Metal oxide resistive memory switching mechanism based on conductive filament properties

Citation: J. Appl. Phys. 110, 124518 (2011); doi: 10.1063/1.3671565
View online: http://dx.doi.org/10.1063/1.3671565
View Table of Contents: http://jap.aip.org/resource/1/JAPIAU/v110/i12
Published by the American Institute of Physics.

Related Articles
Balanced ternary addition using a gated silicon nanowire

Multi-level phase change memory devices with Ge2Sb2Te5 layers separated by a thermal insulating Ta2O5 barrier layer
J. Appl. Phys. 110, 124517 (2011)

Unstable periodic orbits and noise in chaos computing
Chaos 21, 047520 (2011)

Non-volatile high-speed resistance switching nanogap junction memory

Reducing operation current of Ni-doped silicon oxide resistance random access memory by supercritical CO2 fluid treatment

Additional information on J. Appl. Phys.
Journal Homepage: http://jap.aip.org/
Journal Information: http://jap.aip.org/about/about_the_journal
Top downloads: http://jap.aip.org/features/most_downloaded
Information for Authors: http://jap.aip.org/authors
Metal oxide resistive memory switching mechanism based on conductive filament properties

G. Bersuker,1,a) D. C. Gilmer,1 D. Veksler,1 P. Kirsch,1 L. Vandelli,2 A. Padovani,2 L. Larcher,2 K. McKenna,3 A. Shluger,3 V. Iglesias,4 M. Porti,4 and M. Nafria4

1SEMATECH, 257 Fuller Road, Albany, New York 12203 and 2706 Montopolis Drive, Austin, Texas 78741, USA
2DISMI Università di Modena e Reggio Emilia and IU.NET, 42100 Reggio Emilia, Italy
3University College London, WC1E 6BT, United Kingdom
4Universitat Autònoma de Barcelona, Edifici Q, 08193 Bellaterra, Spain

(Received 12 July 2011; accepted 19 November 2011; published online 29 December 2011)

By combining electrical, physical, and transport/atomistic modeling results, this study identifies critical conductive filament (CF) features controlling TiN/HfO2/TiN resistive memory (RRAM) operations. The leakage current through the dielectric is found to be supported by the oxygen vacancies, which tend to segregate at hafnia grain boundaries. We simulate the evolution of a current path during the forming operation employing the multiphonon trap-assisted tunneling (TAT) electron transport model. The forming process is analyzed within the concept of dielectric breakdown, which exhibits much shorter characteristic times than the electroforming process conventionally employed to describe the formation of the conductive filament. The resulting conductive filament is calculated to produce a non-uniform temperature profile along its length during the reset operation, promoting preferential oxidation of the filament tip. A thin dielectric barrier resulting from the CF tip oxidation is found to control filament resistance in the high resistive state. Field-driven dielectric breakdown of this barrier during the set operation restores the filament to its initial low resistive state. These findings point to the critical importance of controlling the filament cross section during forming to achieve low power RRAM cell switching. © 2011 American Institute of Physics. [doi:10.1063/1.3671565]

INTRODUCTION

Among the technology options for non-volatile memory devices, resistive switching memory (RRAM) is currently under consideration because of its superior scaling opportunities.1–3 This technology offers the possibility of reversible changes in the conductivity of a dielectric film between high (HRS) and low (LRS) resistance states by consecutively applying voltage pulses of either different or similar polarities to the metal-insulator-metal (MIM) memory cell. Within the RRAM technology, an approach based on switching between the LRS and HRS by changing the resistance of the conductive filament (CF) in a transition metal oxide dielectric has attracted attention due to its promise of low power and good retention. While the key factors driving the resistive switching in metal oxides have been identified,4–11 there is no sufficiently developed microscopic description of the CF features that enable memory operations. This study is focused on establishing the nanoscale properties of the CF in HfO2-based systems.12

To describe the material changes in a dielectric associated with resistive switching, one must start with the CF formation, which ultimately determines the CF properties. We discuss the CF forming process in HfO2 in terms of dielectric breakdown (BD), an abrupt formation of a localized region between the electrodes, within which the dielectric composition becomes more oxygen-deficient, rendering this region conductive (the CF formation). Dielectric breakdown represents a “weak link” event, controlled by the first conductive path formed between the electrodes; therefore, it is expected to follow Weibull statistics.13 As shown below, the forming voltage values for HfO2 did indeed fit the Weibull distribution.

Within the BD description, the formation of a metal-rich CF requires breaking the Hf-O bonds, with the subsequent diffusion of oxygen ions out of the filament region. To compare with the SiO2 dielectric, which has been thoroughly studied due to its wide range of applications in the semiconductor industry, the oxygen deficiency of the breakdown path has been directly observed by electron energy-loss spectroscopy (EELS) measurements.14 Oxygen ion diffusion in crystalline HfO2 (e.g., “free” O− ions, with an activation energy of 0.3 eV, Ref. 15) is more efficient than oxygen vacancy diffusion (which in effect corresponds to the migration of a fully coordinated bounded oxygen with an activation energy of 1.2 eV and 0.7 eV for V+ and V2+, respectively). Therefore, for consistency, the processes following the forming (BD) event are presented in terms of the movement of oxygen ions rather than oxygen vacancies. Oxygen dissociation is shown to be driven by the applied electric field and elevated temperatures. The former is effective due to the high polarizability of high-k dielectrics, while the latter is caused by electron transport through the existing oxygen

