Defect model for bipolar resistive switching in metal-insulator-metal heterostructures

Duk K. Shin,[1] Gergely T. Zimanyi,[1,*] and O. Heinonen[2,3]

[1]*Physics Department, University of California, Davis, California 95616, USA*
[2]*Materials Science Division, Argonne National Laboratory,*
9700 S. Cass Ave., Building 212, Argonne, Illinois 60439-4806, USA
[3]*Department of Physics, Northwestern University, Evanston, IL 60208-3112*

A model of the bipolar resistance switching observed in many Metal-Insulator-Metal (MIM) system is introduced. The model contains electrons and charged defects. The electronic motions are determined by a system evolution path which reduces the total electronic energy of the system, while the motions of the charged defects are determined by the electric field produced by local charges and potentials at the electrodes. Simulations demonstate that the model can reproduce key experimental observations of resistive switching in MIM systems, such as hysteretic I-V curves and constant resistance as the applied voltage crosses zero. An interface layer with a large concentration of charged defects is necessary for producing the experimental features observed in resistive switching. The presence of such an interface layer leads to electric fields from the charge build-up at the conductive layer next to a high-resistance interface layer, as well as from the potential difference applied to the electrodes, and these fields are the main driving forces for transporting the charge defects which gives rise to the resistive switching.

I. INTRODUCTION

Metal-insulator-metal (MIM) structures in which an insulating layer is sandwiched between two metals can exhibit resistive switching for a wide range of materials choices. There are two classes of resistive switching, bipolar switching and unipolar switching. A general feature common to both classes is that the system can be switched from a high-resistance insulating or semiconducting state to a low-resistance semiconducting or metallic state at some applied voltage $V_{\rm set}$. Systems exhibiting unipolar switching can be returned to a high-resistance state by first reducing the applied voltage to near zero, then applying a voltage $V_{\rm reset} < V_{\rm set}$ with the same polarity as $V_{\rm set}$.[?] In contrast, systems exhibiting bipolar switching can be returned to a high-resistance state by reducing the voltage and reversing it to $V_{\rm reset} \approx -V_{\rm set}$. We will here focus in bipolar switching. By cycling the applied voltage between $V_{\rm reset}$ and $V_{\rm set}$ a hysteretic I-V curve is obtained, as is illustrated in Fig. **??**.

Although the phenomenon has been observed for over forty years[?] , a comprehensive understanding of the underlying mechanism or mechanisms that cause resistive switching is still lacking. Recently, resistive switching in MIM structures has attracted renewed interest due to its potential use in non-volatile memories, termed resistive random access memory (RRAM).[?] Many theoretical models have been proposed to capture and predict resistive switching in MIM structures. Among these models are a drift-diffusion equation based model of ion and electron/hole movement[?] , a model of resistive switching based on interactions between Schottky barrier and oxygen vacancy positions[?] , a model based on energy bands and electron traps[?] , as well as several phenomenological models: metallic filamentary path[?] , electrochemical migration of point defects[? ?] and oxygen vacancy transfer in perovskite-type transition-metal oxides[?] . Here, we propose and analyze a model that includes electrons and charged defects, *e.g.,* oxygen vacancies. In our model the electron motions are determined by choosing the electron paths that minimize the system's total electronic energy. The motion of the charged defects is controlled by the electric fields that arise from electrochemical potential differences. The goal of our work is to qualitatively investigate I-V curves produced using different configurations of MIM structures, *e.g.,* both static and dynamic charged defects, and to establish if any configuration qualitatively reproduces experimentally observed resistive switching behaviors such as I-V hysteresis and low-to-high and high-to-low bipolar resistance switches, and, if so, which combination of material characteristics produces the most desirable resistance switching behavior to serve as a qualitative guide for future experiments.

II. MODEL DESCRIPTION

The model consists of a semiconducting or insulating oxide film sandwiched between two metallic electrodes, which we will call the left and right electrodes. Within the insulator, there is initially a step-like density of charged defects, which we will think of as oxygen vacancies. This could correspond to an insulator with an as-deposited bi-layer structure consisting of a stoichiometric TiO_2 layer and a TiO_{2-x} layer with an intentionally high vacancy density closer to the top electrode. Alternatively, this density vacancy profile can be the result of an electroforming process. The film is modeled by a two-dimensional grid of cells. The cells represent a

* zimanyi@physics.ucdavis.edu

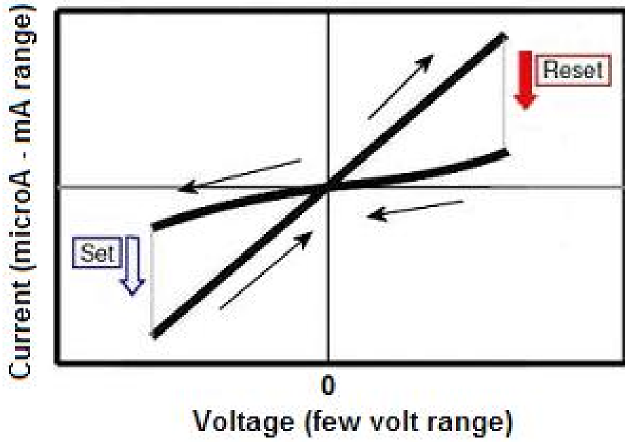

FIG. 1. Schematic I-V diagram showing bipolar resistance. The set operation takes place in one voltage polarity while the reset operation occurs in the opposite polarity.

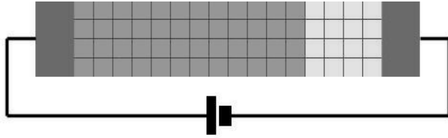

FIG. 2. Figure 5 - Schematic diagram of the model. The left and right blue areas are the electrodes. The vacancy density is high in the green conductive region. In the yellow depletion or interface region the vacancy density is initially low, but varies during the cycling protocol. The external voltage is applied to the left electrode while the right electrode is grounded.

coarse-graining of a real structure on the scale of many unit cells, and a cell can contain hundreds of atoms. A schematic diagram is shown in Fig. **??**.

