Supplementary Information:

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., 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

Method

1. Memristor fabrication:

Standard memristor integrated with CMOS driving circuits:

The CMOS part was fabricated in a standard 180 nm process line in a commercial semiconductor manufacturer with exposed tungsten via at the top. Memristors were processed in the same process line with customized materials / recipes. After tungsten via surface oxide cleaning, Pt bottom electrodes were sputtered and patterned on the vias. Memristor holes were defined by etching through a patterned SiO_2 isolation layer (~ 100nm) and stops at the surface of Pt. Resistive switching layer (HfO_2/Al_2O_3) and top electrode (Ti/Ta) were filled into the etched holes sequentially, where the resistive switching layers were fabricated by atomic layer deposition and the top electrode was fabricated by sputter. Finally, standard aluminum interconnect was made to connect the top electrode to bond pads for electrical testing.

Customized memristor for C-AFM measurement

The customized device was fabricated in a university cleanroom on a Si wafer covered by thermally oxidized SiO_2 (~100nm). The bottom electrode (Ta/Ti) and resistive switching layers (Al_2O_3/HfO_2) were deposited by an AJA sputtering system. The four layers were fabricated continuously in the high-vacuum chamber to avoid oxidation of Ta and Ti. The chip was then patterned and etched to expose part of the bottom electrode. After surface oxide cleaning, Pt was deposited onto the exposed bottom electrode to prevent the exposed bottom electrode from oxidation and serve as the ground contact during C-AFM measurement.

2. Electrical measurement:

Single device measurement

Electrical measurement of the standard memristor (foundry made complete memristor with top electrode) was performed on a Keysight B1500A semiconductor device analyzer equipped with a B1530A waveform generator/fast measurement unit (WGFMU). To realize the algorithm as shown in Fig. S3, we built a program using C# to control the electrical operations of B1500A.

Array measurement

The schematic of 1-transistor-1-memristor array with on-chip driving circuits and the testing setup is shown in Supplementary information Fig. S4 and Fig. S5.

C-AFM measurement

The CAFM was performed by the Bruker Dimension Icon system with a conductive probe (SCM-PIT-V2, 0.01-0.025 Ohm-cm of resistivity) under the contact mode. When performing electrical operations including SET, RESET and read, C-AFM probe was at a fixed position. The filament was first formed with a voltage of 4V. During the in-situ SET/RESET and reading operations for the chosen filament, the setpoint was set at a relatively large number (~80 nm) to increase the strength of the pressing force to make a large contact area between tip and sample surface. The setpoint is a measure of the force applied by the tip to the sample. In contact mode, it is a certain deflection of the cantilever. This deflection is maintained by the feedback electronics, so that the force between the tip and sample is kept constant. When performing the filament morphology mapping, the probe scanned a 50 nm by 50 nm region surrounding the conductive channel. The contact area was set to a small value (~10 nm) for high resolution.

3. First principle calculation

The oxygen interstitial defects' atomic and electronic structure are calculated by the density functional theory (DFT) with the projector augmented wave (PAW) method [1] implemented in Vienna *ab initio* simulation package (VASP) [2]. The generalized gradient approximation (GGA) is employed with the Perdew–Burke–Ernzerhof (PBE) exchange-correlation function [3]. The cut-off energy is set as 400 eV, and the *k*-point mesh is sampled by the Monkhorst-Pack method [4] with a separation of 0.2 rad/Å. The atomic structure of the oxygen interstitial defect is constructed by including one oxygen atom in the $2 \times 2 \times 2$ supercell of the *m*-phase HfO_2 crystal. The initial position of the included oxygen atom is set according to ref. [5], and the atomic configuration is fully relaxed. The force on each atom converges to 0.01 eV/Å, and the electronic energy converges to 10^{-6} eV. The atomic structure, charge distribution of the trap-state, and electronic band structure in Fig. 3 and Fig. S10 are then extracted from DFT calculation results.

Simulation the impact of trapped charge to conductive channel:

We simulate the Coulomb blockade effect through the quantum transport of conduction electron in a cuboid conduction channel, as shown in Fig. 3f. The length of the conductive channel is set as $L = 6$ nm to match the channel length in device. The motion of carriers in the conductive channel is calculated through the effective mass approximation, and the Coulomb blockade effect of the oxygen interstitial defect is simulated by a screened Coulomb potential $V(\vec{r})$ acting on the carriers. Assuming the electric conductance outside the channel is negligible, the quantum transport of electron in the channel can be described by the following equations:

$$\begin{cases} \left[-\frac{\hbar^2}{2m^*}\nabla^2 + V(\vec{r})\right]\psi(x,y,z) = (E - E_c)\psi(x,y,z) \\ V(\vec{r}) = \frac{e^2}{4\pi\epsilon_0\epsilon_r}\sum_i \frac{e^{-|\vec{r}-\vec{r}_i|/\lambda_D}}{|\vec{r}-\vec{r}_i|} \\ \psi|_{x=0} = \psi|_{x=d} = \psi|_{y=0} = \psi|_{y=d} = 0 \end{cases}$$

