

A generic compact model for resistive memories and switches

Received: 2 November 2025

Accepted: 24 February 2026

Published online: 04 April 2026

Cite this article as: Jung S., Kim H., Park J. *et al.* A generic compact model for resistive memories and switches. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-42077-y>

Sangwoo Jung, Hyoseob Kim, Jiseon Park, Sungjun Kim & Min-Hwi Kim

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

A Generic Compact Model for Resistive Memories and Switches

Sangwoo Jung¹, Hyoseob Kim¹, Jiseon Park¹, Sungjun Kim^{2*}, and Min-Hwi Kim^{1, 3*}

¹Department of Intelligent Semiconductor Engineering, Chung-Ang University, Seoul, 06974, Republic of Korea

²Division of Electronics and Electrical Engineering, Dongguk University, Seoul, 04620, Republic of Korea

³School of Electrical and Electronics Engineering, Chung-Ang University, Seoul, 06974, Republic of Korea

*Corresponding author: sungjun@dongguk.edu, minhwi@cau.ac.kr

ABSTRACT

Most previous studies have modeled resistive switching based on physical principles. While physics-driven models provide valuable physical insight, their complexity and device specificity limit scalability and often result in high computational costs in large-scale circuit simulations. To address these limitations, we propose a generic compact modeling methodology in LTspice. This generalized approach constructs compact models for diverse resistive memory technologies using a minimal set of common circuit elements. It integrates nonlinear resistors, capacitors, and switches to reproduce key switching phenomena—including filament formation, polarization-modulated tunneling, and spin-dependent resistance—thereby capturing the essential behaviors of RRAM, TS, FTJ, and MRAM. This generalized framework enables flexible parameter tuning to reflect each device's distinct switching dynamics. Validation using experimental and reference data confirms that the resulting models accurately reproduce the key switching characteristics across all four device classes. Monte Carlo simulations further demonstrate the framework's ability to capture variability arising from parameter fluctuations and cycling effects. The proposed framework scales seamlessly from single-device operation to array-level simulation, enabling performance evaluation of SNN- and BNN-based inference systems and supporting accurate estimation of power consumption. Collectively, these results demonstrate the framework's broad applicability and scalability for large-scale neuromorphic and in-memory computing architectures.

Introduction

The rapid advancement of artificial intelligence and data-driven applications has intensified the demand for scalable and energy-efficient computing architectures. Current transistor-based technologies are facing severe efficiency limits, with the global electricity consumption of information technology projected to reach up to 21% by 2030.¹ To address this impending energy crisis and the scaling limitations, architectures that directly process data within memory—such as in-memory computing (IMC) and neuromorphic systems—are emerging as a promising path to overcome the von Neumann bottleneck.^{2,3}

Emerging memory technologies based on memristive devices are considered key enablers for these paradigms due to their non-volatile storage capability and resistive switching behavior. These properties offer a practical route toward realizing energy-efficient, large-scale computing systems, a potential that is further validated by their growing adoption in industrial embedded applications and cost-competitiveness compared to traditional Flash memories.

Among these technologies, devices such as resistive random-access memory (RRAM), threshold switching (TS) selectors, ferroelectric tunnel junctions (FTJ), and magnetic random-access memory (MRAM) exhibit distinct switching mechanisms suited for non-volatile storage, dynamic selection, and in-memory processing. The unique electrical characteristics and underlying physical mechanisms shown in Fig. 1 have established these devices as promising candidates for diverse next-generation computing platforms. Moreover, their technological readiness is notably advancing, with successful integration into commercial process nodes signaling that they are moving beyond theoretical research into practical implementation.⁴⁻⁷

However, these fundamental operational differences complicate the development of a generalized circuit-level model. Most existing compact models are device-specific or rely heavily on complex physical formulations. This fragmentation increases structural complexity and significantly limits their use in large-scale circuit simulations, thereby hindering reliable exploration of system-level performance and power efficiency.⁸⁻¹⁰

Recent studies have highlighted these challenges, emphasizing the need for a simplified yet broadly applicable modeling framework. To address these challenges, we propose a generic compact modeling methodology in LTspice. This generalized approach uses a minimal set of circuit components to construct distinct compact models for various resistive memory devices. These models (comprising nonlinear resistors, capacitors, and switches) emulate switching dynamics without detailed physical modeling. This methodology achieves both computational efficiency and broad applicability, accurately reproducing key behaviors such as resistance transitions, nonlinear current-voltage characteristics, and threshold-driven switching dynamics.¹¹

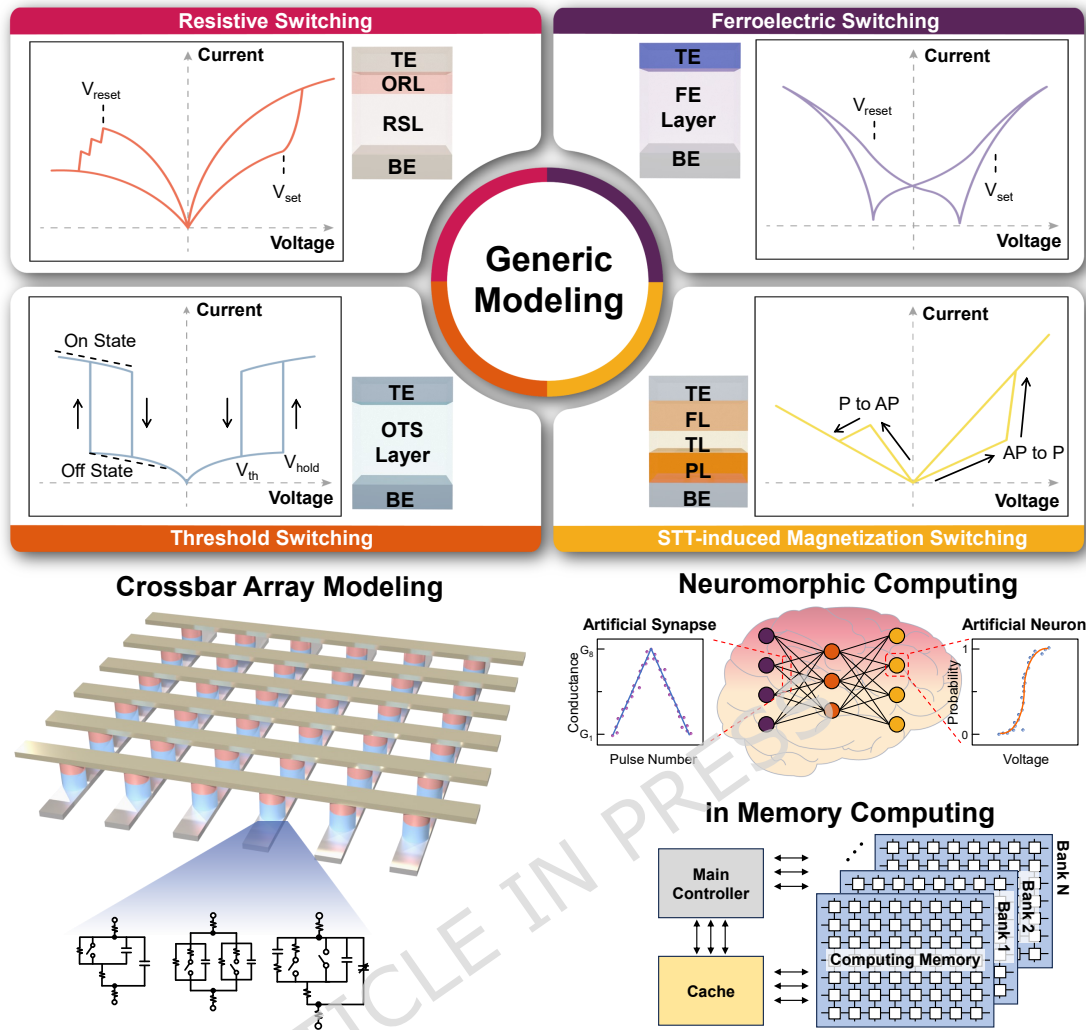


Figure 1. Conceptual overview of the generic modeling framework. It connects the distinct switching mechanisms of RRAM, TS, FTJ, and MRAM and demonstrates their scalability to crossbar array modeling, neuromorphic computing, and in-memory computing architectures.

This common framework scales seamlessly from individual device modeling to array-level simulations and supports practical design and analysis at the system level. The methodology was validated comprehensively at both the device and array levels. The resulting models accurately reproduced the key switching characteristics of all four device types against experimental and reference data. Monte Carlo analysis further confirmed the framework's ability to capture variability arising from fabrication-induced fluctuations and operational noise. Building on this validation, its system-level capability was demonstrated through array-level simulations of SNN- and BNN-based inference systems, enabling accurate evaluation of operational performance and power consumption.¹²

Methods

A generic compact model for RRAM, TS, FTJ, and MRAM was developed in LTspice using manually constructed circuits composed of nonlinear resistors, capacitors, and switches. Device-level validation was performed by comparing simulation results against DC and transient switching characteristics. Detailed procedures for parameter extraction and model adaptation are provided in Supplementary Note 1. For fabricated RRAM (TiN/Hf/HfO_x/Pt/Ti stack) and TS (W/ZnTe/W) devices, these characteristics were experimentally measured at 25°C using a Keysight B1500A (HRSMU/WGFMU units). An overview of the manufacturing process and the stack structure of the fabricated memristive devices used for experimental validation is provided in Supplementary Fig. S1. The FTJ and MRAM models were validated against published reference data from Li et al. and Kim et al., respectively. Stochastic behavior was examined via Monte Carlo analysis, using variability parameters

extracted from experimental cycle-to-cycle measurements of the fabricated RRAM and TS devices. System-level applicability was verified using array-level LTspice implementations. A co-simulation methodology was employed for SNN inference ($784 \times 24 \times 10$), where the 784×24 first layer was simulated in software. Its output spikes were then fed into the 24×10 final layer, which was implemented in LTspice using our RRAM model. Additionally, a BNN convolution system was simulated using an MRAM-based XNOR array to evaluate computational accuracy and power efficiency.

