

# Physics-based modeling of volatile resistive switching memory (RRAM) for crosspoint selector and neuromorphic computing

W. Wang<sup>1</sup>, A. Bricalli<sup>1</sup>, M. Laudato<sup>1</sup>, E. Ambrosi<sup>1</sup>, E. Covi<sup>1</sup>, and D. Ielmini<sup>1\*</sup>

<sup>1</sup>Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano and IU.NET, piazza L. da Vinci 32, 20133, Milano, Italy, \*email: [daniele.ielmini@polimi.it](mailto:daniele.ielmini@polimi.it)

**Abstract**—Volatile resistive switching memory (RRAM) is raising strong interest as potential selector device in crosspoint memory and short-term synapse in neuromorphic computing. To enable the design and simulation of memory and computing circuits with volatile RRAM, compact models are essential. To fill this gap, we present here a novel physics-based analytical model for volatile RRAM based on a detailed study of the switching process by molecular dynamics (MD) and finite-difference method (FDM). The analytical model captures all essential phenomena of volatile RRAM, e.g., threshold/holding voltages, on-off ratio, and size-dependent retention. The model is validated by extensive comparison with data from Ag/SiO<sub>x</sub> RRAM. To support the circuit-level capability of the model, we show simulations of crosspoint arrays and neuromorphic time-correlated learning.

## I. INTRODUCTION

Resistive switching memory (RRAM) devices based on metallic filaments made of Cu and Ag generally offer high on-off ratio ( $> 10^7$ ) [1, 2], steep switching slope ( $< 4$  mV/dec), and high endurance [3]. Although they were originally proposed as non-volatile switching memories [4], it was soon evidenced that they tend to display *volatile* switching due to the spontaneous turn-off within a short retention time from few  $\mu$ s to few ms [5]. Thanks to its high nonlinearity ( $> 10^7$ ), volatile RRAM has been proposed as select device in crosspoint arrays of memories or sensors [6], and as artificial synapse to mimic long/short-term memory phenomena in the brain [7]. However, a comprehensive understanding of the volatile switching mechanism is still missing [8]. Also, accurate compact models are needed to support all potential applications in memory and neuromorphic computing.

In this work, we introduce a new physics-based model for volatile RRAM. First, we study the switching process in volatile RRAM by molecular dynamics (MD) and finite-difference method (FDM) simulations of the spontaneous filament rupture. From this study, we derive a universal equation for size-dependent retention time. By combining the new retention model with a previous RRAM model for voltage-controlled filament growth/dissolution [9], we develop the first physics-based analytical model for volatile RRAM. The model accurately captures all device characteristics, such as DC and AC switching curves and retention. By including the spread of the activation energy, stochastic variations are also described. We finally show simulations of potential applications such as crosspoint selectors and neuromorphic synapses.

## II. VOLATILE BEHAVIOR OF METALLIC FILAMENT

### A. MD Simulations

Fig. 1a shows MD simulations of a nanoscale filament between two electrodes for increasing times at high temperature (800 K) using LAMMPS program [10, 11]. The simulation shows that the filament breaks up spontaneously without any external force after about 200 ps. This is because of the nanoscale “liquid-like pseudo-elasticity” [11], where metallic filaments change shape like a liquid drop to minimize the total surface energy. Atoms in the bulk remain crystallized while surface atomic migration accounts for the shape evolution at the origin of filament disconnection.

### B. Surface-tension Induced Filament Evolution

While highly illustrative of the disconnection process, the MD computational cost and time is not practical for simulating the switching and retention processes for typical temperature and times. FDM simulations were thus carried out by the equation:

$$J_s = - \left( \frac{D_s \gamma \delta^4}{kT} \right) \nabla_s \kappa, \quad (1)$$

where  $J_s$  is surface atomic flux along an arbitrary surface  $s$ ,  $D_s$  is the surface diffusion coefficient,  $\gamma$  is the surface energy,  $\delta$  is the interatomic distance of Ag atoms,  $k$  is the Boltzmann’s constant,  $T$  is the temperature, and  $\kappa$  is the surface curvature [12]. FDM simulations at 300 K in Fig. 1b show that the filament evolution is divided in two phases: (1) The filament diameter  $\phi$  decreases in the filament bottleneck ( $t < 0.2$  ms), (2) the filament is disconnected and the gap length  $g$  increases ( $t > 0.2$  ms). We thus define the retention time  $t_R$  as the time for the filament disconnection ( $\phi = 0$  and  $g = 0$ ). FDM simulations in Fig. 2 show that  $t_R$  strongly depends on the initial filament diameter  $\phi_0$  according to  $t_R \sim \phi_0^4$ .

### C. Modeling of Filament Spontaneous Disconnection

Based on numerical simulations in Fig. 1b, the spontaneous filament disconnection at the origin of the volatile behavior can be modeled by a simple rate equation of the filament diameter  $\phi$ , given by:

$$\left. \frac{d\phi}{dt} \right|_{sp} \propto - \left. \frac{dt_R}{d\phi_0} \right|_{t=t_R, \phi=\phi_0}, \quad (2)$$

where  $dt_R/d\phi_0$  can be obtained from Eq. (1), thus yielding  $d\phi/dt|_{sp} \propto -\phi^{-3}$  at the origin of the  $t_R \sim \phi_0^4$  dependence in Fig. 2.

