Variability-Aware Modeling of Filamentary Oxide-Based Bipolar Resistive Switching Cells Using SPICE Level Compact Models

Christopher Bengel, Anne Siemon, Felix Cüppers, Susanne Hoffmann-Eifert, Alexander Hardtdegen, Moritz von Witzleben, Lena Hellmich, Rainer Waser, *Member, IEEE*, and Stephan Menzel, *Member, IEEE*

***Abstract*— Bipolar resistive switching (BRS) cells based on the valence change mechanism show great potential to enable the design of future non-volatile memory, logic and neuromorphic circuits and architectures. To study these circuits and architectures, accurate compact models are needed, which showcase the most important physical characteristics and lead to their specific experimental behavior. If BRS cells are to be used for computation-in-memory or for neuromorphic computing, their dynamical behavior has to be modeled with special consideration of switching times in SET and RESET. For any realistic assessment, variability has to be considered additionally. This study shows that by extending an existing compact model, which by itself is able to reproduce many different experiments on device behavior critical for the anticipated device purposes, variability found in experimental measurements can be reproduced for important device characteristics such as *I-V* characteristics, endurance behavior and most significantly the SET and RESET kinetics. Furthermore, this enables the study of spatial and temporal variability and its impact on the circuit and system level.**

Manuscript received June 12, 2020; revised July 29, 2020; accepted August 16, 2020. Date of publication August 31, 2020; date of current version December 1, 2020. This work was supported in part by the Deutsche Forschungsgemeinschaft under Grant SFB 917, in part by the European Union's Horizon 2020 Research and Innovation Program through the Project MNEMOSENE under Grant 780215, in part by the Helmholtz Association Initiative and Networking Fund [Advanced Computing Architectures (ACA)] under Project SO-092, and in part by the Federal Ministry of Education and Research (BMBF, Germany) in the Project NEUROTEC [It is based on the Jülich Aachen Research Alliance (JARA-Fit)] under Project 16ES1134 and Project 16ES1133K. This article was recommended by Associate Editor A. M. A. Ali. *(Corresponding author: Christopher Bengel.)*

Christopher Bengel, Anne Siemon, and Moritz von Witzleben are with the Institute für Werkstoffe der Elektrotechnik 2 & JARA FIT, RWTH Aachen University, 52070 Aachen, Germany (e-mail: bengel@iwe.rwth-aachen.de).

Felix Cüppers, Susanne Hoffmann-Eifert, Alexander Hardtdegen, and Stephan Menzel are with the Peter-Grünberg-Institut 7 & 10 & JARA FIT, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany (e-mail: st.menzel@fz-juelich.de).

Lena Hellmich was with the Institute für Werkstoffe der Elektrotechnik 2 & JARA FIT, RWTH Aachen University, 52070 Aachen, Germany. She is now with the Lehrstuhl für Halbleitertechnik und Institut für Halbleitertechnik, RWTH Aachen University, 52070 Aachen, Germany (e-mail: st.menzel@fz-juelich.de).

Rainer Waser is with the Institut für Werkstoffe der Elektrotechnik 2 & JARA-FIT, RWTH Aachen University, 52070 Aachen, Germany, and also with the Peter-Grünberg-Institut 7 & 10 & JARA-FIT, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany.

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSI.2020.3018502

***Index Terms*— ReRAM, BRS, VCM, compact modeling, memristor, resistive switching, variability, Verilog-A.**

I. Introduction

BIPOLAR Resistive Switching cells (BRS cells) based on the valence change mechanism (VCM) are part of an emerging class of resistive devices and are viewed to be promising candidates for future nanoelectronic applications [1]–[3]. They are foreseen to be used as storage-class-memory [4] but also for computation-in-memory [5]–[9], and neuromorphic applications [10], [11]. Neuromorphic computing describes systems that use physical phenomena as computational primitives and represent signals in an analog fashion to mimic biological information processing systems [12]. By virtue of their nonlinearity, multilevel behavior and switching statistics, VCM cells show great promise for such applications. VCM devices, like ReRAM devices in general, consist of a metal - mixed ionic and electronic conductor-metal structure. VCM switching is found in ReRAM devices which are based on oxides like HfO_x, TaO_x, TiO_x or ZrO_x [13], [14]. The original memristor publication from HP [15] also described a VCM cell. The two metal electrodes possess different work functions where the electrode with the higher and lower work functions forms a Schottky-type and an ohmic contact with the oxide, respectively. The high work function electrode is then called (electronically) active electrode (see Fig. 1 bottom electrode) while the opposing electrode is called ohmic electrode (see Fig. 1 top electrode). In its initial state directly after fabrication the oxide is highly insulating. Therefore, a forming step is performed in which through application of a high voltage the oxide is locally reduced and oxygen vacancies are created which decrease the resistance of the device. In filamentary switching systems, this vacancy generation is confined to a small part of the total cell area. The SET and RESET operation in VCM cells is based on the redistribution of oxygen vacancies inside the filament in the vicinity of the active electrode. This insight motivates splitting the filament into a disc and plug region as simplification. The disc is said to be the part of the filament near the active electrode in which the concentration of oxygen vacancies is increased during

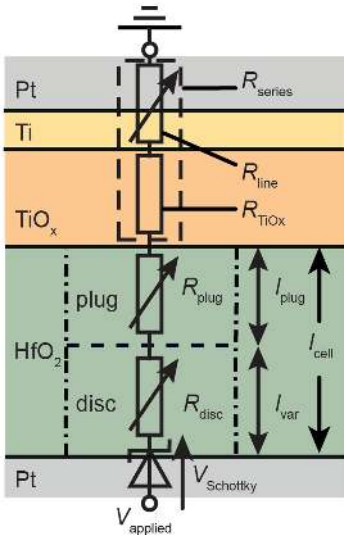

Fig. 1. Equivalent circuit diagram for the electrical model of the $Pt/HfO_2/TiO_x/Pt$ (HOTO) device. The voltage is applied to the bottom electrode. The exact stack properties can be found in [33].

SET and decreased during RESET. It is equivalent to the depletion zone of the Schottky contact [16]–[18]. This makes the disc the part where the actual switching is happening. The plug is assumed to stay well conducting during the switching process and serves as an extension of the ohmic electrode into the oxide towards the active electrode. Disc and plug can be viewed as two resistances in series where only the disc resistance is significantly modified. The SET occurs when a negative potential is applied to the active electrode. This attracts the double positively charged oxygen vacancies from the plug, thereby increasing their concentration near the active electrode and reducing the resistance of the disc. The concentration in the plug is assumed to be large enough that this reduction does not significantly alter the resistance of the plug. After the SET, the device is in the LRS since disc and plug both have a low resistance. The RESET occurs when a positive potential is applied to the active electrode which repels the oxygen vacancies from the disc back into the plug and in turn increases the resistance of the disc. After the RESET the device is in the HRS since the disc now determines the resistance of the disc plug series resistances [17].

One large challenge that hampers the introduction of VCM cells, however, is their very pronounced variability [19]–[24]. For example, the variation in the switching kinetics, this is, the time that is needed to switch a cell between two resistive states at a given voltage, can vary over multiple orders of magnitude [25]. The resistive states also show a strong variation with a factor of 5 - 100 in the high resistive state (HRS) and 2 - 10 in the low resistive state (LRS) [26]. Those properties hinder the development of circuits and make the assessment of their functionality challenging. Variability in resistive devices can be categorized as spatial (device-to-device) variability, which is assumed to stem from process variations, and temporal (cycle-to-cycle) variability. Cycle-to-cycle variability arises due to the stochastic nature of the physical phenomena involved in the resistive switching process. For example, the switching filament was found to form on multiple locations and with different geometrical properties inside the cell representing a source of cycle-to-cycle variability [27]. Variability for filamentary VCM devices has been modeled before [28]–[31]. In [28], the authors use a different set of parameters for the SET and for the RESET process in order to better fit the median *I-V* curves. Additionally, their model considers variability connected to the state variable, which describes the gap distance between a filament tip and the opposing electrode. This means, that the relationship between the device resistance and the state variable is unambiguous. In that case, the response of the cell to an electrical stimuli, like the application of a voltage pulse, will only depend on the one state variable. In [29], in contrast, only device-to-device variability is regarded and cycle-to-cycle variability is neglected. Wald and Kvatinsky [30] considered the influence of variability for the performance of logic operations using the VTEAM model. The VTEAM model is a threshold voltage based behavioral SPICE-level model. For their study they varied empirical fitting parameters. However, the problem with a behavioral model is that the parameters are only faintly connected to the actual device physics, which weakens the predictive capability of their considerations. In [31], variability is introduced into the compact model by drawing the barrier thickness that determines the HRS from a Gaussian distribution. This model, however, has only been verified with regard to quasi static behavior like slow *I-V*-sweeps and the resistance distributions of LRS and HRS during these sweeps.

In [32] we showed a deterministic model, the Jülich Aachen Resistive Switching Tools (JART) VCM v1b model that describes the SET and RESET kinetic behavior with a special focus on the initial states before the SET and the RESET. For [32] the equations have been updated and modified and the parameters have been adjusted to better match the dynamics of SET and RESET processes as well as the *I-V* characteristic with the high resistive state (HRS) and the low resistive state (LRS), compared to JART VCM v1, which was presented in [33].

