frontiers in Physics

ORIGINAL RESEARCH
published: 23 September 2021
doi: 10.3389/fphy.2021.735021

SPICE Simulation of RRAM-Based Cross-Point Arrays Using the Dynamic Memdiode Model

Fernando L. Aguirre[1,2,3]*, *Sebastián M. Pazos*[1,2], *Félix Palumbo*[1,2], *Jordi Suñé*[3] and *Enrique Miranda*[3]*

[1]*Unidad de Investigación y Desarrollo de las Ingenierías, Facultad Regional Buenos Aires, Universidad Tecnológica Nacional, Buenos Aires, Argentina,* [2]*Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina,* [3]*Departament d'Enginyeria Electrònica, Universitat Autònoma de Barcelona, Barcelona, Spain*

We thoroughly investigate the performance of the Dynamic Memdiode Model (DMM) when used for simulating the synaptic weights in large RRAM-based cross-point arrays (CPA) intended for neuromorphic computing. The DMM is in line with Prof. Chua's memristive devices theory, in which the hysteresis phenomenon in electroformed metal-insulator-metal structures is represented by means of two coupled equations: one equation for the current-voltage characteristic of the device based on an extension of the quantum point-contact (QPC) model for dielectric breakdown and a second equation for the memory state, responsible for keeping track of the previous history of the device. By considering *ex-situ* training of the CPA aimed at classifying the handwritten characters of the MNIST database, we evaluate the performance of a Write-Verify iterative scheme for setting the crosspoint conductances to their target values. The total programming time, the programming error, and the inference accuracy obtained with such writing scheme are investigated in depth. The role played by parasitic components such as the line resistance as well as some CPA's particular features like the dynamical range of the memdiodes are discussed. The interrelationship between the frequency and amplitude values of the write pulses is explored in detail. In addition, the effect of the resistance shift for the case of a CPA programmed with no errors is studied for a variety of input signals, providing a design guideline for selecting the appropriate pulse's amplitude and frequency.

Keywords: RRAM, resistive switching, cross-point, memristor, neuromorphic, pattern recognition, write-verify, frequency

OPEN ACCESS

Edited by:
Enrico Piccinini, Applied Materials, Italy

Reviewed by:
Alessandro Cabrini, University of Pavia, Italy
Marco A. Villena, Applied Materials, United States

***Correspondence:**
Fernando L. Aguirre
aguirref@ieee.org
Enrique Miranda
aenrique.miranda@uab.cat

Specialty section:
This article was submitted to Interdisciplinary Physics, a section of the journal Frontiers in Physics

Received: *01 July 2021*
Accepted: *07 September 2021*
Published: *23 September 2021*

Citation:
Aguirre FL, Pazos SM, Palumbo F, Suñé J and Miranda E (2021) SPICE Simulation of RRAM-Based Cross-Point Arrays Using the Dynamic Memdiode Model. Front. Phys. 9:735021. doi: 10.3389/fphy.2021.735021

INTRODUCTION

The Matrix-Vector Multiplication (MVM) method is a key piece for computation in Artificial Neural Networks (ANNs), which have demonstrated outstanding results in the field of pattern recognition and classification, among others [1]. When performed in Von Neumann conventional computing systems, the MVM implies a time complexity of ~$O(n^2)$, n being the number of components, which severely compromises the ANN power consumption and latency. Although highly parallelized CMOS-based implementations enable fast operation (low latency) they fall short in cutting off power consumption, not to mention the silicon area requirements. As no drastic performance improvements can be expected from further technology scaling [2], alternative approaches involving novel architectures and materials are being extensively researched worldwide. Among

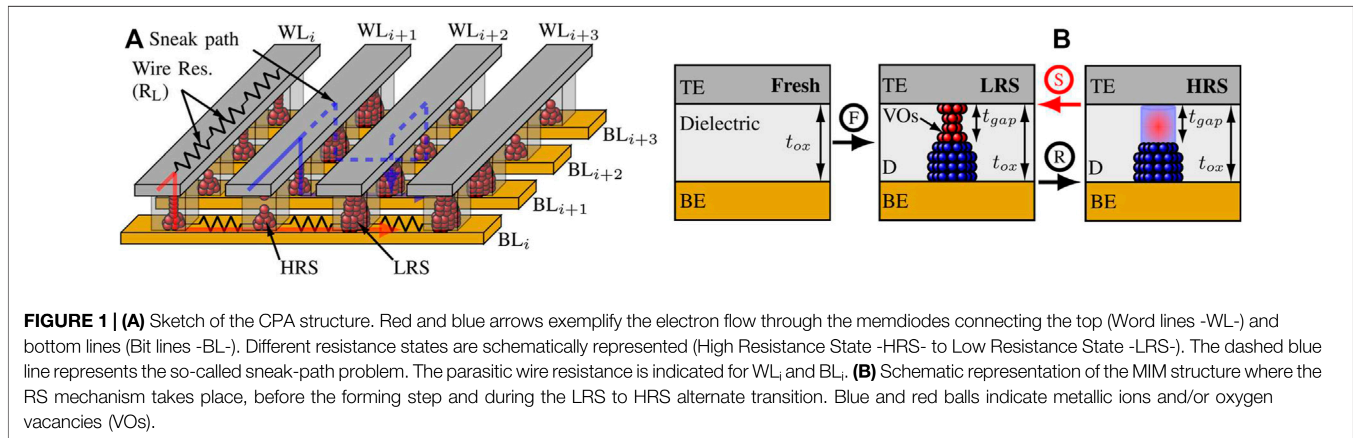

FIGURE 1 | (A) Sketch of the CPA structure. Red and blue arrows exemplify the electron flow through the memdiodes connecting the top (Word lines -WL-) and bottom lines (Bit lines -BL-). Different resistance states are schematically represented (High Resistance State -HRS- to Low Resistance State -LRS-). The dashed blue line represents the so-called sneak-path problem. The parasitic wire resistance is indicated for WL_i and BL_i. **(B)** Schematic representation of the MIM structure where the RS mechanism takes place, before the forming step and during the LRS to HRS alternate transition. Blue and red balls indicate metallic ions and/or oxygen vacancies (VOs).

them, Resistive Random Access Memory (RRAM) or memristor-based Cross-Point Arrays [3–6] (CPAs, see **Figure 1A**) have proven to be able to manage both high speed and low power MVM [7]. Moreover, CPAs can be scaled down to $4F^2$, F being the feature size of the technology node [8], and placed at the Back-End of Line (BEOL), then exploiting the concept of 3D stacking. Memristor-based CPAs for pattern classification have been the focus of previous works [5,9–15] considering not only different architectures but also a variety of device models. Hu et al.[5] reported a simulation-based case study of character recognition using two CPAs with 256 × 26 synapsis (i.e., 256 rows by 26 columns, totaling ~13 k devices) to represent both the positive and negative weights using a Verilog-A nonlinear memristor model [16]. Aiming to reduce both the area and power consumption which arises from having two CPAs, an alternative architecture was considered by Truong et al.[12] (64 × 26, ~1.6 k devices) using the same memristive device model. This model was also successfully tested in voice recognition using a set of CPAs summing up to ~2.5 k memristors [13].

Regardless of the training approach (on-line or off-line, equivalently *in-situ* or *ex-situ*) followed, during the “forward propagation” of a pattern presented to the CPA-based ANN (inference phase) each memristive synapsis located at the CPA intersections drives a current $I_{i,j}$ proportional to the synaptic strength (weight) stored as a conductance value $g_{i,j}$ ($I_{i,j} = g_{i,j} * V_i$). Thereby, setting each memristor in the CPA to the correct conductance value is a crucial and non-trivial task to accomplish. The conductance of the RRAM devices can be changed by applying voltage or current pulses following one of the following approaches: the Write-Verify (or Close Loop Tunning, CLT) [17–19] or the Write-without-Verify (or Open Loop Tunning, OLT) [20–23]. When the CPA’s conductance values require a frequent update, the OLT method is the most appropriate because it preserves the high-speed operation and keeps the hardware overhead to a minimum, at the cost of incurring in a higher writing error. On the contrary, if better controllability of the conductance values is preferred over high-speed operation or if a frequent conductance update is not a major requirement, CLT has been pointed out as the best option. However, in a realistic environment, CPAs have practical limitations such as the line resistances between adjacent cells (R_L), the resistance window of the devices (R_{ON} and R_{OFF}), the sensitivity to the amplitude and frequency of the write pulses, and the device-to-device (D2D) variability as well as the inherent conducting features of the CPAs like the so-called sneak path problem (see **Figure 1A**).

Accordingly, SPICE simulation (or any alternative circuital language) appears as the most suitable approach to realistically investigate and scrutinize the complete system (CPA and control electronics with parasitics) [5,9–13,24]. However, this kind of approach is also constrained to the limitations of the memristor model considered as well as to the size of the CPA mainly because of the high computational requirements [25]. In this regard, great attention has been put in the last years on achieving a simulation tool capable of modelling the wide spectrum of memristive devices found in the literature [26]. This resulted in a variety of models, including simple behavioral models [16,27], device-specific physical-phenomenological models [28] and general phenomenological models (Yakopcic [29], TEAM [30], VTEAM [31], and Eshraghian [32]). Nevertheless, all these models often rely on complex internal equations or on the introduction of artificial window functions that occasionally cause convergence problems [33]. A promising memristor compact model providing high simulation accuracy at reduced computational cost was presented by Miranda et al. in Refs.[34,35]. Its closed-form expression for the current-voltage (I–V) curve (continuous and differentiable) and the recursive nature of the state variable computation based on the Krasnosel’skii-Pokrovskii hysteresis operator [34], makes it suitable for dealing with arbitrary input signals (continuous and discontinuous, differentiable and non-differentiable). Its applicability to the realistic circuital modelling of CPA-based Single (SLP) and Multi Layer Perceptrons (MLP) involving thousands of devices intended for the classification of large pattern datasets was recently demonstrated [14,15]. Although a much simpler approach than the more complex RRAM-based ANNs explored in the literature (MLP [15,19,36,37], Convolutional Neural Networks [38] -CNN-, Spike Neural Networks[39] -SNN-, etc.), the SLP still allows studying and clarifying the ANN limitations caused by parasitic effects and non-idealities occurring in the synaptic layers implemented with

CPAs. Nonetheless, the referred model is in essence a Quasi-Static Model (QMM) because the memory state of the device does not change unless a threshold condition is met. This is a serious limitation for the realistic programming of the CPA.

In this paper, we demonstrate that a novel approach, compatible with the previous one but fully time-dependent, can be used for the SPICE simulation of large-scale memristor-based CPAs intended for pattern recognition tasks without significantly increasing the associated computational cost. The *I-V* characteristic is modelled using an extension of the Quantum Point-Contact (QPC) model for dielectric breakdown [40]. This alternative proposal eliminates the so-called hysteron structure for the memory state as presented in previous works [34,35]. Instead, the DMM generates the hysteretic memory map using a balance differential equation related to the reversible ionic movement. By considering *ex-situ* training of a SLP as a case study and the classification of the handwritten images from the MNIST dataset, we investigate in detail the programming method of the synaptic weights in the CPA (in terms of time and precision). This is a major difference with the works mentioned above, in which the focus was exclusively placed on the results of the inference phase. Here, we pay special attention to the inference accuracy as a function of i) the R_{ON}/R_{OFF} ratio, ii) R_L, iii) the mapping strategy, iv) the write voltage and v) the frequency of the input signal, and last, vi) the particular features of the memristor *I-V* loop, especially the abruptness of the SET transition. To the best of the authors' knowledge, such a detailed and comprehensive study within a unified framework and considering a realistic memristor model has not been published before. The rest of this work is organized as follows: Section *Dynamic Memdiode Model* describes the fundamentals of the DMM: the *I-V* characteristic and the memory equation. Section *Materials and Methods* explains the SLP training, the synaptic weights transfer to the CPA-based implementation, and the electrical simulation procedures. Section *Simulation Results* discusses the obtained simulation results in terms of the aforementioned features, providing useful design considerations and trade-offs. Finally, in Section *Conclusion*, the general conclusions of this paper are presented.

DYNAMIC MEMDIODE MODEL

RRAM devices are based on the Resistive Switching (RS) mechanism, which in the case of CBRAMs and OxRAMs relies on the displacement of metal ions and oxygen vacancies, respectively. The process takes place inside the dielectric film of a metal-insulator-metal (MIM) structure and it is caused by the application of an external electrical stimulus [41–44]. The alternate application of opposite fields originates the completion and destruction (see **Figure 1B**) of a conductive filament (CF) spanning across the insulator, with the consequent change of resistance [45,46] between a high (HRS) and a low (LRS) resistance state. From the electrical viewpoint, the behavior of a memristive device has some major fingerprints [47,48]: i) a pinched hysteresis loop in the *I-V* characteristic, ii) the decrease of the hysteresis lobe area as the input signal frequency increases, and iii) the shrinking of the pinched hysteresis loop to a single-valued function when the frequency tends to infinite. It is also quite common to observe iv) a different *I-V* relationship for HRS and LRS, being linear-exponential in the first case and linear in the second case [46,49], and v) the existence of intermediate conducting states between these two extremes cases (HRS and LRS), which are determined by voltage-controlled redox-reactions. These fingerprints indicate that both the amplitude and frequency effects are interrelated, i.e., the ion/vacancy displacement rates within the insulating layer and the formation and rupture of the CF [50–52]. A model capable of satisfying these constraints was recently proposed by Miranda [53], considering an extension of the conventional memristive approach suggested by Prof. Chua. This proposal again involves a system of two coupled equations, one for the electron transport (transport equation, TE) and a second equation for the memory state of the device (memory equation, ME). For the sake of completeness, the Dynamic Memdiode Model (DDM) is succinctly reviewed in this Section.

Transport Equation

Filamentary conduction models for RRAMs rely on the idea that the current magnitude is somehow related to the presence of a potential barrier or alternatively to a gap along the CF. For example, the QPC model [40,54–56] considers that the current flow through a single nanosized filamentary structure is controlled by the narrowest section along the filament. In terms of the physics of mesoscopic devices, this section has associated a transmission probability like any other barrier or scatterer. Following this idea, Miranda [34] and Patterson et al. [35] proposed a TE based on the similarity of the QPC model expression for the *I-V* curve with two identical opposite-biased diodes in series with a resistor, as shown in the inset of **Figure 2A**. Such a TE allows a progressive transition between an exponential (HRS) and a linear (LRS) curve by simply changing a single parameter of the model $0<\lambda < 1$ called the memory state of the device [34].

