.

Endurance of 2 Mbit based BEOL integrated ReRAM

NILS KOPPERBERG[1], STEFAN WIEFELS[2], KARL HOFMANN[3], JAN OTTERSTEDT[3], (Member, IEEE), DIRK J. WOUTERS[1], (Member, IEEE), RAINER WASER[1,2,4], (Member, IEEE) , AND STEPHAN MENZEL[2], (Senior Member, IEEE)

[1]Institut für Werkstoffe der Elektrotechnik II (IWE II) and JARA-FIT, RWTH Aachen University, 52074 Aachen, Germany
[2]Peter-Grünberg-Institut 7 (PGI-7), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
[3]Infineon Technologies AG, Am Campeon 1-15, 85579 Neubiberg, Germany
[4]Peter-Grünberg-Institut 10 (PGI-10), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

Corresponding author: Nils Kopperberg (e-mail: kopperberg@iwe.rwth-aachen.de).

ABSTRACT In this work, we experimentally characterize the endurance of 2 Mbit resistive switching random access memories (ReRAMs) from a 16 MBit test-chip. Here, very rare failure events where the memory cells become stuck in the low-resistive state (LRS) are observed. As this failure mechanism is the limiting one concerning the endurance of this ReRAM implementation, extensive investigations are conducted and presented. The experimental findings are detailed via a voltage divider model, illustrating why memory cells can become stuck in the LRS. It is proposed, that an insufficient voltage dropping over the cell due to an unfavorable combination of cell- and transistor resistances is responsible for stuck-at-LRS bits. Furthermore, predictions for the origin of these suboptimal combinations are given. Additionally, a one-dimensional Kinetic Monte Carlo (KMC) model that allows a statistical investigation of large numbers of cells with regard to rare random events has been developed. Here, our proposed explanation for the observed failure mechanism is fortified by the simulation and evaluation of the switching process of the memory. All simulations are in very good agreement with the experimental data. Finally, based on our findings, we give suggestions for the improvement of switching algorithms.

**INDEX TERMS** Endurance, KMC, OxRAM, ReRAM, RESET Failure, Stuck at SET, VCM

I. INTRODUCTION

REGARDING the strongly increasing demand for highly scaled, non-volatile memories in many modern applications like smartphones, the research in this field has grown strongly over the last years [1]–[3]. Here, resistive switching random access memories (ReRAMs) are very promising candidates for future industrial applications as well as for neuromorphic computing [4]–[6]. ReRAMs show a lot of great operation features like scalability, fast switching, reliability and complementary metal-oxide-semiconductor (CMOS) compatibility [7]–[9] and thus are expected to replace flash technology in the future [10]–[12].

One of the most promising candidates in the broad field of ReRAM technology are bipolar switching memristive devices based on the valence change mechanism (VCM). VCMs typically consist of a transition metal-oxide as switching layer placed between two metal electrodes [13]–[15]. By applying appropriate voltages to the electrodes, oxygen can be extracted from the metal-oxide layer in a so-called forming step. Thereby, oxygen vacancies are left in the switching layer building a conducting filament through the initially insulating metal-oxide. With suitable applied voltages, the filament can be ruptured or rebuilt depending on the polarity of the voltage and so the device can be switched between a high resistive state (HRS) and a low resistive state (LRS) [16], [17]. The switching from HRS to LRS, in our case with the same voltage polarity as the forming step, is called SET, the opposite process is called RESET. The present state of the cell can be easily read out non-destructively by applying a small voltage to the device.

Independent of the specific application of the ReRAMs, reliability is a major issue that has to be investigated and optimized [18]–[20]. Especially, endurance meaning the ability of huge numbers of faultless consecutive switching cycles is

a)

b)
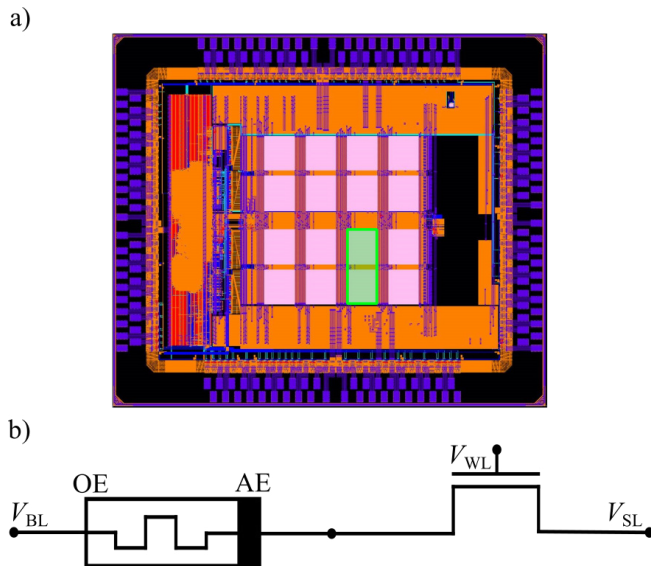
FIGURE 1. a) Image of the 16 MBit test-chip. The cycled block of 2 MBit is highlighted in green. b) Sketch of the typical 1T-1R structure of our VCM cells, with the ReRAM element on the left and the transistor on the right. The AE of the element is connected to the transistor and the OE to the BL. The SET polarity is therefore with a positive voltage V_{BL} at the BL.

of great interest and will therefore be in the focus of our work [21], [22].

In the following, experimental results of 2 MBit VCM cells are presented, predominantly showing great endurance. Nevertheless, looking at the huge statistics with many cells after many switching cycles, a few cells become stuck at the LRS. This rare failure event is the main effect limiting the endurance and is therefore further investigated experimentally. Whereas Yang et al. [23] are connecting this kind of failure mechanism to the presence of a second filament, we present an alternative explanation. In a simple voltage divider model, we propose a too low voltage dropping over the cell to be responsible for the RESET failure. Additionally, a newly developed one-dimensional kinetic Monte Carlo (KMC) model is presented. This model allows us to simulate the RESET behavior with high statistics and with regard to rare random events. These simulations underline our proposed explanation for the physical origin of the experimentally observed failure mechanism.

II. EXPERIMENTAL DETAILS

Our experimental endurance study was conducted on a test-chip with 16 Mbit VCM-type ReRAM, shown in Fig. 1, a). The cells are integrated back-end-of-line (BEOL) in a 1-transistor-1-resistor (1T1R) configuration in a 28 nm CMOS technology. Programming and read out are performed via on-chip circuitry. From this test vehicle, the highlighted block of 2 MBit is cycled.

A schematic of a single 1T-1R cell with the respective biases is shown in Fig. 1, b). The select transistor is opened by applying a voltage V_{WL} to its gate. In order to electroform or SET a bit, a rectangular voltage pulse V_{BL} is applied to the

bitline which is connected to the ohmic electrode (OE) of the ReRAM cell. For the RESET operation, a rectangular voltage pulse V_{SL} is applied to the source of the transistor (while its drain is connected to the active electrode (AE) of the ReRAM cell). To ensure reliable programming, electroforming, SET, and RESET are performed with a program-verify algorithm [24]. Here, the resistance of each bit is determined by a read pulse after each programming step and compared to a target resistance. If the target is reached, the programming operation is terminated and considered successful. Otherwise, the programming step is repeated with increased pulse length, V_{BL}, V_{SL} or V_{WL}. Here, increasing V_{BL} or V_{SL} provides a higher voltage across the ReRAM element during SET or RESET, respectively. As the transistor not only selects the cell to be programmed but also acts as current limiting element, increasing V_{WL} allows for higher current through the device. A good control of this current is crucial during electroforming and SET, as it strongly affects the resulting resistance of the cell [25]–[27]. In contrast, V_{WL} is desired to be comparatively high during RESET in order to allow sufficient current for this process. A current limitation is not needed for this operation, because the increasing cell resistance naturally limits the current.

