Thermal environment impact on HfO_x RRAM operation: a nanoscale thermometry and modeling study

Authors

Matthew P. West[1], Georges Pavlidis[2-3], Robert H. Montgomery[4], Fabia Farlin Athena[5], Muhammad S. Jamil[3], Andrea Centrone[2], Samuel Graham[3,6], Eric M. Vogel[1,]*

Affiliations

[1] Georgia Institute of Technology, School of Materials Science & Engineering, Atlanta, Georgia 30332, United States

[2] Nanoscale Device Characterization Division, Physical Measurement Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland 20899, United States

[3] University of Connecticut, Department of Mechanical Engineering, Stamford, Connecticut 06901, United States

[4] Georgia Institute of Technology, George W. Woodruff School of Mechanical Engineering, Atlanta, Georgia 30332, United States

[5] Georgia Institute of Technology, School of Electrical and Computer Engineering, Atlanta, Georgia 30332, United States

[6] University of Maryland, Department of Mechanical Engineering, College Park, Maryland 20742, United States

[*] Corresponding author, Eric M. Vogel, eric.vogel@mse.gatech.edu

Abstract

As the demand for computing applications capable of processing large datasets increases, there is a growing need for new in-memory computing technologies. Oxide-based resistive random-access memory (RRAM) devices are a promising candidate for such applications because of their industry readiness, endurance, and switching ratio. These analog devices, however, suffer from poor linearity and from asymmetry in their analog resistance change. Various reports have found that the temperature in RRAM devices increase locally by more than 1000 K during operation. Therefore, temperature control is of paramount importance for controlling their resistance. In this study, scanning thermal microscopy is used to map the temperature of Au/Ti/HfO_x/Au devices at a steady power state and to measure temperature dynamics of the top electrode above the filament location during both resistive switching loops and voltage pulsing. These measurements are used to verify the thermal parameters of a multiphysics finite elements model. The model is then used to understand the impact of the thermal conductivities and boundary conductances of the constituent materials on the resistance change during the first reset pulse in RRAM devices. It is found that the resistance change can be reduced significantly when the temperature in the titanium capping layer is reduced. We find that the greatest temperature reduction, and therefore, the lowest resistance change in the device is afforded by capping layers with increased thermal conductivities. This work links thermal properties to the resistance change in RRAM devices providing critical insights for engineering devices with improved switching dynamics.

Introduction

Many emerging technologies rely on understanding, manipulating, and making predictions based on large data sets using various machine learning and artificial intelligence algorithms.[1-5] With traditional computing architectures, these data intensive applications suffer from the von Neumann bottleneck; i.e., the slow shuttling of data back and forth from memory to the central processing unit.[6,7] Technologies that can compute in memory, such as filamentary oxide-based resistive random-access memory (RRAM), have been proposed to overcome this bottleneck and have attracted considerable attention.[6,8] Electrical biasing across their metal-insulator-metal structure generates a conductive bridge, a filament, across the electrodes via the creation of oxygen vacancies in the metal oxide insulator.[9-15] The negatively charged oxygen ions migrate to the positively charged electrode by drift, diffusion, and thermophoresis.[14,16-20] When the bias polarity is reversed, oxygen ions are pushed back into the filament increasing the resistance. The transition from the low resistance state (LRS) to the high resistance state (HRS) is referred to as 'reset' and the transition from the HRS to the LRS is referred to as 'set.' Notably, these devices can achieve a continuum of resistance states between the LRS and HRS based on the polarity, amplitude, and duration of applied voltage pulses.[13,21-24] This resistive switching behavior has been observed in many oxides, including HfO_2[11,15,25,26], TiO_2[9,27,28], and Ta_2O_5/TaO_2[13,14,22,29,30]. This study focuses on a device consisting of a sub-stoichiometric HfO_x ($x \approx 1.85$) active layer and a titanium capping layer due to its known stable operation and industry readiness.[10,12,17,26,31] For example, switching retention times exceeding 10 years ($> 10^5$ cycles) and resistance ratios $> 10^3$ Ω have been reported for Ti/HfO_x-based devices.[10,12] The titanium capping layer, in direct contact with the HfO_x, acts as an oxygen reservoir, i.e., storing oxygen ions that are removed from the filament and returning them back to the filament upon inversion of the bias polarity.[10,12,20] Though this study focuses on

a Ti/HfO_x active material, the results presented here extend to other filamentary oxides that experience large localized temperature increases during operation.

Many challenges need to be overcome to harness the full potential of filamentary RRAM analog neuromorphic devices. These challenges include poor linearity of the resistance change with respect to the number of applied pulses, high variability, and asymmetry of the resistance changes induced by pulses of opposite polarity.[32-34] These non-idealities make it difficult to achieve specific analog resistance states, limiting the computational accuracy and efficiency of neuromorphic circuits.[32-34] Often a large resistance change is observed from the first reset pulse when the device is in the low resistance state. After just a few pulses, the resistance change saturates. This large resistance change likely occurs because of the larger current achieved during the first pulse, which leads to the largest Joule heating and temperature rise.[34] Since the magnitudes of all three ion migration mechanisms (drift, diffusion, and thermophoresis) increase exponentially with temperature; the largest temperature rise also leads to the highest resistance change.[18] Numerous works have studied the effect of the filament thermal environment on the LRS to HRS resistance transition. These studies include introducing thermal enhancement layers/electro-thermal modulation layers[34-36], changing the thermal conductivities (k_{th}) of the electrodes[37], and varying the k_{th} of the substrate[17,18,26]. Past experiments have shown that the thermal environment impacts the resistance change of these devices, however, characterizing the filament temperature during operation remains a challenge. For example, transient thermoreflectance imaging (TTI) was used to characterize and compare RRAM experimental and simulated temperature rises.[18,38] However, TTI typically requires averaging of many images, precluding its use for characterizing transients in RRAM devices in which the resistance change occurs during the course of a single pulse. Similarly, Raman thermometry has measured steady state heating, but its spatial resolution

is insufficient to map localized heating from filamentary devices.[39] The AFM-based Scanning Joule Expansion Microscopy (SJEM), was used to estimate temperature rises in TaO_x devices with nanoscale resolution but SJEM measurements are hardly quantitative, as converting the SJEM signal to temperature is far from trivial.[40] The high spatial and temperature resolution of scanning thermal microscopy (SThM) was recently used to measure the temperature rise of the top electrode (ΔT_{TE}) right above the filament in resistive switching devices during biasing.[39,41-44] Such SThM measurements on HfO_2-based devices with single layer graphene electrodes revealed a filament (≈ 13 nm diameter) reaching a temperature of approximately 1300 °C.[39] Importantly, the authors suggested using low thermal conductivity substrates to reduce the heat transfer between neighboring devices to avoid unintended resistance changes.[39]

Here, expanding on prior works, we developed a comprehensive electro-thermal model to evaluate the impact of the thermal conductivities and of the thermal boundary conductances (G) of the constituent materials and interfaces on RRAM devices' resistance change during analog pulsing. The ΔT_{TE} from the model is compared to the SThM measured ΔT_{TE} during *in-situ* analog pulsing in Ti/HfO_x RRAM devices. In this study we focus on the effects of the thermal environment on first reset pulse since it leads to the largest temperature rise and to the largest resistance change. The results of this parametric analysis suggest that a capping layer with high thermal conductivity and a high G between the capping layer and the top electrode reduce the magnitude of the resistance change from the first pulse. To ensure the attainability of many resistance states in analog devices it is important to avoid large resistance changes from a single pulse. Pairing *in-situ* thermal measurements with a simulation that relate temperature and resistance change, this work offers new insights for optimizing the operation of RRAM devices by engineering the devices' thermal environment.

Experimental Methods

Device fabrication

Hafnium oxide-based neuromorphic devices were fabricated on a silicon substrate with a 300 nm thick SiO_2 surface layer. The bottom electrode was patterned with a negative photoresist using maskless UV photolithography. After development, a titanium adhesion layer (≈ 10 nm) and a gold film (≈ 70 nm) were deposited by electron beam evaporation (6.7×10^{-4} Pa or lower). Lift-off of the photoresist in acetone yielded the patterned bottom electrode. The electrodes were rinsed in acetone, methanol, and isopropyl alcohol and further cleaned with an O_2 plasma descum (8 Pa, 150 W, and 8.3×10^{-7} $m^3 \times s^{-1}$ O_2 flow rate) right before loading the devices into a preheated (250 °C) atomic layer deposition (ALD) chamber. A 5 nm HfO_x layer was deposited with 55 ALD cycles of alternating water vapor for 0.06 s and tetrakis(dimethylamido)hafnium (TDMA-Hf) for 0.25 s. The thickness of HfO_x was confirmed by ellipsometry after deposition. X-ray photoelectron spectroscopy (XPS) revealed a sub-stoichiometric composition ($x \approx 1.85$) of the HfO_x layer. Next, the top electrode was patterned using the same procedure as the bottom electrode. After development, a titanium capping layer (≈ 5 nm) and the gold electrode (≈ 150 nm) were deposited by electron beam evaporation (6.7×10^{-4} Pa or lower). Lift-off in acetone was used to remove the remaining photoresist and obtain the final device pattern. A representative cross-sectional transmission electron microscopy (TEM) image is shown in Figure 1a and an optical micrograph of the 10×10 μm^2 crossbar structure is shown in Figure 1b. Additional details for this fabrication process can be found in Basnet *et al.*[17] and West *et al.*[26].