According to the QPC model, the equation for the *I-V* characteristic of a memdiode *M* is given by **Eq. 1**:

$$I = I_0(\lambda)\left[e^{\beta\alpha(\lambda)(V-IR_S(\lambda))} - e^{-(1-\beta)\alpha(\lambda)(V-IR_S(\lambda))}\right] \quad (1)$$

where $I_0(\lambda) = I_{min}(1-\lambda) + \lambda I_{max}$ is the current amplitude factor, α a parameter related to the particular features of the conduction mechanism, and R_S a series resistance. **Eq. 1** is the result of considering an inverted parabolic potential barrier with height Φ as the tunneling scatterer. Both α and R_S can have the same dependence on λ as I_0. I_{min} and I_{max} are the minimum and maximum values of the current amplitude, respectively, and $0<\beta < 1$ expresses the asymmetry in the potential at the two ends of the CF (in this work $\beta = 1/2$ is assumed). For a monomode ballistic conductor, $I_0 = G_0/\alpha \cdot exp(-\alpha\Phi/e)$ where $G_0 = 2e^2/h$ is the quantum conductance unit (*e* the electron charge and *h* the Planck's constant). As I_0 increases in **Eq. 1** because of the reduction of Φ (widening of the cross-section area of the CF), the *I-V* curve changes its shape from exponential (HRS) to linear (LRS) as experimentally observed for this kind of devices. This

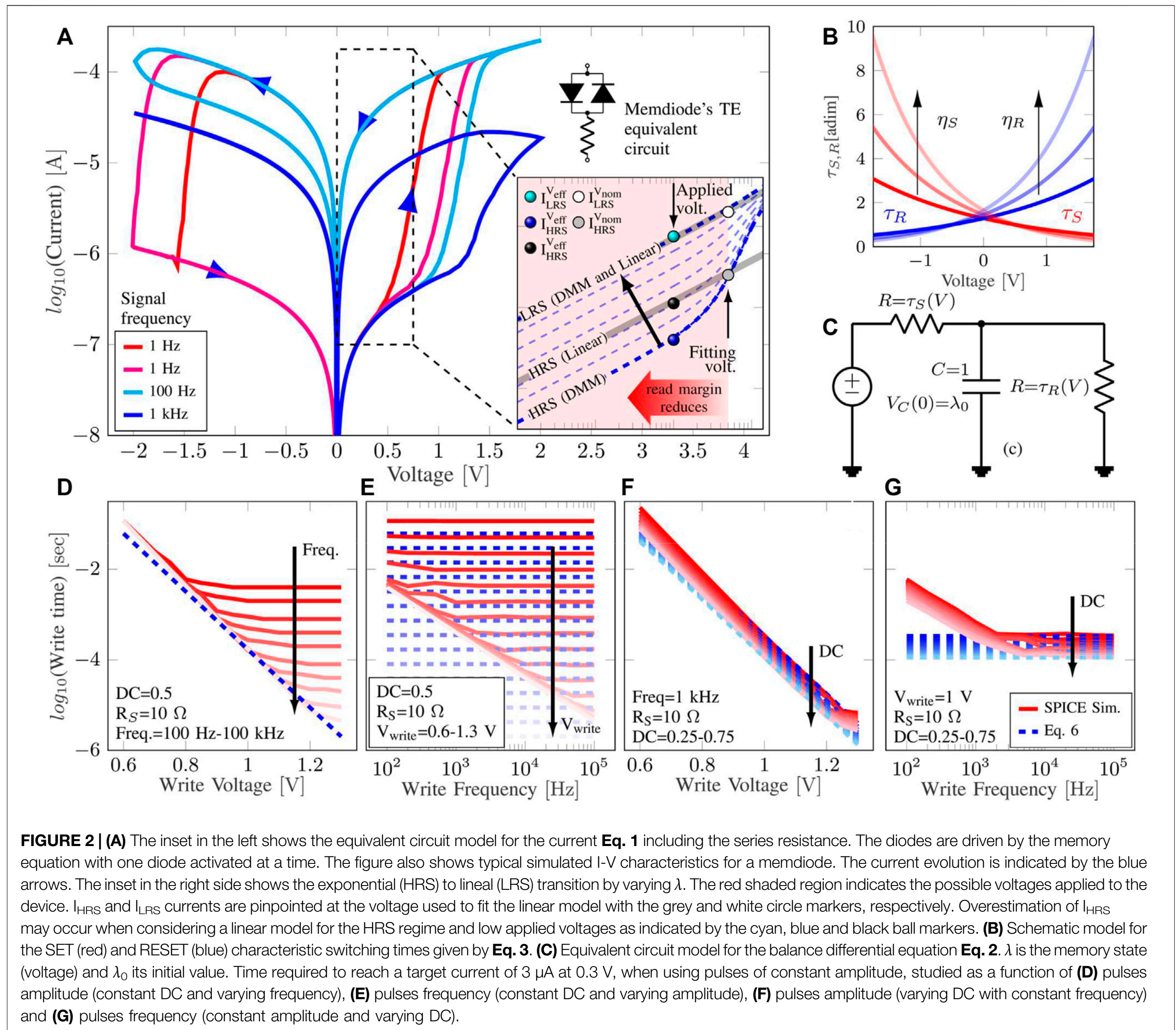

FIGURE 2 | (A) The inset in the left shows the equivalent circuit model for the current **Eq. 1** including the series resistance. The diodes are driven by the memory equation with one diode activated at a time. The figure also shows typical simulated I-V characteristics for a memdiode. The current evolution is indicated by the blue arrows. The inset in the right side shows the exponential (HRS) to lineal (LRS) transition by varying λ. The red shaded region indicates the possible voltages applied to the device. I_{HRS} and I_{LRS} currents are pinpointed at the voltage used to fit the linear model with the grey and white circle markers, respectively. Overestimation of I_{HRS} may occur when considering a linear model for the HRS regime and low applied voltages as indicated by the cyan, blue and black ball markers. **(B)** Schematic model for the SET (red) and RESET (blue) characteristic switching times given by **Eq. 3**. **(C)** Equivalent circuit model for the balance differential equation **Eq. 2**. λ is the memory state (voltage) and λ_0 its initial value. Time required to reach a target current of 3 μA at 0.3 V, when using pulses of constant amplitude, studied as a function of **(D)** pulses amplitude (constant DC and varying frequency), **(E)** pulses frequency (constant DC and varying amplitude), **(F)** pulses amplitude (varying DC with constant frequency) and **(G)** pulses frequency (constant amplitude and varying DC).

occurs because of the voltage drop across R_S and the reduction of the effective applied voltage to the constriction. Notice that **Eq. 1** is equivalent to two opposite-biased ideal diodes (adding a + 1 and a − 1 in the right-hand side of the equation) for which the model was named memdiode. Throughout this work, "memristor" and "memdiode" are used interchangeably to refer to the same device.

Memory Equation

From the modelling viewpoint, the ME is normally a voltage or a current driven differential equation [47] that usually relies on some non-linear dependencies as well as on the so-called window function. Although the introduction of a window function in the ME acting as a boundary condition for λ [32] represented a breakthrough in the modelling of the SET/RESET transitions, the approach was demonstrated not to be exempted from serious mathematical drawbacks [33,57]. The equation discussed here complies with a number of requirements such as dimensional homogeneity, dynamic balance and levelling-off behavior for large opposite biases, with end values 0 and 1 for HRS and LRS, respectively. Physically, the ME expresses the ion/vacancy movement caused by the external applied field. Among the possible candidates, the simplest first order differential equation is given by **Eq. 2**, where $\tau_{S,R}$ are characteristic times linked to the SET and RESET transients (for $V > 0$ and $V < 0$, respectively).

$$\frac{d\lambda}{dt} = \frac{1-\lambda}{\tau_S(V)} - \frac{\lambda}{\tau_R(V)} \tag{2}$$

$$\tau_{S,R}(V) = \tau_{0S,0R}\, e^{-\frac{V}{V_{0S,0R}}} \tag{3}$$

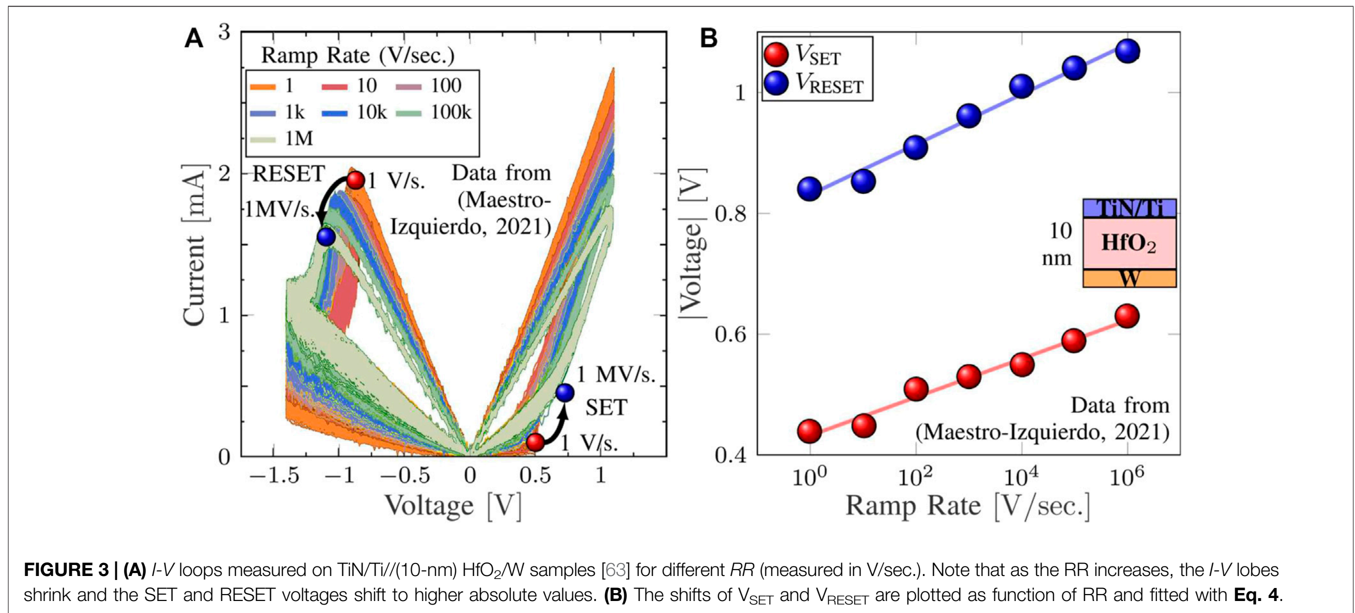

FIGURE 3 | (A) *I-V* loops measured on TiN/Ti//(10-nm) HfO_2/W samples [63] for different *RR* (measured in V/sec.). Note that as the RR increases, the *I-V* lobes shrink and the SET and RESET voltages shift to higher absolute values. **(B)** The shifts of V_{SET} and V_{RESET} are plotted as function of RR and fitted with **Eq. 4**.

In this work, we assume an exponential dependence of $\tau_{S,R}$ on the applied voltage, as expressed by **Eq. 3** and **Figure 2B**. $V_{0S,0R}$ and $\tau_{0S,0R}$ are fitting constants ($V_{0S} > 0$ for SET and $V_{0R} < 0$ for RESET). Although **Eq. 2** has already been used in connection with physical parameters of the CF [58–61], it was first employed as a behavioral memory equation in [62]. A central feature of **Eq. 2** is that it can be represented as an RC circuit with voltage-dependent resistors as illustrated in **Figure 2C**. In this representation, λ corresponds to the voltage drop across the capacitor C_H, whose initial value is λ_0. This representation substitutes the voltage-controlled memory subcircuit considered in many previous memristor models [33].

Switching Dynamics

For a periodic ramped input signal, the combination of **Eqs 1**-**3** results in *I-V* loops such as those illustrated in **Figure 2A**, which starts with HRS (λ = 0) and evolves as indicated by the blue arrows printed on top. The SET and RESET processes triggered at the transition voltages V_{TS} and V_{TR} point out the switching from HRS to LRS and *vice versa*, respectively. The ratio between the LRS and HRS currents measured at a fixed low bias is referred to as the resistive window of the device and when properly normalized to unity represents the memory state window. Note that as the signal frequency increases, V_{TS} and V_{TR} shift toward higher absolute values, as indicated by **Eq. 4** [63]:

$$V_{TS} = V_{0S}\ln(RR) + V_{0S}\ln\left(\frac{\tau_{0S}}{V_{0S}}\right) \tag{4}$$

where *RR* is the input signal's ramp rate [*V(t)* = *RR·t*]. A similar expression holds for V_{TR}. These SET and RESET voltage shifts lead in turn to the mutual approximation of the LRS and HRS curves and thus to the reduction of the hysteresis lobe area. This is in agreement with previously reported experimental measurements of the *I-V* loops as function of frequency, as shown in **Figure 3A**. The evolution of the SET and RESET voltages is plotted separately in **Figure 3B** and fit with **Eq. 4**. The separation between the lines is a consequence of the series resistance effect [63]. In physical terms, the collapse of the memory window is attributed to the incapability of ions/vacancies to follow the input signal [64,65]. From the electrical viewpoint this effect can be directly related to the shift of the switching thresholds [66–68].

For a given *RR*, the HRS (exponential) to LRS (linear) transition as λ is swept from 0 to 1 is detailed in the inset of **Figure 2A** (solid blue lines), altogether with some intermediate states (dashed blue lines). For comparison purposes, a linear model [41,69] (faded-thick grey lines) is also plotted. *sinh*()-based models are omitted here as they require the simultaneous fit of multiple parameters to mimic the smooth linear-exponential to linear transition or even separate expressions for the HRS and LRS regimes [70,71]. Although the two models coincide at low voltages and exhibit a clear linear dependence, significant discrepancies arise as the voltage increases. As can be seen, the linear model cannot capture the departure of the HRS curves at intermediate voltages. As such, if used to fit I_{LRS} and I_{HRS} (the currents in LRS and HRS, respectively) at a nominal read voltage (V_{read}), the linear model leads to an overestimation of the device current when lower voltages are applied. On the contrary, the DMM can accurately describe both the HRS and LRS curves by solely changing a single parameter (λ) in the TE.