III. EXPERIMENTAL RESULTS
A. ENDURANCE
Fig. 2, a) shows the cumulative HRS and LRS distributions read after different cycle numbers up to 500k. SET and RESET are performed using our standard algorithm which comprises multiple steps with increasing pulse length, V_{BL}, V_{SL} and V_{WL}. This results in a very good endurance in the investigated interval of 500k cycles. Whereas the LRS distribution only slightly broadens upon cycling, the HRS distribution is observed to drift towards lower read current, which has a positive effect on the read window between HRS and LRS. However, at higher cycle numbers (300k-500k) a tail appears in the HRS distribution consisting of bits which were not RESET successfully but stuck in the LRS. Although this affects only few ppm of bits, this failure has to be investigated, which is the main subject of this study.

B. STUCK-AT-LRS FAILURE
The inset in Fig. 2, a) shows the number of failed bits in the SET (red) or RESET (blue) operation over the cycle number. Up to 250k cycles, both SET and RESET are comparatively stable, although each cycle carries the chance of single failed bits. Whereas the SET operation remains stable within the tested cycles, an increasing trend of failed bits is observed for the RESET operation. Here, the number of fail-bits increases linearly with the cycle number to a maximum of 22 at 516k cycles, which accounts for approx. 10 ppm.

Despite the significance of this failure mechanism, it affects only very few individual ReRAM cells in the shown experiment. Further, no clusters of stuck-at-LRS bits are observed, but a random distribution of failed bits in the memory block. In order to gain deeper insights of the underlying mechanism,
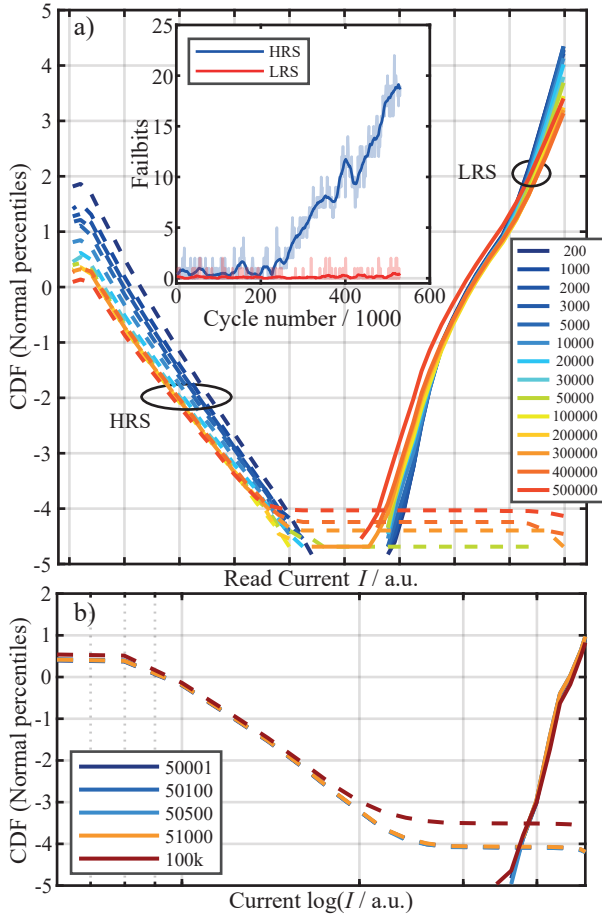
FIGURE 2. a) Cumulative distributions of HRS and LRS with the standard algorithm, showing high endurance of 500k cycles. After 500k cycles, the RESET fails for approx. 10 ppm of bits, leaving them stuck at LRS. The inset depicts the number of failed SET (red) or RESET (blue) operations per cycle. A moving mean is added as a solid line to emphasize the trend. Starting from approx. 250k cycles, the number of fail-bits in HRS increases linearly. b) Adjusted algorithm, intentionally provoking the observed stuck at LRS failure. After 50k cycles, a tail of stuck at LRS bits is observed at -4 σ. The tail grows to -3.5 σ after 100k cycles.

the endurance experiment was repeated with a programming algorithm that was slightly altered in a direction which provokes this failure and thus generates more stuck-at-LRS bits. Cumulative distributions of HRS and LRS after 50k-100k cycles with the suboptimal algorithm are shown in Fig. 2, b). Here, from 50k to 51k the read current of each bit in each cycle was measured directly, resulting in consecutive traces containing the states right after SET or RESET over 1000 cycles.

In order to understand the origin of this failure, exemplary traces out of this data set are shown in Fig. 3. In the five columns, selected bits with different numbers of failed cycles are depicted. The first column (Fig. 3, a), b)) shows reference bits for which no failures have been observed. The remaining traces are sorted by the number of cycles that a bit was stuck in LRS within the monitored 1k cycles. This ranges from only a few cycles in Fig. 3, c), d) to more than 600 cycles in Fig. 3, i), j). The traces show that bits can become stuck in

LRS spontaneously as in Fig. 3, c) or the HRS is observed to drift towards LRS from cycle to cycle until it becomes stuck (c.f. Fig. 3, h)). Although cells can become stuck over several cycles (> 1000), it is not primarily a permanent device failure. The single cells rather end up in a state from where the successful RESET becomes more unlikely. Thus, bits are typically observed to recover after a few or several cycles. Other bits seem to continuously alternate between functioning and faulty, as in Fig. 3, g). It may be noted that the cases in the different columns are not equally likely. RESET failures with long lifetime are much less likely than failures over single or very few cycles.

However, two general observations are made: Firstly, stuck bits usually comprise a comparatively high HRS current before failing or after recovery. Secondly, the stuck-at-LRS bits typically exhibit rather high read current in the problematic LRS. Whereas the first observation might be an indicator how bits become stuck, the second observation of low LRS resistances hints towards a possible explanation of the failure mechanism: In the interplay of ReRAM cell and series resistance of the periphery (including the access transistor), a too low cell resistance might lead to an insufficient electrical field across the cell as the major part drops across the periphery. In the following section we discuss this mechanism using a simple model before we compare our experimental findings to the results of KMC simulations.

C. PHENOMENOLOGICAL MODEL

The origin of the RESET failure can be explained by a simple model. In Fig. 1 an extract of a single memory element in a typical 1T-1R configuration was already presented. The transistor shown on the right can be controlled by the word line voltage V_{WL}, the ReRAM can be read or switched by the applied bit line and source line voltages V_{BL} and V_{SL}. During the RESET, the resistance of the transistor is nearly constant. Therefore, the transistor as well as the line resistances and other parts of the circuitry are summed up to a resistance of the periphery R_{per}. Thus, as can be seen in Fig. 4, a series connection of the ReRAM cell with a cell resistance R_{cell} and the resistance of the periphery R_{per} is taken to represent the 1T-1R element. This results in a voltage divider where the totally applied voltage V_{tot} splits up into the voltage dropping over the periphery V_{per} and the voltage dropping over the ReRAM cell V_{cell}. This voltage V_{cell} is crucial with regard to the switching process, as the switching time depends exponentially on it [28]. According to the voltage divider, V_{cell} can be calculated to