a)Author to whom correspondence should be addressed. Electronic mail: gennadi.bersuker@sematech.org.
vacancies, which is accompanied by energy dissipation that increases the local temperature. The energy dissipation can be described by the multiphonon trap-assisted tunneling (TAT) electron transport mechanism, which explicitly considers structural rearrangement of the dielectric lattice around the trap associated with the electron trapping/detrapping.\textsuperscript{16–18} The TAT description allows the initial stage of the filament formation, when electron transport through the filament is still dominated by hopping, to be simulated. Specifically, it reveals essentially non-uniform heat dissipation across the dielectric thickness since a mismatch between the energies of the injected electrons and electron transferring traps increases when the traps are located deeper in the dielectric (due to band bending under the applied bias). This in turn leads to a non-uniform temperature profile through the conductive path and promotes the formation of a CF, the resistive characteristics of which vary along its length. The CF non-uniformity is shown to play a critical role in HRS/LRS switching by enhancing redox processes at the region of the CF with the highest resistance.

An analysis of the factors controlling the breakage of the chemical bonds suggests that the BD (responsible for the forming) represents a runaway process driven by positive feedback between a temperature increase and defect generation. Electron transport through the defect generates phonons increasing the temperature around the defect position. This enhances the generation of new defects near existing ones, thus increasing the current through these defects as well as increasing the local temperature, etc. A clustering of generated defects may lower the energy barrier for the dissociation of the nearby oxygen in the surrounding lattice, which further enhances defect generation around the BD region (the proposed BD process will be reported in detail elsewhere). Due to the exponential dependence of the bond breakage rate on the electric field and temperature, the BD transient exhibits much shorter characteristic times, on the order of the phonon frequencies, than the electroforming process widely used to describe CF formation.\textsuperscript{1} This allows the bond breakage process to be considered within a “frozen” atomic configuration of the lattice, ignoring in the first approximation the vacancy diffusion, which occurs on a much longer time scale.

This study is organized as follows: We first identify the morphological feature of the HfO\textsubscript{2} dielectric film (which is polycrystalline as deposited at T > 300 C by the atomic layer deposition (ALD) process in the thickness range of >4 nm) and the grain boundaries (GBs) that constitute preferential leakage current paths. During forming, one of these GB conductive paths transforms into a CF. The first principle calculations of the GB structures describe a specific atomic property of the GB (high density of oxygen vacancies) responsible for electron transport through the GB and estimate the values of the electronic parameters of the vacancies. Using these parameter values, the TAT simulations of the MIM cell leakage current through the electrically active vacancies are compared to the experimental data. This TAT electron transport description is then used to calculate the temperature profile around the current path and map the oxygen vacancy generation during forming, which provides insight into the characteristics of the resulting CF filament.

In turn, the CF characteristics determine the temperature profile during reset and, consequently, the region of the CF that is more effectively reoxidized. The TAT simulations demonstrate that the HRS current is controlled by electron transport through a thin dielectric barrier formed by the reoxidized portion of the CF tip next to the forming cathode. A subsequent transition to the LRS is described as an electric field-driven breakdown of this thin dielectric barrier.

In this study, both measurements and electron transport simulations were performed in the DC mode on TiN/HfO\textsubscript{2}/TiN stacks with relatively thin (5–7 nm) dielectric layers, which represent a convenient modeling system while retaining all the essential features of CF-based RRAM structures. Here we do not consider the kinetics of the switching processes, which is critical for describing device switching under pulse operation conditions and understanding the role of the dielectric stoichiometry \textsuperscript{19,20} and electrode materials.\textsuperscript{21,22} Note that the reported experimental data indicates that the processes governing CF formation are materials- (both electrode and dielectric) and forming conditions-specific (e.g., Refs. 23 and 24). Hence, the findings presented here cannot be generalized to the variety of other dielectric materials investigated for RRAM applications\textsuperscript{1} for which the forming process is not necessarily BD-driven.

Conduction through grain boundaries

Atomic force microscopy (AFM)-related techniques such as conductive AFM (CAFM) \textsuperscript{(Refs. 25–29) and Kelvin probe force microscopy (KPFM) allow for nanometer-resolved characterization of the electrical and topographical properties of oxides. Since these techniques work on bare dielectric surfaces, the conductive tip acts as the metal electrode of a nanometer dimension MOS capacitor with an \textasciitilde 300 nm\textsuperscript{2} area.\textsuperscript{30} With CAFM, a current flows through the structure when the (probe tip-sample) system is biased, so that the electrical properties as well as the surface topographical features of devices can be evaluated with nanometer resolution. When a KPFM is used, topography and tip-sample contact potential difference (CPD) maps can be obtained within a similar resolution. Since the CPD signal depends on the amount of charge trapped in the dielectric,\textsuperscript{31} KPFM can be used to obtain information on the charge trapping properties of the structure.

In this study, a gate dielectric stack consisting of a polycrystallized (by a 1000 C anneal) 5 nm atomic layer deposited (ALD) HfO\textsubscript{2} film and a 1 nm SiO\textsubscript{2} interfacial layer grown on a Si substrate was investigated with KPFM and CAFM. Topographical and CPD images of this structure, Figs. 1(a) and 1(b), respectively, correlate perfectly. A clustering of generated defects near existing ones, thus increasing the current through these defects as well as increasing the local temperature, etc. A clustering of generated defects may lower the energy barrier for the dissociation of the nearby oxygen in the surrounding lattice, which further enhances defect generation around the BD region (the proposed BD process will be reported in detail elsewhere).
with scanning tunneling microscopy (STM) analysis. The current-voltage (I-V) characteristics collected with the CAFM over the GBs and grain bodies are qualitatively different, with a much higher current increase at some onset voltages, suggesting some differences in the charge transport mechanisms through the GB and grain body.

