Investigation of the Large Variability of HfO$_2$-Based Resistive Random Access Memory Devices with a Small Current Compliance by a Kinetic Monte Carlo Model

Ching-Jung Chen,* Konstantin Z. Rushchanskii,* and Christoph Jungemann

While scaling down resistive random access memory devices can bring many benefits, it also introduces uncertainties during operation. One example is the wide distribution of the resistances in the low-resistance state (LRS) when the device is connected by a small current compliance (I_c) in the microampere range. During the operation under such small I_c, it is believed that the stochastic migration of oxygen vacancies in the oxide layer plays an important role for the variability. To this regard, the model where oxygen vacancies are treated as point defects and the kinetic Monte Carlo method is applied for the stochastic migration of vacancies is extended. The relation between the macroscopic observations in measurements and the microscopic vacancy distribution are discussed. It turns out that three representative configurations are sufficient to describe the vacancy distribution in the LRS. The large spread of the resistance seen in the cycle-to-cycle statistics is then due to the change from one of the configuration to the other. The origin for the change between configurations is discussed in terms of the anisotropic zero-field energy barrier of the diffusion of vacancies.

1. Introduction

Among all developing nonvolatile memory technologies, the resistive random access memory (ReRAM) has attracted much attention due to features such as fast switching speed, complementary metal–oxide–semiconductor compatibility and good scalability.[1–3] Moreover, the potential applications of neural network computation[4,5] and hardware security[6,7] make it a promising candidate for the next-generation information technologies.

The operation of ReRAM relies on switching between a high-resistance state (HRS) and a low-resistance state (LRS) by reversing the polarity of the applied bias. A variety of oxides, such as HfO$_2$, ZrO$_2$, TiO$_2$, and Ta$_2$O$_5$, have been shown to be suitable for developing filamentary-type valence change memory (VCM), which is a subgroup of the ReRAM technologies.[3] It is widely accepted that a conductive filament (CF), which consists of oxygen vacancies (V_O), is responsible for the LRS while the dissolution or rupture of the CF leads to the HRS.[8–12] To initiate a virgin ReRAM cell, a large voltage is applied, leading to dielectric breakdown in the insulating layer. This step is termed as electroforming or FORMING, and the LRS is reached when the CF connects both electrodes. The ReRAM cell switches to the HRS in the RESET process, when the CF is partially ruptured by an applied bias of opposite polarity. The following LRS is achieved by reversing the polarity again and the gap in the CF is closed during the SET process. During the electroforming and SET process, a current compliance (I_c) is usually attached to avoid thermal runaway.

Despite the general consensus that the switching is based on a CF of oxygen vacancies, the microscopic details are still not fully understood, and various modeling approaches exist. For example, the atomic resolution is lost for models based on continuum approaches,[13,14] while the dynamics of vacancies is not discussed for models treating vacancies as point defects.[15–19] In contrast, the cycle-to-cycle (C2C) statistics of the resistances in the LRS and the HRS have been observed in measurements to follow different power laws[1,20–23] for the HfO$_2$-based devices. That is, the relative standard deviation is proportional to the square root of the median value (μ_R) in the HRS while the exponent varies from 0.5 to 1 in the LRS for a large I_c. In addition, the median resistance over cycles obtained at a small voltage (V_{meas}) has been observed to be proportional to the inverse of the I_c and $V_{trans} = I_c \times \mu_R = 0.4\,V$ is reported.[21] However, both the constant, V_{trans}, and the power law relation in C2C statistics do not hold in the small I_c regime. Instead, an increased $I_c \times \mu_R$ and a leveling off of the relative standard deviation are observed. In contrast to existing simulation works, measurements have

C.-J. Chen, C. Jungemann
Chair of Electromagnetic Theory
RWTH Aachen University
52056 Aachen, Germany
E-mail: cc@ithe.rwth-aachen.de

K. Z. Rushchanskii
Peter Grünberg Institut (PGI-1) and Institute for Advanced Simulation (IAS-1)
Forschungszentrum Jülich and JARA-FIT
52425 Jülich, Germany
E-mail: k.rushchanskii@fz-juelich.de

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/pssa.202300403.

already shown a large C2C variability. To understand the origin of the large C2C variability in the small I_c regime, the dynamics of vacancies treated as point defects should be resolved. On the contrary, recent *ab initio* calculations have shown that local structures such as the CF[24,25] and the grain boundary (GB)[26] would lead to an inhomogeneous and anisotropic vacancy mobility, which are not captured by existing models. Therefore, the structure-induced vacancy mobility in terms of different zero-field energy barriers is considered to extend our physics-based model.

In the framework of our model, the charge transport and the dynamics of vacancy distribution are treated separately. The charge transport model is solved together with the Poisson equation for the quasi-static potential and the Fourier heat equation under stationary conditions for a given vacancy distribution. Once the self-consistent solution is obtained, the potential and temperature distribution inside the device are fixed and the vacancy distribution is updated by the kinetic Monte Carlo (KMC) method.[27–29] That is, the generation, recombination, and diffusion of vacancies are regarded as Poisson processes with rates given by the Arrhenius equation.[30] Only the randomly selected process among all processes can take place and the time step for the evolution is determined in relation to the rate of all processes. To compare with measurements,[21] the simulated device has the structure of a 5 nm thick HfO₂ layer sandwiched by a TiN layer and a hafnium (Hf) layer as the top and bottom electrodes, respectively. While the thickness is the critical dimensionality for filamentary-type VCM devices, the cross section is less relevant, and thus a cross section of $5 \times 5\,nm^2$ is chosen to reduce the computational effort. Furthermore, a crystalline instead of the amorphous structure of the HfO₂ layer is assumed to incorporate parameters obtained from density-functional theory (DFT) calculations.

In this work, different levels of the LRS resistance and their corresponding vacancy distribution are identified. In addition, a possible scenario that drives the change of the distribution, and thus the large C2C variability is proposed. The discussion is centered around the simulation of the specific value, namely, $I_c = 2\,\mu A$.

2. Modeling

In contrast to refs.,[31,32] recent ab initio studies consider the local V_O distribution and show that the neutral charge state is energetically favored under certain conditions.[33,34] In short,

the formation energies of the neutral and the +1 charge states are lower than the +2 state in a vacancy-rich region, while +2 is energetically favored in a vacancy-poor region. We model V_O to be neutrally charged if it is attached to an adjacent V_O in the charge transport direction, or it is connected to a V_O chain with at least two more vacancies shown in **Figure 1**a.

From the formation energy perspective, the most and the second most favored charge states of the V_O differ only by one electron in the most relevant part of the Fermi-level interval for both vacancy-rich and vacancy-poor regions.[33] That is, the +0 and +1 charge states are energetically preferred over the +2 charge state in the vacancy-rich region. Therefore, we assume the V_O can only be in one of the two lower charge states. Similarly, the V_O in the vacancy-poor region is assumed to be either in the +2 or +1 charge state. In this work, we refer to the +0 (+2) charge state as the most favored charge state of a V_O, and it will switch between +0 and +1 (+2 and +1) due to an incoming or outgoing electron in the dynamical process.