$$V_{cell} = V_{tot} \frac{R_{cell}}{R_{cell} + R_{per}}, \tag{1}$$

which therefore depends on three parameters: The first parameter is the externally applied voltage V_{tot}, which is in good approximation constant for all times and all devices. The second parameter is the resistance of the periphery R_{per}, which is constant over time for a single device, but has a certain variance from device to device. The last parameter
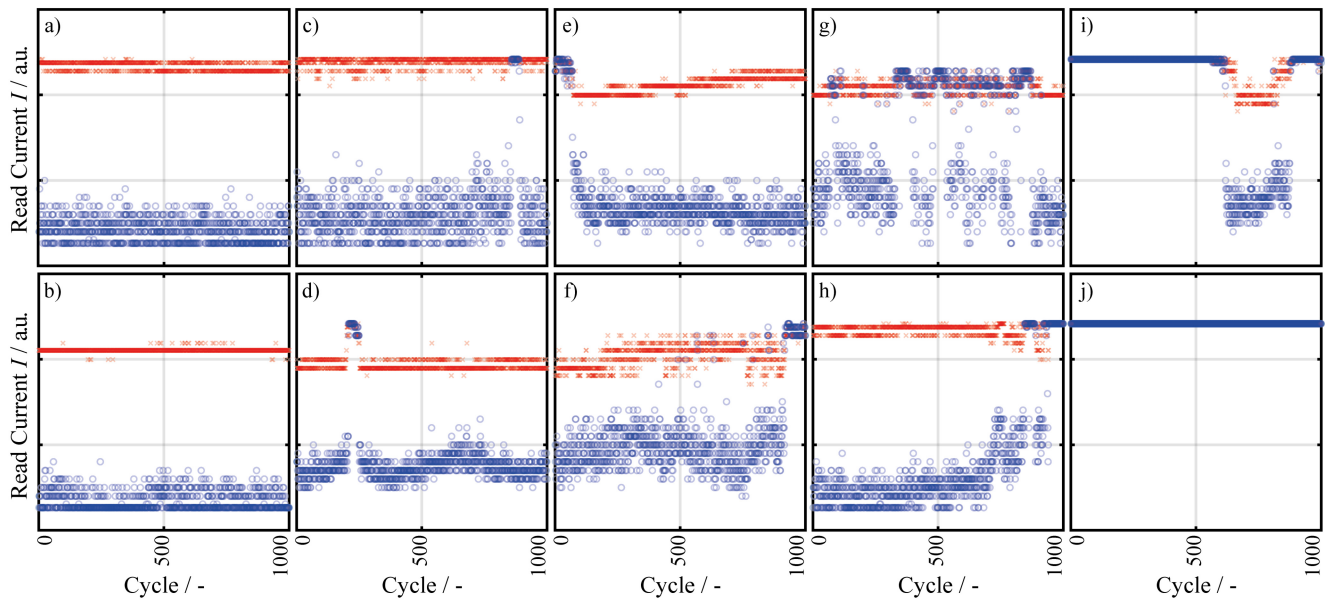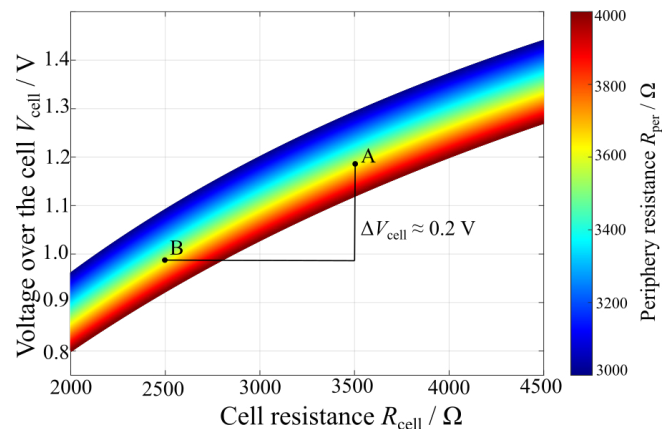
FIGURE 3. Selected traces of HRS (blue) and LRS (red) over 1000 consecutive switching cycles. The data is drawn from cycles 50001-51000 in Fig. 2, b). The traces are sorted column-wise regarding the number of failed RESETs. a), b) show good reference bits which were not observed to be stuck in this interval. c), d) comprise few failed cycles, which increases towards i) with more than 600 and j) with all 1000 failed cycles. Stuck bits can occur spontaneously or appear with a gradually increasing HRS current. The failure is not permanent, but can be recovered. Stuck bits typically come from higher HRS current and comprise high LRS current when they are stuck.
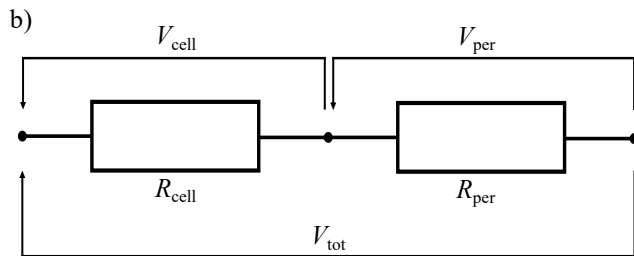
FIGURE 4. Sketch of our model idea, simplifying the typical 1T-1R structure of our VCM cells from Fig. 1 to a voltage divider of two resistances.

FIGURE 5. Cell voltage in dependence of cell resistance and periphery resistance. As an example, a cell A with $R_{per} = 3600\,\Omega$ and $R_{cell} = 3500\,\Omega$ is shown. For cell B, with the same R_{per}, but programmed into a low-ohmic state with $R_{cell} = 2500\,\Omega$, V_{cell} is lowered by approx. $0.2\,\text{V}$.

is the resistance of the cell R_{cell} that varies during switching and is different for each device after every switching cycle. Especially, R_{cell} in the LRS is strongly depending on the preceding SET event. The influence of both varying parameters, R_{cell} and R_{per}, on V_{cell} is shown in Fig. 5. It can be seen that, depending on the possible cell and periphery resistances, V_{cell} can vary significantly. Realistic values were chosen for the shown ranges of R_{cell} and R_{per}. It is observed that V_{cell} can lay in a range from approximately $0.9..1.4\,V$. Exemplarily, a cell with $R_{per} = 3600\,\Omega$ is marked for a common cell resistance of $R_{cell} = 3500\,\Omega$ (A) and a low-ohmic state with $R_{cell} = 2500\,\Omega$ (B) in Fig. 5. Between these two states, a change in V_{cell} of about $0.2\,V$ can be observed. Since it has been demonstrated that a decrease of V_{cell} by 0.1 V can increase the RESET time by one order of magnitude [28], this huge variance in V_{cell} leads to a strongly varying switching behavior from cell to cell and from cycle to cycle.

Most of the cases are not problematic, as the RESET conditions are nicely adjusted. But in rare cases of a very high periphery resistance R_{per} in combination with a very low cell resistance R_{cell}, the voltage dropping over the cell V_{cell} is too low to drive the RESET process within the applied pulse width. This failure process becomes even rarer, as programming algorithms increasing the applied voltage V_{tot} or the time of the switching pulse are able to switch devices that failed or only partly switched in the first try. Only for very extreme configurations, the devices are stuck in the LRS. During the SET operation, the periphery resistance determines to a large extent in which resistance the cell ends

up. The resulting LRS is expected to be in the range of $R_{\text{cell}} \approx R_{\text{per}}$ as the applied voltage would drop mainly across the periphery resistance as soon as the cell resistance becomes significantly smaller. For RESET, the transistor is opened further to ensure having a sufficiently large voltage dropping over the cell. However, the body effect has to be considered. Since the access transistor needs to be operated in two polarities (for SET and RESET), its bulk and source cannot be connected. In SET direction, the voltage is applied at the BL with the transistor source connected to ground. However, during RESET the effective source of the transistor is located at the middle node in Fig. 1, a), between the active electrode of the ReRAM element and the transistor channel. This reduces the effective voltage between gate and source and thus increases the resistance of the channel. As this results in a higher periphery resistance during RESET compared to SET, the body effect strengthens our prediction for the cause of the RESET failure.

