Thousands of conductance levels in memristors monolithically integrated on CMOS

Jianhua Yang (jjoshuay@usc.edu)
University of Southern California https://orcid.org/0000-0003-0671-6010

Mingyi Rao
University of Massachusetts, Amherst

Hao Tang
Massachusetts Institute of Technology

Jiang-Bin Wu
University of Southern California https://orcid.org/0000-0002-8751-7082

Wenhao Song
University of Southern California

Max Zhang
TetraMem Inc.

Wenbo Yin
TetraMem Inc.

Ye Zhuo
University of Southern California

Fatemeh Kiani
University of Massachusetts Amherst

Benjamin Chen
University of Massachusetts Amherst

Hefei Liu
University of Southern California https://orcid.org/0000-0001-6533-7112

Hung-Yu Chen
Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California

Rivu Midya
University of Massachusetts Amherst

Fan Ye
University of Massachusetts Amherst

Hao Jiang
University of Massachusetts Amherst

Zhongrui Wang
The University of Hong Kong https://orcid.org/0000-0003-2264-0677

Mingche Wu

TetraMem Inc.

Miao Hu

Binghamton University https://orcid.org/0000-0002-6370-5564

Han Wang

University of Southern California

Qiangfei Xia

University of Massachusetts Amherst https://orcid.org/0000-0003-1436-8423

Glenn Ge

TetraMem Inc.

Ju Li

Massachusetts Institute of Technology https://orcid.org/0000-0002-7841-8058

Physical Sciences - Article

Keywords:

Posted Date: September 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1939455/v1

Thousands of conductance levels in memristors monolithically integrated on CMOS

Mingyi Rao[1,2,5], Hao Tang[3,5], Jiangbin Wu[4,5], Wenhao Song[4,5], Max Zhang[1], Wenbo Yin[1], Ye Zhuo[4], Fatemeh Kiani[2], Benjamin Chen[2], Hefei Liu[4], Hung-Yu Chen[4], Rivu Midya[2], Fan Ye[2], Hao Jiang[2], Zhongrui Wang[2], Mingche Wu[1], Miao Hu[1], Han Wang[4], Qiangfei Xia[1,2,], Ning Ge[1], Ju Li[3], J. Joshua Yang[1,2,4,*]

[1]TetraMem Inc., Fremont, CA, USA
[2]Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA
[3]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[4]Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA
[5] These authors contributed equally.
[*] Email: jjoshuay@usc.edu

Neural networks based on memristive devices [1-3] have shown potential in substantially improving throughput and energy efficiency for machine learning [4] and artificial intelligence [5], especially in edge applications. [6-19] Because training a neural network model from scratch is very costly, it is impractical to do it individually on billions of memristive neural networks distributed at the edge. A practical approach would be to download the synaptic weights obtained from the cloud training and program them directly into memristors for the commercialization of edge applications (Figure 1a). Some post-tuning in memristor conductance to adapt local situations may follow afterward or during applications. Therefore, a critical requirement on memristors for neural network applications is a high-precision programming ability to guarantee uniform and accurate performance across a massive number of memristive networks. [20-26] That translates into the requirement of many distinguishable conductance levels on each memristive device, not just lab-made devices but more importantly, devices fabricated in foundries. High precision memristors also benefit other neural network applications, such as training and scientific computing. [23, 27] Here we report over 2048 conductance levels, *the largest number among all types of memories ever reported*, achieved with memristors in *fully integrated chips with 256 × 256 memristor arrays monolithically integrated on CMOS circuits in a standard foundry*. We have unearthed the underlying physics that previously limited the number of achievable conductance levels in memristors and developed electrical operation protocols to circumvent such limitations. These results reveal insights into the fundamental understanding of the microscopic picture of memristive switching and provide approaches to enable high-precision memristors for various applications.

Memristive switching devices are known for their relatively large dynamical range of conductance, which can potentially lead to a large number of discrete conductance levels. However, the highest number reported to date has been no more than two hundred.[20]

There are no forbidden conductance states within the dynamical range of the device since a memristor is typically analog and can, in principle, achieve an infinite number of conductance levels. However, the fluctuation commonly observed at each conductance level (Fig.1e) limits the number of distinguishable levels achievable within a specific conductance range. Interestingly, we found that such fluctuation can be substantially suppressed, as shown in Figs. 1e and 1f, by applying appropriate electrical stimuli (termed as 'denoising' processes). Importantly, such denoising process does not require any extra circuitry beyond the normal read and program circuits. We incorporated the denoising process into device tuning algorithms and successfully programmed a commercial-semiconductor-manufacturer-made memristor (Figs. 1b-d) into 2048 conductance levels (Fig. 1g), corresponding to 11-bit resolution, which is the highest in a single cell among all types of memory devices to the best of our knowledge. Conductive atomic force microscopy (C-AFM) was employed to visualize the evolution of conduction channels during programming and denoising processes. We discovered that a normal switching operation (SET or RSET) always ends up with some incomplete conduction channels, which appear as islands or blurry edges along the main conduction channel and are more resistive and less stable than the main conduction channel. First principle calculations suggest that these incomplete channels are unstable phase boundaries, and their conductance is sensitive to trapped charges, contributing to the large fluctuations of each conductance level. We revealed, experimentally and theoretically, that an appropriate voltage in the denoising process either annihilates (weakens) or completes (enhances) these incomplete channels, resulting in a great reduction in fluctuation and a significant increase in memristor precision. The observed phenomena generally exist in memristive switching process with localized conduction channels, and the insights can be applied to most memristive material systems for scientific understanding and technological applications.