The dynamics of the electrons and the oxygen vacancies is driven by energetics, represented by an energy of each cell and of the electrodes. The cell energy has four components: self-charging or capacitive energy, the Coulomb interaction with other cells, site-disorder, and the energy from the external potential. The relative strengths of the energy terms are controlled by scaling constants. These scaling constants are chosen by comparing the $I-V$ output of the model to that of experimental results and selecting the constants that produce $I-V$ outputs that most resembles the experimentally observed dependence of current I on applied voltage V. The site disorder energy is represented by a random site energy term. The scale of the energy terms can be bracketed or fixed by connecting the results of our simulations to the experimental data.

III. EXPERIMENTAL UNDERPINNING

Two important experimental observations have been incorporated in our model. First, recent studies have noted that in many MIM structures in which the insulating layer is a transition-metal oxide, oxygen-ion related defects are essential for the resistance switching[?,?,?,?]. Second, experiments have achieved bipolar switching using asymmetric interface layers: one with a quasi-ohmic contact to the electrode in both high- and low-resistance state, and the other interface layer blocking current from flowing between the electrode and the oxide in the high-resistance state[?]. The blocking interface layer is the active region where the resistive switching occurs. These observations are integrated into our model by introducing a concentration of oxygen vacancies, in the system. In general, a high local oxygen vacancy concentration leads to a low local resistivity. A low vacancy concentration in the interface layer adjacent to the right electrode will then lead to a high local resistivity and a high resistance for the whole system. We will from now on consider a square grid with a total of 64 cells distributed over four rows and 16 columns; the right-most four columns are interface cell columns with variable oxygen vacancy concentration. Initially, two out of four columns near one interface have a low oxygen vacancy concentration and therefore constitute a local high-resistance region. We will explain below how increasing the oxygen vacancy concentration in the initally high-resistance region increases the electron transition probability and therefore increases the current. The amount of oxygen vacancy concentration is represented by oxygen vacancy density $\rho_{\rm ov}$, normalized to 1: $0 \leq \rho_{\rm ov} \leq 1$. Oxygen vacancies can move along the voltage gradient, *i.e.* perpendicularly to the electrode interfaces, so that $\rho_{\rm ov}$ is constant in each column. Left and right electrode work functions are parameterized and added to the energy of each electrode. When work functions are not specified as input parameters, a default value is computed for the work function for each electrode. This default value is the energy slightly higher than the average energy of the adjacent interface cells at zero external voltage. This choice of the work function prevents the electrons in the interface cells from entering the electrodes at zero voltage and makes the system keep a quasi-ohmic interface in the absence of the high-resistance interface cells.

IV. CELL AND ELECTRODE ENERGIES

The energy E_i of insulator cell i is taken as

$$
\begin{aligned}
E_i = & C^{\mathrm{s}}(n_i - N_{\mathrm{avg}})^2 \\
& + \frac{C^{\mathrm{c}}}{2} \sum_{k \neq i} \frac{(n_k - N_{\mathrm{avg}})(n_i - N_{\mathrm{avg}})}{d_{ik}} \\
& + V_i n_i + C^{\mathrm{d}} L_i n_i.
\end{aligned} \tag{1}
$$

Here, n_i represents the number of valence electrons in i^{th} cell, and $N_{\mathrm{avg}} = \langle n_i \rangle$ is the average number of electrons per cell which is computed self-consistently to enforce charge neutrality. The first term on the right-hand side represents the self-charging energy of the cell. This is the largest cell-energy term for coarse grained models, capturing the Coulomb energy cost of placing n_i electrons in the cell. Its magnitude is controlled by the scaling factor $C^{\mathrm{s}} = e^2/C_i$, where the cell capacitance C_i will be taken to be uniform. The second term represents the Coulomb interaction of the electrons in the cell with the electrons in other cells with $C^{\mathrm{c}} = e^2/a$, where a is the average spacing between cells; d_{ik} is dimensionless. The third term represents the potential energy related to the external potential V. The potential V_i of the ith cell is a linear function of current I and includes a resistive scaling factor. The fourth term represents the disorder in each cells energy with L_i a random variable selected from a normal distribution $P(L_i)$ with a width σ, and C^{d} is a scaling factor to control the overall disorder size. The interplay of the energetics of the vacancies and that of the electrons will enter into the model through the dynamics.

V. ELECTRODE PROPERTIES

The two electrodes are modeled as electron reservoirs at different chemical potentials. The number of electrons making transitions between an electrode and adjacent insulator cells is governed by energy balance. The total energy of the left (right) electrode, E^A_{elec} ($A = L, R$) is taken as

$$
E^A_{\mathrm{elec}} = C^{\mathrm{s}} \left(n^A_{\mathrm{elec}}\right)^2 + V^A n^A_{\mathrm{elec}} + W^A_{\mathrm{elec}}, \tag{2}
$$

In Eq. (**??**), n^A_{elec} represents the number of charges in the electrode cell, and is given by

$$
n^A_{\mathrm{elec}} = \frac{C^A_{\mathrm{elec}} V^A}{e} + n^A_{\mathrm{eq}}. \tag{3}
$$

Here, C^A_{elec} is the capacitance of the electrode (scaled by C^{s}), V^A is the external voltage applied to the electrode, and n^A_{eq} is the equilibrium charge in the absence of an external voltage. The first term on the right-hand side of Eq. (**??**) represents a self-charging energy, the second term represents the potential energy, and the third term represents the work function of the metallic electrode, W^A_{elec}.

VI. DYNAMICS

The model simulates the motion of electrons and the oxygen vacancies under an applied external voltage V. Electrons and oxygen vacancies move so as to reduce the total energy of the system until equilibrium or a steady state is reached. Once equilibrium or steady state is reached, system data such as the total current I, the energy landscape and the electron/ion charge distributions are recorded for the given external voltage V. The external voltage V is then changed by a small amount and the dynamical simulation is repeated. This way, the $I-V$ curve for a given set of parameters and initial conditions can be generated.

