

RESEARCH ARTICLE OPEN ACCESS

The Influence of Residual Ion Drift During Programming of Chip-Integrated Nanoscale HfO₂-Based Memristive Devices

Oliver Artner^{1,2}  | Felix Cüppers³  | Seokki Son^{1,2} | Yang Chen^{1,2}  | Xiaohua Liu¹ | Christopher Bengel^{2,4}  | Dennis Nielinger⁵  | André Zambanini⁵  | Stefan Wiefels¹  | Stephan Menzel¹  | Regina Dittmann^{1,2}  | Susanne Hoffmann-Eifert¹ 

¹Peter Grünberg Institute for Electronic Materials (PGI-7), Forschungszentrum Jülich GmbH and JARA-FIT, Jülich, Germany | ²RWTH Aachen University, Faculty of Electrical Engineering and Information Technology, Aachen, Germany | ³JARA-Institute Energy-efficient Information Technology (PGI-10), Forschungszentrum Jülich GmbH and JARA-FIT, Jülich, Germany | ⁴Institute for Materials in Electrical Engineering and Information Technology II and JARA-FIT, RWTH Aachen University of Technology, Aachen, Germany | ⁵Peter Grünberg Institute for Integrated Computing Architectures (PGI-4), Forschungszentrum Jülich GmbH, Jülich, Germany

Correspondence: Oliver Artner (o.solfronk@fz-juelich.de) | Stephan Menzel (st.menzel@fz-juelich.de) | Susanne Hoffmann-Eifert (su.hoffmann@fz-juelich.de)

Received: 6 January 2026 | **Revised:** 3 May 2026 | **Accepted:** 6 May 2026

Keywords: 1T1R structure | filamentary ReRAM | on chip integration | physics-based compact model | transient set analysis | valence change mechanism

ABSTRACT

Filamentary valence change mechanism (VCM)-type memristive devices based on transition metal oxides offer great potential for the realization of energy-efficient analog hardware accelerators used in machine learning and neuromorphic computing. To fully exploit this potential, integration of nanostructured memristive devices with complementary metal oxide semiconductor (CMOS) circuits and multi-level programming are essential prerequisites. In crossbar arrays for in-memory computing, a transistor in series with the VCM device acts as a selector and limits the current in the SET process, which allows programming of different low-resistance states (LRS). However, a discrepancy between the programmed LRS value and the measured conductivity value, G_{LRS} , is often observed, even for VCM devices with linear current–voltage characteristics. In this study, we analyze the physical origin of this effect. Therefore, 100 nm × 100 nm-sized HfO₂-based VCM devices were integrated on foundry-built 180 nm CMOS wafers. Through transient analysis of the device response during the SET, we show that the conductivity of the VCM cell continuously increases for the duration of the SET pulse. This finding could be understood from physical simulation through thermally assisted ion migration in the filament region. This insight can support the development of devices with improved accuracy under multi-state programming.

1 | Introduction

Redox-based resistive random-access memories (ReRAM) and, in particular, filamentary switching valence change mechanism (VCM)-type memristive devices based on transition metal oxides such as HfO₂ and TaO_x have recently received increasing attention for their potential application in energy-efficient accelerators used in neuromorphic computing, image process-

ing, and sensing applications [1–6]. Filamentary VCM devices show fast switching [7, 8], high energy efficiency, and high potential for nanometer-size scaling [9, 10], and they can be integrated into the back-end-of-line (BEOL) of complementary metal oxide semiconductor (CMOS) circuits [11, 12]. After the formation of a conductive filament, the switching behavior of VCM devices is characterized by electric field-induced migration and transfer of oxygen vacancies that is strongly

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Advanced Electronic Materials* published by Wiley-VCH GmbH

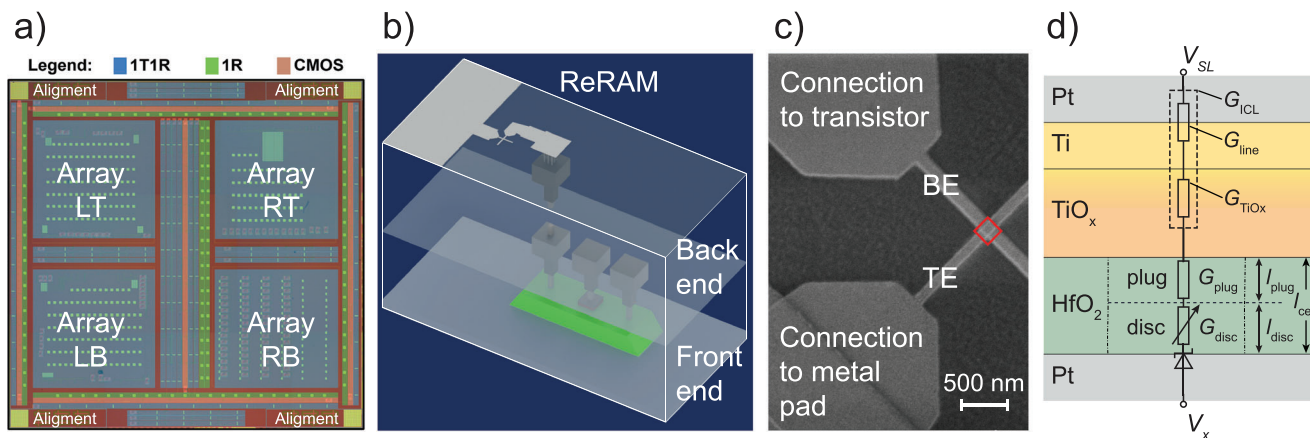


FIGURE 1 | (a) Layout overview of the chip, manufactured in 180 nm CMOS with an edge length of approx. 8 mm × 8 mm. The overlay shows the structure in the chip with three different track types (1T1R in blue, 1R in green, CMOS in orange) and four array structures. (b) Diagram of a nanocrossbar ReRAM device integrated in the back end with a front end foundry transistor, where the active electrode, i.e., the bottom electrode (BE), is connected to the transistor source contact. (c) Scanning electron microscope image of a 100 nm × 100 nm ReRAM device integrated on the CMOS chip. The BE and TE electrodes are connected to the transistor and a measurement pad, respectively. (d) ReRAM stack in the crossbar junction, consisting of 25 nm Pt BE/3 nm HfO₂/3 nm TiO_x/10 nm Ti/20 nm Pt TE. The HfO₂ layer as the switching layer is characterized by its conductive filament, which is defined by the conductive plug, G_{plug} , and the disc area, G_{disc} . The voltage drop across the ReRAM is defined as $(V_{\text{SL}} - V_x)$ according to Figure 2a.

accelerated by Joule heating in the regime of the filament tip [13, 14].

ReRAM devices arranged in arrays can perform computing tasks that are essential for machine learning and neuromorphic computing, such as vector-matrix multiplications very efficiently. One-transistor-one-ReRAM (1T1R) structures are the key elements located at each array node. The use of transistors stems from the need to select the correct device during the programming and reading phase without causing cross-talk to neighboring devices [15, 16]. Another advantage arises when the transistor's ability to control the current is used for the targeted programming of the ReRAM to multiple resistance levels [17–21]. Here, the negative differential resistance (NDR) effect during the SET event of the VCM cell is of particular interest. Recently, the load line analysis that has been introduced for 1-resistor-1-ReRAM (1R1R) structures [22] has been extended to 1T1R structures to investigate the switching behavior of the VCM cell separated from the transistor's characteristic [23–25]. For 1R1R structures, a correlation between the conductivity in LRS, G_{LRS} , and the amplitude of the current used in SET programming, I_{cc} , has been found and has been summarized under the ‘universal switching law’ [26, 27]. Qualitative similar behavior has been reported for 1T1R devices [28–31]. Due to the effect of the ohmic voltage divider in 1R1R structures, the relationship between G_{LRS} and I_{cc} shows a quite high variability. In contrast, in 1T1R structures the maximum value of I_{cc} can be precisely controlled via the voltage applied on the transistor's gate and independently from the voltage applied for switching. Therefore, 1T1R structures are expected to show significantly reduced variability of the G_{LRS} values. However, this is not always the case, which leads to limitations in programming multiple states [11, 20, 25, 32].

In this work, we investigate the switching behavior of filamentary VCM devices in 1T1R configuration, which show an NDR behavior during the SET event. Using a transient analysis, the

voltage on the ReRAM during the SET pulse under constant current load is followed. A detailed analysis reveals a continuous decrease in the voltage drop on the VCM cell following the SET event, which holds despite the constant compliance current. This phenomenon is in detail studied for 1T1R structures consisting of filamentary switching HfO₂-based VCM devices manufactured in the back-end-of-line (BEOL) in series with front-end-of-line (FEOL) transistors of two different sizes (see Figure 1). By comparing the experimental results obtained, further insights into the growth behavior of the conductive filament are gained. The effects of the SET pulse length are discussed in terms of the cycle-to-cycle (c2c) variability of the obtained conductivity state, G_{LRS} , and guidelines for stable operating conditions of the 1T1R elements are derived.

2 | Experiment and Methodology

2.1 | Device Integration on Top of 180 nm CMOS Chips

The ReRAM devices investigated were integrated on top of foundry-built NMOS transistors, using a process compatible with CMOS BEOL manufacturing. For this study, we used a custom-designed silicon chip in 180 nm technology from X-FAB Silicon Foundries SE, for which the processing was stopped before the fifth metal layer (M5). The chip, displayed in Figure 1a, contains transistor structures that allow a fabrication of single 1T1R cells and also arrays of 8 × 4 1T1R structures that enable computation-in-memory (CIM) operations. All test structures in the chip are summarized in the experimental section at the end. The ReRAM crossbar structures are processed with contact pads on the via structures between M4 and M5, building the connection to the front-end transistors as shown in Figure 1b. The fabrication of the 100 nm × 100 nm-sized Pt/3 nm HfO₂/3 nm TiO_x/10 nm Ti/Pt devices was realized in about thirty BEOL-compatible processing steps, which are in detail described in the Experimental Section.

