Supporting Information

Biomimetic-Voltage Nitride-Enhanced Memristor Arrays for Femtojoule In-Memory Biosignal Processing

Zijian Wang[1,2], Zhejia Zhang[1,2], Guobin Zhang[1,2,3], Xuemeng Fan[1,2], Pengtao Li[1,2,3], Baicheng Zhu[1,2], Yi Tong[4], Panpan Zhang[5], Dawei Gao[1,2], Bin Yu[1], Jiuren Zhou[6*], Qing Wan[3*] and Yishu Zhang[1,2,3*]

[1]College of Integrated Circuits, Zhejiang University, ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou, Zhejiang, 311200, China

[2]Zhejiang ICsprout Semiconductor Co., Ltd, Hangzhou, Zhejiang, 310027, China

[3]Yongjiang Laboratory, Ningbo, Zhejiang, 315202, China

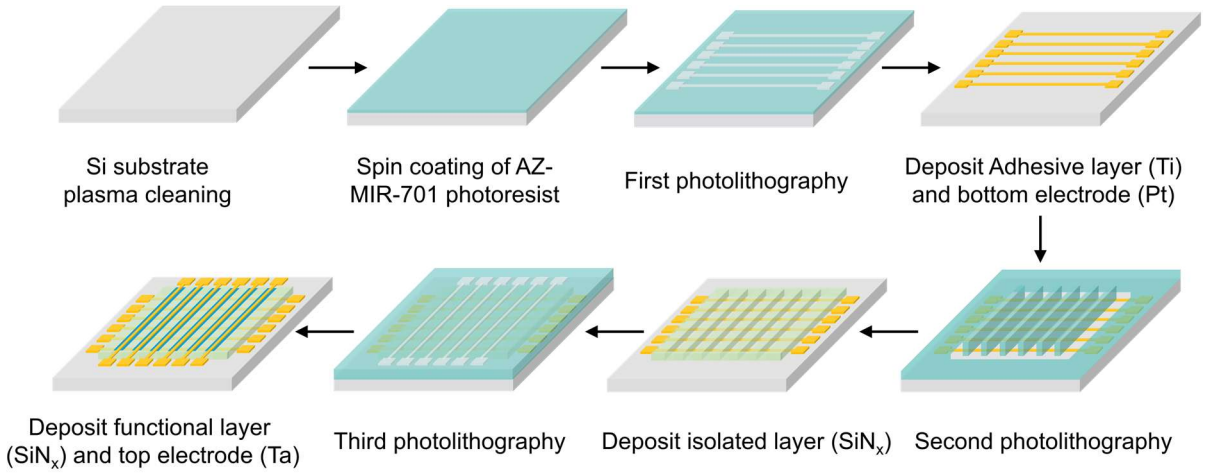
[4]Suzhou lab, Suzhou, Jiangsu, 215123, China

[5]School of Integrated Circuits, Beijing University of Posts and Telecommunications 100876, China

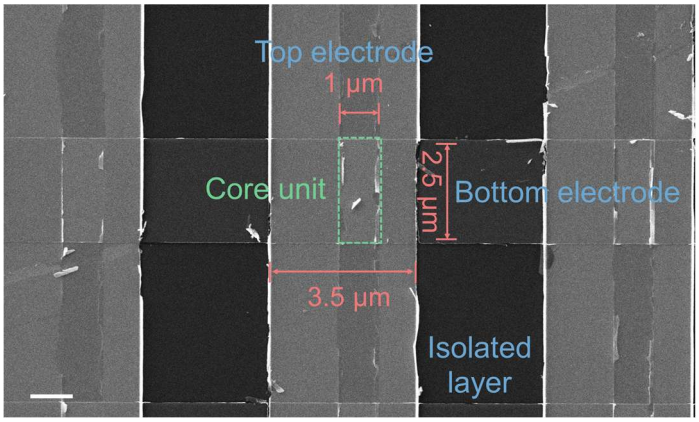
[6]Hangzhou Institute of Technology, Xidian University, Hangzhou, 311231, China.

[*]Corresponding author: zhoujiuren@163.com; qing-wan@ylab.ac.cn; zhangyishu@zju.edu.cn

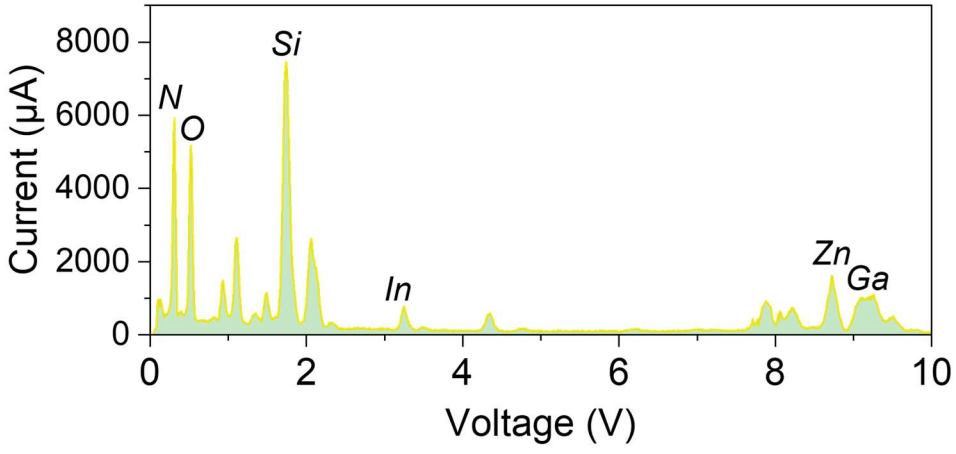
This **Supplementary Information** include:



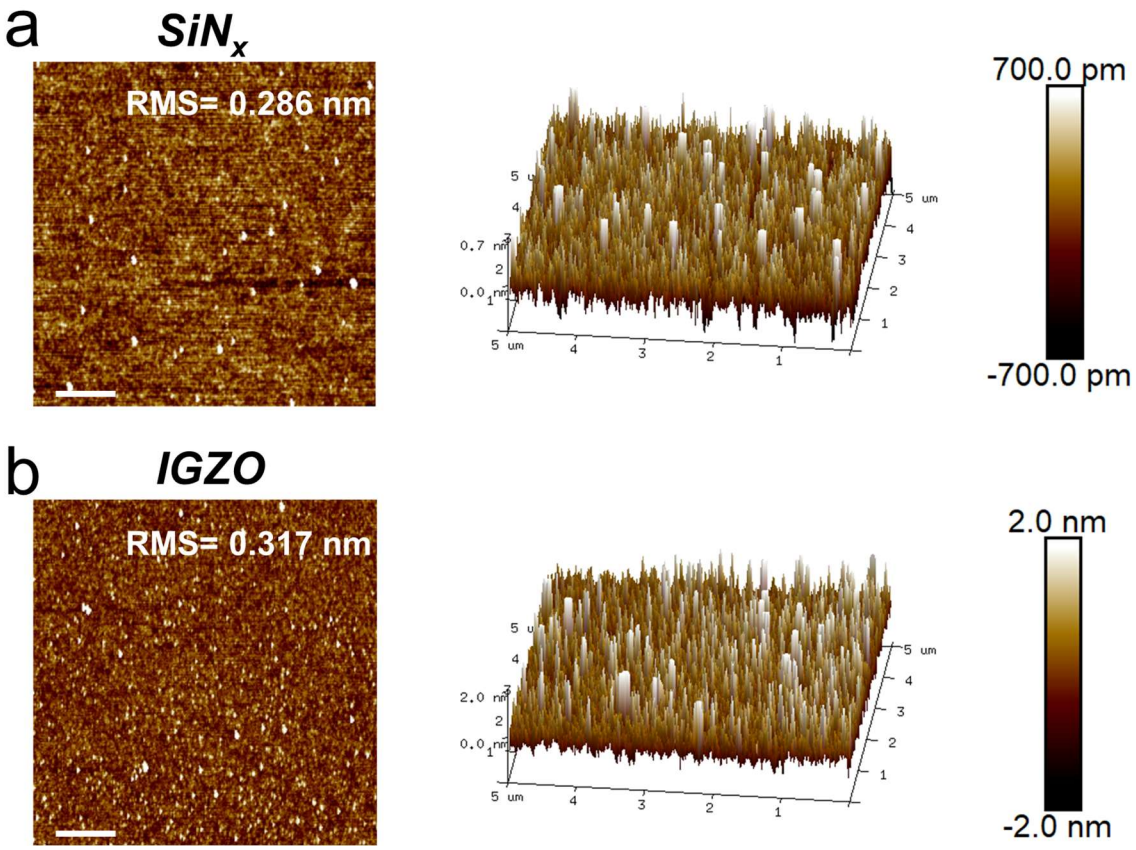
Supplementary Figure 1 The lift-off photolithography process for fabricating the integrated passive array of memristors based on this work. First, a photoresist is spin-coated on a clean SiO_2/p+Si substrate, and the bottom electrode pattern of the array is formed using photolithography. The bottom electrode is then completed through thin film deposition and develop processes. Subsequently, the functional layer, isolated layer and top electrode of the array are fabricated using the same process.



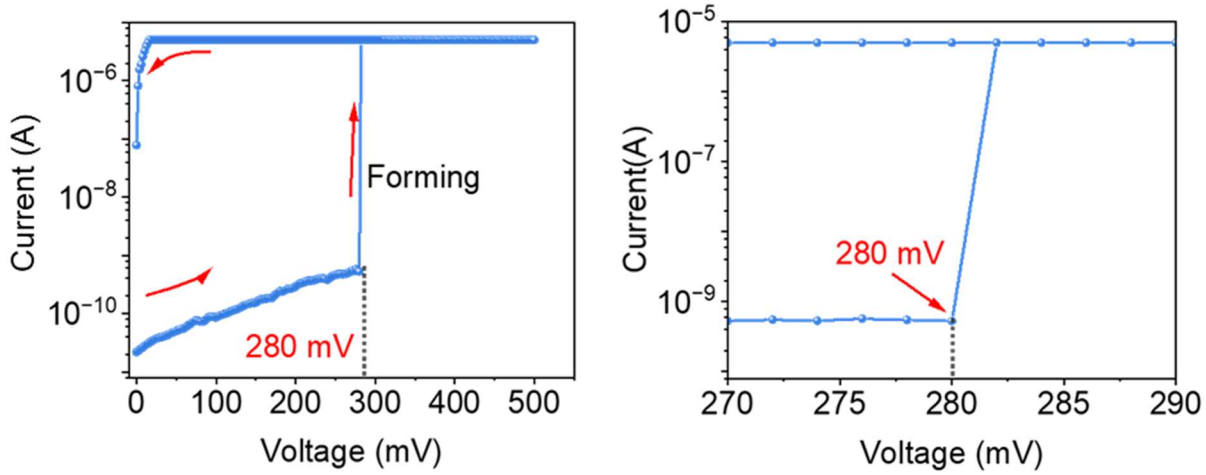
Supplementary Figure 2 Amplified section of the crosspoint positions of the passive array, where the bottom electrode has a line width of 2.5 μm, and the functional layer and top electrode each have a line width of 1 μm (scale bar: 1 μm).



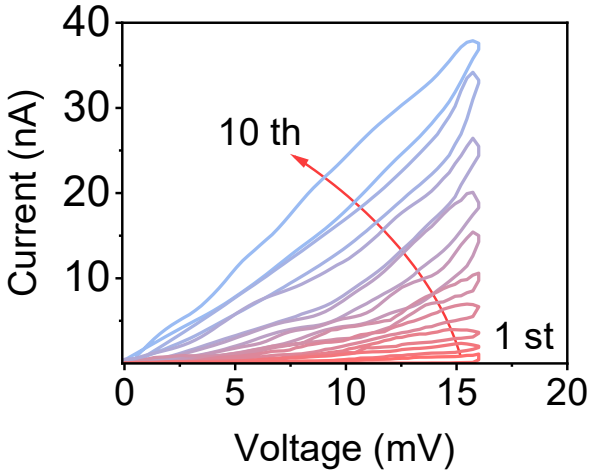
Supplementary Figure 3 The EDS spectrum of the device functional layer, showing peaks for N, Si, In, Zn, Ga, and O, confirming the composition of the SiN_x/IGZO functional layer.



Supplementary Figure 4 Characterization of Device Film Morphology. a AFM characterization of SiN_x and **b** IGZO thin film surfaces, with RMS roughness of 0.286 nm and 0.317 nm, respectively, demonstrating good surface uniformity of the films (scale bar: 100 μm).

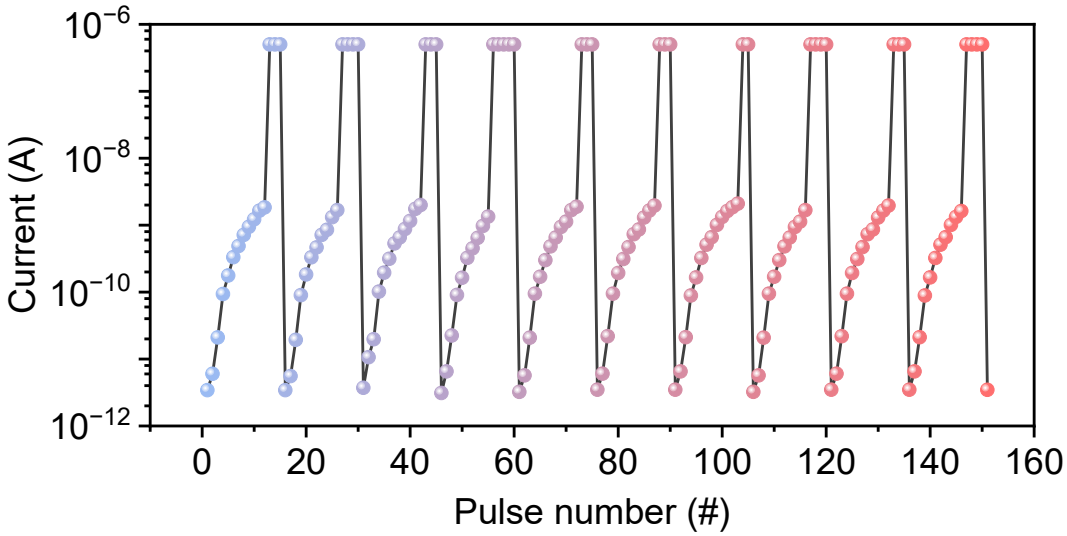


Supplementary Figure 5 I-V curve of the device forming process, with a set voltage of 280 mV.