Memristors used in this study were fabricated on an 8-inch wafer by a commercial semiconductor manufacturer (Fig. 1b). The fabrication details are given in the Method section. Cross-section views of a memristor are shown in Fig. 1c, and the critical resistive switching layers are zoomed-in in Fig. 1d. The device consisting of a Pt bottom electrode, a Ti/Ta top electrode, and a HfO_2/Al_2O_3 bilayer, was fabricated in a 240 nm via above the CMOS peripheral circuitry. The Al_2O_3 and Ti layers are designed to be so thin (<1nm) that they appear as a mixed layer rather than two separate continuous layers. When the bottom electrode is grounded, the device can be switched by applying either a sufficiently positive voltage for SET or a negative voltage for RESET to the top electrode. The fluctuation level (characterized by the standard deviation of a measured current under a constant voltage) after a SET or a RESET operation is distributed in a wide range (Fig. S1). The result shows that an as-programmed state typically has a large fluctuation, which significantly limits the applications of memristors but unfortunately exists in memristive materials generally. [28-31] The data also reveals that a SET operation tends to induce a larger fluctuation in an as-programmed state than a RESET operation. The main contribution of such reading fluctuation is random telegraph noise (RTN) which features step-like transitions between two or more current levels at random time points under a constant reading voltage. Such RTNs generally exist in memristors and even fluctuations that are seemingly not step-like

may in fact be made of RTN noise, [32] which can be revealed only when the measurement sampling rate is higher than the RTN frequency, as shown in Fig. S2. It has been demonstrated previously by simulations that memristor RTNs may be caused by charges occasionally trapping into certain defects and blocking conduction channels via coulomb screening. [29, 33] However, experiments that directly link trapped charges, conduction channel(s), and RTNs are missing, let alone how to remove RTNs. Although a critical issue for memristors in general, it has been unclear how to reduce RTNs in memristors. They are critical not only for understanding the physical origin of memristor RTNs but also for revealing the entire microscopic picture of memristive switching and providing possible solutions to high-precision memristors.

We discovered that the fluctuation level could be greatly reduced by applying small voltage pulses with optimized amplitude and width. One example is given in Fig. 1e, where an as-programmed state with a considerable fluctuation (blue) was stabilized into a low-fluctuation state (red) by denoising pulses. Using a three-level feedback algorithm devised to denoise, as detailed in Fig. S3, a single memristor was tuned into 2048 conductance states between 50 and 4,144 μS, with a 2 μS interval between every two neighboring states. All states were read by a voltage sweeping from 0 to 0.2V, as shown in Fig. 1g. The zoomed-in view of the current-voltage curves is given as the lower inset to Fig. 1g, showing well-distinguishable states and the superb linearity of each state. Three nearest neighboring states after denoising are shown in Fig. 1f, where a constant 0.2V voltage reads each state for 1,000 seconds. The current fluctuation of every state is within 0.4 μA, corresponding to 2 μS in conductance. No significant overlap was observed in the neighboring states. We further adopted the denoising process in the array-level programming of an entire 256 × 256 array using the on-chip circuitry. The experimentally programmed patterns are shown both in Fig. 1g as an upper inset and in Fig. S4. For these demonstrations using the on-chip circuitry, the programming precision was limited by the precision of the on-chip Analog/Digital conversion peripheral circuitry, which was 6-bit (64 levels) in this design. The testing setup and the schematic of the driving circuits are shown in Fig. S5. Further studies show that the denoising operation can also reduce RTNs in other material stacks, e.g., a TaO_x-based memristor, as shown in Fig. S6. Since reading noise has been observed in various resistive switching materials, the above results show that the denoising step is an important, or even essential, process for the training of memristive neural networks as unstable readings lead to incorrect outputs from the neural networks.