D. GENERATION

After the reason for the RESET failure is established, the question arises how the unfortunate combination of low cell resistance and high periphery resistance is generated. One reason is the body effect, as discussed above. During SET, the cell is programmed to a state determined by the series resistance of the periphery. If this resistance increases upon attempted RESET, this process is already more challenging. As another possible origin of the RESET failure, the formation of an additional filament was reported [23]. However, both explanations do not account for the increasing number of failed bits with higher cycle numbers (c.f. inset in Fig. 2, a)). In order to understand the generation of stuck bits, the read currents of the last successful SET and RESET in our experimental data are traced and compared to the probability to become stuck in the following cycle. As depicted in Fig. 6, a) this probability increases for cells with a high read current in HRS, close to the read window. This fits to the qualitative observation in Fig. 3 that problematic bits are often accompanied by high HRS current before becoming stuck or after recovery. Additionally, Fig. 6, a) shows that an LRS read current close to the read window indicates a higher probability to become stuck in the next cycle. This is counter-intuitive, as bits are typically observed to become stuck in a high current LRS. However, note that the figure only contains the last good cycle. Bits that are already in the problematic high current state would not appear here. In both cases (HRS and LRS), the bits close to the read window are likely to become stuck afterwards. This hints towards the applied programming algorithm as a possible cause of bits becoming stuck, as the decision if further programming pulses are executed depends on a threshold current in the read window. It seems likely that in particular those bits that hardly reach the specified threshold endure further attempts by the programming algorithm, where the later steps typically comprise longer pulse width, V_{BL} or V_{WL}.

To understand the impact of the programming algorithm,
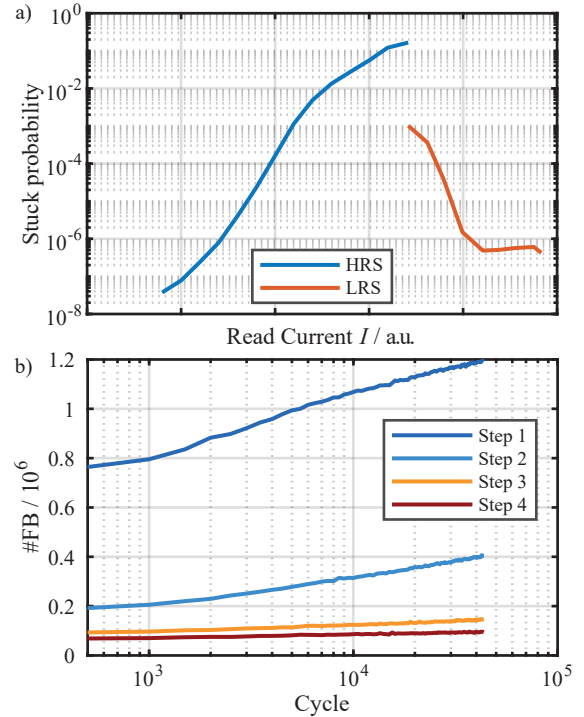
FIGURE 6. Experimental analysis of stuck bit generation. a) Probability of bits to become stuck with the suboptimal algorithm, evaluated regarding read currents in the last good cycle. Bits close to the read window have a higher chance to become stuck. b) Analysis of an exemplary SET algorithm comprising 4 steps with increasing V_{WL}. With each step, the number of failed SET attempts decreases. However, with increasing cycle number, the effectiveness of this algorithm decreases.

an exemplary algorithm for the SET operation is studied experimentally in Fig. 6, b). The algorithm consists of four steps with increasing V_{WL} and thus increasing maximum SET current. Between the steps, all bits are read and the number of bits which failed to reach the threshold current of a successful SET are evaluated. As expected, it can be seen that this number decreases with each executed step. However, the study also shows that the number of fail-bits increases over the number of cycles. Note that here, fail-bits are those which did not SET successfully. This means that the SET operation becomes more difficult at higher cycle numbers. In order to reach a fixed threshold current, the algorithm would therefore more often execute the later steps with higher V_{WL}, which increases the likelihood to end up in the unfortunate combination of $R_{\text{cell}} < R_{\text{per}}$. Especially, an increasing V_{WL} seems to be an operation to more likely generate these bad combinations. It is conceivable that over several steps the limiting factor is the switching kinetics and thus the pulse length and height (V_{BL}). If, nevertheless, the gate voltage V_{WL} is increased, the cell might endure a comparatively high SET current as soon as the switching event occurs.

IV. 1D KMC SIMULATION

To further support our theory and to investigate the origin and behavior of the presented RESET failure on another level, a one-dimensional kinetic Monte Carlo (KMC) simulation has
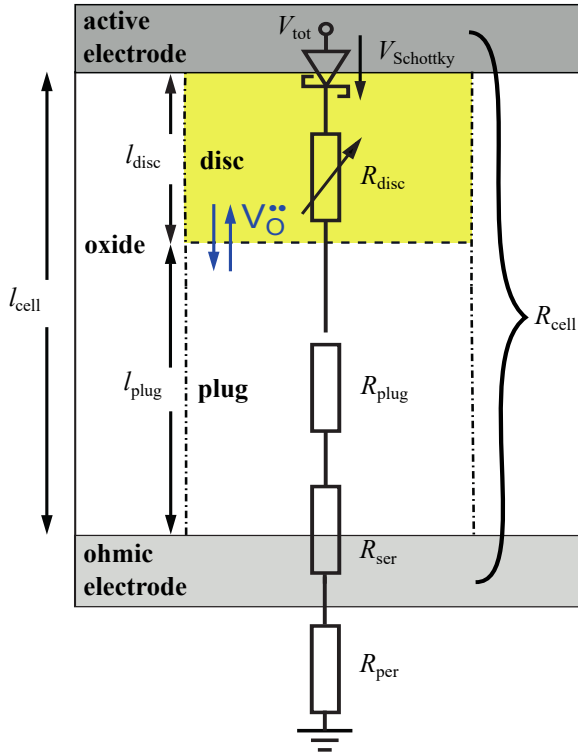
FIGURE 7. Equivalent circuit diagram of the 1D KMC model based on the JART VCM 1.0 by C. La Torre - adapted from [29]

been used. Basically, the JART VCM 1.0 model by C. La Torre [29], [30] was adapted by adding typical KMC methods for the central transition process [31]. On the one hand, the high performance of the compact model allows investigating high cell statistics with a reasonable amount of computation time. On the other hand, the KMC extension enables the investigation of statistical effects with respect to the influence of random processes.

A. MODEL

The physics-based compact model is based on several general assumptions. The model assumes a preexisting filament connecting the two electrodes, which shows a high number of oxygen vacancies. The filament has a time-invariant radius r and is divided into a plug and a disc region with a uniform oxygen vacancy concentration N_{disc} and N_{plug} each. The lengths of both regions l_{disc} and l_{plug} are also constant, whereas the oxygen vacancy concentrations in the regions are the only state variables changing over time. The total number of oxygen vacancies in the cell $N_{cell} = N_{disc} + N_{plug}$ is constant. Oxygen vacancies are treated as doubly positive charged donors that can be moved via drift in an electric field enhanced by temperature, which is assumed constant over the whole filament. Diffusion and thermodiffusion are neglected. Additionally, the metal/oxide interfaces are modelled as Schottky contact at the active top electrode and as ohmic contact at the bottom electrode. An equivalent circuit diagram of the model is presented in Fig. 7. Here, the five el-

ements of our model can be seen, namely the Schottky diode, the disc and plug resistances R_{disc} and R_{plug}, the series resistance R_{ser} at the ohmic electrode and the periphery resistance R_{per} including the transistor. To make the connection to our phenomenological model that was introduced before clear, the first four elements have been summed up to the total cell resistance R_{cell} in Fig. 7.