Concerning the switching dynamics for discontinuous signals, **Figures 2D–G** illustrate the effects of a sequence of equal amplitude voltage pulses (V_{write} = 0.6–1.3 V) and period (indicated as a combination of frequency ranging from 100 to 100 kHz and Duty Cycle—0.2 to 0.8 –) on the time required to reach a given current level. The parameter values for the ME are $V_{0S,0R} = 6.77 \times 10^{-2}$ V and $\tau_{0S,0R} = 8.48 \times 10^{3}$ sec. Of course, these values depend on the switching material under consideration. Such input stimuli replicate those employed in OLT or CLT approaches as proposed in previous works [23]. For

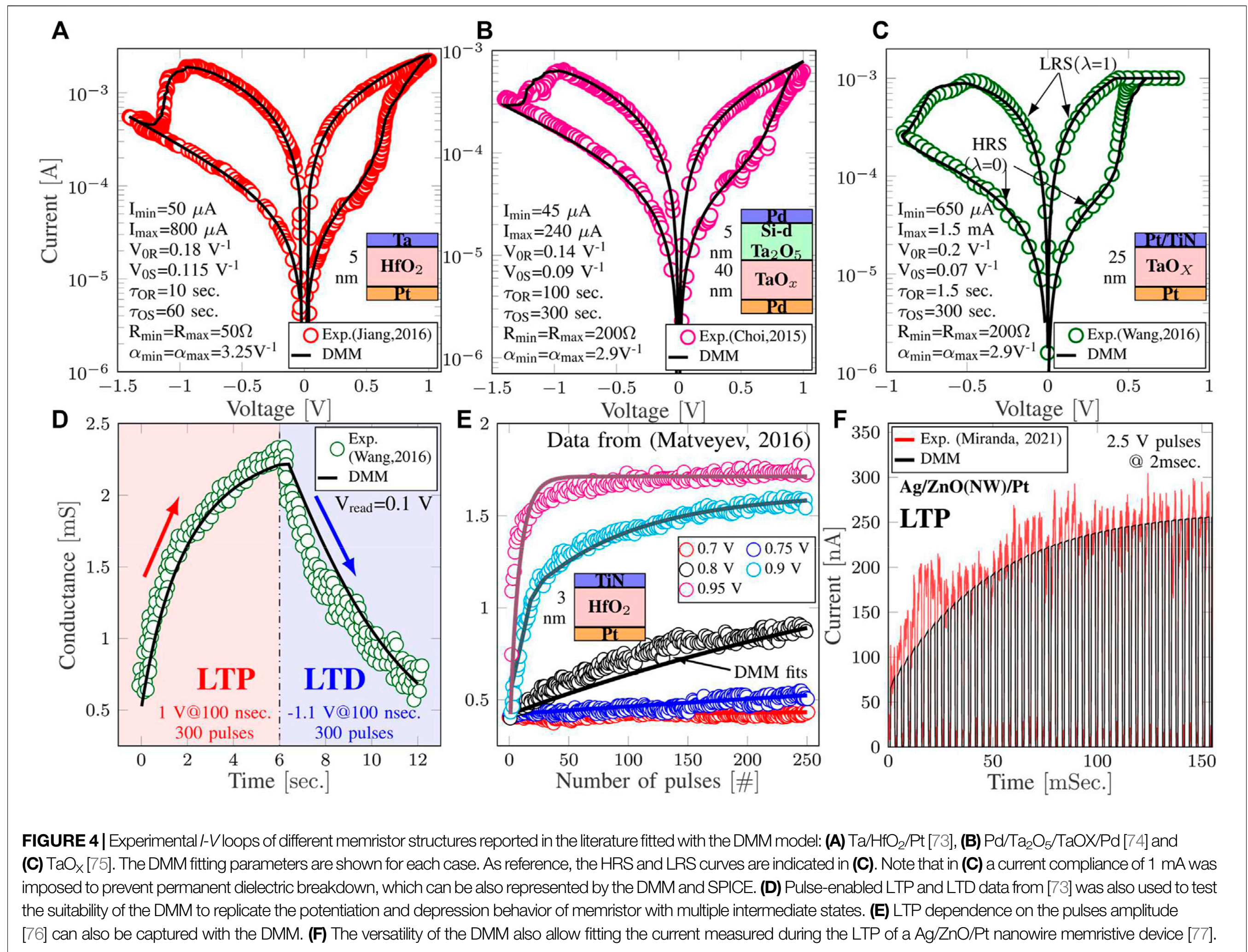

FIGURE 4 | Experimental *I-V* loops of different memristor structures reported in the literature fitted with the DMM model: **(A)** Ta/HfO_2/Pt [73], **(B)** Pd/Ta_2O_5/TaOX/Pd [74] and **(C)** TaO_X [75]. The DMM fitting parameters are shown for each case. As reference, the HRS and LRS curves are indicated in **(C)**. Note that in **(C)** a current compliance of 1 mA was imposed to prevent permanent dielectric breakdown, which can be also represented by the DMM and SPICE. **(D)** Pulse-enabled LTP and LTD data from [73] was also used to test the suitability of the DMM to replicate the potentiation and depression behavior of memristor with multiple intermediate states. **(E)** LTP dependence on the pulses amplitude [76] can also be captured with the DMM. **(F)** The versatility of the DMM also allow fitting the current measured during the LTP of a Ag/ZnO/Pt nanowire memristive device [77].

each pulse, the current increases as a function of the voltage and time due to the so-called potentiation effect in memristors [72]. Knowing the memory state value increase as a function of the number of pulses N with amplitude V applied during a time Δt,

$$\lambda_N = (\lambda_0 - 1)e^{-(\mathrm{N}\Delta t/\tau_S)} + 1 \tag{5}$$

the time required to reach a target current level can be estimated from **Eq. 6** as

$$T_{write} = -\frac{\tau_{0S}e^{-\frac{V}{V_{0S}}}}{DC}\ln\left(\frac{I_{LRS} - I_T}{I_{LRS} - I_{HRS}}\right) \tag{6}$$

where I_{LRS} e I_{HRS} are the maximum and minimum currents measured at the pulse's amplitude V, respectively. DC is the duty cycle. It should be noted that **Eq. 6** is only valid in the linear region of the HRS regime, i.e. for a voltage V lower than V_{TS}. Also, it is assumed that the hysteresis lobe does not collapse in the region of interest. Otherwise the formula is not valid unless the variation of I_{HRS} and I_{LRS} are considered. In this region, the write time increases exponentially as the pulses' amplitude decreases, but remains insensitive to their frequency as shown in **Figures 2D,E**. In addition, notice that as V_{TS} shifts to higher values due to the increase of the input signal frequency (*RR*), so it does the range in which **Eq. 6** holds valid: it goes from 0-~0.8 to 0-~1.2 V when the frequency increases from 100 to 100 kHz (see **Figure 2D**). Similarly, when sweeping the frequency of the pulses, **Eq. 6** holds valid for frequencies above 100 Hz at V_{write} = 0.8 V (the amplitude of the pulses is below V_{TS} for any frequency in the range considered) and for frequencies above 100 kHz at V_{write} = 1.2 V. Rather than the signal frequency, the write time is slightly sensitive to the pulse's width, which is expressed in **Eq. 6**. as a dependence on the signals' DC as shown in **Figures 2F,G**. As it can be seen, when increasing DC, the write time down-shifts as each signal period causes a larger conductance variation.

Experimental Validation of the Dynamic Memdiode Model

The model discussed in the previous Sub-Sections was put under test by fitting experimental data extracted from different

TABLE 1 | Dynamic memdiode model SPICE code.

```
.subckt memdiode p n
  .param H0 = 0 beta = 0.5
  *Transition parameters
  .param T0s = 8.5e + 3 V0s = 6.8e−2 T0r = 1e + 4 V0r = 1e−1
  *I-V parameters
  .param imin = 5e−7 imax = 9.5e−5 alphamin = 1e + 0 alphamax = 1 rsmin = 3.8e + 1
  + rsmax = 3.8e + 1
  *Auxiliary functions
  .param I0(x) = 'imax*x + imin*(1−x)'
  .param A(x) = 'alphamax*x + alphamin*(1−x)'
  .param Rss(x) = 'rsmax*x + rsmin*(1−x)'
  *********H-V*********
  EV A gnd! vol = 1
  RH H A R = 'T0s*exp (-V (p,n)/V0s)'
  RD H gnd! R = 'T0r*exp (V (p,n)/V0r)'
  CH H gnd! C = 1 IC = 'H0'
  *********I-V*********
  RS p D R = 'Rss (V(H))'
  GD D n cur = 'I0(V(H))*(exp (beta*A (V(H))*V (D,n))-exp (-(1-beta)*A (V(H))*V (D,n)))'
.ends memdiode
```

Memdiode script for H-SPICE. p and n are the device terminals. The code is organized in sections: parameter values (sub-divided in transition and I-V parameters), auxiliary functions, memory equation (H-V) and finally I-V characteristic.

published works. In particular, **Figures 4A–C** show the results obtained for different RRAM structures including HfO_2 [73], Ta_2O_5/TaO_X [74] and TaO_X [75] structures, respectively. In all the cases the *I-Vs* were measured at room temperature and under voltage sweeps. The experimental data were fitted using the SPICE model depicted in **Table 1** based on **Eqs 1–3** and applying driving signals as described in the corresponding references. The fitting parameters are listed in each of the panels of **Figure 4** as reference, as well as the details of the stack structure. It should be mentioned that the proposed DMM does not only provide a simple SPICE-compatible implementation for the resistive memory devices but also a versatile one, as it can accurately fit the *I-V* loops experimentally measured in different RRAM devices.

Moreover, we have tested the capability of the DMM to replicate the Long-Term Potentiation (LTP) and Depression (LTD) of memristive synapses used in neuromorphic hardware. This kind of evolutionary behavior is required to achieve gradual conductance changes upon pulse applications. **Figure 4D** shows the LTP and LTD measurements reported in [75] for TiN/(25 nm)TaO_X/Pt -based devices (the experimentally measured *I-V* loop is plotted in **Figure 4C** altogether with the corresponding fit with the DMM) by the application of 300 identical pulses of 1 V (LTP) followed by 300 identical pulses of -1.1 V (LTD). All pulses have the same width (100 ns, t_{ON}) and they are applied every 20 msec During the time in between pulses, a low voltage (0.1 V) pulse is applied to read the conductance (memory) state of the device. These measurement conditions were replicated in the SPICE simulations shown in **Figure 4C**. The simulated LTP and LTD trends are superimposed to the measurements shown in **Figure 4D**, showing a good fit of the experimental trends. There is also a voltage acceleration of the LTP/LTD trends, as reported for instance for the TiN/(3-nm) HfO_2/Pt -based devices measured in [76], and which can be reproduced by the DMM as shown in **Figure 4E**. Finally, the versatility and capability of the DMM to reproduce the evolutionary behavior of memristive devices is shown (see **Figure 4F**) by also fitting the gradual current increase during LTP of Ag/ZnO/Pt based nanowire memristors [77], by considering the same stimuli (pulses with an amplitude of 2.5 V and a duration of roughly 2 msec.). Thereby, the DMM is suitable to model the response of RRAM devices with a large number of incrementally accessible conductance states.

MATERIALS AND METHODS

The basic procedure originally proposed in Ref. [14] for creating and simulating a SLP used as a case study is extended here. The workflow is summarized in the chart depicted in **Figure 5A**. The tasks can be split into two parts: first, the SLP creation, training and circuit-representation using SPICE code generation (MATLAB), and second, the simulation (HSPICE).

Cross-Point Array Based Single Layer Perceptron Creation

First, a software-based SLP of size $n^2 \times m$ (therefore with $n^2 \cdot m$ synapses) is created and trained using a given dataset formed by $n \times n$ px. images (as those represented in **Figure 5B**) distributed into m classes. For the sake of simplicity, *ex-situ* supervised learning is assumed here with the Scaled Conjugate Gradient [78] (SCG) training algorithm. This algorithm provides a good trade-off between accuracy and learning time for the different datasets considered. The possible impact of the learning method was discussed elsewhere [14] where it was demonstrated that no significant statistical differences in terms of the inference accuracy occur. The training procedure generates a $n^2 \times m$ weight matrix $W_M \in \mathbb{R}$, which can be further decomposed as $W_M =$

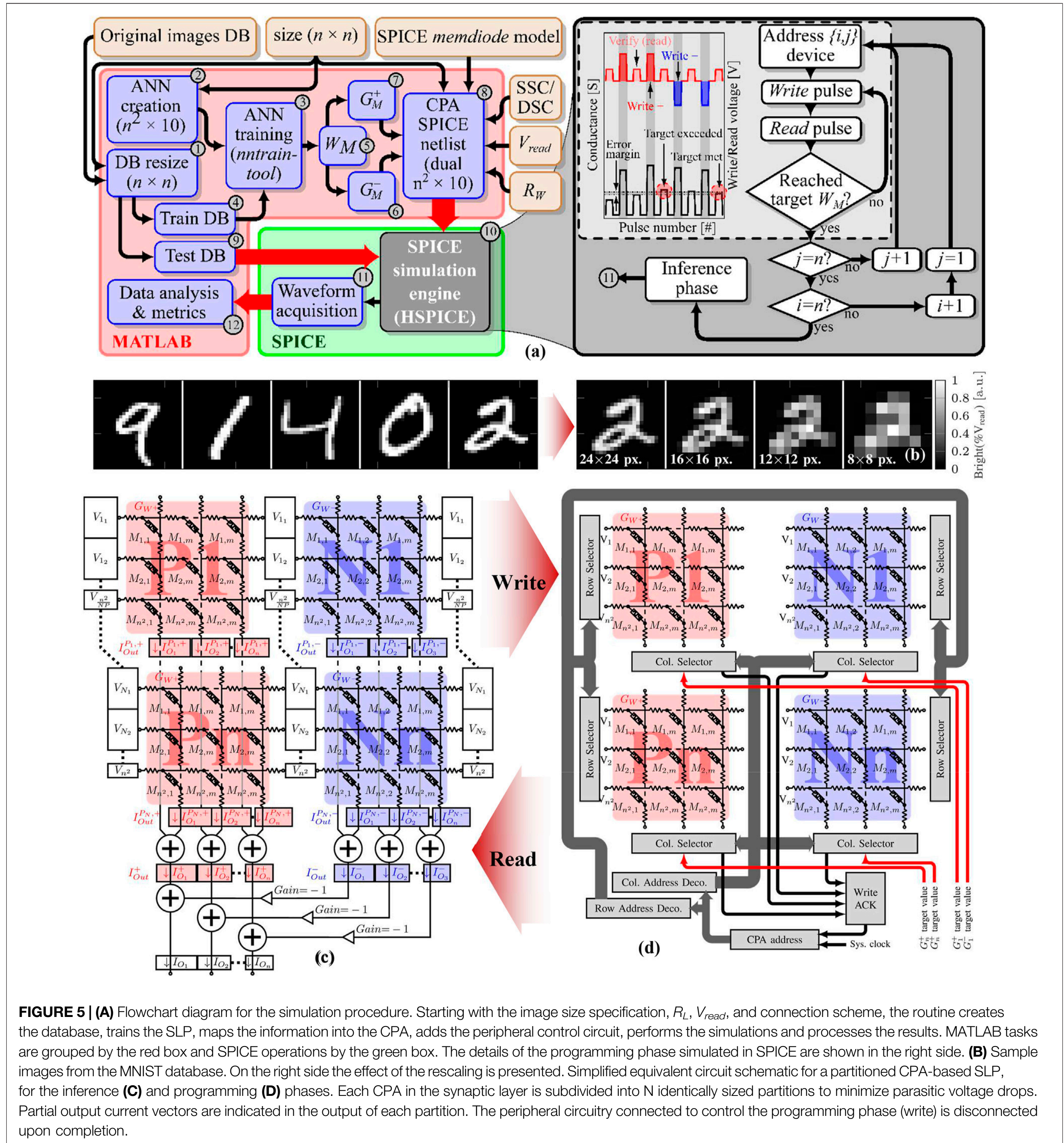

FIGURE 5 | (A) Flowchart diagram for the simulation procedure. Starting with the image size specification, R_L, V_{read}, and connection scheme, the routine creates the database, trains the SLP, maps the information into the CPA, adds the peripheral control circuit, performs the simulations and processes the results. MATLAB tasks are grouped by the red box and SPICE operations by the green box. The details of the programming phase simulated in SPICE are shown in the right side. **(B)** Sample images from the MNIST database. On the right side the effect of the rescaling is presented. Simplified equivalent circuit schematic for a partitioned CPA-based SLP, for the inference **(C)** and programming **(D)** phases. Each CPA in the synaptic layer is subdivided into N identically sized partitions to minimize parasitic voltage drops. Partial output current vectors are indicated in the output of each partition. The peripheral circuitry connected to control the programming phase (write) is disconnected upon completion.