Continuous AFM surface scanning over the same area gradually increases the current through the GBs (with no increase over the grains) and eventually generates BD spots, which are centered at the GBs. The BD spot diameters derived from the CAFM measurements are ~10 nm, which is significantly larger than the GB cross sections estimated to be ~1–2 nm. This indicates that CF formation during BD consumes a significant portion of the adjacent HfO₂ grains, depending on the BD hardness. The maximum of the spatial distribution of the current through the BD spot matches the GB position. Therefore, the CAFM, KPFM, and STM results point to the HfO₂ GBs as the morphology precursors of a BD site.

Structural characteristics of defects at the grain boundaries

To delineate the GB atomic features responsible for its electrical characteristics, several possible GB configurations in the monoclinic HfO₂ were calculated by the first principles (see the methodology in Ref. 34). Empirical pair potentials and the METADISE code were used to generate several prospective low energy boundary configurations. These structures were then optimized using periodic density functional theory to determine the structure with the lowest formation energy. Several possible GB configurations were explored; Fig. 2(a) shows the calculated structure of an oxygen-deficient (101) twin GB in m-HfO₂, which is constructed by mirroring the (101) surface. The boundary is modeled using a 34.37 × 10.35 × 7.87 Å supercell that contains 240 ions. The properties of this boundary are calculated using the PAW method and the PBE functional (VASP code), including plane waves with energies up to 400 eV.

The formation energy of the (101) grain boundary is calculated to be 0.60 J m⁻², which is low since there are no under-coordinated ions and little strain (~2%). The effective diameter of this GB structure is found to be on the order of 1 nm. The lowest energy diffusion paths of oxygen vacancies toward and along the GB are indicated in Fig. 2(b). The activation energy for doubly positively charged O-vacancy diffusion in the bulk of the grain is calculated to be 0.65–0.70 eV (in agreement with Ref. 40), while the energy for neutral vacancies is much higher, ≥ 2 eV, effectively immobilizing them at room temperature. Vacancies are found to be more stable at the GB than in the bulk of a grain, by up to 0.8 eV, suggesting that substantial vacancy segregation at the GBs might be expected. Figure 3 shows the formation of a conductive sub-band, which represents the overlapping localized gap states associated with the unoccupied d-states of the Hf atoms at the GB, projected onto the substrate interface plane. The transfer of injected electrons between these states constitutes a conductive path along the GB, as discussed in connection with the KPFM and CAFM measurements. Thus, positively charged O vacancies segregating at the GBs could be expected to contribute to the electron transport through the HfO₂ dielectric stack.
Leakage current through grain boundaries: Multiphonon TAT model

The current through the HfO₂ dielectric is modeled by applying the statistical multiphonon TAT model, which is shown to accurately reproduce both temperature and voltage dependencies of the current through the dielectric. This model considers that defects assisting the electron transport are positively charged oxygen vacancies located at the GBs, consistent with the KPFM data.

This model takes into account both TAT and direct tunneling (DT) contributions. The electron DT current is computed through the Tsu-Esaki formula; the tunneling probabilities are calculated within the Wentzel-Kramer-Brillouin (WKB) approximation.

The TAT current accounts for the contribution from each defect to the charge transport. The multiple-trap percolation paths are automatically accounted for Ref. on the basis of the randomly generated defect positions and energies.

The electron TAT current through a percolation path, \( I_p \), depends on its slowest trap,

\[
I_p = \frac{q}{\tau_{e,\text{max}} + \tau_{e,\text{max}}},
\]

where \( q \) is the electron charge; \( \tau_{e,\text{max}} \) and \( \tau_{e,\text{max}} \) are the time constants associated with the electron capture and emission by and from the slowest trap in the conductive path, respectively.

In this model, the electron-phonon coupling is taken into account to correctly describe the capture and emission events. As shown in Fig. 4(a), the electron tunneling into a trap is associated with the release of some energy \( \Delta E = n \hbar \omega_0 \) to the lattice, \( n \) being the number of phonons involved, which are assumed to be single frequency \( \omega_0 \) optical phonons. Similarly, the emission of electrons from a trap is associated with the absorption of the energy \( \Delta E = n \hbar \omega_0 \) from the lattice, \( n \) being the number of phonons.

The capture and emission time constants are calculated by accounting for every phonon energy contribution (indexes \( m \) and \( n \)),

\[
\tau_{e,j}^{-1} = \sum_m N_j \left( E_j \right) f_j \left( E_j \right) C_j \rho_T \left( E_j, m, n \right),
\]

where \( E_j \) is either the conduction/valence band edge or the energy level of the \( j \)th trap; \( N_j \) and \( f_j \) are the density of states and Fermi-Dirac occupation probability either at the cathode/anode or at the \( j \)th trap; \( \rho_T \) is the tunneling probability; \( C_j,m \) and \( E_m,n \) are the trap capture and emission rates, which include the carrier-phonon interaction,

\[
C_j,m = c_0 L \left( m \right),
\]

\[
E_m,n = c_0 L \left( n \right) \exp \left( -\frac{n \hbar \omega_0}{kT} \right),
\]

where \( c_0 \) is a constant depending on the electric field and on the capture cross section of the trap (4); \( L \) is the multiphonon transition probability depending on the carrier-phonon coupling.