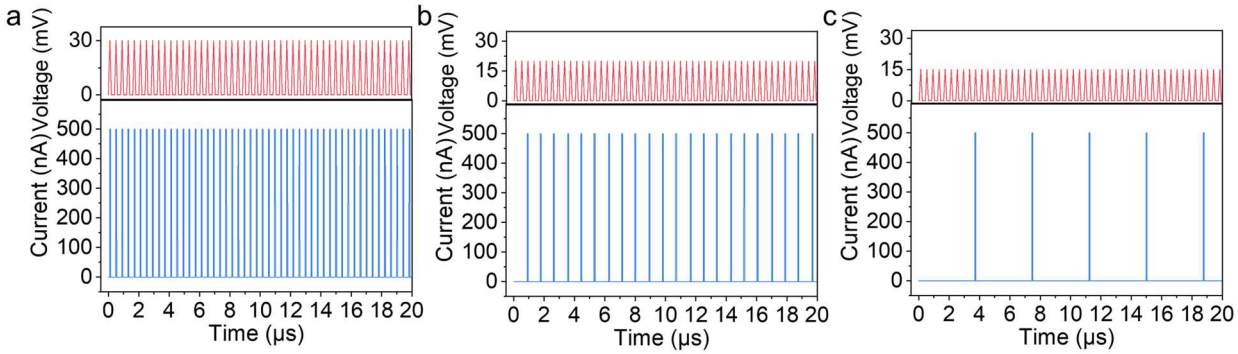


Supplementary Figure 6 The I-V cyclical curve of the device under continuous forward voltage sweep. As the applied voltage approaches the threshold, the conductivity of the device continually increases with each cycle of the applied voltage.

It is noteworthy that the difference in current between the first and tenth scans can reach several times—this cumulative change cannot be explained by external capacitive effects. The underlying physical mechanism originates from the 'memory reconstruction' of the oxygen vacancy-nitrogen vacancy network in the IGZO/SiN_x bilayer structure: each forward voltage scan drives a certain number of vacancies to migrate and accumulate at the heterojunction interface; although the applied voltage magnitude is below the formal set threshold of the device, this subthreshold stimulation can still induce local redistribution of vacancies and partial interconnection, thereby forming an initial conductive pathway in the interface region. The phenomenon of conductive filament conduction gradually enhancing the device's conductivity under continuous scanning has also been widely reported[1–5]. The corresponding expansion of the hysteresis region further verifies the authenticity and effectiveness of this cumulative reconstruction from another perspective: not only does the entire hysteresis loop shift upward (directly reflecting a systematic increase in the average conductivity of the device), but the area enclosed by the hysteresis loop also shows a significant expanding trend, indicating that the device is gradually approaching a 'critical point' from a high-resistance state. Once the accumulated vacancy migration and network reconstruction surpass the critical threshold, subsequent scans will trigger a rapid transition to a low-resistance state, which is completely consistent with the resistance switching behavior ultimately observed in continuous scanning experiments, thus presenting a complete and self-consistent physical picture.



Supplementary Figure 7 Current variations of the device under continuous positive pulse sequences. Under fixed compliance current of 100 nA, triangular pulse sequences with 16 mV amplitude, 500 ns width, and 1 μs with hold voltage of 10 mV, under a fixed compliance current (CC) of 100 nA. After every 15 pulses, the applied voltage is set to zero for 10 ms, after which the next pulse train is tested. The 16 mV sub-threshold stimulation mimics the integrate-and-fire mechanism of biological neurons where individual pulses cannot directly trigger complete resistive switching, but cumulative multi-pulse excitation gradually drives the device toward LRS through progressive oxygen vacancy migration and accumulation at the IGZO/SiN_x interface. The stepwise conductance increase reflects incremental conductive pathway formation, with the device exhibiting volatile behavior and spontaneous return to HRS due to the fragile filament structure under low current compliance, analogous to neuronal leakage current behavior. This cumulative activation process enables sensitive detection of weak signal features essential for neuromorphic audio processing applications.

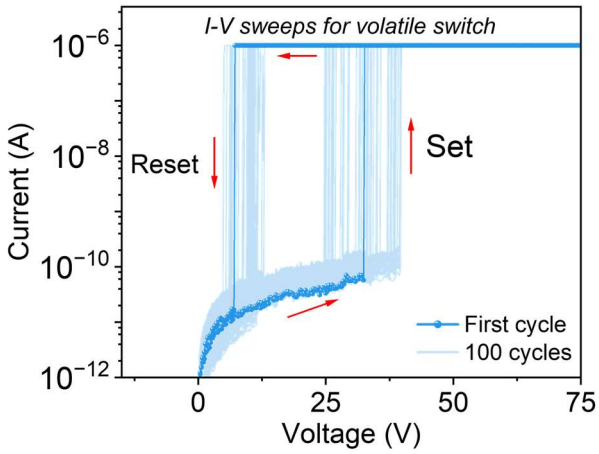


Supplementary Figure 8 Accumulation-type switching behavior of device. With decreasing amplitude of the applied triangular pulse train, **a** device activation transitions from consistent response per pulse at 30 mV, **b** to intermittent response (every 3–4 pulses) at 18 mV, **c** and finally to sparse response (every 10-12 pulses) at 15 mV, demonstrating accumulation switching behavior. CC is defined as 0.5 μA by configuring the pulse testing measurement range. The use of triangular pulses avoids generating enormous transient displacement current spikes and readily avoids high-frequency ringing and voltage overshoot. At the same time, we further compute the constant displacement current arising from capacitive coupling that may mask the device's true response (seen in *Displacement Current from Capacitive Coupling* section in *Supporting Information*).

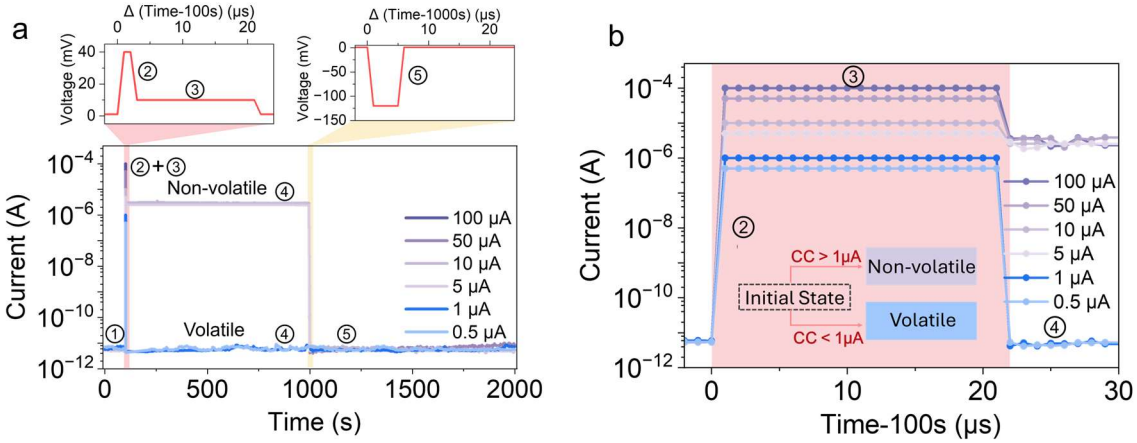
At 30 mV, the device exhibits robust volatile switching characterized by reliable pulse-to-pulse correspondence, indicative of sustained ion migration and efficient charge transfer. This behavior confirms stable conductive filament formation under repeated electrical stimulation.

When the voltage is reduced to 18 mV, activation becomes intermittent under continuous pulse sequences, accompanied by oscillatory charge transfer dynamics. This suggests the device approaches an operational threshold where cumulative pulse energy progressively modulates vacancy migration within the IGZO layer, yet remains insufficient for consistent switching.

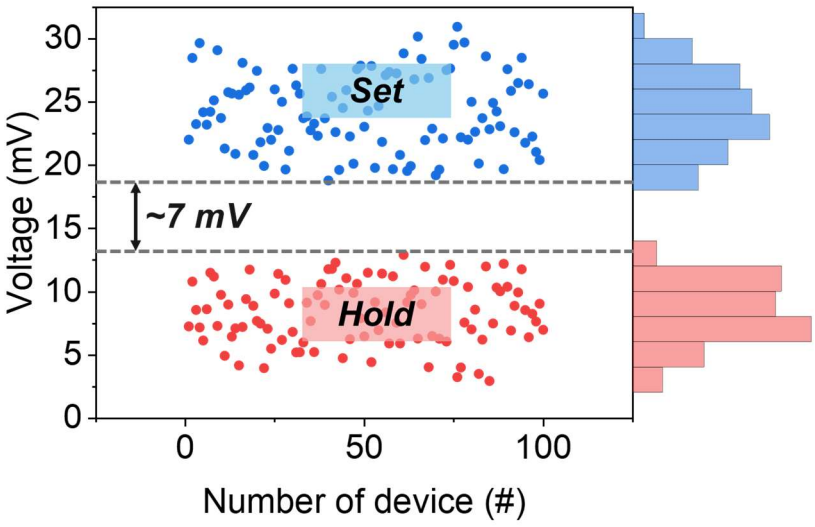
Further reducing the voltage to 15 mV results in sparse, stochastic activation even under extended pulse sequences. Here, only high-frequency, low-amplitude stimuli can trigger limited ion rearrangement and vacancy migration, implying that minimal energy per pulse restricts conductive pathway formation to rare statistical events.



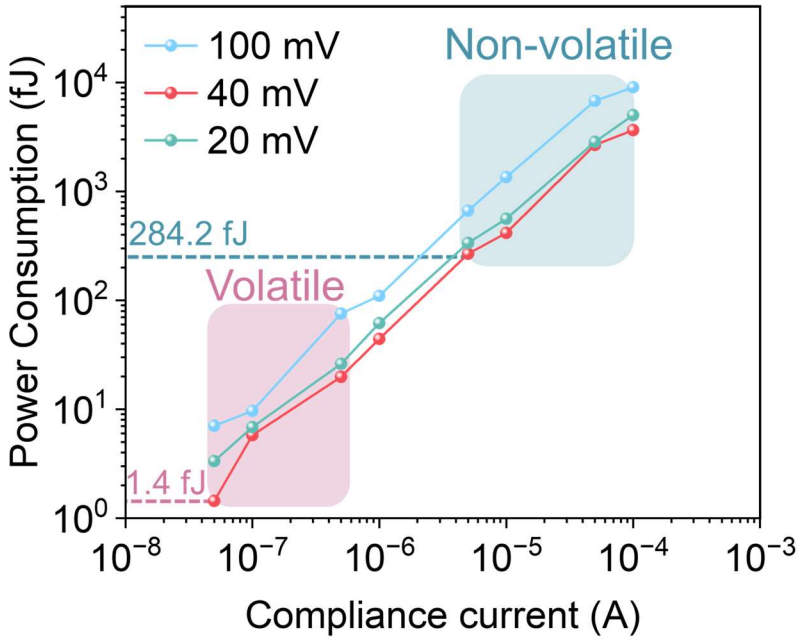
Supplementary Figure 9 I-V characteristics over 100 cycles under the CC of 1 µA, with a minimum set voltage of 22 mV and a hold voltage of 8 mV.



Supplementary Figure 10 Device resistance state retention under different CC. a At stage ①, 1 mV reading voltage is applied to the initial device, maintaining it in HRS. At stage ②, 40 mV reset voltage is applied to switch the device to LRS under different CC. At stage ③, a 10 mV hold voltage stabilizes the device in LRS. At stage ④, the test voltage returns to 1 mV reading voltage, where devices with CC between 100 µA and 5 µA maintain a stable LRS, demonstrating non-volatility (purple curve), while devices with CC from 1 µA to 0.5 µA switch back to HRS, showing volatility (blue curve). At stage ⑤, all devices are switched to LRS with -120 mV reset voltage. **b** The device's resistance state under phases ② and ③. To highlight the device's changes at the microsecond level, the y-axis has been adjusted by subtracting the initial test duration (100 s). During the CC control phase, a 10 mV holding voltage is applied to maintain all devices in the LRS state throughout the compliance current modulation process. This holding voltage serves a critical experimental purpose: it prevents devices that are inherently prone to volatile behavior from randomly returning to HRS during CC control, which would compromise the ability to distinguish between CC-induced volatility and spontaneous volatility due to lack of voltage maintenance. By providing this electrical anchoring state, the holding voltage ensures that compliance current becomes the sole experimental variable, enabling unambiguous attribution of the observed volatile/nonvolatile differences to the specific CC settings rather than other experimental conditions. It shows that devices switching to LRS at lower CC values will self-reset to HRS after the hold voltage is removed, because the conductive filaments generated at low CC values cannot maintain stability and spontaneously rupture upon voltage removal. Thus, by adjusting CC, we can switch the state of device between long-term non-volatile storage and volatile resistance switching.



Supplementary Figure 11 Statistical analysis based on 100 devices randomly selected from a 32×32 array. Each device underwent 10 switching cycles, and its lowest set voltage and highest hold voltage were extracted. A ~7 mV margin between maximum hold voltage and minimum set voltage prevents unintended switching during hold-voltage application.