In this work, the JART VCM v1b model is extended to account for both the device-to-device and the cycle-to-cycle variability. This is in detail achieved by adding (i) a random initialization of the filament parameters and (ii) a random change of the parameters during the simulation. In analogy to the results of [27], the filament geometry and oxygen vacancy concentration in the disc region are varied to achieve variable behavior. The developed SPICE-level model can describe the slow ramped voltage-current measurements as well as the dynamic behavior under square pulse stress using. In addition, it can reproduce the distribution of SET and RESET delays, endurance measurements and *I-V* characteristics, representing a wide range of slow ramped voltage-current measurements to dynamic measurements in contrast to the verification using only slow ramped voltage-current *I-V* sweeps like [31]. It offers a consistent description of SET and RESET by assuming a different conduction mechanism and a different effective thermal resistance for SET and RESET. The state variable, which describes the oxygen vacancy concentration in the vicinity of the so-called active electrode, is not used to model the

variability. Here, the state variables does not have a stochastic component, but the filament parameters are varied randomly. Thus, an ambiguous relation between the state variable and the device state results, as opposed to [28]. This leads to a variable dynamic behavior. Additionally, the dynamic behavior turns out to be different for different filament parameters. The cycle-to-cycle variability is coupled to the switching process itself and since it is a physics-based compact model the meaning of the parameters is closely linked with physics in contrast to earlier behavioral models like [30].

II. Requirements for Memristive Device Models

The utilization of the opportunities enabled by VCM cells remains demanding. This is in part because there exists a gap in understanding between the circuit designers and the device engineers, since circuit designers are classically less concerned with the exact physical phenomena of their circuit components, while device engineers are usually no experts in circuit design. Compact models can help in bridging this gap. Through its equations, a compact model links physical phenomena to experimental data. Its high simulation speed allows performance predictions for circuits and larger systems. In order to do so, however, the models have to be calibrated to more physically accurate ones such as finite element models or Kinetic Monte Carlo models. At the same time, they have to be calibrated to measurements like *I-V* sweeps and SET and RESET dynamic measurements, ideally under different operating conditions [34].

With this calibration, compact models can reliably simulate the actually incurred use cases that appear when the cells are operated in micro chips. If models are only able to reproduce a specific set of measurements, such as slow *I-V* sweeps with time scales of a few seconds per sweep they can hardly be viewed as being useful in circuit simulations. If VCM cells are to be used inside micro chips operating at MHz to GHz frequencies, the models have to reproduce device behavior at these times scales accurately. This requires a special focus on modeling the dynamical behavior. Equations have to be found that are valid over the range of time scales required by micro chip applications. Of course, ideally, their validity should not be limited to a specific time scale.

The most anticipated use case for VCM cells is as synapses in artificial neural networks. Their advantages for this use case are the possibility to perform the inference operation and the weight update operation in the same physical location without needing to transfer the weights from an outside memory which would alleviate the penalty of the memory bottleneck [35]. Additionally, their analog switching capabilities enable multi bit storage per weight [36], which gives them a density advantage compared to resistive technologies that are only binary like MRAM [37]. Their non-volatility gives them an edge over volatile solutions in use cases where a large part of the totally consumed power comes from the standby power. Those neural networks can either be built up from fully passive arrays [38] or from 1T1R arrays [11], which already today enables more advanced architectures. However, those experimental realizations are still in a very early developmental stage compared to CMOS-based counterparts. Problems that arise are variability in the device resistance for high resistance regimes, sneak paths and half select issues, which can lead to a drift of the device state. Further issues are *I-V* nonlinearity, SET and RESET asymmetry and a lack of clarity on the exploitation of stochastic switching behavior [20]. In order to better tackle these issues, understanding the cells will be more useful than progressing by trial and error. Circuit simulations using deterministic device models cannot adequately describe BRS. Therefore, circuit simulations have to take a physically motivated variability of the BRS into account.

III. Device Modeling

Simulations using the deterministic version of the JART VCM v1b model were shown in [32]. Since the exact changes were only indicated in [32], this work is used to describe the deterministic version from [32] as well as the variability version. The JART VCM v1b model therefore exists as a deterministic model and as a variability model. The variability model represents the main contribution of this article. The considered device stack is $Pt/HfO_2/TiO_x/Pt$ fabricated as a nanocrossbar device with a junction of 100 nm x 100 nm. The fabrication and characterization of this device were described earlier [33]. Prior to the electroforming process, the characteristic microstructural properties of the device's oxide layers are properly described by HfO_2 and TiO_x for the hafnium dioxide and the titanium oxide, respectively. For reasons of simplicity, this nomenclature will be kept for the full manuscript. However, it should be clear that the oxygen vacancy concentrations in the plug and disc regime of the electroformed and switched devices correspond to a local change in the stoichiometry, which will significantly deviate from the stoichiometry of the layers in the pristine device.

The compact model assumes that the HfO_2 layer is split into a well conducting plug region and a disc region with a modifiable resistance. The two versions enable a description of the switching behavior on different levels of complexity. In the deterministic version only the change of the oxygen vacancy concentration in the disc "N_{disc}" is considered during switching whereas all other physical parameters are kept constant. The index 'det' denotes the parameter values that relate to this version of the JART VCM v1b model. These are the radius of the filament "r_{det}", the length of the disc region "l_{det}" and the minimum and maximum concentration of oxygen vacancies in the disc "$N_{\mathrm{disc,\,(max/min),\,det}}$". By the variation of the length of the disc region also the length of the plug region is varied since the cell length at any time step is defined as constant. If the deterministic version is to be used, all parameters get the index 'det' instead of the index 'var' in section III-A. Simulations with the deterministic parameter set can be found in [32]. The higher level of complexity in the switching modeling is represented by a variability model. The notation in III-A corresponds to this variant. The notation has been chosen in this way in order to be consistent with the actual Verilog-A code that will be published on our website [39]. In contrast to the deterministic approach, the variability model enables and considers variation of the parameters within predefined ranges

for the maximum and minimum oxygen vacancy concentration "$N_{\mathrm{disc,\,(max/min),\,var}}$" as well as the radius "r_{var}" and the length of the disc "l_{var}" throughout the simulation as described in III-B. By this approach, all the physical equations in III-A are formulated the same way for the deterministic version and for the variability version. On top of these equations, the variability model introduces a variation in certain parameters which are assumed to play an important role for the variability found in these devices in experiments.

A. A Physically Motivated VCM Cell Model

The switching model is based on the movement of oxygen vacancies to and from the metal/insulating metal-oxide interface (here Pt/HfO_2). The concentration of these oxygen vacancies in the interface region (disc) N_{disc} is chosen as state variable. The change in the state variable is calculated as

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

It is altered by the ionic current

$$I_{\mathrm{ion}} = z_{\mathrm{Vo}} e A c_{\mathrm{Vo}} a \nu_{\mathrm{o}} F_{\mathrm{limit}} \cdot \Big(\exp\big(-\frac{\Delta W_{\mathrm{A,\,min}}}{k_{\mathrm{B}} T}\big) - \exp\big(-\frac{\Delta W_{\mathrm{A,\,max}}}{k_{\mathrm{B}} T}\big)\Big), \tag{2}$$

where e is the elementary charge, z_{Vo} is the oxygen vacancy charge number, a is the ionic hopping distance, ν_0 is the attempt frequency, $A = \pi r_{\mathrm{var}}^2$ is the area of the switching filament and k_{B} is Boltzmann's constant. The concentration c_{Vo} is the average concentration of disc and plug and given as

$$c_{\mathrm{Vo}} = \frac{N_{\mathrm{plug}} + N_{\mathrm{disc}}}{2}. \tag{3}$$

The ionic current is limited by the factor F_{limit}. This ensures that the vacancy concentration N_{disc} neither exceeds $N_{\mathrm{disc,\,max}}$ nor falls below $N_{\mathrm{disc,\,min}}$.

$$F_{\mathrm{limit}} = \begin{cases} \left[1 - \left(\frac{N_{\mathrm{disc,\,min}}}{N_{\mathrm{disc}}}\right)^{10}\right] & \text{for } V_{\mathrm{applied}} > 0\ V \\ \left[1 - \left(\frac{N_{\mathrm{disc}}}{N_{\mathrm{disc,\,max}}}\right)^{10}\right] & \text{for } V_{\mathrm{applied}} < 0\ V. \end{cases} \tag{4}$$

The exponent of 10 is chosen such that the ionic current is only influenced when N_{disc} is close to its limits. The temperature T is calculated as

$$T = I \cdot (V_{\mathrm{disc}} + V_{\mathrm{plug}} + V_{\mathrm{schottky}}) \cdot R_{\mathrm{th,\,SET/RESET}} + T_0, \tag{5}$$

with the ambient temperature T_0 and the thermal resistance of the cell $R_{\mathrm{th,\,SET/RESET}}$. I denotes the electrical current through the cell. All simulations were performed with an ambient temperature of 293 K. To achieve a more consistent description of the SET and RESET dynamics a polarity dependent thermal resistance $R_{\mathrm{th,\,SET}}$ and $R_{\mathrm{th,\,RESET}}$ is introduced. $R_{\mathrm{th,\,SET/RESET}}$ is calculated as

$$R_{\mathrm{th,\,SET/RESET}} = R_{\mathrm{th0,\,SET/RESET}} \cdot \frac{{r_{\mathrm{det}}}^2}{{r_{\mathrm{var}}}^2}. \tag{6}$$

Previous versions of the model included the Mott-Gurney law to describe the ionic current. The high electrical fields that arise due to the thin disc region, however, are not sufficiently described by the Mott-Gurney law [40]. Genreith-Schriever and De Souza presented a modification of the Mott-Gurney law for high electrical fields, which is used here. The hopping migration barriers for jumps in the direction of the electrical field (forward jumps) $\Delta W_{\mathrm{A,\,min}}$ and for jumps in the opposite direction of the electrical field (reverse jumps) $\Delta W_{\mathrm{A,\,max}}$ are calculated according to [40] as

$$\Delta W_{\mathrm{A,\,min}} = \Delta W_{\mathrm{A}} \cdot \Big(\sqrt{1-\gamma^2} - \gamma \frac{\pi}{2} + \gamma \cdot \arcsin(\gamma)\Big) \tag{7}$$

and

$$\Delta W_{\mathrm{A,\,max}} = \Delta W_{\mathrm{A}} \cdot \Big(\sqrt{1-\gamma^2} + \gamma \frac{\pi}{2} + \gamma \cdot \arcsin(\gamma)\Big), \tag{8}$$

where γ is defined as

$$\gamma = \frac{e z_{\mathrm{Vo}} a E_{\mathrm{SET/RESET}}}{\Delta W_{\mathrm{A}} \pi}, \tag{9}$$

with the nominal activation barrier ΔW_{A}.