The physics of the carrier-phonon coupling can be illustrated using the configuration coordination diagram in Fig. 4(b), which depicts the total energy of the electronic and vibrational states of the system as a function of the generalized lattice coordinate \( Q \), which accounts for the displacement of lattice atoms to their new equilibrium positions by charge trapping/detrapping. This lattice displacement (i.e., lattice relaxation) introduces an effective barrier to the electron trapping/emission process, \( E_{\text{capt}} \) in Fig. 4(b), which, within the high temperature approximation, can be expressed as

\[
E_{\text{capt}} = \left| \left( E_F - E_T \right) - E_R \right|^2 / 4 \cdot kT \cdot E_T
\]

\( E_F \) is the electrode Fermi level; \( E_T \) and \( E_R \) are the thermal ionization energy and the relaxation energy of the trap, respectively.

The multiphonon TAT model successfully reproduces pre-stress I-V temperature dependency by considering the trap ionization and relaxation energy values matching those calculated for the positively charged oxygen vacancies (\( V^+ \)) at the GBs, see Fig. 5. The electron transport is thus found to proceed through the \( (V^+ + e^{-} \rightarrow V^0 \rightarrow V^+ + e^{-}) \) process. In these simulations, positively charged oxygen vacancies, randomly generated using a Monte Carlo technique, were considered to be uniformly distributed along a one-dimensional GB, within the energy range of \( E_T = 1.7–2.7 \text{ eV} \).

Electrically active defects: Positively charged oxygen vacancies

The TiN/HfO₂/TiN capacitors with 6 nm ALD HfO₂ dielectric were subjected to a constant voltage stress (CVS) in wide voltage/temperature ranges. The CVS was applied either continuously or with periodic interruptions, see Fig. 6. The current during the CVS was found to decrease by a factor of whereas it decreased (although not completely recovering) when the voltage was removed.

To identify the processes responsible for the current change with and without stress interruption, currents were simulated in the voltage/temperature range using the
multiphonon TAT current model. Consistent with the results in the previous section, the current at any given moment during the CVS is described by the \((V^+ + e \rightarrow V^0 \rightarrow e + V^+)\) process. Hence, the current increase should be due to an increase in the \(V^+\) concentration at the GBs. However, the activation energy for the \(V^+\) diffusion in hafnium is too high (>1.5 eV), preventing the \(V^+\) defects in the bulk of the dielectric from moving around and accumulating at the GB under the given experimental conditions. The activation energy for the CVS current increase extracted from the measurement data (Fig. 6) is 0.7 eV, which matches the value obtained from \textit{ab initio} calculations for the \(V^2^+\) diffusion,\(^{34,40}\) typically in hafnia with the relatively high concentration of \(10^{18}–10^{19}/\text{cm}^3\). This indicates that the rate at which the leakage current increases is controlled by the accumulation of \(V^2^+\) defects at the GBs. By capturing an injected electron, \(V^2^+\) can be converted (reversibly) into the electrically active \(V^+\) state. The conversion of \(V^2^+ + e \rightarrow V^+\) (as well as \(V^+\) deactivation) is described in the same theoretical framework of the multiphonon TAT model.

Simulations successfully reproduced the observed current features, Fig. 7, by accounting for two processes running simultaneously: (1) an increase in the concentration of the TAT precursor defects at the GBs, due to random diffusion and segregation of \(V^2^+\) at GBs and (2) the conversion (activation) of these defects into the \(V^+\) state, which can contribute to the TAT current. The activation process is found to be very fast; in addition, at a given bias, a dynamic population of \(V^+\) traps having a certain distribution across the dielectric film is created. Due to a strong dependency of the electron trapping/detrapping probabilities on the energy difference between the trap and incoming electron and on the electron tunneling distance, see Eqs. (2)–(3), the population of the electrically active \(V^+\) traps decreases closer to the anode electrode due to the increased deactivation probability, \(V^+ \rightarrow V^2^+ + e\). Since the characteristic lifetime of a vacancy in the \(V^+\) state is generally much longer than the time to transfer an electron through \(V^+\) (\(10^{-5}–10^{-13}\) s, depending on the \(V^+\) trap distance from the cathode), the \(V^+\) traps were found to effectively support the measured leakage current. On the other hand, the current decrease after voltage has been removed indicates a recovery of some of the \(V^2^+\) defects, which is described as \(V^+\) deactivation.

Forming process: Physical model

After determining that the leakage current in the metal oxide dielectric is controlled by the oxygen vacancies, the next step is describing how this leakage path may transform into a permanent conductive filament in the forming process. Such a description should explicitly include the way new oxygen vacancies are generated around the current path. With a greater concentration of O vacancies, the leakage path develops metallic properties. The \textit{ab initio} calculations demonstrated (Fig. 3) that this eventually results in the formation of the metal-rich CF with ohmic electrical characteristics. Note
that the forming process under the ramp-up voltage conditions used for the forming operation here differs from the constant voltage stress under low voltage discussed above. During CVS, the gradual increase in current in the initial phase of the stress is controlled by diffusion and accumulation of the vacancies at the GBs rather than the generation of new vacancies.