Where m^* is the effective mass of the conductive band of HfO_2, set as $0.11m_e$ according to [6]. E is the eigen energy of the transport electron, set as 0.2 eV above the conduction band minimum E_c estimated by the magnitude of bias voltage of about 0.2 V. The Coulomb potential is the summation of interstitial oxygen atoms located at $\vec{r}_i$, where ϵ_r is the relative dielectric constant (set as 16 according to [7]) and λ_D is the Debye screening length evaluated as $\lambda_D = \sqrt{\frac{\epsilon_0\epsilon_r k_B T}{ne^2}}$ (the temperature T is set as 300 K).

The transport wavefunction with electrons injected from $x = 0$ with unitary amplitude is then calculated with the following boundary conditions:

$$\begin{cases} \psi|_{z=0} = e^{ik_{11}z}\sin\frac{\pi x}{d}\sin\frac{\pi y}{d} + \sum_{m,n} T_{mn}e^{-ik_{mn}z}\sin\frac{m\pi x}{d}\sin\frac{n\pi y}{d} \\ \frac{\partial\psi}{\partial z}|_{z=0} = ike^{ikz}\sin\frac{\pi x}{d}\sin\frac{\pi y}{d} - \sum_{m,n} T_{mn}ik_{mn}e^{-ik_{mn}z}\sin\frac{m\pi x}{d}\sin\frac{n\pi y}{d} \\ \psi|_{z=L} = \sum_{m,n} R_{mn}e^{ik_{mn}z}\sin\frac{m\pi x}{d}\sin\frac{n\pi y}{d} \\ \psi|_{z=L} = \sum_{m,n} R_{mn}ik_{mn}e^{ik_{mn}z}\sin\frac{m\pi x}{d}\sin\frac{n\pi y}{d} \\ k_{mn} = \sqrt{\frac{2m^*}{\hbar^2}(E - E_c) - (m^2 + n^2 - 2)\left(\frac{\pi}{d}\right)^2} \end{cases}$$

The electron transport is then shown by the probability density function of the electron wave function at each cross section of the channel $n(z) = \int|\psi(x,y,z)|^2 dxdy$, reflecting what proportion of the injected electron propagates through the channel. If $\mathrm{n(L)} \simeq 0$, the electron transport is completely blocked; if $\mathrm{n(L)} \simeq 1$, the electron goes through the channel with negligible barrier. Three parameters control the Coulomb blockade: the size of channel d, the carrier density n, and the distance of the oxygen interstitial defect to the channel. These factors lead to the different degrees of Coulomb blockade to the isolated island and main channel, as discussed in the results part.

4. Supporting Figures

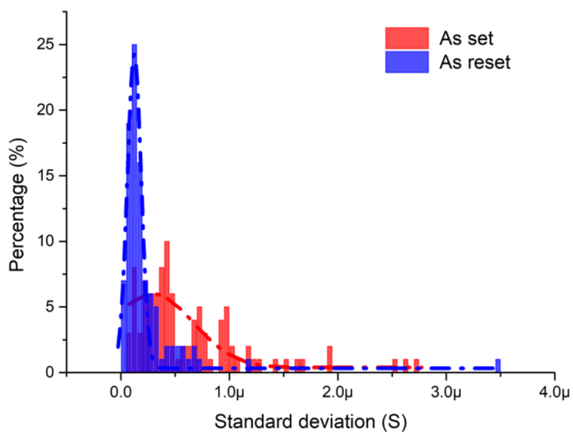

Fig. S1 Statistics of reading standard deviation in an as-programmed memristor. A single memristor was programmed for 200 times, in which 100 programming operations ended with a SET operation and the other 100 programming operations ended with a RESET operation. After each programming operation, the device was read by a constant -0.2V voltage for 5s and the standard deviation of each reading was recorded, which showed a wide distribution from less than 0.2µS to 3.5 µS.

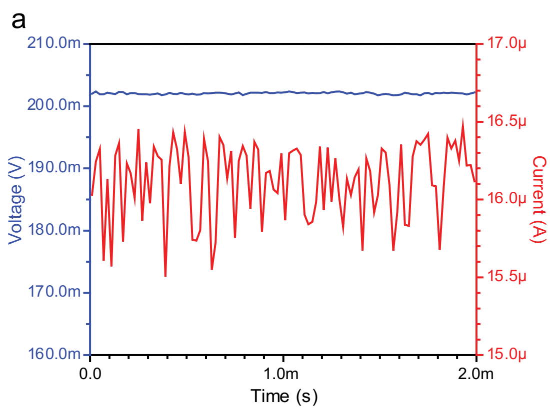

Fig. S2 Readout currents with different sampling intervals of a memristor's state by a constant 0.2V voltage. a) Sampling interval at 20 µs. b) Sampling interval at 2 µs. When the sampling rate is comparable or lower than the frequency of the RTNs, the waveform is distorted, as shown in a). Although being distorted, the noise amplitude is similar to that of the RTNs in b), suggesting that RTN is the main contributor to large amplitude noise in memristors.