As a central part of the model, Kirchhoff's law

$$V_{tot} - [V_{Schottky} + I \cdot (R_{disc} + R_{plug} + R_{ser} + R_{per})] = 0 \quad (2)$$

is solved, with the current being denoted by I. The calculation of R_{disc} and R_{plug} is based on band conduction mechanism and temperature-dependent mobility via

$$R_{disc/plug} = \frac{l_{disc/plug}}{A \cdot z_{Vo} e N_{disc/plug} \mu_{n0}} \exp(\frac{\Delta E_{ac}}{k_b T}), \quad (3)$$

with the cross-sectional area of the filament $A = \pi r^2$, the charge number of the oxygen vacancies relative to the perfect crystal z_{Vo}, the temperature independent prefactor of the mobility μ_{n0} and a small activation energy ΔE_{ac} modelling the temperature dependence of the mobility [29]. The temperature is calculated via

$$T = (V_{disc} + V_{plug}) \cdot I \cdot R_{th,eff} + T_0 \quad (4)$$

with the Joule heating being described by an effective thermal resistance $R_{th,eff}$ and T_0 depicting the ambient temperature. The detailed calculation of $V_{Schottky}$ can also be found in [29].

The oxygen exchange between the disc and the plug region via the ionic current is calculated by typical KMC methods. For both directions, a jump of an oxygen vacancy from the plug to the disc region and vice versa, jumping rates are calculated via

$$R^{f/r} = \nu_0 \cdot \exp(-\frac{\Delta W_A^{f/r}}{k_B T}) \quad (5)$$

where ν_0 denotes the characteristic vibration frequency. For the hopping barrier ΔW_A, a typical value for the present metal-oxide is used and modulated by the Genreith-Schriever approach as predicted in [29] and [32] via

$$\Delta W_A^{f/r} = \Delta W_A \cdot (\sqrt{1 - \gamma^2} \pm \gamma \arcsin \gamma) \quad (6)$$

depending on the jump direction. The factor γ modifies the hopping barrier according to

$$\gamma = \frac{a z_{Vo} e E}{\pi \Delta W_A} \quad (7)$$

with the hopping distance a and the electric field which is calculated by

$$E = \frac{V_{disc} + V_{plug}}{l_{cell}} \quad (8)$$

for the RESET. Then, randomly but weighted by the probabilities, one of the two processes is chosen. The time of the process is calculated as

$$t_{jump} = -\frac{\ln r_1}{R^f + R^r} \quad (9)$$

with r_1 being a random number between 0 and 1. Finally, the simulation time is increased by t_{jump} and the chosen process is executed by updating $N_{\text{disc/plug}}$, which directly influences $R_{\text{disc/plug}}$ and so the current through the ReRAM cell. The simulation parameters are given in table 1.

TABLE 1. Simulation Parameters

Symbol	Value	Symbol	Value
l_{cell}	5 nm	ΔW_{A}	1.2 eV
l_{disc}	0.75 nm	ν_0	2×10^{13} Hz
l_{plug}	4.25 nm	z_{Vo}	$+2$
r	30 nm	μ_{n0}	5×10^{-6} eV
A	$\pi \cdot 900 \, \text{nm}^2$	ΔE_{ac}	$0.08 \, \text{m}^2 \, \text{V}^{-1} \, \text{s}^{-1}$
N_{cell}	8000	$R_{\text{th,eff}}$	$720 \, \Omega$
a	0.25 nm	R_{ser}	$4.24 \times 10^{-6} \, \text{K} \, \text{W}^{-1}$
T_0	293 K		

B. RESULTS

In the first step, the known parameters R_{ser}, l_{cell} as well as several material parameters, are implemented. Furthermore, the unknown values of the filament geometry l_{disc}, l_{plug} and the radius r are chosen in a reasonable way. The number of oxygen vacancies in the plug region N_{plug} is chosen with regard to typical oxygen vacancy densities used in simulations for VCM cells [30], [33]. The key parameter is the number of oxygen vacancies in the disc region N_{disc} and is chosen quite high in view of the LRS being investigated in the beginning. Additionally, for the values of R_{per} a variability known from the experiment and for N_{plug} and N_{disc} an estimated variability ($\sigma \approx 25$) are added as the VCM cells typically show a cell to cell and a cycle to cycle variability leading to a certain width of the current or resistance distributions. The RESET is performed by applying a voltage V_{tot} to the active electrode. Here a pulse sequence of a short read pulse with a low voltage of $V_{\text{read}} = -0.2 \, \text{V}$, a longer RESET pulse of $V_{\text{tot}} = 2.4 \, \text{V}$ and again a short read pulse is applied. The initial LRS current distribution and the HRS current distribution after RESET are presented as dark blue curves of Fig. 8 and fit nicely to the experimental data. We are interested in very rare events that only occur, looking at very high numbers of cells and large numbers of switching cycles in the experiment. Due to the high amount of computing resources needed to simulate even more cells, this method would be very inefficient. Furthermore, increasing the variability of several simulation parameters is not reasonable, as the current distributions of LRS and HRS are very stable in total except the single cells becoming stuck at high currents. From our simple voltage-divider model above, the RESET failure is assumed to occur for bad combinations of R_{per} and R_{cell}. As the properties of R_{per} are well known from the experiment and expected to be stable for a single cell, R_{cell} has to be changed in our simulations. Here, the parameter of choice is the number of oxygen vacancies in the disc region N_{disc}. As can be seen in the color gradient of Fig. 8, N_{disc} has been increased continuously and concomitant with that R_{cell} has been reduced. Thereby, the probability to observe
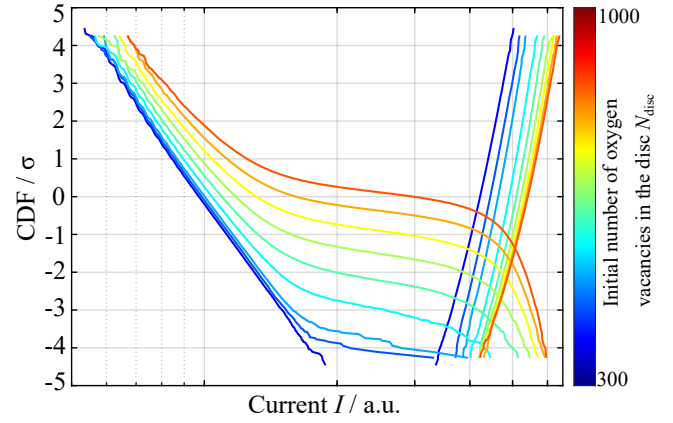

FIGURE 8. Read current distribution of cells programmed to the LRS (right) and after RESET pulse (left). For comparably low numbers of oxygen vacancies initially distributed in the disc region, ideal switching can be observed (dark blue). For higher numbers of oxygen vacancies, more and more badly or even not switched cells occur.

bad combinations of R_{per} and R_{cell} is permanently increased. As a result, more and more bits can be observed that are not completely switched to the HRS in the light blue and green curves of Fig. 8. In the bottom right corner of the orange and red curves, even many cells appear that did not or just slightly changed their resistance after the RESET pulse.