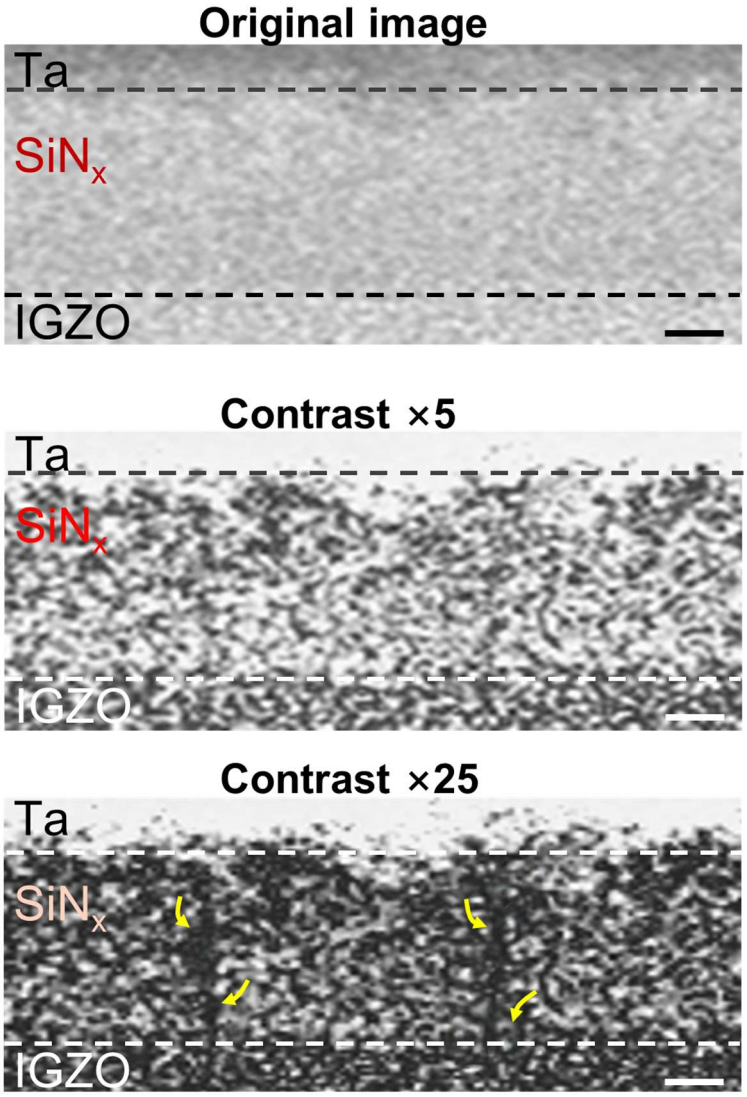
Supplementary Figure 12 The trend of power consumption of the device under different compliance currents and operating voltages for non-volatile switching. Lower compliance currents reduce operational power consumption; however, when combined with lower operating pulse voltages, this leads to an increase in the number of pulses required for switching, which in turn raises overall power consumption. According to the energy formula:

$$E=I_{cc}\cdot V\cdot T$$

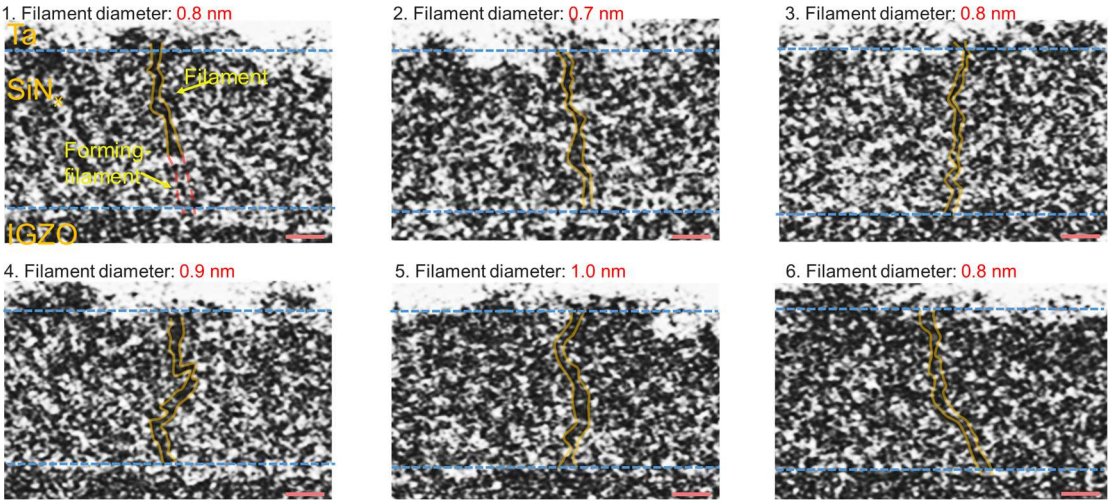
Where I_{cc} is CC, V is the pulse amplitude and T is the switching speed. The device achieves a minimum operating voltage as low as 1.44 fJ for volatile switching (40 mV pulse voltage; 500 nA CC and 72 ns switching speed) and 248.2 fJ for non-volatile switching (40 mV pulse voltage; 5 μA CC and 124 ns switching speed).

Accordingly, deterministic pulsed switching—implemented with 40 mV set pulses—is employed in the recognition region to satisfy stability requirements; whereas the denoising region adopts a low-amplitude accumulative scheme—driven by 15–30 mV pulses along the low-feature-signal pathway—to capture weak features. Through temporal integration, this approach enhances the

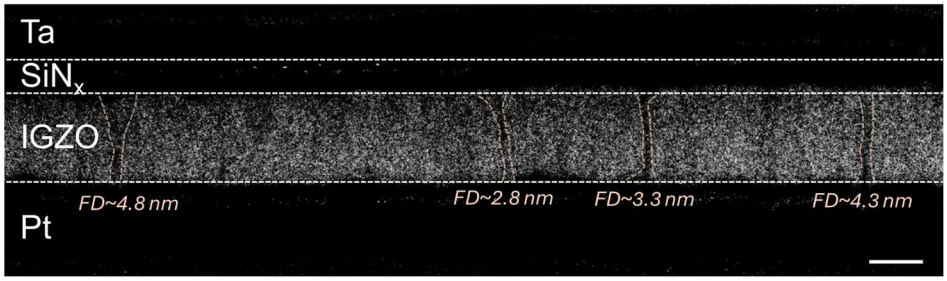
signal-to-noise ratio and feature contrast, thereby achieving a balanced trade-off among system stability, energy efficiency, and functional capability.



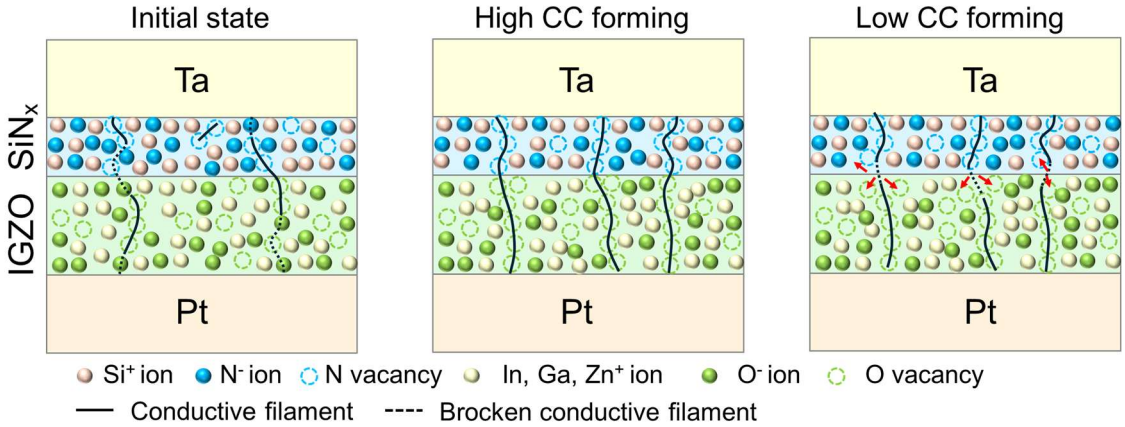
Supplementary Figure 13 TEM cross-sectional characterization of SiN_x layers with varying contrast. a Original TEM characterization of the thin film cross-section; **b** TEM characterization of the SiN_x thin film cross-section after a fivefold increase in contrast, revealing blurred channels; **c** TEM characterization of the SiN_x thin film cross-section after a tenfold increase in contrast, clearly illustrating distinct vacancy channels. Further comprehensive experimental validation of this conclusion is precluded at this stage by the constraints of the available characterization platform (scale bar: 2 nm).



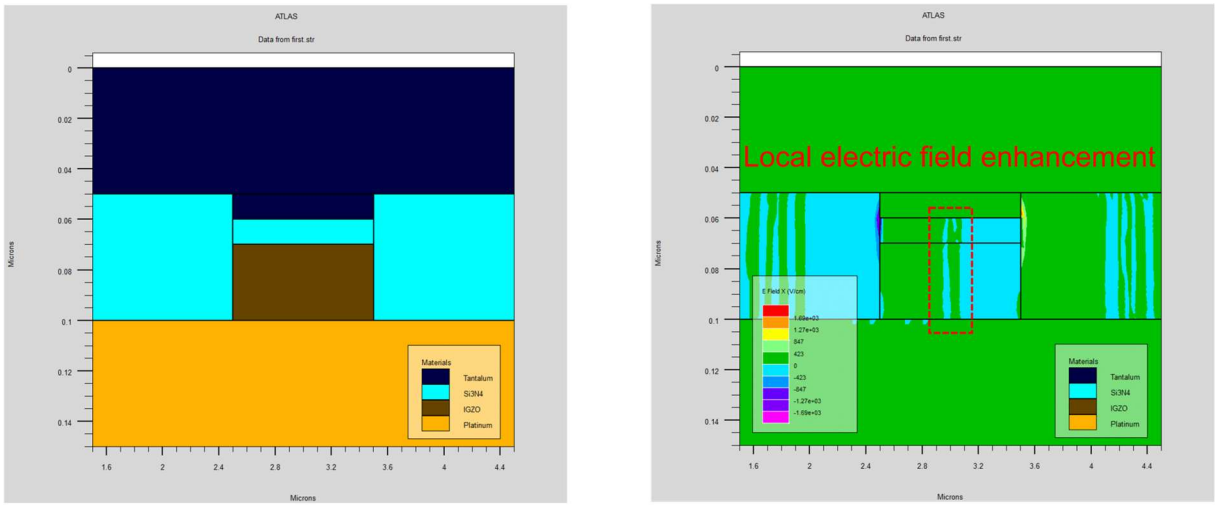
Supplementary Figure 14 Multiple conductive paths exist within the SiNx layer, forming conductive filaments with diameters in the range of 0.7 to 1 nm. A very small fraction have apparent widths as small as about 0.6 nm (scale bar: 2 nm).



Supplementary Figure 15 High-resolution cross-sectional TEM image of the IGZO region. High-contrast filamentary structures extending along the current conduction path with different filament diameter (FD) can be clearly identified at the nanometer scale (scale bar: 15 nm).

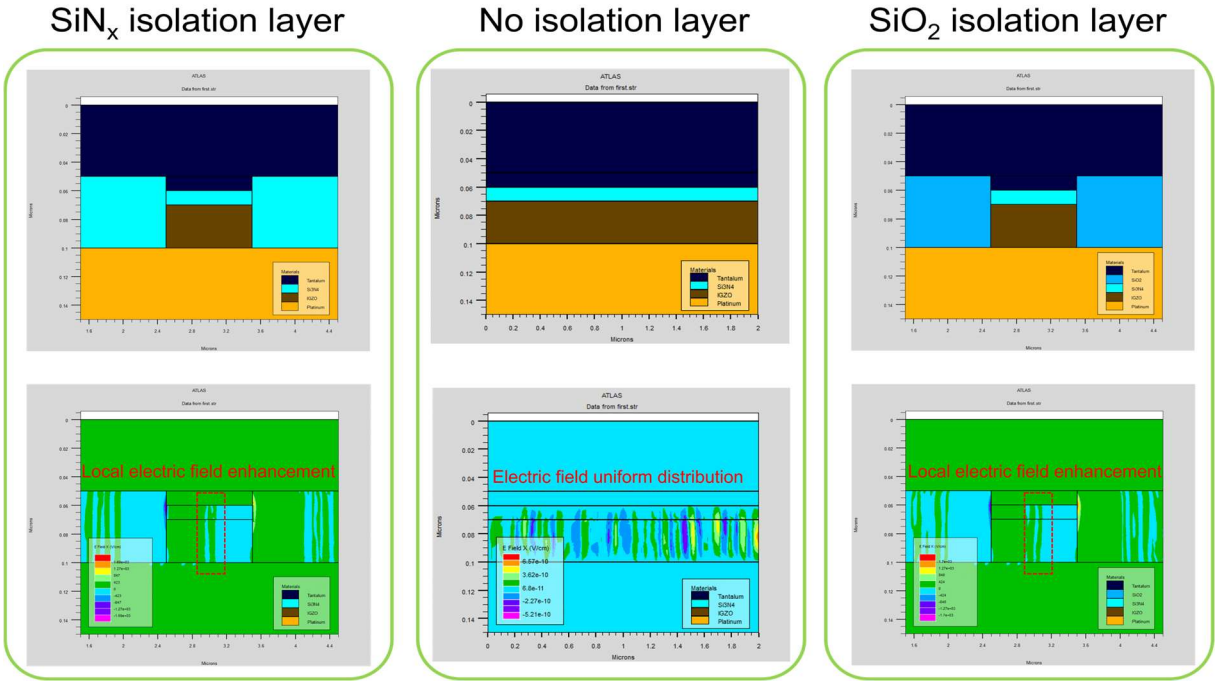


Supplementary Figure 16 The formation and rupture of filaments are regulated by the compliance current (CC), which determines the device's volatility. A higher CC (>1 μA) promotes the growth of robust V_N-rich filaments in SiN_x, supporting stable nonvolatile switching; conversely, a lower CC (≤1 μA) favors V_O dominated metastable filaments in IGZO, which dissipate more readily and thus produce volatile operation.