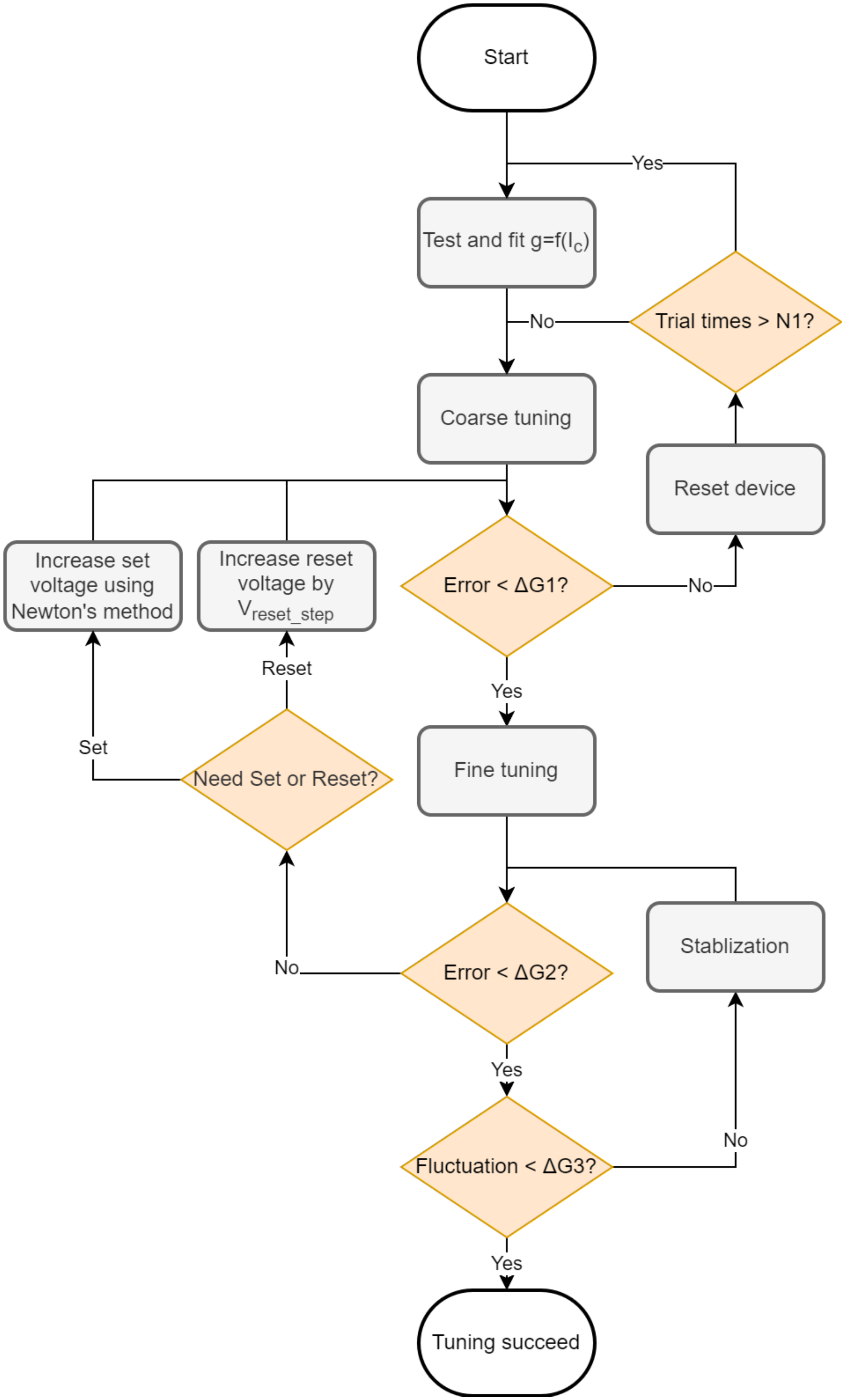
Start
Yes
Test and fit g=f(Ic)
No
Trial times > N1?
Coarse tuning
Reset device
Increase set voltage using Newton's method
Increase reset voltage by Vreset_step
Error < ΔG1?
No
Reset
Yes
Set
Need Set or Reset?
Fine tuning
Stablization
No
Error < ΔG2?
Yes
No
Fluctuation < ΔG3?
Yes
Tuning succeed

Fig. S3 The algorithm of high precision programming. The algorithm was implemented using C# language which could control a Keysight B1500 Semiconductor Device Analyzer by controlling the programming voltage with customized parameters.