$W_M^+ - W_M^-$ based on **Eq. 7** and **Eq. 8**, as proposed in the literature [79]:

$$w^+_{M_{i,j}} = \begin{cases} w_{M_{i,j}}, & w_{M_{i,j}} > 0 \\ 0, & w_{M_{i,j}} \le 0 \end{cases} \tag{7}$$

$$w^-_{M_{i,j}} = \begin{cases} 0, & w_{M_{i,j}} > 0 \\ -w_{M_{i,j}}, & w_{M_{i,j}} \le 0 \end{cases} \tag{8}$$

with both W_M^+ and W_M^- comprising only positive elements. This allows rendering both positive and negative synaptic weights in W_M as well as doubling the dynamic range and reducing the noise and variability susceptibility [80].

In the next step, the conductance matrices G_M^+ and G_M^- (each sized $n^2 \times m$, thus totaling $2{\cdot}n^2{\cdot}m$ synapses in the SLP) to be mapped into the CPAs are calculated by the linear transformation [22,81]:

$$G_M^{+(-)} = (G_{max} - G_{min})W_{M_{Norm}}^{+(-)} + G_{min} \quad (9)$$

where $[G_{min}, G_{max}]$ is a selected conductance range for a linear computation in matrix-vector calculations. For simplicity, we consider $G_{max} = G_{LRS} = 1/R_{ON}$ and $G_{min} = G_{HRS} = 1/R_{OFF}$, with R_{ON} and R_{OFF} defined by the *I-V* characteristic of the memdiode model play evaluated at a given V_{read}. Then $W_{M_{Norm}}^{+(-)}$ is the positive (negative) weight matrix normalized within the range [0,1] by one of the following normalization methods NM_1-NM_2, indicated by **Eq. 10** and **Eq. 11**:

$$MN_1: \quad W_{M_{Norm}}^{+(-)} = \frac{W_M^{+(-)}}{max\{abs\{W_M\}\}} \quad (10)$$

$$MN_2: \quad w_{M_{i,j}} = \begin{cases} 1 & , \; w_{M_{i,j}} > \mu_{W_M} + n\sigma_{W_M} \\ \dfrac{w_{M_{i,j}}}{\mu_{W_M} - n\sigma_{W_M}}, & \mu_{W_M} - n\sigma_{W_M} < w_{M_{i,j}} < 0 \\ \dfrac{w_{M_{i,j}}}{\mu_{W_M} + n\sigma_{W_M}}, & 0 < w_{M_{i,j}} < \mu_{W_M} + n\sigma_{W_M} \\ -1 & , \; w_{M_{i,j}} < \mu_{W_M} - n\sigma_{W_M} \end{cases} \quad (11)$$

$max\{W_M\}$ and $min\{W_M\}$ are the maximum and minimum synaptic weight values in W_M and μ_{W_M}, σ_{W_M} are the mean and standard deviation of the synaptic weights in W_M, respectively. In this way, the W_M^+ and W_M^- matrices are converted to conductance values within the range $[G_{HRS}, G_{LRS}]$.

The subsequent sub-routines write down the circuit netlist for the dual-$n^2 \times m$ memdiode CPA-based SLP, adding the parasitic line resistance (R_L), connection scheme, and control logic necessary for both the CLT and inference phase. Two approaches were followed in order to improve the voltage effectively delivered to the RRAM cells as shown in the simplified equivalent circuit depicted in **Figure 5C**: i) A Dual Side Connection (DSC) scheme and ii) the partitioning of the G_M^+ and G_M^- matrices into *NP* (number of partitions) smaller arrays [14,69,80]. Despite the increased peripheral circuitry complexity, DSC improves the voltage delivery to each synapse [13] by connecting both wordline terminals to the input stimuli. Similarly, the small size of each partition helps reducing the parasitic voltage drop along the line interconnections. Exploding the integrability of the CPA with CMOS circuitry, the vertical interconnects linking the outputs of each partition may be placed under the CPAs, as well as the analogue sensing electronics, thus minimizing the area overhead of the partitioned architecture [80].

To account for both the inference and write phases, the present implementation follows a reconfigurable approach, in which the partitioned CPA-based SLP is alternatively connected to the input/output signals (inference phase, **Figure 5C**) or to the writing stimuli (write phase, **Figure 5D**). The analogue electronics required for the inference phase has been described elsewhere in the literature [14,15]. However, the circuitry needed for the write phase requires a more complex circuital implementation, as the input stimuli are not passed simultaneously to all the CPA inputs, but sequentially. Thereby, as shown in **Figure 5D**, this circuital arrangement involves a CPA address block, the Row/Column address decoders, the Row/Column selectors, and the Write Acknowledge block. All of them have been designed assuming a 130 nm CMOS process from GlobalFoundries.

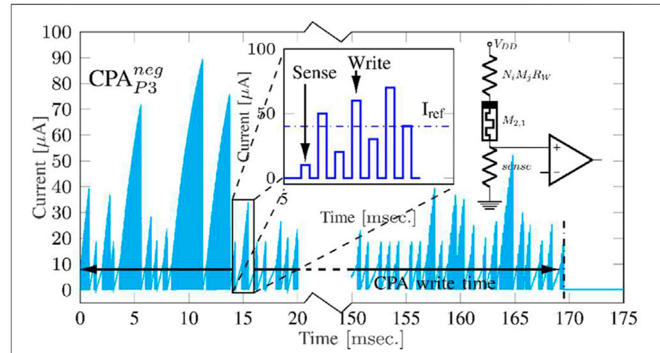

FIGURE 6 | Sensed output current for a SLP partition (one small CPA) during the programming phase following a CLT scheme. The greater the peak, the higher the conductance level being programmed. For this scenario, NM-1 and 8 partitions per CPA were considered. The inset in the centre shows a schematic representation of the current measured during the verify and write pulses as well as the current target. The inset in the right shows a schematic of the equivalent circuit used during the verify phase.

The CPA address block is basically an asynchronous counter that produces n^2/NP *m* output pulses (the number of memdiodes in each partition), which propagate to both the Row and Column decoders. Both decoders consist of another ascending asynchronous counter that counts up to n^2/NP (Row decoder) and m (column decoder) pulses, and outputs a binary code indicating the row and column being addressed, respectively. Then, these binary codes are simultaneously propagated to all the Row/Column selectors. The latter are an array of analogue switches that connect the input node of each CPA Row to V_{write} or V_{read} (for addressing that particular Row during the write procedure), $V_{DD}/2$ (if another row is being addressed) or to V_{input} (when the ANN is operating in the inference phase). The column selector is a similar array that connects the columns output nodes to a sensing amplifier (SA, a transimpedance amplifier coupled to a comparator) if that particular column is being addressed, or $V_{DD}/2$ (if other column is being addressed). Finally, the target current level to be detected by the SA to determine the completion of the write process of a given element $g_{i,j}$ in the G_M^+ and G_M^- matrices is calculated based on the nominal V_{read} and conductance $g_{i,j}$. Once the circuit netlist is written, it is passed to a SPICE simulator which evaluates the voltage and current distributions in the CPA circuit while it is being programmed, and then, as the input images are processed and classified. Finally, the resulting waveforms are passed back to the MATLAB routine for evaluation and metrics extraction.

Write Procedure

During the write operation each memdiode in the CPAs ($M_{i,j}$) is individually addressed and supplied with a train of alternating read and write pulses of amplitude V_{read} and V_{write}, respectively. This causes a gradual increment of the memdiode conductance until a target value is reached. Such addressing procedure is

performed following the $V_{DD}/2$ approach as it minimizes the line disturbance [41]. Within this writing method, the non-addressed rows are set to a constant voltage value $V_{DD}/2$. Similarly, the output node of the column of the addressed memdiode is grounded through the SA, which measures the current flowing out from this column (the other columns are at $V_{DD}/2$). The measured current is proportional to the applied voltage pulses and the memdiode conductance ($g_{i,j}$) plus the parasitic wire resistance corresponding to the addressed device ($M_{i,j}$). This allows to estimate the conductance of the addressed memdiode. This process is represented by the simplified equivalent circuit shown in the inset of **Figure 6**.

The write procedure for the addressed memdiode $M_{i,j}$ begins by sensing the output current during the read pulse of magnitude V_{read}. In case this current is lower than a target value, a write pulse of voltage V_{write} is applied ($V_{write} > V_{read}$), causing an increment of the $M_{i,j}$ conductance ($g_{i,j}$). Then a new read pulse is applied and the current sensed again. This process continues iteratively until the sensed current during the read pulse meets the target value. Once reached, the SA outputs a pulse that indicates the completion of the writing procedure for the addressed memdiode ($M_{i,j}$), stopping the train of read/write pulses and preparing the following devices for the programming step.

It is worth noting that the partitioned architecture allows the simultaneous programming of the $M_{i,j}$ memdiode in all partitions using a simpler control circuit. Let us assume that the devices to be programmed are the $M_{i,j}$ memdiodes of a $2 \times (n^2 \times m)$ RRAM-based SLP implemented with NP partitions (thus totalling 2·NP CPAs), such as the one presented in **Figure 5D**. In this case, the ith output of the row decoder (n^2/NP outputs) will be the only active output, as well as the jth output (m outputs) of the column decoder. Then these output vectors are passed to every Row/Column selector, which therefore simultaneously select the $M_{i,j}$ memdiode in every CPA. This causes all the ith rows to be connected to a train of alternating read and write pulses and all the jth columns to be connected to the partition's SA (each CPA partition has its own SA). All other rows and columns are connected to $V_{DD}/2$. The current flowing through each of the $M_{i,j}$ memdiodes (and therefore out from the jth columns) is sensed by its associated SA until the target conductance value for that $M_{i,j}$ memdiode is achieved. Then the associated SA propagates an acknowledge pulse (ACK) to the Write Acknowledge block. This block waits for the ACK pulses coming from the SAs of every partition. Once all ACK pulses are received, the ith,j^{th} position of all CPAs is considered to be successfully written, and by the time the Write Acknowledge block receives the following system clock pulse, it instructs the CPA address block to address the $M_{i,j+1}$ memdiode and the write sequence starts again. This process continues until the CPA address block has addressed all the memdiode positions in the CPA partitions ($n^2/NP \times m$ positions).

Modified National Institute of Standards and Technology Dataset and Input Stimuli

The MNIST (Modified National Institute of Standards and Technology) dataset of handwritten digits was selected for the training and inference phases considered in this work. This dataset comprises a series of k input feature vectors [$x(k)$ for sample k] and a target output vector [$t(k)$, with 10 dimensions, each corresponding to one digit]. For the classification problem, $t_c(k) = 1$ if sample k belongs to class c, otherwise $t_c(k) = 0$. The input feature vectors ($n^2 \times 1$) are the unrolled grayscale pixel values of thetwo-dimensionalimages ($n \times n$px.). Pixel's brightness is codified in 256 gray levels between 0 (fully OFF, black) and 1 (fully ON, white). The MNIST dataset contains 60,000 training images and 10,000 testing images, both in grayscale and with a 28 × 28 px. resolution [82]. A few examples of these images can be seen in the left panel of **Figure 5B**. The MNIST images were further down sampled using bicubic interpolation to 8 × 8 px. to allow smaller SLPs. This modification reduces the inference accuracy degradation due to the line resistance effect and speeds up the simulations (also reduces the image readability as shown in the right side of **Figure 5B**). Finally, the input stimuli are obtained by scaling the input feature vector by a voltage V_{read}. V_{read} is chosen such as to prevent altering the memdiode state during the inference phase simulation. In this way, during the inference process each of the test images is presented to the CPA as a vector of analog voltages in the range [0, V_{read}].

SIMULATION RESULTS

To quantitatively compare the write performance under different conditions, three different metrics were considered. They are: i) the time required to write the complete CPAs, ii) the inference accuracy achieved after programming the CPAs and, iii) the write error. To quantify the latter, we have considered in this paper the so-called Sum Weight Variation (SWV). Let the weight variation (WV) be equal to $\left|w_{i,j}^{real} - w_{i,j}\right|$ where $w_{i,j}^{real}$ is an element of the matrix $W_{M_{Norm}}^{real}$ with effectively programmed synaptic weights, and $w_{i,j}$ is an element of the matrix $W_{M_{Norm}}$ with ideal synaptic weights. Then, the metric referred to as SWV [83] can be derived to quantify the deviation of $W_{M_{Norm}}^{real}$ matrix from the ideal $W_{M_{Norm}}$ matrix and it is computed as indicated by **Eq. 12**:

$$SWV = \sum_{i=1}^{M} \sum_{j=1}^{N} \left|w_{i,j}^{real} - w_{i,j}\right| \tag{12}$$

where M and N stands for the number of rows and columns in $W_{M_{Norm}}$. From **Eq. 12**, it can be noted that the lower the value of SWV, the lower the error on the mapped weight matrix.