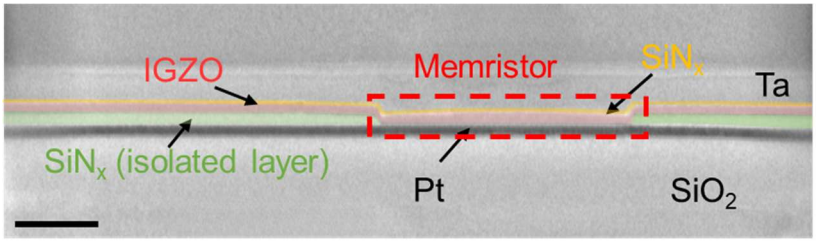
Supplementary Figure 17 The TCAD simulation of our Pt/IGZO/SiN_x/Ta memristor. The result reveals intricate electrophysical dynamics that elucidate the fundamental switching mechanism. By computationally modeling the device's nanoscale architecture, we demonstrate that the unique vertical heterostructure induces highly localized and intensified electric field distributions. Specifically, the interfaces between dissimilar materials (Pt/IGZO, IGZO/SiN_x, and SiN_x/Ta) generate significant dielectric discontinuities, which act as electric field concentration sites. These nanoscale heterojunctions create potential barriers and charge carrier redistribution zones that are critical for resistive switching behavior.

Under modest applied voltages, intense electric field gradients within the device heterostructure drive a cascade of nanoscale processes. Specifically, these gradients initiate selective oxygen vacancy migration and dynamic ionic rearrangement within the IGZO layer, nucleating conductive filaments. Critically, the SiN_x layer mediates this process by concentrating and confining the electric field, thereby enabling precise electrostatic control over filament development. This field-guided mechanism orchestrates the systematic nucleation and spatially confined extension of conductive filamentary networks, transforming inherently stochastic resistive switching into a deterministic process. The vertically confined field profile ensures high spatial selectivity during filament formation, suppressing lateral dispersion and enhancing switching uniformity.

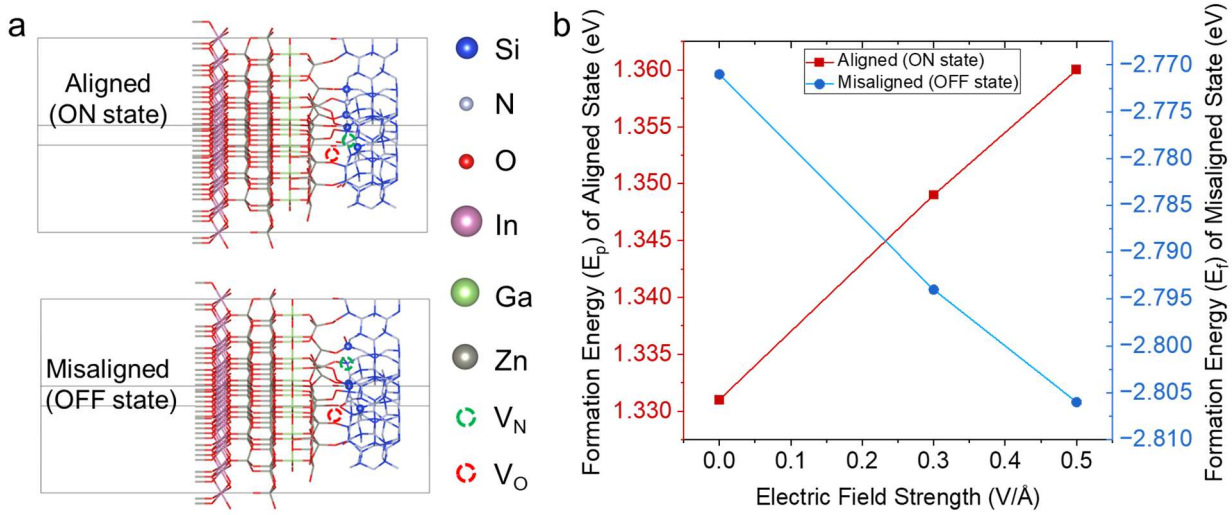


Supplementary Figure 18 Distribution of internal electric fields in the device under different isolation layers. When SiN_x and SiO_2 serve as isolation layers, localized electric field enhancement

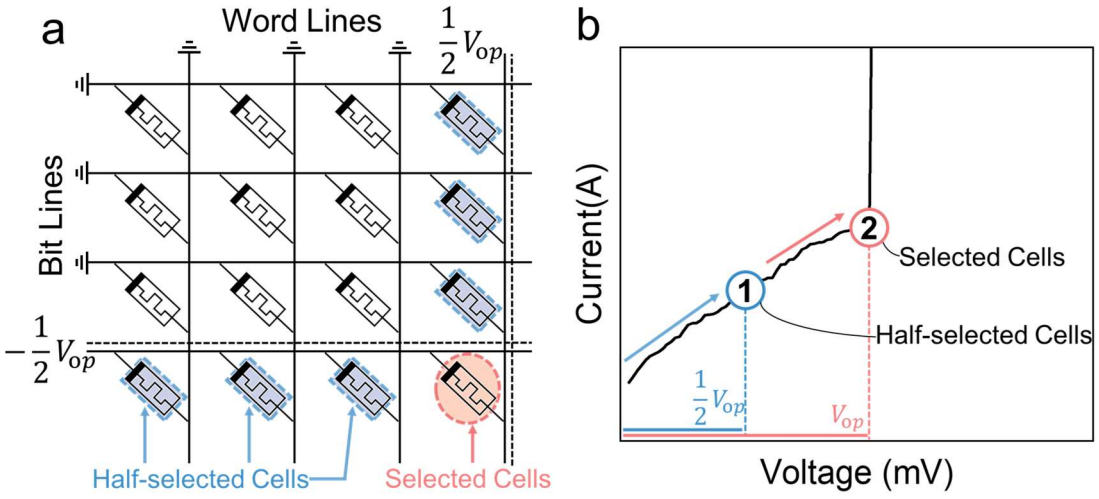
traversing the device electrodes is generated inside the device, whereas when no isolation layer is present, there is no localized field enhancement inside the device.



Supplementary Figure 19 Low magnification TEM cross-sectional image of the fabricated device. The red dashed box contains the memristor prepared in this work, which can be observed to be enveloped by SiN_x isolation layers on both sides (scale bar: 250 nm).

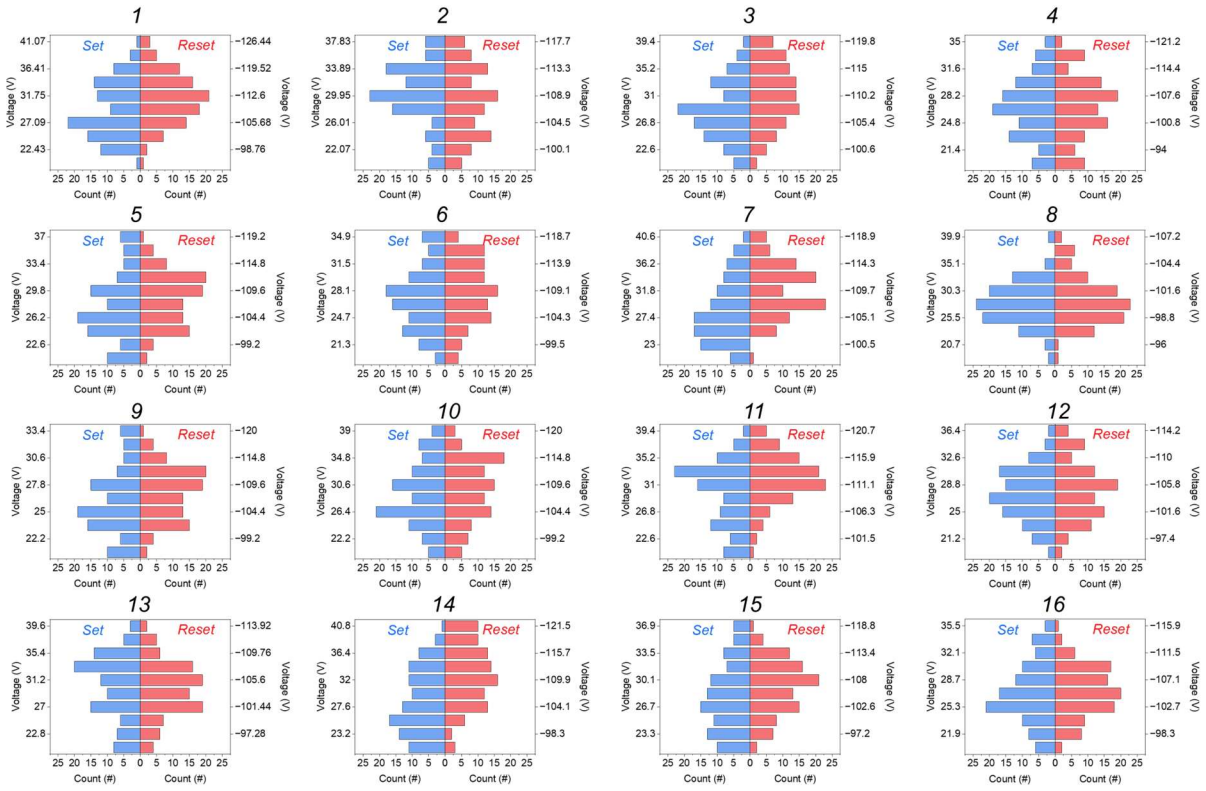


Supplementary Figure 20 DFT calculations of energies for aligned and misaligned states under different electric fields a. Molecular structures of aligned and misaligned states in DFT calculations. **b.** Electric field driven forming and destabilization energy of conductive filament.

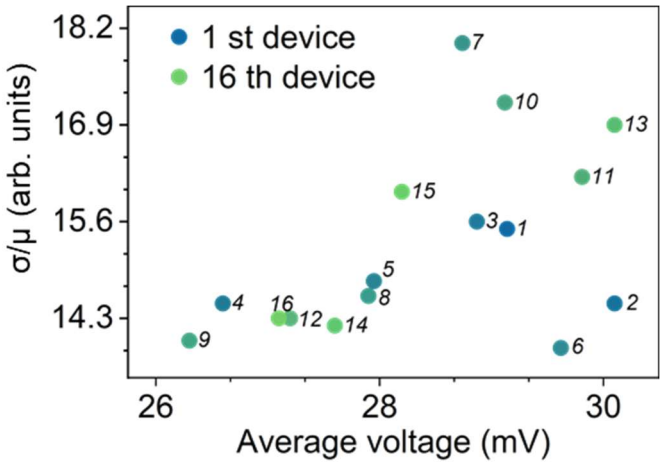


Supplementary Figure 21 1/2 V_{op} voltage distribution scheme to program target cells. a Based on this voltage distribution scheme, positions within the array are categorized into selected cells and half-

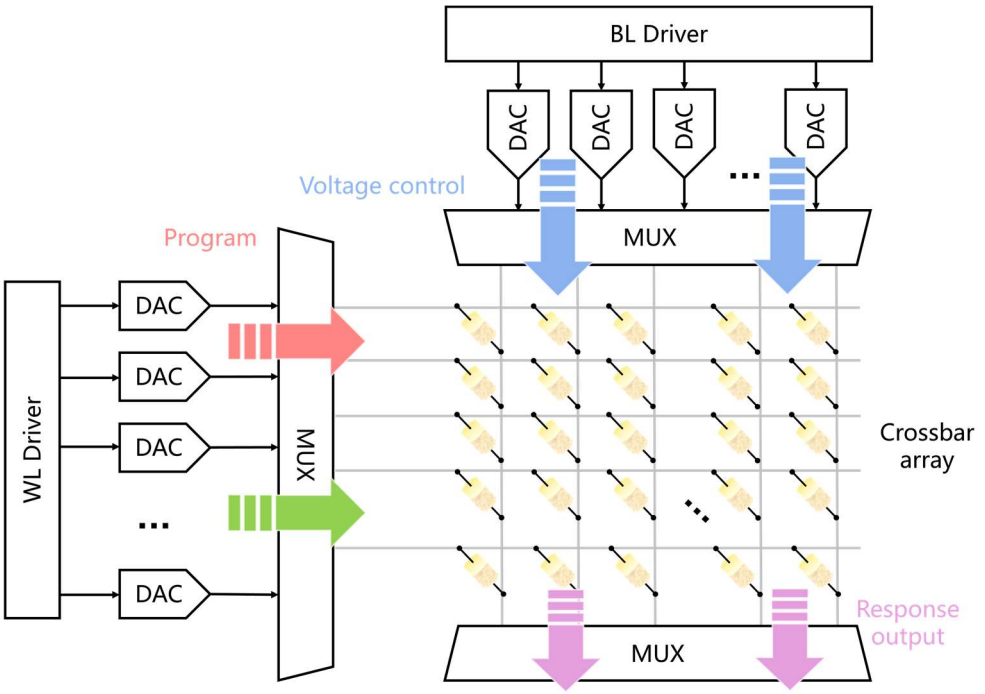
selected cells. Under this scheme, selected cells receive sufficient voltage to meet the required setting criteria, while half-selected cells fail to reach this threshold and thus cannot be successfully programmed. **b** In the I-V graph, the voltage applied to the half-selected cells is represented at position ①, indicating that they did not reach the turn-on voltage and therefore could not be set. On the other hand, the applied voltage for the selected cells is shown at position ②, successfully reaching the turn-on voltage and thus allowing for successful programming.