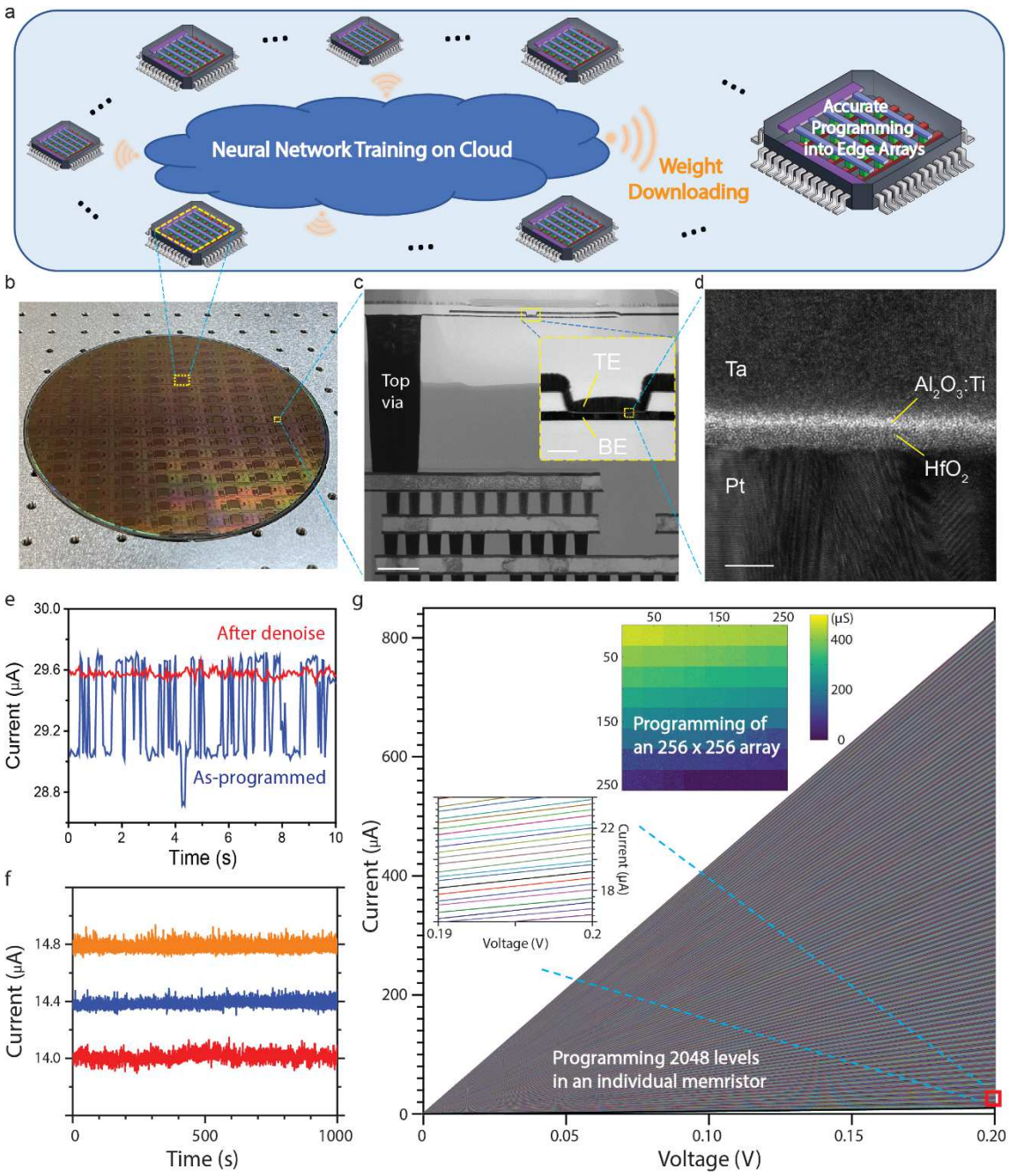

Fig. 1 High precision memristor for neuromorphic computing. a) The most likely scheme of the large-scale application of memristive neural networks for edge computing. Neural network training is performed in the cloud. The obtained weights are downloaded and accurately programmed into a massive number of memristor arrays distributed at the edge, which imposes high-precision requirements on memristive devices. b) The photo of an 8-inch wafer with memristors fabricated by a commercial semiconductor manufacturer. c) HR-TEM image of the cross-section view of a memristor. Pt and Ta serve as bottom and top electrodes, respectively. Scale bar (inset): 1 μm (100 nm). d) The zoomed-in image of

the memristor material stack. Scale bar: 5 nm. e) The as-programmed (blue) and after-denoising (red) currents of a memristor are read by a constant 0.2V voltage. The denoising process eliminated the large amplitude random telegraph noise (RTN) observed in the as-programmed state (see method). f) Zoomed-in view of three nearest neighboring states after denoising. The current of each state was read by a constant 0.2V voltage. No large-amplitude RTN was observed, and all the states can be clearly distinguished. g) An individual memristor on the chip was tuned into 2048 resistance levels by a high-resolution off-chip driving circuitry, and each resistance level was read by a DC voltage sweeping from 0 to 0.2V. The target resistance was set from 50 μS to 4144 μS with 2 μS interval between neighboring levels. All readings at 0.2V are less than 1 μS from the target conductance. The lower inset shows a zoomed-in view of the resistance levels. The upper inset shows experimental results of an entire 256 × 256 array programmed by its 6-bit on-chip circuitry into sixty-four 32 × 32 blocks, and each block is programmed into one of the 64 conductance levels.

Deciphering the underlying reason for the above discoveries is essential for offering a reliable solution to a critical technology problem and understanding the dynamic process of memristive switching. Visualizing the evolution of conduction channels during electrical operations is informative for this purpose.[34-36] We used C-AFM measurement to precisely locate the active conduction channel(s) and scan all the surrounding regions. A customized device was fabricated for the C-AFM measurements. The schematic of its structure is shown in Fig. 2a. To use the Pt-coated C-AFM tip as the top electrode, the device was designed to have a reversed structure of the standard device shown in Fig. 1d. By grounding the bottom electrode and applying a voltage to the top electrode, the device can be operated as our standard device with opposite voltage polarities, i.e., a positive voltage tends to RESET the device, and a negative voltage tends to SET the device. Denoising operations were also successfully performed by C-AFM, as shown in Fig. 2b and Fig. 2c. The conductance scanning results before and after denoising corresponding to the reading results of Fig. 2b (2c) are shown in Fig. 2d (2f) and Fig. 2e (2g), respectively. Comparing the conductance maps in Fig. 2d and Fig. 2e, it is observed that the main part of the conduction channel (the ‘complete’ channel) remains nearly the same while the positive denoising voltage annihilates an island-like channel (the ‘incomplete’ channel). In contrast, the negative denoising voltage (Fig. 2f and Fig. 2g) reduces the noise by increasing the conductivity of the incomplete channels to the level of complete channels. These results indicate that the conductance of an RTN-rich state can be divided into two parts: the base conductance provided by complete channels and the RTN part provided by incomplete channels. These incomplete channels were formed together with complete channels but are smaller in size and conductivity. A memristor can be denoised by eliminating incomplete channels (either removing or completing them). Incomplete channels are more sensitive to voltage stimuli when compared to complete channels, which makes it possible to tune the former without affecting the latter by using appropriate

electrical stimuli. Further studies suggest that such a mechanism is general and can be performed in other material stacks (Fig. S7) as well.