2.1. Charge Transport

The trap-assisted-tunneling (TAT) mechanism is used to model the charge transport process of ReRAM devices driven by an applied voltage.[16–19] It considers two types of charge transport schemes, namely tunneling between vacancies and electrodes, and hopping among vacancies. The quasi-stationary condition applies and the corresponding master equation takes the form

$$(1 - p_i) \sum_{M=A,C} P_{Mi} - p_i \sum_{M=A,C} P_{iM}$$
$$+ (1 - p_i) \sum_{j \neq i} p_j h_{ji} - p_i \sum_{j \neq i} (1 - p_j) h_{ij} = 0 \quad (1)$$

where p_i and p_j denote the electron occupation probabilities of the ith and jth vacancies, respectively. P_{Mi} and P_{iM} are the tunneling probabilities from the electrodes (anode: $M = A$, cathode: $M = C$) to the vacancy site and vice versa given by the Wentzel–Kramers–Brillouin approximation and the Fermi–Dirac distribution.[16,35–37] h_{ij} is the Miller–Abrahams hopping rate[38] from the ith vacancy site to the jth vacancy, which takes the form
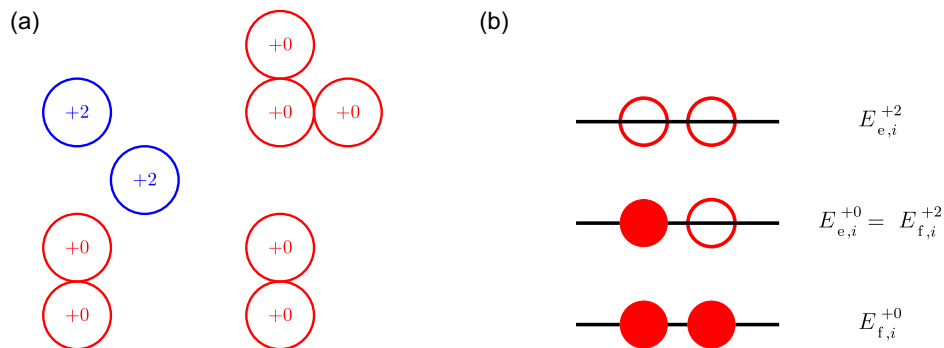

Figure 1. a) The scheme for modeling intrinsic charge states and b) the energy levels of each charge states during the dynamical process.

2300403 (2 of 10)

$$h_{ij} = \begin{cases} \nu_e \exp\left(-\frac{2d_{ij}}{a_0}\gamma\right), & \Delta E_{f,ij} < 0 \\ \nu_e \exp\left(-\frac{2d_{ij}}{a_0}\gamma\right)\exp\left(-\frac{\Delta E_{f,ij}}{k_B T_i}\right), & \Delta E_{f,ij} \geq 0 \end{cases} \quad (2)$$

k_B is the Boltzmann constant and the T_i is the temperature at the ith vacancy site. $\Delta E_{f,ij} = (E_{f,j}^Q - e\phi_j) - (E_{f,i}^Q - e\phi_i)$ is the total energy difference where ϕ_i is the potential and $E_{f,i}^Q$ ($E_{e,i}^Q$) is the filled (empty) energy level of the ith vacancy. Both energy levels depending on the favored charge state are denoted by the superscript, Q. For an intrinsically $+0$ charge state V_O, the empty energy level $E_{e,i}^{+0}$ refers to the same energy as a filled energy level of a $+2$ charged V_O, $E_{f,i}^{+2}$, as illustrated in Figure 1b. The value of $E_{e,i}^{+2}$ and $E_{f,i}^{+2}$ are adapted from ref. [16] while $E_{f,i}^{+0}$ is chosen close to the DFT calculation.[31] ν_e is the attempt frequency and d_{ij} is the distance between the initial and final positions. The localization length a_0 is related to the overlap of electron wave functions at the initial and final site and the solution of the wave function exists under approximation.[39] Here, we keep the analytical form

$$a_0 = \frac{\hbar}{\sqrt{2m^*(E_c - E_{f,i}^Q)}} \quad (3)$$

where m^* is the effective electron mass different from the electron rest mass m_0, $\hbar$ is the reduced Planck constant, and $E_c - E_{f,i}^Q$ refers to the energy separation of the trap level and the conduction band while an extra factor $\gamma = 3/4$ is introduced to match the value used in ref. [16]

The electron current through an ReRAM cell is obtained by summing the tunneling contributions from all vacancies to the anode

$$I^e = e\sum_i [p_i P_{iA} - (1 - p_i)P_{Ai}] \quad (4)$$

2.2. Poisson Equations

The Poisson equation with the space charge at the vacancy sites is used to calculate the quasi-stationary potential. At the top and bottom electrodes, Dirichlet boundary conditions are used fixing the potential to the values 0 and V_{cell}, respectively, while homogeneous Neumann boundary conditions are applied at the other surfaces. In the FORMING and SET processes, the I_c might reduce the voltage across the cell (V_{cell}) from the applied voltage (V_{app}) to avoid an overshoot of the current.

Lastly, the local temperature is obtained by solving the Fourier heat equation under quasi-stationary condition

$$k_{th}\nabla^2 T + q = 0 \quad (5)$$

where k_{th} is the thermal conductivity and q is the local heat dissipation given by Equation (6). The quasi-stationary approximation is made due to the typical thermal response time being less than a nanosecond.[13,40] Similar to the Poisson equation, Dirichlet boundary conditions with $T = 300$ K at both electrodes and Neumann boundary conditions are applied at the other surfaces.

Since the electron transport processes described by the TAT are hopping and tunneling, an electron is considered to absorb or dissipate energy only at the initial or final vacancy site. The power density dissipated at the ith V_O site takes the form

$$q_i = \begin{cases} \frac{1}{e\Omega_i} \sum_{M=A,C} I_{Mi}^e (E_{e,i}^Q - E_{f,i}^Q), & \Delta E_{f,ij} < 0 \\ \frac{1}{e\Omega_i} \sum_{M=A,C} I_{Mi}^e (E_{e,i}^Q - E_{f,i}^Q) \\ \quad + \frac{1}{e\Omega_i}\sum_{j\neq i}(I_{ji}^e - I_{ij}^e)\Delta E_{f,ij}, & \Delta E_{f,ij} \geq 0 \end{cases} \quad (6)$$

where Ω_i is the volume of the grid cell; I_{Mi}^e is the electron currents flowing to the ith vacancy from an electrode; and $I_{ji}^e - I_{ij}^e$ is the net electron current from the jth vacancy. Since electrons involved in the tunneling process are assumed to have the same energy, $E_{e,i}^Q$, and will fall to a lower energy level, $E_{f,i}^Q$, at the vacancy site, the term $I_{Mi}^e(E_{e,i}^Q - E_{f,i}^Q)$ is the power density due to the energy level shift from an empty to a filled state, while the energy dissipation at the electrode is neglected. Similarly, an electron hopping from a vacancy site with the higher energy level will dissipate the energy difference, $\Delta E_{f,ij}$, at its final site. On the contrary, the electron hopping from a lower energy level would absorb the same amount of energy at the initial site to reach the energy level of its final site.