## III. COMPACT MODEL

By combining the voltage-controlled model of set/reset switching in RRAM [9] with the retention model of Eq. (2), we can derive the comprehensive analytical model of volatile RRAM in Tab. I. Here, Eq. (3) describes both voltage-driven set process [9] and spontaneous filament dissolution by  $d\phi/dt|_{sp} \propto -\phi^{-3}$ , the latter accounting for size-dependent retention. The increase of the filament gap  $g$  for  $t > t_R$  is described by Eq. (4), allowing to describe the post-dissolution evolution of the off-state current. The switching mechanism in the volatile RRAM is summarized in Fig. 3, showing the filament growth under an applied voltage, and the phenomena of *disconnection* ( $t < t_R$ ) and *retraction* ( $t > t_R$ ) arising spontaneously. In the model, the RRAM resistance was computed by the series model in Fig. 4a and b, which is analytically captured by Eqs. (5)-(8). Correspondingly, the effective temperature  $T$  during set/reset processes was computed by the Joule heating model in Fig. 4c, described by Eqs. (9)-(11). The model thus describes all essential phenomena of volatile RRAM, including switching and spontaneous disconnection.

## IV. MODEL VALIDATION

### A. DC and AC Characteristics

Experimental data were collected from RRAM devices with an Ag top electrode, a SiO<sub>x</sub> dielectric layer with thickness  $t_{\text{ox}} = 5$  nm, and a bottom electrode made of graphitic carbon [5, 13]. Fig. 5a shows the device structure, while Fig. 5b shows the measured and calculated I-V curves for compliance current  $I_C = 35$   $\mu\text{A}$ . The high on-off ratio ( $> 10^7$ ) and steep switching slope are well captured by the model. As a positive threshold voltage  $V_{T+}$  is applied, the device switches to the on-state by increasing its current from  $10^{-12}$  A to  $\sim 10^{-5}$  A. The compliance current  $I_C$  was limited by an integrated select transistor in the one-transistor/one-resistor (1T1R) structure [13]. As the voltage is reduced below a characteristic holding voltage  $V_{H+}$ , the device spontaneously switches to the off-state by filament disconnection, marked by the steep drop of the current (Fig. 3). Thanks to the bidirectional behavior of the Ag/SiO<sub>x</sub> device [13], switching to the on-state also occurs at negative voltage  $V_{T-}$ , with negative holding voltage  $V_{H-}$ . The bidirectional switching is well captured by the model with the parameters reported in Tab. II. The model could describe the I-V curves for any value of  $I_C$  from 0.1  $\mu\text{A}$  to 80  $\mu\text{A}$ . Fig. 6 shows the measured and calculated  $V_{T+}$ ,  $V_{H+}$ ,  $V_{T-}$  and  $V_{H-}$  as a function of  $I_C$ , supporting the accuracy of the model in a broad range of  $I_C$ .

Fig. 7 shows AC characteristics obtained by applying a triangular pulse of pulse width  $t_p = 100$   $\mu\text{s}$ , followed by a constant read voltage  $V_{\text{read}} = 0.1$  V to monitor the time evolution of device resistance  $R$  (Fig. 7a). The current response (Fig. 7b) indicates a steep switching to the on-state at  $I_C = 20$   $\mu\text{A}$ , while the following read current shows a decay behavior with  $t_R \sim 1.5$  ms. By this technique, we studied  $t_R$  as a function of the pulse width  $t_p$  in both experiments and simulations. Fig. 8 shows that  $t_R$  increases with  $t_p$ , thanks to the gradual increases of  $\phi_0$  with  $t_p$  and the consequently enhanced retention according to Fig. 2.

### B. Cycle-to-cycle Variations

Statistical variations play a dominant role for volatile RRAM devices in both memory and computing applications. Cycle-to-cycle variability of  $V_T$  and  $V_H$  were computed by Monte Carlo simulations where we included a stochastic spread of the activation energy [14]. Fig. 9a shows the distributions of  $V_T$  and  $V_H$  from experiments and calculations, with the assumed spread of energy barriers (inset). The model can also predict the  $t_R$  distributions for various pulse width  $t_p$ , as shown in Fig. 9b.

The relatively long  $t_R \sim 1$  ms of the volatile RRAM device is a concern for its application as selectors, due to half-selected cells remaining in the on-state after read. To capture this scenario, Fig. 10 shows a two-pulse experiment where the first pulse induces switching to the on-state, then a second pulse is used to probe the state of the cell after a certain delay time  $t_D$ . The RRAM device is still found in the on-state ( $V_{T2} = 0$ ) for short time delay  $t_D = 10$   $\mu\text{s}$  (Fig. 10a), whereas the off-state is recovered for a longer delay  $t_D = 10$  ms (Fig. 10b). Fig. 11 reports the threshold voltage  $V_{T2}$  observed in the second pulse as a function of  $t_D$ , indicating a recovery from on-state to off-state in about  $t_R = 1$  ms under both positive and negative voltage. Our model reproduces the experimental results by 1,000 Monte Carlo simulations at each  $t_D$ . Fig. 12 shows the probability of finding the device in the off-state ( $V_{T2} > 0$ ) as a function of delay time.