Results and discussion

Device-level Simulation

Our device-level analysis covers the four representative memory technologies: RRAM, TS devices, FTJ, and spin-transfer torque MRAM (STT-MRAM). For each technology, a distinct compact model was constructed using the proposed methodology, capturing the core physical switching mechanism with a minimal set of circuit components. A detailed summary of the circuit components and parameters for each device is provided in the Supplementary Table S1.

Resistive Random Access Memory Modeling

RRAM exhibits binary switching behavior, alternating between High-Resistance State (HRS) and Low-Resistance State (LRS) under applied voltage bias. This transition originates from the formation and rupture of conductive filaments driven by oxygen ion migration within the HfO_x layer. Under positive bias, oxygen vacancies drift toward the bottom electrode, leading to filament formation and transition to the LRS. In contrast, a negative bias induces localized Joule heating and reoxidation, disrupting the filament and restoring the insulating gap, thereby returning the device to the HRS. Both thermal and field-driven processes govern this bipolar switching mechanism, as shown in Fig. 2a.^{13–15}

Our compact circuit, shown in Fig. 2b, models this filament-based switching mechanism. The circuit consists of a non-linear resistor combined with a series of voltage-controlled switches (S_1 to S_n). As the applied voltage exceeds predefined threshold values, these switches are sequentially activated, replicating the progressive growth of the conductive filament. The electrical characteristics of these switches are governed by a voltage-controlled threshold logic defined by two key parameters: the threshold voltage (V_t) and the holding voltage (V_h), as listed in Supplementary Table S1. Specifically, the switch transitions from the High Resistance State (HRS, R_{off}) to the Low Resistance State (LRS, R_{on}) when the applied voltage exceeds the upper threshold sum ($V > V_t + V_h$). Conversely, the device reverts to the HRS when the voltage drops below the lower threshold difference ($V < V_t - V_h$). In this model, R_c represents the contact resistance at the electrode interface, R_s accounts for the insulating gap resistance in the HRS, and R_f models the conductive filament in the LRS, while C_s and C_i describe the gap and cell capacitances, respectively.

While physical models often use internal state variables (e.g., x in Eq. 1) to describe filament evolution, this approach can increase complexity and computational cost in circuit simulators.

$$R(x) = R_{\text{on}} \cdot \exp(-\lambda x) + R_{\text{off}} \cdot (1 - \exp(-\lambda x)) \quad (1)$$

To maximize SPICE compatibility and efficiency, we therefore implemented a stateless, voltage-dependent exponential resistance model in LTspice, expressed as

$$R = R_0 \cdot \exp\left(-\frac{|V(n1) - V(n2)| - a}{b}\right) \quad (2)$$

where R_0 is a scaling constant, and a and b are fitting parameters calibrated against experimental I – V data. The voltage drop across the non-linear resistor is represented by $|V(n1) - V(n2)|$, where $n1$ and $n2$ denote the positive and negative terminals of the resistive element, respectively.^{16–18}

We applied this expression to model the voltage-dependent behavior of both the insulating gap resistance R_s and filament resistance R_f , reflecting their strong non-linearity under different bias conditions. The exponential form allows for a compact representation of threshold-triggered conduction changes without requiring complex internal state tracking. This strategy achieves both circuit simplicity and an accurate reflection of filamentary switching physics. As shown in Fig. 2c, the simulated I – V characteristics exhibit strong agreement with the measured data obtained from the RRAM device. A deterministic accuracy verification using the best-fitted DC I – V characteristics and a quantitative mean absolute percentage error (MAPE) analysis is provided in Supplementary Fig. S2. The measurement was performed using a double sweep method, where the voltage was swept from -2.5 V to $+2.5\text{ V}$ and then back to -2.5 V —a standard technique for evaluating hysteresis. A compliance current of $100\text{ }\mu\text{A}$ was applied, and a single switching cycle was simulated under the same constraint. The simulated curve closely matches the measured data, successfully reproducing the sharp set transition and the relatively fast reset behavior. These results confirm that the proposed circuit model captures the filamentary resistive switching characteristics of RRAM with strong physical consistency.¹⁹

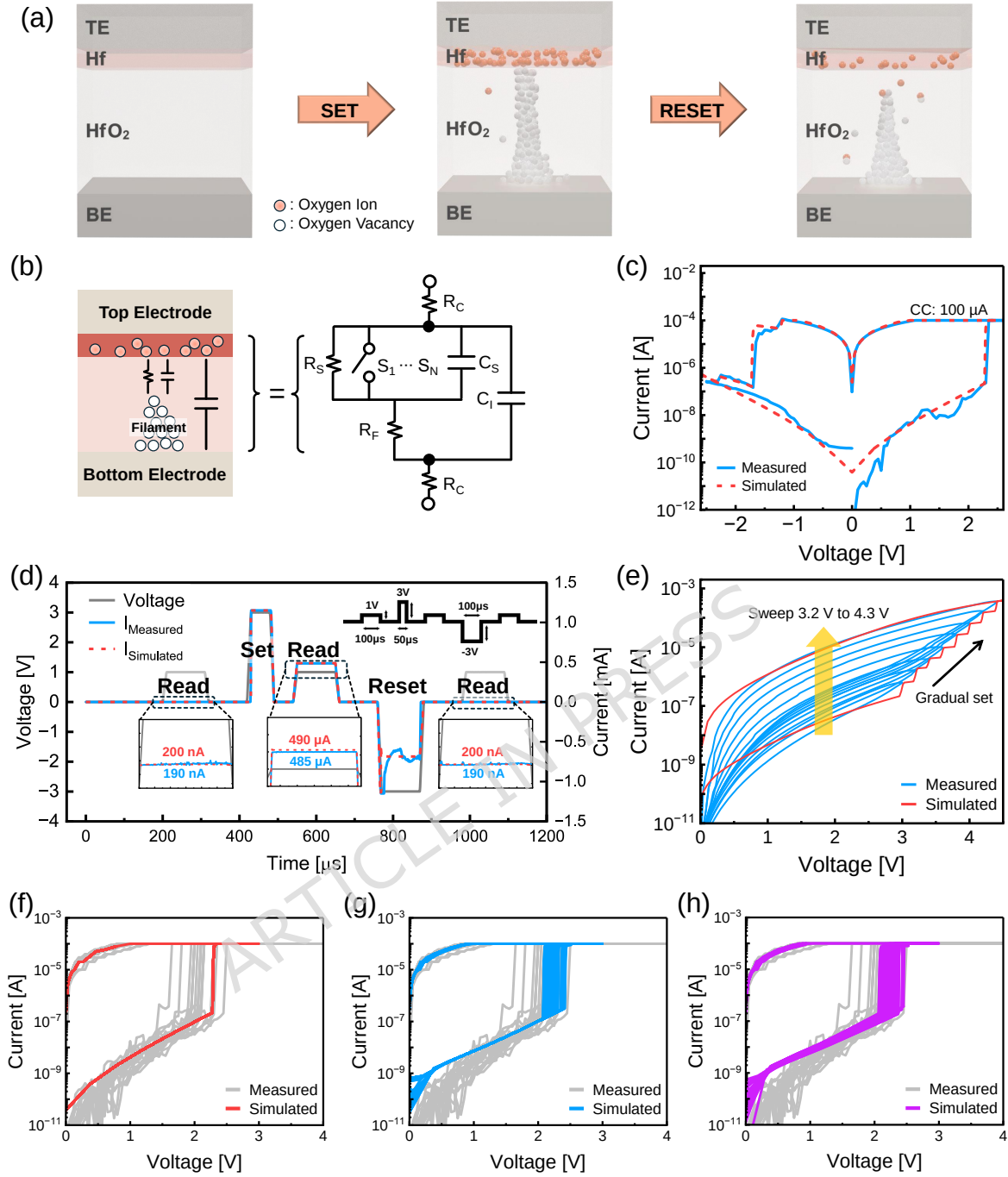


Figure 2. (a) Bipolar resistive switching mechanism in HfO_x-based RRAM, illustrating filament creation and breakdown. (b) Conceptual diagram and equivalent circuit model for the RRAM device. (c) Simulated and measured *I*-*V* characteristics validating consistent switching thresholds (*V*_{set}, *V*_{reset}). (d) Transient response validating dynamic switching behavior under positive and negative pulses. (e) Multi-level (3-bit) operation simulated by controlling the SET voltage. (f) Simulated reference *I*-*V* curve without parameter variation. (g) Monte Carlo simulation with variation in *V*_{set} and *V*_{reset}. (h) Monte Carlo simulation with combined variation in *V*_{set}, *V*_{reset}, *R*_{on}, and *R*_{off}.