The emergence of the CF introduces oxygen-poor regions, where the thermal conductivity is expected to be close to that of the metallic hafnium, $k_{th,Hf} = 23$ W m^{-1}K^{-1}. The thermal conductivity k_{th} is assumed to be linear with respect to the vacancy number density n[13] with

$$k_{th}(n) = \begin{cases} 0.5, & n \leq n_0 \\ 0.5 + \lambda \cdot (n - n_0), & n_0 < n < n_1 \\ 23, & \text{otherwise} \end{cases} \quad (7)$$

where the vacancy number is averaged over a 1 nm cube, $n_0 = 1.0$ nm^{-3} corresponds to one vacancy inside the cube and $n_1 = 29$ nm^{-3} corresponds to the volume such that $k_{th}(n) \approx 23$ W m^{-1} K^{-1}. The thermal conductivity at a low vacancy density $k_{th,ox} = 0.5$ W m^{-1}K^{-1} is chosen from measurements.[41]

After a self-consistent solution of the occupation probability, potential and temperature is obtained, the KMC algorithm is applied for determining the evolution of the vacancy distribution.

2.3. Vacancy Dynamics

The rates of generation and recombination of vacancies at an electrode interface[36,37,42] are given by

$$\begin{aligned} G &= G_0 \exp\left(-\frac{E_G - bF_{ext}}{k_B T}\right) \\ R &= R_0 \exp\left(-\frac{E_R}{k_B T}\right) \end{aligned} \quad (8)$$

where b is the bond polarization adopted from measurements and F_{ext} is the vertical component of the electric field.[37,43,44]

G_0 and R_0 are characteristic vibration frequencies of the generation and recombination processes, respectively. E_G and E_R are the corresponding zero-field activation energies. The generation barrier is lower at a GB.[42,45] We do not consider the generation of vacancies by Frenkel pairs in the oxide,[16,46] because Schie et al.[47] argue that the high energy barrier of the Frenkel pair generation[34,48] together with the weak modification by the electric field[49] make it an unfeasible mechanism for the observed resistive switching.[50–53] In contrast, the metallic electrodes are shown to modify the formation energy of vacancies.[34,54,55] Consequently, vacancies can be generated or annihilated only at the low-work-function electrode. In the case of an intersection of a GB and the interface, the generation barrier is further lowered. The energy barrier is then chosen between $E_{G,0}$ and $E_{G,GB}$. $E_{G,0}$ is chosen to be close to the bond breaking energy obtained from the melting temperature of bulk c-HfO$_2$. $E_{G,GB}$ is adapted from time-dependent dielectric breakdown measurements. The GB is assumed to be composed of several column-shaped regions shown in **Figure 2**, with each column extending from the bottom to the top electrode
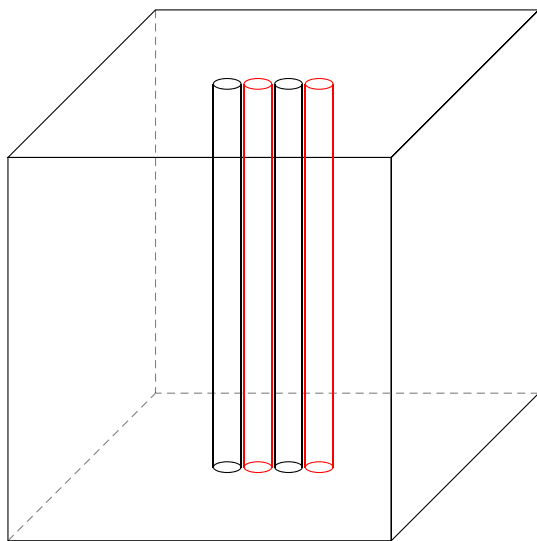

Figure 2. The modulation of energy barriers is visualized by colors where red columns refer to regions with a larger energy barrier change.

interface.[42] Furthermore, physical properties, i.e., the generation barrier and the anisotropic diffusion barrier, are assumed to vary a little. For the column where the generation barrier is modulated stronger, the change of the diffusion barrier is also larger.

The hopping rate inside the oxide takes the form

$$D = D_0 \exp\left(-\frac{E_D - Q_{eff} F_{loc} a/2 - k_B \Delta T/2}{k_B T}\right) \quad (9)$$

where D_0 is the characteristic vibration frequency, E_D is the zero-field activation energy of diffusion, a is the migration distance, and Q_{eff} is the effective charge of the vacancy. F_{loc} is the local electric field calculated from the macroscopic electric field and the self-potential is excluded.[36] While some models include a factor of $\frac{\varepsilon/\varepsilon_0+2}{3}$ for the local electric field, Schie et al.[47] argue that this factor should not be applied in the case of the discussed device. The temperature gradient $\Delta T = T(r') - T(r_0)$ is the driving force coming from the thermal effect, which indicates the preferred migration toward higher-temperature regions.[56] Note that the factor of 1/2 refers to the assumption that the energy barrier peak is located in the middle of its moving path.

With the introduction of different intrinsic charge states, the dependence of the zero-field energy barrier with respect to the charge state is included in the model. In addition, the local structure also plays a role in modulating the energy barrier. More specifically, recent DFT studies have shown an anisotropic energy barrier for vacancies near the filament[24,25] and a GB.[26] The zero-field activation energy has to be distinguished in the parallel and normal direction with respect to the filament direction. For c-HfO$_2$,[24] it is shown that the energy barrier of the V_O leaving the consecutive V_O channel in the normal direction is much larger compared to the one leaving in the parallel direction, see **Figure 3**a. Furthermore, a V_O inside the CF is very mobile, see Figure 3b. The same trends of anisotropic modulation are shown by Duncan et al.[25] for m-HfO$_2$. An anisotropy of the energy barrier is also shown for vacancies near the GB and the trend is similar to the CF effect, i.e., a reduced energy barrier parallel to the GB direction inside the twin GB, and a decreased (increased) energy barrier for moving toward (outward) the GB in the normal direction.[26] However, the lowering in the normal direction is dominant only within about 2 Å, roughly to be the nearest distance of two oxygen atoms. Therefore, only the local
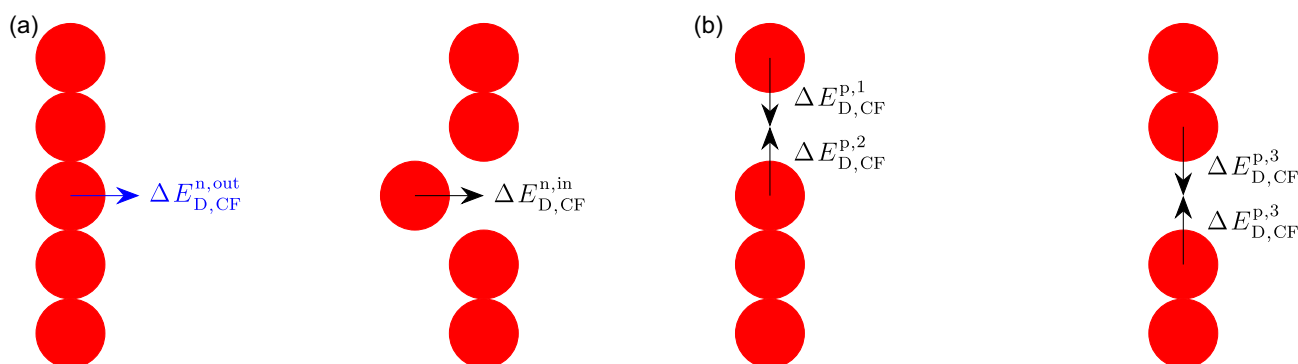