Write Voltage

All three metrics considered in this study show similar trends regardless of the normalization method, R_{ON} and R_{OFF} resistances and line resistance (R_L). When addressing the CPA write time presented in **Figure 7A**, it can be seen that as the voltage of the write pulses decreases, the CPA write time slightly increases until reaching roughly 1.1 V where it increases abruptly. The reason behind this behavior is a consequence of the I-V characteristics of the memristors themselves, in particular the SET transitions. Above the 1.1 V threshold voltage, the reduction of the write pulses amplitude causes the conductance increase produced by

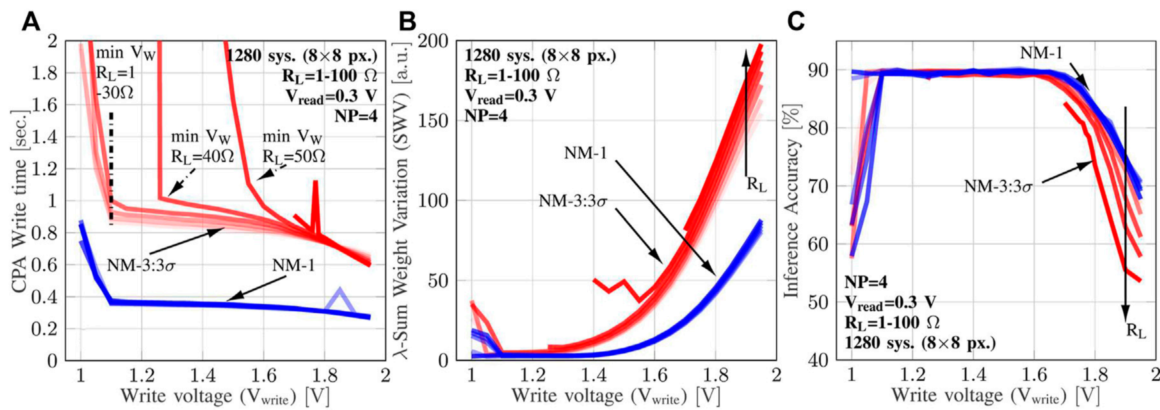

FIGURE 7 | Simulation results for the synaptic weights writing process into the partitioned CPA-based SLP. **(A)** Total Write Time, **(B)** λ-SWV and **(C)** Inference accuracy obtained with the written SLP as a function of the write pulses amplitude. In all cases, the two normalization methods were considered (*NM-1* and *NM-2*) and the line resistance was varied in the range from 1 to 100 Ω.

the application of each pulse to decrease too, thereby requiring more pulses to program each memdiode in the CPA. However, when applying a voltage lower than the 1.1 V threshold, the voltage delivered to a number of cells in the CPA is not enough to alter its conductance due to a reduced read (or write) margin. *RM* (*WM*) is the fraction of the applied input voltage V_{in} (which may be V_{read} or V_{write}) effectively delivered to the memory cells (V_{cell}), *i.e.* V_{cell}/V_{in}.

This interpretation is in close agreement with the results reported in **Figure 7B**, where the SWV metric is represented in terms of the write voltage. Note that as the conductance change produced by each write pulse reduces with the pulse's amplitude, the write procedure has higher control over the programmed conductance, and thereby the write error reduces to a minimum. Nonetheless, when crossing the 1.1 V threshold, a subtle rise in the SWV can be seen, coincident with the abrupt increase of the write time. This could be explained by considering the sneak-path effect. In this context, the programming of certain positions in the CPA is not terminated by the current being driven by the device located in that position, but by parasitic currents loops caused by already set devices located close to the input/output ports. A direct consequence of the write error increase is the degradation of the inference accuracy as shown in **Figure 7C**. Note that both below the 1.1 V threshold and above 1.6 V there is a reduction of the inference accuracy that almost coincides with the increasing writing errors during the memristor programming phase.

Normalization Method

The matrix elements in W_M are in the range $[min\{W_M\}, max\{W_M\}]$. To be mapped to a conductance magnitude in the range $[G_{HRS}, G_{LRS}]$, they must be first normalized to the range $[-1, 1]$. Very often [22,81], such normalization is achieved dividing W_M by the absolute value of the maximum element in W_M (Normalization Method 1, *NM-1*, see **Eq. 10**). As expected, the normalized W_M matrices $W_{M_{N1}}$ preserve the exact same distribution and the $max\{W_M\}/min\{W_M\}$ ratio. Interestingly, for the case of the MNIST images resized to 8 × 8 px. used in this work for benchmarking, ~95% of the elements from $W_{M_{N1}}$ falls within the range [−0.5,0.5]. Thereby, *NM-1* does not exploit the entire dynamic range of the memdiodes as most of the devices will be set to a conductance value in the range $[G_{HRS}, ((G_{LRS} + G_{HRS})/2)]$. To make a better use of the entire memristor conductance range, an alternative approach (*NM-2,* see **Eq. 11**) is considered. In this context, an element $w_{i,j} \in W_M$ has a probability P_i of being within the range $\mu_{W_M} \pm i\sigma_{W_M}$, where μ_{W_M} and σ_{W_M} are the mean and standard deviation of the values of W_M. For i ranging from 1 to 4, ~68.3%, ~95.5%, ~99.7 and ~99.9% of the synaptic weights will be within such range, respectively [84]. Thus, values exceeding such limits are set equal to $\mu_{W_M} \pm i\sigma_{W_M}$ and then W_M is normalized to obtain $W_{M_{N2}}$. This normalization method allows not only a better exploitation of the memristor dynamic range but also minimizes the impact of stuck-at-ON faults, as explained in Ref. [85].

The impact of both *NM-1* and *NM-2* were compared in terms of the proposed metrics in **Figure 7**, with the CPA write time and SWV showing a factor 2 of difference between the two normalization methods. Given that, when considering NM-1, most of the conductance values are mapped close to G_{HRS}, a lower number of write pulses are required to meet the target conductance. This results in shorter writing times, as seen in **Figure 7A**. The lower mean conductance of the synaptic weights resulting from applying NM-1 also results in a lower SWV (See **Figure 7B**). As the conductance of the synapses increases, so does the current flowing from wordlines to bitlines. This not only causes a larger IR-drop in the interconnection lines, but also exacerbates the sneak-path effect too, resulting in a similar scenario to that depicted in section *Simulation Results*. A: the current measured by the sensing amplifier that terminates the writing cycle in a given *{i,j}* position of the CPA is in fact the result of multiple conductive paths. As a consequence, the writing cycle in the *{i,j}* position is terminated before reaching the target conductance, giving rise to a significant increase of the SWV metric. Finally, no significant changes occur in the inference accuracy of the programmed CPA (see **Figure 7C**), except for a

higher sensitivity to the line resistance. This point will be thoroughly discussed in the next section.

Line Resistance (R_L)

Assuming conventional metal wires, the resistance of the interconnect wires (R_L) between two adjacent cells in a $4F^2$ CPA structure can be estimated to be ~4.5, 3, and 1.5 Ω for the 16, 22 and 32 nm technology nodes, respectively [86]. Nevertheless, for the novel technology nodes (10 nm and below) both surface and grain boundary scattering cause a size-dependent resistivity of the Cu wires [87–89] as the mean free path of electrons becomes comparable to the wire dimensions. These two effects are well-known and can be quantified using the Fuchs-Sondheimer (FS) [90] and the Mayadas-Shatzkes (MS) [91] models. They reveal that for highly scaled nodes (~5 nm), R_L can be as large as ~100 kΩ [69]. In this work, we have considered R_L values varying from 1 to 60 Ω.

Interestingly, all the three metrics (CPA Write Time, Inference Accuracy and SWV) show a much higher sensitivity to R_L when considering the *NM-2*. This can be attributed to the CPA's *RM*. *RM* is jointly determined by R_L and the memdiode resistance (R_{memd}, which varies in between R_{OFF} and R_{ON}) [14,41,69]. From a very basic analysis, each memdiode is part of a conductive path between the CPA's input wordline *i* and the output bitline *j*. For a $N \times M$ CPA, the average parasitic resistance associated with such path is $R_L[(N + M)/2 + 1]$ [14,37]. Within this simplified scenario, the V_{cell}/V_{in} ratio can be obtained from the voltage divider between R_{memd} and $R_L[(N + M)/2 + 1]$. This can explain the greater robustness of the *NM-1* against R_L variations: the larger average value of R_{memd} minimizes the IR drop and as a consequence, the impact of R_L. On the other hand, for *NM-2*, the IR drop across R_L significantly limits the minimum write voltage, as for increasing values of R_L the voltage effectively delivered to the CPA cells reduces causing an exponential increase in the CPA write time. This effect also worsens the impact of the sneak-path problem, resulting in larger SWV values and lower inference accuracy (see **Figures 7B,C**, respectively).

Frequency and SET Transition Abruptness

As mentioned in Section *Dynamic Memdiode Model*, the *I-V* characteristic of the memdiode is sensible to the frequency of the input signal [63]. Thereby studying the impact of the write pulses frequency on the CPA write time, SWV and inference accuracy is necessary to complement the trends reported in **Figure 7**. In this regard, the aforementioned metrics were evaluated for different combinations of write pulses voltages and frequencies and for four different *I-V* loops, in which the abruptness of the SET transition is progressively increased. For the sake of simplicity, R_L is kept constant and equal to 10 Ω, as well as the amplitude of the verify pulses (0.3 V) and the read frequency and amplitude (100 kHz and 0.3 V).

In general, the simulation results presented in **Figure 8** indicate that the best performance metrics are obtained for low voltage write pulses of high frequency (upper right corners of the SWV and accuracy metrics). This is a direct consequence of the smaller conductance change induced by each pulse when reducing their amplitude and width. On the contrary, in all the cases, the lower left corner is associated with the highest SWV and consequently with the lowest accuracy, given the lesser controllability of the conductance increase (pulses of high amplitude are applied during longer periods). Nevertheless, such a better controllability comes at the cost of a much higher writing time given that a larger number of pulses is required. This was already seen in the results reported in **Figure 7**.

Beyond the overall trends, the abruptness of the SET transition (see **Figure 8A**) also has a remarkable impact on the evaluation metrics. Note that as the SET transition becomes more abrupt, the SWV increases and consequently, the inference accuracy reduces. The changes can be regarded as an up-shift in the heat-maps presented in **Figure 8**, as no significant variations can be observed as a function of the frequency. This change in the heat-maps can be interpreted once again as a consequence of different conductance modifications caused by the pulses amplitude for the four different SET transitions considered: for a pulse with a given amplitude and width, a more progressive SET transition results in a smaller conductance increment, whereas the conductance modification produced for the same pulse in a memdiode with a steeper SET transition is larger.

Interestingly, apart from modifying the write performance of the conductance values, the frequency and voltage dependences of the memdiode also play a role during the inference phase, mainly because of the resistance shift occurring in the CPA memristors. Depending on the maximum voltage V_{read} used to translate the input pattern into a vector of analogue voltages, and the frequency at which such vectors are presented to the CPA, a resistance shift for the stored synaptic weights may occur (see Section *Switching Dynamics*). Notice that this is ultimately a consequence of the ionic movement inside the dielectric. This is presented in **Figure 9A** in terms of the SWV computed for the control parameter λ after having presented a series of images (x axis) and for different combinations of V_{read} values and input frequencies. The overall trend is that the SWV metric increases as more images are presented to the CPA and this can be minimized both by increasing the input frequency or reducing the read voltage. Such an increase in SWV causes inevitably a reduction of the inference accuracy as illustrated in the heat-maps in **Figures 9B–E**. Note that four cases with different SET abruptness were considered, and similarly to the results presented in **Figure 8**, the more abrupt the SET transition, the higher the sensitivity to the read voltage and the input frequency. As a corollary, it can be seen that V_{read} could be largely increased to improve, for instance, the Signal-to-Noise ratio as long as the input frequency is increased accordingly.

In this regard, it is crucial to consider the stability of the programmed resistive state of the cells. We have thereby studied in detail the distribution of synaptic weights and conductance matrices over time and reported the results in **Figures 9F–T**. This has been carried out by stopping the simulation and reading the value of the state variable λ for each memristor in the CPA (this value corresponds to the voltage of the node H in the model script). When the simulation is stopped, all the nodal voltages are stored and the λ-SWV computed with respect to the original mapped values. Then, the stored nodal voltages are assigned as

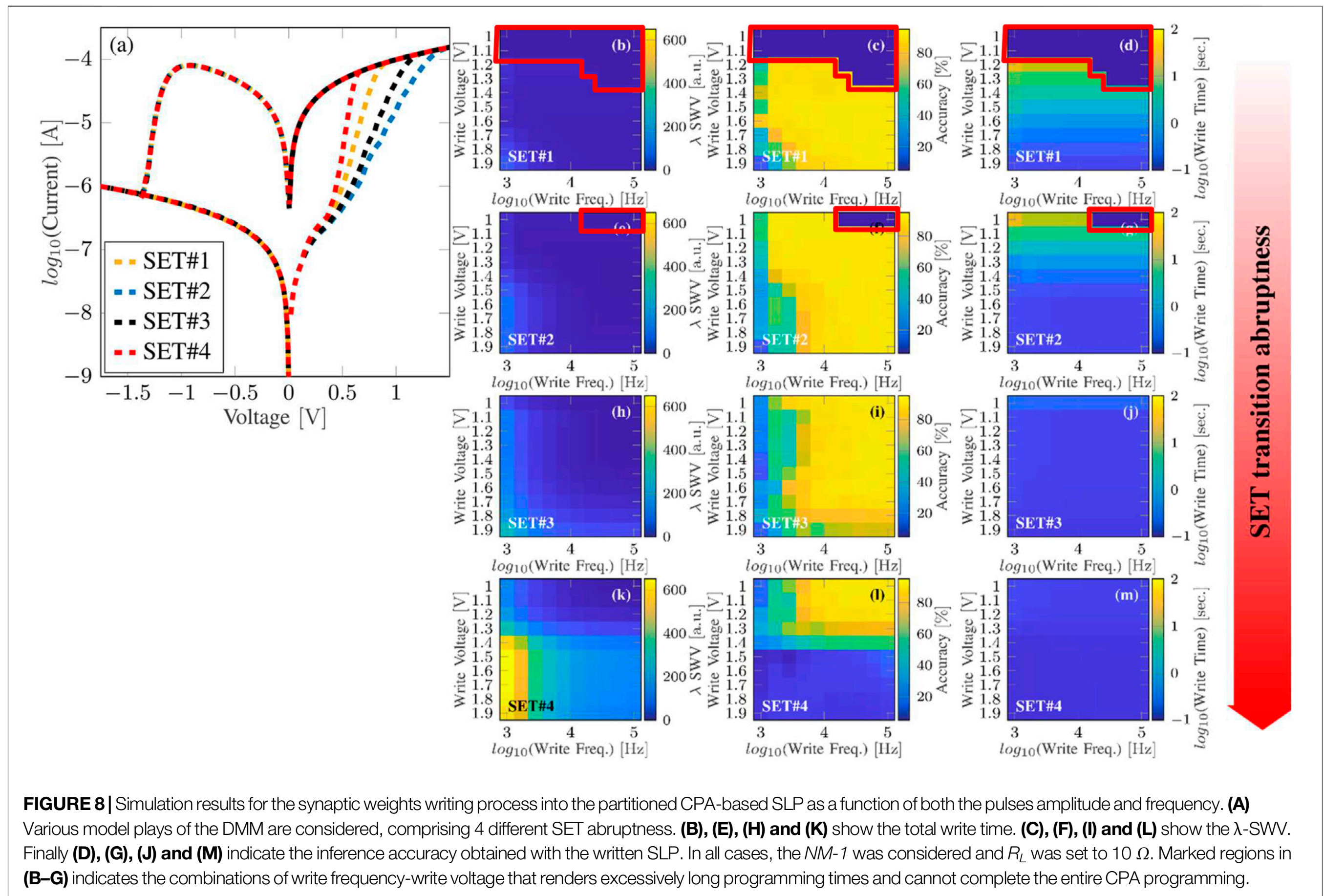

FIGURE 8 | Simulation results for the synaptic weights writing process into the partitioned CPA-based SLP as a function of both the pulses amplitude and frequency. **(A)** Various model plays of the DMM are considered, comprising 4 different SET abruptness. **(B), (E), (H) and (K)** show the total write time. **(C), (F), (I) and (L)** show the λ-SWV. Finally **(D), (G), (J) and (M)** indicate the inference accuracy obtained with the written SLP. In all cases, the *NM-1* was considered and R_L was set to 10 Ω. Marked regions in **(B–G)** indicates the combinations of write frequency-write voltage that renders excessively long programming times and cannot complete the entire CPA programming.