To verify the model's dynamic switching behavior, we performed transient write-read pulse simulations. As shown in Fig. 2d, bipolar voltage pulses (Set and Reset) were applied, followed by small read pulses. The simulated transient current (dashed line) accurately reproduces the abrupt resistance changes and matches the measured read current levels (solid line) after both Set and Reset operations. This demonstrates that the model captures both static and dynamic behaviors of filamentary

RRAM under pulse operation. Additional pulse-timing dependent transient switching results across multiple time domains are provided in Supplementary Fig. S3.²⁰

We implemented multi-level switching behavior by controlling both the number and threshold voltages of the series-connected voltage-controlled switches (S_1 to S_8). As shown in Fig. 2e, each switch activates progressively as the applied voltage exceeds its individual threshold, producing discrete steps in conductance. This mechanism enables a smooth emulation of multiple intermediate resistance states, reflecting partial filament growth or dissolution within the device. During a gradual set voltage sweep from 3.2 V to 4.3 V, eight distinct switching events were sequentially triggered, effectively realizing 3-bit resolution in a single cell. The resulting I - V curve exhibits well-defined conductance plateaus, validating that the model captures voltage-dependent modulation and multi-level switching characteristics. The simulation results closely match experimental measurements and confirm the model's ability to reproduce both gradual and abrupt transitions observed in filamentary RRAM devices.²¹

We evaluated the model's ability to capture device variability through Monte Carlo simulations in LTspice. Figure 2f shows the reference I - V response without parameter variation, demonstrating stable switching behavior. In Fig. 2g, we introduced random variations to the switching threshold voltages, V_{set} and V_{reset} , resulting in observable cycle-to-cycle dispersion near the switching points. Further variation in the resistance parameters, R_{on} and R_{off} , produced I - V curves with a statistical spread and non-linearity consistent with experimentally observed distributions, as shown in Fig. 2h. These results confirm that the model supports variability analysis and replicates the stochastic switching behavior of filamentary RRAM. A quantitative summary of the coefficient-of-variation (CV) statistics for both LRS and HRS (mean, quartiles, and range) is provided in Supplementary Table S2. This statistical agreement was further verified using cumulative distribution functions (CDFs) of V_{set} and read currents at $V_{\text{read}} = 1$ V, as shown in Supplementary Fig. S4. A quantitative comparison of the coefficient of variation (CV) between measurement and simulation is provided in Supplementary Fig. S5.²²

Threshold Switching Device Modeling

The switching behavior illustrated in Fig. 3a originates from field-induced reconfiguration of localized defect states within amorphous chalcogenide materials. During electroforming, a strong electric field generates intrinsic defects such as vacancies and interstitials. These defects subsequently migrate and partially align, forming transient conduction pathways within the insulating matrix. When the applied voltage exceeds the threshold voltage V_t , injected carriers gain sufficient energy to trigger impact ionization and defect-assisted tunneling, resulting in an abrupt transition to a low-resistance state. Unlike resistive switching memories governed by stable filament formation, this conductive state is sustained by continuous carrier injection and field-aligned defect clusters. As the voltage decreases below the holding voltage V_h , the electric field can no longer maintain the defect configuration, the accumulated defects relax into disordered states, and the conduction path collapses. Consequently, the device reverts to a high-resistance state. This volatile switching mechanism confines conduction to voltages above the threshold, effectively suppressing leakage current under subthreshold bias.^{23,24}

The equivalent circuit in Fig. 3b models this volatile behavior using our generalized modeling approach. It consists of the same core elements as the RRAM model—nonlinear resistors, capacitors, and voltage-controlled switches. Although the physical mechanism (defect reconfiguration) differs from that of RRAM (filament formation), the resulting electrical response is phenomenologically analogous: the device abruptly switches ON when V_t is exceeded and switches OFF once the voltage falls below V_h . This voltage-triggered, nonlinear transition between HRS and LRS is functionally similar to the resistive switching in RRAM. Accordingly, we employ the same stateless, voltage-dependent exponential formulation (Eq. 2) to describe the nonlinear resistors (R_s and R_c) in the TS model. To capture the abrupt turn-on and turn-off behavior symmetrically in both polarities, the circuit adopts two asymmetric branches. This configuration successfully reproduces the volatile switching dynamics, and tuning of the parameters (V_t , V_h , and fitting coefficients a and b) enables close agreement with experimental observations.²⁵

As shown in Fig. 3c, the simulated I - V curve exhibits strong agreement with the experimental data of a fabricated TS device. The voltage sweep was performed bidirectionally from +1.5 V to -1.5 V with a compliance current of 10 mA to prevent permanent breakdown. When the voltage exceeds V_t , the simulated current rises sharply, and upon decreasing below V_h , the current drops abruptly, reproducing the rapid collapse of the conduction path. The model clearly reproduces the hysteresis window defined by the V_t - V_h separation, with both the steep turn-on slope and the sudden reset transition accurately matched to the experimental measurements. These results confirm the model's capability to replicate the volatile switching behavior of TS devices.²⁶

Model robustness was further verified by systematically perturbing key switching parameters. Threshold voltages V_t and V_h , as well as resistance values R_{on} and R_{off} , were independently varied to emulate fabrication- and operation-induced fluctuations. Figure 3d shows the reference I - V curve without variation. In Fig. 3e, changes in V_t and V_h shift the switching window horizontally, altering the turn-on and turn-off voltages. In Fig. 3f, variations in R_{on} and R_{off} modify the current level, affecting the ON/OFF conductance ratio. The model maintains consistent switching trends under all conditions, demonstrating its capability to capture the variability observed in TS devices.²⁷

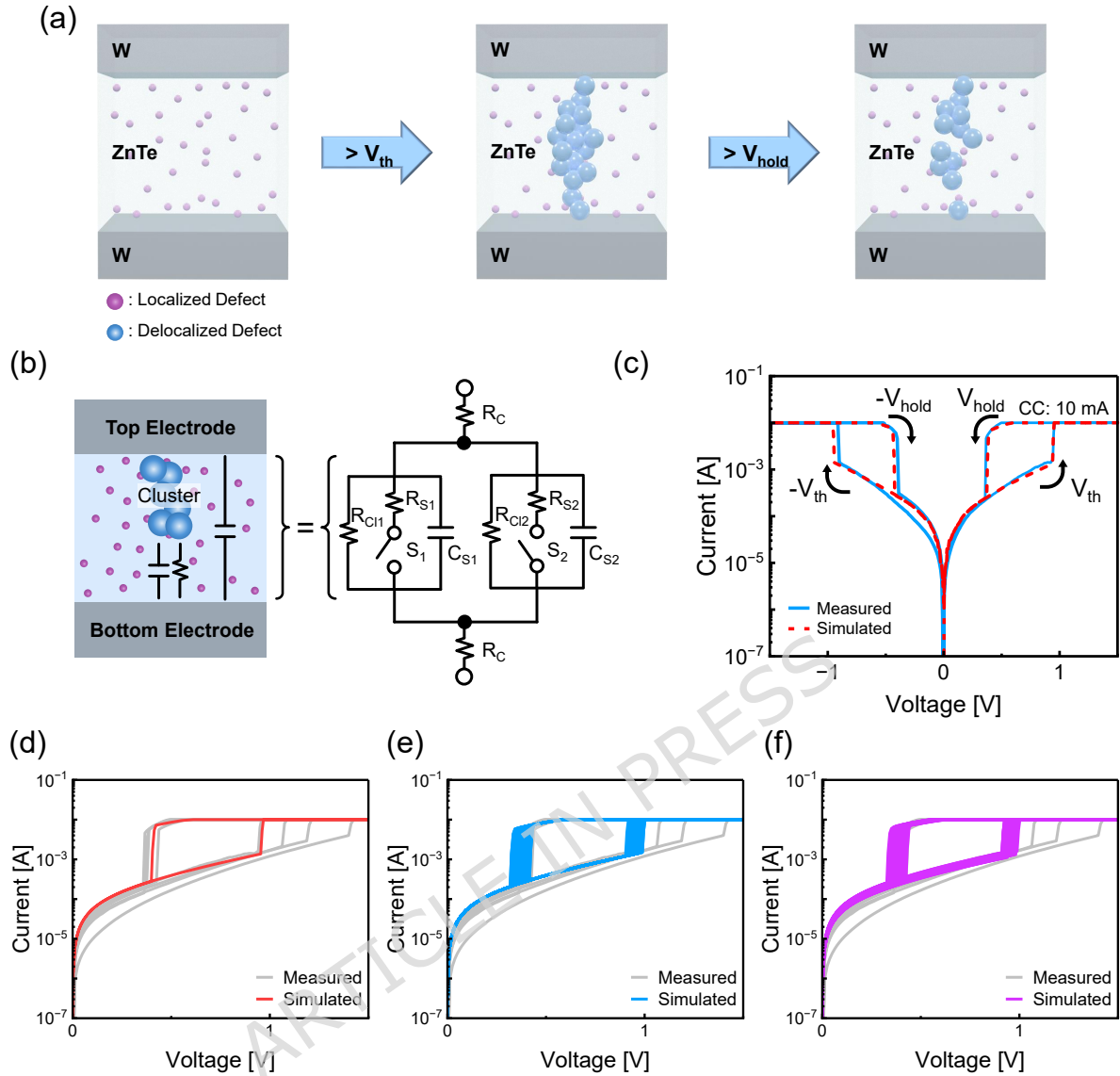


Figure 3. (a) Switching mechanism of a ZnTe-based OTS device, illustrating defect transitions and cluster formation. (b) Conceptual diagram and equivalent circuit model for the OTS device. (c) Simulated and measured $I-V$ characteristics validating the matched threshold (V_t) and holding (V_h) behavior. (d) Simulated reference $I-V$ curve without parameter variation. (e) Monte Carlo simulation with variation in V_t and V_h . (f) Monte Carlo simulation with combined variation in V_t , V_h , R_{on} , and R_{off} .