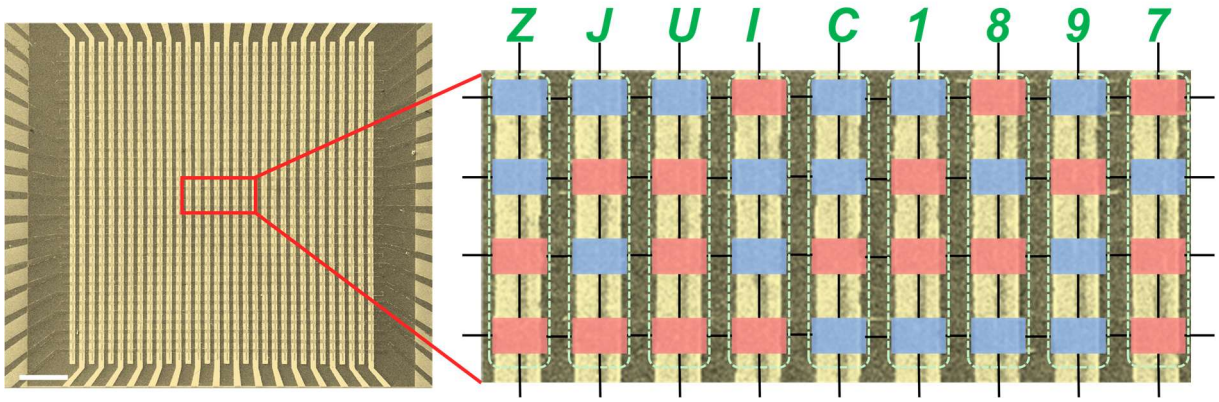
Supplementary Figure 22 Distribution of set and reset voltages for 16 devices in passive array. The slight differences in resistance switching behaviors among various devices in the passive array reflect their uniformity and stability in performance.



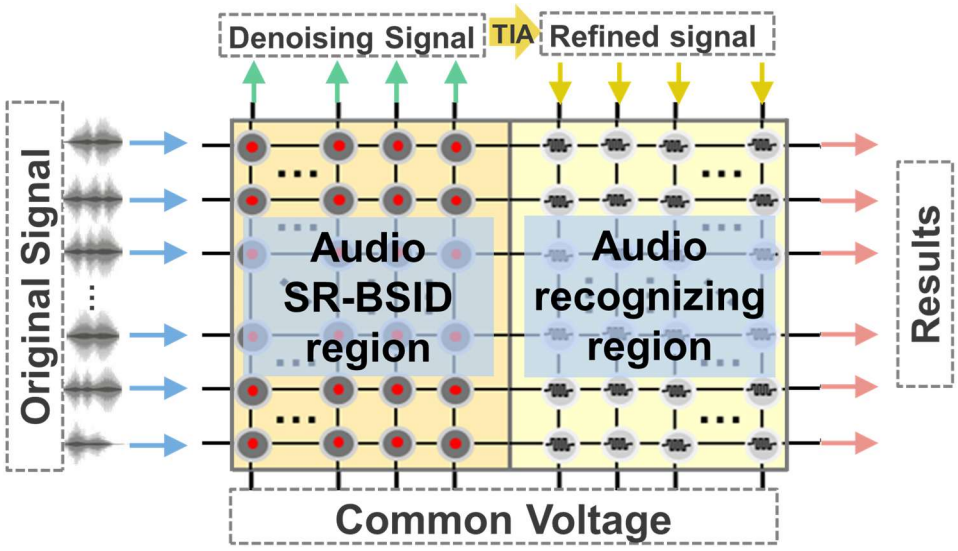
Supplementary Figure 23 Distribution of mean voltages and σ/μ for 16 devices in the passive array. The distribution of σ/μ basically satisfies the requirement of strong noise suppression and weak signal accumulation in the array's denoising region proposed in the *Methods* section.



Supplementary Figure 24 Passive array programming architecture based on DAC and MUX. The information received by the sensing end is converted into a voltage signal by the DAC and then routed through the MUX to select specific channels, generating sequential input pulses. These signals interact within the array and ultimately output responses via the MUX below. The structure of the array facilitates on-chip information processing, enabling data writing and row-wise reading by controlling the rows and columns.



Supplementary Figure 25 Location-specific resistive encoding of alphanumeric information. By leveraging the resistance values at different locations in the encoding array, the information 'Z J U I C 1 8 9 7' can be input into the array based on binary encoding. Specifically, each resistance value at a location represents a particular state, enabling effective encoding of letters and digits (scale bar: 30 μm).

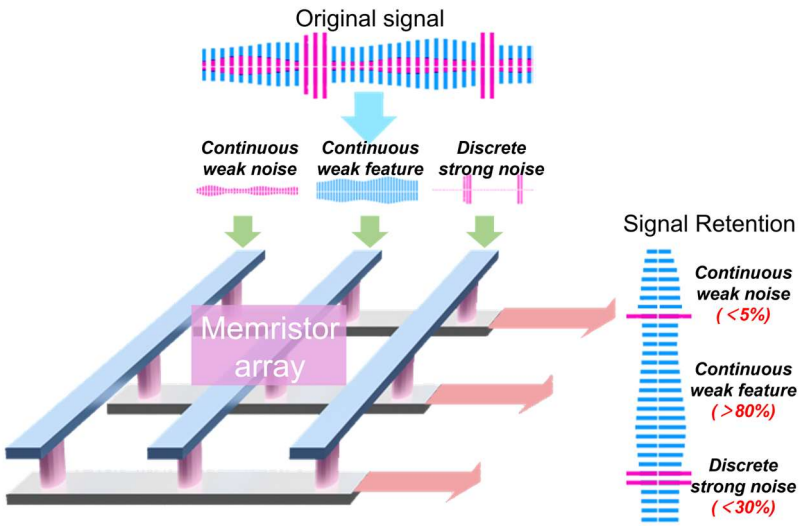


Supplementary Figure 26 Array functional area division, parameter programming, and signal path. For audio recognition, the array is partitioned into two 16×32 sub-arrays. One part is used to implement the SR-BSID process for denoising the audio signal, while the other part is based on FCNN for audio recognition.

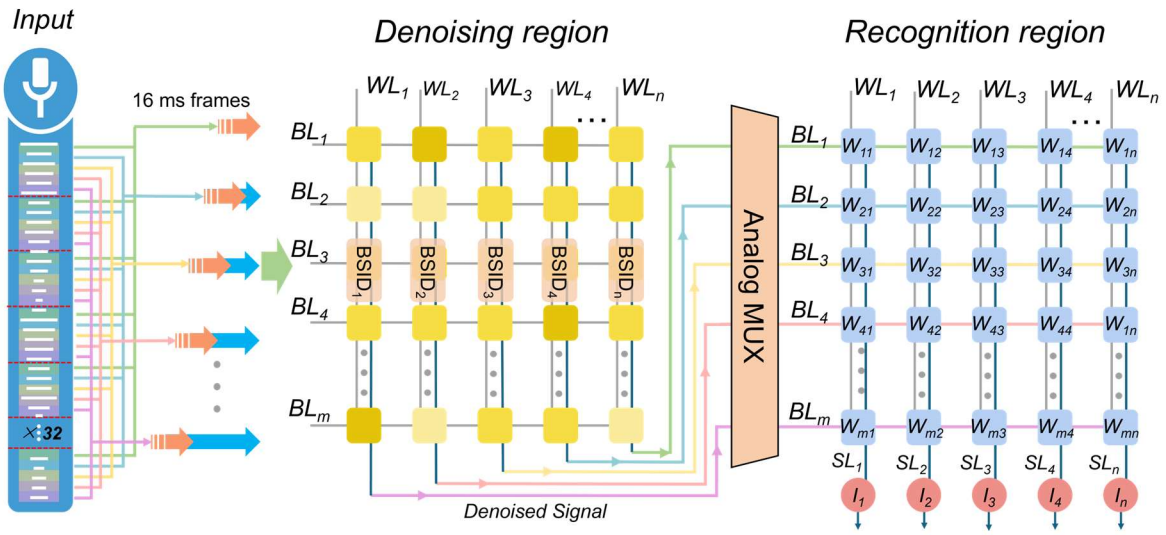
It must be emphasized that crosstalk paths that may arise during array operations must be avoided. A stable crosstalk path requires that, on the same row, column, or diagonal, at least three devices remain in the low-resistance state without any applied voltage and are positioned relative to the target cell so as to form a continuous parasitic current loop. In our design, this condition is fundamentally precluded by two key features:

1. **Volatile Operation in the Denoising Region**: All devices in this region operate in a volatile mode, spontaneously resetting to a high-resistance state once voltage is removed. This eliminates the possibility of maintaining the persistent low-resistance configurations necessary for sustained crosstalk.
2. **Column-Wise Operation Across the Array**: Both during denoising and recognition, all read and write operations are performed on a per-column basis. In this scheme, only devices within the selected column are activated simultaneously. Even if devices in unselected columns remain in a low-resistance state, no complete parasitic path can form with the active column, as illustrated in the accompanying figure. This columnar access method fundamentally alters current flow and isolates the signal path from potential sneak currents at the topology level.

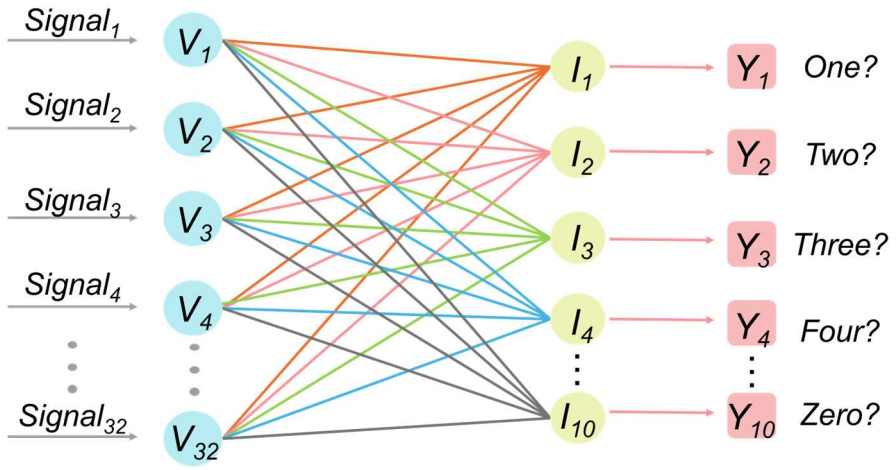
Thus, for the present application and array scale, the combination of volatile device characteristics and column-wise operation provides effective crosstalk suppression without the need for dedicated selectors.



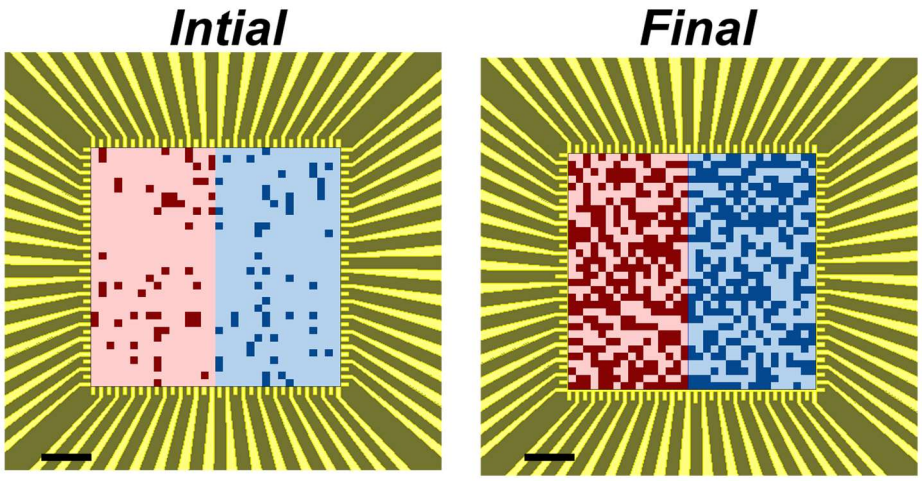
Supplementary Figure 27 Signal processing schematic for noise reduction in a memristor device array. In mixed-signal processing, this system demonstrates high selectivity. Continuous characteristic signals are highly preserved (>80%), while continuous noise and discrete noise are effectively suppressed, with residual rates below 5% and 30% respectively, achieving the goal of extracting target signals from complex backgrounds.



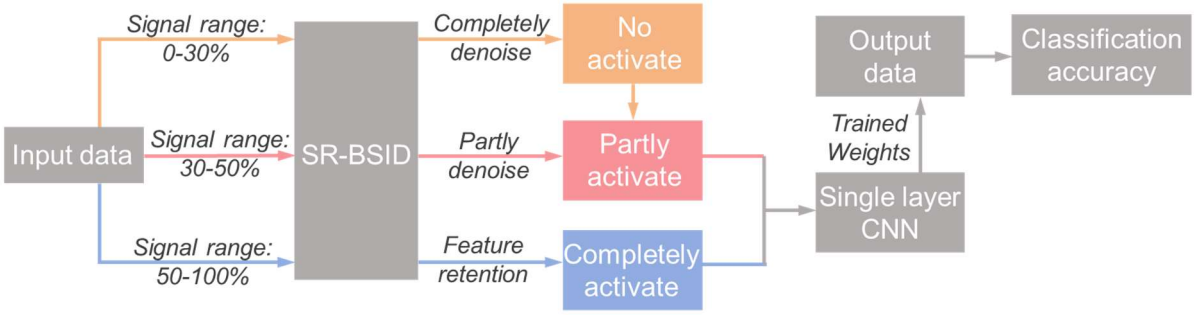
Supplementary Figure 28 Schematic diagram of signal processing flow of audio recognition system with SR-BSID method. Within the time window of the input signal, the pulse train is uniformly divided into 32 equal-duration sub-intervals, each corresponding to one input channel of the recognition array. At the identical temporal position within every sub-interval, a representative 500-microsecond pulse segment is extracted as the feature signal for that channel. This sampling strategy guarantees uniform temporal distribution and preserves the representativeness of the features. After BSID processing, the amplitude information of the extracted 500-microsecond pulse segment is mapped to a voltage range of 10–50 mV. And the cleaned signal is routed through an analog MUX into the recognition array to complete the recognition operation. WL, SL, and BL stand for the word line, source line, and bit line, respectively.