We first constructed a model for the device and used it to coarsely tune the device. The device was gradually set by a 2V DC signal using different compliance currents with an increment of 50 μA from 50 μA to 2 mA. After each set operation, the device was read by a 0.2V voltage. After the 2 mA set current compliance was reached, the device was reset by -1.5V to the initial state. That set/read data collection process was repeated for 5 times to minimize the effect of cycle-to-cycle variation. After all the compliance current-conductance pair data (Ic, g) were collected, they were fitted into a linear model $g = f(I_c)$. When using a transistor as the current limiter, the gate voltage of the transistor could be calculated according to its transfer curve. The device was then set to each target conductance using the compliance current predicted by the model. This was called coarse tuning. If the resulted conductance was within a tolerance $\Delta G1$ of the target, the coarse tuning was successful and the fine tuning was applied next. If the deviation was larger than $\Delta G1$, the device was reset and coarsely tunned again. If the coarse tuning failed after N1 attempts, the estimated model $g = f(I_c)$ was no longer correctly representing the current status of the device, so we re-calibrated it and started over again. The typical value of $\Delta G1$ was 50 μS and N1 was 5.

During the fine tuning, we set/reset the device if the current conductance was smaller/larger than the target. For the set operation, the voltage for the next set operation was calculated using the Newton's method where the gradient was extracted from coarse tuning result, and for the reset operation, the reset voltage amplitude was increased by a fixed step each time. After each set/reset operation, the device conductance was read by a 0.2V DC voltage and compared to the target conductance. If the difference was larger than a predefined tolerance $\Delta G2$ (typically 1 μS), another round of set/reset was performed. Once the difference was smaller than $\Delta G2$, the device current was read at a constant 0.2V signal for a constant of time to measure its fluctuation over time. That fluctuation measurement was done multiple times and each fluctuation measurement was conducted within 10s to save time. If the conductance fluctuation was larger than a predefined value $\Delta G3$ (typically 2μS), a stabilization process took place, which typically used one or more one-positive-one-negative pulse pair (until the noise level is within the tolerance range) at 0.35V. The stabilization voltage may sometime change the average conductance of the device. If the actual-target conductance mismatch was larger than $\Delta G1$ or $\Delta G2$, a corresponding coarse tuning or fine tuning process was applied again.

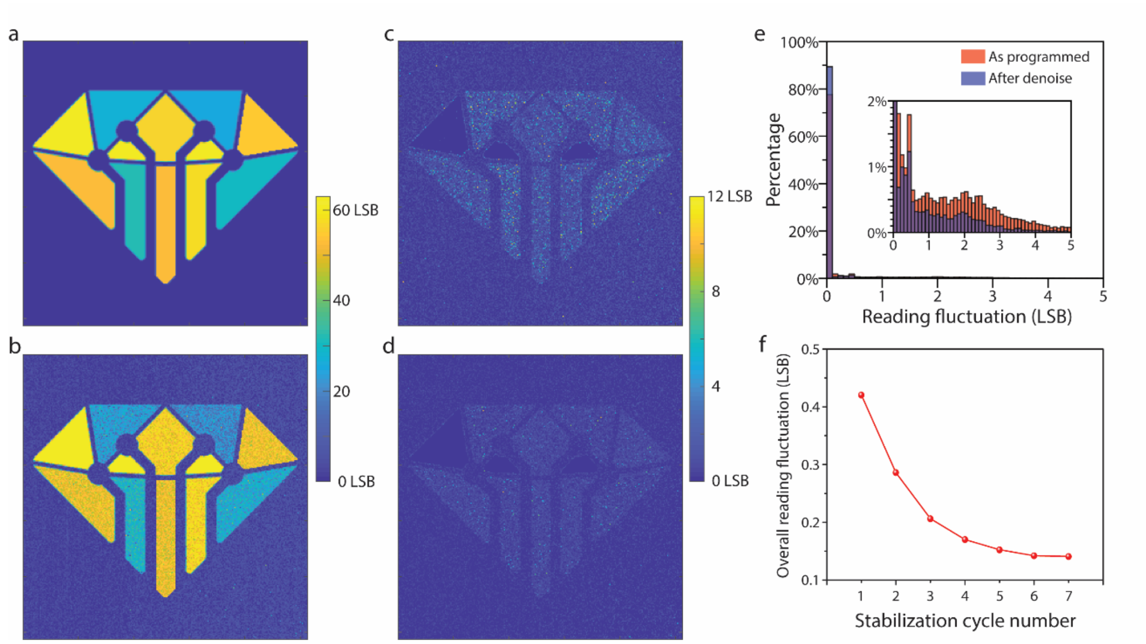

Fig. S4. The pattern programming and denoising in a 256 × 256 1-transistor-1-memristor (1T1M) array. The algorithm is similar to Fig.S3 but performed in a 256 × 256 array. Each array-programming operation is followed by an array denoising operation. In all subplots 1 Least Significant Bit (LSB) is 2.34 µS. a) The target weight map of the programming. b) The programmed weight map. c) The reading noise of devices in the array after the initial programming before the first denoising operation. d) The final reading noise of devices in the array after 7 iterations of programming-denoising, the number of noisy devices significantly decreased compared to those in c). e) The histogram of all device noise level. The inset is a zoomed-in view of the histogram in the range of 0-2%. f) The trend of average reading noise in all 65536 devices in the 7 iterations of programming / denoising.