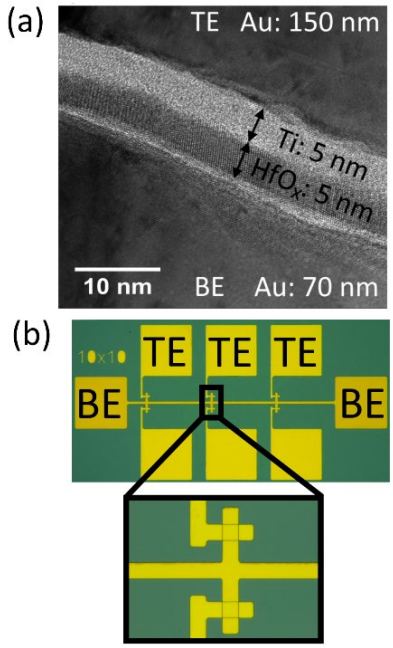


Figure 1. (a) Cross-sectional TEM image of an as-deposited device stack. This confirms the deposition of a ≈ 5 nm titanium capping layer and ≈ 5 nm HfO_x active area. (b) Device structure used to create RRAM crossbars. This design ensures no electrical sneak paths are possible due to the electrical isolation of each device.

Electrical Testing

A positive voltage sweep from 0 to 3.5 V was applied to the top electrode to form the filament while grounding the bottom electrode. A 0.1 mA current compliance ensured no damage to the device. After forming, negative voltage sweeps were applied starting at -0.1 V and gradually increased to -1.5 V in increments of 0.1 V. After reaching the final reset voltage, the device was switched back into the LRS with a positive voltage sweep from 0 to 1.2 V using a current compliance of 1 mA. The device was cycled between the HRS and LRS with -1.5 V and 1.2 V, respectively, to ensure that the device could switch repeatably before thermal and voltage pulse measurements. Pulsing was conducted with the devices starting in the LRS (≈ 550 Ω). Several

negative voltage pulses (1 ms pulse width) were applied to the top electrode to gradually reset the device. Though 1 ms is a relatively long pulse for analog RRAM devices, such long pulse widths mimic biological processes and are used for some neuromorphic implementations.[22,29,45-48]

Scanning Thermal Microscopy

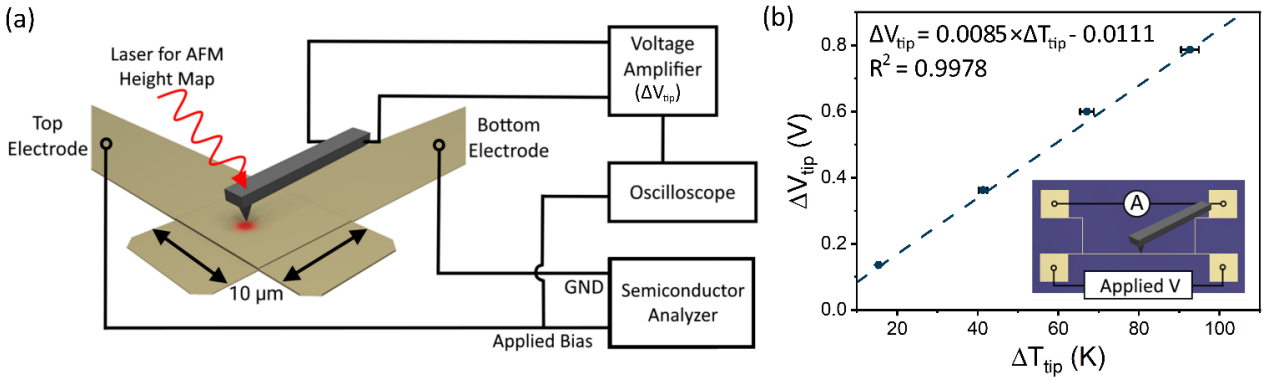


Figure 2. (a) Simplified wiring diagram of the SThM setup. The voltages applied to the device are recorded by the oscilloscope enabling direct correlation with ΔV_{tip}. (b) A four-point probe gold microheater was first placed in an oven to extract its temperature coefficient of resistance (TCR). Next the microheater was used to calibrate the ΔV_{tip} with respect to the temperature of the tip ($\Delta T_{tip} = \Delta T_{microheter\text{-}center}$). The horizontal error bars represent the propagated uncertainty of the ΔT_{tip} derived from the standard error of the averaged microheater resistance and the linear fit uncertainty of the TCR. The vertical error bars represent the standard error of the averaged ΔV_{tip} (with 95 % confidence intervals). A finite element model was used to correct systematic errors associated to heterogeneous heat distribution along the length of the calibrating wire (see the Finite Element Thermal Model for SThM Calibration section in the SI) and the calibration was carried out using the calculated temperature in the center of the heater.

SThM was used to measure with nanoscale resolution i) steady state temperature maps, ii) transient temperatures at selected locations, and iii) to correlate applied voltages to the measured currents and temperatures (Figure 2a). Our custom SThM setup consist of a custom modified

AFM[49], commercially available SThM probe and SThM amplifier (30 kHz bandwidth)[50], a commercially available oscilloscope (200 MHz bandwidth and 2 GS/s sampling rate) and a commercially available semiconductor analyzer. A custom ceramic package was developed to enable wire bonding of the neuromorphic devices and electrical biasing during scanning. Commercially available silicon SThM cantilevers[50] (500 μm long and 50 μm wide) characterized by a 50 nm diameter hollow SiO_2 tip with thermocouple embedded close to the tip apex were used to sense the temperature rise in the sample while in contact with the top electrode.

The thermocouple voltage was amplified (1000x) with a SThM amplifier. For steady state temperature mapping, the amplified SThM voltage was input to an atomic force microscopy (AFM) controller and recorded simultaneously with the AFM height (i.e., topography) channel. This way surface morphological changes (if any) that may be caused by device degradation could be easily detected. After locating the filament position in the SThM map (see for example Fig 2a), the tip was positioned above the filament region to measure, the sample temperature dynamics *in-situ* during set/reset sweeps and pulse biasing. This was accomplished using an oscilloscope triggered by the output bias of the semiconductor analyzer, that also recorded the device bias concurrently.

The SThM tips were calibrated using ≈ 200 nm thick gold microheaters on a SiO_2 (≈ 300 nm)/Si substrate (same as the device substrate). Notably the microheater/tip and top electrode/tip interfaces are made of the same materials. The microheater (1000 μm x 5 μm with two contact pads on each side) enables precise four-point measurements (inset of Figure 2b) of the gold microheater average resistance. First, the gold microheater was placed in an oven and its temperature coefficient of resistance (TCR) was obtained by determining the device resistance as a function of temperature (see SI Figure S1). Next, the microheater was used to calibrate the SThM

tip as follows. The microheater temperature was controlled precisely by using the TCR obtained previously and measuring the heater resistance while supplying a current to it. The SThM probe calibration was obtained by positioning the probe tip at the center of the microheater by measuring the probe voltage (ΔV_{tip}) as a function of the heater ΔT. A linear fit using a Monte Carlo method was used to extract the tip calibration coefficient and its associated uncertainty (Figure 2b).

Since the heater resistance measured as described above is related to a device-wide spatially averaged temperature while the SThM is a local measurement of the temperature in the center of the heater, such calibration may suffer from a systematic errors due to temperature gradients along the microheater length.[51] Therefore a thermal finite element model (FEM) was developed to quantify the temperature gradient along the microheater (see SI Figure S2). The maximum discrepancy between the temperature calculated at the center of the microheater (the spot measured by the SThM probe) and average temperature (based on electrical resistance) was calculated to be 2.41 %. This systematic error was corrected by using the FEM calculated temperatures in heater center in Fig 2b in place of the heater average temperatures. Additional efforts to reduce the measurement uncertainty included positioning the cantilever perpendicular to the microheater (this orientation minimized external indirect heating of the probe tip via heating of the cantilever). Lastly, to ensure the tip does not act as a heat sink when in contact with the surface, the metal heater resistance was monitored before and after engaging the tip onto the surface. No change was detected when engaging the tip, likely due to the tip's low thermal conductivity (similarly, no change in the RRAM device current was observed when engaging the tip).