Ferroelectric Tunnel Junction Modeling

The operation of a Ferroelectric Tunnel Junction (FTJ) relies on ferroelectric polarization modifying the electrostatic potential across its tunneling barrier (Fig. 4a). An applied voltage shifts the conduction band profile, raising or lowering the barrier depending on the polarization direction. During the set operation, the polarization points from the top toward the bottom electrode. This internal electric field lowers the conduction band near the bottom interface and reduces the effective barrier for electrons tunneling from bottom to top. A read voltage of the same polarity therefore increases the tunneling probability and transitions the device into the LRS. Conversely, the reset operation reverses the polarization toward the top electrode, which raises the conduction band and increases both the barrier height and width. This results in a decreased tunneling probability and a transition to the HRS under the same read bias.^{28,29} The interplay between the applied voltage and the polarization-induced internal field governs the readout current by modulating the effective barrier. This dynamic modulation of tunneling probability constitutes the tunneling electroresistance (TER) effect. We validated this mechanism using TCAD simulations with the Centaurus tool, and the results are presented in Supplementary Fig. S6.^{30,31}

The compact model for FTJ behavior (Fig. 4b) is based on this principle. It consists of an exponential tunneling resis-

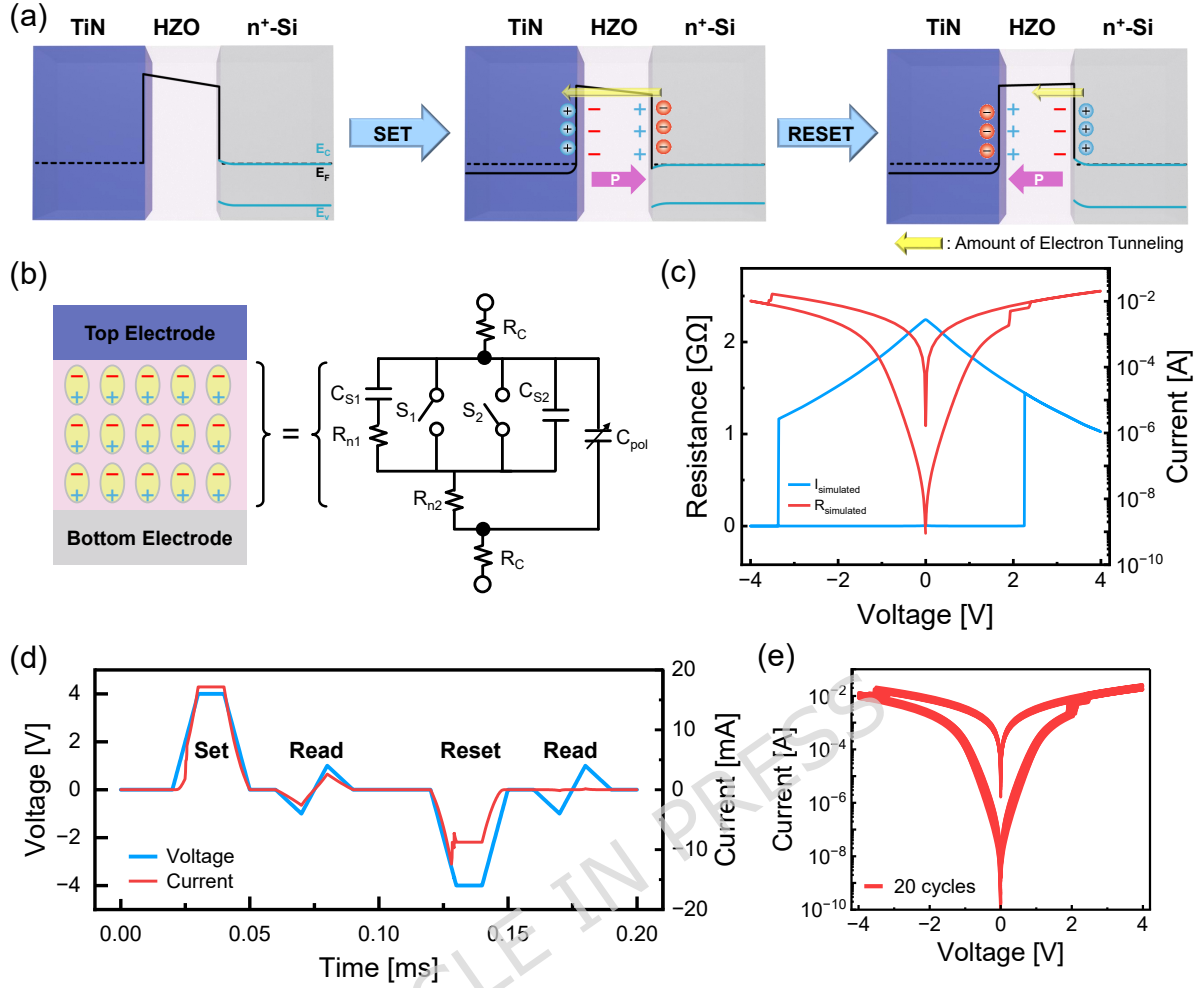


Figure 4. (a) Switching mechanism of the FTJ, illustrating polarization reversal and tunneling barrier modulation. (b) Conceptual diagram and equivalent circuit model for the FTJ device. (c) Simulated I - V and R - V characteristics validating the hysteretic switching behavior. (d) Transient pulse response validating the polarization-dependent TER effect. (e) Monte Carlo simulation with combined variation in switching thresholds and resistance states (R_{on} , R_{off}), confirming model robustness.

tor (R_{FTJ})—representing the combined nonlinear tunneling branches (R_{n1} and R_{n2})—and a voltage-dependent ferroelectric capacitance (C_{pol}) connected in parallel. The tunneling resistance (R_{FTJ}) originates from quantum tunneling, and the WKB approximation (Eq. 3) provides the theoretical foundation:

$$T \propto \exp\left(-\frac{2d}{\hbar}\sqrt{2m\phi_{\text{eff}}}\right) \quad (3)$$

Here, m is the electron effective mass and $\hbar$ is the reduced Planck constant. The tunneling current (I) is proportional to this probability and thus follows the same exponential form simplified to:

$$I \propto \exp\left(-C\sqrt{\phi_{\text{eff}}}\right) \quad (4)$$

where C is a constant incorporating the physical parameters from the previous equation. A crucial aspect of FTJs is that an applied voltage $|V|$ lowers the effective barrier (ϕ_{eff}), which causes the current I to increase exponentially with voltage. Consequently, the resistance $R = V/I$ decreases exponentially as a function of $|V|$. This phenomenological, voltage-dependent exponential behavior, though arising from quantum tunneling, is mathematically analogous to the nonlinear conduction modeled in RRAM and TS. We therefore apply the same stateless, voltage-dependent exponential formulation (Eq. 2) to describe this decay for the resistor R_{FTJ} . Concurrently, the ferroelectric polarization response is modeled by the voltage-dependent capacitance C_{pol} , which is also fitted using the same phenomenological exponential form. This component models the nonlinear

dielectric response (i.e., capacitance increase) as the voltage approaches the coercive region, thereby capturing the hysteretic polarization dynamics. The model's $\exp(-|V|)$ form, applied to both R and C components, directly reproduces the required behavior, where a and b are fitting parameters that scale and shift the exponential resistance curve to match experimental I - V data. The use of absolute voltage ensures symmetric modulation and reflects the bidirectional nature of polarization switching.^{32–35}

As shown in Fig. 4c, the simulated I - V characteristics exhibit strong agreement with reference data from Li *et al.*, obtained under voltage sweeps between -4 V and $+4$ V. The model clearly reproduces the hysteresis and nonlinear conduction. This result confirms that the model correctly represents polarization-modulated tunneling behavior.^{36,37} To further confirm the model's applicability across different FTJ configurations, we additionally validated the model using an extra TiN-based FTJ stack, as shown in Supplementary Fig. S7. Corresponding resistance-voltage (R - V) plots further validate that the model reproduces the abrupt resistance transitions associated with set and reset operations, consistent with the physical mechanisms underlying FTJ switching.

We further analyzed the transient switching behavior of the FTJ under dynamic biasing conditions to demonstrate the TER effect. The applied pulse sequence, shown in Fig. 4d, consisted of a positive 4 V set pulse followed by a -1 V to $+1$ V read sweep. A subsequent negative 4 V reset pulse was then applied, followed by a second identical read sweep. Short stabilization intervals between steps ensured complete polarization transition. The first read revealed a high current, which confirmed the device was in the LRS. In contrast, the read after the reset showed a significantly lower current, indicating a transition to the HRS. This distinct current contrast under identical read conditions is clear evidence of the TER behavior governed by ferroelectric polarization switching.^{38–40}

Finally, we evaluated the FTJ model's variability to assess its robustness against process variations. Twenty Monte Carlo simulations were performed, introducing random variations in switching voltage thresholds and resistance states (R_{on} , R_{off}). These variations emulate physical non-uniformities such as fabrication fluctuations and material imperfections. The resulting I - V curves (Fig. 4e) consistently maintain well-defined hysteresis loops and sharp switching transitions. This result confirms that the model is resilient to variability and accurately preserves the key switching characteristics of FTJs under diverse conditions.⁴¹

Magnetoresistive Random Access Memory Modeling

The magnetic tunnel junction (MTJ) consists of a free layer (FL), a thin insulating tunnel barrier (BL), and a pinned layer (PL). The PL maintains a fixed magnetic orientation, while the FL reverses its magnetization in response to an external magnetic field or a spin-polarized current. The tunneling resistance depends on the relative alignment of the magnetizations of these two layers. In the parallel (P) configuration, minimal electron spin scattering results in the LRS, whereas in the antiparallel (AP) configuration, spin misalignment suppresses tunneling and leads to the HRS. These two stable resistance states form the basis of non-volatile data storage in MRAM. Spin-transfer torque (STT) enables this switching, where a spin-polarized current exerts a torque on the FL to reverse its magnetization, as illustrated in Fig. 5a.^{42,43}

STT-MRAM typically utilizes a one-transistor one-MTJ (1T1M) architecture, where the MTJ is connected in series with an NMOS access transistor (Fig. 5b). Our compact model adheres to this structure and generalized methodology. It combines a current-controlled switch with a voltage-controlled nonlinear resistor. The current-controlled switch emulates the STT mechanism, a fundamentally current-driven process. The voltage-controlled resistor (R_n), which models the non-ohmic tunneling behavior, employs the same stateless, voltage-dependent exponential formulation (Eq. 2) used for the other device models, demonstrating the versatility of the approach.⁴⁴

The MRAM model was validated against the switching behavior of a reference STT device. As shown in Fig. 5c, the simulated I - V characteristics from a ± 0.8 V bidirectional voltage sweep accurately reproduce the reference device behavior. The model produces distinct LRS (P) and HRS (AP) states and correctly captures the abrupt resistance transitions for both Program (AP $\rightarrow$ P) and Erase (P $\rightarrow$ AP) operations, confirming its fundamental accuracy.^{45,46}

The model's dynamic switching behavior was evaluated using the five-step pulse sequence shown in Fig. 5d. This sequence applied a $+0.8$ V Program pulse and a -0.8 V Erase pulse, with intermediate 0.3 V read pulses verifying the resistance state. The simulated transient currents (red line) successfully replicated the behavior of the target device and clearly distinguished between the LRS (approximately $600 \mu\text{A}$) and the HRS (approximately $250 \mu\text{A}$). This confirms that the model accurately captures STT-driven switching at the circuit level without complex magnetic field modeling.