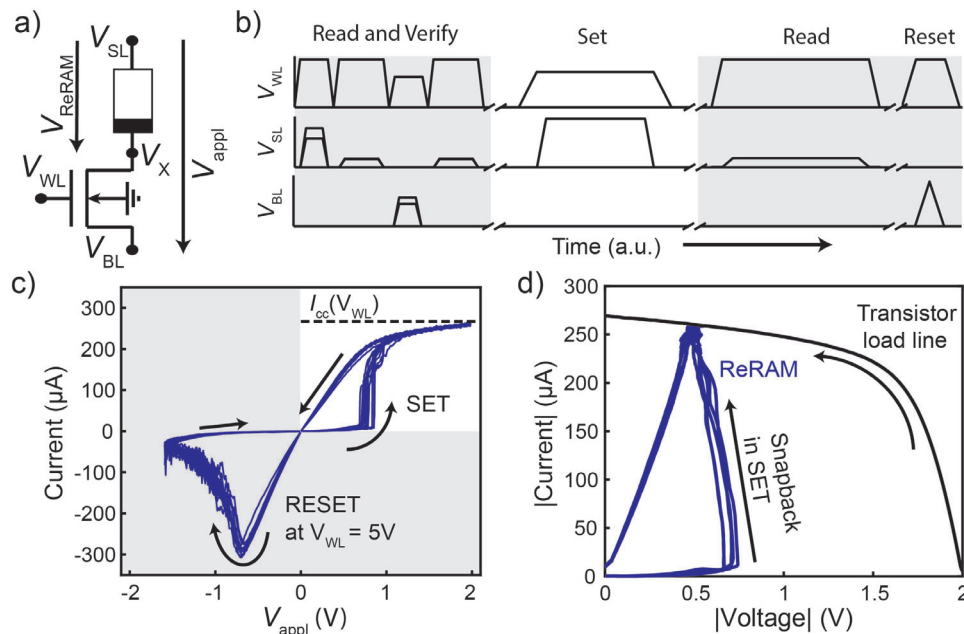


FIGURE 2 | (a) Pictogram of the 1T1R structure with definitions of the voltages: $V_{appl} = V_{SL} - V_{BL}$, V_{WL} is applied between gate and ground, and $V_{ReRAM} = V_{SL} - V_X$. (b) The measurement scheme consists of four blocks: a read-and-verify sequence to program a defined HRS, the transient SET analysis, a read of the programmed LRS, and a reset. (c) Ten switching curves showing measured current versus applied voltage, $I - V_{appl}$, for the set and the reset block. (d) Separated characteristics of ReRAM, $I - V_{ReRAM}$, and transistor load line calculated for the SET events in (c). [Correction added on May 28, 2026, after first Online publication figure 2 was updated in this version].

Processing temperatures did not exceed 350°C and the used materials and chemicals were CMOS-compatible. For the HfO_2 -based VCM devices used here it is important that the resistive switching transition metal oxide layer is sandwiched between a so called ohmic electrode (OE), made from a material of low work function and high oxidation enthalpy (here: Ti), and a chemical inert electrode with high work function called the active electrode (AE, here: Pt) [33]. The TiO_x layer serves as an oxygen vacancy reservoir, which moderates the oxygen-exchange between the HfO_2 switching layer and the Ti OE [34, 35] and enables gradual-type resistance change under defined operation parameters [36]. The resulting 1T1R structure with the AE of the ReRAM cell connected to the source of the transistor can be seen in Figure 1b–d. In this work, NMOS transistors specified for up to 5 V and of the size $W/L = 20$ ($W = 10 \mu\text{m}$, $L = 500 \text{ nm}$), as well as $W/L = 1$ ($W = 10 \mu\text{m}$, $L = 10 \mu\text{m}$) were used. It is worth noting that the structuring method used in this work is somewhat generic and also works with other materials, e.g., TiN instead of Pt, and other processing steps such as deep ultra violet (DUV) lithography.

2.2 | Measurement Scheme and Parameter Analysis

The experiments were performed using a Keithley 4200A SCS equipped with two 2-channel pulse measure units (4225-PMUs) and four remote amplifiers (4225-RPMs). The strategy for bipolar switching of the 1T1R structures (see Figure 2a,b) considers that the used NMOS transistors require positive voltage signals with reference to ground at the bulk contact. To set the ReRAM from a high resistance state (HRS) to a specific LRS, a positive voltage V_{SL} is applied to the source line while the bit line is grounded,

$V_{BL} = 0 \text{ V}$. The I_{CC} -value is controlled by the word line voltage V_{WL} on the transistor's gate. For the reset process, a positive voltage is applied at the bit line ($V_{BL} > 0 \text{ V}$) while keeping the source line grounded ($V_{SL} = 0 \text{ V}$) with the transistor fully opened ($V_{WL} = 5 \text{ V}$). Prior to reproducible resistive switching, as-fabricated ReRAMs required an electroforming step, which was carried out at a current compliance of $325 \mu\text{A}$ ($V_{WL} = 1.4 \text{ V}$) by applying a positive voltage sweep with rate of 1 kV s^{-1} at SL. Typical forming voltages were found at about $(3.14 \pm 0.19) \text{ V}$ for 70 devices. To ensure proper device functionality, one hundred voltage sweeps consisting of set and reset operations were performed prior to the transient SET analysis.

The experimental switching scheme is shown in Figure 2b and is made up of three distinct parts. The first part represents a read-and-verify scheme consisting of set, reset, and read pulses to initialize the ReRAM in a well-defined HRS of about 125 to 175 k Ω . The second part is the transient SET analysis performed under controlled compliance current given by V_{WL} and programmed by sweeping the voltage signal V_{SL} at a rate of 10 kV s^{-1} to a maximum voltage of $V_{appl,max}$ with a pulse-like hold time, t_{hold} , at the maximum. The various waveforms are further explained in the different subsections. In the third part, a low voltage pulse is applied in SET direction to read the LRS conductivity value G_{LRS} . Lastly, a reset is applied before the switching sequence is repeated.

The voltage that actually drops across the ReRAM cell at each point of the switching curve is important for analyzing the SET-switching behavior of the ReRAM operating in any configuration. For 1T1R configurations where the ReRAM is in series with an NMOS transistor as shown in Figure 2a, the $I - V_{ReRAM}$ curve can be calculated based on the "transistor load line" concept [25],

which considers the ReRAM as a variable load of the transistor. From the measured current and actual transfer characteristic of the transistor, the voltage divided over transistor and ReRAM cell can be calculated. The formulas used are:

$$V_{\text{ReRAM}}(t) = V_{\text{appl}}(t) - V_{\text{DS}}(t) \quad (1)$$

$$V_{\text{DS}}(t) = f(I_{\text{total}}(t), V_{\text{Gate,eff}}(t)) \quad (2)$$

$$V_{\text{Gate,eff}} = V_{\text{WL}} + \Delta V_{\text{Gate}} \quad (3)$$

where $V_{\text{appl}} = V_{\text{SL}} - V_{\text{BL}}$ is the voltage applied on the 1T1R structure, $V_{\text{DS}} = V_{\text{x}} - V_{\text{BL}}$ the voltage on the transistor, and $I_{\text{total}}(t)$ the measured total current at any time t . The transistor data are obtained from reference measurements and small deviations between various transistors on a chip are considered by ΔV_{Gate} .

3 | Experimental Results

3.1 | SET Behavior of Filamentary VCM Devices Integrated in 1T1R Structures

The bipolar switching behavior of filamentary VCM devices is characterized by an abrupt SET event originating from a temperature-enhanced drift-diffusion of oxygen vacancies. The high electronic current density, which is reached in the filament volume causes Joule heating that enhances the migration of oxygen vacancies under the applied electric field. For HfO_2 -based cells, the SET polarity results in a reduction of the hafnium oxide in the filament gap, thereby increasing the local conductivity [13, 37]. This self-accelerated process ends when $I_{\text{total}}(t)$ reaches I_{CC} that is defined by the experiment. Simple setups use either an external current compliance of the parameter analyzer or a passive resistor in series with the ReRAM. It has been shown that the compliance current can be significantly distorted by the discharge of parasitic capacitance, which can lead to high current peaks [38]. Therefore, the universal switching law postulated in ref.[37], which suggested a correlation between the maximum current during SET-programming and G_{LRS} , could be proven qualitatively, but not with the accuracy required for storing of multi levels. The co-integration of ReRAM and transistor on the same silicon chip enables parasitic capacitance to be reduced to a minimum [39, 40]. In 1T1R structures the I_{CC} -value in SET is controlled by V_{WL} in combination with V_{DS} . With this experimental setting, different levels of LRS can be programmed using the transistor as current limiter. Limitations in the accuracy of programming 1T1R elements, which can arise from an unsuitable choice of transistor and ReRAM device, range from increased variability of the LRS resistance due to fluctuations in the transistor current to incomplete reset when the LRS value is in the range of the impedance of the fully open transistor [25, 39, 41, 42]. To prevent overloading of the ReRAM, programming/verification schemes were introduced, such as the incremental step pulse program and verify algorithm (ISPVA) and the incremental gate voltage and verify algorithm (IGVVA) [43].

This work begins at a more fundamental level, namely understanding the highly dynamic SET-switching behavior of filamen-

tary VCM devices in series with a nonlinear active device. The SET-switching behavior of the ReRAM cell was derived from the transient analysis of the current response of the 1T1R structure under defined programming schemes shown in Figure 2b. The programming sequence is repeatedly applied starting with the read/verification step of the HRS. The total current is continuously measured. By arranging it in an I - V_{appl} diagram, the switching curve shown in Figure 2c is obtained.

It should be noted that the graph shows the two half-cycles of SET and subsequent reset, while the next complete cycle begins with a read/verify step for the HRS. The I - V_{appl} curves show an abrupt set at approximately $V_{\text{appl}} = 0.8$ V and a gradual reset starting at approximately $V_{\text{appl}} = -0.8$ V. For completeness, the reset curves are shown here, while the following analysis focuses on the SET behavior and the programming of defined LRS.

For the SET programming shown in Figure 2d, the voltage runs through $V_{\text{SL}} = (0.0$ V; 2.0 V; 0.0 V), while the compliance current I_{CC} of approximately 265 μA is controlled by the transistor at $V_{\text{WL}} = 1.4$ V. The contributions of the ReRAM and the transistor can be separated from each other using the algorithm summarized in Equations (1) to (3) [25]. The results are shown in Figure 2d. Here, the absolute value of the voltage refers to either V_{ReRAM} or V_{DS} . In Figure 2d, the SET event also appears at about $|V_{\text{ReRAM}}| = 0.8$ V, as the applied voltage mainly drops across the ReRAM cell in HRS. During the abrupt SET event, the voltage across the ReRAM cell collapses, which appears in the I - $|V_{\text{ReRAM}}|$ curve as snapback. In contrast, this behavior is masked in the I - V_{appl} plot. With the beginning of the snapback, the total current through the 1T1R structure is controlled by the transistor. The I - V_{appl} curve follows the transistor's characteristic until V_{DS} falls below the onset voltage of the transistor. In this range, the ReRAM in LRS acts as a low-resistance load in series with the transistor, which is evident from the linear part in the I - $|V_{\text{ReRAM}}|$ plot in Figure 2c.

3.2 | ReRAM SET Behavior under Different Programming Conditions

The SET-switching characteristics of nanoscale $\text{Pt}/\text{HfO}_2/\text{TiO}_2/\text{Ti}/\text{Pt}$ devices integrated into 1T1R structures were subsequently investigated for two I/O transistors of different width to length ratios, namely $W/L = 20$ and $W/L = 1$. In addition, variations in gate voltage V_{WL} and applied voltage V_{appl} were performed in the experiments. The programming schemes, the SET characteristics of ReRAM (50 cycles), and the respective transistor characteristics are depicted in Figure 3a-f. Furthermore, Figure 3g-i shows the conductivity values of the ReRAM cells in LRS, measured with the transistor gate fully open, i.e., at $V_{\text{WL}} = 5.0$ V, and at a read voltage of $V_{\text{read}} = 0.2$ V applied between SL and BL (c.f. Figure 2b) "Read").