Supplementary Figure 29 Schematic diagram of the readout layer for biological audio signal classification in a single-layer FCNN.



Supplementary Figure 30 Changes in activated site distribution from early training to convergence. During the cold start initialization, only 13.4% of the weighted loci are active at the beginning (scale bar: 30 μm).



Supplementary Figure 31 The scheme of training and testing processes. Input data segmented into 0-30%, 30-50%, and 50-100% signal-intensity tiers undergoes hierarchical processing by the SR-BSID module: the weak-signal tier (0-30%) undergoes complete denoising, the mid-signal tier (30-50%) receives partial denoising with core feature retention, while the strong-signal tier (50-100%) deactivates denoising and enables feature enhancement; the processed data then trains weight parameters through a single-layer CNN, ultimately generating classification accuracy metrics.

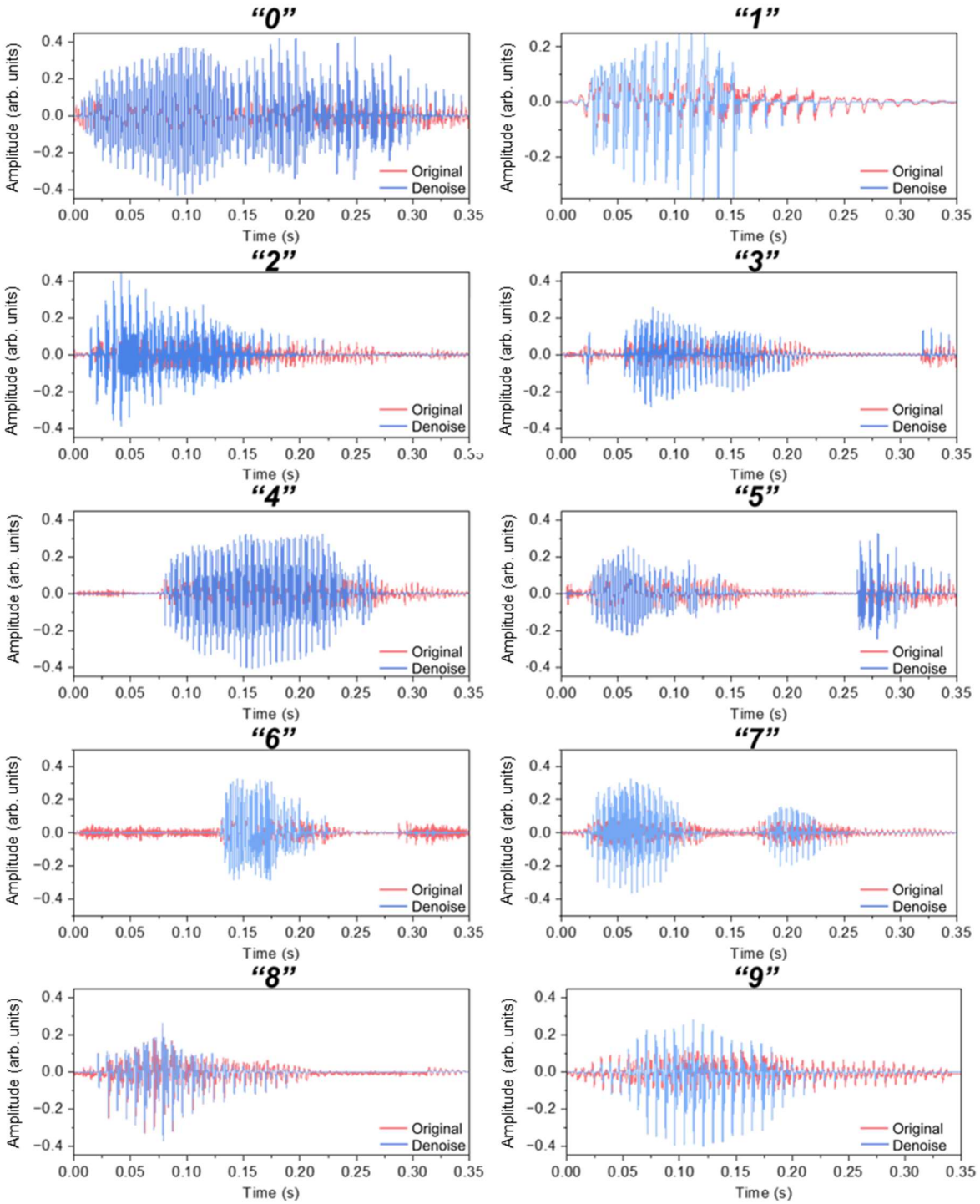
Our signal strength classification strategy is established based on the amplitude distribution characteristics of audio signals after normalization processing. As described in the manuscript, raw audio signals are mapped to an amplitude range of -0.5 to 0.5 through standardization processing.

Through statistical analysis of extensive training samples, we observed that normalized audio signals exhibit distinct amplitude-stratified distributions: strong characteristic signals (such as fundamental frequencies and primary formants) are predominantly concentrated within the 0.2-0.4 amplitude range, while weak characteristic signals and noise components are primarily distributed in the sub-0.2 interval. Based on this distribution characteristic, our denoising region specifically targets audio signals below 0.2 for segmented processing and establishes a three-tier strength classification system.

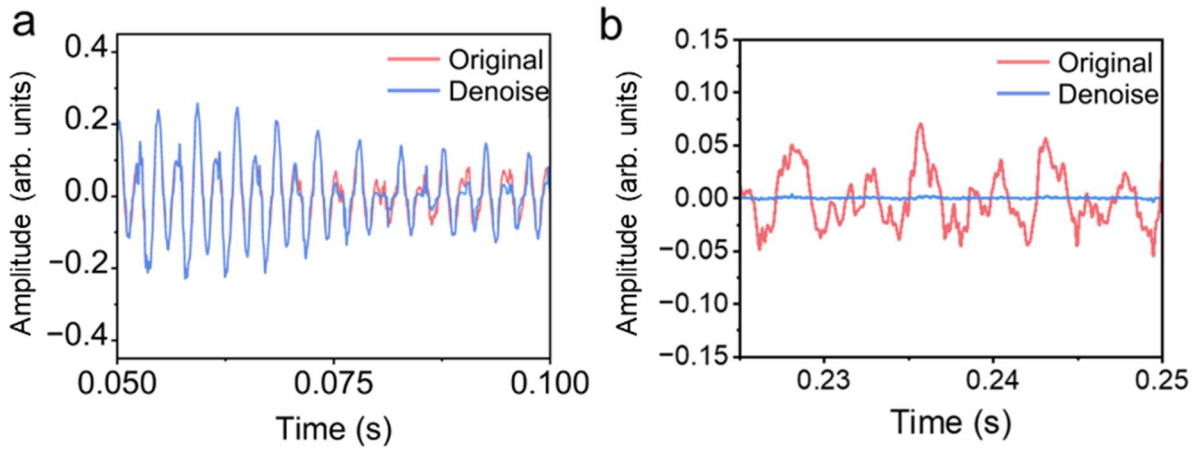
Definition and Processing Strategy for '0–30%' Strength Range: This range corresponds to normalized amplitude absolute values $|\text{amplitude}| \leq 0.06$, primarily encompassing background noise and random interference components. These signals typically exhibit irregular random fluctuations, lack evident spectral structural features, and demonstrate no significant statistical association with any specific audio categories. From a time-series perspective, these signals present typical white noise or colored noise characteristics, with rapidly decaying autocorrelation functions and relatively flat power spectral densities. For signals within this range, we employ a complete filtering strategy: since these weak signals have amplitudes far below the cumulative excitation threshold of memristive devices, they cannot trigger device state switching even after prolonged integration, thus being naturally filtered at the hardware level. This physics-based automatic filtering mechanism achieves zero additional power consumption noise suppression, avoiding the computational overhead of traditional digital filters.

Definition and Processing Strategy for '30–50%' Strength Range: This range corresponds to $0.06 < |\text{amplitude}| \leq 0.10$, containing the most critical weak characteristic signal segments. Although these signals have relatively low amplitudes, they typically carry important audio recognition information and category discrimination features. For instance, in digit speech recognition tasks, certain transitional phonemes, secondary formants, or pronunciation variation details often fall within this amplitude range; while easily confused with background noise when analyzed individually, their presence is decisive for distinguishing similar digits (such as '5' and '7', '6' and '8'). For this range, we implement a selective retention strategy based on cumulative excitation: while single low-amplitude pulses are insufficient to directly trigger memristor state switching, when identical features repeatedly appear in the temporal dimension, they gradually drive the device toward the activation threshold through cumulative effects within the device. When accumulated excitation energy exceeds the critical value, irreversible resistance state transition is triggered, thereby achieving intelligent capture of weak but relevant features. The physical foundation of this mechanism lies in the progressive formation process of conductive filaments in memristors, simulating the 'integrate-and-fire' behavior of biological neurons.

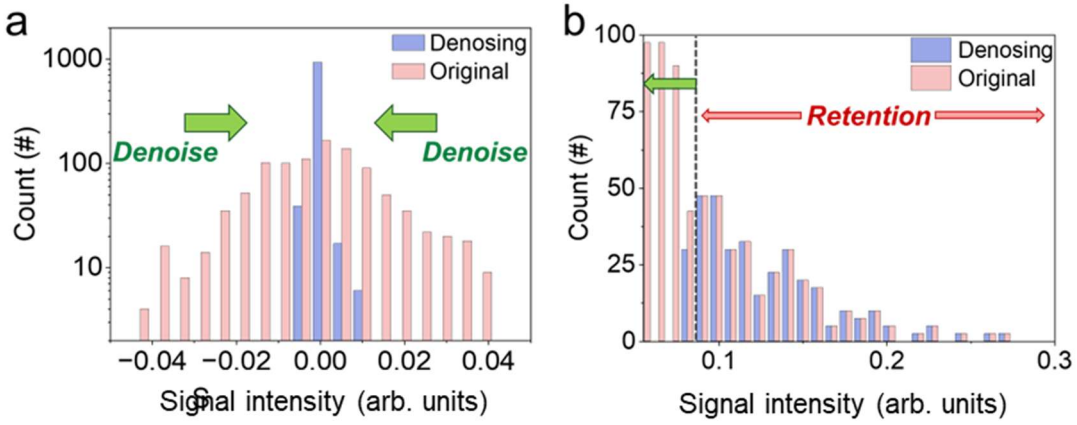
Definition and Processing Strategy for '>50%' Strength Range: This range corresponds to $|\text{amplitude}| > 0.10$, primarily comprising significant characteristic signals. These high-amplitude components typically correspond to core spectral features of audio signals, such as fundamental frequencies and primary formants in speech signals, or melodic components in musical signals, possessing distinct spectral structures and strong categorical associations. From an information-theoretic perspective, these strong signals carry the primary identification information of target audio categories, exhibiting high mutual information content and low uncertainty. For signals in this range, we adopt a direct retention and enhancement processing strategy: since these strong signals can directly activate memristors and switch them to stable low-resistance states without requiring accumulation processes, rapid response and high-fidelity transmission can be achieved. Simultaneously, appropriate voltage amplification further enhances their decision weight in final recognition results, ensuring full utilization of critical feature information.



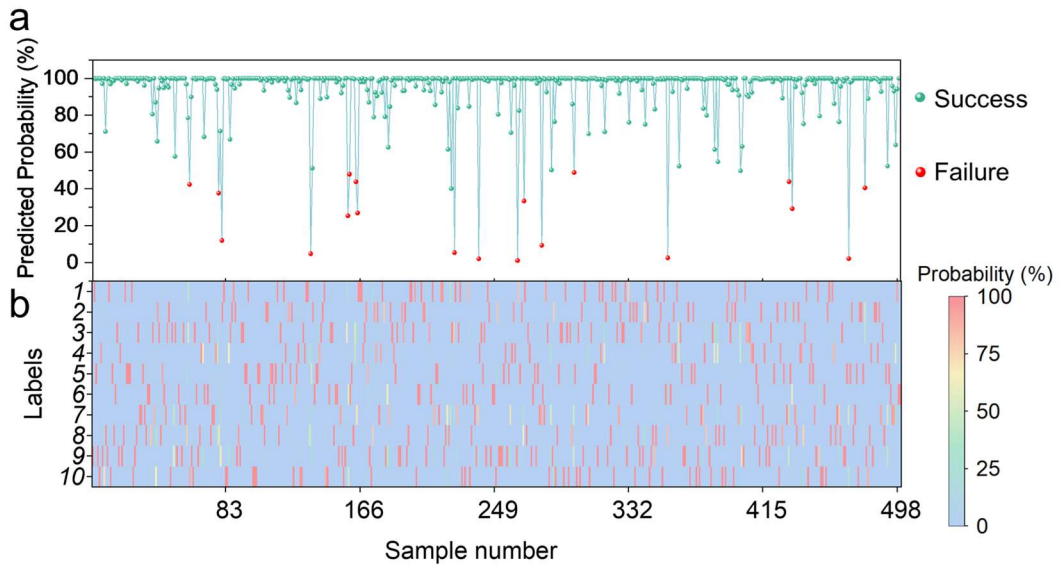
Supplementary Figure 32 Original and Denoised Audio Signals of Digits '0' to '9'. Strong signal data is effectively retained, while noisy and chaotic interference signals are effectively suppressed.