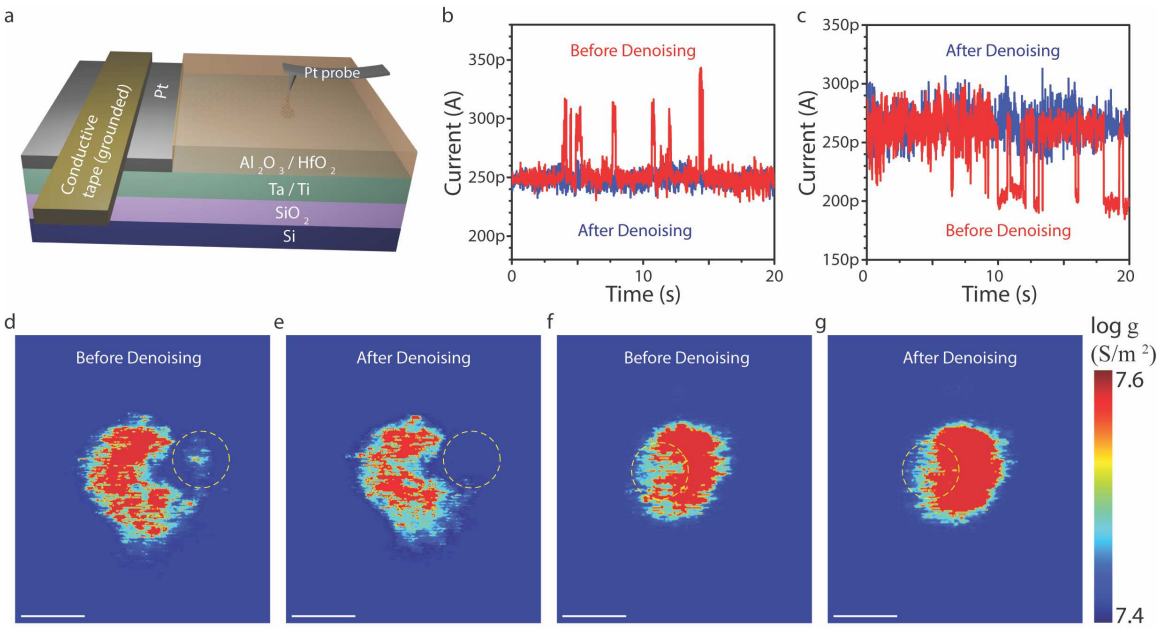

Fig. 2 Direct observation of the evolution of conduction channels in the denoising process through conductive atomic force microscope (C-AFM). a) A schematic of the customized memristor structure and C-AFM testing setup. C-AFM probe played the role of the top electrode in the customized device. Since Ta is easily oxidized in air and not practical to be used as the probe material, a Pt probe was adopted, which served the same role as that of the bottom Pt electrode of a standard memristor we used. To maintain the material stack of a standard memristor, the customized memristor has a reversed structure. b) The current readings by 0.1V voltage before (red) and after (blue) a denoising process by a sub-threshold RESET voltage. c) The current readings by 0.1V voltage before (red) and after (blue) a denoising process by a sub-threshold SET voltage. d) Conductance map measured by C-AFM scanning corresponding to the before-denoising state (red) in b). e) Conductance map corresponding to the after-denoising state (blue) in b). f) Conductance map measured by C-AFM scanning corresponding to the before-denoising state (red) in c). g) Conductance map corresponding to the after-denoising state (blue) in c). All scale bar: 10 nm.

To understand the mechanism of denoising, we studied the microscopic origin of RTNs in memristors. A critical question is whether RTN is induced by an 'atomic effect' or 'electronic effect'. As shown in Fig. S8, incomplete channels are consistently observed in a C-AFM scanning whenever RTN is observed. Once incomplete channels are

eliminated, RTN disappears. Such result indicates that RTN is a phenomenon in company with incomplete channels rather than being induced by the transition process between incomplete and complete channels. The transition between incomplete and complete channels is typically caused by atomic motion or rearrangement, leading to phases (or states) with different conduction levels. Moreover, from an energy point of view, typically only one of the phases is most thermodynamically favored, which implies that if RTN is caused by a frequent shifting of the incomplete part of the conduction channel among two or more phases, it would exhibit a trend of gradually stabilizing towards one of the phases. This, however, has never been observed experimentally (e.g., in Fig. S9). Binary-mannered RTNs if caused by atomic effect may only occur as a result of external stimulations, as reported in [37]. Therefore, the conductance jump in RTNs we observed is likely not caused by atomic motion but by the fact that the electrical conduction of the incomplete conduction channels can be frequently blocked when the system assumes certain electronic structures. The RTN amplitude and frequency offer important clues for detailed understanding. The amplitude of the RTN signals is related to the total conductance of incomplete channels and should be positively proportional to the reading current level. The frequency is associated with the time the specific electronic structure sustains in the system. Without changing the atomic structure, the likely reason for changes in electronic structure is charge trapping/de-trapping. For verification, three different reading voltages are sequentially applied to a memristor (with no denoising process applied). The current profile is shown in Fig. 3a, which displays three current plateaus. The RTN is observed at each plateau with an amplitude proportional to the amplitude of the total current. The observed RTNs have an evident bistable character, suggesting that only one incomplete channel is being modulated between 'blocked'/'unblocked' states. The residence time at the higher current level of the bistable states of each RTN event, τ_1, is statistically inversely related to the absolute value of the applied voltage, while the residence time at the lower current level, τ_0, shows no significant relation with the applied voltage, as shown in the solid curves in Fig. 2b. With these observations, we construct a microscopic model using HfO_2 as an example to explain these properties. Combining the statistical analysis of RTN currents with the atomistic simulations, we identify that interstitial oxygen defects are responsible for the observed RTNs (defect structure and electronic structure shown in the insets to Fig. 3b). An interstitial oxygen defect hosts a localized electronic state, acting as a trap for electrons. The trap state is an antibonding orbital of the interstitial oxygen and the adjacent oxygen, which is unoccupied in a neutral defect. After capturing an electron, the defect becomes negatively charged, and the trap state is occupied by a spin-up electron (Fig. S10). As the occupation of the antibonding orbital significantly weakens the bond between the interstitial oxygen and its adjacent oxygen atom, evident atomic relaxation happens during the electron trapping (Fig. 3b). This leads to a large atomic relaxation energy of 2.62 eV and a long characteristic time of the RTN. The electron trapping affects both the electrostatic potential and the local strain of the material, both of which have an impact on the conduction channel. The potential energy of electrons in the conduction channel is raised by the electrostatic repulsion, giving rise to an energy barrier. Therefore, the conduction channel can be partially or entirely blocked by the charged defect,

depending on the conduction channel size and oxygen vacancy concentration. Therefore, the neutral state and negatively charged state correspond to the high current level and low current level in the RTN, respectively. With this picture, τ_1 and τ_0 are simulated by the electron capture rate ($1/\tau_c$, $\tau_c = \tau_1$) and emission rate ($1/\tau_e$, $\tau_e = \tau_0$) of the defects.