initial conditions to the corresponding nodes and the SPICE simulation is resumed starting from the previous processed image. Note that we have considered accelerated stress conditions, V_{read} and Frequency values of 800 mV and 1 kHz. Under these circumstances the trends can be obtained in reduced times (the simulated time is 10 s for 10,000 images). Normal operation voltage (approx. 200 mV) and frequency (above 1 MHz) should ensure that the weights remain stable over a long period of time, otherwise, re-tuning of the synaptic weight shall be performed periodically. From **Figures 9F–T**, the statistical distribution of the programmed weights shift towards G_{LRS} as the number of images increases (i.e., over time). When considering the changes in the conductance matrices associated with these shifts in the statistical distribution of the synaptic weights, it can be seen that a certain pattern emerges, with the upper and lower region of the CPA arrangement showing no changes. Moreover, equidistant stripes of devices with no variations in middle of highly potentiated devices are observed. Such changes can be explained considering the particular features of the dataset analyzed in this work (MNIST). In the MNIST data-base, pixels located at the edges of the images are normally OFF. When unrolling the $n{\times}n$ images into $n^2 \times 1$ column vectors, the pixels from the left border of the image (normally unactive) occupies the first n positions of the $n^2 \times 1$ column vector, and those on the right border, the last n positions. Then the firsts and lasts n rows of the CPA arrangement are never stressed, so that they are always biased with approx. 0 V. Then, moving forward with the same reasoning, the upper and lower pixels of the middle columns of the $n \times n$ images are also normally unactive. When feeding the CPA arrangement with a $n^2 \times 1$ column vector, this results in equally spaced unbiased rows, which thereby suffers no variations of the weights stored in the corresponding memristors.

Finally, it is worth recalling that even for the case of high stability of the programmed memristor devices, a possible device-to-device (D2D) variability and the consequent programming error need to be considered. The intrinsic stochastic nature of the switching mechanism can induce a high degree of variability [92]. Such variability in the memristive structures also increases the possibility of a deviation of the weighted sum from the target conductance value [93]. The variability of the conductance states G_{LRS} and G_{HRS} across a matrix is largely influenced by the choice of the stack's materials (e.g., single material HfO_X vs. bilayer HfO_X + TaO_X) [94,95], as well as by the device scaling. Extreme scaling seems to reduce variability likely because of a reduction of the area where the switching occurs [96]. Furthermore, the conductance value set during the CLT procedure also exhibits variability [19,93].

The normalized device variability can be expressed as σ/μ, where σ is the standard deviation and μ the mean value of the LRS and HRS

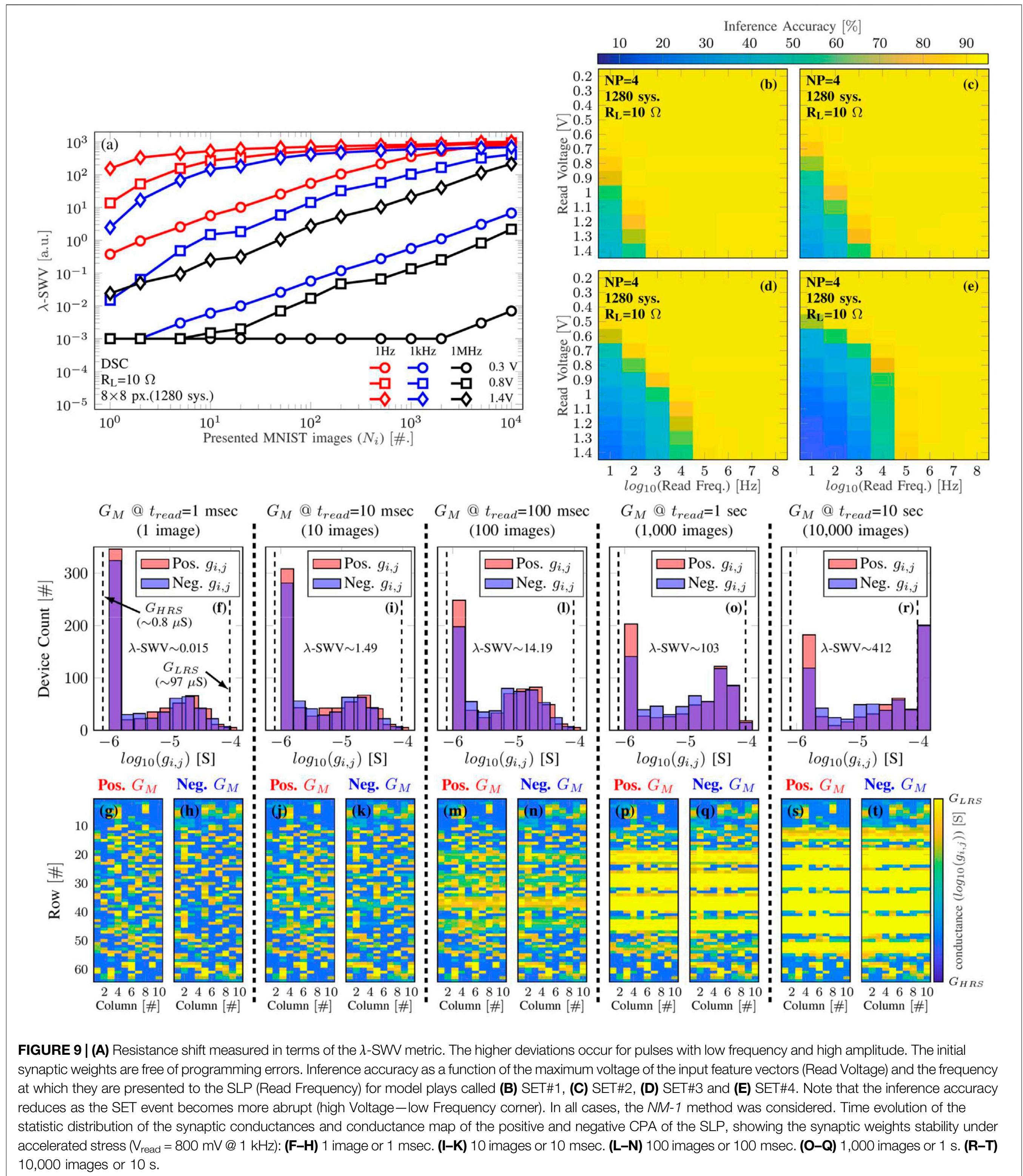

FIGURE 9 | (A) Resistance shift measured in terms of the λ-SWV metric. The higher deviations occur for pulses with low frequency and high amplitude. The initial synaptic weights are free of programming errors. Inference accuracy as a function of the maximum voltage of the input feature vectors (Read Voltage) and the frequency at which they are presented to the SLP (Read Frequency) for model plays called **(B)** SET#1, **(C)** SET#2, **(D)** SET#3 and **(E)** SET#4. Note that the inference accuracy reduces as the SET event becomes more abrupt (high Voltage—low Frequency corner). In all cases, the *NM-1* method was considered. Time evolution of the statistic distribution of the synaptic conductances and conductance map of the positive and negative CPA of the SLP, showing the synaptic weights stability under accelerated stress (V_{read} = 800 mV @ 1 kHz): **(F–H)** 1 image or 1 msec. **(I–K)** 10 images or 10 msec. **(L–N)** 100 images or 100 msec. **(O–Q)** 1,000 images or 1 s. **(R–T)** 10,000 images or 10 s.

conductance (G_{LRS} and G_{HRS}) distributions. The DMM is suitable to study the device variability in terms of the G_{LRS} and G_{HRS} dispersion, as well as due to errors in the conductance programming, expressed as a variability of the DMM control parameter λ (see **Figure 10A**). To do so, we made a small amendment in the SPICE subcircuit representation of the DMM (see **Table 1**), and defined the parameters λ, $I_{\min}$ and $I_{\max}$ in terms of the *gauss*() SPICE function. For instance, for λ, we define H as *param* H =

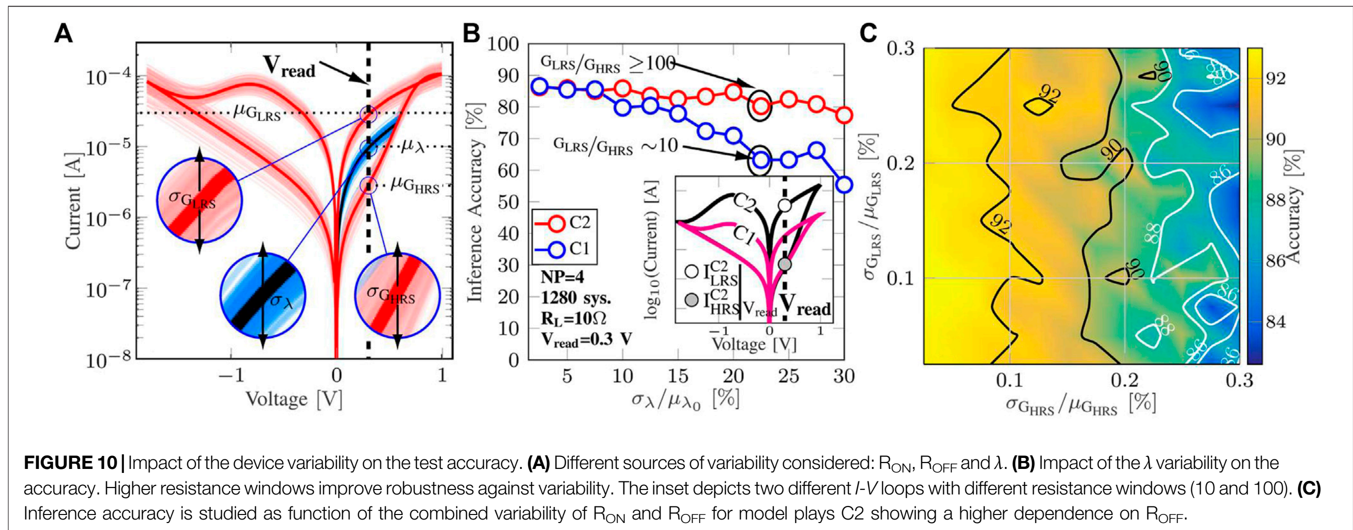

FIGURE 10 | Impact of the device variability on the test accuracy. **(A)** Different sources of variability considered: R_{ON}, R_{OFF} and λ. **(B)** Impact of the λ variability on the accuracy. Higher resistance windows improve robustness against variability. The inset depicts two different *I-V* loops with different resistance windows (10 and 100). **(C)** Inference accuracy is studied as function of the combined variability of R_{ON} and R_{OFF} for model plays C2 showing a higher dependence on R_{OFF}.

gauss(H_0,H_var,3), being *H_0* and *H_var* two arguments passed to the subcircuit representing the nominal value of λ and its relative variability specified for 3 sigma, respectively. Similar variability is assumed for $I_{\min}$ and $I_{\max}$ parameters, which introduces dispersion in the G_{LRS} and G_{HRS} values. Because of the D2D variability among devices in the CPA, Monte Carlo (MC) simulations were performed, assuming different levels of variability for all devices. In each MC run, the characteristic (stored memory state for the resistive window) of each device was individually altered following a Gaussian distribution around the nominal value with increasing dispersion. Subsequently, the CPA was used to classify the images from the MNIST dataset. 10 MC runs were considered for each value of σ/μ.

In **Figure 10B**, the influence of the control parameter λ variability ($\sigma_\lambda/\mu_\lambda$ ranging from 0 to 30%) over the inference accuracy is studied for two different G_{LRS}/G_{HRS} ratios (10 and 100) achieved by two model plays (namely C1 and C2, shown in the inset of **Figure 10B**) and considering $R_L = 10\ \Omega$. The image size remains 8×8 px. and 4 partitions were used (that is, 4 CPAs of size 16×10 for each polarity of synaptic weights). For simplicity, no variability in the major *I-V* loop is considered ($\sigma_{\mathrm{GHRS}} = \sigma_{\mathrm{GLRS}} = 0$). Two trends are clearly observed. On one hand, the model play having a G_{LRS}/G_{HRS} ratio equal to or greater than 100 (C2) exhibits a very reduced sensitivity to λ variations (accuracy loss is below 5% for variabilities up to 30%). On the other hand, there is a sustained accuracy reduction for model play C1 over the same range of $\sigma_\lambda/\mu_\lambda$. Given that the case of model play C2 is more robust to λ variations it was then selected to be thoroughly studied by considering the joint variability of G_{LRS} and G_{HRS} ($\sigma_{\mathrm{GLRS}}/\mu_{\mathrm{GLRS}}$ and $\sigma_{\mathrm{GHRS}}/\mu_{\mathrm{GHRS}}$, respectively). As the variability is normally higher in HRS than in LRS [38], they were swept independently, resulting in the accuracy map illustrated in **Figure 10C**. The trend observed in **Figure 10B** is repeated in **Figure 10C**. Interestingly, $\sigma_{\mathrm{GHRS}}/\mu_{\mathrm{GHRS}}$ has a higher impact on the inference accuracy, likely due to a higher number of memdiodes mapped close to the G_{HRS} value.

CONCLUSION

In this work, we have demonstrated the viability of the Dynamic Memdiode compact Model (DMM) for realistic SPICE simulations of large RRAM-based Cross-Point Arrays (CPA) intended for neuromorphic applications. A single layer perceptron (SLP) and the MNIST database of greyscale, handwritten digits were considered as case study. Although a simplistic approach when compared with more sophisticated multi-layer Artificial Neural Networks (ANN), the SLP allows studying and clarifying the ANN limitations during both the writing and inference phases caused by parasitic effects and non-idealities occurring in the synaptic layers implemented with CPAs.