Multiphysics model

The model self consistently solves the partial differential equations for the conservation of oxygen vacancies (Equation 1), conservation of current (Equation 2), and conservation of thermal energy (Equation 3) using commercially available multiphysics software.[18]

$$\partial n_v/\partial t + \nabla \cdot (v_v n_v) = \nabla \cdot (D_v \nabla n_v) + \nabla \cdot (S_v D_v n_v \nabla T) \qquad \textbf{Equation 1}$$

$$\nabla \cdot (\sigma \nabla \psi) = 0 \qquad \textbf{Equation 2}$$

$$\partial T/\partial t = (1/\rho c_p)\, \nabla \cdot (k_{th} \nabla T) + \sigma/\rho c_p\, (\nabla \psi)^2 \qquad \textbf{Equation 3}$$

In Equation 1, n_v [m^{-3}] represents the concentration of oxygen vacancies, v_v [$m \times s^{-1}$] is the oxygen vacancy drift velocity, D_v [$m^2 \times s^{-1}$] is the diffusion coefficient, S_v [K^{-1}] is the Soret (or thermophoresis) coefficient, and T [K] is temperature. In Equation 2, σ [$S \times m^{-1}$] represents electrical conductivity and ψ [V] is the voltage potential. Lastly, in Equation 3, ρ [$kg \times m^{-3}$] is the density, c_p [$J \times kg^{-1} \times K^{-1}$] is the heat capacity, and k_{th} [$W \times m^{-1} \times K^{-1}$] is the thermal conductivity. In the filament region of the device, the thermal and electrical conductivity of the material are dependent on n_v according to Equation S1 and S2 of the SI and are not varied in this study. The diffusion coefficient, drift velocity, and thermophoresis coefficient equations used for the filament are given in Equation 4-6.

$$D_v = 0.5 a^2 f \exp(-E_a/k_b T) \qquad \textbf{Equation 4}$$

$$v_v = af \exp(-E_a/k_B T) \times \sinh(qa\psi/k_B T) \qquad \textbf{Equation 5}$$

$$S_v = -E_a/k_B T^2 \qquad \textbf{Equation 6}$$

In these equations, a [m] represents the hopping distance for a vacancy, f [Hz] is the hoping frequency, E_a [J] is the activation energy for vacancies, k_B [$J \times K^{-1}$] is the Boltzmann constant, and

q [C] is the charge of an electron. This FEM multiphysics model is unable to simulate the forming process, therefore, an initial condition of the filament must be assumed. Here a filament with an inverted conical shape and a diameter varying along its length from 6 nm to 10 nm was used, consistent with previous work.[52-55] A cross-sectional image of the initial vacancy distribution is shown in Fig S3. The model also assumes that the device is radially axisymmetric with respect to the center of the filament. The only variables that are changed in this study are the thermal conductivities of various materials and the thermal boundary conductances between the layers. In the model, the interface of the bottom of the substrate with the environment is fixed to room temperature. This boundary was found to have little impact on both the simulated device temperature and the resistance change, thus fixing it to room temperature is a reasonable approximation. All other environmental boundaries of the device assume a convective heat flux with still air using a heat transfer coefficient of 5 $W \times m^{-2} \times K^{-1}$.[56] The values of all other material properties are listed in Table S2. More detailed explanation of this electro-thermal model can be found in the study by Pahinkar *et al.*[18].

Results and Discussion

Scanning thermal microscopy was used to locate the filament, perform *in-situ* thermal measurements of ΔT_{TE} (top electrode temperature) during voltage sweeps and voltage pulsing, and to calibrate a comprehensive electro-thermal model. After the filament formation step and multiple cycling between the LRS and HRS, the device was held at a steady state power of 1 mW while scanning the SThM tip. Note that in these conditions, the voltage is too low to induce resistance changes in the device. A representative SThM map (Figure 3a) shows that heat generation is localized to one small filament or filament area, with a maximum ΔT_{TE} of only 2.79 ± 0.08 K. The temperature rise in this device is modest because the filament is buried under a 150 nm thick top

electrode and much of the heat is dissipated through the substrate.[17,18,26] Simulations under the same conditions (1 mW steady state power) show a very good agreement with the experimental data (see Figure 3a).

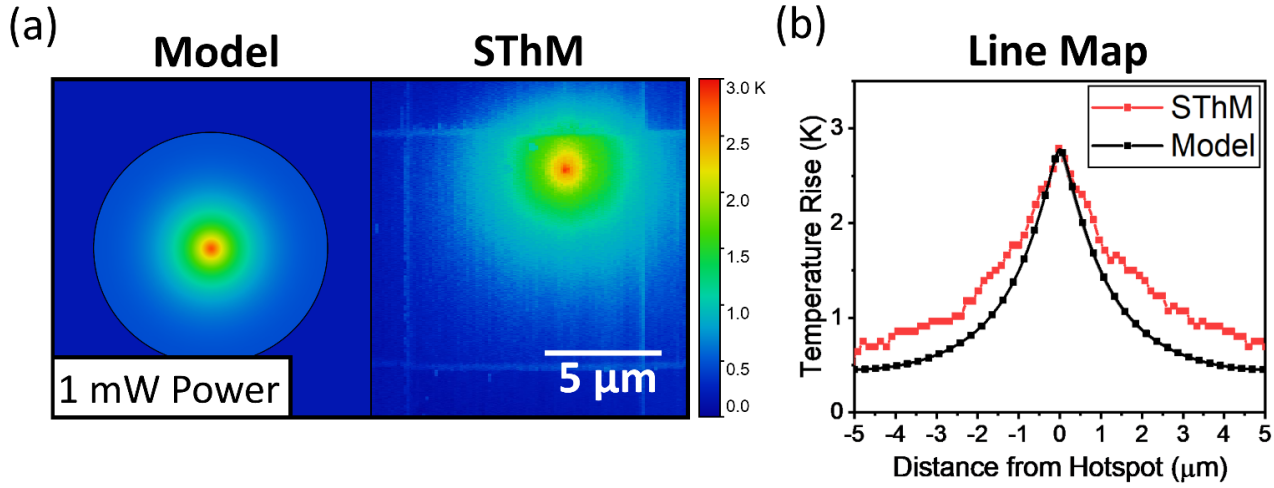


Figure 3. (a) Maps of the top electrode temperature obtained with the multiphyics model (left) and SThM (right) showing good agreement when a constant 1 mW of power goes through the filament. (b) Top electrode temperature profile passing through center of the hot spot for simulation and SThM.

The model captures the lateral heat dissipation reasonably well, although the full width at half maximum (FWHM) of the line profile across the hot spot does not exactly match the experimental data. In contrast to the perfectly centrosymmetric hot spot of the model, the measured hot spot is slightly asymmetric and has a larger FWHM, possibly because of the asymmetric heat transfer that results from the formation of the filament near the edge of the device. The relevant thermal parameters that yield such agreement are listed in Table 1.

Materials	Thermal Conductivity ($W \times m^{-1} \times K^{-1}$)
Au Top & Bottom Electrode	310[57]
Ti Capping Layer	4[58]
Filament	10 - 23
Non-Filament HfO_x	2[59]
Substrate	138[17]
Interfaces	**Thermal Boundary Conductance ($GW \times m^{-2} \times K^{-1}$)**
Ti Capping/ Au Top Electrode	100
Au Bottom Electrode/ Substrate	80

Table 1. Thermal properties used in the model to obtain good agreement with the experiment. The filament thermal conductivity varies with the number of vacancies (Equation S1 of the SI) but are assumed to not vary with temperature.

Next, the SThM tip was positioned directly on the hot spot (i.e., above the filament) and held in contact with the sample while the device was cycled between the LRS and HRS to verify that the filament survived the temperature mapping procedure and still exhibits resistive switching. This allows tracking of the top electrode temperature (ΔT_{TE}) *in-situ,* directly above the filament during both voltage sweeping and pulsing enabling direct correlation with the current and applied voltage which were recorded with the same oscilloscope. Figure 4 shows a typical set/reset hysteresis loop with the measured current (left axis) and ΔT_{TE} (right axis). As expected, the temperature rise is strongly correlated with the current. During negative biasing the device starts in the LRS. As the magnitude of the voltage increases, ΔT_{TE} also increases due to the rising current and Joule heating through the device. At ≈ -0.5 V, the device starts to reset since the filament has reached a critical temperature and the bias is large enough to push oxygen ions from the titanium capping layer into the filament by drift. This causes an increase in the resistance (i.e., decrease in

the current), which lowers ΔT_{TE} even as the magnitude of the bias is increased. At the start of positive biasing the temperature rise is very small because the current is very low in the HRS. The low ΔT_{TE} measured from 0 V to ≈ 0.8 V shows that the measured tip voltage is not affected by the voltage applied to the device. At ≈ 0.8 V, there is an abrupt increase in the current and ΔT_{TE} which in turn promotes drift, diffusion, and thermophoresis of oxygen vacancies into the filament.

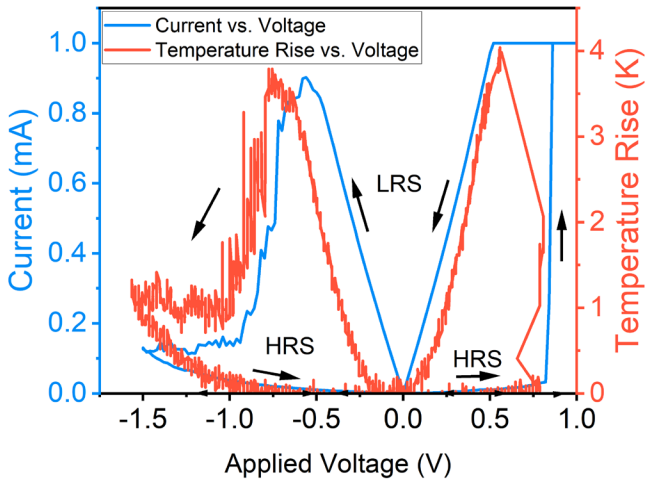


Figure 4. Set/reset hysteresis loop of a Ti/$HfO_{1.85}$ RRAM device with simultaneous measurement of the applied voltage, current, and local ΔT_{TE}. The temperature rise is well correlated to the current because the heat is primarily generated in the filament via Joule heating. The negligible ΔT_{TE} when the device is in the HRS proves that the SThM tip voltage is not impacted by the voltage applied to the device. A current compliance of 1 mA is used on the set side to ensure the device does not permanently breakdown.