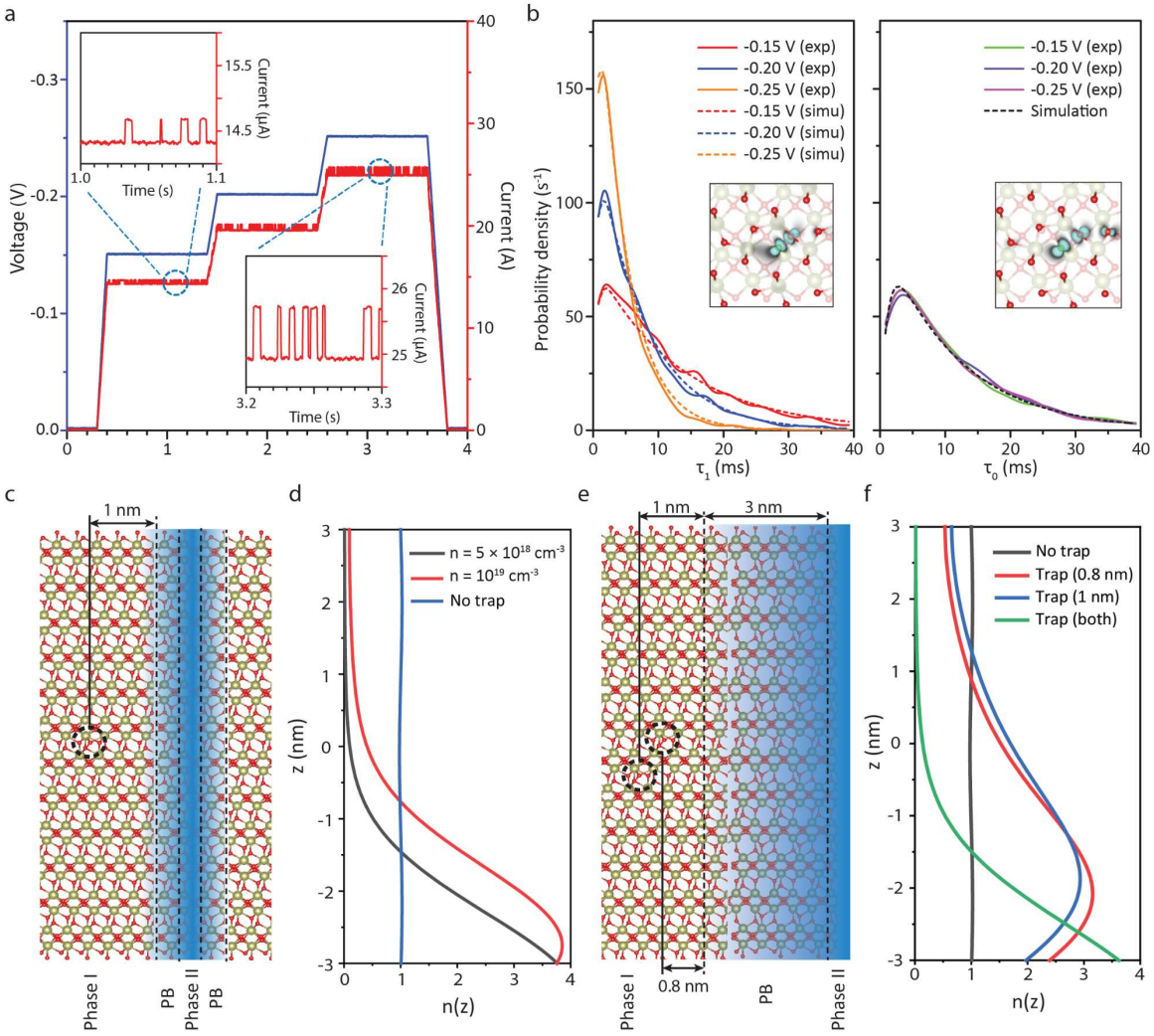

Fig 3. Origin of the RTN in HfO_2 memristor device. (a) Applied electric voltage and measured current in the device as a function of time. Three voltage plateaus (-0.15, -0.20 and -0.25 V) are applied, and two-level RTNs are observed. (b) Simulation results of the probability density of the lifetime of the oxygen interstitial defect at neutral ($f(\tau_1)$) and -1 charged states ($f(\tau_0)$) at the three-voltage plateaus, which agree well with the experimentally observed statistical distributions of τ_1 and τ_0. The insets show the charge density function of the electron trap orbital by the grey-scale map (color from grey to white plotted as the background) and the iso-surface of $\rho = 0.015\ e/Å^3$ is marked cyan. The balls represent atoms: the green ball is the interstitial oxygen atom, red balls are other oxygen atoms, and light-yellow balls are Hf atoms. (c) The schematic where an interstitial oxygen defect (marked by a black dashed circle) is 1 nm away from an island-like incomplete channel. The incomplete channel is formed by a conductive phase region (phase

II) and the phase boundary region (PB). (d) Transport electron wave function corresponding to (c). z denotes the position of the channel along the electron transport direction (from -3nm to 3nm), and n(z) shows the normalized integration of the transport electron wave function on the plane perpendicular to the z direction, which indicates the electrical conduction at each z position. The black and red curves are $n(z)$ when the carrier density in the channel is 5×10^{18} or 10^{19} cm^{-3} with one electron trapped at the oxygen interstitial defect, respectively, and the blue line is $n(z)$ with no electron trapped. (e) The schematic where two interstitial oxygen defects (marked by dashed circles) are away from an incomplete channel that is attached to the main conduction channel. The PB region is 3 nm in width in this case. (f) Transport electron wave function corresponding to (e). The red/blue lines represent $n(z)$ when one electron is trapped in the defect 0.8/1 nm away from the channel, respectively, and the green/black lines are $n(z)$ when both/none defects trap electrons. The carrier density in the channel for the simulation is 5×10^{18} cm^{-3}.

The value of τ_c and τ_e are computed in a similar method as reported in ref. [38], where the material-specific parameters were chosen according to our device information and first principle calculation results, including the atomic relaxation energy during electron trapping and the thermal activation energy of the trapped electron. The distribution functions of τ_0 and τ_1 are derived for different reading voltages (the delay effect of the current response to the electron trapping is included by an empirical coefficient, see Supplementary Information S.11 for details). We can see that the simulated distributions are consistent with those extracted from the experimental RTN measurements. The large E_{REL} is the primary reason for the long τ_0 and τ_1. In comparison, the characteristic time of oxygen vacancy and Ta substitution defects are both on the order of nanoseconds because of their small E_{REL}. As a result, their capture rate and emission rate are much faster than oxygen interstitial defects, which is why they do not induce noticeable RTNs in the measurements of memristors. The trend of τ_0 and τ_1 can be intuitively understood as follows. Because the electron capture rate of the defect is positively related to the electric current density through the defect, τ_1 (τ_c) is inversely related to the absolute value of the reading voltage. In comparison, the electron emission process is not substantially influenced by the current level, so τ_0 (τ_e) is approximately independent of the applied voltage. Finally, the incomplete channel blocking process was modeled as shown in Fig. 3c-f. According to C-AFM experiments, the device region can be classified into three phases: the non-conductive phase (phase I), the conductive phase (phase II), and the phase between them, which features an intermediate conductance (phase boundary, PB). During the programming / denoising operations, these PB regions form or disappear, accompanying the observation of RTN and its annihilation, indicating that some RTN-inducing incomplete channels are located in these PB regions. Fig. 3c shows the schematic of the case where an interstitial oxygen defect is 1 nm away from an incomplete channel whose width is 1 nm. This corresponds to the 'island' case observed in C-AFM where the RTN is accompanied by a very thin incomplete channel isolated from the main channel. The transport electron wave functions $\psi(x, y, z)$ with / without a trapped charge are plotted