The electrical field during the SET is calculated as

$$E_{\mathrm{SET}} = \frac{V_{\mathrm{disc}}}{l_{\mathrm{var}}}, \tag{10}$$

and during the RESET as

$$E_{\mathrm{RESET}} = \frac{V_{\mathrm{schottky}} + V_{\mathrm{disc}} + V_{\mathrm{plug}}}{l_{\mathrm{cell}}} = \frac{V_{\mathrm{cell}}}{l_{\mathrm{cell}}}, \tag{11}$$

where V_{disc}, V_{plug} and V_{schottky} are the voltages dropping over R_{disc}, R_{plug} and the Schottky diode, which is found at the Pt to HfO_2 interface.

Fig. 1 shows the equivalent circuit diagram that is described by the equations in this chapter. The voltage is applied to the bottom electrode, which leads to a negative voltage for the SET and a positive voltage for the RESET operation. The total length of the cell l_{cell} is defined as the varying lengths of the disc l_{var} plus the length of the plug l_{plug}.

By applying a voltage, the oxygen vacancy concentration in the disc N_{disc} can be controlled between N_{min} and N_{max}, which label the minimum and maximum oxygen vacancy concentration in the disc. In this model the plug is considered to be an infinite reservoir of oxygen vacancies, that can be exchanged with the disc. Therefore, the vacancy concentration in the plug is constant and represented by N_{plug}.

The resistance of TiO_x/Ti/Pt is summarized in the series resistance R_{series}. It consists of a constant resistance for the TiO_x ($R_{\mathrm{TiO_x}}$) and a current dependent resistance of the lines contacting the device (R_{line}), which has been calibrated to measurements of line test structures. In total the series resistance is calculated similarly to [41] as

$$\begin{aligned} R_{\mathrm{series}} &= R_{\mathrm{TiO_x}} + R_{\mathrm{line}} \\ &= R_{\mathrm{TiO_x}} + R_0 \cdot (1 + \alpha_{\mathrm{line}} R_0 I^2 R_{\mathrm{th,\,line}}), \end{aligned} \tag{12}$$

where R_0 represents the resistance of the lines for the limit of zero current, α_{line} represents the temperature coefficient of the line and $R_{\mathrm{th,\,line}}$ represents the thermal resistance of the lines. The difference in the resistance is around 100 Ω from $I = 0$ A to $I = 700$ μA and therefore might seem negligible. Other groups have shown, however, that even such a small difference

in the series resistance of the cell will lead to a change in the switching time of multiple orders or magnitude [42], indicating that it can not be neglected.

The disc and plug resistances are given by

$$R_{\text{plug/disc}} = \frac{l_{\text{plug/var}}}{z_{\text{Vo}} e A N_{\text{plug/disc}} \mu_{\text{n}}}, \tag{13}$$

where μ_{n} represents the electron mobility.

The Pt/HfO_2 interface is described by a Schottky diode, possessing an electrostatic barrier that is varied according to N_{disc}. The current through this barrier is described as

$$\begin{aligned} I_{\text{schottky, V<0}} &= -\sqrt{\pi W_{00} e \cdot \left(- V_{\text{Schottky}} + \frac{\phi_{\text{Bn}}}{\cosh^2(\frac{W_{00}}{k_{\text{B}} T})} \right)} \\ &\quad \cdot \exp\left(\frac{-e\phi_{\text{Bn}}}{W_0}\right)\left(\exp\left(\frac{-e V_{\text{Schottky}}}{\varepsilon'}\right) - 1\right) A A^* \frac{T}{k_{\text{B}}} \end{aligned} \tag{14}$$

for the SET process and

$$\begin{aligned} I_{\text{schottky, V>0}} &= A A^* T^2 \exp\left(\frac{-e\phi_{\text{Bn}}}{k_{\text{B}} T}\right) \\ &\quad \cdot \left(\exp\left(\frac{e V_{\text{schottky}}}{k_{\text{B}} T}\right) - 1\right) \end{aligned} \tag{15}$$

for the RESET process with A^* being the Richardson constant, $e\phi_{\text{Bn}}$ being the effective electrostatic barrier height and V_{schottky} being the voltage across the barrier. W_{00} is given by

$$W_{00} = \frac{eh}{4\pi} \sqrt{\frac{z_{\text{Vo}} N_{\text{disc}}}{m^* \varepsilon}} \tag{16}$$

and W_0 as

$$W_0 = \frac{W_{00}}{\tanh\left(\frac{W_{00}}{k_{\text{B}} T}\right)} \tag{17}$$

and ε' as

$$\varepsilon' = \frac{W_{00}}{\frac{W_{00}}{k_{\text{B}} T} - \tanh(\frac{W_{00}}{k_{\text{B}} T})}. \tag{18}$$

Here, h refers to Planck's constant and ε to the dielectric permittivity of the oxide. The applied voltage leads to an electric field, which lowers the electrostatic barrier $e\phi_{\text{Bn0}}$ to the effective electrostatic barrier $e\phi_{\text{Bn}}$

$$e\phi_{\text{Bn}} = e\phi_{\text{Bn0}} - e\sqrt[4]{\frac{e^3 \, z_{\text{Vo}} N_{\text{disc}} \cdot (\phi_{\text{Bn0}} - \phi_{\text{n}} - V_{\text{Schottky}})}{8\pi^2 \varepsilon_{\phi_{\text{B}}}^3}}, \tag{19}$$

in which $e\phi_{\text{n}}$ defines the energy difference between conduction band and Fermi level. $\varepsilon_{\phi\text{B}}$ describes the permittivity related to the process of image-force barrier lowering.

The parameter values are listed in Table I. They are chosen to fit the SET and RESET kinetic as well as the *I-V* characteristic together with the values of HRS and LRS and they are the same as used in [32].

TABLE I
DETERMINISTIC MODEL PARAMETERS

$A_{\text{det}} = \pi r^2 = 6.36 \cdot 10^{-15}$ m^2	$r_{\text{det}} = 45$ nm
$N_{\text{disc, min, det}} = 0.008 \cdot 10^{26}$ m^{-3}	$N_{\text{disc, max, det}} = N_{\text{plug}} = 20 \cdot 10^{26}$ m^{-3}
$l_{\text{cell}} = 3$ nm	$l_{\text{det}} = 0.4$ nm
$l_{\text{plug}} = 2.6$ nm	$a = 0.25$ nm
$R_{\text{series}} = 1.37$ kΩ (I = 0 μA)	$R_{\text{series}} = 1.46$ kΩ (I = 700 μA)
$R_{\text{line}} = 719$ Ω (I = 0 μA)	$R_{\text{line}} = 810$ Ω (I = 700 μA)
$R_{\text{TiOx}} = 650$ Ω	$\Delta W_{\text{A}} = 1.35$ eV
$\nu_0 = 2 \cdot 10^{13}$ Hz	$\mu_{\text{n}} = 4 \cdot 10^{-6}$ m^2/Vs
$R_{\text{th0, SET}} = 15.72 \cdot 10^6$ K/W	$R_{\text{th0, RESET}} = 4.24 \cdot 10^6$ K/W
$e\phi_{\text{Bn0}} = 0.18$ eV	$e\phi_{\text{n}} = 0.1$ eV
$R_{\text{th,line}} = 90471.47$ K/W	$R_0 = 719.24$ Ω
$\alpha_{\text{line}} = 3.92 \cdot 10^{-3}$ 1/K	$m^* = 9.11 \cdot 10^{-31}$ kg
$e = 1.6 \cdot 10^{-19}$ C	$T_0 = 293$ K
$A^* = 6.01 \cdot 10^5$ A/m^{2}K^2	$z_{\text{Vo}} = 2$
$k_{\text{B}} = 1.38 \cdot 10^{-23}$ J/K	$\varepsilon_0 = 8.854 \cdot 10^{-12}$ As/Vm
$\varepsilon_{\phi_{\text{B}}} = 5.5 \cdot \varepsilon_0$	$\varepsilon = 17 \cdot \varepsilon_0$
$h = 6.626 \cdot 10^{-34}$ Js	

B. A Physically Motivated Variability Model

Baeumer *et al.* [27] showed that cycle-to-cycle variability is influenced by the possibility of multiple resistance configurations which have different active filaments. The different filaments have different oxygen vacancy concentrations and different geometries. The variate parameters are chosen accordingly as

- $N_{\text{disc, max}}$ and $N_{\text{disc, min}}$, the maximal and minimal oxygen vacancy concentration in the disc,
- r, the radius of the disc and plug filament and
- l, the length of the disc,

all connected to the filament geometry or the bounds for N_{disc}.

1) Device-to-Device Variability: Before the initialisation of the circuit, a random set of parameters is drawn from a truncated Gaussian distribution for every cell independently. A Gaussian distribution over x has a nonzero probability for all x, therefore a truncated Gaussian distribution is used to restrict the model parameter values to reasonable ranges. The different parameters have values of different orders of magnitude. While the radius and the length are both in the range of nanometers (10^{-9} m), the bounds for the oxygen vacancy concentration are in the range of (10^{26} m^{-3}). In order to obtain a reasonable behavioral model, we choose, arbitrarily, to apply the same relative standard deviation for all parameters. In this way, device-to-device variability is realized.