The creation of a new oxygen vacancy (as opposed to accumulation of the pre-existing vacancies created during fabrication) is energy-consuming, requiring chemical bonds to be broken (oxygen dissociation) and the oxygen ion to be repositioned outside the CF. The final position of this oxygen ion, which strongly affects the energy barrier in the new vacancy formation process, will be discussed elsewhere. Within the effective activation energy description,\textsuperscript{47,48} bond breakage is considered to be assisted by the local electric field and the temperature around the bond, Fig. 8. The corresponding defect generation rate, \( G \), can be approximated by an Arrhenius expression,\textsuperscript{48}

\[
G = G_0 \exp\left[-\frac{(E_A - bF)}{k_B T}\right], \tag{7}
\]

where \( F \) is the applied field; \( T \) is the local temperature near the specific oxygen atom; \( k_B \) is the Boltzmann’s constant; \( G_0 \) is a rate constant, which depends on intrinsic material properties, that is derived by fitting the experimental data; \( b \) is the bond polarization factor.\textsuperscript{47} \( E_A \) is the effective activation energy of the defect generation, which represents the energy barrier for the relocation of the oxygen ion from its regular position to some energetically favorable site, probably outside the CF. Note that the energy of this oxygen vacancy generation, which leaves the sub-lattice of the metal ions mostly intact, may be expected to be less than the energy required for material melting, \( k_B T_{\text{mol}} \) which can result in a truly amorphous material structure. The value of \( E_A = 4.4 \) eV was estimated from the BD voltages measured on the MIM HfO\(_2\) structures under a fast voltage ramp-up condition by comparing the defect generation rate \( G(F, T) \), which is assumed to be constant, at different temperatures and voltages. Despite such high activation energy, vacancy generation in high-k materials can be very effective since the local field at the metal-oxygen bonds is dramatically enhanced (compared to the applied external field) due to strong bond polarizability, as reflected in the high value of \( b \sim 90 \) e\( \text{Å} \) in HfO\(_2\).\textsuperscript{47}

Besides the dependency on applied voltage, the defect generation rate (7) increases at elevated temperatures. The strong dependency of defect generation on temperature singles out the conductive paths in the dielectric as regions of preferential defect generation. Indeed, electron transfer through a trap causes the atomic displacements of the surrounding lattice, which is associated with the phonon emission/absorption, thus modulating the temperature around the current path. Higher local temperatures associated with the leakage currents along the GBs cause BD spots to be located at the GBs, as observed in CAFM.\textsuperscript{32} An additional factor contributing to a higher defect generation rate around the conductive path is that the generation of an oxygen vacancy distorts the surrounding lattice, which lowers the activation energy for new oxygen vacancy generation at nearby sites. The exponential dependence of the defect generation rate on the electric field and temperature combined with the dependency of temperature and activation energy on the local defect density forms a positive feedback loop driving the runaway defect generation process (e.g., the dielectric breakdown transient). Higher defect density at a given location leads to higher local temperature/lower activation energy, which in turn enhances the defect generation rate. The runaway process was simulated using the TAT current and heat dissipation descriptions discussed below (these simulations do not account for the reduction of activation energy due to high local defect density).

### Simulations of the forming process

The above physical model describing the dielectric breakdown transient was implemented in a program, the schematic flow diagram of which is shown in Fig. 9. The model is able to reproduce the current increase observed during the ramped-up voltage stress; it also accounts for the temperature increase induced by the power dissipation and the temperature-assisted generation of new defects.

The current simulation is based on the multiphonon TAT charge transport model that is used to calculate the current through GB defects. The energy released/adsorbed associated with the phonon exchange with the lattice is calculated for each trap and subsequently used to calculate the 3 D map of the power dissipated around the conductive path in the dielectric, \( P \). The power dissipation map is then used to calculate the temperature map in the RRAM cell, see Fig. 10. To this end, the 3 D Fourier’s heat flow equation is solved by properly accounting for the heat flow between a GB and the surrounding lattice and electrodes,\textsuperscript{33}

\[
P(x, y, z) = k_{\text{TH}} \cdot \nabla^2 T(x, y, z), \tag{8}
\]

where \( k_{\text{TH}} \) is the thermal conductivity of the HfO\(_2\). The heat transport through the metal electrode at the GB interfaces to

![FIG. 8. (Color online) Schematic of the dependency of the chemical bond energy of the generalized coordinate of the atomic configuration forming the bond. Under the applied electric field, \( F_{\text{ox}} \), the probability of the bond breakage process increases due to reduction of the activation barrier, \( E_a \), caused by the bond polarization, \( bF_{\text{ox}} \), and by temperature population of the excited vibrational states.](image-url)
the external environment is modeled by two finite thermal resistances as described in Ref. 49.

The generation of vacancies during forming is simulated using the iteration loop shown in Fig. 9. The current flow was simulated at each voltage step during the applied voltage ramp, starting with some initial random distribution of the vacancies along the GBs, and the corresponding power dissipation map and resultant temperature profile were calculated. Based on the latter, the random generation of new vacancies was calculated using the defect generation rate \( G \) (Eq. (7)) and Monte Carlo techniques. The entire sequence of the TAT current calculation was then repeated until a predefined current compliance value was reached. Here we ignored the vacancy diffusion, which, according to our estimates, is expected to be negligible during the BD transient time.