in Fig. 3d by the probability density at each cross-section of the channel $n(z) = \int |\psi(x,y,z)|^2 dxdy$ (z is the axis along the channel). This reflects what proportion of the injected electron propagates through the channel. If $n(L) \simeq 0$ (L is the z coordinate of the terminal of the channel), the electron transport is completely blocked, corresponding to a low conductivity; if $n(L) \simeq 1$, the electron goes through the channel with a negligible barrier, corresponding to high conductivity. To mimic the case where there are different percentages of phase || , two charge carrier densities were used for the simulations. The results suggest that the incomplete channel is entirely blocked at a lower charge carrier density (lightly doped with oxygen vacancies, corresponding to less phase ||) and mostly blocked at a higher charge carrier density (heavily doped, corresponding to more phase ||). Fig. 3e corresponds to another common case as observed in C-AFM, where the incomplete channel is attached to the main channel. We simulated the case where there are multiple charge traps (interstitial oxygens) and different trapping scenarios. The corresponding result in Fig. 3f shows that the trapped charge close to the incomplete channel tends to have a bigger impact on conductance than the one far away. It is also observed that the impact of multiple charge traps can accumulate and lead to a major change of conductance as the thick PB region is completely blocked in this case. It can be further inferred that two or more (N) charge trapping defects can lead to complex RTN patterns with a maximum of 2^N levels, which is consistent with previous reports. [39, 40]

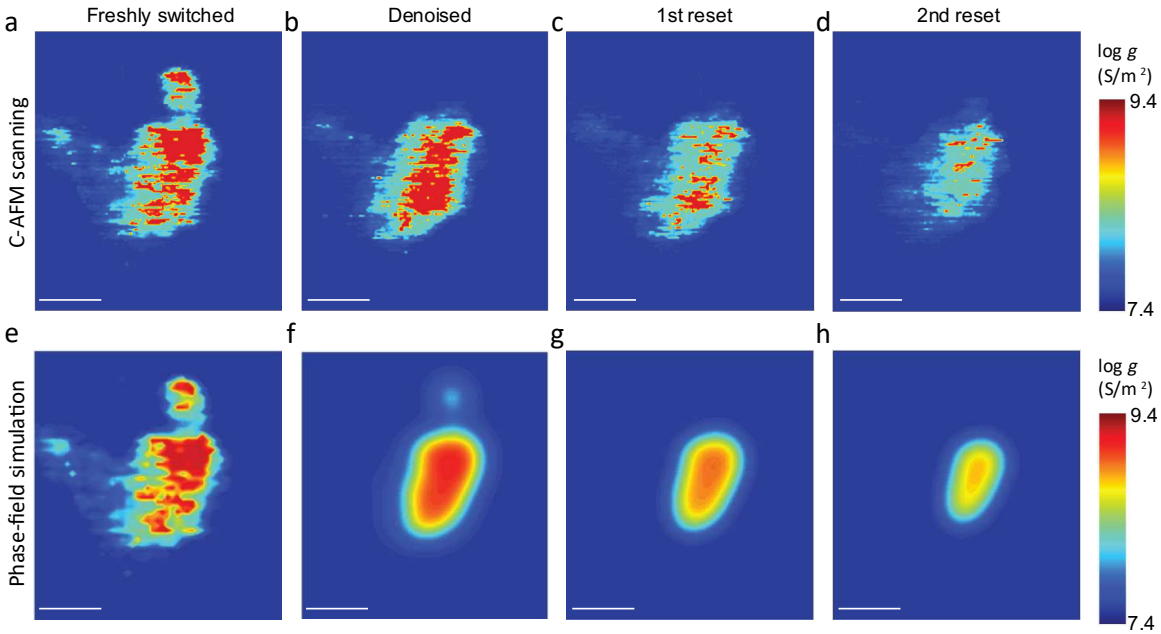

Fig 4. Mechanism of denoising using sub-threshold voltage identified through C-AFM measurement and phase-field theory simulation. From left to right, the conduction channel is first denoised by a 0.2 V voltage and then RESET with a 0.5 V voltage twice. The dynamics of conductive and insulating phase fields are simulated based on the phase transition energy pathway from the first principle calculation. We hypothesize that the

conductive and insulating phases are the orthorhombic phase with a high concentration of oxygen vacancy and the monoclinic phase without oxygen vacancy, respectively. The denoising process is captured by the phase-field relaxation, where the island of the incomplete channel disappears, and the phase boundary sharpens.

As the RTN originates from the incomplete conduction channels, the denoising process is associated with the disappearance of both the island and the blurry boundary of the main channel. The reason why a 'sub-threshold' voltage that is much smaller than the SET or RSET voltages, can decrease the RTN is explained by the phase-field relaxation, as shown in Fig. 4. For this specific material system, the relatively conductive and insulating phases (phase || and phase | in Fig. 3) are the orthorhombic (o) and monoclinic (m) phases of HfO_2, as the o phase is stabilized through a high concentration of oxygen vacancy.[41] The denoising voltage provides a driven force for the phase relaxation by both the temperature effects and the current-induced force, enabling the system to relax towards an equilibrium state. The free energy F and equation of motion of the system are as follows:

$$\Delta F = \int \left[\Delta f_0(\eta) + \frac{1}{2} K (\nabla \eta)^2 \right] dV$$

$$\frac{1}{\alpha} \frac{\partial \eta(r)}{\partial t} = -\alpha \frac{\delta \Delta F[\eta]}{\partial \eta(r)} = -\frac{\partial \Delta f_0}{\partial \eta} + K \nabla^2 \eta$$