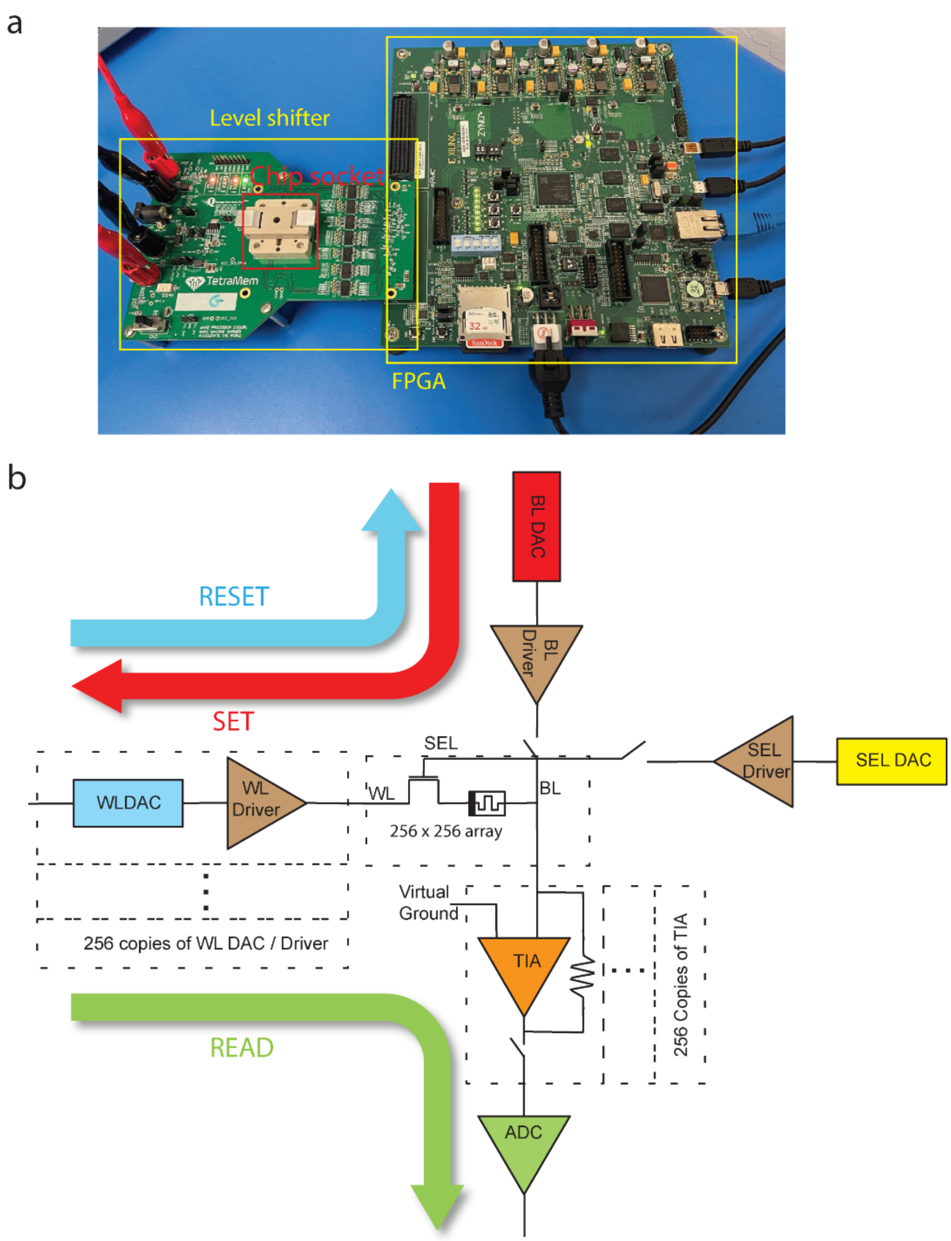

Fig. S5. The testing environment and the schematic of the 1T1R array with its driving circuits. a) The testing set up. The 1T1R array and its driving circuits are controlled by a field programmable gate array (FPGA) through a voltage level shifter. b) The schematic of the 1T1R array and its on-chip driving circuits. Device reading and writing voltage are controlled by on-chip digital-analog-converters (DACs). Current reading is realized by a transimpedance amplifier (TIA) and an analog-digital-converter (ADC). The current flow direction in the three basic operations (SET, RESET and READ) are marked by arrows.