Figure 3. The modulation of the diffusion barrier due to the conductive filament (CF) effect in different directions. a) The barrier is increased (decreased) for the vacancy leaving (entering) the CF in the normal direction. b) The energy barrier is getting lower for a V_O being deeper inside the CF with the lowest energy barrier being $E_{D,min} = 0.51$ eV. This value sets the minimal zero-field energy barrier for migration.

structure at the initial and the final sites are considered while other local structures in a larger distance are neglected. The anisotropic zero-field energy barrier takes the form

$$E_{\mathrm{D}} = E_{\mathrm{D},0}^Q + \Delta E_{\mathrm{D,str}}^{\mathrm{d}} \tag{10}$$

where $E_{\mathrm{D},0}^Q$ refers to different charge states and $\Delta E_{\mathrm{D,str}}^{\mathrm{d}}$ is the anisotropic energy barrier modification in the normal or parallel direction arising from different local structures, i.e., CF or GB. Note that the modification is assumed to be identical for different charge states. Due to the preference of vacancy aggregation at the GB, the condition of a V_{O} being inside a GB and a CF does not exclude each other and makes the $\Delta E_{\mathrm{D,str}}^{\mathrm{d}}$ ambiguous. In such cases, the one with the larger change of energy barrier is chosen since the changes of zero-field energy barrier result from the local lattice distortion. On the contrary, superposition of individual modification is employed based on the assumption that the lattice distortion in the middle of the migration path is induced by both local structures at the initial and final sites. Important parameters are summarized in **Table 1**.

Table 1. Parameters used in this work.

Symbol	Value [eV]	Symbol	Value
$E_{\mathrm{G},0}$	5.7	$E_{\mathrm{D,min}}$	0.51 eV[24]
$E_{\mathrm{G,GB}}$	4.1	D_0, G_0, R_0	70 THz[42]
$E_{\mathrm{R},0}$	1.1	ν_e	70 THz
$E_{\mathrm{D},0}^{+2}$	0.7[42]	b	90 eÅ
$E_{\mathrm{D},0}^{+0}$	1.1	λ	$8 \times 10^{26}\,\mathrm{W\,K^{-1}\,m^{-4}}$
$\Delta E_{\mathrm{D,GB}}^{\mathrm{n,in}}$	−0.2	$k_{\mathrm{th,ox}}$	$0.5\,\mathrm{W\,K^{-1}\,m^{-1}}$[41]
$\Delta E_{\mathrm{D,GB}}^{\mathrm{n,out}}$	0.2	ε	$29\varepsilon_0$[58,59]
$\Delta E_{\mathrm{D,CF}}^{\mathrm{n,in}}$	−0.3	m^*	$0.1\,m_0$[16]
$\Delta E_{\mathrm{D,CF}}^{\mathrm{n,out}}$	0.3	$E_c - E_{e,i}^{+2}$	1.83 eV[16]
$\Delta E_{\mathrm{D,CF}}^{\mathrm{p,1}}$	−0.4	$E_c - E_{f,i}^{+2}$	1.97 eV[16]
$\Delta E_{\mathrm{D,CF}}^{\mathrm{p,2}}$	−0.43	$E_c - E_{f,i}^{0}$	2.11 eV
$\Delta E_{\mathrm{D,CF}}^{\mathrm{p,3}}$	−0.59	a	2.5 Å

Note that the electric field has a different impact on the diffusion, generation, and recombination processes. We interpret the physical scheme behind the generation and recombination being ion transfer processes and vacancies are treated as charged ions in the diffusion process. In terms of ion diffusion, the peak of the energy barrier is assumed at the middle of the migration path in most cases. In contrast, the energy barrier modified by the electric field $-\alpha b F_{\mathrm{ext}}$ and $(1-\alpha)b F_{\mathrm{ext}}$ for the generation and recombination processes are adopted from empirical relations. To match the abrupt increase of the current in the SET process and the much more gradual decrease of the current in the RESET process seen in experiments,[22,51] the asymmetry factor α is chosen to be 1.

3. Simulation Results

Preliminary to the switching of different resistance states, a cycle including the FORMING process followed by a RESET process is simulated. It starts with only one V_{O} inside the GB, which refers to a fresh device. Results do not depend on initial vacancy numbers provided that a filament does not preexist. The maximum V_{app} of 2.5, 1.5, and −1.5 V are used for the FORMING, SET, and RESET processes, respectively, and the same I_{c} is applied in the FORMING and SET process, which are aligned with the setup in measurements.[21]

Figure 4 shows the simulation result of a device for a very low I_{c} of $I_{\mathrm{c}} = 2\,\mu A$ for 100 switching cycles, where the FORMING process is excluded. The LRS resistance and HRS resistance are obtained by V_{meas}/I, where $V_{\mathrm{meas}} = V_{\mathrm{cell}} = -0.1\,V$ at the early and late stage of the RESET process, respectively. The probability plot of the cumulative distribution function (CDF) with respect to the log scale of the resistance is shown in Figure 4a. Steep straight lines are observed in the large I_{c} regime in ref. [21] and the resistances follow the log-normal distribution. However, the measured CDF is more gradual and bents for small I_{c}, which is captured by our simulation. In addition, the lowest LRS resistance as well as the spread of roughly two orders of magnitude are in a good agreement with the measurements for the same I_{c}.[21] For the HRS resistance, the overlap with the higher value of the LRS resistance is also seen in our simulation results. To check if the change of the resistance is stochastic over the cycles, the resistances are plotted over 100 cycles. No fixed pattern of the
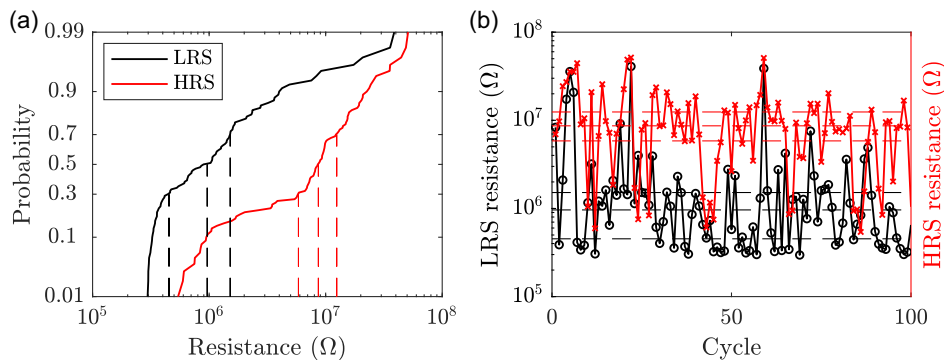