Various influences of the transistor can be derived from the experiments. First, the results for the variation of gate voltage at $V_{\text{appl,max}} = 2.0$ V for the transistor with $W/L = 20$ are discussed, as shown in Figure 3a,d,g. The I - V_{ReRAM} curves in Figure 3d show a SET voltage of approximately 0.8 V, comparable snapback behavior, and comparable values for $V_{\text{ReRAM,hold}}$ for different values of $V_{\text{WL}} = (1.0 - 1.4)$ V. Figure 3g shows the read conductance

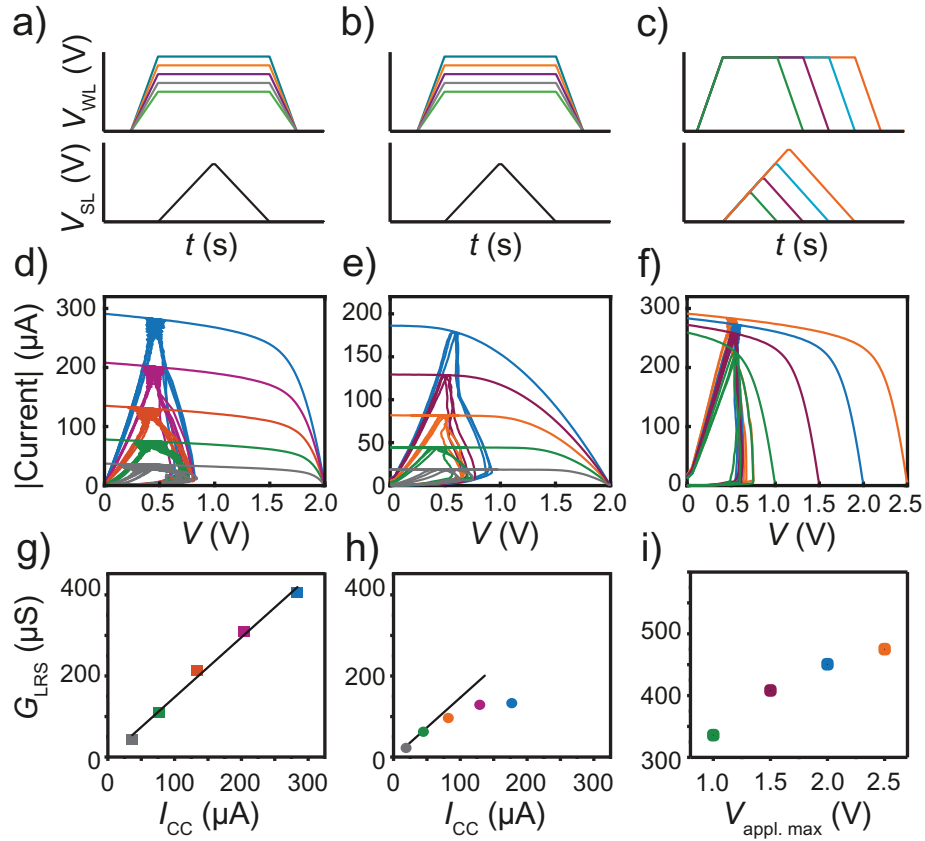


FIGURE 3 | (a–c) show the applied voltage diagrams for both the word line voltage V_{WL} and the source line voltage V_{SL} for the tested programming conditions for both transistors: (a) and (c) refer to $W/L = 20$, (b) to $W/L = 1$. (a) and (b) show a variation of V_{WL} , while in (c) V_{SL} is changed. (d,e) and (f) show the corresponding I - V_{ReRAM} plots with 4 out of 50 repetitions plotted per switching condition, together with the relevant transistor output characteristic fixed at $V_{\text{appl,max}}$. In (g–i) the resulting mean conductivity values are plotted against the mean compliance current for (g) and (h), and against $V_{\text{appl,max}}$ in (i). Furthermore, the linear regression from (g) is transferred to (h).

values G_{LRS} of the ReRAM cell after programming as a function of the maximum current during the SET event, namely I_{CC} . A linear correlation is found up to conductance values of about 400 μS .

A corresponding experiment was performed for the transistor with $W/L = 1$ and the respective results are shown in Figure 3e,h. While the data for G_{LRS} versus I_{CC} in Figure 3h follow the same straight line as in Figure 3g, the achievable conductivity values reach saturation at around 120 μS for larger values of I_{CC} . The origin of this is clearly shown by the separation of the ReRAM and transistor behavior in Figure 3e. For $V_{\text{WL}} > 2.5$ V that corresponds to values of $I_{\text{CC}} > 100$ μA , there is an interruption of the snapback at current values that are significantly below the current limit defined by the transistor in saturation. Due to the moderate increase in transistor current with increasing V_{DS} , current limiting for SET begins within the transition range of the transistor characteristic curve. Another reason for the saturation of the G_{LRS} values is the non-negligible resistance of the transistor in read mode, which limits the maximum value of the total conductance.

The effects discussed above contribute to the results which are shown in Figure 3f,i. The experiment described in Figure 3c is complementary to the one in Figure 3a where the maximum applied voltage is varied between 1.0 and 2.5 V at constant gate voltage. Surprisingly, the results summarized in Figure 3i show

a slight increase of the G_{LRS} values with increasing $V_{\text{appl,max}}$. The issue for the green curve, $V_{\text{appl,max}} = 1.0$ V, in Figure 3i is, that the snapback of the SET process is stopped at a current below the transistor's saturation current due to a too low value of V_{DS} . However, this argument can't explain the increase in G_{LRS} when increasing V_{appl} from 1.5 V to 2.0 V to 2.5 V. For all three experiments the SET event stopped in the transistor's saturation regime. The measured increase in G_{LRS} by approximately 20 % is larger than expected from the slight increase of the saturation current. The main difference which becomes obvious from a comparison of the design of experiments, this is comparing Figure 3a,c, is the total time that is spent for SET. At a fixed slew rate, the half cycle for SET takes longer in time for larger values of $V_{\text{appl,max}}$ compared to smaller ones. Keep in mind that the SET snapback curves appear very much identical in Figure 3f, and that the ReRAM in LRS is under controlled current until the voltage is reduced to 0 V. Further investigations on this finding will be shown and discussed in the next subsection.

3.3 | Influence of Time Under Constant Load on the Final Cell State

To analyze the influence of the SET programming time under constant load on the conductivity of the ReRAM in LRS, time-dependent experiments were performed. Following the

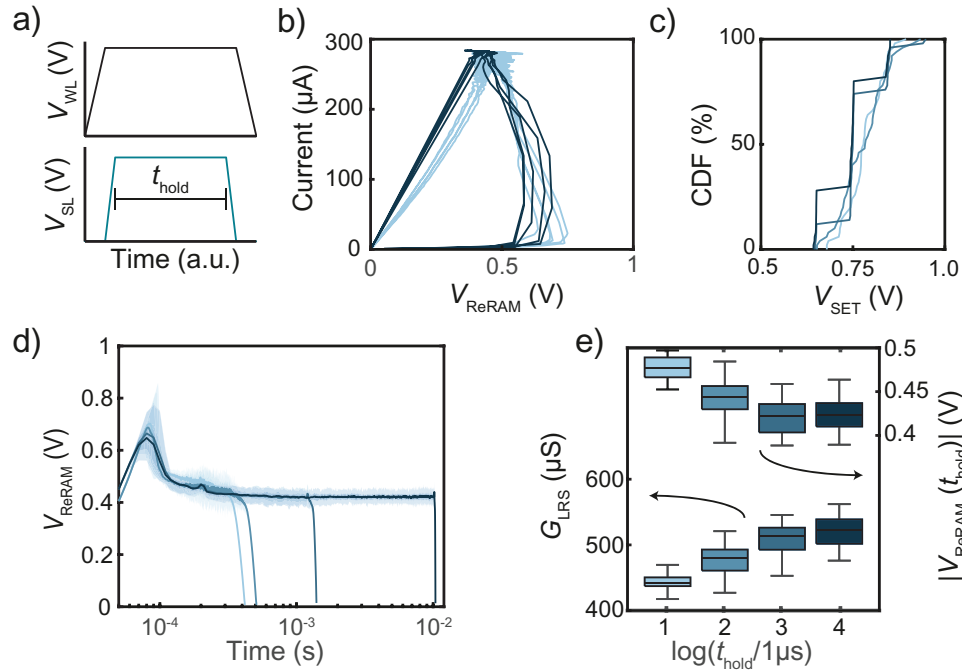


FIGURE 4 | (a) Voltage scheme applied to the source line and the word line of the 1T1R structure while keeping the bit line and the bulk of the transistor grounded. The hold time is varied between $10 \mu s$, $100 \mu s$, $1 ms$ and $10 ms$. (b) 2 calculated $I-V_{ReRAM}$ examples from the $10 \mu s$ (light blue) and the $10 ms$ (dark blue) case. (c) Probability of V_{SET} for all four different cases. (d) $V_{ReRAM}-t$ curves of the different hold time experiments. The darkness of the shade represents the different hold times at the maximum applied voltage. (e) Box plot of the G_{LRS} values extracted at $0.2 V$ applied on the BL and $5 V$ applied on the WL, and the voltage drop on the ReRAM extracted from the $V_{ReRAM}-t$ before the transistor turns off.

programming sequence in Figure 3b, the SET was modified for variable hold times ranging from $10 \mu s$ to $10 ms$, which is realized by a trapezoidal voltage waveform shown in Figure 4a. The experiments shown in Figure 4 were performed with the following parameters: transistor $W/L = 20$, $V_{WL} = 1.4 V$, $V_{SL,max} = 2.0 V$ with a slew rate of $10 kV s^{-1}$. Figure 4b shows $I-V_{ReRAM}$ curves in light and dark blue determined for values of t_{hold} equal to $10 \mu s$ and $10 ms$, respectively. The difference close to the SET event is attributed to a difference in the sampling rate, which was insufficient to resolve the SET event properly for the experiment at $t_{hold} = 10 ms$. In order to confirm that the SET events were comparable, the distributions of the SET voltages extracted for 50 experiments each are shown in Figure 4c. The various sampling rates of 10^6 samples/s and 10^5 samples/s are obvious, while both distribution functions show good agreement overall. The snapback of both $I-V_{ReRAM}$ curves in Figure 4b stops at constant current defined by I_{CC} . The important difference is evident in the different values of the hold voltage, where the snapback effect ceases and, consequently, in the different slopes of the $I-V_{ReRAM}$ curves as they approach zero. This effect is further supported by the trend of the read values G_{LRS} , which increase with increasing programming time under constant load, as shown in Figure 4e.