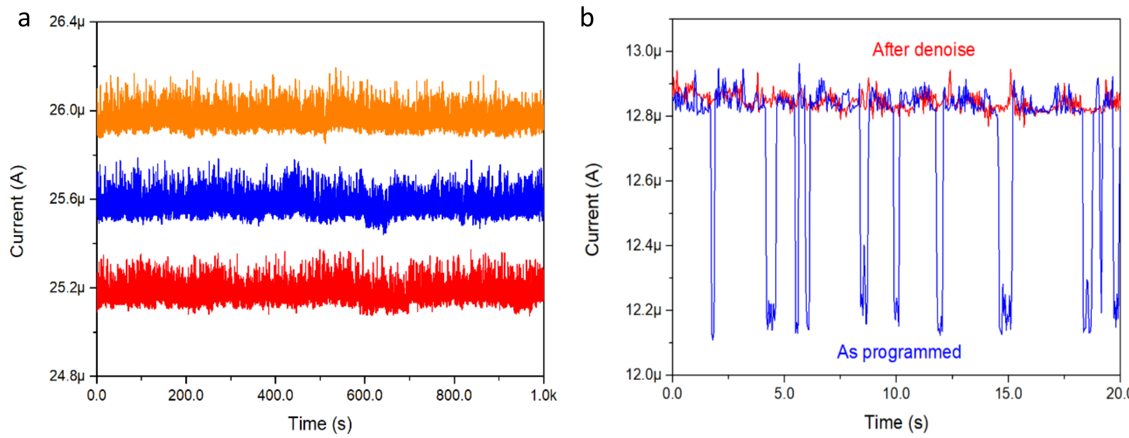

Fig. S6. The denoising effect of TaO_x based device. The device was fabricated by a similar process to the device as shown in Fig.1 except that the HfO_2 layer was replaced by TaO_x layer. The device was read at a constant 0.2V voltage. a) By using a similar denoise operation, the device reading fluctuation can be reduced to below 2 µS. b) A comparison between the device reading when it is as-programmed (blue) and after denoise b).

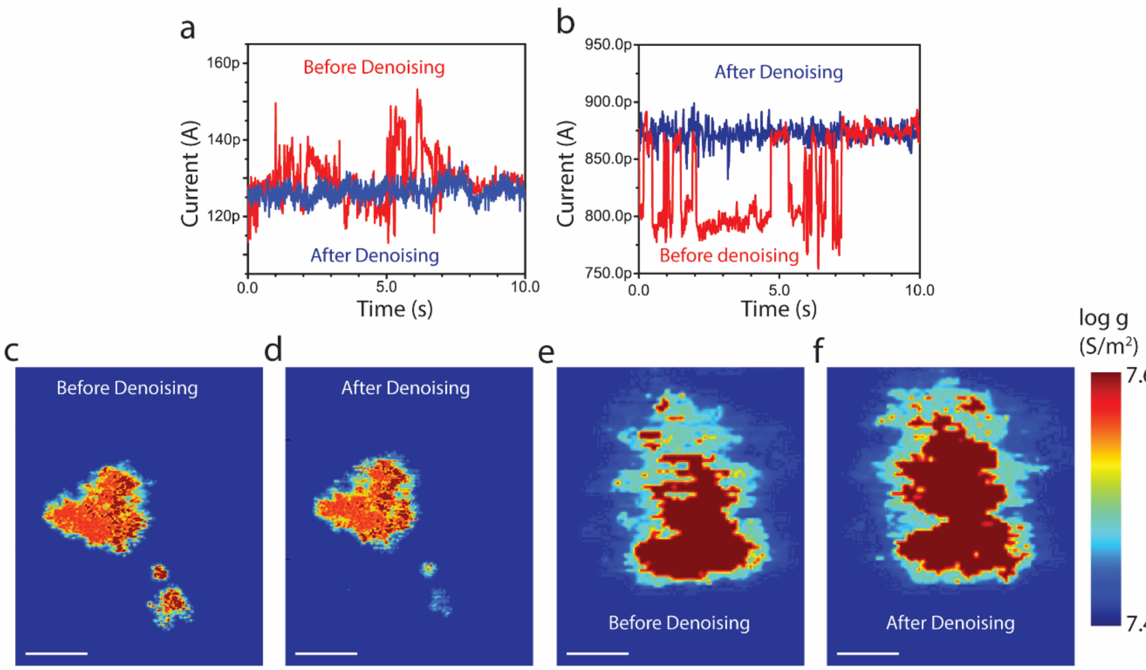

Fig. S7 C-AFM denoise result in a device with material stack Pt/HfO2/Ta. a) The current readings by 0.1V voltage before (red) and after (blue) a successful denoise process by "below switching threshold" RESET voltage. b) The current readings by -0.1V voltage before (red) and after (blue) a successful denoise process by "below switching threshold" SET voltage. c) Conductance map measured by C-AFM scanning corresponding to the before-denoise state (red) in a). d) Conductance map corresponding to the after-denoise state (blue) in a). e) Conductance map measured by C-AFM scanning corresponding to the before-denoise state (red) in b). f) Conductance map corresponding to the after-denoise state (blue) in b). All scale bar: 10 nm.