VI.1. Electron dynamics

The steady state electron distribution is obtained for a given external voltage V by first computing $n^A_{\mathrm{elec}}(V)$, from which the electrode energies $E^A_{\mathrm{elec}}(n_{\mathrm{elec}})$ can be computed. The voltage V_i for each column in the insulator is then computed, and the energy E_i of each cell is updated using the new V_i. Electron transitions between an electrode and the insulator are then calculated, as well as transitrions between neighboring cells within a maximum range range. We will here include transitions both between near-neighbor cells and next-nearest neighbors. These steps are then repeated until equilibrium or steady state is reached within a specified numerical tolerance. This protocol determines the steady state electron current I corresponding to the applied voltage V. The probability of an electron making a transition from one cell to another cell is computed by comparing the change in total system energy ΔE if an electron from cell i moves to cell k, and also by coupling the oxygen vacancy density to the transition probability. If i and k are nearest neigbors, the transition is allowed if $\mathcal{P}(\rho_{\mathrm{ov}}) \exp\left[-\gamma d_{ik} - \beta \Delta E\right] > x$, where x denotes a random number drawn from a rectangular distribution on (0,1]. Here, β is the effective inverse temperature of the system, γ represents the spatial decay rate of the localized electronic states (an effective inverse localization length), and $d_{ik} = 0$ if i and k are nearest neighbors. The prefactor $\mathcal{P}(\rho_{\mathrm{ov}}) = \exp[-\delta(1-\rho_{\mathrm{ov}}, i)] \exp[-\delta(1-\rho_{\mathrm{ov}}, k)]$ i f either cell is a vacancy cell, and unity otherwise. The parameter δ is chosen so that total transition probability into a cell with unit (saturated) oxygen vacancy concentration is ten times higher than total transition probability into the same cell with zero oxygen vacancy ($\rho_{\mathrm{ov}} = 0$). This choice of resistance ratio of 10 is motivated by experiments. The exponential form of the transition probability comes from the fact that the transitions are tunneling processes. Electrode-cell tranistions using an analogous algorithm, comparing the total energy of the system before and after a possible transition between an electrode and an adjacent interface cell in either direction.

VI.2. Oxygen Vacancy Transport Rules

As discussed in the introductory sections, it appears that dynamics of the oxygen vacancies must be included in order to reproduce experimental resuts. In our mode, vacancies are driven by the local electric field and also feel the effect of disorder. In order to have a small number of parameters, we used the simple picture in which the vacancies are driven by the gradient of the cell energies $\Delta E_{ij} = E_i - E_j$, which should be an acceptable first approximation. This approximation is equivalent to the picture of Rozenberg[?] , where the energy gradients driving the vacancies are generated by the electrons, and the back-fields generated by the vacancies for the electrons are not included. For the dynamics, we used the Mott-Guerney picture[?] of the ions performing activated jumps over an energy barrier in both directions according to the law of the detailed balance:

$$v \approx fa\mathrm{e}^{-\beta U_A} \sinh\left(\frac{\beta qEa}{2}\right) = \begin{cases} \mu E & E \ll E_0 \\ \mu E_0 \mathrm{e}^{\frac{E}{E_0}} & E \sim E_0 \end{cases}, \tag{4}$$

where v is the average ion drift velocity, f is the frequency of escape attempts, q is the ion charge, while $\mu = qfa\exp(-\beta U_A)$ is the mobility at small electric fields, $E_0 = 2k_\mathrm{B}T/(qa)$ is the characteristic field for a particular mobile atom in the crystal, which is typically about 1 MV/cm for T = 300K, and U_A is the energy barrier. Given the low operating voltages of the devices, we will use the Mott-Guerney formula in the low-field, linear response limit. In practice, since the vacancies move only one cell in our simulation, the above Mott-Guerney formula for the vacancy current is implemented by making the number of the oxygen vacancies in cell i transitioning to the adjacent cell j proportional to the energy gradient.

We employ a coarse graining to the vacancy number n_ov. We represent the oxygen vacancies in terms of their density ρ_ov and allow only discreet changes of the density. Finally, we adopt a maximum vacancy density to represent the constraints on the vacancy density. These arise from the initial sample preparation with a particular oxygen density as well as chemical energetics and Coulomb penalties not preferring high vacancy densities. We normalize the oxygen vacancy density so that its maximum is unity and it can change only in steps of 0.1 Importantly, the oxygen vacancy density directly affects the transport of electrons as the vacancies dramatically impact the resistivity encountered by the electrons as described earlier. Finally, the electrodes are assumed to be completely blocking for vacancies, so that vacancies are not allowed to move into an electrode.

An equilibrium or steady state is achieved for each applied external voltage V by performing sweeps over all cells in the systems. In each sweep, transitions from one cell i to all other ones and to the electrodes are considered and performed until cell i is in steady state with the other cells and the electrodes. Similarly, allowed oxygen vacancy movements between cell i and the others are considered. The sweeps are then repeated at least 100 times and until the current from the insulator from one electrode is equal to the current from the other electrode to the system within numerical tolerances.

VII. RESULTS AND DISCUSSION

To generate an $I - V$ curve, an external voltage is applied to one electrode with the other electrode grounded. The voltage will be referred to as positive when the left electrode has a potential higher than the right electrode and negative when it is lower. The voltage is changed in uniform steps starting from zero until the current suddenly changes by a large amount, *i.e.* a switching event occurs. The magnitude of the voltage is then decreased in the same uniform steps through zero to negative values until a reverse switching event occurs. From this, the voltage protocol is completed by starting to increase the voltage back to positive values. Our model has a five-dimensional parameter space spanned by C^s, C^c , C^d, N_avg, and W_elec. We have explored this five-dimensional space extensively comparing the simulated $I - V$-curves to key features of the experimentally determined $I - V$ curves of RRAM devices. This allowed us to determine how each parameter affects the $I - V$ hysteresis. Many $I - V$ curves did not meet key criteria for RRAM applications. For example, for some regions of parameter space, the $I - V$ curves did not exhibit hysteresis. In other regions, the $I - V$ curves switched from the high resistance (R_Hi) state to the low resistance (R_Lo) state at a switching voltage, but switched back to R_Hi state before the voltage reached zero on the descending branch of the voltage protocol. Such systems obviously fail to be candidates for non-volatile memory, as they do not retain their resistance state once the write-voltage is removed. Yet others had a curvature so that the values of the resistance on the R_Lo branch kept increasing and would span the resistance in the R_Hi state. Such systems are obviously not good for memory applications. The parameter regions where the system exhibits proper non-volatile memory behavior were identified by performing systematic two-dimensional slices of the five dimensional parameter space and creating contour maps of the resistance ratios on the two branches of the $I - V$ curves as they crossed zero voltage.