To resolve the behavior of the ReRAM cell after the snapback in SET, the voltage drop on the ReRAM is plotted as a function of time for the duration of the programming pulse. The transient behavior $V_{ReRAM}(t)$ is shown in Figure 4d. Analog to a box-plot representation, the lighter and darker shaded lines represent the 5th and 95th percentiles and the 25th and 75th percentiles of

all $V_{ReRAM}(t)$ -curves, respectively. The mean $V_{ReRAM}(t)$ -curves for various t_{hold} are shown as solid lines. From Figure 4d it can be seen that the SET time t_{SET} defined by the maximum of $V_{ReRAM}(t)$ is quite comparable for all switching events, which is expected for experiments performed under the same conditions. The snapback in Figure 4b appears as a drop in $V_{ReRAM}(t)$, which follows a stretched exponential decay function over time in Figure 4d. The region that shows the effect of different hold times follows afterwards. Figure 4d clearly shows, that $V_{ReRAM}(t)$ is not constant, but decreases slightly over time at a constant applied voltage $V_{SL}(t) = \text{const.}$ and under constant I_{CC} . Figure 4d summarizes transient measurements for all four hold times as can be seen from the drop in $V_{ReRAM}(t)$ at the end of the $V_{SL}(t)$ pulse. The seamless transition between the different measurement conditions provides further indication that the temporal change of the conductivity value programmed under constant load by means of 1T1R structures has a physical origin inherent to the filamentary VCM-type switching. The read values G_{LRS} programmed for different values of t_{hold} are shown in Figure 4e (left axis) in combination with the voltage at the ReRAM at the end of the programming pulse $V_{ReRAM}(t_{hold})$. This clearly proves the correlation between the decrease in V_{ReRAM} and the increase G_{LRS} with increasing t_{hold} , even under constant load, I_{CC} , defined by the saturation current of the series-connected transistor. From these findings, it could be concluded that the SET-process, which is characterized by voltage- and temperature-controlled drift-diffusion of oxygen vacancies, did not stop abruptly after the snapback had occurred. On the contrary, after controlled switching to LRS, the conductivity of the ReRAM cell increases steadily as long as current flows.

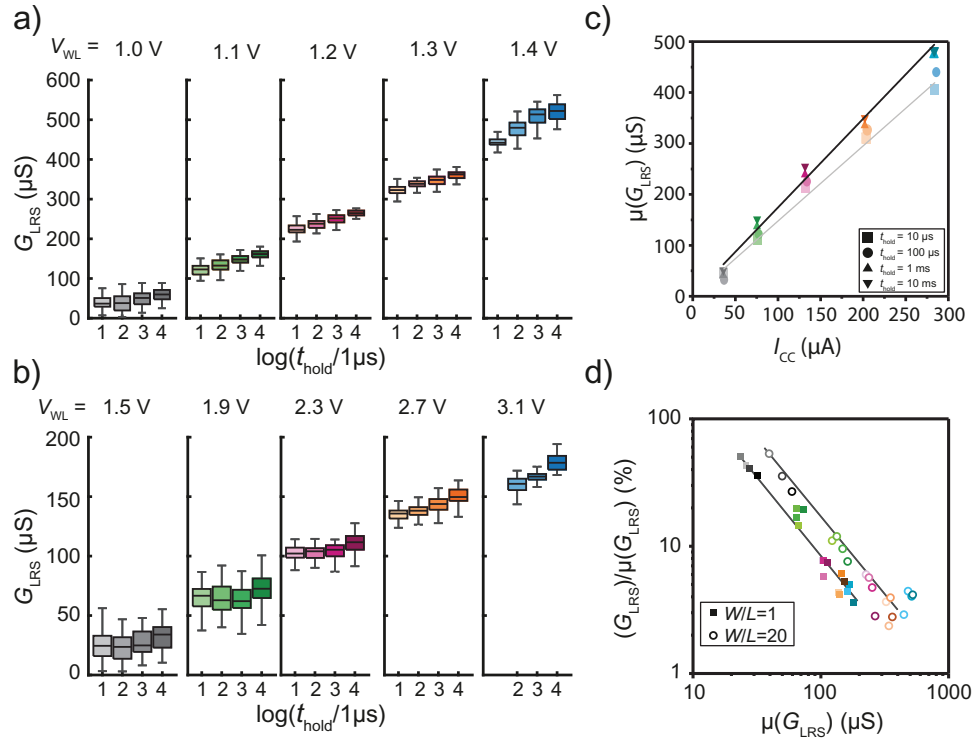


FIGURE 5 | (a) and (b) show the G_{LRS} read values of 1T1R structures for transistors with $W/L = 20$ and $W/L = 1$, respectively, obtained from SET experiments (c.f. Figure 4) performed with various gate voltages V_{WL} and four different hold times t_{hold} , namely 10 μs, 100 μs, 1 ms, and 10 ms. The box plots summarize the statistical distributions from 50 repetitions each. In (c), the mean values $\mu(G_{LRS})$ for the 1T1R structure with $W/L = 20$ are plotted over the I_{CC} values (corresponding to the V_{WL} -values) with t_{hold} coded in the symbols according to the legend. (d) shows the coefficient of variation, namely $\sigma(G_{LRS})/\mu(G_{LRS})$, versus the mean value. The colors of the open and closed symbols follow the codes in (a) and (b), respectively.

3.4 | Influence of Transistor Characteristic and Current Range

Transient analyzes of the ReRAM response during programming into the LRS under constant load, controlled by a transistor in series, which were performed for two different transistors, $W/L = 20$ and $W/L = 1$, for different compliance currents $20 \mu A \leq I_{CC} \leq 300 \mu A$, controlled by V_{WL} , and for different hold times with $10 \mu s \leq t_{hold} \leq 10$ ms, are shown in Figure 5a,b and illustrate the G_{LRS} read values measured at 0.2 V as box plots, each comprising 50 repetitions of the programming algorithm in Figure 2b. Furthermore, Figure 5c shows G_{LRS} as a function of the measured I_{CC} values for the transistor with $W/L = 20$. An increase in G_{LRS} with increasing length of the programming pulse can be observed in all cases, regardless of the transistor characteristics and current ranges. Plotting a linear regression for data points with the same hold time results in straight lines with slightly different slopes, as shown in Figure 5c for the two hold times of 10 μs and 10 ms. As the pulse length increases, the slope of the line increases, which corresponds to a decrease in $V_{ReRAM}(t_{hold})$ as t_{hold} increases. Thus, the length of the programming pulse t_{hold} affects the conductance change ΔG_{LRS} obtained for a certain ΔI_{CC} .

In addition, the cycle-to-cycle variability of the programmed LRS is examined using the coefficient of variation CoV, which is defined as the ratio of the variability σ and the mean value μ : $CoV(G_{LRS}) = \sigma(G_{LRS})/\mu(G_{LRS})$. For the measured values of G_{LRS} in Figure 5d shows the percentage CoV-value as a function the

mean value of G_{LRS} . In the double logarithmic representation, there is an almost linear relationship between the cycle-to-cycle $CoV(G_{LRS})$ and $\mu(G_{LRS})$. Two observations can be made: First, two straight lines can be seen, corresponding to the two transistors with different gate lengths. Second, the values determined for a particular transistor lie almost entirely on the straight line, regardless of the length of the SET pulse. Similar behavior can be found in refs. [39, 44–46]. According to the explanations in ref. [44], this suggests a uniform mechanism that controls the SET process under constant load and the variability of the programmed states.

3.5 | Influence of Current Changes During SET Programming

The analyzes discussed in the previous sections show a continuous slight increase in the conductivity of the ReRAM cell after the main SET process, which continued during the duration of the programming pulse. The voltage drop across the ReRAM cell after snapback, in combination with the programming current, provides sufficient power to maintain a moderate ion displacement. Even at low compliance currents, no “freezing” of the ion current was observed. This raises the question of how low the current has to be that progressive ion displacement in the time domain of the SET pulse is no longer significant. The 1T1R structure, with its flexibility to control the programming voltage and current independently, offers the possibility to conduct experiments to answer this question.

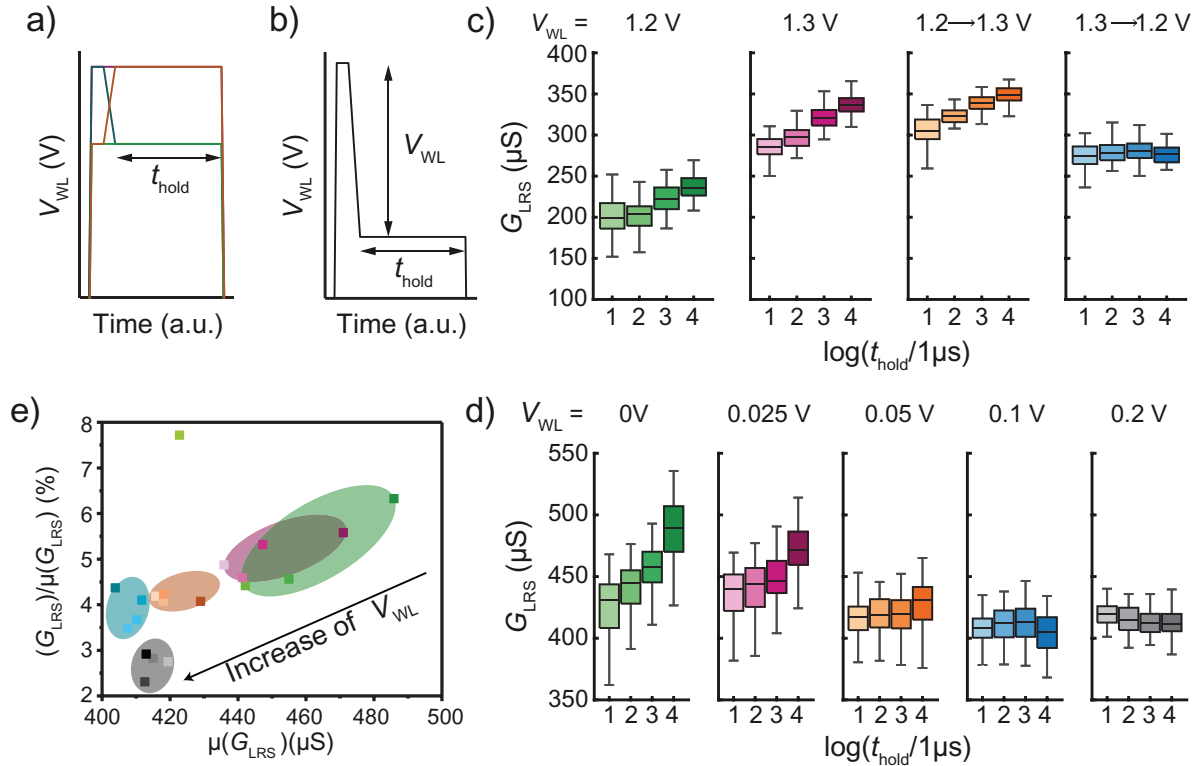


FIGURE 6 | (a) and (b) show the voltage schemes applied on the word line. When the source line hits the maximum applied voltage of 2 V, V_{WL} is varied. Afterwards, the voltages are applied for 10 μs , 100 μs , 1 ms and 10 ms, respectively. (c) and (d) show the extracted $G_{LRS,read}$ box plots of the different cases. (e) shows the cycle-to-cycle coefficient of variation for the five different cases in (d).