Supplementary Figure 33 Time-frequency analysis before and after denoising of signals with different intensities. a Retention of high-intensity feature signals; **b** Low-intensity noise suppression effect;

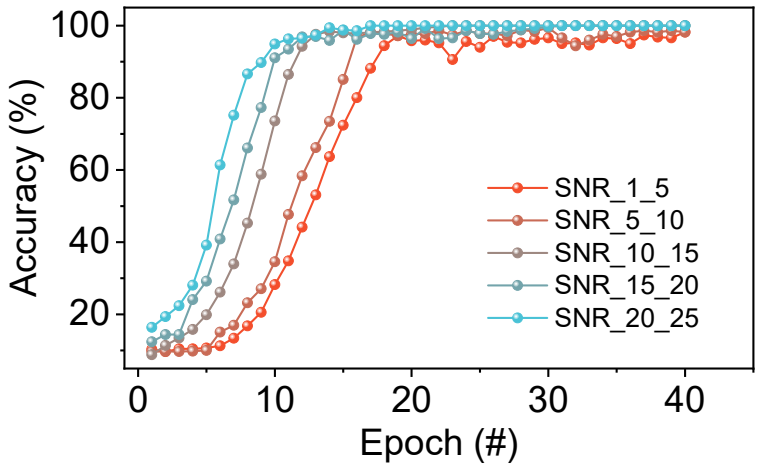


Supplementary Figure 34 Intensity distribution of signals before and after denoising. a Weak signals (<0.07), typically noise, are largely suppressed to 0–0.01, **b** while strong feature signals (>0.1) are fully preserved.



Supplementary Figure 35 Distribution of prediction results for 500 sample points after 10 epochs. a The predicted probabilities of the true labels by the system, where green points represent successful predictions, the majority of which exceed 99%, while red points represent failed predictions, accounting

for 3.8% of the total sample size. **b** Distribution map of recognition probabilities for ten labels across the 500 samples.



Supplementary Figure 36 Post-training correlation between recognition accuracy and epochs for varying SNR audio signals. Lower SNR (1-5 dB) signals show greater training difficulty, yet maintain robust recognition accuracy after SR-BSID processing.

Supplementary Table 1 Comparison of key metrics among state-of-the-art devices[6–22]

Structure	Set voltage (V)	Voltage slope (mV/dec)	On/off ratio	Retention (s)	Power consumption	Speed	Ref
Pt/Cu_2S/Pt	0.9	2	> 100	> 10^3	90 nJ*	0.1 ms	[6]
Sn/HfO_2/Cr/Pt	4	5.75	> 10^3	Volatile	20 fJ*#	17 ns^	[7]
Pt/TiO_2/Pt	1.4	< 0.5	> 10^2	Volatile	1 μJ*	30 ns	[8]
Al/PVP/Al	2.01	23.7	10^5	Volatile	20 nJ*	2.1 μs	[9]
Pt/V_2O_5/Pt	0.4	< 0.5	> 10	Volatile	8 pJ	17 ns	[10]
PMMA/$CSCu_2I_3$/ITO	0.23	6.2	10^4	Volatile	10 nJ*	Not report	[11]
W/Cu_2S/W	0.1^	< 5	10^5	Volatile	100 pJ*	100 ns	[12]
Cr/Au/h-Bn/Ag	0.026	Not report	10^{8}#	10^4	72 pJ	200 ns	[13]
Ag/SnO_x/SnSe	0.4	Not report	> 10^3	> 10^5	2 nJ*	1 ms	[14]
Al/Cu/GaO_x/Au	0.48	Not report	4.9×10^3	> 10^4	6 mJ*	Not report	[15]
ITO/Bi:SnO_2/TiN	0.25	Not report	> 10	10^4	16 μW	Not report	[16]
Ag/SiO_2/Ta_2O_5/Pt	0.19	Not report	> 10^3	10^5	Not report	Not report	[17]
Ti/h-BN/Au	1.73	Not report	> 50	Not report	< 2 pJ	120 ps#	[18]
Ag/$BA_2MA_5Pb_6I_{19}$/Pt	0.15	Not report	10^7	10^5	150 μW	< 35 ns	[19]
Au/$CuInS_2$/Cu	0.6	Not report	> 100	10^4	10 nW	Not report	[20]
Ag/PFC-73/ITO	0.86	Not report	> 10^3	> 10^3	Not report	Not report	[21]
Ag/TaO_x/ZnO/Pt	0.177	0.221^	10^5	Volatile	Not report	100 ns	[22]
Ta/SiNx/IGZO/Pt	0.019#	0.032#	10^{6}^	10^{6}#	248.2 fJ^	72 ns	This work

***Calculated by parameter in paper; # Best performance; ^Second best performance**

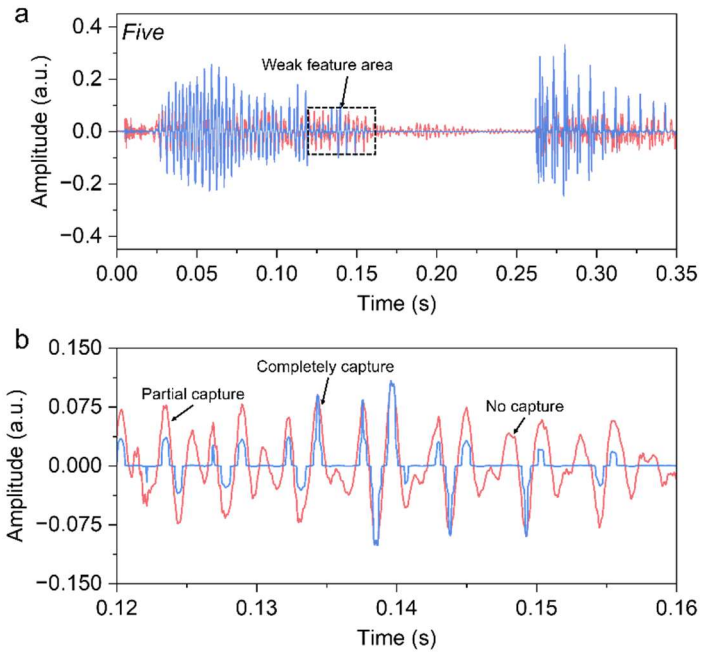
Cumulative capture of weak signals

Our defined weak but relevant features specifically refer to frequency domain components that belong to the target audio label's feature bands but have relatively low amplitudes. These signals are often easily mixed with background noise and mistakenly judged or discarded by traditional signal processing methods. Taking the audio for the number '5' as an example, shown in **Supplementary Figure 36**, such features typically manifest as:

(1) Signal amplitudes slightly higher than the background noise floor, but with distinct frequency domain structural characteristics (weak feature area in **Supplementary Figure 36a**);

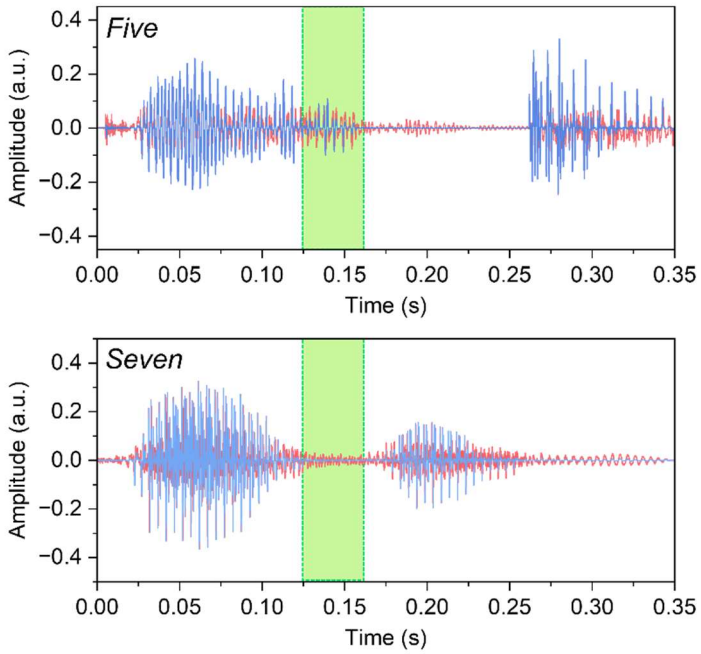
(2) The signal exhibits a continuous sequence characteristic with relatively stable amplitude fluctuations. This continuity allows it to progressively activate memristive devices through an accumulated excitation mechanism, enabling the array system to implement a hierarchical feature-capture strategy, shown in **Supplementary Figure 36b**: strong continuous signals are fully retained (Complete capture), medium-strength signals—after device activation—retain only their key spectral components (Partial capture),

and overly weak signals are effectively filtered out (No capture), thereby achieving adaptive adjustment of the recognition weighting for weak feature signals.



Supplementary Figure 36 Selective denoising of audio spectrograms via intensity-based signal processing. a Audio spectrogram of the digit '5', where the red line represents the original signal and the blue line represents the denoised signal. Signals in the weak-feature band have lower intensity. **b** During denoising, signals in the weak-feature band that have continuous high intensity are fully preserved (fully captured); signals of moderate intensity—after device activation—retain only their key spectral components (partially captured); while signals that are too weak are effectively filtered out (not captured).

(3) intermittent appearance in the time domain, but statistically highly correlated with specific audio categories; For example, in recognizing the digit '5', besides the main audio peaks, there exist some transitional frequency bands with smaller amplitudes, shown in **Supplementary Figure 37**. Although these weak features of green region are insufficient to independently trigger recognition, their presence or absence directly affects the discrimination accuracy between '5' and similar digits (such as '7').



Supplementary Figure 37 Weak-feature-enhanced classification discriminability. When recognizing the digit '5', besides the main audio peaks there are also some lower-amplitude transitional frequency bands (green areas); although these weaker features are not sufficient to trigger recognition on their own, their presence or absence affects the accuracy of distinguishing '5' from similar digits (such as '7').

Although difficult to quantify precisely, the denoising-zone array's capture of weak, continuous signals depends fundamentally on two key parameters—signal strength (amplitude) and continuity—ultimately enabling on-chip discrimination and capture of weak characteristic signals. The process can be viewed as a temporal integration of the duration of each regional sequence, weighted by the signal strength of the weak-feature region; when the integral reaches a preset threshold, the weak feature is deemed 'relevant' and retained. This procedure is expressed by the following simplified formula:

$$\int_{t_0}^{t_0+T} A(t) \cdot W(t) dt > \theta \tag{1}$$

where *A(t)* denotes the signal amplitude at time *t*, *W(t)* is a weighting function (reflecting the signal's relevance to the target label), *T* is the duration of the weak-feature sequence, and θ is the device's cumulative activation threshold. Only when the integral exceeds θ will the corresponding memristor switch state, thereby achieving 'memory' and retention of that weak feature.

This device-physical-characteristic-based discrimination mechanism has an inherent adaptive capability: transient signals with high amplitude but short duration requires higher instantaneous amplitudes to reach the integration threshold, while relatively low-amplitude but long-duration continuous signals can reach the same activation condition through temporal accumulation. This mechanism enables our system to flexibly capture distinct weak-feature patterns across different types of audio signals without predefining a single uniform quantization parameter. It is precisely this hardware-level intelligent adaptability that allows our denoising area to effectively extract faint but classification-relevant audio features that are difficult for conventional digital signal-processing methods to capture, thereby significantly improving recognition performance in noisy environments.

Comparison Methodology

The establishment of energy consumption comparison methodology requires strict experimental data and standardized procedures. Our memristor array achieves a total energy consumption of 130 pJ for complete audio recognition tasks over 15 training epochs. To achieve fair energy efficiency comparison, we adopted standardized energy efficiency evaluation metrics:

$$\eta_{energy} = \frac{TOPS}{W} = \frac{N_{OPS} \times f_{op}}{P_{total}} \tag{2}$$

where N_{ops} represents the number of operations per unit time, f_{op} is the operation frequency, and P_{total} is the total power consumption. For the CPU benchmark, Intel Core i7 5820K exhibits measured power consumption of approximately 140 W during audio convolution processing[23,24]. Our comparative data shows CPU total energy consumption of 3 mJ corresponding to 20 training epochs. For the FPGA benchmark establishment, we referenced the power consumption data of Virtex 6 at 155 MHz operating frequency[25], as well as the detailed power analysis of Zybo Z7 provided in reference (~1.5W)[26], where static power consumption is approximately 120 mW and dynamic power consumption can reach 1.405 W during complex audio algorithm execution. Our FPGA comparative data shows total energy consumption of 0.5 mJ corresponding to 18 training epochs. For the total energy consumption of the platform based on our memristor array, the core energy consumption of the memristor array when executing a single recognition operation is $E_{\text{memristor array}}$ = 248.2 fJ. Considering that our 32×32 memristor array requires 10 μs for a single operation, the average power consumption of the memristor array is $P_{\text{memristor array}} = E_{\text{memristor array}} \times \text{frequency} = 248.2 \times 10^{-15}\ \text{J} \times 10^{5}\ \text{Hz} = 0.2482 \times 10^{-5}\ \text{W} = 2.482$ μW. We use sixteen AD7928BRUZ-REEL7 Analog-to-Digital Converters (ADCs). According to their datasheets, at 3V supply voltage and 1 MSPS sampling frequency, the typical power consumption of a single chip is 1.8 mW, so the total power consumption of the ADCs is $P_{ADC} = 16 \times 1.8\ \text{mW} = 28.8\ \text{mW}$. We use sixteen AD5328BRUZ-REEL7 Digital-to-Analog Converters (DACs). According to their datasheets, at 3 V supply voltage, their typical power consumption is 11.2 mW. Additionally, we use eight ADG804YRMZ-REEL7 multiplexers, with extremely low typical power consumption of only 0.8 μW (0.0001 mW). Considering the total power consumption of these main peripheral circuits, the total power consumption of the current test platform is approximately $P_{\text{peripherals}} = P_{ADC} + P_{DAC} + P_{MUX} = 28.8\ \text{mW} + 11.2\ \text{mW} + 0.0008\ \text{mW} \approx 40\ \text{mW}$. Therefore, the system-level total power consumption of the

current test platform when executing a single recognition operation is approximately the sum of the core power consumption of the memristor array and the peripheral circuit power consumption: $P_{system\ total} = P_{memristor\ array} + P_{peripherals} = 0.0002482$ mW + 40 mW ≈ 40 mW. Furthermore, the system-level energy consumption of the current test platform when executing a single recognition operation is $E_{system\ total} = P_{system\ total} \times$ single array scanning time = 40 mW × 32 μs = 4×10^{-2} W × 3.2×10^{-5} s = 128 nJ. After 15 epochs, it is 1.92 μJ. We estimated that traditional CPU performing equivalent recognition tasks have energy consumption of approximately $E_{CPU} \approx 3$ mJ, $E_{FPGA} \approx 0.5$ mJ. Based on the system-level energy consumption $E_{system\ total}$ of our current test platform, the energy consumption advantages are: $E_{CPU} / E_{system\ total} = 3$ mJ / 1.92 μJ ≈ 1562 times; $E_{FPGA} / E_{system\ total} = 0.5$ mJ / 6 μJ ≈ 260 times.