VII.1. Electron-only dynamics

One of the key goals of our project was to determine the minimal or paradigmatic model that captures the bipolar switching behavior adequately. To explore whether or not the model can exhibit hysteresis without including dynamics of the oxygen vacancies, we performed simulations with the oxygen vacancies frozen.

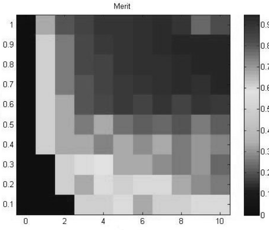

FIG. 3. Contour plot of the RRAM resistance ratio $R_{\rm hi}/R_{\rm lo}$. The horizontal axis is $N_{\rm ave}$, and the vertical axis is the self-energy factor $C^{\rm s}$.

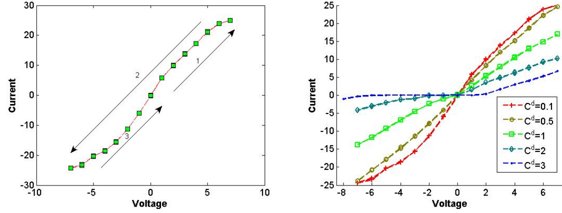

FIG. 4. Left panel: Typical I-V curve obtained with $\rho_{\rm ov} = 1$ and the oxygen vacancies cannot move. Here, $C^{\rm s} = 0.3$, $C^{\rm c}$=0.1, and $N_{\rm avg}$=10. No hysteresis is observed. Right panel: I-V curves for the same set of parameters but with different disorder energies $E^{\rm d}$. No hysteresis is observed. As the disorder energy increases, a larger voltage is required to produce a given current as electrons get trapped in cells with low energy due to disorder.

The left panel of Fig. **??** shows a typical $I-V$ curve for the case of $\rho_{\rm ov} = 1$ in every cell and frozen oxygen vacancies. When the voltage was cycled according to the above protocol, no hysteresis was observed in the $I-V$ curve. The right panel shows a a number of $I-V$ curves for different values of the disorder parameter $C^{\rm d}$. No hysteresis was observed. Similar runs of simulations were performed in other one- and two-dimensional slices of the $(C^{\rm s}, C^{\rm c}, C^{\rm d}, N_{\rm avg})$ parameter space, all with negative results. Therefore, we conclude that our model does not exhibit hysteresis in the limit of including only electron dynamics.

VII.2. Dynamics of both electrons and vacancies

Figure **??** shows a typical $I-V$ curve obtained when dynamics of both electrons and oxygen vacancies are included. Hysteresis proved to be a persistent feature in wide segments of the parameter space. Moreover, the two branches of these hysteretic $I-V$ curves cross zero voltage with markedly different slopes. In other words these models do retain their resistance states after the write voltage is removed. Therefore, these models seem to reproduce the two basic phenomena of bipolar resistive switching. In what follows, we describe our exploration of the parameter space and the various classes of

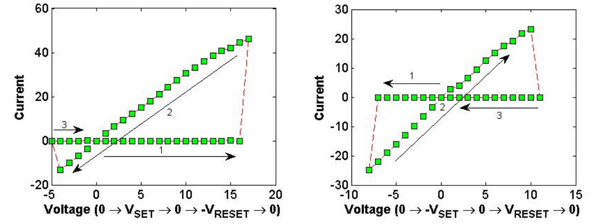

FIG. 5. $I-V$ curves from the model with interface layers containing mobile oxygen vacancies. Left panel: the voltage sweep starts with positive polarity. $V_{\rm set} = 16$ and $V_{\rm reset} = -5$. Here, $C^{\rm s} = 0.3$, $C^{\rm c} = 0.1$, $C^{\rm d} = 0.5$, $N_{\rm avg} = 30$, and disorder $\sigma = 5$. Right panel: the voltage sweep starts with negative polarity. $V_{\rm set} = -7$ and $V_{\rm reset} = 11$. The same parameters as the left panel are used.

switching that we have observed.

Depending on the sense of cycling of the applied voltage, the $I-V$ curve assumes different shapes. When a voltage cycle starts with positive polarity ($0 \to V_{\rm set} \to 0 \to -V_{\rm reset}$), an asymmetric $I-V$ curve results, as shown in the left panel of Fig. **??**. Furthermore, the $I-V$ curve typically has a larger set voltage $V_{\rm set}$ and a smaller reset voltage $V_{\rm reset}$. In contrast, when a voltage cycle starts with negative polarity ($0 \to -V_{\rm set} \to 0 \to V_{\rm reset}$), the resulting asymmetric $I-V$ curve (the right panel of Fig. **??**) has a larger $V_{\rm reset}$ than $V_{\rm set}$. These $I-V$ curves are qualitatively similar to the experimental result obtained ones for an asymmetric Pt/$SrZrO_3$/$SrRuO_3$ MIM structure[?] . Below we analyze the physics of these two cycling protocols in detail. We show that for both senses of the cycling, the processes dominating the physics take place in the interface layer and the bulk region immediately next to it.

The following analysis is for the case of a voltage cycle starting with negative polarity. Negative voltage is realized by applying a positive voltage to the right electrode while grounding the left electrode. The positive voltage is increased from 0 to $V_{\rm set}$, defined as the voltage where the current increases abruptly. For voltage $V < V_{\rm set}$, a low initial, or as-prepared concentration of oxygen vacancies in half of the interface cells near the right electrode (columns 13 and 14 in the right panel of Fig. **??**) prevents electrons from transitioning into these interface cells, as the low oxygen vacancy density suppresses transitions as described above above. Also, a thick interface region with low oxygen vacancy concentration prevents electrons from tunneling through the interface region directly to the electrode. As a result, the system is in a high-resistance state during this low-voltage portion of the voltage cycle or protocol. As the voltage of the right electrode increases, the electrons in the bulk cells of columns 1 through 12 are increasingly attracted towards the right electrode. But since they cannot enter the interface region itself, the electrons pile up in the cells just left of the interface region next to the right electrode, as shown in the middle of the left panel of Fig. **??**.