Finally, the variability of the MRAM model was assessed through twenty transient simulation cycles with random perturbations applied to key parameters. Specifically, variations were introduced to the LRS and HRS resistances (R_{on} , R_{off}) and the LTspice model parameters for the current-controlled switch (I_t , I_h), emulating process-induced non-uniformities. As illustrated in Fig. 5e, the resulting current traces remain clearly separated. This demonstrates the model's robustness and predictive stability under parameter variations.⁴⁷

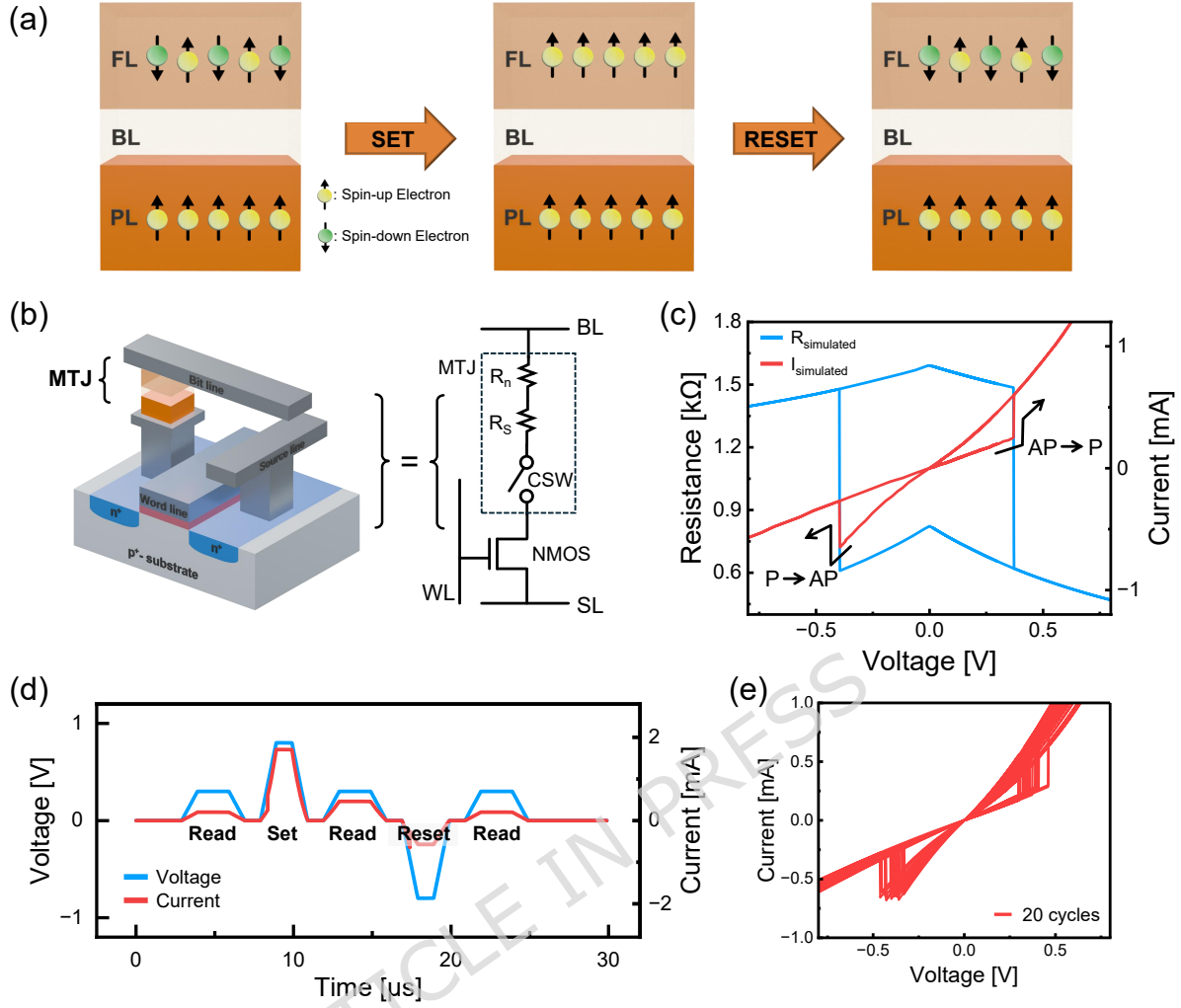


Figure 5. (a) STT-MRAM switching mechanism, illustrating spin alignment change modulating tunneling resistance. (b) Conceptual diagram and equivalent circuit model for the 1T1M STT-MRAM device. (c) Simulated I - V and R - V characteristics validating the Program (AP $\rightarrow$ P) and Erase (P $\rightarrow$ AP) transitions. (d) Transient pulse response validating STT-driven resistance change. (e) Monte Carlo simulation with variation in R_{on} , R_{off} , and the LTspice current-switch parameters (I_t , I_h), confirming model robustness.

Array-Level Simulation

The validated device-level models form the basis for subsequent array-level simulations. To demonstrate the system-level applicability of our methodology, we extended the RRAM and MRAM models, as representative examples, into array-level configurations. These array-level simulations enable the evaluation of fundamental computing operations such as Vector-Matrix Multiplication (VMM) and Multiply-Accumulate (MAC), which are key functions in neuromorphic and in-memory computing systems. This approach allows for the estimation of dynamic power consumption under realistic inference conditions, accounting for both device-level switching and interconnect effects. This supports a comprehensive assessment of array performance, including functionality, energy efficiency, and scalability, and demonstrates the practical applicability of our modeling framework.

RRAM Array Simulation with SNN Inference

To perform array-level computing simulations using the proposed generic model, we simulated array operations based on an SNN inference scheme, as illustrated in Fig. 6a.⁴⁸ We employed an off-chip training approach, where the weight map and input encoding were prepared in software and then transferred to the array-level simulation. First, to obtain the weight map, we implemented a $784 \times 24 \times 10$ artificial neural network (ANN) in PyTorch. The network was trained on the MNIST dataset with a batch size of 32 for 100 epochs. ReLU was used as the activation function and stochastic gradient descent