In Fig. 9, the read current of all cells in the LRS before the RESET pulse I_{LRS} and after the RESET pulse I_{RESET} are presented. It can be seen that the cell with lower I_{LRS} can be easily switched to the HRS. At higher I_{LRS}, many cells can be observed that do not or only partly switch towards the HRS. From the color of the data points, N_{disc} can be read out
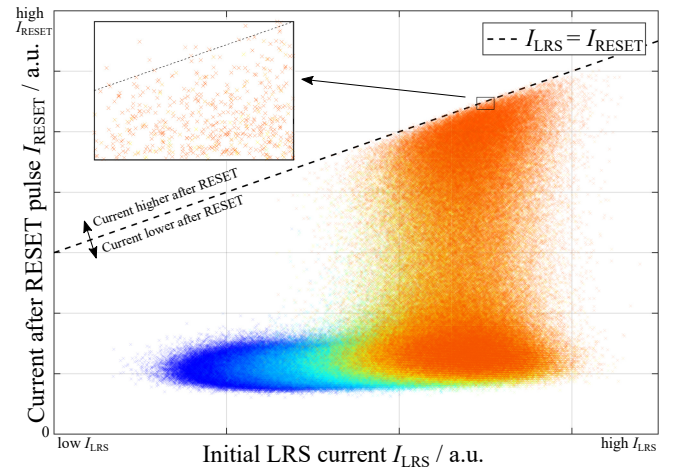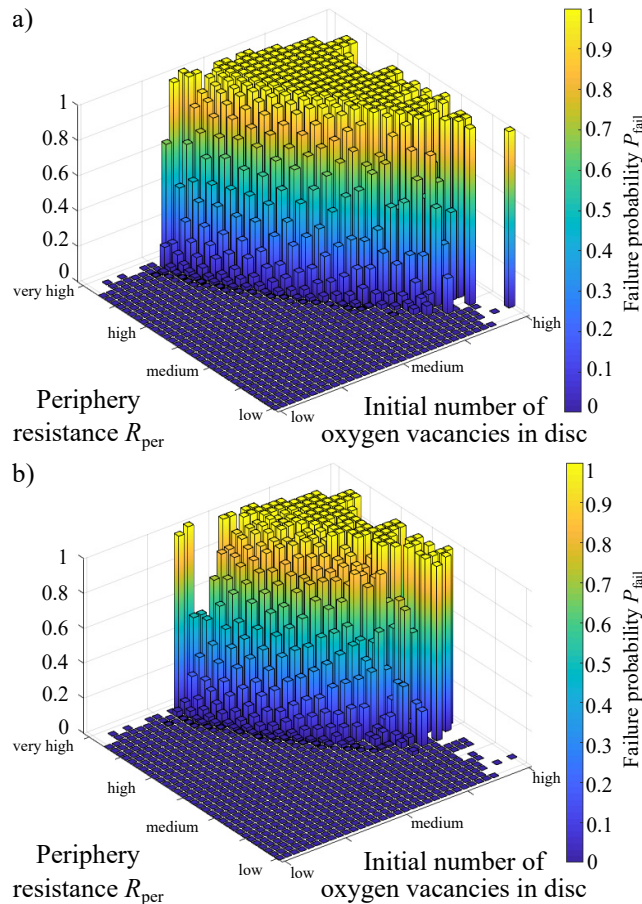

FIGURE 9. Read current of all cells presented in figure 8 in the initial LRS state before the RESET pulse and after the RESET pulse. The color of the presented cells corresponds to the ones in figure 8. The dashed line represents $I_{\text{LRS}} = I_{\text{RESET}}$, meaning the cells below have a lower and the cells above have a higher current after the RESET pulse. The current of the cells on the dashed lines did not change. The cells with lower N_{disc} and lower I_{LRS} in average can be switched to the HRS much better than the ones with higher N_{disc} and I_{LRS}. The inset shows a zoom-in, where also a few cells with higher current after the RESET pulse can be observed.
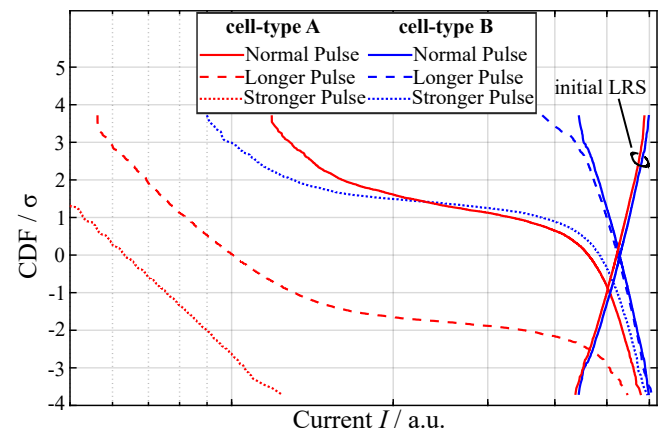
similar to the color gradient in Fig. 8 before. Here again, it is visible that the cells with high N_{disc} are much more prone to the RESET failure.

In a second step, we have a closer look at the properties of the cells that have not or only been partly switched from the LRS towards the HRS. In Fig. 10, the probability of a cell to not switch properly is presented in dependence on R_{per} and N_{disc}. N_{disc} is strongly correlated to the resistance of the cell R_{cell}, leading to a lower R_{cell} the higher N_{disc} is. In Fig. 10, a), all cells that did not switch properly to the HRS are presented. It can be seen that at least a combination of a medium N_{disc} and a high R_{per} or a high N_{disc} and a medium R_{per} is needed to obtain the failure mechanism. In Fig. 10, b) only cells that are stuck very deep in the LRS above a high current level are shown. To obtain these bits that do not or hardly switch, a combination of high N_{disc} and high or very high R_{per} is necessary, which fits well to our simple model idea from above. In the next step, we want to have a closer look at typical cells that only partly switched and typical cells that did not switch under regular

conditions. Hence, cells with a medium N_{disc} and a medium to high R_{per} on the one hand (cell-type A) and cells with a high N_{disc} and a very high R_{per} on the other hand (cell-type B) are taken. Now, additionally to the regular switching conditions, the switching behavior of the cells under longer switching pulses or switching pulses with higher V_{tot} are investigated. Thus, both cell-types are pulsed with a second pulse after the first normal RESET pulse. The second pulse either has an increased pulse time or an increased applied voltage $V_{tot} = 2.6\,V$. The current distributions of the initial LRS distributions as well as the current distributions after the normal and the second longer or stronger RESET pulse are presented in Fig. 11. As expected, it can be seen that the cells of type A (red) can partly be switched towards HRS, whereas the cells of type B (blue) are nearly completely stuck in the LRS after a normal pulse (solid line). With a second, longer pulse (dashed line), the majority of the cells of type A can be switched to the HRS, whereas the cells of type B remain stuck in the LRS. Alternatively, with a second stronger pulse, nearly all cells of type A can be switched to the HRS and even some cells of type B switch towards the HRS. On the one hand, this explains why in the experiment as in Fig. 2, a) nearly no cells occur, that only partly switched towards the HRS. The cells, that initially only partly switch towards the HRS, can be completely switched via additional pulses by a programming algorithm. The cells, that cannot even be switched via additional pulses, are typically stuck at a deep LRS state. On the other hand, the possibility to recover many of the cells initially stuck at the LRS via programming algorithms explains the rare occurrence of this failure mechanism in the experiment. Furthermore, it can be mentioned that the increase of V_{tot} during the programming algorithm seems to be more effective than increasing the pulse time.

a)

b)

FIGURE 10. Fail probabilities of the cells in dependence on the periphery resistance R_{per} and the initial number of oxygen vacancies in the disc region N_{disc}. In a) all cells that did not completely switch to the HRS are shown, in b) only cells that are stuck at very high currents are presented. The white panels indicate that no cell with these properties exists, whereas the dark blue panels show that all cells with these properties switched to the HRS.