Accordingly, these Interface-Adjacent (IA) cells become increasingly negatively charged. The electrostatic potential from the excess negative electrons accumulated in the IA cells and the increasingly positive voltage applied at the right electrode combine to exert an increasing energy gradient from right to left for the positively charged oxygen vacancies inside the interface. Since the vacancy configuration was formed by minimizing its energy, the vacancies are experiencing a local energy minimum and do not immediately move in response to the increasing energy gradient. However, in a depinning-like manner, at a particular voltage the energy gradient becomes large enough to move some oxygen vacancies to the left, as illustrated in the middle and lower panel of Fig. **??**. This vacancy drift and rearrangement reduces the width of the region with low oxygen vacancy density and reduces the barrier width against which the electrons pile up in the IA cells, allowing them to enter the left region of the interface. While we did not explicitly simulate this, we speculate that if the exponential portion of the Mott-Guerney law for the vacancy drift would have been kept in our simulations, this switch could have been made even more dramatic. But it is noted that this further enhancement of the sharpness of switching would require the Mott-Guerney characteristic field E_0 to fall within the range of the field corresponding to the presently studied V_{set} voltage. Otherwise the system may switch at a lower voltage, stabilizing itself and never reaching the E_0 field.

The process of vacancy drift and rearrangement has two main effects. First, it forms an increasingly sharply defined step or wall in the vacancy concentration, and second, this vacancy-density wall moves towards the right electrode, in effect narrowing the poorly conducting low-vacancy region. Both of these phenomena, the wall formation and the wall propagation, are also exhibited in the simulations by Strukov, Borghetti, and Willams[?] . Our model includes one more physical process not present in Ref. **?** . As the poorly conducting low vacancy interface region narrows, the probability of electrons successfully tunneling through it increases. The opening of the tunneling transport channel through the right interface region allows the piled-up electrons to successfully reach the right electrode. One consequence of this punch-through is that the electrons can traverse the entire interface experiencing less resistance: the system switches into the low resistance state. This is due in part to the fact that while the resistance increases in the part of the interface region where the oxygen vacancy density is reduced, this is more than compensated for by opening a tunneling transport channel as the width of this interface layer is reduced. Second, as the excess electron density in the IA cells is discharged or dissipated, the driving force for further vacancy movement is diminished: the wall of the vacancy density stops moving and the system is trapped in a stable steady state. For example, Fig. **??** illustrates the case when the regions low in oxygen vacancies are narrowed down to single cell thickness and alternated with high vacancy interface regions. It is also reassuring that the process is self-limiting: the vacancy density cannot increase in all interface cells because conservation of the number of vacancies.

In order to re-set the system to a high-resistance state, the magnitude of the voltage applied to the right electrode is reduced from V_{set} to zero. Subsequently, the polarity of the overall applied voltage is switched as a voltage is applied to the left electrode, while keeping the right electrode grounded. The left-electrode voltage is increased in magnitude in equal steps from 0 to a voltage V_{reset}, where the value of the current drops abruptly. As the magnitude of the external voltage on the left electrode is increased, the electrons entering from the right electrode can efficiently tunnel through the low vacancy density interface regions (columns 13 and 15 of Fig. **??**) and the system remains in a low-resistance state. However, the low vacancy regions in the interface cells have more and more pronounced effects on the electron landscape in the interface region for the following reasons. First, the tunneling probability across the low vacancy density regions (column 13 and 15) is decidedly smaller than the direct transport from the high vacancy density cells as well as from the IA cells in column 12 towards the left electrode. Therefore, the electron density in a cell becomes dependent on the vacancy densities in neighboring cells. Second, the low-vacancy cells maintain approximately the same number of electrons as the magnitude of the voltage is increased, while high-vacancy cells become more and more positively charged. Accordingly, there will be a smaller number of electrons in the interface regions with high oxygen vacancy concentration (columns 14 and 16 in Fig. **??**), compared to the cells with low oxygen vacancy concentration (columns 13 and 15 in Fig. **??**). This reduces the electron density in the IA (col. 12) as well as the number of high-vacancy interface cells, leading to the buildup of an effectively positive charge in these regions (columns). As a consequence, the low-vacancy cells become more and more negative in comparison, though these cells do not acquire many electrons. The resulting charge inhomogeneity becomes more pronounced as the magnitude of the external voltage is increased and creates an energy gradient or field that exerts a force on the positively charged vacancies directed towards the low-vacancy region. This gradient eventually overcomes the restoring force near the local energy minimum in which the system is located and becomes large enough to move the excess oxygen vacancies. The step or wall in the vacancy distribution then becomes less sharp and it moves to the left. This phenomenon was observed in Ref. **?** . The motion of the oxygen vacancies increases the thickness of the high resistance region thickness as shown in Fig. **??**, and the probability of tunneling across this region drops exponentially, rapidly deactivating this transport channel. This deactivation increases the resistivity considerably: the system switches back to a high-resistance

state. The magnitude of $V_{\rm set}$ is typically smaller than that of $V_{\rm reset}$, because for the SET process electrons pile up outside the interface region on the left and thus create a large electric field/energy gradient acting on the vacancies. Therefore, a smaller external voltage is sufficient to move the vacancies. In contrast, for the RESET process the more negatively-charged low-vacancy cells are scattered around the more positively charged cells rich in oxygen vacancies. As a consequence, the source of the electric field generated by the electrons is spatially more diffuse. Therefore, during the RESET process the electron-generated electric field can be less effective for moving the vacancies and thus a higher external voltage is needed to induce the RESET. The magnitude of $V_{\rm set}$ can be somewhat smaller than $V_{\rm reset}$, as in the right panel of the above Fig. **??**, or much smaller as in the left panel of the Fig. **??**. The precise difference depends on the details of the charge distribution and vacancy density profile, as well as on the model parameters.