## V. APPLICATIONS AS SELECT DEVICE

Our model can be used to simulate the volatile RRAM within a one-selector/one-RRAM (1S1R) element of a crosspoint array (Fig. 13a). Fig. 13b shows the calculated I-V curves for the memory element, obtained by our previous non-volatile RRAM model [15]. Fig. 13c shows the combined I-V curve of the 1S1R, clearly indicating memory set/reset transitions and selector on/off switching. By utilizing a V/2 reading scheme (Fig. 14a) with read voltage  $V_R = 2.5$  V, the *static* window between a selected low-resistance state (A in Fig. 13d) and a half-selected state (B in Fig. 13d) is  $\sim 10^7$ , which translates in a feasible  $N \times N$  array with  $N \sim 10^7$  (Fig. 14b) [16]. However,

*dynamic* simulations in Fig. 14c show that the read margin strongly decreases at increasing  $t_R$ , due to half-selected cells being recently accessed (C in Fig. 13d) and remaining in the on-state due to the long retention time.

## VI. APPLICATIONS IN NEUROMORPHIC COMPUTING

While limiting the applications in crosspoint memories, the retention time in the few-ms range well matches with the biological timescale for short-term plasticity, thus making volatile RRAM suitable for neuromorphic computing [17]. Fig. 15 shows the current response of a volatile RRAM to a train of spikes with frequency  $f_{\text{spike}} = 2$  kHz using a series resistor compliance ( $R_C = 10$  k $\Omega$ ). The spiking stimulation results in paired-pulse facilitation (PPF), namely a gradual increase of the RRAM conductance due to filament growth during each spike. Reducing  $f_{\text{spike}}$  to 250 Hz in Fig. 16 causes the transition to paired-pulse depression (PPD) [7], as filament disconnection dominates over growth. The resulting spike-rate dependent plasticity (SRDP) is shown in Fig. 17a, highlighting the conditions for PPF and PPD.

The spiking frequency also affects the retention time, thus enabling tunable long-term memory (LTM,  $t_R > 20$  ms) and short-term memory (STM,  $t_R < 10$  ms) regimes (Fig. 17b). Unlike long-term plasticity, LTM is a transient memory which enables coupling of sensory information received at any time to the one received earlier, thus supporting speech recognition [18] and sequence learning [19]. The STM and LTM concepts are demonstrated in Fig. 18: high-frequency learning of the image ‘5’ results in LTM (a), while low-frequency learning results in STM (b). A newly submitted pattern ‘0’ can thus be coupled to the previous one ‘5’ (Fig. 18c) which is essential for the time-correlated pattern learning in a spatio-temporal spiking neural network [19], while no sequence recognition is possible in the case of STM (Fig. 18d).

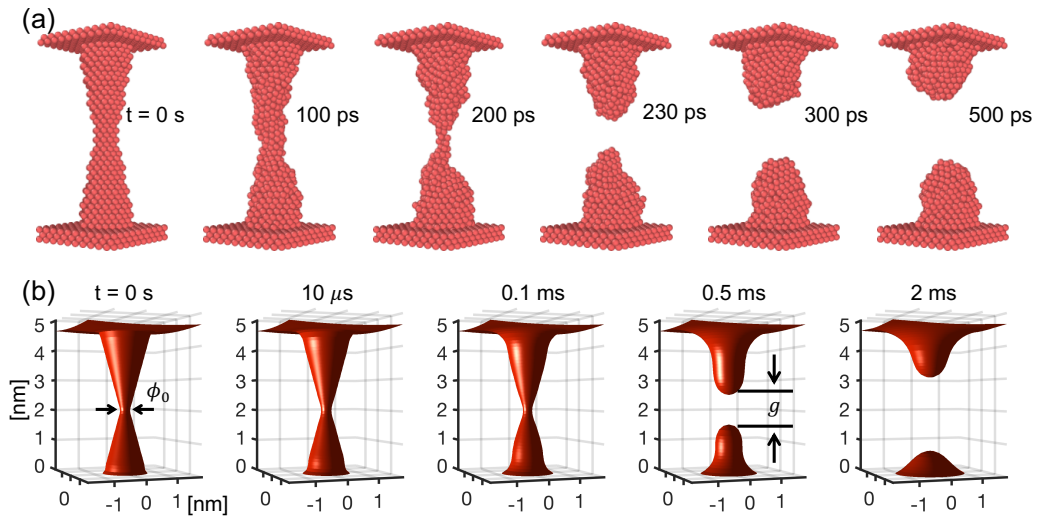
## VII. CONCLUSION

In summary, we presented a physics-based analytical model for volatile RRAM, capturing DC and AC characteristics including switching and retention, and their stochastic variations. The model is validated against experimental data and deployed for the circuit-level simulation of crosspoint arrays and neuromorphic systems.

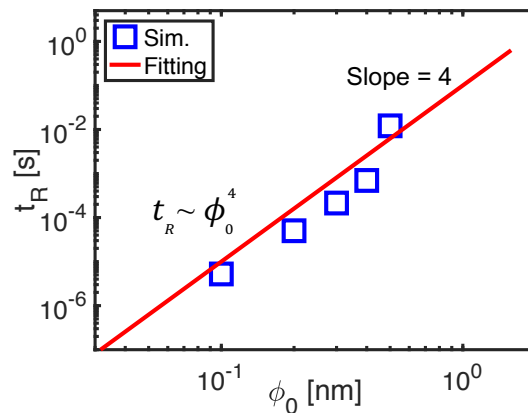
This work has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 648635).