The DMM model considers a Transport Equation (TE) based on an extension of the QPC model. The extension consists in replacing the current amplitude factor for an ideal monomode ballistic conductor with a more general amplitude factor that overcomes the physical limitation of a single tunnelling barrier. The capability of the DMM's Memory Equation (ME) of generating the hysteretic memory map by itself without *ad-hoc* definitions, allows the device to be set to a given conductance level by applying pulses of a constant amplitude, as experimentally observed for some materials. This property was exploited to implement a Closed-Loop-Tunning (CLT) approach for the programming of each RRAM devices in the simulated CPAs. Thereby, in this paper not only the classification accuracy was investigated but also the time required to program each electronic synapsis of the CPA-based SLP. The study also involved computing the error committed during such process.

It was found that the normalization method (NM) significantly modifies these metrics, and that a NM targeted to exploit the entire dynamic range of the RRAM devices will likely produce higher write times and errors. Similarly, the line resistance reduces the voltage effectively delivered to the RRAM cells causing an increase of the write time and eventually causing the write pulses not to meet the switching conditions. Such conditions change with the frequency of the writing pulses, adding further complexity both to the write and inference phases. For the write phase, the increase of the SET transition voltage for higher frequencies implies that the write

voltage should be changed accordingly with the frequency to avoid incurring into excessively long programming times or intolerable writing errors that at the end affect the classification performance.

APPENDIX

The SPICE script for the DMM used for all simulations shown in this paper is reported in **Table 1**. A(x) and Rss(x) stands for the α and R memdiode parameters, which are a function of the memory state. The memory state λ is represented by V(H), with H0 being the initial state. Parameters modelling the HRS-LRS transition are T0s, T0r, V0s, and V0r for τ_{0S}, τ_{0R}, V_{0S}, and V_{0R}, respectively. A Voltage-controlled current source is used to implement **Eq. 1** (GD and resistor RS), while **Eq. 3** is modeled through voltage controlled resistors (RH and RD) and a voltage source and capacitor of fixed value (EV and CH). The antiparallel diodes are modelled by the controlled current source GD in the script. beta defines whether the conduction is symmetric with positive and negative applied voltages (beta = 0.5) or not (beta≠0.5). The model is written in terms of the HSPICE syntax.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

FA developed the framework for the CPA-based ANN construction, training and simulation, as well as designed and performed the simulations. EM and JS developed the DMM model. All authors discussed the results. FA and EM wrote the main manuscript. FA and SP prepared the figures. All authors reviewed and accepted the manuscript. FP, JS and EM secured the funding and provided the necessary resources.

FUNDING

This work was supported in part by the Argentine Ministerio de Ciencia, Tecnología e Innovación (MINCyT) under Contract PICTE 2018-0192, Contract PICT 2016/0579, and Contract PME 2015-0196; in part by the CONICET under Project PIP-11220130100077CO; and in part by the UTN-FRBA under Project PID-UTN EIUTIBA4395TC3, Project PID-UTN CCUTIBA4764TC, Project PID-UTN MATUNBA4936, Project PID-UTN CCUTNBA5182, and Project PID-UTN CCUTNBA6615. The work of JS and EM was supported by the TEC2017-84321-C4-4-R project (Spanish Ministerio de Ciencia e Innovación). This work is supported by the EMPIR 20FUN06 MEMQuD project with funds from the EMPIR program co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation program. S.P. is currently also with the Physical Sciences and Engineering Division of the King Abdullah University of Science and Technology.

REFERENCES

1. Lecun Y, Bengio Y, and Hinton G. Deep Learning. *Nature* (2015) 521:436–44. doi:10.1038/nature14539
2. International Technology Roadmap for Semiconductors (ITRS). *Edition 2.0*. Available at: https://www.semiconductors.org/wp-content/uploads/2018/06/0_2015-ITRS-2.0-Executive-Report-1.pdf (Accessed June 1, 2021) (2015).
3. Freitas RF, and Wilcke WW. Storage-class Memory: The Next Storage System Technology. *IBM J Res Dev* (2008) 52:439–47. doi:10.1147/rd.524.0439
4. Yu S, Gao B, Fang Z, Yu H, Kang J, and Wong H-SP. A Low Energy Oxide-Based Electronic Synaptic Device for Neuromorphic Visual Systems with Tolerance to Device Variation. *Adv Mater* (2013) 25:1774–9. doi:10.1002/adma.201203680
5. Hu M, Li H, Chen Y, Wu Q, Rose GS, and Linderman RW. Memristor Crossbar-Based Neuromorphic Computing System: A Case Study. *IEEE Trans Neural Networks Learn Syst* (2014) 25:1864–78. doi:10.1109/TNNLS.2013.2296777
6. Upadhyay NK, Joshi S, and Yang JJ. Synaptic Electronics and Neuromorphic Computing. *Sci China Inf Sci* (2016) 59:061404. doi:10.1007/s11432-016-5565-1
7. Wang Y, Tang T, Xia L, Li B, Gu P, Li H, et al. Energy Efficient RRAM Spiking Neural Network for Real Time Classification. In: Proceedings of the ACM Great Lakes Symposium on VLSI, GLSVLSI. New York, New York, USA: Association for Computing Machinery (2015). p. 189–94. doi:10.1145/2742060.2743756
8. Sasago Y, Kinoshita M, Morikawa T, Kurotsuchi K, Hanzawa S, Mine T, et al. Cross-point Phase Change Memory with 4F2 Cell Size Driven by Low-Contact-Resistivity Poly-Si Diode. In: Digest of Technical Papers - Symposium on VLSI Technology, June 15-17 2009 Kyoto, Japan: IEEE (2009). p. 24–5.
9. Park S, Kim H, Choo M, Noh J, Sheri A, Jung S, et al. RRAM-based Synapse for Neuromorphic System with Pattern Recognition Function. In: Technical Digest - International Electron Devices Meeting. New Jersey: IEDM (2012). doi:10.1109/IEDM.2012.6479016
10. Ham SJ, Mo HS, and Min KS. Low-Power VDD/3 Write Scheme with Inversion Coding Circuit for Complementary Memristor Array. *IEEE Trans Nanotechnol* (2013) 12:851–7. doi:10.1109/TNANO.2013.2274529
11. Truong S, Ham S-J, and Min K-S. Neuromorphic Crossbar Circuit with Nanoscale Filamentary-Switching Binary Memristors for Speech Recognition. *Nanoscale Res Lett* (2014) 9:629. doi:10.1186/1556-276X-9-629
12. Truong SN, and Min KS. New Memristor-Based Crossbar Array Architecture with 50-% Area Reduction and 48-% Power Saving for Matrix-Vector Multiplication of Analog Neuromorphic Computing. *J Semicond Technol Sci* (2014) 14:356–63. doi:10.5573/JSTS.2014.14.3.356
13. Truong SN, Shin SH, Byeon SD, Song JS, and Min KS. New Twin Crossbar Architecture of Binary Memristors for Low-Power Image Recognition with Discrete Cosine Transform. *IEEE Trans Nanotechnol* (2015) 14:1104–11. doi:10.1109/TNANO.2015.2473666
14. Aguirre FL, Pazos SM, Palumbo F, Suñé J, and Miranda E. Application of the Quasi-Static Memdiode Model in Cross-Point Arrays for Large Dataset Pattern Recognition. *IEEE Access* (2020) 8:202174–93. doi:10.1109/ACCESS.2020.3035638
15. Aguirre FL, Gomez NM, Pazos SM, Palumbo F, Suñé J, and Miranda E. Minimization of the Line Resistance Impact on Memdiode-Based Simulations of Multilayer Perceptron Arrays Applied to Pattern Recognition. *J Low Power Electron Appl* (2021) 11:9. doi:10.3390/jlpea11010009
16. Strukov DB, Snider GS, Stewart DR, and Williams RS. The Missing Memristor Found. *Nature* (2008) 453:80–3. doi:10.1038/nature06932
17. Papandreou N, Pozidis H, Pantazi A, Sebastian A, Breitwisch M, Lam C, et al. Programming Algorithms for Multilevel Phase-Change Memory.

Proc - IEEE Int Symp Circuits Syst (2011) 329–32. doi:10.1109/ISCAS.2011.5937569
18. Yao P, Wu H, Gao B, Eryilmaz SB, Huang X, Zhang W, et al. Face Classification Using Electronic Synapses. *Nat Commun* (2017) 8:1–8. doi:10.1038/ncomms15199
19. Milo V, Zambelli C, Olivo P, Pérez E, Mahadevaiah KM, et al. Multilevel HfO2-Based RRAM Devices for Low-Power Neuromorphic Networks. *APL Mater* (2019) 7. doi:10.1063/1.5108650
20. Yu S, Chen PY, Cao Y, Xia L, Wang Y, and Wu H. Scaling-up Resistive Synaptic Arrays for Neuro-Inspired Architecture: Challenges and prospect. In: Technical Digest - International Electron Devices Meeting, IEDM. Piscataway, NJ: Institute of Electrical and Electronics Engineers Inc. (2015). p. 17.3.1–17.3.4. doi:10.1109/IEDM.2015.7409718
21. Gokmen T, and Vlasov Y. Acceleration of Deep Neural Network Training with Resistive Cross-point Devices: Design Considerations. *Front Neurosci* (2016) 10:19. doi:10.3389/fnins.2016.00333
22. Hu M, Strachan JP, Li Z, Grafals EM, Davila N, Graves C, et al. Dot-product Engine for Neuromorphic Computing. In: DAC '16: Proceedings of the 53rd Annual Design Automation Conference. New York, NY, USA: Association for Computing Machinery (2016). p. 1–6. doi:10.1145/2897937.2898010
23. Woo J, Moon K, Song J, Lee S, Kwak M, Park J, et al. Improved Synaptic Behavior under Identical Pulses Using AlOx/HfO2 Bilayer RRAM Array for Neuromorphic Systems. *IEEE Electron Device Lett* (2016) 37:994–7. doi:10.1109/LED.2016.2582859
24. Liu B, Li H, Chen Y, Li X, Huang T, Wu Q, et al. Reduction and IR-Drop Compensations Techniques for Reliable Neuromorphic Computing Systems. In: IEEE/ACM Int. Conf. Comput. Des. Dig. Tech. Pap. ICCAD 2015-Janua, 2-6 November 2014 San José, CA: IEEE (2015). p. 63–70. doi:10.1109/ICCAD.2014.7001330
25. Yakopcic C, Hasan R, Taha TM, McLean MR, and Palmer D. Efficacy of Memristive Crossbars for Neuromorphic Processors. *Proc Int Jt Conf Neural Networks* (2014) 15–20. doi:10.1109/IJCNN.2014.6889807
26. Panda D, Sahu PP, and Tseng TY. A Collective Study on Modeling and Simulation of Resistive Random Access Memory. *Nanoscale Res Lett* (2018) 13. doi:10.1186/s11671-017-2419-8
27. Prodromakis T, Peh BP, Papavassiliou C, and Toumazou C. A Versatile Memristor Model with Nonlinear Dopant Kinetics. *IEEE Trans Electron Devices* (2011) 58:3099–105. doi:10.1109/TED.2011.2158004
28. Merrikh Bayat F, Hoskins B, and Strukov DB. Phenomenological Modeling of Memristive Devices. *Appl Phys A Mater Sci Process* (2015) 118:779–86. doi:10.1007/s00339-015-8993-7
29. Yakopcic C, Taha TM, Subramanyam G, and Pino RE. Generalized Memristive Device SPICE Model and its Application in Circuit Design. *IEEE Trans Comput Des Integr Circuits Syst* (2013) 32:1201–14. doi:10.1109/TCAD.2013.2252057
30. Kvatinsky S, Friedman EG, Kolodny A, and Weiser UC. TEAM: Threshold Adaptive Memristor Model. *IEEE Trans Circuits Syst Regul Pap* (2013) 60: 211–21. doi:10.1109/TCSI.2012.2215714
31. Kvatinsky S, Ramadan M, Friedman EG, and Kolodny A. VTEAM: A General Model for Voltage-Controlled Memristors. *IEEE Trans Circuits Syst Express Briefs* (2015) 62:786–90. doi:10.1109/TCSII.2015.2433536
32. Eshraghian K, Kavehei O, Cho KR, Chappell JM, Iqbal A, Al-Sarawi SF, et al. Memristive Device Fundamentals and Modeling: Applications to Circuits and Systems Simulation. *Proc IEEE* (2012) 100:1991–2007. doi:10.1109/JPROC.2012.2188770
33. Biolek Z, Biolek D, Biolkova V, and Kolka Z. Reliable Modeling of Ideal Generic Memristors via State-Space Transformation. *Radioengineering* (2015) 24:393–407. doi:10.13164/re.2015.0393
34. Miranda E. Compact Model for the Major and Minor Hysteretic I-V Loops in Nonlinear Memristive Devices. *IEEE Trans Nanotechnol* (2015) 14:787–9. doi:10.1109/TNANO.2015.2455235
35. Patterson GA, Sune J, and Miranda E. Voltage-Driven Hysteresis Model for Resistive Switching: SPICE Modeling and Circuit Applications. *IEEE Trans Comput Des Integr Circuits Syst* (2017) 36:2044–51. doi:10.1109/TCAD.2017.2756561
36. Burr GW, Shelby RM, Sidler S, Di Nolfo C, Jang J, Boybat I, et al. Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165 000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element. *IEEE Trans Electron Devices* (2015) 62:3498–507. doi:10.1109/TED.2015.2439635
37. Li C, Belkin D, Li Y, Yan P, Hu M, Ge N, et al. Efficient and Self-Adaptive *In-Situ* Learning in Multilayer Memristor Neural Networks. *Nat Commun* (2018) 9:1–8. doi:10.1038/s41467-018-04484-2
38. Dong Z, Zhou Z, Li Z, Liu C, Huang P, Liu L, et al. Convolutional Neural Networks Based on RRAM Devices for Image Recognition and Online Learning Tasks. *IEEE Trans Electron Devices* (2019) 66:793–801. doi:10.1109/TED.2018.2882779
39. Querlioz D, Bichler O, Dollfus P, and Gamrat C. Immunity to Device Variations in a Spiking Neural Network with Memristive Nanodevices. *IEEE Trans Nanotechnol* (2013) 12:288–95. doi:10.1109/TNANO.2013.2250995
40. Miranda E, Walczyk C, Wenger C, and Schroeder T. Model for the Resistive Switching Effect in HfO2 MIM Structures Based on the Transmission Properties of Narrow Constrictions. *IEEE Electron Device Lett* (2010) 31: 609–11. doi:10.1109/LED.2010.2046310
41. Chen A. A Comprehensive Crossbar Array Model with Solutions for Line Resistance and Nonlinear Device Characteristics. *IEEE Trans Electron Devices* (2013) 60:1318–26. doi:10.1109/TED.2013.2246791
42. Lee AR, Bae YC, Im HS, and Hong JP. Complementary Resistive Switching Mechanism in Ti-Based Triple TiO X/TiN/TiO X and TiO X/TiO X N Y/TiO X Matrix. *Appl Surf Sci* (2013) 274:85–8. doi:10.1016/j.apsusc.2013.02.100
43. Duan WJ, Wang JB, Zhong XL, Song HJ, and Li B. Complementary Resistive Switching in Single sandwich Structure for Crossbar Memory Arrays. *J Appl Phys* (2016) 120:084502. doi:10.1063/1.4961222
44. Yang M, Wang H, Ma X, Gao H, and Hao Y. Voltage-amplitude-controlled Complementary and Self-Compliance Bipolar Resistive Switching of Slender Filaments in Pt/HfO 2/HfO X/Pt Memory Devices. *J Vac Sci Technol B, Nanotechnol Microelectron Mater Process Meas Phenom* (2017) 35:032203. doi:10.1116/1.4983193
45. Fröhlich K, Kundrata I, Blaho M, Precner M, Tapajna M, Klimo M, et al. Hafnium Oxide and Tantalum Oxide Based Resistive Switching Structures for Realization of Minimum and Maximum Functions. *J Appl Phys* (2018) 124: 152109. doi:10.1063/1.5025802
46. Aguirre FL, Rodriguez-Fernandez A, Pazos SM, Sune J, Miranda E, and Palumbo F. Study on the Connection between the Set Transient in RRAMs and the Progressive Breakdown of Thin Oxides. *IEEE Trans Electron Devices* (2019) 66:3349–55. doi:10.1109/ted.2019.2922555
47. Chua L. Resistance Switching Memories Are Memristors. *Appl Phys A* (2011) 102:765–83. doi:10.1007/s00339-011-6264-9
48. Adhikari SP, Sah MP, Kim H, and Chua LO. Three Fingerprints of Memristor. *IEEE Trans Circuits Syst* (2013) 60:3008–21. doi:10.1109/TCSI.2013.2256171
49. Ielmini D. Resistive Switching Memories Based on Metal Oxides: Mechanisms, Reliability and Scaling. *Semicond Sci Technol* (2016) 31:063002. doi:10.1088/0268-1242/31/6/063002
50. Waser R, Dittmann R, Staikov C, and Szot K. Redox-based Resistive Switching Memories Nanoionic Mechanisms, Prospects, and Challenges. *Adv Mater* (2009) 21:2632–63. doi:10.1002/adma.200900375
51. Larentis S, Nardi F, Balatti S, Ielmini D, and Gilmer DC. Bipolar-switching Model of RRAM by Field- and Temperature-Activated Ion Migration. In: 2012 4th IEEE International Memory Workshop, IMW 2012, 20-23 may 2012 Milan, Italy: IEEE (2012). doi:10.1109/IMW.2012.6213648
52. Padovani A, Larcher L, Pirrotta O, Vandelli L, and Bersuker G. Microscopic Modeling of HfOx RRAM Operations: From Forming to Switching. *IEEE Trans Electron Devices* (2015) 62:1998–2006. doi:10.1109/TED.2015.2418114
53. Miranda E, Member IS, Suñé J, and Fellow I. *Fundamentals and SPICE Implementation of the Dynamic Memdiode Model for Bipolar Resistive Switching Devices* (2020). doi:10.36227/techrxiv.12479426
54. Miranda E, and Sune J. Analytic Modeling of Leakage Current through Multiple Breakdown Paths in SiO/sub 2/Films. In: 2001 IEEE International Reliability Physics Symposium Proceedings. 39th Annual