Therefore, experiments were designed to increase and decrease the compliance current between two levels, $I_{CC,low}$ and $I_{CC,high}$ during the SET pulse by controlling the gate voltage, as shown in the $V_{WL}(t)$ -profiles in Figure 6a,b. The specified programming scheme begins with a given value of $V_{WL,1}$. When $V_{SL}(t) = 2.0$ V is reached (c.f. Figure 2b), the gate voltage is changed to $V_{WL,2}$, which is kept constant for the remaining pulse duration of 10 μs , 100 μs , 1 ms, and 10 ms in total. The experiments shown in Figure 6a,c deal with four different waveform variants, each of which was tested for 50 cycles on the same device: (1) constant $V_{WL} = 1.2$ V (green), (2) constant $V_{WL} = 1.3$ V (pink), (3) increase of V_{WL} from 1.2 to 1.3 V (orange) and (4) decrease of V_{WL} from 1.3 to 1.2 V (blue). The corresponding box plots for the measured values of G_{LRS} in Figure 6c show the following: Experiments (1) and (2) conducted as references for $I_{CC,low}$ and $I_{CC,high}$, respectively, show the same trend as the previous experiments (c.f. Figure 5a). The results of experiment (3) show G_{LRS} values that are close to the values in experiment (2), even though the SET-event was performed with a lower current $I_{CC,low}$ until the end of the snapback and then increased to $I_{CC,high}$. This suggests that increasing the compliance current towards the end of the SET pulse amplifies the change in LRS due to the hold time. In experiment (4), the compliance current was reduced to the low value $I_{CC,low}$ as soon as $= 2.0$ V was reached. The resulting values of G_{LRS} are slightly lower than the lowest values in experiments (2) and (3), but significantly higher than the highest values in experiment (1). The experiments in Figure 6a,c show that the LRS of the ReRAM cell is always influenced by the highest current occurring during the SET pulse. An increase in G_{LRS} is observed when the current is increased after the main SET event, but

still within the SET pulse. Conversely, however, the continuous increase in G_{LRS} during the duration of the SET pulse can be suppressed if the compliance current is reduced immediately after the SET event within the SET pulse. A detailed analysis of the results of experiment (4) in Figure 6c was performed using the experiments described in Figure 6b,d. Here, the series starts with a set at $V_{WL} = 1.4$ V, i.e., at an increased $I_{CC,high}$ compared to the previous experiment. The G_{LRS} values obtained for the reference waveform, in which is kept constant at 1.4 V, are shown in green in Figure 6d. The following waveforms (pink, orange, blue, gray) consider different values for ΔV_{WL} , namely 25 mV, 50 mV, 0.1 V, and 0.2 V, respectively. The results in Figure 6d show that the increase in G_{LRS} with increasing t_{hold} can already be attenuated if the gate voltage is reduced by 25 mV after the SET event. A larger reduction in gate voltage leads to a stronger slowdown in the increase. From a reduction of approximately 0.1 V, the recorded G_{LRS} values are almost constant for all pulse lengths.

These experiments show that LRS states with precisely defined G_{LRS} values can be programmed by controlled reduction of the total current after the SET event, but still within the SET pulse. The improved cycle-to-cycle reproducibility of LRS becomes apparent when looking at the CoV-values for the G_{LRS} data from Figure 6d, which are shown in Figure 6e. Analogous to Figure 5d, the percentage ratio $\sigma(G_{LRS})/\mu(G_{LRS})$ is plotted above the $\mu(G_{LRS})$. Contrary to the increasing cycle-to-cycle CoV with decreasing $\mu(G_{LRS})$ observed in Figure 5d, the set programming described in Figure 6b led to reduced CoV-values with increasing ΔV_{WL} , even for lower values of $\mu(G_{LRS})$. The results underscore how critical it is for the reproducibility of the written LRS value from

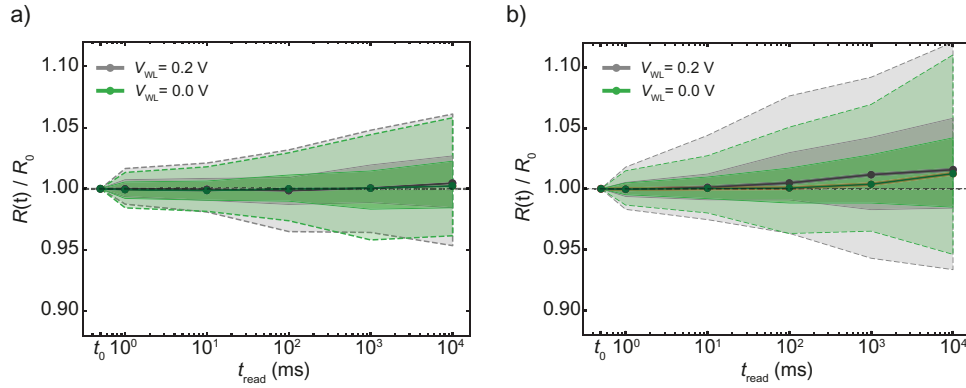


FIGURE 7 | Stability measurements performed for the two differently programmed LRS, namely without and with reduction of the WL voltage, this is $\Delta V_{WL} = 0.0$ V in green and $\Delta V_{WL} = 0.2$ V in grey color, respectively: (a) short term stability measurements and (b) read disturb measurements.

cycle-to-cycle to control ion migration in the filament during the SET-switching process.

3.6 | Effect on the Stability of the Programmed Conductance States

The previous section demonstrated that low-resistance states programmed by reducing the gate voltage of the transistor after the SET event exhibit smaller cycle-to-cycle variations than those programmed by reducing the source-drain voltage while keeping the gate voltage constant. However, the abrupt reduction in the programming current could lead to a less well-controlled filament shape. To analyze the stability of the various programmed states, additional measurements of short-term stability and read disturb were therefore performed, both for standard programmed states and for states achieved by reducing the transistor's gate voltage.

In the first case, a voltage sweep as seen in Figure 4a with one of three wordline voltages $V_{WL} = 1.1$ V to $V_{WL} = 1.3$ V and a hold time of 100 μ s is used. In the second case, the proposed algorithm as plotted in Figure 6b is applied, using the same V_{WL} values and hold time. For both, short term stability and read disturb measurements, a read pulse of 0.2 V and $t_{read} = 10$ μ s is applied directly after programming while keeping the transistor fully open ($V_{WL} = 5$ V). Afterwards, read pulses are applied every decade between 10 ms and 10s. Between the read pulses, the voltage is held at 0 V in case of the short-term stability and at 0.2 V for the read disturb characterization. To test the stability in both set and reset direction, the pulses are applied on either BL or SL. The resulting resistance states are compared to their original states R_0 at t_0 (after the programming) for each time step. Each V_{WL} and read operation in the two directions is repeated 100 times. The analyzed results are plotted in Figure 7 in a manner similar to a box plot. Here, the lighter (outer) area represents the 5th and 95th percentiles, while the darker (inner) shaded area represents the 25th and 75th percentiles, respectively, of all read measurements. The mean values are given by the solid circles connected by the solid lines. Regarding short term stability of the written states the distributions obtained by the two programming methods shown in Figure 7a almost overlap. However, regarding the read disturb behavior, the states programmed w/o reduction

of the WL voltage show a reduced widening of the distribution compared to the states that were written with reduction of the WL voltage. The experiments indicate that the relaxation-like process of filament formation that is obtained in standard 1T1R programming scheme lead to filaments with an increased stability by about 2% compared to states programmed by reduction of the WL voltage.

4 | Physical Understanding of the Pulse Time-Dependent LRS in 1T1R

The SET process in filamentary switching VCM-type devices is understood from a thermally accelerated drift-diffusion of ionic defects in the area of the filament due to Joule heating [13]. The thermal amplification supports the nonlinear relationship between switching time and applied voltage, enabling extremely short switching times down to the ps range [7, 36]. Unfortunately, the runaway of the current during SET causes significant cycle-to-cycle variability in LRS for single-step programmed ReRAM devices. To mitigate severe current overshoots, compliance is required to limit the maximum current during SET operation. In 1T1R structures, compliance is determined by the total current controlled by the transistor in series. However, the results of this study on the transient SET-behavior of on chip-integrated nanoscale ReRAM cells show a clear switching aftereffect in the form of a continuous increase in the conductivity of the ReRAM cell as long as the SET pulse continues and the compliance current flows through the cell, albeit at a lower voltage. The experimental results suggest the influence of a residual ion drift, which persists even at reduced voltage. The cause of this change in conductivity is likely to be a combination of the residual temperature after the SET event and the load exerted on the cell. This finding is supported by results from a physical model. For this purpose, the simulation results for an alternative filament-like VCM switching system, specifically Pt/TaO_x/Ta/Pt components, which were discussed in detail in reference [47], are considered below for a qualitative comparison, using the simulation parameters as given in Table 1 of reference [47].

The influence of temperature in the filament region on the switching behavior of ReRAM cells integrated into 1T1R structures was simulated using the recently introduced advanced

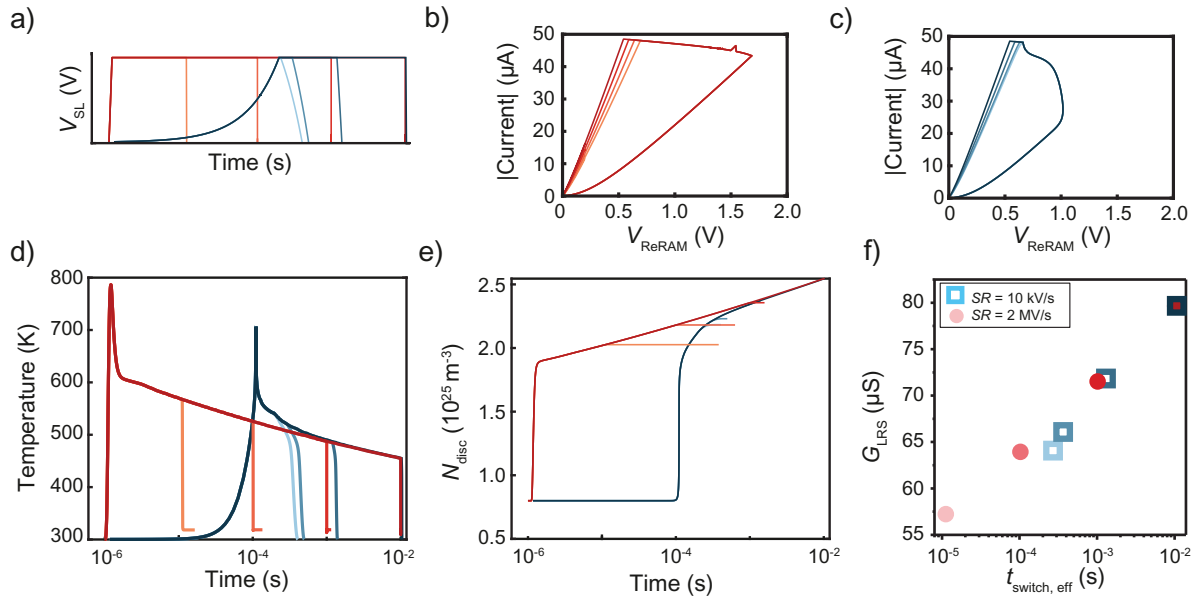


FIGURE 8 | Simulations of hold time experiments with two different sweep rates (SR): 10 kVs^{-1} in blue and 2 MVs^{-1} in red with $I_{CC} = 50 \text{ } \mu\text{A}$. The length of t_{hold} is shown in the shade of the color, where a darker color means a longer hold time. (a) Voltage scheme for both sweep rates. (b) and (c) $I_{\text{total}} - V_{\text{ReRAM}}$ curves for the two sweep rates, respectively. (d) Transient temperature curves for both simulations, showing the highest temperature at V_{SET} with a continuous decrease with increasing time afterwards. (e) Transient concentrations of oxygen vacancies in the disc region, showing a sudden increase at V_{SET} with a continuous increase with time afterwards. (f) Calculated G_{LRS} versus the effective switching time $t_{\text{switch,eff}}$.