Next, we analyze the opposite voltage cycling protocol, when the external voltage applied to the left electrode is increased from zero with the right electrode grounded. As before, an initial large thickness of the low oxygen vacancy region in the left half of the interface region prevents electrons from tunneling through the interface to the electrode. As the external voltage applied to the left electrode increases, the electrons in the system move to the left and into the left electrode, but few electrons enter the system from the right electrode since the thick oxygen vacancy deficient interface region acts as an efficient barrier against tunneling. The right-most column with high oxygen vacancy density cells (column 14 in Fig. **??**) becomes increasingly electron-deficient, building up an effective positive charge density (hole density). As the external voltage increases further, the charge density discrepancy between the interface cells with high oxygen vacancy density (column 14) and those with low oxygen vacancy density (column 15) increases. Eventually, at $V_{\rm set}$, the energy gradient force and the external positive voltage from the left electrode become large enough to push the oxygen vacancies away from the electron-depleted positively charged cells (column 14) to the right into the less positively charged interface cells (column 15). This reduces the thickness of the region with lowoxygen vacancy density and allows the electrons to tunnel through the barrier, making the system conductive. The process is self-limiting. Once the oxygen vacancy density becomes large enough in cells close to the electrode, tunneling sets in, and the system transitions into the low resistance state. At this voltage the electron depletion layer disappears, and with it a large part of the force driving the vacancy movement. As a result, the vacancy distribution does not change any further.

In order to re-set the system to a high-resistance state, the magnitude of the voltage applied to the left electrode is reduced from $V_{\rm set}$ to zero. Subsequently, a positive voltage is applied to the right electrode while keeping the left electrode grounded. The positive voltage applied to the right electrode is increased in magnitude in equal steps from 0 to a voltage $V_{\rm reset}$, where the value of the current drops abruptly. In this case, though the fields induced by the inter-cell electron density gradient play a role in moving the oxygen vacancies, the fields are weak in comparison to the strong electric repulsion from the close-by positively polarized right electrode. The electrode eventually repels the oxygen vacancies away to the left at $V_{\rm reset}$. We have in this case not observed any large differences in electron density between the columns before $V_{\rm reset}$, as shown in Fig. **??**, indicating that inter-cell electron density gradient plays a minor role in this reset process. At $V_{\rm reset}$, motion of the oxygen vacancies widens the high-resistance region and most electrons cannot tunnel through the region: the current drops abruptly and the system switches to the high-resistance state.

The analysis shows that the resistance switching is mainly controlled by the oxygen vacancy movements in the active interface region. The close proximity of the positively charged electrode plays a prominent role in repelling the oxygen vacancies away from the electrode, speeding up the set or reset process, depending on the polarity of the cycle. Also, the initial electron density composition of low-vacancy high-resistance (LVHR) cells plays a big role in determining the critical voltages $V_{\rm set}$ or $V_{\rm reset}$. For example, in the case that the electron density increases in cells with high oxygen vacancy density as the external voltage increases, an electron density gradient will arise more quickly between these cells

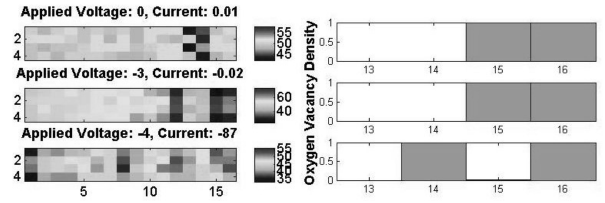

FIG. 6. (Color online) $C^{\rm s}=0.3$, $C^{\rm c}=0.1$, $C^{\rm d}$, and $N_{\rm avg}=50$. Left panel: 4×16 cell matrices showing each cells density of electrons for three different applied voltages to the right electrode. Applied voltages have a multiplicative factor of 10; e.g., the applied voltage of -4 shown in the bottom diagram means -40. The right-most four columns are interface cell columns. The fifth column of cells from the right is adjacent to the interface cells. For a voltage below a threshold voltage $V_{\rm set}$, the cells in the fifth column have a high density of electrons. As the electrode voltage reaches a threshold value $V_{\rm set}$, the system reaches a low-resistance state and this charge build up dissipates to the right. Right panel: The oxygen vacancy concentration (shown in the green color) for the four interface columns (this concentration is uniform within each column) for the above three applied voltages. The four numbers below each box indicate the column numbers. As the applied voltage reaches the threshold value $V_{\rm set}$, the oxygen vacancies move to the left, leaving only the right-most column with low oxygen vacancy concentration.

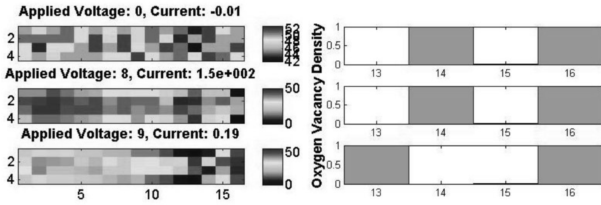

FIG. 7. (Color online) Left panel: The electron concentration as in Fig. ??, but with the external potential applied to the left electrode. As the voltage is increased, a large difference in electron density difference develops between high oxygen vacancy cells and low oxygen vacancy cells. Right panel: The oxygen vacancy concentration for the left three applied voltages. At $V_{\rm reset}$, oxygen vacancies in column 14 are pulled into column 13 due to the energy gradient between the two columns, originating from the growing electron density difference. The width of the high resistance region increases, the system reaches a high-resistance state and the current drops abruptly.

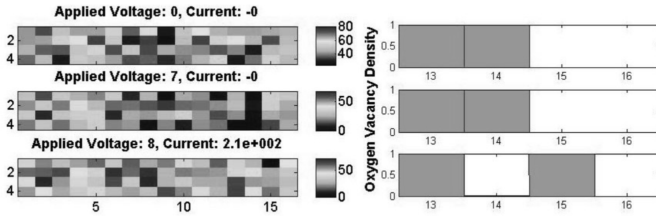

FIG. 8. (Color online) $C^{\rm s} = 0.3$, $C^{\rm c} = 0.1$, $C^{\rm d} = 1$, and $N_{\rm avg} = 50$. Left panel: 4×16 cell matrices showing each cells density of electrons for three different applied voltages to the left electrode. Applied voltages have a multiplicative factor of 10; e.g., the applied voltage of 8 shown in the bottom diagram means 80. The right-most four columns are interface cell columns. Note that for the voltage 7 which is just below a threshold voltage $V_{\rm set}$, the third right interface column is almost depleted of electrons and is very charge positive. A large energy gradient develops between the column and the adjacent low oxygen vacancy column. Right panel: The oxygen vacancy concentration for the four interface columns (this concentration is uniform within each column) for the above three applied voltages.