An example of the simulation of the temperature and oxygen vacancy defect maps during forming is shown in Fig. 10 for TiN/5 nm HfO2/TiN cells. The calculated temperature profile depends strongly on the boundary conditions for metal electrodes, which are the primary thermal sink (thermal conductivity of the HfO2 dielectric is negligible in comparison), specifically for anodes, closer to which most heat dissipation occurs. In this simulation, the electrodes are connected to the ambient at room temperature through finite thermal resistances. In actual devices, the top electrode is separated from the outer surface by a thick layer of passivating dielectric, which dramatically increases the electrode temperature and, subsequently, the density of generated oxygen vacancies in the dielectric.

Simulations of the forming process in a 5 nm HfO2 RRAM cell using a set of 30 randomly selected initial trap distributions along the GB allowed a statistical distribution of the forming voltage values, Fig. 11, to be generated, showing excellent agreement with the measured data.
filament (out of 1000 simultaneously simulated parallel paths), Fig. 11. Figure 12 shows the defect density profile across the dielectric thickness in the GB region calculated at different times during forming, which demonstrates that the conduction filament is essentially non-uniform across the dielectric thickness. During electron transfer, more energy is released at the traps in the dielectric region close to the anode, which is caused by a greater energy release during each electron trapping event. This is due to a greater mismatch between the trap energy ($E_T$) and the energy of the injected electrons ($E_F$) under the applied electric field ($F$), $E_F - E_T - x \cdot F$, where $x$ is the distance from the cathode. Since the relaxation energy of the $V^+$ vacancies in HfO$_2$ is rather high, the traps with greater energy mismatch values (e.g., larger distances $x$ from the cathode) experience smaller electron trapping barriers ($E_{capt}$ in Eq. (6)) and, subsequently, higher trapping/detrapping rates. The greater energy dissipation closer to the anode electrode raises the temperature in that region, which boosts the local generation of oxygen vacancies, so that a larger portion of the surrounding dielectric is converted into a metal-rich filament.

Note that intrinsic limitations of the TAT description, which is adequate when the vacancies are not too dense and the trapped electrons can still be considered localized, preventing it from simulating the filament formation all the way through the ohmic conduction state. Nevertheless, the region closer to the anode can still be expected to exhibit lower resistance because of the higher density of oxygen vacancies, see Fig. 12. Such filament non-uniformity strongly affects RRAM switching operations.

**Reset process: The filament change from low to high resistance**

The DC bipolar operations of the TiN/5 nm HfO$_2$/TiN RRAM cells, Fig. 13, were simulated using the multiphonon TAT model, extended to include the ohmic charge transport in the CF region where a metallic sub-band was formed by high oxygen vacancy concentration. Simulation of the current in the pre-forming state assumed that electron transport is assisted by the $V^+$ traps, as was demonstrated in Fig. 7.

After forming, the current is purely ohmic, as follows from the linear I-V dependency at low voltages during reset, Fig. 14, as well as an observed increase of the resistance with temperature. This indicates that the electrodes are connected by a continuous conductive path of an essentially metallic nature since even a rather thin remaining dielectric layer would result in deviation from the metallic behavior (as shown below in the case of ≤1 nm dielectric barrier, see Fig. 15). Thus, negative oxygen ions, released during the filament formation, which might get stored near the interface with the forming anode, must reside outside of the filament region. This out-diffusion of the oxygen ions toward the surrounding dielectric is driven by the density gradient of the “free” oxygen within and outside of the filament (assisted by high temperature in the filament region during the forming).

At higher voltage values, the current starts to deviate from the ohmic behavior, which can be attributed to an
assisting the electron transport, with the resistance profile in Eq. (10) is in good agreement with the measured current, Fig. 14. Here the resistance of the R(z) is given by:

\[ R(z) = R_0(z)(1 + \frac{z}{T - T_0}) \]

where \( R_0(z) \) and \( z \) are the CF resistance at ambient temperature, which depends on the position across the dielectric \( z \) and the resistance temperature coefficient, respectively.

Based on the simulation results, the forming process is assumed to create a non-uniform filament exhibiting higher resistance next to the forming cathode electrode. This can be due to a non-uniform profile of either the density of the oxygen vacancies or CF geometric cross section through the CF length. The composition and geometry profiles of the CF in LRS are not currently known; therefore, both options should be examined, although they have the same effect on the CF resistance. The cone-shape CF was considered in Ref. 12; here we focus on the effects of a non-uniform composition.

To simulate a current during reset, in the first approximation the CF resistance is considered to be inversely proportional to the density of the oxygen vacancies \( n_T(z) \) assisting the electron transport,

\[ R_0(z) = \frac{R_p}{n_T(z)} \]

where \( R_p \) is a fitting parameter. The reset current simulated with the resistance profile in Eq. (10) is in good agreement with the measured current, Fig. 14. Here the resistance of the CF near the forming cathode is assumed to be \( 6 \times 10^4 \) greater than the resistance next to the forming anode electrode based on the ratio of the calculated defect densities in the corresponding regions of the filament (Fig. 12). Since the applied voltage during reset mostly drops across the higher resistance portion of the filament next to the forming cathode, more power is dissipated in this region (see insert of Fig. 14), effectively heating it. The temperature profile \( T(x) \) along the CF length calculated self-consistently within the framework of the model is shown in the inset of Fig. 14. This result is similar to that obtained under the assumption of a varying CF cross section along its length.[12] At high reset voltage values, the calculated temperature is found to be high enough to reoxidize a CF tip (as well as potentially some partial CF body oxidation); this is accompanied by a dramatic increase in current noise, indicating that the CF is undergoing structural changes. Such reoxidation is expected to form a dielectric barrier at the CF interface with the electrode, subject to an adequate supply of oxygen ions field-driven to the anode electrode.