compact model for filamentary switching VCM-type memristive devices. The “JART VCM R_{th} Compact Model” [47] is an extension of the established “JART VCM v1b” model, which has already been successfully used to describe the switching behavior of $\text{HfO}_2/\text{TiO}_x$ -based nano-crosspoint devices [48]. As an extension, the “JART VCM R_{th} Compact Model” takes into account a state-dependent effective thermal resistance $R_{\text{th,eff}}$ that depends on the actual oxygen vacancy concentration in the disc area, namely N_{disc} . In detail

$$R_{\text{th,eff}}(t) = \frac{\beta}{\alpha + \gamma N_{\text{disc}}^2(t)} \quad (4)$$

with the coefficients α , β , and γ . This $R_{\text{th,eff}}$ affects the local temperature T in the area of the disc, which is calculated from the power dissipation due to Joule heating within the area of the filament:

$$T(t) = V_{\text{ReRAM}}(t) \cdot I_{\text{total}}(t) \cdot R_{\text{th,eff}}(t) + T_0 \quad (5)$$

with $V_{\text{ReRAM}} = V_{\text{disc}} + V_{\text{plug}} + V_{\text{Schottky}}$ (see Figure 1d) and $I_{\text{total}} = I_{\text{elec}} + I_{\text{ion}}$. The change in N_{disc} is defined by the ionic current I_{ion} , which depends exponentially on the temperature T , and is therefore influenced by changes of the state-dependent $R_{\text{th,eff}}$. Conversely, a change in N_{disc} is fed back into $R_{\text{th,eff}}$, causing the state-dependent variables in the model to mutually affect each other. The complete equation system is given in the original work [47]. For the simulations in this work, the same parameter set as in the original work was used.

Figure 8 shows the results for two different sweep rates, namely 10 kVs^{-1} (blue) and 2 MVs^{-1} (red). The slower sweep is close to the experimental setting, while the faster sweep rate is chosen to reduce the influence of the rise time with respect to the hold

time. The various voltage signals $V_{\text{SL}}(t)$ for the two sweep rates and four values of t_{hold} are given in Figure 8a. The $I_{\text{total}} - V_{\text{ReRAM}}$ curves in Figure 8b,c clearly show the characteristic snapback-effect of filamentary VCM devices. The different SET voltages obtained for different sweep rates is due to the nonlinear SET kinetics behavior. Furthermore, different conductance states for the various t_{hold} -values can already be seen in the $I_{\text{total}} - V_{\text{ReRAM}}$ curves after snapback. These simulation results correspond very well with the experimental observations. The simulation now offers the possibility to extract the time behavior of physical parameters during the SET process such as the temperature in the filament $T(t)$ and the oxygen vacancy concentration in the disc region $N_{\text{disc}}(t)$, which are shown in Figure 8d,e, respectively. During snapback, an abrupt rise in T and in N_{disc} can be observed. More important, however, is the steady decline of $T(t)$ and the steady increase of $N_{\text{disc}}(t)$ after the snapback, which continue until the end of the programming pulse. Interestingly, the different slew rates have only a minor influence on the final values of N_{disc} for longer pulses. Figure 8f shows the resulting G_{LRS} values as a function of the effective SET time, $t_{\text{switch,eff}}$. The increase in G_{LRS} with increasing $t_{\text{switch,eff}}$ is in good agreement with the experimental findings shown in Figures 5a,b.

5 | Discussion

The SET switching behavior of ReRAM devices integrated with transistors in 1T1R structures on the same chip is considered to be ideally controlled. Integration on the same chip reduces capacitive charging effects that can lead to unwanted current spikes during SET events. Furthermore, a transistor operated in saturation acts as a nearly ideal current limiter. Based on the idea of ideal current compliance, the universal switching law [37]

formulates a direct correlation between the compliance current I_{CC} during SET and the resulting conductivity G_{LRS} of the ReRAM in LRS. However, often, a significant cycle-to-cycle variability in LRS is observed limiting the number of distinguishable multilevel states, for example, in CIM applications.

To investigate and understand the transient behavior during the SET switching event of standard filamentary switching VCM-type memristive devices, a CMOS-chip has been designed and manufactured based on standard 180 nm bulk silicon technology, which offers the possibility for integrating nanoscale crossbar-structured memristive devices in a BEOL process on top of the chip with differently sized BEOL NMOS transistors. This work about the transient analysis of the SET behavior of filamentary VCM-type devices used nanoscale Pt/HfO₂/TiO_x/Ti/Pt cells integrated into 1T1R structures on chip using two different transistors, namely $W/L = 20$ and $W/L = 1$. The $I-V_{ReRAM}$ curves, derived by the algorithm in ref.[25], showed the characteristic snapback effect in SET, which is understood by field- and temperature-enhanced acceleration of the ion migration. A closer examination of the region behind the snapback revealed a steady drop in the voltage across the ReRAM, which continued for the duration of the programming pulse at constant current, I_{CC} , controlled by the voltage at the transistor gate. At the circuit level, this continuous change in the state of the ReRAM cell during SET under constant load is reflected in the dependency of the read value of the conductivity in LRS on the SET pulse length or the hold time: $G_{LRS}(t_{hold})$.

An important finding of this study is that an influence of residual ion drift after the snapback was observed at I_{CC} -values in the range of approximately 20 to 600 μ A. This indicates a common physical mechanism and suggests that there is no threshold value below which the $G_{LRS}(t_{hold})$ dependence is no longer observed. Considering the cycle-to-cycle variability of the conductivity in LRS programmed by single-pulse SET events, a linear decrease of $CoV(G_{LRS})$ with increasing $\mu(G_{LRS})$ on the log-log scale has been found, regardless of t_{hold} or I_{CC} . These findings suggest a uniform mechanism that controls the set process under constant load. Hence, SET events with lower current compliance, which lead to more fragile filaments, offer more possibilities for different ion configurations, and, therefore an increased cycle-to-cycle variability of G_{LRS} , in agreement with previous reports [39, 44–46].

The time-dependent investigation of the ReRAM behavior under the SET pulse clearly shows that ion migration within the hot filament continues after the snapback effect in the SET, which is often mistakenly referred to as a set event in $I-V_{appl}$ -curves, even at the then reduced voltage across the ReRAM cell, until V_{appl} is reduced to such an extent that the transistor no longer supplies any current.

In contrast, the residual ion drift after snapback in the SET could be successfully reduced by a controlled reduction of the total current during the SET pulse immediately after the SET event. In this way, it was even possible to achieve a reduction of the $CoV(G_{LRS})$ values of approximately 50 %, from around 6 % to around 3%, even though the $\mu(G_{LRS})$ values were reduced from around 480 mS to around 410 mS at the same time. The significantly lower variation from cycle-to-cycle is accompanied by a slight decrease in the stability of the filaments obtained

using the programming method that reduces the WL voltage compared to standard programming. Nevertheless, the reduction in variation by approximately 50% outweighs the loss of stability of about 2%.

The proposed programming scheme can be reproduced by the use of a write termination circuit, i.e., a circuit that can accurately sense the resistance change of a memristor and terminate the voltage pulse once such a resistance change occurs [49–54]. Next to the reduction of resistance variability, a reduction of the required switching energy can be achieved at the cost of a larger area needed to implement the circuit. Implementation in large memristive arrays helps to reduce the ratio percentage of memristive array and write circuits, where the impact of the write circuit on the memory area can be reduced down to 10% for a 400 x 4 array [49]. To be able accurately program the resistance states, the latency of the circuit should be as small as possible. In case of the write termination circuit in ref.[50], a latency of 2.8 ns has been achieved. Similar circuits show a latency of 2.2 ns [54], 1.8 ns [52] and 5 ns [53]. All referenced write termination circuits, which are based on technology nodes between 22 and 150 nm, show a latency below 10 ns. Therefore, the transition to the here employed technology node should not significantly increase latency. Comparing the latency of these devices to the proposed current reduction algorithm, where approximately 100 μ s are between the snapback process and the reduction of the voltage applied on the wordline, it can be assumed that the write termination is quick enough to implement the used algorithm, and thereby decrease the cycle-to-cycle variability of the memristor.

Qualitatively, the experimental results were well-matched by simulation results using a very recently reported compact model, namely the “JART VCM R_{th} Compact Model” [47], which takes into account a state-dependent effective thermal resistance $R_{th,eff}$ that depends on N_{disc} . This correlation supports the physical description that the growth of the filament in the SET process continues beyond the abrupt switching event as long as the SET pulse is applied. Due to the voltage divider between the ReRAM cell and the transistor, the value of the maximum applied voltage, V_{appl} , is not decisive as long as the transistor is in its saturation regime during the SET event. However, in addition to the value of I_{CC} , the duration of the pulse t_{hold} is important for the programmed G_{LRS} . As the simulation does not give a direct information of the filament length and radius due to the restrictions of the compact model, changes in the filament size are expressed by changes of oxygen vacancy concentration in the disc region. After the snapback, the concentration steadily increases linearly, showing further growth of the filament in the vertical direction from the OE toward the AE, explaining the increasing conductance over time as a direct result of the decreasing disc length. In comparison, due to the reduction of the current and thereby the shut off of the switching process in the proposed programming scheme, the uncontrolled and unintentional filament growth in the vertical direction is cut short, reducing its influence on the variability of the programmed state, while keeping the stability similar to normal programmed states.