Fig. 2 shows the variation in the variability of the SET delay, which is defined as the time difference between applying a voltage pulse and the time at which the current through the cell has the highest slope. Here, a voltage pulse of -1.05 V with a rise time of 1 ns was applied (see black curve). It can be observed that the SET delay varies from a minimum of around 4 ns (red curve) up to around 20 μs (light green curves), showing a variation factor of more than

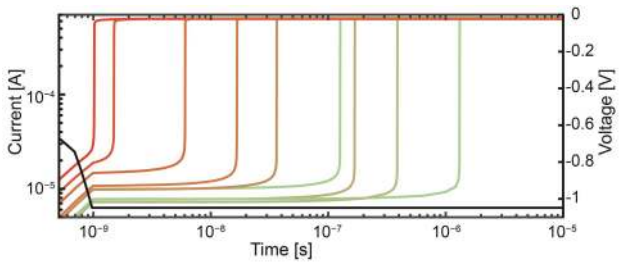

Fig. 2. The applied voltage (black curve) is shown together with the current transients (colored curves) for nine differently initialized cells. It can be observed that although the SET switching delay, which is defined as the time difference between applying a voltage pulse and the time at which the current through the cell has the highest slope varies by three orders of magnitude the general shape of the curves stays the same. Additionally it can be observed that a smaller HRS leads to faster switching than a larger HRS.

three orders of magnitude. In general, it is observed that higher initial HRS lead to a longer SET delay, which is consistent with experimental work [32], [42] and [43]. Even though there is a huge variation in the delay, the shapes of the curves still look very similar and are consistent with previous studies [25], [42] and [43]. This suggests that the underlying physical processes do not change but are merely accelerated. The transients can be divided into two parts. First a slow and linear increase of the current, which arises from an electrical field induced accumulation of oxygen vacancies at the active electrode in unison with an increase of the local temperature due to the higher conductivity. Secondly an abrupt thermal runaway process which is triggered by the positive feedback of temperature increase and increased mobility of the oxygen vacancies [25] and [42]. This suggests that the physical mechanisms making up the model are stable to a certain extent of parameter variation since their "general behavior" is not changed if those parameters are varied.

2) Cycle-to-Cycle Variability: To model cycle-to-cycle variability, the variable parameters are changed before every SET or RESET voltage pulse. This change has a variable but confined step size whose maximum is chosen to be 10 % of the current value. This maximum value can be changed to control the strength of the cycle-to-cycle variability. In order to change the parameters in a stochastic way, a random walk model in one dimension with equal probabilities for increasing or decreasing the parameter is chosen. Random walk is a mathematical model for describing random movement. It is a stochastic process with discrete time steps, where every step is a Bernoulli process that either increases or decreases the value that it describes. Each step is random in the sense that each step is independent of the previous step (e.g. [44]). Considering the parameter ξ in the step k with a maximum change in every step of $\Delta\xi_{\max}$. $\Delta\xi_{\max}$ is chosen here as 90 % for $N_{\text{disc, min, var}}$ and as 10 % for the other parameters. Choosing a larger step size for $N_{\text{disc, min, var}}$ enables a larger variation in the HRS between subsequent cycles which corresponds with the experimental observations. The determination of the next random walk parameter is described as

$$\xi_k = \xi_{k\text{-}1} \cdot (1 \pm \Delta\xi_{\max} \cdot P), \tag{20}$$

where P is a random number between zero and one, which randomly scales the step size between zero and $\Delta\xi_{\max}$. The direction of the change is described by the $\pm$ and has a 50 % chance to be plus and 50 % to be minus. Before the actual SPICE-level simulation, all the needed calculations are performed in MATLAB and written to a text file, which is read in by the circuit simulator (Spectre) as a parameter set file. This file contains one column for each variable parameter (four per cell) and additionally a column with the points in time at which the new values should be adopted. The points in time are chosen such that they align with the start of the SET/RESET pulses. The circuit netlist is then generated with placeholder variables (wildcards) for the variability parameters. During the circuit simulation the parameter file is read by Spectre and its values are inserted into the model. While $N_{\text{disc, max, var}}$ and $N_{\text{disc, min, var}}$ are directly inserted into the model, l_{var} and r_{var} are coupled with the change of the state variable.

For negative voltages (SET direction)

$$r_{\text{var}} = r_{\text{old}} + (r_{\text{new}} - r_{\text{old}})\left(\frac{N_{\text{disc}} - N_{\text{disc, old}}}{N_{\text{disc, max, var}} - N_{\text{disc, old}}}\right) \tag{21}$$

and

$$l_{\text{var}} = l_{\text{old}} + (l_{\text{new}} - l_{\text{old}})\left(\frac{N_{\text{disc}} - N_{\text{disc, old}}}{N_{\text{disc, max, var}} - N_{\text{disc, old}}}\right), \tag{22}$$

holds. Likewise, for positive voltages (RESET direction)

$$r_{\text{var}} = r_{\text{old}} + (r_{\text{new}} - r_{\text{old}})\left(\frac{N_{\text{disc, old}} - N_{\text{disc}}}{N_{\text{disc, old}} - N_{\text{disc, min, var}}}\right) \tag{23}$$

and

$$l_{\text{var}} = l_{\text{old}} + (l_{\text{new}} - l_{\text{old}})\left(\frac{N_{\text{disc, old}} - N_{\text{disc}}}{N_{\text{disc, old}} - N_{\text{disc, min, var}}}\right) \tag{24}$$

are used. The 'var' index denotes the parameter that connects to the physical equations, while the 'old' index denotes the last parameter being used right before the application of a SET or RESET voltage and the 'new' index denotes the new value that was determined by the random walk algorithm. 'old' and 'new' are therefore only connected to the variability model and only influence the model through their connection to 'var'. One effect of this coupling is that r_{var} and l_{var} only change if the state of the cell N_{disc} changes. In addition, only intermediate values for r_{var} and l_{var} are attained if the switching is incomplete, i. e. the state variable does not switch to $N_{\text{disc, min, var}}$ or $N_{\text{disc, max, var}}$. Complete switching is defined here as switching into the limits $N_{\text{disc, min, var}}$ or $N_{\text{disc, max, var}}$.

Fig. 3 shows the functionality of the cycle-to-cycle variability implementation during an *I-V* sweep simulation with a current compliance of -400 μA. Fig. 3 (a) shows the applied voltage signal, which is swept between −1.3 V and +1.3 V with a sweep rate of 1 V/s. Fig. 3 (b) shows the resulting current through the cell. At the first dashed vertical line 1, a SET occurs. It is characterized by an abrupt absolute increase of the current, which is limited to −400 μA. The RESET

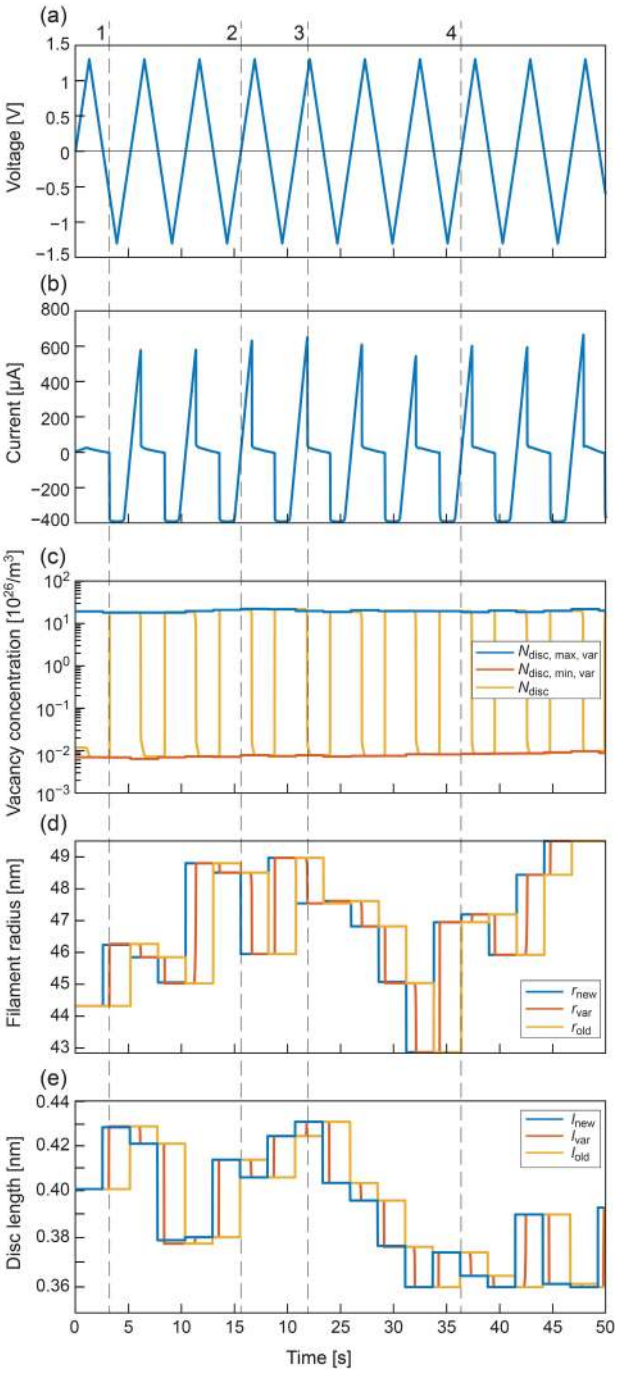

Fig. 3. The functionality of the cycle-to-cycle variability is shown during an *I-V* sweep simulation with a current compliance of −400 µA. The graphs (a) to (e) show different properties of the simulations during the simulation with aligned time axes. The applied voltage is swept between −1.3 V and +1.3 V with a sweep rate of 1 V/s (a). (b) shows the resulting current at the same time points. (c) shows the model parameters N_{disc}, $N_{\text{disc, max, var}}$ and $N_{\text{disc, min, var}}$, (d) shows r_{new}, l_{var} and r_{old} and (e) shows the parameters l_{new}, l_{var} and l_{old}.