Voltage pulsing was initiated with the devices in the LRS (≈ 550 Ω). The first pulse is defined as the pulse after the device is put back in the LRS. Figure 5 shows a typical ΔT_{TE} measured for the first -0.7 V and -1 V pulses on the same device. Between these two pulses, the device is put back in the LRS with a 0 V to 1.2 V sweep (1 mA current compliance). It is noted that due to a

limitation of the pulsing equipment the voltage pulse is not a square pulse, and the peak voltage is only held for a short time, see figure 5 (right axis). As expected, the temperature rise during the -0.7 V pulse (2.16 K ± 0.06 K) is smaller than the ΔT_{TE} obtained for the -1 V pulse (2.62 K ± 0.07 K). Additionally, while temperature rises monotonically during the -0.7 V pulse, for the -1 V pulse the temperature starts decreasing while the voltage is still increasing. We interpret this as follows. During the larger -1 V pulse, the resistance increases quickly due to the higher peak temperature which drastically increases the drift, diffusion, and thermophoresis of oxygen vacancies since these processes have an exponential dependence on the temperature (Equation 4 - Equation 6). In addition, RRAM devices are very sensitive to the applied voltage via the exponential relationship with ionic drift (Equation 5). These two mechanisms act together, quickly increasing the resistance (lowering the current), leading to a reduction in ΔT_{TE} with time. Consequently, the -0.7 V pulse caused the filament to increase in resistance by only ≈ 285 Ω, while the -1 V pulse caused a resistance change of ≈ 5120 Ω.

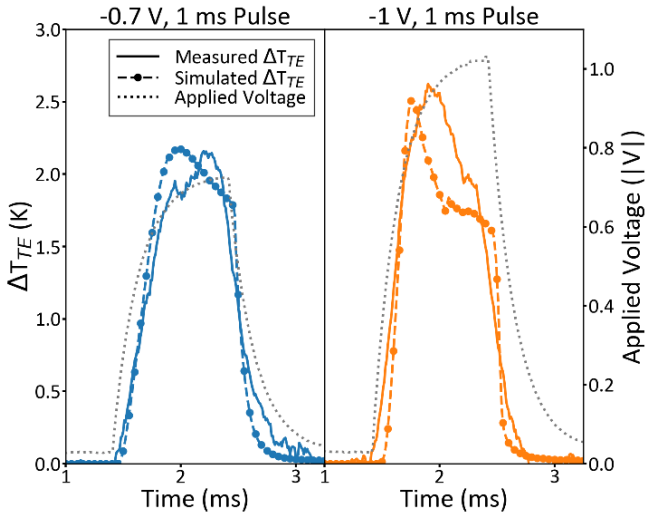


Figure 5. Comparison between the ΔT_{TE} measured in the experiment on top of the filament and the ΔT_{TE} obtained by the model during voltage pulsing under identical conditions (-0.7 V, left and

-1 V, right). The model predicts the maximum temperature rise accurately and captures the temperature dynamic trends.

The voltage pulses from the experiment were applied to the model using the same materials parameters that yielded good agreement for the steady state comparison in Figure 3. Similarly, Figure 5, shows good agreement between the modeled and measured ΔT_{TE} for both the -0.7 V and -1 V amplitude pulses. For the -0.7 V pulse, the ΔT_{TE} predicted by the model is 2.18 K, very close to the measured value (≈ 2.2 K). The maximum simulated ΔT_{TE} for the -1 V pulse is 2.52 K compared to the measured value of ≈ 2.6 K, well within the uncertainty of the measurement (≈ 0.07 K). Besides the peak temperatures, the modeled ΔT_{TE} follows reasonably well the measured temperature dynamics. That is, the -1 V pulse amplitude causes a fast reduction in temperature during the pulse compared to the lower amplitude pulse. There is, however, a noticeable reduction in ΔT_{TE} for the -0.7 V pulse in the simulation, which is not observed experimentally (Figure 5). This can be attributed to the model's overestimation of the resistance change (≈ 1250 Ω) for the -0.7 V pulse compared to just ≈ 285 Ω in the experiment. The estimated resistance change (≈ 4830 Ω) has a much better agreement with the experiment (≈ 5120 Ω) for the -1 V pulse. Given the large variability of RRAM devices observed in practice, capturing the exact resistance changes in the model is very difficult. The simulation, however, estimates the ΔT_{TE} well and yields the correct trends for the resistance change. This allows for systematically assessing the impacts of various material thermal properties and interfaces on the estimated trends in resistance change.

The same simulation can also be used to estimate the relationship between the measured ΔT_{TE} and the rise in filament temperature. The plot in Figure 6 compares three different estimated filament region temperatures to the estimated ΔT_{TE} for powers ranging from 17 μW to 737 μW.

The average filament temperature is defined as the average temperature within the 5 nm filament radius and 5 nm filament height. Similarly, the average titanium capping layer temperature is defined as the average temperature above the filament in the same volume (depicted in Figure 6 inset). To understand the heat transfer between the filament and the capping layer, the average temperature of their interface is also plotted. According to the linear fit in Figure 6, a ΔT_{TE} of just ≈ 2 K to 3 K (depending on biasing condition), corresponds to a temperature rise of the filament/capping layer interface between ≈ 883 K and ≈ 1330 K, respectively. For example, ΔT_{TE} measured by SThM for the -1 V pulse (2.62 K ± 0.07 K) correspond to a rise of the filament/capping layer interface by 1160 K ± 31 K. This filament temperature is similar to other estimates in previous RRAM studies.[18,25] We can estimate the rise in filament temperature with the fits in Figure 6. For material parameters validated by the steady state experiments (Figure 3) the average temperature of the capping layer and filament are nearly identical. The relationship between these two temperatures and the resistance change will be discussed in a later section of this paper. These large locally confined temperature rises greatly impact the mobility of oxygen vacancies by means of drift, diffusion, and thermophoresis. Therefore, determining which thermal properties of the materials and interfaces in the device affect the device resistance change the most is critical to engineer and improve the device characteristics, particularly with regards to reducing the amount of resistance change from the first reset pulse.

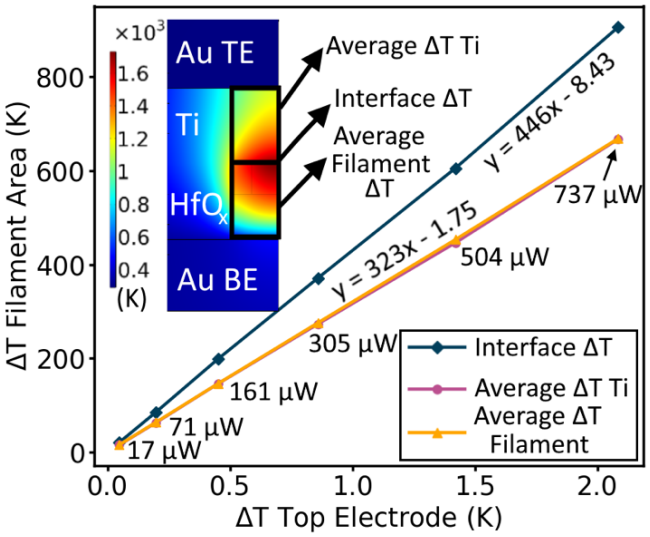
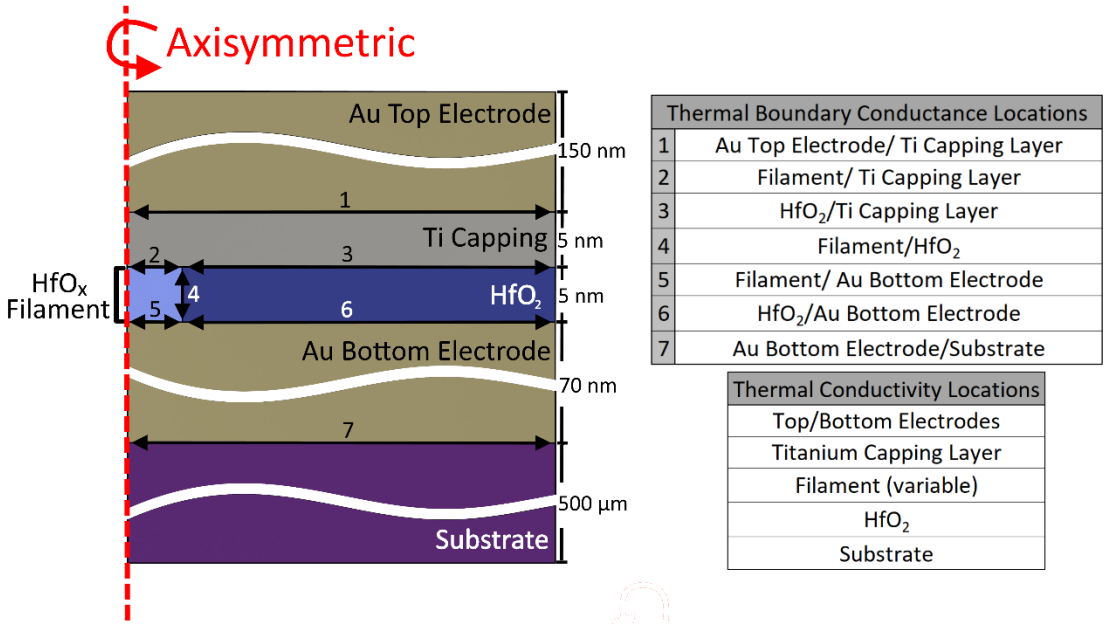


Figure 6. Average capping layer temperature (pink), filament temperature (yellow), and temperature of their interface (blue) compared to the ΔT_{TE} as a function of the device applied power. The inset shows the localized temperature generation in the HfO_x filament area and defines the average capping layer, filament, and interface temperatures. This graph demonstrates how a relatively low top electrode temperatures relates to large increases in the filament temperature deep into the device structure.