## REFERENCES

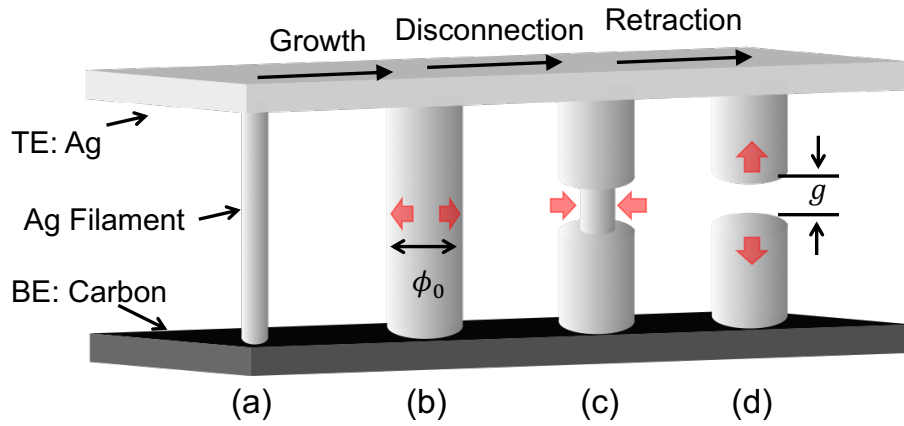
- [1] Y. D. Zhao, *et al.*, IEEE IEDM, **2016**, 7.6. [2] Q. Luo, *et al.*, IEEE IEDM, **2015**, 10.4. [3] X. Zhao, *et al.*, Adv. Mater., **30**, 1705193 (2018). [4] M. N. Kozicki, *et al.*, IEEE Trans. Nanotechnology, **3**, 331 (2005). [5] A. Bricalli, *et al.*, IEEE Trans. Electron Devices, **65**, 122 (2018). [6] M. Wang, *et al.*, Adv. Mater., 1802516 (2018). [7] Z. Wang, *et al.*, Nat. Mater., **16**, 101 (2017). [8] N. Shukla, *et al.*, IEEE IEDM, **2017**, 4.3. [9] D. Ielmini, IEEE Trans. Electron Devices, **58**, 4309 (2011). [10] S. Plimpton, J. Comput. Phys., **117**, 1 (1995). [11] J. Sun, *et al.*, Nat. Mater., **13**, 1007 (2014). [12] F. A. Nichols, *et al.*, J. Appl. Phys., **36**, 1826 (1965). [13] A. Bricalli, *et al.*, IEEE IEDM, **2016**, 4.3. [14] S. Ambrogio, *et al.*, IEEE Trans. Electron Devices, **61**, 2920 (2014). [15] S. Ambrogio, *et al.*, IEEE Trans. Electron Devices, **61**, 2378 (2014). [16] J.-J. Huang, *et al.*, IEEE Electron Device Lett., **32**, 1427 (2011). [17] T. Ohno, *et al.*, Nat. Mater., **10**, 591 (2011). [18] S. Hochreiter, *et al.*, Neural Comput., **9**, 1735 (1997). [19] W. Wang, *et al.*, Sci. Adv., **4**, eaat4752 (2018).



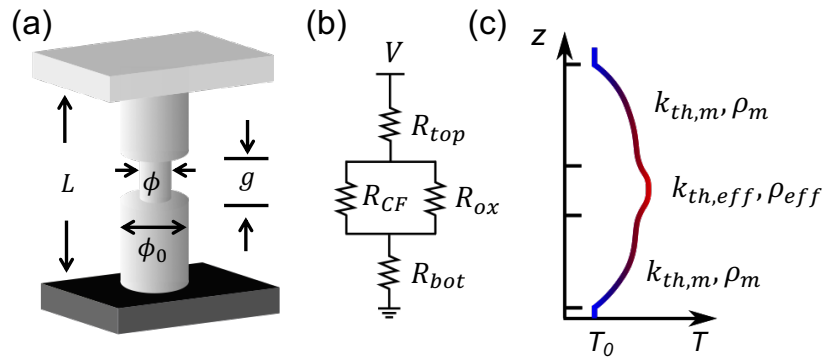
**Fig. 1.** (a) Accelerated (800K) molecular dynamic (MD) simulation of the relaxation process of a nanoscale filament between two electrodes. Spontaneous filament opening is observed due to surface diffusion. (b) Finite-difference method (FDM) simulation of the filament evolution.



**Fig. 2.** The calculated retention time of a metallic filament, indicating a strong dependence on initial filament diameter  $\phi_0$ .



**Fig. 3.** Schematic view of the complete analytical model of volatile RRAM including: (a)-(b) growth of the filament, (b)-(c) disconnection, i.e., diameter decrease in the bottle neck, and (c)-(d) retraction, i.e., increase of gap length  $g$ .