**CF characteristics in HRS**

To understand the major factors controlling HRS properties, we simulated the HRS I-V temperature dependency. To reproduce the experimental data, one must assume the existence of dielectric barrier next to the forming cathode electrode, in accordance with the suggestion that the CF tip is oxidized during reset; the electron transport through this barrier limits the current flow. In the simulations, the required barrier thickness is 0.9–1.0 nm, and the barrier band offset and \( k \) value match those of HfO2. The current temperature dependence is controlled by the multiphonon TAT process through the \( 2 \times 10^2 \) cm\(^{-3} \) dense \( V^+ \) defects in the barrier layer, Fig. 15. The energy of the contributing traps exhibits a wide distribution, 1.7–2.7 eV, matching the values obtained for the electrically active \( V^+ \) traps, which control the current in the fresh HfO2, Fig. 5. A wide distribution of the trap parameters should be expected because of the disordered structure of the grain boundaries (HfO2 retains a nanocrystalline structure \( \geq 1.4 \) nm thick[51]). We may conclude that the HRS is associated with the CF separated from the forming cathode electrode by a thin HfO2 dielectric barrier. The electron transport through this barrier, which includes both direct tunneling and TAT through the traps in the barrier, determines the current in the HRS.

To verify the hypothesis of preferred oxidation near the forming cathode electrode of an asymmetric filament, RRAM devices were fabricated such that a portion of the HfO2 adjacent to the forming cathode was replaced by highly sub-stoichiometric HfO\(_{2-x}\). Indeed, these devices do not reset to a HRS since the dielectric oxygen deficiency prevents a barrier from forming next to this electrode. However, reversing this stack (HfO\(_{2-x}\) away from and HfO2 adjacent to the forming cathode) makes resetting to a HRS not only possible, but highly reproducible (> \( 10^6 \) endurance cycles). The model for the HRS neglects the possibility of partial reoxidation of the entire CF body during reset. Whether such an extension of the model is warranted depends on the...
physical characterization results of the CF stoichiometry in high and low resistive states, which are currently unavailable.

To assess the factors controlling the CF parameters, three different current compliance limits, 1.2 mA, 0.16 mA, and 0.03 mA, Fig. 16, were used for the forming process. For these samples in the HRS, the TAT model was used to reproduce the measured I-V data, Fig. 17, using the energy parameters values for the V\(^+\) defects as in the simulations in Fig. 13. The barrier thickness, l\(_{\text{bar}}\), was kept constant at about 0.9 nm for all three cases. Assuming the same V\(^+\) vacancy density of \(10^{21}/\text{cm}^3\) in the barrier layers formed during reset in all three cases (the same \(\rho\) value was used to simulate the HRS currents in Fig. 15), the differences in the HRS current values can be assigned to the differences in the CF radius \(r\). This is consistent with the observations that higher currents are usually associated with larger BD spots.\(^{30}\) The extracted radius of the CF barrier for the smallest current compliance limit of 0.03 mA, \(r \approx 1\) nm, is consistent with that of the GB (\(\approx 1\)–2 nm), which the current flows through before forming. For 1.2 mA and 0.16 mA current compliance limits, the \(r\) values are estimated to be on the order of 4 nm and 2.5 nm, respectively. These radius values have been used to estimate the average resistivity \(\rho\) of the CFs in the LRS, \(\rho = l_{\text{CF}}/n_r^2\), where \(l_{\text{CF}}\) is the CF length and \(R\) is the average LRS ohmic resistance values of 0.9 kOhm, 6 kOhm, and 34 kOhm measured for the forming current compliance limits, Fig. 13. For all three cases, the average CF resistivity is estimated to be \((0.4–1.0) \times 10^{-2} \, \Omega\cdot\text{cm}\), compared to \(3 \times 10^{-5} \, \Omega\cdot\text{cm}\) for pure Hf metal and \(10^7 \, \Omega\cdot\text{cm}\) for HfO\(_2\). Thus, these data indicate that the CF in the LRS consists of highly sub-stoichiometric HfO\(_x\), the composition of which is roughly independent of the current compliance limit. This supports the assumption that the compliance limit controls the CF effective radius rather than its stoichiometry. Lower LRS currents correlate to lower HRS currents since currents in both states are linked through the CF cross section, which is defined (for a given cell structure) by the power dissipation during forming; the released power is determined by the magnitude of the forming voltage and maximum current (which may exceed a pre-set current compliance limit as a result of the current overshoot effect\(^{32}\)).

**Set process**

In the HRS, during the set process, most of the applied voltage drops over the dielectric barrier formed at the higher resistance portion of the CF during reset. The electric field profile along the filament at the set switching condition for the data in Fig. 13 (set voltage \(= 0.6\) V) was calculated, Fig. 18. Since the voltage drop along the low resistance portion of the CF (outside the barrier) is negligible, the forming voltage drops across the thin barrier layer inducing a high electric field, which exceeds the HfO\(_2\) dielectric strength of \(\sim 4–6\) MV/cm.\(^{47}\) The field-driven breakage of Hf-O bonds

---

**FIG. 16.** (Color online) The I-V characteristics of the RRAM devices (10 switching cycles) operated using the 1T1R configuration (a transistor in series with the RRAM cell controls the current compliance), which were formed with different current compliance limits.