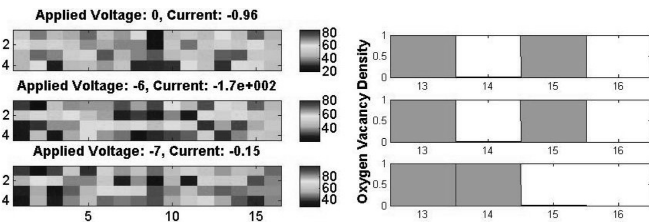

FIG. 9. (Color online) The electron concentration as in Fig. ??, but with the external potential applied to the right electrode. Right panel: The oxygen vacancy concentration for the left three applied voltages. At $V_{\rm reset}$, oxygen vacancies in column 15 are pushed into the column 14 due to the repulsion from the positively polarized right electrode.

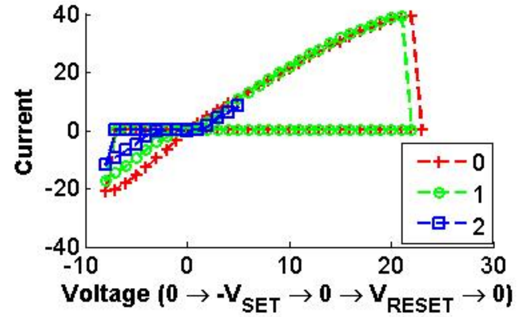

FIG. 10. I-V curves for different work functions. For the case of a work function of magnitude 2, the simulation stopped at an applied voltage of 5 because the system had already been re-set to a high resistance state. As a consequence, the system did not maintain low resistance as the voltage crosses zero.

and neighboring LVHR cells with lower electron density than if neighboring LVHR have high electron density, and the system reaches a critical voltage more quickly.

VIII. EXPLORING THE MODEL PARAMETER SPACE

In addition to the interface interactions, the insulator material properties also impact the characteristics of the resistive switching. We have explored a range of model parameters $(C^{\rm s}, C^{\rm c}, C^{\rm d})$ and choose a set of parameters that give $I - V$ characteristics similar to those observed experimentally. From this selected set of model parameters, we evaluated the impact of each parameter on the switching behavior. The simulations show that the system loses a low-resistance zero-voltage crossing feature, and therefore also its usefulness in non-volatile memory applications, when the electrode work function is higher than a threshold, as shown in Fig.??. This suggests that it may be advantageous to utilize insulators and electrodes whose Fermi levels are comparably close to each other for memory applications. Another downside of the higher electrode work function is that it requires a higher set voltage $V_{\rm set}$.

When the disorder factor $C^{\rm d}$ is varied from 0 to 5 while keeping the other parameters constant, the resulting $I - V$ curves are shown in Fig. ??in which a dc voltage is applied from 0 to $V_{\rm set}$, and then to 0 and to $V_{\rm reset}$. A higher disorder factor creates a smaller hysteresis and increases the set voltage while decreasing the reset voltage. For the highest disorder factor of five, the current can even stop flowing before the voltage reaches zero. The $I - V$ curves in Fig. ?? exhibit a slight jaggedness because of the built-in randomness in the electron transitions. However, the overall qualitative effect of changing the disorder factor can be clearly seen.

Increasing the disorder also has the effect of making the electron distribution more inhomogeneous. Cells in valleys of the disorder energy landscape accumulate more electrons while cells in peaks of the disorder en-

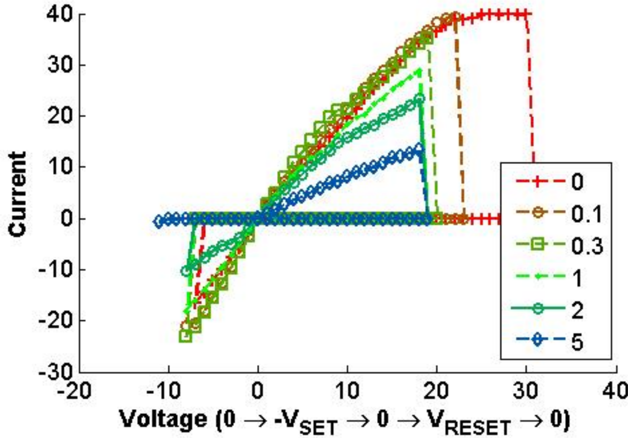

FIG. 11. $I-V$ curves for different values of the disorder factor $C^{\rm d}$.

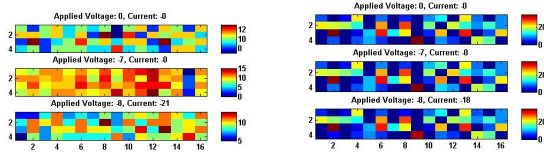

FIG. 12. (Color online) Left panel: Snapshot of the electron densities in a low-disorder system. Right panel: snapshot of the electron density in a high-disorder system.

ergy landscape acquire much fewer electrons. As shown in Fig. **??**, the lower-disorder system has more evenly distributed electrons for the different applied voltages. This enables more easily electron transitions and allows the system to retain its resistance state across zero voltage. As the higher color contrasts among cells in Fig. **??** show, the higher disorder system leaves patches of cells with a large number of electrons while other cells have scarce electrons. A higher disorder factors represent deeper traps that localize more electrons. This has the adverse effect of inhibiting electron transitions and the system quickly returns to a high-resistance state as the voltage is reduced after $V_{\rm set}$. In effect, high disorder has the tendency of re-trapping or localizing the electrons as the voltage is lowered, preventing the system to be on the low-resistance branch at zero voltage and thus undermining the non-volatile memory function. In summary, the disorder has the following effects. (a) Increasing disorder increases $V_{\rm set}$, (b) increasing disorder decreases $V_{\rm reset}$, and (c) there is a maximum disorder above which the system is in a high-resistance state near zero voltage. From these observations, for non-volatile memory applications a lower disorder is preferred since the device can accommodate a lower write voltage while its higher reset voltage has a better retention characteristic against a small electrical stress. Thus, devices having an insulator layer film with a smaller number and size of impurity traps are preferred. When the self energy factor $C^{\rm s}$ is varied from 0.1 to 1 while keeping the other parameters constant, the resulting $I-V$ curves are shown in Fig. **??**. When the self energy factor is lower than a threshold, the system loses a low-resistance zero voltage crossing, and the simulation is stopped at a voltage of 5: the case of $C^{\rm s}=0.1$. Increasing $C^{\rm s}$ has a favorable effect of reducing $V_{\rm set}$ slightly, but an adverse effect of increasing $V_{\rm reset}$. Increasing $C^{\rm s}$ has the effect of limiting accumulation of electrons in a given cell; the requirement of a minimum threshold $C^{\rm s}$ can be met by using an insulator film with small grains.