(Cat. No.00CH37167). Piscataway, NJ: IEEE (2001). p. 367–79. doi:10.1109/RELPHY.2001.922929
55. Sune J, Miranda E, Nafria M, and Aymerich X. Point Contact Conduction at the Oxide Breakdown of MOS Devices. In: Technical Digest - International Electron Devices Meeting. Piscataway, NJ: IEEE (1998). p. 191–4. doi:10.1109/iedm.1998.746318
56. Miranda E, and Suñé J. Electron Transport through Broken Down Ultra-thin SiO2 Layers in MOS Devices. *Microelectron Reliab* (2004) 44:1–23. doi:10.1016/j.microrel.2003.08.005
57. Biolek D, Biolek Z, Biolkova V, and Kolka Z. Modeling of TiO2memristor: from Analytic to Numerical Analyses. *Semicond Sci Technol* (2014) 29: 125008–7. doi:10.1088/0268-1242/29/12/125008
58. Bocquet M, Deleruyelle D, Aziza H, Muller C, Portal J-M, Cabout T, et al. Robust Compact Model for Bipolar Oxide-Based Resistive Switching Memories. *IEEE Trans Electron Devices* (2014) 61:674–81. doi:10.1109/TED.2013.2296793
59. Blonkowski S, and Cabout T. Bipolar Resistive Switching from Liquid Helium to Room Temperature. *J Phys D: Appl Phys* (2015) 48:345101. doi:10.1088/0022-3727/48/34/345101
60. González-Cordero G, Roldan JB, Jiménez-Molinos F, Suñé J, Long S, and Liu M. A New Compact Model for Bipolar RRAMs Based on Truncated-Cone Conductive Filaments - A Verilog-A Approach. *Semicond Sci Technol* (2016) 31:115013. doi:10.1088/0268-1242/31/11/115013
61. Wang Z, Ambrogio S, Balatti S, Sills S, Calderoni A, Ramaswamy N, et al. Postcycling Degradation in Metal-Oxide Bipolar Resistive Switching Memory. *IEEE Trans Electron Devices* (2016) 63:4279–87. doi:10.1109/TED.2016.2604370
62. Miranda E, and Sune J. Memristive State Equation for Bipolar Resistive Switching Devices Based on a Dynamic Balance Model and its Equivalent Circuit Representation. *IEEE Trans Nanotechnol* (2020) 19:837–40. doi:10.1109/TNANO.2020.3039391
63. Maestro-Izquierdo M, Gonzalez MB, Campabadal F, Sune J, and Miranda E. A New Perspective towards the Understanding of the Frequency-dependent Behavior of Memristive Devices. *IEEE Electron Device Lett* (2021) 42:565–8. doi:10.1109/LED.2021.3063239
64. Sah MP, Kim H, and Chua LO. Brains Are Made of Memristors. *IEEE Circuits Syst Mag* (2014) 14:12–36. doi:10.1109/MCAS.2013.2296414
65. Campbell KA, Drake KT, and Barney Smith EH. Pulse Shape and Timing Dependence on the Spike-Timing Dependent Plasticity Response of Ion-Conducting Memristors as Synapses. *Front Bioeng Biotechnol* (2016) 4:97. doi:10.3389/fbioe.2016.00097
66. Dongale TD, Patil KP, Gaikwad PK, and Kamat RK. Investigating Conduction Mechanism and Frequency Dependency of Nanostructured Memristor Device. *Mater Sci Semicond Process* (2015) 38:228–33. doi:10.1016/j.mssp.2015.04.033
67. Dongale TD, Khot KV, Mohite SV, Desai ND, Shinde SS, Patil VL, et al. Effect of Write Voltage and Frequency on the Reliability Aspects of Memristor-Based RRAM. *Int Nano Lett* (2017) 7:209–16. doi:10.1007/s40089-017-0217-z
68. Eshraghian JK, Kang S-M, Baek S, Orchard G, Iu HH-C, and Lei W. Analog Weights in ReRAM DNN Accelerators. *Proc 2019 IEEE Int Conf Artif Intell Circuits Syst AICAS* (2019) 2019:267–71. doi:10.1109/AICAS.2019.8771550
69. Liang J, Yeh S, Simon Wong S, and Philip Wong HS. Effect of Wordline/bitline Scaling on the Performance, Energy Consumption, and Reliability of Cross-point Memory Array. *ACM J Emerg Technol Comput Syst* (2013) 9:1–14. doi:10.1145/2422094.2422103
70. Choi W, Moon K, Kwak M, Sung C, Lee J, Song J, et al. Hardware Implementation of Neural Network Using Pre-programmed Resistive Device for Pattern Recognition. *Solid State Electron* (2019) 153:79–83. doi:10.1016/j.sse.2018.12.018
71. Han R, Huang P, Zhao Y, Cui X, Liu X, and Kang J. Efficient Evaluation Model Including Interconnect Resistance Effect for Large Scale RRAM Crossbar Array Matrix Computing. *Sci China Inf Sci* (2019) 62:1–11. doi:10.1007/s11432-018-9555-8
72. Miranda E, and Suñé J. Memristors for Neuromorphic Circuits and Artificial Intelligence Applications. *Materials (Basel)* (2020) 13:938. doi:10.3390/ma13040938
73. Jiang H, Han L, Lin P, Wang Z, Jang MH, Wu Q, et al. Sub-10 Nm Ta Channel Responsible for Superior Performance of a HfO2 Memristor. *Sci Rep* (2016) 6: 1–8. doi:10.1038/srep28525
74. Choi S, Sheridan P, and Lu WD. Data Clustering Using Memristor Networks. *Sci Rep* (2015) 5:1–10. doi:10.1038/srep10492
75. Wang Z, Yin M, Zhang T, Cai Y, Wang Y, Yang Y, et al. Engineering Incremental Resistive Switching in TaO: X Based Memristors for Brain-Inspired Computing. *Nanoscale* (2016) 8:14015–22. doi:10.1039/c6nr00476h
76. Matveyev Y, Kirtaev R, Fetisova A, Zakharchenko S, Negrov D, and Zenkevich A. Crossbar Nanoscale HfO2-Based Electronic Synapses. *Nanoscale Res Lett* (2016) 11:1–5. doi:10.1186/s11671-016-1360-6
77. Miranda E, Milano G, and Ricciardi C. Modeling of Short-Term Synaptic Plasticity Effects in ZnO Nanowire-Based Memristors Using a Potentiation-Depression Rate Balance Equation. *IEEE Trans Nanotechnol* (2020) 19:609–12. doi:10.1109/TNANO.2020.3009734
78. Møller MF. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning. *Neural Networks* (1993) 6:525–33. doi:10.1016/S0893-6080(05)80056-5
79. Prezioso M, Merrikh-Bayat F, Hoskins BD, Adam GC, Likharev KK, and Strukov DB. Training and Operation of an Integrated Neuromorphic Network Based on Metal-Oxide Memristors. *Nature* (2015) 521:61–4. doi:10.1038/nature14441
80. Fouda ME, Lee S, Lee J, Eltawil A, and Kurdahi F. Mask Technique for Fast and Efficient Training of Binary Resistive Crossbar Arrays. *IEEE Trans Nanotechnol* (2019) 18:704–16. doi:10.1109/tnano.2019.2927493
81. Liu C, Hu M, Strachan JP, and Li HH. Rescuing Memristor-Based Neuromorphic Design with High Defects. In: Proceedings - Design Automation Conference. Piscataway, NJ: Institute of Electrical and Electronics Engineers Inc. (2017). doi:10.1145/3061639.3062310
82. LeCun Y, Cortes C, and Burges CJC. MNIST Handwritten Digit Database, Yann LeCun, Corinna Cortes and Chris Burges (1998). Available at: http://yann.lecun.com/exdb/mnist/(Accessed June 1, 2019).
83. Liu B, Li H, Chen Y, Li X, Wu Q, and Huang T (2015). Vortex: Variation-Aware Training for Memristor X-Bar. Proc. - Des. Autom. Conf. 2015-July, 8-12 June San Francisco, USA: IEEE, 1–6. doi:10.1145/2744769.2744930
84. Montgomery DC, and Runger GC. *Applied Statistics and Probability for Engineers*. Hoboken, NJ: John Wiley & Sons (2010).
85. Aguirre FL, Pazos SM, Palumbo F, Suñé J, and Miranda EA. Assessment and Improvement of the Pattern Recognition Performance of Memdiode-Based Cross-Point Arrays with Randomly Distributed Stuck-At-Faults. *MDPI Electronics (Submitted Rev)* (2021).
86. Lee YK, Jeon JW, Park E-S, Yoo C, Kim W, Ha M, et al. Matrix Mapping on Crossbar Memory Arrays with Resistive Interconnects and its Use in In-Memory Compression of Biosignals. *Micromachines* (2019) 10:306. doi:10.3390/mi10050306
87. Rossnagel SM, and Kuan TS. Alteration of Cu Conductivity in the Size Effect Regime. In: Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures. American Vacuum SocietyAVS (2004). p. 240–7. doi:10.1116/1.1642639
88. Steinhögl W, Schindler G, Steinlesberger G, Traving M, and Engelhardt M. Comprehensive Study of the Resistivity of Copper Wires with Lateral Dimensions of 100 Nm and Smaller. *J Appl Phys* (2005) 97:023706. doi:10.1063/1.1834982
89. Josell D, Brongersma SH, and Tőkei Z. Size-Dependent Resistivity in Nanoscale Interconnects. *Annu Rev Mater Res* (2009) 39:231–54. doi:10.1146/annurev-matsci-082908-145415
90. Fuchs K. The Conductivity of Thin Metallic Films According to the Electron Theory of Metals. *Math Proc Cambridge Philos Soc* (1938) 34:100–8. doi:10.1017/S0305004100019952
91. Mayadas AF, and Shatzkes M. Electrical-resistivity Model for Polycrystalline Films: The Case of Arbitrary Reflection at External Surfaces. *Phys Rev B* (1970) 1:1382–9. doi:10.1103/PhysRevB.1.1382
92. Adam GC, Khiat A, and Prodromakis T. Challenges Hindering Memristive Neuromorphic Hardware from Going Mainstream. *Nat Commun* (2018) 9: 1–4. doi:10.1038/s41467-018-07565-4

93. Yi W, Kim Y, and Kim JJ. Effect of Device Variation on Mapping Binary Neural Network to Memristor Crossbar Array. *Proc 2019 Des Autom Test Eur Conf Exhib DATE* (2019) 2019:320–3. doi:10.23919/DATE.2019.8714817
94. Chen A, and Lin MR. Variability of Resistive Switching Memories and its Impact on Crossbar Array Performance. *IEEE Int Reliability Phys Symp Proc* (2011). doi:10.1109/IRPS.2011.5784590
95. Luo Q, Xu X, Gong T, Lv H, Dong D, Ma H, et al. 8-Layers 3D Vertical RRAM with Excellent Scalability towards Storage Class Memory Applications. In: Technical Digest - International Electron Devices Meeting, IEDM. Piscataway, NJ: Institute of Electrical and Electronics Engineers Inc. (2018). p. 2.7.1–2.7.4. doi:10.1109/IEDM.2017.8268315
96. Pi S, Lin P, and Xia Q. Cross point Arrays of 8 Nm × 8 Nm Memristive Devices Fabricated with Nanoimprint Lithography. *J Vac Sci Technol B Microelectron Nanom Struct* (2013) 31:06FA02. doi:10.1116/1.4827021

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Aguirre, Pazos, Palumbo, Suñé and Miranda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.