The model was used to independently understand the impact of thermal boundary conductances and thermal conductivities on the predicted resistance change of HfO_x-based filamentary RRAM. There are seven thermal boundaries in the simulated device geometry and five material thermal conductivities as illustrated in Figure 7.



Thermal Boundary Conductance Locations	
1	Au Top Electrode/ Ti Capping Layer
2	Filament/ Ti Capping Layer
3	HfO_2/Ti Capping Layer
4	Filament/HfO_2
5	Filament/ Au Bottom Electrode
6	HfO_2/Au Bottom Electrode
7	Au Bottom Electrode/Substrate

Thermal Conductivity Locations
Top/Bottom Electrodes
Titanium Capping Layer
Filament (variable)
HfO_2
Substrate

Figure 7. Simulated model geometry annotated with thermal conductivity and boundary locations varied in the parametric studies. The thermal conductivity of the filament is variable with the concentration of oxygen vacancies and is not changed in any of the parametric studies (Equation S1). Each of these thermal conductivities and interfacial thermal conductances is changed independently while all other parameters are constant according to Table S2.

The thermal conductivity of the titanium capping layer was varied between 0.5 $W \times m^{-1} \times K^{-1}$ and 17 $W \times m^{-1} \times K^{-1}$ spanning the range between TiO_2 and titanium metal, respectively.[58,60] The electrode thermal conductivity was varied between 50 and 300 $W \times m^{-1} \times K^{-1}$ to represent a wide variety of metals, including gold and gold thin films.[57] The thermal conductivity of non-filamentary HfO_2 was varied from 0.1 $W \times m^{-1} \times K^{-1}$ to 2 $W \times m^{-1} \times K^{-1}$ which encompass the typical range for HfO_2 thin films and substoichiometric films.[59] The substrate thermal conductivity was varied between 1 $W \times m^{-1} \times K^{-1}$ and 300 $W \times m^{-1} \times K^{-1}$ to include many possible substrate materials.[18] Since not all G values have been measured previously, the G of each interface was varied over 5 orders of magnitude from 1×10^5 $W \times m^{-2} \times K^{-1}$ to 1×10^9 $W \times m^{-2} \times K^{-1}$, i.e., within the ranges typical

of metal/oxide interfaces and encompassing the values assumed in other SThM RRAM works.[39,59,61] For all of these parameters, the resistance was measured with a -0.1 V, 1 ms read pulse before and after the application of a square -0.7 V, 1 ms reset pulse.

Figure 8a shows that the thermal conductivity of the capping layer ($k_{th,capping}$) has the largest impact on the calculated resistance change, ranging from 215 Ω for $k_{th,capping}$ = 17 $W \times m^{-1} \times K^{-1}$ (typical for bulk titanium metal) to 4473 Ω for 0.5 $W \times m^{-1} \times K^{-1}$ (typical for TiO_2). Over this range, the thermal conductivity of the capping layer has the effect to reduce the average filament temperature rise from 1283 K ($k_{th,capping}$ = 0.5 $W \times m^{-1} \times K^{-1}$) to 938 K ($k_{th,capping}$ = 17 $W \times m^{-1} \times K^{-1}$), a 27 % reduction. Furthermore, the average capping layer temperature decreases even more significantly from 973 K to 581 K, a 40 % reduction, when $k_{th,capping}$ is increased. The impact of the capping layer temperature and filament temperature with respect to resistance change will be discussed later. The data in Figure 8a suggests that the thermal conductivity of the electrodes and of the substrate have a marginal impact on the device resistance change, compared to the impact of $k_{th,capping}$. Though the electrodes and substrate have significant impact on the overall heat dissipation from the device, they have marginal impact on the local temperature in the regions where the migration of ions occurs. Additionally, Figure 8a shows that the thermal conductivity of the material laterally next to the filament, in this case the non-filament HfO_2, has a large impact on the resistance change. Although increasing the thermal conductivity of this layer would reduce the resistance change, only few materials are known to have high thermal conductivity, but a very low electrical conductivity. The latter is a strict requirement for RRAM devices because if the electrical conductivity is too large, the leakage current would prevent the formation of the filament in the active (HfO_2) layer.

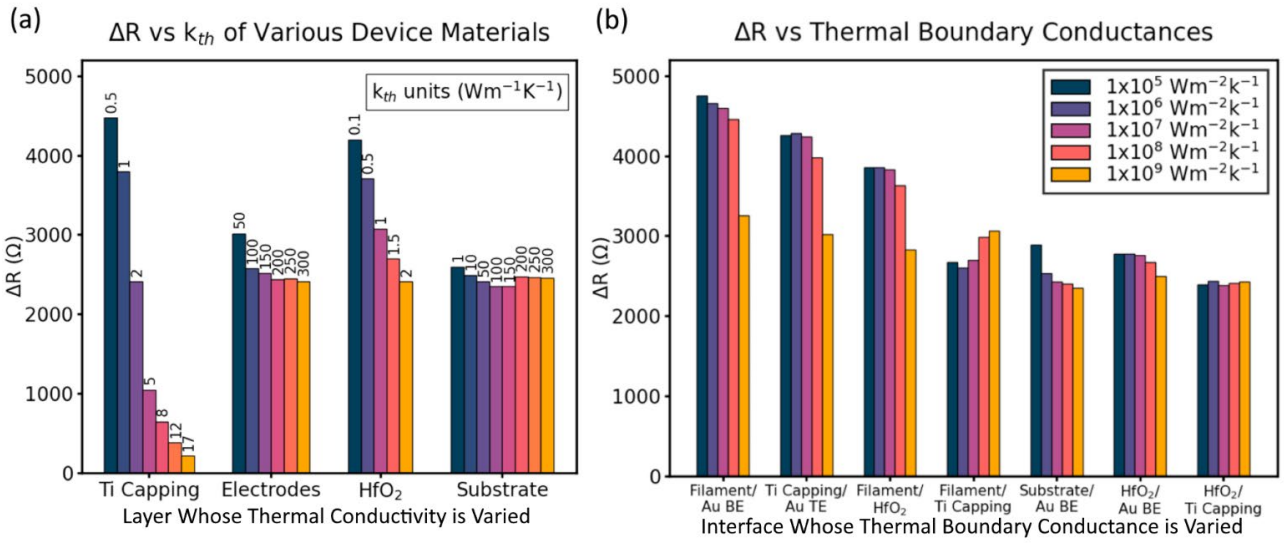


Figure 8. (a) Change in RRAM device resistance due to a single -0.7 V, 1 ms pulse as a function of varied thermal conductivities of the capping layer, electrodes, surrounding HfO_2, and substrate. The thermal conductivity of the capping layer has the greatest impact on the resistance change. (b) Change in RRAM device resistance due to a single -0.7 V, 1 ms pulse as a function of the thermal boundary conductance (G) magnitude between the device interfaces. Results suggest that a low G for the filament/Au BE interface has the greatest impact on the resistance change and that the Ti Capping/Au TE interface is the only interface that is not in direct contact with the filament that impacts the resistance change.

Of the seven boundary locations, only three have significant impacts on the device resistance change: the filament/Au bottom electrode interface, the titanium capping layer/Au top electrode interface, and the filament/surrounding HfO_2 interface (Figure 8b). The only influential boundary not in contact with the filament is the titanium capping layer to gold top electrode. A low G between these layers traps the heat within the capping layer, slowing dissipation into the top electrode. Since the capping layer is the device oxygen reservoir and the T strongly affects the migration of oxygen ions, maintaining a high temperature in the capping layer induces large

changes in the device resistance. Most of the temperature rise is generated by the filament, not the capping layer. Therefore, the transfer and dissipation of heat to/from the capping layer crucially affects its temperature rise. If the heat from the filament is allowed to dissipate quickly, the capping layer temperature may be too low for any significant migration of oxygen ions, leading to small resistance changes.