Figure 4. Simulation results of a) the cumulative distribution function and b) the LRS resistance and HRS resistance versus cycles. Dashed lines in (a) and (b) are the 30%, 50%, and 70% values, following the same definition as measurements.
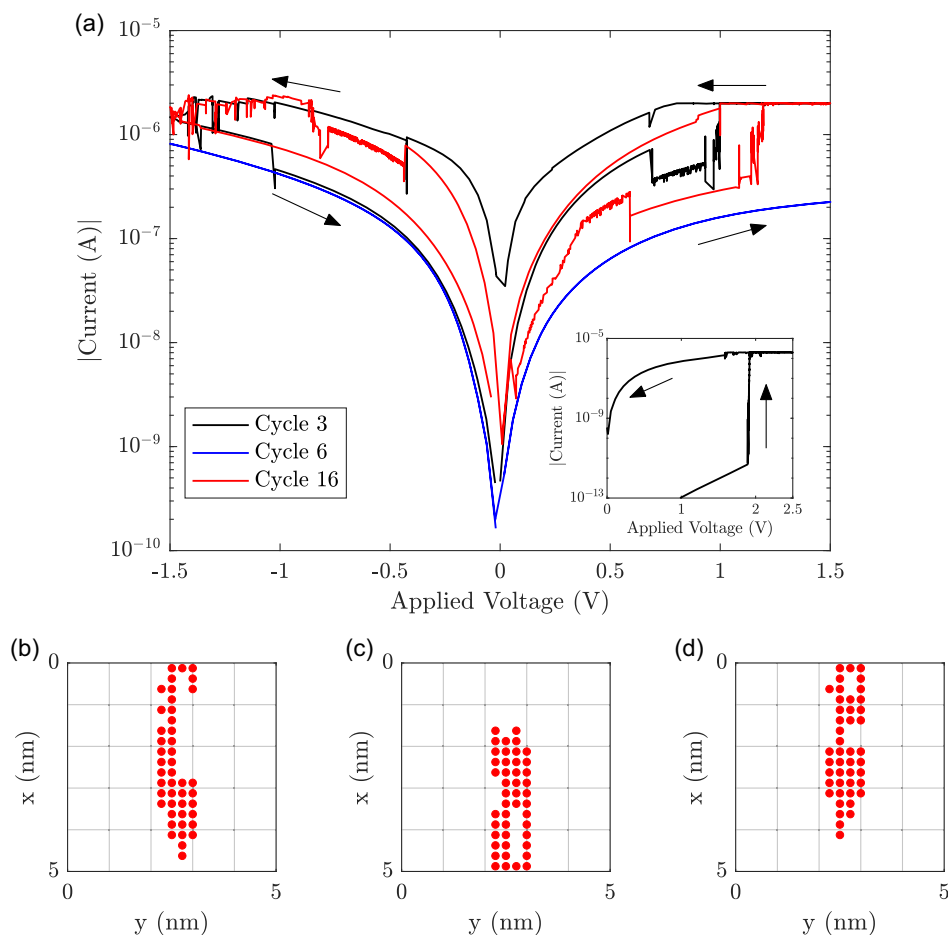
ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

physica
status
solidi
a
www.pss-a.com

Figure 5. a) The current–voltage (I–V) characteristic curves for three representative cycles in the main figure and the FORMING process in the inset. b–d) The vacancy distributions along the grain boundary plane of the third, sixth, and sixteenth cycle at $V_{meas} = -0.1\,V$ for the LRS resistance, respectively.

resistance occurs in Figure 4b. It thus suggests that the vacancy migration is stochastic during the dynamical process.

To further investigate the relation between the resistance and the microscopic vacancy distributions, three cycles with a low, median, and high LRS resistance are selected. **Figure 5**a shows current–voltage (I–V) characteristic curves, where the voltage being applied to the bottom electrode is plotted against the absolute value of current, and the corresponding vacancy distributions when the resistances are measured are shown in Figure 5b–d. It is clearly seen that the existence of the gap has a large impact on the LRS resistance as expected. In Figure 5b, a conductive path between the two electrodes appears and thus the LRS resistance is in the low resistance regime. However, the conductive path is not composed of a single column of vacancies. Instead, the connection between both electrodes is due to two groups of vacancies, i.e., the more mobile and less mobile vacancies belonging to the stronger and weaker GB effect shown in Figure 2, respectively. The dynamics during the SET process is demonstrated in **Figure 6**. In the early stage of the SET process, the gap near the top interface is first filled by the more mobile vacancies and thus an abrupt increase of the current is observed. The migration of vacancies will slow down due to the decrease of V_{cell}, while the low resistance is obtained if

the less mobile vacancies can stay in the bottom region. For a large enough I_c, the mobile vacancies are surrounded by a large amount of less mobile vacancies generated in the FORMING stage and the migration will be stopped. This leads to the constant resistivity for large I_c. We then interpret the constant V_{trans} as a result of Poisson statistics, namely $A \propto I_c$[22] where A is the cross section of the CF.

In contrast, the constant $V_{trans} = I_c \times \mu_R$ does not hold for a small I_c[21] and a higher median LRS resistance is obtained. The larger median resistance of the measurement, $\mu_R > R_0 \equiv V_{trans}/I_c$, implies that the resistivity is not a constant due to the interruption of the conductive path, such as a rupture of the CF or a large gap after the SET process. It can be seen that the large gap still exists at the top interface during the SET process in Figure 5c, suggesting that even the mobile vacancies are not driven by the applied voltage. As a consequence, a failure of the SET operation happens and the resistance is still in the HRS level, i.e., $R \gg R_0$ with the vacancy distribution similar to the one corresponding to a high HRS resistance shown in **Figure 7**a. If the gap persists for multiple cycles, a FORMING process is required to make the switching take place again. However, this is not observed in the simulation and the resistance is reduced after a few cycles (Figure 4b). At the cycle where the resistance is close to the median
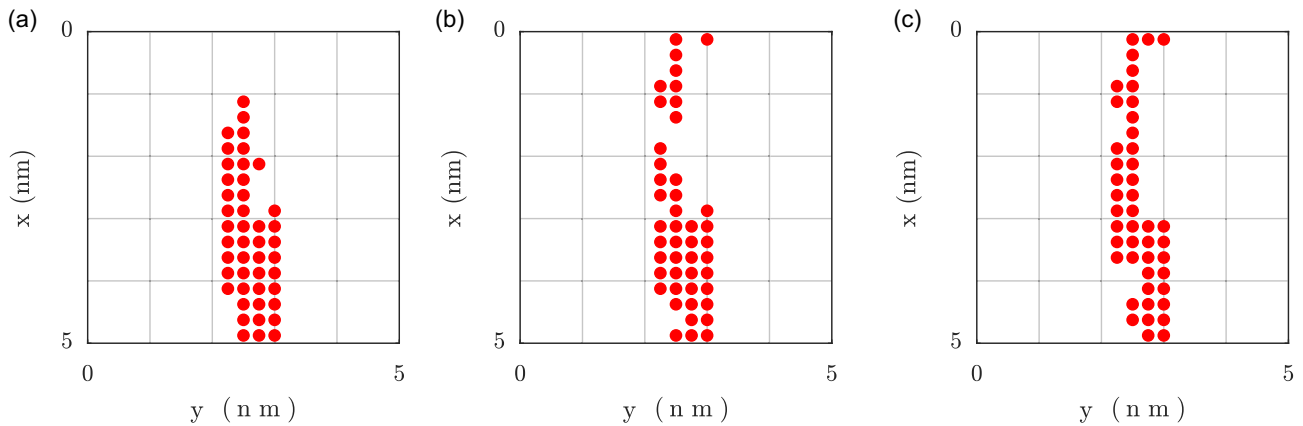
ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