can be observed at the third dashed vertical line 3. It is characterized by an initially abrupt decrease of the current from a high positive value. Since the current compliance used in Fig. 3 was −400 µA for the SET the RESET is more abrupt than the RESET in Fig. 4 (b) which was simulated with a lower current compliance of −100 µA. This agrees with the results from [33]. Fig. 3 (c) shows the oxygen vacancy concentration N_{disc} and its maximum and minimum values, $N_{\text{disc,max,var}}$ and $N_{\text{disc,min,var}}$. For a SET (vertical line 1), the oxygen vacancy concentration abruptly increases, which corresponds to a decrease of the device resistance characteristic for the SET. During the RESET (vertical line 3), the oxygen vacancy concentration first decreases abruptly and then changes more gradually as the RESET switching slows down when higher resistances are reached. The parameters $N_{\text{disc, max, var}}$ and $N_{\text{disc, min, var}}$ are used to confine N_{disc}. They change whenever the voltage polarity changes. If 0 V is crossed, the conditions change from SET conditions ($V_{\text{applied}} < 0$ V) to RESET conditions ($V_{\text{applied}} > 0$ V) and back. The vertical lines 2 and 4 indicate points in time when 0 V is crossed (see Fig. 3 (a)) and the values of $N_{\text{disc, max, var}}$ and $N_{\text{disc, min, var}}$ change (cf. Fig. 3 (c)). Fig. 3 (d) shows the parameters connected to the filament radius r_{new}, r_{var} and r_{old}. In the JART VCM v1 b model, the filament radius is only used to calculate the filament area. As described above, r_{var} denotes this radius. It changes towards r_{new} in accordance with equations (21) and (23) as shown in Fig. 3 (d). The values of r_{new} and r_{old} always change when the polarity switches (see vertical lines 2 and 4). At these instants r_{old} changes to r_{var} and r_{new} changes to the corresponding value from the parameter set file. Fig. 3 (e) shows the parameters connected to the disc length l_{new}, l_{var} and l_{old}. Their behavior corresponds to the one observed in Fig. 3 (d). In our simulation setup we generate the netlists and simulation stimuli and perform the evaluation of results in MATLAB in an automated fashion. In this way we synchronize the parameter file with the SET and RESET events. It is also automated to work with the different simulations shown in this article.

The first few rows of the parameter set file, which was used to simulate Fig. 3, can be seen below

```
pset1 paramset{
time        Ndiscmin0   Ndiscmax0   rnew0     lnew0   ...
0           0.0076      20.5        44.3e-9   0.401   ...
2.6         0.0070      19          46.1e-9   0.428   ...
5.2         0.0068      19.3        45.9e-9   0.42    ...
7.8         0.0071      20          44.9e-9   0.379   ...
...         ...         ...         ...       ...     ...
}.
```

The number 0 at the end of the parameter name rnew0, lnew0 is used to distinguish multiple devices when they are simulated at the same time. The scales in this example deviate from the actual scales i. e. Ndiscmin0 is 0.0076, Ndiscmax0 is 19 and lnew0 is 0.401. The scaling of these values by 10^{26}, 10^{26} and 10^{-9} is performed in the equations of the compact model in order to improve the convergence behavior. This file actually corresponds to the first two *I-V* sweeps of Fig. 3. The first row of values shows the initial device parameters that are valid for the first half of the first sweep from 0 V to 1.3 V and back to 0 V (RESET direction). The second row of parameters shows the new values that can be reached during the second half of the first sweep from 0 V to −1.3 V and back to 0 V (SET direction) and therefore starts at 2.6 s. The third row of parameters shows the new values that can be reached during the first half of the second sweep from 0 V to 1.3 V and back to 0 V (RESET direction) and therefore starts at 5.2 s,

while the fourth row of parameters again shows the values to be reached during the second half of the second sweep (SET direction). To sum up, two rows of parameters are needed for one *I-V* sweep; one for the SET branch and one for the RESET branch.

3) Understanding the Variability Parameters Influence on the Compact Model Behavior: In order to successfully work with this compact model and to exploit its variability capabilities, it is necessary to understand the influence of the different variability parameters on the compact model behavior. As an example, it will be analyzed how the variability parameters influence the SET delay. The variation in this characteristic quantity is displayed in Fig. 2. The influence of the variability parameters can be understood by considering the model equations. The thermal resistance of the cell R_{th} is proportional to $1/r^2$ (see Eq. (6)), meaning that decreasing the radius will increase the thermal resistance, which in turn will increase the temperature of the filament according to Eq. (5). This temperature increase has a very strong impact on the switching speed [41]. For the influence of the disc length on the SET delay it is found that the electrical field during the SET (Eq. (10)) and the ionic current (see Eq. (2)) are inversely proportional to the disc length. Therefore, a shorter disc length will lead to a higher electrical field during the SET and to a higher ionic current, which both increase the switching speed. If the initial oxygen vacancy concentration in the disc is elevated, the initial resistance is smaller, which results in a higher current at a specific voltage. This higher current results in a higher filament temperature due to Joule heating, which in turn speeds up the switching process significantly. To conclude, the minimum (maximum) SET delay is achieved for the minimum (maximum) values of r and l and for the maximum (minimum) initial oxygen vacancy concentration N_{disc}. A deterministic simulation with these parameters will result in the fastest (slowest) possible device.

This kind of analysis can be performed for different cell characteristics like the ranges of LRS and HRS as well as for the RESET dynamics and represents a strength of using physically motivated compact models. Not only material characteristics like Schottky barriers and ion hopping barriers but also fabrication characteristics such as layer thicknesses and series resistances can be analysed for their influence on the characteristics of a single cell and also on the circuit behavior.

IV. Results

A. Validation of the Model

To determine the validity of the variability model, it was fitted to experimental data of a $Pt/HfO_2/TiO_2/Pt$ device [33]. In the first step of the fitting process the deterministic version of the model has been fitted simultaneously to all the experiments shown in [32]. This was done manually by modifying the physical parameters such as the disc length or the filament radius in order to achieve the best combined agreement between the measurements and the deterministic simulations. During the fitting process special care is taken to keep the fitting parameters in physically reasonable ranges. The fitting of the variability model is achieved by tuning the parameters for device-to-device and cycle-to-cycle variability independently for each experiment and in addition by choosing the parameters according to the experimental results. Ideally, different simulations would use the same parameters. However the fit between simulation and experiment can be optimised by a more thorough tuning of the parameters to the different experiments. Still the first aim was to have a consistent set of parameters and to only tweak some to a specific experiment. It can be seen that this was achieved for the filament radius and disc length while the ranges for the maximum and minimum oxygen vacancy concentration are changed for the different experiments. These choices will be explained in the context of the corresponding experiments. The range of the variability parameters, i. e. the truncation for the device-to-device variability, are listed in Table II. In each cell entry, the first value specifies the lower truncation value, the second one specifies the mean of the underlying Gaussian distribution and the last value specifies the upper truncation value. The width of the underlying Gaussian distribution, which is expressed by the *RSD*, was chosen as 0.5 for all model parameters for the simulations shown in Fig. 4. The *RSD* enables to tune the device-to-device variability independently from the cycle-to-cycle variability. The *I-V* sweeps in Fig. 4 (b) and the endurance in Fig. 4 (h) showcase one device over multiple SET and RESET cycles. Therefore, device-to-device variability only plays a role in the initial choice of parameters while cycle-to-cycle variability is responsible for the observed variation. On the other hand the SET and RESET kinetics in Fig. 4 (d) and (f) are simulated by initialising a number of cells and subjecting each of them to the same stimulation for one (SET kinetic) or two (RESET kinetic) switching cycles. The variability in these simulations therefore mainly depends on the device-to-device variability and thereby directly on the choice of the *RSD*. The value was therefore motivated by these two simulations. The value of $N_{\mathrm{min,\,var}}$ was chosen in a different way. In [32] (Fig. 2 (d) therein), it was shown that the JART VCM v1 b model could reasonably well match the relationship between the range of initial states and the range of SET times. The initial experimental resistance values used were in the range of 100 kΩ to 500 kΩ. In order to represent the variation in the SET kinetics, the compact model needs a variation in the initial state between 50 kΩ and 340 kΩ. For range 1, the maximum $N_{\mathrm{min,\,var}}$ value was adapted to correspond to a resistance of around 50 kΩ and the minimum $N_{\mathrm{min,\,var}}$ value corresponds to a resistance of around 340 kΩ. Both resistance values are evaluated at a read voltage of −0.1 V.