A closer look at the varied thermal boundary conductance of the filament and capping layer interface ($G_{\text{filament/capping}}$) is needed to further understand the importance of the capping layer temperature. $G_{\text{filament/capping}}$ is the only interface in Figure 8b that shows an increase in the device resistance with an increase in G. This can be explained by $G_{\text{filament/capping}}$ having a proportionally small effect on the filament temperature rise (from ≈ 1398 K to ≈ 1293 K) but a much larger effect on the capping layer temperature rise (from ≈ 354 K to ≈ 714 K), which in turn has a large effect on resistance change (Figure 9a). Figure 9b shows the vacancy distributions after the reset pulse for the cases with $G_{\text{filament/capping}}$ of 1×10^5 $\text{W}\times\text{m}^{-2}\times\text{K}^{-1}$ and 1×10^9 $\text{W}\times\text{m}^{-2}\times\text{K}^{-1}$. For low $G_{\text{filament/capping}}$, a higher concentration of vacancies remains in the top portion of the filament and accumulates just above the filament in the titanium layer. In contrast, the high $G_{\text{filament/capping}}$ allows for higher temperatures in the titanium layer, facilitating the migration of vacancies further from the filament. Remarkably, this suggests that reducing the capping layer temperature and confining the heat to the filament can reduce the large resistance change typically observed from the first reset pulse. Furthermore, by identifying the thermal properties with the greatest effect on the device resistance change, this work also identifies the thermal properties that should be measured with greatest possible accuracy to aid in the engineering of RRAM devices. Therefore, this work calls for the deveopment of novel measurement techniques capable of measuring thermal properties with nanoscale spatial resolution and high accuracy.[62]

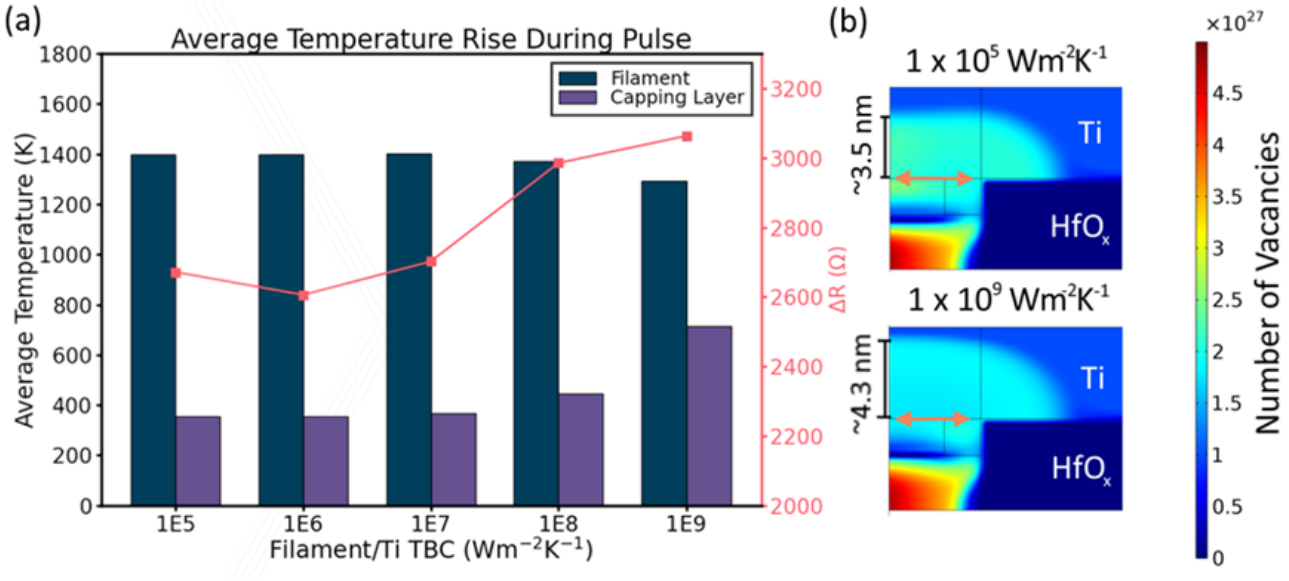


Figure 9. (a) Average filament temperature, capping layer temperature, and resistance change for varying $G_{\text{filament/capping}}$. The resistance change does not start increasing until the heat is able to transfer from the filament to the capping layer. (b) Vacancy distribution after the -0.7 V, 1 ms simulated pulse for $G_{\text{filament/capping}}$ of 1×10^5 W×m^{-2}×K^{-1}and 1×10^9 W×m^{-2}×K^{-1}. With a low $G_{\text{filament/capping}}$ (top) a higher concentration of vacancies ($\approx 2.5 \times 10^{27}$ compared to $\approx 2 \times 10^{27}$) is trapped near the top of the filament leading to a reduced change in resistance.

Conclusions

In this work we analyze the effects of the thermal environment on the magnitude of resistance changes in oxide-based RRAM devices during the first reset pulse. SThM was used to measure the top electrode temperature during steady state measurements providing the basis for reasonable estimates of the thermal conductivities and thermal boundary conductances in our RRAM device. The estimated thermal properties were validated by comparing a comprehensive electro-thermal model with top electrode temperature measured *in-situ* with SThM during voltage

pulsing. Further, parametric studies of thermal conductivities and thermal boundary conductances in HfO_x neuromorphic devices identify the thermal conductivity of the capping layer is the most important thermal property in these devices. A low thermal conductivity capping layer, a low thermal boundary conductance between the capping layer and the top electrode, and a high thermal boundary conductance between the filament and the capping layer all promote large resistance changes in the device due to the resulting higher capping layer temperature. Therefore, the novel guidance provided by this work is to thermal engineer oxide-based RRAM devices with a low capping layer temperature to mitigate undesirable large resistance changes during the first reset pulse. Furthermore, by identifying which thermal properties most critically affect the device behavior, we also highlight proper targets for novel, nanoscale, low uncertainty thermal property measurements of materials and interfaces that may change between resistance states.

Acknowledgements

This work was supported by the Air Force Office of Scientific Research MURI entitled, "Cross-disciplinary Electronic-ionic Research Enabling Biologically Realistic Autonomous Learning (CEREBRAL)" under Award No. FA9550-18-1-0024. This work was performed in part at the Georgia Tech Institute for Electronics and Nanotechnology, a member of the National Nanotechnology Coordinated Infrastructure (NNCI), which is supported by the National Science Foundation (No. ECCS-2025462). This material is based upon the work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE- 1650044.

Supplementary material

Supporting material contains S1-S3, Table S1-S2, and Equations S1-S3. More information on the SThM calibration procedure and multiphysics model are available in the supporting information.

References

[1] P. P. Shinde and S. Shah, 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 1-6 (2018).

[2] J. T. McCoy and L. Auret, Minerals Engineering 132, 95-109 (2019).

[3] M. W. Libbrecht and W. S. Noble, Nat Rev Genet 16, 321-332 (2015).

[4] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, Computational and Structural Biotechnology Journal 13, 8-17 (2015).

[5] J. Stilgoe, Soc Stud Sci 48, 25-56 (2018).

[6] X. Zou, S. Xu, X. Chen, L. Yan, and Y. Han, Science China Information Sciences 64, 160404 (2021).

[7] N. K. Upadhyay, H. Jiang, Z. Wang, S. Asapu, Q. Xia, and J. Joshua Yang, Advanced Materials Technologies 4,, 1800589 (2019).

[8] E. J. Fuller, S. T. Keene, A. Melianas, Z. Wang, S. Agarwal, Y. Li, Y. Tuchman, C. D. James, M. J. Marinella, J. J. Yang, A. Salleo, and A. A. Talin, Science 364, 570-574 (2019).

[9] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, Nature Letters 453, 80-83 (2008).

[10] H. Y. Lee, P. S. Chen, T. Y. Wu, Y. S. Chen, C. C. Wang, P. J. Tzeng, C. H. Lin, F. Chen, C. H. Lien, and M.-J. Tsai, 2008 IEEE International Electron Devices Meeting, 1-4 (2008).

[11] G. Bersuker, D. Veksler, D. M. Nminibapiel, P. R. Shrestha, J. P. Campbell, J. T. Ryan, H. Baumgart, M. S. Mason, and K. P. Cheung, Journal of Computational Electronics 16, 1085-1094 (2017).

[12] S. Z. Rahaman, Y. D. Lin, H. Y. Lee, Y. S. Chen, P. S. Chen, W. S. Chen, C. H. Hsu, K. H. Tsai, M. J. Tsai, and P. H. Wang, Langmuir 33, 4654-4665 (2017).

[13] Y. Jeong, S. Kim, and W. D. Lu, Applied Physics Letters 107, 173105 (2015).

[14] S. Kim, S. Choi, and W. Lu, ACS Nano 8, 2369-2376 (2014).

[15] G. Bersuker, D. C. Gilmer, D. Veksler, P. Kirsch, L. Vandelli, A. Padovani, L. Larcher, K. McKenna, A. Shluger, V. Iglesias, M. Porti, and M. Nafría, Journal of Applied Physics 110, 124518 (2011).

[16] L. Vandelli, A. Padovani, L. Larcher, G. Broglia, G. Ori, M. Montorsi, G. Bersuker, and P. Pavan, 2011 International Electron Devices Meeting, 17.5.1-17.5.4 (2011).

[17] P. Basnet, D. G. Pahinkar, M. P. West, C. J. Perini, S. Graham, and E. M. Vogel, Journal of Materials Chemistry C 8, 5092-5101 (2020).

[18] D. G. Pahinkar, P. Basnet, M. P. West, B. Zivasatienraj, A. Weidenbach, W. A. Doolittle, E. Vogel, and S. Graham, AIP Advances 10, 035127 (2020).

[19] S. Menzel, U. Böttger, M. Wimmer, and M. Salinga, Advanced Functional Materials 25, 6306-6325 (2015).

[20] S. Dirkmann, J. Kaiser, C. Wenger, and T. Mussenbrock, ACS Appl Mater Interfaces 10, 14857-14868 (2018).

[21] L. Gao, P.-Y. Chen, and S. Yu, IEEE Electron Device Letters 36, 1157-1159 (2015).