Additionally, the simulation results allow a deeper understanding of the behavior of the cell during the pulse. During the hold time, the electric field inside the memristor follows the voltage

dropping on the memristor, as a constant plug length is assumed in the simulation while the concentration of oxygen vacancies increases in the disc area. Simultaneously, the elevated temperature in the cell results in the observed ion/conductance drift, as shown by calculating the ionic current under constant electrical field at room temperature T_{amb} (like in the case of an applied read pulse) and at an elevated temperature T_{hold} (the temperature given in Figure 8). The following relation is obtained: $\frac{I_{\text{ion}}(T_{\text{hold}}, E)}{I_{\text{ion}}(T_{\text{amb}}, E)} = \exp\left[\frac{\Delta W_a}{k_B} \cdot \left(\frac{1}{T_{\text{amb}}} - \frac{1}{T_{\text{hold}}}\right)\right] \sim 130 \cdot 10^6$, using $\Delta W_a = 1.3$ eV, $T_{\text{amb}} = 293.16$ K and $T_{\text{hold}} = 460$ K [47]. Due to the exponential increase of ionic current with temperature under constant electrical field, controlling the temperature after the SET process is of utmost importance to decrease variability by unwanted filamentary growth.

6 | Conclusion

In this work, the time-dependent change of the ReRAM conductivity during SET programming was analyzed for 1T1R structures. For this purpose, nano crossbar ReRAM cells were integrated on a chip designed and manufactured in 180 nm, 5 V silicon bulk technology. For two different transistor characteristics and for programming currents between 20 and 600 μA , it was shown that the LRS conductivity of the ReRAM cell increases with increasing time under constant load. By calculating the voltage drop on the ReRAM, it could be shown that the increasing conductance relates to a decreasing V_{ReRAM} with a stretched exponential decay behavior under constant current compliance. The physical understanding is by means of residual ion drift under reduced voltage enabled by elevated temperature in the filament area, which is caused by the decaying high temperature of the abrupt SET and Joule heating due to the constant current on the cell. These aftereffects result in LRS conductivity values that depend on the pulse length of the SET pulse for filamentary VCM devices operated in 1T1R structures. This hypothesis was validated by simulations performed using the JART VCM R_{th} Compact Model. The present work underlines the important role the local temperature in the filament area has on the cycle-to-cycle variability of LRS conductivity values programmed for ReRAM cells integrated in 1T1R structures. In addition to the physical understanding of the transient behavior of ReRAM devices during SET programming in a 1T1R configuration, a programming algorithm is proposed in which the current through the device is reduced by decreasing the gate voltage immediately after reaching the maximum applied voltage. By this, a reduction in the $\text{CoV}(G_{\text{LRS}})$ by approximately 50 % was achieved. This underscores that for filamentary-switching VCM cells, temperature control during the SET is an important parameter for optimizing cycle-to-cycle variability in VCM-type ReRAM devices.

7 | Materials and Methods/Experimental Section

7.1 | Test Chip Design

The chip was designed in a 180 nm 5 V CMOS technology from X-FAB Silicon Foundries SE. Its focus was to provide a test platform for various memristor-CMOS integration tests, hence it

contains different test structures. Contacting the chip was done through the vias connecting metal layers 4 and 5. The processing was stopped before metal 5 deposition, exposing via blocks of $2.34 \mu\text{m} \times 2.34 \mu\text{m}$. The post processing integrated then bigger pad structures of $80 \mu\text{m} \times 80 \mu\text{m}$ for positioning of probe tips and smaller structures of $5 \mu\text{m} \times 5 \mu\text{m}$ to provide an interface to the ReRAM structures.

Two different groups for pad layouts exist on the chip, as it can be seen in Figure 1a: an array structure having the pads oriented in a square of $3 \text{ mm} \times 3 \text{ mm}$ with 16 pads per edge, resulting in 64 contacts per array structure. Each structure allowed memristive crossbar arrays with $x \times y$ 1T1R combinations in a matrix configuration. LT consisted of four 1T1R subarrays (three 8×4 and one 7×4), all with the same W/L ratio in one subarray, but with different ratios between subarrays. LB consists of a 3×6 dual-bitline array, a 4×8 pseudo-crossbar array, an 8×4 1T1R array and an 8×4 nano-1T1R array with minimal transistor dimensions. RT was a single 31×16 1T1R array. RB consisted of four 16×1 line arrays with different transistor sizes. In all arrays, the empty areas were used for 1R cells.

The remaining space on the chip was filled with groups of four pads for additional device testings, bundled together in tracks of different lengths and orientations to optimally fill the remaining space. Three categories of tracks are included: CMOS test structures enable an insight into variations and post-processing effects on all areas of the chip. They included two ring oscillators of different eigenfrequency, a constant gm cell, four resistive strings, as well as differently dimensioned PMOS and NMOS transistors as they were used in the circuits and the 1T1R combinations. Of these single transistors, four core PMOS and NMOS (1.8 V) and five I/O PMOS and NMOS (5.0 V) were used. In the next track category, single memristor (1R) tracks integrate up to 22 pad pairs to characterized the memristor itself without any active CMOS component and just metal connections between the two contact structures. Finally, the majority of tracks were devoted to one memristor - one transistor combinations (1T1R) with different transistor dimensions and up to 11 1T1R structures per track. Some of the 1T1R structures included a simple ESD structure at the gate, protecting the transistors from electrostatic post-processing effects. All transistors intended to interface a memristor use I/O transistors specified for voltages up to 5 V. In total, the chip contains 47 CMOS test structures, 378 1R structures, and 950 1T1R structures.

7.2 | Memristor Fabrication

The VCM devices were fabricated instead of metal layer 5 of the 180 nm CMOS chip by X-FAB Silicon Foundries SE. The process started with the fabrication of contact pads, partially on the tungsten via connecting to the transistors underneath, and e-beam markers from DC sputtering of 5 nm Ta and 50 nm Pt at room temperature. Patterning was done by laser-lithography using a mask-less aligner and the structuring was done by a lift-off process in acetone. Afterwards, the $100 \text{ nm} \times 100 \text{ nm}$ sized filamentary VCM cells were fabricated between the contact pads. To this end, first the bottom electrodes (BE)

were fabricated by DC sputtering of 5 nm Ta and 25 nm Pt at room temperature, patterning by e-beam lithography and structuring with a room temperature RIBE process using argon. Afterwards, the switching oxide from 3 nm HfO₂ and 3 nm TiO_x layers was deposited by atomic layer deposition (ALD) in a FlexAl system of Oxford Plasma Technologies at a table temperature of 300°C. The used process for the HfO₂ was an oxygen plasma process with tetrakis(ethylmethylamino)hafnium (TEMAH) as Hf-precursor, while the TiO_x layer was deposited from tetrakis(dimethylamino)titanium (TDMAT) and water vapor. Lastly, the TE metal, namely 10 nm Ti and 20 nm Pt, was deposited by DC sputtering. The switching oxide and the TE were again patterned by an e-beam lithography step and structured using a RIBE process. Lastly, connector bars were fabricated by a lift-off process to establish electrical connection between the electrodes and the metal pads. Therefore, resist was patterned by a lithography step, 5 nm Ta and 50 nm Pt were subsequently deposited via DC sputtering and structures were obtained by a lift-off process based on acetone. During the fabrication, CMOS compatible chemicals and temperatures were used in order to preserve the transistors. The resulting 1T1R structure had the active electrode (Pt) of the VCM cell connected to the source of the NMOS transistor.

Author Contributions

O.A., F.C., and S.H.-E. worked on the conceptualization, the experimental design, and analysis of resulting data. O.A. fabricated, characterized and tested the devices. The CMOS chip was designed and taped-out by D.N., C.B., and A.Z., and the electrical characterization was supported by X.L., Y.C., and S.W. Compact model simulations were carried out by S.S. and S.M. The results were discussed between all authors. O.A. and S.H.-E. wrote the manuscript. All authors reviewed and edited the manuscript.

Acknowledgments

This work was supported in part by the Federal Ministry of Research, Technology and Space (BMFT, Germany) in the projects NEUROTEC-II (project numbers: 16ME0398K and 16ME0399) and NeuroSys (project number: 03ZU2106AB) and by Deutsche Forschungsgemeinschaft within TRR 404 Active-3D (project number: 528378584). The authors thank H. Sultani, D. Storelli, G. Potemkin, S. Shoja, M. Gerst, M. Deckers (all PGI-7), B. Bennemann (PGI-9), V. Christ, P. Schubert (PGI-4), S. Trelenkamp, F. Lentz, T. Wessels (HNF), and C. Sudarshan (PGI-14) for technical support and helpful discussions.

Open access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

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

Ethics Declarations

This study does not involve human participants, animal subjects, or sensitive data. No specific ethical or inclusivity considerations were required.

References

1. A. Mehonic, D. Ielmini, K. Roy, et al., "Roadmap to Neuromorphic Computing with Emerging Technologies," *APL Materials* 12, no. 10 (October 2024): 109201.
2. Y. Huang, T. Ando, A. Sebastian, M.-F. Chang, J. J. Yang, and Q. Xia, "Memristor-Based Hardware Accelerators for Artificial Intelligence," *Nature Reviews Electrical Engineering* 1 (April 2024): 286–299.
3. W. Choi, T. Stecconi, D. F. Falcone, et al., "Update Disturbance-Resilient Analog RERAM Crossbar Arrays for In-Memory Deep Learning Accelerators," *Advanced Science* n/a, no. n/a (2025): e04578.
4. W. Wan, R. Kubendran, C. Schaefer, et al., "A Compute-in-Memory Chip Based on Resistive Random-Access Memory," *Nature* 608 (August 2022): 504–512.
5. D. V. Christensen, R. Dittmann, B. Linares-Barranco, et al., "2022 Roadmap on Neuromorphic Computing and Engineering," *Neuromorphic Computing and Engineering* 2, no. 2 (May 2022): 022501.
6. S. Brivio, S. Spiga, and D. Ielmini, "HfO₂-Based Resistive Switching Memory Devices for Neuromorphic Computing," *Neuromorphic Computing and Engineering* 2, no. 4 (October 2022): 042001.
7. U. Böttger, M. von Witzleben, V. Havel, et al., "Picosecond Multilevel Resistive Switching in Tantalum Oxide Thin Films," *Scientific Reports* 10 (October 2020): 16391.
8. M. Csontos, Y. Horst, N. J. Olalla, et al., "Picosecond Time-Scale Resistive Switching Monitored in Real-Time," *Advanced Electronic Materials* 9, no. 6 (2023): 2201104.
9. B. Govoreanu, G. Kar, Y.-Y. Chen, et al., "10 × 10 nm² Hf/HfO_x Crossbar Resistive RAM with Excellent Performance, Reliability and Low-Energy Operation," in *2011 International Electron Devices Meeting* (2011), 31.6.1–31.6.4.
10. S. Pi, C. Li, H. Jiang, et al., "Memristor Crossbar Arrays with 6-nm Half-Pitch and 2-nm Critical Dimension," *Nature Nanotechnology* 14 (November 2019): 35–39.
11. X. Sheng, C. E. Graves, S. Kumar, et al., "Low-Conductance and Multilevel CMOS-Integrated Nanoscale Oxide Memristors," *Advanced Electronic Materials* 5, no. 9 (2019): 1800876.
12. D. Ielmini and G. Pedretti, "Resistive Switching Random-Access Memory (RRAM): Applications and Requirements for Memory and Computing," *Chemical Reviews* 125, no. 12 (2025): 5584–5625.
13. 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," *Advanced Functional Materials* 21, no. 23 (2011): 4487–4492.
14. R. Dittmann, A. Sarantopoulos, C. Bengel, et al., "Engineering the Kinetics of Redox-Based Memristive Devices for Neuromorphic Computing," in *2023 International Electron Devices Meeting (IEDM)* (2023), 1–4.
15. D. Schön and S. Menzel, "Spatio-Temporal Correlations in Memristive Crossbar Arrays Due to Thermal Effects," *Advanced Functional Materials* 33, no. 22 (2023): 2213943.
16. S. Pande, B. Chakrabarti, and A. Chakravorty, "Thermal Crosstalk Analysis in RERAM Passive Crossbar Arrays," in *2024 37th International Conference on VLSI Design and 2024 23rd International Conference on Embedded Systems (VLSID)* (2024): 390–395.
17. M. Rao, H. Tang, J. Wu, et al., "Thousands of Conductance Levels in Memristors Integrated on CMOS," *Nature* 615, no. 7954 (2023): 823–829.
18. E. Pérez, A. Grossi, C. Zambelli, P. Olivo, R. Roelofs, and C. Wenger, "Reduction of the Cell-to-Cell Variability in Hf_{1-x}Al_xO_y-Based RRAM Arrays by Using Program Algorithms," *IEEE Electron Device Letters* 38 (2017): 175–178.
19. V. Milo, A. Glukhov, E. Pérez, et al., "Accurate Program/Verify Schemes of Resistive Switching Memory (RRAM) for In-Memory Neural