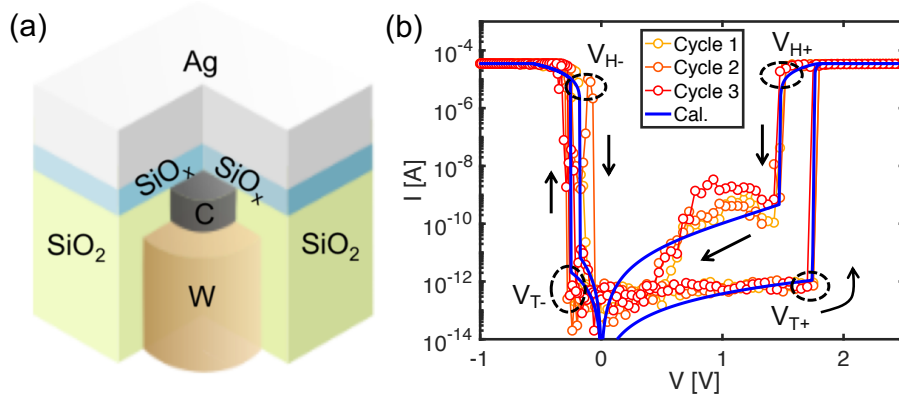
**Fig. 4.** Schematic views of morphological parameters of the filament (a), equivalent electrical circuit for resistance calculation (b), and parameters for Joule heating equations to obtain the filament temperature (c).

**Table I.** Equations used in the model

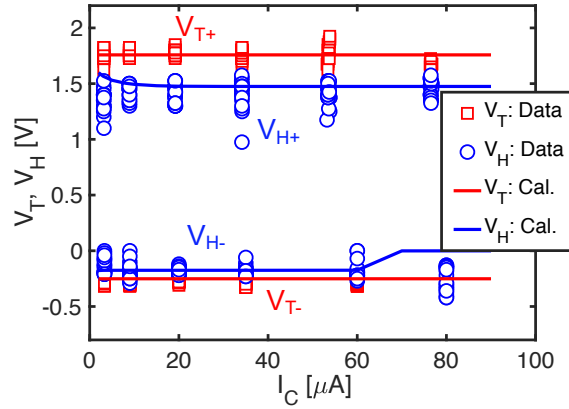
(3)	$\frac{d\phi}{dt} = Ae^{-\frac{E_{A0}-\alpha qV}{kT}} - \frac{C}{\phi^3}e^{-\frac{E_{A1}}{kT}}, \phi > \phi_a$	$A, C$ : prefactors $E_{A0}, E_{A1}$ : activation energy
(4)	$g = L(1 - e^{-t/\tau_{rt}}), \phi < \phi_a$	$\phi_a$ : single atom filament diameter $\tau_{rt}$ : stub retraction time constant
(5)	$R = R_{top} + R_{bot} + R_{CF}    R_{ox}$	
(6)	$R_{top} + R_{bot} \approx \rho_m \frac{4(L-g)}{\phi_0^2}, R_{CF} \approx \rho_m \frac{4L}{\phi^2}, R_{ox} \approx \rho_{ox} \frac{4g}{(\phi_0 - \phi)^2}$	
(7)	$\rho_m = \rho_{m0} \frac{1-p l_f}{1+p \phi}$	$\rho_{m0}$ : bulk metal resistivity
(8)	$\rho_{ox} = \frac{\rho_{ox0}}{1+\gamma F}$	$\rho_{ox}$ : oxide resistivity
(9)	$T = T_0 + \frac{J^2 \rho_m}{8 k_{th}} (L^2 - g^2) + \frac{J^2 \rho_{eff}}{8 k_{th,eff}} g^2$	$T_0$ : ambient temperature $k_{th,m}$ : metal thermal conductivity
(10)	$k_{th,eff} = \phi^2 / \phi_0^2 k_{th,m} + (1 - \phi^2 / \phi_0^2) k_{th,ox}$	$k_{th,ox}$ : oxide thermal conductivity
(11)	$\rho_{eff} = \phi_0^2 (\rho_m / \phi^2    \rho_{ox} / (\phi_0^2 - \phi^2))$	

**Table II.** Relevant parameters for the calculation

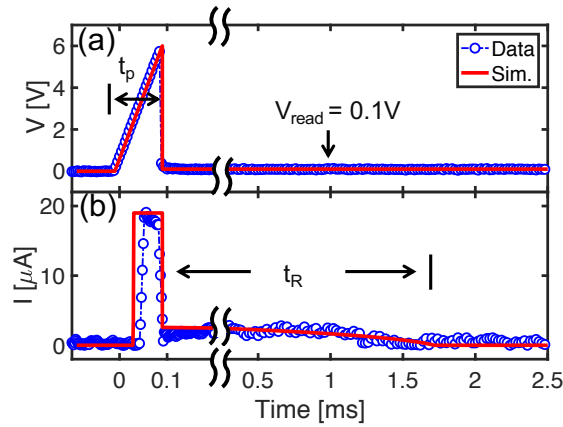
Param.	Value (Pos.)	Value (Neg.)	Param.	Value	Param.	Value
$\alpha$	0.3	0.62	A	$1 \times 10^3$ m/s	$\rho_{ox}$	$2 \times 10^4 \Omega \cdot m$
$E_{A0}$	0.88 eV	0.58 eV	C	$4 \times 10^{-22}$ m <sup>4</sup> /s	$k_{th,m}$	$5 \times 10^3$ W/(m·K)
$E_{A1}$	0.59 eV	0.56 eV	$\rho_{m0}$	$2 \mu\Omega m$	$k_{th,ox}$	$1$ W/(m·K)