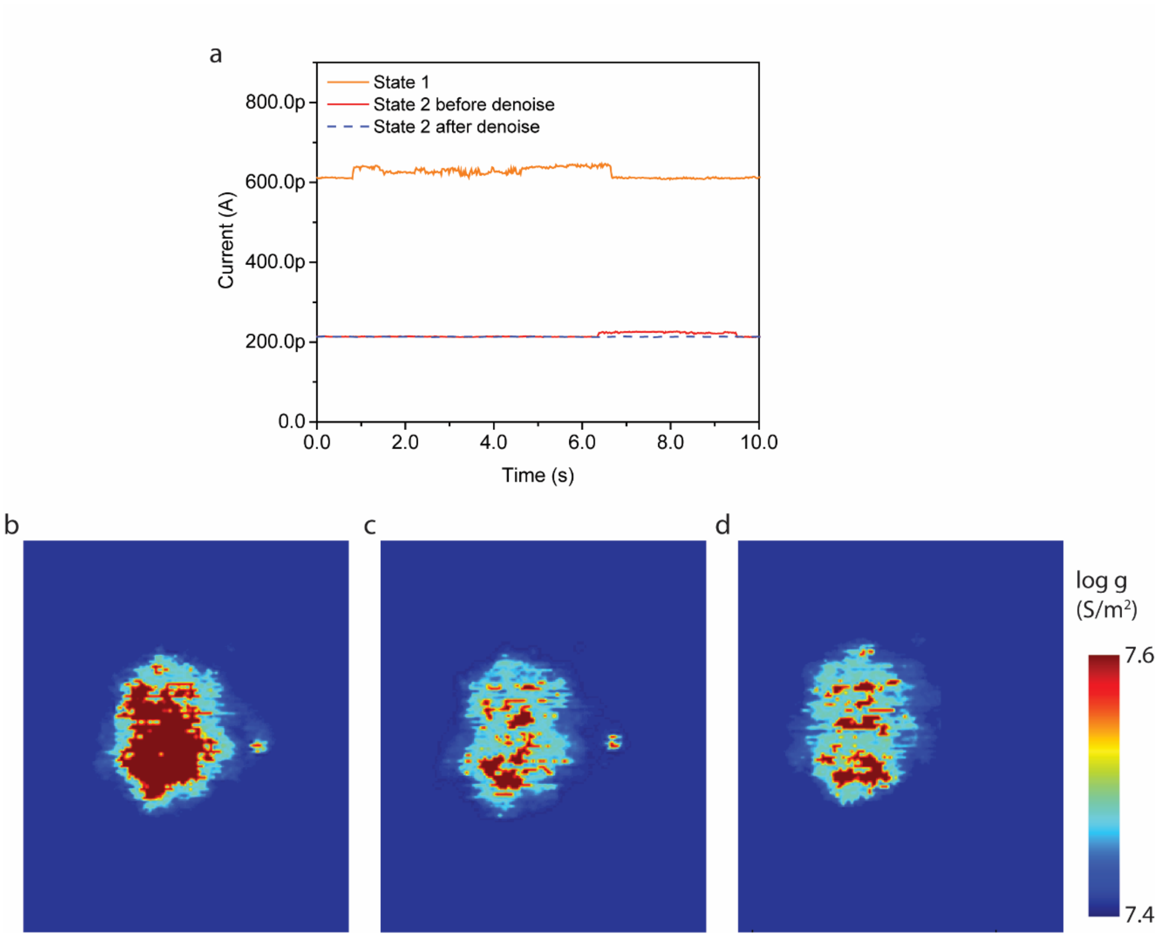

Fig. S8 RTNs only appear when an incomplete filament exists. a) the current reading by C-AFM of a noisy resistance state (state 1, plotted in orange solid line), the state reset from state 1 (state 2, plotted in red solid line) and the stabilized state 2 (plotted in blue dashed line). b) Conductance map scanning of the filament region at state 1. c) Conductance map scanning of the filament region at state 2 before stabilization. d) Conductance map scanning of the filament region at state 2 after stabilization.

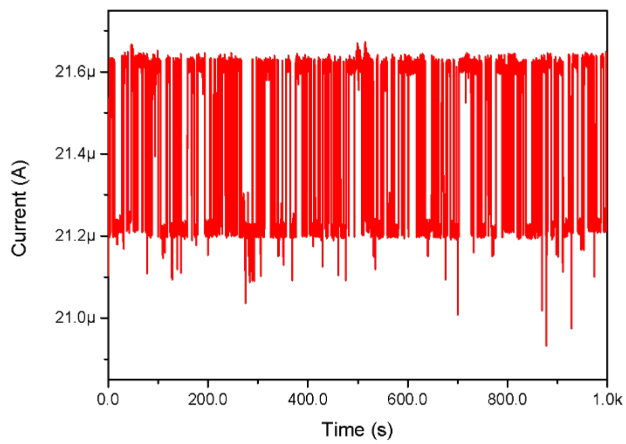

Fig. S9 RTNs of a standard memristor measured for 1000s measurement by a constant 0.2V voltage. The result indicates that without an external voltage to denoise, memristor RTNs can persist for a long time without any trend to stabilize itself into either of the two states of the fluctuation.