TABLE II
Parameter Ranges for Figure 4

	$N_{\mathrm{min,\,var}}$ $[10^{23}\mathrm{m}^{-3}]$	$N_{\mathrm{max,\,var}}$ $[10^{26}\mathrm{m}^{-3}]$	r_{var} [nm]	l_{var} [nm]
(b)	4 / 8 / 16	0.10 / 0.15 / 0.2	40.5 / 45 / 49.5	0.36 / 0.4 / 0.44
(d)	4 / 8 / 16	18 / 20 / 22	40.5 / 45 / 49.5	0.36 / 0.4 / 0.44
(f)	4 / 8 / 16	18 / 20 / 22	40.5 / 45 / 49.5	0.36 / 0.4 / 0.44
(h)	4 / 8 / 25	0.39 / 0.4 / 0.41	40.5 / 45 / 49.5	0.36 / 0.4 / 0.44

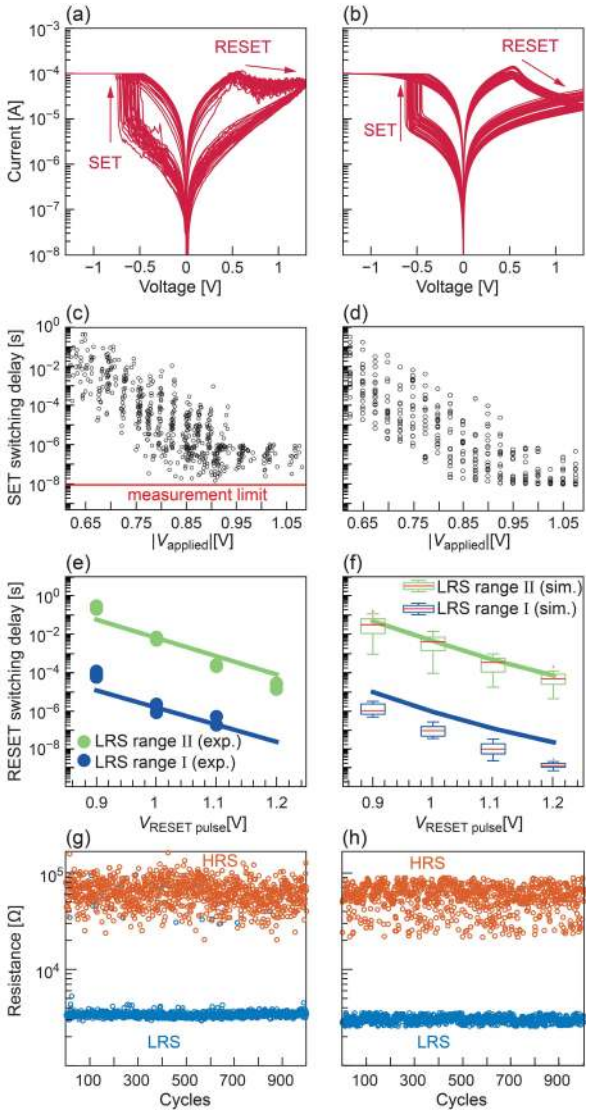

Fig. 4. Measured device characteristics alongside results obtained using the extended compact model. All simulations here were performed using the ranges from Table II. First row shows measured (a) and simulated (b) voltage sourced *I-V* sweeps. Second row shows the measured (c) and simulated (d) SET kinetics and the third row shows the measured (e) and simulated (f) RESET kinetics. The blue circles represent the experimental RESET kinetics starting from LRS between 1.85 kΩ and 2.22 kΩ (LRS range I (experimental)) while the green points represent the RESET kinetics starting from LRS between 1.52 kΩ and 1.67 kΩ (LRS range II (exp.)). The solid lines represent the RESET kinetics of the deterministic model using the parameters from Table I and the initial LRS as detailed in [32]. The blue box plots show the RESET kinetics for the variability model for an LRS range between 1.92 kΩ and 2.04 kΩ (LRS range I (simulated)) while the green box plots show the RESET kinetics of the variability model for an LRS range between 1.58 kΩ and 1.67 kΩ (LRS range II (sim.)). The fourth row ((g) and (h)) shows the measured and simulated endurance behavior over 1000 cycles. Generally, a good qualitative agreement between measurement and simulation is observed not only for device properties like LRS, HRS, SET and RESET voltages and switching kinetics but also for the statistical behavior of these properties.

Characteristic figures used to describe VCM cells include the *I-V* characteristic, which can be used to determine the range for LRS and HRS as well as the ranges of SET and RESET voltages. Experimental *I-V* sweeps for a current compliance of 100 μA are shown in Fig. 4 (a) whereas Fig. 4 (b) shows the simulation result. Both sweeps were performed using the same sweep rate (1 V/s) and the same stop voltages for SET and RESET of ±1.3 V. The SET voltages are between −0.45 V and −0.75 V, while the RESET voltages are between 0.45 V and 0.65 V. The RESET voltage has been defined as the inflection point when, starting from the LRS, the current first begins to decrease with an increasing voltage. For this simulation it is important to match SET and RESET voltage together with the HRS and LRS states. The possible ranges for HRS and LRS can be controlled by choosing the right $N_{\text{min, var}}$ and $N_{\text{max, var}}$ and by correctly controlling their variation over time. The LRS is in the range of 7 kΩ - 10 kΩ for most of the the cycles (80 %) and between 4 kΩ - 8 kΩ for the simulations. The HRS is between 50 kΩ and 500 kΩ in experiment and simulation. For voltages < 0.5 V the HRS matches very well between measurement and simulation. This match is highly important as reading voltages < 0.5 V are typically employed to prevent unintended switching. When measuring and simulating *I-V* sweeps, it is important to look at multiple switching events since the resistance state at the beginning of a voltage sweep or pulse has a strong influence on the switching voltage as well as on the state that is reached after the switching, e. g., [28], [32], [33], [45] and [46]. Here, 30 *I-V* sweeps were measured and 50 *I-V* sweeps were simulated. Comparing multiple sweeps reveals an increase in the SET voltage for higher HRS states, both, in the experimental data and the simulation data.

Fig. 4 (c) shows the measured SET delays, i. e. the time required to switch the cell into the LRS for a voltage pulse of a specific height between −0.6 V and −1.1 V. It is highly non-linear and shows a large variation of up to 5 orders of magnitude at each voltage, which presents a large challenge for circuit designers willing to employ VCM cells. The nonlinearity as well as the variability has been identified as a general property of filamentary valence change mechanism based systems such as STO [25], HfO_x [32] and TaO_x [42], which makes it one of the most relevant properties to consider in the modeling of these devices. Fig. 4 (d) shows the simulated SET kinetics. The variability model reproduces the non-linearity as well as the multiple orders of magnitude variation in the SET time. Here, it has to be noted that the resolution limit of the measurement setup was 10 ns. Therefore, switching events faster than 10 ns are not observed.

Fig. 4 (e) shows a more complex experiment, which could be reproduced using the proposed model, namely the RESET kinetics for different initial LRS. The experimental results are displayed with points. The solid lines show the results simulation results without the variability model and the box plots show the simulation results including the variability model. The reader is referred to [32] for a detailed description of the experimental procedure. Now, the same procedure is simulated using the compact model with variability. The colors are chosen consistent with the colors in [32]. The blue points represent RESET events from a LRS between 1.85 kΩ and 2.22 kΩ (LRS range I (experimental)), while the green points represent RESET events from an LRS between 1.52 kΩ and 1.67 kΩ (LRS range II (exp.)). These LRS were programmed using 10 μs pulses with −0.9 V (LRS range I) or −1.2 V (LRS range II). The solid lines represent the simulated deterministic

RESET kinetic from an initial state chosen from inside this range. Fig. 4 (f) shows the deterministic RESET kinetics (solid lines) as well as the RESET kinetics obtained by using the variability model and employing the same procedure as for the experimental RESET kinetic. For this simulation, the cells were first initialised in the HRS and then subsequently SET using the same pulse voltage and pulse duration as in the experiment, but with an additional current compliance to limit the switching. This current compliance was used to achieve the same range of resistances as in the experimental data. The resulting LRS ranges are 1.92 kΩ - 2.04 kΩ (LRS range I (simulated)) for the blue box plots and 1.58 kΩ - 1.67 kΩ (LRS range II (sim.)) for the green box plots. After the cells were programmed to these LRS ranges, a RESET voltage was applied with different voltages ranging from 0.9 V to 1.2 V until the cells switched. The resulting switching times, which were defined as the point in time at which the current drops below 300 μA, are comparable with the experimental data. It can be observed, that although the variability in the simulations is larger than in the experiments, still the experimental values are in most cases within the simulated RESET time range. The resulting variability observed in the simulations is significantly larger for the LRS between 1.52 kΩ up to 1.67 kΩ, and only slightly larger for the LRS between 1.85 kΩ and 2.22 kΩ. Only a few measurement points, however, are provided for each combination of range of LRS values and RESET voltage. It can therefore be envisioned that conducting more successful measurements will lead to larger variability as well. The RESET switching also shows a strong nonlinear dependence on the applied voltage as well as a strong dependence on the initial LRS and variability which was also previously reported by [42]. An increase in the RESET voltages for smaller LRS has also been shown by other groups like [45] and [47]. This is analogous to a longer delay for smaller initial LRS at a constant voltage.

Fig. 4 (g) shows the measured endurance characteristic over 1000 SET-RESET cycles. Initially, the resistance of the cell was read out at 0.1 V while in the HRS. Afterwards, a negative voltage pulse was applied to SET the cell. After the SET pulse, the cell was read again at 0.1 V, which was followed by an positive RESET pulse and a last read pulse at 0.1 V. The SET pulses were the same for measurement and simulation, in particular −0.8 V applied for 1 μs. The RESET pulses had a length of 5 μs and a pulse amplitude of 1.2 V in the experiment and 1.5 V in the simulation. The rise and fall times of the pulses were in the range of 1 ns up to 10 ns and can therefore be neglected compared to the pulse lengths. The RESET pulse voltage for the simulation was chosen slightly above the experimental voltage to better reach the range of HRS in the experiment as a smaller voltage would have lead to smaller HRS values and less RESET switching events. The LRS turns out to be very stable at around 3 kΩ, while the HRS varies between 30 kΩ and 100 kΩ with only a few outliers above 100 kΩ. Fig. 4 (h) shows the simulated endurance behavior over 1000 cycles. The LRS again is rather stable with a slightly larger variability than the experiment around 3 kΩ, while the HRS varies between 20 kΩ and 100 kΩ. The agreement of the resistance states between experiment and simulation was again achieved by choosing the right $N_{\mathrm{min,\,var}}$ and $N_{\mathrm{max,\,var}}$ ranges.

In conclusion, except for the LRS resistance there exists very little deviation between the ranges observed in experiment and simulation. The mismatch in the LRS range can be explained by the speed of the switching transition, which is faster in the model. The measured devices have a transition time between 1 ms and 100 ms at −0.6 V (see [32] supplementary Fig. 1), suggesting an even slower transition at lower voltages, while the transition time in the compact model is 50 ns at −0.5 V (see [32] Fig. 6). It should be noted that the compact model has been optimized to match the delay times rather than the transition times.

