semicond. Sci. Technol.* **2017**, 32, 093006.
- [37] Y. Zhao, P. Huang, Z. Chen, C. Liu, H. Li, B. Chen, W. Ma, F. Zhang, B. Gao, X. Liu, J. Kang, *IEEE Trans. Electron Devices* **2016**, 63, 1524.
- [38] W. Wu, H. Wu, B. Gao, N. Deng, S. Yu, H. Qian, *IEEE Electron Device Lett.* **2017**, 38, 1019.
- [39] S. Kim, Y. Abbas, Y. R. Jeon, A. S. Sokolov, B. Ku, C. Choi, *Nanotechnology* **2018**, 29, 415204.
- [40] C. M. M. Rosário, B. Thöner, A. Schönhals, S. Menzel, A. Meledin, N. P. Barradas, E. Alves, J. Mayer, M. Wuttig, R. Waser, N. A. Sobolev, D. J. Wouters, *Nanoscale* **2019**, 11, 16978.
- [41] T. Heisig, K. Lange, A. Gutsche, K. T. Goß, S. Hamsch, A. Locatelli, T. O. Montes, F. Genuzio, S. Menzel, R. Dittmann, *Adv. Electron. Mater.* **2022**, 2100936.
- [42] M. J. Rasch, D. Moreda, T. Gokmen, M. Le Gallo, F. Carta, C. Goldberg, K. El Maghraoui, A. Sebastian, V. Narayanan, in *2021 IEEE 3rd Int. Conf. on Artificial Intelligence Circuits and Systems (AICAS)*, IEEE, Piscataway, NJ **2021**.
- [43] L. Deng, *IEEE Signal Process Mag.* **2012**, 29, 141.
- [44] R. Sun, K. Makise, W. Qiu, H. Terai, Z. Wang, *IEEE Trans. Appl. Supercond.* **2015**, 25, 3.
- [45] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Mater.* **2013**, 1, 011002.
- [46] M. Halter, L. Bégon-Lours, V. Bragaglia, M. Sousa, B. J. Offrein, S. Abel, M. Luisier, J. Fompeyrine, *ACS Appl. Mater.* **2020**, 12, 17725.
- [47] L. Goux, A. Fantini, A. Redolfi, C. Y. Chen, F. F. Shi, R. Degraeve, Y. Y. Chen, T. Witters, G. Groeseneken, M. Jurczak, in *2014 Symp. on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, IEEE, Piscataway, NJ **2014**.
- [48] D. Ielmini, *Semicond. Sci. Technol.* **2016**, 31, 063002.
- [49] D. C. Gilmer, G. Bersuker, H. Y. Park, C. Park, B. Butcher, W. Wang, P. D. Kirsch, R. Jammy, in *2011 3rd IEEE Int. Memory Workshop, IMW 2011*, IEEE, Piscataway, NJ **2011**.
- [50] K.-J. Lee, L.-W. Wang, T.-K. Chiang, Y.-H. Wang, *Materials* **2015**, 8, 7191.
- [51] K. Szot, W. Speier, G. Bihlmayer, R. Waser, *Nat. Mater.* **2006**, 5, 312.
- [52] D. Cooper, C. Baumer, N. Bernier, A. Marchewka, C. L. Torre, R. E. Dunin-Borkowski, S. Menzel, R. Waser, R. Dittmann, *Adv. Mater.* **2017**, 29, 1700212.
- [53] A. Marchewka, R. Waser, S. Menzel, in *Int. Conf. on Simulation of Semiconductor Processes and Devices, SISPAD*, IEEE, Piscataway, NJ **2015**, pp. 297–300.
- [54] A. Schönhals, D. Wouters, A. Marchewka, T. Breuer, K. Skaja, V. Rana, S. Menzel, R. Waser, in *2015 IEEE 7th International Memory Workshop*, IEEE, Piscataway, NJ **2015**.
- [55] H. B. Michaelson, *J. Appl. Phys.* **2008**, 48, 4729.
- [56] L. P. B. Lima, H. F. W. Dekkers, J. G. Lisoni, J. A. Diniz, S. V. Elshocht, S. D. Gendt, *J. Appl. Phys.* **2014**, 115, 074504.
- [57] D. R. Lide, *CRC Handbook of Chemistry and Physics*, Vol. 85, CRC Press, Boca Raton, FL **2004**.
- [58] H. L. Skriver, N. M. Rosengaard, *Phys. Rev. B* **1992**, 46, 7157.
- [59] S. Tirano, L. Perniola, J. Buckley, J. Cluzel, V. Jousseau, C. Muller, D. Deleruyelle, B. De Salvo, G. Reimbold, *Microelectron. Eng.* **2011**, 88, 1129.
- [60] A. Sawa, R. Meyer, in *Resistive Switching* (Eds. D. Ielmini, R. Waser), Wiley, New York, NY **2016**, p. 457.

- [61] A. Kindsmuller, A. Schonhals, S. Menzel, R. Dittmann, R. Waser, D. J. Wouters, in *NVMTS 2018- Non-Volatile Memory Technology Symposium 2018*, IEEE, Piscataway, NJ **2019**.
- [62] Z. Fang, X. P. Wang, J. Sohn, B. B. Weng, Z. P. Zhang, Z. X. Chen, Y. Z. Tang, G. Q. Lo, J. Provine, S. S. Wong, H. S. Wong, D. L. Kwong, *IEEE Electron Device Lett.* **2014**, 35, 912.
- [63] C. D. Landon, R. H. Wilke, M. T. Brumbach, G. L. Brennecke, M. Blea-Kirby, J. F. Ihlefeld, M. J. Marinella, T. E. Beechem, *Appl. Phys. Lett.* **2015**, 107, 02.
- [64] D. C. Sekar, B. Bateman, U. Raghuram, S. Bowyer, Y. Bai, M. Calarrudo, P. Swab, J. Wu, S. Nguyen, N. Mishra, R. Meyer, M. Kellam, B. Haukness, C. Chevallier, H. Wu, H. Qian, F. Kreupl, G. Bronner, in *Technical Digest - International Electron Devices Meeting (IEDM)*, IEEE, Piscataway, NJ **2015**.
- [65] D. B. Strukov, F. Alibart, R. Stanley Williams, *Appl. Phys. A: Mater. Sci. Process.* **2012**, 107, 509.
- [66] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, et al., *Nat. Methods* **2020**, 17, 261.
- [67] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, *Proc. IEEE* **1998**, 86, 2278.
- [68] Z. I. Botev, D. P. Kroese, R. Y. Rubinstein, P. L'Ecuyer, *Handbook of Statistics*, Vol. 31, Elsevier, New York **2013**, pp. 35–59.
- [69] T. Gokmen, W. Haensch, *Front. Neurosci.* **2020**, 14, 103.
- [70] S. Ambrogio, V. Milo, Z. Q. Wang, S. Balatti, D. Ielmini, *IEEE Electron Device Lett.* **2016**, 37, 1268.