Process node standardization for area comparison is crucial to ensure fair evaluation. Area efficiency between different process nodes requires correction through standardized scaling factors:

$$\alpha_{scaling} = \left(\frac{\lambda_{ref}}{\lambda_{target}}\right)^2 \times \beta_{density} \tag{3}$$

where λ represents the process feature size and $\beta_{density}$ is the density correction factor considering actual layout efficiency differences. Our memristor array has an actual chip area of 3 mm², including complete peripheral circuits. For the CPU benchmark, based on our comparative data, the effective computational area is approximately 100 mm². For the FPGA benchmark, detailed resource utilization statistics is provided[25]: in audio processing applications, low actual utilization rates is reported[26], with LUT usage rate of only 3.11% and register usage rate of 2.25%. Our FPGA comparative data shows an effective area of 15 mm², which is consistent with the low utilization phenomenon reported in literature. Therefore, our memristor array demonstrates approximately 33.3× and 5× area advantages compared to CPU and FPGA, respectively.

Latency comparison requires consideration of complete data path characteristics across different computing architectures. Our memristor array achieves a latency of 0.86 ms, including complete end-to-end processing time. For CPU latency, Intel Core i7 processors demonstrate measured latency of approximately 2.3 ms in similar complexity audio processing tasks[23], while our comparative data shows 20 ms, which includes the complete training and inference process. For FPGA latency, measured performance shows 0.6 s latency primarily contributed by software algorithms[26] while our FPGA comparative data shows 2 ms, reflecting the effect of hardware acceleration. Therefore, our memristor solution achieves approximately 23.3× and 2.3× latency advantages compared to CPU and FPGA, respectively.

Theoretical analysis of throughput performance requires consideration of the combined effects of parallelism and pipeline efficiency. We calculate the throughput based on the following model:

$$\Phi_{throughout} = \frac{N_{parallel} \times f_{clock}}{CPI_{effective}} \times \eta_{pipelline} \tag{4}$$

where $N_{parallel}$ is the number of parallel processing units, f_{clock} is the clock frequency, $CPI_{effective}$ is the effective cycles per instruction, and $\eta_{pipeline}$ is the pipeline efficiency. Our memristor array achieves a throughput of 5000 fps, while CPU and FPGA achieve 1000 fps and 3000 fps, respectively. FPGA implementation at 155 MHz operating frequency achieves actual throughput of approximately 620 MMAC/s, primarily limited by memory bandwidth and algorithm complexity[25]. The ARM+FPGA hybrid architecture achieves effective throughput of approximately 167 MMAC/s when processing complex audio algorithms, confirming performance bottlenecks in hybrid architectures for complex computational tasks[26].

Power density analysis provides another important dimension for performance evaluation. We adopted standardized power density metrics to quantify energy efficiency across different platforms:

$$\rho_{power} = \frac{P_{dynamic}}{A_{effective}} \times \frac{1}{\eta_{\text{utilization}}} \tag{5}$$

where $P_{dynamic}$ is the dynamic power consumption, $A_{effective}$ is the effective computational area, and $\eta_{utilization}$ is the resource utilization efficiency. FPGA dynamic power consumption reaches approximately 1.28W with effective area utilization of about 5-10%, yielding power density significantly higher than our memristor implementation[25]. Detailed power breakdown analysis shows

ARM processor power density of approximately 1.32 W/mm^2 and FPGA logic power density of approximately 0.086 W/mm^2, providing important references for our comparative analysis[26].
Benchmark consistency verification ensures scientific rigor and reproducibility of the comparison. All performance tests were conducted on identical audio datasets and with matched classification accuracy targets. Measurement conditions were standardized following established methods: all power consumption tests were performed at room temperature with the power supply voltage maintained within nominal ranges[25]. We adopted a sampling frequency of 8 kHz to eliminate sampling-rate effects on performance comparisons[26]. We adopted similar precision validation methodologies, referencing the high-accuracy benchmark where FPGA achieved correlation exceeding 99%[25]. By establishing a complete performance evaluation framework and standardized computational methodologies, we ensure fairness and scientific rigor across computing platforms. This rigorous, literature-benchmarked comparison, combined with theoretical computational model verification, supports the scientific validity and practical significance of our performance advantage conclusions.

Displacement Current from Capacitive Coupling

For our devices (area 2.5 μm²) with a multi-layer stack comprising 8 nm SiN_x and 20 nm IGZO, the total capacitance can be calculated as a series combination of the two dielectric layers:

$$\frac{1}{C_{total}} = \frac{d_{SiN}}{\varepsilon_0 \varepsilon_r^{SiN} A} + \frac{d_{IGZO}}{\varepsilon_0 \varepsilon_r^{IGZO} A} \tag{6}$$

Substituting literature values for the relative permittivity of SiN (7.5) and IGZO (8), we obtain:

$$C_{SiN} = \frac{8.85 \times 10^{-12} \times 7.5 \times 2.5 \times 10^{-12}}{8 \times 10^{-9}} \approx 20.78 \text{ fF} \tag{7}$$

$$C_{IGZO} = \frac{8.85 \times 10^{-1} \times 8 \times 2.5 \times 10^{-12}}{20 \times 10^{-9}} \approx 8.85 \text{ fF} \tag{8}$$

$$C_{total} \approx 6.2 \text{ fF} \tag{9}$$

Under our triangular pulse conditions (1 μs pulse width, 50 mV amplitude), the voltage change rate is:

$$\frac{dV}{dt} = \frac{50 \text{ mV}}{0.5 \text{ μs}} = 10^5 \text{ V/s} \tag{10}$$

The resulting steady-state displacement current is:

$$I_d = C \cdot \frac{dV}{dt} = 6.2 \text{ fF} \times 10^5 \text{ V/s} \approx 0.62 \text{ nA} \tag{11}$$

Since the device switching occurs at the nanosecond level and this value is far below the set compliance currents, this displacement current maintains a strictly constant value throughout the pulse duration and has a negligible effect on the dynamic switching behavior of devices in the array. In sharp contrast, the transient current spikes generated by square pulses introduce substantial high-frequency interference and significant transient deviations, severely compromising the reliability of low-current measurements. Therefore, the fundamental advantage of triangular pulses is that they smooth the voltage ramp, eliminating the transient current spikes arising from voltage discontinuities, thereby enabling us to accurately capture the true, gradually-evolving resistive switching process of the device. The steady-state displacement current, being a small and predictable constant background, does not obscure or distort the device response.

Supplementary References

1. Zhu, Y. *et al.* The filaments control for tunning digital resistive switching in data storage application and analog behavior as an artificial synapse with CsPbBr3-based memristor. *Applied Physics Letters* **124**, 063504 (2024).
2. Dai, X. *et al.* Artificial synapse based on a tri-layer AlN/AlScN/AlN stacked memristor for neuromorphic computing. *Nano Energy* **124**, 109473 (2024).
3. Li, L.-H. *et al.* Multilevel switching in Mg-doped HfOx memristor through the mutual-ion effect. *Applied Physics Letters* **119**, 153505 (2021).
4. Mahata, C. *et al.* Forming-Free Tunable Analog Switching in WOx/TaOx Heterojunction for Emulating Electronic Synapses. *Materials* **15**, 8858 (2022).
5. Xing, S. *et al.* Reliable Multistate RRAM Devices for Reconfigurable CAM and IMC Applications. *IEEE Journal of the Electron Devices Society* **13**, 383–389 (2025).
6. Mou, N. I., Zhang, Y., Pai, P. & Tabib-Azar, M. [1]Steep sub-threshold current slope (∼2 mV/dec) Pt/Cu2S/Pt gated memristor with lon/Ioff > 100. *Solid-State Electronics* **127**, 20–25 (2017).
7. McGuire, F. A. *et al.* [5]Sustained Sub-60 mV/decade Switching via the Negative Capacitance Effect in MoS2 Transistors. *ACS Publications* https://pubs.acs.org/doi/full/10.1021/acs.nanolett.7b01584 (2017) doi:10.1021/acs.nanolett.7b01584.
8. Hua, Q. *et al.* [6]A Threshold Switching Selector Based on Highly Ordered Ag Nanodots for X-Point Memory Applications. https://doi.org/10.1002/advs.201900024 doi:10.1002/advs.201900024.
9. Sonde, S. *et al.* [7]Silicon compatible Sn-based resistive switching memory. https://doi.org/10.1039/C8NR01540F doi:10.1039/C8NR01540F.
10. Zhou, X. *et al.* [8]A high performance electroformed single-crystallite VO2 threshold switch. https://doi.org/10.1039/C9NR08364B doi:10.1039/C9NR08364B.
11. Huang, X. *et al.* [9]Steep-slope field-effect transistors with AlGaN/GaN HEMT and oxide-based threshold switching device. *Nanotechnology* **30**, 215201 (2019).
12. Wang, Y., Yuan, Q., He, N., Sun, Y. & Wen, D. [10]Ovonic threshold switching device and its application to logic gate function and steep slope in field effect transistors. *Surfaces and Interfaces* **30**, 101895 (2022).
13. Song, B., Xu, H., Liu, S., Liu, H. & Li, Q. [15]Threshold Switching Behavior of Ag-SiTe-Based Selector Device and Annealing Effect on its Characteristics. *IEEE J. Electron Devices Soc.* **6**, 674–679 (2018).
14. Song, B. *et al.* [16]An ovonic threshold switching selector based on Se-rich GeSe chalcogenide. *Applied Physics A* **125**, 1–6 (2019).
15. Speckbacher, M. *et al.* [17]Directed Assembly of Nanoparticle Threshold-Selector Arrays. *Adv Elect Materials* **5**, 1900098 (2019).
16. Liu, Y. *et al.* A Robust and Low-Power Bismuth Doped Tin Oxide Memristor Derived from Coaxial Conductive Filaments. *Small* **16**, (2020).
17. Guo, X. *et al.* SiO2/Ta2O5 heterojunction ECM memristors: physical nature of their low voltage operation with high stability and uniformity. *Nanoscale* **12**, 4320–4327 (2020).
18. Teja Nibhanupudi, S. S. *et al.* Ultra-fast switching memristors based on two-dimensional materials. *Nat Commun* **15**, 2334 (2024).
19. Kim, S. J. *et al.* Reliable and Robust Two-Dimensional Perovskite Memristors for Flexible-Resistive Random-Access Memory Array. *ACS Nano* **18**, 28131–28141 (2024).
20. Chen, Z.-J. *et al.* A flexible artificial synapse based on the two-dimensional CuInS2 memristor for neural morphology calculation. *Materials Science in Semiconductor Processing* **188**, 109203 (2025).
21. Yang, X. *et al.* Solution-Processed Hydrogen-Bonded Organic Framework Nanofilms for High-Performance Resistive Memory Devices. *Advanced Materials* **35**, 2305344 (2023).
22. Li, P. *et al.* Dual-layer volatile memristor with ultralow voltage slope. *Applied Physics Letters* **124**, (2024).

23. Misic, M., Nikolov, D. & Tomasevic, M. Analysis of CPU and GPU implementations of convolution reverb effect. *Telfor* **8**, 121–126 (2016).
24. Intel® Core™ i7-5820K Processor (15M Cache, up to 3.60 GHz) - Product Specifications. *Intel* https://www.intel.com/content/www/us/en/products/sku/82932/intel-core-i75820k-processor-15m-cache-up-to-3-60-ghz/specifications.html.
25. Wassi, G., Iloga, S., Romain, O. & Granado, B. FPGA-based real-time MFCC extraction for automatic audio indexing on FM broadcast data. in *2015 Conference on Design and Architectures for Signal and Image Processing (DASIP)* 1–6 (2015). doi:10.1109/DASIP.2015.7367253.
26. Vaca, K., Jefferies, M. M. & Yang, X. An Open Audio Processing Platform with Zync FPGA. in *2019 IEEE International Symposium on Measurement and Control in Robotics (ISMCR)* D1-2-1-D1-2–6 (2019). doi:10.1109/ISMCR47492.2019.8955662.