B. Applications of a Variability Model

The way the variability has been implemented, makes it easy to adapt it to fit different systems. In addition, it enables to test the robustness of circuits and architectures against different amounts of variability. Fig. 5 analyses the switching behavior of 100 cells initialised in the HRS that are subjected to pulses of different lengths (from 10 ns to 10 ms) at −0.8 V. This represents cells from the red rectangle in Fig. 5 (a). When considering this plot multiple effects can be recognized. The y-axis displays the cell number from 1 to 100 while the x-axis displays the resistance at the end of the pulse. The left part of the x-axis from 1.58 kΩ to 1.7 kΩ corresponds to the LRS, while the right part between 20 kΩ and 200 kΩ corresponds to the HRS. The different pulse lengths are encoded by colors. If the light green points (1 μs pulse lengths) are considered 35 cells are on the left side of the plot, while 65 cells are on the right side of the plot. This means that after the application of −0.8 V for 1 μs, 35 cells (equivalent to 35 % of the total number of cells) have switched to the LRS, while 65 cells (or 65 % of the cells) still remain in the HRS. By reading the plot in this way, the switching probabilities at a specific voltage for a specific pulse length can be determined. Another feature that can be extracted for each pulse length is the range of resistance states that are reached after the pulse. The 35 cells that have switched during the 1 μs pulse switched to LRS between 1.61 kΩ and 1.67 kΩ while the cells that did not switch had resistances between 30 kΩ and 180 kΩ. For pulse lengths of 10 ns or shorter and for pulse lengths of 10 ms or longer the switching is deterministic even with the variability. Between those two pulse lengths, the SET becomes stochastic since the cells only switch with a specific probability. The plot also shows that LRS and HRS are well separated. Only for the fastest pulses data points between 1.7 kΩ and 20 kΩ appear. The switching is binary for this voltage as the pulse lengths is longer than the transition time, which measures only the time to switch without the initial delay. As pointed out by Cueppers *et al.*, an analog switching accessing also the intermediate resistance range becomes only possible if the pulse length is smaller than the transition time at this specific pulse amplitude [32]. Another feature that can be observed is that longer pulses lead to smaller LRS states. Nardi *et al.* [48] reported that for a wide range of materials the resistance after the SET was reduced if the maximum

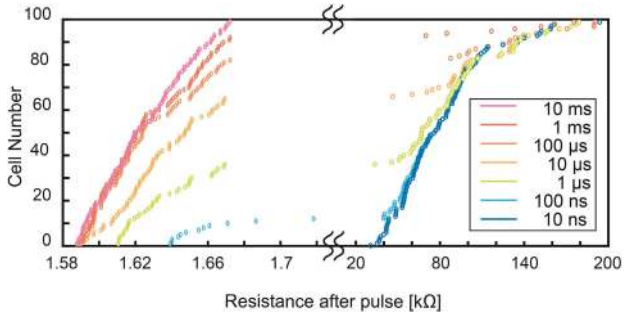

Fig. 5. Change in the switching behavior for different pulse durations for 100 cells is shown.

TABLE III
SUMMARY OF THE SWITCHING BEHAVIOR FOR DIFFERENT PULSE LENGTHS

Pulse Length	SET Probability	LRS Range [kΩ]	HRS Range [kΩ]
10 ns	0 %	-	30 - 190
100 ns	11 %	1.64 - 1.72	40 - 190
1 µs	35 %	1.61 - 1.67	30 - 180
10 µs	65 %	1.59 - 1.67	40 - 140
100 µs	82 %	1.59 - 1.67	80 - 190
1 ms	91 %	1.59 - 1.67	60 - 180
10 ms	100 %	1.59 - 1.67	-

current during the SET was higher. This higher maximum SET current corresponds to a stronger switching and is in that sense equivalent to a longer application of a voltage pulse. Also the application of a longer SET pulse increases the SET probability which is consistent with the reporting of other groups [49]. In the left part of Fig. 5 the range of LRS changes to lower resistances. While the range lies between 1.59 kΩ and 1.67 kΩ for a pulse length of 10 ms, it lies between 1.64 kΩ and 1.72 kΩ for a pulse length of 100 ns. Table III summarizes the switching probabilities, HRS and LRS ranges for the different pulse durations. The column SET probability shows, what percentage of the 100 cells has switched after the application of the voltage pulse. As can be observed from the resistances after the pulse the switching is binary which makes it easy to determine whether a cell has switched or not. The third and the fourth column both show the resistance at the end of the applied voltage pulse read at the pulse voltage (−0.8 V). For example at a pulse length of 100 ns the SET probability is 11 % which means that the LRS range for that row represents the range of device resistances of the 11 % of cells that have switched while the HRS range for that row represents the range of device resistances of the 89 % of cells that did not switch.

V. CONCLUSION

In this work, our JART VCM v1b model was extended by a physically motivated variability model based on the statistical variability of the switching filament. The resulting model can either be used in a deterministic fashion or with variability. The extended model exhibits device-to-device and cycle-to-cycle variability. It shows a good agreement with the experimentally observed variability considering transient behavior, non-linearity and statistics. The complete package of deterministic and variability model offers an new quality of consistency between different types of experiments and simulations. Additionally, since it has been show, that the deterministic model offers a description of filamentary based bipolar resistive switches with different transition metal oxides, such as STO [50], the variability model should be applicable to these systems as well. The proposed variability model can easily be tuned to the amount of observed variability. In addition, it enables the design of variability-aware circuits including VCM cells. A deeper understanding of the impacts of variability on a specific concept can also be used to communicate to device engineers the requirements which their devices have to fulfill for that concept to work in experiments.

Our compact model allows a reasonable reproduction of experiments that require both the quasi static properties to match as well as the dynamic properties. To the knowledge of the authors there exists no compact model that has been able to replicate a comparable number of measurements in combination with variability.

REFERENCES

[1] J. Borghetti, G. S. Snider, P. J. Kuekes, J. J. Yang, D. R. Stewart, and R. S. Williams, "'Memristive' switches enable 'stateful' logic operations via material implication," *Nature*, vol. 464, no. 7290, pp. 873–876, Apr. 2010.

[2] S. Kvatinsky, E. G. Friedman, A. Kolodny, and U. C. Weiser, "The desired memristor for circuit designers," *IEEE Circuits Syst. Mag.*, vol. 13, no. 2, pp. 17–22, 2nd Quart., 2013.

[3] J.-W. Jang, S. Park, Y.-H. Jeong, and H. Hwang, "ReRAM-based synaptic device for neuromorphic computing," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Jun. 2014, pp. 1054–1057.

[4] Y. Chen and C. Petti, "ReRAM technology evolution for storage class memory application," in *Proc. 46th Eur. Solid-State Device Res. Conf. (ESSDERC)*, Sep. 2016, pp. 432–435.

[5] A. Siemon *et al.*, "Realization of Boolean logic functionality using redox-based memristive devices," *Adv. Funct. Mater.*, vol. 25, no. 40, pp. 6414–6423, Oct. 2015.

[6] A. Siemon, D. Wouters, S. Hamdioui, and S. Menzel, "Memristive device modeling and circuit design exploration for computation-in-memory," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2019, pp. 1–5.

[7] K. M. Kim *et al.*, "Single-cell stateful logic using a dual-bit memristor," *Phys. Status Solidi (RRL)-Rapid Res. Lett.*, vol. 13, no. 3, Mar. 2019, Art. no. 1800629.

[8] N. Xu, K. J. Yoon, K. M. Kim, L. Fang, and C. S. Hwang, "Fully functional logic-in-memory operations based on a reconfigurable finite-state machine using a single memristor," *Adv. Electron. Mater.*, vol. 5, no. 1, Jan. 2019, Art. no. 1800775.

[9] L. Xie *et al.*, "Scouting logic: A novel memristor-based logic design for resistive computing," in *Proc. IEEE Comput. Soc. Annu. Symp. VLSI (ISVLSI)*, Jul. 2017, pp. 176–181.

[10] M. Hu *et al.*, "Memristor-based analog computation and neural network classification with a dot product engine," *Adv. Mater.*, vol. 30, no. 9, Mar. 2018, Art. no. 1705914.

[11] C. Li *et al.*, "Long short-term memory networks in memristor crossbar arrays," *Nature Mach. Intell.*, vol. 1, no. 1, pp. 49–57, Jan. 2019.

[12] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct. 1990.

[13] R. Waser, R. Dittmann, G. Staikov, and K. Szot, "Redox-based resistive switching memories–nanoionic mechanisms, prospects, and challenges," *Adv. Mater.*, vol. 21, nos. 25–26, pp. 2632–2663, Jul. 2009.

[14] H.-S. P. Wong *et al.*, "Metal–oxide RRAM," *Proc. IEEE*, vol. 100, no. 6, pp. 1951–1970, Jun. 2012.

[15] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80–83, May 2008.

[16] C. Funck and S. Menzel, "An atomistic view on the Schottky barrier lowering applied to $SrTiO_3$/Pt contacts," *AIP Adv.*, vol. 9, no. 4, Apr. 2019, Art. no. 045116.