**FIG. 17.** (Color online) The measured (symbols) and TAT simulated (lines) HRS current density in the samples formed with different compliance limits \(I_{\text{cl}}\) in Fig. 16.

**FIG. 18.** (Color online) The electric field profile along the filament axis (x-x) when the anode voltage is equal to the set voltage (= 0.6 V) in the example in Fig. 13.
associated with the dielectric breakdown results in a high current flow through the broken barrier, which triggers the process switching the memory cell to the LRS. We suggest that such a process, for which a numerical description has not been yet developed, completely dissolves the dielectric barrier and restores a continuous metal-like CF between the electrodes. The effectiveness of the barrier reduction depends on the maximum temperature in the barrier region. Under the set voltage and compliance current, the maximum temperature is determined by the CF barrier dimensions (it increases with shorter CF radii) and thermo conductive properties of the memory cell (materials and architecture). An incomplete barrier reduction may lead to higher LRS resistance, which might increase with each set/reset cycle, impairing device endurance.

**CONCLUSION**

The description proposed here for the LRS/HRS switching mechanisms of HfO2-based RRAM devices identifies the forming process as critical to defining the CF properties, which in turn control the switching characteristics of the resistive memory cell. The proposed model considers CF formation to be an abrupt process (e.g., a dielectric breakdown), which includes oxygen dissociation from the Hf-O bonds (the oxygen vacancy generation) and subsequent out-diffusion of the “free” oxygen ions out of the filament region driven by their density gradient. Consequently, from the standpoint of electron transport and corresponding power dissipation, which controls the oxygen dissociation process, the intrinsic time scale of this CF formation is in the ps range, as opposed to an electroforming formalism (which was applied to a variety of systems, e.g., TiO2, Ref. 53) that requires longer times. Following the vacancy generation, the subsequent out-diffusion of the dissociated oxygen ions may also affect the CF geometry and stoichiometry, which are defined by the temperature-electric field distribution around the conductive path at the moment of dielectric breakdown.

Based on the electrical and physical characterization and transport/atomistic modeling results from this study, the CF features critical to controlling switching in the HfO2 RRAM were identified. In a fresh state, the leakage current through the dielectric is described by the TAT conduction through the oxygen vacancy defects in their positive charge state. When the defects are characterized by a relatively high lattice relaxation energy (as for the $V^+$ defects in HfO2), the energy barrier associated with the electron trapping/detrapping tends to decrease and, accordingly, the electron transfer rate increases with greater trap distances from the injecting electrode (when the energy of the injected electron approaches the value of the trap relaxation energy). As an important consequence of such phonon-assisted electron transport, more energy dissipates closer to the anode electrode, inducing a corresponding temperature increase along the conduction path. Subsequently, during the forming operation when a sufficiently high voltage is reached, more new vacancies are generated closer to the anode electrode, assisted by the higher temperature. As a result, the forming process is found to generate a highly oxygen-deficient, essentially non-uniform (with respect to its stoichiometry and/or shape) filament with its more resistive end next to the forming cathode.

These CF characteristics determine its HRS and LRS properties. In the LRS, the CF is highly oxygen-deficient, exhibiting a metallic behavior with non-uniform resistance across the dielectric thickness. During the reset operation, the simulated temperature profile along the CF points to the possibility of effective reoxidation of the more resistive (either narrower or less oxygen-deficient) tip of the filament, thus creating a dielectric barrier layer of about 1 nm. The electron transport through this barrier, which includes both direct tunneling and trapping/detrapping through the oxygen vacancy defects in the barrier layer, determines the HRS current. During the set operation, the applied voltage drops over this barrier layer, resulting in a high electric field in the thin dielectric. The dielectric breakdown of this barrier layer, which occurs when the electric field exceeds the dielectric intrinsic breakdown strength, converts the CF into the LRS.

To summarize, while the transition to the HRS requires reaching sufficient local temperature in the more resistive portion of the filament to assure its effective oxidation, the set operation is primarily controlled by the electric field in the filament.

Since oxygen vacancies tend to accumulate at the GBs, these GBs represent a natural conduction path, which is likely to represent a weak link in the dielectric where the breakdown can occur during forming. Therefore, GBs facilitate a dielectric breakdown at lower voltages, which results in less energy dissipation ($\approx I^2V$) around the conduction path and, correspondingly, higher resistance of the resultant CFs (either due to the smaller radius or lower oxygen vacancy density). Even more important is that a lower forming voltage lowers the transient current associated with the charging of the parasitic capacitors during abrupt forming (a difficult to control component of the so-called current overshoot, which can be only partially suppressed by connecting a transistor in series with the RRAM cell52,54), further increasing CF resistivity. Higher CF resistance promotes the Joule heating required for CF reoxidation at lower reset currents, thus enabling low power cell switching. Whether GBs is the essential morphology feature for the formation of the switchable CF in hafnium-based RRAM, as dielectric oxygen deficiency is known to be,19 is not yet clear.

In addition to these CF properties, successful switching also requires sufficient oxygen for these redox processes. Since the availability of oxygen ions depends on the electric field, which directs negative oxygen ions toward/away from a CF tip, bipolar operations in transition metal oxide-based RRAMs are expected to be more favorable than a unipolar mode of operation.