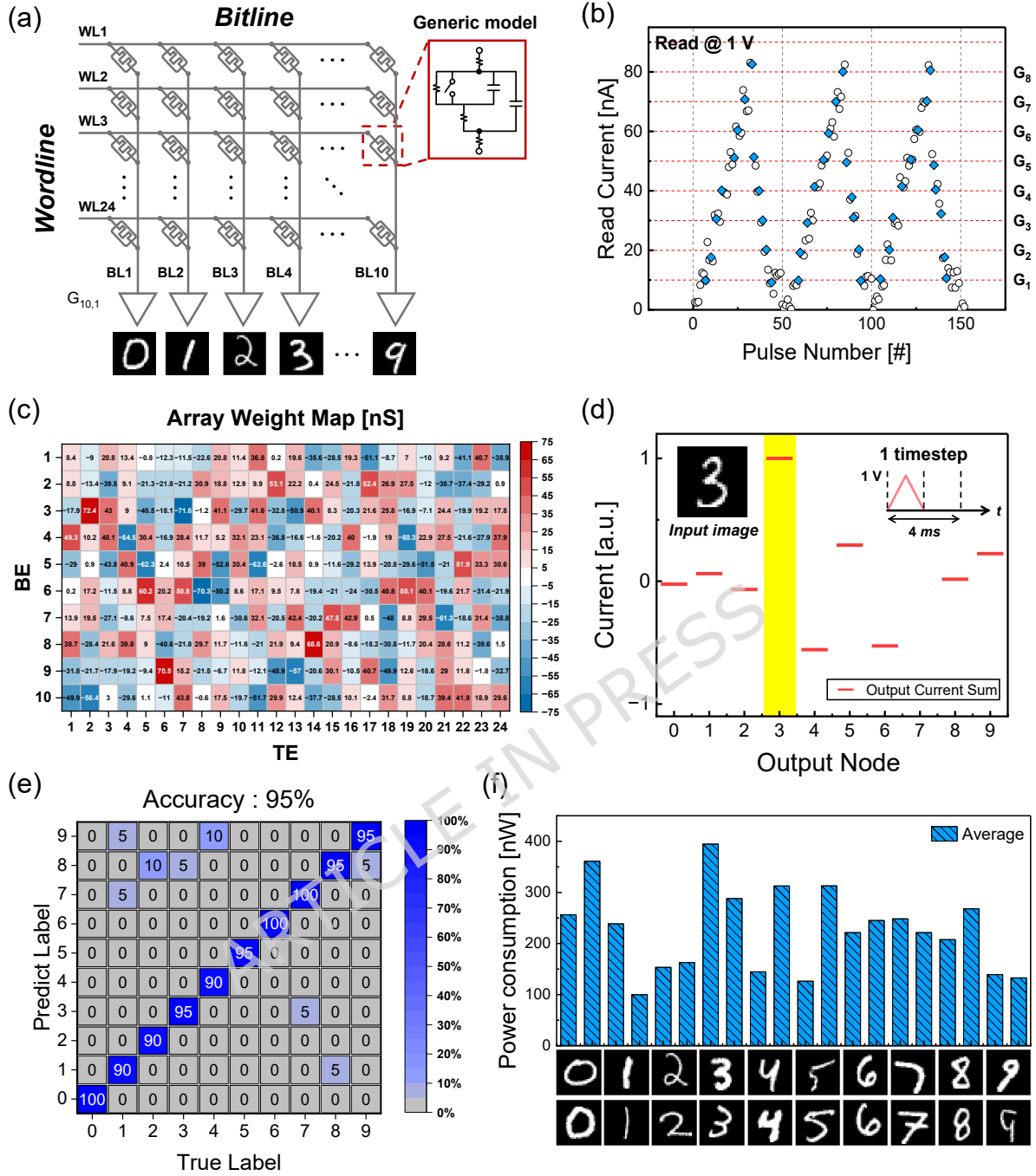


Figure 6. (a) Schematic of the RRAM-based SNN array co-simulation, implementing the 24×10 final layer with the generic RRAM model. (b) ISPP simulation (blue dots) fitted to the measured 3-bit (8-level) potentiation and depression characteristics (white dots), representing the conductance states. (c) Simulated 24×10 weight map after conductance programming. (d) Inference result for the input digit '3', with the highest current at the correct output node. (e) Confusion matrix of the SNN co-simulation over 200 test samples. (f) Average power consumption per digit class during MNIST inference.

(SGD) was employed as the optimizer. The weights were initialized using He initialization. After training, the weights were quantized to 3-bit precision. The quantized model achieved approximately 95% classification accuracy on the MNIST test set (10,000 images). Next, the trained ANN was converted to an SNN using a rate-coding scheme in software. Pixel intensities were normalized to the range $[0, 1]$ and encoded into spike trains over 25 time steps, where the spike rate was proportional to the normalized intensity. The spikes were applied to the first layer (784×24). The resulting currents were summed through

vector-matrix multiplication (VMM) using the trained weights and processed by an integrate-and-fire mechanism. The second layer (24×10) performed VMM in the same manner and generated output currents. Finally, the extracted weight map was transferred into the array-level simulation as the cell conductance states. The input spike trains were applied to the wordlines to execute VMM, and the output currents were read out from the bitlines and accumulated over the time window. The predicted label was determined by the bitline with the maximum accumulated current.

Each RRAM cell in the array was required to represent 3-bit weights. To achieve this, we adopted the generic model of multi-level RRAM, as presented in Fig. 2e. To verify the conductance range of this model, an Incremental Step Pulse Programming (ISPP) simulation was conducted. As shown in Fig. 6b, set pulses (3 V to 4.2 V in 25 mV increments) and reset pulses (−1 V to −1.6 V in 50 mV increments) were sequentially applied, and intermediate read pulses were used to verify the resulting states. Eight conductance levels (G1–G8) were formed within the range of 10 to 80 nA, with intervals of 10 nA, closely matching the multi-level characteristics observed at the device in Fig. 2e.

Additionally, to reflect the intrinsic variations of actual RRAM devices in set/reset behavior, Gaussian-distributed errors were embedded into the model. The programming-error distributions across conductance levels (G1–G8) and the validity of using a Gaussian noise source with fixed parameters ($\mu = 0$, $\sigma = 1.064$ nS) are provided in Supplementary Fig. S8. Weight transfer within the array was performed using $2 \mu\text{s}$ triangular pulses and a repeated read-verify scheme. To ensure precise tuning of each cell's conductance to its target value, we applied the Incremental Step Pulse Verification and Adjustment (ISPVA) method. As shown in Fig. 6c, approximately 94% of the cells converged within ± 2 nS of their target, while the remaining cells fell within ± 4 nS, confirming accurate and high-precision weight mapping.⁴⁹

Subsequently, inference simulations were carried out by applying input spike trains of 1 V amplitude and 4 ms period over 25 time steps to the array. As illustrated in Fig. 6d, an input image of the digit '3' was used, and the output neuron corresponding to the correct label exhibited the highest integrated current, confirming correct classification.

Moreover, classification accuracy was evaluated using 200 MNIST test samples, achieving approximately 95% accuracy (Fig. 6e). Finally, power analysis during inference (Fig. 6f) confirmed that the proposed model quantitatively reproduces realistic array-level behavior, validating its predictive capability for neuromorphic computing simulations. A scalability analysis of the relative simulation time for processing 200 input samples across various array configurations is provided in Supplementary Fig. S9.^{50,51}

The power consumed by a single RRAM cell was defined as the product of the applied voltage and the cell current. For an array with m WLs and n BLs, when a voltage was applied to a given WL, the total current through that WL was equal to the sum of the currents through the n connected cells (one per BL). This relationship indicated that the power consumption of a single WL could be expressed as the sum of the power dissipated by the cells connected to that WL. Accordingly, the power consumed on the i -th WL during inference was calculated as in Eq. (5).

$$\sum_{j=1}^n \int_0^T P_{ij}(t) dt = P_{tot,i} \quad (5)$$

The total power consumption of the array was then defined as the sum of the power consumed across all m WLs. The average power consumption was obtained by dividing the total energy consumption by the total inference time T , as expressed in Eq. (6).

$$\frac{1}{T} \sum_{i=1}^m P_{tot,i} = P_{AVG,array} \quad (6)$$

MRAM Array Simulation with BNN Inference

To demonstrate array-level computation using the STT-MRAM model, we simulated a BNN-based in-memory convolution operation. The validation combined software pre-processing with an LTspice array simulation. The target architecture (Fig. 7a) integrates an STT-MRAM sub-array with peripheral WL and BL drivers to enable parallel MAC operations.⁵²

A 2T-2J bitcell configuration was employed to perform hardware-level XNOR operations suitable for binarized inference (Fig. 7b). Each bitcell consists of two MTJs paired differentially, one storing the parallel state (representing a digital 1) and the other the anti-parallel state (representing a digital 0). When differential input signals are applied to the BL and BL_B lines, XNOR logic operations naturally emerge. The resulting currents are then summed through the shared source line (SL) to accumulate the MAC operation across multiple bitcells.⁵³

The simulation workflow (Fig. 7c) combined software pre-processing with the LTspice simulation. Three 3×3 convolution kernels and a sample input image (a dog) were binarized (per RGB channel) and flattened into 1D vectors in software. Subsequently, these vectors were used in the LTspice simulation, where the three 9-element kernel vectors were programmed into the three rows of a 3×9 sub-array model. The flattened image vectors were then applied sequentially as differential signals to the nine columns (BLs) to perform the convolution. This process was executed for 900 timesteps to generate a 30×30 feature

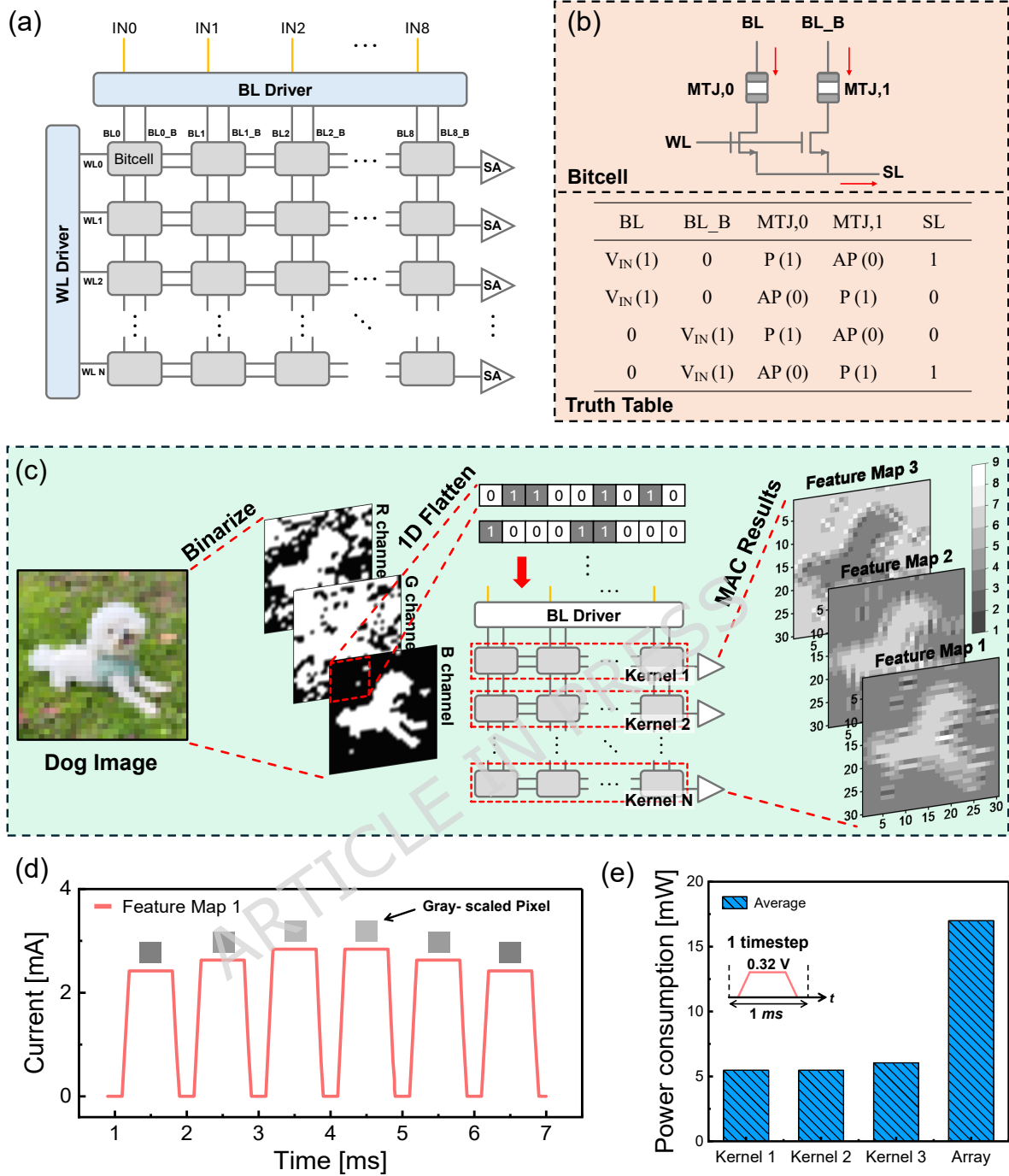


Figure 7. (a) Schematic of the STT-MRAM sub-array for BNN in-memory computation. (b) Circuit schematic of the 2T-2J differential bitcell and its XNOR truth table. (c) Co-simulation workflow, illustrating software pre-processing (Binarize, Flatten) and mapping to the LTspice array model. (d) Simulated output current waveform for Feature Map 1, where each current level corresponds to a single gray-scaled pixel value in the output map. (e) Average power consumption of the 3×9 STT-MRAM sub-array, detailing power for three kernels and the total array.