[22] X. Li, H. Wu, B. Gao, W. Wu, D. Wu, N. Deng, J. Cai, and H. Qian, Nanotechnology 27, 305201 (2016).

[23] F. F. Athena, M. P. West, P. Basnet, J. Hah, Q. Jiang, W.-C. Lee, and E. M. Vogel, Journal of Applied Physics 131, 204901 (2022).

[24] J. Hah, M. P. West, F. F. Athena, R. Hanus, E. M. Vogel, and S. Graham, Journal of Materials Science 57, 9299-9311 (2022).

[25] S. R. Bradley, G. Bersuker, and A. L. Shluger, Journal of Physics: Condensed Matter 27, 415401 (2015).

[26] M. P. West, P. Basnet, D. G. Pahinkar, R. H. Montgomery, S. Graham, and E. M. Vogel, Applied Physics Letters 116, 063504 (2020).

[27] H. Y. Jeong, Y. I. Kim, J. Y. Lee, and S. Y. Choi, Nanotechnology 21, 115203 (2010).

[28] D. Carta, I. Salaoru, A. Khiat, A. Regoutz, C. Mitterbauer, N. M. Harrison, and T. Prodromakis, ACS Appl Mater Interfaces 8, 19605-19611 (2016).

[29] Z. Wang, M. Yin, T. Zhang, Y. Cai, Y. Wang, Y. Yang, and R. Huang, Nanoscale 8, 14015-14022 (2016).

[30] Y. D. Zhao, P. Huang, Z. Chen, C. Liu, H. T. Li, W. J. Ma, B. Gao, X. Y. Liu, and J. F. Kang, 2015 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), 285-288 (2015).

[31] Z. Fang, X. P. Wang, J. Sohn, B. B. Weng, Z. P. Zhang, Z. X. Chen, Y. Z. Tang, G. Q. Lo, J. Provine, S. S. Wong, H. S. P. Wong, and D. L. Kwong, IEEE Electron Device Letters 35, 912-914 (2014).

[32] M. J. Marinella, S. Agarwal, A. Hsia, I. Richter, R. Jacobs-Gedrim, J. Niroula, S. J. Plimpton, E. Ipek, and C. D. James, IEEE Journal on Emerging and Selected Topics in Circuits and Systems 8, 86-101 (2018).

[33] S. Agarwal, S. J. Plimpton, D. R. Hughart, A. H. Hsia, I. Richter, J. A. Cox, C. D. James, and M. J. Marinella, 2016 International Joint Conference on Neural Networks (IJCNN), 929-938 (2016).

[34] W. Wu, H. Wu, B. Gao, P. Yao, X. Zhang, X. Peng, S. Yu, and H. Qian, 2018 IEEE Symposium on VLSI Technology, 103-104 (2018).

[35] W. Wu, H. Wu, B. Gao, N. Deng, S. Yu, and H. Qian, IEEE Electron Device Letters 38, 1019-1022 (2017).

[36] Y. Xi, J. Tang, B. Gao, F. Xu, X. Li, Y. Lu, H. Qian, and H. Wu, IEEE Transactions on Electron Devices 69, 4254-4258 (2022).

[37] S. Kim, S. J. Kim, K. M. Kim, S. R. Lee, M. Chang, E. Cho, Y. B. Kim, C. J. Kim, U. I. Chung, and I. K. Yoo, Sci Rep 3, 1680 (2013).

[38] S. K. Nandi, S. K. Nath, A. E. El-Helou, S. Li, T. Ratcliff, M. Uenuma, P. E. Raad, and R. G. Elliman, ACS Applied Materials & Interfaces 12, 8422-8428 (2020).

[39] S. Deshmukh, M. M. Rojo, E. Yalon, S. Vaziri, C. Koroglu, R. Islam, R. A. Iglesias, K. Saraswat, and E. Pop, Science Advances 8, eabk1514 (2022).

[40] J. M. Goodwill, G. Ramer, D. Li, B. D. Hoskins, G. Pavlidis, J. J. McClelland, A. Centrone, J. A. Bain, and M. Skowronski, Nature Communications 10, 1628 (2019).

[41] I. M. Datye, M. M. Rojo, E. Yalon, S. Deshmukh, M. J. Mleczko, and E. Pop, Nano Letters 20, 1461-1467 (2020).

[42] S. Deshmukh, M. M. Rojo, E. Yalon, S. Vaziri, and E. Pop, 2018 76th Device Research Conference (DRC), 1-2 (2018).

[43] S. K. Nandi, E. Puyoo, S. K. Nath, D. Albertini, N. Baboux, S. K. Das, T. Ratcliff, and R. G. Elliman, ACS Applied Materials & Interfaces 14, 29025-29031 (2022).

[44] R. Islam, S. Qin, S. Deshmukh, Z. Yu, C. Köroğlu, A. I. Khan, K. Schauble, K. C. Saraswat, E. Pop, and H. S. P. Wong, Applied Physics Letters 121, 082103 (2022).

[45] T. Chang, S.-H. Jo, and W. Lu, ACS Nano 5, 7669-7676 (2011).

[46] A. Subramaniam, K. D. Cantley, G. Bersuker, D. C. Gilmer, and E. M. Vogel, IEEE Transactions on Nanotechnology 12, 450-459 (2013).

[47] T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski, and M. Aono, Nature Materials 10, 591-595 (2011).

[48] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, Nano Letters 10, 1297-1301 (2010).

[49] D. S. Jakob, A. Centrone, Analalytical Chemistry 94, 15564-15569 (2022).

[50] G. S. Shekhawat, S. Ramachandran, H. J. Sharahi, S. Sarkar, K. Hujsak, Y. Li, K. Hagglund, S. Kim, G. Aden, A. Chand, and V. P. Dravid, ACS Nano 12, 1760-1767 (2018).

[51] G. Pavlidis, S. Pavlidis, E. R. Heller, E. A. Moore, R. Vetury, and S. Graham, IEEE Transactions on Electron Devices 64, 78-83 (2017).

[52] N. Raghavan, D. D. Frey, M. Bosman, and K. L. Pey, Microelectronics Reliability 55, 1422-1426 (2015).

[53] E. A. Ganykina, A. A. Rezanov, and Y. S. Gornev, Proc. of SPIE 12157, 121570B (2022).

[54] C. H. Wu, S. K. Lin, C. H. Pan, P. H. Chen, W. Y. Lin, T. C. Chang, T. M. Tsai, Y. L. Xu, C. C. Shih, Y. S. Lin, W. C. Chen, M. H. Wang, S. D. Zhang, and S. M. Sze, IEEE Electron Device Letters 39, 1163-1166 (2018).

[55] Y. Hou, U. Celano, L. Goux, L. Liu, R. Degraeve, Y. Cheng, J. Kang, M. Jurczak, and W. Vandervorst, Applied Physics Letters 109, 023508 (2016).

[56] M. Vollmer, European Journal of Physics 30, 1063-1084 (2009).

[57] G. Langer, J. Hartmann, and M. Reichling, Review of Scientific Instruments 68, 1510-1513 (1997).

[58] D. J. Kim, D. S. Kim, S. Cho, S. W. Kim, S. H. Lee, and J. C. Kim, International Journal of Thermophysics 25, 281-289 (2004).

[59] M. A. Panzer, M. Shandalov, J. A. Rowlette, Y. Oshima, Y. W. Chen, P. C. McIntyre, and K. E. Goodson, IEEE Electron Device Letters 30, 1269-1271 (2009).

[60] D. Thuau, I. Koymen, and R. Cheung, Microelectronic Engineering 88, 2408-2412 (2011).

[61] E. A. Scott, J. T. Gaskins, S. W. King, and P. E. Hopkins, APL Materials 6, 058302 (2018).

[62] M. Wang, G. Ramer, D. J. Perez-Morelo, G. Pavlidis, J. J. Schwartz, L. Yu, R. Ilic, V. A. Aksyuk, and A. Centrone, Nano Letters 22, 4325-4332 (2022).

Supplemental Material for

Thermal environment impact on HfO_x RRAM operation: a nanoscale thermometry and modeling study

Temperature Coefficient of Resistance

The calibration of the scanning thermal microscopy (SThM) tip voltage to temperature was conducted by measuring the surface temperature of gold microheaters. The temperature of the microheater was controlled by accurately varying the metal line resistance, R_H, with a 4-point measurement configuration. To accomplish this, the temperature coefficient of resistance (TCR) of the gold microheater was extracted by measuring the metal line resistance as a function of the oven temperature, T_{oven} (Figure S1). A linear fit using a Monte Carlo method (that accounts for the uncertainty associated with the oven temperature and the resistance measurement) was implemented to extract the TCR.