physica
status
solidi a

www.pss-a.com

Figure 6. a,b) The vacancy distributions before and after I_c is reached in the third cycle. The large gap at the top region is filled by vacancies after the I_c is reached. c) The minor gap is filled by further migration.
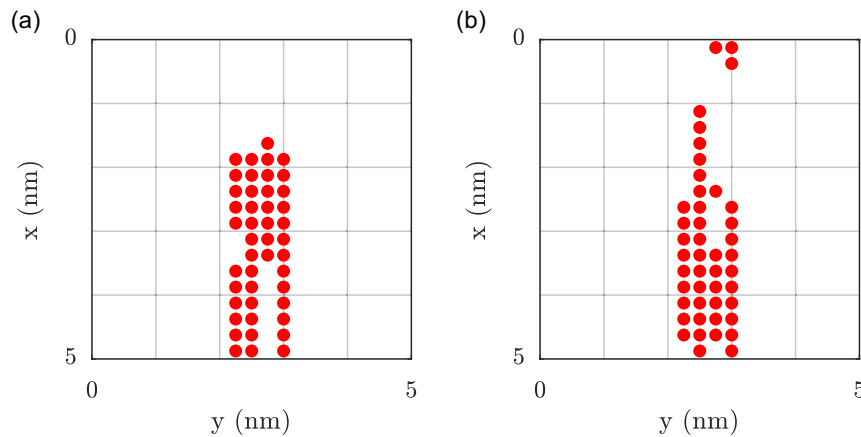

Figure 7. The vacancy distributions at $V_{meas} = -0.1\,V$ for the HRS resistance. a) The vacancy distribution of the high HRS resistance is similar to the one corresponds to the SET failure while b) the longer column of vacancies together with residual vacancies at the top interface contribute to a lower HRS.

value, the broken connection to the bottom electrode by less mobile vacancies is seen in Figure 5 d. Interestingly, the vacancy distribution of $R > R_0$ in the LRS can also be observed in the HRS in Figure 7b. The difference is that the position of the minor gap is close to the top interface due to the opposite polarity.

While different vacancy distributions at V_{meas} for each resistance level have been shown, the widespread of LRS resistance suggests the vacancy distribution is changing from one type to the other type instead of staying the same. The switching mechanism between different types of vacancy distribution can be explained by two processes, i.e., the increasing and decreasing of the LRS resistance compared to the previous cycle. **Figure 8** shows the scenario responsible for the change from a low LRS resistance to a higher value in the next cycle. In the early stage of the RESET process, vacancies migration toward the bottom interface is shown in Figure 8a,b. In general, it leads to formation of a gap near the top electrode followed by the appearance of the hottest region around the interface of the CF and the gap. In this region, the escape of vacancies from the longest column is possible with the aid of the high temperature while the highest temperature is decreasing as the gap grows. As a consequence, the length of shorter vacancy columns are increased

which results in the top vacancy of the shorter column entering the high-temperature region in the next SET operation shown in Figure 8c. Once this vacancy migrates upward, the second vacancy in this column will also move upward due to the reduced energy barrier of the CF effect, see Figure 8d. In this case, the vacancies below will follow this migration pattern and thus leave a gap at the bottom interface in the end of the SET operation shown in Figure 8e.

On the contrary, **Figure 9** shows the opposite situation where the LRS resistance is lower in the next cycle. In this case, the shorter columns roughly reach the middle height in the beginning of the RESET process shown in Figure 9a. After mobile vacancies fill the gap at the bottom interface, the escape of vacancies in the longer column cannot take place because neighboring sites are already occupied by vacancies. Alternatively, a column of vacancies will reach the bottom region easily after the downward migration of the lowest vacancy which is similar to the SET operation in the previous case. In Figure 9b, the high-temperature region is seen around the bottom of shorter column of vacancies, which suggests that the described migration is possible during the RESET process. As a result, most vacancies are in the bottom region in the late stage of the RESET process as
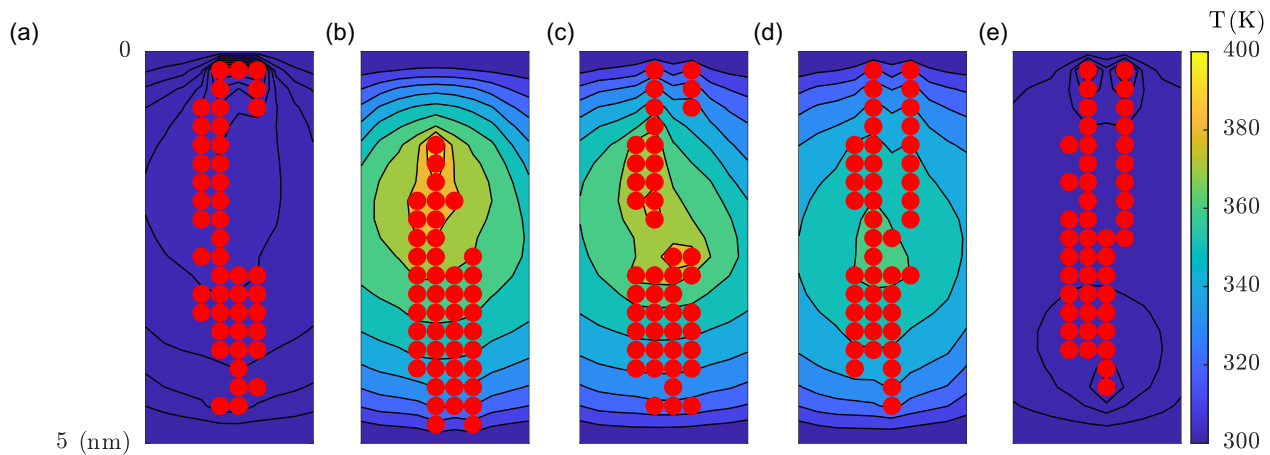
2300403 (7 of 10)