- Network Circuits,” *IEEE Transactions on Electron Devices* 68 (2021): 3832–3837.
20. A. Baroni, A. Glukhov, E. Pérez, et al., “An Energy-Efficient In-Memory Computing Architecture for Survival Data Analysis Based on Resistive Switching Memories,” *Frontiers in Neuroscience* 16 (2022): 932270.
 21. M. Hu, C. E. Graves, C. Li, et al., “Memristor-Based Analog Computation and Neural Network Classification with a Dot Product Engine,” *Advanced Materials* 30, no. 9 (2018): 1705914.
 22. A. Fantini, D. J. Wouters, R. Degraeve, et al., “Intrinsic Switching Behavior in HfO₂ RRAM by Fast Electrical Measurements on Novel 2R Test Structures,” in *2012 4th IEEE International Memory Workshop (IMW)* (2012), 1–4.
 23. A. Glukhov, V. Milo, A. Baroni, et al., “Statistical Model of Program/Verify Algorithms in Resistive-Switching Memories for In-Memory Neural Network Accelerators,” in *2022 IEEE International Reliability Physics Symposium (IRPS)* (2022), 3C.3–1–3C.3–7.
 24. C. Bengel, K. Zhang, J. Mohr, et al., “Tailor-Made Synaptic Dynamics Based on Memristive Devices,” *Frontiers in Electronic Materials* 3 (2023): 1061269.
 25. X. Liu, C. Bengel, F. Cüppers, et al., “Effect of Transistor Transfer Characteristics on the Programming Process in 1T1R Configuration,” *IEEE Transactions on Electron Devices* 71, no. 4 (2024): 2423–2430.
 26. D. Ielmini, F. Nardi, and S. Balatti, “Evidence for Voltage-Driven Set/Reset Processes in Bipolar Switching RRAM,” *IEEE Transactions on Electron Devices* 59, no. 8 (2012): 2049–2056.
 27. 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,” *Applied Physics Letters* 117, no. 26 (December 2020): 262902.
 28. R. Degraeve, A. Fantini, N. Raghavan, et al., “Modeling RRAM Set/Reset Statistics Resulting in Guidelines for Optimized Operation,” in *2013 Symposium on VLSI Technology* (2013), T98–T99.
 29. A. Fantini, L. Goux, R. Degraeve, et al., “Intrinsic Switching Variability in HfO₂ RRAM,” in *2013 5th IEEE International Memory Workshop* (2013), 30–33.
 30. D. Walczyk, C. Walczyk, T. Schroeder, et al., “Resistive Switching Characteristics of CMOS Embedded HfO₂-Based 1T1R Cells,” *Microelectronic Engineering* 88, no. 7 (2011): 1133–1135.
 31. W. G. Bennett, N. C. Hooten, R. D. Schrimpf, et al., “Single- and Multiple-Event Induced Upsets in HfO₂/Hf 1T1R RRAM,” *IEEE Transactions on Nuclear Science* 61, no. 4 (2014): 1717–1725.
 32. C. Bengel, J. Mohr, S. Wiefels, et al., “Reliability Aspects of Binary Vector-Matrix-Multiplications Using RRAM Devices,” *Neuromorphic Computing and Engineering* 2, no. 3 (June 2022): 034001.
 33. R. Dittmann, S. Menzel, and R. Waser, “Nanoionic Memristive Phenomena in Metal Oxides: The Valence Change Mechanism,” *Advances in Physics* 70, no. 2 (2021): 155–349.
 34. V. V. Afanas'ev, A. Stesmans, L. Pantisano, et al., “TiN_x/HfO₂ Interface Dipole Induced by Oxygen Scavenging,” *Applied Physics Letters* 98, no. 13 (March 2011): 132901.
 35. 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 HfO₂/TiO_x Bilayer ReRAM Cells,” *IEEE Transactions on Electron Devices* 65, no. 8 (2018): 3229–3236.
 36. F. Cüppers, S. Menzel, C. Bengel, et al., “Exploiting the Switching Dynamics of HfO₂-Based ReRAM Devices for Reliable Analog Memristive Behavior,” *APL Materials* 7, no. 9 (2019): 091105/1–9.
 37. D. Ielmini, “Modeling the Universal Set/Reset Characteristics of Bipolar RRAM by Field- and Temperature-Driven Filament Growth,” *IEEE Transactions on Electron Devices* 58, no. 12 (2011): 4309–4317.
 38. T. Hennen, E. Wichmann, R. Waser, D. J. Wouters, and D. Bedau, “Stabilizing Amplifier with a Programmable Load Line for Characterization of Nanodevices with Negative Differential Resistance,” *Review of Scientific Instruments* 93, no. 2 (February 2022): 024705.
 39. A. Fantini, L. Goux, R. Degraeve, et al., “Intrinsic Switching Variability in HfO₂ RRAM,” in *2013 5th IEEE International Memory Workshop* (2013), 30–33.
 40. K. Kinoshita, K. Tsunoda, Y. Sato, et al., “Reduction in the Reset Current in a Resistive Random Access Memory Consisting of NiO_x Brought About by Reducing a Parasitic Capacitance,” *Applied Physics Letters* 93, no. 3 (July 2008): 033506.
 41. E. Perez, M. K. Mahadevaiah, E. P. Quesada, and C. Wenger, “Variability and Energy Consumption Tradeoffs in Multilevel Programming of RRAM Arrays,” *IEEE Transactions on Electron Devices* 68 (2021): 2693–2698.
 42. A. Grossi, E. Vianello, C. Zambelli, et al., “Experimental Investigation of 4-kb RRAM Arrays Programming Conditions Suitable for TCAM,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 26, no. 12 (2018): 2599–2607.
 43. A. Baroni, A. Glukhov, E. Pérez, et al., “Low Conductance State Drift Characterization and Mitigation in Resistive Switching Memories (RRAM) for Artificial Neural Networks,” *IEEE Transactions on Device and Materials Reliability* 22, no. 3 (2022): 340–347.
 44. S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini, “Understanding Switching Variability and Random Telegraph Noise in Resistive RAM,” in *2013 IEEE International Electron Devices Meeting* (IEEE, 2013), 31.5.1–31.5.4.
 45. S. Balatti, S. Ambrogio, D. C. Gilmer, and D. Ielmini, “Set Variability and Failure Induced by Complementary Switching in Bipolar RRAM,” *IEEE Electron Device Letters* 34, no. 7 (2013): 861–863.
 46. S. Pande, S. Balanethiram, B. Chakrabarti, and A. Chakravorty, “A Physics-Based Compact Model of Thermal Resistance in RRAMs,” *Solid-State Electronics* 204 (2023): 108636.
 47. S. Son, A. Bende, D. Schön, et al., “A Comprehensive Compact Model for Multilevel Switching in TaO_x-Based Memristive 1T–1R Cells,” *IEEE Access* 13 (2025): 127280–127291.
 48. C. Bengel, A. Siemon, F. Cüppers, et al., “Variability-Aware Modeling of Filamentary Oxide-Based Bipolar Resistive Switching Cells Using SPICE Level Compact Models,” *IEEE Transactions on Circuits and Systems I: Regular Papers* 67, no. 12 (2020): 4618–4630.
 49. M. Caselli and A. Boni, “A Write System for Compact RRAM Memory Arrays Based on F-1T1R,” in *2024 IEEE International Symposium on Circuits and Systems (ISCAS)* (2024): 1–5.
 50. A. Levisse, M. Bocquet, M. Rios, et al., “Write Termination Circuits for RRAM: A Holistic Approach From Technology to Application Considerations,” *IEEE Access* 8 (2020): 109297–109308.
 51. P. Jain, U. Arslan, M. Sekhar, et al., “13.2 A 3.6 Mb 10.1 Mb/mm² Embedded Non-Volatile ReRAM Macro in 22 nm FinFET Technology with Adaptive Forming/Set/Reset Schemes Yielding Down to 0.5 V with Sensing Time of 5 ns at 0.7 V,” in *2019 IEEE International Solid-State Circuits Conference (ISSCC)* (2019), 212–214.
 52. X. Xue, W. Jian, J. Yang, et al., “A 0.13 μ m 8 Mb Logic-Based Cu_xSi_yO ReRAM With Self-Adaptive Operation for Yield Enhancement and Power Reduction,” *IEEE Journal of Solid-State Circuits* 48, no. 5 (May 2013): 1315–1322.
 53. M.-F. Chang, J.-J. Wu, T.-F. Chien, et al., “Low V_{DDmin} Swing-Sample-and-Couple Sense Amplifier and Energy-Efficient Self-Boost-Write-Termination Scheme for Embedded ReRAM Macros Against Resistance and Switch-Time Variations,” *IEEE Journal of Solid-State Circuits* 50, no. 11 (2015): 2786–2795.
 54. W.-H. Chen, W.-J. Lin, L.-Y. Lai, et al., “A 16 Mb Dual-Mode ReRAM Macro with Sub-14 ns Computing-in-Memory and Memory Functions Enabled by Self-Write Termination Scheme,” in *2017 IEEE International Electron Devices Meeting (IEDM)* (2017), 28.2.1–28.2.4.