FIGURE 11. On the right, the LRS current distributions of cells of type A (red) and B (blue) are presented. The solid lines show the current distributions after a normal RESET pulse, as before. As expected, cell-type A only partly switches towards HRS and cell-type B is completely stuck at LRS. With a second longer pulse (dashed line) or pulse with higher voltage (dotted line) the RESET is tried again. The stronger pulse has a much higher impact than the longer pulse. Cell-type A has a higher chance to reach the HRS with a second pulse, whereas type B tends to stuck in the LRS.

Although we are able to reproduce and explain most of the experimentally observed properties of the stuck-at-LRS failure mechanism, our phenomenological model and our 1D KMC model are limited. So far, our models cannot thoroughly explain the increase of the read current of single cells during RESET when becoming stuck. In the 1D KMC model, an increase of the current can indeed be observed due to single oxygen vacancies jumping randomly from the plug to the disc region during the RESET. But, in comparison with the experiment, this increase in current is too low. Thus, we propose to consider and investigate additional effects like diffusion, thermodiffusion or an additional oxygen exchange at the AE in future works [20], [28], [34]. Furthermore, in our model, so far, we assumed N_{cell} to be constant. Nevertheless, e.g., due to oxygen exchange at the interface of the OE and the oxide or thermodiffusion, N_{cell} could increase during cycling. This comes along with an increased probability of reaching unfortunate combinations of R_{cell} and R_{per}. Once reaching such a bad combination with high currents, thermodiffusion or oxygen exchange are even more encouraged. This would explain why the observed failure events increase over time in the experiment.

V. CONCLUSION

In our work, we presented the generally great endurance of 2 MBit ReRAM based on VCM integrated in a 1T-1R configuration in $28\,\mathrm{nm}$ CMOS technology. Nevertheless, looking at high cell statistics and high switching cycle numbers, a rare failure mechanism is observed after RESET. As the failure events increase over the cycle number, we investigated them in more detail. Hence, a phenomenological model simplifying the ReRAM structure into a voltage divider consisting of a cell and a periphery resistance was introduced. With this simple model, the RESET failure was proposed to occur for very rare and unlucky combinations of low cell and high periphery resistances. In the next step, we dealt with the question of how this unlucky combinations turn up and why they increase over time. Here, the SET event and the dedicated program verify algorithm play an important role. In the last step, a 1D KMC model was introduced. The high performance of compact modelling was combined with the ability to investigate random processes by the integration of KMC methods. This allowed us to simulate the RESET process with huge statistics and with special regard to rare processes. In our simulation, the RESET process could be nicely reproduced, and the failure events could be provoked by increasing the number of oxygen vacancies in the disc region. We saw that, as we assumed, cells with a bad combination of cell and periphery resistance are prone to this failure mechanism. Finally, we looked at the impact of possible programming algorithms and the outcome of longer or stronger RESET pulses. These programming algorithms have to be further optimized to prevent permanent RESET failures of the cells.

REFERENCES

[1] R. Dittmann, S. Menzel, and R. Waser, "Nanoionic memristive phenomena in metal oxides: the valence change mechanism," *Advances in Physics*, vol. 70, no. 2, pp. 155–349, 2021.

[2] Y. Chen, "Reram: History, status, and future," *IEEE Transactions on Electron Devices*, vol. 67, no. 4, pp. 1420–1433, 2020.

[3] Z. Wang, H. Wu, G. W. Burr, C. S. Hwang, K. L. Wang, Q. Xia, and J. J. Yang, "Resistive switching materials for information processing," *Nat. Rev. Mater.*, vol. 5, no. 3, pp. 173–195, 2020.

[4] D. V. Christensen *et al.*, "2022 roadmap on neuromorphic computing and engineering," *Neuromorphic Computing and Engineering*, vol. 2, no. 2, p. 022501, 2022.

[5] S. Aldana, P. García-Fernández, R. Romero-Zaliz, M. González, F. Jiménez-Molinos, F. Gómez-Campos, F. Campabadal, and J. Roldán, "Resistive switching in hfo2 based valence change memories, a comprehensive 3d kinetic monte carlo approach," *J. Phys. D Appl. Phys.*, vol. 53, no. 22, p. 225106, 2020.

[6] G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sanches, I. Boybat, M. L. Gallo, K. Moon, J. Woo, H. Hwang, and Y. Leblebici, "Neuromorphic computing using non-volatile memory," *Advances in Physics: X*, vol. 2, no. 1, pp. 89–124, 2017.

[7] R. Waser, *Nanotechnology, Volume 3: Information Technology.* Wiley-VCH, Weinheim, 2008.

[8] C. Yang, C. Wu, M. Yang, W. Wang, M. Yang, T. Chien, V. Fan, S. Tsai, Y. Lee, W. Chu, and A. Hung, "Industrially applicable read disturb model and performance on mega-bit 28nm embedded rram," in *2020 IEEE Symposium on VLSI Technology*, 2020, pp. 1–2.

[9] F. Pan, S. Gao, C. Chen, C. Song, and F. Zeng, "Recent progress in resistive random access memories: Materials, switching mechanisms, and performance," *Mater. Sci. Eng. R-Rep.*, vol. 83, pp. 1–59, 2014.

[10] D. S. Jeong, R. Thomas, R. S. Katiyar, J. F. Scott, H. Kohlstedt, A. Petraru, and C. S. Hwang, "Emerging memories: resistive switching mechanisms and current status," *Rep. Prog. Phys.*, vol. 75, no. 7, p. 76502, 2012.

[11] K. M. Kim, D. S. Jeong, and C. S. Hwang, "Nanofilamentary resistive switching in binary oxide system; a review on the present status and outlook," *Nanotechnology*, vol. 22, no. 25, p. 254002, 2011.

[12] S. Slesazeck and T. Mikolajick, "Nanoscale resistive switching memory devices: a review," *Nanotechnology*, vol. 30, no. 35, p. 352003, 2019.

[13] R. Waser, R. Bruchhaus, and S. Menzel, "Redox-based resistive switching memories," in *Nanoelectronics and Information Technology (3rd ed.).* Hoboken, NJ, USA: Wiley-VCH, 2012, pp. 683–710.

[14] S. Fukuyama, K. Maeda, S. Matsuda, K. Takeuchi, and R. Yasuhara, "Suppression of endurance-stressed data-retention failures of 40nm taoxbased reram," in *2018 IEEE International Reliability Physics Symposium (IRPS)*, 2018, pp. P–MY.4–1–P–MY.4–5.

[15] L. Goux, X. P. Wang, Y. Y. Chen, L. Pantisano, N. Jossart, B. Govoreanu, J. A. Kittl, M. Jurczak, L. Altimime, and D. J. Wouters, "Roles and effects of tin and pt electrodes in resistive-switching hfo2 systems," *Electrochem. Solid-State Lett.*, vol. 14, no. 6, p. H244, 2011.

[16] S. Menzel and J.-H. Hur, *Modeling the VCM- and ECM-Type Switching Kinetics.* Wiley-VCH, 2016, pp. 395–436.

[17] D. Cooper, C. Baeumer, N. Bernier, A. Marchewka, C. L. Torre, R. E. Dunin-Borkowski, S. Menzel, R. Waser, and R. Dittmann, "Anomalous resistance hysteresis in oxide reram: Oxygen evolution and reincorporation revealed by in situ tem," *Adv. Mater.*, vol. 29, no. 23, p. 1700212, 2017.

[18] A. Fantini, G. Gorine, R. Degraeve, L. Goux, C. Chen, A. Redolfi, S. Clima, A. Cabrini, G. Torelli, and M. Jurczak, "Intrinsic program instability in hfo2 rram and consequences on program algorithms," in *2015 IEEE International Electron Devices Meeting (IEDM)*, 2015, pp. 7.5.1–7.5.4.