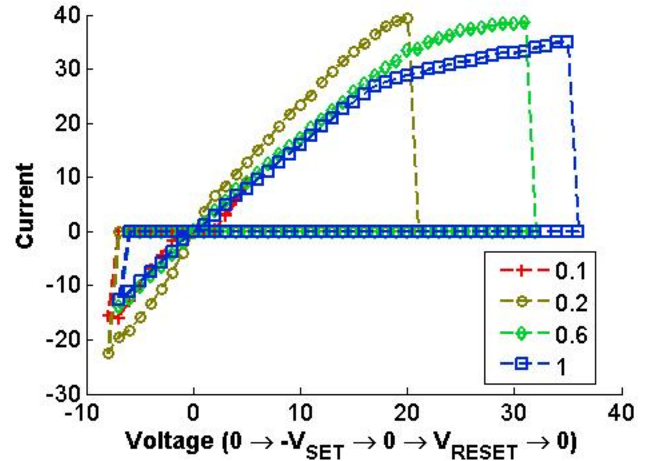

FIG. 13. $I-V$ curves for different self energy factors $C^{\rm s}$. For the case of $C^{\rm s}=0.1$, the system loses its low resistance zero voltage crossing characteristic and the simulation was stopped at an applied voltage of 5.

When the coulomb factor $C^{\rm c}$ is varied, above a threshold value, the system is unable to retain its low resistance state approaching zero voltage after $V_{\rm reset}$. This sets a maximum limit on $C^{\rm c}$ for RRAM applications. Below the threshold $C^{\rm c}$, increasing $C^{\rm c}$ increases $V_{\rm set}$ but decreasing $V_{\rm reset}$ (Fig. **??**).

IX. CONCLUSIONS

We have constructed a model dynamics of both electrons and charged defects (here thought of as oxygen vacancies) for MIM systems that is capable of exhibiting bipolar resistive switching. In the model the electrons impact the vacancy dynamics by generating an electric field for them, whereas the vacancies impact the cell-to-cell transport of the electrons. The energetics of our coarse-grained model included the self-charging energy of each grain, their Coulomb interaction, a disorder term and the electric field generated by the external applied voltage. Our simulations of the model demonstrated the following: (*i*) The presence of mobile vacancies that support the electron transport is necessary to produce hysteretic $I-V$ curves with resistive switch-

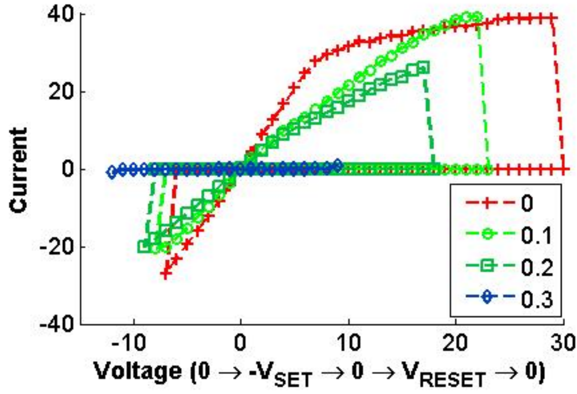

FIG. 14. $I - V$ curves different values of Coulomb factors C^{c}. For the case of $C^{\mathrm{c}} = 0.3$, the system did not maintain a low-resistance state as the applied voltage crossed zero and the simulation was stopped at an applied voltage of 5.

ing between high- and low-resistance states. The mobility of the vacancies is especially important in the interface layer adjacent to the electrode. (*ii*) As an external voltage is applied, an excess electron density builds up adjacent to the interface region or electrons are depleted in the interface region, depending on the voltage polarity. The excess electrons or the electron depletion generate an internal field that enhances the effect of the external voltage. Upon reaching a V_{set} voltage, the combined effect of the external and internal fields is capable of moving out the vacancies from their local energy minima. (*iii*) During the set process the vacancies move to break the insulating interface layer. This can take the form of distributing oxygen vacancy rich cells so that a conductive channel is made. This distribution breaks up an insulating region into a combination of narrower insulating and conducting regions. The narrowing of the insulating interface region eventually activates the parallel transport channel of the electrons tunneling through the narrowed insulating interface region. This parallel channel allows the excess electrons to reach the electrode efficiently, switching the system into the low resistance state. (*iv*) This switching process is self-arresting: as the exodus of the electrons depletes the buildup or the electrons move into the electron depleted region, the internal field gets weakened, stopping the further movement of the vacancies. Therefore, the conductive path formed is persistent for a finite range of the external voltages above V_{set}. However, if the voltage keeps increasing, all the vacancies will be pushed into one region, leaving a thick insulating region, switching the system back to a high-resistance state. (*v*) The reset process is analogous, with the difference that the internal and the external fields push the oxygen vacancies to make a thicker insulating interface region. (*vi*) The retention of the low-resistance states is more robust for samples with fewer disorders. This avoids the disorder re-trapping, or localizing the electrons prematurely on the descending branch of the $I - V$ curve. (*vii*) Relatively small grains are preferred to increase the self energy factors C^{s} above a minimum threshold to achieve the retention of the resistive state at zero voltage. (*viii*) The strength of the Coulomb energy interaction should be limited to below a threshold value again for retaining the resistive state across zero voltage. This can be done by, *e.g.*, utilizing materials that have a higher dielectric constant. (*ix*) A smaller work function between the electrode and the adjacent interface cells is preferred. Our model is able to account for the main features of the observed resistance switching behavior of MIM structures. The model and the simulations on the small samples did not reproduce secondary features such as wavy $I - V$ shapes. Such details hopefully can be captured by incorporating additional physics, such as ionic traps, temperature effects and Joule heating.

ACKNOWLEDGEMENTS

O.H. acknowledges support from Argonne National Laboratory. Argonne National Laboratory is operated under Contract No. DE-AC02-06CH11357 by UChicago Argonne, LLC.