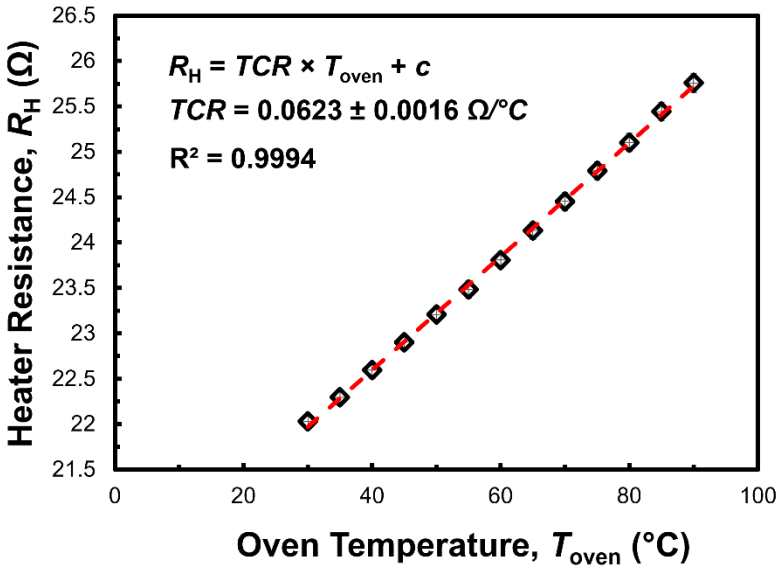


Figure S10. Measured resistance (4-point configuration) of a ≈ 200 nm thick gold microheater as a function of the oven temperature. A linear fit using a Monte Carlo method was used to extract

the temperature coefficient of resistance (TCR). The horizontal error bars are smaller than the dot size and correspond to the nominal oven temperature uncertainty according to the manufacturer specifications. The vertical error bars represent the standard error of the average device resistance (with 95% confidence intervals).

Finite Element Thermal Model for SThM Calibration

Since the TCR is obtained with a global, i.e. average, measurement across the length of the microheater but the SThM calibration is conducted with a local measurement in the microheater mid-point, we used a thermal finite element model (FEM) to assess the microheater temperature heterogeneities. The model consisted of a 200 µm thick, 1000 µm long and 5 µm wide gold (k_{Gold} = 315 W/mK) microheater, a 300 nm thick SiO_2 (k_{SiO2} = 1.1 W/mK) layer and 10 µm thick Si (k_{Si} = 148 W/mK) substrate. The heating was modelled as a constant surface heat flux across the 1000 µm × 5 µm area of the gold microheater A line temperature probe was applied along the length (1000 µm) of the microheater along the median, to quantify the spatial temperature distribution. The extracted temperature profile was averaged to mimic the average temperature estimated using the microheater metal resistance. The heat flux was adjusted accordingly to the five temperature settings (spatially averaged) used for the SThM tip calibration. For each calibration temperature, the temperature profile from the center of the microheater (SThM tip measurement location) to the edge was extracted from the FEM model (plotted in Fig. S2).

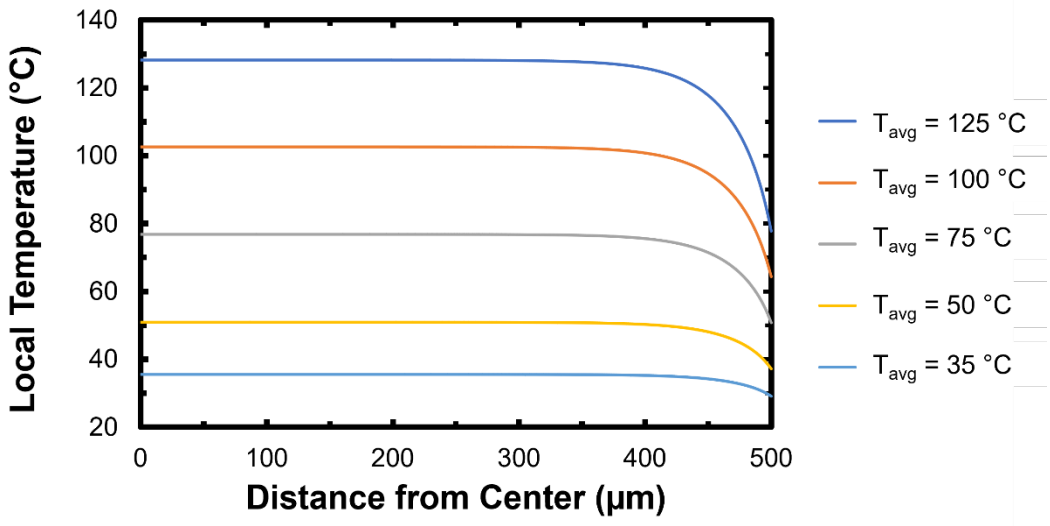


Figure S11. Spatial temperature profiles extracted from the thermal Finite Element Model (FEM) used to simulate Joule heating in the gold microheater. The heating was modelled for five different average temperatures (T_{avg}) corresponding to the conditions used for calibrating the SThM tips.

Based on this model, the maximum discrepancy between the maximum temperature at the center of the microheater (SThM tip measurement location) and the average line temperature (used to control the microheater temperature) was estimated to be 2.54% (Table S1). To correct for this systematic error, for the SThM calibration (Fig 2b) we used FEM calculated temperatures in the center of the wire..

Table S2. Averaged simulated temperature across the heater length compared to the maximum temperature of the heater.

Nominal Average Temperature Rise (°C)	**Max. Temperature (°C)**	**Average Temperature from Simulation (°C)**	**Discrepancy (%)**
35	35.54	35.14	1.16
50	50.91	50.05	1.73

75	76.82	75.17	2.19
100	102.61	100.20	2.41
125	128.19	125.01	2.54

FEM Parameters

$$k_{th,filament} = k_{th,HfO_x} + \frac{(k_{th,Hf} - k_{th,HfOx})}{n_{v,max}} n_v \quad \text{(Equation S7)}$$

$$\sigma_{filament} = \frac{n_v}{n_{v,max}} \sigma_{Hf} e^{\frac{-E_{AC}}{k_B T}} \quad \text{(Equation S8)}$$

$$E_{AC} = E_{AC,max} e^{\frac{-5n_v}{n_{v,max}}} \quad \text{(Equation S9)}$$

As described in the text, these equations are used to estimate the electrical and thermal conductivity of the filament based on the local concentration of oxygen vacancies. The constants used in these equations are listed in Table S2 below.

Table S3. FEM model parameters

Model Parameters		
Parameter	Value	Units
$n_{v,max}$	4.75×10^{27}	m^{-3}
F	2×10^{10}	Hz
a	1×10^{-10}	m
$E_{a,Ti}$	0.6	eV
$E_{a,HfOx}$	1.5	eV
$E_{AC,max}$	0.35	eV
ρ_{Ti}	4506	$kg*m^{-3}$
ρ_{HfOx}	9680	$kg*m^{-3}$
$C_{p,Ti}$	540	$J*kg^{-1}*K^{-1}$
$C_{p,HfOx}$	120	$J*kg^{-1}*K^{-1}$
$k_{th,HfOx}$	10	$W*m^{-1}*K^{-1}$
$k_{th,Hf}$	23	$W*m^{-1}*K^{-1}$
$k_{th,HfO2}$	2	$W*m^{-1}*K^{-1}$
$k_{th,Ti}$	4	$W*m^{-1}*K^{-1}$

Ti/Au TE_{TBC}	100 x 10^6	$W*m^{-2}*K^{-1}$
$Au/SiO_{2,TBC}$	80 x 10^6	$W*m^{-2}*K^{-1}$
σ_{Hf}	2.5 x 10^5	$S*m^{-1}$
σ_{Ti}	6 x 10^6	$S*m^{-1}$

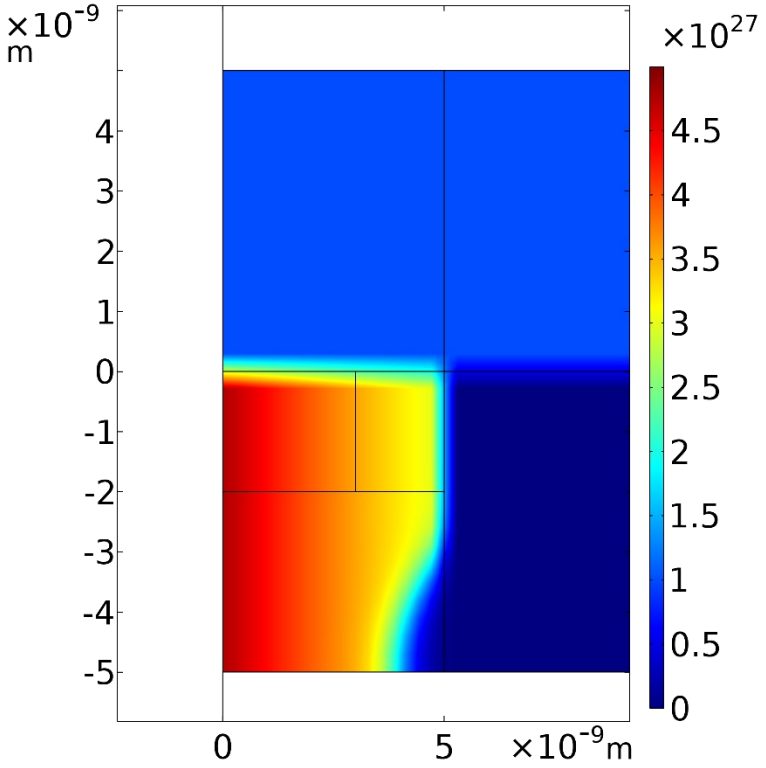


Figure S12. Initial vacancy distribution for the device in the low resistance state. The model is unable to simulate the forming process. A shape with a smaller diameter near the bottom electrode is assumed. This filament has proven to match the electrical and thermal behavior of our devices well.