Figure 8. 2D maps of the temperature distribution, where vacancies are shown in red filled circles, in the process of a lower LRS resistance becoming a higher LRS resistance in the next cycle. a,b) Configurations of the early and the subsequent stages of the RESET process. c–e) Configurations of the middle and the subsequent stages of the SET process in the next cycle.
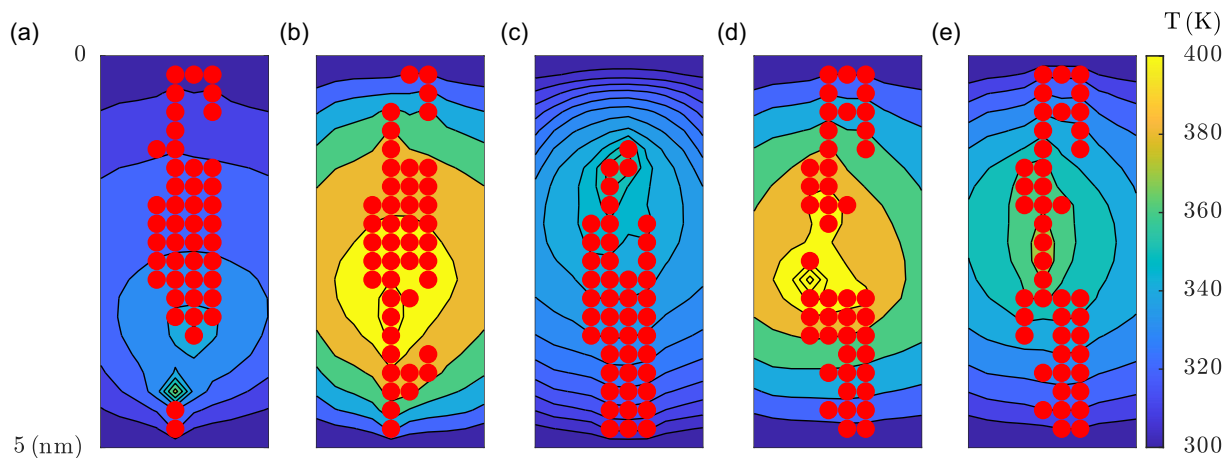
Figure 9. 2D maps of the temperature distribution, where vacancies are shown in red filled circles, in the process of a higher LRS resistance becoming a lower LRS resistance in the next cycle. a–c) Configurations of the early and the subsequent stages of the RESET process. d,e) Configurations of the middle and the subsequent stages of the SET process in the next cycle.

shown in Figure 9c. With more vacancies in the bottom region, the top vacancy of the shorter column can be away from the high-temperature region or the temperature is low enough to trigger the migration of this vacancy (Figure 9e) and thus the CF effect is suppressed during the SET operation. In addition, vacancies outside of the longer CF might be attracted into it during the migration process shown in Figure 9d. In this way, the connection of both electrodes by the conductive path composed of both mobile and less mobile vacancies is recovered and the LRS resistance is lower.

4. Conclusion

In this article, we extend our model and discuss the specific value, $I_c = 2 \mu A$, in the small current regime. An extra charge state is introduced to capture the influence of the local structure,

i.e., the CF, and different zero-field energy barriers are applied to the corresponding intrinsic charge states. In addition, the GB, which has an impact on both the generation and migration processes, is considered and the spatial distribution of it is assumed. The modulation of zero-field activation energies due to the CF and GB are taken into account and the impacts to the vacancy dynamics, which leads to the observed statistical measures, are discussed.

It turns out that the large variation of the LRS resistance is due to three types of vacancy distributions. The SET failure is the result of a large gap created in the RESET process, for which even the mobile vacancies cannot fill the gap because of the ensuing low temperature. In the case of a successful SET operation, the position of less mobile vacancies is critical after more mobile ones reach the top interface. That is, the connection to the bottom electrode due to less mobile vacancies leads to the low resistance $R \approx V_{trans}/I_c$. With the lowest LRS resistance $R = 3 \times 10^5 \, \Omega$

taken from the simulation result, we obtain $V_{trans} = 0.6\,V$, which is comparable to the experimental value, $0.4\,V$. In addition, the LRS C2C variability is close to the measurements. Furthermore, the dynamics for the switching between different types of vacancy distributions is discussed. It turns out that the position of the high-temperature region in both SET and RESET processes is critical. If the first vacancy of a consecutive vacancy column is moved, the other vacancies left behind tend to move as well due to the CF effect. The relative position of consecutive vacancies together with the exchange of vacancies between long and short filaments leads to the large variation of the LRS resistance from C2C.

This work explores a potential scenario for the switching instability in the small I_c regime by using $I_c = 2\,\mu A$ with the assumed GB distribution. The stoichiometry in the GB area is critical to the switching dynamics. Under the assumption of a single GB with a large area, simulation results indicate a strong reduction of both the LRS resistance and the C2C variability due to the introduction of excessive vacancies in the FORMING stage. In contrast, a single GB with a small area leads to a region filling with vacancies due to the anisotropic diffusion barrier within the GB area. As a result, the gap is small and thus the LRS resistance is always low. The specific vale of the I_c is chosen since it is the lowest value used in the measurement and we hypothesize the proposed scenario as an elementary process originated from a single GB area. Similar to the assumption of multiple GB,[19] the inclusion of multiple GB areas provides a possible approach to extend the proposed model. The generalization of different I_c together with the multiple GB area remains for future works.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Research data are not shared.

[1] D. Ielmini, *Semicond. Sci. Technol.* **2016**, *31*, 063002.

[2] G. D. Wilk, R. M. Wallace, J. M. Anthony, *J. Appl. Phys.* **2001**, *89*, 5243.

[3] F. Zahoor, T. Z. A. Zulkifli, F. A. Khanday, *Nanoscale Res. Lett.* **2020**, *15*, 90.

[4] P. Chi, S. Li, C. Xu, T. Zhang, J. Zhao, Y. Liu, Y. Wang, Y. Xie, in *2016 ACM/IEEE 43rd Annual Int. Symp. on Computer Architecture (ISCA)*, **2016**, pp. 27–39.

[5] Y. Long, T. Na, S. Mukhopadhyay, *IEEE Trans. Very Large Scale Integr. VLSI Syst.* **2018**, *26*, 2781.

[6] A. Chen, *IEEE Electron Device Lett.* **2015**, *36*, 138.

[7] R. Liu, H. Wu, Y. Pang, H. Qian, S. Yu, *IEEE Electron Device Lett.* **2015**, *36*, 1380.

[8] R. Waser, R. Dittmann, G. Staikov, K. Szot, *Adv. Mater.* **2009**, *21*, 2632.

[9] Y. Zhang, G.-Q. Mao, X. Zhao, Y. Li, M. Zhang, Z. Wu, W. Wu, H.-J. Sun, Y. Guo, L. Wang, X. Zhang, Q. Liu, H. Lv, K.-H. Xue, G. Xu, X. S. Miao, S. Long, M. Liu, *Nat. Commun.* **2021**, *12*, 7232.

[10] J.-Y. Chen, C.-W. Huang, C.-H. Chiu, Y.-T. Huang, W.-W. Wu, *Adv. Mater.* **2015**, *27*, 5028.

[11] T. Perevalov, V. Aliev, V. Gritsenko, A. Saraev, V. Kaichev, *Microelectron. Eng.* **2013**, *109*, 21.

[12] R. Dittmann, S. Menzel, R. Waser, *Adv. Phys.* **2021**, *70*, 155.

[13] S. Larentis, F. Nardi, S. Balatti, D. C. Gilmer, D. Ielmini, *IEEE Trans. Electron Devices* **2012**, *59*, 2468.

[14] S. Kim, S. Choi, W. Lu, *ACS Nano* **2014**, *8*, 2369.