[19] F. Puglisi, L. Larcher, A. Padovani, and P. Pavan, "A complete statistical investigation of rtn in hfo2-based rram in high resistive state," *IEEE Trans. Electron Devices*, vol. 62, no. 8, pp. 2606–2613, 2015.

[20] N. Kopperberg, S. Wiefels, S. Liberda, R. Waser, and S. Menzel, "A consistent model for short-term instability and long-term retention in filamentary oxide-based memristive devices," *ACS Appl. Mater. Interfaces*, vol. 13, no. 48, pp. 58 066–58 075, 2021.

[21] M. Lanza, R. Waser, D. Ielmini, J. J. Yang, L. Goux, J. Suñe, A. J. Kenyon, A. Mehonic, S. Spiga, V. Rana, S. Wiefels, S. Menzel, I. Valov, M. A. Villena, E. Miranda, X. Jing, F. Campabadal, M. B. Gonzalez, F. Aguirre, F. Palumbo, K. Zhu, J. B. Roldan, F. M. Puglisi, L. Larcher, T. Hou, T. Prodromakis, Y. Yang, P. Huang, T. Wan, Y. Chai, K. L. Pey, N. Raghavan, S. Dueñas, T. Wang, Q. Xia, and S. Pazos, "Standards for the

characterization of endurance in resistive switching devices," *ACS Nano*, vol. 15, no. 11, p. 17214–17231, 2021.

[22] S. Fukuyama, A. Hayakawa, R. Yasuhara, S. Matsuda, H. Kinoshita, and K. Takeuchi, "Comprehensive analysis of data-retention and endurance trade-off of 40nm taox-based reram," in *2019 IEEE International Reliability Physics Symposium (IRPS)*, 2019, pp. 1–6.

[23] C.-F. Yang, C.-Y. Wu, M.-C. Shih, M.-T. Yang, M.-H. Yang, T.-C. Wu, Y.-T.and Chien, C.-W. Lai, S.-C. Tsai, W.-T. Chu *et al.*, "Demonstration of high endurance capability on mega-bit rram macro and model of ppm level failures," in *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, 2022, pp. 318–319.

[24] C. Peters, F. Adler, K. Hofmann, and J. Otterstedt, "Reliability of 28nm embedded rram for consumer and industrial products," in *2022 IEEE International Memory Workshop (IMW)*, 2022, pp. 1–3.

[25] D. Ielmini, "Modeling the universal set/reset characteristics of bipolar rram by field- and temperature-driven filament growth," *IEEE Trans. Electron. Devices*, vol. 58, no. 12, pp. 4309–4317, 2011.

[26] T. Hennen, E. Wichmann, A. Elias, J. Lille, O. Mosendz, R. Waser, D. Wouters, and D. Bedau, "Current-limiting amplifier for high speed measurement of resistive switching data," *Rev. Sci. Instrum.*, vol. 92, p. 054701, 2021.

[27] H. Tseng, T. Chang, Y. Wu, S. Wu, J. Huang, Y. Chen, J. Yang, T. Lin, S. Sze, M. Tsai, Y. Wang, and A. Chu, "Impact of electroforming current on self-compliance resistive switching in an ito/gd:siox/tin structure," *IEEE Electron Device Lett.*, vol. 34, no. 7, pp. 858–860, 2013.

[28] M. von Witzleben, S. Wiefels, A. Kindsmüller, P. Stasner, F. Berg, F. Cüppers, S. Hoffmann-Eifert, R. Waser, S. Menzel, and U. Böttger, "Intrinsic reset speed limit of valence change memories," *ACS Appl. Electron. Mater.*, vol. 3, no. 12, pp. 5563–5572, 2021.

[29] C. L. Torre, "Physics-based compact modeling of valence-change-based resistive switching devices," Ph.D. dissertation, RWTH Aachen University, Aachen, Germany, 2019.

[30] C. L. Torre, A. F. Zurhelle, and S. Menzel, "Compact modelling of resistive switching devices based on the valence change mechanism," in *2019 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2019.

[31] S. Menzel, B. Wolf, S. Tappertzhofen, I. Valov, U. Böttger, and R. Waser, "Statistical modeling of electrochemical metallization memory cells," in *2014 6th IEEE International Memory Workshop (IMW), Taipeh Taiwan*, 2014, pp. 1–4.

[32] A. R. Genreith-Schriever and R. A. De Souza, "Field-enhanced ion transport in solids: Reexamination with molecular dynamics simulations," *Phys. Rev. B*, vol. 94, p. 224304, Dec 2016.

[33] C. Bengel, A. Siemon, V. Rana, and S. Menzel, "Implementation of multinary lukasiewicz logic using memristive devices," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021, pp. 1–5.

[34] D. B. Strukov, F. Alibart, and R. S. Williams, "Thermophoresis/diffusion as a plausible mechanism for unipolar resistive switching in metal-oxide-metal memristors," *Appl. Phys. A Mater. Sci. Process.*, vol. 107, no. 3, pp. 509–518, 2012.

NILS KOPPERBERG was born in Gummersbach, Germany. He received the B.Sc. and the M.Sc. degree in physics with a focus on nanoelectronics from RWTH Aachen University, Aachen, Germany, in 2016 and 2019, respectively. There, he is currently pursuing the Ph.D. degree in electrical engineering, focussing on the physical modelling and the simulation of ReRAM. Mainly the reliability, statistical effects and failure mechanisms are in the scope of his work.

STEFAN WIEFELS was born in Grevenbroich, Germany. He received the M.Sc. degree in materials science and the Ph.D. degree in electrical engineering and information technology from RWTH Aachen University, Aachen, Germany, in 2016 and 2021, respectively. His current research focuses on the electrical characterization of resistive random-access memory devices with respect to their reliability and their application in neuromorphic circuits.

KARL HOFMANN was born in Mannheim, Germany. He has studied physics and received his PhD from RWTH Aachen in 2001. Karl joined Infineon Technologies as a development engineer and published several papers on device reliability. Currently, he is focusing on the development of emerging non-volatile memories and is a member of the technical committee of the International Memory Workshop (IMW).

JAN OTTERSTEDT (Member, IEEE) was born in Bremen, Germany. He has studied electrical engineering at the University of Hannover, Germany, and in 1997 received his Dr.-Ing. degree. Afterwards, Jan has joined the semiconductor group of Siemens, which later became the Infineon Technology AG. Since more than 15 years, he is responsible for concept engineering for embedded non-volatile memories. Since 2006, Jan also lectures on "Testing Digital Circuits" at the Technische Universität München (TUM).

DIRK J. WOUTERS (Member, IEEE) received the master's and Ph.D. degrees in electrical engineering from the University of Leuven, Leuven, Belgium, in 1982 and 1989, respectively. In 2014, he joined the Institute of Electronic Materials, RWTH Aachen University, Aachen, Germany, where he focused on the research of metal-oxide-based RRAM.

RAINER WASER (Member, IEEE) received the Ph.D. degree in physical chemistry from the University of Darmstadt, Darmstadt, Germany, in 1984. In 1992, he joined the Faculty of Electrical Engineering and Information Technology, RWTH Aachen University, Aachen, Germany as Professor and in 2007 as a Director of the Institute of Solid State Research, Forschungszentrum Jülich, Jülich, Germany, in 1997. He was a recipient of the prestigious Gottfried Wilhelm Leibniz Preis in 2014.

STEPHAN MENZEL (Senior Member, IEEE) was born in Bremen, Germany. He received the Diploma degree and the Ph.D. degree (*summa cum laude*) in electrical engineering from RWTH Aachen University, Aachen, Germany, in 2005 and 2012, respectively. He is currently Senior Researcher with the Peter-Grünberg-Insitut 7, Forschungszentrum Jülich GmbH, Jülich, Germany, where he is also leading the Simulation Group.

• • •