Where η is the order parameter (here use the monoclinic angle) describing the transition from *m* to *o* phase, Δf_0 is the free energy density for a system with a certain order parameter, and K is the gradient energy parameter. The energy density Δf_0 is derived from the first principle calculations. Using the phase-field simulation, we derive a similar behavior as observed by the C-AFM: after denoising, the island disappears, and the boundary of the main channel sharpens. The disappearance of the islands is driven by the gradient energy, which evanishes the islands whose size is below the critical nuclei. The sharpening of the boundary is driven by the energy barrier between the two phases, where the high energy boundary region is reduced. During the RESET process, the conduction channel shrinks in size, and its conductivity also decreases, as oxygen vacancy is drifted away from the switching-active region by the strong voltage. Intuitively, it can be understood as follows. The incomplete conduction channels, namely the islands and boundary regions in a freshly switched state, are 'frozen' in a highly non-equilibrium state because they are always formed at the end of the SET or RESET voltage pulse and do not have a chance (sufficient time) to reach the same stable state as the more 'mature' complete channel region formed earlier. Therefore, they are prone to change (either being completed or annihilated), which can be activated by a sub-threshold voltage. On the other hand, different from the complete main conduction channel, electron transport of incomplete channels can be readily blocked by trapped charges as shown in Fig. 3, making them the main source of RTN noises. The situation is more severe for a conductance state obtained by a SET switching process as

conduction channel creation and growth are a positive feedback process, which happens faster and faster and leaves no time for maturation of the newly formed conduction channels before the end of each switching pulse. Although the specific phases involved may be different for different oxide systems, the approach used here and the conclusions drawn are generally applicable.

In summary, we have achieved 2048 conductance levels in memristors, over an order of magnitude higher than previous demonstrations and the highest among all known memories. Importantly, these were obtained in memristors of a fully integrated chip fabricated in a commercial foundry. We have revealed the root cause of conductance fluctuations in memristors through experimental and theoretical studies and devised an electrical operation protocol to denoise the memristors for high precision operations. The denoising process has been successfully operated on the entire 256 × 256 crossbar using the on-chip driving circuitry designed for regular reading and programming without any extra hardware overhead. These results not only provide critical insights into the microscopic picture of the memristive switching process but also represent a leap forward in commercializing memristor technology as hardware accelerators of machine learning and artificial intelligence for edge applications.

Acknowledgements J.J.Y., W.H.S. and Y.Z. were partially supported by a subcontract (GR1055585 53-4502-0003) from the University of Massachusetts Amherst, with the sponsor being TetraMem Inc. J.J.Y also serves as co-founder and paid consultant of TetraMem Inc. R. M., Q. X., and J. J. Y. were also partially supported by the Air Force Office of Scientific Research (AFOSR) through the MURI program under contract no. FA9550-19-1-0213 and by the National Science Foundation under contract no. 2023752. The authors thank Alan Tan for proofreading the manuscript.

Author contributions J.J.Y. and M.R. conceived the concept. J.J.Y and Q.X. supervised the entire project. J.J.Y., M.R., Q.X., H.T., J.W., and W.S. designed the experiments and simulations. M.R., M.Z., R.M., and H.J fabricated the devices. M.R., W.S., Y.Z., B.C., and Z.W. made electrical measurements. H.T., M.R., and J.L. designed and carried out the simulation. J.W., M.R., H.L., H.C., and H.W. designed and carried out the CAFM studies. W.Y., F.K., F.Y., Z.W., M.W., M.H., Q.X., N.G., and J.J.Y. helped with experiments and data analysis. M.R., H.T., and J.J.Y wrote the paper. All authors discussed the results and implications and commented on the manuscript at all stages.

Competing interests The authors declare no competing financial interests.

Reference

[1] L. Chua, "Memristor-the missing circuit element," *IEEE Transactions on circuit theory,* vol. 18, no. 5, pp. 507-519, 1971.

[2] I. Valov, R. Waser, J. R. Jameson, and M. N. Kozicki, "Electrochemical metallization memories—fundamentals, applications, prospects," *Nanotechnology,* vol. 22, no. 25, p. 254003, 2011.