[15] S. Yu, X. Guan, H.-S. P. Wong, *Appl. Phys. Lett.* **2011**, *99*, 063507.

[16] X. Guan, S. Yu, H.-S. P. Wong, *IEEE Trans. Electron Devices* **2012**, *59*, 1172.

[17] S. Dirkmann, J. Kaiser, C. Wenger, T. Mussenbrock, *ACS Appl. Mater. Interfaces* **2018**, *10*, 14857.

[18] G. Jegert, A. Kersch, W. Weinreich, U. Schröder, P. Lugli, *Appl. Phys. Lett.* **2010**, *96*, 062113.

[19] L. Vandelli, A. Padovani, G. Bersuker, D. Gilmer, P. Pavan, L. Larcher, in *2011 3rd IEEE Int. Memory Workshop (IMW)*, IEEE, Piscataway, NJ **2011**, pp. 1–4.

[20] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, D. Ielmini, *IEEE Trans. Electron Devices* **2014**, *61*, 2912.

[21] A. Fantini, L. Goux, R. Degraeve, D. Wouters, N. Raghavan, G. Kar, A. Belmonte, Y.-Y. Chen, B. Govoreanu, M. Jurczak, in *2013 5th IEEE Int. Memory Workshop,* IEEE, Piscataway, NJ **2013** pp. 30–33.

[22] S. Balatti, S. Ambrogio, D. C. Gilmer, D. Ielmini, *IEEE Electron Device Lett.* **2013**, *34*, 861.

[23] E. Wu, R.-P. Vollertsen, *IEEE Trans. Electron Devices* **2002**, *49*, 2131.

[24] K. Rushchanskii, S. Blügel, M. Lezaic, *Phys. Rev. Mater.* **2018**, *2*, 115002.

[25] D. Duncan, B. Magyari-Köpe, Y. Nishi, *IEEE Electron Device Lett.* **2016**, *37*, 400.

[26] K. McKenna, A. Shluger, *Appl. Phys. Lett.* **2009**, *95*, 222111.

[27] R. Nieminen, A. Jansen, *Appl. Catal., A* **1997**, *160*, 99.

[28] D. T. Gillespie, *J. Comput. Phys.* **1976**, *22*, 403.

[29] A. U. Modak, M. T. Lusk, *Solid State Ionics* **2005**, *176*, 2181.

[30] S. Arrhenius, *Z. Phys. Chem.* **1889**, *4U*, 226.

[31] K. Xiong, J. Robertson, M. C. Gibson, S. J. Clark, *Appl. Phys. Lett.* **2005**, *87*, 183505.

[32] J. Robertson, O. Sharia, A. A. Demkov, *Appl. Phys. Lett.* **2007**, *91*, 132912.

[33] K. Kamiya, M. Y. Yang, T. Nagata, S.-G. Park, B. Magyari-Köpe, T. Chikyow, K. Yamada, M. Niwa, Y. Nishi, K. Shiraishi, *Phys. Rev. B* **2013**, *87*, 155201.

[34] Y. Guo, J. Robertson, *Appl. Phys. Lett.* **2014**, *105*, 223516.

[35] B. Gao, B. Sun, H. Zhang, L. Liu, X. Liu, R. Han, J. Kang, B. Yu, *IEEE Electron Device Lett.* **2009**, *30*, 1326.

[36] E. Abbaspour, S. Menzel, C. Jungemann, in *2015 Int. Conf. on Simulation of Semiconductor Processes and Devices (SISPAD)*, **2015**, pp. 293–296.

[37] E. Abbaspour, S. Menzel, A. Hardtdegen, S. Hoffmann-Eifert, C. Jungemann, *IEEE Trans. Nanotechnol.* **2018**, *17*, 1181.

[38] A. Miller, E. Abrahams, *Phys. Rev.* **1960**, *120*, 745.

[39] I. Lundström, C. Svensson, *J. Appl. Phys.* **1972**, *43*, 5045.

[40] D. Ielmini, *IEEE Trans. Electron. Devices* **2011**, *58*, 4309.

[41] M. A. Panzer, M. Shandalov, J. A. Rowlette, Y. Oshima, Y. W. Chen, P. C. McIntyre, K. E. Goodson, *IEEE Electron. Device Lett.* **2009**, *30*, 1269.

[42] B. Butcher, G. Bersuker, L. Vandelli, A. Padovani, L. Larcher, A. Kalantarian, R. Geer, D. Gilmer, in *2013 5th IEEE Int. Memory Workshop*, IEEE, Piscataway, NJ **2013** pp. 52–55.

[43] J. McPherson, J.-Y. Kim, A. Shanware, H. Mogul, *Appl. Phys. Lett.* **2003**, *82*, 2121.

[44] J. W. McPherson, H. C. Mogul, *J. Appl. Phys.* **1998**, *84*, 1513.

[45] K.-H. Xue, P. Blaise, L. R. C. Fonseca, G. Molas, E. Vianello, B. Traoré, B. De Salvo, G. Ghibaudo, Y. Nishi, *Appl. Phys. Lett.* **2013**, *102*, 201908.

[46] A. Padovani, L. Larcher, O. Pirrotta, L. Vandelli, G. Bersuker, *IEEE Trans. Electron Devices* **2015**, *62*, 1998.

[47] M. Schie, S. Menzel, J. Robertson, R. Waser, R. A. De Souza, *Phys. Rev. Mater.* **2018**, *2*, 035002.

[48] A. O'Hara, G. Bersuker, A. A. Demkov, *J. Appl. Phys.* **2014**, *115*, 183703.

[49] M. Youssef, K. J. Van Vliet, B. Yildiz, *Phys. Rev. Lett.* **2017**, *119*, 126002.

[50] F. Puglisi, L. Larcher, G. Bersuker, A. Padovani, P. Pavan, *IEEE Electron. Device Lett.* **2013**, *34*, 387.

[51] A. Padovani, L. Larcher, P. Padovani, C. Cagli, B. De Salvo, in *2012 4th IEEE Int. Memory Workshop*, IEEE, Piscataway, NJ **2012** pp. 1–4.

[52] F. Nardi, S. Larentis, S. Balatti, D. C. Gilmer, D. Ielmini, *IEEE Trans. Electron. Devices* **2012**, *59*, 2461.

[53] P. Jančovič, B. Hudec, E. Dobročka, J. Dérer, J. Fedor, K. Fröhlich, *Appl. Surf. Sci.* **2014**, *312*, 112.

[54] W. Kim, S. Menzel, D. J. Wouters, Y. Guo, J. Robertson, B. Roesgen, R. Waser, V. Rana, *Nanoscale* **2016**, *8*, 17774.

[55] J. Robertson, S. J. Clark, *Phys. Rev. B* **2011**, *83*, 075205.

[56] D. Strukov, F. Alibart, S. Williams, *Appl. Phys. A* **2012**, *107*, 509.

[57] Jülich Supercomputing Centre, *JLSRF* **2018**, *4*, A132.

[58] X. Zhao, D. Vanderbilt, *Phys. Rev. B* **2002**, *65*, 233106.

[59] X. Zhao, D. Vanderbilt, *MRS Online Proc. Libr.* **2002**, *747*, T5.2/N7.2.