[17] C. Funck *et al.*, "A theoretical and experimental view on the temperature dependence of the electronic conduction through a Schottky barrier in a resistively switching $SrTiO_3$-based memory cell," *Adv. Electron. Mater.*, vol. 4, no. 7, Jul. 2018, Art. no. 1800062.
[18] A. Marchewka, R. Waser, and S. Menzel, "A 2D axisymmetric dynamic drift-diffusion model for numerical simulation of resistive switching phenomena in metal oxides," in *Proc. Int. Conf. Simulation Semiconductor Processes Devices (SISPAD)*, Sep. 2016, pp. 145–148.
[19] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Understanding switching variability and random telegraph noise in resistive RAM," in *IEDM Tech. Dig.*, Dec. 2013, pp. 1–4.
[20] Q. Xia and J. J. Yang, "Memristive crossbar arrays for brain-inspired computing," *Nature Mater.*, vol. 18, no. 4, pp. 309–323, Apr. 2019.
[21] C. Zambelli *et al.*, "Statistical analysis of resistive switching characteristics in ReRAM test arrays," in *Proc. Int. Conf. Microelectron. Test Struct. (ICMTS)*, Mar. 2014, pp. 27–31.
[22] H. Aziza, P. Canet, J. Postel-Pellerin, M. Moreau, J.-M. Portal, and M. Bocquet, "ReRAM ON/OFF resistance ratio degradation due to line resistance combined with device variability in 28 nm FDSOI technology," in *Proc. Joint Int. EUROSOI Workshop Int. Conf. Ultimate Integr. Silicon (EUROSOI-ULIS)*, Apr. 2017, pp. 35–38.
[23] S. Yu, X. Guan, and H.-S.-P. Wong, "On the stochastic nature of resistive switching in metal oxide RRAM: Physical modeling, Monte Carlo simulation, and experimental characterization," in *IEDM Tech. Dig.*, Dec. 2011, pp. 1–4.
[24] A. Chen, "Failure analysis of resistive switching devices," in *Proc. IEEE Int. Rel. Phys. Symp.*, May 2010, pp. 84–88.
[25] K. Fleck, C. La Torre, N. Aslam, S. Hoffmann-Eifert, U. Böttger, and S. Menzel, "Uniting gradual and abrupt set processes in resistive switching oxides," *Phys. Rev. A, Gen. Phys.*, vol. 6, no. 6, Dec. 2016.
[26] A. Chen and M.-R. Lin, "Variability of resistive switching memories and its impact on crossbar array performance," in *Proc. Int. Rel. Phys. Symp.*, Apr. 2011, pp. 1–4.
[27] C. Baeumer *et al.*, "Subfilamentary networks cause cycle-to-cycle variability in memristive devices," *ACS Nano*, vol. 11, no. 7, pp. 6921–6929, Jul. 2017.
[28] Z. Jiang *et al.*, "A compact model for metal–oxide resistive random access memory with experiment verification," *IEEE Trans. Electron Devices*, vol. 63, no. 5, pp. 1884–1892, May 2016.
[29] M. Bocquet *et al.*, "Compact modeling solutions for oxide-based resistive switching memories (OxRAAM)," *J. Low Power Electron. Appl.*, vol. 4, no. 1, pp. 1–4, 2014.
[30] N. Wald and S. Kvatinsky, "Understanding the influence of device, circuit and environmental variations on real processing in memristive memory using memristor aided logic," *Microelectron. J.*, vol. 86, pp. 22–33, Apr. 2019.
[31] F. M. Puglisi, L. Larcher, A. Padovani, and P. Pavan, "Bipolar resistive RAM based on HfO_2: Physics, compact modeling, and variability control," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 2, pp. 171–184, Jun. 2016.
[32] F. Cüppers *et al.*, "Exploiting the switching dynamics of HfO_2-based ReRAM devices for reliable analog memristive behavior," *APL Mater.*, vol. 7, no. 9, Sep. 2019, Art. no. 091105.
[33] A. Hardtdegen, C. La Torre, F. Cuppers, S. Menzel, R. Waser, and S. Hoffmann-Eifert, "Improved switching stability and the effect of an internal series resistor in HfO_2/TiO_x bilayer ReRAM cells," *IEEE Trans. Electron Devices*, vol. 65, no. 8, pp. 3229–3236, Aug. 2018.
[34] M. Lanza *et al.*, "Recommended methods to study resistive switching devices," *Adv. Electron. Mater.*, vol. 5, no. 1, 2019, Art. no. 1800143.
[35] I. Boybat *et al.*, "Multi-ReRAM synapses for artificial neural network training," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2019, pp. 1–5.
[36] A. Schonhals, J. Mohr, D. J. Wouters, R. Waser, and S. Menzel, "3-bit resistive RAM write-read scheme based on complementary switching mechanism," *IEEE Electron Device Lett.*, vol. 38, no. 4, pp. 449–452, Apr. 2017.
[37] D. Apalkov, B. Dieny, and J. M. Slaughter, "Magnetoresistive random access memory," *Proc. IEEE*, vol. 104, no. 10, pp. 1796–1830, Oct. 2016.
[38] F. M. Bayat, M. Prezioso, B. Chakrabarti, H. Nili, I. Kataeva, and D. Strukov, "Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits," *Nature Commun.*, vol. 9, no. 1, Dec. 2018, Art. no. 2331.
[39] S. Menzel and C. Bengel. *JART—Juelich Aachen Resistive Switching Tools*. Accessed: Apr. 2020. [Online]. Available: http://www.emrl.de/JART.html
[40] A. R. Genreith-Schriever and R. A. De Souza, "Field-enhanced ion transport in solids: Reexamination with molecular dynamics simulations," *Phys. Rev. B, Condens. Matter*, vol. 94, no. 22, Dec. 2016, Art. no. 224304.
[41] M. von Witzleben *et al.*, "Thermal effects on the I-V characteristics of filamentary VCM based ReRAM-cells using a nanometer-sized heater," in *Proc. 17th Non-Volatile Memory Technol. Symp. (NVMTS)*, Aug. 2017, pp. 1–5.
[42] J. P. Strachan *et al.*, "State dynamics and modeling of tantalum oxide memristors," *IEEE Trans. Electron Devices*, vol. 60, no. 7, pp. 2194–2202, Jul. 2013.
[43] T. Diokh *et al.*, "Investigation of the impact of the oxide thickness and RESET conditions on disturb in HfO_2-RRAM integrated in a 65 nm CMOS technology," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Apr. 2013, pp. 1–4.
[44] F. Spitzer, "Principles of random walk," in *Graduate Texts in Mathematics*, 2nd ed. New York, NY, USA: Springer, 2001.
[45] H. Y. Lee *et al.*, "Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM," in *IEDM Tech. Dig.*, Dec. 2008, pp. 1–4.
[46] Y. Nishi, S. Menzel, K. Fleck, U. Bottger, and R. Waser, "Origin of the SET kinetics of the resistive switching in tantalum oxide thin films," *IEEE Electron Device Lett.*, vol. 35, no. 2, pp. 259–261, Feb. 2014.
[47] J. Kwon *et al.*, "Transient thermometry and high-resolution transmission electron microscopy analysis of filamentary resistive switches," *ACS Appl. Mater. Interfaces*, vol. 8, no. 31, pp. 20176–20184, Aug. 2016.
[48] F. Nardi, S. Larentis, S. Balatti, D. C. Gilmer, and D. Ielmini, "Resistive switching by voltage-driven ion migration in bipolar RRAM—Part I: Experimental study," *IEEE Trans. Electron Devices*, vol. 59, no. 9, pp. 2461–2467, Sep. 2012.
[49] K. Moon *et al.*, "RRAM-based synapse devices for neuromorphic systems," *Faraday Discuss.*, vol. 213, pp. 421–451, Feb. 2019.
[50] C. La Torre, "Physics-based compact modeling of valence-change-based resistive switching devices," Ph.D. dissertation, RWTH Aachen, Aachen, Germany, 2019.

Christopher Bengel received the B.Sc. and M.Sc. degrees in electrical engineering from RWTH Aachen University, Aachen, Germany, in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree with the IWE 2 RWTH Aachen, Aachen, with a focus on modeling resistive switching devices for computation-in-memory.

Anne Siemon was born in Neuss, Germany, in 1986. She received the Diploma degree in electrical engineering from RWTH Aachen University, Aachen, Germany, in 2013. She is currently pursuing the Ph.D. degree with the IWE 2 RWTH Aachen, Aachen. Her research interests include circuit modeling of resistive switches and advanced logic concepts based on resistive switches.

Felix Cüppers received the B.Sc. and M.Sc. degrees in materials science from RWTH Aachen University, Aachen, Germany, in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with the Peter-Grünberg-Institut 10, Forschungszentrum Jülich, Jülich, Germany, with focus on resistive switching devices in neuromorphic applications.

Susanne Hoffmann-Eifert received the Diploma degree in physics and the Ph.D. degree in electrical engineering from RWTH Aachen University, Aachen, Germany, in 1998. She is currently a Senior Researcher with the Peter-Grünberg-Institut, Forschungszentrum Jülich GmbH, Germany. Her research interests include thin film growth and device physics with current focus on resistive switching devices for application in neuromorphic computing.

Alexander Hardtdegen received the B.Sc. and M.Sc. degrees in physics from RWTH Aachen University, Aachen, Germany, in 2013 and 2015, respectively. He is currently pursuing the Ph.D. degree with Forschungszentrum Jülich, Jülich, Germany, with the focus on HfO_2-based resistive switching memories.

Moritz von Witzleben was born in Trier, Germany, in 1990. He received the M.Sc. degree in physics from RWTH Aachen University in 2015, where he is currently pursuing the Ph.D. degree in electronical engineering. His research focuses on the characterization of the switching kinetics of ReRAM devices at sub-nanosecond time scales.

Lena Hellmich received the M.Sc. degree in materials science from RWTH Aachen University in 2018, where she is currently pursuing the Ph.D. degree with the Institute of Semiconductor Electronics. Her current field of research contains reconfigurable transistors with dopant free contacts and 2D materials.

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 a Professor and the 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 (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 a Senior Researcher with the Peter-Grünberg-Institut 7, Forschungszentrum Jülich GmbH, Jülich, Germany, where he is also leading the Simulation Group.