map. As shown in Fig. 7d, the accumulated SL current from each timestep was post-processed and mapped to a grayscale pixel, successfully forming the output feature maps.⁵⁴

Finally, the system's power consumption was analyzed. Each input pulse applied via the BL driver consisted of a 1 ms pulse width at 0.32 V amplitude, executed over 900 timesteps. As shown in Fig. 7e, Kernel 1 and Kernel 2 consumed approximately

5.47 mW each, while Kernel 3 consumed approximately 6 mW. The total power consumption across the 3×9 sub-array accumulated to approximately 17 mW. This array-level simulation confirms that the proposed generic model can be readily extended to support complex BNN-based in-memory computations and enables quantitative analysis of power characteristics under realistic inference workloads.

Conclusion

This work presented a generic compact modeling methodology for RRAM, TS, FTJ, and MRAM. This framework employs a generalized approach, using a minimal set of components (nonlinear resistors, capacitors, and voltage- or current-controlled switches) to construct distinct sub-models. This configuration accurately reproduces device characteristics without requiring device-specific physical equations.

At the device level, the resulting models successfully reflect the intrinsic switching mechanism of each corresponding memory. We demonstrated that the same voltage-dependent resistance formulation (Eq. 2) could phenomenologically capture filament-based switching (RRAM), defect-driven volatility (TS), polarization-modulated tunneling (FTJ), and spin-dependent tunneling (MRAM). All models demonstrated strong consistency with experimental and reference data.

To evaluate scalability, this framework was extended to array-level simulations. The RRAM model was used to implement a multi-level SNN for MNIST inference, and the MRAM model was used to create a 2T-2J differential XNOR array for BNN convolution. Both array simulations demonstrated correct inference behavior and validated the models' predictive capabilities at the system level.

Nevertheless, the proposed behavioral modeling approach entails inherent limitations compared to physics-based simulations. Specifically, the adoption of a simplified switching mechanism implies that complex transient dynamics—such as the precise pulse-width dependence of switching voltages or the gradual evolution of conductance—may not be fully captured without specific parameter tuning. Furthermore, while the model incorporates the temperature dependence of conduction, the thermal reset mechanism is abstracted as a threshold-triggered event to avoid the complexity of explicitly solving differential heat equations. Relevant temperature-dependent I - V characteristics and compact model fitting results are provided in Supplementary Fig. S10 and Supplementary Note 2. As a phenomenological model, it necessitates calibration against measured data, and simulations predicting unmeasured physical regimes beyond the characterized range may exhibit increased deviation.

However, these abstractions are intentional and prerequisite for achieving scalability. Unlike conventional physical models, which often involve computationally intensive equations (e.g., dynamic heat distributions), our methodology abstracts these transient physical details to generate compact models suitable for direct integration into circuit-level configurations. This approach enables the efficient simulation of large-scale inference workloads and the accurate estimation of system-level power consumption. Ultimately, this framework effectively bridges the gap between single-device physics and array-level system performance, providing a robust platform for the design and evaluation of heterogeneous memory architectures. A comprehensive feature-level comparison between the proposed generic model and representative compact models is provided in Supplementary Table S4. A quantitative runtime benchmark for Set/Reset operations of a 32×32 synaptic array, including comparisons to representative physics-based models, is summarized in Supplementary Table S3.

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

1. Lanza, M. *et al.* The growing memristor industry. *Nature* **640**, 613–622, DOI: <https://doi.org/10.1038/s41586-025-08733-5> (2025).
2. Menghani, G. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Comput. Surv.* **55**, 1–37, DOI: <https://doi.org/10.1145/3578938> (2023).
3. Roy, K., Jaiswal, A. & Panda, P. Towards spike-based machine intelligence with neuromorphic computing. *Nature* **575**, 607–617, DOI: <https://doi.org/10.1038/s41586-019-1677-2> (2019).
4. Wong, H.-S. P. *et al.* Metal–oxide rram. *Proc. IEEE* **100**, 1951–1970, DOI: <https://doi.org/10.1109/JPROC.2012.2190369> (2012).
5. Apalkov, D. *et al.* Spin-transfer torque magnetic random access memory (stt-mram). *ACM J. on Emerg. Technol. Comput. Syst. (JETC)* **9**, 1–35, DOI: <https://doi.org/10.1145/2463585.2463589> (2013).

6. Soni, R. *et al.* Giant electrode effect on tunnelling electroresistance in ferroelectric tunnel junctions. *Nat. Commun.* **5**, 5414, DOI: <https://doi.org/10.1038/ncomms6414> (2014).
7. Zhao, Z. *et al.* Chalcogenide ovonic threshold switching selector. *Nano-Micro Lett.* **16**, 81, DOI: <https://doi.org/10.1007/s40820-023-01289-x> (2024).
8. Yu, S. Neuro-inspired computing with emerging nonvolatile memorys. *Proc. IEEE* **106**, 260–285, DOI: <https://doi.org/10.1109/JPROC.2018.2790840> (2018).
9. Mannocci, P. *et al.* In-memory computing with emerging memory devices: Status and outlook. *APL Mach. Learn.* **1**, DOI: <https://doi.org/10.1063/5.0136403> (2023).
10. Nabil, A., Kumar, T. N. & Almurib, H. A. F. A generic volatile memristor model. *Phys. Scripta* **99**, 125031, DOI: <https://doi.org/10.1088/1402-4896/ad92e8> (2024).
11. Kvatinisky, S., Friedman, E. G., Kolodny, A. & Weiser, U. C. Team: Threshold adaptive memristor model. *IEEE transactions on circuits systems I: regular papers* **60**, 211–221, DOI: <https://doi.org/10.1109/TCSI.2012.2215714> (2012).
12. Prezioso, M. *et al.* Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61–64, DOI: <https://doi.org/10.1038/nature14441> (2015).
13. Chen, Y. Reram: History, status, and future. *IEEE Transactions on Electron Devices* **67**, 1420–1433, DOI: <https://doi.org/10.1109/TED.2019.2961505> (2020).
14. Zahoor, F., Azni Zulkifli, T. Z. & Khanday, F. A. Resistive random access memory (rram): an overview of materials, switching mechanism, performance, multilevel cell (mlc) storage, modeling, and applications. *Nanoscale research letters* **15**, 1–26, DOI: <https://doi.org/10.1186/s11671-020-03299-9> (2020).
15. Kim, H., Kim, S., Cho, J.-Y., Lee, S.-H. & Kim, M.-H. Investigation of dual memory behavior in rram: coexistence of resistive and capacitive switching phenomena. *Sci. reports* **15**, 12867, DOI: <http://doi.org/10.1038/s41598-025-96782-1> (2025).
16. Merrikh Bayat, F., Hoskins, B. & Strukov, D. B. Phenomenological modeling of memristive devices. *Appl. Phys. A* **118**, 779–786, DOI: <http://doi.org/10.1007/s00339-015-8993-7> (2015).
17. Jiang, Z. *et al.* A compact model for metal–oxide resistive random access memory with experiment verification. *IEEE Transactions on Electron Devices* **63**, 1884–1892, DOI: <https://doi.org/10.1109/TED.2016.2545412> (2016).
18. Yarragolla, S. *et al.* Identifying and understanding the nonlinear behavior of memristive devices. *Sci. Reports* **14**, 31633, DOI: <https://doi.org/10.1038/s41598-024-80568-y> (2024).
19. Liao, Y. *et al.* A compact model of analog rram with device and array nonideal effects for neuromorphic systems. *IEEE Transactions on Electron Devices* **67**, 1593–1599, DOI: <http://doi.org/10.1109/TED.2020.2975314> (2020).
20. D'Agostino, S. *Single neuron SNN with Memristor Generated Delays for Real-Time Analysis of Temporal Signals*. Ph.D. thesis, Politecnico di Torino (2022).
21. Swathi, S. & Angappane, S. Low power multilevel resistive switching in titanium oxide-based rram devices by interface engineering. *J. Sci. Adv. Mater. Devices* **6**, 601–610, DOI: <https://doi.org/10.1016/j.jsamd.2021.08.002> (2021).
22. Yarragolla, S., Hemke, T. & Mussenbrock, T. A generic compact and stochastic model for non-filamentary analog resistive switching devices. In *2023 12th International Conference on Modern Circuits and Systems Technologies (MOCASST)*, 1–4, DOI: <http://doi.org/10.1109/MOCASST57943.2023.10176574> (IEEE, 2023).
23. Degraeve, R. *et al.* Modeling and spectroscopy of ovonic threshold switching defects. In *2021 IEEE International Reliability Physics Symposium (IRPS)*, 1–5, DOI: <http://doi.org/10.1109/IRPS46558.2021.9405114> (IEEE, 2021).
24. Ravsher, T. *et al.* Polarity-dependent threshold voltage shift in ovonic threshold switches: Challenges and opportunities. In *2021 IEEE International Electron Devices Meeting (IEDM)*, 28–4, DOI: <http://doi.org/10.1109/IEDM19574.2021.9720649> (IEEE, 2021).
25. Kim, S., Heo, J., Kim, S. & Kim, M.-H. Dual functionality of nbo x memristors for synaptic and neuronal emulations in advanced neuromorphic systems. *J. Mater. Chem. C* **12**, 16294–16308, DOI: <http://doi.org/10.1039/D4TC03212H> (2024).
26. Heo, J., Kim, S., Kim, S. & Kim, M.-H. Configurable synaptic and stochastic neuronal functions in znTe-based memristor for an rbm neural network. *Adv. Sci.* **11**, 2405768, DOI: <https://doi.org/10.1002/advs.202405768> (2024).
27. Saludes-Tapia, M. *et al.* Tunability properties and compact modeling of hfo 2-based complementary resistive switches using a three-terminal subcircuit. *IEEE Transactions on Electron Devices* **68**, 5981–5988, DOI: <http://doi.org/10.1109/TED.2021.3117488> (2021).