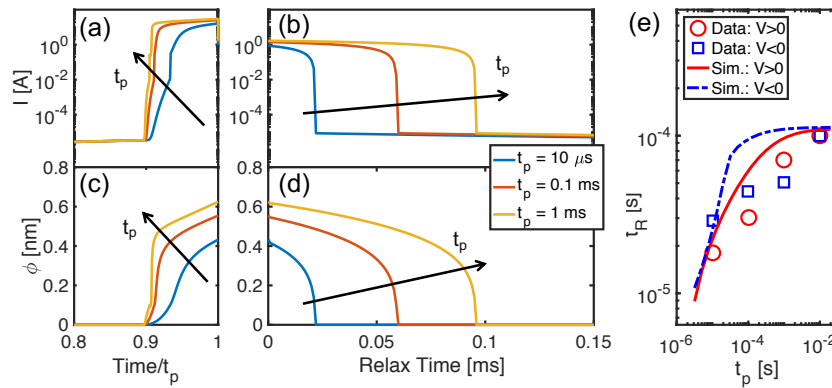
**Fig. 5.** (a) Device structure of the Ag-based volatile switching memory [5]. (b) Experimental and calculated DC I-V curves of the volatile RRAM device. Both filament disconnection and retraction are visible as abrupt on-state current decrease and gradual off-state current decrease, respectively.



**Fig. 6.** Experimental and calculated threshold voltage ( $V_{T+}$  and  $V_{T-}$ ) and holding voltage ( $V_{H+}$  and  $V_{H-}$ ) as a function of compliance current  $I_C$ .

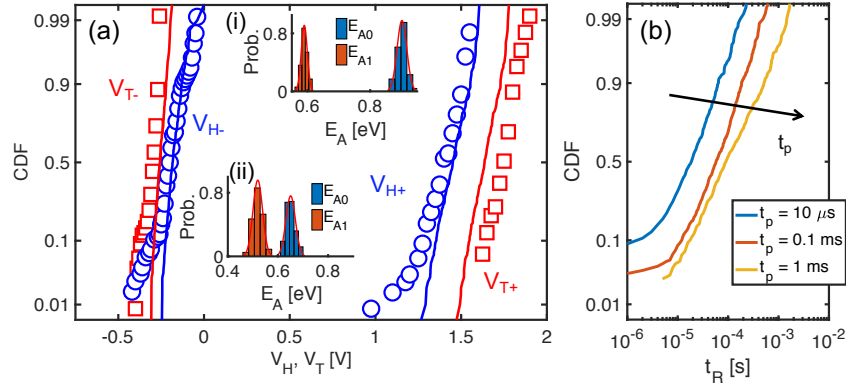


**Fig. 7.** Measured and calculated voltage (a) and current (b) indicating switching during the triangular pulse and retention/relaxation after  $t_R$  from the end of the pulse.

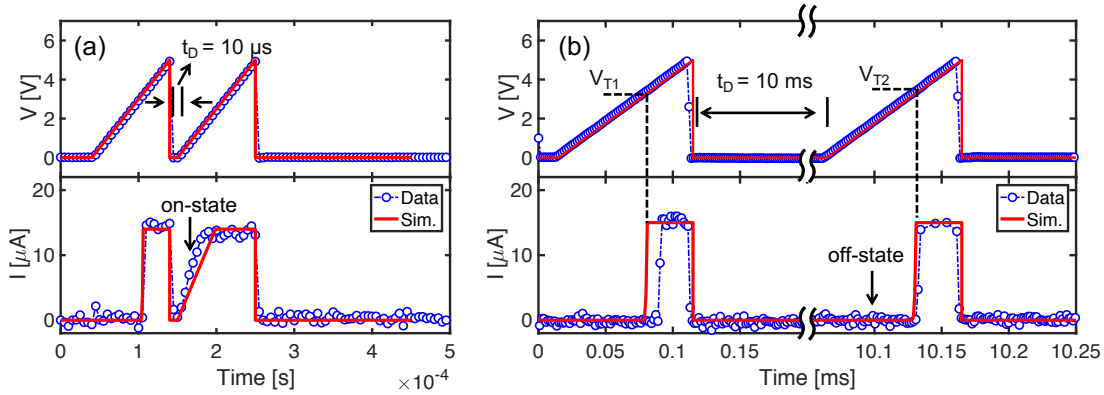


**Fig. 8.** Calculated current response (a, b) and filament diameter (c, d) for pulse experiments similar to Fig. 7, during either the triangular pulse (a, c) or the read phase (b, d) for various pulse width  $t_p$ . (e) Calculated  $t_R$  as a function of  $t_p$  for both positive and negative pulses, compared to experimental data from [13].