[3] Y. Yang and R. Huang, "Probing memristive switching in nanoionic devices," *Nature Electronics,* vol. 1, no. 5, pp. 274-287, 2018.
[4] W. Wen, C. Wu, Y. Wang, Y. Chen, and H. Li, "Learning structured sparsity in deep neural networks," *Advances in neural information processing systems,* vol. 29, 2016.
[5] S. Kumar, X. Wang, J. P. Strachan, Y. Yang, and W. D. Lu, "Dynamical memristors for higher-complexity neuromorphic computing," *Nature Reviews Materials,* pp. 1-17, 2022.
[6] C.-X. Xue *et al.*, "A CMOS-integrated compute-in-memory macro based on resistive random-access memory for AI edge devices," *Nature Electronics,* vol. 4, no. 1, pp. 81-90, 2021.
[7] M. Lanza *et al.*, "Memristive technologies for data storage, computation, encryption, and radio-frequency communication," *Science,* vol. 376, no. 6597, p. eabj9979, 2022.
[8] P. Yao *et al.*, "Fully hardware-implemented memristor convolutional neural network," *Nature,* vol. 577, no. 7792, pp. 641-646, 2020.
[9] W. Zhang *et al.*, "Neuro-inspired computing chips," *Nature electronics,* vol. 3, no. 7, pp. 371-382, 2020.
[10] D. Ielmini and H.-S. P. Wong, "In-memory computing with resistive switching devices," *Nature electronics,* vol. 1, no. 6, pp. 333-343, 2018.
[11] M. A. Zidan, J. P. Strachan, and W. D. Lu, "The future of electronics based on memristive systems," *Nature electronics,* vol. 1, no. 1, pp. 22-29, 2018.
[12] S. Yu, "Neuro-inspired computing with emerging nonvolatile memorys," *Proceedings of the IEEE,* vol. 106, no. 2, pp. 260-285, 2018.
[13] S. Jung *et al.*, "A crossbar array of magnetoresistive memory devices for in-memory computing," *Nature,* vol. 601, no. 7892, pp. 211-216, 2022.
[14] V. K. Sangwan and M. C. Hersam, "Neuromorphic nanoelectronic materials," *Nature nanotechnology,* vol. 15, no. 7, pp. 517-528, 2020.
[15] G. W. Burr, "A role for analogue memory in AI hardware," *Nature Machine Intelligence,* vol. 1, no. 1, pp. 10-11, 2019.
[16] S. Chen *et al.*, "Wafer-scale integration of two-dimensional materials in high-density memristive crossbar arrays for artificial neural networks," *Nature Electronics,* vol. 3, no. 10, pp. 638-645, 2020.
[17] E. J. Fuller *et al.*, "Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing," *Science,* vol. 364, no. 6440, pp. 570-574, 2019.
[18] C. Choi *et al.*, "Reconfigurable heterogeneous integration using stackable chips with embedded artificial intelligence," *Nature Electronics,* pp. 1-8, 2022.
[19] D.-H. Lim, S. Wu, R. Zhao, J.-H. Lee, H. Jeong, and L. Shi, "Spontaneous sparse learning for PCM-based memristor neural networks," *Nature communications,* vol. 12, no. 1, pp. 1-14, 2021.
[20] Y. Sun *et al.*, "A Ti/AlO x/TaO x/Pt analog synapse for memristive neural network," *IEEE Electron Device Letters,* vol. 39, no. 9, pp. 1298-1301, 2018.
[21] S. Stathopoulos *et al.*, "Multibit memory operation of metal-oxide bi-layer memristors," *Scientific reports,* vol. 7, no. 1, pp. 1-7, 2017.
[22] H. Kim, M. Mahmoodi, H. Nili, and D. B. Strukov, "4K-memristor analog-grade passive crossbar circuit," *Nature communications,* vol. 12, no. 1, pp. 1-11, 2021.
[23] M. A. Zidan *et al.*, "A general memristor-based partial differential equation solver," *Nature Electronics,* vol. 1, no. 7, pp. 411-420, 2018.
[24] C. Li *et al.*, "Analogue signal and image processing with large memristor crossbars," *Nature electronics,* vol. 1, no. 1, pp. 52-59, 2018.

[25] C. Mackin *et al.*, "Optimised weight programming for analogue memory-based deep neural networks," *Nature Communications,* vol. 13, no. 1, pp. 1-12, 2022.
[26] S. Choi *et al.*, "SiGe epitaxial memory for neuromorphic computing with reproducible high performance based on engineered dislocations," *Nature materials,* vol. 17, no. 4, pp. 335-340, 2018.
[27] P. Yao *et al.*, "Face classification using electronic synapses," *Nature communications,* vol. 8, no. 1, pp. 1-8, 2017.
[28] S. Choi, Y. Yang, and W. Lu, "Random telegraph noise and resistance switching analysis of oxide based resistive memory," *Nanoscale,* vol. 6, no. 1, pp. 400-404, 2014.
[29] D. Ielmini, F. Nardi, and C. Cagli, "Resistance-dependent amplitude of random telegraph-signal noise in resistive switching memories," *Applied Physics Letters,* vol. 96, no. 5, p. 053503, 2010.
[30] F. M. Puglisi, P. Pavan, A. Padovani, L. Larcher, and G. Bersuker, "Random telegraph signal noise properties of HfOx RRAM in high resistive state," in *2012 Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, 2012: IEEE, pp. 274-277.
[31] J.-K. Lee *et al.*, "Extraction of trap location and energy from random telegraph noise in amorphous TiO x resistance random access memories," *Applied Physics Letters,* vol. 98, no. 14, p. 143502, 2011.
[32] F. M. Puglisi, A. Padovani, L. Larcher, and P. Pavan, "Random telegraph noise: Measurement, data analysis, and interpretation," in *2017 IEEE 24th International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, 2017: IEEE, pp. 1-9.
[33] F. M. Puglisi, N. Zagni, L. Larcher, and P. Pavan, "Random telegraph noise in resistive random access memories: Compact modeling and advanced circuit design," *IEEE Transactions on Electron Devices,* vol. 65, no. 7, pp. 2964-2972, 2018.
[34] Y. Yang, X. Zhang, L. Qin, Q. Zeng, X. Qiu, and R. Huang, "Probing nanoscale oxygen ion motion in memristive systems," *Nature communications,* vol. 8, no. 1, pp. 1-10, 2017.
[35] A. Ranjan, N. Raghavan, K. Shubhakar, S. J. O'Shea, and K. L. Pey, "Random Telegraph Noise Nano-spectroscopy in High-κ Dielectrics Using Scanning Probe Microscopy Techniques," in *Noise in Nanoscale Semiconductor Devices*: Springer, 2020, pp. 417-440.
[36] F. Hui and M. Lanza, "Scanning probe microscopy for advanced nanoelectronics," *Nature electronics,* vol. 2, no. 6, pp. 221-229, 2019.
[37] S. Brivio, J. Frascaroli, E. Covi, and S. Spiga, "Stimulated ionic telegraph noise in filamentary memristive devices," *Scientific reports,* vol. 9, no. 1, pp. 1-9, 2019.
[38] F. M. Puglisi, L. Larcher, A. Padovani, and P. Pavan, "A complete statistical investigation of RTN in HfO 2-based RRAM in high resistive state," *IEEE Transactions on Electron Devices,* vol. 62, no. 8, pp. 2606-2613, 2015.
[39] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Statistical fluctuations in HfO x resistive-switching memory: Part II—Random telegraph noise," *IEEE Transactions on Electron Devices,* vol. 61, no. 8, pp. 2920-2927, 2014.
[40] T. Becker *et al.*, "An electrical model for trap coupling effects on random telegraph noise," *IEEE Electron Device Letters,* vol. 41, no. 10, pp. 1596-1599, 2020.
[41] Y. Zhou *et al.*, "The effects of oxygen vacancies on ferroelectric phase transition of HfO2-based thin film from first-principle," *Computational Materials Science,* vol. 167, pp. 143-150, 2019.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SIsubmitted.pdf