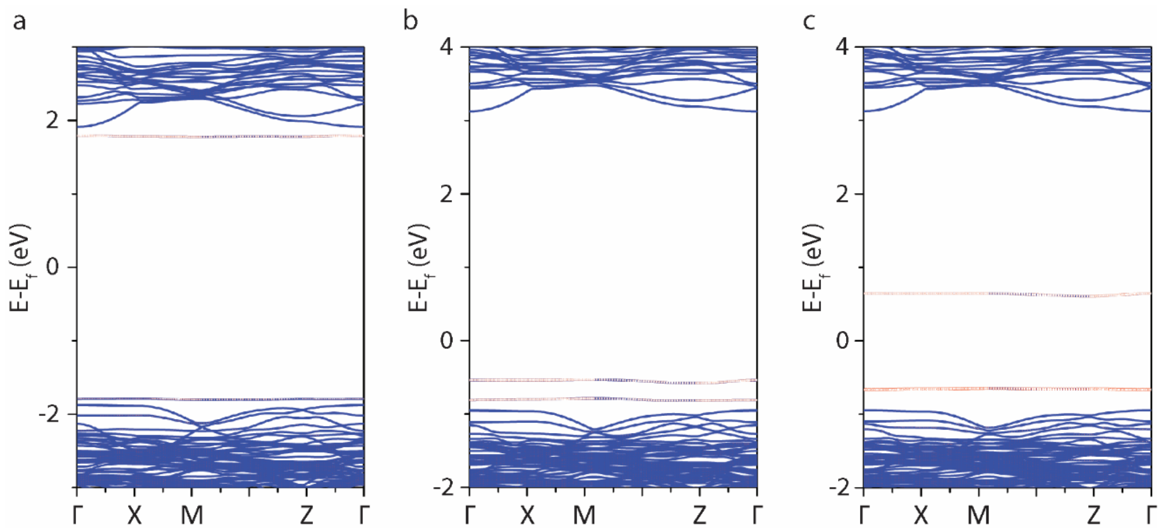

Fig. S10. The band structure of HfO_2 with an oxygen interstitial defect with/without a trapped electron. (a) The band structure without a trapped electron. The red dots represent the projection weight to the interstitial oxygen atom. (b) and (c) Spin-polarized electronic band structure of the defect with one trapped electron. (b) Spin up. (c) Spin down.

S11

It can be observed the distribution curves of τ_0 and τ_1 first increase with the time then decrease. We defined an empirical coefficient τ_d and modified the distribution function as:

$$f_c(t) = \frac{1}{(\tau_c - \tau_d)} e^{-\frac{t}{\tau_c}} (1 - e^{-\frac{\tau_c - \tau_d}{\tau_c \tau_d} t})$$

$$f_e(t) = \frac{1}{(\tau_e - \tau_d)} e^{-\frac{t}{\tau_e}} (1 - e^{-\frac{\tau_e - \tau_d}{\tau_e \tau_d} t})$$

Where τ_c, τ_e are the average intrinsic capture time and emission time, respectively, of the oxygen interstitial defects, functions f_c, and f_e are the simulated distribution functions of the capture time and emission time, respectively, τ_d is the time delay, which is set as 1 ms and 0.5 ms for f_e and f_c, respectively.

Reference

[1] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," *Physical review b,* vol. 59, no. 3, p. 1758, 1999.

[2] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Physical review B,* vol. 54, no. 16, p. 11169, 1996.

[3] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Physical review letters,* vol. 77, no. 18, p. 3865, 1996.

[4] H. J. Monkhorst and J. D. Pack, "Special points for Brillouin-zone integrations," *Physical review B,* vol. 13, no. 12, p. 5188, 1976.

[5] J. Lyons, A. Janotti, and C. Van de Walle, "The role of oxygen-related defects and hydrogen impurities in HfO2 and ZrO2," *Microelectronic engineering,* vol. 88, no. 7, pp. 1452-1456, 2011.

[6] S. Monaghan, P. Hurley, K. Cherkaoui, M. Negara, and A. Schenk, "Determination of electron effective mass and electron affinity in HfO2 using MOS and MOSFET structures," *Solid-State Electronics,* vol. 53, no. 4, pp. 438-444, 2009.

[7] X. Zhao and D. Vanderbilt, "First-principles study of structural, vibrational, and lattice dielectric properties of hafnium oxide," *Physical Review B,* vol. 65, no. 23, p. 233106, 2002.