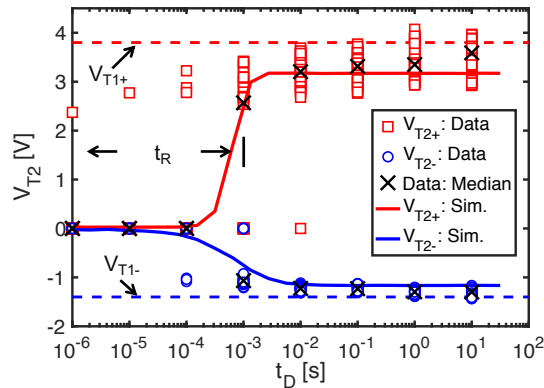




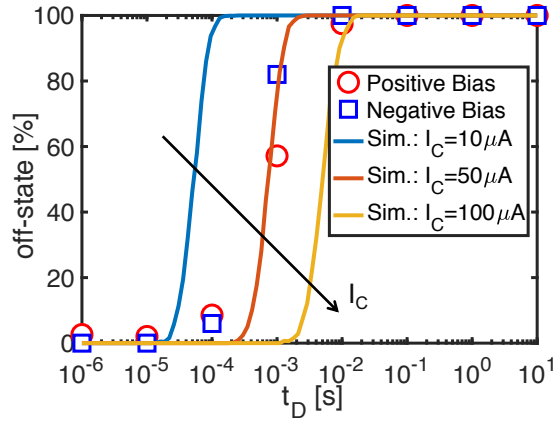
**Fig. 9.** (a) Distributions of measured (symbols) and calculated (lines) threshold voltage and holding voltage of the volatile RRAM, with insets showing the distributions of activation energies for positive bias (i) and negative bias (ii). (b) Distribution of the calculated retention time  $t_R$  for increasing pulse width  $t_p$ .



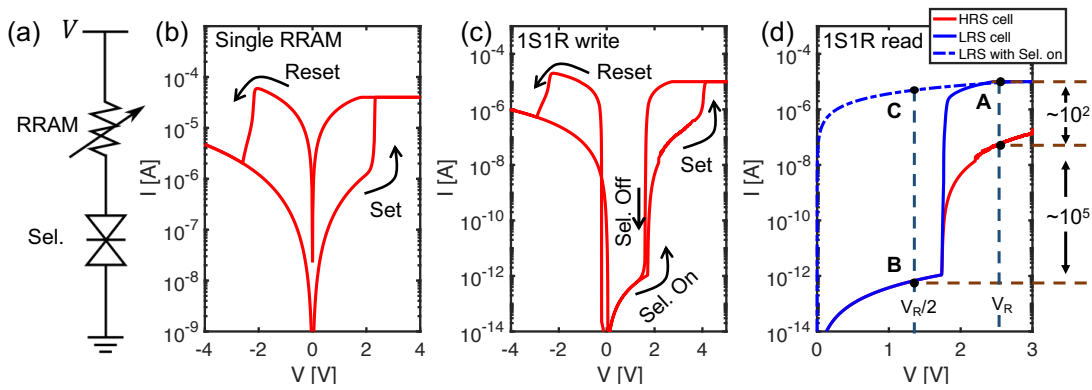
**Fig. 10.** Measured and calculated voltage and current for two-pulse experiments with delay time  $t_D = 10 \mu\text{s}$  (a) and  $10 \text{ ms}$  (b). The device is still found in the on-state for relatively short  $t_D$ , whereas the off-state is recovered for relatively long  $t_D$ , as revealed by the threshold voltage in the 2<sup>nd</sup> pulse  $V_{T2}$  being similar to the first one  $V_{T1}$ .



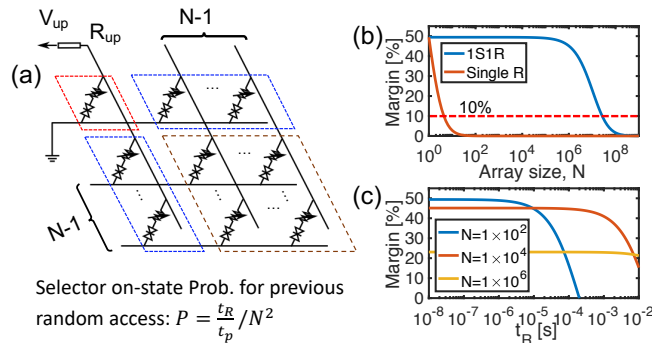
**Fig. 11.** Measured and calculated threshold voltage in the 2<sup>nd</sup> pulse for positive/negative pulses as a function of  $t_D$  indicating retention time  $t_R$  of about  $1 \text{ ms}$ .



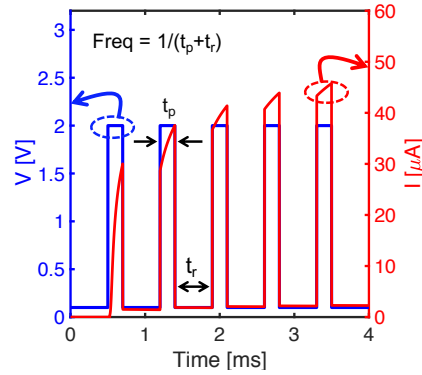
**Fig. 12.** Probability of finding the device in the off-state in correspondence of the 2<sup>nd</sup> pulse as a function of  $t_D$  in a two-pulse experiment as in Fig. 10 for increasing  $I_C$ .



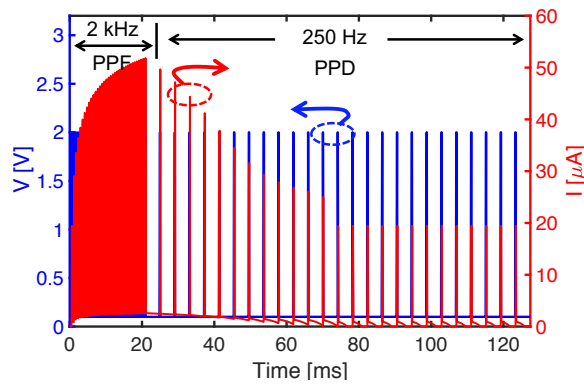
**Fig. 13.** (a) 1S1R cell configuration. (b) Calculated I-V curve of single non-volatile RRAM with  $I_C = 40 \mu A$  by our previous RRAM analytical model [15]. (c) Calculated I-V curve of a 1S1R structure by combining the non-volatile RRAM analytical model and the developed volatile RRAM model. (d) Nonlinearities of the 1S1R cell to calculate read margins.