28. Weinberg, Z. Tunneling of electrons from si into thermally grown sio₂. *Solid-State Electron.* **20**, 11–18, DOI: [https://doi.org/10.1016/0038-1101\(77\)90027-2](https://doi.org/10.1016/0038-1101(77)90027-2) (1977).
29. Chen, Y.-F., Hsu, L.-W., Hu, C.-W., Lai, G.-T. & Wu, Y.-H. Enhanced tunneling electro-resistance ratio for ferroelectric tunnel junctions by engineering metal work function. *IEEE Electron Device Lett.* **43**, 208–211, DOI: <http://doi.org/10.1109/LED.2021.3133577> (2021).
30. Barbot, J. *Electrical Characterization of Ferroelectric Tunnel Junctions based on Hafnium Oxide for Neuromorphic Applications*. Ph.D. thesis, Université Grenoble Alpes (2023). HAL Id: tel-04529177.
31. Hsiang, K.-Y. *et al.* Bilayer-based antiferroelectric hfnro₂ tunneling junction with high tunneling electroresistance and multilevel nonvolatile memory. *IEEE Electron Device Lett.* **42**, 1464–1467, DOI: <http://doi.org/10.1109/LED.2021.3107940> (2021).
32. Li, Y. *et al.* High-efficient and comprehensive modeling of mfim ferroelectric tunnel junctions for non-volatile/volatile applications. In *2024 IEEE International Memory Workshop (IMW)*, 1–4, DOI: <http://doi.org/10.1109/IMW59701.2024.10536918> (IEEE, 2024).
33. Jose, A. P. & Teh, Y.-K. Ferroelectric capacitor modelling in ltspice. In *2022 IEEE 15th Dallas Circuit And System Conference (DCAS)*, 1–4, DOI: <http://doi.org/10.1109/DCAS53974.2022.9845649> (IEEE, 2022).
34. Si, M. *et al.* A novel scalable energy-efficient synaptic device: Crossbar ferroelectric semiconductor junction. In *2019 IEEE International Electron Devices Meeting (IEDM)*, 6–6, DOI: <http://doi.org/10.1109/IEDM19573.2019.8993622> (IEEE, 2019).
35. Hwang, J., Goh, Y. & Jeon, S. Effect of insertion of dielectric layer on the performance of hafnia ferroelectric devices. *IEEE Transactions on Electron Devices* **68**, 841–845, DOI: <http://doi.org/10.1109/TED.2020.3043728> (2020).
36. Brinkman, W., Dynes, R. & Rowell, J. Tunneling conductance of asymmetrical barriers. *J. applied physics* **41**, 1915–1921, DOI: <https://doi.org/10.1063/1.1659141> (1970).
37. Li, J. *et al.* Reproducible ultrathin ferroelectric domain switching for high-performance neuromorphic computing. *Adv. Mater.* **32**, 1905764, DOI: <https://doi.org/10.1002/adma.201905764> (2020).
38. Huang, H.-H. *et al.* Performance enhancement and transient current response of ferroelectric tunnel junction: A theoretical study. *IEEE Transactions on Electron Devices* **69**, 4686–4692, DOI: <https://doi.org/10.1109/TED.2022.3184654> (2022).
39. Guo, R. *et al.* Functional ferroelectric tunnel junctions on silicon. *Sci. reports* **5**, 12576, DOI: <https://doi.org/10.1038/srep12576> (2015).
40. Garcia, V. & Bibes, M. Ferroelectric tunnel junctions for information storage and processing. *Nat. Nanotechnol.* **9**, 487–495, DOI: <https://doi.org/10.1038/ncomms5289> (2014).
41. Tuomisto, N., Zugarramurdi, A. & Puska, M. J. Modeling of electron tunneling through a tilted potential barrier. *J. Appl. Phys.* **121**, 134304, DOI: <https://doi.org/10.1063/1.4979533> (2017).
42. Zhang, Y. *et al.* Compact model of subvolume mtj and its design application at nanoscale technology nodes. *IEEE Transactions on Electron Devices* **62**, 2048–2055, DOI: <https://doi.org/10.1109/TED.2015.2414721> (2015).
43. Torunbalci, M. M., Upadhyaya, P., Bhave, S. A. & Camsari, K. Y. Modular compact modeling of mtj devices. *IEEE Transactions on Electron Devices* **65**, 4628–4634, DOI: <https://doi.org/10.1109/TED.2018.2863538> (2018).
44. Yoon, I. & Raychowdhury, A. Modeling and analysis of magnetic field induced coupling on embedded stt-mram arrays. *IEEE Transactions on Comput. Des. Integr. Circuits Syst.* **37**, 337–349, DOI: <https://doi.org/10.1109/TCAD.2017.2697963> (2018).
45. Kim, J. *et al.* A technology-agnostic mtj spice model with user-defined dimensions for stt-mram scalability studies. In *2015 IEEE Custom Integrated Circuits Conference (CICC)*, 1–4, DOI: <https://doi.org/10.1109/CICC.2015.7338407> (2015).
46. Prajapati, S., Verma, S., Kulkarni, A. A. & Kaushik, B. K. Modeling of a magnetic tunnel junction for a multilevel stt-mram cell. *IEEE Transactions on Nanotechnol.* **18**, 1005–1014, DOI: <https://doi.org/10.1109/TNANO.2018.2875491> (2019).
47. Pai, M. Y. X. *et al.* Analysis of the electrical characteristics for compact spice modelling of stt-mtj device with physical parameters variation. In *2024 IEEE International Conference on Semiconductor Electronics (ICSE)*, 5–8, DOI: <https://doi.org/10.1109/ICSE62991.2024.10681337> (2024).
48. Wu, H., Judd, P., Zhang, X., Isaev, M. & Micikevicius, P. Integer quantization for deep learning inference: Principles and empirical evaluation. *arXiv preprint arXiv:2004.09602* DOI: <https://doi.org/10.48550/arXiv.2004.09602> (2020).

49. Kim, M.-H., Lee, S.-H., Kim, S. & Park, B.-G. A fast weight transfer method for real-time online learning in rram-based neuromorphic system. *IEEE Access* **10**, 37030–37038, DOI: <https://doi.org/10.1109/ACCESS.2022.3157333> (2022).
50. Zhao, H. *et al.* Memristor-based signal processing for edge computing. *Tsinghua Sci. Technol.* **27**, 455–471, DOI: <https://doi.org/10.26599/TST.2021.9010043> (2022).
51. Hu, M. *et al.* Memristor-based analog computation and neural network classification with a dot product engine. *Adv. Mater.* **30**, 1705914, DOI: <https://doi.org/10.1002/adma.201705914> (2018).
52. Pham, T.-N., Trinh, Q.-K., Chang, I.-J. & Alioto, M. Stt-bnn: A novel stt-mram in-memory computing macro for binary neural networks. *IEEE J. on Emerg. Sel. Top. Circuits Syst.* **12**, 569–579, DOI: <https://doi.org/10.1109/JETCAS.2022.3169759> (2022).
53. Pham, T.-N., Trinh, Q.-K., Chang, I.-J. & Alioto, M. Stt-mram architecture with parallel accumulator for in-memory binary neural networks. In *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–5, DOI: <https://doi.org/10.1109/ISCAS51556.2021.9401695> (2021).
54. Nguyen, V.-T., Trinh, Q.-K., Zhang, R. & Nakashima, Y. Stt-bsnn: An in-memory deep binary spiking neural network based on stt-mram. *IEEE Access* **9**, 151373–151385, DOI: <https://doi.org/10.1109/ACCESS.2021.3125685> (2021).

Author contributions

S. Jung and H. Kim contributed equally to this work. S. Jung and H. Kim are responsible for the conceptualization, device measurement, circuit/software simulation, data curation, formal analysis, and writing of the original draft. J. Park performed the TCAD simulation. S. Kim provided writing review and editing, and supervision. M.-H. Kim handled the conceptualization, funding acquisition, supervision, validation, and writing review and editing. All authors reviewed the manuscript.

Funding

This research was supported in part by National R&D Program through the National Research Foundation of Korea (NRF) funded by Ministry of Science and ICT (RS-2023-00217539), in part by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE)(P0023718, Inorganic Light-emitting Display Expert Training Program for Display Technology Transition) and in part by the Chung-Ang University Research Scholarship Grants in 2024.

Ethics declarations

This article does not contain any studies with human participants or animals performed by any of the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information