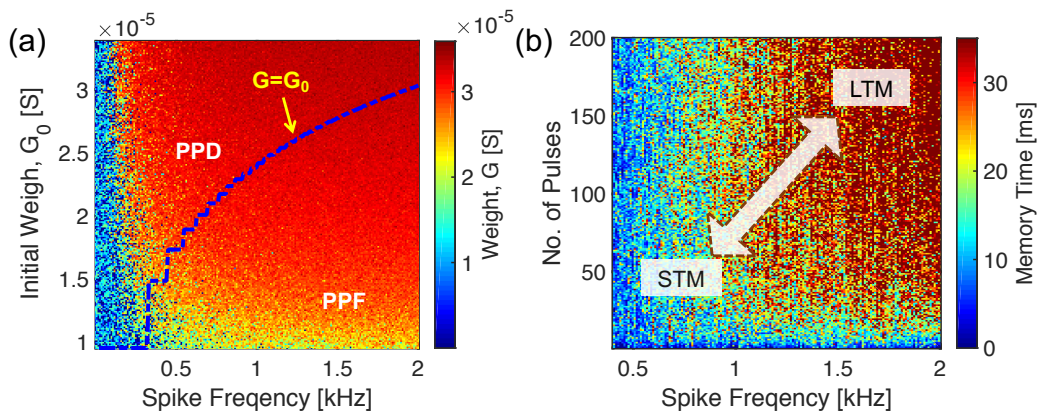
**Fig. 14.** (a) 1S1R crosspoint array with volatile RRAM selectors. (b) Static read margin as a function of array size. The size of the array is projected to be  $N \sim 10^7$ . (c) Dynamic read margin, where random access causes some selectors to remain in the on-state while read, thus strongly reducing the read margin for  $t_p = 10$  ns.



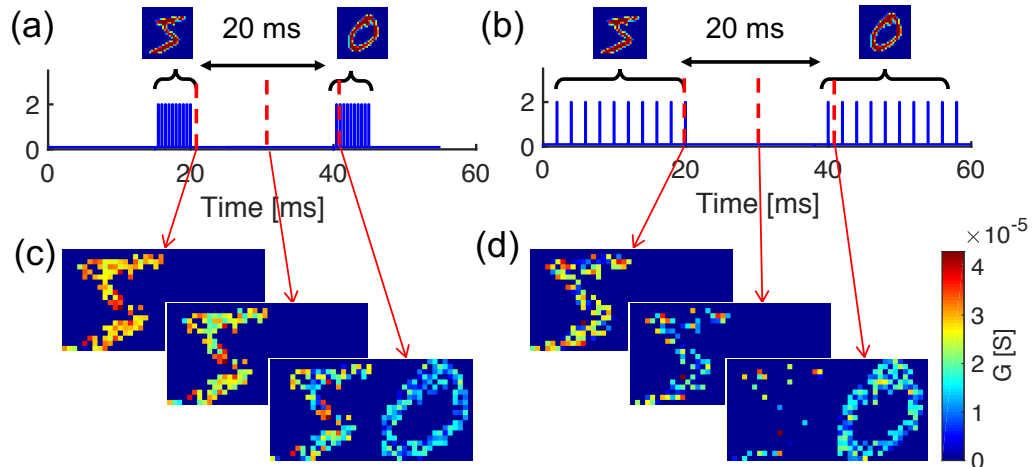
**Fig. 15.** A train of pulses (spikes) induce gradual increase of current of the volatile RRAM. Relatively low pulse amplitude (2 V) is applied to limit the switching speed.



**Fig. 16.** A train of spikes causes paired-pulse facilitation (PPF) for relatively high frequency  $f_{\text{spike}} = 2 \text{ kHz}$ , and paired-pulse depression (PPD) for relatively low frequency  $f_{\text{spike}} = 250 \text{ Hz}$ . Similar experimental result has been reported in Ag-based volatile RRAM [7].



**Fig. 17.** (a) Color map of synaptic weight  $G$  as a function of spike frequency  $f_{\text{spike}}$  and initial weight  $G_0$  after a train of 10 spikes, indicating the regimes for PPF and PPD, separated by the line of stable weight  $G = G_0$ . (b) Color map of retention time  $t_R$  as a function of  $f_{\text{spike}}$  and number of spikes, indicating regimes for long-term memory ( $t_R > 20 \text{ ms}$ ) and short-term memory ( $t_R < 10 \text{ ms}$ ).



**Fig. 18.** Simulation of visual sequence recognition in a synaptic array where '5' and '0' are sequentially submitted to recognize the number '50' (a, b). (c) Spiking at high  $f_{\text{spike}}$  leads to LTM, i.e. new pattern '0', can be linked to the previous one '5'. (d) Spiking at low  $f_{\text{spike}}$  leads to STM, which prevents the recognition of the full sequence due to lack of LTM.